Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Understanding and Designing for Control in Camera Operation

vorgelegt von

Axel Hösl

München, den 13.6.2019

Erstgutachter: Prof. Dr. Andreas Butz
Zweitgutachter: Prof. Dr. Marc Christie

Mündliche Prüfung: 19.7.2019

Axel Hösl
Understanding and Designing for Control in Camera Operation
Dissertation, Ludwig-Maximilians-Universität München, 2019

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Understanding and Designing for Control in Camera Operation

vorgelegt von
Axel Hösl

München, den 13.6.2019

Acknowledgements

A long journey now comes to an end and it is hard to put into words how deeply thankful I feel and how much I want to say ‘thank you, you are the best!’ to so many people - to those who encouraged my first steps on this path, to those who joined in during its various stages and to those who even after this comes to an end are here to stay. I have no doubt in my mind that this formative experience of mine was only possible because of the great and enabling atmosphere that I encountered at the LMU Mediainformatics group. Therefore, let me first thank my advisor **Andreas Butz** and **Heinrich Hußmann** who I see as the main architects building up this amazing environment. For sure (and for the better), I am no longer the same person who I was when entering the office of **Andreas Butz** for the first time in 2013 and leaving it for the last time as a part of the research staff in 2018. His continuous guidance and mentorship allowed me to grow with the challenges laying ahead of me, even when things did not turn out as intended. Despite busy times and complicated situations, he was always there to help - thank you so much for it! **Heinrich Hußmann** was not supervising my research directly but this was no reason for him to not have an open ear, supportive advice or the right questions at hand. Further contributing to the spirit of the lab, it is fair to say that **Florian Alt** and **Alexander Wiethoff** led outstandingly by example in showing us how it’s done.

Also outside of the LMU, I was lucky to meet only the kindest people. A special thanks is reserved for **Marc Christie** who not only serves as an external examiner for this thesis but who also hosted me for a week in his lab, eventually leading to a close encounter with the Bogdanoff brothers. Thank you for letting me see our shared research interests from your perspective and for connecting me with other interested researchers. Additionally, I want to thank **Andreas Fitz** for his collaboration in our shared projects and for showing me the state-of-the art from a manufacturer’s point of view. Also, **Stephan Vorbrugg** is nothing short of deserving a very big ‘thank you’! You introduced me thoroughly to the world of cinematography through the eyes of a camera operator and your input fundamentally shaped how I appreciate not only the practice but also the art.

Even a prime environment as found in this group is nothing without great people contributing to the realisation of its potential. Foremost, I want to thank the good souls of the lab. Thank you, **Franziska Schwamb**, for your kindness and your support. Further, thank you, **Rainer Fink**, for keeping the group operational and your dry humour which is second to none. Next, I want to thank all colleagues of mine who also work(ed) in the group. Thank you, for your positive attitudes, your support, companionship and even friendship as I am proud to say. **Fabian Hennecke** supervised my Diploma thesis connecting me to applied research for the first time. Further, **Doris Hausen** also provided me with practical research insights on research as well as photography from the start. **Julie Wagner** guided my early research eventually leading to fundamental insights and **Bernhard Slawik** instilled some of his never ending knowledge on tinkering and 3D printing in me. You all impacted me initially at making my very first steps into the world of research, thank you for that! Also, thank you, **Alina Hang** for your timeless aphorisms, **Sarah Tausch** for all the hilariously derailed conversations and **Simon Stusak** for never being short of a joke about it. Sharing a coffee with you made any day in the lab brighter. Also, thank you, **Mariam Hassib** for your continuously encouraging and uplifting comments. Given his humorous personal takes on HCI and beyond, also **Henri Palleis** made all of this so much more enjoyable. Similarly, in sharing his unequivocal points of view, **Christian Mai** made my day on a regular basis. In case you want to see a great example of someone conducting research with ease and competence simply go to **Daniel Buschek**, but be aware, he might spark a discussion too interesting to leave. Further, it might be quickly joined by **Hanna Schneider** who is always up for extracting deeper insights of any research topic and **Renate Häuslschmid** who easily lets joking remarks on her martial arts background pass as she knows what she is capable of doing to you. Thanks to, **Mohamed Khamis** for putting in so much great energy into our collaborations, to **Bastian Pfleging** for sharing all his insights from finding and talking to the right people and to **Sarah Aragon Bartsch** who flawlessly made the transition from supervised student to colleague. Similarly, I am glad that I got to know **Malin Eiband** first as a student and then as a colleague. Further, thank you, **Maria Fysaraki** for introducing the lab folks to all the Greek specialities you brought. **Ceenu George** - please keep your down-to-earth attitude! Also, thank you, **Sylvia Rothe** for sharing your insights as a documentary filmmaker and your research perspective.

A very special 'thank you' is reserved for my long time office mates **Emanuel von Zezschwitz** and **Tobias Seitz** who made A206 and 437 some much more than just room numbers. It was daily fun guaranteed, vitality projected onto research, intellectual sparring and always new. **Emanuel von Zezschwitz** is by far one of the most persevering people that I know and the worst thing about him is that he handles all the workload that he shoulders with a smile and positive attitude deserving of only the best. **Tobias Seitz** not only shares the same passion for the peculiarities of daily life, culture and the net, he is further also the best person to be around as he is fluent in immediately understanding people and positively impacting anyone near him. Thank you both for being the best!

The greatest thanks of all however is reserved for my family. I deeply thank all of you for your understanding, support and care. Especially, the support of my parents **Ruth Hösl** and **Michael Hösl** and my sister **Francis Hösl** knows no boundaries. Words can hardly express how grateful and lucky I am. You supported me from day one on and never stopped since, despite me being not always the easiest person to deal with especially in the later stages of writing this document. While conducting the work leading up to this thesis and writing it up was a great part of my life, it does by no means compare to when **Nadine Dengler** and **Greta Dengler** came into it, for you are so much more than part of it - you are my life. Thank you for all the love and all the joy in sharing all the great moments we were able to experience so far. Most certainly, there are way more to come and to support **Greta** in her first as well as both of you in your further steps is one of the greatest pleasures of my life!

Opening Remarks

This thesis mostly concentrates on camera motion, its dynamic on-screen effects and the tools that are used for its creation behind the scenes. However, the static nature of text often makes it hard to provide an accurate representation. Relevant yet subtle aspects can get lost. On the one hand, this document is intended to be self-containing, on the other first-hand experience of these on-screen effects is crucial. In weighing up, how we could provide assisting material regarding the visual peculiarities of camera motion, we decided to offer supporting video with this document in the form of external links. The individual links are provided as references throughout the text and compiled as a complete list in Chapter IV.

To further maintain an easy reading experience, we additionally provide these links on the page margins co-located with their appearance in the text. The links are represented by a ▶ icon that is clickable in the electronic version of this thesis.

Additionally, we introduced further icons on the margins next to their appearance in the text of this document. These represent appearances of what we believe to be important aspects for the following text. For example, the ? icon, represents a primary research question. All list of all icons and their meaning is attached below:

- ? A primary research question, mostly on a higher level of abstraction
- ! A research idea, also mostly referring to a more general or fundamental idea
- ▶ An external link a video referenced in the text

Further, we added a page with summarising insights at the end of a chapter when appropriate. These pages always appear on a right-hand side (or on an odd page). Additionally, they are marked on the top with a **Insights** symbol (as on the right-hand side of this double page). Therefore, when looking for these insights, the reader can flip through the pages of this document, look for the header on the top right and hence hopefully find them more easily.

Insights

- Videos in the form of external links are provided with this document and listed in the Appendix (Chapter IV).
- They are marked with a (clickable) ▶ icon on the page margin co-located with a reference in the text
- Further icons are used on page margins. A ? indicates a primary (research) question and ! a primary (research) idea.

Statement of Collaboration

This thesis reports on research that I conducted between October 2013 and September 2018. The work would have been impossible without the collaboration of others. As I want to acknowledge the efforts of the collaborators, the scientific plural will be applied throughout this thesis. Chapters 1, 2, 4, 5, 9, and 10 are written exclusively for publication in this thesis. In contrast, Chapters **3**, **6**, **7** and **8** reference research that was previously published in co-authored papers presented at international peer-reviewed conferences. Furthermore, some of the presented work was compiled based on theses and practical projects conducted by supervised students which were carried out under my continuous guidance. The publications are attributed to at the beginning of each chapter. Details on how they differ from this thesis are found below along with attributions to student work.

3 - Users: Insights from Cinematographers

Section 3.2 is in parts based on a paper that was published at INTERACT '17 [122] with me being the leading author of the co-authored paper. While the research idea originated from me, the sampling of the participants and conducting the online survey was part of the Master's thesis of Thomas Burghart [288]. Regarding the data collected in the survey, I was responsible for the analysis ahead of the paper publication.

Changes: *While the publication only describes derived results (distinction of user roles), this thesis further describes the procedure and provides more detailed summaries of the results in written (Section 3.2.1), visual (Figure 3.1 and Figure 3.2) and tabulised form (Table 3.2.1).*

Sections 3.3 and 3.5 contain research results which were published at CHI '15 [124] with me being the leading author of the co-authored paper. While the research idea was developed together with Julie Wagner, the sampling of the participants and conducting the interviews was part of the Bachelor's thesis of Theresa Mücke [299]. I was responsible for the analysis of the seminars as represented in the paper and this thesis. To analyse the interviews, we used a Grounded Theory informed (open) coding process incorporating Theresa Mücke, Julie Wagner and me as coders making us all responsible for the process and the results.

Changes: *While the paper and this thesis detail our process and its results similarly, this thesis extends our conclusions by presenting a set of proto-user journeys and proto-personas in Section 3.6 that are not found in the paper.*

Material presented in Section 3.4 also stems from the Bachelor's thesis of Theresa Mücke [299] with me conducting the analysis of the contextual inquiries as presented in this thesis.

6 - Prototyping Toolkit and Evaluation Framework

Section 6.1 refers to research that was previously published in multiple papers. Those were presented at INTERACT ’17 [122, 550], UIST ’17 [142, 474] and the DroNet ’18 workshop (co-located with MobiSys ’18) [79]. For the co-authored papers presented at INTERACT ’17 and the DroNet ’18 workshop, I was the leading author. For the UIST ’17 paper, Mohamed Khamis was the first author and likewise, for the DroNet ’18 paper, it was Andreas Ellwanger. While the research idea for the INTERACT ’17 paper originated from me, the practical work presented was part of the Bachelor’s thesis of Patrick Mörwald [298]. Similarly, for the DroNet ’18 paper, the primary research idea was derived in collaboration with Andreas Ellwanger with practical research being part of his Bachelor’s thesis [296]. In contrast, for the UIST ’17 paper, Mohamed Khamis was the main ideator with practical aspects being part of the Master’s theses of Alexander Klimczak [290] and Martin Reiss [291]. In Section 6.1 material from further collaborations can be found as of the Master’s thesis of Thomas Burghart [288], John-Louis Gao [289] and a practical course conducted by Ludwig Trotter, Andrea Attwenger and Maximilian Körner [302].

Changes: *While the INTERACT ’17 paper generally reports on the development of our physical prototyping platform, this thesis provides the necessary technical details required for its reproduction. Further, in discussing our prototyping toolkit from a more general perspective in Section 6.1, we refer to the UIST ’17 and DroNet ’18 publications as examples.*

Section 6.2 consists partly of ideas that were also echoed in publications presented at CHI ’17 [123] and the co-located MICI ’17 workshop [121] with me being the leading author of the co-authored papers. While the research ideas of these publications originated from me, the presented user study of the paper presented at CHI ’17 was conducted in a supervised practical course taken by Florian Lehmann, Christina Rosenmöller and Phuong Anh Vu [301]. Further, material presented in Section 6.2.2 was in part derived from supervised seminar reports compiled by Daniel Seliger [304] and Katharina Sachmann [303]. Additionally, the user studies presented in Sections 6.2.4 and 6.2.5 were part of the Bachelor’s thesis of Korbinian Blanz [293].

7 - Basic Explorations

Sections 7.1, 7.2 and 7.3 entail material which was also part of a paper which was presented at INTERACT ’17 [122] with me being the leading author of the co-authored paper. While the research ideas originated from me, the implementation and user studies were conducted as parts of the Bachelor’s theses of Patrick Mörwald [298], Philipp Burgdorf [294] and Elisabeth Dreßler [295].

Changes: *In contrast to the paper, this thesis locates the presented design within the radar-chart inspired visualisation introduced in Section 5.2. Additionally, the statistical analysis as presented in this thesis was done separately in R based on our R-templates as found in Appendix IV (in contrast to analysis via SPSS as in the papers).*

8 - Prototyping and Evaluating Design Alternatives

Section 8.1 is in parts based on papers which were presented at INTERACT ’17 [120] and CVMP ’16 [119] with me being the leading author of the co-authored papers. While the research idea originated from me, practical work was part of the Master’s thesis of Sarah Aragon Bartsch [287].

Section 8.2 entails material that was already presented in a paper at INTERACT ’17 [118] with me being the leading author of the co-authored paper. While the research idea originated from me, the practical work was part of the Bachelor’s thesis of Mujo Alic [292].

Section 8.3 reports on material that was created as part of the Bachelor’s thesis of Andreas Ellwanger [296] and the subsequent publication [79] which was presented at the DroNet ’18 workshop (co-located with MobiSys ’18).

Section 8.4 is in parts based on the work reported on in the Bachelor’s thesis of Filip Hristov [297]. Further, it references work which was presented at CHI ’17 [123] and the MICI ’17 workshop [121] (co-located with CHI ’17) with me being the leading author of the co-authored papers. While the research ideas of these publications were initiated by me, the presented user study of the CHI ’17 paper was conducted in a supervised practical course taken by Florian Lehmann, Christina Rosenmöller and Phuong Anh Vu [301].

Finally, Section 8.5 presents content derived from the Bachelor’s thesis of Michael Puriss [300].

For all of the above-mentioned theses in Chapter 8, the research idea generally originated from me, and I contributed to the design of the user studies while the supervised students were responsible for the sampling of the participants and conducting the studies.

Changes: *Further, in contrast to the published papers, this thesis locates the presented designs within our radar-chart inspired visualisation introduced in Section 5.2. Additionally, the statistical analyses as presented in this thesis was done separately in R based on our R-templates as found in Appendix IV (in contrast to a statistical analysis via SPSS as in the papers).*

Abstract

Cinematographers often use supportive tools to craft desired camera moves. Recent technological advances added new tools to the palette such as gimbals, drones or robots. The combination of motor-driven actuation, computer vision and machine learning in such systems also rendered new interaction techniques possible. In particular, a content-based interaction style was introduced in addition to the established axis-based style. On the one hand, content-based cocreation between humans and automated systems made it easier to reach high level goals. On the other hand however, the increased use of automation also introduced negative side effects. Creatives usually want to feel in control during executing the camera motion and in the end as the authors of the recorded shots. While automation can assist experts or enable novices, it unfortunately also takes away desired control from operators. Thus, if we want to support cinematographers with new tools and interaction techniques the following question arises: How should we design interfaces for camera motion control that, despite being increasingly automated, provide cinematographers with an experience of control?

Camera control has been studied for decades, especially in virtual environments. Applying content-based interaction to physical environments opens up new design opportunities but also faces, less researched, domain-specific challenges. To suit the needs of cinematographers, designs need to be crafted with care. In particular, they must adapt to constraints of recordings on location. This makes an interplay with established practices essential. Previous work has mainly focused on a technology-centered understanding of camera travel which consequently influenced the design of camera control systems. In contrast, this thesis, contributes to the understanding of the motives of cinematographers, how they operate on set and provides a user-centered foundation informing cinematography specific research and design.

The contribution of this thesis is threefold: First, we present ethnographic studies on expert users and their shooting practices on location. These studies highlight the challenges of introducing automation to a creative task (assistance vs feeling in control). Second, we report on a domain specific prototyping toolkit for in-situ deployment. The toolkit provides open source software for low cost replication enabling the exploration of design alternatives. To better inform design decisions, we further introduce an evaluation framework for estimating the resulting quality and sense of control. By extending established methodologies with a recent neuroscientific technique, it provides data on explicit as well as implicit levels and is designed to be applicable to other domains of HCI. Third, we present evaluations of designs based on our toolkit and framework. We explored a dynamic interplay of manual control with various degrees of automation. Further, we examined different content-based interaction styles. Here, occlusion due to graphical elements was found and addressed by exploring visual reduction strategies and mid-air gestures. Our studies demonstrate that high degrees of quality and sense of control are achievable with our tools that also support creativity and established practices.

Zusammenfassung

Kameraleute nutzen traditionell gezielt Hilfsmittel um kontrollierte Kamerabewegungen zu ermöglichen. Der technische Fortschritt hat hierbei unlängst zum Entstehen neuer Werkzeugen wie Gimbals, Drohnen oder Robotern beigetragen. Dabei wurden durch eine Kombination von Motorisierung, Computer-Vision und Machine-Learning auch neue Interaktionstechniken eingeführt. Neben dem etablierten achsenbasierten Stil wurde nun auch ein inhaltsbasierter Interaktionsstil ermöglicht. Einerseits vereinfachte dieser die Arbeit, andererseits aber folgten dieser (Teil-)Automatisierung auch unerwünschte Nebeneffekte. Grundsätzlich wollen sich Kameraleute während der Kamerabewegung kontinuierlich in Kontrolle und am Ende als Autoren der Aufnahmen fühlen. Während Automatisierung hierbei Experten unterstützen und Anfänger befähigen kann, führt sie unweigerlich auch zu einem gewissen Verlust an gewünschter Kontrolle. Wenn wir Kamerabewegung mit neuen Werkzeugen unterstützen wollen, stellt sich uns daher die Frage: Wie sollten wir diese Werkzeuge gestalten damit sie, trotz fortschreitender Automatisierung ein Gefühl von Kontrolle vermitteln?

In der Vergangenheit wurde Kamerakontrolle bereits eingehend erforscht, allerdings vermehrt im virtuellen Raum. Die Anwendung inhaltsbasierter Kontrolle im physikalischen Raum trifft jedoch auf weniger erforschte domänenspezifische Herausforderungen welche gleichzeitig auch neue Gestaltungsmöglichkeiten eröffnen. Um dabei auf Nutzerbedürfnisse einzugehen, müssen sich Schnittstellen zum Beispiel an diese Einschränkungen anpassen können und ein Zusammenspiel mit bestehenden Praktiken erlauben. Bisherige Forschung fokussierte sich oftmals auf ein technisches Verständnis von Kamerafahrten, was sich auch in der Schnittstellengestaltung niederschlug. Im Gegensatz dazu trägt diese Arbeit zu einem besseren Verständnis der Motive und Praktiken von Kameraleuten bei und bildet eine Grundlage zur Forschung und Gestaltung von Nutzerschnittstellen.

Diese Arbeit präsentiert dazu konkret drei Beiträge: Zuerst beschreiben wir ethnographische Studien über Experten und deren Praktiken. Sie zeigen vor allem die Herausforderungen von Automatisierung bei Kreativaufgaben auf (Assistenz vs. Kontrollgefühl). Zweitens, stellen wir ein Prototyping-Toolkit vor, dass für den Einsatz im Feld geeignet ist. Das Toolkit stellt Software für eine Replikation quelloffen bereit und erleichtert somit die Exploration von Designprototypen. Um Fragen zu deren Gestaltung besser beantworten zu können, stellen wir ebenfalls ein Evaluations-Framework vor, das vor allem Kontrollqualität und -gefühl bestimmt. Darin erweitern wir etablierte Ansätze um eine neurowissenschaftliche Methodik, um Daten explizit wie implizit erheben zu können. Drittens, präsentieren wir Designs und deren Evaluation aufbauend auf unserem Toolkit und Framework. Die Alternativen untersuchen Kontrolle bei verschiedenen Automatisierungsgraden und inhaltsbasierten Interaktionen. Auftretende Verdeckung durch graphische Elemente, wurde dabei durch visuelle Reduzierung und Mid-Air Gesten kompensiert. Unsere Studien implizieren hohe Grade an Kontrollqualität und -gefühl bei unseren Ansätzen, die zudem kreatives Arbeiten und bestehende Praktiken unterstützen.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two Dimensional
3D	Three Dimensional
AA	Adaptive Automation
Ah	Amper hour
ANOVA	Analysis of Variance
API	Application Programming Interface
AR	Augmented Reality
BR	Bayerischer Rundfunk (eng. Bavarian Broadcasting)
CSI	Creativity Support Index
CST	Creativity Support Tool
CV	Computer Vision
DIT	Digital Image Technician
DIY	Do It Yourself
DOF	Depth of Field
DOP	Director of Photography
DRT	Detection-Response Task
DSLR	Digital Single Lens Reflex
EEG	Electroencephalogram
EMG	Electromyography
fps	frames per second
GPS	Global Positioning System
HCI	Human-Computer Interaction
HD	High Definition
HDMI	High Definition Multimedia Interface
HFF	Hochschule für Fernsehen und Film München (eng. University of Television and Film Munich)
HMD	Head-Mounted Display
Hz	Hertz
IEE	Interval Estimation Errors
IR	Infra-Red
LED	Light Emitting Diode
LOA	Level of Automation
m	metre
Mdn	Median
MICI	Mixed-Initiative Creative Interfaces
mm	millimetre
Mn	Mean
MoCo	Motion Control
ms	milliseconds
Nm	Newton metre
PD	Progressive Disclosure
POV	Point of View
PR	Progressive Reduction
PTZ	Pan-Tilt-Zoom
px	pixel
RO	Research Objective
RQ	Research Question
RTSP	Real-Time Streaming Protocol
SA	Situational Awareness
SCS	Sense of Control Scale
SDLP	Standard Deviation of Lateral Position
SUS	System Usability Scale
TLX	Task Load Index
TV	Television

UART Universal Asynchronous Receiver Transmitter

UI User Interface

V Volt

VR Virtual Reality

Wh Watt hour

STEVE MCQUEEN (DIRECTOR)

The camera movement should be like a cat jumping onto a table – with just enough amount of effort and that's it. That's enough.

MIKE FIGGIS (FILMMAKER)

The function of camera movement is to assist the storytelling. That's all it is. It cannot be there just to demonstrate itself.

BRUCE DERN (ACTOR)

In Hitchcock's eyes the movement was dramatic, not the acting. When he wanted the audience to be moved, he moved the camera.

1

Introduction

What to expect?

- Introduction to historic and humanistic aspects of cinematographic production
- Closer look at recent developments and relevant aspects with a Human-Computer Interaction lens
- Motivating research questions along with further details on our research approaches

What to take away?

- A sense of how our research is motivated from a general human as well as a specialised HCI perspective along with further information on our research agenda and practice

1.1 Motivation

"I learned that the first technology appeared in the form of stone tools, 2.6 million years ago. First entertainment comes evidence from flutes that are 35,000 years old, and evidence for first design comes 75,000 years old - beads."

(Zeresenay Alemseged, Paleoanthropologist)

Human use of technology in general and dedicated tools in particular has a long tradition. Those foremost physical devices eventually evolved into today's highly specialised, technically advanced and sometimes also computer-supported tools. During this evolutionary process, tools were not only applied for purely utilitarian purposes, but also for crafting other tools that in turn might be used for socio-cultural and entertainment purposes as pointed out by Alemseged [419], for instance. Within this developmental trajectory spanning from ancient to modern day's utilities, entertainment devices were also subject to evolutionary processes that are rooted in their distant past and eventually shaped their contemporary functions and forms. This process likewise included enhancement of their capabilities due to the integration of computation. In the more recent past, this might also be especially the case for the upcoming and gradual refinement of a class of tools that allows for capturing visual images. That first enabled photography and later also videography along with the further addition of recording audio and even more recently also depth information.

The motives as to why human beings are inclined to make creative use of such tools that they craft, in general, are manifold. Among the many potential explanations, one particular which seems especially relevant for cinematography might be found in a general psychological function that it fulfils; more specifically the narrative properties that it entails. As pointed out by Iser [128], human beings seem to be subject to anthropological dispositions towards a need for fiction (German: Fiktionsbedürfniss). He, in particular, investigates how this need for fiction might be considered a constituting factor in the emergence of literature. His fundamental idea was also expanded upon and transferred to other domains of creative expression as by Ventarola [265], for instance. While she makes the case for a reiterated theory of fiction explicitly based on examples of literature, paintings and music, to us, it seems reasonable to expand its explanatory scope also to cinematography.

In detail, Ventarola understands the human need for fiction as a motivating factor. In the case of cinematography, the process of translating this motivation into expressed behaviour is bound by a set of domain-specific limitations. For instance, using imaging tools to craft creative material is characterised by technical constraints and requirements that need to be met to ensure a 'successful' production (in terms of purely technical capturing). Additionally, on a different layer, how these aspects are met further affects other properties such as, among others, the semantics implicitly. Therefore, in creating the content, it is not only vital to ensure that a recording is crafted with care technically, but also stylistically. In choosing between several options that generally enable a sound technical recording, it is important to understand how those options simultaneously affect other characteristics and how these effects might be experienced and interpreted by the audience. Some of these characteristics of the cinematic repertoire might be subsumed as the visual aspects. The capabilities required to infer or read those only from the looking at the results of the production process might further be referred to as a form of visual literacy. In cinematography, visual literacy is especially bound to understanding the peculiarities that stem from opting for certain types of cameras, lenses, and tools for setting the camera in motion during a recording as well as the styles of their operation. However, to be knowledgable about those and to be able to infer those from a motion picture is not necessarily commonplace as the following quote by Scorsese [497] illustrates:

> *"What it [watching movies] made me realize was that there was an intelligence, another kind of intelligence, that was trying to tell a story through where the director, the writer, the cinematographer, where they were focusing your eyes, you know. Whether it was [that] the camera may be at an extremely low angle, looking up at you, the use of the lens, the size of the lens. I began to understand certain lenses did interpret the story differently. A longer lens crushed everything together and made it flat. A wider lens stretched everything and sometimes distorted it, especially if camera movement - I learned looking at particularly Welles' films and William Wyler, too. A wide-angle lens, but Wyler used his wide-angle lens in a very strong, steady image, but Welles used that wide-angle lens, 18 mm as it turns out, very often to move along the walls, to move along a hall - and you really felt - I felt - as if the camera was flying, as if the story was flying by, you know. I didn't know why until I kept seeing the films again and again, and as I began to know a little more about what filmmaking was like and what cameras did - and I still didn't know who made the pictures, you know - but I was beginning to understand that there are certain tools you use, and those tools become part of a vocabulary that's just as valid as that vocabulary that is used in literature and our language."*
>
> (Martin Scorsese, Director)

Besides the drive to craft and tell a story motivated by aiming at fulfilling a human need for fiction, also other motivating factors might be involved. As the realisation process is characterised by restrictions that are intrinsic to the nature of the (technical) recording procedures, the options that one could theoretically choose from are indeed limited. However, by the same token, these technical aspects and limitations open up a new space of opportunities in terms of how one adapts to these limiting challenges in general and how one uses them creatively as stylistic devices formed out of such a visual vocabulary as mentioned above. In exploring the potential of resulting solution-spaces and in manifesting a particular solution, it is necessary to also operate the associated tools skillfully. Being (or feeling as) the author of the results and mastering the technicalities in operating the devices might constitute further rewarding factors as summarised by Csikszentmihalyi and colleagues [62]:

"The idea that the ability to operate effectively in the environment fulfils a primary need is not new in psychology. In Germany, Groos (1901) and Bühler (1930) elaborated the concept of Funktionlust or 'activity pleasure' which Piager (1952) included in the earliest stages of sensorimotor development as the 'pleasure of being a cause' that drove infants to experiment. In more recent psychological thought, Hebb (1955) and Berlyne (1960) focused on the nervous system's need for optimal levels of stimulation to explain exploratory behavior and the seeking of novelty, while White (1959) and deCharms (1968) focused on people's need to feel in control, to be the causal agents of their actions. Later Deci and Ryan (Deci 1971; Deci and Ryan 1985) elaborated on this line of argument by suggesting that both competence and autonomy were innate psychological needs that must be satisfied for psychological growth and well-being."

(Mihály Csíkszentmihályi, Psychologist)

Aiming at fulfilling the need for fiction and expressing a particular (cinematic) form of activity pleasure might be fundamental in providing explanation models as to why people gravitate towards cinematic productions. Further, one might add that in the production process, (at least the potential for) a further motivating factor might be entailed with (the chance of) experiencing a flow state [62]:

"The basic conclusion was that, in all the various groups studied, the respondents reported a very similar subjective experience that they enjoyed so much that they were willing to go to great lengths to experience it again. This we eventually called the 'flow experience' (...). Flow is a subjective state that people report when they are completely involved in something to the point of forgetting time, fatigue, and everything else but the activity itself. It is what we feel when we read a well-crafted novel or play a good game of squash, or take part in a stimulating conversation. The defining feature of flow is intense experiential involvement in moment-to-moment activity. Attention is fully invested in the task at hand, and the person functions at his or her fullest capacity."

One peculiarity of this flow state is its relation to the human sense of control as laid out further [62]:

"During flow, we typically experience a sense of control—or, more precisely, a lack of anxiety about losing control that is typical of many situations in normal life. This sense of control is also reported in activities that involve serious risks, such as hang gliding, rock climbing, and race car driving—activities that to an outsider would seem to be much more potentially dangerous than the affairs of everyday life. Yet these activities are structured to provide the participant with the means to reduce the margin of error to as close to zero as possible. Rock climbers, for example, insist that their hair-raising exploits are safer than crossing a busy street in Chicago, because, on the rock face, they can foresee every eventuality, whereas when crossing the street, they are at the mercy of fate. The sense of control respondents describe thus reflects the possibility, rather

than the actuality, of control. Worrying about whether we can succeed at what we are doing—on the job, in relationships, even in crossing a busy street—is one of the major sources of psychic entropy in everyday life, and its reduction during flow is one of the reasons such an experience becomes enjoyable and thus rewarding."

In the sense of the above-mentioned quote, one might think of (the sense of) control rather as a relative and subjective phenomenon than an absolute. Exerting it might even be an enjoyable and rewarding experience. What any exertion of control requires is (at least) the necessary capabilities as well as the (proper) tools. How the tools themselves are shaped, however, might be another influencing factor affecting this subjective experience. Therefore, 'getting the right design' and 'getting the design right' [29] is of central interest in their development as pointed out by Norman [196].

"So for the moment, let me move from a discussion of theories of action and conceptual models and speak of the qualitative nature of human-computer interaction. The details of the interaction matter, ease of use matters, but I want more than correct details, more than a system that is easy to learn or to use: I want a system that is enjoyable to use. This is an important, dominating design philosophy, easier to say than to do. It implies developing systems that provide a strong sense of understanding and control. This means tools that reveal their underlying conceptual model and allow for interaction, tools that emphasize comfort, ease, and pleasure of use: for what Illich (1973) has called convivial fools. A major factor in this debate is the feeling of control that the user has over the operations that are being performed. A 'powerful', 'intelligent' system can lead to the well documented problems of 'overautomation', causing the user to be a passive observer of operations, no longer in control of either what operations take place, or of how they are done. On the other hand, systems that are not sufficiently powerful or intelligent can leave too large a gap in the mappings from intention to action execution and from system state to psychological interpretation. The result is that operation and interpretation are complex and difficult, and the user again feels out of control, distanced from the system."

(Don Norman, Cognitive Scientist)

In sum, one might conclude that to provide users with (cinematic) tools that as stated earlier allow for aiming at the fulfilment of a need for fiction, the expression of activity pleasure and potentially the experience of flow, designing with care is central. Based on the notions that Norman puts forth in the above-mentioned quote, one might further suggest that in designing (cinematic) User Interfaces (UI) it is specifically vital to provide users with a sense of control. How to implement it is not trivial as it requires to strike a balance as one can infer from his Daedalean advice: designers should support users to a low enough degree so that they do not suffer from the effects of over-automation, yet also to a large enough degree so that they do not feel distanced from the system or out of control simultaneously.

1.2 Research Objectives and Questions

As part of our work was conducted within two projects that were funded via the Central Innovation Programme for Small and Medium-sized Enterprises (KF2953003 and KF2953004) of the German Federal Ministry for Economic Affairs and Energy focusing on cinematic Motion Control (MoCo) tools and their UIs, we examined the domain more closely. The projects investigated the development of novel tools and UIs together with Panther as an industrial partner. Based on this experience, we concluded that besides the detailed aspects that we focused on initially, a further and deeper overall understanding was necessary and subsequently broadened the scope of our research.

With the ideas of the previous section in mind, we started examining the current technological developments. To provide the reader with a brief introduction to cinematic MoCo tools, we would like to offer some pointers ahead of presenting our preliminary results. For instance, the visual effects of novel tools such as drones can be seen in the music video *I won't let you down* (2014) of the Rock/Pop band OK Go [504]. Regarding the user and design aspects that we touched above, however, the making-off [503] to the video might be considered a more interesting resource as it allows for a closer look behind the scenes depicting the complexities of the production process. Yet, as it only showcases the use of drones, one may also find additional material that displays a greater variety of the cinematic tool palette helpful. A demonstration of a variety of tools that support varying degrees of freedom and shot types is found in a video that was recorded in parallel to *Hard Target* (1993) [488].

In our examination of the status-quo, we found that contemporary devices make increasingly use of computation. As a consequence, new types of tools emerged that in parallel became increasingly affordable and easier accessible. In contrast to the majority of traditional mechanic tools, current systems offer (semi-)autonomous operation, assistance functions or remote operation along with novel interaction styles, for instance. However, simultaneously, we found that the understanding of users and designs based on a thorough empirical investigation as represented in the literature is also still limited (to a degree). From these initial findings we derived two major Research Objectives (ROs) that we wanted to accomplish in our broadened research efforts:

RO1: Understanding (Users and Usage Context) :
To come to a better understanding of users and their working context. Especially of their usage patterns regarding established support tools to derive general design insights as well as opportunities for future developments

RO2: Designing for Control (in Camera Operation):
To explore novel designs and to examine them focussing regarding aspects of perceived control. As we also found that the evaluation practice in general as reported in the literature is quite heterogeneous, we also aimed for incorporating a more structured way in studying our designs

Research Questions

As we not only aim at fostering further understanding but also at contributing novel design ideas and user studies investigating their effects (based on the notions introduced above), one might put our overarching empirical research motives into the following high-level Research Question (RQ):

How should we design UIs for cinematic MoCo tools that, despite being increasingly automated, provide cinematographers with an experience of control and how can we estimate how well we did?

While this question might be mainly helpful in framing our research in a broader sense, it also might be considered as being too unspecific to be practically helpful. Therefore, we also derived several sub-questions that focus more on detailed aspects of the priorly stated research objectives and that are meant to guide our research efforts in the individual projects that are presented in the following:

RQ1 (RO1): What should researchers know about users and their practice to be able to empathise?

RQ2 (RO1): What are central user motives specific to cinematic camera operation?

RQ3 (RO1): Which principles can we derive from RQ2 that might be leveraged by Human-Computer Interaction (HCI) research and that are relevant to the design of systems and UIs?

RQ4 (RO2): How should researchers prototype and evaluate their ideas (based on RQ3)?

RQ5 (RO2): What can we learn from empirical studies examining instances of the design principles investigated in RQ3 based on prototypes conforming to the learned lessons in RQ4?

These RQs are not necessarily arbitrarily selected as they are generally inspired by conforming to the Design Thinking framework [24] and might be mapped onto it as follows:

RQ1 and RQ2 aim at providing answers that might enable researchers and designers to empathise with their users which is often considered an important first step in exploring new design opportunities. However, the answers to RQ2 should also help in further defining the problem(s) at hand. In our case, in particular, the answers to RQ2 helped identify a problem-space as well as hint at first ideas towards inferring a potential solution-space. Next, the results of investigating RQ3 should help to gather insights that might foster the ideation process more practically by building on top of the insights gathered by addressing RQ2. To promote an externalisation of those ideas, the insights derived from addressing RQ4 should be integrated into our research and design process. Further, as a thorough evaluation practice is also considered a key part in the Design Thinking framework, addressing RQ5 should not only help in inferring evidence-based learnings but in turn also in once more understanding the problems at hand a bit better as well as in once more informing the ideation process which might be supportive of coming up with new and potentially even better design ideas.

1.3 Approach

As we used a variety of perspectives and methods besides Design Thinking in our research, it is not necessarily trivial to find one single label to accurately describe it. Therefore, we would like to offer several perspectives on how one might conceptualise our efforts. The most generic way of framing our work might be by associating it with a set of already established conceptualisations and practices that we applied. Consequently, in an initial step, one could describe it as follows:

In a User-Centred Design process, we make use of Grounded Theory approaches that are embedded within a Design Thinking framework to inform designs which we subsequently evaluate empirically in user studies based on statistical analyses.

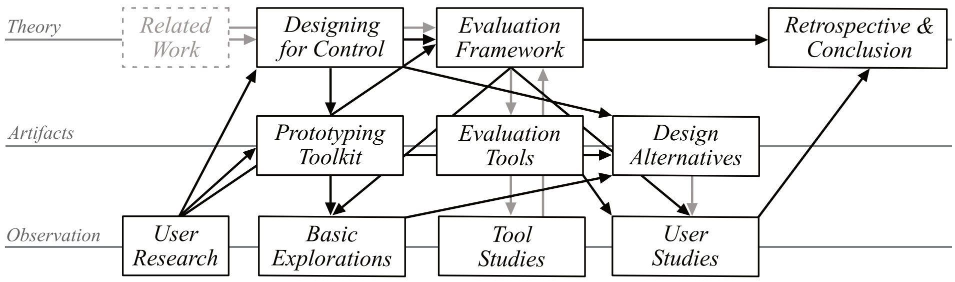


Figure 1.1: Introduction: Our efforts represented based on the framework of Mackay and Fayard. Bold arrows mark interactions between chapters of this thesis and light arrows progressions within conceptual units.

As this statement is not necessarily indicative of the details of our research approach(es), one might further refine it, for instance, based on the propositions of Mackay and Fayard [167]. We introduce their ideas more extensively in Section 7.4. Briefly summarised, they advocate for a mixed-methods approach in HCI research that should aim at a triangulation between theory, the design of artefacts and (user) observation. Besides the introduction of the concept, they also provide a visualisation that displays how individual projects (or steps in a research project as in their case) might be mapped onto the triangulation process. Adapting their initial visualisation, we mapped our projects presented later in this thesis onto their propositions (Figure 1.1). Integrating their perspective in written form into our initial statement, one might characterise our work as follows:

In a mixed-methods User-Centred Design process that triangulates between theory, artefacts and observations, we make use of Grounded Theory to infer theoretical propositions on designing for control and on an evaluation framework which is informed by initial observational user research. Both propositions are further embedded within a Design Thinking framework and are used to guide the design of artefacts, in particular, a prototyping toolkit as well as instances of UI design alternatives. The created artefacts are evaluated using our proposed evaluation framework in further observational user studies and the collected data analysed statistically to update our theoretical understanding.

1.4 Contributions

Similar to the previous section, also our contributions might be related to existing conceptualisations. For instance, they might be framed based on the ideas of Wobbrock and Kientz [277] who suggested that contributions in HCI can be classified as either empirical, artefact, methodological, theoretical, benchmark/dataset, survey or opinion. Based on the generic statement of the previous section expanded by their classification one could also describe our efforts as follows:

In a User-Centred Design process in which we empirically investigate users and analyse the gathered data with Grounded Theory approaches, we provide insights that contribute to a deeper theoretical understanding. Embedded within a Design Thinking framework, our initial insights inform the designs of novel artefacts such as our prototyping toolkit or our design alternatives for cinematic UIs. Based on a proposed evaluation framework that we further contribute methodologically, the artefacts are empirically examined in user studies. Statistical analyses of the herein collected data, in turn, allows for inferences enabling informed opinions in opting for design choices.

Further, one could think of expanding this statement by merging it with the description of our approach based on the ideas of Mackay and Fayard. However, as such further layering might quickly become too complex to be parsed easily, we rather prefer to summarise our contributions as follows:

In sum, our research efforts aim at gathering information and forming abstractions that might be useful in approximating an answer to our RQs raised before. For us, this entails improving our understanding of users and their usage contexts via ethnographic studies as well as studying design alternatives promoting an experience of control empirically based on our suggested prototyping toolkit. As this requires the ability to estimate the quality of our proposed solutions, we further investigate a set of (to us) relevant metrics (derived from initial user research) and how to integrate those into an evaluation framework that supports us in drawing inferences - or put in a nutshell:

Semi-automated tools provide benefits to cinematographers, but also take away desired control. To counteract associated negative experiences, designing for de facto and perceived control becomes necessary. This thesis establishes a user-centred foundation for understanding experts, tools and usage patterns. Methodologically, it contributes a domain-specific prototyping toolkit for in-situ deployment and an evaluation framework applicable to various domains within HCI. On an empirical level, it summons the results of a structured exploration of alternative prototypical designs that are embedded in a mixed-methods approach triangulating between theory, artefacts and observations.

While these summaries are more specific than our initial generic statement, they still only represent more of a top-level view on our work. To also provide a more detailed description, we complied a more extensive reflection on our contributions in Section 10.2 at the end of this thesis.

1.5 Thesis Structure

In the previous section describing our research approach, we used the triangulation framework of Mackay and Fayard to frame this thesis. One could also represent its structure differently by relating it to our initially presented ROs and RQs. Given this approach, one could say that Part I - Perspectives and Fundamentals is associated with aspects relating to RO1 (Understanding). In particular, RQ1 (prerequisites for empathising) is addressed in Chapters 2 (Technology) and 3 (Users). Here, we outline the historical development of cameras and tools for camera motion from its initial stages until its recent transition into its digital era and present an analysis of contemporary systems and their associated UIs in Chapter 2. Further, Chapter 3 addresses RQ1 extensively by presenting a literature review (covering practices, shot types and narrative functions) and reports on our initial investigations in form of an online survey, participation in film-school seminars and contextual inquiries. Later in the chapter, also aspects addressing RQ2 (central motives) are presented with, for example, the analysis of expert user interviews which are reported in Section 3.5. Further information relating to RQs1, 2 and 3 (principles that can be leveraged by HCI) are in part presented in Chapter 4 (Research) with a thorough review of the relevant related work on the state-of-the-art in research on automation, camera control and the human sense of control. Further details relating to RQ3 can be found in Chapter 5 (Design) where we present our propositions in terms of an explanatory framework informing the process of Designing for Control as well as specific design challenges and design variables that we consider relevant in designing for camera control.

Likewise, Part II - Applied Research, may be mapped onto addressing RO2 (Designing for Control). In detail, answers to RQ4 (prototyping and evaluation) can be found in Chapter 6 (Prototyping Toolkit and Evaluation Framework) where we investigated prototyping options (paper, hybrid, virtual and physical) as well as our proposed evaluation framework integrating explicit and implicit measures (on workload, sense of control and creativity support). RQ5 (insights from empirical studies) might be addressed in parts by a description of our research projects as presented in Chapter 8. In detail, we report on five projects with TrackLine investigating a technique for refining content-based interaction in cinematic UIs, Progressive Reduction examining the use of visual reduction strategies, Axis+Content, where we integrate manual and content-based control with a visual reduction strategy, Mid-Air Gestures, where we examine the outsourcing of functions to mid-air gesture to further save screen-space and In-the-Wild which aims at identifying and addressing problems in current UIs.

Finally, in Part III Retrospective and Conclusion, further answers regarding RQ5 are provided taking on a broader view in Chapter 9 - Retrospective presenting a summarising retrospective discussion from the perspectives we introduced earlier to structure Part I (Technology, Users, Research and Design). Subsequently, we present conclusive statements of our overall insights in Chapter 10 - Conclusion accompanied by remarks on the limitations of our efforts, our contributions and ideas for future work.

Insights

- An innate human need for fiction lets a great number of people gravitate towards media in general and historically more recently so towards cinematography. While this form makes substantial use of visuals, it characteristically further integrates various other types of media which together unfold over time. This integration of modalities makes it a fertile ground for engaging storytelling which seems to be one of the most primordial and powerful devices for propagating information and fiction across time, cultures and media and thus strongly contributes to satisfying this intrinsic need.

- Similar to other media, (storytelling in) cinematography is encompassed by a set of stylistic devices and production practices. As they are ideated and implemented behind the scenes, those are often not perceived consciously by most consumers. Nonetheless, they are vital to experts and enthusiasts that want to tell a story cinematically. Advancing the craft and bringing more options to the table, novel tools for these user groups emerged recently due to the integration of computing power.

I

Perspectives and Fundamentals

STEVEN SPIELBERG (DIRECTOR)

My film-making really began with technology. It began through technology, not through telling stories, because my 8mm movie camera was the way into whatever I decided to do.

LUCY WALKER (DIRECTOR)

With portable cameras and affordable data and non-linear digital editing, I think this is a golden age of documentary film-making. These new technologies mean we can make complicated, beautifully crafted and cinematic films about real-life stories.

2

Technology: Digital Cinematography

What to expect?

- Introduction to the domain
- Chronology of camera systems
- Presentation of the status quo of digital movie cameras and motion control systems
- Analysis of advancing factors and user interfaces

What to take away?

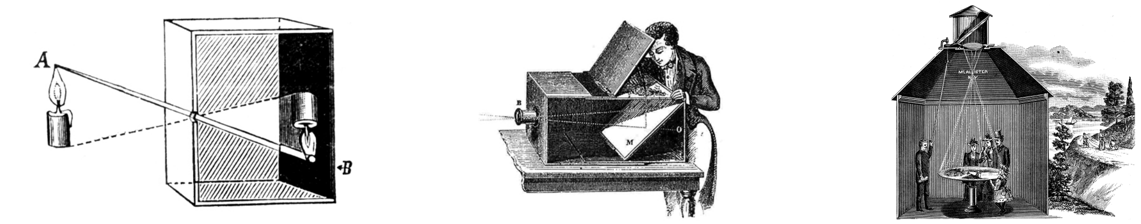
- An understanding of what is and what could be, in particular regarding the technologies, systems and potential for future development

2.1 Cameras

The focus of this thesis is primarily on designing and evaluating user interfaces for camera motion control tools. To introduce the domain first, we will briefly summarize the technological principles, origins and chronological advances of photographic, cinematographic and motion control systems. This should *enable an understanding of the status quo and help to motivate future advances*. The information presented below is mainly based on the work of Marchesi [170].

The Camera Obscura

The genesis of today's digital video cameras follows a tradition of analogue systems, the origins of which can be traced back over two millennia. Its most fundamental form is the camera obscura[1]. One of its first descriptions can be found in the writings of Chinese philosopher Mozi [131] who lived in 4th century Before Christ (BC). While religious and artistic usage of the phenomenon is



(a) Camera obscura principle [400] *(b)* In form of a box [401] *(c)* In form of hut [409]

Figure 2.1: Technology: Principle and applications of the camera obscura

already assumed by human beings living as early as the Paleolithic (as proposed by Gatton and colleagues [96, 97]), it was increasingly used in the centuries following its first mentions. Here, it often served the purpose of a support tool for naturalistic drawing (Figure 2.1b) or that of a study apparatus to examine the nature and mechanics of light (Figure 2.1c). For instance, Euclid (~300 BC) is assumed to have used it for observations preceding his book Optics [28] that presents his conclusions of a rectilinear expansion of light and its underlying mathematics. Following the approach of controlled experimental testing and inductive reasoning, Arab physicist Ibn al-Haytham utilized it in his experiments. He collected empirical data and summarized his studies and insights in his Book of Optics written from 1011 to 1021 Anno Domini (AD) [229, 230]. His work provided a fundamental framework and influenced much later research on vision and optics, as for example by Leonardo da Vinci (1452–1519) [210] or by Johannes Kepler (1571-1630) who used it to inform his theory of the retinal image [158].

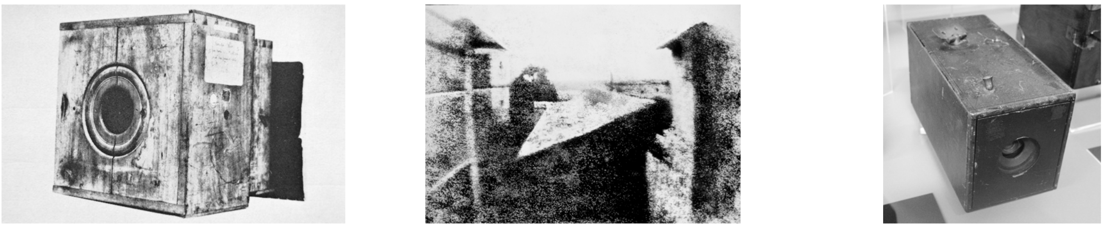
[1] Also referred to as cubiculum obscurum, cubiculum tenebricosum, conclave obscurum, locus obscurus [93] or pinhole camera [220].

Its working principle is relatively straightforward. In a camera obscura, light enters through a small pinhole into an otherwise closed environment. The surface on the opposite side captures it, where a vertically and horizontally mirrored image appears (Figure 2.1a). Even in this simple setup, fundamental photographic properties can already be altered. Given a pinhole of a fixed size, moving the capturing surface can change how much of the scene is displayed. In consequence, this change leads to different angles of view and, for instance, can be observed today when cameras make use of variously sized films or sensors. If the size of the pinhole is varied, the position of the focal plane, where the image appears most sharply, is affected. The resulting distance is also referred to as the focal length and can today be changed by using different lenses for instance. Manipulating either the size of the pinhole or the position of the image screen (or both) also affects the brightness of the image. Today this relationship is referred to as the f-stop ratio. In a camera that uses light-sensitive material or sensors to capture images, this would consequently affect the exposure time. Adjusting such camera properties as position, focal length and angles of view consciously and in congruence to the semantic attributes of a scene is the foundation of stylistic devices in cinematographic storytelling (Section 3.1).

Towards Photographic Cameras

While such basic photographic properties could already be observed and altered, still gradual improvement was necessary to evolve from the camera obscura towards today's cameras. In 1550, Hieronymus Cardanus added a converging lens to the pinhole and observed a brighter and sharper image. Daniele Barbaro assessed in 1568 that the introduction of an aperture could additionally be used to increase the sharpness of the resulting image further. Athanasius Kircher describes a portable version of the camera obscura in 1646. Various findings that would incrementally advance the understanding of photosensitive materials and their reactions appeared in the 17^{th} and 18^{th} century. These observations and developments were all necessary to provide the foundation of photographic processes. Building on top of them, Nicéphore Niépce tried to capture a photograph on carneous silver paper with a modified camera obscura in 1816, but he remained unsuccessful with the image's fixation. In 1822, he used a different approach and was thus able to create a permanent image for the first time. He used an asphalt-covered glass plate and referred to the process as heliography. With a revisited version of his earlier setup (Figure 2.2a), he could capture what is commonly believed to be the earliest surviving photograph of a real-world scene in 1826 (Figure 2.2b). It was taken with an asphalt-covered tin-plate and needed an exposure time of more than eight hours. The same year Niépce met Louis Daguerre, and in 1829 both agreed on a partnership contract to further advance photographic processes. In 1835, Daguerre found a method to capture positive images that worked with much shorter exposure times and by 1837 he was able also to fixate these images. Putting it all together, he provided a process, namely the Daguerreotype [468], that could take permanent images within minutes using an advanced camera ▶
obscura. In parallel, William Talbot worked on a negative-positive process based on a carneous silver layer and was able to fixate the image with sodium chloride in 1835. In 1840, Alexander Wolcott built

a camera based on the Daguerreotype process that was granted the first US patent for a photographic camera. He used a wooden camera body and added a concave mirror and an adjustable image plate to its interior. It allowed to focus an image by adjusting the screen's position [492]. In 1841, Peter Voigtländer built a metal camera producing Daguerreotypes. Fuelled by further advances in the photo-



(a) The camera obscura used by Niépce for his experiments [335]

(b) View from the Window at Le Gras (1826) by Niépce [415]

(c) Kodak Nr. 1 roll-film camera manufactured from 1889-1895 [406]

Figure 2.2: Technology: First photograph and examples of positive and negative-positive process cameras

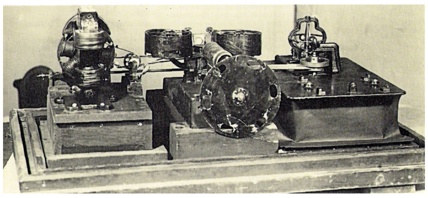
chemistry of the negative-positive process, George Eastman founded Kodak and produced roll film by 1888. Kodak, at the time, would not only provide the roll films but also robust camera hardware such as the Kodak Box and Nr.1 models (Figure 2.2c). These models could be loaded with Kodak films and once exposed, could be handed back to develop image positives. By offering cameras, roll film and a development process, Kodak enabled the era of amateur photography or as put in its advertisement slogan *"You Press the Button, We Do the Rest"*.

Towards Cinematographic Cameras

Concurrently, inventors and engineers were also interested in capturing movements. A system that could be considered an early predecessor towards cinematographic cameras was invented in 1845 by Francis Ronald and deployed at the Kew Observatory [222]. Driven by a clockwork, the system used a photo camera to record values of observation tools minute by minute for periods of 12 or 24 hours. It allowed documenting changes without a person needing to be present and can be considered an early surveillance system. The first patent regarding a genuine cinematographic system was applied for in 1876 by Wordsworth Donisthorpe [334]. It was intended to capture images with a rate of multiple exposed frames per second (fps). While it was one of the very first to be patented, yet another was used to record the material often considered to be the first surviving motion picture. The recording is referred to as the Roundhay Garden Scene [429] and was captured with the LPCCP Type-1 MkII developed by Louis Le Prince in 1888 (Figure 2.3a). A multitude of systems was developed in the years to follow with varying degrees of success. While the Chronophotographic Camera engineered by William Friese-Green was perceived as unconvincing due to its low frame rate and unreliability by the contemporary press [22], also more promising systems were proposed. Among those were the

Kinetographic Camera (1891) by William Dickson (Figure 2.3b), who worked for Thomas Edison, the Pleograph by Kazimierz Prószyński (1894) and the Cinématograph (Figure 2.3c) by brothers Auguste and Louis Lumière (1895). As Edison favoured a separate device for presentation (the Zograscope), the Kinetographic Camera was intended as a single-purpose device only capable of recordings. The

(a) LPCCP Type-1 MkII by Le Prince (1888) [417]



(b) The Kinetograph by Dickson and Edison (1891) [402]

(c) The Cinématograph by Auguste and Louis Lumière (1895) [405]

Figure 2.3: Technology: Early cinematographic systems of the late 19th century

Pleograph and later the Cinématograph, on the contrary, could also be used for projections (and even as developers or copiers). The Cinématograph was also built as a portable device and thus allowed to record exterior scenes such as horse trick riding [539], blacksmiths working [528], a gardener tricked into sprinkling himself [534] or people bathing in the sea [532]. The portable nature of the system also fostered public projections. Given the technological advances of the early 1890s, systems produced by the mid-1890s were already stable enough to be offered as mass-market devices, and projections of the recorded material were able to attract a larger audience. The first camera that could be operated hand-held, the Aeroscope also by Prószyński, was developed in 1909. The same year, Bell & Howell used the 35 millimetre (mm) film as the standard format for their Kinodrome projectors, film perforators, and the cameras making the format the standard of this emerging industry. With their Standard 2709 camera model [320] they also introduced the first all-metal video camera. Still, decades of technical improvements were necessary to produce modern cameras as used in major feature film production, but its technical foundations were laid out in its fundamentals at the beginning of the 20th century. Engineers increased the capabilities of systems incrementally, and so it became possible to shoot colour film and to capture audio along with the video. From 1935 on, the Bolex H 16 [315] was sold and became one of the best-selling cameras of its time. It similarly did for cinematography what the Kodak Box did for photography. Being a reliable and versatile hand-held consumer camera, it enabled many aspiring filmmakers, among those also the young Steven Spielberg. In the 1960s, the Super 8 format and cameras were sold and further fostered amateur cinematography. In that regard, these cameras can be seen as the equivalent of modern-day Digital Single Lens Reflex (DSLR) cameras. Also in the professional domain, technical limits were pushed which led to the introduction of the Super and Ultra Panavision 70 formats and correspondent cameras starting in 1959.

Towards Digital Cinematography

Noticeably, with the introduction of digital systems, a significant caesura occurred. Starting with the 1980s, Sony provided the High Definition Video System (HDVS) format for recording motion pictures electronically and with it, cameras using that standard. One of these models was used to shoot *Julia and Julia* (1987) which was the first feature film to be recorded on an electronic medium [342]. *Windhorse* (1999) became the first movie to only use electronic media as well as only digital post-processing. While theatres at the time still kept mainly projecting from negative film, the movie was therefore converted and distributed as 35mm film. As, in general, electronic cameras could not yet replace analogue equipment in the professional domain, this was a common practice. Here also, gradual improvement was necessary for the technology to mature and become adopted. A critical impact was observable when fully digital cameras appeared in the professional domain. In contrast to their predecessors, they would use digital sensors and store the recordings on digital media. This enabled, easier editing, storage, copying, distribution and integration with Computer Generated Images (CGI). By 2009, the first movie mainly shot digitally was awarded an Academy Award for Best Cinematography [363] with *Slumdog Millionaire* (2008). With the slow transition of professional production tools towards digital, also motion picture distribution and projection would follow and make the shift towards digital formats. In 1999, only four theatres installed digital projectors for screenings of *Star Wars: Episode I – The Phantom Menace* (1999). By 2016, already 98.2% of cinemas around the globe have converted to digital distribution and screening technologies as analysts claim [333]. In parallel, also new digital non-linear distribution platforms emerged such as Netflix with over 100 million subscribers by 2017 [552], YouTube with over 600 million unique monthly users by 2016 [375] or Vimeo with over 800 million subscribers in 2017 [376]. So, summing up the recent historical developments in cinematographic production, distribution and consumption, one could say that starting with the 2010s filmmaking transitioned into its digital era.

2.1.1 Digital Cameras: Timeline and Taxonomy

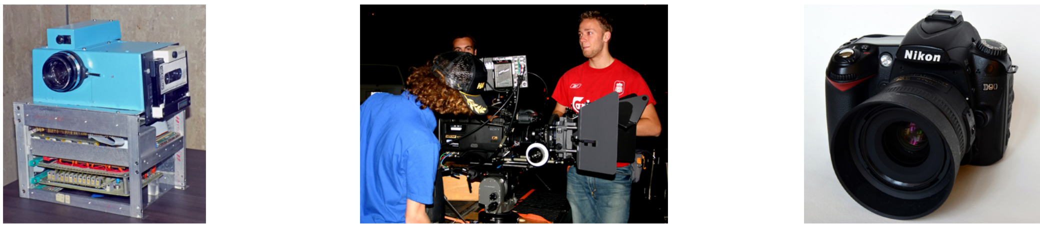
For this thesis digital video cameras, especially in the context of professional cinematography, are of primary interest. Viewed from an HCI perspective, filmmaking is a niche application domain with specific requirements, in particular regarding the design of cameras and motion control tools. In the professional realm, these can differ from regular consumer products quite substantially. To ground our understanding of these systems and their designs, we present how digital cameras developed over time and also the contemporary state-of-the-art. Additionally, we describe a taxonomy that offers a differentiation of cameras based on their fundamental aspects and usage domains.

Timeline

The history of electronic imaging started in 1897 with the invention of the Cathode Ray Tube (CRT) by Ferdinand Braun and the experiments on intentionally deflecting cathode rays by Joseph Thomson. They provided the fundamentals of using the CRT for television broadcasting and reception. In 1906, Max Dieckmann and Gustav Glage produced first raster images in a CRT. Boris Rosing already used the technology as a receiver for a video signal and could display basic geometric shapes in 1907. In 1908, Alan Campbell-Swinton theoretically outlined the concept of transmitting television wirelessly using a CRT in Nature magazine but remained unsuccessful in experiments trying to realise this vision together with George Minchin and J. C. M. Stanton (1926). The same year, 1926, Kálmán Tihanyi developed an electronic Television (TV) system, the Radioskop, based on a camera tube and enhanced by charge storage that helped in overcoming the low light-sensitivity of the existing devices. Modern TV devices still used the principle he developed. Also in 1926, Kenjiro Takayanagi publicly displayed the first working TV with a 40-line resolution using a CRT display as a receiver, but did not apply for a patent. Earlier, in 1921, 14-year-old Philo Farnsworth envisioned scanning images with a beam of electrons in consecutive rows [234]. Following his idea, he built a camera that was able to capture the first electronic picture in 1927. In cooperation with Vladimir Zworykin, he was able to put together a working TV using a CRT and wireless transmission of the signals in 1934. Their invention was the foundation of stable electronic TV recording, broadcasting and receiving that later became the standard TV format and devices of their time. Besides broadcasting, there was also a great interest in storing these electronic transmissions. For this purpose, consequently, various types of storage media became built and evaluated. Magnetic tapes were a promising technology and further became a mass medium in the early 1980s. Therefore, the following generation of cameras was extended by storage capabilities on tape. These also enabled the development of hand-held systems such as Sonys Beta-Movie (1975) or HDVS (1984). Using magnetic tape not only allowed recording already available broadcasts but also many amateur videographers to record and replay their content. Additionally, an alternative distribution format emerged with rental systems. While the mentioned cameras would capture and store the recordings already electronically, yet they still often remained analogue either regarding the signals or the storage processes and formats.

Going beyond electronic-analogue systems, the mentioned shift towards digital systems started in the 1960s with the introduction of digital sensors and storages. In 1961, Eugene Lally already started working on a mosaic photosensor capturing digital images at the Jet Propulsion Laboratory. Bell Laboratories were able to demonstrate a first Charge-Coupled-Device (CCD) that George Smith and Willard Boyle had developed in 1969 [347]. It was at first thought of as a digital storage device but was also observed to be light-sensitive. For their invention, they were later awarded the Nobel Prize in Physics (2009) [554]. The CCD allowed converting light into digital signals and was a central component that enabled digital imaging. Although at first, they could only capture 100×100 pixels, they improved the technology and by 1978 were able to develop a 500×500 pixel array. Michael

Tompsett brought the light-sensitive property of CCDs to further application and built a first solid-state video camera on its basis in 1970. However, the system was not able to also save the images digitally due to technical constraints. In 1972, Thomas McCord and James Westphal addressed this issue in their Vidicon system. It captured images via a 256 × 256 array and further saved them as 8-bit files



(a) First digital camera by Steven Sasson (1975) [392]

(b) Camera of the Sony CineAlta line (from 1997) [414]

(c) First DSLR with video capabilities, Nikon D90 (2008) [408]

Figure 2.4: Technology: First digital versions of video cameras

digitally on magnetic tape [174]. The Vidicon weighed about 10 kilogram (kg) and needed a separate apparatus functioning as the recorder. Thus, it needed to be considered a stationary system. By 1973, CCDs started to become commercially available and were produced for example by Fairchild Imaging. They used the sensor in their Modell MV-100 TV camera which was one of the first digital cameras being available. It only weighed about 170 grams but also only used 100 × 100 pixels and hence was limited to surveillance or medical applications. The first portable digital movie camera was developed in 1975 by Steven Sasson at Kodak (Figure 2.4a). The system used the Fairchild CCD, weighed about 4 kg and was able to record black and white material digitally on a cassette tape. The same year, Bell Laboratories built the first CCD-based camera with a resolution suitable for broadcast television and Bryce Bayer invented the Bayer Colour Filter Array that enabled the capturing of colour images. Also in 1975, a new sensor technology based on a Metal-Oxide-Semiconductor (MOS) was firstly commercially available with the Cromemcos Cyclops Camera. However, with a resolution of 32 × 32 pixels, it was not yet suited as a mass-market product. In the late 1980s, camera manufacturers rolled out models with higher resolutions and larger storage capacities. In 1981, the Sony Mavica (Magnetic Video Camera), a video camera which recorded images on a floppy disk, was presented. Roughly ten years later, also the introduction of digital cinematographic cameras could be observed, as with, for example, the Sony HDCAM format and cameras. Further technical improvements followed in the 1990s, especially with upcoming of Compact Flash memory cards (1994) that allowed building cameras way smaller. The use of digital technology in audio-visual media production became increasingly ubiquitous and, remarkably, *Vidocq* (2001), shot with a Sony HDW-F900 (Figure 2.4b), was the first all major digital production [344]. In parallel, DSLR cameras for photography were increasingly produced in the 2000s, for instance, in 2003, Canon manufactured one of the first affordable DSLRs cameras for non-professionals with the Digital Rebel. In 2006, Nikon started transitioning to the production of only digital models. Their D90 model (Figure 2.4c) was the first DSLR that became extended by the ability to also record movies in 2008 and thus enabled enthusiast photographers and videographers alike. For the particular application domain of outdoor sport, the development of action cameras also emerged in the early 2000s, as with the GoPro models in 2002. These allowed recording with similar resolutions as professional cameras but could

(a) The Red One camera a first major digital camera (2007) [411]

(b) The Arri Alexa often used professionally (2010) [398]

(c) The Lytro cinema for light-field recordings (2016) [396]

Figure 2.5: Technology: Digital cameras for professional motion picture and light-field recordings

be built at very small sizes. Hence, they allowed the exploration of novel Point of Views (POV) and the easy integration into supportive MoCo tools.

While manufacturers increasingly offered digital cameras, cinematographic professionals still often preferred the proven and tested analogue cameras. Digital cameras focusing on addressing the needs of professionals challenged this. In 2007, Red Digital Cinema presented their first digital model with the Red One (Figure 2.5a). It was able to record with a 4K resolution at 30 fps, 120 fps at 2K and could provide a cinematographic look regarding the Depth of Field (DOF) that operators often found missing so far. The National Broadcasting Association awarded it the Star Award and Mario Award for Most Innovative Camera [316]. It also was competitively priced and in hindsight became an essential milestone in the transition of production tools towards digital. Soon after, in 2010, even more, established manufacturers such as Arri offered digital cinematographic cameras, as with, for example, their Alexa model (Figure 2.5b). With the release of these models, more and more professionals accepted digital technology for film recording, and thus they had a significant impact. Since then the process of gradual improvement has started. In 2008, also live-action Three Dimensional (3D) digital recordings became possible with the Fusion Camera System which was used to record of *Journey to the Center of the Earth* (2008). Manufacturers released various cameras, such as the Blackmagic Cinema Camera, the Sony CineAlta series, the Panavision Genesis and further models of Red and Arri. Among those are also the open-source cameras of apertus (AXIOM Alpha, 2013). The EU funded project [551] set out to make professional cinematographic recordings available to a broader audience in an affordable and customizable way. Advancing technical possibilities further in 2012, Lytro released the first commercially available light-field camera. In the past, cameras would create images only by capturing information on light intensity and colour. In contrast to this approach, light-field cameras were engineered to additionally register information about the direction that light is travelling in space. The additional information allows a variable DOF, faster recording, increased low-light performance and to automatically generate 3D images. The created 3D scenes can also be integrated with CGI or Virtual Reality (VR) more easily. In 2016, they presented their cinema line (Figure 2.5c). The system consists of a camera with up to 755 RAW Megapixels and 300 fps and a connected server processing the light-field data. This setup allowed refocusing and adjusting the DOF, the frame-rate and the shutter angle in post-production and automatically can generate a depth map of the scene.

Taxonomy

In the section above, we described various types of cameras and their historical developments up to the status quo. An essential differentiation can be made between professional and non-professional devices. However, due to the variance in available cameras especially in the modern days' digital era, this binary separation is not necessarily accurate. *For further reference* and to provide a more technology-based classification, we would like to offer a more refined taxonomy of cameras below. It can be used to classify cameras based on their technical characteristics, limitations, strengths and intended user group. We also provide contemporary examples for each class.

Mobile Device Cameras

Today's smartphones are ubiquitously available and feature cameras capable of recording video. As designers intended use by a majority of amateur users, they are hardly used in professional environments regarding final recordings. They can be characterised by their relatively low costs and resolutions that often are below those of more advanced systems. As a consequence of their intended use by amateurs, frequently control options for aperture, DOF or shutter speeds are hidden from the users. UI elements are rarely built as hardware elements[2] and the available parameters can be changed via software UI components that are displayed on top of the camera feed. Examples of these can be found in most modern-day mobile devices ranging from smartphones to tablets with varying designs.

Action Cameras

The small size and low costs of action cameras make them particularly attractive for a broad range of users. Historically, they were primarily intended for outdoor sports enthusiasts but nowadays are also often used as multi-purpose cameras for general documentary purposes. Due to their intended sports use case, they are built to be small, robust and sometimes even waterproof. The cameras resolutions are often comparable to professional equipment however their sensors are smaller which affects the DOF. Additionally, they are also limited in other regards. Usually, a fixed wide-angle lens is attached which does in consequence only allow to record a subset of possible shot types and leads to a particular look that is not necessarily cinematic. Also, their dynamic range is lower compared to

[2] Sometimes existing buttons dedicated for a different primary use (such as the volume buttons) become reused, and overloaded.

professional equipment. For their control, often a small number of hardware buttons are provided, but most parameters need to be remotely controlled mostly via mobile devices. While this type of cameras is frequently associated with GoPro models, also further manufacturers such as Sony, Polaroid, Ricoh or Garmin offer similar models.

Digital Single Lens Reflex and Mirrorless Cameras

Especially in the domain of photography, DSLR and mirrorless cameras are intended for users varying from beginners to professionals. Regarding videography, they increasingly optimised performance so that nowadays they provide resolutions comparable to professional movie cameras and incorporate even full-frame sensors. High performance at small to medium body sizes makes them particularly interesting for users with an interest in creating a broad range of visual content (photo, video, indoors, outdoors). One of the primary benefits is that they offer to interchange lenses that allow recording a broader diversity of shot types including particular cinematographic ones. These lenses often also allow adjusting the focus during video recordings intentionally which is commonly used as a stylistic device in filmmaking. Furthermore, external devices can be connected that support a more professional record of the audio material. In general, these cameras have a more extensive set of hardware controls and allow users to control all essential photographic parameters. Still, sometimes the set of control options is restricted[3]. While more traditional photo companies such as Canon, Nikon or Olympus are more associated with DSLR models, companies such as Sony, Fuji or Panasonic could establish well-accepted models especially in the domain of mirrorless cameras.

Camcorders

DSLR cameras were by design primarily intended for photography and later extended by video capturing. Camcorders, on the contrary, were designed mainly for videography and then extended by still imaging capabilities. Originally, camcorders were produced as a combination of video cameras and tape recorders. At first, they would save analogue signals to the cassettes, but increasingly also supported to keep digital representations. Over time, further, the storage medium changed from cartridges to external memory cards or internal flash memory. In general, they can be operated shoulder-mounted or hand-held with the primary purpose of recording video, but many of the contemporary devices also allow to take still images. The devices mostly offer a set of interchangeable lenses and were targeting consumers as well as professionals. The resolution of the devices is comparable to advanced systems, and more recent models are also capable of 3D recordings. Camcorders also tend to provide more hardware UI elements, especially in professional models. Similar to DSLR and mirrorless cameras, they also allow to change the aperture, focus or zoom level of a lens during the recording and to connect external devices for audio recordings or reviews of the camera feed on larger displays. While in the consumer segment various manufacturers such as Panasonic, Canon or Sony offer products and established market shares, in the professional segment Sony camcorders seem to be used most often.

[3] For some manufacturers, third-party firmware sometimes tries to provide access, e.g. Magic Lantern for Canon Eos cameras [356].

Professional Video Cameras

For professional recordings of digital cinematography, fully digital cameras emerged in the past decade. As stated above, the introduction of the Red One [368] had quite an impact in 2007. Since then, the manufactures improved their models such that they are currently offering systems capable of more than 35-megapixel resolutions, a dynamic range of over 17 stops and frame-rates of up to 240 fps (with 2K resolution) with sensor sizes similar to a (35 mm) full-frame sensor. Systems such as the Red Weapon [369], Arri Amira [310] or the BlackMagic Ursa [314] also have a relatively small form factor, the Red Weapon, for instance, only weighs about 1.5 kg. Due to their comparatively small form factor, such cameras can be used in pairs for the recording of stereoscopic 3D films or can be attached to motion controls tools that allow stabilising and guiding the camera movement in a dynamic scene. These cameras frequently offer hardware buttons to change settings and simultaneously offer remote control via additional hardware or software. The systems support interchangeable lenses, a rich set of parameters customizable by the users, provide ports to attach further audio equipment or reviewing monitors. Sometimes the systems also support wireless connections for control or review devices.

Light-field Cameras

Compared to the other classes of cameras, the market for light-field-cameras is quite different. With Lytro, only one major producer started to offer the cameras commercially. The targeted user groups varied. In the beginning, Lytro designed for a broad audience with its Original Lytro camera (2012). The camera would only feature one hardware button as the following manipulation was intended to be made on a computer. The camera featured a 10.5 mega-ray sensor was still quite low compared to later models. The resulting Two Dimensional (2D) images were interpolated to a resolution of about 1 megapixel. The second-generation Illum models (2014) were tailored to more advanced users and had a form similar to a DSLR. It featured a 40 mega-ray sensor resulting in about 4 megapixel 2D images. It was also equipped with more than only one button, but the emphasis regarding UI elements remained on the post-production. With their cinema line (2016), they addressed a professional context, especially in combination with VR or CGI. To operate a light-field system, an additional server is required, and traditional camera operators would not necessarily be accustomed to operating these systems instantaneously.

2.1.2 Advancing Factors

Throughout the history of cameras, some factors influenced the progress and acceptance of technologies more strongly than others. Ahead of the transition towards digital systems, these can, in general, be found in form factors, quality of the recording and associated costs. While these are also relevant to the introduction of electronic and digital systems, an additional relevant consideration emerged with connectivity to other devices.

Form Factors

Regarding the form factors of the devices, a reoccurring pattern is that cameras that could provide a robust exterior were favoured. This preference could be seen, for instance, with the different public reactions towards the Chronophotographic Camera and the Kinetographic Camera. Also, smaller sizes and lower weights are essential especially in the phases when certain types of cameras became mass-market devices and popular with amateur and enthusiast users. For example, this was central for the introduction of the Kodak Box or Super 8 video cameras regarding analogue systems but also for early digital DSLR systems. Moreover, a common theme seems to be that the more a system is tailored towards expert users, the more hardware UI elements are preferred. These not only allow quick access to primary features but also enable an eyes-free control [199] when focusing on the viewfinder, for instance. The inclination towards hardware UI components on a camera's exterior could already be found in analogue systems, but it grew notably with upcoming of electronic or digital systems such as the CineAlta camcorders, Panavision cameras and today's digital successors Arri or Red Digital.

Recording Quality

Sometimes the quality of the recording is prioritised over form factors. Choices along this trade-off in favour of recording quality became apparent in professional-grade systems, such as the Panavision cameras. These systems were large and heavy but could provide an imaging quality that was particularly wanted. Recording quality was a requirement in the transition to digital. Thus, the characteristics of sensors in terms of resolution and regards to its size became central aspects based on which systems would be rated. The results needed to be comparable to the quality of the analogue film to become accepted on a professional level. For digital systems, further, the sensing quality in the domains of frame-rates, depth-sensing or light-field sensing are nowadays central factors. For the quality of the resulting images, not only the sensor is crucial but also the lenses in front of the sensor. In general, systems that offered an ecosystem of interchangeable lenses, especially in the premium segment were more likely to be accepted. Therefore, continuing to implement existing lens mount formats or adapters for analogue lenses on newer mount systems was important. With the ability to use established high-quality lenses also the more subjective aspect of an individual cinematic look could be addressed. The look was relevant to camera operators and often found missing. Consequently, systems that allowed to create this look were preferred over other models.

Costs

Low costs are naturally attractive for a multitude of users generally speaking. For a new technology to become accepted it needed to be priced at least competitively. In the beginning, digital systems were already developed, their impact was still low as the prices were quite high and the imaging quality not necessarily comparable to the results of analogue cameras. The competitive pricing of the Red One models, however, enabled it to become accepted. The growing acceptance not only provided the manufacturer with the resulting revenue but as a side product opened the market for devices of this type in general and hence provided a foundation for the switch to digital.

Connectivity

One of the major benefits digital that cameras can provide is their high degree of connectivity to other devices. For camera recordings, this can entail a multitude of devices being designed for particular subtasks of cinematographic production. Among those, one can find external monitors for reviewing the resulting images on larger displays on set and in real-time. Small displays, for instance, make it difficult to adequately judge the position of the focal plane and the resulting DOF. The camera feed can be streamed to multiple devices and not only provide the camera department with the resulting image but as well the director or the make-up and costume department. Nearly all departments are interested in the final look of their work already during the shooting to be able to adapt quickly. Digitally, already some post-production steps like colour grading[4] for instance (see also 3.1.1), can be visualised or be already conducted on set. For the safety of the recorded material copies of all the digital data can already be made on location and by now are even often required by film-insurances. Digital systems often allow extending the capabilities of former mechanical systems or humans. Due to wireless connections, cameras can be controlled remotely, which helps to bring cameras to POVs that are hard to get to or dangerous for operators. Bringing cameras to novel POVs and crafting the motion in-between can nowadays not only be controlled manually via remote control but also be guided by smart systems that analyse the camera feed and triggering the actuation of a supportive MoCo tool.

[4]An intentional colour correction process used for artistic effect. It often involves cooling or warming the colour temperature in congruence to the semantic properties of a scene. It can also incorporate the use of motion tracking and masking software.

2.2 Camera Motion

In the early days of cinematography, locating the camera at a central spot facing a scene of a theatrical play was a popular approach, and hence the camera was hardly moved at all. Motion pictures were still a developing medium, and artistic exploration commonly second to contemporary theatre conventions. Further examination of camera motion led to its recognition and application as a cinematographic stylistic device. Until today, it often serves the purpose of depicting and advancing the story and is mostly applied subtly. Thus, camera motion itself is frequently not perceived consciously by the audience, but its on-screen effects still play a vital role in filmmaking. It has advanced quite substantially in its contemporary form and along with it led to a multitude of tools for its support.

2.2.1 Motion Control Systems: Timeline and Taxonomy

In their experiments, operators at first tried to position the camera at different locations. Later on, they looked for ways to enable controlled transitions between these. Along with exploring camera motion during a recording, supportive tools were developed to enable the then-new moves. The devices came in various designs to assist with different types of movements. Traditionally, they were built as purely mechanical systems that would mainly constrain the motion to some degree. With tracks, for instance, the motion path would become restricted to a linear path. Such restrictions allowed operators to exert more control over the overall movement and further engineering fostered the quality of the motion regarding smooth transitions most. In addition to improvements regarding mechanical engineering, modern tools became enhanced by computation and with it sometimes diverging concepts for their control. Similar to the previous section, we want to ground our understanding of these systems. Therefore, we describe how they developed over time and present state-of-the-art examples. Also, we want to *offer a taxonomy for further reference that enables a differentiation based on fundamental aspects and usage domains*. As the UIs for controlling MoCo tools are a central aspect of this thesis, we further focus on describing existing approaches and design patterns in this section.

Timeline

In cinematic practice, more dynamic film-making techniques were pioneered at the beginning of the 20$_{th}$ century. Extending the stationary tripods by rotating mounts can be considered one of the first approaches introducing camera motion. It started in the late 1800s and enabled movements such as pans and tilts. Placing a camera on tracks allowed additional new shot types. These were to move it either towards or away[5], in parallel, diagonally or around a scene. The use of such shots started in the early 1900s and was technically enabled by devices that would combine a moving platform, wheels and tracks. One of the first appearances of the so-called dolly shots could be

[5] Also referred to as a dolly-in or dolly-out.

seen in *Cabirira* (1914) [486] or *David Harum* (1915). Also, the effects of sequences of different shot types were explored more profoundly. For example, David Griffith varied different shot types and became particularly famous for his extensive use of close-up shots, as seen in his movie *Intolerance* (1916) [428]. In the shooting of the film, technical problems and limitations due to the

(a) Example use of the unchained camera technique [391]

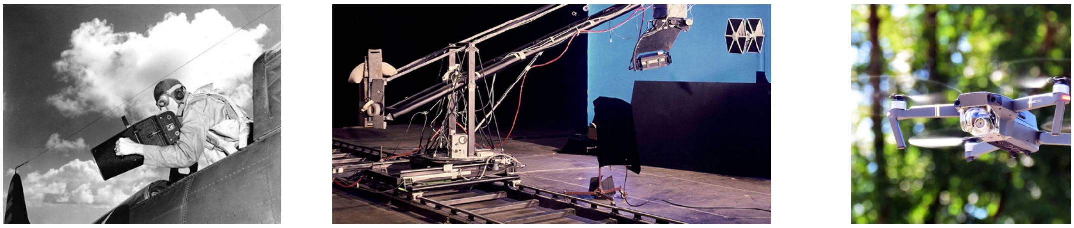
(b) Nouvelle vague directors Jean-Luc Godard and Raoul Coutard [394]

(c) Garrett Brown with Steadicam in Rocky (1976) [393]

Figure 2.6: Technology: Examples of moving the camera in different eras (1920s - 1970s)

dolly's construction became apparent. Those were later addressed in a design by Allan Dwan who combined the dolly with a crane and thus enabled the first crane shots. As now, moves could also be carried out in the vertical axis, it allowed transitioning from close shots into long shots seamlessly. Cinematographers continued to experiment further and in the 1920s began to look for other ways of setting a camera in motion. Therefore, sometimes cameras were attached to people or to moving objects, but due to the massive weight of the cameras, these approaches (and hand-held shots in general) were still rarely used. In 1924, Friedrich Murnau displayed an elaborate use of various shot types in combination with refined camera motion. His technique was later referred to as the unchained camera and was applied in *The Last Laugh* (1924), for instance. It is still regarded as one of the most influential stylistic innovations in 20^{th}-century cinematography and provided a theoretical as well as a practical foundation for its further development. The development of smaller and lighter cameras led to further exploration and new motion support devices. Originating in the 1950s, for example, the French nouvelle vague cinema made extensive use of the newly available hand-held shots. These, however, were sometimes quite unsteady especially when the camera operator would move while shooting. To enable not only hand-held but also smooth stabilised recordings, Garret Brown developed the Steadicam support tool in the 1970s. The system was mounted onto the operator, and so allowed to isolate the motion of the camera from the operator's motion on the one hand and on the other did not require any tracks or dollies. In feature films, it was firstly used in *Bound for Glory* (1976), *Rocky* (1976) [525] or *Marathon Man* (1976).

The same year, also the first MoCo system incorporating computation was developed with the Dykstraflex. It enabled to precisely control camera motion for special effects shots. This precision was a requirement for applying the Bluescreen technique with analogue film material in high-quality productions. The system was built incorporating Transistor-Transistor Logic (TTL) chips which were invented in 1961 and established a foundation for the development of integrated circuits since. Actuated by motors that were driven by the TTLs, the Dykstraflex allowed exerting precise control over the roll, pan, tilt, swing, boom, traverse, track, lens focus, motor drive, shutter control, and to



(a) Fairchild F-1 aerial camera system (ca. 1942) [416]

(b) The Dykstraflex on set [397]

(c) The DJI Mavic quadrocopter drone [404]

Figure 2.7: Technology: Aerial recordings and motion control applications in cinematography

save and repeat motions. The first application of the system was in the shooting for *Star Wars: A New Hope* (1977) [524] and was awarded an Academy Award for Best Visual Effects in 1978 (John Dykstra, Al Miller and Jerry Jeffress).

Recording images and controlling the transition between several POVs was of interest not only in theatres, studios or more generally speaking on ground level but also above the ground. This domain is referred to as aerial photography or cinematography, and its roots trace back to the first photographs of Gaspar Tournachon (also known as Nadar) and his images taken from a balloon[6]. The motion of a balloon is hard to steer, and thus more technical progress was necessary to establish the technology for a more controlled recording process. Experimentation began with utilising kites, rockets or pigeons as carriers of mainly photo cameras. One of the first cameras recording motion pictures was mounted to an aircraft in 1909, and the material was used for the movie *Wilbur Wright und seine Flugmaschine* (1909) [549]. Affected by the first world war, it followed a period of military use mainly for reconnaissance and surveillance. Also, early remote-controlled aeroplanes were used by the military. Reginald Denny, for instance, founded the Radioplane Company that would take military Remote Control (RC) planes and develop it further into commercial systems. After the war, aerial photography was also offered commercially by Sherman Fairchild who used overlapping photos for the creation of a map of Manhattan in the early 1920s. In parallel, Cecille B. DeMille started to acquire aircrafts for cinematographic recordings in 1921. Until the mid-1940s the use of aeroplanes for aerial recordings was the norm. In 1945, a helicopter was firstly used for a feature film recording in *The Bandit of Sherwood Forest* (1945). As aeroplanes by design need to be always in motion, they did not allow several cinematographic shot types such as the dolly-in or dolly-out.

[6] Unfortunately none of the pictures survived. Therefore, the first surviving aerial photo is *Boston as the Eagle and the Wild Goose See It* (1860) [399] by James Wallace.

Helicopters could provide this to operators and hence were often preferred. So far aerial shot needed to be carried out with an operator on board. The need for also carrying a person with the aircraft would change with the introduction of RC helicopters in the early 1970s. A first commercially available system was presented in 1972 with the Bell Huey Cobra by Dieter Schlüter and soon be adapted

(a) Iris by Bot&Dolly [390] *(b)* Edelkrone SliderPlus [413] *(c)* DJI Osmo gimbal [403]

Figure 2.8: Technology: Advanced motion control tools supporting and automating camera motion

by cinematographers. As first RC helicopters were still complicated to operate further technical improvements were added over time [454]. By 2014 the developers of the Helicam system, Gifford Hooper and Philip George, were awarded an Academy Award for Technical Achievement [362]. Their system was highly manoeuvrable even at high-speeds but still needed trained operators. The ongoing technical improvement that made helicopters more stable and comfortable to operate did eventually also lead to systems that could be controlled without much training. These systems are based on a different design, namely the multicopter. Such devices feature four (quadrocopter) or more rotors and hence are more stable during operation. These systems are colloquially also referred to as drones and recently became quite widespread for amateur and enthusiast use. While first quadrocopter designs can be traced back to Étienne Œhmichen in 1920 [471], the development of today's popular robotic drones started in the 2010s and recordings were used for major feature films such as *Skyfall* (2012), *The Wolf of Wall Street* (2013) and *Spectre* (2015) [522]. For enthusiast users, in particular, in the domain of outdoor sports, action cameras became often attached to consumer drones. The drones sometimes became extended by Computer Vision (CV) which allowed to follow a person while conducting sports automatically, as with the HexoPlus [482] or the Vertical Studio [543] system. As such CV capability was not existing in all drone models, it also became available as a separate retrofit device with Percepto [508]. The device contains a small computation unit and can be attached to the control board of a drone. This way, it allows automatic steering of a drone based on optical sensing.

Besides aviation also other technologies were adapted as motion support systems. In the 1970s in particular industrial robots with six electromechanically driven axes were developed for repetitive subtasks in assembling and welding. For instance, as one of the first, the Kuka Famulus was presented in 1973. By the early 2010s, a variety of tools was already available that could be mounted at the ends of the robotic arms and also a camera mount followed. This way, also the large and heavy cameras often used in high-quality productions could be attached and their motion-controlled in six degrees of freedom. In 2012, Bot&Dollys Iris system that was based on a Kuka industrial robot [449] ▶ emerged. Here, the six axes robot was additionally mounted onto a motor-driven platform attached to a track on which it could be moved. Material shot with the system appeared in the movies *Gravity* (2013) [480] and *Man of Tai Chi* (2013) [490]. Also, other manufacturers such as Mark ▶ Roberts Motion Control [500] or igus provide industrial robots for carrying out precise and repetitive ▶ ▶ camera control. Igus, for instance, provided the robots for the TV studio of German ARD public network news broadcast *Tagesschau* (1952 - today) [458]. Industrial robots also not only could be ▶ used to execute existing shot types and moves but also enabled new shot types. While high speed or slow-motion recordings were essentially made possible by cameras that could record at higher framerates, it remained hard to set them in motion. Carrying out such movements would have required an operator also to move and control the camera at very high speeds. At certain speeds, this was outside human capabilities. Robots, such as the Spike system by TheMarmalade [535] or Mark ▶ Roberts Bolt CineBot system [447] were primarily designed for such high-speed tasks and allowed ▶ slow-motion shots to incorporate camera motion that would become quite popular in advertising.

Camera robots, in general, can also come in other forms than the six axes robots. Rather than adopting technology from separate fields, a typical approach was also to extend the already existing cinematic support tools. Platforms on tracks were one of the earliest of these instruments. Similar to the features of the Dykstraflex or other robot systems, gradually the platforms were extended by motors. Also, motor-driven three axes mounts, so-called pan-tilt(-roll) systems or remote heads, were added to those platforms. Modern-day devices of this kind such as the CineDrive by Kessler [433] ▶ or the SliderPlus with MotionKit [502] by Edelkrone provide control over camera translation ▶ and rotation and remote control via mobile devices. Also, existing mechanical camera-cranes (and the helicopters mentioned above) were extended with such remote heads. For its control, such setups would need at least two operators. One responsible for the necessary control of the machine (helicopter, crane or other systems) and one for the remote-head or camera control. These traditional user roles became challenged by further technological developments. Similar to the progress observable in drone control, also for camera cranes, remote heads were extended by CV capabilities. This way one operator could delegate some task to a computational system and therefore was capable of controlling the crane and head at the same time, as with the Stype kit [436] for example. ▶

These new MoCo systems shared the same problem that already appeared in nouvelle vague cinema: missing stabilisation during camera motion. Using a Steadicam, however, was not possible with remote-heads, drones or industrial robots. To address the shaky images stabilisation was provided by new technology with three axes gimbals. The systems are based on the Cardan suspension[7] and an electronic system capable of sensing a deviation from its normal position. The system then actuates

[7] A device consisting of three rings that are mounted with orthogonal pivot axes onto each other. This construction restricts the rotation of an object in every single axis. In consequence, the object can rotate independently from the surrounding structure.

the motors to counteract this aberration. Gimbals can come in different forms so that the deviations of human operators, as with the DJI Ronin [461], or MoCo systems can be counterbalanced equally. Gimbals are also often found as extensions to today's remote-heads to provide smooth images but do further exist in small sizes to stabilise action cameras or mobile devices, as in the DJI Osmo [487].

Taxonomy

We described the historical development of MoCo tools above. These devices can most fundamentally be differentiated between purely mechanical and motor-driven or computer-controlled systems. Due to the variety of instruments that includes mechanical machines, computationally extended systems and adapted technologies from other domains, *we would like to offer a more detailed classification for further* reference below. Our taxonomy is based on technical aspects, the degrees of freedom that a system allows and traditional cinematographic terms and classes. New classes due to the adaptation of novel technology, for instance, were added.

Pan-Tilt(-Roll) Systems (Remote Heads)

As presented in the previous section, pan-tilt systems allow the camera to be rotated. Rotation is mainly carried out around the central axes of the camera, more precisely around the central axes of the sensor of the camera. In mechanical systems, this could be provided by tripod mounts that are operated manually via physical handles. In physical systems carrying out smooth continuous pans or following a moving object passing by still requires practice by an operator. Electronically controlled systems allow remote control which minimises the transfer of unwanted motion by an operator and also make constant panning easier because it can be programmed and controlled parametrically. Sometimes these systems also provide the ability to roll the camera. Such systems are then referred to as pan-tilt-roll. However, as the on-screen effects of rolling are quite uncommon to the regular human visual experience, such shot-types are only occasionally used as a stylistic device in professional productions. The common pan-tilt systems, in contrast, are similar to the regular visual experience and therefore can be found in various contexts including surveillance cameras, amateur films, sports broadcasting, teleconferencing and cinematography. Consumer systems are mostly mounted on tripods or walls and sometimes also incorporate lens controls, mostly zooming and focusing, and hence are also called Pan-Tilt-Zoom (PTZ) cameras. These are designed as general-purpose tools and mainly for various application domains outside cinematography. Commercially available general-purpose cameras are the Axis PTZ systems for example. They often can be remotely controlled via an Application Programming Interface (API) from an Internet Protocol (IP) network [443]. The Soloshot PTZ system is one of the fewer examples where such a system has been designed especially for cinematography. In general, cinematic pan-tilt systems face different requirements and prerequisites due to their particular application domain and thus differ in size and designs. They are also often referred to differently by the term remote heads and are available from manufacturers such as Arri with the SRH-3 [485], Panther with the Maximus 7 [424] and also various others. They are predominantly mounted on camera cranes and remote-controlled by a dedicated operator. Further, they are often designed to carry an additional payload to the camera. Most importantly for additional units to allow control over lens settings such as the position of the focal plane, the opening of the iris or the level of zoom. The combined change of the pan-tilt position and the level of zoom in one move is an often-used technique for establishing a scene or in sports broadcasting. For basketball broadcasting also computationally supported devices exist, such as the KeeMotion system [346]. It was designed to frame the scene, set the focus and track the players while moving automatically [511]. The system incorporates CV and Machine Learning (ML) to carry out these tasks on its own. It tracks individual players and identifies movement patterns among groups of players. From these patterns, it derives triggers that start camera motion, for example, a pan to the opposing basket when an interception of the ball is registered. ▶

Sliders

Sliders fundamentally consist of rails and carriages. The carriage is attached to the rail, and its platform can carry a camera mount or remote head. This design enables a controlled translation of the camera by restricting the motion to a linear path determined by the rail. In combination with a remote head also more complex motion paths integrating translation and rotation can be achieved. Variations of the basic slider design exist, sometimes wheels are attached to the carriage other times the carriages directly slides on the rails. Sliders also come in varying sizes to account for the weight of the mounted cameras which can be mobile devices, DSLR or professional-grade cameras. Depending on how the system is set up, horizontal, vertical and also diagonal shots are possible. Dollying (in or out) is only possible with restrictions as the rail most likely becomes visible in the shot at some point which is mostly regarded as unwanted. Traditionally the motion would be carried out by an operator that would manually move the carriage and operate the remote head directly or via handles. Modern-day systems are motor-driven and technically quite advanced compared to purely mechanical systems. Examples of such systems are the already mentioned Edelkrone SliderPlus, the Kessler CineDrive or the Axis360. As they are not only motor-driven but also digitally controlled, they often can be remotely controlled by hardware controllers or mobile devices. The carried out motions can also be recorded and replayed including lens controls. Computational control not only provided to carry out traditional panning and tracking shots more smoothly and repeatable, but also enabled new shot types such as time-lapse as well as slow-motion shots depending on the system.

Cable Cameras

As rails can become a limiting factor to the length of a shot, also alternative designs have been explored. To bridge wider gaps and also to bridge impassable terrain, cable cameras were developed. Those are based on the principle that the carriage is attached to a cable that is fixated on two points mostly in elevated positions. Again, the platform of the carriage can enclose a tripod mount or a remote head. One of the early uses of cable cameras in feature films can be seen in *Soy Cuba* (1964) [465]. ▶
These systems also come in various sizes to account for the weight of the cameras and the addition-

ally connected control units. Originally, such systems were mainly used in professional productions. A modern-day professional system is the CableCam by CamPilots for instance. While cable cameras sometimes also come with platforms that also allow carrying human operators, mostly they are remotely controlled, and the systems only carry the cameras. For their control dedicated hardware controllers are an often found solution in professional systems. Over time also amateurs or enthusiasts systems were released that were cheaper, smaller in size and controlled via mobile devices. These enthusiast systems are designed to move DSLR or mirrorless cameras, and an example can be found in the Syrp Genie [379]. Also, variations of the basic design have been proposed. In the domain of sports or event broadcasting the SpiderCam system [373] is found. It is a variation of the traditional cable camera concept as it is attached to multiple cables, often two / four, instead of only one. They are mounted at opposing points allowing it to manoeuvre faster and more freely above of the playfield.

Cranes (Boom / Jib)

In its most basic form, cranes consist of a jib arm that rotates around a boom point. The boom point often is in the centre of the jib arm, but designs can vary. This construction allows exploiting physical properties that lead to mechanical advantage. This advantage, in consequence, enables to move heavy payloads with less effort. It also enables to elevate the payload. Depending on the design the payload can move vertically and horizontally in several degrees of freedom. In cinematography, cranes allow a vaster range of motion in vertical shots and also provide increased flexibility when mounted on a moving base. Crucial in their design is that the camera mount located on the end of the jib arm is flexible enough such that one can keep the framing of an object (for example on ground level) while the boom is moved upward. Crane designs, similar to other tools, vary mostly in regards to size and whether one or multiple human operators can be positioned on the platform at the end of the crane. In well-engineered systems due to the placement of the weights and counterweights, a single crane operator can manoeuvre a camera team and professional camera equipment at a reasonable effort manually. In remote controlled systems the platform for human operators is often omitted resulting in smaller and more lightweight systems. Regarding the timing and the potential framing options, the camera operator in both types of systems is highly dependent on the performance of the crane operator. Also, advanced modern systems can be found in various designs. The Technocrane [378], for example, features a telescopic arm and wheeled platform similar to the base of a dolly (see next paragraph). Consisting of telescoping arm, remote head, adjustable centre pillar and wheeled base it is a versatile machine for use in studios and at outdoor locations. A different approach has been taken by the Stype kit system [377]. It is a retrofit device that can be attached to existing purely mechanical cranes. It allows to steer the head remotely and use CV to delegate the tracking of objects.

Dollies

While dollies have a similar purpose as cranes, they make use of a different underlying design. They also use a base platform with multiple, usually four, wheels. The designs vary, for example, in high-quality systems the steering behaviour of the wheels can be changed between crab and steer. While some use hydraulic jib arms (with boom points at the end of the arms) to elevate the camera, the column dollies, rely on a telescoping centre pillar that is electronically driven to adjust its vertical position. The electronics allow to change the characteristics of the motion, so, for instance, the acceleration or deceleration[8] can be finely tuned. Also, similar to a crane, a dolly is controlled by multiple operators with dedicated tasks. The grip steers the dolly and the camera operator on the dolly the camera motion. Remote-controlled dollies are rare, so most are still operated manually. To allow the recording of smooth shots dollies can also be set up to run on tracks. This way, also challenging terrain can be compensated for or (half-)circles can be precisely followed. An example of a modern-day dolly with a central pillar is the Classic Plus by Panther [364]. It was one of the first with an electronically driven centre pillar, and meanwhile, also a wireless remote controller can be used for its control. The Hustler IV by Champman [319] is an example of a hydraulic jib arm system. One of the few examples of a remotely controlled dolly that is also capable of saving and repeating motions is the Arri Motion system [311]. It is an extension to a Panther column dolly that can control the dolly as well as an attached remote head and lens controls. For non-professionals, also smaller devices exist. These can range from simple wooden platforms with attached skate wheels for mobile devices to steerable platforms with tripods attached that also carry DSLR cameras.

Industrial Robots

Robots initially intended for the automating of technical tasks have been adapted for use as cinematic motion control tools. They are attractive to directors and operators as they allow to move the camera very accurately. The blueprint of their design at the moment does not change much as they are mainly based on six-axes robots. This design allows moving the camera at a minimum of three degrees of freedom each in rotational and also in translational movements. Extended systems can also provide additional degrees of freedom. The precision that robot systems offer is often applied in advertisement recordings. Further, the systems also can track their position and movement and hence allow to easily fit its physical camera motion to the motion of a virtual camera when working with CGI. This automatic fitting of camera positions makes the creation of scenes incorporating CGI faster and requires less effort in post-production. The precision of these systems is also exploited in high-speed recordings. Here human operators face a natural limitation. While the design of the six-arm robots is universal across various implementations, they can differ in size. While larger systems are often used as stand-alone devices, smaller sized arms, in particular, can be used as an extension to existing dollies. The Cmocos system [321] is an example of this combination that allows integrating this novel technology into existing tools and practices. Similar to the Arri Motion systems, it is a motion control system that is mounted on a Panther column dolly allowing to carry out, record, track, save and repeat camera motion. In comparison to the Arri Motion system, however, it offers more degrees of freedom.

[8] Also referred to as ramping or easing in and out.

Unmanned Aerial Vehicles (Drones)

Amateurs, enthusiasts and professionals alike use Unmanned Aerial Vehicles (UAV). Due to their small form factor and high manoeuvrability, they can support recording in studios as well as outdoors. While the quadrocopter design is found most commonly, there is also a less frequently used design, the hexacopter, which is mostly used for carrying the more massive professional-grade equipment. Usually, there is no platform for human operators on the drone, so they are nearly always remotely controlled. Compared to earlier tools such as RC helicopters, they can more easily be controlled by untrained users. However, they would still need some time to master precise control. Theoretically, drones can carry out nearly all types of shots that sliders, cranes, dollies or robots offer, but in practice, they are most often used for aerial cinematography outdoors. Some reasons might be that the motion paths are not as smooth or stable as with a system running on tracks, they produce noises that negatively affect the audio recordings and that they are, at the moment, not capable of high-speed recordings. Similar to other novel tools, drones support a variety of traditional shot types but also established an increased use of a characteristic shot type with the top-shot. While the top-shot was already possible in the early days of aerial cinematography, its use remained limited. Its application as a stylistic device, however, increased with the more widespread use of drones.

Stabilisers (Gimbals)

Camera stabilisation can be achieved mechanically, as with a Steadicam system, or electronically steered, as with modern gimbals. For both types of systems, a Cardan suspension is the central element as it allows to isolate the operator’s movement from the cameras. In Steadicam systems, for example, operators wear a harness. An iso-elastic arm is attached to the vest on one end and on the other end to the so-called sled. This is a post that holds a gimbal, the camera, which determines the weight to be balanced and a counterweight plus additional devices. When the systems and the weights are properly adjusted, slight manipulations to the sled can already cause wide pans for instance, and even with the operator’s hands completely removed the camera would stay in place. The systems enabled to set the camera free in a controlled and smooth way but also face some limitations. As it is mounted to one particular operator, it cannot be handed over to another operator which can hinder the recording of long sequences. Due to its working principle, it results in a weight that requires substantial effort and physical training from an operator carrying it. In contrast, the modern systems based on electronic stabilisation are more lightweight, and further do not require to be attached to one operator. While mechanical systems require a distinct handling practice and afford some hands to be used for specific tasks, electronic systems are more flexible and allow to switch between one-handed and two-handed use or also vertical use. Their electronic components also offer wireless remote control. Therefore, two operators can share the workload, or the systems can be attached to other MoCo devices such as UAVs. The Ronin by DJI is an example of such systems capable of carrying DSLR cameras but also professional-grade cameras such as the RED Epic.

Further Designs

Exploring the possibilities of remote-controlled MoCo tools, various attempts in engineering the right design have been undertaken. In this process, various technologies stemming from other domains became adopted. Over time, some were used more broadly, like the ones presented above, and others remained niche devices. To touch also less frequently used MoCo systems, we want to mention some of the explored designs. For instance, the HoverCam system [453] was intended to provide smooth low-level shots independent of the nature of the underground surfaces which could vary from water to solid even and uneven ground. The system combines a small RC hovercraft and gimbal and can carry a payload of 5 kg at a speed of 50 kilometre (km) per hour. The system was presented by CampPilots, but its development was discontinued in the meantime. CamPilots focussed more on further exploring other designs as with their TurtleCam [318] for instance. This system is a combination of a medium-size RC car with all-terrain wheels and gimbal. The system is capable of achieving a maximum speed of up to 90 km per hours. Further, they also offer hands-free Segway recordings. Here, a Segway is be operated hands-free by a camera operator [422]. With his hands, the operator then controls a camera, that is stabilised by a Steadicam or other gimbal system while simultaneously steering the Segway.

2.2.2 Advancing Factors

While the advancing factors applying to cameras (form, quality, costs and connectivity) certainly were also important for the progress of MoCo tools, we want to emphasise more domain-specific and cinematography related factors, in this section. We see them mainly in the areas of the exploration of camera motion as a stylistic device, the technological advances in microelectronics, the integration with CGI and CV and the user-centred and cinematography-specific design of UIs.

Motion as Stylistic Device

As pointed out by Nielsen [194], camera movement does not serve one singular purpose at a time but rather serves multiple at once (see Section 3.1). These cannot necessarily be separated from one another. Most likely, one's role or perspective determines the motivation and focus. So, for example, directors or writers need camera movement to support the storytelling and to advance the storyline. Thus, they apply it out of narrative motivation. In contrast, for operators, camera movement is their craft. So they more likely tend towards identifying challenging shots and strive to overcome challenges to advance their craft and skill. In this process, their work is more focused on the visual aspects of camera work. Overall, they often follow a compositional motivation. Further motivations by different additional stakeholders such as the producers are also reasonable to assume. Identifying the motivation helps in understanding why camera motion is used. However, to better grasp how it is used on a larger scale and how its use evolved requires more and also more objective information. To identify common patterns in cinematography regarding camera motion, therefore also statistical data collection and analysis has been conducted as by Salt, Bacher or Crisp. Their applied study methodologies are often based on selecting a set of movies from different periods and classifying shots with camera motion into classes[9] and counting the number per 500 shots. While these studies enable the identification of trends, one needs to be aware that in examining the same film they come to different distributions regarding the classes and different absolute numbers for the number of shots as observable in their analyses of *Caught* (1949). This variation can be explained by the challenges of post-hoc analyses. When only examining the on-screen effects of camera motion, it sometimes becomes hard to precisely identify a combination of camera travel and pans or tilts as found in interviews with camera operators also conducted by Nielsen. While the classification and numbers vary still a general trend which identifies an increased use of camera movement can be observed. In conclusion, one can say that the data collected suggests that cinematographers use camera motion more often, which can be explained by the various motives of the stakeholders mentioned above and an interest in advancing the craft. The increased application also fuelled the advance of enabling technologies. As pointed out in an interview with Alfonso Cuarón, the director of *Gravity* (2013), he waited several years to shoot the film as the development of an advanced industrial robot system was a necessary prerequisite [322]. Vice versa, the emergence of new technologies motivated exploring the newly available shot types.

Advances in Microelectronics

The upcoming of microelectronics was a major technical milestone as also sketched out in Section 2.2.1. The electronic components were initially used for the pure operation of the motors that actuate the camera in standalone systems. By now, microelectronics also integrate further functionalities mainly regarding connectivity. For MoCo devices, some of the most influential features were the introduction of remote control interfaces but also the tracking of the position and the parameters of the camera. Electronics capable of wireless communication enabled remote control and allowed the use of various controllers suited for different contexts. Due to the wireless communication, also multiple dedicated controllers for subtasks could be connected, and hence the workload shared among operators. Systems could also be synchronised or motion control executed autonomously. Being able to track the position of the camera and its parameters further helped to bridge the gap into virtual environments more easily. The motion path of a camera could be recorded and mapped on to a virtual camera which provided a first link between physical camera work and CGI.

Integration with Computer-Generated Images

The gathered metadata became increasingly important for digital special effects. Special effects were already used when only physical cameras and film was available. For instance, the bluescreen-technique allowed to record an object in front of a blue-screen which then could be altered to any desired background in post-processing. This process needed to be conducted manually for each recor-

[9] Classes: pans, tilts, pans with tilt, tracking shots, tracking shots with pan, tracking shots with pan and tilt and crane movements.

ded frame. Today's green-screen technique follows the same basic principle but is conducted digitally. The transition to digital enabled the use of 3D generated graphics as backgrounds. However, when the real camera is moved, the virtual camera needs to be moved accordingly. This mapping needs to be done to account for the change in the perspective and the field of view. While this process can be done manually as in the early days of CGI, it is a time-consuming process. Advanced MoCo systems which track the necessary metadata can forward it to the virtual system so that this process can be automated. This metadata can not only be used to change the background and adjust the field of view but also in combination with Motion Caputre (MoCap) data. To animate characters more realistically often markers are attached to actors. These can be used to track their skeleton or particular muscles when an actor moves. The collected data can then be used to animate virtual avatars. This can even include inferring the motion of soft tissue [164] or how (virtual) clothing is affected by a moving body [215]. ▶
To generate a fitting render, in the end, the 3D environment, the changing camera position and the ▶
tracked actors need all to be combined and continuously updated.

Integration with Computer-Vision

With increased processing power implemented in MoCo devices also features beyond remote control and integration with CGI became possible. The computational power allowed to analyse the camera stream sometimes also in real-time. Advanced CV algorithms emerged and enabled the identification and tracking of objects. In the domain of cinematography, this data helped to define triggers which start a recording or camera motion and enabled assistive and autonomous systems. CV capable systems can manoeuvre on their own and follow a person or object similar to the work of a human operator. In their steering, they can also integrate cinematographic aspects. While surveillance systems for instance also can exploit CV to follow someone, they tend to frame the person of interest right at the centre of the image. Cinematic framing techniques differ from this approach for aesthetic or narrative reasons. Applying these techniques also let the results appear less robotic which is often favoured by producers and consumers. CV can further be used as a safety measure. In particular, industrial robots can rapidly exert great force. If an object or a human bystander stands in a motion path when the motion is carried out, they might be severely damaged or injured. Assistive features analysing the scene and surrounding beforehand can help to minimise casualties on set. Such safety precautions are not limited to industrial robots but also implemented in other systems such as drones.

User-Interface Design

The design of UIs as well could bring benefits to the progress of MoCo tools. For expert systems, especially industrial robots, new ways of controlling the systems were necessary. In an interview with the developers of the Spike system, the Lead Developer Christian Fritz mentioned the importance of software that could be handled cinematically [535]. Traditionally industrial robots were controlled ▶

and programmed by teach pendant devices[10]. As these require programming skills and often work waypoint and coordinated based, they are difficult to operate in a cinematographic context. Also for less high-end systems, the upcoming of new types of UIs made interaction easier. For enthusiasts, UIs on mobile devices enabled a large audience. As they consistently adapted to principles users were accustomed to they provided a low entrance barrier. Mobile devices also provide displays with enough resolution to present the camera stream. This feature on the one hand help in the steering of the system and on the other hand allowed novices to work with CV visually.

2.2.3 User Interfaces

MoCo UIs are a specific subdomain of HCI particular to cinematography. We, therefore, want to introduce them in more detail. The design of modern UIs is influenced by a tradition of mechanic tools and their controllers as pointed out in Section 2.2.1. Based on the taxonomy presented in Section 2.2.1, we provide examples for UIs of each class and describe the underlying control concepts.

Remote Head User Interfaces

For remote heads, we identified basic components that afford user control. Those are handwheels, foot pedals, buttons, knobs, joysticks and touchscreens. In designing reasonable interactions with these UI components, also functions need to be mapped. For remote head control, these functions are most often pan, tilt and roll. As explained earlier often also remote lens control devices are attached to remote heads. For these devices, the mapped functions mostly are controlling focus, iris and zoom. For some systems, additional features are implemented and controlled via an additional dedicated element. These could offer, for instance, to record or replay a camera move, to stop the motion, or to change the framerate of the camera. The underlying concepts of control are slightly varying. For the control of the camera orientation via joystick or handwheels, continuous control models are mostly used. Sometimes the operator also can change the used transfer function that maps one's input motion onto the output motion. By choosing different functions, the steering can be more sensitive or more robust. For lens control, also continuous control via rotary dials (focusing) or parametric control (aperture) is often used. The mentioned basic UI components are often integrated into one control panel. This panel can have various forms ranging from a tabletop console, hand grips or bars to a so-called panbar that mimics the physical aspects of camera operation on a tripod the most [516].

Slider User Interfaces

For sliders also physical components are often found UI solutions. Similar to remote heads, buttons, knobs and joysticks are used for continuous control. Some systems also use different concepts such as parametric control. For instance, the Edelkrone Motion Kit uses physical movement for input and

[10] Sometimes also referred to as programming pad or teach(ing) pad.

a mobile device for feedback and control [533]. The system offers full manual control or assisted parametric motion control. Here, a user can determine a parameter as speed for instance instead of manually steering it. Also, further approaches are taken in assisted and motor-driven slider systems that provide virtual control environments. The Kessler systems, for example, can be controlled via a dedicated Operating System (OS) [491]. Alternatively, Kessler offers control via UIs that run on tablets that afford different UIs. Virtual environments allow to switch between continuous and parametric control more easily and also offer more refined parametric control options such as the definition and manipulation of ramping curves by displaying the curve and allowing users to manipulate it.

Cable Camera User Interfaces

For expert cable camera systems such as the SpiderCam, a combination of remote head and slider controls are used. The system is designed to be used by at least two operators with one being responsible for the positioning the camera and one responsible for controlling its orientation [555, 523]. The operators are provided with a physical panel that offers physical buttons, knobs and joysticks for continuous control. In contrast, different UI designs are used for enthusiast users using smaller and more portable systems such as the Syrp Genie [527]. Here, a combination of physical buttons and a display supporting parametric control is provided, sometimes also applications on mobile devices can be used to input the parameters.

Crane User Interfaces

Similar to the previous classes, the components used in crane control are also mainly physical. For control over the column or a telescopic arm mainly buttons or rocker switches are used for continuous control. Extensions as the Stype kit can introduce different UIs such as joysticks and a panel of buttons that can be used for discreet or parametric control [519]. Mobile devices for the control of cranes are hardly found at the moment.

Dollie User Interfaces

The fundamental components of dolly control a very similar to previously mentioned systems. Due to the variation of the basic engineering of the systems, dollies show some peculiarities. If electronically driven a combination of buttons and rocker switches is often used for continuous control [507]. For hydraulically engineered systems, however, the main controller is a valve. Here, by controlling the pressure, the vertical position of the dolly platform can be changed.

Industrial Robot User Interfaces

For industrial robots, the control concepts are primarily determined by the programming language. As for the use of teach-pads programming is a necessary skill. However, regarding cinematography higher-level control concepts have evolved additionally. Although this reduces complexity, nonetheless their operation requires skill and training. For systems such as Iris or Spike, software for their control has been developed on top of 3D applications [535]. For their development often the Motion Builder Plugin for Maya serves as a framework. They incorporated various forms of control concepts regarding 3D-interaction, timeline control and keyframing. While they are primarily designed for a classic keyboard and mouse desktop settings, they can also offer different interaction modalities such as pen-based interaction.

Drone User Interfaces

Similar to the presented tools, also for drones the most often found solutions are hardware controllers featuring joysticks and buttons for continuous remote control. To display the camera stream and to provide additional features they are sometimes used in combination with virtual UIs on mobile devices. In this case, the physical controllers afford holders for a mobile device to be integrated into one control panel. Alternatively, the mobile devices sometimes also emulate the physical controllers provided by the controllers. Here, mostly software-joysticks and virtual buttons and knobs are utilized instead of physical ones following the same paradigm of continuous control. But in virtual environments also different paradigms were proposed and used. Integrating multi-touch displays that display the camera-stream, virtual UIs and CV capabilities allowed to transition to various forms of new control paradigms [462], one of which is content-based control. In these designs, the operators interact with the system on a higher level of abstraction and can take over manually if needed [544]. Also, parametric control concepts are implemented as in autonomous systems such as the Percepto extension [508]. Here, parameters for safety distances, geo-fencing, selected shot types and others can be determined ahead and determine the motions that the systems carry out on their own.

Gimbal User Interfaces

For electronically controlled gimbals the UI options are limited as the gimbal itself is often carried by one operator using both hands. To change between different modes or to adjust the direction of the camera mostly a combination of joystick and button that is supposed to be operated by the thumb of the operator. Pushing and pressing are the main ways of interaction, and audio signals are used to provide feedback [460]. More options can be accessed if a remote controller for a second operator is used. These devices can come in different forms ranging from physical joystick and button RCs to kinetic controllers such as the Link by ACR [306] which resembles the concepts presented in the Sun Starfire video prototype [261] for instance.

2.3 Summary

Cameras and motion support systems have a rich tradition of analogue systems preceding today's developments as presented in the timelines in Section 2.1.1 and Section 2.2.1. In conclusion, one could say that cinematography transitioned into its digital era in all of its central domains such as production, distribution and consumption. In retrospection, the propositions that Samuel Goldway put forward in his article *Hollywood in the Television Age* (1949) [101] became (more or less) a reality. In this early essay he suggested that to be able to compete with TV, the film industry had three options. First, to have their own TV stations, second, to deliver first-run movies to homes via telephone wires (in an early version of pay-TV) and third to develop large-screen theatre ”TVs” so that one production can be transmitted by leased wires to thousands of theatres at the same time. In 1949, none of these options was pursued, but nowadays, the options two and three became a technical reality. The transition to digital also included a change in production tools. This change incorporated digital cameras and digital theatre projections on a professional level. But also on enthusiasts level when considered with cameras, self-distribution platforms and post-production software. Nowadays, for all levels of users hardware and software is available that can enable novices and assist experts. The availability of enabling technology can be considered an essential catalyst for aspiring film-makers as echoed in the opening quote of Steven Spielberg reflecting on his early days ahead of this chapter. More recently, this change also affected MoCo technology and again enabled and supported various user groups. For their control, on the one hand, the UIs are consistent with their traditional origins, on the other hand, the potential of today's computational capabilities are explored but not fully exploited. This further exploration of the newly incorporated computation power still requires the development and evaluation of new types of user interfaces.

Insights

- Cinematography has a long tradition and the underlying technologies now matured in such a way that it entered its digital era on a larger scale; covering not only camera systems but more recently also motion control tools.
- Alongside this shift, however, also the traditional notion of distinct user groups became more and more challenged. Various forms of amateur, enthusiast, semi-professional and professional groups resulted as a consequence. Further, traditionally trained camera professionals might not necessarily be experts for newly emerging motion control tools.
- Also, in-the-wild, the potential of the implemented computing power is not yet fully exploited. Especially regarding the control of complex camera work on a higher level of abstraction. Well designed user interfaces might further enable users of all groups to access technologies such as computer vision and can hence assist them without requiring years of practice regarding their control.

MARSHALL MCLUHAN (SOCIOLOGIST)

A point of view can be a dangerous luxury when substituted for insight and understanding.

THOMAS CARLYLE (PHILOSOPHER)

Nothing is more terrible than activity without insight.

EDWARD T. HALL (ANTHROPOLOGIST)

The information is in the people, not in your head.

3

Users: Insights from Cinematographers

What to expect?

- Presentation of user roles and tool application
- Observations from seminars and contextual inquiries
- Findings from expert user interviews

What to take away?

- A further understanding of camera work in practice together with some lessons we learned regarding the various methodologies

Attribution: This chapter references research that we previously published at INTERACT ’17 [122] and CHI ’15 [124].

Our Statement of Collaboration details the differences between the paper(s) and this chapter

3.1 Literature Review

As mentioned in Chapter 2, carrying out camera motion is not trivial. Consequently, managing the inherent complexity led to the emergence of different user roles and responsibilities. Also, supporting the act of carrying out particular camera motions led to specific tools. Thus, the developed tools and technologies were also shaped to meet the user's needs. Further, their application is not arbitrary and can serve various functions in nuanced ways. Wanting to know more about the details of these aspects led to a first primary question: *What can we learn from literature regarding users, their behaviour and reasoning?* Below, we will present details on these topics that are intended to help answer the question and provide findings from selected literature in greater detail.

3.1.1 User Roles and Responsibilities

The working environment on a film set is hierarchically organised regarding the authority over decisions. On the top, the power is centralised at the position of the director. The tasks that are associated with the job entail managing all aspects of film production including the artistic vision, technical realisation and crew organisation. The director is in charge of picking the cast and crew members and guides them through all phases of the shooting. Therefore, the director is also administrating the camera department. To also share the workload, some freedom regarding the decision-making is also granted to the heads of the departments on the next lower level in the hierarchy. That also applies to the head of the camera department who is also often referred to as the Director of Photography (DOP). The DOP can then further delegated subtasks to the subordinate team(s). As the various subtasks might require specialised knowledge, training or experience, consequently distinct user-roles and working profiles within the camera department have emerged. Below we list the different user roles in the camera department of a film set hierarchy and provide details (based on [317]).

Director of Photography

The DOP sometimes also referred to as Cinematographer (as by the American Society of Cinematographers [312]) is, in accordance with the director, responsible for the visual design aspects of a film and coordinating the subordinate crew(s). In particular, this entails the artistic and aesthetic vision as well as its technical realisation. The DOP determines and checks the aesthetic and technical parameters of the recordings incorporating composition, movement and lighting. Given the creative nature of the medium and necessity for teamwork, DOPs do not work isolated from the other departments. In consequence, discourse with the director and other departments regarding details and overlaps between the departments is typical. Despite the hierarchical structure in general, this creative discourse requires the DOP sometimes to convince the director to change a decision. How this necessary discourse is handled most often depends on the persons involved and is bound by the framework that the script provides.

The DOP is engaged in the preparation, production and post-production stages of the recording. In the preparation phase, this can incorporate reading the script and associated material, talking to the involved departments (directing, production, props, etc.), scouting locations, laying out the recording workload, determining the technical staff and sometimes already taking test shots. During production, this shifts towards deciding the structure of the shot sequences, types, angles and movements. That again is planned out together with the director. For each of the resulting shots, the parameters are set and checked. That includes the camera position, the angle of view, the composition, the movement of the camera, the lens and aperture setting. Additionally, the DOP is responsible for the lighting of the scenes. The decisions are made together with the chief electrician who will consequently set up the stage as discussed. The DOP is also involved in picking the takes (if multiple were shot), recording with the special effects in mind, supervising the technical aspects of the workflow (often together with the Digital Image Technician (DIT)), making sure the production stays within the financial budget and teaching and empowering the crewmembers. In the post-production phase, the DOP supervises the special effects, the colour grading and the final cut.

Camera Operator

Once the director and the DOP came to a consensus regarding the shot sequences, angles and movements. The details are discussed with the camera operators, as they are responsible for setting up the equipment and the recording. So, for operators, it is essential to know how to take the given details and set up the equipment accordingly. This entails knowledge on cutting techniques, camera and acting axes, a basic understanding of dramaturgy. Further, it is also central that they know their gear so that they can operate it correctly. That includes knowledge of the various film or sensor formats and the camera lenses. They have to keep an eye on the stage to make sure that lighting does not irridate, or that the shadows of the camera are visible in the shot. They further have to coordinate the choreography between actors and camera motion, manage that all sequences are shot and coordinate with the other departments. Most importantly they also check all the settings before a scene is recorded and rate the take after it finished.

Camera Motion Operator

The prerequisites for being an operator for camera motion tools are similar to the ones for being a camera operator. Frequently, the operation of such devices as mentioned in Section 2.2, requires specific knowledge and extensive training. Thus, some camera operators specialise in their operation, practice for years regularly and take workshops to advance their skill. In contrast to a camera operator who mainly is booked for a whole production, a camera motion control operator is scheduled for a limited number of days. Therefore, the motion operator needs to adapt to the situation and the routines on set quickly. As the operator is the foremost expert regarding the expressiveness of the tool possibilities and limitations need to be discussed with the director and the DOP.

First Assistant

The primary task is of the first assistant is to adjust the focus and hence the position is sometimes also referred to as focus puller. In particular, this includes setting the focus, following, relocating and maintaining associated technical parameters. The position requires the first assistant to be knowledgeable about the principles of photography, optics, filters, films and sensors as well as recording and transmission techniques. Further, the focus puller needs to handle cameras and accessories, handle timecodes and check if light irridates. In the preparation phase, the focus puller needs to read the script and discuss the individual scenes with the DOP. They subsequently need to derive a list of the required technical devices and define the requirements of recording and the workflow. The later is often done together with the DIT. The focus puller also makes sure that all devices are on set and that they work correctly ahead of a recording. During production, the assistant is responsible for the setup and the relocation of the devices for each scene which is often handled together with the second assistant. Technical setup refers to the mounting of lenses, filters, inserting recording media, functionality checks and lens cleaning. Additionally, this covers the setting of the camera[1] and lens parameters, the connecting of additional devices (for example for the transmission of the camera stream). Of course, carrying out the primary task of adjusting the focus correctly is paramount during a recording. In the post-production phase responsibilities are shared with the second assistant, the DIT and the data assistant, as the foremost task is to ensure that the record data is complete and that technical devices are cleaned and eventually handed back to the equipment rental.

Second Assistant

The main objective of the second assistant is administering and organising the technical devices. Additionally, assisting the first assistant during the production phases is essential. Hence, second assistants need to be knowledgeable in the same areas as the first assistant. In the preparation phase, second assistants manage the details regarding the logistics of the technical devices and acquire expendable materials among other aspects. During production, they assist in preparing the equipment and check with the first assistant as well as the DIT. They support in connecting additional devices, measuring distances for pulling the focus correctly and set markers if necessary. Also, they handle the clapperboard, talk to the continuity department, handle the recorded media and help with the transportation of the equipment. After a recording, they also assist in the first assistant and the DIT in wrapping up the material and the return of rental devices.

Digital Image Technician

With the introduction of digital recording techniques, also new technology-specific positions were established. One of these jobs is the DIT. Fundamentally, a DIT is involved in all production phases and is responsible for quality assurance and safe data management of the digitally recorded data. During

[1] Interestingly from a UI perspective, this also includes to set customizable buttons to the preferences of the operator.

the preparation phase, a DIT discusses technical options with the DOP, plans the digital workflow and therefore also talks to the post-production staff ahead of the shooting. That, in particular, might entail test shots with post-production effects applied, so that the shots are set up right for the post-production steps. Together with the assistants, technical devices are selected and parameters discussed. In the production, the DOP and the DIT collectively decide on the visual characteristics. The DIT checks the material for visible fuzz and makes sure it correctly exposed to be used for a given post-production technique such as chroma-keying. Sometimes the DIT already applies techniques that alter the visual appearance, such as colour grading, on set (in accordance with the DOP). Similar to the other assistants, the DIT is also responsible for the setting up the equipment for each scene and to relocate it to the next. During the post-production period, the DIT checks the material, if needed creates copies and hands the data over to the post-production staff.

Data Assistant

A data assistant is primarily taking charge of the tasks concerned with saving, copying and verifying the copied data. That is often necessary as digital storage media for cameras can be reused once a (full and accurate) copy has been made. Correct and verified copying and storage of the data requires careful effort and time. The proper storage is also often made mandatory by insurances of cinematic productions. Given this special subset of tasks, the data assistant mainly supports the first assistant and the DIT who sometimes also take on responsibilities regarding data storage and safety. The storage procedure entails making multiple copies from one source on multiple separate storages and verifying the copying process via checksums. This can already be carried out on set. The data assistant additionally makes sure that the data is handed over to the post-production staff. Overall, the data assistant needs to be knowledgeable about contemporary (digital) camera systems and their storage media, computers and peripheral devices, specific software for the (safe) copying process, the various media formats, timecodes and post-production processes. In the preparation phase, the data assistant suggests suitable hardware and software, agrees with the first assistant and the DIT on the workflow and helps with test recordings. During the production, the reliabilities are setting up the technical equipment, making safe copies, testing its integrity and documenting the copies, in addition to assisting the first assistant and the DIT. After the shooting, the data assistant might be asked to create more copies and hand over the copies to the post-production crew.

Colourist

On some sets also a colourist is found. If present, the particular task of the colourist is to take care of the colour grading (in accordance with the DOP and the DIT). The grading should suit the mood of a particular scene and create a visual consistency across similar scenes. The grading can be part of the preparation, as in test recording, the actual shootings and after the shooting. Consequently, knowledge of the used software is necessary and the grading process needs to fit into the existing workflow.

3.1.2 Application of Motion Support Tools

As introduced in Section 2.2, different tools are used to create specific types of shots. While, in theory, tools could be operated in enumerable ways, in practice, some types have emerged that consistently and successfully supported narrative and compositional functions. To answer the question *How are tools predominantly used?*, we provide an overview of such a set of often used shots below. In particular, we cover, which tools might be applied and why they are used (based on [177]).

Pan Shot

In a panning motion, the camera is swivelled horizontally. Usually, it is mounted stationary, so for executing a pan often a tripod or remote head is used, but also handheld operation is possible. The rotation revolves around the vertical axis, and the on-screen effects of this type of movement imitate the visual experience of a person turning their head. This type of shot can be used as a narrative device as details of the scene are presented sequentially, to follow a particular person or object, to introduce the location or to shift the focus from a specific person or object to another [506]. In contrast to individual shots and cuts in between, the integrity of real-time and space or an essential acting scene remains preserved in such as continuous motion. For instance, panning back and forth could be used to depict two people arguing and increase with the growing intensity of the argument.

Tilt Shot

The tilt shot is similar to the pan shot. It differs from a pan as the direction of the movement is vertical. Thus, the camera pivots around the horizontal axis. Consequently, it imitates the visual experience of someone looking upwards or respectively downwards. Also for the tilt, a tripod or remote head or handheld operation is used. Similar motives as for the pan apply to this technique, however establishing a scene or a character is more likely the motive then following a person [506]. Again, in contrast to cutting, a tilt maintains the integrity of time and space in a scene but also puts more emphasis on it than stationary shots. Consequently, it should be motivated by the story.

Dolly Shot

The on-screen effects of dolly shots result from putting the camera on a wheeled platform and pushing the platform evenly while the focal length is not altered. When pushing the camera towards a scene, theoretically also a zoom could be used alternatively. However, the on-screen effects of a dolly-in resemble the human visual experience of walking towards a subject to a more substantial degree than zooming. The motivation for dollying-in or out can be manifold; often it is applied to reveal or conceal particular aspects of the scene or as a comment [451]. Both movements can be used for dramatic effect. While moving in is more associated with tension, suspense or intensity, moving out is more associated with increasing emotional distance, loneliness or despair after a tragic event.

Dolly Zoom Shot

For representing a moment with great dramatic effect also a particular sub-form of the dolly shot is available. It is referred to as a dolly zoom or sometimes also the vertigo effect. As suggested by the name, it is a combination of dolly movement and change in focal length. Two types of combinations are possible. When the dolly moves in, the lens zooms out or vice versa when the dolly moves out the lens zooms in. The goal of each of the combinations is that a person or object within the frame of the camera has the same size in the frame despite the changes in the optical parameters. So, the dolly move and the zoom must be operated in synchronicity. Given that operators cannot perform the movement correctly in sync, the visual effect vanishes. Due to the great dramatic effect that this motion creates, it most often used to depict sudden and substantially revelations of a character, an intensely emotional episode of rage, anger, fear or obsession [530]. The faster the move is carried out ▸ the more emphasis it puts on the motive.

Tracking Shot

Given a wheeled platform, shots can not only be recorded by setting up the tracks in such a way that the camera moves in or out. Alternatively, the tracks and camera can be oriented so that the camera films the scene while running parallel to it. Tracking shots can be taken using a dolly [451], but are ▸ not limited to this device. Also sliders and handheld or stabilised systems are considerable options. Again, the motives for applying a tracking shot can be manifold. Often they are used for following a character and/or to establish a scene. In establishing a scene, they can also be combined with a zoom to further focus on a particular person or object. Similar to a pan, a tracking shot can also be used to take the focus of attention away from one subject and shift it towards another.

Stabilised Shot

Setting up systems such as tripods, sliders or dollies can at times not be feasible due to the environment of the scene or due to the intention of the shot. In such cases, handheld operation is often the preferred alternative. However, this operation style can, if not taken measures against it, also often result in shaky recordings. Stabiliser systems can dampen the shakiness that otherwise becomes visible in the recorded material. They can either be mounted onto the operator, such as the Steadicam or can be operated handheld as in the case of gimbals. Applying stabilizers allows for recording shots that have similar on-screen effects as pan, tilt, tracking or dolly shots and further, depending on the system, enable 360° shots around a character [525]. Additionally, the fluidity of the motion is fairly unique to ▸ stabilized systems. One of the strong motives for using a stabilised shot is to take the performance of a single character in one take while the actor can move freely. This can help to build tension or put an emphasis on the idea that the particular scene is significant.

Crane Shot

As mentioned in Section 2.2, cranes allow for horizontal and vertical movements or a combination of both [452]. Due to their construction, cranes afford vertical shots predominantly. The upward movements generally support the slow revealing of a location. That might especially be interesting if the area is of grand scale (as often found in outdoor scenes). Another popular approach is to transition from a close up framing into a wider framing continuously. Naturally, this can be operated in reverse when transitioning from a broader framing into a closer one. While a crane can be used stand-alone, it is often used in combination with moving platforms such as dollies as this provides multiple degrees of freedom in moving the camera [430] increasing the expressiveness of the resulting movements. Besides being used to establish a location or coming close to a subject of interest, crane shots can be used as a narrative tool to reveal that a change in perspective with relation to the story.

Sequence Shot

Among the most elaborate forms of camera operation is the sequence shot. It is not limited to the use of a single tool or shot as often presented above. The sequence shot is mostly a long take covering continuous action, sometimes also by multiple actors [538]. The camera can be handed over between multiple operators or mounted onto several motion support devices during its recording. The sequence shot can also be planed to transition between different framings. It can be used in a motivated setting where it follows a particular actor but then switch to an autonomous camera style where the action is merely documented. Consequently, it can be used to convey an important story point, create a dramatic effect, build tension or focus on an actor's performance.

3.1.3 Narrative Functions

We mentioned that narrative functions are central to directors and operators in their decision making. However, given the techniques presented above, combinations could be made in uncountable ways. This led us to the question of *Given the various possibilities why is one movement chosen over another and what are the underlying top-level motives that lead to the decisions?* To further provide details on these questions, we present an overview of narrative functions as put forth in the taxonomy by Nielsen [194] below. While this taxonomy tries to foster understanding about the varying aspects of narrative functions, it needs to be mentioned, that camera movement most likely does not only serve a single purpose, they rather need to be understood as multi-layered or functions that *"multitask"* [194].

Orientation

Camera movement can be used to orient the viewer spatially in several regards. First, movement can increase the depth or volume. By the nature of the medium, film is a 2D representation of a 3D scene. Consequently, some of the information becomes omitted mainly regarding the depth of the scene or the volume of an object. However, moving the camera can account for that to some degree. When the camera is actuated, more depth cues, for example in the background, can be revealed to the viewer due to movement parallax. Thus, a viewer can more easily estimate the depth relationships and derive a sense of closeness or distance. Additionally, moving around an object or passing by an object, for instance during a dolly-in, can provide more visual cues on the volume of an object and hence promote this spatial aspect in the viewing experience. Second, the viewer can be oriented towards story information that is represented by a particular person or object. In this sense, the camera and movement can be understood as a pointer that directs the attention of the viewer. For instance, when a person is followed using a tracking shot or if an important character is singled out applying a dolly or crane shot. Third, camera motion can serve as a vehicle to describe the scope of the action. Emphasis is not put on shifting the viewer's attention, but rather to establish the spatial or social environment. For example, a crane shot can be used to depict the vastness of space as in the graveyard scene of *The Good, the Bad and The Ugly* (1966) [531] or the Steadicam shot in *Boogie Nights* (1997) [448] that introduces the quasi-family that the main character joins.

Pacing

Camera movement and along with it (dramatic) pauses between movements, can give a movie a particular rhythm. Further, this rhythm can also be paced differently at times. So, fast sequences of intense movement and frequent changes result in a perceived faster pace than long sequence shots that are slowly operated and scarcely changing. But not only the speed of the camera movement contributes to the pace but also how it relates to the scene and to which degree it does contribute to the transfer of information to the viewer. Simply put, the more information can be transferred to the viewer in a shorter period, the higher is its pace. Consequently, camera movement can be used to maintain an already existing pace, slow a given pace down or increase it. For example, the opening scene of *Magnolia* (1999) makes frequent use of dolly-ins to increase the pace of the sequence.

Inflection

While dolly-ins might set a particular pacing, for example, the very same movements can at the same time also carry an emotional quality. In this regard, motion, in general, can result in inflecting a commenting or valuative attitude towards what is depicted. So, camera motion cannot only be used to orient the viewer in the sense of a pointer and suggest where to look but also in the sense of a comment that suggests how to look (or how to read the scene). While commenting is one way that inflection can be motivated, connecting to the emotional state of an actor might be another. An actor who is an intense emotional state might be filmed with a shaky handheld technique. Here, how much the camera moves might be connected to the increasing level of emotional intensity the character experiences. Alternatively, shots can be taken using the Dutch tilt to indicate a change of perspective in a character as in the restaurant scene in *Mission Impossible* (1996) [509].

Focalisation

While for the function of inflection the camera movement can be symbolically anchored onto an actor, it does not yet take on the (first person) perspective of the character directly. Taking on the POV of a character can in its most direct form, result in shots that show the field of view of the character. This approach is referred to as focalisation[2] and an example of this can be found in the cemetery scene of *Vertigo* (1958) [545]. Beyond a pure optical POV also in this approach, the camera movement can additionally be attached to the emotional state of character (as mentioned above) and is then referred to as an affected POV. Sometimes, the camera is moved in a POV fashion, but the character through the eyes of which the audience can see has not been introduced in the movie before. This form of focalisation is also referred to as invoked presence. Generally, the shots are often operated hand-held.

Reflexive

For Nielsen, camera movement used for reflexive purposes is rather a collective term. In his understanding, the term refers to a motion that is not primarily serving a narrative function. Its purpose is rather to appeal to the audience or to foster engagement in parallel to the storyline. Here, camera motion can be seen as aesthetic visuals which can take on many forms. Most noticeably they can be characterised by showcasing a *"virtuosity of transport"* [194]. In the history of cinema, there are plenty of examples of such engaging camera movements. For example, in *Citizen Kane* (1941) the camera flies through the bar sign [455], in *Matrix* (1999) time seems to stop as the camera revolves around a character [498] or in the opening scene of *Lord of the Rings - Two Towers* (2002) [478] where the camera seems to fly through a mountain. Reflexive camera movement cannot only be found in the form of a single highly elaborate shot but also patterns of shots. In this sense, a particular motion can be used as a motif, for example for depicting a character, which reoccurs throughout the movie. Given multiple characters, there can also be a multitude of such motifs. Also, given a consistent use of a motif, they can create an expectation of sameness. After a motif is successfully established, this expectation can be broken, to create a moment of surprise (that might relate to the story).

Abstract

The motivations for camera movements presented so far were mostly connected to the story, the craft or the audience. Beyond these considerations, motion can also be used for representing an abstract idea. The more abstract the idea is, the more plentiful the possibilities for interpretation become and hence the harder it gets to give a precise model of explanation or reasoning. As Nielsen put it "*Abstract concepts and ideas are sometimes attributed to camera movements by critics overstretching their hermeneutic muscle but in other cases filmmakers have genuinely invested their movement with abstract functions.*" [194]. Thus, often an analysis of individual scenes of a movie are necessary, the interpretations might vary and the findings might be hard to generalise.

[2]The term was introduced to describe forms of narration in literature [99] and was later adapted to the field of film studies [194].

3.2 Online Survey

Based on the information provided by the literature we wanted to know more regarding the question *Do operators behave and reason in practice as the literature suggests?*. Therefore, we collected first-hand knowledge of expert users and the usage context by conducting an online survey. We wanted to investigate how experts use tools actually in their work, in contrast to what we conceptualised that they should do theoretically. Further, we wanted to know more, not only about device usage in general but more specifically also about motion control tools. Due to their recent emergence, the literature on MoCo tools and associated user aspects is still limited. We opted for an online survey as it can be a quick way to gather insight. We approached professionals that already displayed the use of MoCo tools in their portfolio so that we could assume that participants knew the domain.

3.2.1 Procedure and Results

To identify usage patterns in creative practice, we sampled twenty cinematographers that displayed their work via personal portfolios and blogs online and who were using MoCo devices in their productions. Due to our conjunction with an industrial partner manufacturing slider and dolly systems, we specifically focused on cinematographers using motor-driven sliders in this survey.

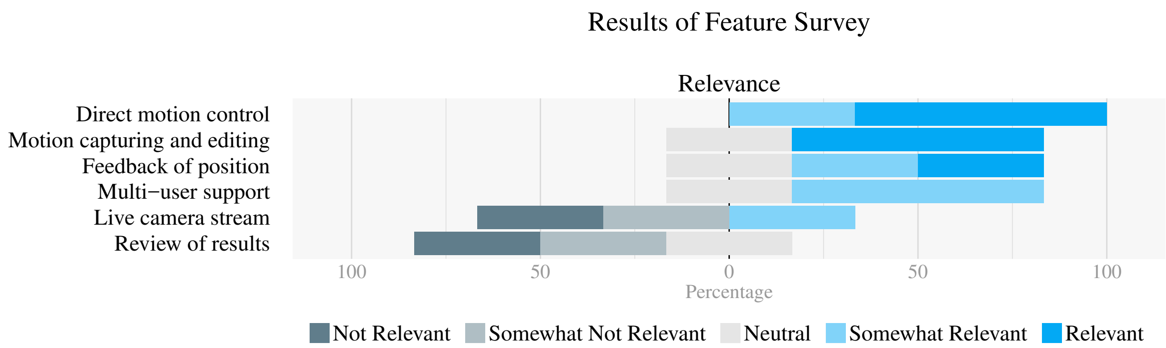


Figure 3.1: Users: Results of the survey and features of user interfaces for motion control

We approached them via the contact addresses they provided on their websites and asked them to participate in our survey. They were informed about the purpose of the survey which was to help to derive insights on how cinematographers use MoCo tools in the field and what relevance a set of six features was in their design. The feature set was derived from our analysis of status quo slider systems presented in Section 2.2. The features either were missing in status quo designs or newly emerging and hardly researched. We also provided a link to the survey so that the sampled cinematographers could conduct it at a time of their choice and asked them to share the link with further suitable cinematographers. The survey consisted of three open questions asking for user roles, tasks and design requirements. Open questions could be answered via text fields and were not limited in length. Six 5-scale Likert items followed asking on the relevance of specific features of MoCo UIs such as live camera stream, multi-user support and others. All questions and Likert items of the survey are provided with this thesis (Appendix 12.1). From the twenty survey requests only three were answered, and the questionnaire filled out. Therefore, the quantitative data collected via the Likert items are hard to interpret as personal preferences (positive and negative) of the participants cannot necessarily be generalised. The results of the Likert items are presented in Figure 3.1. Due to the low return rate, we focused more on the analysis of the qualitative statements. Here, we could identify mentions of single-user and multi-user scenarios. More precisely *"a mix of both - single-user and sometimes a collaborative"* as put by one participant. We further looked for explanations as to why one particular type of approach was preferred. From the answers, we could identify that the complexity of the scene and capabilities of the tools were additional factors determining whether one or multiple persons become involved or as another participant mentioned: *"a simpler controller system will often be controlled (only) by the camera operator"*.

In summary, user roles as described in the literature could also be found in our survey. Additionally, we found that operators work in single as well as multi-user environments depending on the context. Single user setups are mainly used in low budget or documentary productions such as images films or wild-life recordings where a full crew might be disturbing [499]. On the contrary, in larger productions, more complex shots are wanted, and hence the workload is shared among the crew. Therefore, the camera or MoCo operator is not the only one in authority of the camera movement. Consequently, close communication between the different departments and crew members is necessary [423]. Naturally, both approaches face advantages and disadvantages which are rendered more precisely in Table 3.2.1.

Dimension	Single User Environment	Multi User Environment
Roles taken on at once	Director, DOP, Operator	Operator
Preparation	Rough, On-Scene Decisions	Detailed planning and scheduling
Equipment	Mobile and lightweight	Heavy and stationary
Tasks	Multiple at once	Delegated to multiple people
Advantages	Free and flexible process	Sharing of the workload
Disadvantages	Multitasking creates high workload	Good communication and management necessary

Table 3.2.1: Users: Online Survey on user roles and environmental properties

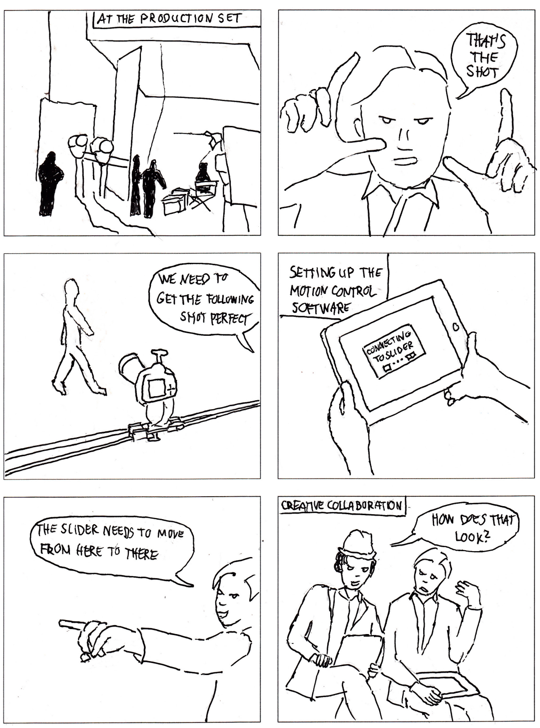
3.2.2 Lessons Learned

Despite the low return rate, we could find that MoCo tools are used in single as well as multi-user environments. While multi-user setups use human-human task delegation, in a single user setup one operator is responsible for carrying out multiple tasks at once. Therefore, further delegation to assistance systems seems a reasonable approach. As it is difficult to generalise from our results, we decided to use our findings as a foundation for storyboards of the two usage scenarios (Figure 3.2). The storyboards were used as a basis for interviews and brainstormings in later ideation phases.

We further figured that to get more profound feedback from experts we needed to refine our methodology and use additional approaches. So, to increase the return rate we derived the following propositions that we also integrated into the later research phases further presented in this thesis.

Expect low return rates and adapt...

- ... by providing a greater incentive
- ... by handing out more questionnaires including expert users groups on social media
- ... by shortening the questionnaires to adapt to busy schedules in one-to-one communication
- ... by observing and interviewing experts at work



(a) Single User Storyboard

(b) Multi-User Storyboard

Figure 3.2: Users: Storyboards for single user (Figure 3.2a) and multi-user scenarios (Figure 3.2b) from [288]

3.3 Film-School Seminars

Given the findings of the previous section (3.2), we took our own advice and set out to visit experts to gather more insight and a chance to observe some of their practices at work. Observing experts at work was not only supposed to deepen our understanding gained from the literature but also to investigate the question *When we observe people in practice, can we find evidence that they actually behave the way they suggest when asked (as in our survey)?*

3.3.1 Environments and Observations

We reached out to the camera operator and teacher Stephan Vorbrugg [425]. He invited us to join a seminar he gave on the theory of camera movement while he was teaching at Hochschule für Fernsehen und Film München (eng. University of Television and Film Munich) (HFF) Munich. At HFF, we additionally were invited to be part of a practical seminar on camera movement held by Leo Borchard. Here students and we were able to gather first-hand experiences in the operation of mechanical camera motion tools and other film production equipment.

Theoretical Seminar

In the 1-week seminar, HFF camera students were asked to identify production techniques of well-crafted camera motion appearing in feature films from different cinematographic periods spanning from the beginning of the 20th century up to contemporary examples. Some of the scenes presented and discussed were (among many others) from the movies *Touch of Evil* (1958), *Fargo* (1996), *Boogie Nights* (1997), *Festen* (1998) or *Soaked in Bleach* (2015). For some examples also scenes from the making of were presented. Such screening helped to understand how the scenes were shot and how the results looked in consequence. A further objective was to derive and discuss the underlying motivation of directors, DOPs and camera operators. Inferring a motive was needed to answer questions on why they chose this particular combination of lenses, angles and movements or why they might have decided against other reasonable options. An additional aspect of the seminar was to decipher the connection between the camera work and the story and to interpret the relevance of the movement in advancing the storyline. For example, in some cases, that camera motion could be tied to the emotional state of the leading actor. So the more the actor was feeling uneasy that more the camera would shake subtly. In other cases, the camera could be detached from the content of a scene and take on the role of neutral portrayer of the event unfolding. To better illustrate the value and content of this seminar, we recommend to the reader to view an excerpt of the series *Breaking Bad* (2008-2013). In particular, a bit from the opening of the final episode of the series. A clip with this scene is provided electronically [450]. In this clip, the raw film material was extended by a text overlay. The text features an analysis of the camera movement applied and describes the tools and techniques used in the recording. It also hints at the motivation of the shot and how it is connected to the storyline of the episode. The scene exemplifies how a simple pan and dolly shot for instance can be used for dramatic effect with a simple gesture and advance the storytelling drastically by changing the narrative context. Further insight on the motives of a director regarding a particular style of camera motion can be seen in an interview with Alejandro González Iñárritu, the director of *Birdman* (2014) that is also provided electronically [547].

Practical Seminar

The practical seminar was conducted at a studio at HFF. The students were provided professional-grade camera and camera motion tools. Among those were a pan-tilt head, dolly, indoor crane, outdoor crane and a monitor displaying a wireless transmission of the camera stream. Over the course of one week, the students were first introduced to the tools by Leo Borchard. Following the introduction,

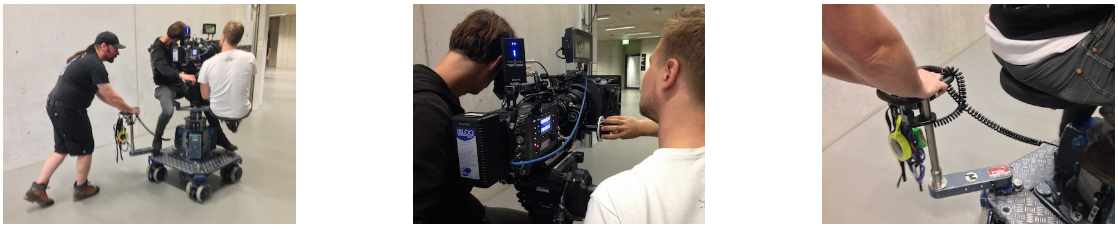
(a) The camera students taking on the various user roles of actor, camera operator and first assitant

(b) Camera operator and first assistant (left) levitated by the crane grip (right) on the rear end of a crane

(c) Camera crane used to move the camera and operators in an outside environment

Figure 3.3: Users: Camera students exposed to various tools for camera motion such as a dolly (Figure 3.3a) and camera cranes for indoor (Figure 3.3b) and outdoor use (Figure 3.3c)

they would set up the lighting of the scene. After that, the students were exposed to operating the tools on their own. In the process, some students would take on the role of actors while others would take the role of a camera operator, first assistant or grip. For the massive cranes, a professional grip was asked to operate the jib arm at the end opposing the camera platform. The scenes that were rehearsed and recorded often featured one actor moving around a two-story stage setup. The goal of the camera operators was to keep a near or close up shot of the actor's face at all time during the scene. Given a typical walking speed of the actor, a telelens attached to the camera and the motion of the camera platform (with dolly or crane), the execution is not trivial. Further, the first assistant was asked to continuously keep the focus on the eyes of the lead actor. So the operators not only had to account for the actor's motion but also for their own movement when focussing. Thus, before carrying out a shot, the students took the time to discuss and pre-visualise how the recording should play out, at which time or event they transition from one motion into another, where to start and where to end the move. After a set of five to six trials, Leo Borchard would review the results and give feedback to the students on what and how to improve before the next student could take a chance in the recording of the scene. The provided tools mentioned above were available at a minimum of one day so that there was enough time to shoot multiple scenes with one device. The next day usually a different device was introduced and explored. For the dolly, we also got the chance to take on the role of a camera operator our selves. However, given the same number of trials as the camera students, we could not execute the



(a) The camera operator (middle), the first assistant (right) and the dolly grip (left) on set

(b) Camera operator (left) adjusts the framing while the first assistant (right) focuses

(c) The dolly grip (left) pushes the dolly smoothly towards the scene and controls the wheels

Figure 3.4: Users: A camera crew recording a scene using a dolly as a camera movement tool (images from [124])

task successfully even once. Mostly we were unable to keep the face of the actor within the recorded frame. This first-hand experience allowed us to understand better the peculiarities and challenges of maintaining a proper framing. Figure 3.4 illustrates one of the scenes that were shot during the seminar. In this scene, the lead-actor would first come around a corner talking on the phone. The camera crew would already wait for her and anticipate her appearance (Figure 3.4a). Once she appeared, the camera crew followed her at a given distance to keep a constant framing. Simultaneously, the dolly grip started pushing the dolly (Figure 3.4c) and the camera operator adjusted the pan and tilt of the camera to frame her slightly to the right while the first assistant kept the focal plane on her eyes (Figure 3.4b). At one place the actor would stop surprised by a message on the phone while the camera continued to move further away from her. This choreography behind the camera reduces the workload of the operators but is also error-prone: if one crew member fails with the timing for examples, the scene needs to be reshot. However, in unforeseen situations, in this scene, when the actor suddenly stops earlier than scheduled, the crew can still react and manage to get a good shot.

3.3.2 Lessons Learned

Over the course of the theoretical seminar, we were granted a glimpse behind the curtain of film production in the camera department from a theoretical perspective. We gained the ability to roughly deduce how a scene might have been recorded and which tools were likely to be used in the process only from looking closely at the on-screen effects of camera movement. We further could take an educated guess in inferring the applied tools and the motives of cinematographers.

In the practical seminar, we could observe that the students would take on the different user roles of a camera operator, first assistant and grip as described in the literature. We could further witness how close and detailed the communication between the operators is to make sure the shot plays out well. They dedicated a similar amount of time to the planning of the scene as they needed for its recording. This process is not only time-consuming but also error-prone. Once one of the operators could not manage a task correctly or is out of synchronicity with the others, they would abort and start anew. We further got an in-depth introduction to existing mechanic tools, their principles of operation and human aspects in their control. For instance, in wide pans, the operators can often end in an uncomfortable position. In this position, they are often likely to be unable to afford the necessary body control to keep the framing stable and smooth, especially when the last bit of the scene takes some time. One workaround that was introduced is to start in an uncomfortable position and then to change with the pan into a more comfortable one to provide a better recording. Additionally, we could self-experience the challenging aspects of camera motion control.

3.4 Contextual Inquiries

In our user research, we wanted to make sure that we are not only exposed to students learning their craft in a teaching environment. Thus, we reached out further to camera operators. As a result, we were invited to join two camera crews at two different production sites during professional recordings (Figure 3.5). We used the opportunity for contextual inquiries on the practices of professionals on production location. Additionally, we wanted to incorporate the aspects regarding the question of *When we observe people in real-life production rather than a teaching environment, can we find evidence that they behave the way they suggest when asked and further, how they were taught?*

3.4.1 Environments and Observations

At the two sets, different kinds of material were recorded. The first location was the production site of the TV series *Dahoam is Dahoam* (2007-today) televised by the regional public broadcaster Bayerischer Rundfunk (eng. Bavarian Broadcasting) (BR). The production site is quite large with roughly 15.000 m^2 and located in Dachau near Munich. The set was built in 2007 and mimics the scene of a typical local Bavarian village. For example, it allows the recording of outdoor scenes, as in front of the façade of a church, as well as indoor scenes as in the local pub. The day of our visit, several outdoor scenes were scheduled to be recorded. Two camera crews would record the scene from different angles (Figure 3.5a). While a secondary unit would shoot from a stationary tripod, the primary unit would use a dolly to move the camera. The location would not only provide the scenes but a place for the staff to prepare, storage facilities for the camera department and a tent for the first steps in screening and post-production. These steps entail checking whether the material was recorded at all, is copied, matches the schedule, sometimes the director would already screen how the material plays out. In the beginning, the camera teams were handed out the actualised version of today's schedule. While the shooting schedule would be propagated earlier, sometimes last-minute changes are made. Then the camera teams would meet and discuss the details of the plan, once the

(a) Primary camera unit with dolly (right) and secondary camera unit (left)

(b) Camera operator (left) framing the scene with first assistant pulling the focus (right)

Figure 3.5: Users: Observing experts recording a TV series (Figure 3.5a) and a feature film (Figure 3.5b)

necessary details were figured out and consensus achieved, the camera operator of each unit would address further low-level issues with their assistants. Once they settled the details, the team would go to the first location and set up their equipment. Here again, each operator would instruct the assistants and sometimes ask them to try out different alternatives and asked them to report back. Once the motion devices, tripods, cameras, lenses and settings were all prepared by the assistants, the operators would go to the equipment and check whether the setup would work out for them. If necessary, they would ask the assistant to set it up differently or decide that they are good to go. Having set up their equipment, but still, before the shooting, the operators would check-in with the heads of the other departments such as lighting or audio and see whether they were also ready or whether changes in their setup were to be expected. In case that a department was ready for the shooting, they would also report back to the director who would ultimately signal all departments that the recording will start in several minutes. Once all departments were on their supposed locations and ready to shoot, the director would ask the team behind the scenes to start the recording and the actors in front of the camera to start playing. After all takes were recorded, the camera operator would already move to the next location, while the director often would debrief the actors and talk to them about the interpretation of the characters.

The second set was the shooting of an outdoor scene of the feature film *Thumb* (2016). For this recording, the location would only be available for the day of the shooting. Thus, all the gear had to be brought to the location and set up ahead. Similar to our previous visit, we could observe that first, the operators would discuss the setup with their assistants and then delegate particular tasks to them. Once those were done, the operators would check and adapt if necessary. When all of the gear was installed and prepared for the shooting, they would report back to the director who then would give instructions on when to start. During this visit, we could more closely follow the discussing between the operator and the first assistant (Figure 3.5b). Ahead of talking to the director, they would play out the camera motion multiple times to identify challenging parts and to synchronise their actions. They rehearsed the scene several times to make details such as timing, framing and focussing work. Sometimes during the rehearsal, they would also give feedback to assistants and actors so that they know not to block specific areas. They also recorded the camera move and reviewed the material without the actors. This way they made sure that they would not overlook something. The film material was reviewed on a separate monitor and was also used as a shared frame of reference. When discussing the mentioned details, both operators would often refer to parts of the recordings to enable the other operator to understand better what was meant by a particular remark.

3.4.2 Lessons Learned

From our observations, we could derive several insights. Independent of a fixed or temporary shooting location, the setup and relocation of equipment is time-consuming and happens often. Therefore, easy and quick setup is crucial for everyday work and vice versa, cumbersome set-ups are only reasonable in a fixed environment.

Also, the margin for errors in each department is minimal. A film crew, therefore, applies various approaches to make sure they minimise the likelihood of errors to appear. For each step of the production (ahead, during and after) they use different methods. In the beginning planning, pre-visualisation and schedules are used to make sure that once set the necessary equipment is on location and everybody knows what is scheduled to be shot. Ahead of the recording each department sets up and checks the equipment and the settings so that they are good to go. After the take, a short review of the material is done as well as a backup of the recording. Due to the low margin, tools not only need to be quick to set up but also stable during use.

To avoid errors but also shape an image creatively, operators would constantly check the framing and the focus of the camera resulting material. Unsurprisingly the very result of their work is central to a camera crew. Therefore, naturally, they want to make sure that they can provide the needed quality and flawlessness.

Beyond quality assurance, the camera stream is of further central importance. Similar to the seminar visits (Section 3.3), we could see that communication within a department is essential. On the sets visited this time, we could further discover that it is also vital among the departments. In our observations, we could see in-depth that to communicate properly, a shared frame of reference is essential. At multiple occasions, a look through the viewfinder, a review of the stream of the camera or a replay of the material shot was used as a foundation that enabled the emergence of such a shared frame of reference. The options mentioned were also used by people outside the camera crew. Directors need to see how a scene plays out and talk to or with actors about it. The camera crew members of various departments needed to see how movement unfolds and what they perhaps overlooked to decide on changes or to synchronise better. In consequence, 'seeing what the camera sees', is important in all major steps in shooting: preparation, during recording and after a take. Additionally, it is of relevance for multiple stakeholders involved in the production.

3.5 Expert Interviews

The goal of our previous approaches was to gain insight into the practices of working as a cinematographic camera operator. The review of selected literature (Section 3.1) could provide us with a framework of ideas and practices. The seminars (Section 3.3) offered us first-hand experiences regarding the motives of operators and their practices. Our contextual inquiries (Section 3.4) allowed us to come to a deeper understanding of the collaboration between the departments and the importance of setting up the details right within each. Reflecting on our approaches so far we found that our open explorations led to a general understanding, but for more in-depth insights we found a more structured approach complementing our efforts missing.

Therefore, we opted for expert interviews on background, challenges and the particularities of novel tools and automation in cinematographic movie production. For the interviews, we chose a semi-structured approach and used methods proposed by Grounded Theory [252] for their analysis. The underlying question of this step in our research regarding user ethnography was: if we ask operators in more extensive interviews that are informed by our empirical insights from previous efforts, *Can we identify central and common aspects (relevant for UI design) among operators provided that we use a structured approach in their analysis?*.

3.5.1 Participants and Procedure

Two of our recruited participants were female and seven male. The participants' years of age ranged from 22 to 63 with a median age of 28 years. We also asked our participants how many days of shooting they attended in the past 12 months. Ranging from 25 to 144, the average number was 65 days. Given that we also included students in our sampling, who were required to attend a curriculum of lectures and courses, we further divided this item. We split all participants further into a non-student group and a student group. The non-student group had 80 days of shooting experience in the past 12 months on average and the student group 58 days. We also asked how many years of professional expertise in film production our participants had in total. Overall, two of the participants had professional experience for more than 15 years, two participants for ten years or more (but below 15) and five participants for more than three years (but below 10). No operator had less than three years of experience. In summary, one could say that given the total number of 9 participants, we could manage to sample operators with varying backgrounds and years of age and shooting practice.

Aiming at a more structured approach, we chose the form of semi-structured interviews. Therefore, we prepared a set of questions ahead so that all interviews could establish a high enough degree of comparativeness (Appendix 12.2). However, we also wanted to acknowledge and go along with the answers provided by the participants. For these situations, we had no particular questions prepared; in these situations, we would mostly ask the participants to explain an issue in more detail or to expand on the connection to a point they mentioned earlier that seemed valuable to further investigate.

Over the course of three weeks, we carried out the interviews. They took circa 90 minutes each and were conducted at locations proposed by the participants. That would allow them to take part the easiest regarding their schedules and also in an environment that they regarded as comfortable and appropriate. Ahead of the actual interview, we would inform the participants on the purpose and details, and that they could leave at any time. We further asked them for their consent to take part in the interview and its recording. Having declared consent, we asked the participants to answer questions regarding demographics and their prior experience in camera and motion tool operation. The questionnaire then would start with some simple warm-up questions such as *"What is your definition of fun behind the camera?"*. After the warm-up phase, we would transition into the details of a shooting that they would characterise as a best-case scenario. Here we wanted to know more about what is essential to make things go the right way and what criteria operators would use to evaluate that things went well. In the following, we would also ask on the details of a day of shooting that can be considered the worst-case. Here, vice versa we wanted to know when and how things went wrong and whether operators would use the same (or inverted) criteria to evaluate the situation as in the best-case they described earlier.

These two expansive parts of the interview were followed by smaller sets of questions asking on the details of the practices in preparation, communication and camera operation. As humans tend to judge their behaviour different than the behaviour of others, we asked to describe what aspects they could observe in the practice of others that would cause unnecessary problems. After the interview was finished, we would thank the participants and shortly debrief them.

3.5.2 Analysis and Results

We recorded our questions and the answers of participants during the interview with an audio recorder. After all, interviews were conducted, we transcribed the audio recordings to text documents using the software f4 [327]. Using the transcripts as a basis, we applied the methodology of open coding. Thus, we would take the answers apart into chunks of mentions that would only relate to a particular issue or sub-issue. These singular mentions would further be clustered into a group if they would share a greater common theme. These heaps of remarks that described a particular aspect of an overarching theme were then assigned a name/category that would represent the identified topic. Additionally, we would count the total number of individual mentions and the mentions per category. Categories that would result in a higher overall number of mentions would consequently be regarded as a more relevant topic to the participants (given our questions). To minimise the degree of arbitrariness in our coding process, we used two initial coders who would conduct the process of identifying specific mentions and grouping. The results of both procedures were given to a third meta-coder who would harmonise both results and provide the final results.

In inspecting the interviews, we identified several mentions concerning social aspects at work, as in the following example *"Of course, the director agrees with the camera operator! He actually chooses his camera operator, so that they can work well together."* (P08). Another instance was *"Technical (and) aesthetic expertise is one thing, but the ability to work in a team is another."* (P01). However, for our research, the tools and practices in camera operation and their relation to user interface design were the central study object. Therefore, we further focused on mentions that would relate to it. Following this premise, we identified seven main issues that were repeatedly discussed throughout the interviews. We present these in more detail below. For all main issues, we will also provide quantitative data along with each. The quantitative data was used to derive a degree of relevance as explained above. As report format for the data, we chose to present the following two numbers in brackets for each main issue. The first number references the number of subjects mentioning an issue; the second number will equal the total number of mentions for each class.

Preparation (9 / 65)

During preparation, the departments discuss the story, the camera viewpoints the locations and logistics and try to come to a consensus if the opinions are diverging. This phase is characterised by intense effort in communication between and among the departments. A multitude of tools is used to illustrate and communicate ideas. Besides the script that is often used, we additionally found mentions of storyboards, mood boards and floorplans. Not all tools are used by all crew members as illustrated in one statement by P07: *"Storyboard, floorplan, all sorts of things. (...) Sometimes there is a storyboard department that draws that. But a floorplan is done by the camera operators themselves and the shot list is written following their conceptions. (...) And a grip who makes drawing or notes how this is handled in staging and with a crane, dolly or Steadicam."*. Overall it is worth mentioning that the importance of preparation was reported by every participant.

Low Error Tolerance (9 / 50)

The answers to the interviews suggested that in general if an error occurs the shot is stopped, and the scene is set up again before a new shot is taken. However, there is also a slight margin of error that is being tolerated to stay within the overall budget of time and money. Therefore, operators will continue recording although they see minor compositional insufficiencies. In such situations, they speculate that the audience does not necessarily register such small mistakes. Given for example a scene with an intense or dramatic point in the story might allow for some minor issues as the attention of the audience might be mainly on the characters and the unfolding of the story. Producing shots that entail errors is so common that every interviewee would mention it. As P09 for example: *"So we rehearsed and shot that several times because something always was not right."*.

Need for Improvisation (9 / 39)

Advanced film productions undertake a remarkable effort regarding preparation. However, not all details can be prepared ahead. That can, for instance, be attributed to the type of the movie such as documentaries or simply to the fact that actors can never precisely play a scene identically even if rehearsed. Although the degree of unpredictability might vary, all operators had in common that they have to adapt at some point. As P07 put it *"It is rare that everything works smoothly (...) the problems can come from any direction: technical, personal, weather, motive or acting"*. Consequently, independent of the quality of the preparation, operators need to react spontaneously.

Constraints (9 / 38)

We also found a set of constraints that would influence the way the operators would approach a shooting. We grouped the mentions into six groups all referring to a different constraint. The three most obvious ones were time, money and space. These are quite unsurprising as they are not particular to the field of cinematography but rather apply to any business effort. In the words of P07 *"Staging, dolly/crane shots and then it's calculated how much the motive is beside the rent. The operator checks whether it is doable at all."*. Operators not only have to weigh in aspects regarding their craft but also regarding their budget. Additionally, also the desired aesthetics of a movie, the rhythm of sequence that a scene is surrounded by in the final movie and the zeitgeist can affect the selection of techniques and tools. P04 summed it up in the following statement: *"The era of the grand gesture is over."* Given our semi-structured form, we could ask him to expand on this particular point. More broadly speaking what the comment referred to was the notion that in contemporary cinema the cutting rates have increased (this is also a finding of the statistical analysis presented in Section 2.2.2). As the total movie lengths remained rather constant, the lengths of the scenes needed to be shortened. Most remarkably this can be seen in the genre of contemporary action movies. Additionally, the rhythm of a movie is paced by the cutting rate. In consequence, if the cutting rate is as high as in contemporary examples, it is hard to include a scene with elaborate continuous camera motion without interruption[3].

High Workload (6 / 29)

As mentioned before, the margin of error was described as low by the participants and mistakes are easily made. One of the major contributing factors seems to be the complexity of moving the camera purposefully in multiple dimensions while adjusting the focus accordingly and in synchronicity with the action before the camera. Regarding the tasks of focus pulling, we found multiple mentions on curved motions. Keeping the focal plane seemingly steady, so for example on the eyes of the actor, is rather hard once the camera and the actor move on separate curved motion paths. Therefore estimating the right distance is not trivial due to the fact the resulting distance is changing continuously and affected by two variables at once. Further, if an open aperture is used for a shallow DOF, the depth of the focal plane is in particular small. Considering the complexity of the overall task P05 remarked: *"This planning of camera movements (...) this is so difficult, I think it comes with experience."*

Context Dependency (5 / 42)

Not only the constraints pointed out above are influencing which tools are used is which approach is favoured. Also, the context is relevant for the operators in their decision-making. From our sampled answers we could overall identify five contexts that would lead to different choices the operators would take to record material. As pointed out by P03: *"Working towards a certain picture and very technical preparation is common when it is a commercial."* We list the context and important aspects for the operators below:

Scenic: The environment can be characterised by intense preparation, the shots can be repeated as they are often staged, frequently provide high monetary budgets and hence can make extensive use of tools.

Documentary: Opportunities for preparation are limited to a degree. Also, most actions are not repeatable, and productions are mainly only provided a medium or low monetary budget. Due to budget limits and the need to bring all the equipment to sometimes also remote locations result in restricted use of motion support tools.

TV Studio: In a TV context there is room for preparation and need for scheduling. Shots can be repeated, unless there is no live broadcast, depending on the format, the budget varies between medium and low, in contemporary formats highly automated tools are already in use, as they can contribute to the visual identity of a format due to their repeatable movement.

[3] Challenging this idea, instances can be found that take an opposing approach. *Birdman* (2014) [446] ▶ is recorded so that it looks as if it was one long take. There are cuts, but they are hidden very well. Most remarkably, *Victoria* (2015) is a feature film that was actually recorded in a single take. However, such examples of a counter tendency are rather singular instances.

Advertisement: As aesthetics are highly valued, great effort is put in preparation as well as the actual recording. That leads to shots that are often repeated before the desired quality is achieved. Although advertisement clips are rather short, the high quality often requires a high to medium budget and hence here also the use of highly automated tools is already likely.

Sports: The broadcasting of sports also makes a great amount of preparation necessary. Although high budgets and the use of highly automated tools is common, the results vary less. That is often due to a restricted number of angles. These can result either from wanting to show as much of a playfield as possible to provide an overview of the game or due to regulations by sports federations which issue the broadcasting licences.

Being in Control (4 / 12)

Experiencing agency over the situation and the tools involved is an important and recurring theme among the interviewed operators. Being able to change settings quickly and most preferably manually is most likely connected to the above-identified need for improvisation: *"Every camera operator who looks through the eye-piece of three cameras or so does need this control or supremacy of the one who is in charge of the look."* (P02). For tasks in manual operation, operators also tend to prefer physical controls that provide haptic feedback and solid tools. The later seems to be likely to be connected to the low error tolerance already found previously in our interviews.

3.5.3 Lessons Learned

Besides the social aspects that we did not further dive into, we could identify seven areas that are important to operators. While this list is most likely not extensive, it is intended to help in formulating higher-level abstractions regarding the central aspects of camera work. Therefore, we would further propose that our identified classes could also be grouped further into abstractions of even higher levels. We suggest that preparation, low error tolerance, constraints and context which afford decisions by the operators are grouped into environmental factors. Once decisions are made based on these environmental factors, (assisting with the) high workload, (satisfying) the need for improvisation and (providing the sense) of being in control need to be addressed. These issues could be grouped as (system and UI) design factors which affect the way the actual camera operation is carried out.

Additionally, we found that regarding novel tools operators tend to be hesitant. This is likely to be connected to the low margin of error we encountered. So, to avoid user rejection, even early prototypes will most likely need to provide a large enough degree of sophistication. Similar reactions were found by Garrett Brown when introducing his Steadicam prototypes in the late 1970s [525]. ▶

3.6 Summary

As we could derive from the literature in Section 3.1, moving the camera is a complex task. To deal with that complexity, task division and delegation is often applied. Thus, different user roles emerged that came with different responsibilities. To fulfil a role properly, expertise regarding the particular tools is necessary. When investigating in the wild practices in Section 3.2, we found that this approach is mainly followed in larger productions as, especially in low budget productions, hiring multiple operators is not feasible. Consequently, one operator, often supported by technologically advanced tools, takes on all major roles at once. In our seminar visits described in Section 3.3, we were sensitised that camera motion is not applied arbitrarily. Quite the contrary, why certain angles, motion paths and techniques are used depends on multiple layers of motives. Often with the reason being to serve a narrative or a compositional function. A certain angle or move, however, could also be a homage to another movie or is used to create a specific mood. As mentioned in the literature, we also experienced first-hand that carrying out a move is a complex task. Operators need to train intensely to master that skill, but also communicate intensely with others. To establish and maintain a shared frame of reference, they often use the recorded material as pointed out in Section 3.4. That helps to communicate with crew members within and outside the camera department. Whether it is a live stream or a replay is not crucial, however being able to review the material on a big enough screen is favoured. This contrasts our finding from Section 3.2 where a live camera stream was not rated as majorly important. Here, our question was also only concerned with slider UIs. As other tools such as drones (see Section 2.2.3) make intense use of live camera streams, we suggest that there are good reasons to emphasise a live stream, such as increased general Situational Awareness (SA) [82] and action-feedback association [270], while acknowledging that some tools might benefit more than others from the feature. In our interviews in 3.5, we found that environmental and design factors contribute the most to the decisions and experiences of operators. As we focus on system and UI design in this thesis, we regard the following issues as most essential: (assisting with the) high workload, (satisfying) the need for improvisation and (supporting a feeling of) being in control.

Overall, we had exposure to more operators than the ones who ultimately took part in the studies. Given the still limited total number of operators observed, we found that there are some user groups among the operators that would share particular points of view, practices or other traits. From our exposure, we would not go as far as saying that this already is sufficient to enable the construction of personas, especially when considering the detailed propositions by Alan Cooper [56]. Still, we wanted to conceptualise our understanding of that matter. Therefore, we opted for restricting our efforts to the creation of meta- respectively proto-personas that we list below. They should only serve as a reference for the commonalities we saw within sets of users and distinctions towards other sets. As they rather describe commonalities than individual details, we refer to them as meta.

Settled Traditionalist: Permanent camera operator, who usually passed the mid-40s. Was trained using analogue tools and experienced the shift towards digital in his working life. Strongly trusts in own ability, hesitant towards novel tools; jokingly remarks that the most important part of any automation is the off-button.

Experienced Teacher: Usually has passed the mid-30s and had a fair share of exposure to various domains of cinematographic production. Can provide a rich portfolio incorporating feature-film, short film, TV series, advertisement, image film, music video production and more. Is open towards novelty in production techniques but needs a convincing idea of how they can advance the craft or enable new perspectives. Contributes own ideas and seeks to empower others.

Settled Creative: Makes a living being a camera operator but also has other creative endeavours going on in parallel. Is open to new tools and approaches, but also has a preferred personal style of working. Prioritises the creative aspects and the content of a scene (and how it relates to the overall story) over the technical sophistication in the production. Thinks that great movies can be recorded with simple techniques when the story and creative aspects are strong enough.

Diligent Worker: Has advanced knowledge of cinematographic productions and the underlying means of production. Is further interested in the peculiarities of a broad range of camera and motion support tools and hence picks up novel operating techniques quickly. Is open towards novel approaches and technologies but also aspires to record material that resembles the work of personal role models in cinematic history. Is almost always industrious, calm and present at work, even when things go wrong. Thinks that even if several attempts failed, they will get the shot right next time (or the time after that).

Exploring Student: Has a high interest but so far only limited background in camera operation. The great interest leads to a high degree of openness towards all kinds of approaches and tools. Maybe has multiple conflicting ideas on how to approach a particular shot, but is willing to put in the effort of learning the craft, trying out different options one-self and discuss the results.

Intuitive Artist: Has an artistic background that does not necessarily involve training in camera operation. Wants to try out all sorts of media to work with. Often uses a Do It Yourself (DIY) approach in directing a film and operating the camera. Does not necessarily compare own work to the work of masters in cinematography, but instead strives to realise an individual vision that was derived intuitively without caring much about traditional compositional rules or user roles.

Associated with the differences in types of users, we consequently figured that it is likely that they approach camera and tool operation and achieving mastery of it following different paths. If we were to put these paths into a form that would allow us to identify shared commonalities and distinctions between different approaches, we would suggest using the form of proto-user journeys. Relating them to already existing user journeys and ways of thinking about them, we propose two primary proto-journeys given our studies. On the one hand, the expert journey and on the other hand the journey of walk-up and use. We expand on both proto-journeys below.

Expert journey: Is a path of striving for mastery without a predefined end. Higher-level mastery requires the achievement of lower-level mastery of various subdomains. The process is characterised by years of learning the history of the domain, identifying its fundamental principles and acquiring the capabilities in multiple areas of translating abstract ideas into reality. In particular domain knowledge in the operation of cameras and motion support tools is a central aspect that particularly covers operator training. Once a great enough degree of mastery is achieved in some subdomains, the journey can continue on a higher level. Here it is characterised by establishing a personal vision and attitude. The mastered subdisciplines and knowledge of universal principles allow the operators to derive novel and unique solutions. These can be understood as new solutions for situations that already have been encountered and properly addressed before (but now in a different way) but also as solutions for unknown situations that have never been encountered before.

Walk-up and use: For a smaller part of the sampled participants a traditional walk-up and use scenario would apply. This sort of user journey would not necessarily be based on a phase of extensive prior training. Occasionally, the intuitive artist would use that kind of approach. However, it is worth mentioning that in our estimation, a traditionally trained expert in camera motion is not always also an expert in operating newly emerging tools. Given that most of the proto-user types we identified are open to exploring new tools in techniques, the walk-up and use proto-user journey should also be part of the design process of new tools and UIs even if they are tailored for use by experts.

Insights

- Integrating experts and enthusiasts in research on computationally extended cinematic tools seems reasonable as they are very conscientious of why and how tools are applied in practice. For this, observations on location and personal appointments seem more beneficial than remote approaches.

- As experts often have busy schedules one needs to put in dedicated effort or cope with approximating methodologies. Interviewing them, we identified that camera operators face a high workload and a low tolerance for errors. Computational systems could assist by increasing the level of automation or the intelligence of systems. However, operators also want to be or, at least, feel in control and need to improvise. Thus, systems should also be designed incorporating adaptive strategies regarding their assistance.

- Additionally, in such novel systems, designers should stay consistent with existing hierarchies, user roles and task delegation practices. Acknowledging those patterns, one automated sub-system should only assist in a task that a dedicated operator would execute. Thus, for each of the traditional tasks, one sub-system would be required. Such a sub-system would further allow accounting for the insight above. In an adaptive system, remote operators would still be able to take over and improvise once a task is delegated.

PETER AGRE (BIOLOGIST)

In science, one should use all available resources to solve difficult problems. One of our most powerful resources is the insight of our colleagues.

STEPHEN HAWKING (PHYSICIST)

As scientists, we step on the shoulders of science, building on the work that has come before us - aiming to inspire a new generation of young scientists to continue once we are gone.

EDSGER DIJKSTRA (COMPUTER SCIENTIST)

Computer science is no more about computers than astronomy is about telescopes.

4

Research: Insights from Related Work

What to expect?

- Trade-offs in designing automated systems
- Interaction styles for camera control
- The human sense of control and its application in HCI

What to take away?

- A more fundamental understanding of the relationships between automation, camera control and the human sense of control

4.1 Automation

In Chapter 3 we established the idea that assisting technologies might provide benefits to operators, especially in single-user environments. Here, automation seems particularly promising. In general, it has already been researched for decades in prior work. To provide further details on why the initial idea seems viable to us, we present selected work that is linked to the question of *What is fundamentally shaping the human-automation relationship?* Both actors at play, humans and machines, can contribute different benefits to reasonable cooperation. Therefore, we discuss work that identified the strengths of each below. Identifying these strengths led to a set of fundamental system design patterns aiming at exploiting and combining them. These designs introduced automation benefits in a human-centred fashion but also side-effects that inevitably came along with it. Therefore, we additionally present strategies and designs that try to reconcile benefits and downsides by addressing the emerging side-effects. In general, camera operation is a task that is characterised by technical as well as creative aspects. Automation in creative tasks, however, faces different challenges than traditional industrial automation that is often the main interest of research and literature. Thus, we also dive into the peculiarities and challenges of applying automation to creative work.

4.1.1 Human-Automation Relationship

Automated systems and associated user interfaces have taken on various forms in the past as automation could be applied to various application domains with great success such as aviation, industrial production or health-care [15, 66, 205, 241]. Central aspects that explain the success across various domains might be found in the benefits that automation can offer regarding safety, reliability, economy, and comfort [15]. Over time, varying designs were used to assist humans or even to outperform them. Below, we will outline in more detail, how humans and automated systems can complement each other best and how the different designs can be classified. In general, classifications of such kinds can help to inform the designs of novel systems or to adapt better to a given environment. We will also look into problems arising as side-effects especially in creative domains and how to address them.

Complementing Benefits and Function Allocation

As mentioned above, automation has been researched for decades. As early as the 1950s, researchers investigated the nature of the then promising technology and its relationship to humans. For example in 1951, Fitts [87] proposed that the relationship could be understood as complementary based on research regarding air trafficking systems. His fundamental point was that both parties have certain advantages over another (Figure 4.1). Preceding his description of the particular domains of strength for each party, Fitts remarked that *"Human tasks should provide activity"* [87] and that *"Human tasks* *should be intrinsically interesting"* [87]. Given these presuppositions, he suggested that humans could outperform the then-contemporary systems in the following areas.

Sensory Functions: Fitts remarks that (within the bounds of the human sensory apparatus) often only a minimum amount of energy is necessary for the detection of stimuli. He presents the human visual and auditory systems as examples that are capable of registering subtle cues or stimuli with low effort. (He also acknowledges that machines can sense outside these human bounds.)

Perceptual Abilities: In addition to the previous point, humans further show capabilities of stimulus generalisation. This term describes the phenomenon that humans can identify a particular object and sort it into a (larger) describing class even if the object has never been encountered before. This process is furthermore independent of changes in illumination and other variation such as shape, form or size, For example, the designs of chairs can vary widely, but even entirely new designs can be identified as items of the class of chairs.

Flexibility: Humans share the ability to improvise and adapt flexibly to occurring problems. While Fitts found that the psychological understanding was still limited at the time, (so far) machines were unable to compete with humans regarding flexible and general problem-solving.

Judgement and Selective Recall: According to Fitts, judgement is a capacity uniquely observed in humans. It is characterised by being able to store vast amounts of information and retrieving appropriate parts out of this long-term storage at the right time. Further, it can also be applied in a generalised way, meaning that proper judgement regarding a situation never encountered before can be (in parts) informed by past experiences.

Reasoning: Fitts sees the advantages of humans especially in the domain of inductive reasoning. He refers in particular to conceptualising a new set of principles based on a set of empirical data. He further acknowledges that there are machine advantages regarding deductive reasoning, however.

In summary, Fitts sees distinct human strengths in their ability to adapt to changing or even unforeseen circumstances (based on inductive generalisation and reasoning). In contrast, he also remarks that automated systems can perform better than their human counterparts, in particular, in the following domains.

Speed and Power: Humans face natural limitations as to how fast they can orient themselves and respond to a change in the environment even when provided with training and only a limited set of choices (see also Hick's Law [112]). Machines, however, can be engineered to respond in a matter of milliseconds over even nanoseconds. Further, they can be constructed to excel great force in fast, smooth and precise ways which is also hard for humans.

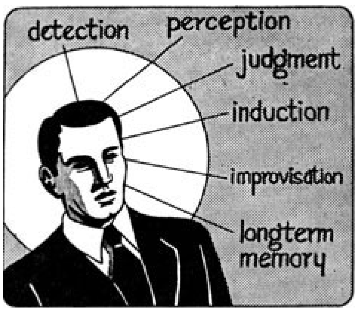
Rountine Work: As machines do not face the limitations of becoming bored or inattentive, they can produce consistent and repeatable results with fewer errors (especially when actions need to be carried out fast).

Computation: As, by design, computers are built to compute, they tend to outperform humans in this domain. Especially in complex situations which can entail the aspect of deductive reasoning mentioned above. Given a set of rules and postulates a computer can, for instance, derive all possible deductions and then reject invalid ones.

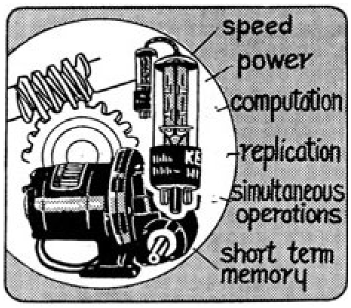
Short-Term Storage: Also, regarding the capabilities of storing information short-term, machines seem to outperform humans. On the one hand, they can completely erase a set of information to prepare for a new task which human operators hardly can do. On the other hand, it is harder to assert that a human has registered provided information. Therefore, precautionary measures then become necessary for error prevention.

Simultaneous Activities: Humans further face limitations regarding how many tasks they can take on in parallel. Leaving out vegetative functions such as breathing, humans often can only carry out a single primary cognitive task properly. Only once the tasks are over-learned, multiple can be taken on in parallel by switching between them quickly.

Fitts concludes that the proposed list should only be understood as a starting point and that further research is surely needed. Similar work has been done along those lines since and became summarised under the term of function allocation research. The results are sometimes also referred to as Men-Are-Better-At/Machines-Are-Better-At (MABA-MABA) lists.



(a) Advantages of human operators over machines



(b) Advantages of machines over humans

Figure 4.1: Research: Representation of Fitts' MABA-MABA list displaying human (Figure 4.1a) and vice versa machine (Figure 4.1b) strengths [87]

Since Fitts' list, further work on function allocation was presented as by Chapanis (1965) [41], Edwards and Lees (1972) [77], Mertes and Jenney (1974) [178], Swain and Guttman (1983) [253], Sheridan (1987) [242], McGuire (1990) [176] or Kaber (2017) [135]. Although Fitts' list was primarily intended only as a starting point, over decades, researchers still keep on referring to this initial list.

De Winter and colleagues [65] discuss this finding and propose that some reasons for its longevity might be found in its fulfilment of certain criteria for assessing scientific theories; namely plausibility, explanatory adequacy, interpretability, simplicity, descriptive adequacy, and generalisability.

Levels of Automation

The idea of what and how to allocate became subsequently refined by further researchers since its initial proposal by Fitts. As a consequence, concepts beyond binary function allocation [23, 81, 83, 273] have emerged, for example in the form of gradual approaches. Early mentions of such ideas can already be found in Fitts' paper, suggesting four possible user roles[1] or by Bright [23] who understood automation as unfolding through 17 levels of competency and that different sorts of workload, skill, and training are required by the different levels. Sheridan and Verplank [243], for instance, further expanded on the idea and provided a model in which full manual control and full automation mark the ends of a spectrum. Further, on this spectrum, various intermediate Level of Automation (LOA) can be found. For the design of systems, the LOA is intended to be adjusted to fit a given environment, task and operator skill-set best. Based on their model, they proposed a set of ten levels (Table 4.1.1).

Levels of Automation	
1.	Human does it all
2.	Computer offers alternatives
3.	Computer narrows alternatives down to a few
4.	Computer suggests a recommended alternative
5.	Computer executes alternative if human approves
6.	Computer executes alternative; human can veto
7.	Computer executes alternative and informs human
8.	Computer executes selected alternative and informs human only if asked
9.	Computer executes selected alternative and informs human only if it decides to
10.	Computer acts entirely autonomously

Table 4.1.1: Research: Levels of automation as proposed by Sheridan and Verplank [243]

Other researchers such as Parasuraman [202] pointed out that gradual approaches face difficulties as well. Decision-making, providing suggestions and autonomous action-taking require a dedicated goal. However, not all tasks that can be automated necessarily also have a goal to it that is well-defined. Foremost, such tasks can be found in the domains of sensing and information analysis, which do not require any suggestion or action to be proposed or executed. Thus, Parasuraman suggested a further approach as a refinement. The proposed technique built upon the fundamentals of human information processing which can be modelled and decomposed into four steps: sensory processing, perception and/or working memory, decision-making and response selection [207].

[1] Fully automatic control, automatic control with human monitoring, semi-automatic control supplemented by human control of critical functions and primary control by human operators who would be assisted by effective data analysis, transmission and display [87].

Sensory Processing: Information of various sources is acquired and registered including the position and orientation of receptors, pre-processing of input and selective attention.

Perception / Working Memory: Information located in the working memory is perceived consciously and might be manipulated including rehearsal, integration and inference.

Decision Making: Cognitive processes as mentioned above are precursors to choosing a particular decision out of a set of possible decision alternatives.

Response Selection: To manifest a response, the necessary actions are determined and implemented in accordance with the decision taken above.

This model was translated by Parasuraman [207] into functions that a computation-based system could execute. Therefore, also four (different) essential steps are defined in his model: information acquisition, information analysis, decision and action selection and action implementation.

Information Acquisition: Sensing and registration of input data including mechanically moving the sensors, organising the information stream to some criteria, highlighting parts of the incoming information or filtering of particular information.

Information Analysis: Inferential processes including extrapolation over time leading to predictions, integrating multiple sensor values into a fused variable or even providing context-dependent summaries for users.

Decision and Action Selection: A decision and its following action sequence are selected from the potential decision and action alternatives. Selection functions as well can be designed according to the model of Sheridan and Verplank.

Action Implementation: The necessary steps for acting out the chosen action sequence are executed by the automated system. The actuators and actions are highly dependent on the capabilities of the system.

For each LOA of the Sheridan-Verplank model, the proposed information processing technique can be used as a foundation informing medium to high-level actions to attain the overall goals. Additionally, for a particular LOA, the four steps of this technique, however, can also be assigned various LOA themselves. Thus, for example, a system with an overall medium LOA can assign a low LOA regarding information acquisition and analysis (consequently involving an operator more), while assigning a high LOA to action selection and implementation within the same system (Figure 4.2).

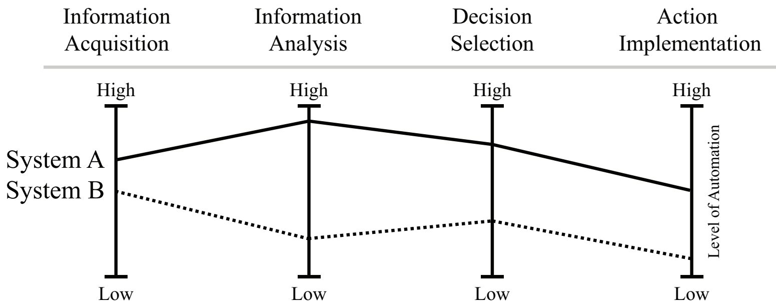


Figure 4.2: Research: Information processing based on Parasuraman et al. [207]. For two exemplary systems, different LOA were assigned varying for each of the steps of the proposed information processing model.

Trading and Sharing

Strategies that allocate based on complementing benefits as proposed by the Fitts list have also been referred to as comparison allocation. In general, there are two further types of allocation: leftover allocation and economic allocation [127]. The former means that any function that can be executed by a machine will be allocated to it and human operators are assigned what is left (maximise automation). The latter refers to the practice of assigning all functions to the system as long as it is economically reasonable, if costs are higher than human labour, the tasks are assigned to humans.

As an alternative framework of thinking about the allocation of control, Sheridan [242] introduced the concepts of trading or sharing it (Figure 4.3). Here, sharing refers to simultaneous cooperation between humans and systems regarding one specific function. He differentiated three types of sharing:

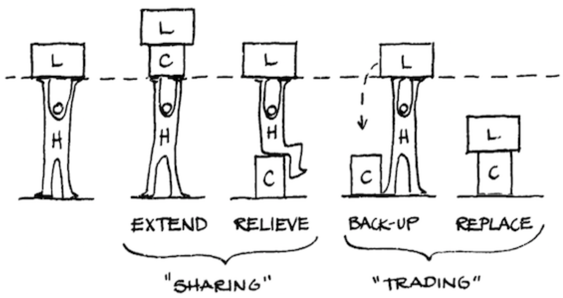


Figure 4.3: Research: Roles of computers proposed by Sheridan [243] (L = load or task, H = human and C = computer)

extension, relief and partitioning. In extension, each of the agents helps the other to extend their own capabilities. Regarding relief, the system executes a task for the operator to ease the workload. In partitioning, tasks are delegated to each agent according to their capabilities. Trading is rooted in the idea that one of the agents is exclusively in charge of a function. Who is in charge can change over time. Therefore, decisions on when to hand over control and to whom are needed. Most importantly also who is responsible for those decisions (human or machine).

Ironies

Automation is not free from challenges and Fitts' 1951 paper already hinted at those. Later, in 1983, Bainbridge [8] took a retrospective look at the benefits and the side-effects of introducing automation in industrial processes and suggested in hindsight that some ironies accompany automation. In her work, she refers to ironies as a *"combination of circumstances, the result of which is the direct opposite of what might be expected"* [8].

Bainbridge sees an irony in the fact that system designers want to automate processes because human operators are to some degree unreliable and inefficient. However, for tasks they cannot manage to automate they, again, rely on human operators. Further, in automated systems, only limited sets of tasks remain for human operators. These can be mainly assigned to one of two categories: monitoring or taking over. Take-overs face ironies regarding physical and cognitive aspects. Physical skill degrades if it is not used regularly ('use it or lose it'). However, the fundamental point of automation is to minimise physical task execution by operators (inevitably leading to degradation). Similar observations have been made regarding cognitive aspects. For operators, it is important to retrieve the 'right' knowledge at the 'right' time from the long-term memory. However, the knowledge that is stored in the long-term memory depends on what operators frequently do. So given that many tasks are automated, the particular knowledge needed in a take-over situation might be hard to retrieve. Also, if a system recognises a failure, usually, it sends a take-over request to a human operator. In these cases, a quick and accurate response by an operator is needed. The more the skill-set has degraded, the harder this gets for operators. In safety-critical situations, this can introduce additional risk. However, risk minimisation was the reason to introduce automation in the first place. In general, monitoring tasks appear as simple tasks: operators check whether the process is executed properly and intervene if necessary. However, in practice monitoring can easily become a tedious task, especially if no intervention is necessary for longer periods. Vigilance studies [169] suggest that, even for motivated operators, maintaining visual attention and focus on information source where only little changes already becomes increasingly harder after 30 minutes. In consequence, it seems reasonable to assist operators with alarms. This introduces novel ironies. How can operators monitor whether the alarms work properly? More fundamentally, monitoring is predicated on the idea that systems that execute tasks better than operators. However, ironically, operators are asked to observe the quality of this execution. The more complex the tasks that are automated get, the more this becomes a design challenge. So for instance, systems can take decisions and implement actions much more rapidly than humans. However, to enable operators to observe the quality of the process, the systems need to inform operators at a (slowed down) rate that is humanly comprehensible. What is exemplified here regarding speed also applies to levels of complexity or the number of variables considered in the reasoning (or the combination of both). In consequence, this renders real-time monitoring hardly feasible. In summary, monitoring can be a tedious task yet very responsible and at the same time. Further, it is hard to learn and especially maintain what is needed to live up to that responsibility.

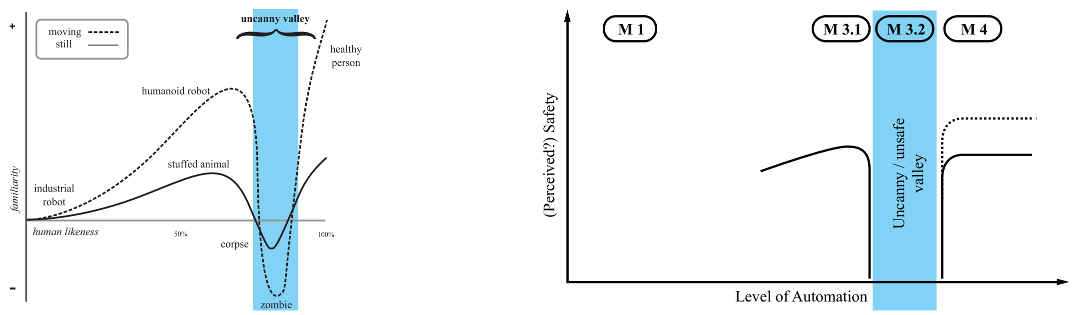
She also proposes solutions targeting the identified ironies. These can take on multiple forms: such as alarms on alarms (forwarding the ironies to a different level), displaying target values or providing analogue backup systems if electronic devices fail. Systems could also be slowed down to foster human comprehension (which however reduces the overall efficiency of the system), introduce fake signals to increase vigilance (which was observed to diminish trust in the system) or avoid such problems by shutting down immediately (which has its own problems as is not always feasible because rapid shutdowns can cause massive risks). Also, operator training is proposed (which often makes simulators necessary due to safety reasons, however simulators ironically cannot prepare for unforeseen real-world problems). So, in conclusion, Bainbridge states *"Perhaps the final irony is that it is the most successful automated systems, with rare need for manual intervention, which may need the greatest investment in human operator training"* [8].

Criticism

Beyond pointing out shortcomings inherent to automation, also more severe criticism was formulated regarding the approaches of function allocation based on intermediate LOA and information processing models. Dekker and Woods [67], for example, critique the fundamentals of these propositions. Thus, they propose that the main question that researchers and designers ought to answer about humans and automated systems should rather be *"how do we make them get along together"* [67] instead of *"who does what"* [67]. They expand on their point of view by suggesting that the very idea of function allocation is not feasible conceptually. They unfold their argument along the following lines. As high-level functions are composed of lower-level sub-functions, a function can be compartmentalised in smaller and smaller fractions. However, how many layers there are or should be is not self-evident and mostly arbitrary. So, it remains undefined how one should reason whether and when to intervene or not or whether to hand over the control to the system or not. Additionally, function allocations to systems also have side-effects. They are likely to introduce new tasks or challenges along the way that have not existed before but need to be addressed. In consequence, human operators are always involved in any given task anyways. Further, at the bottom of such a function hierarchy, one usually needs an underlying fundamental model of information processing as the models proposed by Parasuraman [207]. Regarding systems predicated on such models, Dekker and Woods point out that they ultimately all face the parametrisation problem. In short, automated systems cannot be provided with access to all relevant world parameters by design and hence are ill-equipped for problem-solving in all contexts. This is also a compounded problem, as, first of all, one would need to know what the relevant parameters are, which even humans often do not. Secondly, one would need to make sure that these all at least can be registered by machines which might sometimes be a hard technical problem itself. Thirdly, this would only include situations where solving a problem based on parametrisation is wanted, however in a creative context that might not even be the point fundamentally. Consequently, no actual way would reasonably allow a full substitution. Therefore, human operators and automated systems need to cooperate anyway. Further, they conclude that researchers and designers should therefore rather focus on how systems can make their reasoning and actions transparent and how users can contribute to the reasoning and the carrying out of actions.

The Uncanny Valley of Automation

Besides automation capabilities, ironies and fundamental critique, there are also further noteworthy perspectives in particular regarding the User Experience (UX). Here, Flemisch and colleagues [88] describe the notion of an uncanny valley of automation. They are mainly concerned with driving and



(a) The uncanny valley in robotics based on [188, 407] *(b)* The uncanny valley in automation based on [88]

Figure 4.4: Research: Uncanny valleys in robotics (Figure 4.4a) and automation (Figure 4.4b)

as crashes often can be an unwanted consequence resulting from a crossing of human and machine control also refer to it as an unsafe valley. Similar to the uncanny valley in robotics coined by Mori [188, 189], they refer to a phenomenon of irruption regarding the perceived (positive) quality aspects of systems. Beyond this area of irruption, however, the perceived positive aspects are also observed to stabilise and even increase again. In general, researchers try to identify what causes such a valley and to figure out solutions that keep the promising aspects accessible that are beyond it.

It is worthwhile to note that both valleys are fundamentally different. In human-robot relationships, the sense of familiarity can be irrupted depending on the human likeness of the robot (Figure 4.4a). In contrast, in automated systems safety and/or trust (the authors seem not clear on this themselves) can be irrupted due to the LOA (Figure 4.4b). Despite their varying nature, similar strategies can be used to address uncanny (or unsafe) valleys. In the past, researchers and designers proposed two major strategies to do so. The first strategy can be summarised as preventing users from ‘falling into the abyss’ [237]. This means in particular that designs make sure that system designs only incorporate low to medium degrees of human likeness or LOA. Regarding automation, such strategies primarily aimed at keeping operators in the loop [83] and avoided taking away too much control from operators. While these approaches can show benefits, they also face the shortcoming that higher-level benefits and interactions are rendered inaccessible. In the automotive domain, for example, this would exclude highly automated driving. To make benefits on the other side of valley accessible also a second set of strategies have been proposed. In the case of automation, researchers suggested that it is not the particular LOA per se that causes irruption but rather the transitions between the levels (and side-effects that come along). Therefore, the strategies addressing this issue focus on securing those transitions. Transitions most fundamentally can go from 'left of the valley to the right' or from low to high LOA or vice versa.

For example, 'down' transitions[2] can be assisted by interlocked transitions [115]. These make sure that the operator is aware of the surrounding and wants to take over control. This can be done by requesting to conduct a low-risk task. To further make sure that irruptions are minimised, also transitions aim at landing in a low-risk state. After the successful execution of the low-risk task and the shift into a proper state, the interlocked transition is completed. Additionally, such transitions can be secured via a backup strategy by the system.

Creative Work

Application of automation and computation have of been intensely thought of in the context of industrial tasks. These can be characterised most profoundly, first, by a well-defined high-level goal that allows for further division into smaller actionable items and second by the possibility to properly parametrise the problem domain given the dedicated goal. Work in the creative domain, however, is not necessarily predicated on these assumptions. Thus, incorporating automation and computation is sometimes used for different motives or in diverging forms. Below, we illustrate why and how automation and computation still can be reasonably applied even in the creative domain.

Creativity is a phenomenon that is hard to put into commonly accepted terms. Various definitions can be found in the literature many of which emphasise different aspects. For this thesis, we will not dive too much into the depths of these discussions and therefore will only provide brief definitions and distinctions. Those are not meant to provide the definitional precision and all-encompassing understanding that the phenomenon and the surrounding discourse deserves. Instead, they are intended to provide a definitional minimum enabling the following discussion concerning computational approaches (while leaving the question of what precisely is creativity aside for now). Among the many ways one could try to frame creativity, we found the distinction between generative and adaptive as helpful. The proposition was introduced by Bown [20] and the definitions are quoted below.

Generative creativity: An instance of a system creating new patterns or behaviours regardless of the benefit to that system. There is an explanation for the creative outcome, but not a reason.

Adaptive creativity: An instance of a system creating new patterns or behaviours to the benefit of that system. The creative outcome can be explained in terms of its ability to satisfy a function.

[2] Shifts from a higher to a lower LOA (or from the right of the valley to left of it).

In general, generative creativity is often associated with evolutionary processes in biology, when machines generate alternative solutions by varying a set of affecting variables or when one seeks to generically examines new plants to determine their medical effects. More broadly speaking, generative creativity makes innovations possible, but can also be paraphrased as 'solution looking for a problem'.

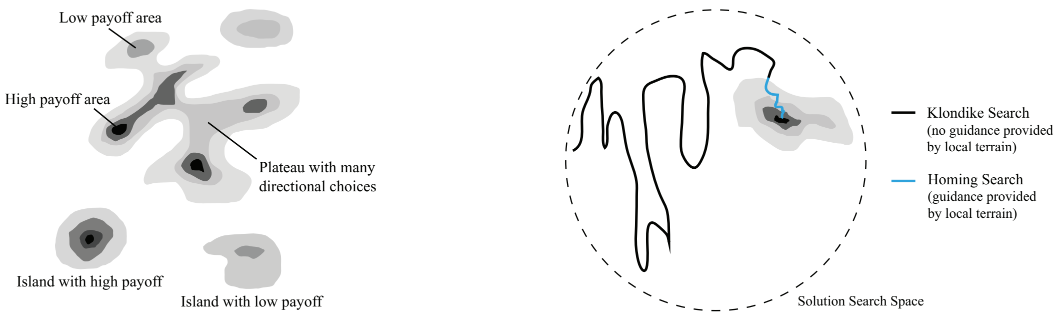


Figure 4.5: Research: A Perkins [211] Klondike space (left) and varying search strategies (right)

In contrast, adaptive creativity is often associated with human creativity. It faces difficulty in its evaluation and hardships in adequately defining it. It is disparate from generative creativity as it is intended to provide value for others in relation to a predefined goal. Therefore, it often follows the path of problem identification and subsequent solution generation. While the process sounds trivial, it is often hard to come up with the solutions. Part of what makes the process difficult is that local optimisation strategies do not necessarily work, partly due to the parametrisation problem. Thus, different strategies need to be applied to this problem domain. Fostering further understanding of such situations, Perkins [211] introduced a topographic analogy with Klondike spaces (Figure 4.5) to describe creative processes. He thought of them as an exploration through a territory of possibilities looking for regions of reward. For such explorations, he also identified some peculiarities. For example, existing solutions often need to be understood as a plateau which does not provide the necessary indication of where to look for better solutions. Thus, they can become an oasis that is hard to leave. Further, novel and better fitting solutions are rare and can mostly only be found in isolated places. Those are hard to discover as indicators of where one should look are missing. Also, global strategies, such as, most simply, brute force, are also not very useful as the space of possible (but not viable) solutions is too vast for linear search. In summary, it seems that one should think of *"creative search and discovery as an explorative process, as opposed to an optimisation"* [175]. This idea includes that it is common to take a path through a creative space that is not optimising incrementally towards one single fitness measure. It needs to be understood as a complex trajectory composed of intermediate goals. Each intermediate goal might hint at which pathway is to be taken next and is itself the result of a creative process. Regarding generative aspects, computational systems can generate more solutions in a more structured way than humans could. However, identifying viable and meaningful solutions is the domain of human strength. In the pursuit of creative exploration and the identification of intermediate goals, it is in human-machine cooperation that new and purposeful solutions might be found.

4.1.2 Workload-Unpredictability Trade-Off

As suggested in the sections above, the application of automation is not free from (side-)effects. Some are desired, as the benefits it provides, others are unwanted as habituation, skill-degradation or

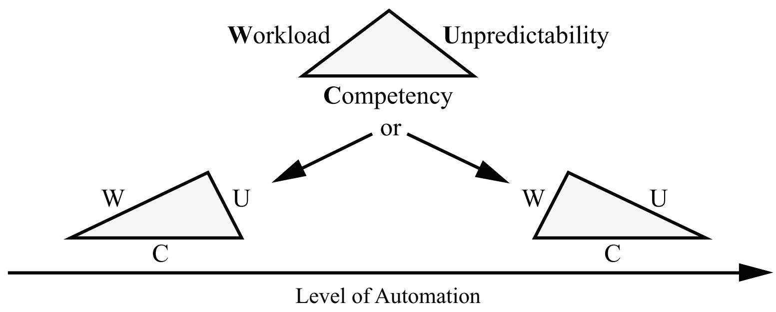


Figure 4.6: Research: Workload-Unpredictability Trade-Off based on Miller et al. [180]

newly arising safety issues. Besides the above mentioned, the chosen LOA is accompanied by further side-effects. Those are concerned with a trade-off between the reduction in operator workload and the (un)predictability of the results as suggested by Miller and Parasuraman [180].

Low LOAs are often associated with a high degree of operator involvement or even full manual control. Given a human-machine system with a sufficient level of competency to execute an assigned task, the more operators do manually, the more control they can exert over details during the process. Being able to execute a high-level task and to control low-level details allows operators to make accurate predictions of how the results of their work may look. However, human working speed and accuracy are limited compared to machines and hence higher LOAs might be applied to increase throughput and quality. With higher LOAs more control over the details is delegated to the system. This delegation yet has effects on the operators' ability to predict the results, therefore the unpredictability increases. This trade-off between reducing workload and increasing unpredictability (of the results) is inherent to automation and varies with the system design, especially the chosen LOA [179]. In Figure 4.6 instances of this trade-off are represented as triangles. Each side of the triangle is associated with either workload, unpredictability or competency. For simplification, we assume that competency is constant. The sides associated with workload or unpredictability and can vary in length. The longer the side the higher is the associated item. Depending on the LOA in the system design, the triangles or respectively the trade-off take various forms. Sometimes, also reduction of both can be achieved at the same time by a better overall design ('shortening the height of the triangle'). The increased unpredictability of the results itself can have side-effects on further aspects of a users experience. For this thesis, we foremost focus on consequences on the feeling and judgement of control as we will elaborate more upon in Section 4.3.

4.1.3 Human-Centred and Adaptive Automation

Early automation engineering often followed a technology-centred approach, mainly focusing on maximising automation as characterised by the idea of left-over function allocation. This idea was based on the premise that the approach would also maximise overall performance. Taking into account the negative side-effects mentioned so far, again further adaptations were considered. On the one hand, refinements such as LOAs were introduced. On the other hand, also a more fundamental paradigm shift towards human-centred automation followed.

Human-Centred Automation

This shift, for example, was advocated for by Billings [15] and aimed at fostering cooperation between users and machines to control and manage complex systems. One of the design goals was in particular to create systems that not only keep the operator in the loop [136] (to avoid the side effects of high LOAs) but also to provide them with meaningful tasks instead of a subset of complex to handle task fragments (echoing Fitts' initial propositions). Also, in the end, the outcome should be considered a team effort of an integrated human-machine system. Thus, it became additionally relevant to look for other evaluation criteria than performance alone as in early approaches. For instance, the operators' perceived workload or situational awareness became increasingly relevant.

Adaptive Automation

At the time of Bainbridge's work on the ironies of automation in 1983, systems mostly had a fixed LOA that was determined at design and implementation time. To account for shortcomings, subsequently further advances to the concept have been proposed under the terms of Adaptive Automation (AA) [204, 225, 226], flexible interactions [179] or dynamic function allocation [58]. Rouse [226] proposed the idea already in 1976 and introduced the notion that the LOA of a system could be changed during runtime. However, technology and systems needed to mature before it also became technically feasible and beneficially implemented in the 1990s. Since then several studies suggested that the adaptive approach is capable of minimizing workload, increasing performance [113, 137, 187, 217] and addressing downsides as skill degradation if designed properly [138, 203, 206]; mainly as it allowed adapting to the requirements of operators at different stages of a process. Ahead of the advent of AA, shortly summarised, existing methods so far answered two fundamental questions. First, "what should one automate?", which was addressed by function allocation. Second, "how much should one automate?" which was answered by the introduction of LOAs. "When should one automate?" was a question still unanswered with AA trying to provide answers. With its proposition, one of the primary concerns was to reconcile users with systems in a human-centred fashion even further. Instead of allocating functions to one agent in a fixed way, AA proposes to reallocate functions during runtime based on conditions. While there is a multitude of reasonable conditions [233], we briefly want to introduce four of them below.

Critical Events: Transition in function allocations occur when a critical event such as malfunctions are detected [114].

Performance: Given that human performance as in monitoring tasks is detected to decrease, a transition is triggered [203].

Psychophysiology: Data on the psychophysiologcial responses of an operator such as workload estimation based on Electroencephalogram (EEG) are used to infer a reallocation [32].

Behaviour Modelling: To ensure that actions carried out by users and machines conform to a predefined model, function allocation is adapted if necessary [227].

Studies on AA used for the control of complex systems suggested that they are particularly useful for certain task types and with certain durations of allocating control either to operators or the system. Parasuraman [203], for instance, found that performance in task monitoring could be increased by periodic allocations between operators and systems when compared to static automation (every 10 min for 10 min). Complementing the findings, Hilburn and colleagues [114] found that AA also performed better than purely manual monitoring.

Interaction of Levels of Automation and Adaptive Automation

However, AA studies have mainly been concerned with function allocation at the extremes. Comparisons were often drawn between high versus low LOA implementations. As sketched out earlier, LOA yet is a gradual concept. Including the missing intermediate LOAs in studying an AA system, Kaber and Endsely [137] determined the arising interaction effects in a dynamic control task empirically. Given an objective incorporating a primary and a secondary task and a system with varying combinations of LOAs and AA, they collected data on task performance, workload and SA of operators. Analysing the gathered data, they reported the following effects. Overall, the LOA seems to be predominantly responsible for performance and SA. Notably, performance in primary tasks was stronger influenced by SA than occurring workload. The variations in the AA strategy have only little effect on primary task performance. However, they affect the secondary task performance and perceived workload (as more of the primary task became automated). Further, the authors suggested that the effects of combining LOAs and AA are not *"additive"*. The best-performing LOA and the best-performing AA level in combination, did not add up to the best overall performance. It rather seems that one should choose a LOA × AA combination based on which aspect one wants to maximise. Regarding the performance of the human-machine system, best results seemed to be achieved at a low LOA (3) with a high AA level (12 min on and 8 min off), even when compared to full automation. Human ability in strategising and planning here seemed to a substantial contributing factor. However, the combination also resulted in lower SA (at modest workload). In contrast, higher levels of SA were found at combinations involving intermediate LOAs. So in conclusion, it seemed that only one of the parameters can be maximised (if not compensated for as by varying LOAs across the AA cycles).

4.2 Camera Control

We sketched out the developmental timeline of cinematographic support tools, user interfaces and their application in crafting camera motion in Chapters 2 and 3. While science often informs the underlying technologies and designs, so far we did not provide more information covering the research literature itself. Utilising camera motion is not limited to cinematography as various lines of computer science and HCI research show. In general, the literature provides fundamental insights for virtual environments as in VR, and game design applications, but also for physical environments such as teleconferencing or CV informed motion control. Most commonly it is probably used in the domain of 3D interaction. In this section, we thus want to *introduce fundamentals of camera control and research on virtual and physical environments* focusing on 3D interaction and cinematic application.

4.2.1 Virtual Camera Control

For the conceptualisations and working processes, we described in Chapter 3 there is also a counterpart in the literature originating from the field of automated 3D graphics generation. Here, Seligmann [239] introduced a four stage pipeline regarding graphical presentation as detailed below:

Generation of Communicative Goal: What is the intention of the image? What should it accomplish? The intention, for example, could be to explain, to convince or to call to action.

Selection of Presentation Strategy: According to the determined goal, a strategy and visual effects are selected. For example, the strategy could be to put visual emphasis on a particular object (no specification is given how this might be achieved).

Selection of Presentation Method: Given the goal and the strategy, the means of presentation are determined. In the provided example, additional light sources could be installed to shed a spotlight on the object to emphasise it.

Image Generation: Based on the presented preparation, the visual representation is created. This process includes camera (and lighting) work that satisfies the visual properties best, but also the arrangement of the scene and selection from alternative but similar shots.

The vocabulary used by the experts as found in our studies differed from Seligmann's list, yet the underlying concepts seem quite alike. In particular, the first two points address the preparation phase and the motivation behind a shot as discussed in our seminar visits in Section 3.3. The third point is closely connected to the selection of the tools and the shots types they afford as shown in Section 2.2.1. The last point is mostly concerned with controlling the actual movement. Regarding the methods of control, some styles and techniques were already briefly introduced in Section 2.2.3 and will be further expanded below.

Camera motion can also be thought of as an interactive process in a dialogue between users and systems. In this regard, Jankowski and Hachet [129] provided a survey of interaction techniques. They see three major domains of interaction in a 3D environment: navigation, selection/manipulation and system control. As only navigation is primarily concerned with camera control, we will only present those aspects in detail below. In general, navigation can be characterised as the motor task of viewpoint manipulation which also includes the cognitive task of wayfinding. Different researchers such as Mackinlay et al. [168], Bowman et al. [19] and Tan et al. [255] have proposed several classifications of camera motion. Jankowski and Hachet took those classifications and mapped them onto each other so that one comprehensive list emerged as found below. It is based on the work of Mackinlay et al. and integrates the propositions of Bowman et al. and Tan et al.

General Movement: Used for exploration as in walking through a scene (is associated with the exploration goal of Bowman et al. and Tan et al.).

Targeted Movement: A target is specified, and the movement aims at reaching the target as in getting closer to a detail of a model (is associated with the search goal of Bowman et al. and Tan et al.)

Specified Coordinate Movement: A position and orientation is specified, and the movement aims at reaching the specific position of that viewpoint as reaching a particular POV in a scene (no association with the work of Bowman et al. and Tan et al.).

Specified Trajectory Movement: Motion is carried out on a position and orientation trajectory as in cinematic camera movement (is associated with the inspection goal of Bowman et al. and Tan et al.).

For cinematography, specified trajectory movement seems to be the predominant motion style. However, as identified in our expert interviews in Section 3.5, operators need to be able to improvise. Therefore, it is vital that tools also support general movement features (additional details will also be discussed in Section 8.1). To cover both domains, we will first present an overview of general movement techniques that is followed by techniques for trajectory movement.

General Movement

Jankowski and Hachet see three areas of control regarding general movement: rotate-pan-zoom control, screen-space methods and walking/driving/flying metaphors. The first area consists of fundamental techniques such as direct control over rotation and translation as already mentioned in Section 2.2.3. Please note that a pan in a cinematographer's vocabulary might refer to a rotation whereas the authors here refer to a translation (which this thesis refers to as a slide). Also, zoom is referred to as a translation (which this thesis refers to as a dolly shot).

Beyond the very fundamentals, also additional approaches were proposed. For instance, Gleicher and Witkin [100] introduced screen-space methods with their so-called through-the-lens techniques. Their propositions are based on a different paradigm than the traditional rotation and translation control. It allows users to steer the camera by interacting directly with the image and the objects rendered on the user's viewport. In their system, users can, for example, select parts of an object to define a constraint. Two of such points can determine an axis around which a later rotational move ▶ revolves [536]. Their work inspired further work as by Christie and Hosobe [52] or Reisman and colleagues [219]. Christie and Hosobe extended the approach by using virtual composition primitives and the ability to also change the lighting of a scene by interacting through the viewport. Reisman ▶ and colleagues translated the initial idea to gesture input for multitouch devices [420]. For instance, two figures of one hand resting on a surface would define a custom axis. Interaction with the other hand would trigger operations such as rotating a touched object around the previously defined axis.

Further, also various techniques based on metaphors such as walking, driving or flying were proposed for navigation through a virtual 3D environment. The walking metaphor, for instance, is commonly used and most widespread in the form of a first-person perspective in computer games. Lécuyer and colleagues [149] studied it and found that when moving the camera position, a general oscillating motion in combination with a compensating motion enhanced the experience of walking for users, especially when the focal point would be kept in the same place (similar to human oculomotor compensation). This application reminds us very much of the notion of hand-held first-person perspective or POV shots that was introduced in Section 3.1.3. Especially in the form of affected POV shots, where the camera movement would be coupled to a character's emotional state that is used particularly to increase intensity or to invoke a sense of presence. Multiple such metaphors could also be used in combination as Drucker and colleagues [74] show with their CINEMA system.

Specified Trajectory Movement

The introduced techniques so far allow users to roam around freely. In a cinematic context, this at times can be unwanted. Therefore, more guiding techniques were proposed. Those limit the user freedom to a certain extent. However, the upside is that when implemented properly this approach helps to present (only) relevant or compelling information or objects, to provide an overview of the scene and easy to learn paths. The latter can help to avoid disorientation and address the *"lost in cyberspace"*-problem [129]. Guided or constrained navigation was introduced by Galyean [95] with the river analogy that aimed at a more narrative presentation style. Her primary suggestion was to think of the navigational pathway as a river flowing through a landscape with the user onboard of a boat that floats along with the current. On the one hand, the user is mainly moved by the current, on the other, there are also some control options a user still holds such as getting closer to a particular detail in sight. Beyond Galyean's implementation in a museum exhibit, several systems subsequently ▶ incorporated her initial idea as the StyleCam system presented by Burtnyk and colleagues [26, 526].

The system focused on presenting details of car models in a visually appealing way. The idea was later refined as with the ShowMotion [27, 518] system which was intended for instance for presentations where it should allow to adapt, for instance, to questions from the audience.

Cinematic Movement

Additionally, techniques specifically tailored to (virtual) cinematography have been put forward. Those often make use of existing cinematic techniques as presented in Section 3.1. For instance, Christianson and colleagues [51] transferred existing practices into a formal language, the Declarative Camera Control Language, that allows specifying shots by the actors' positions and movements on the screen abstractly. Similarly, abstract high-level control of camera positions and transitions have been used in the work of He and colleagues [110]. In their virtual cinematographer system, they aimed at automating virtual camera placement that is informed by often used cinematic solutions, so-called idioms. Here, users could freely join and roam around an online in a VR environment. For rendering a visual representation, (virtual) cameras were automatically placed and transitioned between visually appealing spots based on abstract instructions. This way, for instance, conversations between two or three people were framed. Similarly, automatic camera placement that aims at framing the player avatar in an unoccluded way informed by cinematic idioms was implemented in games as by Li and Cheng [152]. Jankowski and Hachet described techniques for camera control in 3D environments from a more general perspective that would also entail cinematic use. In contrast, Christie and Olivier [53] presented a survey focussing primarily on camera control for cinematographic applications. What is noteworthy about their survey is their problem estimation and reasoning on issues in camera control, classification of control techniques and their remarks on the importance of tool expressiveness. In their estimation, there are three main issues in camera control. First, it is hard for users to operate all seven degrees of freedom at once (three for translation, three for rotation and one for the focal-plane). Second, when thinking about camera movement as path planning, it is a PSPACE hard problem the complexity of which is exponential in its degrees of freedom. Third, the framing is dependent on the content of the scene which includes geometric, perceptional and aesthetic qualities that are hard to model. To satisfy such requirements, tools need to provide a high degree of expressiveness. To characterise the surveyed tools, they classified control techniques as found below:

Interactive Approaches: Those fundamentally provide a mapping from the dimensions of UI devices to camera parameters. Their complexity depends on the usage.

Reactive Approaches: Camera movement is executed by a system that is driven by a continuous control loop that is informed by the image stream.

Generalized Approaches: Higher-level control that allows declaring the motion path and properties abstractly without requiring direct control.

For Christie and Olivier expressiveness is a central aspect that such tools should provide. It *"describes the possible properties offered to the user to control a virtual camera and in such characterizes the declarative power of the camera control system"* [53]. While there are many ways that system could support expressiveness, they see four major common aspects: the range of properties, the nature of properties, the required level of abstraction of the scene and the extensibility.

4.2.2 Physical Camera Control

Variation in the ways of camera control is not limited to virtual environments. Also in physical environments cameras with attached MoCo systems are put to use in various forms. Some example UIs were already presented in Section 2.2.3. Moreover, Chen and Carr [46], for instance, compiled a survey in 2014 detailing the developments in autonomous systems made in the past twenty years. In their examination, they distinguish between three vital steps that nearly all systems incorporated: planning, controlling and selecting. Similar to our remarks on the automation literature in Section 4.1, these steps can be understood within a framework of questions that a system addresses. So, the question of "Where should a camera look?" is essentially associated with planning. "How should cameras move?" is addressed in the domain of control. Finally, when considering multiple cameras placed in a scene "Which camera stream at what point in time should be displayed to the audience?" is what is addressed in selecting. What Chen and Carr also lay out in more detail is that for planning and selecting, the fundamentals and techniques are quite similar in virtual and physical environments, however, when it comes to controlling, virtual and physical cameras differ substantially. This difference is in parts due to the fact that real cameras require feedback loops that incorporate sensor data that can be noisy at times. In contrast, noise is absent in virtual cameras which additionally have by default access to additional information such as hierarchical and semantic properties of the scene as well as future goals or states. Regarding user roles presented in Section 3.1.1 planning and selecting can be understood as the tasks that are done by directors and DOPs in the pre and post-production phase of movie production (or live in a TV broadcast) while the operators working the cameras handle controlling.

Planning

Planning strategies have been implemented in various systems. For example in stationary PTZ cameras or remote heads. In this case, what any algorithm should provide as a result on a technical level are pan and tilt angles as well as zoom factors. How those are derived can vary depending on the system architecture and CV features. Pinhanez and Pentland [214] provided one of the earliest of such autonomous systems for a cooking show. Here, wide-angle cameras were used to provide an overview of the scene. A processing unit used the video streams and was provided additional context with the script. Based on these pieces of information events (expected based on the script) were identified via CV. High-quality TV cameras were then subsequently instructed based on the derived understanding of the scene. A human TV director, however, would still instruct the system 'where to look' by calling the shots via a speech interface (*"close-up on chef"*). Similar systems have been used in various contexts ranging from an autonomously moving robot photographer [30, 31] for social events to lecture recordings [173, 280, 284]. What is common across these approaches is the use of a static physical camera in the architecture that informs a further dynamic component for creating the results. While Yokoi and Fujiyoshi [284] use a virtual camera[3] to frame a subregion of interest, for instance, Wulff and colleagues [280, 281, 495] drive a dynamic PTZ camera to render the resulting material. While Pinhanez and Pentland still relied on a human operator, Wulff and colleagues additionally implemented a virtual director component that would take the planning decisions based on information provided by a tracker module overlooking the scene. Besides lectures which can be characterised as a somewhat static environment, also more complex situations have been taken on as various examples in automated sports recording show. Ariki and colleagues [6] tracked the position of soccer players and the ball using a stationary High Definition (HD) camera. Clipping results to Standard Definition (SD), a virtual camera was used to create the resulting shots including virtual panning and zooming. Planning decisions were derived from the tracking and fused with event detection (as for free kicks). For the various situations, a fixed set of framing rules was implemented in the system. Further extending the underlying approach, Chen and De Vleeschouwer [45] presented a system where a network of high-resolution cameras placed at different spots of the court was used to cover a basketball game. Again, one camera would overlook the overall court and gameplay. Based on an analysis of its stream, a directing component would select a particular camera as the image source for the broadcast. Similar to the work of Ariki and colleagues, a subregion of the whole camera stream was picked to provide a proper framing that focused on relevant and salient features. For broadcasting basketball also further approaches have been taken as by Carr and colleagues [38]. Instead of one or multiple stationary physical cameras, they used a single dynamic PTZ camera. In their system, they use an approach they refer to as a hybrid camera. First, the planning component would orient the camera towards a purposeful and aesthetically informed general direction. In a second step also a virtual camera would further refine the framing as already implemented by preceding systems.

Controlling

In autonomous systems, moving the camera from A to B is directed by the planning algorithm and often takes the form of a regulatory process. It needs to meet the goals of smooth motion between those points at an appropriate speed while maintaining precise timing throughout a sequence. Besides their technical and regulatory nature, the underlying algorithms can additionally be informed by patterns of human operation. For example, Kato and colleagues [141] identified such characteristics for operators

[3]For clarification: Virtual cameras can also be used outside fully virtual environments based on actual recorded (not rendered) footage. Subregions with the dimensions of the results can be defined within the confines of the source material given a recording with a stationary camera and higher resolution than the resulting material. For instance, panning moves transitioning between those subregions can be rendered post-hoc. This process results in similar on-screen effects than physical motion although none was carried out.

in cooking shows and sports programs. They found that for panning moves different speed levels are appropriate. For easing into a motion, speeds can be higher than for deceleration when easing out of it. This is interesting as in an uninformed design it is likely that one would set the speed for easing in and out at the same value by default. In a follow-up study, Kato and colleagues [140] further suggested that this is mainly true for simple moves. For more complex moves that incorporate the following of a talent who moves unpredictably, those rules do not necessarily apply. Further, the control of virtual cameras used on actual footage and motor-driven cameras also differs. While virtual cameras are rendered post-hoc, smooth camera motion does not need additional hardware but more importantly can be achieved by rendering algorithms that have access to all the material. This access makes it possible also to look up future states and hence to optimise the motion path such that proper framing is maintained throughout the motion. In contrast, physical actuation is based on a feedback loop that has only access to past states and the current frame. Control and steering approaches based on this set of information are also described in the literature on robotics and referred to as visual servoing (see also Section 4.2.3). The underlying fundamental principles of such approaches is that the algorithm established where the camera is looking, where it should be looking and incrementally tries to minimise the delta between these states.

Selecting

As shot selection is not the primary concern of this thesis, we will only touch it briefly. Decisions on "which camera stream should be broadcasted when?" should be predicated upon the communicative goals and aesthetics aspects. To increase the quality of the results, some systems use rules derived from human operators. For instance, Liu and colleagues [162] interviewed experts and determined rules such as *"avoid jump cuts"* or *"switch after a maximum duration"* that were implemented in their lecture recording system switching between three POVs. Execution of such rules can also be derived from low-level tracking data as in the case of Doubek and colleagues [73]. They used multiple fixed cameras to record a talent walking through an office. Further, they derived a set of rules from cinematographic literature as *"use a long shot while the talent moves"* or *"switch to a medium shot while the talent stops"*. A viewpoint score was determined based on those rules and the tracking data, that would lead to the selection of a particular camera. Also, other approaches informing the choice have been tried. Wang and colleagues [268] used a Hidden Markov Model for a soccer broadcasting system. The same technology was also applied by Chen and colleagues [45] in a system for basketball games. They also let users select salient objects, the size and visibility of which they then tried to make part of the process. More recently, Chen and colleagues [44] used an approach that was only data-driven. They trained a random forest classifier for a field hockey application. The system would then recommend the best option to a human director. Older broadcasts were used as training material for the classifier featuring ball visibility, player positions and camera PTZ settings. The classifier could further be trained to offer also various choices regarding the directing style.

4.2.3 Computer Vision and Motion Control

Except for full manual control, all higher LOAs provide some form of assistance by a system. As summarised by Chen and Carr [46], for real-time physical camera control, this is often predicated upon a particular control or feedback loop. This loop most commonly incorporates some form of sensing (most fundamentally the visual stream of the camera, but also other modalities have been used in combination with it), some form of (cinematic) reasoning and some form of actuation. How camera travel could be informed by cinematic attributes was already sketched out earlier in Section 4.2. Thus, we focus more on sensing and actuation below. From a technical perspective, CV for sensing and MoCo for actuation are commonly at the heart of such systems. Therefore, we want to expand a fair bit on the underlying technical details of these core components based on selected literature. In general, robotic servo control (a term similar to motion control) can be differentiated into position-based or image-based visual servoing approaches as proposed by Chaumette and Hutchinson [42, 43]. Referring to similar phenomena yet applying a different vocabulary, Bonin-Font and colleagues [18] distinguished between map-based and mapless navigation. For the following remarks on such techniques, we will use the terms suggested by Bonin-Font and colleagues. Also, as navigation in reference to a coordinate system might not be the primary use case for cinematic camera operation, we thus will present mapless navigation techniques more in detail. According to Bonin-Font and colleagues, mapless navigation majorly consists of reactive techniques incorporating visual cues. Such systems do not require a previous understanding of the surroundings and derive decisions from their perception of the unknown environment. The cues can originate from a visual analysis of the camera stream across multiple frames while simultaneously predominantly no global representation exists describing the environment. The authors further distinguish between the following techniques for mapless navigation.

Optical Flow-based Navigation

A set of features or objects are traced across a sequence of image frames. This way, apparent motion of such features can be derived. This image analysis technique was pioneered by Horn and Schunk [125] as well as Lucas and Kanade [166]. For two consecutive images, the direction and magnitude of movement regarding translation and rotation of a particular feature are computed. The resulting values are often stored in a vector for each pixel. Its direction describes the movement of the pixel across the images and its norm is affected by the speed of the movement. To provide a visual representation, the resulting vectors are often displayed on top of the visual material [431, 505]. To minimise computational costs, sometimes the optical flow is only calculated on priorly determined image features such as corners or edges [107, 244]. Based on the way it is calculated, the own motion of a system is perceived as relative to the field of view. Consequently, (real-world) static objects seem to move in relation to a moving system. Vice versa, for static cameras, it can also be used to identify moving objects [63]. In terms of cinematography, it can further be used for post-hoc software

▶ video stabilisation as in the SteadyFlow system presented by Liu and colleagues [163, 426]. The approach has also inherent limitations. UAVs steered based on optical flow approaches run into problems when flying at low altitudes or high velocities as the accuracy of the optical flow estimation decreases. Srinivasan and colleagues [247], for instance, addressed this issue by introducing a mirror in front of the camera. Due to its placement, the authors could reduce the speed of movement and hence increase accuracy again. Also, multiple cameras can be used in combination. This approach is often used for so-called insect-like steering and motion control of UAVs [248]. For instance, it can mimic the behaviour of bees by comparing the magnitude of the optical flow of a left and a right camera looking at walls on each side. The steering behaviour subsequently aims at harmonising the flow magnitudes and hence steer the UAV in the centre between the walls.

Appearance-based Navigation

For this method a two-step process is necessary. First, a set of images is recorded and stored. The images are subsequently used as templates and annotated with localisation information or control commands for each location. In the second step, the robot navigates autonomously. To do so, it needs to self-localise by comparing the stored images to a currently captured image. For the self-localisation basically, two approaches can be followed: model-based and view-based. The former uses previously defined models of objects to identify features in the environment. The latter uses no additional information and hence only image matching algorithms are applied. Examples can be found in the work of Matsumoto and colleagues [172] or Jones and colleagues [132]. As displayed in the ▶ work of Bürki and colleagues [25, 439], appearance-based approaches can, for instance, be used for the steering of a car. The approach, in particular, is also reported to be more resilient towards changes in appearance due to varying illumination.

Visual Characteristics Extraction-based Navigation

Robot navigation that also includes obstacle avoidance can alternatively be based on the extraction of qualitative image characteristics. Similar to appearance-based approaches, one can differentiate between systems that are model-based (relying on prior information) or sensor-based (without using prior information). In general, these systems often follow behaviour-based steering approaches ▶ as explored by Arkin and colleagues [7, 445]. Due to this approach, they rely little on abstract internal representations of the environment. Mostly they base steering decisions on sensor input such as distances, location coordinates, speed or contact time with obstacles. Sometimes also information about past events is recorded and informs subsequent decisions. This principle of a bottom-up closely coupling of sensor input to action selection is sometimes also referred to as subsumption architecture. While the architecture has its benefits regarding real-time action selection and execution especially for low-level tasks, it faces difficulties in incorporating higher-level goals and abstractions by design.

Feature Tracking-based Navigation

Further, navigation techniques can use feature tracking as a foundation to derive steering decisions. The techniques are predicated on the idea, that particular features can be extracted from the visual stream such as corners [89, 107, 224, 244], edges [34, 68], ridges [106, 144], blobs [160, 275] or even reasonable suggestions on the scale of unknown objects [159] that help guiding the steering behaviour of the system. Generally, systems divide the tracking task into two subtasks: motion detection and feature matching [262]. First, in motion detection, a region in the following frame is detected where a selected feature is likely to appear. Second, the selected feature is tried to be identified in this particular region. Additionally, a further (separate) component needs to take the information into account, derive steering instructions and translate them into executable actions.

Various techniques have been proposed and will be outlined below. For example, Pears and Liang [209] used homographies to track corners of the ground plane in an indoor setting with their proposed H-based tracker that guides steering instructions. Feature tracking per se does not offer any form of obstacle avoidance. Based on their H-based tracker, the authors could also incorporate this characteristic in the steering decisions [153]. As mentioned above, for algorithmic navigation of UAVs at high velocity is a tricky task. Addressing this issue in an alternative way than Srinivasan and colleagues, Ollero and colleagues [201] proposed a tracking algorithm that is based on a homography matrix that compensates the motion of the systems and that also detects objects to inform means of avoidance. To advance the underlying feature extraction algorithms robustly and efficiently was the aim of many researchers. Regarding such developments, one of the more outstanding techniques that have been proposed was the Scale Invariant Feature Transform (SIFT) method by Lowe [165]. As the name suggests, these methods are not affected by image scaling. Further, they are also quite robust against illumination and changes in the POV of the camera. Due to its robustness SIFT can be not only navigation but also for global localisation [238]. In parts, inspired by the SIFT approach is the Speeded Up Robust Features (SURF) technique proposed by Bay and colleagues [11]. The SURF algorithm can be used for similar purposes but is claimed to perform faster and more robust regarding image transformations than SIFT [143]. More recently, further advances were proposed such as Oriented FAST and Rotated BRIEF (ORB) [228]. The algorithm aims at faster performance than SIFT but is only unaffected by rotation. It further builds on the previously introduced algorithms FAST [224] (for corner detection) and BRIEF [33] (for feature detection). Additionally, a further approach was proposed with KAZE (and its accelerated A-KAZE variant) [2, 3, 434, 489]. Chien and colleagues [48], ▶
conducted a comparison of the mentioned algorithms. Their results suggested that while ORB is the ▶
cheapest to compute, it is also less accurate. The highest levels of accuracy can be achieved by using SIFT or SURF. However, those are (comparatively) computationally expensive. In contrast, A-KAZE seems to be fairly able to balance accuracy and computational costs.

Computer Vision Capabilities and Applications

Since we established some fundamental aspects of CV approaches, we further want to give pointers regarding the question "what can these algorithms accomplish?" Therefore, we briefly provide a selection of exemplary contemporary systems. Those showcase applications that are based on such techniques as introduced above and that also potentially could find broader application in cinematic contexts. As we will further discuss their role in potential future developments in Chapter 9, we also present systems that are concerned with cinematic aspects beyond camera motion.

Face Detection and Hand Gesture Recognition

As Karam [139] pointed out, hand gestures are particularly often used to communicate compared to other body parts. Therefore, they represent aspects of naturalness in human-human as well as human-computer interaction. Hand gesture recognition thus can also contribute to the control of MoCo support tools. Detailing the fundamentals and approaches to visual recognition of hand gestures, Rautaray and Agrawal [218] recently presented a survey. Similar to other CV techniques, hand recognition can also be distinguished between model-based and appearance-based approaches. Model-based techniques inform an internal 3D model of a hand by an analysis of the image stream, while appearance-based approaches only rely on information they can extract from the visual stream. In implemented systems hand gestures are often used as a substitute or additional modality to previously existing input techniques as mouse and keyboard for example. Depending on the application domain, Rautaray and Agrawal suggested that they can fulfil varying purposes. For tablet applications, they primarily aim at providing an easy-to-use and independent interface. In contrast, for games, medical applications or in Augmented Reality (AR) environments, the goal is to control the movement and orientation of the POV or to manipulate objects. Further, in mobile and pervasive computing, the intended purpose is to provide eyes-free interaction and hence to enable users to focus their visual attention on a primary task. The three mentioned motivations can also be transferred to UIs of cinematic MoCo as far as we are concerned. Gestures can be used for system control as in signalling a UAV to return to the operator or to instruct how to frame a shot. Applicability and implementing of ▶ this idea have been investigated as by Cauchard and colleagues [39, 466] who present the result of an elicitation study that derived a gesture set for drone control. For instance, Nagi and colleagues [191, ▶ 542] implemented a drone system that is controlled via a combination of hand gestures that are identified by colour tracking and face poses that are derived from the OpenCV [21] implementation of the Viola-Jones face detector [267] (which is based on a cascade Haar classifier). Further, Monajjemi ▶ and colleagues [182, 483] used CV for face detection and hand gesture recognition for the control of multiple UAVs. For face detection, they also used the OpenCV Viola-Jones face detector. Gesture recognition was implemented by detecting the magnitude of optical flow in fixed regions near the face ▶ of a user. Notably, Kyriazis and Argyros [146, 515] presented a technique for hand tracking that can even differentiate between individual fingers and that can handle occlusion when one hand is in front ▶ of the other. Similar techniques were integrated into commercial products as the DJI Mavic [463].

Face Recognition for Talent Identification

While face detection as briefly mentioned above allows identifying a (not further specified) face in a camera stream, it does not allow identifying and track a particular one. In a cinematic context, it is necessary to be able to focus on or dolly-in on a particular actor or a particular (star) player in a sports broadcast. For humans, this process of identifying an individual face is fairly straightforward and hence it might seem like a simple problem to address, however from a technical perspective it is complex. Fundamentally, a generic system needs to cover three main areas: face detection and tracking, feature extraction and face recognition [285]. Face detection is used to decide whether there is a face at all in the material. Several techniques can be distinguished, for example, Yang [282] and colleagues provide a classification that is well accepted (into knowledge-based, feature invariant, template matching and appearance-based methods). Subsequently, detected faces need to be tracked across the sequence of images a video stream. This can also be done in various ways such as head tracking, feature tracking, image or model-based tracking. Zhao and colleagues [285] also provide a classification for these types of techniques. The idea behind the step of feature extraction is to get a set of parameters that allow identifying an individual face (with an acceptable error rate). Various feature extraction algorithms were already discussed above. When a set of features is extracted it is necessary to classify the image. This process allows to match the feature set and appearing faces in a camera stream. The matching can be hard due to challenges like changing illumination, facial expressions or head poses. Therefore, learning algorithms are involved in aiming at a more accurate and robust classification. The implementation of face recognition can leverage existing high-level APIs as the free OpenFace [9, 457] software or commercial products by Aimetis [308, 437] or Firefly [329, 476]. ▶ Such systems can also be informed by existing images of an actor so that a feature set can be extracted ▶ from an offline image and compared to the detected faces in a stream. ▶

Pose Estimation

While a rich body of work is concerned with face detection and recognition, also other CV approaches are pursued. In cinematography, a full-frame view of an actor's face that could provide a good foundation for face detection might be generated by a close-up shot. However, for reasons of storytelling also other shot types might be used as medium or long shots that display most or all parts of the body of the actor. In such situations, also other CV approaches are viable. For instance, we want to focus on deriving the human pose based on the visual material below. A survey on pose estimation was compiled by Moeslund and Granum [181] describing the underlying techniques. Similar to face recognition pose estimation is a multi-step process that consists of initialisation, tracking, pose estimation, and recognition. Also, similar to previously mentioned techniques, pose estimation can be informed by an underlying 3D model or be derived purely from the visual data. Applying an approach that is based on optical flow, Zuffi and colleagues [286] describe a system that identifies upper body poses and can differentiate between various body parts such as arms or the chest. They used it to extract poses from existing material as, for example, the TV series *Friends* (1994-2004) [470]. ▶

Camera Motion Estimation

Not only items visible in the images can be detected, but also the 3D motion path of the camera itself. Tools as the Voodoo Camera Tracker [385] provided freely by the University of Hannover extracts features using the Harris corner detector [107] and SIFT to calculate to camera motion path of a given image sequence. The derived motion paths can then be used to inform the generation of computationally proposed paths as those often face certain issues. On the one hand, a general audience is accustomed to the camera paths of human operators and computationally derived ones differ from those in their fundamental structure. On the other, concerned rather with details, the aesthetics differ as real motion is to a limited degree jerky, shaky or varying in speed. For instance, Sanokho and colleagues [232] used the mentioned approach to further refine virtual camera paths to incorporate a higher degree of realism. In our estimation, the approach could also be used to inform the motion paths of physical camera motion to counteract motion that is perceived as too robotic.

Emotional State Estimation

In addition to the aforementioned aspects, there are also qualitative aspects that one can extract from a video. Researchers in medical and psychological domains as well as researchers in the domain of affective computing are particularly interested in identifying emotional states. For instance, for ▶ the psychotherapeutic context, DeVault and colleagues [70, 251, 520] presented a system that aimed at deriving emotional states from the non-verbal behaviour of a person. The states are estimated by sensing facial expressions, body posture, acoustic features, linguistic patterns and higher-level behaviour descriptors (such as fidgeting). Further, in the medical supervision of premature infants, doctors want to check the heart rate, however, attaching devices to the infants in the incubators is ▶ unwanted. To derive the heart-rate alternatively, Wu and colleagues [279, 510] provide a technique they refer to as video magnification. Here, some aspects such as slight changes in skin colour due to a heart pulse, are amplified and hence made visible to a viewer or made detectable to a system. Their technique was also used on cinematic footage as scenes from *The Dark Knight Rises* (2012) to identify whether the actor was merely simulating emotions or experiencing them while acting a ▶ dramatic scene [501]. As previously introduced, certain operation techniques try to link the camera motion to the emotional state of the actor. For systems, therefore, a way of sensing the emotional state would be necessary. Its implementation could be done similarly to the discussed examples.

Content Reenactment

In cinematic post-production, it is common to dub the original tone for distribution in foreign countries. Given the different pronunciations, this often leads to asynchronous situations of the facial expressions and the tone. Remarkably, it also possible, to change the facial expression and articula- ▶ tion of a person as Thies and colleagues showed [259, 475]. They used the face of another person as a source to extract the relevant features and mapped those to the face of the original person. This can be carried out in real-time, hence, what a person says on-screen can be changed even in a live broadcast.

4.2.4 Human-Machine Interplay

As outlined in the Section (4.1), automation and creative work can be perceived as contradictory as well as complimentary at times. Thus, balancing benefits and downsides is not trivial. Further, creative work is also strongly domain-dependent. Therefore, top-down generic solutions need to be adapted to the domains and individual solutions for particular situations need to be developed

(a) TopoSketch from [272] *(b)* Director's Lens from [161] *(c)* Stype kit from [432]

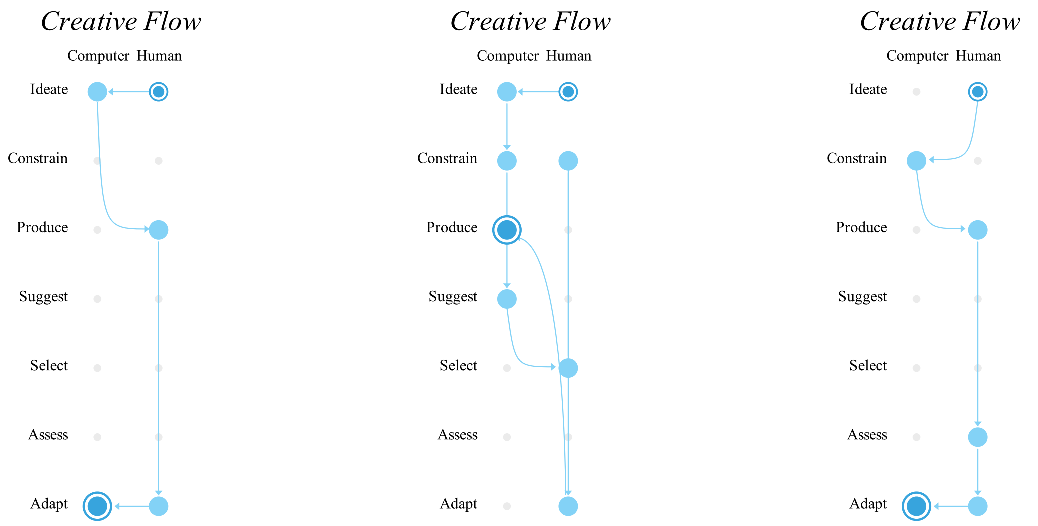
Figure 4.7: Research: Mixed-Initiative Creative Interfaces (MICI) example systems that focus on cinematic tools

bottom-up. A similar situation arises when Artificial Intelligence (AI) is used in addition to automation. For systems incorporating AI in creative work, Deterding and colleagues [69] coined the term MICI[4]. For the authors, such systems are characterised by a close interactive loop incorporating both agents, users and systems, where the systems contribute non-deterministic input to a creative process. As of the domain dependence, a variety of MICIs emerged independently from various fields in the past. On the one hand, this makes it somewhat hard to keep track of such systems. On the other, the uniqueness of the approaches makes it also slightly hard to compare such systems. To address both issues, Deterding and colleagues compiled a library that is available electronically [351] collecting MICIs and characterising them based on a structure of (also proposed) shared traits of the creative flow. For each tool, they identified where the initiative originates from (human or system) across a set of subprocesses. The selected subprocesses of interest were: ideating, constraining, producing, suggesting, selecting, assessing and adapting. Given the particular pattern collected for one tool, they visualised its characteristic as a graph.

The library also captures systems for film production or animation (Figure 4.7). For instance, in TopoSketch by White and Low [272, 537], a user provides an image of a face that is then processed by neural networks that offer possible animations. It displays the animations to the user in a grid who then can hover over the grid items. The motion path of the mouse will then be taken as a basis for generating the resulting animation (Figure 4.7a). Given the above-introduced traits, this system can be characterised along the following lines. A human initiates the ideation, and the system contributes to it. Subsequently, the system relies on human input for producing the results. Then both agents are involved in a final adapting subprocess before the system provides a result (Figure 4.8a).

[4] Prior, Yannakakis and colleagues already introduced to a similar term with mixed-initiative co-creativity [283].

However, the items found in this library hardly focus on camera motion (at least at the time of the writing of this thesis). To provide information also on systems that focus on such a human-machine interplay in creative cinematic co-creation, we want to present briefly two (non-AI) systems in more detail below and characterise them harnessing the proposed structure.



(a) Creative flow of TopoSketch *(b)* Creative flow of Director's Lens *(c)* Creative flow of Stype kit

Figure 4.8: Research: Analysis of the creative flow of the tools presented in Figure 4.7 based on Deterding et al. [351]. A darker blue dot marks the start of the overall interaction process, light blue dots mark subprocesses taken on by the actors at play and a darker blue dot with an additional circle around it mark the end of the interaction

With Director's Lens, Lino and colleagues [161, 529] presented a system that suggests camera viewpoints to a user based on a cinematic reasoning and visual criteria provided by the user (Figure 4.7b). Concerning the presented characterisation of the creative flow, one could say that a user initiates the ideation process which is continued by the system. In particular, it constrains the overall options (in detail, by filtering out viewpoints where actors are occluded for instance) and suggests POVs that the user can then select from. Next, a user might either constrain the recommendations further or select one of the presented options which are subsequently further adapter before a final result is produced. In summary, the system provides a higher degree of assistance as it assists in constraining and suggesting (Figure 4.8b). In contrast, Stanciu and Oh [249, 250] implemented and studied a system that also supports users but to a lesser degree of assistance. They extended the remote head of a camera crane with a device that would automatically keep a point in the image at the same location despite the motion of a human operator. Their studies preceded the commercial Stype kit [377, 519] implementation (Figure 4.7c). Regarding the MICI characteristics, one could say that a user initiates the creative process that is then constrained by the system. Further, human and system input contribute to producing and adapting ahead of creating the final recording (Figure 4.8c).

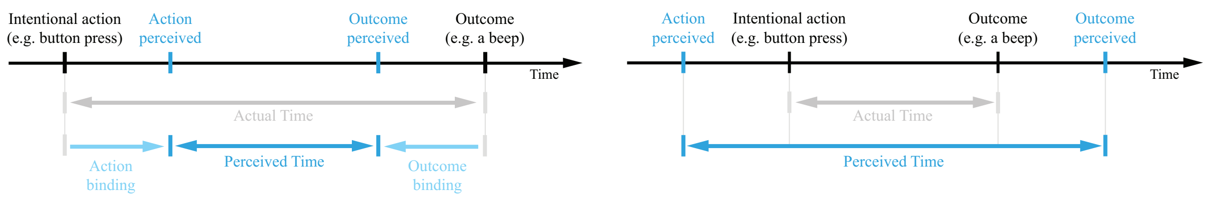
We want to emphasise that aspects concerning automation are discussed in the literature to an extent that clear distinctions and implications can be drawn for the design of systems as in Sheridan and Verplank's 10-LOA table, for example. Regarding creative work, however, it is harder to find such explicit propositions for the design of systems. For example, general advice regarding design principles (*"low threshold, high ceiling, and wide walls"* or *"support many paths and many styles"*) was suggested by Resnik and colleagues [221]. To some degree, a more distinct conceptualisation is presented here with the creative flow characterisation of Deterding and colleagues. While the original characterisation of the creative flow of MICIs was intended for tools incorporating AI, it can also be transferred to (semi-)automated tools for cinematic production in our estimation. We chose two examples that differed regarding the LOA as well as the implementation environment. Independent of their varying characteristics, the creative flow characterisation could be reasonably applied. Thinking about such (mixed-initiative) support tools for creative work in terms of creative flow can not only help to characterise tools in hindsight but also to inform the design of future tools.

4.3 Human Sense of Control

Norman [196, 198] famously transferred the notion of the Gulf of Execution and the Gulf of Evaluation from psychology to HCI. These concepts provided a reliable framework of how one could think about user-centred system design and its effects. Briefly summarised, bridging the Gulf of Execution entails all necessary actions that a user needs to execute on a given system to achieve a particular goal. In contrast, bridging the Gulf of Evaluation sums up all necessary steps of sensing and understanding the status of a system and its relation to the achievement of a particular goal. The related work presented in Sections 4.1 and 4.2 so far, was mainly concerned with different facets of bridging the Gulf of Execution. However, aspects regarding the Gulf of Evaluation were only briefly touched. Therefore, in this section, we want to *introduce insights on how different designs affect the perception of users, in particular regarding how much they feel in control* based on selected literature.

4.3.1 Background

Whether human beings are (pre-)determined or free in their will is a question that concerned philosophers for a long time. While various schools of thinkers on each side of the argument often tried to think their way through it based on introspection and logic reasoning, Libet and colleagues [155] tried a more empirical approach. In a controlled experiment they equipped the participants with EEG sensors measuring cerebral activity and asked them to sit opposite a modified clock (the so-called Libet-clock). During the experiment, they asked the participants to flex their right wrist voluntary at a time of their choice. Once they (were aware that they) had consciously decided to act, they should look at the clock and report the displayed value. Matching the cerebral activity data with the reports of the participants, Libet and colleagues found a reoccurring pattern. Cerebral activity (preceding consciousness) could be measured ahead of the participants being aware of making a choice. So basically one could predict that participants were going to make a decision already before they were aware of it themselves. While the data gathered in their experiment also could not settle the question once and for all [154], it introduced two major contributions. First, it gave an additional notion to a long-fought discussion. It seems that there might be corridors in which we are 'freer' in our choices (as in long-term planning) and corridors narrowing down towards pre-determination (as in short-term action implementation). Second, it introduced a methodology that was used to further inquire about the sense of control as we will present below. The relation between a free will or conscious intentions and a Sense of Agency (SOA)[5] or sense of control are not obvious or self-evident. As put by Haggard [548], *"recent findings suggest that the conscious experience of intending to act arises from preparation for action in frontal and parietal brain areas. Intentional actions also involve a strong sense of agency, a sense of controlling events in the external world. Both intention and agency result from the brain processes for predictive motor control, not merely from retrospective inference."* [104].



(a) Intentional binding for voluntary actions

(b) Intentional binding for involuntary actions

Figure 4.9: Research: Intentional binding for voluntary and involuntary actions based on Haggard et al. [105]

On a fundamental level, Haggard and colleagues [105] further studied the association of (voluntary) actions and its effects. They used a variation of the methodology proposed by Libet and colleagues to conduct a controlled experiment. In short, what they found was that participants who were experiencing a SOA are subject to a phenomenon they refer to as intentional binding. Given an action sequence for a simple button press that causes a subsequent outcome such as a beep played by the system, there is a factual timespan between action and outcome. However, when experiencing SOA, participants tend to report shorter timespans (Figure 4.9). Their explanation model goes along the following lines. Associating a motor action to one's intent is an activity that cognitively requires some (minor) amount of time (action binding). Likewise, associating a sensory input to one's action and intention also requires some (minor) amount of time (outcome binding). Oversimply put, in the participants experience it appears as if those periods are subtracted from the timespan of the overall experience. Therefore, they feel as if a shorter period has passed and consequently report back shorter estimates. This model can

[5]To be precise, SOA is often defined as the *"experience of controlling both one's body and the external environment"* [156].

be applied to carrying out voluntary actions (Figure 4.9a). For involuntary actions the opposite seems true; (Figure 4.9b).

Roughly ten years after the experiment by Haggard and colleagues, Moore and Obhi [185] compiled a literature survey regarding intentional binding and SOA. As in particular critique regarding the Libet-clock methodology manifested, they reported on different approaches that were proposed as the direct interval-estimation method [60, 186]. Here, participants are directly asked to estimate the interval between an action and an outcome event omitting the Libet-clock from the experiment. The methodologies proposed were increasingly used to further investigate the question of "where does a sense of agency originate from?" So far, two notable theoretical positions emerged. The first proposes that SOA stems from a prediction process. Prediction, for instance, is necessary for controlling the human motor system. A forward-dynamic model is used to predict the dynamics of body movement. In addition, a forward sensory model is used to predict the sensory consequences of movement. Based on those predictor models, Blakemore and colleagues [16] proposed a comparator model of the SOA. They suggested that the predicted and actual sensory consequences of movement are compared. When the comparison results in a match, a SOA arises as a result. Contrary to this, the second proposition is that a SOA emerges as a consequence of retrospective inference instead of forward-prediction as suggested by Wegner and Wheatley [269]. In short, the idea is that *"a general-purpose inferential mechanism uses sensory information to establish the causal origins of actions and their effects"* [185]. Thus, the experience of a SOA emerges if there is an intention ahead of the action, if intention and action are consistent and if it is the most plausible cause for the action. In conclusion, however, it seems that actually, both aspects contribute to the emergence of a SOA as suggested by a study conducted by Moore and Haggard [184]. Investigating the issue further, Moore, Wegner and Haggard [186] conducted a follow-up study integrating priming of the participants into the study methodology. What they concluded from their experiment was that it is likely that SOA is derived from agency cues based on their reliability [183]. Forward prediction and retrospective inference might be two sources of such cues. However, there may be additional cues which can factor in and that are still unknown so far.

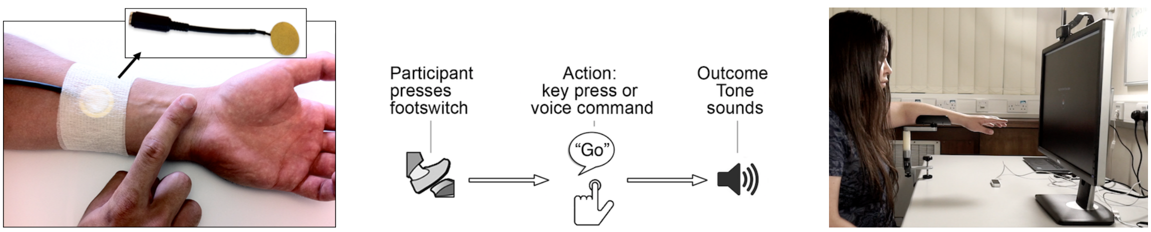
The direct estimation method is also interesting for use in HCI research for multiple reasons. It allows for collecting parametric data implicitly. Thus, one can not only make a binary distinction between whether participants felt in control or not, but also estimate the degree to which they felt more or less in control. Further, as not much additional equipment is needed for its execution, it can easily be integrated into existing methodologies. It is important to note that similar to the notion that our capacity for free will can be thought of as short-term and long-term corridors, a similar distinction was proposed for agency as well. Synofzik and colleagues [254] suggested that a distinction between the feeling of agency and the judgement of agency is appropriate. Here, the feeling of agency is associated with the implicit and pre-reflective low-level action aspects, while the judgement of agency refers to the explicit judgement and attribution of agency regarding conceptual and goal-oriented aspects.

4.3.2 Application

Beyond its philosophical and psychological fundamentals, control is also at the very heart of HCI. Independent of the different subdisciplines, the herein emerging systems and the underlying variety of ideas, the different approaches to HCI have in common that they are all concerned with providing interfaces and feedback to exert control in one way or another. As Sheridan dryly put it in 1978, *"Control means to make a thing do what is desired. There are two main problems in control: 1) to decide what is desired; 2) to make the thing do it"* [243]. More distinctly, Shneiderman [245] proposed in his eight golden rules that designers should *"strive for an internal locus of control"* [245]. Especially for this proposition, it is necessary to determine the effects of design alternatives. For this, determining SOA via direct interval estimation, for example, seems to be a promising approach complementing existing methodologies. However, the attention it received in the HCI literature so far still is limited. Some of these HCI publications and their findings are presented below. For instance, Limerick and colleagues [156] provided a survey on the experience of agency in HCI. Their survey conveyed that the presented studies often compared the effects regarding different input modalities or computer assistance for a given task or tool. Therefore, we will also use this differentiation below.

User Interfaces

Coyle and colleagues [60] used intentional binding determined via a Libet-clock to investigate various interfaces. In one of their presented studies, they compared traditional keyboard input with a novel skin-based input prototype [521] (Figure 4.10a). For the interfaces, either a button press or a tap on a prepared arm would trigger an audio output. Based on their data they concluded that intentional binding was greater for skin-based input. So, one could paraphrase the results as participants felt more in control due to this input modality. More generally, one could also say that different input devices were observed to cause varying experiences of agency using intentional binding as a metric.



(a) Skin-based input [60] *(b)* Speech interface [157] *(c)* Mid-Air gestures [57]

Figure 4.10: Research: UIs of varying input modalities that were studied regarding SOA

Similar to the work of Coyle and colleagues, Limerick and colleagues [157, 469] compared keyboard input to speech input by estimating intentional binding using the Libet-clock approach (Figure 4.10b). They report on two studies they conducted and concluded that their speech interface led to a diminished SOA compared to keyboard input.

In addition to skin-based and speech input also mid-air gesture input was investigated regarding the arising SOA as in the work of Cornelio Martinez and colleagues [57, 435]. They also compared mid-air gestures to keyboard input using a Libet-clock to determine intentional binding in two studies (Figure 4.10c). In the first study, they found that in general keyboard and gesture input led to intentional binding. For the mid-air gesture condition, they also tested two different feedback mechanisms which used either visual or auditory feedback. Based on the results of these conditions, the authors suggested that an increased SOA is experienced with auditory feedback compared to visual. They further extended the system with haptic feedback in a second study. The results indicated that higher intentional binding was perceived with haptic feedback than visual feedback.

The explanation models that several authors offer follow similar lines of argumentation. In the work of Coyle and colleagues, a higher degree of motor activity might be factoring in substantially leading to an increased SOA. Vice versa, the apparent minor degree of motor activity in the speech interface condition studied by Limerick and colleagues might be contributing to the diminished SOA. Likewise, SOA could be increased by incorporating haptic feedback as reported by Cornelio Martinez and colleagues. While here still motor activity due to the mid-air gestures might factor in, associated perception systems that register and evaluate the haptic input additionally seem to play a substantial role. This explanation model might also apply for the studied skin-based input.

Automation

Researchers also investigated different LOAs and their effects on SOA. For instance, Berberian and colleagues [13] tested a system with four LOAs ranging from full manual to full automation at the extremes and two intermediate levels. The study task was to observe flight plans and to avoid future conflicts with other aeroplanes. Thus, participants were asked to decide on an appropriate reaction (larger goal) and execute it via a button-based UI (low-level actions). The system provided visual and auditory cues as feedback representing a successful execution. For measuring SOA, the direct interval estimation method was applied. What the authors found was that with increasing LOAs interval estimations increased. They subsequently suggested that hence higher LOAs led to a diminished SOA.

In their paper on skin-based input, Coyle and colleagues [60], also report on one study examining how SOA is affected by computer assistance. In particular, they looked at a pointing task where participants were asked to point and click presented targets with an off-the-shelf mouse. They also compared four LOAs with no assistance and high assistance at the extremes and two intermediate levels (mild and medium). For measuring SOA, they also used the direct interval estimation approach. From the collected data the authors derived that at a mild LOA the participants still experienced a SOA. For the higher LOAs, this experience seemed to be lost, however, although here the goal was still successfully achieved.

Summarising the presented studies, Limerick and colleagues concluded that SOA might be *"a graded experience in situations where the line between voluntary and assisted action is gradually blurred"* [156]. This conclusion seems to be in line with prior findings regarding sensorimotor prediction models. Explanation models offered propose that with an increasing LOA, prediction models become less accurate at predicting the next sensory state. Thus, congruence between the predicted and the sensed state (as suggested by the comparator model) is reduced which led to a diminished SOA. In consequence, the authors suggested that measuring SOA should be incorporated into the practices of UI design and evaluation so that it can be identified when SOA becomes disrupted.

4.3.3 Predominance

So far, the presented studies were mainly concerned with estimating the effects on the SOA of the participants subject to either varying UIs or varying LOAs using intentional binding approaches. Overall the results suggested that an increased motor activity accompanies a higher SOA. However, due to the measurement techniques applied and the methodologies of the studies, the subsequent interpretation of the results only relates to the previously introduced notion of a feeling of agency. In contrast, for estimating the judgement of agency concerning a high-level goal other measurements such as verbal reports would be additionally necessary.

Incorporating these aspects, Kumar and Srinivasan [145] compared a joystick and button setup in a pointing task. In their experiment, they used three levels of additional random movement resulting in a full, a medium and a low control condition. They measured SOA implicitly and explicitly. For the implicit measures, they used the direct interval estimation approach, for the explicit they asked for a verbal report on a seven point-scale regarding the sense of authorship based on work by Ebert and Wegner [76]. Based on the gathered data one could say that for unsuccessful trials, where participants missed targets and consequently not achieved the high-level goal, intentional binding decreased in relation to the automation provided. This finding is consistent with the prior findings by Coyle and colleagues and Berberian and colleagues. However, in successful trials, this pattern did not emerge.

Feeling and judgement of agency are distinct concepts by definition, however, in reality, they are hard to separate. Often systems involve low-level actions and high-level goals, especially when systems embrace a human-centred approach. Therefore, which of both aspects matters more to users is interesting to inquire. Following this question, Wen and colleagues [270] created ambiguous situations on purpose incorporating both aspects. They conducted a controlled user study where the participants were asked to execute a pointing task via a keyboard interface. The experimenters introduced delays (100, 400, and 700 ms) to manipulate the difficulty of the task. They also provided two LOAs with full manual control and mild assistance. In the assisted condition the system would ignore erroneous commands, which would, therefore, lead to better overall performance but weakened the association between action and feedback. The authors measured the sense of control via verbal self-reports on a nine-point scale (Figure 4.11). Analysing their data, they found that SOA increased in the assisted condition although a substantial amount of user commands were not executed. Interpreting their findings, they suggest that *"(...) both factors influenced the sense of agency, but based on task performance, the goal-directed inference played a dominant role in the judgment of sense of agency when the action-feedback association was uncertain"* [270].

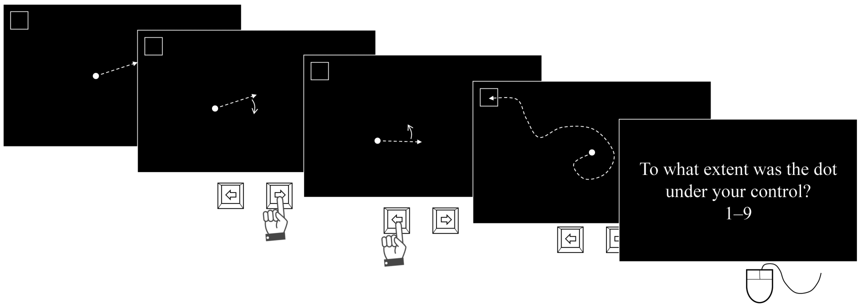


Figure 4.11: Research: Flow of the experiment regarding predominance in sense of control from Wen et al. [395]

4.4 Summary

In Section 4.1.1, we introduced the mutual benefits that humans and machines can contribute. Being able to delegate some tasks to either agent, also different delegation and allocation strategies emerged. Those also introduced further side-effects. As such a set of ironies, risks (such as skill degradation) and in particular an uncanny valley of automation was identified by various researchers. We emphasised that the concept of (varying) levels of automation is of central importance for the description and design of human-automation systems. We further sketched out how addressing observed shortcomings led to the idea of LOAs being static transitioning to be thought of as dynamically shifting during runtime. We also focused on how, inevitably, with the introduction of different LOAs in systems a trade-off between workload and unpredictability arises that needs to be taken into account. Besides, in the domain of creative work, automation faces particular challenges that most likely need to be incorporated in the design of (semi-)automated systems in particular for camera control.

Different approaches to how cameras can be controlled were detailed in Section 4.2. Most fundamentally, we distinguished between camera control in virtual and physical environments. In the virtual domain, one can find techniques for control over general movement, specified trajectories and particular cinematic approaches. Here, in particular, the content-based techniques proposed by Gleicher and Witkin are noteworthy as they minimise errors and a need for training in camera operation. Further, they can be extended to control other cinematic parameters as Christie and Hosobe showed and can be translated to mobile devices as Reisman and colleagues reported. These aspects make them particularly interesting also for use in physical environments. It was further emphasised that it is not always the case that any virtual technique can be directly translated into the physical world. We detailed how physical systems are fundamentally set up, where they differ from virtual implementations and how basic techniques (such as computer vision) are used to drive such systems. CV techniques are not entirely novel in cinematic production. Their capabilities and how they are already used was also presented.

Taking into account the findings mentioned in the previous sections, we present research on the phenomenon of a sense of control or agency in Section 4.3. The LOA affects this SOA on system design level especially regarding the degree of actual control that is required by the user. As we identified priorly as in Section 3.5, expert camera operators especially want to be or at least feel in control when they are crafting their creative content. How to design for control, however, is also not trivial as various mechanisms seem to be at play simultaneously as psychological researchers suggest. On the one hand, it seems that a feeling of control emerges due to motor activity of users, so to speak when users control more manually. On the other hand, it seems that a judgement of control is derived from matching a current operational status to the achievement of high-level goals, which is often supported by automation. By nature, these approaches seem somewhat contrary and consequently need to be balanced by a system properly. Therefore, one needs to carefully design systems such that one avoids ambiguous situations between them and make sure that one is not unnecessarily sacrificed for maximising the other. In conclusion, evaluation in this regard seems particularly necessary, however, so far it is not commonly found in the literature.

Insights

- The research literature already describes various approaches of how assisted camera control could be implemented. It indicates that techniques for camera control in virtual environments can sometimes (but not always) be transferred to the physical world. Especially content-based techniques are interesting as they are capable of reducing the margin of error and require less training. Additionally, the approaches also translate to mobile devices that are already used (as for drone control). Under the hood, often computer vision is driving such techniques, and it is increasingly leveraged in various domains of cinematic production.

- For the implementation of such techniques into actual systems, overall adaptive automation seems a promising design approach for multiple reasons as it allows to integrate the benefits of both agents, to adapt to unforeseen environments and to incorporate human-centred aspects. However, designs incorporating the full triad of adaptive automation, human-centred aspects (such as feeling in control) and a computer-supported adaptive creative exploration are not free from pitfalls.

- Those, for instance, can arise due to higher levels of automation and abstraction in system control which affect the perception of when and how users feel in control. This perception can be affected at the level of the feeling or the judgement of control. Fundamentally, both seem to tap into different neural processes, and hence ambiguous situations can arise. Such situations bare the risk of leaving the user behind with a negative sense of not being in control.

JASON SILVA (DIRECTOR)

Creativity and insight almost always involve an experience of acute pattern recognition: the eureka moment in which we perceive the interconnection between disparate concepts or ideas to reveal something new.

5

Design: Connecting Technology, Users and Research

What to expect?

- Making the relationships between users and the current state of technology apparent
- Extracting design variables for user interface designs

What to take away?

- A conceptual framework contributing to a more holistic understanding of control in HCI in general and physical cinematic camera control in particular

5.1 Propositions

What we have been trying to do so far was to lay out the necessary foundation that enables us to speak about the peculiarities of cinematic camera control and possibly about viable future designs. At some points, it hopefully shined through that the introduced aspects are not isolated from each other but rather are intertwined. The idea behind this chapter is two-fold. First, we want to further *make the connections between insights from users, research and technology apparent*. This approach, at best, can contribute to a more holistic conceptual framework that possibly can help to add further detail to the understanding of the relationship between actual and perceived control in HCI. Second, we want to *extract a set of design variables* that can be used to describe cinematic camera control UIs in hindsight but also allow inferring new designs and research opportunities.

5.1.1 Designing for Control: An Explanatory Framework

While, for example, Fitts [87] already anticipated challenges arising due to automation in 1951, his statements did not go along with an explanation model focusing on why the problems might arise in detail. His statements were informed by the psychological literature of his time and a set of (then available) empirical observations. A more detailed summary of further empirical findings followed, for instance, with the work of Bainbridge [8]. While she presented arising issues and ironies and also contributed ideas on how to resolve them, we think of her contributions mainly as diagnosing symptoms and offering cures to those. Going deeper regarding the level of analysis, Miller and Parasuraman [179] contributed to a better understanding of the chain of aspects factoring in that might be causing those symptoms. In their point of view, the introduction of LOAs inevitably came with a workload-unpredictability trade-off. Please note that the terms prediction and (un)predictability are used in the discourse on computer science and on psychology referring to varying definitions. Therefore, we want to present the understanding of Miller and Parasuraman as closely as we can by quoting it directly: *"Unpredictability refers to the inability of the human to know exactly what the automation will do when. Unpredictability is a consequence of the human not personally taking all actions in the system – of not being "in control" directly and immediately. Unpredictability is inversely related to the workload required to maintain awareness of the actions being taken by the system. It impacts overall situation awareness by reducing awareness of those specific aspects of the situation that pertain to the understanding of the automation behaviours and the system functions they control."* [179]. While the authors here try to get further into the roots of the cause, they acknowledge and point at the psychological aspects of not being 'in control' that play into it. However, the mechanisms underlying this phenomenon are not presented in further detail (and rightfully so, as they might not necessarily be of central interest for their contributions). As in our work, we think of further understanding as vital. Therefore, we try our best at following that trail further below.

So far, the history of ideas that unfolded covering the issues presented above is akin to a top-down approach as far as we are concerned. Challenges or symptoms presented itself to researchers and step by step they aimed at classifying, finding commonalities among them, presenting solutions or cures by exploring reasons deductively as to why those might arise and subsequently put their explanation models to empirical tests. In parallel, similar to an inductive bottom-up approach, the concept of neural intentional binding processes affecting the perception of agency emerged and was transferred from one field to another. Initially, it was used by Libet and colleagues [155] to inquire on the nature of human 'free will'. The technique they came up with was later used to investigate patients with schizophrenic personality disorders. They observed that patients sometimes claimed that they were not the ones behaving in a certain way; instead it was done through them. Researchers examined how one could tell the difference between someone who actually experienced such a situation and somebody who was not. Estimating intentional binding researchers such as Haggard, Moore and others [105, 184] found a way to infer implicitly and objectively whether someone would feel as the author of one's action and to which degree [13]. Together with the underlying conceptual models, those were transferred to HCI as by Coyle, Limerick and others [60, 157]. Summarising the HCI literature regarding SOA, Limerick and colleagues [156] report on studies that relate intentional binding effects to automated systems. While this relationship became introduced, it did not take into account the full scale of detail that has been put forth in research on automation.

Below, we try our best to bring the findings of both domains further together and fill in some blind spots that we see in matching up the top-down deductive findings in automation and HCI literature with the bottom-up inductive explanation framework found in the psychological literature. Limerick and colleagues already did a great job at relating both domains: *"With regard to the cognitive basis for the sense of agency, the finding that there is a graded loss of sense of agency with increasing assistance is potentially consistent with our current understanding of sensorimotor prediction. For example, increasing assistance may result in internal sensorimotor predictive models becoming less accurate at predicting the next sensory state; this could, therefore, give rise to reduced congruence between the predicted sensory state and that actual sensory state."* [156]. Although not mentioned in their paper, we want to make it apparent that we here see a close connection to the work of Miller and Parasuraman. For us, it seems, as if both propositions mark the ends of distinct threads of research that could be beneficially tied together. In our estimation, the (un)predictability of systems referred to by Miller and Parasuraman interlocks with the deeper levels of analysis and understanding that underlie the SOA and intentional binding literature. In other words, it (potentially) represents the overlap of what computer scientists might refer to as the (un-)predictability that arises for users due to system design its effects on the one side. On the other side, it might be caused by what psychologists refer to as a reduced congruence between sensed states and prediction models regarding motor activity that are embedded within the users affecting their perception of the system.

This overlap can also be used to derive a coherent explanation for the observed uncanny valley of automation (Section 4.1.1). Flemisch and colleagues [88] who adapted the term to automated driving, offer the following scissors-metaphor as an explanation by analogy: *"Another possible metaphor depicts a pair of scissors: One blade represents the operator's take-over capabilities, the other the automation's availability. The first blade is slowly closing when the operator's take-over capability decreases with increasing trust. As soon as the automation availability decreases temporarily, the movement of the two blades is cutting off the operator from the control of the process, e.g., of the vehicle."* [88]. Using the intentional binding model and the workload-unpredictability model as a foundation, one could add that (regarding the first blade) with increasing LOA the unpredictability of a system's results also increases. That consequently decreases the reliability of motor and sensory state prediction models of the driver which consequently causes intentional binding and along with it the SOA to decrease. Taking into account the predominance aspects proposed by Wen and colleagues [270], however, (for the second blade) one could say that at the same time, a SOA emerging as a consequence of a large enough extent of proper judgement of agency has not (yet) increased enough or is diminished to a significant enough degree so that it does not dominate in this situation. In consequence, as both SOA aspects are pronounced too weakly, this leads to an uncanny situation for users or in the words of Flemisch and colleagues 'cuts off operators from the control process'.

Solutions for resolving such situations might further be built on top of other work that we presented. For instance, Hoeger and others [115] proposed interlocked transitions as a solution for avoiding critical situations in LOA shifts. As their primary interest was to provide secure car operation, they evaluated their designs using safety as a metric. In contrast, we suggest that the concept of interlocked transitions could be (potentially) reasonably transferred to a more extensive set of designs, including those that are not primarily concerned with driving assuming a SOA informed explanation model. While this might be easier conceivable for industrial processes, we would further suggest that it might also translate to creative work as in our case. Deterding and colleagues [69], for instance, proposed conceptualising and visualising the creative flow of MICIs. For us, this concept and representation resembles the idea of shifts in the LOAs in AA but instead deals with the peculiarities of creative work. Here also, the aim is to make it apparent when agency shifts. Along with those shifts, potential hazards regarding agency or a sense of authorship can be reasonably anticipated. Looking at the concept with a SOA lens, we suggest that for the apparent shifts, potential problems should be investigated and tried to be addressed. For example, most obviously, this might be the case when a computer takes-over the initiative (arrow in the illustration). Adapted concepts that aim at translating interlocked transitions to MICIs might further be helpful. Investigating such issues at the cross-section of LOAs / MICIs and SOA leads to fundamental questions as raised by Limerick and colleagues: *"what happens to a person's sense of agency when they voluntarily initiate an action, but a computer then steps in to complete the action? This agentive ambiguity in interactions with intelligent technologies also presents interesting challenges for research into the sense of agency."* [156].

Regarding SOA, the distinction between a feeling and a judgement of agency is central. From a psychological perspective, it is necessary as they are different phenomena with different neural processes associated with each term. So far, we focused on the feeling of agency. We aimed at associating the concepts of a loss of intentional binding on a non-conceptual level and a loss of control at a more conscious level that, for instance, potentially can arise at intermediate LOAs by suggesting that the former propagates to the latter (when the judgement of agency is not dominating). More generally speaking, we tried to connect SOA and LOA regarding the feeling of agency. In contrast, for a judgement of agency we have not provided further explanatory models or literature in particular. Thus, we want to additionally associate this concept with a cornerstone of the HCI literature below. As mentioned earlier, Norman [196] introduced the idea of the Gulf of Execution and the Gulf of Evaluation in HCI. Here, the Gulf of Execution is concerned with all necessary steps for carrying out actions: *"intention formation, specifying the action sequence, executing the action, and, finally, making contact with the input mechanisms of the interface"* [196]. Making contact is, in particular, influenced by the LOA, the modalities provided by the system and the UI that afford (low-level) action taking. Further, the Gulf of Evaluation is concerned with sensing and identifying the systems status and mapping that status to a larger high-level goal. In Section 4.3, we mentioned that SOA is somewhat related to the Gulf of Evaluation. This statement was a bit oversimplifying, and therefore we want to expand on the details in this paragraph. On further examination, it seems to us that in essence, a SOA is a (by-)product of whole execution-evaluation cycles. Feeling and judgement of agency, however, relate differently to the characteristics of those cycles. A feeling of agency might arise in a sequence of smaller or shorter cycles that provide proper feedback more immediately. If, for instance, controllers for continuous control elements and the mappings between UI elements and the conceptual model of the system are well-designed and provide immediate feedback, it is less likely that the congruence between the internal prediction models and the sensory state models diminishes. Thus, in such a system, the action-feedback association is close, and a feeling of agency is more likely to arise. In other words, *"Rather than being explicitly aware of the motor representations involved in the comparator process, we experience self-agency by a rather diffuse sense of a coherent, harmonious ongoing flow of anticipations and sensory feedback"* [254]. In a vaster interaction process, those smaller cycles might be nested in larger cycles where users work to achieve a higher goal and where the feedback is less immediate. Multiple smaller cycles might be needed to accomplish larger milestones of which multiple are needed to achieve a high-level goal or as put by Norman: *"Real activity does not progress as a simple sequence of stages. Stages appear out of order, some may be skipped, some repeated."* [196]. However, not every action in a smaller cycle might affect the feedback regarding the overall goal. We would suggest that such larger cycles relate to the emergence of a judgement of agency. A properly designed system makes it easier for the users to evaluate the status of the system and its relation to the status of the process given an overarching goal. Thus, in such a system, a (positive) judgement of agency is more likely to arise or, as phrased by Synofzik and colleagues, *"the pre-conceptual basic feeling of agency is further processed by conceptual capacities and belief stances to form an attribution of agency"* [254].

In sum, we suggest that it is the characteristics of these cycles and the dynamics between them that affect both aspects of agency. In particular, while, in shorter cycles, immediate feedback is necessary for establishing a proper action-feedback association and hence a feeling of agency, in larger cycles, more feedback regarding conceptual aspects and its relation to the status of the process is necessary for the emergence of a judgement of agency.

5.1.2 Designing for Camera Control: Challenges

As we elaborated in Section 4.1.1, the design of computer-assisted support tools for creative work is not free from challenges. A general one, for instance, is that the vocabularies used by experts and researchers differ although they refer to similar or the same phenomena. Without matching those, it can be harder to associate existing practices and research findings that can guide the design. We tried to match up the aspects that we found already so far as in Section 4.2. In this section, however, we want to focus distinctly on aspects that can more directly challenge or translate into design decisions below. The compilation of those challenges is by no means meant to be all-encompassing. It is derived from the insights arising in our user research (Chapter 3), and the related work presented (Chapter 4) and is meant to (merely) guide the following design process.

Diverging User Groups and Criteria for Rejection

As detailed in Section 3.6 there are diverging user groups with different expectations and skillsets regarding the control of the camera and MoCo systems and hence have different criteria for rejecting designs. For novices, camera movement used as a stylistic device and the steering the devices might be somewhat of new concepts. Therefore, one probably should support a walk-up and use scenario and offer some degree of assistance at best on devices novices already own. For them, expert tools easily seem as too hard to learn or too complicated. In contrast, enthusiasts and experts favour haptic control elements and manual control options as they emphasise a feeling of control. For this group, however, one should design for knowledgeable power-users that are used to managing different hardware and control concepts. Experts potentially are also more willing to train with new tools as they are sometimes also new to specific tools, but by the same token also more hesitant towards assistance.

Challenge 1: *“Getting the right design and the design right”* [29] for a particular user group

Haptic Properties on Mobile Devices

Design trade-offs as mentioned in Section 4.1.2 not only become apparent on a system design level. Similarly, they appear in the design of UIs as well. As a platform, mobile devices are in particular attractive due to their relatively large display sizes and their versatility and flexibility in use. Additionally, assistance features and content-based approaches have been implemented on them already in the past as presented in Section 4.2.1. However, they also afford fewer haptic properties such as physical buttons or joysticks as they make predominantly use of multi-touch interaction. Touch-based interactions have been observed to result in a diminished sense of precision and occlusion [12]. As mentioned above, experts strongly favour haptic properties. Therefore, mixed designs incorporating physical controllers and mobile devices are used more frequently in-the-wild (see Section 3.6).

Challenge 2: Design on mobile devices to account for the diminished precision and/or plan for a mixed setup incorporating physical controllers

Occlusion in Content-based Approaches

The large display sizes allow presenting the camera stream to operators which is considered vital feedback for low-level control [82, 240] as well as estimating the quality of the material and coordinating the team regarding high-level goals (see Section 3.6). However, to some degree, UI elements are presented as an overlay on top of the video stream and therefore occlude parts of it. For non-content-related features, UI elements can theoretically be placed at uncritical locations as black areas due to letter-boxing. However, for content-based control features, this gets increasingly harder as it seems reasonable to place UI elements next to associated objects or locations [223]. This phenomenon has been investigated before, addressing the issue in various ways, for instance, by adapting symbols in camera-based [36] or map-based UIs [195]. Also, occlusion can arise when the fingers occlude the screen during the interaction [12] but requires different solutions.

Challenge 3: Minimise occlusion due to UI elements without unnecessarily sacrificing usability

Prototyping

To foster fast prototyping and evaluation, Nielsen [193] famously proposed a set of discount usability engineering methods. These encourage the use of paper prototyping, scenarios and heuristic evaluation. In our estimation, it is not perfectly clear how these approaches can be translated best to experts users in camera motion. On the one hand, in other domains, experts were already observed to be hesitant towards paper prototypes [116, 117] as they are very early drafts of systems, yet they are used to operating state-of-the-art systems with a low margin for errors on a regular basis and sometimes even with time and budget restrictions. On the other hand, the details of how the camera moves are important but also very hard to translate into a (semi-)static paper prototype. Paper prototypes are the first go-to option for most designers and researchers early in the design process. If those are not necessarily as useful in this environment, alternative approaches should be considered.

Challenge 4: Find viable ways of prototyping that are expressive enough for camera motion

5.1.3 Designing for Camera Control: Variables

To better conceptualise the design of map-based information visualisations, Bertin [14] introduced his concept of visual variables. Those describe individual visual properties such as position, shape, size and others that could be used to encode particular information and how these could be combined to encode more complex information. While his ideas were conceptualised for purely analogue maps, his propositions were also considered by researchers for digital information visualisations as, for instance, by Carpendale [37]. For our work, a particular focus only on visual variables seems mostly too fine-grained as the design challenges mentioned above also tap into various areas outside of the visual design. Therefore, we suggest translating the underlying concept from visual variables, to a more abstract set of design variables (for the design of camera control UIs) that is derived from the challenges mentioned above and related work presented earlier. Those will be detailed below and could, for instance, alternatively be understood as dimensions in a design space. However, as our selected items are not meant to be all-encompassing, this possibly might only result in an arbitrarily selected sub-space. Thus, we suggest referring to these items (only) as (combinable) design variables.

User Expertise

The transition of cinematic production into its digital era went along with a greater differentiation regarding the user groups. Addressing the resulting and above mentioned **Challenge 1**, we can use our analysis of existing UIs in 2.2.3 together with the meta-personas in 3.6 to conceptualise this differentiation. Thus, regarding design implications, we propose following scale (Figure 5.1):

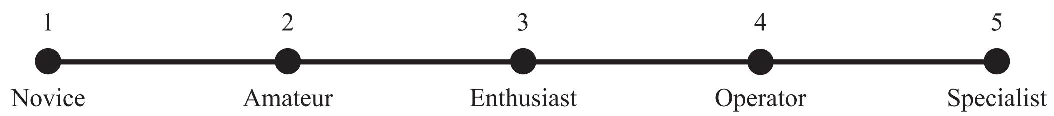


Figure 5.1: **Design Variable 1**: Level of user expertise

(1) Novice: Is new to cinematography and camera operation and has an interest in and access to some means of cinematic production

(2) Amateur: Similar to the 'Intuitive Artist'; has some background in creative work but not particularly in camera motion

(3) Enthusiast: Could be an 'Exploring Student'; already has some experience in cinematography and tool operation and is open to new tools as well

(4) Operator (General Expert): Could be a 'Settled Traditionalist', 'Settled Creative', a 'Diligent Worker' or an 'Experienced Teacher' who has mastered several aspects of the craft

(5) Specialist (Specialised Operator): Beyond mastering the fundamentals, a specialist focused on a particular tool above other aspects; might be what a 'Diligent Worker' strives for

Sensorimotor Activity

The degree of motor activity that operators can exert seems to be important on multiple levels such as sensorimotor prediction models, haptic feedback and proprioception, perceived affordances, precision and others. To better address **Challenge 2**, we aim at establishing a design relevant scale of it below, based on the work presented in 4.2.1 and the studies mentioned in 4.3.2 (Figure 5.2):

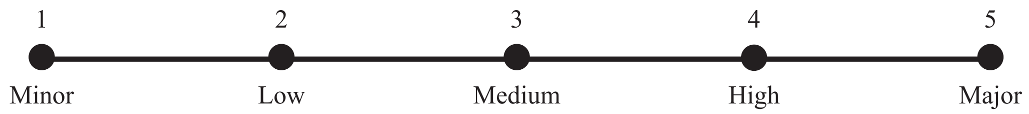


Figure 5.2: **Design Variable 2**: Level of motor activity

(1) Minor Level of Sensorimotor Activity: As in abstract declaration or initial (parametric) setup only without continuous control (as the Declarative Camera Control Language by Christianson and colleagues [51])

(2) Low Level of Sensorimotor Activity: As in speech interfaces mainly focusing on abstract parametric control (as in the cooking show system by Pinhanez and Pentland [214])

(3) Medium Level of Sensorimotor Activity: As in mid-air gestures interfaces incorporating hand-eye coordination and proprioception more pronounced (as Nagi and colleagues [191] presented for the control of a drone)

(4) High Level of Sensorimotor Activity: As in continuous remote control via touch-interfaces but still facing a diminished sense of precision (as often found in the wild, see 2.2.3)

(5) Major Level of Sensorimotor Activity: As in continuous remote control via physical interfaces or full manual manipulation of physical items (often used in the wild, see 2.2.3 and 3.3)

Video Stream Saliency (User Interface Occlusion)

As mentioned above in **Challenge 3** and in 3.4, an unoccluded view on the camera stream is important for multiple reasons. Displaying visual UI elements on top of the stream can become necessary, and a design trade-off result from it. How a design handles this trade-off affects the saliency of the underlying video stream. For this thesis, we suggest the following scale along which designs can be classified regarding the visibility of the captured material (Figure 5.3):

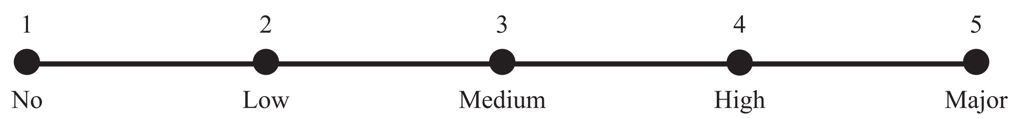


Figure 5.3: **Design Variable 3**: Level of video stream saliency

(1) No Video Stream Saliency: No video stream is displayed and only a UI is visible, as in the Edelkrone Motion Kit [533]

(2) Low Video Stream Saliency: The UI is presented more prominently than video stream, as (sometimes) in control apps of PTZ cameras, such as the Axis Companion [444]

(3) Medium Video Stream Saliency: Elements of the UI and the video stream are presented equally prominent as in the drone UI of Galvane and colleagues [94, 421]

(4) High Video Stream Saliency: The video is displayed more prominently than the UI, as for most professional grade drones (see 2.2.3)

(5) Major Video Stream Saliency: Systems that actively try to maximise the saliency of the video stream by strategies such as Progressive Disclosure [260, 361]

Automation

We tried to provide an overview of the literature on automation and introduced the importance of LOAs in its design in 4.1. While Sheridan and Verplank proposed ten levels for HCI, the Society of Automotive Engineers [370] suggests using six levels for interacting with (partially) automated cars. While the extremes are obvious (full manual control and full automation), the gradation seems more debatable for each context. For this thesis, we suggest a five-level scale (Figure 5.4) for the design of cinematic camera motion (while full automation is viable, it is outside our focus on active users):

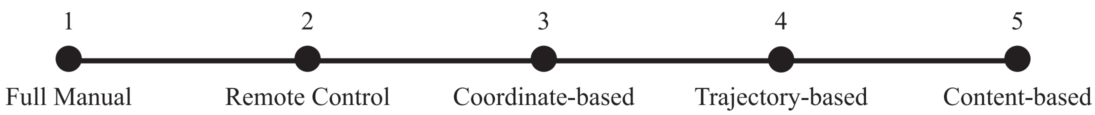


Figure 5.4: **Design Variable 4**: Level of automation

(1) Full Manual (Continuous) Control: No motorisation or assistance is used. Operators physically move the tools continuously as mentioned in 3.3 and 3.4. It could be categorised as tools for General or Targeted Movement based on the taxonomy of Mackinlay and colleagues [168]

(2) (Continuous) Remote Control: The operators are physically detached from the tools. They use continuous control interfaces such as joysticks or wheels [338, 484] to steer the systems as detailed in 2.2.3. Those could also be categorised as tools for General or Targeted Movement based on Mackinlay and colleagues [168]

(3) Coordinate-based Control: A particular set of coordinates is specified or saved. The system then executes the necessary actuation steps to reach the coordinates. It can be associated with Specified Coordinate Movement as proposed by Mackinlay and colleagues [168]. It is, for instance, implemented in the Supertechno TechnoDolly [556, 496] (among other features)

(4) Trajectory-based or Keyframe-based Control: A trajectory (incorporating a larger number of coordinates) is required as an input by the users. It can encode details on the position and the orientation and be characterised by the Specified Trajectory Movement of Mackinlay and colleagues [168]. For operating such a system, for instance, Gebhardt and colleagues [98, 438] used a pen-based UI. Alternatively, users can input the position and orientation at individual locations and points in time. The system automatically interpolates the trajectory between those keyframes which often can be further edited as in the Cmocos or TechnoDolly systems (2.2.3)

(5) Content-based Control: As detailed in 2.2.1 and 4.2 control can also be exerted in a more abstract form that takes into account an analysis of the content of a scene and derives control decision based on it. In the work of Christie and Olivier [53], this could be categorised as Interactive Approaches and can be found in systems such as Vertical AI or Percepto (2.2.3)

(-) Autonomous Camera Planning and Controlling: The system derives decisions on planning, steering and selecting independently as proposed by Chen and Carr [46]. Alternative such behaviour could be understood as Reactive or Generalized Approaches in the terminology of Christie and Olivier [53]. It is implemented in drones such as the HexoPlus (2.2.3)

Please note that in practice it is common that systems offer multiple ways for their control. Mixtures can, for example, be found in Spike (remote pen-based control and trajectory-based control) or the TechnoDolly (remote joystick and wheel control, coordinate and trajectory-based control).

Prototype Fidelity

Using prototypes at different levels of fidelity is already well-established in HCI. To address **Challenge 4**, we want to briefly provide a scale of it for prototyping cinematic UIs that seems reasonable for us (Figure 5.5). The items are influenced by our experiences later presented in Chapter 6.

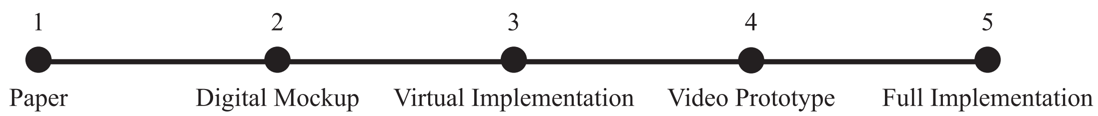


Figure 5.5: **Design Variable 5**: Level of prototype fidelity

(1) Paper: (Semi-)static low-resolution prototypes that make mainly use of paper.

(2) Digital Mockup: A digital and clickable implementation at medium resolution

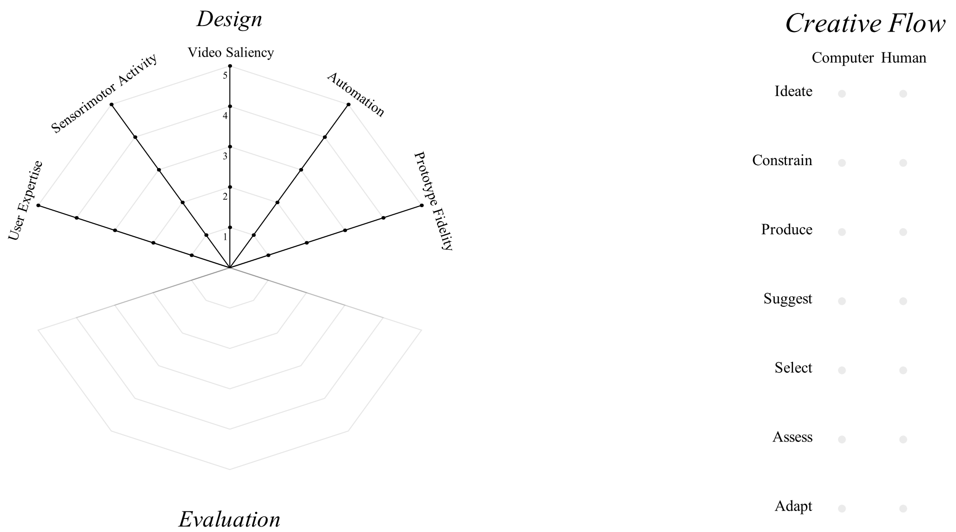
(3) Virtual Implementation: A medium resolution implementation in a virtual 3D environment

(4) Video Prototype: It showcases a system at high resolution but is hardly interactive

(5) Full Implementation: Prototypical implementation of the necessary hardware and software system components for use in realistic circumstances

Representing Design Variables (and Evaluation Metrics)

Different prototypes can be understood as different instances of the design variables with values varying along each of the presented scales. For reasons of simplicity, we suggest using a single representation that entails the introduced variables for later use in this thesis. Therefore, we will introduce a visual representation that fuses the presented singular scales into a radar chart below.



(a) Design variables template

(b) Creative flow template

Figure 5.6: Design Variables: Templates for their representation (Figure 5.6a) and the creative flow based on Deterding et al. [69] (Figure 5.6b)

As the effects of alternative designs of prototypes can have varying effects regarding evaluation criteria, it is, in our estimation, vital to represent not only design aspects but also their effects given specific evaluation metrics in the chart. The essential criteria in our work will later be presented in Chapter 6. In summary, including design and evaluation criteria in one representation, the various aspects can be illustrated based on the form of a decagon. The upper half can be used to represent the instantiation of the design variables and the lower half to represent the results regarding the evaluation criteria. An exemplary template is found in Figure 5.6. We will use this template to represent the properties of our prototypes later described in Chapters 7 and 8. Please note that this approach mainly represents static properties. However, as mentioned earlier, for instance, the LOA and dynamic changes during runtime concerning who is in charge are also fundamental. Thus, one might additionally consider using the creative flow representation of Deterding and colleagues [69] as an additional template to depict these aspects. In sum, both templates can be reasonably used to characterise prototypes regarding their central static design characteristics (and the resulting effects) and the central dynamic changes in control as far as we are concerned. So that in consequence, (hopefully) the vital design aspects can already be estimated based on these abstract representations.

5.2 Summary

In Section 5.1.1, we presented the outlines of an explanatory framework for designing for control that is intended to add further detail to bridging the gaps between the literature on automation, HCI and SOA. Here, we emphasised, in particular, the connection between the LOA in the design of (semi-)automated systems and how it can affect the user experience on the level of SOA. We presented how a reduced congruence of internally predicted and sensed agency cues emerging at intermediate LOAs could be responsible for what is often empirically observed and referred to as a "*loss of control*" or uncanny situations. We also mentioned, how this is not the only mechanism at play and how alternative cognitive processes that keep track of the process of achieving higher-level goals can also interfere. Further, we introduced the idea of interlocked transitions (that are one way of addressing such uncanny situations in automated driving) that could be transferred for use in creative work; or in other words how this approach could help to bridge the uncanny valley of automation in creative environments. In particular, we mentioned how those could relate to an analysis of the creative flow which is common, for example, to conceptualise MICIs. Relating the HCI and SOA literature more closely, we also sketched out in Section 5.1.1 how the congruence of internal prediction and sensory models or lack thereof relate to the dynamic characteristics of smaller cycles (executing low-level actions) that might be embedded in larger cycles (achieving high-level goals) in bringing the Gulf of Execution and the Gulf of Evaluation.

Bringing together insights from our analysis of status-quo technologies (Chapter 2), our user research (Chapter 3) and from related work (Chapter 4), we derived a set of four design challenges in Section 5.1.2: designing for diverging user groups, addressing missing haptic properties when using mobile devices, trying to minimise occlusion of the video stream and finding proper means of (rapid) prototyping that feature a large enough degree of expressiveness regarding camera motion. It seems to us that these particular challenges need to be adequately addressed in the design of systems for camera control that put an emphasis on using technology to support operators in the control of human-centred systems that further incorporate existing practices.

From those challenges and already introduced related work on automation, we further extracted a set of five design variables in Section 5.1.3: user expertise, motor activity, video saliency, automation and prototype fidelity. Instantiations of those could each take on one of five values, in our estimation. For each variable, we based our proposed gradation on particular examples of established systems or reports found in the literature. However, we also want to point out that those values are located on a gradual scale. So, one could also argue that a more fine-grained gradation leading to a higher number of values is reasonable or necessary. Similarly, one could also make the case for floating-point values located between two integers.

It is further necessary to mention though that these design challenges and variables are not the only ones one could pick. Even based on our prior estimations, further challenges could be selected that could be added or substituted for a particular one. We chose this selection as this thesis is predicated on the idea of supporting (knowledgeable) active users. Further, based on our findings, the entailed aspects seemed to be the areas where due to poor design things could go wrong quickly. By designing properly in those areas, however, user rejection might be avoidable and human and machine benefits beneficially reconciled. However, still, even if designs address the issues well, that does not mean that there are no other influencing factors to be expected leading to rejections by users.

For easier reference, we also suggested in Section 5.1.3 to use a visual representation in form of a radar chart that is derived from a decagon depicting the mentioned design variables. Although there are only five design variables presented, we proposed the decagon with a later extension in mind depicting evaluation results (Chapter 6). However, as this approach is only capable of illustrating static design aspects, we further suggested using it in combination with a creative flow diagram as proposed by Deterding and colleagues. In contrast, this type of diagram is also able to capture the dynamic changes in allocating control and/or initiative. As mentioned earlier, those dynamic changes can be used as markers suggesting potential hazards.

In conclusion, the use of representations as suggested above might serve two major functions. First, it might be used as a visual aid that helps to make core aspects of a system fundamentally apparent and also quick and easy to parse. As a consequence of (successfully) fulfilling this first function, it might further enable comparisons between prototypes. A set of distinct dimensions as proposed might enable a common 'language' that enables to talk about distinctions and commonalities more easily and precisely in general. Further, the suggested gradation might also allow for estimating to which degree prototypes are similar or different in relation to a metric or a set of metrics.

Insights

- The degree to which users feel in control or not can potentially be explained by the concept of intentional binding or lack thereof. Potentially, it can push the envelope regarding explanation models presented so far in the great body of work concerned with the design and evaluation of levels of automation in system design and their effects on the experience of users. Fundamentally, it possibly can provide reasoning without too many gaps from empirical observations to the abstractions formed on the design of systems explaining trade-offs emerging therein up to the perceptions by users and the neural underpinnings embedded in them.

- As adaptive automation systems that support being in control and computer-supported adaptive creative exploration can take on a plethora of designs, it not necessarily obvious how these can or should be shaped. To be a bit more precise regarding the details, we derived a set of design variables based on insights we collected in our analysis of status quo technologies, our research on user requirements and our literature review. While, in general, such variables could be used to describe a more extensive set of tools for creative work, we focused on a particular subset, namely user expertise, motor activity, video saliency, automation and prototype fidelity in our work. Those are considered to be relevant to cinematic camera operation by active users. We suggested that those variables can each take on one of the five values that we also proposed. Thus, different designs can be characterised by their instantiation of each variable regarding these static design aspects. For illustrating their dynamic aspects, we suggested the use of the creative flow diagrams as proposed by Deterding and colleagues.

II

Applied Research

JONATHAN IVE (DESIGNER)

At the start of the process the idea is just a thought - very fragile and exclusive. When the first physical manifestation is created everything changes. It is no longer exclusive, now it involves a lot of people.

ALAN KAY (COMPUTER SCIENTIST)

People who are really serious about software should make their own hardware.

ROBERT D. HARE (PSYCHOLOGIST)

Science cannot progress without reliable and accurate measurement of what it is you are trying to study. The key is measurement, simple as that.

6

Prototyping Toolkit and Evaluation Framework

What to expect?

- Overview of our approaches to prototyping
- An introduction to our framework and analysis tool for evaluating cinematic user interfaces

What to take away?

- The fundamentals of our approaches towards designing for and evaluating user interfaces for cinematic camera work

Attribution: This chapter references research that we previously published at INTERACT '17 [122], UIST '17 [142], CHI '17 [123] and the co-located MICI '17 workshop [121] as well as the DroNet '18 workshop (co-located with MobiSys '18) [79].

Our Statement of Collaboration details the differences between the paper(s) and this chapter

6.1 Prototyping Toolkit

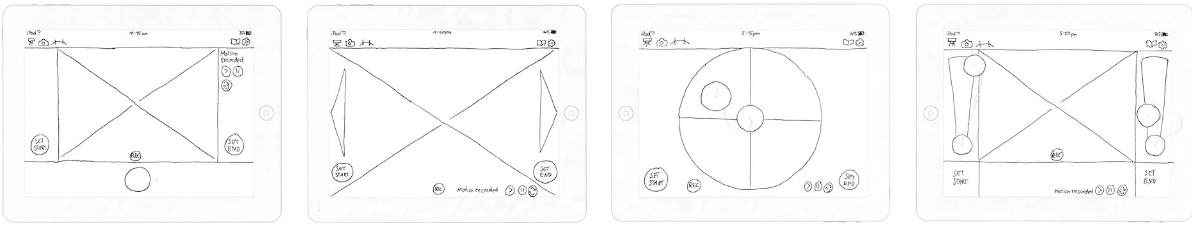
For the projects presented in Chapters 7 and 8 we used varying approaches for their prototyping and implementation. It is commonly known that different types of prototypes allow getting different feedback, but also require different efforts regarding their implementation. In Section 5.1.3, we established that cinematic camera control faces some particular challenges concerning prototyping. Therefore, we think of the prototype fidelity as one of the more important aspects of our work and hence included it in our selection of the design variables. For each of the later on presented projects, we weighed up the alternatives and consequently chose a particular style of a prototype. How suited the varying approaches were, however, was unclear when we started. Therefore, we want to *present our prototyping approaches in a summarising overview along with some lessons we learned*.

6.1.1 Paper

As mentioned in Section 5.1.2, paper prototypes are often a first go-to option when trying out different concepts. While this form is easy to implement, we expected some shortcomings using it with camera experts. However, it was unclear to us, whether a pure paper-only prototype could be beneficially extended by some digital and dynamic aspects resulting in a hybrid format. Potentially, this would allow for a rapid approach while addressing some additional requirements that appear due to the peculiarities of camera motion. Therefore, we will report on some of our experiences with a traditional paper-only prototype and an extended paper-plus prototype that is used on top of a tablet.

Paper Prototypes

As we cooperated with an industrial partner that manufactures sliders and dollies, we aimed at exploring designs for remote control of such particular systems. As a starting point for the design process, we chose a slider as the targeted support tool. Its control is less complex due to the reduced degrees of freedom that one needs to control. Therefore, getting a fundamental understanding of what might be reasonable design choices and what might not be, could and should be established early on before moving on to more complex tools. To come to such an understanding more rapidly, we opted for paper-prototypes. As we laid out in Section 5.1.2, in general, mobile devices are attractive to choose as a hardware platform for to multiple reasons in cinematic production and are already used by many manufacturers (see also Section 2.2.3). To inform the utility of varying basic design approaches for our particular use case but also to inform projects beyond the scope of controlling sliders, we created paper-prototypes of several UI design alternatives having tablets as a hardware platform in mind (Figure 6.1). Those diverged regarding the UI elements used for controlling the position and movement direction of the camera or more precisely the sledge of the slider. A detailed list of the options is found below:



(a) Software-Joystick *(b)* Buttons *(c)* 2D Software-Joystick *(d)* Throttle

Figure 6.1: Prototyping Toolkit: Paper prototypes

(Centred) Software-Joystick: Controls direction (left / right) and speed (distance to centre) of the slider simultaneously (Figure 6.1a)

Buttons: Left and right buttons that control the direction of the movement (Figure 6.1b)

2D Software-Joystick: Integration of direction control (left / right) and speed control (up / down) in a different joystick layout (Figure 6.1c)

Throttle: Vertical speed controls for each direction (Figure 6.1d)

Evaluation

We invited eight participants (4 male, 4 female) to ask them about their preference regarding the design alternatives. The Median (Mdn) age of the participants was 25.5 with ages ranging from 21 to 60. At first, we informed the participants regarding the purpose of the study and details of the procedure. After providing general information on the study, we asked them for their consent before

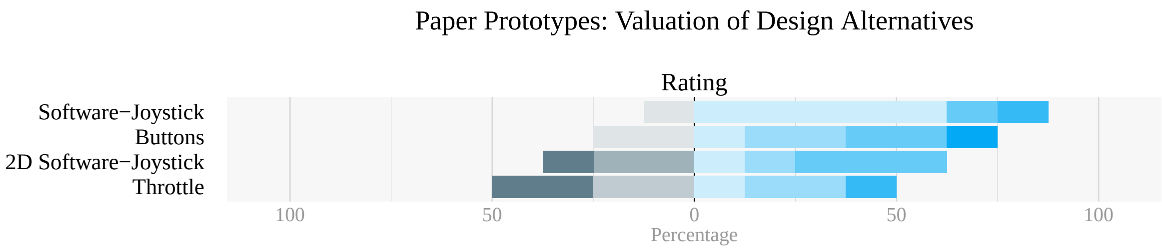


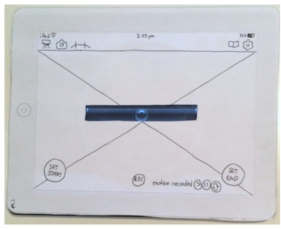
Figure 6.2: Prototyping Toolkit: Results of evaluating paper prototypes

continuing. Having declared consent, we subsequently presented the alternatives to them in counter-balanced order for each participant. For each of the trials, the participants were asked to steer the system given the provided controls, to change some settings and to think aloud while executing the study tasks. Given the limitations of paper, the actual steering was not presented on the prototype, overall resembling more of a cognitive walkthrough. After being exposed to one prototype, the participants were asked to rate it on a neutrally labelled [40] 10-item scale and were subsequently asked for qualitative feedback. The results of the ratings are presented in Figure 6.2.

PaperPlus Prototypes

As mentioned above, we also introduced a hybrid approach where a tablet would provide some dynamic elements to enrich a paper-only prototype. In our case, we used it for two design alternatives (Figure 6.3). In the case of 'Panning' (Figure 6.3a), we used it to explicitly address the problem of being unable to integrate camera motion in a paper prototype. So, in this prototype, we used a

(a) Panning



(b) Level

Figure 6.3: Prototyping Toolkit: Paper and tablet hybrid prototype

pre-installed photo app to display a panoramic image. By panning with one's finger over the display, one was able to control the movement. This form of interaction provides a direct mapping of the necessary input action (for example, moving the finger from left to right) and the result (moving the camera from right to left; dragging the image below one's finger so to speak). In the second example, 'Level' (Figure 6.3b), we wanted to exploit the technical ability to register the tilt of a tablet. The direction of the tilt should be used for controlling the movement direction of the slider and the magnitude of the tilt for controlling its speed. For this control style, we used the spirit level as a metaphor which we displayed in the centre of the tablet. With this design approach, we wanted to acknowledge that looking at the tablet might not be wanted some of the time, maybe operators rather want to the check the scene directly or on an external display while controlling the slider in an eyes-free style.

Evaluation

In addition to the paper-only prototypes, we also presented these two design alternatives in the study described above. We also collected feedback on the preference of the participants via the same 10-item scale. The plots of the results can be found in Figure 6.4.

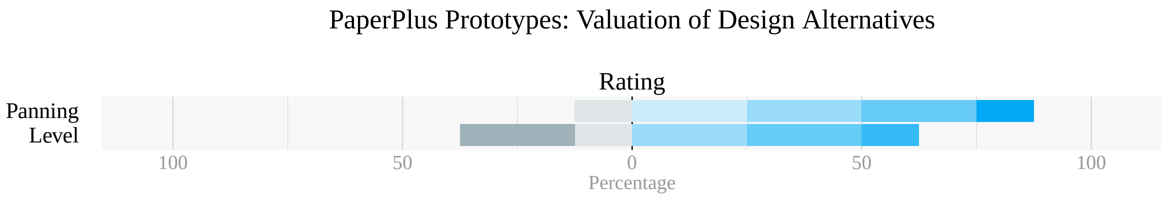


Figure 6.4: Prototyping Toolkit: Results of evaluating PaperPlus prototypes

Data Analysis

We further tested for statistical difference applying non-parametric tests. We tested paper and hybrid prototypes as two separate groups as the conditions varied regarding the proposed design and the implementation. We used Friedman's Test [90, 91] as an omnibus test for both groups. For the paper-only group, we found a significant main effect ($\chi^2(3) = 11.1621$, p $\leq$ .011) as well as for the hybrid group ($\chi^2(3) = 4$, p $\leq$ .046). For posthoc pairwise comparison, we also used a non-parametric test. While Wilcoxon's Signed Rank Test [274] is sometimes suggested as a non-parametric test for pairwise comparison as by Fields and Hole [86], its use after a Friedman's test is also controversial. Alternatives are recommended, for instance, by Sachs [231] such as Conover's [54] or Nemenyi's Test [192]. Thus, we will use Conover's Test for pairwise comparison (with Bonferroni correction) for this analysis and later in this work. Our data analysis was conducted in R (we provide a template of the script in the Appendix; see 12.3). For the paper-only group, we found differences for the Software-Joystick (Mdn $= 6$) × Throttle (Mdn $= 5$, p $\leq$ .024), Buttons (Mdn $= 7$) × 2D Software-Joystick (Mdn $= 6.5$, p $\leq$ .011) and Buttons × Throttle (p $\leq$ .001) comparisons. As the hybrid group only consisted of two conditions, we did not use a further posthoc test. In conclusion, for the paper-only group, a software-joystick or a button seems to be preferred by the participants while panning was preferred in the hybrid group. Integrating the data that we collected with both prototypes, we can generate an overall plot for visual comparison as in Figure 6.5. Here also, the three preferred designs are found in the upper three ranks. Consequently, those should be considered for later exploration.

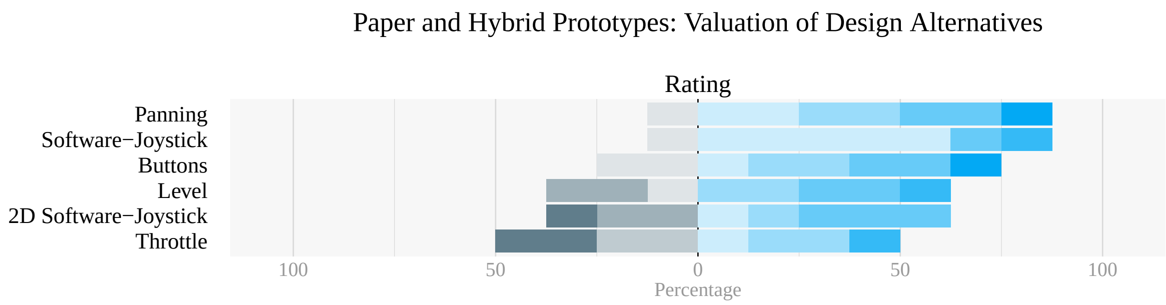


Figure 6.5: Prototyping Toolkit: Evaluation of paper and PaperPlus prototypes

Discussion

While we gathered some insight into user preferences, we also started with questions regarding the use of hybrid prototypes. While in our estimation the difference was not immense, they helped to make an erroneous assumption of ours visible. For the implementation of the panning interaction, we used a standard view of a panoramic image. One of the more advanced users critiqued this particularly: *"this is a pan and not a slide"*. So, while we aimed at addressing the cinematic peculiarities, we picked the wrong approach for a slider. In consequence, we would, therefore, recommend using an implementation with a higher degree of prototype fidelity to address the issues better.

6.1.2 Virtual

To follow up on our last recommendation, but also to address the challenge of incorporating details of the camera motion in prototypes generally, we also used further forms of prototyping with a higher degree of fidelity and report on their peculiarities. In detail, we used two different approaches: a mockup tool and a virtual 3D environment (Figure 6.6). Both environments have in common that they provide UI components built-in, that they can be implemented on a desktop or laptop computer and that they support the automatic export of applications that run on tablets out-of-the-box.

UI Mockup

For the project described in Section 8.3, Axis+Content, we applied the mockup and prototyping tool Framer [330, 517] (Figure 6.6a). The tool allows creating responsive UI mockups that can run across various platforms for mobile devices such as most prominently iOS and Android. Besides a library of UI components and animations that can be used out of the box, it supports the use of custom code.

Discussion

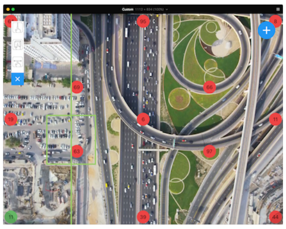
It is a rich environment for the development of UI prototypes at high fidelity and resolution. The ability to integrate custom code made it a helpful tool for our use case as it allowed us to import original cinematic footage. The implementation of a low latency streaming was a bit cumbersome but could be realised stable enough for use in a user study. While we just used one particular shot type, footage for varying shot types can be easily integrated. However, what this approach lacks is 'real' interaction where the displayed camera stream would change in reaction to the user input. For the project we applied it, this was not necessary, but obviously, this can render it useless for other projects.

3D Environment

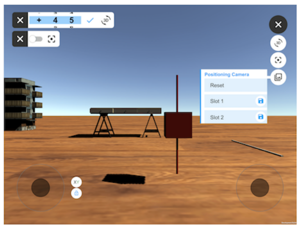
Only a more elaborate implementation can address the latter shortcoming mentioned above. For this purpose, we chose implementations in a virtual 3D environment. While it requires more effort than a mockup tool, the effort is still less compared to implementing a physical one, hence we opted for an implementation based on Unity3D. The environment allows to create cinematic footage even in real-time as displayed at SIGGRAPH '18 [384, 541], provides a library of UI components [540] and the integration of custom source code.

Discussion

This approach provided us with a powerful environment for creating cinematic footage as well as cinematic UIs where the camera stream can also change in reaction to user input. We used in for multiple projects as described in Section 8.1 (TrackLine) or Section 8.2 (Progressive Reduction, Figure 6.6b).



(a) Framer Mockup [79]



(b) Unity3D implementation [79]

Figure 6.6: Prototyping Toolkit: Examples of our virtual prototyping approaches

6.1.3 Physical

While virtual approaches require less implementation effort, they also do not allow collecting data in-situ. Only a physical setup can provide a suited platform for this purpose. It also allows integrating the testing or studying process into the routines that operators already follow on set. This option, however, also requires the most effort regarding implementation as for novel prototypes existing systems often do not offer an open interface. Trying to keep the effort as low as possible also in physical prototypes, we started out using an off-the-shelf PTZ camera. To further provide prototypes also for use on set, we also implemented a semi-automatic slider system (for lateral motion) that also can mount a gimbal (for rotational motion).

Pan-Tilt-Zoom Camera

First, we used an Axis 214 PTZ camera and the provided API to create various web-based and Android-based applications. Both types would be used on an Android tablet, displayed the camera-stream and forwarded touch-events and touch-gestures to the API. In such a setup, users could, for instance, change the panning-angle by applying a pan-gesture on the tablet or steer toward a set of pre-programmed coordinates by using pre-defined gestures.

Discussion

While the implementation was rather straight-forward, we encountered several shortcomings. The set of commands provided by the API was either restricted to coordinate-based control (allowing a smooth transition from start to given coordinates) or, for custom camera travels, required a sequence of multiple angle adjustments to reach a target destination using a form of direct control. For each these steps, however, the system would unavoidably integrate acceleration and deceleration. So that one move from start to end would be chopped into multiple short staccato moves, which is highly undesirable in terms of control and of resulting camera motion. Additionally, the video stream could only be delivered in low resolution and high latencies which led to a poor overall experience although all devices were connected to the same local network. With all these shortcomings emerging, we eventually did not consider further following this approach for prototyping.

Slider

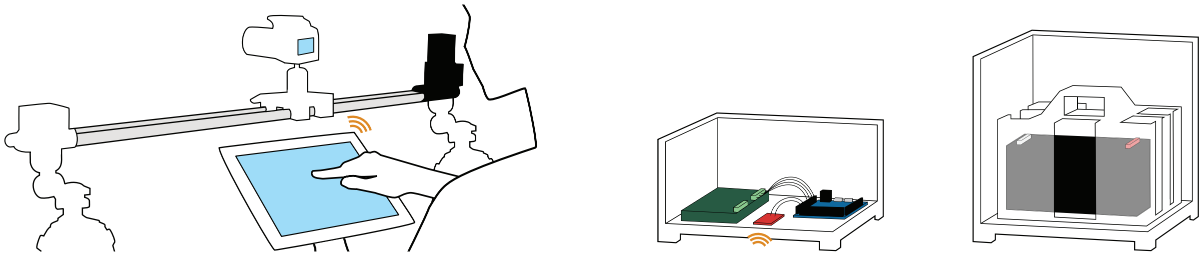
Instead, we decided to implement a slider system based on easily available parts and open source software with the intention in mind that it should be stable enough for shooting on location. The system was also designed to serve as a prototyping platform that allows connecting various UIs wirelessly (Figure 6.7). Further, we intended to focus primarily on the single-user use case we identified earlier in Section 3.2. To inquire more about the requirements, we interviewed a professional operator regarding technical hardware and UI aspects. The operator had more than ten years of experience in camera operation as a professional who was also responsible for the cinematography of major feature films. Additionally, he was a consultant to our industrial partner regarding the production and usability aspects of new cinematographic equipment. Therefore, he was knowledgeable in the domain of tool operation as well as tool production. We derived the list of requirements found below from a 45 minutes long semi-structured interview.

Hardware Requirements	**User Interface Requirements**
Stable for on-set use	Wireless remote control
Smooth motion with stable images	Speed is controlled manually
Runs at constant speed	Programming of moves
Bounces between ends repeatedly	Programming of ramps
Stops before it hits an end	Moves can be repeated precisely
Offers time-lapse recordings	Moves can be saved, loaded and edited
Payload of 20 kg (hor.) and 6 kg (vert.)	Displays the camera stream live

Table 6.1.1: Prototyping Toolkit: Expert requirements for our physical prototyping platform

Actuation Unit: Rail, Slider and Motor

To translate the requirements properly into a physical prototype we teamed up with a mechanical engineer. Together we determined the necessary motor torque and acceleration for actuating the payload of 20 kg horizontally and 6 kg vertically. Before the parts were selected, a safety coefficient of 2.0 was added to the calculation. Based on the calculated parameters, a NEMA 23 stepper motor [339] with 2 Newton metre (Nm) holding torque was chosen. Its acceleration depends on the given voltage (see the curve in [339]). It moves a slider attached to a 1.5 metre (m) long aluminium rail [340] (Figure 6.7a). For tripod attachment, the rail is furnished with ¼ inch tapped holes at both ends and the centre. Both ends are covered with custom-designed 3D printed brackets. These carry the motor, the coupling [352], two gear wheels [355] and four ball-bearings [354] needed for a 9 mm HTD [353] toothed belt transition from the motor-axis to the slider. They also carry two limit switches to sense the position of the slider near the ends for its safety deceleration. One case also contains Infra-Red (IR) Light Emitting Diodes (LEDs) for remote triggering of the camera release in time-lapse recordings.



(a) Actuation Unit *(b)* Control Unit *(c)* Power Unit

Figure 6.7: Prototyping Toolkit: Actuation, Control and Power units of our slider system

Control Unit: Motor Driver and Control

A Leadshine DM556 [348] hardware motor-driver powers the motor. For smooth motion, the driver supports micro-stepping up to a factor of 256. It is actuated by an Arduino Mega [309] with a Bluetooth Low Energy module [307] attached for remote control, an EEPROM extension for a permanent memory of settings and a low-pass filter circuit for smoothing signals from the limit switches. All mentioned parts – except the motor – are covered by a 3D-printed case (Figure 6.7b).

Power Unit: Line Current or Batteries

The system can either be powered by line current through a Mean Well SP-320 power-supply [358] or by two lead-acid batteries (12 Volt (V) / 7.2 Amper hour (Ah)) (Figure 6.7c). The batteries are covered by a further 3D-printed case with a built-in display showing the battery status. The case of the power unit has the same size in width and length as the control units. It further has fitting slots for the control units feet, so that the cases can be securely stacked on top of each other.

Software: Arduino Code, Libraries and Wireless Protocol

The Arduino runs the AccelStepper [305] library as software motor-driver and implements the features determined as requirements. To calibrate the system, the slider is moved until it hits a limit switch on each side. Counting the executed motor steps in-between, the total length (in steps) is determined. By ongoing step-counting during use, the slider position is updated. For time-lapse, the Multi-Camera IR library [386] is integrated. It remotely triggers a camera release by emitting signals through the IR LEDs and supports various manufacturers. Via Bluetooth, a bi-directional connection over Serial and a protocol for wireless control is provided. All of the source files of the Arduino firmware, the OpenSCAD files of the 3D printed brackets and wiring diagrams of the control unit are provided electronically [336].

Discussion

We conducted an exploration of our system on set with expert users in multiple shootings. We report on our findings in detail in Section 7.1.

Gimbal

To further enable moves based on rotations, we subsequently used a gimbal as the hardware platform (instead of a PTZ system). Most fundamentally, our gimbal system incorporates two cameras. One that is used for recording and a second one for steering the system. The second camera generates a stream of raw data that is forwarded to an encoder. The encoded stream is then further forwarded to a steering controller and a UI for display. On the UI, users can also input steering commands which are incorporated by the steering controller. Together with an analysis of the visual stream, the user input results in actionable commands that are subsequently executed by a further actuation unit. The video stream analysis incorporates various tracking algorithms for identifying objects of interest and steering algorithms deriving commands framing the objects some of which we presented earlier in Section 4.2.3. The video analysis is executed on an external data processor for reasons of performance.

Hardware: BaseCam 32bit Controller and Raspberry Pi

As a platform, we used the CameTV 7800 gimbal. The stabiliser is based on a three-axis gimbal design which is capable of carrying a DSLR camera with lenses and remote-controlled units attached. The axis-stabilisation is controlled by one brushless motor for each axis which is powered by a 12 V lithium polymer battery. The battery further powers a controller board. In our case, a BaseCam Electronics 32bit gimbal controller board (BGC32) [313] is used. It runs an open-source firmware and provides the necessary interface that we need for steering based on visual properties. We attached a further tracking unit to the gimbal which runs on an ARM-based single-board computer (Raspberry Pi 3 [366]) and uses an 8-megapixel tracking camera (IMX219 sensor [372]). Those components are powered by a separate 33 Watt hour (Wh) battery pack. The Raspberry Pi and the controller board can communicate via a Serial interface (Universal Asynchronous Receiver Transmitter (UART)). To alter the different logic levels of the boards (5 V for the BGC32 and 3.3 V Raspberry Pi), we use a BSS138 logic level converter. The Raspberry Pi further communicates with other systems via Wi-Fi and transmits the live image and tracking data to the monitoring UI.

Software: Streaming, Analysis and Actuation

The software implementation consists of a server, a tracking, a streaming and a Serial API component. The components are implemented in Python 2.7. Further, we developed a cross-platform UI based on current web technologies. The server is built on the Tornado web framework [382] providing input and output interfaces. In particular, it provides endpoints for the video streaming, the REST-API and the web-socket server. It is also responsible for the thread management and serves as a controller between the other components. In addition, it manages the properties of the system and hosts the on-board tracking and steering component. A command-line interface is used for its control, for instance, to set up basic parameters such as the network port, image dimension, compression quality, frame-rate and camera type.

Our first implementation of the encoding unit was based on FFMPEG [328] to generate a MPEG-1 stream for further distribution via web-sockets. This approach, however, was unsatisfactory regarding multiple aspects such as the low image quality and resolution (640×480 pixel (px)), the low frame-rate (< 10 fps) and too much latency (> 3000 milliseconds (ms)). Consequently, we opted for an alternative implementation applying a custom MJPEG streaming module running on top of the Tornado framework. Our implementation can now be characterised by low computational costs, intra-frame compression that fosters high image quality and robustness in fast-changing motions. Therefore, we considered it well suited for processing on low-performance hardware such as the Raspberry Pi and a cinematic context. Our second implementation stably handles up to 30 fps at full HD resolution (1920×1080 px). The MJPEG format further is highly compatible with a broad range of browsers and clients.

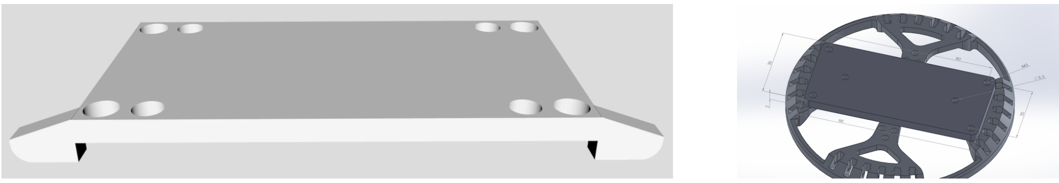
We use three types of tracking: colour tracking, face detection (based on the Viola-Jones Haar classifier) and face recognition. From those, we drive the coordinates of objects of interest. All are implemented in the OpenCV library (v2.4). Additional improvement could be achieved by incorporating a Local Binary Patterns Histogram classifier for face recognition. The computing performance on our hardware could further be increased by creating a custom OpenCV build that allows exploiting the hardware acceleration features of the Broadcom BCM283 (ARM Cortex A53) chipset on the Raspberry Pi. All sensing data is processed within the steering controller which forwards control commands to the gimbal controller board via UART. Since available API implementations for the UART interface were either incompatible with our setup or unsatisfactory regarding stability and functionality, we implemented a custom serial API based on Python 2.7. It follows the SimpleBGC API specification (v2.5) and supports SimpleBGC 32bit and 8bit controller boards alike. Our implementation allows for fine-grained control of movement regarding velocity and speed as well as the control of absolute angles (using a zero-point measure defined during the boot of the system) and relative angles (based on the actual position) for the front-to-back (roll), side-to-side (pitch) and vertical axis (yaw).

Discussion

While the implementation was challenging, we could manage to increase the overall performance to an acceptable level. Most prominently, we could handle full HD stream at low latencies which is critical. Given our implementation so far, two further areas for improvement emerged. First, the performance could be further increased by externalising the stream processing to the more powerful devices. Secondly, instead of a camera recording motion pictures also other capturing devices such as eye-trackers or motion sensors could be mounted. Therefore, our platform can be of use in various other research areas than just cinematic use cases.

Combining Slider and Gimbal

To be further able to attach our gimbal system to our slider, we asked a damper manufacturer [323] for a custom part (Figure 6.8b). The part is made from aluminium and based on a combination of a wire rope isolator design and our slider baseplate (Figure 6.8a). The drilling of the top platform is fitted so that we can mount our gimbal (upside down) while the drilling of the bottom platform matches our existing slider system. Both platforms are connected via a wire rope which not only functions as a pure connection but additionally isolates vibrations. Altogether, the part not only serves as an adapter between the tools but also contributes to smoother motion.



(a) Baseplate   *(b)* Baseplate with damper

Figure 6.8: Prototyping Toolkit: Our baseplate (Figure 6.8a) and baseplate with damper extension (Figure 6.8b)

6.1.4 Lessons Learned

The different prototyping approaches led to some insights and below a summary of the most important one is found. A more detailed discussion of our prototyping toolkit is found in Section 7.1.

Lesson 1: Hybrid prototypes can add fidelity and help with prototyping a restricted set of moves. However, already for similar but different moves such as pans and slides, they can be confusing.

Lesson 2: Virtual prototypes, in contrast, are not limited regarding the range of moves. They provide a good balance of effort and outcome and can be applied to a cinematic context (better than paper prototypes). However, they cannot be used for live shootings on location.

Lesson 3: Physical prototypes, can address the latter, but can also be cumbersome to implement; especially as low latencies are a requirement that is hard to meet at times.

Lesson 4: Utilising a relatively simple mechanical adapter that connects a slider and a gimbal already allows to build a platform capable of translating (1 DOF) and rotating the camera (3 DOF). This enables a greater variety of shots making the tools similarly expressive as virtual prototypes which is central as mentioned in Section 4.2.

Lesson 5: Our platform is not limited to prototyping in a cinematic context. For instance, as presented by Khamis and colleagues [142], it can be used to enhance gaze-based interactions when an eye-tracker is mounted instead of a video camera.

6.2 Evaluation Framework

The projects presented in Chapters 7 and 8 are not only built on our prototyping toolkit mentioned above but also evaluated based on a framework that we present below. Evaluations of cinematic systems and UIs have been previously carried out but in very heterogeneous ways. Due to this diversity on the one hand and a (too strong) focus on a subjective interpretation on the other, the results become hard to compare at times. The diverse evaluation approaches can, for example, take the form of visual documentation (often a printed image sequence), an emphasis on estimating the technical capabilities of a system but also user tests. For instance, visual documentation is often found but, in the end, requires the viewer to form personal judgements that hardly can be analysed in a structured way. *Our approach seeks to understand cinematic tools and UIs from various perspectives, evaluate them accordingly and to integrate the perspectives into a larger framework that fosters comparability*.

6.2.1 Perspectives

As introduced, evaluating cinematic tools can be done based on various perspectives. Mainly based on related work already mentioned in Chapter 4, we identified some of the most common perspectives and their associated methodological approaches below. For example, understanding the tools as a means for 3D navigation, the evaluation criteria already used in this domain can be applied. For instance, Jankowski and Hachet [129] also summarised evaluation techniques. Their summary incorporates usability inspection methods such as cognitive walk-through and/or heuristic evaluation and usability testing methods such as viewpoint control comparisons and/or semi-structured interviews. As introduced in Section 4.1.3, one could understand the tools alternatively as systems incorporating a particular or varying degree of adaptive automation. Thus, the workload-unpredictability trade-off suggested by Miller and Parasuraman [180] should be of key interest in evaluating a system. While estimating workload could be done using methodologies that have already been commonly used in the past, estimating unpredictability might be more difficult or of minor interest per se. As detailed in Section 5.1.1, in our estimation, unpredictability can also affect the sense of control of users. Consequently, estimating its effect could be a more resourceful approach in evaluating the latter part of the trade-off. Further, one could also see a device as a Creativity Support Tool (CST) [221] or MICIs [69] as mentioned in Section 4.2.4. In that regard, they can either be seen from a more general perspective or a domain-specific one. Understood as a more general CST, a questionnaire-based approach as detailed in the following section (6.2.2) seems reasonable. However, when understood as a domain-specific tool, also domain-specific evaluation methods become necessary. For cinematographic support tools, this could be done in various forms such as visual documentation. As those are hard to compare, we presented an alternative approach that stems from evaluation automotive UIs below. It aims at deriving objective measures for comparing alternatives in a cinematic context.

6.2.2 Methods and Measurements

Depending on the chosen perspective, varying evaluation criteria become of central interest. Along with the different criteria, a selection of different methods and measurements follows. So, for the perspective of 3D navigation, criteria estimating the quality of control might be central. Seeing tools, however, rather as systems incorporating a degree of AA, determining workload and sense of control can be most relevant. When they are understood as CSTs, consequently measuring the level of creativity support is vital. For each of the mentioned criteria, we will detail the methods and measurements below. In studying systems, those are not mutually exclusive and can also be used in combination.

Quality of Control

The (level of) quality of control is usually a measure that is derived from a structured and goal-oriented task. Varying tasks and measures outside a cinematic use case can already be found in the literature. While task-completion times, error-rates or margins of error are often used, additional measures acknowledging peculiarities of cinematography might also be considered.

Existing Methodologies

For instance, Stanciu and Oh [250] asked participants to use an unassisted and their assisted camera crane to follow a target that was mounted on an industrial robot and moving on a trajectory resembling the figure eight on its side (or infinity symbol). They report on the experiment using image sequences of the results. Authors such as Hulens and colleagues [126] used a target mounted on an automatically driven slider to test the capability of their motorised camera slider framing the target according to the Rule of Thirds. They present their results in the form of a video [494]. While those approaches already describe suited tasks, the representation of the results hardly quantifiable and comparable.

Adaptation

To provide a quantifiable and easier to compare method, we adopted an approach from the automotive domain: the Standard Deviation of Lateral Position (SDLP) as proposed by Verster and Roth [266]. Its basic idea goes along the following lines. While executing a driving task, participants are likely to deviate from the ideal position, which is often the centre of the lane (Figure 6.9a). These deviations can be measured and collected over time. Integrating the deviations one can infer the quality of (driving) control. The closer the trajectory of the participant matches the ideal, the higher the quality of control and vice versa. We adapted this approach to a particular technique of cinematic production. Instead of staying close to the lane centre, our adaption is based on the Rule of Thirds. While it can be seen as an aesthetic heuristic, it can also be understood as a goal-oriented operation task. Similar to the SDLP, one can estimate how close, for instance, a face is to the ideal position, which is the first third in movement direction within the image space (Figure 6.9b). Consequently, one can be derived a score for the quality of control when collecting data over time: the smaller the distance from that

(a) SDLP: deviations (blue) from lane center *(b)* Cinematic-SDLP: deviations from desired framing

Figure 6.9: Evaluation Framework: The SDLP approach and our cinematic-SDLP adaptation

ideal position, the higher the quality of control. We used the approach to estimate quality of control, for instance, in projects reported on in Section 7.2 (Exploration 2) or Section 8.1 (TrackLine).

Additionally, the quality of control can also be derived differently. One can also use the number of retakes that are necessary as a data source. As, in practice, time and budget are limited, it is highly desirable to use only a minimum number of shots to capture a scene. We report on our use of this approach (together with the presented cinematic SDLP adaption) also in Section 8.1 (TrackLine).

Workload and Sense of Control

In a larger picture, the effects of the design of a system given the workload-unpredictability trade-off are substantial. Measuring the effects can be divided into metrics of workload and sense of control. For evaluating both, various methodologies exist. Those most profoundly differentiate into explicit and implicit measurements techniques. As both approaches have peculiar advantages and shortcomings, we would recommend an informed choice and hence report on a range of options.

Explicit Measurements

Since its development in 1988, the NASA Task Load Index (TLX) [108, 109], is a questionnaire that is still often chosen by HCI researchers today. In short, it asks participants to rate a system using 20-item scales across six dimensions: mental demand, physical demand, temporal demand, performance, effort and frustration. Participants are also asked to weigh the importance of different dimensions. This step can also be omitted which is then often referred to as a Raw-TLX format. A comparative study presented by Moroney and colleagues [190] could not establish that reports differ due to the variation in the report format (Full or Raw TLX). Based on either approach, summarising the data, an overall taskload score is calculated which can be used to compare design alternatives.

Similarly, a sense of control can be estimated using a questionnaire. For instance, Dong and colleagues [72] provide a Sense of Control Scale (SCS). They derived the gradation of the scale and the wording from a user-centred design process. In contrast to the TLX, the SCS is only measuring on one dimension. Thus, its form can make statistical comparison harder as the scale might not be considered equidistant which in consequence does not fulfil requirements of parametric testing [35].

Despite a thorough search in the literature (as reported by Seliger [304]), we were unable to find a comprehensive list or set of dimensions that would allow constructing of a multi-dimensional sense of control score. Though we found a broad range [64] of well-established questionnaires on similar aspects such as the locus of control [271] or the sense of authorship [1].

Implicit Measurements

Asking directly on certain items can reveal the study interest and lead to interferences due to reporting bias or social desirability bias. If the measurements are taken continuously throughout the study, the carried-out task needs to be interrupted also leading to biased data [61]. Further, if the data is collected only after executing a task, it can be skewed as it represents rather an 'average' of the experience omitting its peaks and valleys, sometimes it might only be able to collect the data after a delay further influencing a sample. Therefore, deriving data from other sources during task execution is attractive to researchers. For workload estimation, this can take on the form of measuring physiological reactions such as eye activity [263], pupil diameter [213] or the heart rate [276]. However, these approaches can require much effort in setting up the sensors properly to measure data accurately. Thus, also alternatives are used such as Detection-Response Tasks (DRTs) [55]. This approach is based on a stimulus-response model. Participants are given a primary task such as steering a car and, as a secondary task, are asked to react to the flashing of a light source (as fast as possible) usually by pressing a button. The time needed is logged, and the workload inferred: the longer it takes the participants to register the visual cue and react to it, the larger the overall workload.

For the sense of control, such measurements can be taken via the interval-estimation approach as detailed in Section 5.1.1. As the method of asking participants to estimate the interval between two cues is also fundamentally based on stimulus-response model, it would be desirable if both measurements could be integrated into a single sequence. This would allow gathering data on both aspects central to the workload-unpredictability trade-off implicitly in one go. This approach would entail the ability to measure multiple data points continuously during a study. We gathered experiences with such setups and report on it in Sections 6.2.4 and 8.1.

Creativity Support

For the evaluation of CSTs various recommendations and domain-specific approaches and limitations exist. As proposed by Hewett and colleagues *"one should use more than one methodology to allow one to compensate for the weakness of a single methodology"* [111]. They further suggest triangulating across a selection of experimental, biographical and contextual methods and promote that methods should be combined where ever possible to answer the questions that invoke the evaluation process in the first place. An example of such an approach can, for instance, be found in the work of Terry and Mynatt [258]. They designed two CSTs and evaluated them with different approaches: they first used a lab study with a dedicated goal-oriented task that allowed inferring tool performance. Then they conducted another study with an open-ended creative task that allowed estimating tool exploration and expressiveness. Further, they paired participants to study collaborative aspects.

Explicit Measurements

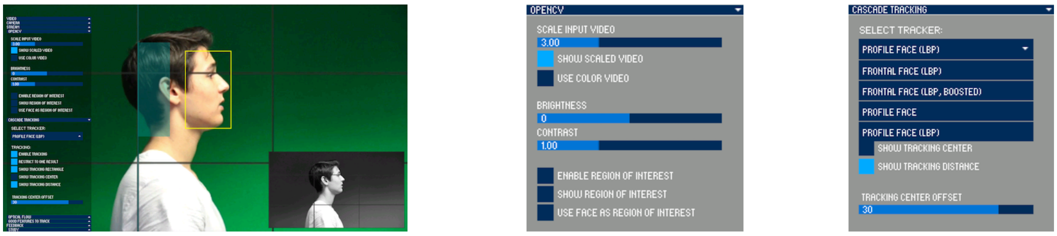
Collecting data by asking study participants directly can also be considered in such a process. For example, Terry and Mynatt additionally used the TLX to determine the workload of their tools. To estimate creative aspects, Cherry and Latilupe [47] provide a further relevant questionnaire with the Creativity Support Index (CSI). They derived the basic design from the TLX. Therefore, they also use 20-item scales across six dimensions. In their case, those are collaboration, enjoyment, exploration, expressiveness, immersion and results worth effort. Following the fundamentals of the TLX, the items can be weighed pairwise, and an overall creativity-support score can be calculated. Also similar to the work of Dong and colleagues, they used a user-centred process to determine the wording.

Implicit Measurements

One of the prime implicit measures in creative contexts is probably plainly evaluating the results of the process. However, as mentioned earlier, such an approach is subject to personal bias and also might be more concerned with the outcome than the tools. Focusing on tool evaluation, we conducted a literature survey and inquired on how one could apply further implicit measures that contribute to a more objective perspective. The survey is reported by Sachmann [303] and tries to address the particular question of how the CSI could be strengthened by adding implicit measurements for each of the proposed dimensions (following the recommendation of Hewett and colleagues). In evaluating collaboration, one could draw from methods used to evaluate Computer-Supported Collaborative Work. Here, for instance, one could additionally use observational techniques regarding groupware observational user testing as suggested by Gutwin and Greenberg [103]. Concerning enjoyment, implicit measures derived from facial expressions [78] or EEG data could be used. To clarify, regarding creativity and brain activity, there is a common notion that the right brain hemisphere and creativity are linked. As this is a misconception, using EEG sensors in such scenarios is not to be recommended. However, Pope and colleagues [216] identified three frequency bands (alpha, beta and theta) that interact in engaging tasks which might be used as a proxy measure. The literature further indicates that it is the different peculiarities of the creative tasks (such as composing music or looking at a painting) that seem to influence with specific brain areas are disinhibited [71]. Concerned with exploration, one could use the approach of Terry and Mynatt as a blueprint. They also observed and logged how much of the available tools participants actually used and for how long to derive an implicit measure. Alternatively, the level of exploration can also be reflected by the number of tries or results created as proposed by Andolina and colleagues [4] or Tausch and colleagues [256]. In contrast, how immersive a tool is, can be estimated using eye-tracking data as suggested by Cox and colleagues [59] or Jenett and colleagues [130]. Regarding expressiveness and results worth effort, we hardly found options for evaluation besides evaluating the results or asking users explicitly.

6.2.3 Evaluation Tools

For our two main propositions regarding evaluation, measuring the quality of control (via an adapted cinematic-SDLP) and measuring sense of control (via an integrated implicit workload and intentional binding estimation approach that is based on an adapted DRT), we developed a set of assistive tools. For each approach, we introduce our tools below.



(a) Distance estimation *(b)* OpenCV settings *(c)* Supported cascades

Figure 6.10: Evaluation Framework: Our analysis tool for a cinematic-SDLP with CV settings and cascades

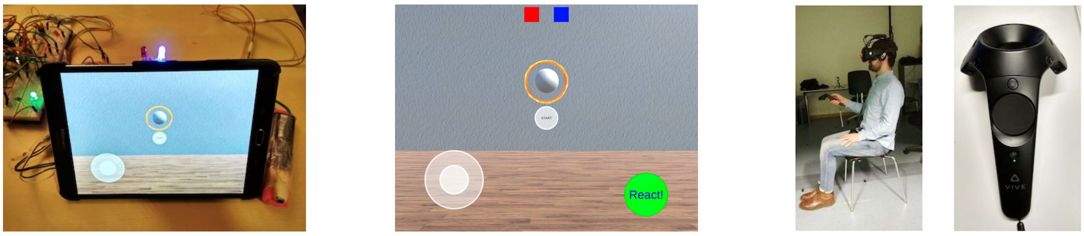
Estimating Quality of Control

Developing an apparatus for our cinematic-SDLP approach (Figure 6.10), we started with particular assumptions in mind: far, medium or close shots of a body or a face of a single person walking by who is framed using a tracking shot. Restricting the use cases to a set that only involves shots of a single person made the tool implementation less cumbersome. These assumptions made sure that a person can be expected in the visual material, but also only one. Further, no other object detection algorithms need to be supported besides body or face detection. In future work, the tracking of one particular face among multiple could be addressed by incorporating face detection as mentioned in Section 4.2.3. Further, tracking cascades for other objects can easily be integrated. After a video is loaded, the user can select a tracker from the following cascades: frontal faces, profile faces, eyes, ears, mouths, full bodies, upper bodies, lower bodies, pedestrians (Figure 6.10c). The walking directing is automatically determined and hence the relevant third inferred. The program provides visual feedback regarding the offset to the ideal position and automatically logs the values into a .csv file for later analysis. To increase performance with large video files, the face-detection is performed on a subsampled stream of the video (black and white in Figure 6.10a). The scale factor can also be changed from in the UI (Figure 6.10b). To accommodate for interference due to the illumination, the contrast and brightness of the subsampled stream can also be changed. To further increase performance it offers to define a region of interest. Image analysis is subsequently only performed in this area of the image space. As our tool (which is also available electronically [337]) is implemented in Processing[1], the calculations could additionally be parallelised and outsourced to a graphics board (via JCUDA [345]).

[1]Processing (v3.0.2) with custom OpenCV library (v3.1.0)

Estimating Sense of Control

We thought of our evaluation tool for estimating sense of control as a study tool that should be able to serve as a tool for general use in HCI research. To meet the objective, we developed three tools instances each concerned with a different control modality as it is commonly found in HCI. Each tool implements the same evaluation principle (see Section 4.3) but in varying environments (Figure 6.11).



(a) Physical button *(b)* Virtual button *(c)* HMD and VR controller

Figure 6.11: Evaluation Framework: Study tools for estimating sense of control in different environments

As we wanted to integrate implicit interval estimation into a DRT, we first developed an instance that was modelled after a common (physical) DRT setup (Figure 6.11a). We built it on top of an Arduino-based implementation with open-source software as provided electronically by Krause and colleagues [383]. Our second implementation translates the basic idea onto a touch-interface displayed on a tablet. In particular, we chose a Unity3D-based environment (Figure 6.11b) as it is closest to our cinematic use cases and prototyping toolkit. Thirdly, we chose an implementation that is done in a VR environment (Figure 6.11c) that is also based on Unity3D but requires a Head-Mounted Display (HMD) and physical VR controller. Overall, with this selection of environments we intended to account (in parts) for the broad range found in the HCI discourse.

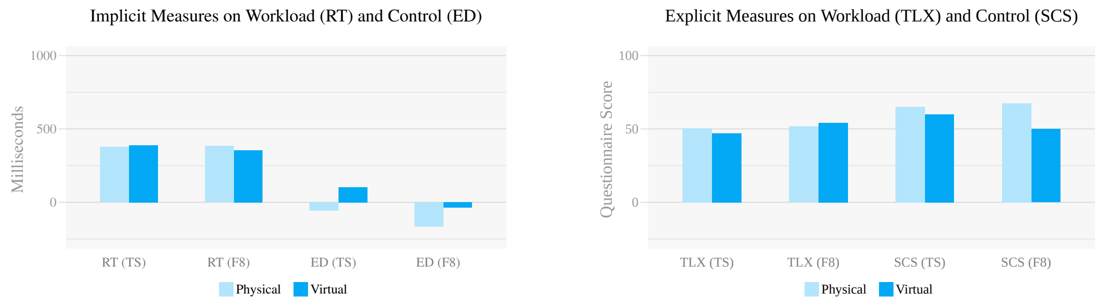
Implementing DRTs in different environments is no novelty. For instance, Conti and colleagues [55] used different setups implementing driving simulators and compared them in a user study. During the study they asked the participants to execute the same study task with the different implementations, to see where differences emergence due to varying implementations. Shortly summarised, they could not identify an effect on the reactions of the participants or diverging interpretations after data analysis due to the changing conditions.

Following their approach, we also conducted studies comparing our implementation with comparable study tasks to check for differences in the results. We used a set of two tasks as presented in the related work: a Rule of Thirds tracking shot as by Hulens and colleagues (with an additional change in movement direction) and a figure eight on its side trajectory as in the work of Stanciu and colleagues. We report on the procedure and results in Section 6.2.4 below.

6.2.4 User Study 1: Physical and Virtual Interfaces

While our first tool was merely intended to be an analysis tool, subsequently its usefulness depends, most fundamentally, on the quality of the measurements that it produces. Those, in turn, depend directly on the quality of the technical implementation. To enable external inspections regarding quality aspects, we made the source files openly available [337]. Also, as it is mainly a technical matter, we refrained from conducting further user studies examining the tool itself. However, as mentioned above, we conducted further studies investigating our implicit sense of control tools similar to the example of Conti and colleagues and report on our studies below.

We first conducted a user study comparing our physical and our purely tablet-based approach. We planned to test our VR condition in a second study that should be informed by the first. The two tasks were chosen as they require controlling different degrees of freedom from the user and therefore should cause varying and distinguishable degrees of workload. We recruited 15 participants (13 male, 2 female). The Mdn age of the participants was 24, with ages ranging from 22 to 31. We asked them to report on how experienced they think they were with tablets: the Mdn was 4 (on a scale from 0 to 6). Further, two of the participants were left-handed. For the physical condition, we built an Arduino-DRT as provided by Krause and colleagues and extended it by implementing the features necessary for the measuring of intentional binding. As a LED was already used to provide the visual cue for the DRT, we attached another one for the second visual cue. A physical button was provided to the participants to indicate their registration of the first cue. The microcontroller would register a button-press and also log the resulting reaction time. In the virtual condition, the LEDs would be translated to rectangles that were displayed at the top of the tablet screen. Ahead of conducting the study, we welcomed the participants, introduced them to its details and handed out a consent form. Having declared consent, we handed out a demographic questionnaire and provided a five-minute training phase with the interfaces that also contained interval estimation trials. Subsequently, we exposed the participants to all interfaces (physical, virtual) and tasks (tracking shot (TS), figure eight (F8)) in counter-balanced sequences while logging the implicit data. Integrating the DRT approach, a visual cue was presented at random intervals that participants were instructed to react to as fast as possible. After presenting cue and registering a press of the reaction button, a second (pseudo-)random interval followed after which another visual cue would appear (implementing the intentional binding measure), and participants were asked to estimate the interval that has passed. For each task, the reaction time, the actual time interval and the estimated time interval was recorded. After each trial, we asked the participants to also explicitly rate each interface on questionnaires determining workload (Raw-TLX) and sense of control (SCS adapted to a 20-item scale to be consistent with the TLX format). In contrast to our previous mentions, we further did not use the CSI as our task was only goal-oriented and did not focus on creative exploration or outcome.



(a) Data from implicit measures (adapted) *(b)* Data from explicit measures

Figure 6.12: Evaluation Framework: Results of the study in a physical and virtual environment

Regarding implicit measures on workload, we applied parametric significance testing as the data measured (times) can be regarded as equidistant. As we varied user interfaces and study tasks in a within-subjects study design, we analysed the data with a two-way repeated-measures Analysis of Variance (ANOVA). As for the non-parametric tests, we conducted the analysis in R and a template of the script we used is also found in the Appendix (12.4). We found no significant main effect for a user interfaces × task interaction ($F = 0.211$, $p \leq .647$). Further, we only found a significant main effect among the user interfaces ($F = 404.616$, $p \leq .001$). Pairwise post-hoc comparisons using Tukey's Test (within the same task) indicate a difference between the UIs in each task ($p \leq .001$ for both with $z_{TS} = 14.457$ and $z_{F8} = 13.962$ [$\text{Mean}(\text{Mn})_P(TS) = 376$ms, $\text{Mn}_V(TS) = 698$ms, $\text{Mn}_P(F8) = 384$ms and $\text{Mn}_V(F8) = 664$ms]). However, in the physical condition participants could already place a finger on top of the button without pressing it. This behaviour was not possible in the virtual condition as it would immediately result in a button press. Consequently, the participants needed to move their finger to the button. Lee and colleagues [150] who investigated finger stroke time estimates for touchscreen-based devices in-depth refer to this movement as tapping and propose that it takes 310 ms to execute it on average. We used this value to adjust our data for the virtual condition and again applied significance testing. Accounting for the tapping movement, we no longer found significant differences between the UIs. As the explicit measures on workload were collected with the TLX, they result in an overall workload-score that can be regarded as equidistant. Thus, we also analysed it with a two-way repeated-measures ANOVA. We found no significant differences.

The implicit measures that we took on control were also tested using the same approach. We found no significant main effect for a user interfaces × task interaction ($F = .598$, $p \leq 0.440$). Further, we found a significant main effect for the UIs ($F = 31.154$, $p \leq .001$) and the tasks ($F = 25.937$, $p \leq .001$) individually. Pairwise post-hoc comparisons using Tukey's Test (within the same task) indicate differences between the UIs in both tasks ($p_{TS} \leq .001$ with $z_{TS} = 4.506$ [$\text{Mn}_P(TS) = $ -55ms, $\text{Mn}_V(TS) = 104$ms]; $p_{F8} \leq .003$ with $z_{F8} = 3.448$ [$\text{Mn}_P(F8) = $ -165m, $\text{Mn}_V(F8) = $ -37ms]).

Regarding explicit measures, we used non-parametric testing as the (one-dimensional) SCS cannot necessarily be regarded as an equidistant scale. Therefore, we tested the UIs for each task separately with Friedman's Test. For each of the two tasks we found no main effect ($\chi^2_{\mathrm{TS}}(1) = .077$, p $\leq .782$; $\chi^2_{\mathrm{F8}}(1) = 1.143$, p $\leq .285$).

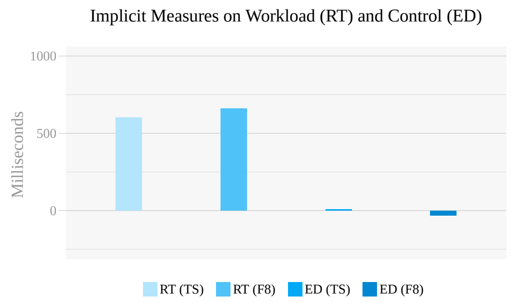
Further, we conducted correlation analyses concerning the different interfaces and the different measures. Focussing on workload and using Pearson's correlation, for the physical input condition the result was .265 for the tracking shot. For the figure-eight trajectory, it was -.281. In the virtual condition, the result was .616 for the tracking shot and .238 for the F8-task. Similarly, we calculated the values concerned with the sense of control. Regarding Pearson's correlation, for the physical input condition, we determined a correlation of -.134 for the tracking shot and of .353 for the figure-eight trajectory task. For the virtual controller, it was .088 for the tracking shot and .187 for F8.

6.2.5 User Study 2: Virtual and Virtual Reality Interfaces

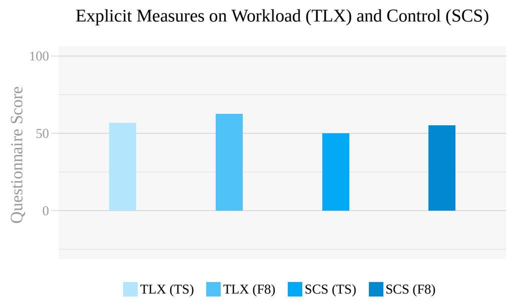
As introduced earlier, we conducted a second study concerned with a VR environment and associated controllers. As the control style is per se hard to compare to our previous tested interfaces, we only compared the tasks and measures within the VR environment. Fundamentally following the procedure of the first study, this time, we recruited 15 participants (9 male, 6 female). The Mdn age of the participants was 24, with ages ranging from 20 to 27. We asked them to report on how experienced they think they were with VR environments: the Mdn was 4 (on a scale from 1 to 10). Further, one of the participants was left-handed. We used Unity3D for the implementation of both task we used previously and again collected data implicitly and explicitly.

For each measure, we tested the implicitly collected data for a difference between the tasks using a paired t-Test and found no significant difference. For the explicit scores we only found a difference for the TLX scores (t = -4.0665, df = 14, p $\leq .002$).

As in the previous study, we calculated the correlation values regarding implicit and explicit measures. Regarding workload, it was -.168 for the tracking shot and -.068 for the F8 trajectory. Concerning sense of control, it was -.334 for the tracking shot and .412 for the F8 task.



(a) Data from implicit measures



(b) Data from explicit measures

Figure 6.13: Evaluation Framework: Results of the study in a virtual reality environment

6.2.6 Lessons Learned

Our studies on the various evaluation approaches led to some insights and below we summarise the most important learnings:

Lesson 6: While established and commonly used methodologies for evaluating cinematic UIs are hard to find, we can draw from established practices of other domains.

Lesson 7: Integrating various relevant perspectives, a set of central evaluation criteria might entail workload, sense of control and creativity support.

Lesson 8: Adapting the SDLP, a measure on the quality of control that integrates seamlessly into a native cinematic task can be derived, and post-hoc data extraction be automated.

Lesson 9: Further, explicit and implicit techniques should be used side by side to counteract the shortcomings of either approach.

Lesson 10: While such an approach seems to be well-suited for workload and sense of control, implicit measures for creativity support are difficult to take and interpret. Therefore, in this work, we further will only focus on explicit measures.

Lesson 11: In our experiments, both ways of measuring (where applicable) would yield results that led to similar interpretations, while implicit measures seemed to be more sensitive (as indicated by related work).

Lesson 12: However, correlations between both measures were only in some cases $\geq .5$ and especially the sense of control measurements seems to be the least correlated. Therefore, the feeling of control (implicit measure) and judgement of control (explicit measure) should be distinctly discussed (also in line with related work). It is noteworthy that explicit and implicit measures for workload are similar, but regarding control, those are more distinct. While they are considered to represent the same measure for workload, this most likely is not the case for control.

Based on the ideas introduced at the beginning of this section and the above-mentioned lessons learned, we derived a set of evaluation criteria to complement our template of design variables presented in Section 5.1.3 (Figure 6.14). Therefore, we suggest including the following criteria in the representation: workload, feeling of control, judgement of control, creativity support and performance.

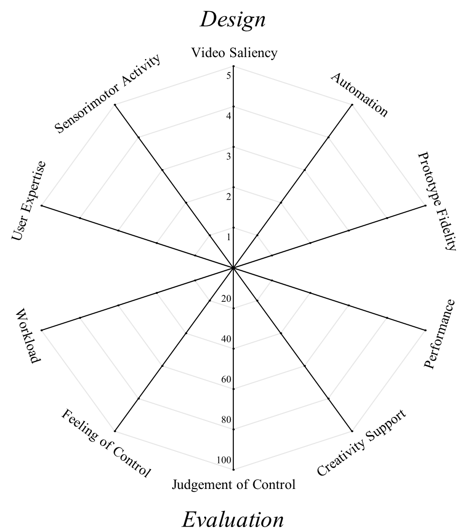


Figure 6.14: Design Variables: Final template representing design variables and evaluation criteria

We further suggest mapping the data collected on the feeling of control to a scale ranging from 0-100. The underlying data is time measured in milliseconds and has no natural end to its scale. However, as for a positive feeling of control only negative estimates are relevant, we suggest omitting positive values as they indicate a (great) lack of control, mapping 0 ms to 0 on the scale and the values of -1 to 300 ms to the range of 1 to 100 to fit the scale.

Similarly, the data representing performance can either be taken from the TLX if measured on an explicit level or from the adapted SDLP. For the latter case, we suggest mapping a value of the difference to the ideal position of 0 to the value of 100 on the used scale as it represents the best possible performance. Regarding the value of 0 on the scale, we suggest mapping it to a difference of 200 px as (depending on current resolutions), an offset of 200 px can, in our estimation, be considered outside the margin of error tolerable for a tracking shot.

These propositions for re-mappings to a particular scale might seem cumbersome to some extent. Not remapping the scales would make it necessary to additionally print the scales next to each axis that is different from the 0-100 scale that the explicit measures share. To avoid visual clutter in the diagram, we consequently suggest this approach.

6.3 Summary

In Section 6.1, we provided an overview of the different prototyping approaches that we applied (Paper, PaperPlus, Virtual and Physical). For each approach, we tried to establish for which type of feedback they were relevant to us and which shortcomings we encountered in applying them.

As expressiveness is a vital aspect of cinematic tools, we developed virtual and physical tools that allow moving the camera in multiple degrees of freedom. While this is easier implemented in virtual systems, the additional effort required for physical applications allows for user testing with experts on location. To also enable other researchers, we, therefore, provide detailed information on the architecture and the implemented parts that (hopefully) fosters reproduction of our physical setups. Further, we shared the software source code that runs on our prototype openly. As we also intended our tools to be purposefully used also outside the cinematographic domain, we encouraged reproduction and customisation. For example, successful reproduction and extension of our system can be found in the work of Khamis and colleagues [142]. We additionally conducted studies to investigate the effects of novel approaches that we took and compiled a list of lessons we learned in implementing and evaluating our systems.

Further, as the evaluation approaches of design alternatives of cinematic UIs are quite heterogeneous and hard to compare, we tried to provide an evaluation framework that potentially fosters a higher degree of comparability across a set of tested design alternatives (Section 6.2). As we think cinematic support tools can (among other things) be understood as tools for 3D navigation, systems implementing adaptive automation and/or creativity support tools, we integrated those perspectives into our evaluation framework. Therefore, we suggested using measures estimating vital aspects for each perspective: performance via quality of control (3D navigation), workload and sense of control (adaptive automation) and creativity support. For each measure, we further tried to provide techniques that would allow measuring relevant data explicitly and implicitly as far as applicable. As both approaches yield particular benefits and shortcomings, they are best used in combination. Further, as tools helping with measurements and/or analysis were not available for every approach we proposed, we implement the missing assistive tools. To foster their use more generally in HCI, we implemented our approaches in various environments, studied the effects of the various implementations and reported on the results in this section.

Insights

- For translating design ideas into prototypes, we explored paper, paper/tablet hybrids, virtual 3D environments and physical setups. While we found paper to be feasible for laying out basic UI patterns, it revealed major shortcomings in a cinematic context. A hybrid approach using paper on top of a tablet helped, but its use should be limited to a particular set of moves (e.g. pans not slides). Virtual prototypes helped to overcome such limitations, and for UI tests in controlled environments, they can provide great benefit with a good trade-off between effort and outcome. However, they can hardly be used in-situ. Only a physical implementation, can address the latter issue. As their development is challenging and commercial products often lack open interfaces, we opted for a custom implementation that we shared openly encouraging reproduction.

- To complement our propositions regarding design variables, we derived a set of evaluation criteria suited for cinematic UIs, namely: workload, feeling of control, judgement of control, creativity support and performance. For each, we presented explicit and implicit measures. Integrating both types is recommended as a complementary use might help to address the shortcomings of either approach being used individually. We further implemented assistive tools intended for general use in HCI that we also share. We studied our approach in varying environments and could not establish effects on the measurements results due to the changing implementations. However, while explicit and implicit measures should be used complementary, one needs to be cautious concerning the sense of control. As the feeling and judgement of it might tap into different neural processes, the measured results might not lead to identical inferences.

ISAAC NEWTON (PHYSICIST)

The best and safest method of philosophizing seems to be first to inquire diligently into the properties of things, and establishing those properties by experiments, and then to proceed more slowly to hypotheses for the explanation of them.

VENKATRAMAN RAMAKRISHNAN (BIOLOGIST)

Science is curiosity, testing and experimenting.

SALLY RIDE (PHYSICIST AND ASTRONAUT)

Science is fun. Science is curiosity. We all have natural curiosity. Science is a process of investigating. It’s posing questions and coming up with a method. It’s delving in.

7

Basic Explorations

What to expect?

- Exploration of our physical prototype on set
- Use in commercial shootings and in the lab
- Studies on various LOAs and the review of results

What to take away?

- Insights from empirical observations on the quality of our prototyping toolkit and on the effects of integrating automation in the design and of adding a phase for reviewing the results in the study procedure

Attribution: This chapter references research that we previously published at INTERACT '17 [122].

Our Statement of Collaboration details the differences between the paper(s) and this chapter

7.1 Exploration 1: Prototyping Toolkit on Set

For the most part, we tested the features that were well-defined empirically during the development. Those were mainly concerned with speed changes, bouncing, safety stops and handling of the payload. Additionally, we thoroughly made sure that they were implemented and working in terms of hardware

(a) Vertical setup *(b)* Horizontal setup *(c)* Ground level setup

Figure 7.1: Exploration 1: Field evaluation of our physical prototype with an expert camera operator in a museum

and firmware as those features were essential. However, in particular two of the requirements we identified were too vaguely defined to be evaluated during the implementation phase. Those were *"stability for use on set"* and *"smoothness of the camera motion (suited for a professional context)"*. *To further determine whether both aspects were also met, we asked a professional camera operator (who was unaware of the identified requirements) to evaluate our setup during assigned shootings.*

7.1.1 Environment and Recordings

To estimate our system regarding the requirements mentioned above, we conducted an expert evaluation on set. Concerning stability, we found that our system worked stably in five assignment shootings for exhibition films in a modern art museum (Figure 7.1). As multiple exhibitions were captured for the assignment, the system needed to be frequently moved between multiple locations with changing points of view, movement directions and speeds. During the recordings, an increased variety of shots was captured on purpose to collect a greater number of alternatives which one could choose from in the post-production. In our case, this was supported as we could already set up the next scene while recording one automatically. This was contrary to the traditional approach of determining fixed shots in the pre-production phase. Therefore, following this alternative approach, it was used to record a broad variety of shots such as horizontal shots at waist level (Figure 7.1b) and ground level (Figure 7.1c), straight (Figure 7.1a) or inclined vertical shots (Figure 7.2c). Overall, our system was used for indoor as well as outdoor recordings (Figure 7.2). Also, different cameras were mounted and moved at a constant speed and automatically bounced between both ends for longer periods of time. The limit switches performed well for calibration and safety measures. Moves and ramps were programmed remotely and wirelessly by a second trained person via Bluetooth (Figure 7.1b).

7.1.2 Results

The footage that was captured with our prototype was used for the production of six exhibitions films that were approved and published by the museum [440, 441, 442, 472, 473, 481]. As the institution sets high standards regarding the aesthetic and technical quality of its representation, we see it as an indicator that our system also met the second requirement (smooth motion). Further, only one recording was scheduled at the beginning of this field test and evaluation. The additional sessions were only initiated by the operator after the results of the first screenings were evaluated. Consequently, we take this as a further indication that the requirement of being suited for a professional context was met. After all recording sessions were finished, we had a summarising debrief with the operator. While he was satisfied with the stability of the system for the most part, also some issues were mentioned where improvements seemed necessary. One of his main concerns was that saving time while recording is vital. Therefore, the system should be extended by shortcuts. In particular, the system should provide shortcuts to positions on the rails such as the left or right end, the centre or at half the distance between the centre and the end. This would be useful as we would often start at one position to frame the beginning of a sequence and then move to end to set up the framing of the end image of the scene. For this repetitive process, we needed to enter the respective coordinates explicitly each time. Further, after setting up the rough parameters of the move, subsequently, we would often try several speed settings to check how the motion properties play out in the material and to create footage to select from in the post-production. Interestingly, we observed a more natural behaviour of people being recorded. This was brought to our attention by the operator when he reviewed the material. Due to the automated system and its wireless control, we could distance ourselves from the physical setup during the recording which might have led to this effect. We had not considered effects on the people in front of the camera ahead of the recordings.

(a) Inclined setup outdoor *(b)* Vertical shot outdoor *(c)* Inclined vertical shot

Figure 7.2: Exploration 1: Outdoor use of our system recording vertical and vertically inclined shots

While the slider was remote-controlled, this was done by a second trained person using a command-line interface (see also Figure 7.3). We primarily focused on testing the hardware and firmware and postponed UI evaluations at that point in time. Those were tested later-on, and we report on it in the following sections of this chapter and more detailed in Chapter 8.

7.1.3 Lessons Learned

From the various shootings we derived some lessons that we learned and that we want to briefly summarise below along with a representation of the system based on our design variables (Figure 7.3):

Lesson 1: We fundamentally could meet the requirements of stability and smoothness

Lesson 2: Shortcuts for position and speed are to be incorporated

Lesson 3: Automated systems might affect not only operators but also recorded persons

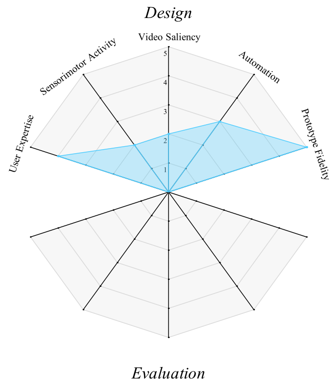


Figure 7.3: Exploration 1: Diagram of our proposed design variables as found in our first exploration (there is no plot for the evaluation as it was qualitative)

7.2 Exploration 2: User Interfaces and System Assistance

Addressing the shortcoming of not incorporating end-user UIs so far, we set out another study in a controlled environment that would focus on the particular effects that those might introduce. We used two UIs with varying LOAs that we compared to a fully manual baseline condition without motorisation.

7.2.1 User Interfaces

We used a tablet as a hardware platform and built interfaces that would display the camera stream and UI components to address the requirements that we previously identified (Section 6.1.3). The different control styles that we used for the remote control UIs were derived from our earlier studies conducted with the paper and paper/tablet hybrid prototypes (Section 6.1.1). Thus, we investigated three conditions: Full Manual control, Software-Joystick and Panning. As the UI concepts were already introduced in more detail earlier, we will only briefly describe them below (see also Figure 7.5).

Full Manual: In this condition the participants would actuate the sledge of the slider by hand (equalling level 1 on our automation scale presented in Section 5.1.3)

Software-Joystick: A status-quo design used to control the motion of the slider horizontally also used to control speed (UI elements might occlude the video stream, equalling level 2)

Panning: While the mapping is similar to the software-joystick, the active area is extended to the whole touch-screen (UI elements might occlude the video stream, but overall less are necessary, equalling level 2)

7.2.2 User Study: Comparing Various Levels of Automation

We conducted the study in a controlled environment and invited 18 participants (14 male, 4 female) with a Mdn of 23.5 years of age, overall ranging between 21 and 31 years. Four participants reported prior knowledge in tools for camera motion.

For the study, we used a within-subjects design and therefore each participant was exposed to all UI conditions. The sequences were counterbalanced based on a Latin Square design to avoid ordering effects.

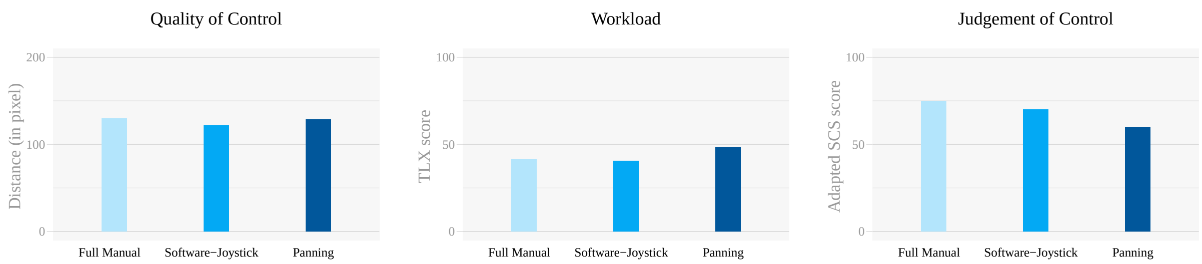
As a platform, we used an unmotorised slider for the Full Manual condition and our motorised prototype introduced in Section 6.1 for the other conditions. Both rails were mounted on tripods (at the same height) and placed in front of each other. As the remotely controlled conditions would display the video stream of the camera, a certain delay was to be expected. To provide the participants in the manual condition with the same delay, we mounted a smartphone on its sledge displaying the video stream. Here, we used the same setup that we used in other conditions. In detail, a Canon Eos 60D captured the scene, and a Teradek Cube 255 connected via High Definition Multimedia Interface (HDMI) was used as an encoder and transmitter. The device encoded a stream using the Real-Time Streaming Protocol (RTSP) which was propagated via WiFi for display on the tablet and smartphone. Additionally, it was recorded at a resolution of 1920×1080 px for post-hoc analysis.

Ahead of the study, we welcomed the participants and briefed them about its procedure. We further explained which data was recorded and how it was handled. We also handed out consent forms, and only once consent was declared, we would hand out a demographic questionnaire and present an introductory video. As the study task, we asked the participants to perform a tracking shot of a person walking by. We also showed an introductory video that would display how the results should look. After the video finished, the participants were exposed to the first UI condition. For each condition, the study task was executed 10 times, and the material was recorded for later analysis. At the end of each trial, we asked the participants to fill in the Raw-TLX that was extended by an adapted version of the SCS (20-item scale). After all trials, we additionally conducted a semi-structured interview investigating user preferences and shortcomings of the presented conditions.

7.2.3 Results

Based on the sampled data, we can provide estimates for the quality of control, workload and judgement of control (Figure 7.4). To derive inferences regarding the quality of control, we used the video analysis tool described in Section 6.2.3. From the log files, we retrieved the distances to the ideal position (in px). We conducted a Shapiro-Wilk to test for normality of distribution which revealed a significant difference between the groups ($W \leq 0.799$, $p \leq .001$). Thus, the normality of the data cannot be assumed, and we consequently used Friedman's Test to test for differences between the groups. We found no significant difference ($\chi^2(2) = 1.143$, $p \leq .179$) with Mdn values of 130 px for Full Manual, 122 px for Software-Joystick and 129 px for Panning (Figure 7.4a).

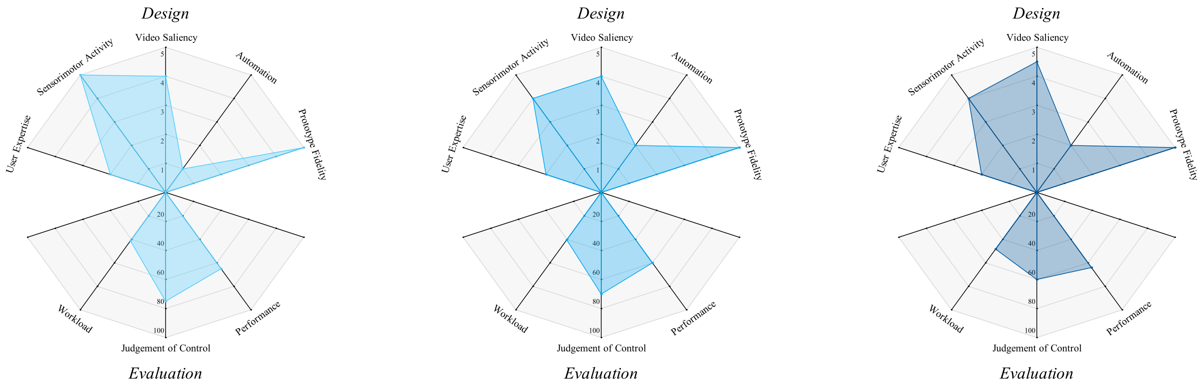
To stay consistent with the analysis procedure on quality of control, we also used Friedman's Test to test for differences between the UI conditions regarding workload. We therefore analysed the overall workload scores as determined by the TLX and found no significant main effect ($\chi^2(2) = 7.00$, $p \leq .030$) with Mdn values of 41 for Full Manual, 40 for Software-Joystick and 48 for Panning (Figure 7.4b).



(a) Quality of Control *(b)* Workload *(c)* Judgement of Control

Figure 7.4: Exploration 2: Results of our exploration of manual and remote control UIs integrating a low LOA

Concerning the judgement of control, we also test for differences between the conditions applying Friedman's Test. We found no significant difference ($\chi^2(2) = 5.03$, $p \leq .081$) with Mdn values of 75 for Full Manual, 70 for Software-Joystick and 60 for Panning (Figure 7.4c).



(a) Full Manual *(b)* Software-Joystick *(c)* Panning

Figure 7.5: Diagram of our proposed design variables and evaluation criteria as found in our second exploration

Compared to a fully manual control baseline, we could not identify any effects on the quality or judgement of control introduced by our remote control UI. Additionally, we could not find any effect on workload as well. These findings seem to be in line with prior findings as by Miller and colleagues [179].

Regarding workload, this was surprising to us as the physical demand is one of the sampled dimensions of the TLX and could be expected to be rated differently given our full manual baseline (which should require more physical effort). Therefore, we investigated the TLX data in further detail. Focusing exclusively on the dimension of physical demand, we found a significant difference when testing with Friedman's Test ($\chi^2(2) = 29.6$, p $\leq .001$). Thus, we conducted post-hoc pairwise comparisons applying Bonferroni correction and found differences between all UIs with p $\leq .003$ or less for each; with a Mdn value of 62.5 for Full Manual, 15 for Software-Joystick and 30 for Panning.

In conclusion, when only focusing on physical demand, we can observe that the remote control approaches could minimise the workload. However, this effect was not strong enough to affect the overall workload score, given our sample, sample size and measurement tool.

7.2.4 Lessons Learned

Lesson 4: No (negative) effects on the quality and judgement of control or workload were found due to the introduced LOA were found; however, this also does not equal proof.

Lesson 5: As we could identify a difference that is to be expected regarding physical demand, our methodology does not seem too coarse (which could explain why we found no difference).

7.3 Exploration 3: Integrating Reviewing the Results in the Evaluation Process

In our first study, we found no difference between the conditions on all the main metrics. This outcome could be attributed to various causes: too coarse measurement tools, a too low degree of automation or an unawareness of the consequences on the quality of control by the participants. As mentioned above in Lesson Learned 5, a possible explanation as to why we did not find a difference regarding the judgement of control conditions could be the coarseness of the measurement tool. However, as we found a difference in physical demand using the same report format, that reasoning could also be called into question. A higher level of automation was not included on purpose and can only be addressed by including other UI concepts. Additionally, we also did not identify effects on the quality of control. However, we also only looked at deviations from the ideal position. Further aspects decreasing the overall quality would be essential to identify as they could influence the judgement of control indirectly [270]. The latter might be possible since we did not include a review phase of results. However, reviewing results is a common practice that we observed in the field earlier. Therefore, to realise that one did perform better or worse with a particular UI might require such a phase. Further, it might be additionally better judged on a larger external display than the smaller internal ones. To look into this idea a bit further, we checked the recordings of the first study and found that shaking and jerky motion was (subjectively) much more visible in the Full Manual condition, but as participants were hardly exposed to these in detail, it probably did not strongly affect their reports on judgement of control much. Possibly an additional review phase during the study (on a larger screen) might allow the participants to judge their own performance better and might affect their ratings.

7.3.1 User Interfaces

To investigate the effects of an increased level of automation and the integration of a review phase, we conducted a follow-up study, that would incorporate conditions addressing these issues. This investigation might help to avoid misconceptions and thus provide a more externally valid result. To address the level of automation issue, we used two UIs in this study (Figure 7.7). As a baseline condition for remote control, we again used the Software-Joystick (this would equal level 2 on our level automation scale). Additionally, we used a keyframe-based prototype (equalling level 4). In both prototypes, we displayed the camera stream on the tablet. In the keyframe-based condition, participants could move the tablet freely in the room similar to a remote viewfinder. They then could select certain positions as keyframes and the motion-controlled slider would then derive one resulting motion path including all the keyframes.

For the technical implementation of such an approach, known points of references are needed. These can be provided several ways for instance via optical tracking or synchronising the tablet with a virtual model of the study room that is updated when the tablet is moved. As such an implementation is technically complex, we opted for a Wizard-of-Oz implementation. We were only interested in the effects on the perceptions of the participants which did not require a full implementation (at least for this study) in our estimation. So, to create the illusion of a system capable of such features, we asked the participants only to select keyframes situated on particular predefined points. We introduced these points at the beginning of the study by showing sample images. This approach allowed us to use a preprogrammed motion path that would be executed after user input.

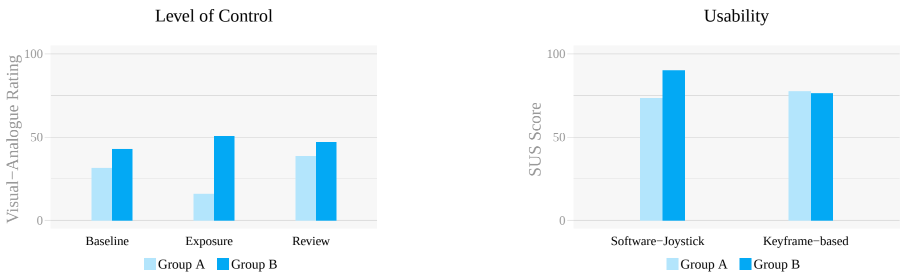
7.3.2 User Study: Examining the Effect of Reviewing Results

To conduct the study, we invited 12 participants (8 male, 4 female) with a Mdn age of 24 years, in total ranging from 21 to 32. None of the participants reported on prior knowledge with tools for camera motion.
Ahead of it, we welcomed the participants and provided information on the procedure of the study and how the collected data it was handled. We then handed out a consent form and would only continue after a participant declared consent. After that, we handed out a demographic questionnaire and showed the participants the images mentioned above that would represent a set of points of interest that they should capture by moving the camera. Each participant was exposed to all UI conditions. We made two groups and counterbalanced the order of exposure for each. For later reference: Group A was first exposed to the Keyframe-based UI and Group B to the Software-Joystick.

Before we started with the first trial, we took a baseline measurement asking the participants on their preferred level of control on a visual-analogue scale. We used a visual-analogue scale to address the issue that a 20-item scale might be considered too coarse. The captions on the ends of the scale read *"I control the system and the results manually"* and *"The system controls the results"*. We subsequently exposed the participants to both conditions and took another measurement. So, this time they could consider their rating in reference to the varying levels of automation. To be able to identify whether problems arise due to one UI, we asked the participants the fill out the System Usability Scale (SUS) questionnaire after being exposed to a condition. Next, we provided the participants with a review phase where they could watch the resulting material on a larger external display. After this, we took a final measurement which could be rated in reference to their perception of the quality of control with each UI. Having finished all trials and measurements, we debriefed the participants. As of the Wizard-of-Oz style, we did not consider measurements on the quality of control as in the prior study.

7.3.3 Results

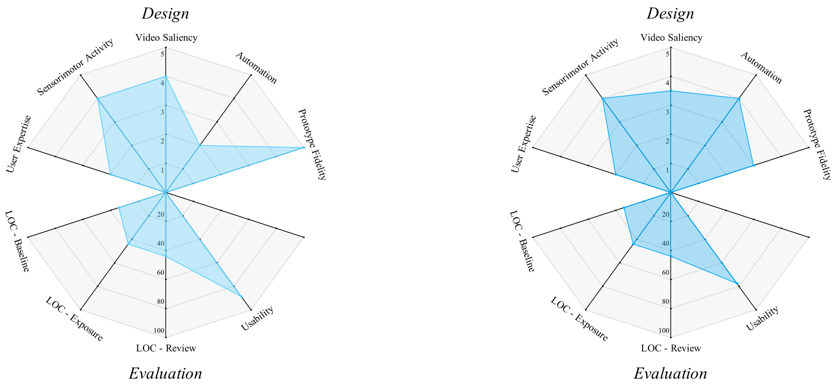
As the data on the preferred level of control was collected with a visual-analogue scale, it fulfils the requirements for parametric testing. Therefore, we first conducted a Shapiro-Wilk Test. As it showed no significant main effect ($W = .970$, $p \leq .414$), a normal distribution of the data can be assumed. However, to stay consistent with our analysis methodology of the first study, we also used non-parametric testing. Thus, we applied Friedman's Test and found no significant difference between the three measurements ($\chi^2(2) = .359, p \leq .084$) with a Mdn of 34 for Baseline, 44 after Exposure and 44 after the Review Phase. We plotted the data in Figure 7.6a and were surprised regarding the differences in ratings regarding after the exposure in respect to the group assignments. To further investigate possible interactions between the measurements and the assigned groups, we conducted a Scheirer-Ray-Hare Test [235]. We found no Measurement × Group interaction effect ($H = 1.767, p \leq .413$) or between the measurements ($H = .208, p \leq .901$). However, we found a significant difference regarding the groups ($H = 4.991$, $p \leq .025$). To investigate this finding further, we conducted a post-hoc analysis applying Conover's Test. We found no differences for the Baseline measurement ($p \leq .201$) and the Review measurement ($p \leq .248$). However, we found a significant effect regarding the Exposure measurement ($p \leq .014$; Bonferroni correction to $\alpha^* = .016$).



(a) Level of Control *(b)* Usability

Figure 7.6: Exploration 3: Results of our exploration of remote control UIs integrating a low and medium LOA

Similarly, we tested the sampled data on Usability (Figure 7.6b). We found a significant main effect using Friedman's Test ($\chi^2(1) = 4.456, p \leq .035$) with a Mdn value of 88.75 for Software-Joystick and of 77.50 for the keyframe-based UI (Figure 7.6b). To further investigate whether it is due to effects regarding the UIs or the assignment to a group, we also used the non-parametric Scheirer-Ray-Hare Test. We found no significant effects regarding a Concept × Group interaction ($H = 2.629, p \leq .105$) or among the groups ($H = 1.211$,$p \leq .271$). However, we found a significant difference regarding the UIs ($H = 3.877$,$p \leq .049$). To additionally make sure we did not miss differences among the groups, we again checked with Conover's Test and again could not identify a significant difference ($p_{Joystick} \leq .230$, $p_{Keyframe} \leq .630$). As there were only two groups, no post-hoc tests were applied.



(a) Software-Joystick *(b)* Keyframe-based

Figure 7.7: Exploration 3: Diagram of our proposed design variables and evaluation criteria

From this study, one can infer two major propositions in our estimation: First, the introduction of a review phase did not affect the ratings of the participants regarding their preferred level of control. While we could not identify an effect in this study, we would yet strongly recommend review phases in general based on our estimation of established best practices in the field.

Second, in this study, it seemed that the order in which the interfaces were presented resulted in greater differences than we would have expected: Cohen's d $= -1.542$ (large; with Hedges correction applied). While this could be due to some aspects in our sample of participants that we were unaware of, alternatively this could hint an effect to be taken into account regarding presenting UIs with varying LOA. In short, a biasing effect might occur when an assistance system is introduced first and only after the manual condition which might tilt the resulting data towards an increasing down-rating of the manual option. However, as only more studies on this issue can help to point to an answer, we will not speculate on either option regarding our presented study.

7.3.4 Lessons Learned

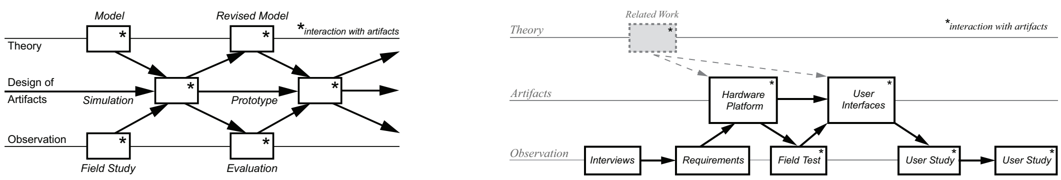
Lesson 6: Introducing a review phase could not be observed to affect the ratings of participants. Although we would still recommend them, they might not necessarily be indispensable when testing UI alternatives in controlled environments.

Lesson 7: The presentation sequence of designs with varying LOA might be more affecting than estimated (at least given the study mentioned above). However, this issue needs further thorough investigation (which might potentially be able to attribute it to the participant sampling). In any case, randomisation and counterbalancing are central (as it is already existing best-practice).

7.4 Summary

As we elaborated priorly, camera motion can be conceptualised as a task carried out synchronously by multiple operators in a well-defined workflow with a low margin of error. A similar environment was found by Mackay and Fayard [167] when they designed for air-traffic controllers. Generalising from their approach and experiences, they proposed a particular framework for HCI research that proposes a triangulation between theory, the design of artefacts and observation. They emphasise that given that HCI is an interdisciplinary field, the use of a limited number of approaches is potentially a threat to the generalisability of the gathered findings.

Therefore, they are in favour of a triangulation that applies various methods in combination. Their argumentation unfolds along the following lines. In lab studies, the conditions and variables can be isolated and controlled. However, this leads to an artificial environment which might incline study participants to act differently. Vice versa, people behave naturally in the field, but it is difficult to establish cause and effect relationships because of interfering factors. Thus, to be able to understand the investigated the subject matter thoroughly one needs to make use of the various available approaches. Similar to the propositions on creative tools by Hewett and colleagues (Section 6.2.2), Mackay and Fayard therefore advocate for a mixed-methods approach (Figure 7.8a).



(a) as by Mackay and Fayard

(b) as in our work

Figure 7.8: Basic Explorations: Triangulation across a set of mixed methods as proposed by Mackay and Fayard (Figure 7.8a) and as conducted in our efforts (Figure 7.8b)

In our work, we also applied such a mixed-methods approach across the design and evaluation processes presented so far (Figure 7.8b). To provide a conclusive overview, we present the proposition of Mackay and Fayard and our particular approach in Figure 7.8. As the overall framework covers aspects regarding theory, artefacts and observations, we want to briefly summarise already introduced aspects in our work. In terms of theory, we presented an overview of related work on human-automation interplay, UI design and particular aspects of cinematic camera control in Chapter 4. Similarly, we reported on user observations preceding the development of our ‘artefact’ in Chapter 3 and Section 6.1. In that Section, we also introduced our rational for the design of our artefact.

This section now reports on our observations in a field test, the design of further artefacts (the tested UIs) and further observations. Regarding the field test, it is worthwhile to note that camera operators are used to high-quality equipment and therefore using a prototype on an assigned shooting is likely to be considered with caution.

This was mirrored in so far as the following recordings were only considered after the results of the first were screened and found to provide a high enough degree of quality. Additionally, we found that given the benefits of automation, operators tend to explore and record more variations of the same shot, for example, the speed was varied or the location was changed more often. This was done to provide more material to choose from in the post-processing where the rhythm of the underlying music score, of spoken text and the filmed material, needed to match but were not already produced by the time of the recording.

Further, natural interaction processes and phenomena such as *"unexpected use"* emerge primarily in the wild as, for instance, referenced in the work of Marshall and colleagues [171]. In our observations, the most surprising and noteworthy finding in that regard was that (untrained) people tend to display a more natural behaviour in front of the camera when filmed by an automated tool instead of a film crew.

Concerned with the additional observations, those were made in controlled environments. This was done to investigate distinct research questions: first, what are the effects of introducing automation in the design and, second, what are the effects of introducing a review phase in the evaluation.

In conclusion, we could not find any (negative) effects on the quality of control, workload and judgement of control. Although we could find a reduction in physical demand due to the introduced automation, this did not affect the overall workload score. Given our previously made experiences, this surprised us, especially when concerned with the judgement of control. As much more jerky movements were recorded in the manual control condition which participants might have been unaware of, we tested exposing the participants to their recorded results. Based on our data, we could not identify an effect due to a review phase but still would recommend it as it is an established practice.

Insights

- To investigate the properties of our physical implementation, we took it to a field test. Together with a professional camera operator, it was used on five assigned shootings. The material captured at those recordings was later used in various exhibition films that were published online. This gave us confidence that the quality of our implementation is at least to a large enough degree comparable to professional tools. Therefore, possible bad ratings of UIs using it as a platform, must not be simply attributed to poor implementation. Most surprisingly, we found that the use of tools such as our platform, affects not only the recording procedures as it allows automating certain subprocesses but also the people being recorded; especially untrained people tend to behave more naturally.

- We further looked into how different control styles and UI designs influence the quality of control, the perceived workload and the judgement of control. Our data suggest that continuous remote control approaches do not perform differently than full manual control. This was surprising, as we expected full manual control to be rated highest at least regarding the judgement of control. We investigated the issue further and proposed that potentially an unawareness of the consequences might be factoring in as manual recordings showed a higher degree of shaky and jerky motion. Consequently, we suggested introducing a review phase that would give participants the chance to judge the quality of their results. We measured the effect that such a phase might have on the preferred level of control. Given our data, we found no influence. However, although it does not seem to be indispensable for controlled studies, we still recommend a review phase as we observed it to be a vital procedure on set.

KARL PEARSON (MATHEMATICIAN)

Statistics is the grammar of science.

RICHARD HAMMING (MATHEMATICIAN)

The purpose of computing is insight, not numbers.

VINCENT VAN GOGH (ARTIST)

Great things are done by a series of small things brought together.

8

Prototyping and Evaluating Design Alternatives

What to expect?

- User studies on design alternatives with varying LOAs, visual designs and input modalities
- Observations of use in laboratory environments as well as in the wild

What to take away?

- A set of insights from our empirical observations regarding individual design aspects that can be used as a foundation to inform design decisions on cinematic user interfaces

Attribution: This chapter references research that we previously published at CVMP ’16 [119], INTERACT ’17 [118, 120], CHI ’17 [123] and the co-located MICI ’17 workshop [121] as well as the DroNet ’18 workshop (co-located with MobiSys ’18) [79].

Our Statement of Collaboration details the differences between the paper(s) and this chapter

8.1 TrackLine: Refining Content-Based Interaction for Motion Control

As introduced in Section 4.2.3, computer vision can be used to inform the steering of semi-automated camera motion control tools. With the use of CV and delegation to a steering component, alternative ways of defining how the results should look like become necessary as a consequence. In contrast, to the existing approach of continuous remote control, for example, the framing could be defined through through-the-lens approaches (as introduced in Section 4.2.1). However, such approaches are accompanied by their own unique design challenges. Most frequently, they require objects to be within the screen-space to afford interactions. However, in cinematic practice expecting actors or objects to enter from outside the screen-space is quite common, but such behaviour is hardly supported in existing UIs. To address this particular issue, *we investigated an approach that would address the issue of supporting interaction for off-screen items in a content-based style*. It can be understood as an extension or further refinement to existing approaches.

8.1.1 Background

Systems incorporating such content-based control approaches often use a touch-to-track interaction style for acquiring user input. This interaction style often requires users to select a person or object of interest in the video stream by tapping. In the case of cinematic framing tasks, the systems subsequently often track and follow the selected item. How the details of the framing are handled varies across different implementations of this fundamental principle. While some systems automatically frame the selected item in the centre of the viewport, others use the Rule of Thirds, and again others require further user input. Nonetheless, what all these approaches have in common is that they entangle item selection, timing and sometimes also framing in one interaction. While this entanglement can make it a fast interaction, it can also make it a somewhat inaccurate and error-prone approach. The occurring errors often can be attributed to the property that the interactions need to be timed accurately. An item to be selected might move out of the frame before being selected (miss), or when selected, it might be at the wrong moment (bad framing). The latter frequently makes it necessary to perform select-and-correct moves to readjust the framing properly. For example, this can be achieved by a dragging the item to its intended location in the image-space. Independent of the underlying interaction style, such moves are carried out while recording. Consequently, these correcting manoeuvres are also recorded. This can render the material unsuited for further processing at times. The origins of such errors lie within the general issues of manual selection of moving items on touch devices. Here, Chiu and colleagues [49] suggest that, in general, performance can be expected to decrease with an increased movement speed of the items to be selected. In cinematic production, this might become a more significant issue for recordings of car commercials, sports or in high-speed shots.

8.1.2 Concept and Prototypes

To address these issues, we investigated an alternative design that we refer to as TrackLine. Most fundamentally, it untangles the interactions that are often combined in existing touch-to-track designs. In particular, TrackLine offers operators to preset the desired tracking position and to delegate the timing of the camera motion to an assisting subsystem (Figure 8.1).

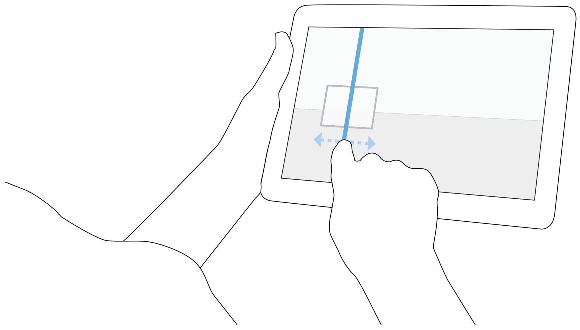
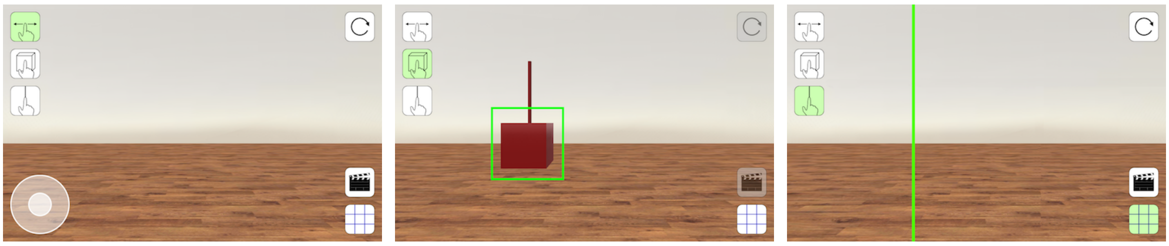


Figure 8.1: TrackLine: An adjustable line (blue) serving as a motion trigger overlays the camera feed (grey)

In detail, a moveable vertical line that is displayed on top of a video stream is used as a motion trigger. Its position is adjustable via drag-and-drop and can be set ahead of a recording. As soon as the CV analysis registers that a (recognisable) item intersects with the line, the camera motion subsystem is triggered and starts following it. During this tracking and following, the item is kept in the same position within the image space that is represented by TrackLine. In general, as the framing position can be set in advance, select-and-correct moves can be avoided. Additionally, even for moving items the error-rate can be expected to be minimised as, for fast objects, overall performance is no longer bound to human reaction and execution time.



(a) Software-Joystick *(b)* touch-to-track *(c)* TrackLine

Figure 8.2: TrackLine: User interface conditions of the study

We tested our proposed TrackLine design (Figure 8.2c) empirically and compared it to state-of-the-art designs (Figure 8.2) such as the Software-Joystick (Figure 8.2a) and touch-to-track (Figure 8.2b) as describe earlier (see also Diagrams in Figure 8.5).

8.1.3 User Study: Comparison to Status-Quo Designs

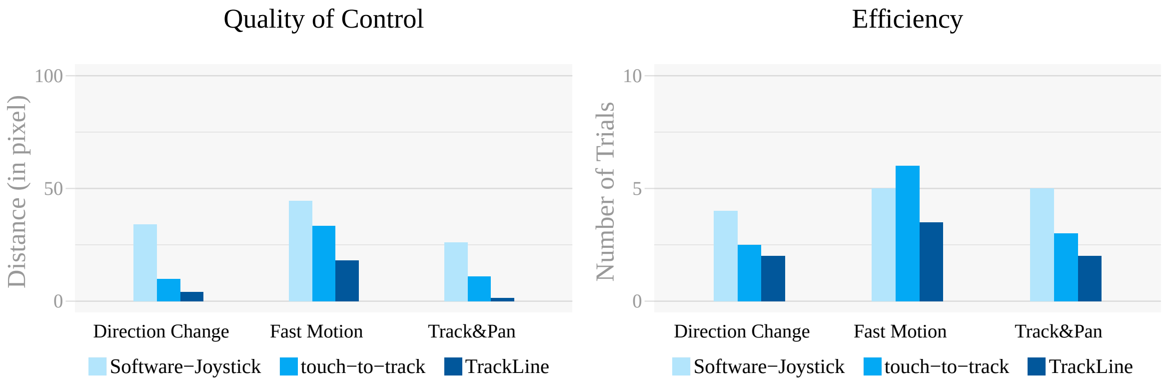
To evaluate our approach ahead of a full hardware and software implementation, at least conceptionally, we developed a prototype in a virtual 3D environment (as introduced in Section 6.1.2). In particular, our implementation was done in Unity 5.3 and ran on an off-the-shelf Android tablet. Regarding the UI conditions, with the Software-Joystick, participants could move the camera horizontally. With touch-to-track, the automatic following was triggered by tapping on the moving item to be selected at the desired location in the image-space. TrackLine allowed setting up the position in advance as detailed above. For both automated tracking methods, also visual feedback was provided. Here, the outline of a green box would appear around an item once it was selected (Figure 8.2b).

We invited 12 participants (9 male, 3 female) with years of age ranging from 21 to 27 and a Mdn of 24 years. All participants reported being proficient with touch devices, and three reported acquaintance with camera operation. We used a within-subjects study design with the independent variables being the user interface and study task. Both were counter-balanced with a Latin-Square design. We used quantitative as well as qualitative measurements and asked the participants to carry out different tasks. In particular, three task variations were asked to be performed: Direction Change (cube would change movement direction), Fast Motion (cube would move at higher velocity) and Track&Pan (participants were prompted to transition from a tracking shot into a panning shot). To enable panning in the Software-Joystick condition, we provided a second joystick in the UI controlling the rotation of the camera in the Track&Pan task. For the assisted techniques, the users only had to tap on the background to trigger the transition. For all tasks, a moving red cube should be followed and framed according to the Rule of Thirds. As common in camera systems, users could overlay a Rule-of-Thirds-grid by using a dedicated button. For further support, the centre of the cube was also marked with a pole. After a user-initiated button press starting a countdown of three seconds, the cube would start moving and enter from outside the image-space. To infer quality of control, we used our adapted SDLP measuring technique (see Section 6.2.2). Additionally, we recorded the Number of Trials as a measure regarding efficiency. Concerning the latter, we asked the participants to repeat a shot as many times as they deemed necessary to ensure that further improvement was unlikely.

We first welcomed the participants, explained the details of the study and handed out a consent form. Only after declaring consent, we continued by handing out a demographic questionnaire. Subsequently, an introduction to the tested applications was provided, and participants could familiarise themselves with the conditions in a training task for five minutes. After this training phase, the details of each task were explained ahead of each trial session. To be consistent with usage in real-life scenarios, we asked the participants to carry the tasks out while standing. After finishing a trial, the participants were asked to fill in a questionnaire asking on Efficiency, Ease of Use, Learnability and Comfort. After all trials, we asked the participants on the strengths and weaknesses of the UIs.

8.1.4 Results and Discussion

For the data regarding Quality of Control and Number of Trials, we conducted Shapiro-Wilk Tests which showed significance ($W_{QoC} = .508$ / $W_{NoT} = .887$, $p \leq .001$ for both). Therefore, the normality of the data cannot be assumed, and thus we used non-parametric tests for our analysis.



(a) Quality of Control

(b) Efficiency

Figure 8.3: TrackLine: Results of the study regarding objective data

Regarding Quality of Control, we found a Mdn distance of 34 px for the Software-Joystick, 10 px for touch-to-track and 4 px for TrackLine in the Direction Change task. Further, we found Mdn distances of 46 px for the Software-Joystick, 36 px for touch-to-track and 18 px for TrackLine in the Fast Motion task. Lastly, in the Track&Pan task, we identified Mdn distances of 26 px for Software-Joystick, 11 px for touch-to-track and 2 px for TrackLine (Figure 8.3a).

We conducted a Scheirer-Ray-Hare Test and found no indication of an interaction between the tested UIs and Study Tasks ($H(4) = 4.870$, $p \leq .301$). However, analysing both items individually suggested a significant difference among each with $p \leq .001$ for both; with $H(2)_{UI} = 53.632$ and $H(2)_{Task} = 27.178$.

As we were mainly interested in differences among the UIs, we further conducted post-hoc analysis applying Conover's Test for each comparison of UIs. We found significant differences among all three conditions with $p \leq .008$ or less for each, except for the comparison between the Software-Joystick and touch-to-track for the Fast Motion task with $p \leq .089$.

In conclusion, our data analysis suggests that TrackLine can help to increase the overall Quality of Control compared to the use of Software-Joystick or touch-to-track.

Further, we identified a Mdn Number of Trials of 4 for Software-Joystick, 2.5 for touch-to-track and 2 for TrackLine required by the participants in the Direction Change task. For the Fast Motion task, we found that a Mdn number of 5 trials was needed when using the Software-Joystick, 7 when using touch-to-track and 3.5 with TrackLine. In the last task, Track&Pan, we found Mdn numbers of 5 trials for Software-Joystick, 3 for touch-to-track and 2 for TrackLine (Figure 8.3b).

To test for differences, we also used the Scheirer-Ray-Hare Test. We found no indication for a UI × Study task interaction ($H(4) = 7.326, p \leq .120$). Analysing both items separately suggests significant differences in both groups separately, with $p \leq .002$ ($H(2)_{UI} = 27.477$ and $H(2)_{Task} = 12.936$).

Using Conover's Test for post-hoc analysis of the Direction Change task, we found significant differences with $p \leq .001$ except for the comparison between touch-to-track and TrackLine ($p \leq .070$). For the Fast Motion task, we could identify a significant difference only when comparing touch-to-track to TrackLine ($p \leq .001$). Further, for Track&Pan, we found significant differences with $p \leq .001$ for all comparisons except for touch-to-track × TrackLine.

In summary, assisted approaches such as touch-to-track and TrackLine could reduce the Number of Trials required by users compared to the Software-Joystick. Given our study and gathered data, TrackLine could outperform touch-to-track in terms of efficiency only in the Fast Motion task.

The data on the Efficiency, Ease of use, Learnability and Comfort was collected via 5-item rating scales. With a Mdn of 2, the Software-Joystick was rated low in Efficiency and Ease of Use, but also as comfortable with a Mdn of 4.5. Touch-to-track was considered to be efficient and easy to use with a Mdn of 4, but also rated the least comfortable UI with a Mdn of 3.5. TrackLine was judged to be very efficient, easy to use and comfortable with a Mdn of 5 for each dimension (Figure 8.4).

In the final semi-structured interviews concluding the study, participants pointed to further improvements. Most notably, they critiqued that the assisted approaches lacked expressiveness and did not support exploration. *"I liked the joystick best, because I had the most control, even if the technique is potentially more imprecise than the TrackLine"* (P09). The same participant also acknowledged that *"If this had been a real recording, it would have become pretty jerky"*. Further, multiple participants also suggested that the Software-Joystick required more cognitive resources, stating that it was harder to react because they were preoccupied with focusing on the cube and timing their actions.

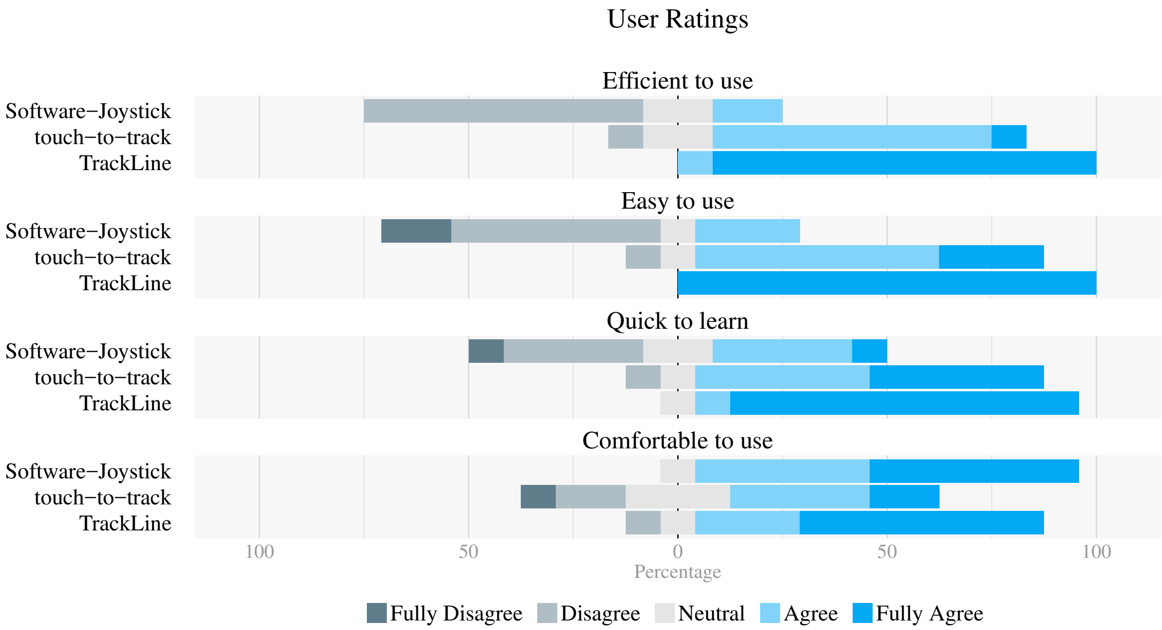


Figure 8.4: TrackLine: Results of the study regarding subjective data

8.1.5 Lessons Learned

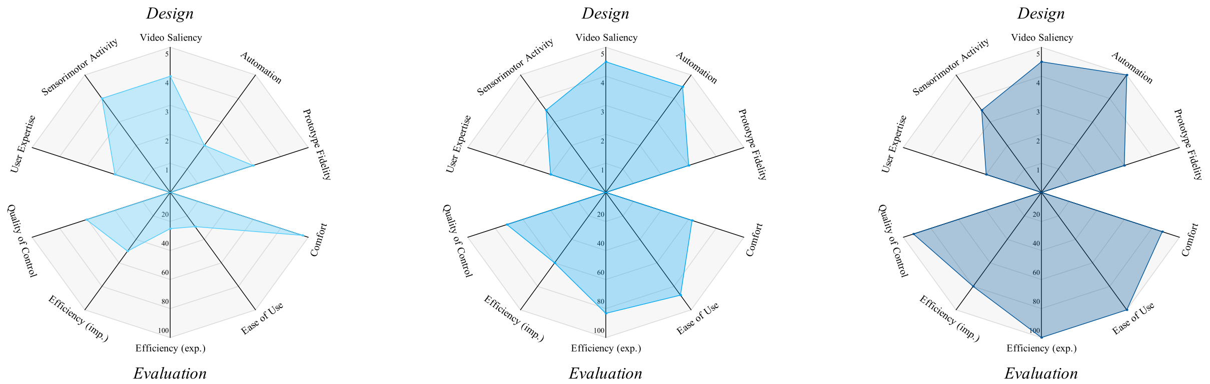
From our study, we could take away some lessons that we want to list below briefly:

Lesson 1: We observed that assisted approaches could increase performance compared to remote manual control based on implicit measures and qualitative statements.

Lesson 2: Among the tested assisted approaches our proposed TrackLine design could outperform touch-to-track for faster-moving items (as in car commercials for example).

Lesson 3: Participants especially like the freedom and expressiveness associated with remote manual UIs. Therefore, assisted approaches should be used to complement such designs.

Our conceptual implementation also faces several shortcomings and limitations. For instance, it only supports a horizontal movement. So, motion triggers expecting off-screen elements that appear vertically cannot be handled. Also, inclined motion triggers or lines representing motion triggers that do not fill the whole width or height are not supported. For a conceptual implementation, these aspects were not crucial to us and an implementation integrating a more substantial degree of user control does not seem far-fetched. For a more elaborate implementation, we suggest using a two-finger gesture that also to draw an arbitrary line at a user-chosen location with a user-chosen inclination; similar to the interaction designs of touch-based through-the-lens control as presented by Reisman and colleagues [219] (see Section 4.2.1).



(a) Software-Joystick *(b)* touch-to-track *(c)* TrackLine

Figure 8.5: Diagrams of the tested user interfaces in the TrackLine study

Additionally, so far we only proposed motion triggers in the form of a line. However, also more advanced geometrical shapes can be used as motion triggers. Based on two-finger gestures, also shapes such as rectangles could be defined that serves as a region of interest. Within such a region a system might then check for conditions to be met and subsequently informs and triggers a motion-control subsystem. Beyond rectangles, also other (higher-dimensional) polygons are imaginable (if they show to be reasonable choices).

Further, we only expected an object of a single class to appear from outside the viewport. We, therefore, did not provide further differentiation on which particular class this object might belong. For more practical implementation, it seems reasonable to allow for further classification such as (individual) persons, animals, cars or else. Such an implementation seems to be a two-part problem at the minimum; with the first being the underlying capability to differentiate a set of relevant classes and the second being designing access and control for end-users. While the first one rather seems to be a CV focus software-engineering task, the second is more of an UI design issue. Future (user-centred and conceptional) work addressing the latter, however, can be used to inform the first as it might point to what relevant classes are necessary for cinematic tasks (in contrast to general-purpose CV).

Please note that, for the diagrams in Figure 8.5c, the values regarding Quality of Control and Efficiency were inverted (100-*value*). Thus, a smaller SDLP deviation would represent a larger degree of Quality of Control. Likewise, a smaller Number of Trials would represent a greater degree of Efficiency. Further, we scaled the values from their original scaling of 1-10 to 1-100 to fit the diagram.

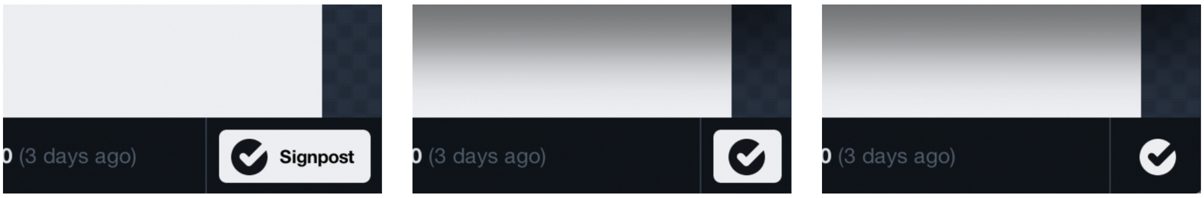
8.2 Progressive Reduction: Reducing the Visual Footprint of User Interfaces

As indicated in the previous section, automation and manual control UIs should complement each other. While this could increase overall performance without involuntary diminishing experiential aspects, this would most likely also lead to a greater degree of occlusion of a video stream that is displayed on a canvas below the UI layer. This might be most strongly notable in touch-based devices using camera-based UIs, as more features need to be displayed. As introduced earlier, maximising the visibility of the camera stream is also a paramount consideration for such control and cinematic UI designs. To address all the issues mentioned above simultaneously, it seems reasonable to further *include design strategies that (in addition) aim at minimising the displayed UI elements*. Generally, ! such an objective can be achieved in various ways that we want to outline below.

8.2.1 Background and Related Work

A well-established concept used for de-cluttering a UI is Progressive Disclosure (PD). It is described, for instance, by Nielsen [361] or Tidwell [260] and is following the fundamental idea of minimising the number of visual elements that are displayed. It does so by disclosing or 'hiding' UI features into a collapsed or off-screen menu. The items being hidden are not arbitrarily selected. Items that are used more frequently remain undisclosed, whereas less frequently used items are disclosed in a menu. As those are only visible when the menu is opened, screen-space is 'saved'.

As control elements, especially continuous control elements, are vital for basic functioning and steering of a system, disclosure of the associated UI elements into a menu might not be a suitable strategy. Especially, when things go wrong immediate access to these controls is of crucial importance. Further, the user experience might suffer when central UI elements are hidden early-on, for instance, as the perceived affordances [197] are taken away along with the hidden elements. Consequently, further approaches that aim at minimising the visibility of UI elements have been proposed. For example, a recently discussed alternative approach and/or extension to PD is Progressive Reduction (PR) [324, 332]. While the main idea behind PD is to reduce the number of displayed items, PR suggests minimising the footprint regarding the visual appearance of the items. In particular, the visual variables as presented by Bertin [14] and later Carpendale [37] can be used as a fundamental framework to derive characteristics that can be minimised, such as size, form, opacity and others. An additional notion to PR (and PD for that matter) is that this process does not happen abruptly, but gradually (Figure 8.6). In the case of PR, this can entail time-spans of weeks to months. Such gradual adaptation allows for the exploitation of a user's ability to learn particular functions, their abstract and visual representation as well as their location on the screen. So by the time that a further reduction step is taken, users already know to a greater degree where to look for a particular function and roughly which visual pattern to expect. In PR approaches, also a user's learning curve is taken into account. Should a user return to a system after a longer period of not using it, the overall level of reduction might be reverted by one step, to support reorientation within the UI and to counteract an experience decay. Up to now, in general, it has received only limited attention in the HCI literature (under this term), which makes a further inquiry on the subject matter necessary.



(a) Reduction level 0 *(b)* Reduction level 1 *(c)* Reduction level 2

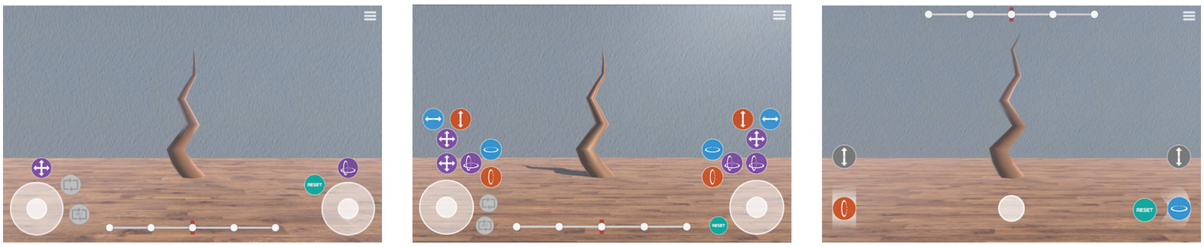
Figure 8.6: Examples of three levels of Progressive Reduction where an interface element gradually changes its appearance into a more minimalistic version (from [332])

However, the general idea of dynamic UI adaption has been investigated priorly in different areas and approaches. For instance, much work was conducted in the field of AR. Here, Carmo and colleagues [36] adapted symbols that were used to provide additional information in outdoor environments to adapt to changes in backgrounds and illumination. Similarly, Gabbard and colleagues [92] exemplified the use of adaptation strategies to enhance the readability of text that was overlayed on camera streams as an AR feature compensating for varying backgrounds. Further, Julier and colleagues [134] proposed an adaptation strategy to filter information based on temporal, distance and angle cues. Adaptation further is applied in map-based interfaces as presented by Nivala and colleagues [195] who adapt the rendering of maps taking into account certain user needs and context.

What seems to be common across these approaches is that they either try to adapt to changing environments (background and illumination) and/or to separate the 'more relevant' from the 'less relevant' (which is not trivial). The latter approach often aims at a more minimalistic UI. Such minimalist UIs can by nature of their design only provide a smaller number of cues on what they represent conceptionally or where to look for certain features. For instance, gradual approaches such as PR also exploit a user's spatial memory to some extent. UI elements of a particular function do not change their position; only their visual representation. Therefore, usability aspects become less negatively affected. Exploiting spatial memory also has been used priorly as in the work of Lepinski and colleagues [151] proposing and investigating Marking Menus or Gustafson and colleagues [102] suggesting and studying their approach of Imaginary Interfaces.

8.2.2 User Study 1: Exploring Design Alternatives

To investigate the feasibility of integrating PR into the design of UIs for camera motion control, we conducted a controlled experiment. To make sure, that we would test this UI adaptation strategy with a design that was perceived positively by users, we first developed three design alternatives, that were tested with users, and used the most promising for further studying.



(a) Software-Joysticks *(b)* Joysticks with (open) pie menus *(c)* Central Knob

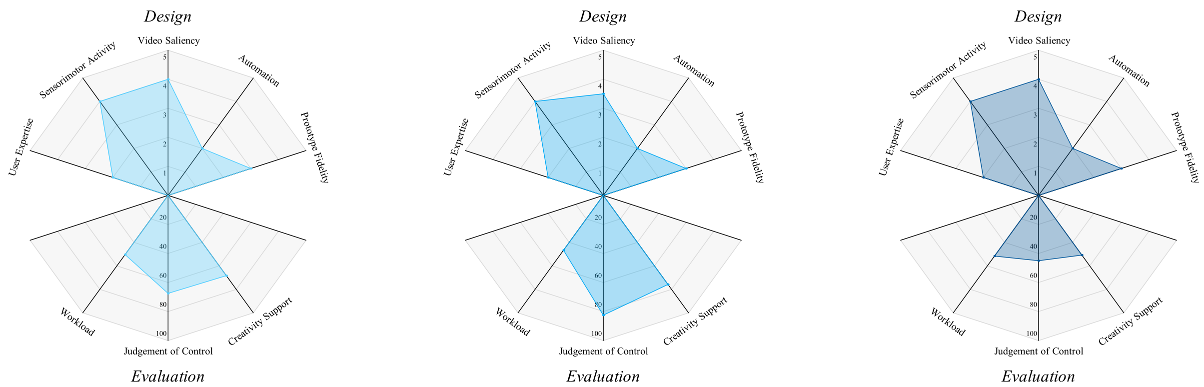
Figure 8.7: Design alternatives we tested ahead of implementing a PR adaptation strategy (from [292])

Designs and Prototypes

Therefore, we implemented prototypes of our design alternatives in Unity3D (Figures 8.7 and 8.8) which would run on an off-the-shelf Android tablet. As a remote control baseline, we again opted for a Software-Joystick (Figure 8.7a). In detail, we used two: one allocated to translational control (left) and another allocated to rotational control (right). Each could be used to control movement in two dimensions (left-right, up-down).

As an alternative, we implemented a design that we refer to as Extended Software-Joystick (Figure 8.7b). In detail, the design would extend the traditional joystick design by pie menus that are located above the joysticks. We chose this approach based on the insights presented by Lepinski and colleagues [151]. In their work on marking menus, they suggest that the structure of pie menus lends itself well to mental and physical learning (for example through muscle memory). In consequence, once the locations, the position within the menu and the representation and functionality were (over-)learned, less visual cues need to be displayed. This, in particular, might further interlock well with our primary idea of minimising the visual footprint of UI elements.

In our case, the pie menus would provide more freedom to the users as they can choose for themselves which movement control they want to allocate to a particular joystick. So, they could use a traditional mapping as in the regular Software-Joystick UI but also switch the mapping between the joysticks or use only a subset of the controls such as only left-right control with one joystick.



(a) Software-Joystick *(b)* Extended Joystick *(c)* Central Knob

Figure 8.8: Diagrams of the design variables and evaluation criteria regarding the PR design alternatives

As a further alternative, we considered a novel approach that decoupled translational and rotational control. Our contextual inquiries into cinematographic practice inspired this design. We observed that the different axes would sometimes be controlled by different operators [124].

We found that the first camera operator would delegate translational moves to an assistant. The main operator would then go on and take only care of controlling the rotation of the camera. We translated this finding into our design by introducing an assistance system. In the centre, we displayed a horizontally restricted software-joystick (Figure 8.7c). This could be used to control the translation of the camera manually. Similar to our early investigations, we mapped directional and speed control to the joystick (see Section 6.1.1).

Once the movement direction and speed was fitting, users could save the settings by dragging upward or downward from the software-joystick. After a dwell-time the progress of which was indicated by a vertical bar, these would be locked and the motion continued (Figure 8.9). Thus, users were free to focus on controlling the rotation with the two UI elements present at the left and right border of the screen. In detail, an upward swipe would save the movement direction and speed as it is while a downward swipe would additionally trigger a speed bracketing function.

To make the alternatives comparable to each other, we also provided this functionality in the other conditions (next to the left joystick).

For all UIs, the button size was at least 11 × 11 mm as recommended [208, 357, 359]. Further, in all designs, we added a horizontal line representing the rail of a slider with five dots on it, representing shortcuts to positions on this rail as found in our expert evaluation (Section 7.1).

(a) Swipe-Lock starting *(b)* Swipe-Lock loading *(c)* Swipe-Lock locked

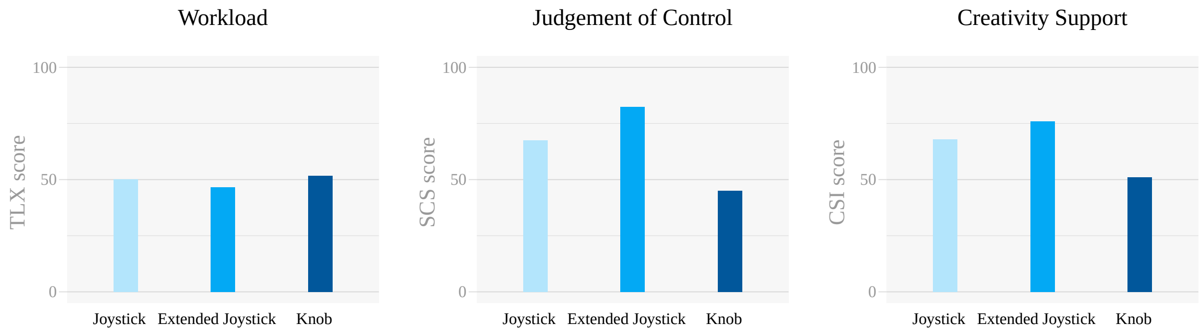
Figure 8.9: Progressive Reduction: Central Knob with Swipe-Lock (from [292])

User Study and Results

To examine which of the proposed designs might be the most promising, we set up a controlled user study. Therefore, we invited 10 participants (7 male, 3 female) with years of age ranging from 20 to 37 with a Mdn of 26 years. All participants reported to be acquainted with touch devices, and none had prior training in cinematic motion control tools; one was left-handed. For this study, we applied a within-subject design testing our three design alternatives with three study tasks. Each participant was exposed to the independent variables in a unique counter-balanced order to counteract potential learning or order effects.

While the studied UIs were already introduced above, we want to provide further details regarding the study tasks. Over the course of the study, we asked the participants to execute the following tasks: Product Shot (controlling translation and rotation of the camera simultaneously; similar as used in Section 8.1), Figure Eight (an eight on its side or infinity symbol trajectory as already used in Section 6.2) and Free-Roaming (participants were free to explore a scene). In addition to previously used tasks, we opted for the Free-Roaming task, as it somewhat mimics a first exploration of possible camera angles and transitions on a new set. Additionally, it allows for more expressive use. Supporting expressiveness in cinematic and creative tools is essential as pointed out by Christie and Olivier [53] or Hewett and colleagues [111] and as found in our TackLine study (Section 8.1). For each UI, we measured Workload, Creativity Support and Judgement of Control applying the TLX, the CSI and the SCS which we introduced already in Section 6.2.

Ahead of the study, we welcomed the participants and informed them about the procedure and its details. We then handed out a consent form and only continued if consent was given. Subsequently, we handed out a demographic questionnaire. Next, we introduced the first design and asked the participants to carry out the study tasks based on the (counter-balanced) order they were presented. To mimic a real-life use case, we further asked them to do so while standing. Having completed all tasks with a UI, we handed out the TLX, CSI and the SCS questionnaires and asked the participants to fill them in. This procedure was repeated until all tasks were executed with all interfaces. Next, we additionally asked the participants on their preferences and the usability of the UIs in a semi-structured interview. After the interview, we debriefed the participants.



(a) Workload *(b)* Judgement of Control *(c)* Creativity Support

Figure 8.10: Progressive Reduction: Questionnaire scores of the design alternatives

For analysing the collected data (Figure 8.10), we used non-parametric tests. As an omnibus test, we used Friedman's Test with Conover's as a post-hoc test (with Bonferroni correction applied).

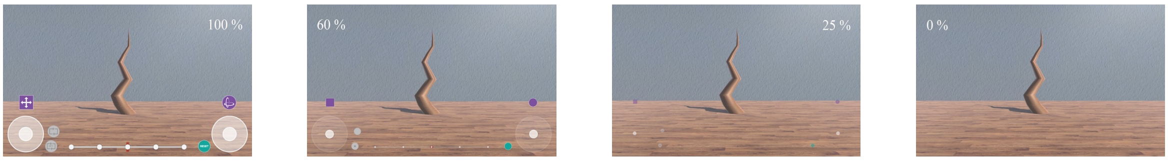
Regarding Workload, we found no significant difference among the conditions ($\chi^2(2) = 2.0$, p $\leq$ .368). Also, for Creativity Support, we found no significant main effect ($\chi^2(2) = 3.128$, p $\leq$ .209). But, concerned with Judgement of Control, we could identify a significant main effect ($\chi^2(2) = 8.359$, p $\leq$ .015).

Post-hoc pairwise comparison indicated that the Extended Software-Joystick (Mdn $= 82.5$) compared to the Central Knob (Mdn $= 45$) was judged significantly higher in terms of Control (p $\leq .004$). Comparing the Extended Software-Joystick to the Software-Joystick (Mdn $= 67.5$), we could not identify a significant difference (p $\leq .097$). Comparing the Software-Joystick to the Central Knob also showed no significance (p $\leq .081$).

In conclusion, we found the Extended Software-Joystick to be most promising out of multiple reasons. First, it reached the highest absolute score concerned with Judgement of Control (Mdn $= 82.5$). Second, it could significantly outperform the Central Knob. Third, also in the semi-structured interviews participants stated that they preferred the Extended Software-Joystick.

8.2.3 Reduction Strategies and Levels

Consequently, we opted for the Extended Software-Joystick design as a foundation on top of which we implemented different PR strategies. Differences in strategies can emerge from the fact that visual variables can be reduced in various ways over time. For example, in our designs, we only minimised the appearance of UI elements regarding their size, opacity, form and visibility (Figure 8.11).



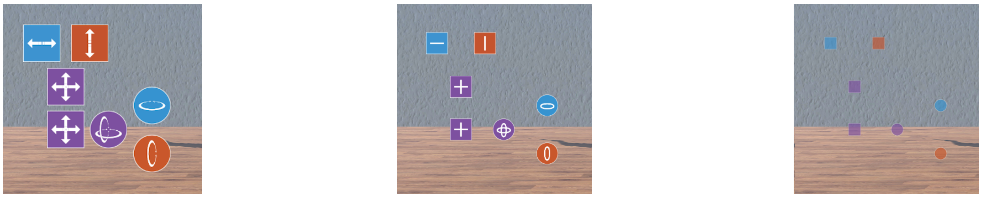
(a) UI Visibility 100% *(b)* UI Visibility 60% *(c)* UI Visibility 25% *(d)* UI Visibility 0%

Figure 8.11: Progressive Reduction: Levels over time (from [292])

In particular, we used two distinct strategies, referred to as Icons First (Figure 8.12) and Background First (Figure 8.13). In Icons First, the icons used to reference the functions in the pie menu of the Extended Software-Joystick would gradually shrink and fade out first. In detail, they would fade out earlier than the backgrounds of the UIs items. This requires users to internalise which function is placed where to some extent, but once (over-)learned provides a larger degree of (rough spatial) visual cues where interactive areas are located. Especially, when compared to our Background First approach. Here, the background would fade out earlier than the icons. Vice versa, the functionality referencing icons remains longer visible, yet the boundaries of the interactive areas are visible only for a shorter period of time. Therefore, also the background would be less occluded overall. For both conditions, we would minimize the size of the buttons over time along three major steps (Figure 8.11): regular-sized UI elements (Level 0), intermediate reduction step (Level 1) and invisible (Level 2). We introduced the intermediate reduction step mostly to smooth the reduction transitions and to make icons easier identifiable at small sizes. For the latter, we changed the original icon appearance, for example, 'cross with arrows' to simplified version such as a simple 'plus' when the overall visibility was lower than 75%. Only after this level, our distinct strategies would be applied and depending on it either the icons or the backgrounds would gradually disappear. To counteract a decrease in performance due to smaller target sizes, as one could expect based on Fitts' Law, we kept the actual touch-sensitive areas of the UI elements at their original dimensions of at least 11 × 11 mm.

8.2.4 User Study 2: Examining Reduction Strategies

To investigate potential effects due to PR in general and the different PR strategies in particular, we set up a second user study also in a controlled environment. To make sure that we would isolate the effects of the PR strategies from the study tasks peculiarities, we used only one task: a modified



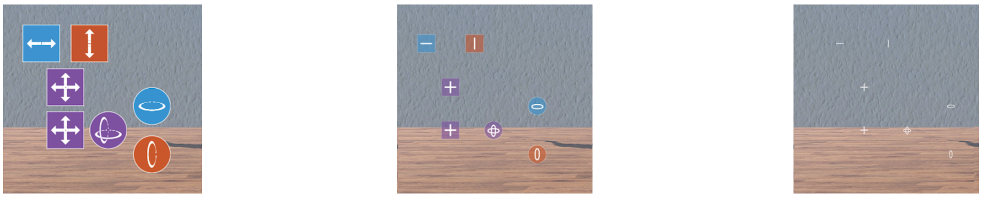
(a) Icons First (Level 0) *(b)* Icons First (Level 1) *(c)* Icons First (Level 2)

Figure 8.12: Progressive Reduction: Levels over time for Icons First (from [292])

version of the Free-Roaming task, that we already used in the previous study. We modified it to a Prompted Roaming task. Here, we displayed three additional boxes in the upper half of the screen. Those would provide icons prompting the settings of that participants should choose from the pie menu of the Extended Software-Joystick design. The boxes were vertically aligned and equally spread horizontally. The left box would show prompts regarding the left joystick. Equally, the right box would display prompts for the right joystick. The centre box displayed which type of motion should be executed in the roaming task. To make sure, that the UI elements are frequently used, the prompts change every ten seconds. The sequence of prompts was determined ahead of the study in a pseudo-random fashion. This way we made sure that unreasonable combinations of settings and motion were not prompted as, for example, moving up with one joystick while simultaneously moving down with the other while both are steering the translation of the camera.

We used a between-groups design to minimise interference due to learning effects. Potentially, if a participant experienced PR already once, ratings on a second exposure might be biased, especially in the later reduction stages. Therefore, each participant was only exposed to one strategy. We invited a total number of 18 participants (11 male, 7 female) with years of ranging from 19 to 33 with a Mdn of 22.5 years. All participants reported to be experienced with touch devices and that they had no prior training in cinematographic camera motion control tools; two were left-handed.

First, we welcomed the participants, detailed general information about the study and its procedure and asked for the consent of the participants. Once consent was declared, we would go ahead and hand out a demographic questionnaire. Subsequently, they were exposed to the Extended Software-Joystick with an additional PR strategy incorporated and asked to carry the Prompted Roaming task. As in the previous study, the participants were asked to carry out the task while standing. For the first 5 minutes, the UI would remain in its initial state. Then the PR would gradually start and would further progress to the next reduction level every 5 minutes until the final level was reached (being an invisible or Imaginary UI [102]). At this level, the task was executed for another 5 minutes. Having carried out the Prompted Roaming for a period of 5 minutes, we asked the participants to rate their Judgement of Control using our proposed adapted version of the SCS. To also collect data on Workload and Creativity Support, we also asked the participants to fill out the TLX and the CSI. However, to



(a) Backgrounds First (Level 0) *(b)* Backgrounds First (Level 1) *(c)* Icons First (Level 2)

Figure 8.13: Progressive Reduction: Levels over time for Backgrounds First (from [292])

minimise interruptions during the study, we only collected data using the latter two questionnaires at given points of interest in time: in particular after 5, 25 and 35 minutes. So, overall 35 minutes was needed to carry out the Prompted Roaming with all stages of PR. After that period, we debriefed and thanked the participants.

8.2.5 Results and Discussion

We used non-parametric tests as in the prior study. Also, Bonferroni correction was applied to the post-hoc tests to compensate for pairwise comparisons which were only conducted after a significant main effect was found.

Overall, we found no significant main effect for the reduction strategies for Judgement of Control (Mdn_{IF}=75, Mdn_{BF}=65, $Z \leq .00$, $p \leq 1.000$), Workload (Mdn_{IF}= 49, Mdn_{BF}=49, $Z \leq .036$, $p \leq .971$) and Creativity Support (Mdn_{IF}=66, Mdn_{BF}=55, $Z \leq .361$, $p \leq .718$).

We conducted further significance tests comparing the data gathered after 5, 25 and 35 minutes for each dimension. For Workload, we found a significant main effect ($\chi^2(2)$=20.111, $p \leq .001$). Post-hoc pairwise comparison using Conover's Test (with Bonferroni correction) indicated significant differences between all sampled points in time: with 5 × 25 ($p \leq .004$, Hedge's $g = -.609$), 5 × 35 ($p \leq .001$, $g = -1.311$) and 25 × 35 ($p \leq .001$, $g = -1.119$). For Judgement of Control, we also found a significant main effect ($\chi^2(2)$=30.629, $p \leq .001$). Post-hoc pairwise comparison showed a significant difference between all measurements with 5 × 25 ($p \leq .001$, $g = .817$), 5 × 35 ($p \leq .001$, $g = 2.381$) and 25 × 35 ($p \leq .001$, $g = 1.765$). For Creativity Support, we also found a main effect ($\chi^2(2)$=13.914, $p \leq .001$). Post-hoc pairwise comparison also showed a significant difference between all measurements with $p \leq .001$ with 5 × 25 ($p \leq .001$, $g = .575$), 5 × 35 ($p \leq .001$, $g = .918$) and 25 × 35 ($p \leq .001$, $g = .743$).

Given our study setup and its results, several questions and shortcomings come to mind; some of which we want to address below. A first question might be "when precisely and potentially why does

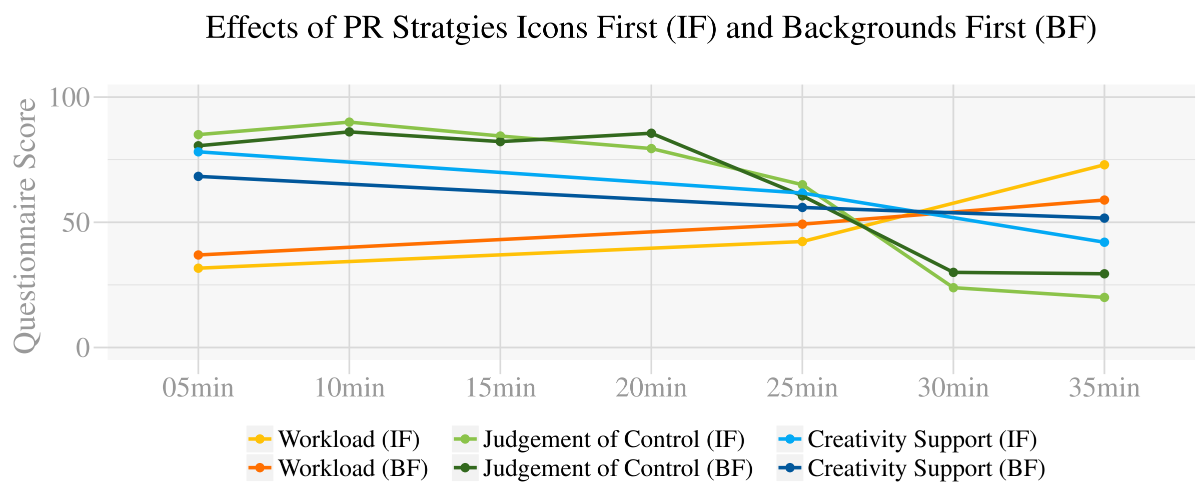


Figure 8.14: Progressive Reduction: Questionnaire scores of reduction strategies

performance diminish at a particular point?". Between our sampled data points of 25 and 30 minutes, Judgement of Control diminishes rapidly while simultaneously, Creativity Support is negatively affected and Workload increases substantially. Given that, to some extent, all our measures seem to be affected at this point, a major design issue seems to arise in the reduction process. As an explanation model explaining what might cause this effect, we want to offer the following ideas. At this point, the UI elements are, in principle, still visible but rather small. Their size is below the minimum sizes for UI element that is deemed necessary (roughly 10 × 10 mm [208]). They are also only visible in principle as one's fingers occlude the elements. Similar observations have been made when investigating the effects of (finger) occlusion or target size of UI elements. However, we did not minimise the hitboxes of the targets, only the size of their visual appearance. So, it seems to us that not only the actual target sizes affect the performance of users, but also the (a priori) perceived target sizes. For our purpose, this might give room for a potential solution mediating between UI reduction and performance or experience aspects. It might entail introducing further feedback representing the hitboxes, for example by displaying the outline of the hitbox. This way, users might be informed about their size being constant.

The generalisability of our results is also limited especially regarding the following issues. While PR is often considered a long-term strategy, we conducted a short-term study over a course of 35 minutes. While this surely limits the ability to generalise from our results, we expect potential positive effects to be similar or even increase their effect, while negative effects might also be similar or less affected due to the longer phases of learning and adaption in a long-term deployment.

In comparing our two reduction strategies we could not identify a significant difference. However, we did not include changes in background and illumination. Adding these aspects to the study setup might potentially affect the results. To minimise the visual footprint of the UI to the fullest, one could recommend our Backgrounds First strategy given our data collected so far. Should changes in backgrounds and illumination change this estimation there is also a potential fall-back strategy, which is to limit the reduction to size, opacity and icon complexity resulting in somewhat of a mix of our tested strategies.

Further, our tested design consisted mainly of a manual control option, the Extended Software-Joystick with a limited number of assisting functions. However, options such as our TrackLine design were not integrated. Consequently, further mixed designs with complementing manual and assisted options should be considered as well to maximise generalisability.

8.2.6 Lessons Learned

From both studies mentioned above, we could infer some lessons that we want to briefly share below:

Lesson 1: The Extended Software-Joystick might be a worthwhile alternative to the (fixed) Software-Joystick as it is traditionally implemented in many systems in the wild. It provides benefits such as an increased level of user freedom in terms of steering behaviour, and can potentially minimise errors by only reacting to a subset of the user input as in (only) horizontal control. Its downside is that it requires additional menus to allocate such behaviour which adds complexity to some extent and further camera stream occlusion.

Lesson 2: To account for UI occlusion PR can be successfully used, generally speaking. Our experiments indicate that up to a certain point regarding the reduction level the main metrics are only affected to a minor degree. Beyond that point, however, major disruption is observable.

Lesson 3: As no difference emerged due to our study on different reduction strategies (Icons First and Background First) one could opt for the Backgrounds First strategy to reduce UI occlusion of the background to a greater extent.

Lesson 4: While the observed major disruption could be attributed to a decrease in target size, and it is important to notice that only the visual cues of the targets were minimised, but not the actual hitboxes. So, to us, it seems that interaction with UI targets is not only subject to psycho-motoric aspects such as conceptualised by Fitts' Law but further also by perceptual cues.

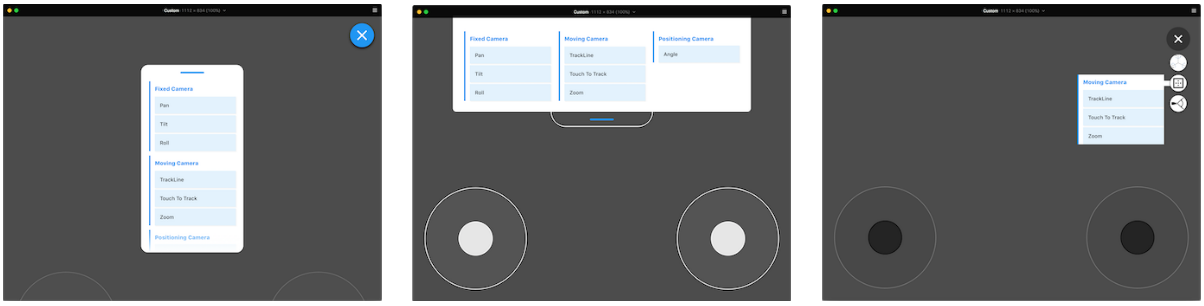
8.3 Axis + Content: Integrating Manual and Content-Based Control

Considering the two previously reported projects, some further limitations come to mind. In the TrackLine study, we focused mainly on the assistance aspects while in our Progressive Reduction study, we focused mainly on manual control elements. Further, in both projects, we hardly changed the backgrounds. Regarding our Progressive Reduction study, it also remained unanswered whether a gradual reduction is beneficial in itself or whether one could alternative already start with a UI that is reduced somewhat close to the identified major disruption point. Consequently, in this project, we address these issues by *implementing a UI that would be tested with changes in the underlying background scene, and use manual and assisted control elements complementary investigating how gradual reduction compares to an already minimised UI.*

8.3.1 Background

Similar to our previous attempts, we aimed at developing and investigating a UI that uses a camera stream as a base layer and tries to maximise its visibility while providing a necessary minimum of UI elements. We divided this main objective into smaller sub-objectives to be able to handle the various design variables at separate stages during this process.

Therefore, the first sub-objective was to define a menu structure that combined manual and assisted control features. During the first initial sketches, it appeared to us as if such a menu could quickly become complicated. Consequently, we opted for a logical grouping of similar features to support operators in selecting menu items. Additionally, the UI should be designed to minimise the visual footprint of inevitable occlusion. This goal should be achieved by applying a mix of multiple methods: disclosure into menus, (progressive) reduction of the visual appearance of UI items and animated transitions translating UI elements off-screen. As potentially there is an innumerable set of design variations that fulfil these prerequisites, we decided to restrict our efforts to a particular subset. As a reasonable subset for our purpose, we decided to mainly use already established and commonly used design patterns as they are found on mobile devices in the wild. Further, we aimed at implementing those using a rapid prototyping approach, and to evaluate them in a user study regarding UI occlusion. The second sub-objective was to investigate the effects of progressive and fixed reduction based on the most promising design emerging from the evaluation mentioned above. Consequently, a second study examining that design alternative should be conducted. In contrast to the first evaluation that focused on investigating UI occlusion based on a rapid prototyping approach, the second study should examine a fully functional prototype, and its evaluation metrics should be based on our priorly established evaluation framework.



(a) Draggable Menu *(b)* Top-fixed Menu *(c)* Icon-based Menu

Figure 8.15: Axis+Content: Design alternatives explored in the study (from [79])

8.3.2 User Study 1: Exploring Design Alternatives

To explore fundamental alternatives, we implemented three designs in Framer (v102). The alternatives differed most regarding the menu structures (Figure 8.15) and are outlined below.

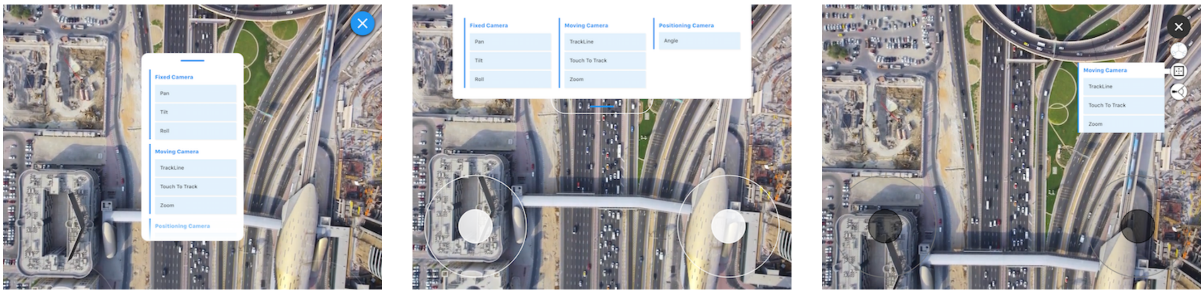
Draggable Menu: As a first option we used a Draggable Menu (Figure 8.15a) that opens after a button press and simultaneously moves the software-joysticks off-screen. This should minimise the occlusion of the background and increase the users' freedom to move the menu to a less disturbing location. Further, the menu contains a draggable list of items in a vertical layout that are hierarchically organised.

Top-fixed Menu: Secondly, we used a Top-fixed Menu (Figure 8.15b) that was attached to the top of the screen. It was hidden by default and could be dragged down to display its contents. For this design, we used a horizontal layout for the items. The design of notifications inspired this design as they are common in many contemporary mobile operating systems.

Icon-based Menu: The third design, Icon-based Menu (Figure 8.15c), would also open after pressing a button that was fixed to the top right corner. After the press, a set of clickable icons appeared representing the top-level hierarchy. A further click on one of the items would further open a sub-menu displaying the disclosed contents of a particular lower level in the hierarchy.

To evaluate the different UIs, we recruited 15 participants (12 male, 3 female) for a user study with years of age ranging from 21 to 34 and a Mdn age of 24 years. All participants reported prior experience with tablets, one experience with camera control tools; further, one was left-handed.

We conducted the study in a controlled environment using a within-subjects design. Therefore, we exposed each participant to all UI conditions. To minimise learning or order effects, we randomised the exposure sequence of the conditions. In contrast to Figure 8.15, we added an animated video clip as a background. For this, we chose an aerial top shot recording that would provide changes in background mimicking a real-life usage (Figure 8.16).



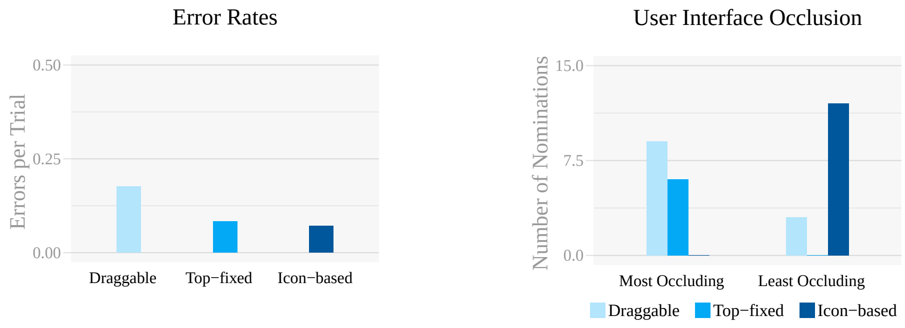
(a) Draggable Menu *(b)* Top-fixed Menu *(c)* Icon-based Menu

Figure 8.16: Axis+Content: Design alternatives with background stream (from [296])

First, we welcomed the participants, provided them with a brief introduction to the study procedure and handed out consent forms. Only once consent was given, we continued by handing out a demographic questionnaire. Next, we exposed the participants to the first UI condition. We provided the participants with a quick 30 second period of free exploration before we continued. After this initial exploration, we asked them to carry out a set of instructions that were given to them. These incorporated common UI interactions such as opening or closing the menu, moving a menu dialogue to a particular on-screen location or adjusting, for instance, the position of the joysticks or the TrackLine. We confronted the participants with such instructions for two minutes. Simultaneously, we instructed the participants to observe a set of 13 bubbles that would appear on top of the background. The set consisted of multiple red bubbles and one green bubble each of which would show a number in its centre. The layout of the bubbles was not random, and we placed them important at presumably important screen locations for camera operators (centre, corners, edges). Also, we gave the participants the task of identifying the green bubble and reporting back its number as quickly as possible. Per trial, the bubbles were visible for 1.5 seconds, which is a retention time recommended by Eng and colleagues [85] for visual search tasks and would appear after a random interval of 6.5 to 10 seconds. Next, the following UI condition was presented and the procedure repeated until the instructions were carried out with all UIs. Following the third condition, we conducted a semi-structured interview on occlusion and personal preferences regarding the UIs, their handling and asked the participants to rank the alternatives. Finally, we debriefed and thanked the participants.

Applying the approach as mentioned above, we collected data on occlusion which we derived from objective as well as subjective measures. The objective data was derived from the Error Rate in the reports of the participants. This rate can be determined by matching their answers with the correct answers.

We conducted a statistical analysis of our gathered data regarding the Error Rates (Figure 8.17). To test for main effects, we applied Friedman's Test. For post-hoc pairwise comparisons, we used Conover's Test with Bonferroni correction which was only executed after finding a significant



(a) Error Rates *(b)* Top-fixed Menu

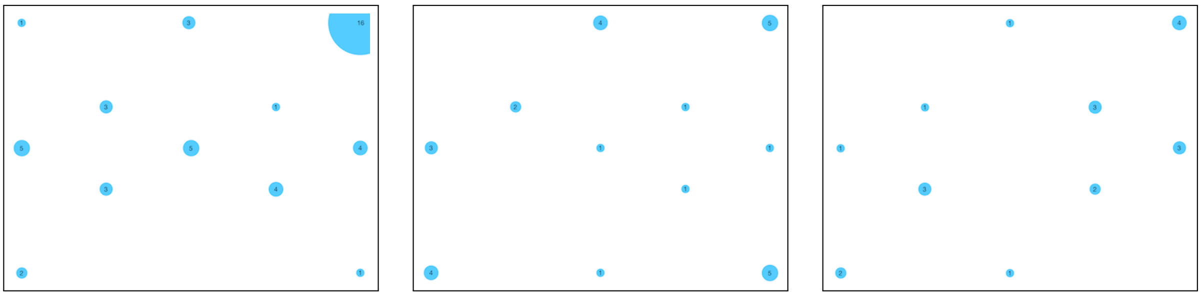
Figure 8.17: Axis+Content: Results of evaluating design alternatives

effect with the omnibus test. We calculated the Error Rate for each UI condition per participant concerning the number of trials. The latter varied based depending on the random intervals from 12 to 19 trials. Examining the resulting dataset with Friedman's test revealed a significant main effect. ($\chi^2(2) = 6.107, p \leq .047$). Post-hoc analysis suggests significant differences between the Draggable Menu ($Mdn = .177$) and the Top-fixed Menu ($Mdn = .083, p \leq .001$). Further, a difference is also indicated between the Draggable Menu and Icon-based Menu ($Mdn = .071, p \leq .001$). However, no effect was found ($p \leq 1.000$) comparing the Top-fixed Menu to the Icon-based Menu (Figure 8.17a).

We further plotted the screen locations and frequencies for each UI condition (Figure 8.18). Here, a blue bubble, in general, represents that errors were made at this particular location; its size and number inside refer to the total number of errors.

Additionally, we investigated the data collected in our semi-structured interviews. Here, 9 participants reported that they thought that the Draggable Menu was most occluding the underlying camera stream. Whereas, 6 participants thought that the Top-fixed Menu led to the greatest degree of occlusion. None of the participants opted for the Icon-based Menu. We also turned the question around and asked on which design led to the least degree of occlusion. Here, 12 participants chose the Icon-based Menu, 3 chose Draggable Menu and 0 chose the Top-fixed Menu (Figure 8.17b).

In summary, our study suggests that the Icon-based Menu can lead to a decrease in the Error Rate compared to the Draggable Menu and that it is additionally also perceived as to be the least occluding compared to both alternative designs (the Draggable Menu and the Top-fixed Menu).



(a) Draggable Menu *(b)* Top-fixed Menu *(c)* Icon-based Menu

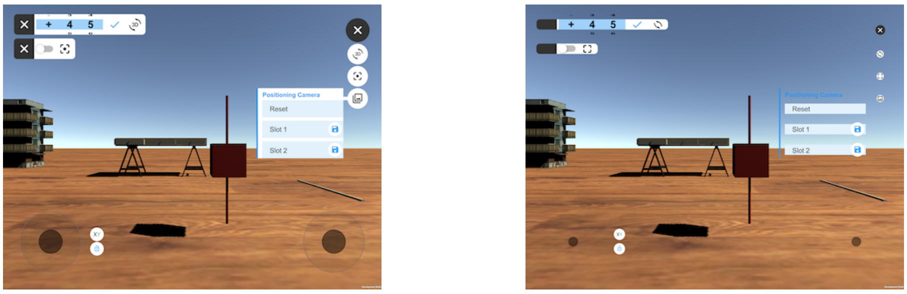
Figure 8.18: Axis+Content: Error distributions for our design alternatives (from [296])

8.3.3 User Study 2: Progressive and Static Reduction

After our initial implementation in Framer, we developed a follow-up prototype in a virtual environment using Unity3D (v 2017.2). We additionally improved the original design regarding several aspects that we derived from the semi-structured interview. Now, the main menu could also be dragged and dropped as the Draggable Menu tested in the previous study. This addition should account for the fact that important details could also be located in the upper right corner which if occluded can increase the Error Rate as it became apparent in plots of the locations of the errors (Figure 8.18a). Further, we exchanged the icons representing the top-level items of the nested hierarchy for more distinct and clear ones. Concerning sub-menus and the contained menu items in lower levels of the hierarchy, we replaced some features that we included in the previous design to get a more sound feature set. Finally, to further minimise the visual footprint, we incorporated Progressive Reduction into our iterated Icon-based Menu design (Figure 8.19).

To investigate the effects especially of the latter aspect of the design iteration, we conducted another user study. We invited 24 participants (20 male, 4 female) with years of age ranging from 19 to 31 with a Mdn of 23 years. Two of the participants reported only a limited degree of previous experience with tablets. Four of them reported previous use of camera motion tools; one was left-handed. We opted for a between-groups design for this study, to better isolate the effects of PR from learning or order effects (similar to Section 8.2). In total, we used 3 groups with different instances of our prototype regarding the reduction implementation. The Baseline group would use the prototype without any reduction. Secondly, we used a Progressive Reduction group. Here the visual UI elements would gradually minimise over time. For this study, we capped the reduction ahead of the disrupting point previously identified. Thirdly, we used a Static Reduction group that would already start the reduction level that the PR group would end at and that would not reduce any further.

To provide a large enough degree of variation, we used three different tasks that would require the participants to utilise manual as well as automation control features. In the first task, we used three different product shot scenarios similar to the ones used in Section 8.1. When executing this task,



(a) Initial reduction level (Baseline) *(b)* Final reduction level (Static)

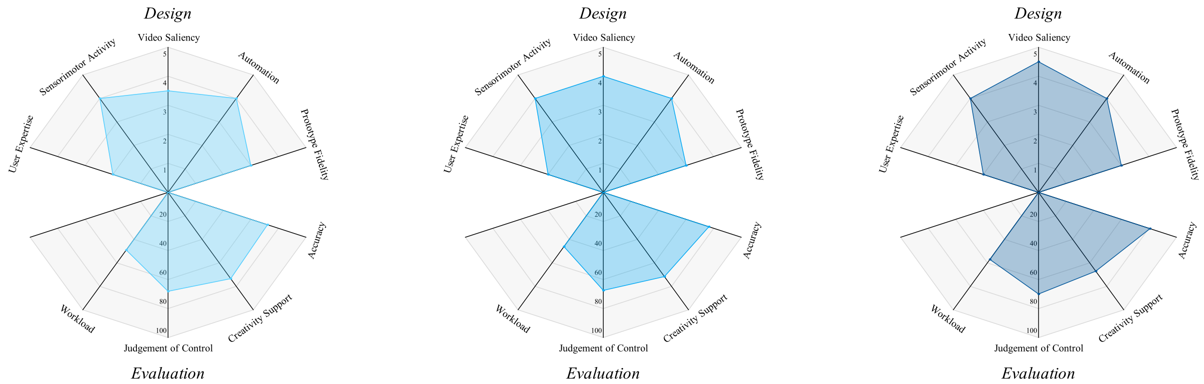
Figure 8.19: Axis+Content: Levels of reduction of our iterated Icon-based Menu design (from [79])

we displayed a grid overlaying the background supporting a framing according to the Rule of Thirds. After a short countdown, one out of a total of three off-screen cubes entered the frame on the left. The participants were asked to frame each incoming cube at the first horizontal intersection with the grid in the direction of the movement. The cubes had varying properties. So, the first cube would slowly enter the screen and continue to move horizontally to the right and revert its movement direction twice (Direction Change). The second cube would move faster without any direction changes (Fast Motion). The third cube would follow the trajectory of an eight on its side (Figure 8). We used these tasks already prior to this study as described in Sections 8.1 and 8.2.

For the second task, we asked the participants to position the camera at three predefined positions within the 3D environment. As a visual prompt and support for the participants, we printed out the resulting view from these predefined positions and hung them on a wall at eye level. Once they were finished with their positioning and framing, we logged the position of the virtual camera and continued with the next frame until all were finished.

The third task required the participants to steer the camera regarding translation on the x and y-axes and the z-axis. To enable such steering behaviour participants needed to change options in the advanced settings menu. Again, a cube entered from a left off-screen position and moved horizontally towards the centre of the screen. Next, it transitioned into a movement along the z-axis away from the centre (towards the horizon). The task was threefold for the participants. At first, we asked them to follow it maintaining a certain distance. Then, we reset the scene and instructed the participants to follow the cube from a Birds-eye perspective (downwards shot at an angle of 45° as often used with drones). Finally, we prompted the participants to additionally orbit the cube (360° movement trajectory around the cube) while keeping it in the centre of the screen and maintaining the perspective.

Beginning the study, we first welcomed the participants, provided them with the necessary information regarding the study, its procedure and how the data is handled. Subsequently, we asked for their consent and only continued once consent was given. We handed out a demographic questionnaire.



(a) Baseline *(b)* Progressive Reduction *(c)* Static Reduction

Figure 8.20: Axis+Content: Diagrams of the design variables and evaluation criteria of the alternatives

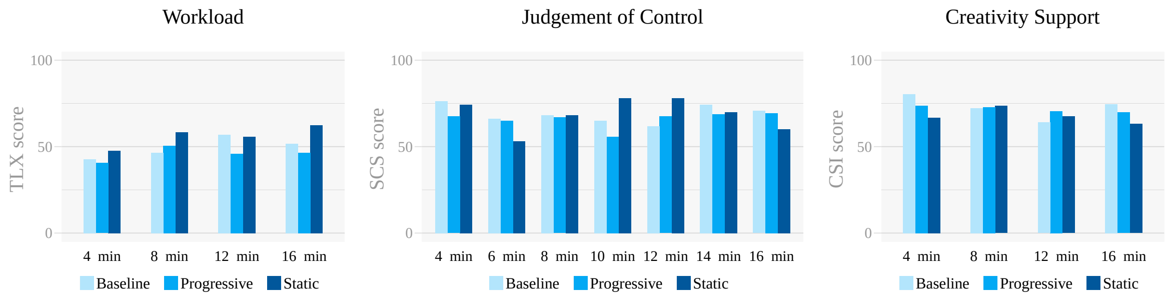
Next, we provided each participant with a four-minute-long guided exploration phase of the assigned UI. Subsequently, we asked the participants to carry out the study tasks mentioned above. For the course of the study, we asked them to do so while standing as we did in earlier studies.

We collected data on Workload, Judgement of Control and Creativity Support using the questionnaires mentioned in Section 6.2. We sampled the first data points after the exploration phase. Then the first task was carried out provided a four-minute-long window. After two minutes, we would sample data regarding Judgement of Control, after four minutes on all the above-mentioned metrics (Figure 8.21). Next, the second task was introduced and the procedure repeated until all tasks were carried out. Subsequently, we conducted a semi-structured interview asking on UI occlusion and usability of the UI. Finally, we debriefed and thanked the participants. Additionally, we gathered data regarding the Quality of Control in terms of Accuracy of all three adopted positions in the second task. Here, we determined the Accuracy by calculating the Euclidean Distance from the ideal position to the participants' position. Here, we followed an idea similar to our adapted SDLP as mentioned in Section 6.2. Shortly summarised, the shorter the distance to the ideal position, the better the Accuracy.

8.3.4 Results and Analysis

To test for main effects regarding Workload, we used the Scheirer-Ray-Hare Test. We found no indication for a UI × Time ($H(6) = 2.064, p \leq .914$) interaction. Analysing the items individually, we found no significant effects, with $H(2)_{UI} = 5.115, p_{UI} \leq .077$ and $H(2)_{Time} = 3.821, p_{Time} \leq .914$.

Concerned with Judgement of Control, we also found no indication for a UI × Time ($H(6) = 12.135$, $p \leq .435$) interaction. Analysing the items individually, we also found no significant differences, with $H(2)_{UI} = 1.476, p_{UI} \leq .478$ and $H(2)_{Time} = 6.389, p_{Time} \leq .381$.



(a) Workload *(b)* Judgement of Control *(c)* Creativity Support

Figure 8.21: Axis+Content: Results of measuring Workload, Judgement of Control and Creativity Support

Further, focusing on Creativity Support, we again found no indication for a UI × Time ($H(6) = 5.680$, $p \leq .460$) interaction. Analysing the items individually, we also found no significant differences, with $H(2)_{UI} = 2.423$, $p_{UI} \leq .298$ and $H(2)_{Time} = 2.897$, $p_{Time} \leq .408$.

We also conducted the Scheirer-Ray-Hare Test for our data on Accuracy. We also found no indication for a UI × Position ($H(4) = 1.233$, $p \leq .873$) interaction. Analysing the items individually, we found no significant differences for the UIs with $H(2)_{UI} = 1.732$, $p_{UI} \leq .421$. However, we found a main effect for the different positions with $H(2)_{Pos} = 6.19$, and $p_{Pos} \leq .045$. Post-hoc analysis using Conover's Test in combination with Bonferroni correction, however, did not further indicate significant differences.

Based on our data, one could argue that Progressive Reduction was not needed in the case of our second study. Thus using the statically reduced design would already be sufficient. However, we would further argue that this does not yet imply that a gradual approach yields no benefits in general. We used a feature set that we think of as reasonable and realistic, however surely one could add further features, which might make a gradual reduction process more beneficial. Further, we did only conduct a short-term study in a controlled environment; thus long-term effects (potentially studied in the wild) cannot be derived from our approach.

An additional limitation is that we did not compare our result to a real-life status-quo design. However, we would argue that at least our methodological approach can be easily transferred to contemporary frameworks that, for instance, run on drones as used by DJI. So, in conclusion, we designed and evaluated touch-based user interfaces, which integrate manual controls with assistance features. The fundamental principles tested in the design alternatives can be transferred to controlling various camera motion tools such as drones, booms or sliders.

We conducted this research with the central premise of minimising occlusion and visual clutter in mind by combining various approaches such as the layout per se, disclosure, animated transitions and (progressive) reduction. As a result of our process, we can present a functional UI prototype that addresses the issues that we initially posed, and might be used as a foundation to inform the design of contemporary or future systems.

8.3.5 Lessons Learned

From both studies mentioned above, we could take away some lessons that we want to detail below:

Lesson 1: The Icon-based Menu was preferred by participants as it was perceived as the least occluding compared to other alternatives and less error-prone than the Draggable Menu.

Lesson 2: Iterating the UI design, we integrated manual control elements, automation features and provided a menu that also was designed for a minimal visual footprint. Evaluating the resulting design, we found no difference between further additionally implemented gradual or static reduction strategies.

Lesson 3: For us, this infers that if a minimal UI is the paramount objective, one could recommend already starting with a minimal UI (as long as it is above the critical point mentioned in Section 8.2). However, if the objective is to get to a minimal UI while maintaining experiential aspects, we would recommend further studying gradual reduction, especially in long-term usage contexts.

8.4 Mid-Air Gestures: Elicitation and Evaluation of Gestures for Cinematography

In addition to our previously examined strategies for minimising the visual UI footprint, also other approaches can be considered outside the confines that tablet displays establish. For instance, these could be the use of external physical controllers (as already used in practice) or alternatively the application of mid-air gestures. These control styles afford haptic or proprioceptive feedback which on the one hand is (somewhat) missing in touch-based UIs and on the other hand are observed to be positively factoring in regarding the emergence of a sense of agency. Regarding these aspects, one could say that they not only can contribute to minimalist visual UI appearances (by disclosing particular functions into a different input modality) but simultaneously also to promoting a sense of control. As both aspects are central in our work, therefore *we further investigated how mid-air gestures might be designed and applied for the context of controlling cinematic tools*. In particular, we focussed on the cinematic use of drones as detailed below.

8.4.1 Background and Related Work

There are different ways of technically implementing gesture recognition and control. Besides CV-based approaches that we already introduced in Section 4.2.3, also dedicated external devices can be used. Two devices of this kind recently emerged and are discussed below.

The Myo system mainly consists of an armband that collects data on muscle activity via Electromyography (EMG) and device motion using an accelerometer and a gyroscope. The data of these sources are fused, and arm motion and hand gestures are inferred. The system also provides an API which makes it attractive for application developers. For instance, there exist applications that use a Myo armband to control a Parrot AR.Drone [331, 360, 477]. However, to the best of our knowledge, prototypes exploring this interaction style in more detail, especially beyond predefined gesture motion by the manufacturer or simple navigational commands are hardly used and researched. An alternative technological approach for gesture recognition is the Leap Motion. It makes use of two wide-angle lens cameras and three infrared LEDs for capturing and tracking hand motion. Similar to the Myo, the system offers an API supporting the development of custom applications and some implementations for controlling a Parrot AR.Drone are documented [349, 493].

The fact the additional equipment is necessary might be considered a hindrance to the adaptation of mid-air gestures by a broader audience. Alternative technical approaches address this issue. Contemporary research investigates the use of wearable devices, most prominently, smartwatches, for the recognition of mid-air gestures. For instance, Laput and colleagues [148, 546] examined the use of bio-acoustic hardware to collect data for classifying hand gestures (flicks, claps, scratches or taps) in combination with on-device motion tracking (accelerometer) by increasing the sampling rate of an off the shelf smartwatch to 4000 Hertz (Hz) (priorly 100 Hz). Their approach seems promising considering the accuracy of detecting and classifying gestures or activity patterns. However, regarding its technical properties, it still seems battery draining at the moment. Nonetheless, assuming that following implementations might achieve similar results using different and less draining approaches and that, in general, smartwatches potentially become more commonly used devices like smartphones or tablets, increased use of mid-air gesture interaction does not seem as far-fetched.

Beyond technical aspects, related work is also concerned with exploring the gestures themselves. ▶ For instance, Cauchard and colleagues [39, 467] conducted a user study and derived a set of basic mostly navigational commands for the control of drones. In detail, they conducted an elicitation study deriving gestures as well as speech commands via a Wizard-of-Oz apparatus. They did not instruct the participants on which modality to use and subsequently compared their usage patterns. They found that gestures were preferred compared to speech instructions. However, 26% of interactions used both modalities suggesting that speech was used along with (some) gestures. Further, regarding the underlying metaphors that users would choose they also found reoccurring patterns: interacting with the drone as a person or as a pet.

Replicating the approach of Cauchard and colleagues, E and colleagues [75] used an elicitation study to derive a set of gestures and speech commands. In contrast, to the first study which only sampled participants from the United States of America, in this study they sampled from a Chinese population. One of the major objectives of the study was also to investigate whether culture-specific difference might affect the result (using the work of Cauchard and colleagues for comparison). Their findings suggest that in both studies the majority of the participants used more gestural commands that speech instructions. However, Chinese participants chose a multi-modal command (gesture and speech in combination) more often (in 45% of the interactions).

What both studies have in common in terms of the methodology is that they use an approach that is referred to as elicitation. One of the earliest mentions that emphasise a direct derivation of gestures from the natural gestural inclinations that users display is found in the work of Wobbrock and colleagues [278]. While they are focussed on gestures for interacting with touch-sensitive surfaces, their approach had the same goal in mind: to infer a set of reasonable gestures from the users' observed behaviour. As a means to come to such a result, they displayed the effect of gestures to the study participants and ask them to perform the gesture that would cause the effect. The authors recorded the gestures provided, formed a taxonomy and matched similar gestures. To come to a more detailed understanding they further derived agreement scores, indicating which types of gestures were used most commonly. The most promising would eventually be considered for the resulting gesture set. What is also noteworthy is that the authors developed a gesture set on their own ahead of conducting the study. Although they were knowledgeable of the domain, their results only covered roughly 60% of the gesture variability compared to the set that resulted from the study. In later work, Vatavu and Wobbrock [264] extended the method of deriving the agreement score such that additionally disagreement and co-agreement was inferred.

Also, other methodological approaches are reported in the literature. For instance, Obaid and colleagues [200] also aimed at deriving a gesture set and taxonomy for the navigation of drones applying elicitation. In contrast to the presented work, they exposed the participants to videos showing the effects of the drone movement. The gestures were subsequently executed by the participants and picked up by a 'wizard' who controlled a drone-based on these commands.

Within the bounds of today's technological constraints, gesture control is already implemented in some commercially available drones. For example, the DJI Spark or the Hover Camera integrated full-body or arm and hand gesture recognition. Contemporary systems often incorporate pre-programmed functions that are triggered by a gesture such as the tracking and following of a particular person (usually the operator or owner). Additional features that might be triggered could be, for instance, an orbiting shot around the person.

In the case of the DJI Spark, a smartphone or a physical controller that is connected to a smartphone can be used as an interface in general. Once the system is set to gesture mode, it will try to recognise a predefined gesture set by analysing the video stream [464]. For example, if a user ▶ waves at the drone, it starts following that person. Further, detailed aspects of how the drone should move while following can be subsequently set and edited on the smartphone. Alternatively, users can also use one of their palms to basically steer the drone (left, right, up, down). Video recording can also be started by extending one's arm and lifting it; roughly at a 45° angle. While recording, another wave at drone can be used to trigger the 'fly away' command resulting in a wider shot.

Also, more advanced systems are more likely to be used in a cinematic environment. These sometimes already feature such a gesture mode as the DJI Mavic Pro or the Phantom 4, for example. Similar to the Spark, however, the predefined gesture sets are rather limited and not necessarily designed focusing on a cinematic context. They rather can be seen as a general-purpose instruction set. The systems further feature a purely CV-based approach, therefore the speed of the gesture recognition seems to be limited and it is an open question how the accuracy of the recognition is affected by an even larger gesture set (as desirable for specific cinematic use).

Pre-Study: Exploring the Effects of Mid-Air Gestures

Our initial ethnographic studies suggested that providing users with a sense of control is important. Further, the literature on sense of control and/agency indicates that putting an emphasis on physical activity during the interaction with systems might even have a positive effect. Therefore, we investigated the effects that the introduction of mid-air gestures in a cinematic UI might have in a controlled user study [123].

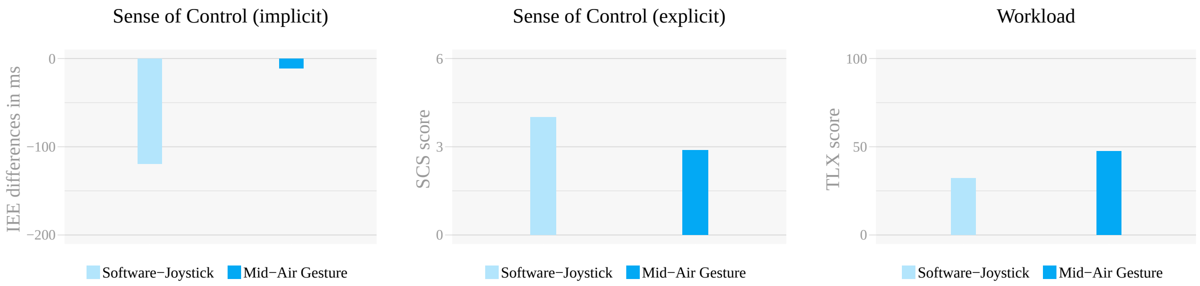
(a) Introduction *(b)* No Subject

Figure 8.22: Mid-Air Gestures: Pre-Study UI conditions

In particular, we compared a traditional Software-Joystick to a Mid-Air Gesture UI (Figure 8.22). The horizontal position of the software joystick controlled the direction and the speed of our slide that we used to facilitate camera motions. The Mid-Air gesture condition was implemented using a Leap Motion for controlling the direction and speed via the pitch of the hand and forearm and mapped on to control commands for the slider.

For conducting the study, we recruited 8 participants (4 male, 4 female) between 25 and 34 years with a Mdn of 26.5 years. To avoid interference due to the handedness of the participants in the Mid-Air Gesture condition, we restricted our sampling to right-handed people. All participants reported no prior experience in professional camera work with two participants indicating use of a gesture control interface before.

We used a within-subjects study design in a controlled environment. The participants were exposed to the UI conditions in counterbalanced sequences to avoid learning or order effects. During the study, a portrait photograph was projected onto a wall and filmed by a camera mounted on our motion controlled slider. The photograph moved from left to right and participants were asked to follow it with the camera while framing it in the centre of the screen, mimicking a tracking shot. The setup resembled a studio film environment with a moving actor and a MoCo tool involved. For the display of the photograph, we used a short distance wide-angle projector. The projection was captured with a DSLR camera and transmitted wirelessly to a tablet to provide visual feedback. To further support framing the photograph at the centre continuously, a line indicated the centre position was displayed on top of the video stream.



(a) IEE differences *(b)* Sense of Control *(c)* Workload

Figure 8.23: Mid-Air Gestures: Results

As we measured sense of control implicitly via intentional binding, our procedure was modelled after the reports found in the literature as by Coyle and Limerick and colleagues [60, 157, 270]. We further adapted it to a camera motion task. After welcoming the participants, we informed them about the study details and asked them to sign a consent form. Next, the participants were presented the first UI. Ahead of each trial, a short training phase (< 5 min) was given for each condition and the study task was explained. After each training phase, also an interval estimation training phase was conducted. The participants were given two sequential audio stimuli and asked to guess the amount of time between them and subsequently, the correct answers were revealed.

Next, the assigned trial started and the participants used a UI to steer the MoCo slider to follow the moving image and frame it in the centre. After filming for 7 to 13 seconds, and after an additional random interval an audio-visual cue appeared that signalled the participants to stop. With an action-effect delay of 500 ms, 800 ms or 1500 ms a second cue was presented indicating that the recording successfully stopped. The exact intervals were taken from related work [13, 60, 254].

Subsequently, the participants were asked to estimate the interval (implicit measurement). Each task and measurement was repeated ten times for each condition. After finishing the tasks with one UI, questionnaires on the Sense of Control (explicit measurement via SCS), Workload (Raw TLX) and demographics were handed out. After filling in the questionnaires for all interfaces, the participants were debriefed and thanked for their participation.

Regarding the data on Interval Estimation Errors (IEE) and Workload, Shapiro-Wilk Tests were non-significant, therefore normality can be assumed. Still, negative covariance could lead to a loss of statistical power for paired samples. We thus determined the covariances and found that all data sets (including the SCS) have a positive correlation between both conditions. As further an overestimated effect size could occur, due to a correlation between related samples we corrected the effect size according to Lakens [147].

Sense of Control (Interval Estimation Errors)

Assuming normality, we conducted a repeated-measures ANOVA to compare IEEs. We found a significant difference for Software-Joystick (Mn $=$ -183.19, SD $=$ 216.75) and Mid-Air Gesture (Mn $=$ -11.04, SD $=$ 187.15) with $F(1,7) = 9.28$ and $p \leq .019$. These results suggest that the interaction style affected the sense of control of the participants. More specifically, our results suggest that it decreases for the Gestures.

Sense of Control (SCS)

To compare the data collected in both conditions we ran a Wilcoxon-Signed-Rank test and found a significant difference for Software-Joystick (Mdn $=$ 4) and Mid-Air Gesture (Mdn $=$ 2.5) with $Z = -1.98$ and $p \leq .047$. This suggests that also that sense of control decreases for Gestures.

Workload (Raw TLX)

We also sampled workload via the Raw TLX and derived a task load score. For both index scores, we conducted a repeated-measures ANOVA. We found a significant difference in the scores for Software-Joystick (Mn $=$ -32.13, SD $=$ 11.74) and Mid-Air Gesture control (Mn $=$ 47.38, SD $=$ 15.4) with $F(1,7) = 10.31$ and $p \leq .015$. These results suggest that the interaction style affects the perceived workload, in particular, it increases for Gestures.

Discussion

Contrary to our initial hypothesis, Mid-Air Gestures could not increase the perceived Sense of Control, at least for continuous control. This was emergent from the explicit as well as the implicit data and supported by our Workload measurement. However, the p-value in the explicit data analysis ($p \leq .047$) is derived from a single 6-item rating scale and barely below the alpha level of 0.05. To rely only on this measurement alone, would not necessarily provide strong support for our conclusion. However, we can take the implicitly collected data into account as well which further supports our prior conclusion of a measurable significant difference. In conclusion, these results do not equal a general verdict against Mid-Air Gestures in our estimation. Rather, they might still be suitable for discrete control operations, such as starting, pausing or stopping camera motion or triggering automated tasks (e.g., moving closer towards the recorded subject). Although Gestures could not outperform a Software-Joystick regarding Sense of Control, they still provide the benefit of not occluding the screen, allow for triggering moves and coarse adjustments in an eyes-free interaction style.

At the moment, however, feature sets as supported by current commercial systems are rather limited and their (potential) benefits consequently harder to study. To enable and support further studying that also focuses on discrete control, we thus investigated how systems featuring gesture control for triggering higher level content-based control should be shaped by conducting an elicitation study.

8.4.2 User Study 1: Elicitation of Candidate Gesture Sets

Independent of questions regarding a technological realisation of such gesture control, we investigated how a gesture set that would focus on use in cinematic production could be reasonably designed. Based on the propositions of Wobbrock and colleagues, we used an elicitation study to derive a gesture set.

The procedure of our elicitation study followed the approach used by Obaid and colleagues. We exposed the participants to a video depicting the movement of the drone and asked them to come up with a suitable gesture for triggering such steering behaviour. For each gesture, we additionally collected subjective data as Cauchard and colleagues did in their study. We extended their approach and consequently asked on the following items:

Suitability: How suitable was the gesture for the current action? (1 (not suitable at all) - 7 (very suitable)).

Ease of Execution: How easy could you could perform the gesture (1 (very hard) - 7 (very easy)).

Naturalness: How natural was the execution of the gesture (1 (not natural at all) - 7 (very natural)).

Based on the gathered gestures and data, we built a taxonomy similar to the process of Obaid and colleagues. We then derived agreement scores (and descriptive statistics) to identify the most frequently used gestures for each particular inspected use case.

To test to a reasonable set of use cases, we first picked a set of cinematic shots that were relevant to drone cinematography. These covered basic steering as well as the emulation of existing tools that require expensive hardware tools. The latter could be the case in mimicking slides (left or right translation), dolly shots (forward or backward translation) or crane shots (up or down vertical elevation combined with left, right, up or down rotation), for instance. To cover these in a structured way, we classified the shots regarding their movement characteristics below.

Slide (left, right): A left or right translation move of the drone.

Dolly (forward, backward): A particular subject of interest is tracked and followed by the drone while framing the front or alternatively the back.

Crane (upward, downward): The drone mimics an establishing crane shot either moving down from an elevated position or moving up from a lower position.

Further, we distinguished between three different contexts as a foundation for our elicitation. The different contexts differ from each other regarding the referent or subject of interest that is captured.

No Subject: A drone might record the landscape or establish the environment of a scene. No subjects of interest are identified or being filmed.

Third Person: A particular subject of interest is given that a drone follows and films.

Self: The subject of interest that a drone follows and films is the operator or owner.

A full list of the fundamental use cases that we used for our study was subsequently derived by asking on a gesture for each possible combination of the two aspects (movement characteristic × context) detailed above.

To classify potential candidate sets and gestures, we built a taxonomy. It was based on the taxonomy provided by Obaid and colleagues and was extended to better fit a cinematic environment. As a result, it covers the following classes: gesture type, gesture form, body parts involved and determination of the subject of interest.

The first three classes are already presented in more detail by Obaid and colleagues and will therefore only briefly be summarised below. First gesture types can be distinguished into four sub-classes:

Deictic: Reference to a direction or position such as pointing at a specific location or swiping left or right.

Iconic: Visual illustration of a symbolic representation of an object (physical, spatial or temporal) such as representing the drone with one's hand and moving it representing the motion trajectory.

Metaphoric: Abstract representation of a concept or real-world property such as putting one's palms open expressing that one is unable to find an answer ('being empty handed' so to speak).

Emblematic: Is independent of visual or abstract representations and requires to be learned and can be culture-specific such as closing all fingers except for the index and second finger which can refer to 'V for victory' or 'peace' if the palm faces away from the body; else it can be a symbol of rude dismissal.

Second, gesture form might further be classified regarding characteristics of it movement:

Static: The core gesturing phase does not incorporate movement, while the preparation phase might entail some motion to start from a resting position or to retract to it.

Dynamic: The core gesturing phase does incorporate movement; just as in preparing or retracting.

Thirdly, which body parts are involved can be categorised in various ways for (cinematic) gesture control the following seems reasonable:

None: If now gesture is performed also no body part is involved.

One Hand: Gestures that are based on the use of only one specific hand.

Two Hands: Gestures that require the simultaneous use of two hands.

Full Body: At least one additional body part alongside the two hands is incorporated.

Additionally, we introduced the class of determination of the subject of interest, that can be further distinguished into four sub-classes:

None: No indication of a subject of interest; gestures control the drone.

Point: The operator points directly at the subject of interest.

Icon: A finger or hand is used as a visual representation of the subject of interest.

Automatic: The subject of interest is determined independently by the drone itself.

For the extension of the taxonomy, we also considered further classes as camera control (is the camera of the drone-controlled separately) and gesture movement (is a gesture performed in a single stroke phase or is a stroke phase being repeated). We excluded those, as they were not suitable and generally applicable enough to apply to the usage scenarios we wanted to elicit.

To conduct the study, we invited 15 participants (9 male and 6 female) with years of age ranging from 19 to 35 and a Mdn of 22 years. All participants reported to be right-handed and were born or raised in Germany. Previous experience with gesture control was reported by three participants, with drones by four and with other motion control tools also by three. We used a within-subject study design to elicit gesture candidates, so we would collect gestures for all usage scenario mentioned above from each participant. First, we welcomed the participants and informed them about the study and its procedure. Next, we asked the participants for their consent and only once a participant consented we would continue. We did so by handing out the demographic questionnaire. Following the questionnaire, the videos were grouped by the referent or subject of interest and presented in random order to counteract learning or order effects. For each subject of interest, three videos would we presented: first, an introductory video depicting the trajectory of the drone, second, the footage recorded by the drone and third, footage of a subject of interest being followed by the drone (Figure 8.24). After each video (except for the introductory video), the participants were asked to perform a gesture that would trigger the depicted drone behaviour. Having performed the gesture, we collected subjective data on Suitability, Ease of Execution and Naturalness as presented above. Subsequently, the participants were asked to perform a gesture instructing the drone to film themselves in reference to the displayed video and to rate their the gesture as done priorly.

In total, we collected 230 gestures that were suitable for later analysis. We needed to exclude some as the participants misunderstood the referent video. We analysed the data regarding the frequencies, distributions and agreement scores regarding the classes of the taxonomy mentioned above for each subject of interest.

(a) Introduction

(b) No Subject

(c) Third Person / Self

Figure 8.24: Mid-Air Gestures: Videos used for the elicitation study

If we look at gesture use across all investigated contexts, we find that Iconic (49.1 %) and Deictic (30.9 %) gestures were chosen most frequently (Table 8.4.1). Only a minority of the participants used an Emblematic (10.9 %) or a Metaphoric one (9.1 %).

Gesture Type	Frequency	Percentage
Deictic	71	30.9 %
Iconic	113	49.1 %
Emblematic	25	10.9 %
Metaphoric	21	9.1 %

Table 8.4.1: Mid-Air Gestures: Gesture types

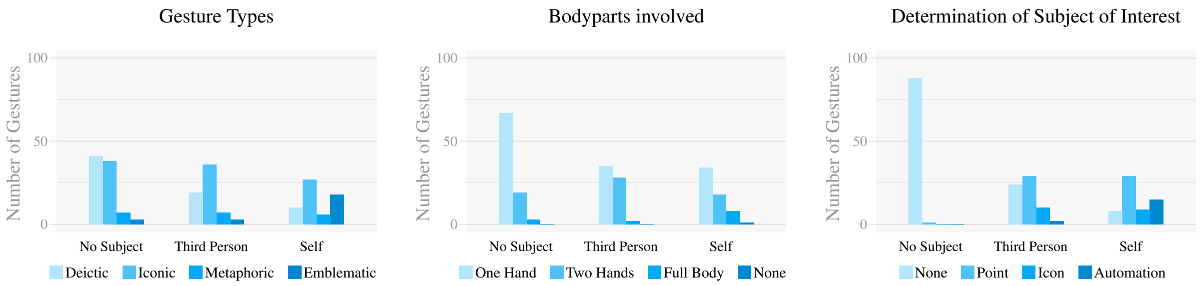
Referent	Agreement	Disagreement
No Subject	.485	.515
Third Person	.407	.593
Self	.393	.607
Overall	.432	.568

Table 8.4.2: Mid-Air Gestures: Agreement scores

Regarding the forms of the gestures, we observed that overall the minority of the used gestures were Static (14.8 %); given the binary division of this class, the majority were consequently Dynamic (85.2 %). Further, the overall majority were executed with One (63.0 %) or Two Hands (30.9 %).

The agreement scores for the gesture types being chosen, can be summarised as follows[1]: The Mn agreement score for all subjects of interest was $AR = .432$ and the Mn disagreement score was $DR = .568$. Similar scores were found for No Subject and Third Person; a higher disagreement score was only found in the Self condition with $DR = .607$ (Table 8.4.2).

[1]Total agreement between participants is referenced by 1 and no agreement at all by 0.



(a) Gesture Types *(b)* Bodyparts *(c)* Subject of Interest

Figure 8.25: Mid-Air Gestures: Results of the analysis per referent

Gesture Types per Referent

When No Subject was filmed, the participants mostly chose Deictic (46.1 %) or Iconic (42.7 %) gestures (with Metaphoric being 7.9 % and Emblematic 3.4 %, see also Figure 8.25a).

Once a particular subject was to be filmed in a Third Person context, the frequency distributions changed. Most participants opted for Iconic gestures (55.4 %). As the second frequently used option participants chose Deictic gestures (29.2 %). Metaphoric (10.8 %) and Emblematic (4.6 %) gestures were less frequently proposed.

For filming one Self, the majority of the participants used Iconic gestures (44.3 %). The second frequently chosen option was Emblematic gestures (29.5%), Deictic (16.4%) and Metaphoric (9.8%) gestures were used less frequently.

Bodyparts Involved per Referent

When the participants were asked on gestures for filming No Subject, most participants used One Hand (75.3 %) or Two Hands (21.3 %). More than two hands (Full Body 3.4 %) or None (0.0 %) were hardly chosen (Figure 8.25b).

To film a Third Person, again most participants preferred to use One Hand (53.8 %) or Two Hands (43.1 %). Full Body was hardly used (3.1 %); no gesture at all never (0.0 %).

To reference a shot of one Self, the majority used One Hand (55.7 %) or Two Hands (29.5 %). Full Body was sometimes used (13.1 %); None were hardly used (1.6 %).

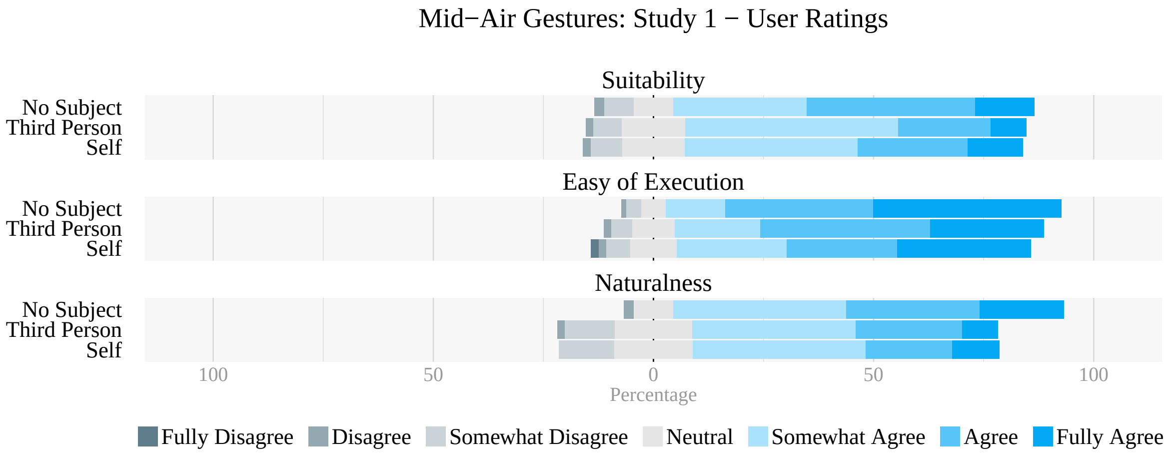


Figure 8.26: Mid-Air Gestures: Results of the first study regarding user ratings

Determining the Subject of Interest

When No Subject was prompted to be filmed, most participants also did not use any indication (98.9 %), hardly any used pointing (1.1 %). The other options were not used at all (Automation and Icon each 0.0 %, see also Figure 8.25b).

Provided with a subject that should be framed in a Third Person perspective, Pointing was chosen most frequently (44.6 %). This was followed by no indication (36.9 %); some participants used an Icon (15.4%) while hardly any relied on the Automation (3.1 %).

To frame one Self, it was most popular to use Pointing (47.5 %). Following that, participants relied on Automation (24.6 %). Some participants would rather use an Icon (14.8 %) or None (13.1 %).

Agreement Scores

The highest Mn Agreement Score (AR = .485) was found for No Subject. In particular, the highest score was found for gestures performing and Crane Shot (Upward) with AR = .673. Surprisingly, the lowest score was identified for Slides from left to right (AR = .364), however, for Slides from right to left the score was medium-high at AR = .509.

For filming a Third Person, that Mn Agreement Score was AR = .407. While participants agreed the least on Dolly Shots (forward) AR = .309, they agreed the most on Orbiting a person (AR = .527).

For filming one Self, we identified an Agreement Score of AR = .393 with the lowest rating for Self-filming positioning the drone on one's left side (AR = .182). The highest rating was found for Orbiting oneself with AR = .655.

Subjective Measures

The user ratings on measures such as Suitability, Ease of Execution and Naturalness (Figure 8.26) for all referents were rated quite high. Similar to the other measures, gestures for contexts without referents seem to be more natural compared to the other conditions. However, given that the participants rated their own creations the overall results also is unsurprising. We took this measure to serve as a baseline for comparison with data from our second study (see details below).

Further Findings

In general, participants tended towards hardly using Metaphoric gestures. Some did and came up with interesting propositions. For example, P6 used an imaginary lasso to pull the drone closer to change the framing in a Dolly Shot. P13 mimicked a drone with both hands in front of his chest tilting to indicate a left or right Slide. Further, P5 used his hands to represent a camera. Both hands formed a frame which also could be used for framing a Third Person. One could also consider this gesture as Emblematic, as it is or at least was commonly used to evaluate how a scene might look.

Emblematic gestures were also used, for instance, by P1 and P7 who formed a “Y” by putting both arms up. This can be considered culture-specific or learned as the gesture is used in DJI systems[2].

8.4.3 Construction of Candidate Gesture Sets

The lower agreement scores for the Third Person and Self conditions led us to the construction of a candidate gesture set for each of theses cinematic contexts. For its construction, we used the taxonomy that we introduced earlier as a foundation. The subjective measurements that we collected were not used to inform the construction process as they might be biased given the high assessments for the gestures proposed by the participants.

Below, we introduce each candidate gesture set in more detail. For each gesture, we further relate to aspects of a potential underlying mental model as they were proposed by Peshkova and colleagues [212].

No Subject

For No Subject being filmed, we mainly chose Deictic gestures for navigating the drone (referring to the movement direction or position) that were mostly Dynamic. Only for Establishing Shots, we used Iconic gestures with the reason being that the majority of the participants opted for those (66.7 %).

[2] The gestures itself originates from signalling helicopter pilots. Here, putting both arms up forming a ‘Y’ references ‘Yes’ (Yes, please land, we need help). In contrast, putting only one arm up, hence forming an ‘N’ would mean ‘No’ (No, do not land).

Navigate Drone (Nav): One executes a swiping gesture to control the drone with the direction of the gesture being mapped onto the direction of the drone's movement. Only for Establishing Shots one represents the drone with a hand. Here, the trajectory of the drone is indicated by the movement of the gesture. These gestures correspond to the *"super-power"* metaphor from the "*instrumented"* mental model where one pulls and pushes the drone in the desired direction.

Third Person

Regarding framing a scene with a Third Person perspective, we focused on Dynamic Iconic gestures as we often observed that users would use one of their hands to represent the drone and moving it to indicate the drone's motion. We also took into account the variance of how people indicated a subject of interest.

Imitative Set (IM): One represents the drone with a hand with the palm and fingers straight and pointing in the direction to be filmed. The movement of the hand is mapped onto the drone's movement. These gestures correspond to the *"imitative"* class of mental models where users imitate the flight of the drone.

Imitative Set with Pointing (IM&P): Users first point at the subject of interest. Next, they use a hand to represent the drone's movement. This set contains a mixture of the *"imitative"* class (representing the drone with a hand) and the *"intelligent"* class (pointing at the subject to be focussed).

Indicate Object and Drone's Position (IO&P): Users take one hand to represent the subject of interest with the index finger pointing up. Additionally, they use the other hand with the index finger pointing down indicating the position of the drone relative to the indicated subject of interest. This metaphor is part of the *"intelligent"* class of mental models where users rely on certain assistance and automation features of a system.

Indicate Object and Movement (IO&M): One represents the subject of interest with a hand by pointing the index finger up. Further, the drone is represented by pointing the other index finger down indicating the position according to the subject of interest. Next, one executes the movement of the drone in relation to the moving subject. This metaphor also can be related to the "*intelligent"* class of mental models, as it incorporates abstract concepts; in particular, depicting the subject of interest with one hand and the drone with the other.

Self

For the Self condition, we relied entirely only on Iconic gestures. We further distinguished between two distinct gesture sets that would differ regarding how a subject of interest is determined: first, Pointing at oneself and subsequent specification of the position of the drone (also by Pointing) and second, Pointing at oneself and relying on automation for the subsequent positioning of the drone.

Point & Automation (P&A): The users point at themselves to indicate to the drone to focus on them. They subsequently rely only on assistance or automation features to handle framing and following them when in motion.

Point & Indication (P&I): One again starts with pointing at oneself and next specifies on which side the drone should follow. This, for instance, can be done by pointing on one's left or right shoulder.

To both conditions focusing on Iconic gestures (Third Person and Self), we added a gesture for orbiting around the subject of interest. While the feature itself is already implemented in off the shelf systems, we hardly found any gesture command in place that would trigger this move.

8.4.4 User Study 2: Evaluation of the Gesture Sets

Below, we report on a further exploration of our candidate gestures sets. We conducted a second user study to investigate how our gesture sets were rated by a population that was not involved in its development. Further, we wanted to examine which and how differences in the ratings would emerge. Therefore, for each gesture, we compared the data collected in this study with the data from the previous study.

We invited 12 participants (6 male and 6 female) who were not part of the first study with years of ranging from 17 to 37 a Mdn age of 22. All participants reported to be right-handed and were born or raised in Germany. Previous experience with gesture control was reported by two participants; with drones was indicated by one and other motion control tools by three participants. We conducted this follow up study in the same environment as the first one using a projector to display the drone and gesture videos. We recorded the participants during the study which took about 30 minutes for its completion. First, we welcomed the participants and informed them about the details of the study. Next, we asked for their consent to take part in the study and only if given we would continue. We went ahead by handing out a demographic questionnaire. Once it was filled in, we started with the data collection on the gesture sets. Similar to the first study we used a within-subject design to investigate the gesture sets. Those were split into three groups (No Subject, Third Person and Self) with each group containing the different candidate gesture sets as described above. For each participant, we randomised the sequence of exposure regarding these conditions. Thus, the participants were exposed to the videos of the gestures of one set. Those were followed by the individual videos regarding a particular referent video (the same we used in the first study). The participants were then asked to perform the gesture they were introduced to themselves and to rate it given the same three metrics that we used in the first study (Suitability, Ease of Execution, Naturalness). The procedure was repeated until all gestures for all conditions were executed.

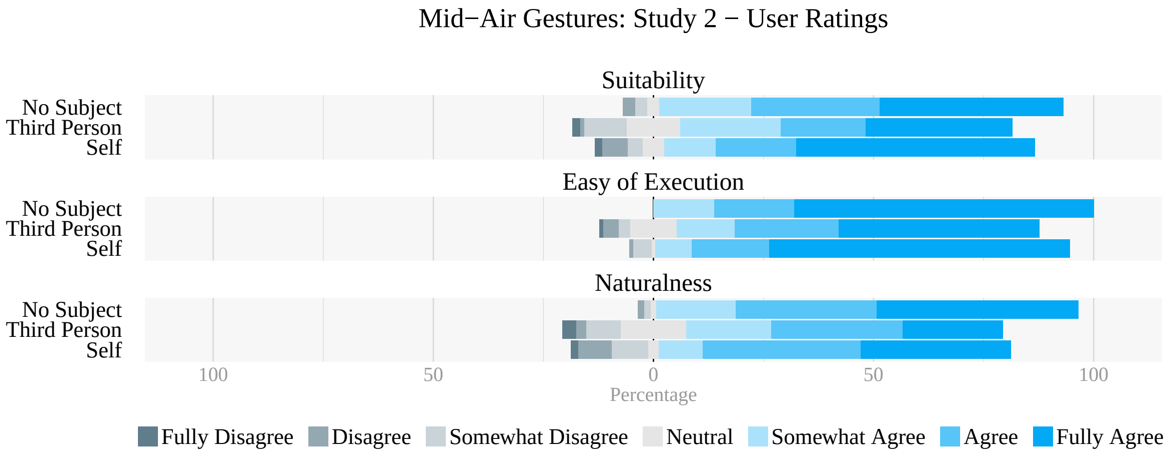


Figure 8.27: Mid-Air Gestures: Results of the second study regarding user ratings

8.4.5 Results and Discussion

As each participant assessed 35 gestures, we consequently collected 420 ratings in total. First, we will provide an overview of overall findings, as we did in our report on the first study, which is subsequently followed by a further analysis for each condition.

In general, we found that the participants rated the Ease of Execution of the gestures the highest (Mdn $= 7$). The second-highest ratings were found regarding the Suitability and Naturalness (each with Mdn $= 6$). The distribution was similar compared to the first study with Ease of Execution (Mdn $= 6$) and Suitability and Naturalness (each with Mdn $= 5$).

Measurements	Correlation Study 1	Correlation Study 2
Naturalness × Suitability	.641	.794
Naturalness × Ease of Execution	.441	.298
Suitability × Ease of Execution	.438	.281

Table 8.4.3: Mid-Air Gestures: Results of the analysis of measurement correlations

Further, we determined a strong positive correlation for Naturalness and Suitability utilising Pearson's Correlation with r $= .794$. Naturalness and Ease of Execution seem to be less correlated with r $= .298$. Similarly, Suitability and Ease of Execution are less correlated with r $= .281$. A comparison with the first user study is found in Table 8.4.3.

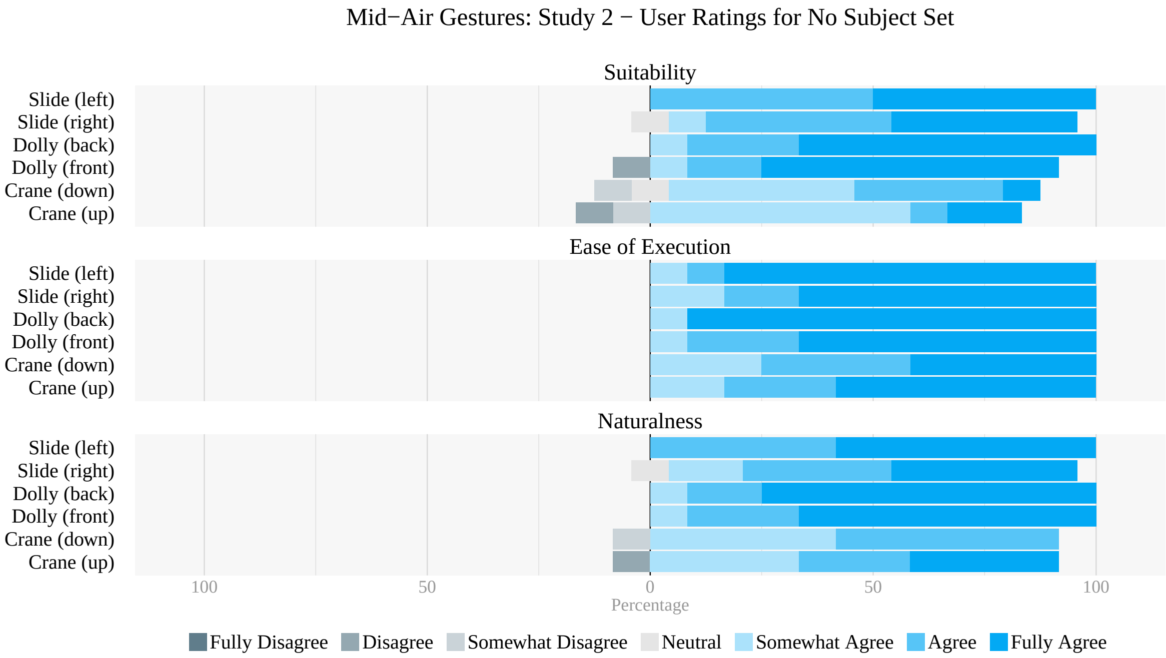


Figure 8.28: Mid-Air Gestures: Results of the No Subject Set

No Subject

For the No Subject condition, we only evaluated one gesture set as it focused primarily on navigational drone control using Deictic gestures. The set mainly consisted of swiping gestures, pushing and pulling the drone forwards and backward gestures. Additionally, two Iconic gestures for performing Establishing Shots by representing the drone and its trajectory with one hand.

All gestures in this set were rated rather high in all three dimensions with a Mdn of 6 for Suitability (first study: Mdn $= 6$), 7 for Ease of Execution (first study: Mdn $= 6$) and 6 for Naturalness (first study: Mdn $= 5$). For this condition, a lower correlation between Naturalness and Suitability was observed with r $= .604$.

The Deictic Gestures were rated mostly rated higher compared to the Iconic gestures for Establishing Shots except for Ease of Execution (Figure 8.28). Vice versa gestures for Establishing Shots were rated the lower for Suitability (Mdn $= 5$ for both (up and down)) and Naturalness ($\text{Mdn}_{\text{down}} = 5.5$ and $\text{Mdn}_{\text{up}} = 6$) but, as mentioned, higher in Ease of Execution ($\text{Mdn}_{\text{down}} = 6$ and $\text{Mdn}_{\text{up}} = 7$).

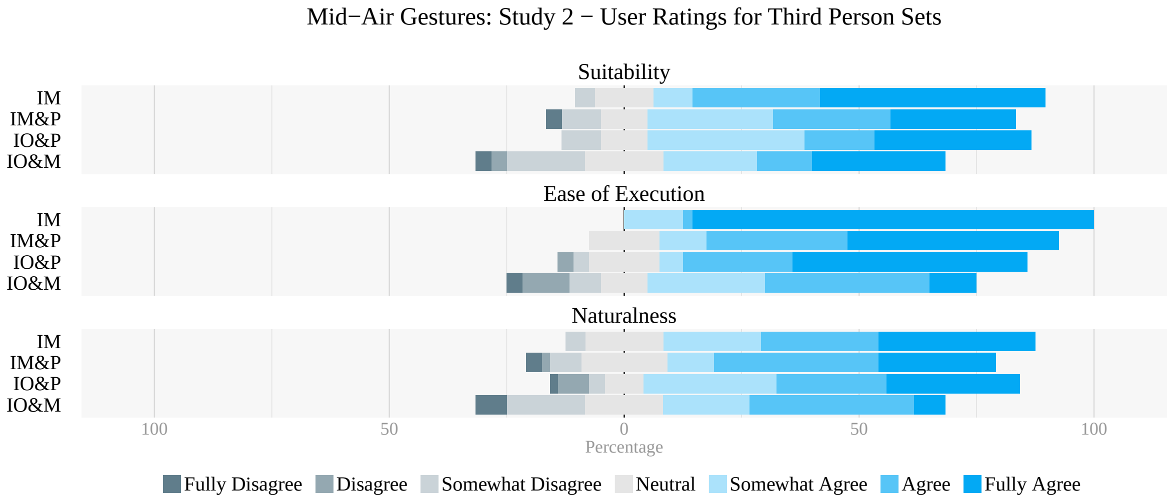


Figure 8.29: Mid-Air Gestures: Results of the No Subject Sets

Third Person

For a Third Person context, the gesture set with the highest ratings overall was the Imitative Set with a Mdn of 6 for Suitability, 7 for Ease of Execution and 6 for Naturalness (Figure 8.29).

Imitative Set (IM): Regarding the individual gestures, the user ratings of the IM set can be characterised as hardly varying except for the dimension of Naturalness. Here, the gestures for drone Translation received the lowest ratings with a Mdn of 5.5 for each, left and right, motion. Further, the overall observed correlation between Suitability and Naturalness was slightly lower in this set compared to the other gesture sets in this context ($r = .692$).

Imitative Set with Pointing (IM&P): This set was overall rated similar to the IM set. However, the Mdn rating for Ease of Execution scoring a 6 was lower compared to IM. Here, in particular, the Orbiting gesture received the highest rating with a Mdn of 7 for Suitability and Ease of Execution and 6 for Naturalness. In general, the correlation between Suitability and Naturalness was also high in this set ($r = .853$).

Indicate Object and Movement (IO&M): Participants rated this set with the lowest scores overall. Similar to the IM&P condition, the Orbiting gestures received the highest ratings with a Mdn of 6.5 for Suitability and 6 for Ease of Execution and Naturalness. We further also observed a high degree of correlation between Suitability and Naturalness ($r = .883$).

Indicate Object and Drone's Position (IO&P): This Static gesture set was overall rated higher than its Dynamic counterpart. Similar to the other sets the gestures referencing lateral movement scored uniformly and the Orbiting gesture was rated the highest with a Mdn of 6.5 for Suitability and Naturalness as well as 7 for Ease of Execution. Again, a high correlation between Suitability and Naturalness was observed ($r = .755$).

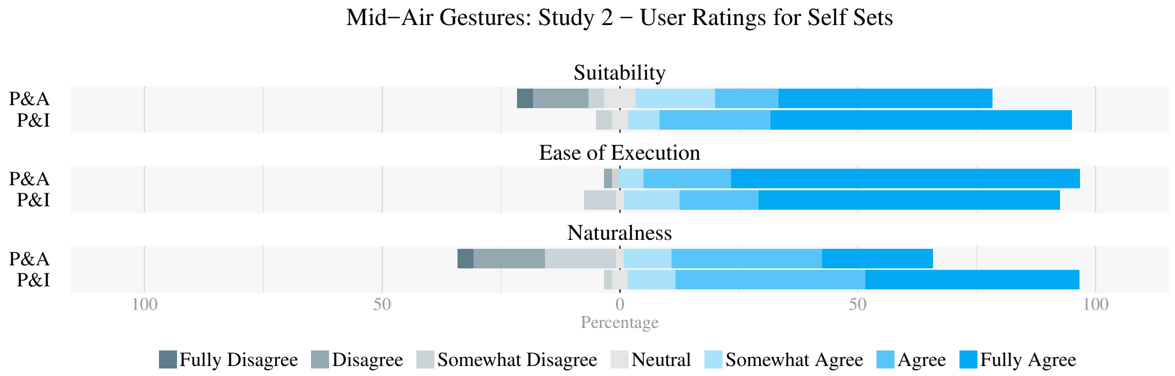


Figure 8.30: Mid-Air Gestures: Results of the Self Sets

Self

For the two sets tested in the Self condition, Pointing and Indication was preferred by the participants based on their ratings. Apparently, as it was more suitable (Mdn = 7) compared to Pointing and Automation (Mdn = 6). Further, also for this condition, we found a high correlation between Suitability and Naturalness ($r_{P\&A} = .829$ and $r_{P\&I} = .601$).

Pointing&Automation (P&A): In terms of Ease of Execution, the various gestures were rated rather uniformly at a Mdn of 7. For Suitability and Naturalness, almost all gestures received lower scores except for Tracking (front) and Orbit.

Pointing&Indication (P&I): This set was also rated uniformly at a Mdn of 7 regarding Ease of Execution. But for Suitability and Naturalness, the majority of gestures was rated higher (compared to P&A) with the Orbiting gestures being the highest rated with a Mdn of 7 for all metrics.

Discussion

Below we will discuss and compare the two studies in further detail. Most apparently, we found in both studies a high correlation between Suitability and Naturalness, although it is pronounced weaker in the second study.

Overall, the evaluated gesture sets were perceived as easy to execute. The often found a high correlation between Suitability and Naturalness can be indicative of a similar or shared underlying measure (or at least that participants perceived the questions in that way). However, further research regarding the nature of this finding is surely necessary to determine to which degree both dimensions are similar and to which degree they are distinct (and in consequence to be able to answer whether they point towards the same phenomenon).

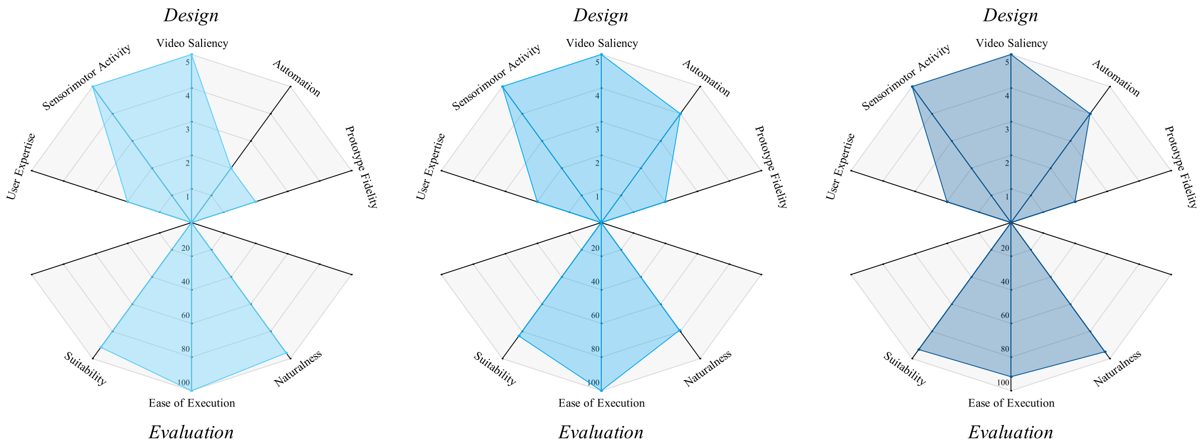
For filming a Third Person, gesture sets that are mimicking the movement of the drone are preferred. Their counterpart, which relies on first pointing at a subject and subsequently mimicking the movement of the drone seem to be less preferred. A similar phenomenon is mentioned by Wobbrock and colleagues [278] who suggest that users might tend towards simplifying interaction metaphors (and hence prefer gesture set that conforms to this preconception). The other tested sets (IO&P and IO&M) were overall given low ratings. The ratings might be due to the complexity of the gestures. While the participants could indicate the subject of interest quite easily, it was harder for them to indicate the positioning of the drone (for example on one's left or right side). A potential difficulty may lie within misalignments between different points of view where 'my left' and 'your left' might be different for the operator and the drone.

Although some gestures were rated similarly in both studies, they in comparison provide further insight into the gesture set preferences and the trade-offs between the functionalities and capabilities of gesture sets in relation to their Suitability, Ease of Execution and Naturalness. Both candidate gesture sets of the Self condition exemplify that seemingly a trade-off between the functionality, the Ease of Execution and Suitability of a gesture set exists. Gestures incorporating pointing at oneself and relying on automation (as in the P&A set) are mainly easy to perform but also face a disadvantage. Their functional versatility is limited which in consequence may make them perceived as more unnatural and less suited for complicated tasks.

Unexpectedly, the participants rated their own gestures overall lower than (neutral) participants did in the second evaluation study. This pattern was consistent across the various contexts that we tested. For the No Subject conditions, the overall ratings of the participants diverged only a little comparing both studies. Concerned with the Third Person perspective one could summarise that regarding all metrics the gesture sets were overall rated higher compared to the first study. Further, for the Self condition, it is noteworthy that in the first study both conditions (P&A and P&I) were rated similarly regarding Suitability and Naturalness. In the second study, however, P&I was rated higher compared to P&A. Overall, this was surprising to us as we hypothesised that the creators of the gestures would rate their own particular creations higher than users who were not involved in the creation process. One potential explanation could be a novelty effect that would affect the ratings (given that mid-air gestures are contemporarily still a less commonly used form of interaction with computational systems). Another explanation might be a bias due to social desirability.

Additionally, the participants did not receive any feedback on the quality of the motion execution. While this was not our focus in eliciting the gestures, it is important to note, that in a fully functional technical implementation the quality of gesture recognition and the executing of the communicated motion paths might additionally affect the perceptions and ratings of users.

From the various evaluated gestures sets, one can also derive representations based on our proposed template regarding design variables with adapted evaluation criteria as in Figure 8.31. For the No



(a) No Subject: Nav *(b)* Third Person: IM *(c)* Self: P&I

Figure 8.31: Mid-Air Gestures: Diagrams

Subject condition, only one set was tested (Nav). Concerning the Third Person condition, the highest-rated set was IM and likewise, for the Self condition, it was P&I.

8.4.6 Lessons Learned

From both studies described above, we could take away some lessons that we want to share below:

Lesson 1: For pure navigation purposes when No Subject is being filmed, Deictic gestures mapping the drone's motion to the movement of one's hand were preferred by users.

Lesson 2: For cinematic contexts such as Self filming or filming a Third Person, users mainly preferred gestures involving One Hand. For filming a Third Person, they rated emulating the movement of the drone with one hand highest. In the case of Self filming, first pointing at oneself and subsequently at the position from which the drone should film was the highest-rated option.

Lesson 3: Overall, the participants preferred simple gestures over complex gestures. Complexity might additionally be introduced by different points of view when users and drones face each other. Confusion due to the missing mapping between 'my left' and 'your left' is highly unwanted.

Lesson 4: Surprisingly, given an elicitation setup as in our study, we found that participants would be self-critical and underestimated the gestures they would come up with (compared to a neutral control group).

8.5 In-the-Wild: Examining and Addressing Information Display in Current Systems

For the projects presented above, we developed designs that assisted users while providing them with a sense of control and minimising the visual footprint of their UIs. So far, addressing the latter aspect, we mainly used technical approaches such as altering the visual representations or outsourcing the interactions to a different input modality (minimising the need for them to be displayed on the screen). Further, one could also opt for other strategies regarding the visual representation, for example, by filtering the displayed items and rearranging their placement based on a given set of priorities as proposed by related work described in Section 8.2.1. In contrast to the already examined technical approaches that fundamentally treat all items the same way, such filtering approaches are actively rating, for instance, the importance of items and filtering them or changing their visual appearance according to their rating. For such approaches, one of the central questions is "Which aspects are used to inform the rating process?". In general, these approaches might be informed by the usage context or be curated based on domain-specific knowledge (or both). Curation, in particular, might be useful as it might further contribute to increasing the sense of control by focusing on displaying important items (or displaying more important items more prominently) while at the same time to reducing the visual footprint by filtering less important items (or displaying less important items less prominently). To investigate the effects of such an approach it is vital to provide a solid base on top of which the curation process can be built. To derive a set of relevant insights that help to derive such a base, our primary idea, therefore, was to use investigations in-the-wild with experts. Thus, we chose to *closely examine existing contemporary systems for drone operation, identifying challenging design aspects, offering possible solutions on a high-resolution level and evaluating those with experts*.

8.5.1 Background and Status Quo Designs

Steering a flying aerial vehicle is not an entirely new endeavour per se. A long tradition of research concerned with aircraft piloting and its design aspects can be found in the literature [15, 32, 87, 114, 203, 204, 236]. Although, there is a great body of work concerned with manned as well as unmanned aerial vehicle operation, the insights described in this literature have not been entirely translated to the design of UIs for cinematic drones. To us, it seems as if the design of UIs for drone control, at least regarding UIs on mobile devices are rather influenced by general design practices for mobile applications or mobile games. Additionally, the steering of cinematic drones adds another layer of complexity to an already complex task. Besides the primary task of navigating the drone, the operators additionally need to make sure that they are framing the filmed material properly. To provide a foundation for introducing relevant aspects found in research concerned with various aspects of the design and evaluation of UIs for aviation, we want to introduce selected relevant work below.

When comparing contemporary aircraft UIs and cinematic drone UIs, one of the most predominant differences one could identify is probably the varying layouts and patterns that are used in their design. In the case of aircraft UIs, most contemporary systems are using a distinct layout: the Basic-T arrangement (or derivations) of primary flight instruments as, for example, pointed out by Schuivens [236].

(a) Basic-T (full) *(b)* Basic-T (shape) *(c)* Primary flight display

Figure 8.32: In-the-Wild: Basic-T arrangement (Figure 8.32a from [412]) and T-shape (Figure 8.32b from [300]) as used in 'Iron Cockpits' and a primary flight display (Figure 8.32c from [410]) as used in 'Glass Cockpits'

In the Basic-T arrangement, the following six instruments are grouped into one panel: airspeed indicator, artificial horizon (attitude indication), altimeter, horizontal situation indicator, turn and bank indicator, vertical speed indicator. In older aircraft cockpits, so-called 'Iron Cockpits', such a panel would incorporate physical analogue instruments (Figure 8.32). Here, the T-shape is formed using the first four instruments of the list above; the remaining two are used to provide necessary additional information and hence are place at less prominent locations in this layout (Figure 8.32c). In modern cockpits, so-called 'Glass Cockpits', multifunctional displays with digital instruments would replace their physical analogue predecessors and lead to different overall designs which still followed this initial basic T-convention (Figure 8.32c). Such multifunctional panels are often referred to as the primary flight displays and combine the four instruments that would build up the T-shape into one digital instrument. Further, the placement of each individual instrument is kept consistent with their original location in the physical T-model.

One of the prime reasons for the selection and arrangement of these instruments according to their relative importance is to maximise Situational Awareness. The concept has been briefly touched earlier in this thesis and will be a bit expanded upon below. Overall, one could summarise that SA contributes to the overall safety and hence it is emphasised in the design of cockpits and interfaces. Examples of this idea are found in assistance systems for landing the aircraft, for instance. During landing, crosswinds are often encountered and require the full concentration of the pilot operating the aircraft. An assistive system would recognise the landing attempt and regularly report the altitude of the airplay to the pilot via an audio channel. This frees up the pilot from checking the values on the displays and hence allows for concentrating on the steering while maintaining proper SA.

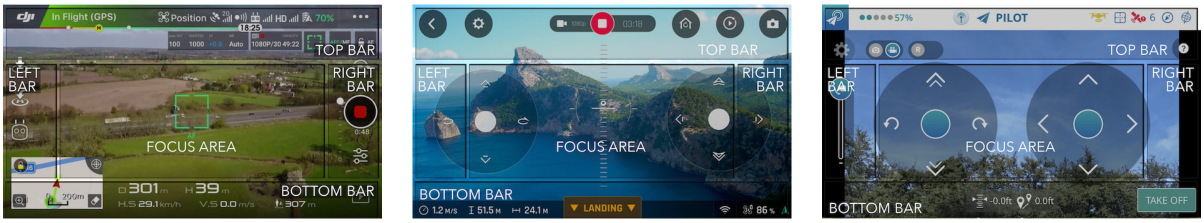
Transferring the context to cinematic drones, SA might further entail aspects surrounding the aesthetic framing and scene-specific dramaturgy that operators also have to keep in mind when recording a scene using drones besides the similarly necessary steering aspects as they are found in aviation.

Analysis of Status Quo Drone User Interfaces

Expanding on Section 2.2.3, we want to provide a more detailed overview of the status quo by further analysing contemporary systems for cinematic drone control. We investigated a selection of different UIs by various manufacturers regarding their commonalities or prime differences and subsequently their advantages or disadvantages. Concerning the manufacturers, we chose to examine systems from DJI, Parrot and Yuneec (Figure 8.33). The manufacturers also roll out various mobile applications that help to steer their drones. Contemporarily, the available applications are the Go 2 [326, 387] for DJI, the Freeflight Pro 3 [365, 388] for Parrot and the Breeze Cam 4 [389] for Yuneec. Ahead of visual analysis of the UIs, we exposed ourselves to self-experience with the drones and the applications by using them in the wild. Here, we noticed that the tested UIs share a similar fundamental layout: all made use of top, left, right, bottom bars and a focus area at the centre of the screen. However, where and how a particular set of information was placed and displayed varied across the applications.

For the DJI Go app, information regarding the connections to the Global Positioning System (GPS) satellites, the remote control, the image transition rate, the battery status and flight mode can be found in the top bar (Figure 8.33a). Further, it displays the battery status in form of a colour gradient at the bottom edge of the top bar. The right bar entails the buttons for camera control options such as record and release. Additional information on the recording status and properties is displayed in the top bar and the focus area. The left bar holds functions for piloting the drone such as starting, landing or returning to a home location. Supporting the navigation, a map referencing the drone's position is displayed in the lower-left corner of the display. The bottom bar is used to present information on the altitude, distance to the controller and flight speed. Here, all information is represented in written form. In the focus area, the properties of the auto-focus function are displayed. In summary, the display of information and functionalities are concentrated at the edges of the screen, and hence the focus area is mainly kept free from UI elements occluding the underlying camera stream. Overall, the presented information seems densely packed which makes it harder to retrieve a particular bit of information at times.

In contrast, the Parrot Freeflight Pro app presents fewer visual elements and consequently seems less cluttered (Figure 8.33b). The top bar holds the camera controls as for starting a recording and information on the properties of the recordings. Additionally, it covers functions for general settings (left) and returning to home (right). The left and right bars provide space for a software-joystick that is placed on each side of the screen (left: rotation, right: translation). Flight information regarding altitude, distance to the controller and flight speed is displayed at the bottom bar in written form along with information on the quality of the connection and the battery status. To further support the operators, the focus area also encodes information, for instance, on the flight speed in a visually abstracted form. In conclusion, similar to the DJI Go, the Freeflight Pro app aims at keeping the centre free and thus placing UI elements at the edges of the screen. Further, due to the fewer elements that are presented and the increased use of UI transparency visual clutter is additionally avoided.



(a) DJI Go   *(b)* Parrot Freeflight Pro   *(c)* Yuneec Breeze Cam

Figure 8.33: In-the-Wild: Comparison of status quo designs

The Breeze Cam app uses the top bar to present most of the displayed information (Figure 8.33c). Similar to the DJI Go app, it provides information on the connection between drone and controller, the flight mode and the GPS or battery status. Often combinations of icons and text are used to encode the information. Given that the importance of particular information changes, for example, when the quality of the connection decreases, the colour of the item also changes (from black to yellow or eventually red). Further, similar to the Freeflight Pro app, the left and right bars are used to display software-joysticks. Additionally, a slider controlling the rotation of the camera gimbal is placed on the left bar. Functionalities for drone control such as starting or landing are found in the bottom bar. Here, also additional information on altitude and distance are presented.

Overall, we found that the three examined UIs followed a similar blueprint. A major difference one could still point at was that the DJI Go app was designed for use with a further physical controller in mind whereas the other ones could be used as stand-alone software. The top and bottom bars were used most often to display vital information and the focus areas were mainly only used for a small number of UI elements. In general, one could also say that in terms of UI design aspects the manufacturers mainly focussed on a consistent look. This visual homogeneity however made it difficult to spot a particular piece of information, especially when an increased number of items were displayed. In the DJI Go app and the Breeze Cam app, we also found adaptations to this issue. Here, depending on the status, items such as most prominently the battery status display would change their visual representation when a critical status was detected. This was mainly done by changing the colour from black or green to yellow or red.

8.5.2 User Study 1: Expert Survey

So far, given our self-experience, we mainly found that systems seem to be built along similar lines and that visual clutter sometimes is hindering information retrieval. However, we did not incorporate the perspective of expert users who are more used to drone UIs and proficient in their use. To also derive insights from their perspective, we set up an expert survey with the goal in mind to identify what experts might perceive as critical problems in contemporary systems.

In Section 3.2.2, we already provided a list of lessons that we learned for conducting research with expert cinematographers. To recapitulate, the list contained the following two statements. We suggested expecting low return rates and adapting *"by handing out more questionnaires and hence include expert users groups on social media"* or *"by shortening the questionnaires to adapt to busy schedules in one to one communication"*.

So for this expert study, we followed our own suggestion by keeping the questionnaires short and handing them out to professionals whom we could contact personally as well as via social media. Therefore, the questionnaire would only cover the five questions that are listed below:

1. What is your level of expertise in operating drones?
2. Do you own a drone?
3. What general difficulties do you notice when operating drones?
4. What is particularly hard to implement or difficult to use?
5. What limitations do you notice when operating a drone?

In sum, we sent our 20 invitations to take part in the user study. The return-rate was about 60% or 12 in absolute numbers (11 male, 1 female). While the majority of the sample (10 participants) rated themselves as experts, two judged their level expertise as being a beginner. This might be due to reaching out to social media groups where filtering potential participants is hard to be done precisely. However, all participants owned a drone, so a minimum viable level of exposure to drone operation can be expected even by the beginners.

We initially contacted the participants via e-mail, private message or social media messages introducing them to the study idea and asking for their consent in taking part. Once they agreed to take part in the study, we sent them a file containing an introduction to the study details in written form and the questionnaire presented above. We further offered the participants two ways of providing a response: either in written form or as an audio message. The latter might better fit into a busy work schedule where one can quickly record the answers during a break and send them via a messaging service. Each participant was asked to answer the questions within two weeks. From the responses we could identify several major domains where difficulties were experienced by the experts. For instance, many answers focused on the legal situation or technical properties concerning the basic quadrocopter hardware design and its consequences regarding flying behaviour. Further, the answers of the beginners and the experts clearly varied. While beginners would mostly comment on technical difficulties and hardships in learning and adapting to different modes of control, the experts would more precisely focus on detailed aspects of the designs. A set of remarks that we found being particularly relevant for the design of the UIs is listed below:

- Not enough customisable buttons to which pre-programmed steering behaviour or shot types can be assigned to (such as 360° view or follow mode)
- If a combination of hardware controller and smartphone is used, information is often displayed redundantly which was perceived as irritating
- Flight properties such as altitude, distance or flight speed should be displayed complementary
- Too much information is displayed on small screen spaces
- The representation of the elements does not mirror their importance for the operation (more important and less important items are visually represented equally)

Overall, the results are similar to our previous findings. First, we found a difference regarding the issues that beginner and expert user groups perceive as important. While beginners seem to have issues with various control aspects in general, experts focus more on detailed aspects that could be improved. Second, as found in our earlier expert interviews presented in Section 3.5, environmental factors such as legal aspects are important but also difficult to handle by the operators. Third, SA and workload seem to be as central for drone operation as they are for aviation in general, although the drone operators might be inclined to use a different vocabulary to refer to the concept of SA (such as *"the UI requires too much attention"* or *"the design displaying the flight parameters is confusing"*). For our work, the most relevant category of feedback is concerned with the details of the designs of the UIs. Here, the expert comments are a good resource for examining the problem space. Abstracting from our findings in this survey, we suggest that there is room for improvement regarding the UI design for cinematic drones regarding the following aspects:

- Provide customisable controls
- Minimise visual clutter by reducing the display of redundant information
- Increase visual clarity by displaying flight properties complementary (as in aircraft aviation) and altering the visual representation of UI items based on their (relative) importance

8.5.3 User Study 2: Expert Evaluation of Design Alternatives

Taking into account the insight of our analysis of status quo designs together with the results of the expert survey, we started to build prototypes that would address the issues that were revealed. As customisable controls are an already established concept especially in photo and video cameras (as introduced in Section 3.1.1) that also was already briefly explored in Section 8.2, we focused on the aspect of minimising visual clutter and increasing visual clarity. As the latter was mentioned in the context of a complementary display of primary flight parameters, we especially emphasised addressing this aspect with the different designs. To explore the alternatives, we implemented three designs as video-prototypes as detailed below and illustrated in Figures 8.34 and 8.36.

(a) Baseline *(b)* Alternative 1 *(c)* Alternative 2

Figure 8.34: In-the-Wild: Comparison of design alternatives

Baseline: This design can be used as a neutral control condition for user studies as in our case. Here, we fundamentally used the DJI Go application as a template regarding feature selection, representation and location (Figure 8.34a). To minimise visual clutter, we removed some information that was priorly displayed in the top bar and emphasised the display of primary flight parameters such as altitude or distance and placed UI items mainly at the edges of the screen. We opted for this approach as we found in Study 1 that it seems to be hard to retrieve important navigational information at times and also to keep the centre free from occluding elements as much as possible. To further increase visual clarity, we changed the battery status display to an icon in battery shape. We chose a set of pre-recorded drone scenes from prior shootings as a background video material. Please note, that the values presented, for example, regarding altitude change in synchronicity with the motion of the drone as seen in the background video material.

Alternative 1: T-model inspired: The second design was inspired by T-model adaptations of modern 'Glass Cockpit' layouts and representations (Figure 8.34b). These are designed with the goal in mind to increase SA and visual clarity. This approach would address our earlier finding that operators might be better supported with more complementary designs. We adapted the T-Model principle and hence placed a distance meter on the left and an altimeter on the right. Both encode the information visually (bar) and in written form (value). As a maximum height is set by law, the altimeter bar changes its colour to yellow above ¾ of the maximum height.

At the top, we displayed the battery status (along with further information such as GPS connection). Unlike aeroplanes, drones can rotate around their vertical axis. As contemporary systems lack representation for this motion, we included a ring display with a movable plate that visualises this intrinsic rotation. The speed and (translational) movement direction of the drone is encoded via a further element also at the centre of the screen. A moveable dot in the centre represents both aspects. The speed is represented by the distance of the dot from the centre with the maximum speed being represented by the greater circle mentioned above. The direction of the translational movement is mapped to the joystick's position on the controller and hence the position of the dot is moved depending on the operator's input.

Alternative 2: Less occluding UI: By design, the previous approach introduced more visual elements in the focus area. However, keeping this area as free from occluding elements as possible is one major concern we found consistently in existing devices. To further minimise the visual footprint in the central area, we developed another design alternative. Here, we kept the distance indicator from the Baseline design, the altimeter and the rotation indicator from the design of Alternative 1 and additionally integrated a variometer with the altimeter (Figure 8.34c). Further, the display of the battery status was moved to the bottom bar.

To evaluate the proposed designs, we recruited 7 experts (6 male, 1 female) for a user study with years of age ranging from 24 to 35 and a Mdn of 28. All participants reported to be right-handed, owned a drone and had between 6 months and 3 years experience of working in cinematography. Experience with drone operation varied between 10 to 50 hours (4 participants) and 50 to 100 hours (3 participants). Additional aviation experience was reported by 1 participant who also mentioned 280 flight hours in commercial aircrafts and who also held a commercial piloting licence.

In contrast to our initial survey, we arranged fixed appointments for this study. For each appointment, we would first introduce the study details to the participants and asked for their consent. Only once consent was given, we continued by handing our a demographic questionnaire. Next, each participant was exposed to one of the three video prototypes. For each prototype we compiled multiple scenes featuring different manoeuvrers. We randomised the sequence of the video prototypes as well as the various scenes based on Latin Square designs. After exposure to one of the designs, we collected data. First, we would start with some warm-up questions examining how accurately the participants were perceiving (or at least could recall) detailed aspects of the video prototype (see below).

- How would you characterise the flying behaviour of the drone?
- Which elements of the user interface were irritating or confusing?
- For how long will the drone be able to continue flying?
- How did the altitude change?
- How far away was the drone?

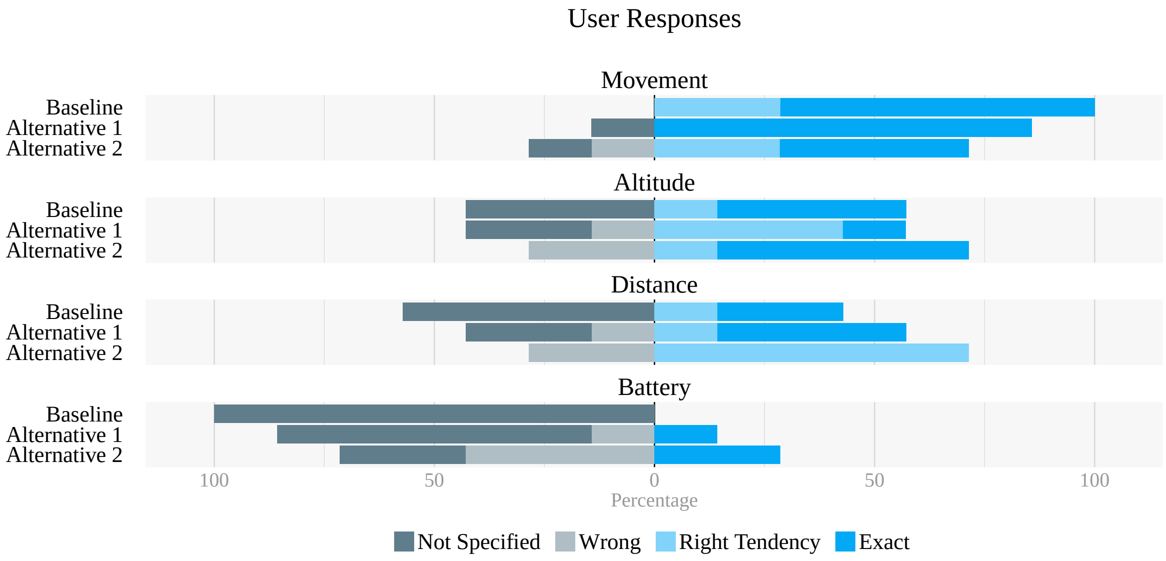


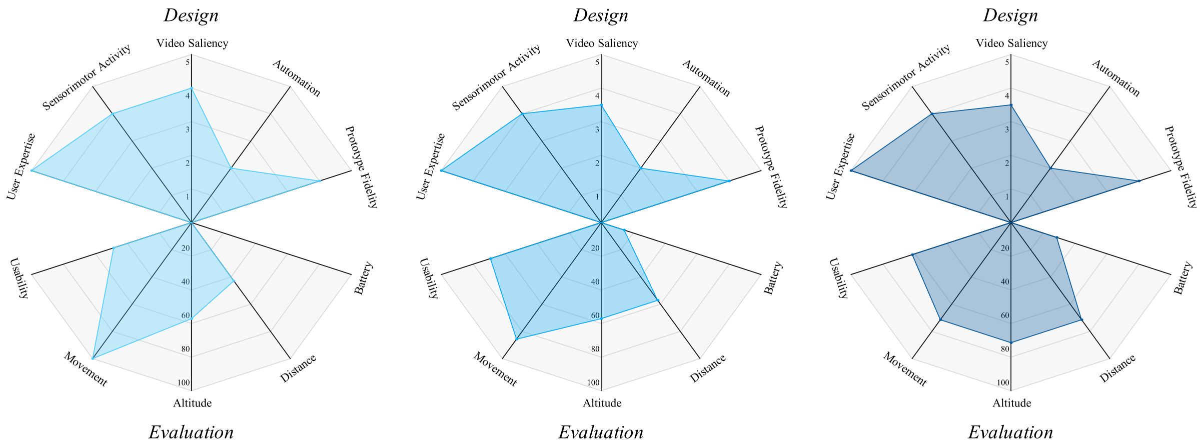
Figure 8.35: In-the-Wild: User Responses regarding perceptional questions

To further measure aspects of Usability, we asked the participants to fill in the SUS questionnaire Here, we omitted the last question as it did not apply to our designs and used an adjective scale as proposed by Bangor and colleagues [10]. After collecting data on all conditions, we continued with a semi-structured interview investigating general feedback, occurring problems or missing features.

8.5.4 Results and Discussion

The first set of questions was centred around perceptional aspects and ranged from general design issues to questions asking on particular navigational aspects. For the latter category, we conducted a more structured analysis. We grouped the answers into following classes: Movement, Altitude, Distance and Battery status. For each class, we evaluated how accurate the answer was. Here, we clustered the answers into four groups: Not Specified, Wrong, Right Tendency, Exact. The overall results for each design are illustrated in Figure 8.35 and discussed below.

Regarding the perception of the drone's Movement, our data indicate that most participants could recall the aspect most accurately for the baseline design overall. The maximum number of exact recalls, however, was found for the T-Model inspired design alternative. Focussing on recalling the primary flight aspects, Design Alternative 2 resulted in the largest number of exact reports followed by the Baseline design. Concerned with the Distance value, the maximum exact reports were found for Alternative 1 followed by the Baseline condition. However, the latter also led to the most Not Specified answers followed by Alternative 1. None were found for the second Alternative.



(a) Baseline *(b)* Alternative 1 *(c)* Alternative 2

Figure 8.36: In-the-Wild: Design variables and evaluation criteria

In terms of Battery status, Design Alternative 2 resulted in the most exact answers. However, overall all the tested designs performed weakly (with only Not Specified answers for the Baseline condition).

We analysed the collected data on usability applying Friedman's Test. We found no significant difference among the conditions ($\chi^2(2) = .857$, p $\leq$ .651) with a Mn value of 48.67 (Poor, Grade F, Acceptability: Not Acceptable) for the Baseline, 69.25 (OK, Grade D, Acceptability: Marginal High) for Alternative 1 and 61.73 (OK, Grade D, Acceptability: Marginal Low) for Alternative 2 (Figure 8.37).

From the semi-structured interviews with derived further insights. Concerned with the Baseline condition, participants positively mentioned the *"minimalistic design"* which included fundamentally *"all important information"* but also seemed to be *"difficult to understand"* as *"elements were not immediately recognisable"* (P5,P6). Especially *"Once you stare at the numbers, you can no longer perceive the remaining motion and the surroundings of the drone"* (P1).

Participants felt that Design Alternative 1 *"requires less time to understand it"* (P7) and provides the most clearly arranged representation of the important information as of the visual bar representation on the left and right edges of the screen. However, the location indicator with integrated variometer in the centre of the screen was also subject to criticism as its function was unclear to three participants, occluded the background camera stream or was in general sense referred to as unpleasant.

Regarding Design Alternative 2, four participants mentioned positively that it was the most straight-forward design and that information display was also representing the importance of the items. Also, the grouping together of the variometer and the altimeter was perceived as more beneficial compared to the first alternative design. However, we also found negative remarks on the

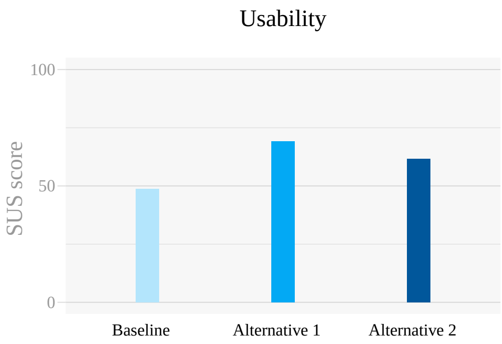


Figure 8.37: In the Wild: SUS scores

Figure 8.38: In-the-Wild: Final design iteration

display of the battery status which seemed to be hard to parse for some participants. Additionally, variations in the representations of the information on distance and altitude were considered as inconsistent and hence confusing by some participants.

Overall, we found mixed feedback for each design alternative. Some aspects in each design were regarded as promising and beneficial alternatives, however, none of our propositions would combine only the best options into one design. The details that were well-perceived were the visual representations for the distance meter and altimeter, a rotation indicator and graphical speed indicator, additional colour coding of information and the grouping of the variometer with the altimeter. However, not all designs featured a clear representation and UI layout, especially the display of the battery status was often criticised as too small or too hard to parse.

We additionally asked on features that participants found missing. Here, we found mentions of a speedometer, some participants explicitly asked for a display similarly to car cockpits. Additionally, some mentioned the addition of functions for automated camera motion directly in the UI.

8.5.5 Lessons Learned

To account for the mixed results of the evaluation, we decided to translate our findings into a further final design iteration. Here, we aimed at integrating the aspects that worked well into one design and at reshaping some details based on individual mentions that we further found in our interviews.

As the bars that represent, distance and altitude were well perceived in general, we integrated them in our final iteration (Figure 8.38). Additionally, participants found that colour-coding information was beneficial and that their initial design was inconsistent, we addressed both aspects in our final design. We chose the combination of altimeter and variometer as it was the preferred option suggested in our study. However, we decreased the size of the inclination indicator of the variometer to save screen space and to avoid drawing too much attention. We further reduced the opacity of the central rotation and speed indicator. We positioned the battery status display at the centre of the top bar and made it more prominent and colour-coded its status. Additionally, we provided a time estimate representing the flight time that is left depending on the battery status. Finally, we reduced the size and opacity of the map in the lower-left as some participants reported to be distracted by it.

8.6 Summary

In Section 8.1 we reported on a user study investigating our TrackLine approach to support interactions for camera motion in relation to off-screen objects. From our study data, we inferred that TrackLine performed better compared to a manual control option. In particular, for fast-moving objects, it could additionally outperform a contemporary assistance approach that was based on touch-to-track interactions. We also took away from the study that while the support by TrackLine increased performance, users found the freedom and expressive of manual control missing.

We especially addressed the latter aspect of the project reported on in Section 8.2. In a first user study, we explored and compared three design alternatives for controlling the basic translation and rotation of a camera. In detail, our data suggested that the Extended Software-Joystick might entail additional benefits compared to a traditional software-joystick layout such as more user freedom and potentially fewer errors. Overall, however, it also introduces more UI elements which further occlude an underlying camera stream. To address this shortcoming, we set out a second user study examining the effects of different Progressive Reduction strategies. Here, our results indicate that PR helps to reduce the visual footprint of a UI in general at low costs regarding additional workload or a diminished judgement of control. However, there is also a limitation on how far this can be reasonably applied. In the cases we studied, the major aspect factoring in is suspected to be the sizes of the hitboxes; or more precisely, the hitbox sizes as perceived or expected by the participants. Beyond this general finding regarding PR, we found no indication that the different reduction strategies we investigated had any (measurable) effect.

In both prior projects, we mainly focused on either assisted control or manual control in isolation. However, as laid out in Chapter 5, various good reasons suggest designing for a balance between assistance and manual control. In Section 8.3, we, therefore, focused on a project that would integrate both design approaches into one UI. Similar as in the second project, we first conducted an initial user study to explore the effects of different design alternatives. Here, we found that the Icon-based Menu was the preferred option as participants perceived it as the least occluding and less error-prone option. Again, we implemented PR to further minimise the visual footprint of the UI. To investigate whether a gradual reduction provides a particular benefit over static (non-reduced and reduced) approaches, we conducted a second user study. In our study, we found no indication for a (significant) difference between the gradual and the static reduction approach. However, a gradual reduction approach might be beneficial for other reasons such as long-term effects or user experience aspects that we did not measure in our study.

Also, other approaches contributing to a reduction of the visual footprint of a UI are imaginable. In Section 8.4, we followed the idea of outsourcing features that would require a display on the screen to mid-air gestures. To investigate, which types of cinematic shots might lend themselves to gesture interaction and particularly how to design them in detail. In a first elicitation study, we asked participants to come up with suggestions for navigating cinematic drones with mid-air gestures. Our study results indicated that when No Subject is being filmed Deictic gestures were preferred by most participants. For cinematic of Self filming or filming a Third Person, gestures that could be executed with One Hand were chosen most often. For filming a Third Person, mimicking the movement of the drone with one's hand was rated highest. To encode Self filming, first pointing at oneself and then using a One Hand gesture was the most often used approach. To examine, whether participants who were not involved in the elicitation process would rate the gestures similarly, we conducted a second user study. We found that surprisingly, users would rate the elicited gestures better than the participants who initially came up with it and were asked to rate their own ideas.

So far, the projects that we presented focused on UI designs that mainly used virtual implementations in 3D environments (or a Wizard-of-Oz approach). While this was helpful to quickly examine novel ideas and approaches, it incorporated the contemporary status quo only minimally. To address this shortcoming, we identified problems arising in current systems that are used in-the-wild via an expert survey and reported the results in Section 8.5. Our interviews suggested that customisable buttons to which pre-programmed functions could be allocated to are missing. Further, information is not optimally displayed as it is perceived as unnecessarily redundant, overly complicated, visually cluttered or designed without the importance of particular items in mind. We developed two design alternatives inspired by design patterns stemming from aviation. To examine the effects and to collect feedback on our propositions, we conducted an expert user study. We exposed the participants to video prototypes depicting use in different contexts incorporating actual cinematic drone footage. We found that overall, our T-model inspired design approaches were the most promising in principle, but also that they all had several shortcomings. Therefore, we integrated the propositions made by the experts and iterated our design in a final design cycle.

Insights

- In greater detail, we examined status quo designs and found room for improvement regarding various issues: support for off-screen items, visual occlusion, limited gesture sets, customisable controls and missing navigation features. To address the identified issues, we conducted five research projects proposing design alternatives and evaluating their effects based on our earlier work. While, for instance, our TrackLine approach could increase performance for situations where off-screen items were entering a scene, it also led to a decrease in expressiveness overall. Therefore, we further investigated combinations of axis and such content-based control designs. Here, especially our proposed Extended Software-Joystick provided further customizability and hence was well perceived by participants. To counteract visual occlusion, we examined various approaches. We investigated how reducing the visual representation of UI items might be beneficially integrated. We found, that minimising a set of visual variables is generally supportive. However, our data was not indicative of short-term benefits for Progressive Reduction in particular. Further, interviews with drone experts revealed that for use in-the-wild revisiting current designs with the goal of presenting more relevant information more prominently in mind (and vice versa) is desirable. Putting this proposition into design practice, we proposed a video prototype closely modelled after current systems and integrating missing navigation features. We also explored the general design space of mid-air gestures focusing on cinematic application in an elicitation study. The results indicated that a distinction between navigational and cinematic commands into deictic and iconic gestures seems to be in accordance with the mental models and their expressions by users.

III

Retrospective and Conclusion

EDWARD TELLER (PHYSICIST)

The science of today is the technology of tomorrow.

ELBERT HUBBARD (PHILOSOPHER)

One machine can do the work of fifty ordinary men. No machine can do the work of one extraordinary man.

9

Retrospective

What to expect?

- Recap of our insights and essential learned lessons
- Discussion of our findings taking on the perspectives introduced in Part I

What to take away?

- Inferences from the different perspectives regarding what our results might mean and which future consequences they might potentially bear

The Opening Remarks so far revealed (only) one major purpose of the insights placed at the end of each chapter which was to summarise the most important ideas and findings of an individual chapter and to implicitly mark its end. Thus, in a broader sense one could also refer to them as a supportive narrative device. As introduced in Section 3.1.3, narrative and stylistic devices in cinematography are often used for multiple purposes at once. The same way, we overloaded these brief insights at the end of each chapter in this thesis. In addition to their above-mentioned immediate purpose, we crafted them to produce a coherent text if put in subsequent order. Therefore, we want to present a retrospective summary of our work reported in this thesis by daisy-chaining the insights already presented for the individual chapters. Given that we were able to present our brief concluding insights per chapter in these brief summaries convincingly, an overarching and coherent retrospective summary of our overarching body of work should emerge.

Summary from Insights

An innate human need for fiction lets a great number of people gravitate towards media in general and historically more recently so towards cinematography. While this form makes substantial use of visuals, it characteristically further integrates various other types of media which together unfold over time. This integration of modalities makes it a fertile ground for engaging storytelling which seems to be one of the most primordial and powerful devices for propagating information and fiction across time, cultures and media and thus strongly contributes to satisfying this intrinsic need. Similar to other media, (storytelling in) cinematography is encompassed by a set of stylistic devices and production practices. As they are ideated and implemented behind the scenes, those are often not perceived consciously by most consumers. Nonetheless, they are vital to experts and enthusiasts that want to tell a story cinematically. Advancing the craft and bringing more options to the table, novel tools for these user groups emerged recently due to an integration of computing power.

Cinematography has a long tradition and the underlying technologies now matured in such a way that it entered its digital era on a larger scale; covering not only camera systems but more recently also motion control tools. Alongside this shift, however, also the traditional notion of distinct user groups became more and more challenged. Various forms of amateur, enthusiast, semi-professional and professional groups resulted as a consequence. Further, traditionally trained camera professionals might not necessarily be experts for newly emerging motion control tools. Also, in-the-wild, the potential of the implemented computing power is not yet fully exploited. Especially regarding the control of complex camera work on a higher level of abstraction. Well designed user interfaces might further enable users of all groups to access technologies such as computer vision and can hence assist them without requiring years of practice regarding their control.

Integrating experts and enthusiasts in research on computationally extended cinematic tools seems reasonable as they are very conscientious of why and how tools are applied in practice. For this, observations on location and personal appointments seem more beneficial than remote approaches. As experts often have busy schedules one needs to put in dedicated effort or cope with approximating methodologies. Interviewing them, we identified that camera operators face a high workload and a low tolerance for errors. Computational systems could assist by increasing the level of automation or the intelligence of systems. However, operators also want to be or, at least, feel in control and need to improvise. Thus, systems should also be designed incorporating adaptive strategies regarding their assistance. Additionally, in such novel systems, designers should stay consistent with existing hierarchies, user roles and task delegation practices. Acknowledging those patterns, one automated sub-system should only assist in a task that a dedicated operator would execute. Thus, for each of the traditional tasks, one sub-system would be required. Such a sub-system would further allow accounting for the insight above. In an adaptive system, remote operators would still be able to take over and improvise once a task is delegated.

Chapter 3 - Users

The research literature already describes various approaches of how assisted camera control could be implemented. It indicates that techniques for camera control in virtual environments can sometimes (but not always) be transferred to the physical world. Especially content-based techniques are interesting as they are capable of reducing the margin of error and require less training. Additionally, the approaches also translate to mobile devices that are already used (as for drone control). Under the hood, often computer vision is driving such techniques, and it is increasingly leveraged in various domains of cinematic production. For the implementation of such techniques into actual systems, overall adaptive automation seems a promising design approach for multiple reasons as it allows to integrate the benefits of both agents, to adapt to unforeseen environments and to incorporate human-centred aspects. However, designs incorporating the full triad of adaptive automation, human-centred aspects (such as feeling in control) and a computer-supported adaptive creative exploration are not free from pitfalls. Those, for instance, can arise due to higher levels of automation and abstraction in system control which affect the perception of when and how users feel in control. This perception can be affected at the level of the feeling or the judgement of control. Fundamentally, both seem to tap into different neural processes, and hence ambiguous situations can arise. Such situations bare the risk of leaving the user behind with a negative sense of not being in control.

Chapter 4 - Research

The degree to which users feel in control or not can potentially be explained by the concept of intentional binding or lack thereof. Potentially, it can push the envelope regarding explanation models presented so far in the great body of work concerned with the design and evaluation of levels of automation in system design and their effects on the experience of users. Fundamentally, it possibly can provide reasoning without too many gaps from empirical observations to the abstractions formed on the design of systems explaining trade-offs emerging therein up to the perceptions by users and the neural underpinnings embedded in them. As adaptive automation systems that support being in control and computer-supported adaptive creative exploration can take on a plethora of designs, it not necessarily obvious how these can or should be shaped. To be a bit more precise regarding the details, we derived a set of design variables based on insights we collected in our analysis of status quo technologies, our research on user requirements and our literature review. While, in general, such variables could be used to describe a more extensive set of tools for creative work, we focused on a particular subset, namely user expertise, motor activity, video saliency, automation and prototype fidelity in our work. Those are considered to be relevant to cinematic camera operation by active users. We suggested that those variables can each take on one of the five values that we also proposed. Thus, different designs can be characterised by their instantiation of each variable regarding these static design aspects. For illustrating their dynamic aspects, we suggested the use of the creative flow diagrams as proposed by Deterding and colleagues.

For translating design ideas into prototypes, we explored paper, paper/tablet hybrids, virtual 3D environments and physical setups. While we found paper to be feasible for laying out basic UI patterns, it revealed major shortcomings in a cinematic context. A hybrid approach using paper on top of a tablet helped, but its use should be limited to a particular set of moves (e.g. pans not slides). Virtual prototypes helped to overcome such limitations, and for UI tests in controlled environments, they can provide great benefit with a good trade-off between effort and outcome. However, they can hardly be used in-situ. Only a physical implementation, can address the latter issue. As their development is challenging and commercial products often lack open interfaces, we opted for a custom implementation that we shared openly encouraging reproduction. To complement our propositions regarding design variables, we derived a set of evaluation criteria suited for cinematic UIs, namely: workload, feeling of control, judgement of control, creativity support and performance. For each, we presented explicit and implicit measures. Integrating both types is recommended as a complementary use might help to address the shortcomings of either approach being used individually. We further implemented assistive tools intended for general use in HCI that we also share. We studied our approach in varying environments and could not establish effects on the measurements results due to the changing implementations. However, while explicit and implicit measures should be used complementary, one needs to be cautious concerning the sense of control. As the feeling and judgement of it might tap into different neural processes, the measured results might not lead to identical inferences.

To investigate the properties of our physical implementation, we took it to a field test. Together with a professional camera operator, it was used on five assigned shootings. The material captured at those recordings was later used in various exhibition films that were published online. This gave us confidence that the quality of our implementation is at least to a large enough degree comparable to professional tools. Therefore, possible bad ratings of UIs using it as a platform, must not be simply attributed to poor implementation. Most surprisingly, we found that the use of tools such as our platform, affects not only the recording procedures as it allows automating certain subprocesses, but also the people being recorded; especially untrained people tend to behave more naturally. We further looked into how different control styles and UI designs influence the quality of control, the perceived workload and the judgement of control. Our data suggests that continuous remote control approaches do not perform different than full manual control. This was surprising, as we expected full manual control to be rated highest at least regarding the judgement of control. We investigated the issue further and proposed that potentially an unawareness of the consequences might be factoring in as manual recordings showed a higher degree of shaky and jerky motion. Consequently, we suggested introducing a review phase that would give participants the chance to judge the quality of their results. We measured the effect that such a phase might have on the preferred level of control. Given our data, we found no influence. However, although it does not seem to be indispensable for controlled studies, we still recommend a review phase as we observed it to be a vital procedure on set.

In greater detail, we examined status quo designs and found room for improvement regarding various issues: support for off-screen items, visual occlusion, limited gesture sets, customisable controls and missing navigation features. To address the identified issues, we conducted five research projects proposing design alternatives and evaluating their effects based on our earlier work. While, for instance, our TrackLine approach could increase performance for situations where off-screen items were entering a scene, it also led to a decrease in expressiveness overall. Therefore, we further investigated combinations of axis and such content-based control designs. Here, especially our proposed Extended Software-Joystick provided further customisability and hence was well perceived by participants. To counteract visual occlusion, we examined various approaches. We investigated how reducing the visual representation of UI items might be beneficially integrated. We found, that minimising a set of visual variables is generally supportive. However, our data was not indicative of short-term benefits for Progressive Reduction in particular. Further, interviews with drone experts revealed that for use in-the-wild revisiting current designs with the goal of presenting more relevant information more prominently in mind (and vice versa) is desirable. Putting this proposition into design practice, we proposed a video prototype closely modelled after current systems and integrating missing navigation features. We also explored the general design space of mid-air gestures focusing on cinematic application in an elicitation study. The results indicated that a distinction between navigational and cinematic commands into deictic and iconic gestures seems to be in accordance with the mental models and their expressions by users.

After this recap, we want to elaborate on our interpretations and potential consequences of our findings taking on the perspectives of Technology, Users, Research and Design that we already took on in Part I: Perspectives and Fundamentals.

9.1 Technology

In Chapter 2, we outlined how the historical development of camera and camera motion systems progressed over time. We did this having the idea of revealing an abstract developmental trajectory in mind. However, as seen in the past, technological advancement is not always a straight-forward path without complications, hardships in adaptation or difficulties in achieving economic goals. Discussions on how the developments might advance and affect the landscape in the future are nonetheless interesting and worthwhile. We hope to be able to contribute to some aspects of this discussion. Here, most of our insights on how technology might progress in the future are informed by a combination of a (cautious) extrapolation from the observable past, our analysis of general advancing factors and statements from expert camera operators as found in our interviews. Combining these insights, we deduce a general expectation of a further increase regarding the use of digital systems for recordings and camera motion in the future. How this vague and general expectation might translate into reality however is more speculative. In the near future, we expect a continuation regarding aspects that are connected with advancing factors such as costs, connectivity, integration with CGI and CV and the design of UIs.

The question of how systems might be built in the near future still remains hard to answer. Extrapolating from the past while considering advancing factors and the developmental trajectories of other domains such as (semi-)automated cars, however, there are a few guesses that we would like to suggest. Our first proposition is that for expert and enthusiast systems alike, the implementation and use of sensors are likely to keep evolving. Here, we want to offer two more detailed suggestions in particular. First, beyond CV that is already commonly used to inform the steering behaviour, other types of sensors might be incorporated that provided greater resolution and allow for a more accurate modelling of the surroundings of a system as already implemented in (semi-)automated cars. Especially, for powerful systems such as industrial robots, such an advanced sensor-fusion approach might not only be reasonable in terms of increasing the quality of motion regarding cinematic properties but also for the safety of the staff as well as of the equipment. For instance, such additions might also help in other systems like drones to avoid collisions with birds, static objects or other drones during flight. Second, especially for use in expert environments, even closer integration with CGI or more generally speaking the post-processing might be a considerable future development. So far, the camera motion of physical tools is tracked and matched onto a virtual camera to integrate CGI material. In real-time scenarios where actors might also be tracked, integrating conditional statements such as ‘if X happens, move the camera along path Y’ as often used for virtual or game environments, might be considered. This might help to deal with the variance real-time human acting brings to the scene and to further streamline the production process especially for fast-paced scenes or complex choreographies.

While tools for physical camera motion such as drones and gimbals are likely to be further advanced and adapted by various user groups from experts to amateurs, other technologies might face greater hardships. For instance, Lytro, one of the few manufacturers for cinematic light field cameras, ceased operations over the course of working on the research projects that led up to this thesis (in detail by March 2018). As this surely marks a downturn for their technology, it does not necessarily indicate that it is a dead-end entirely. On the one hand, the company and its know-how were bought up by Google. How it might be used in the future is still unknown to the public at the moment. On the other hand, the core technology is still used in other domains like industrial applications, for example, by manufacturers such as Raytrix [367] to advance machine vision for applications as 3D microscopy. So, potentially, once economic requirements are met and products can be more easily mass-produced or integrated into commonly used devices, a revival of the technology in the context of cinematography is not ultimately to be ruled out.

While motion-controlled systems behind the camera are already a standard option for cinematographers, similar technology might also be used for different purposes in front of it. For example, Disney is exploring the use of humanoid robots as substitutes for stuntmen with their *“Stuntronics”* systems [325, 459]. For the purpose of car commercials, the agency The Mill, developed a car rig named Blackbird [380, 427] which serves as a foundation on top of which the car to be advertised is rendered in the post-processing. The rig features a wheelbase the length and width of which is adjustable so that it can serve as a surrogate for a broad variety of cars. It also records its surroundings with cameras and laser scanners for later integration in CGI or even VR applications such as the exploration of different models in a showroom. Increased connectivity linking the devices in front of and behind the camera might also render novel shot types (more easily) possible that are hard to record at the moment. For instance, for capturing car races, drones have already operated mimicking a style that is comparable to the ‘dramatic’ motion paths that a virtual camera in racing games would follow [479]. Greater connectivity and increased computational power might further fuel achieving complex shots that hardly possible given human operation skill alone, especially for situations that are less predictable than predefined racing courses.

Regarding the more distant future, we also encountered an interesting and somewhat provocative idea suggested by one of the interviewed experts. Assuming that light field cameras are available and that a reasonable fusion of the data even of multiple sources is possible, he suggested that consequently, one could build a wall or an array of light-field cameras, and posed the question *“Is physical camera motion in constrained contexts as in-studio recordings then necessary at all?”*. Is, alternatively, a future conceivable where the camera motion, even for physical recordings, is finally only a product of post-production decisions and subsequent rendering?

9.2 Users

Similarly, as we did in terms of technology, we learned a lot regarding the users from our interviews with experts and enthusiasts. Most fundamentally, we observed a shift in user groups due to the recent changes in the means of production. Formally, there was a clearer distinction between larger teams of experts cooperating on a set for the production of feature films and amateurs. From our investigations, we assume that the expert user group increased in absolute numbers over time, however, with the growth being mainly linked to the general growing number of productions. This group favours physical tools and controllers and is sometimes a bit sceptical regarding the introduction of novel technologies especially regarding camera motion. More recently, fuelled by greater availability of affordable tools in combination with self-distribution platforms and social media, a growing semi-professional enthusiast group formed. In contrast to experts, they often work in smaller teams and on smaller projects. While some underwent a traditional education and training in camera work, others are mainly self-taught. They tend towards being less sceptical of novel technology as it is a major factor enabling them. Similarly, a greater group of amateurs who also make use of semi-professionals tools is observable. Most fundamentally, they differ from the other groups in terms of their motives; they do not necessarily have professional motives or making profit in mind when creating content. Regarding their UI preferences, they are similar to the enthusiast group. Additionally, they also tend more towards filming themselves than traditional camera operators or enthusiasts. These categorisations are derived from our surveys and interviews, our observations on set and our own recording sessions with experts. We (conservatively) derived the classes and descriptions mentioned above from these data sources that are mainly to be regarded as qualitative findings. However, it is vital to note, that we can only present limited data regarding the actual absolute numbers of users making up each group. Such a quantitative statistic would surely be helpful to convincingly present the main point more substantiated by hard data. Future research and subsequently derived knowledge in that direction might not only be worthwhile to gather from a purely scientific point of view. Most profoundly, as sketched above, we estimate that there are consequences to be expected from this shift in the user base. The consequences are mainly driven by changes in technology as well as the surrounding (cinematic) landscape. While we hope to contribute to a better understanding of it with our work, we also acknowledge that we mainly provide preliminary results and hence further research is recommended. This is not only relevant for HCI research in general in order to increase knowledge and understanding, but also for manufacturers that aim at holding or increasing market shares. As far as we can tell, the consequences entail that for the various user groups different design options need to be considered. These seem mainly necessary as an according adaptation of the systems and UIs to the changing landscape and user population.

Similar to the previous section, we not only want to take a retrospective look at our insights on users but also speculate about their future. This seems to be a sensitive subject as we also found in our interviews. So far, we discussed the positive aspects of the increased use of novel technology on film sets extensively, however especially experts also see their jobs potentially at risk. How concerned

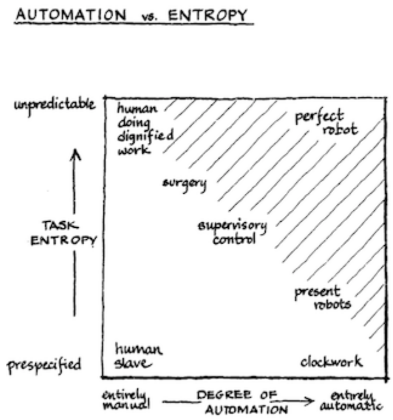


Figure 9.1: Retrospective: Sheridan's ideas on Automation vs Entropy (from [242])

experts are, differs quite a bit. While some experience general discomfort thinking about autonomous recordings in various areas, others are more relaxed. The first group mainly sees their everyday jobs at risk, such as recordings for image-films, advertisements or studio recordings. While these jobs are not necessarily the most wanted jobs for camera operators overall, they at least provide a sufficient foundation for earning a living. More prestigious jobs are surely welcome options, but they are also more scarce, competitive and irregularly available. The latter group thinks of the situation along the following lines. While surely more advanced technology challenges traditional jobs, at the same time increased demand for specialist operators who know how to set up the systems and how to deal with the challenges that arise on set might emerge. While it is surely hard to predict the future of camera operators, we want to add a further perspective. Already in 1987, Sheridan [242] proposed following graph illustrating the relationship of automation and entropy (Figure 9.1). Shortly summarised, it suggests that the more predictable a situation is the easier it is to automate the task. However, the more unpredictable a task the more you either need a *"perfect robot"* (which might be intensely difficult to build) or the more likely it is that the work that remains being done by humans is *"dignified work"*. This might further translate to cinematography along the following lines. While jobs might be increasingly challenged by automated tools, the jobs that require human operators are actually the jobs that operators want to work on. Some jobs that still are operated manually today might also be shifted towards jobs with *"supervisory control"* which might contribute to enabling operators to earn a living. So overall, the situation does not seem totally unpromising. However, uncertainty remains on how much of the *"dignified work"* jobs come into being in the future, how much supervisory jobs allow to earn a living with both aspects being highly dependent on the growth in movie productions overall. However, independent of such economic aspects, it seems also noteworthy to us, that full automation is hardly achieved in the near future. Human creativity is and will remain a significant aspect in cinematic production, consequently, we would argue for a real-life need of involving active users; especially when considering the parametrisation problem in the context of creative endeavours where it remains unclear what the parameters should be and whether parametrisation might even be desirable.

9.3 Research

In terms of generalisable research contributions, potentially most of our relevant work is concerned with evaluation approaches. Here, we first proposed a set of three central metrics with Workload, Sense of Control and Creativity Support that might be especially relevant for estimating cinematic UIs but that could also be applied to other semi-automated creative tools. Where applicable, we discussed and investigated the use of explicit as well as implicit methods for their evaluation.

For estimating Workload, a great body of work is already described in the literature. Here, explicit methods are well-established [108, 109]. Implicit methods such as DRTs are more recent developments [55]. For adaption to the domain of cinematic camera operation, we investigated their use in virtual UIs. We found that by using implementations resembling the original setup more closely, the results were more likely to lead to similar inferences when compared to the explicit measurements (Section 6.2.4). We also examined the use of overloading existing UI components which consequently led to designs that diverged to a greater degree from the original setup but kept the fundamental principle intact. For some study setups (as in the TrackLine study [287, 293] in particular), we used visual elements that would already be presented on the UI to prompt user action as the visual cues for the DRT and logged the response times. In summary, when comparing the implicit data with the explicit data that we also collected, we found that they were more likely to lead to different inferences. In detail, the workload derived from explicit data was more sensitive than that from the implicit data which was particularly bothering. Overall, the main reason for the latter approach was to investigate whether such implicit data collection during the interaction would allow for integrating an ‘invisible’ DRT. As the results were not necessarily convincing so far, it seems to us that while desirable such an approach needs to be crafted and investigated with more care and rigour.

For Sense of Control, we investigated the literature regarding explicit tools for data collection [304]. We found that a limited number of questionnaires is available and also used some of them in our work [79, 118, 123]. However, more generally speaking, we also found that extensive research on the (sub-)domains of control seems to be missing. Similar to the domains of workload as found in the TLX such further refinement would be desirable and help to better understand the phenomenon and the associated effects of and on UIs. Concerned with implicit measures, we particularly investigated the use of a rather recently developed methodology based on estimating intentional binding effects [121, 123, 156, 185]. In our studies, that focussed on an application on a particular usage domain (in contrast to the abstract experiments reported in the literature), we (also) found that explicit and implicit measures would point towards similar inferences; with implicit measures being more sensitive (which is to be expected). However, the literature also points out, that what is measured by both methods might be the result of different neural processes. In consequence, their interpretation needs to take into account this particular notion.

For Creativity Support, it is noteworthy that the literature on explicit measurements provides one detailed questionnaire with the CSI that is similar to the TLX, and hence is also comprised out of eight dimensions [47]. Regarding implicit measures, we also investigated the literature [303], suggesting that for some dimensions mentioned in the CSI at least some proxy measures could be considered. However, these might be subject to criticism as of their proxy nature and also the non-deterministic aspects that might come along with interpreting these measures. Consequently, one of the main sources of implicit evaluations might be the results of the creative process [111]. However, this approach might also be criticised as judging or interpreting the creative results might also be affected by subjective valuation. We found that implicit evaluation in the creative domain is (comparatively) less well-established, and therefore we did not include it in our work.

As mentioned above, in future work, a careful exploration of implicit cinematic DRTs is surely a worthwhile direction. Similarly, collecting more data on the use of implicit measures utilising the intentional binding technique in applied environments is desirable. As these recommendations seem to be rather obvious directions for future work, we instead want to expand more on an additional and more distinct direction below. In addition to the above-mentioned metrics or as a substitute for Creativity Support, one could also consider Situational Awareness to become a central aspect. Especially, as for its estimation, a great body of explicit as well as implicit methods already exists [80, 82, 84, 137, 257]. In our work, we figured that the creative aspects that operators need to focus on during steering systems need to somehow be represented in the evaluation and consequently opted for creativity support. However, alternatively, measuring the SA regarding the aesthetic 'situation' could also be considered. This approach, however, would first require a definition of the (sub-)aspects are to be included in what could be regarded as an aesthetic understanding of a scene and along with suitable measurement tools. Once such a definition might be found, the latter could be adopted from traditional (often aviation-centred) tools that are already described in the literature. It is particularly noteworthy that various methods for estimating SA are broadly discussed in the literature; with Endsley and colleagues [84] arguing in favour of implicit approaches, for instance. While their argumentation seems solid for established practices, it is in our eyes not self-evident that this automatically applies to what might be considered as SA in a creative environment.

A more fundamental question and even more so challenging research question might also arise from the mentioned observations. Having the notion in mind, that implicit and explicit evaluation methods might actually tap into similar but distinct neural processes (for Sense of Control), the questions is how (if at all) does this translate to the other metrics. In the construction of most of the methods of their measurement, the fundamental assumptions of what is measured are derived from statistically estimating correlations between the measurements. However, only because they are correlated does not necessarily mean that the measurements are also measuring data stemming for the same neural process (only that the results are similar enough to be correlating to a large enough degree).

9.4 Design

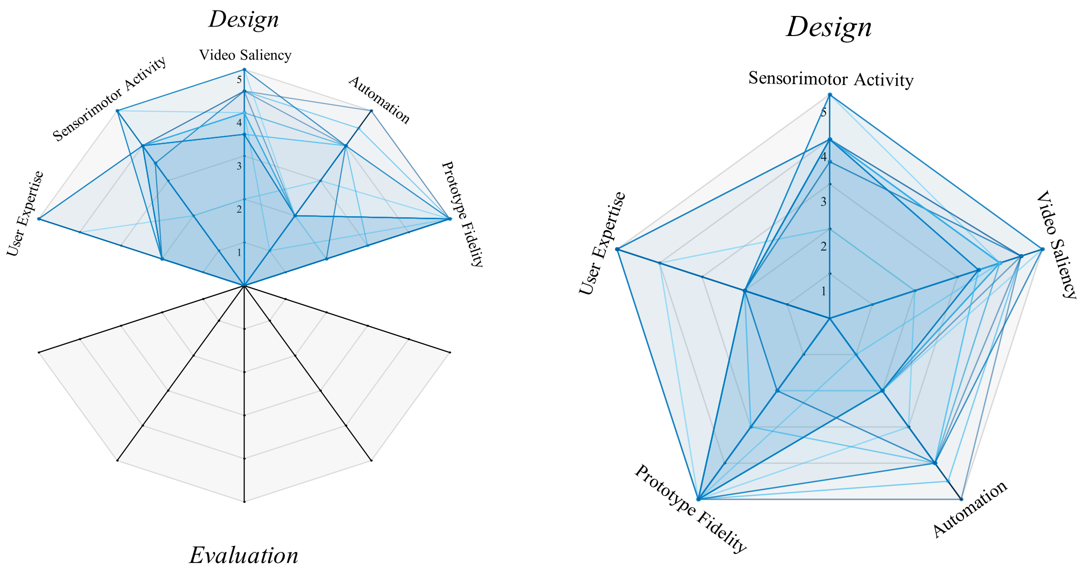
Looking back at our design process, one could characterise it, generally speaking, as a user-centred mixed-methods approach as it was already recommended by Mackay and Fayard [167]. It is particularly noteworthy that because of the mix in methods, we were able to learn lessons on various levels. First, we derived evidence-based insights regarding design decisions and alternatives as presented above in Chapters 7 and 8. Second, and perhaps more importantly in the long run, we also came to a better understanding of the design space as well as the underlying processes and practices. To avoid reiterating the former, we instead focus on discussing the latter aspect below.

As already mentioned when reporting on the literature in Chapter 4, designing for the full triad of adaptive automation, human-centred aspects (such as feeling in control) and a computer-supported adaptive creative exploration is not trivial. Ideally, one would prefer to maximise all aspects simultaneously, however, due to side-effects of individual design aspects or simply task-specific aspects this might not always be possible. Based on this premise, one important fundamental question might be: "Is a reconciliation of these aspects possible at all?" From our experiments, we would infer that while maximising all aspect at the same time is very difficult, promoting one such as control or introducing one such as assistance is possible without introducing sacrifices in other areas at least. A further follow-up question might consequently be "So, are there any good reasons for prioritising these aspects in a particular order?" While we do not want to put out a final answer on the issue, we would give the following recommendation based on our approach and given our particular user sample:

1. Designing for Control
2. Incorporating Assistance
3. Supporting Creative Exploration

We suggest this particular order with the main reason being that exerting control, on the one hand, seems to be at the very heart of interacting with (cinematic) UIs and on the other hand as creative exploration is implied by it to a large enough degree as long as proper control is provided and a skilled operator puts in the focus, effort and dedication. Yet, for particular use cases such as untrained user groups or specialist setups as the Spike robot, one could also argue for a different set of priorities. If considered seriously, this might even influence the hardware design of the systems. This, however, might be more relevant for a narrow and distinct group of specialists and hence for reasons of maintaining general applicability, we opted for our recommendation mentioned above.

Further, our thinking was fundamentally based on the design space that we conceptualised early on. It seems worthwhile to discuss some of its pros and cons. To recap, it was formed along the dimensions of user expertise (whom to design for), motor activity (manual operation during interactions),



(a) Plot based on our initial template *(b)* Pentagram plot of our design variables

Figure 9.2: Retrospective: Plots of all our instantiated design variables

video saliency (when the camera stream was displayed below a UI layer), the level of automation (or assistance) and the prototype fidelity. These dimensions were not arbitrary selected as they were informed via a bottom-up approach. Therefore, they mark central traits for designing and evaluating cinematic UIs in our opinion, but one could surely make the case that other dimensions are also worthwhile to include; especially when considering that a bottom-up approach as in our case is not necessarily deterministic. Consequently, one of the major questions arising is “What should one consider additionally?”. What we found missing to some extent was a degree of refinement or resolution that would allow us to better distinguish between design alternatives. Especially for the instantiated design variables, as in studies on design alternatives, the differences between the alternatives are hard to identify based on the data plots alone. On the one hand, this might be done by additionally adding dynamic design properties (creative flow) as proposed earlier or on the other hand by additional static properties more fine-tuned to addressing the issue at hand.

Regarding the individual dimensions, in hindsight, one of our greatest resources was the variance of our prototypes. It fundamentally enabled us, to run such a mixed-methods approach in the first place. Further, it allowed us to get into contact with experts in much more depth on sets; but also to conduct studies in controlled environments. The other design variables served us well as indicators guiding our design explorations. In Figure 9.2a, we plotted the instantiated design variables of all of our designs in one plot based on our initial template. For reasons of clarity one could alternatively plot the design variables on a pentagram (Figure 9.2b). One could summarise that except for designs featuring low values of user expertise, motor activity and video saliency, we extensively explored the resulting design variables.

MARSHALL MCLUHAN (SOCIOLOGIST)

It is the framework which changes with each new technology and not just the picture within the frame.

EVGENY MOROZOV (WRITER)

Technology changes all the time; human nature, hardly ever.

10

Conclusion

What to expect?

- A summary of our limitations and contributions
- Comments regarding future work
- Closing remarks and a broader outlook on the potential future of cinematography taking on a technology-centred perspective

What to take away?

- The key findings of our work, their scope of validity and a more far-reaching speculation on the future of digital cinematic production techniques

10.1 Limitations

> *"The idea is to try to give all the information to help others to judge the value of your contribution; not just the information that leads to judgement in one particular direction or another."*
>
> (Richard Feynman, Physicist)

Following this idea as put by Feynman, we want to discuss the limitations that also accompany our work below. Although we followed established methodologies and practices to come to our inferences, the choices we opted for are per se also not free from shortcomings.

User Perspective: Non-deterministic approach

Limitations regarding the generalisability of our approach may be found, most fundamentally, in the non-deterministic nature of the Grounded Theory-based approach that we applied in our initial work. We used observations and interviews to extract key features from investigating the work of camera operators especially in their relationship to assisting tools. Although we used a methodology that aimed at increasing the validity of the results by requiring multiple coders to come to a common interpretation, yet a different set of interviewees, might come up with a different set of answers and a different set of coders might come to different inferences. On the one hand, such an approach is commonly accepted for the exploration of novel fields. On the other, it also strongly influences the work that is later built on top of its findings as in our case the derived design variables or the evaluation metrics. In that regard, our approach and the insights that we got by applying it, are consequently limited as they mainly reflect the opinions of the experts that we interviewed or put more accurately, our interpretations of their answers majorly influenced what we focused on in our later work.

Prototyping Toolkit: Software integration of slider and gimbal

Concerned with the hardware, our slider and gimbal setup can be connected via our custom-made damper. However, regarding the software, a great amount of work is still needed to seamlessly integrate both systems especially as they run on different firmware. Ideally, the software to be developed should support the steering of smooth compound movements by one (and/or multiple) controller(s) along with (semi-)automated features. As this undertaking is quite intensive in terms of programming effort and the relation to its scientific contribution is not necessarily easy to predict, we opted for prioritising other projects. Given their often self-contained nature, these required less implementation work and consequently allowed us to prioritise studying the user-centred issues at hand more thoroughly. However, when looking at the prototyping toolkit with the eyes of a practitioner, implementing the capabilities for compound movements that integrate translation, rotation and image stabilisation is highly desirable. Further, this wish-list might also include features and other real-life requirements as detailed further below that we did not particularly focus on in our work yet.

Evaluation Framework: Implicit measures for real-life use

Regarding evaluation practices, we already discussed above that our set of metrics is derived from our initial interviews and hence one could also consider others given different initial results. Further, we also already discussed earlier (Section 9.3) that instead of focusing on Creativity Support, one could also consider using measures from investigating Situational Awareness adapted to a creative or as in our case a cinematic use case.

Additionally, we believe that for use in real-life, the proposed methodologies for measuring implicitly might not necessarily provide the desired stability and robustness for seamless integration into the production process yet. To gather data implicitly on, for instance, workload or sense of control, it seems to be crucial to follow established practices quite precisely and thoroughly in our estimation. However, in a production context, this might not always be possible due to timely and budgetary constraints. Therefore, their reliance on controlled environments might be somewhat limiting for putting them to good use in the field; at least in their current form.

Design Alternatives: Study limitations

In terms of methodological concerns limiting the generalisability of our study results, it is necessary to note that we mainly studied groups with a limited number of participants. Smaller sample sizes are not necessarily always problematic as pointed out by Nielsen in his work on discount usability engineering [193]. However, especially for rather recent developments such as the work on intentional binding effects that try to estimate psychological effects, larger sample sizes are desirable. Rigour in terms of participant sampling in psychological research is crucial and often requires larger samples for valid outcomes. These are often a consequence of a priori power analyses that take into account a previously determined effect size. Especially small effect sizes make larger samples necessary for results to show up in a later statistical analysis of the effect (if they are there at all). Further, this rigour also includes recruiting a representative sample regarding the research question to allow for a generalisable interpretation of the resulting data. In our case, we often either sampled a small number of experts as sampling this population is in itself already a somewhat hard endeavour, or we used medium-sized groups that mainly consisted of a student population that is accustomed to digital technology and used to being exposed to novel designs. These sample populations mark somewhat of a bias when compared to an ideal representative sample with large sample sizes. Further, in terms of study objects, we often used virtual implementations to conduct our studies. These allowed us to iterate more rapidly while also being able to log user data automatically. However, for reliable inferences on how our initial findings might generalise to a context without controlled environments and how the tested concepts might perform on top of established designs, separate studies under field conditions are still necessary in our estimation.

Design Alternatives: Integration of the learned lessons

As pointed out in Section 9.4, we explored a large portion of the to us relevant combinations of the design variables that we proposed earlier (Section 5.1.3). For each project that we described, we also emphasised the lessons when experimenting with a new design alternative. For some projects, we were able to include previous lessons that we learned into the follow-up designs. So far, our approach of mainly individual projects allowed us to isolate the studied variables as well as excluding (unwanted) interfering variables. However, although we tried, especially in our study described in Section 8.3, this approach consequently did not always allow us to test systems that would include all of our proposed design alternatives at once, as one or more features might have been left out such as gesture control for some projects. So in consequence, an evaluation of a system incorporating the explored design alternatives and our learned lessons all together would be desirable. While such an overarching implementation could be more easily implemented virtually, its implementation in a physical setup that could be applied our realistic circumstances would be much more beneficial as pointed out in the previous section. However, the latter would require a physical setup that integrates our slider and gimbal as mentioned above. Acquiring the necessary resources for such an intense effort, however, was out of our scope so far.

Design Alternatives: Integration of further real-life requirements

Finally, in regard to our chosen design alternatives, our approaches are additionally limited in the sense that they exclude some existing real-life use cases and practices that we observed to be central during a production. For instance, adjusting the position of the focal plane should also be considered as part of an ideal system for use on set as indicated earlier above. Most likely, as observed by us in practice already, a holistic system for (remote and/or assisted) control of camera properties, orientation and the focus should ideally also integrate more elaborate task-delegation models as well as multi-user support for assisted manual control and/or take-overs. Picking up on our remarks on an ideal system mentioned above once more, such a system should be built on top of an advanced version of our physical platform, integrate previous lessons learned in one system and further include missing features as sketched above. However, such a system additionally might not only be considered to support the production process but also provide means to prepare the steps following it. Here, we just scratched the surface (Section 7.3) of offering capabilities for supporting the post-production process already on set such as features for screening the results and potentially exploring early steps of the editing process. In particular, operators might already pre-select certain scenes on the set, work on a rough first cut and/or explore the combination of their visuals with a musical score that might be used in combination in order to check whether their rhythms match.

As a further real-life requirement, one might also consider robustness under varying (unforeseen) conditions. Here, gathering long-term experiences could be especially relevant. However, our projects show additional limitations as they only study short-term use of our proposed design alternatives.

10.2 Contributions

In our estimation, this thesis provides contributions that can be attributed to three major areas: ethnographic studies, prototyping and evaluation methodologies and UI design studies. *For each contribution area, we will summarise our findings and emphasise their (potential) value* for the HCI research community, especially when studying UIs in cinematic contexts.

Ethnographic Studies

We used online surveys, expert interviews, and contextual inquiries to investigate experts and their practice. As especially ethnographic studies regarding recently developed cinematic MoCo tools from an HCI perspective are hardly found in the literature, our findings contribute to a deeper fundamental understanding of the user population applying such tools. Most prominently, we distinguished between environmental and design factors that seem to play important roles in the work of camera operators. In detail, we classified preparation, low error tolerance, constraints and context as environmental factors while acknowledging that this list might not be extensive. As these areas are less likely addressable by HCI efforts, we instead focussed on the area of design factors, especially in terms of system and UI design features. Here, we found that designs that afford (assisting with the) high workload, (satisfying) the need for improvisation and (providing the sense) of being in control seem to be vital to operators. Further, we found that given environmental factors such as low error tolerance, experts in camera operation might be more hesitant towards technological novelty. So, most likely, design prototypes need to provide a large enough degree of sophistication. Alternatively, researchers might consider focusing on user groups that are more open to novel tools such as enthusiasts. Regarding user groups, our investigations suggest that one could generally classify users as novices, amateurs, enthusiasts and (general or specialised) operators. For each of the classes, we provide proto-personas such as the Intuitive Artist, the Exploring Student, the Settled Creative, the Settled Traditionalist or the Experienced Teacher. We further indicated that the user journey for each might rather be characterised as a walk-up-and-use scenario or an expert-journey while acknowledging that even experts face limitations in certain areas for which again supporting a walk-and-use scenario might be reasonable.

To manage the workload associated with handling the complexity on set, we found that especially in larger productions, task-delegation is often applied. In the camera motion department, the camera operator therefore often only controls the orientation of the camera while a grip takes care of its translation and a focus puller adjusts the focus continually. In order to synchronise and to plan ahead, the crew members of the camera department, often refer to previously recorded or live-stream camera material using it as a (shared) frame of reference supporting their discussions. In contrast, in smaller productions, one operator tries to take on multiple roles at once and where possible makes use of assistance system that would allow delegating one of the above-mentioned roles to a (sub-)system.

We also provide basic information on what user roles are usually found on (larger) sets (director of photography, camera operator, camera motion operator, assistant, digital image technician, data assistant and colourist), which shot types involving camera motion are usually applied (pan, tilt, dolly, dolly zoom, tracking, stabilised, crane and sequence) and for which narrative motivation a shot type might be picked (orientation, pacing, inflection, focalisation, reflexive or abstract). However, besides the narrative functions, also additional layers of motivation might be considered such as compositional aspects. Overall, we additionally found that camera experts are not necessarily an easily accessible population (with sometimes low return rates). Consequently, either time-intensive personal appointments or coping with approximating methodologies is required by researchers.

Prototyping Toolkit and Evaluation Framework

In prototyping for cinematic UIs a trade-off seems apparent between the interests of researchers and operators. As above mentioned, operators usually prefer more sophisticated prototypes which often require greater implementation efforts. On the contrary, researchers often prefer approximating rapid prototypes that aim at minimising implementation efforts while still allowing to gather the necessary data. Our prototyping approaches and toolkit provide researchers with a range of options (paper, mockup, virtual, physical, video) that allow adjusting the prototype fidelity to the individual requirements of a given research project. Especially our most sophisticated option, the physical platform, is provided with the goal of low-cost replication in mind in order to support the exploration of various designs with different levels of automation. This platform already showed to be stable for professional use [122]. Further, by helping to overcome the limitation of existing commercial systems that do hardly support connecting custom UIs, it might foster the exploration of new designs in cooperation with users, even on set. Additionally, even use outside the domain of cinematography has been documented as in enabling motorised eye tracking [142]. In summary, our efforts hopefully help to resolve this trade-off by providing a multitude of options, even sophisticated ones, so that the emphasis of further research efforts can be concentrated on studying novel designs.

Similarly, also our evaluation framework aims at supporting research on cinematic UIs. On a general level, one could say that it might help foremost by providing orientation; especially as the literature on evaluating cinematic UIs is rather diverse and lacks established and/or comparable approaches. In contrast to the variation in approaches as found in the literature, measures taken as emphasised by our framework aim at enabling comparisons, potentially even meta-analyses over a longer period. Our fundamental approach of taking multiple measures follows existing recommendations [167, 111] and is, in its core metrics (workload, (quality and sense of) control and creativity support), informed by prior user research. Thus, in detail one could say that our approach is particularly tailored towards evaluating cinematic UIs by balancing performance, experience and creative aspects, yet use in other HCIs domains is emphasised, especially when only applying individual components such as estimating sense of control (explicitly and/or implicitly which we tested in physical, virtual and VR settings).

Controlled Studies Exploring Design Alternatives

We used the insights from our ethnographic studies together with our prototyping toolkit and evaluation framework to test design alternatives of various projects. As our ethnographic studies indicated, there are several opportunities for better designs. For instance, assistance systems might help operators to deal with the workload they are handling. While a multitude of such systems already is implemented in existing devices, they are often only considering steering the motion of a MoCo tool without regard to the content that it is filming. In contrast to such approaches, we tested a design alternative with TrackLine that implements a content-based approach. Evaluating our approach, we found that it helps to improve performance, but also that it lacks expressiveness when used only by itself. On a more general level, this might be indicative of two things: first, that integration with manual control options is to be recommended, and second, that further exploration of the many others forms of content-based assistance systems might be a worthwhile direction for further research.

In further projects, we foremost followed our first recommendation and investigated the effects and side-effects of integrating different control paradigms into a single system. The first issue one might encounter in doing so is that with more features, more UI elements become necessary which leads to more unwanted occlusion of an underlying camera-stream as often found in content-based approaches. To resolve this trade-off between assistance and occlusion (to a degree), we examined a progressive UI reduction strategy that minimises the appearance of the UI elements and thus the occurring occlusion. This strategy might also be especially useful for further reasons. It might support mediating between a (potentially) self-explanatory UI when fully displayed and a less self-explanatory, but minimal UI when reduced. Similarly, it might also support the learning process by mediating between the recalling of visual elements when using the UI initially and the use of muscle memory that requires fewer visual cues in later reduction stages. Therefore, in conclusion, one might conceptualise it, more generally, even as a mediator between a walk-up-and-use user journey at initial stages and an expert-journey when minimised in later stages. Given the studies that we conducted, we came to the inferences that while occlusion is minimised, negative effects become observable beyond a certain level of reduction. Beyond this point, further reduction is not recommended. This is in line with prior recommendations on minimal sizes of UI elements.

So far, our inferences and insights mainly emerged from studying design alternatives in isolation. However, we initially found that integrating assistance and manual control is crucial. Therefore, we considered further studies on designs that focus more on this interplay as necessary. In our investigations on designs that combine axis and content-based control, we also applied progressive reduction following our prior propositions. Evaluating the resulting prototype which brings together all the above-mentioned design aspects, we found that a gradual reduction was not particularly necessary in our case. However, we also only looked at short-term effects and a limited number of tasks. Overall, the emerging take away might be akin to the proposition that if only a minimal

UI is the central design goal than starting with a reduced design is possible, however given the priorly mentioned user journeys and the mediating role that a gradual reduction approach can play in it, also such experiential aspects need to be taken into consideration when choosing a design approach.

Our ethnographic studies additionally suggested that providing users with a sense of control is important. For increasing a sense of control and/or agency, prior research points towards the idea of emphasising physical activity in interacting with systems (due to underlying psychological effects such as intentional binding). For the domain of cinematic UIs, this might be translated into systems by integrating mid-air gesture control, for instance. This control approach additionally interlocks with our initial goal of reducing the visual footprint of UIs, as it allows outsourcing certain features which in turn do not require to be displayed. Therefore, we investigated it in more detail [123]. Given these study results, however, this has to be called into question; at least for the study conditions that we investigated as we particularly studied continuous control. In this condition, a traditional software-joystick was observed to perform better. However, we did not study gestures that would trigger assistance functions. For this condition study results may vary. At the moment, however, feature sets as supported by current commercial systems are very limited and their (potential) benefits thus harder to study. To enable and support further studying on the subject matter, we, therefore, investigated how systems featuring gesture control for triggering higher level content-based control in cinematic UIs should be shaped by conducting an elicitation study. Here, we mainly learned that distinguishing between strictly navigational commands and content-based commands seems reasonable and that these command types might be mapped onto different classes of gestures. Here, we suggest that navigational commands are mapped onto deictic gestures and content-based commands to iconic or emblematic gestures (depending on a further classification by context such as the referent of a gesture). Additionally, we also found that participants were more self-critical than expected as they were observed to rate the gestures they came up with lower than a neutral control group.

To account for the fact that we mainly examined design alternatives in virtual environments and controlled experiments, we also conducted a further project with experts in more real-life circumstances. First, we identified problems intrinsic to contemporary systems: customisable buttons are missing and the visual layout and design used for representing information and interactive elements are often cluttered not representing the importance of certain features. Addressing these shortcomings, we developed design alternatives in the form of video prototypes. We exposed the experts to them and collected their feedback. One of our designs seemed to be the most promising. It was based on an approach that is commonly used in commercial aviation UIs, the T-model. In detail, we adapted its principle by placing a distance meter to the left and an altimeter to the right and a speed and direction indicator in the centre. However, also shortcomings of the design were revealed through our evaluation which we improved upon in a further design cycle. The result is intended to support in informing further iterations of status quo in-the-wild designs.

Conclusion

In conclusion, we would like to suggest that our work offers two major areas of insights and contributions to the HCI research community with the first being an increased understanding. A deeper understanding as emergent from our work might entail the domains of understanding (expert) camera operators and usage contexts of MoCo tools, understanding the design space (and/or variables) regarding recent (design and research) opportunities as well as understanding approaches to their user-centred evaluation. For the latter, we especially provide insights into a more profound understanding of the phenomenon of control on a perceptual level. This addresses, in particular, one of the major shortcomings that we identified. Despite being a core part that is ubiquitously used and immanent to nearly any HCI domain in one shape or another, a profound and deep (psychological and/or neurological) understanding of sense of control (and/or agency) and of the effects that various designs have on it is still limited. Acknowledging that the current state of research is in itself also not without limitations, we emphasised the (potentially) explanatory link between neuroscientific fundamental research (based on the model of intentional binding) and fundamental as well as applied research on computing systems that are concerned with automated systems and/or UIs as referred to in the literature and as described in this thesis.

Second, we offer insights regarding the design of UIs for cinematic MoCo tools for camera work. Here, we transferred the above-mentioned fundamental principles of estimating sense of control to an applied context that includes systems and studies that focus on exploring our proposed design variables. What we take away from our basic explorations and controlled user studies might be briefly summarized as follows. To increase control in terms of quality of control it might be considered worthwhile to integrate assistance systems. These should lay out their basic designs based on the principle of adaptive automation while respecting the task separation practices of existing user roles. In order to better support a proper feeling and/or judgement of control, UI designs should be crafted with care. Here, we suggest that integrating various control paradigms into one system might be generally supportive. On a technical level, this integration might be enabled by applying content-based approaches as they allow to seamlessly combine an underlying camera stream with UI elements displayed on top. This paradigm is not only already established (in other domains) in the research literature but is also found in-the-wild. It supports interacting with scene-specific elements that are found in a camera stream (such as actors or various inanimate objects) to coordinate the movement of the camera in relation to (compositional or even context-dependent) properties of these elements. This linking of the camera motion to the content of a scene is, for instance, applied in tracking shots and is, generally speaking, in accordance with the mental models of (expert) users. The above-mentioned integration of various control styles is also important as providing manual control options needs to be considered a further central design requirement. While, in general, assistance functions might be considered as distinct from manual controls, also a (more) fluid integration can be achieved. Here, revisiting or extending existing designs might be a worthwhile option as showcased by our extended software-joystick, for instance. Given such a multitude of control options recommended being supported, further emphasis should be put on minimising the visual footprint of the resulting UIs. To achieve this goal, several supportive strategies might be applied: manual control elements might (initially and for the most part) transition to an off-screen location, (progressive) disclosure and/or reduction of UI elements might be used to de-clutter the UI or a (limited) number of features might be outsourced to mid-air gestures. To further foster a sense of control, also rethinking established designs should be considered. In rethinking especially the visual representations, it might not only be beneficial to consider minimising the visual footprint but also to adapt (existing) layouts and designs (in parts) to the relative importance of individual bits of information that are displayed in a UI. This piece of recommendation might be considered to be in opposition to displaying all information in the same way to emphasise consistency which is generally (and for good reasons) recommended. However, in the particular case of expert systems for cinematic camera work, a limited number of vital and central bits of information that become displayed in a way that somewhat breaks the general convention of consistent display of UI elements might actually be beneficial for the overall (expert) user experience. We also emphasise that these need to be selected and designed with care and caution. Consequently, in designing for control in camera operation, designers should strive for a balance between assisting to increase performance (where applicable), supporting creative exploration by providing expressive tools and providing fallback solutions to minimise the risk of leaving users behind with a feeling of being out of control while also displaying the camera stream as large and as unoccluded as possible.

10.3 Future Work

> *"Perhaps one day we will have machines that can cope with approximate task descriptions, but in the meantime, we have to be very prissy about how we tell computers to do things."*
>
> (Richard Feynman, Physicist)

While our propositions might contribute to getting a step closer towards systems that allow for approximate input, they still require users to be detailed about how they convey their intentions to systems; although users might be able to be less detailed when applying our propositions compared to existing solutions. Nonetheless, further necessary steps advancing this line of research are imaginable and could potentially contribute to getting even closer towards a future of cinematic systems that might be able to reasonably interpret (more) approximate input. Our ideas on the particular subject matter of future work with regard to designing and researching cinematic UIs entail three major areas - interactions that focus on movements, interactions that focus on motivations as well as efforts advancing the evaluation framework - and are detailed below.

Interactions for Movements (Content-based)

In our TrackLine study, we tested a design that, generally speaking, aims at providing a UI that helps to align (existing) movement-based techniques with scene-specific aspects (mostly in terms of composition). In reference to the famous *"Put-That-There"* interface of Bolt [17], one might also conceptualise it as a 'Keep-That-There' approach. However, we only built a system that would focus on tracking shots. As detailed in the taxonomy in Section 2.2.1, many other shot types that focus apply different basic movements might be used as a foundation to implement such content-based interactions. Exploring the resulting design space that unfolds by investigating (all the) designs that allow controlling camera motion in terms of content-based moves for each of the traditionally used shot types individually, might be a resourceful approach for future work. Additionally, to us, it seems to be especially necessary to finally integrate those (multiple individual) solutions into one coherent set of interactions given the insights gathered based on a design and evaluation process as sketched out above. The resulting set of interactions might further benefit from aiming at a minimal and consistent design. In particular, reoccurring steps in the interactions should be identified and implemented using a shared set of common underlying design patterns. Further, given its inherent support of a multitude of shot types, such a UI could provide the expressiveness that such a tool for creative work requires.

Interactions for Motivations

Going beyond the exploration of a (rich and expressive) set of 'Keep-That-There' designs that focus mainly on composition, also exploring more high-level approaches should be considered as a worthwhile direction for future work. Similar to the above-mentioned design and evaluation approach, an exploration based on a given set of options might be a reasonable start. Here, an informing foundation might be found in the taxonomy of narrative functions presented in Section 3.1.3 (based on the work of Nielsen [194]). While there are enumerable ways imaginable that integrate this basic idea into UIs, we want to offer some suggestions. In general, we emphasise that designs should only be presented upon user request or based on user input that informs how much assistance they want. The reason being for this suggestion is to accommodate the preferences of experts and creatives that already have a particular plan or vision in mind. Also, creatives might want to explore the options and their intuitions themselves first, before being confronted with other ideas. However, when needed, supportive systems might provide such additional propositions. When requested, those should be grouped by narrative function. Once a particular narrative function is selected, a further set of (shot and motion) options translating the function into actual movement might be presented in more detail; potentially along with a limited number of adjustable settings. Following this selection, the system might subsequently offer a preview, allow for further adjustments and/or let the user return to the overall set of options. Additionally, one might consider the exploration of approaches that let users stack narrative functions and/or arrange them on a timeline to organise a sequence of multiple functions (or shots).

Advancing the Evaluation Framework

As indicated by our limitations section presented earlier, our proposed evaluation framework might also benefit from further research efforts. Referencing the introductory quote for this chapter by Morozov, it might also provide a long-term benefit, as it allows to estimate the properties of tools that are subject to technological change in ways that seem to be of vital importance independent of the applied technology. Here, for us, two major lines of future work seem emergent from our work so far. As a first set of options, one might consider advancing the framework in terms of adapting Situational Awareness measures and further in terms of stability and robustness in order to enable implicit measurements on set during shootings. For example, SA measures that are adapted to camera operation might include estimating (the awareness of) compositional features. Those might entail whether image corners are occluded, the horizon is levelled or how the figure-to-ground relationship unfolds. Besides compositional also storytelling aspects might be used such as the matching the position and/or size of the characters in the scene to their relevance for the story (however those might be hard to generalise). Regarding stability and robustness, further controlled studies with larger samples might be a reasonable choice. Such studies might focus on investigating the relationship with explicit measures more closely and examine how the design of the implicit measurement apparatus needs to be changed for (stable) use on set.

Overall, future work in the above-mentioned directions can be understood as advancements within the cinematic context. As a second set of options also application outside this context is to be considered. Here, further adaption to other domains is most likely beneficial in our estimation. In detail, a multitude of contemporary calls for paper or conference workshops aims at investigating the overlap of HCI and (explainable) AI. Here, the research interest often focuses on the effects on SOA in particular. In short, one might infer from the findings gathered so far that, given the recent advances in AI (that now allow for a closer interactions cycles with users), similar findings as to the ironies of automation (see Bainbridge [8]) seem to become emergent. While in detail the, let's say, ironies of AI systems might diverge from the effects of (over-)automation, both conceptualisations share some common underlying similarities: for instance, (both types of) systems aiming at providing a reduction in workload, but in the end, lead to additional workload due to supervision, error prevention or management, (human) interpretation of the results and/or maintenance. Further, at this particular overlap, providing a large enough degree of SOA becomes a necessary factor to make the systems usable (and potentially enjoyable) by people. Consequently, measuring the effects of design alternatives is of central interest. However, given the broad application of AI systems, it becomes increasingly apparent that a deep(er) understanding of the mechanics of the human SOA so far is limited and needs further detailed investigation. While our propositions might provide further initial steps toward this goal, also further adaptations to individual application contexts and systems are most likely necessary in such future work. Given the simple fundamental stimulus-response model that the measuring of SOA is based upon, this is most likely feasible in our estimation.

10.4 Closing Remarks

So far we made sure to only extrapolate cautiously from the developmental trajectory of the past into the (near) future and only on the basis of our empirical findings. To mirror our efforts of Chapter 2, where we took a deeper look into the more distant (technological) past of cinematic systems, *this section of Closing Remarks is dedicated to a more speculative long-term extrapolation that focuses on the development of technology and its potential effects on the future of (digital) cinema*. To avoid becoming overly speculative, we would like to fundamentally suggest that however cinematic production and consumption might look like in the future, it might strongly depend on which (technological) answers will be found regarding the following three questions (and potentially even more). Therefore, we will present our (potentially arbitrary) ideas along with the details regarding these three guiding questions.

How Physical is Future Cinematic Production?

Currently, there exists a broad variety of promising, not yet fully developed and still mainly isolated digital technologies that bring novel options to the table of cinematic production. So, for instance, CGIs of characters look more realistic by taking into account the effects of movement on body tissues as investigated in the work of Black and colleagues [164, 215], neural networks that are capable of voice cloning are emerging [5] and allow even for post-hoc manipulations of existing video material [259] are added to the palette of cinematic or video tools. Picking up on the provocative question of one of our interviewees earlier (see Section 9.1), who asked *“How physical is the future of camera motion given the ability of recording light-fields?”*, we would like to expand on the question a bit further. In production, one of the first things to be developed is a script. Here, for instance, the script of the short film *Sunspring* (2016) [343, 492], was already written exclusively by an AI system. Then the script and screenplay need to be translated into an actual motion picture usually either by physical recordings or by rendering CGIs. While it is ‘easier’ to imagine a CGI based film to be produced by an AI system, also physical recording is not out of this picture entirely. Systems as the Stuntronics robots mimicking human-like features might be recorded by a computer-controlled camera robot. Subsequently, the robots might be used as a base for rendering CGI on top, similar to the approach of car rigs such as the Blackbird. The material rendered on top might stem from a library of human actors that was scanned previously. The library might entail scans of the actors in motion, including the movement of their body tissue and a recording of their voice that might be fed into a voice cloning system. Alternatively, such ‘human rig’ robots might be recorded by an array of light field cameras, rendering recording camera motion obsolete and hence making it more easily accessible to an AI supported post-production system. Altering the visual appearance in the post-production is already an area where the application of neural network has been explored, as, for the film *Come Swim* (2017) [341, 133, 456]. Here, the visuals were altered by the system to mimic the look of an impressionist painting made by its director and writer Kirsten Stewart.

Such systems currently are evolving as individual components, that potentially might mature in the future. In a more mature future state, fusing these systems into a larger production suite seems reasonably imaginable. Given a scenario as sketched out above, one might especially consider the importance of the role that AI might consequently get. Based on our depiction above, it might be implied that many of the steps of the production might be automated and delegated to AI systems only. While there is no shortage of cinematic representations of autonomously acting AI systems that take over priorly human domains, in our estimation, its more likely that such systems might assist and guide humans in the creative process of cinematic production using AI enriched production environments. So, while there is still a human component in the production cycle, a central question remains: *"How physical is the future of cinematic production?"* While AI and robot systems might be used for reoccurring and labour-intensive task and/or dangerous recordings, the physical presence of actors and recording systems might overall decline yet remain an integral part of the production process. However, how much of the production might be transferred to the virtual domain might further be one of the determining factors that shapes the future of cinematic production on a larger scale; especially when considering its various conceptualisations as being a business, a workplace, a storytelling vehicle, a means of entertainment and a form of art.

How Linear is Future Cinematic Consumption?

In terms of the non-linear or on-demand distribution of cinematic content, we already approximated the status quo earlier in Section 2.1. However, there is also a different kind of non-linearity that one could consider regarding the consumption of cinematic content. For reasons of better differentiation, it might alternatively better to refer to it as the degree of interactivity within the storytelling. Interactive storytelling in movies has already been explored priorly, but somehow hardly evolved into a mass-market phenomenon; despite the increasing technological capabilities that are currently available. Overall, to us, it seems that so far interactive cinema seems to be a somewhat uncanny spot between straight storytelling which observably has a multitude of strengths as well as a long tradition and (technologically advanced) forms of interactive storytelling as found to be very popular in (already interactive) games. Given its rare use in-the-wild, it might be reasonable to expect a similarly rare use in the future. In our estimation, however, it is at the moment hard to properly estimate the effect that new ways of consuming cinematic content such as HMDs might have on future cinema and its storytelling, especially when considering their particular interactive properties. In summary, it seems to us that two fundamental conceptualisations about cinematic content might be formed that underlie the phenomenon of telling stories cinematically. On the one hand, each new technology such as digital recordings, streaming on apps running on multi-touch tablets or more recently HMDs have their own properties and challenges that require unique adaptations in the ways that content is crafted and delivered. However, on the other hand, straight or linear storytelling seems to be a (more or less) constant feature that remains stable across the various technological platforms or even media (books vs. movies). While both seem to be somewhat mutually exclusive, at the same time interactive technologies are increasingly used, but linear non-interactive storytelling is still the paramount paradigm. One could argue, in a binary sense, that ‘interactive cinema’ is no longer cinema, as it is traditionally conceptualised. Rather, it is a new form in need of a new definition or term. In contrast to such a binary classification, we would think of the overall issue rather in terms of degrees of interactivity. Consequently, it is, in our estimation, more a question of to which degree interactivity is integrated into future cinema. Whatever such a future might bring, it might be a further strong influencing factor that might change the way that content is produced and also how it is consumed. However, as stated above, given the slow increase of its use so far, it seems reasonable to only further expect a similar slow increase in the future; that is unless a major change in the overall experience might proceed it. Such a major influencing factor could be, for instance, imagined when the amount of personalisation that is incorporated in the development of the cinematic content is increased, as we will outline below.

How Personalised is Future Cinema?

Interactive storytelling as indicated above is often understood as requiring (mainly) explicit interactions. For instance, users can select a different story path for a character or choose a particular ending. In our opinion, however, also more implicit ways of integrating central human aspects regarding commonly shared as well as individual human aspects and traits are to be considered. On the one hand, those are already incorporated in the writing of movie scripts as well as their selection for publication. In consequence, such human-adapted narrative motives might also be used to guide future systems in crafting a story (and its cinematic representation). On the other hand, these general human aspects that guide the storytelling might be further enriched by a greater understanding of the unique aspects of an individual consumer. To specify more precisely what we are referring to, we want to introduce some details below.

The above-introduced integration of general human aspects with regard to storytelling has a long tradition, however, an abstract representation of it has only emerged more recently compared its extensive application. Narratives that focus on one central character or a ‘hero’ follow a theme that is found in the traditions of many cultures and even goes back to very early tribal tales and mythologies. In his book, *“The Hero with a Thousand Faces”*, Joseph Campbell traces back a multitude of these mythological representations of the hero across multiple cultures (where they are thought to be emerging independently from each other). Despite their independent emergence, these stories share a common pattern of how the journey of the hero unfolds. Thinking about it from the perspective of a Grounded Theory approach, one might say that Campbell looked a broad dataset of (independent) stories and classified their shared traits in a bottom-up way somewhat akin to an open coding approach. What he found was that the stories investigated seemed to have a similar underlying blueprint, a story pattern that might be divided into three acts, each of which is composed of circa 5 to 6 narrative motives and twists that in their sequence make the story appealing, captivating and relevant to a human audience. His finding of such independent yet very similar narrative structures further suggested to Campbell that the story of the hero's journey is deeply rooted into human beings (or more precisely, deeply associated with human consciousness). For him, it might potentially also be the most central of the stories being told by humans. Therefore, he refers to the hero's journey also as the monomyth. Campbell and others, such as the psychologist Neumann (student of Carl Gustav Jung), even classify religious narrations as of Gautama Buddha or Jesus Christ as instances of it. Although narrations of the hero's journey have a long tradition and vast application, they hardly seem to get out of style. This might be indicative of their greater value to human beings that is hardly lost even if being exposed to such type of narrations multiple times in one's life. The underlying narrative structure is not only used in ancient mythology and central religious texts but is also found in screenplays of modern cinema. For example, George Lucas used Campbell's work as a foundation that inspired the storytelling of his first trilogy of Star Wars movies [374]. Further, also more recent films such as *The Matrix* (1999) might be interpreted as an example of the hero's journey as suggested by Richardson [50].

Using an underlying blueprint to create a captivating story is also found in the work of Blake Snyder [246]. In his book, *"Save the Cat!"*, Snyder provides classifications of stories and lays out central points in storytelling that make a movie script appealing to a general audience. Based on his suggestions, a hero should *"save the cat"* [512] early on in the movie (when the character is introduced) to show the audience the good qualities associated with this person and to emphasise relatability. Following this initial moment, he also proposes to tell a story that is embedded into a structure of three acts that employ certain narrative motives and twists along this plot blueprint to make them more appealing to a human audience. In detail, he introduces a set of 15 so-called beats that refer to these particular points in the narration [513, 514]. Along with the beats, he also suggests how much per cent of the overall story one should plan in for these particular points in the story. In conclusion, he suggests that a hero story needs to employ these 15 narrative points to make a story interesting and captivating for a human audience. His work seemingly shows similarities to the work of Campbell, however, this particular approach to structuring storytelling is strongly focussed on writing screenplays intended for movies, especially (but not exclusively) for a larger scale production context.

The above-mentioned ways of informing cinematic production might be considered as stories being personalised based on what other storytellers found to be working and/or what other consumers liked in the past, so to speak. A modern-day version of coming to a similar kind of understanding of what others liked in the past regarding storytelling can be found in two example systems that we want to introduce shortly. First, ScriptBook [371] is a company that provides an AI system for screenplay analysis that aims at identifying the proclivity for a screenplay to become an economic success.

Their systems were trained on a dataset of film scripts released between 1970 and 2016. Based on an interview with the company's CEO, Nadira Azermai, their systems are three times more accurate than human expertise (however sample size and population sample is not mentioned) [418]. Second, Netflix Chief Content Officer Ted Sarandos together with others analysed aggregated user data to successfully inform investment decisions. In particular, they analysed data collected from natural user behaviour such as ratings, viewing history, the popularity of shows but also associated data as which producers or actor were featured. As a result, they found that users might be interested in a political drama series depicting a senator. Therefore, they suggested investing in a US adaptation of the BBC mini-series *House of Cards* (1990). As the data suggested that people who already liked that series also liked actor Kevin Spacey and director David Fincher, Netflix chose to cast both [350].

What we would like to suggest is that a future system might be capable of assisting the storytelling process not only based on training data aggregated from multiple users. However, that similar as in the first Star Wars trilogy, the storytelling, that precedes any camera motion that is in dialogue with the story, might be crafted by or with the assistance of a system that is not only trained on existing scripts (as done in ScriptBook). Yet, their training might include abstract knowledge on how to write a story to make it relevant to humans. Further, it might also be enriched by an understanding of the more detailed watching behaviour. Such an adaption to fundamental human proclivities regarding storytelling might allow entertaining a mass-market given its general applicability, however, it might also be considered roughly as somewhat of a 'one size fits all' approach. Especially, the hero's journey or the save the cat approach is quite commonly used for that matter. However, advanced systems might also allow for further personalisation to the individual. The notion of a hero's journey was introduced by Carl Gustav Jung in his work on what he refers to as archetypes. Jung not only conceptualised mythologies as an abstract representation of such archetypes that are detached from life but also considered his own life to be an example of the hero's journey. So, if a system was capable of producing content technically as sketched out in the first section of these Closing Remarks, and was knowledgable about telling stories that incorporate such abstract knowledge on the human condition in general, these capabilities might be fused with further data on the path of an individual (hero) in particular. Current systems are already aiming at (very simplistic versions of) personalised 'movies'. For instance, current mobile phones (featuring rather simple capabilities) offer to generate a short movie from a set of photos on demand that roughly approximates an animated retrospect look into one's life or at least its photographic representation. Further, they allow adjusting the soundtrack setting to create a particular mood. However, it is directly relying on captured material and does not feature any peculiar narrative. Extrapolating in the future, an evolved system of that fundamental idea, might be capable of creating a (short) movie, based on one's genre preferences, that takes into account recent personal life circumstances and might be able to match them to a relevant storytelling based on archetypal motives that might be further transferred to the rendering of an advanced cinematic representation. As of today, imagining such a representation to incorporate human aspects in terms of camera motion is already feasible as research in that area demonstrates. Sanokho and colleagues [232], for instance, trace camera motion graphs from existing visual material. Given more advanced versions of this approach and a subsequently derived library of camera paths taken by human operators enriched by knowledge on meta aspects such as their motivation, might be further contributing to high-quality results. Such a potential personalised movie (or series of movies) might carry the capacity for providing a captivating and immersive personalised content. Together with the above-mentioned propositions following the other questions, such a personalised development of cinematic context might be imagined to go along with (some degree of) interactive storytelling the cinematic representation of which is mainly virtually rendered and that finally becomes consumed by people on increasingly immersive devices. So that in the end, an immersive experience is not only fostered by the design of the external devices but also with regard to the internal aspects such as the narration of captivating content from a general human as well as an individual perspective.

One of the key questions that remain when imagining such a system is which user data might be used (or needed) in order to inform a personalised narrative and cinematic representation. To inform personalised storytelling, most likely accumulated data might be required from a multitude of sources. Among those, one might find continuously measured physiological data such as heart-rate or pupil diameters which might serve as bodily markers for emotionally loaded encounters in one's life. Further one might bring those together with a person's schedule, location data and/or occurrences in personal relationships especially when digitally communicated. Also, preferences and behavioural patterns in media consumption might be a reasonable and rich data source to be included. Beyond such direct implicit information one might also consider profiling a person based on the Big Five personality trait model that might only be used to predict the likelihood of an individual's preference towards certain genres, actors, directors or camera operators but also to match the profile with similarly profiled persons to extend the repertoire to options to be included.

In summary, although it is highly speculative to predict the future of cinematic production and consumption, one might consider that the technological answers found to all of the above (and potentially even more) questions are contributing to the manifestation of one particular out of the many possible futures. Going along with this process, we expect some forms of advanced systems that might assist in the production process in order to make it more efficient especially in regard to reoccurring tasks. This, in particular, might entail the possibility of human beings to focus on the creative aspects as those might be the hardest domain to be delegated to machines. The work on such creative aspects might, on the one hand, be guided by systems incorporating abstract general storytelling and cinematic knowledge and on the other be capable of adapting those general principles to the creation of personalised content taking into account personal trait aspects, preferences and potentially even encounters of everyday lives. In consequence, they might allow for crafting meaningful content presented in captivating ways that are created on-demand and displayed in increasingly immersive environments.

IV

References and Appendix

References

Bibliography

[1] Henk Aarts, Ruud Custers, and Daniel M. Wegner. 'On the Inference of Personal Authorship: Enhancing Experienced Agency by Priming Effect Information'. In: *Consciousness and Cognition* 14.3 (2005), pp. 439–458.

[2] Pablo F. Alcantarilla, Adrien Bartoli, and Andrew J. Davison. 'KAZE Features'. In: *Computer Vision – ECCV 2012*. Ed. by Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid. Springer Berlin Heidelberg, 2012, pp. 214–227. ISBN: 978-3-642-33783-3. DOI: 10.1007/978-3-642-33783-3_16.

[3] Pablo F. Alcantarilla, Jesus Nuevo, and Adrien Bartoli. 'Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces'. In: *Proceedings British Machine Vision Conference 2013*. Ed. by Tilo Burghardt, Dima Damen, Walterio Mayol-Cuevas, and Majid Mirmehdi. BMVC '13. Bristol, UK: BMVA Press, 2013, pp. 13.1–13.11. DOI: 10.5244/C.27.13.

[4] Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. 'InspirationWall: Supporting Idea Generation Through Automatic Information Exploration'. In: *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. C&C '15. Glasgow, United Kingdom: ACM, 2015, pp. 103–106. ISBN: 978-1-4503-3598-0.

[5] Sercan Arik, Jitong Chen, Kainan Peng, Wei Ping, and Yanqi Zhou. 'Neural Voice Cloning with a Few Samples'. In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc., 2018, pp. 10019–10029.

[6] Yasuo Ariki, Shintaro Kubota, and Masahito Kumano. 'Automatic Production System of Soccer Sports Video by Digital Camera Work Based on Situation Recognition'. In: *Eighth IEEE International Symposium on Multimedia*. Eighth IEEE International Symposium on Multimedia. ISM '06. San Diego, CA, USA: IEEE, 2006, pp. 851–860. ISBN: 0-7695-2746-9. DOI: 10.1109/ISM.2006.37.

[7] Ronald C. Arkin. *Behavior-Based Robotics*. USA: MIT Press, 1998. ISBN: 0-262-01165-4.

[8] Lisanne Bainbridge. 'Ironies of Automation'. In: *Proceedings of the IFAC/IFIP/IFORS/IEA Conference, Baden-Baden, Federal Republic of Germany, 27–29 September 1982*. Ed. by J. E. Rijnsdorp. Oxford, UK: Pergamon, 1983, pp. 129–135. ISBN: 978-0-08-029348-6. DOI: 10.1016/B978-0-08-029348-6.50026-9.

[9] T. Baltrušaitis, P. Robinson, and L. P. Morency. 'Openface: An Open Source Facial Behavior Analysis Toolkit'. In: *2016 IEEE Winter Conference on Applications of Computer Vision*. 2016

IEEE Winter Conference on Applications of Computer Vision. WACV '16. Lake Placid, NY, USA: IEEE, 2016, pp. 1–10. ISBN: 978-1-5090-0641-0. DOI: 10.1109/WACV.2016.7477553.

[10] Aaron Bangor, Philip Kortum, and James Miller. 'Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale'. In: *Journal of Usability Studies* 4.3 (2009), pp. 114–123.

[11] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 'SURF: Speeded Up Robust Features'. In: *Computer Vision – ECCV 2006*. Ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. ECCV '06. Springer Berlin Heidelberg, 2006, pp. 404–417. ISBN: 978-3-540-33833-8. DOI: 10.1007/11744023_32.

[12] Hrvoje Benko and Daniel Wigdor. 'Imprecision, Inaccuracy, and Frustration: The Tale of Touch Input'. In: *Tabletops - Horizontal Interactive Displays*. Ed. by Christian Müller-Tomfelde. London: Springer London, 2010, pp. 249–275. ISBN: 978-1-84996-113-4. DOI: 10.1007/978-1-84996-113-4_11.

[13] Bruno Berberian, Jean-Christophe Sarrazin, Patrick Le Blaye, and Patrick Haggard. 'Automation Technology and Sense of Control: A Window on Human Agency'. In: *PLoS One* 7.3 (2012), e34075. DOI: 10.1371/journal.pone.0034075.

[14] Jacques Bertin. *Sémiologie Graphique: Les Diagrammes-Les Réseaux-Les Cartes*. Paris, France: Mouton and Co, 1967. 431 pp.

[15] Charles E. Billings. 'Aviation Automation: The Search for a Human-Centered Approach'. In: *Ergonomics in Design* 6.4 (1997). Ed. by John A. Wise, pp. 31–31. ISSN: 1064-8046. DOI: 10.1177/106480469800600410.

[16] Sarah-Jayne Blakemore, Daniel M. Wolpert, and Christopher D. Frith. 'Abnormalities in the Awareness of Action'. In: *Trends in Cognitive Sciences* 6.6 (2002), pp. 237–242. ISSN: 1364-6613. DOI: 10.1016/S1364-6613(02)01907-1.

[17] Richard A. Bolt. '"Put-That-There": Voice and Gesture at the Graphics Interface'. In: *SIGGRAPH Comput. Graph.* 14.3 (1980), pp. 262–270. DOI: 10.1145/800250.807503.

[18] Francisco Bonin-Font, Alberto Ortiz, and Gabriel Oliver. 'Visual Navigation for Mobile Robots: A Survey'. In: *Journal of Intelligent and Robotic Systems* 53.3 (2008), pp. 263–296. ISSN: 0921-0296. DOI: 10.1007/s10846-008-9235-4.

[19] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola Jr, and Ivan Poupyrev. 'An Introduction to 3-D User Interface Design'. In: *Presence: Teleoperators & Virtual Environments* 10.1 (2001), pp. 96–108. DOI: 10.1162/105474601750182342.

[20] Oliver Bown. 'Generative and Adaptive Creativity: A Unified Approach to Creativity in Nature, Humans and Machines'. In: *Computers and Creativity*. Ed. by Jon McCormack and Mark d'Inverno. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 361–381. ISBN: 978-3-642-31727-9. DOI: 10.1007/978-3-642-31727-9_14.

[21] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. 1st ed. Sebastopol, CA, USA: O'Reilly Media, 2008. ISBN: 978-0-596-51613-0.

[22] Marta Braun. *Picturing Time: The Work of Etienne-Jules Marey (1830–1904)*. Chicago, Illinois, USA: University of Chicago Press, 1992. ISBN: 0-226-07173-1.

[23] James R. Bright. 'Does Automation Raise Skill Requirements?' In: *Harvard Business Review* (1958), p. 18.

[24] Tim Brown. 'Design Thinking'. In: *Harvard Business Review* 86.6 (2008), pp. 84–92. URL: `https://fusesocial.ca/wp-content/uploads/sites/2/2018/06/Design-Thinking.pdf` (visited on 06/12/2019).

[25] Mathias Bürki, Igor Gilitschenski, Elena Stumm, Roland Siegwart, and Juan Nieto. 'Appearance-Based Landmark Selection for Efficient Long-Term Visual Localization'. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS '16. Daejeon, South Korea: IEEE, 2016, pp. 4137–4143. ISBN: 978-1-5090-3762-9. DOI: 10.1109/IROS.2016.7759609.

[26] Nicholas Burtnyk, Azam Khan, George Fitzmaurice, Ravin Balakrishnan, and Gordon Kurtenbach. 'StyleCam: Interactive Stylized 3D Navigation Using Integrated Spatial & Temporal Controls'. In: *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*. UIST '02. Paris, France: ACM, 2002, pp. 101–110. ISBN: 1-58113-488-6. DOI: 10.1145/571985.572000.

[27] Nicolas Burtnyk, Azam Khan, George Fitzmaurice, and Gordon Kurtenbach. 'ShowMotion: Camera Motion Based 3D Design Review'. In: *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games*. I3D '06. Redwood City, California: ACM, 2006, pp. 167–174. ISBN: 1-59593-295-X. DOI: 10.1145/1111411.1111442.

[28] Harry Ediwn Burton. 'The Optics of Euclid'. In: *Jounal of the Optical Society of America* 35.5 (1945), pp. 357–372. ISSN: 0030-3941. DOI: 10.1364/JOSA.35.000357.

[29] Bill Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. 1st ed. Burlington: Morgan Kaufmann, 2007. 448 pp. ISBN: 978-0-12-374037-3. DOI: 10.1016/B978-012374037-3/50046-8.

[30] Zachary Byers, Michael Dixon, Kevin Goodier, Cindy M. Grimm, and William D. Smart. 'An Autonomous Robot Photographer'. In: *Proceedings 2003 IEEE/RSJ International Con-*

ference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453). Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453). Vol. 3. Las Vegas, NV, USA: IEEE, 2003, pp. 2636–2641. DOI: 10.1109/IROS.2003.1249268.

[31] Zachary Byers, Michael Dixon, William D. Smart, and Cindy M. Grimm. 'Say Cheese! Experiences with a Robot Photographer'. In: *AI Magazine* 25.3 (2004), pp. 37–46.

[32] Evan A. Byrne and Raja Parasuraman. 'Psychophysiology and Adaptive Automation'. In: *Psychophysiology of Workload* 42.3 (1996), pp. 249–268. ISSN: 0301-0511. DOI: 10.1016/0301-0511(95)05161-9.

[33] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. 'BRIEF: Binary Robust Independent Elementary Features'. In: *Computer Vision – ECCV 2010*. Ed. by Kostas Daniilidis, Petros Maragos, and Nikos Paragios. Springer Berlin Heidelberg, 2010, pp. 778–792. ISBN: 978-3-642-15561-1. DOI: 10.1007/978-3-642-15561-1_56.

[34] John Canny. 'A Computational Approach to Edge Detection'. In: *Readings in Computer Vision*. San Francisco, CA, USA: Morgan Kaufmann, 1987, pp. 184–203. ISBN: 978-0-08-051581-6. DOI: 10.1016/B978-0-08-051581-6.50024-6.

[35] James Carifio and Rocco J Perla. 'Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends About Likert Scales and Likert Response Formats and Their Antidotes'. In: *Journal of Social Sciences* 3.3 (2007), pp. 106–116.

[36] Maria B. Carmo, Ana P. Afonso, António Ferreira, Ana P. Cláudio, and Edgar Montez. 'Symbol Adaptation Assessment in Outdoor Augmented Reality'. In: *2014 International Conference on Computer Graphics Theory and Applications*. GRAPP '14. Lisbon, Portugal: IEEE, 2014, pp. 1–10. ISBN: 978-989-758-078-9.

[37] Sheelagh Carpendale. 'Considering Visual Variables as a Basis for Information Visualisation'. In: 2003. URL: `https://prism.ucalgary.ca/handle/1880/45758` (visited on 06/12/2019).

[38] Peter Carr, Michael Mistry, and Iain Matthews. 'Hybrid Robotic/Virtual Pan-Tilt-Zoom Cameras for Autonomous Event Recording'. In: *Proceedings of the 21st ACM International Conference on Multimedia*. MM '13. Barcelona, Spain: ACM, 2013, pp. 193–202. ISBN: 978-1-4503-2404-5. DOI: 10.1145/2502081.2502086.

[39] Jessica R. Cauchard, L. E. Jane, Kevin Y. Zhai, and James A. Landay. 'Drone & Me: An Exploration into Natural Human-Drone Interaction'. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. New York, NY, USA: ACM, 2015, pp. 361–365. ISBN: 978-1-4503-3574-4. DOI: 10.1145/2750858.2805823.

[40] Federica Cena and Fabiana Vernero. 'A Study on User Preferential Choices about Rating Scales'. In: *International Journal of Technology and Human Interaction* 11.1 (2015), pp. 33–54. ISSN: 1548-3908. DOI: 10.4018/ijthi.2015010103.

[41] Alphonse Chapanis. 'On the Allocation of Functions Between Men and Machines'. In: *Occupational Psychology* 39.1 (1965), pp. 1–11.

[42] François Chaumette and Seth Hutchinson. 'Visual Servo Control Part I: Basic Approaches'. In: *IEEE Robotics & Automation Magazine* 13.4 (2006), pp. 82–90. ISSN: 1070-9932. DOI: 10.1109/MRA.2006.250573.

[43] François Chaumette and Seth Hutchinson. 'Visual Servo Control Part II: Advanced Approaches'. In: *IEEE Robotics & Automation Magazine* 14.1 (2007), pp. 109–118. ISSN: 1070-9932. DOI: 10.1109/MRA.2007.339609.

[44] Christine Chen, Oliver Wang, Simon Heinzle, Peter Carr, Aljoscha Smolic, and Markus Gross. 'Computational Sports Broadcasting: Automated Director Assistance for Live Sports'. In: *2013 IEEE International Conference on Multimedia and Expo (ICME)*. 2013 IEEE International Conference on Multimedia and Expo (ICME). San Jose, CA, USA: IEEE, –July 19, 2013, pp. 1–6. ISBN: 978-1-4799-0015-2. DOI: 10.1109/ICME.2013.6607445.

[45] Fan Chen and Christophe De Vleeschouwer. 'Personalized Production of Basketball Videos from Multi-Sensored Data Under Limited Display Resolution'. In: *Special Issue on Multi-Camera and Multi-Modal Sensor Fusion* 114.6 (2010), pp. 667–680. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2010.01.005.

[46] Jianhui Chen and Peter Carr. 'Autonomous Camera Systems: A Survey'. In: *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI Workshops. Québec City, Québec, Canada, 2014, pp. 18–22. URL: `http://www.aaai.org/ocs/index.php/WS/AAAIW14/paper/view/8733` (visited on 06/12/2019).

[47] Erin Cherry and Celine Latulipe. 'Quantifying the Creativity Support of Digital Tools Through the Creativity Support Index'. In: *ACM Transactions on Computer-Human Interaction*. TOCHI '14 21.4 (2014), p. 21. DOI: 10.1145/2617588.

[48] Hsiang-Jen Chien, Chen-Chi Chuang, Chia-Yen Chen, and Reinhard Klette. 'When to Use What Feature? SIFT, SURF, ORB, or A-KAZE Features for Monocular Visual Odometry'. In: *2016 International Conference on Image and Vision Computing New Zealand*. 2016 International Conference on Image and Vision Computing New Zealand. IVCNZ '16. Palmerston North, New Zealand: IEEE, 2016, pp. 1–6. ISBN: 978-1-5090-2747-7. DOI: 10.1109/IVCNZ.2016.7804434.

[49] Tien-Ting Chiu, Kuu-Young Young, Shang-Hwa Hsu, Chun-Ling Lin, Chin-Teng Lin, Bing-Shiang Yang, and Zong-Ren Huang. 'A Study of Fitts' Law on Goal-Directed Aiming Task with Moving Targets'. In: *Perceptual and Motor Skills* 113.1 (2011), pp. 339–352. DOI: 10.2466/05.06.25.PMS.113.4.339-352.

[50] Chris Richardson. 'The Matrix as the Hero's Journey'. In: *Quest* 91.6 (2003), pp. 220–225. URL: `https://www.theosophical.org/publications/quest-magazine/42-publications/quest-magazine/1507-the-matrix-as-the-heros-journey` (visited on 06/12/2019).

[51] David B. Christianson, Sean E. Anderson, Li-wei He, David H. Salesin, Daniel S. Weld, and Michael F. Cohen. 'Declarative Camera Control for Automatic Cinematography'. In: *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 1*. AAAI'96. Portland, Oregon: AAAI Press, 1996, pp. 148–155. ISBN: 0-262-51091-X. URL: `http://dl.acm.org/citation.cfm?id=1892875.1892897` (visited on 06/12/2019).

[52] Marc Christie and Hiroshi Hosobe. 'Through-the-Lens Cinematography'. In: *Smart Graphics: 6th International Symposium, SG 2006, Vancouver, Canada, July 23-25, 2006, Proceedings*. Ed. by Andreas Butz, Brian Fisher, Antonio Krüger, and Patrick Olivier. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 147–159. ISBN: 978-3-540-36295-1. DOI: 10.1007/11795018_14.

[53] Marc Christie, Patrick Olivier, and Jean-Marie Normand. 'Camera Control in Computer Graphics'. In: *Computer Graphics Forum*. Vol. 27. 8. Blackwell Publishing Ltd, 2008, pp. 2197–2218. DOI: 10.1111/j.1467-8659.2008.01181.x.

[54] William J. Conover. *Practical Nonparametric Statistics*. 3rd ed. USA: Wiley, 1999. 584 pp. ISBN: 978-0-471-16068-7.

[55] Antonia S. Conti, Carsten Dlugosch, and Klaus Bengler. 'Detection Response Tasks: How Do Different Settings Compare?' In: *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. AutomotiveUI '12. Portsmouth, New Hampshire: ACM, 2012, pp. 257–260. ISBN: 978-1-4503-1751-1. DOI: 10.1145/2390256.2390298.

[56] Alan Cooper, Robert Reimann, David Cronin, and Christopher Noessel. *About Face: The Essentials of Interaction Design*. 4th ed. Indianapolis, IN, USA: Wiley, 2014. 720 pp. ISBN: 978-1-118-76657-6.

[57] Patricia Ivette Cornelio Martinez, Silvana De Pirro, Chi Thanh Vi, and Sriram Subramanian. 'Agency in Mid-Air Interfaces'. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: ACM, 2017, pp. 2426–2439. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025457.

[58] Gregory M. Corso and Megan M. Moloney. ‘Human Performance, Dynamic Function Allocation and Transfer of Training’. In: *Proceedings of the 5th International Conference on Human Aspects of Advanced Manufacturing*. Ed. by R. Koubek and W. Karwowski. Vol. Manufacturing Agility and Hybrid Automation - I. Louisville, KY, USA: IEA Press, 1996.

[59] Anna L. Cox, Paul Cairns, Nadia Berthouze, and Charlene Jennett. ‘The Use of Eyetracking for Measuring Immersion’. In: *Workshop on What Have Eye Movements Told Us so Far, and What Is Next*. London, UK: UCL Interaction Centre, Department of Psychology, 2006, pp. 26–29.

[60] David Coyle, James W. Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. ‘I Did That! Measuring Users’ Experience of Agency in Their Own Actions’. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’12. New York, NY, USA: ACM, 2012, pp. 2025–2034. DOI: 10.1145/2207676.2208350.

[61] Mihaly Csikszentmihalyi. ‘Flow and the Psychology of Discovery and Invention’. In: *New York: Harper Collins* (1996).

[62] Mihaly Csikszentmihalyi, Sami Abuhamdeh, and Jeanne Nakamura. ‘Flow’. In: *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Ed. by Mihaly Csikszentmihalyi. Dordrecht: Springer Netherlands, 2014, pp. 227–238. ISBN: 978-94-017-9088-8. DOI: 10.1007/978-94-017-9088-8_15.

[63] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. ‘Detecting Moving Objects, Ghosts, and Shadows in Video Streams’. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.10 (2003), pp. 1337–1342. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2003.1233909.

[64] Beth Arburn Davis. *Development and Validation of a Scale of Perceived Control Across Multiple Domains*. Philadelphia College of Osteopathic Medicine, 2004.

[65] Joost C. F. de Winter and Dimitra Dodou. ‘Why the Fitts List Has Persisted throughout the History of Function Allocation’. In: *Cognition, Technology & Work* 16.1 (2014), pp. 1–11. ISSN: 1435-5566. DOI: 10.1007/s10111-011-0188-1.

[66] Asaf Degani. *Taming Hal: Designing Interfaces Beyond 2001*. 1st ed. New York, NY, USA: Palgrave Macmillan US, 2003. 312 pp. ISBN: 978-0-312-29574-5. DOI: 10.1057/9781403982520.

[67] Sidney W. A. Dekker and David D. Woods. ‘MABA-MABA or Abracadabra? Progress on Human–Automation Co-Ordination’. In: *Cognition, Technology & Work* 4.4 (2002), pp. 240–244. ISSN: 1435-5558. DOI: 10.1007/s101110200022.

[68] Rachid Deriche. ‘Using Canny’s Criteria to Derive a Recursively Implemented Optimal Edge Detector’. In: *International Journal of Computer Vision* 1.2 (1987), pp. 167–187. ISSN: 1573-1405. DOI: 10.1007/BF00123164.

[69] Sebastian Deterding, Jonathan Hook, Rebecca Fiebrink, Marco Gillies, Jeremy Gow, Memo Akten, Gillian Smith, Antonios Liapis, and Kate Compton. ‘Mixed-Initiative Creative Interfaces’. In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI ’17. Denver, Colorado, USA: ACM, 2017, pp. 628–635. ISBN: 978-1-4503-4656-6. DOI: 10.1145/3027063.3027072.

[70] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. ‘SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support’. In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*. AAMAS ’14. Paris, France: International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068. ISBN: 978-1-4503-2738-1. URL: http://dl.acm.org/citation.cfm?id=2615731.2617415.

[71] Arne Dietrich and Riam Kanso. ‘A Review of EEG, ERG, and Neuroimaging Studies of Creativity and Insight’. In: *Psychological Bulletin* 136.5 (2010), p. 822. DOI: 10.1037/a0019749.

[72] Mia Y. Dong, Kristian Sandberg, Bo M. Bibby, Michael N. Pedersen, and Morten Overgaard. ‘The Development of a Sense of Control Scale’. In: *Frontiers in Psychology* 6 (2015), p. 1733. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.01733.

[73] Petr Doubek, Indra Geys, Tomáš Svoboda, and Luc Van Gool. ‘Cinematographic Rules Applied to a Camera Network’. In: *Omnivis2004: The Fifth Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*. Prague, Czech Republic: Czech Technical University. 2004, pp. 17–29.

[74] Steven M. Drucker, Tinsley A. Galyean, and David Zeltzer. ‘CINEMA: A System for Procedural Camera Movements’. In: *Proceedings of the 1992 Symposium on Interactive 3D Graphics*. I3D ’92. Cambridge, MA, USA: ACM, 1992, pp. 67–70. ISBN: 0-89791-467-8. DOI: 10.1145/147156.147166.

[75] Jane L. E, Ilene L. E, James A. Landay, and Jessica R. Cauchard. ‘Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques’. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. Denver, Colorado, USA: ACM, 2017, pp. 6794–6799. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025755.

[76] Jeffrey P. Ebert and Daniel M. Wegner. 'Time Warp: Authorship Shapes the Perceived Timing of Actions and Events'. In: *Consciousness and Cognition* 19.1 (2010), pp. 481–489. ISSN: 1053-8100. DOI: 10.1016/j.concog.2009.10.002.

[77] Elwyn Edwards and Frank P. Lees. *Man and Computer in Process Control*. Institution of Chemical Engineers, 1973.

[78] Paul Ekman. 'Facial Expressions of Emotion: New Findings, New Questions'. In: *Psychological Science* 3.1 (1992), pp. 34–38. ISSN: 0956-7976. DOI: 10.1111/j.1467-9280.1992.tb00253.x.

[79] Andreas Ellwanger, **Axel Hoesl**, and Andreas Butz. 'Axis- Plus Content-Based Control for Camera Drones: Design and Evaluation of User Interface Concepts'. In: *Proceedings of the 4th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications*. DroNet '18. Munich, Germany: ACM, 2018, pp. 9–14. ISBN: 978-1-4503-5839-2. DOI: 10.1145/3213526.3213529.

[80] Mica R. Endsley. 'Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task'. In: *Ergonomics* 42.3 (1999), pp. 462–492.

[81] Mica R. Endsley. 'The Application of Human Factors to the Development of Expert Systems for Advanced Cockpits'. In: *Proceedings of the Human Factors Society Annual Meeting* 31.12 (1987), pp. 1388–1392. ISSN: 0163-5182. DOI: 10.1177/154193128703101219.

[82] Mica R. Endsley and Debra G. Jones. *Designing for Situation Awareness: An Approach to User-Centered Design*. 2nd ed. Boca Raton, FL, USA: CRC Press, 2016. ISBN: 978-1-4200-6358-5.

[83] Mica R. Endsley and Esin O. Kiris. 'The Out-of-the-Loop Performance Problem and Level of Control in Automation'. In: *Human Factors* 37.2 (1995), pp. 381–394. ISSN: 0018-7208. DOI: 10.1518/001872095779064555.

[84] Mica R. Endsley, Stephen J. Selcon, Thomas D. Hardiman, and Darryl G. Croft. 'A Comparative Analysis of SAGAT and SART for Evaluations of Situation Awareness'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 42.1 (1998), pp. 82–86. ISSN: 1541-9312. DOI: 10.1177/154193129804200119.

[85] Hing Yee Eng, Diyu Chen, and Yuhong Jiang. 'Visual Working Memory for Simple and Complex Visual Stimuli'. In: *Psychonomic Bulletin & Review* 12.6 (2005), pp. 1127–1133. ISSN: 1531-5320. DOI: 10.3758/BF03206454.

[86] Andy Field and Graham Hole. *How to Design and Report Experiments*. Trowbridge, GB: Sage, 2003. ISBN: 0-7619-7382-6.

[87] Paul M. Fitts. *Human Engineering for an Effective Air-Navigation and Traffic-Control System*. National Research Council, Division of Anthropology and Psychology, Committee on Aviation Psychology, 1951.

[88] Frank Flemisch, Eugen Altendorf, Yigiterkut Canpolat, Gina Weßel, Marcel Baltzer, Daniel Lopez, Nicolas D. Herzberger, Gudrun M. I. Voß, Maximilian Schwalm, and Paul Schutte. 'Uncanny and Unsafe Valley of Assistance and Automation: First Sketch and Application to Vehicle Automation'. In: *Advances in Ergonomic Design of Systems, Products and Processes*. Ed. by Christopher M. Schlick, Sönke Duckwitz, Frank Flemisch, Martin Frenz, Sinem Kuz, Alexander Mertens, and Susanne Mütze-Niewöhner. Springer Berlin Heidelberg, 2017, pp. 319–334. ISBN: 978-3-662-53305-5. DOI: 10.1007/978-3-662-53305-5_23.

[89] Wolfgang Förstner and Eberhard Gülch. 'A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features'. In: *Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*. 1987, pp. 281–305.

[90] Milton Friedman. 'A Correction: The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance'. In: *Journal of the American Statistical Association* 34.205 (1939), pp. 109–109. ISSN: 0162-1459. DOI: 10.1080/01621459.1939.10502372.

[91] Milton Friedman. 'The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance'. In: *Journal of the American Statistical Association* 32.200 (1937), pp. 675–701. ISSN: 0162-1459. DOI: 10.1080/01621459.1937.10503522.

[92] Joseph L. Gabbard, J. Edward Swan, Deborah Hix, Si-Jung Kim, and Greg Fitch. 'Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications'. In: *2007 IEEE Virtual Reality Conference*. 2007 IEEE Virtual Reality Conference. Charlotte, NC, USA: IEEE, 2007, pp. 35–42. ISBN: 1-4244-0905-5. DOI: 10.1109/VR.2007.352461.

[93] Henry P. Gage and Simon H. Gage. *Optic Projection, Principles, Installation, and Use of the Magic Lantern, Projection Microscope, Reflecting Lantern, Moving Picture Machine*. Optic Projection: Principles, Installation and Use of the Magic Lantern, Projection Microscope, Reflecting Lantern, Moving Picture Machine. Comstock, 1914. 731 pp.

[94] Quentin Galvane, Julien Fleureau, Francois-Louis Tariolle, and Philippe Guillotel. 'Automated Cinematography with Unmanned Aerial Vehicles'. In: *Eurographics Workshop on Intelligent Cinematography and Editing*. Ed. by M. Christie, Q. Galvane, A. Jhala, and R. Ronfard. WICED '16. The Eurographics Association, 2016. ISBN: 978-3-03868-005-5. DOI: 10.2312/wiced.20161097.

[95] Tinsley A. Galyean. 'Guided Navigation of Virtual Environments'. In: *Proceedings of the 1995 Symposium on Interactive 3D Graphics*. I3D '95. Monterey, CA, USA: ACM, 1995, 103–ff. ISBN: 0-89791-736-7. DOI: 10.1145/199404.199421.

[96] Mark Gatton and Leah Carreon. 'Probability and the Origin of Art: Simulations of the Paleo-Camera Theory'. In: *Journal of Applied Mathematics* 4 (2011).

[97] Mark Gatton, Leah Carreon, Madison Cawein, Walter Brock, and Valerie Scott. 'The Camera Obscura and the Origin of Art: The Case for Image Projection in the Paleolithic'. In: *Official Proceedings of the XV World Congress of the Union Internationale Des Sciences Préhistoriques et Protohistoriques*. UISPP. Oxford: Archaeopress, 2010.

[98] Christoph Gebhardt, Benjamin Hepp, Tobias Nägeli, Stefan Stevšić, and Otmar Hilliges. 'Airways: Optimization-Based Planning of Quadrotor Trajectories According to High-Level User Goals'. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. Santa Clara, California, USA: ACM, 2016, pp. 2508–2519. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858353.

[99] Gerard Genette. *Narrative Discourse*. Ithaca, USA: Cornell University Press, 1983. ISBN: 978-0-8014-9259-4.

[100] Michael Gleicher and Andrew Witkin. 'Through-the-lens Camera Control'. In: *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '92. New York, NY, USA: ACM, 1992, pp. 331–340. ISBN: 0-89791-479-1. DOI: 10.1145/133994.134088.

[101] Samuel Goldwyn. 'Hollywood in the Television Age'. In: *Hollywood Quarterly* 4.2 (1949), pp. 145–151. ISSN: 15490076. DOI: 10.2307/1209678.

[102] Sean Gustafson, Christian Holz, and Patrick Baudisch. 'Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device'. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. UIST '11. Santa Barbara, CA, USA: ACM, 2011, pp. 283–292. ISBN: 978-1-4503-0716-1. DOI: 10.1145/2047196.2047233.

[103] Carl Gutwin and Saul Greenberg. 'The Mechanics of Collaboration: Developing Low Cost Usability Evaluation Methods for Shared Workspaces'. In: *Proceedings IEEE 9th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*. Proceedings IEEE 9th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. WET ICE '00. Gaithersburg, MD, USA: IEEE, –June 16, 2000, pp. 98–103. ISBN: 0-7695-0798-0. DOI: 10.1109/ENABL.2000.883711.

[104] Patrick Haggard. 'Conscious Intention and Motor Cognition'. In: *Trends in Cognitive Sciences* 9.6 (2005), pp. 290–295. ISSN: 1364-6613. DOI: 10.1016/j.tics.2005.04.012.

[105] Patrick Haggard, Sam Clark, and Jeri Kalogeras. 'Voluntary Action and Conscious Awareness'. In: *Nature Neuroscience* 5 (2002), p. 382. DOI: 10.1038/nn827.

[106] Robert M. Haralick. 'Ridges and Valleys on Digital Images'. In: *Computer Vision, Graphics, and Image Processing* 22.1 (1983), pp. 28–38. ISSN: 0734-189X. DOI: 10.1016/0734-189X(83)90094-4.

[107] Chris Harris and Mike Stephens. 'A Combined Corner and Edge Detector'. In: *Proceedings of Fourth Alvey Vision Conference*. 1988, pp. 147–151.

[108] Sandra G. Hart. 'NASA-Task Load Index (NASA-TLX) - 20 Years Later'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (2006), pp. 904–908. DOI: 10.1177/154193120605000909.

[109] Sandra G. Hart and Lowell E. Staveland. 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research'. In: *Human Mental Workload*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Advances in Psychology. North-Holland, 1988, pp. 139–183.

[110] Li-wei He, Michael F. Cohen, and David H. Salesin. 'The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing'. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 217–224. ISBN: 0-89791-746-4. DOI: 10.1145/237170.237259.

[111] Tom Hewett, Mary Czerwinkski, Michael Terry, Jay Nunamaker, Linda Candy, Bill Kules, and Elisabeth Sylvan. 'Creativity Support Tool Evaluation Methods and Metrics'. In: *NSF Workshop Report on Creativity Support Tools*. Washington, DC, USA, 2005.

[112] William E. Hick. 'On the Rate of Gain of Information'. In: *Quarterly Journal of Experimental Psychology* 4.1 (1952), pp. 11–26.

[113] Brian Hilburn, Peter G. Jorna, Evan A. Byrne, and Raja Parasuraman. 'The Effect of Adaptive Air Traffic Control Decision Aiding on Controller Mental Workload'. In: *Human-Automation Interaction: Research and Practice* (1997), pp. 84–91.

[114] Brian Hilburn, Robert Molloy, Dick Wong, and Raja Parasuraman. 'Operator Versus Computer Control of Adaptive Automation'. In: *Proceedings of the 7th International Symposium on Aviation Psychology*. Ed. by R. S. Jensen and D. Neumeister. Columbus, OH, USA, 1993, pp. 161–166.

[115] Reiner Hoeger, Angelos Amditis, Martin Kunert, Alfred Hoess, Hans-Peter Krueger, Arne Bartels, and Achim Beutner. 'Highly Automated Vehicles for Intelligent Transport: Haveit Approach'. In: ITS World Congress. New York, NY, USA, 2008.

[116] **Axel Hoesl**. *Katastrophenschutz Am Curve*. With envis precisely, Katastrophenschutz des Bayerischen Staatsministerium des Inneren, Integrierte Leitstelle Ingolstadt (in German). Project Thesis. Munich, Germany: LMU Munich, 2012.

[117] **Axel Hoesl**. *Promoting Stress-Resistance - Catastrophe-Management on a Curved Display*. With envis precisely, Katastrophenschutz des Bayerischen Staatsministerium des Inneren, Integrierte Leitstelle Ingolstadt. Diploma Thesis. Munich, Germany: LMU Munich, 2013.

[118] **Axel Hoesl**, Mujo Alic, and Andreas Butz. 'On the Effects of Progressive Reduction as Adaptation Strategy for a Camera-Based Cinematographic User Interface'. In: *Proceedings of The 16th IFIP TC.13 International Conference on Human-Computer Interaction*. INTERACT '17. Mumbai, India, 2017, pp. 513–522. DOI: 10.1007/978-3-319-67744-6_32.

[119] **Axel Hoesl**, Sarah Aragon Bartsch, Philipp Burgdorf, and Andreas Butz. 'TrackLine: Refining Touch-to-Track Interaction for Camera Motion Control on Mobile Devices'. In: *Printed Proceedings of the 13th European Conference on Visual Media Production*. CVMP '16. London, UK, 2016.

[120] **Axel Hoesl**, Sarah Aragon Bartsch, and Andreas Butz. 'TrackLine: Refining Touch-to-Track Interaction for Camera Motion Control on Mobile Devices'. In: *Proceedings of The 16th IFIP TC.13 International Conference on Human-Computer Interaction*. INTERACT '17. Mumbai, India, 2017, pp. 283–292. DOI: 10.1007/978-3-319-67684-5_17.

[121] **Axel Hoesl** and Andreas Butz. 'Sense of Authorship and Agency in Computational Creativity Support'. In: *Proceedings of the Workshop on Mixed-Initiative Creative Interfaces (MICI'17) in Conjunction with CHI'17*. MICI '17. Denver, CO, USA: ACM, 2017. URL: `http://ceur-ws.org/Vol-1907` (visited on 06/12/2019).

[122] **Axel Hoesl**, Patrick Mörwald, Philipp Burgdorf, Elisabeth Dreßler, and Andreas Butz. 'You've Got the Moves, We've Got the Motion - Understanding and Designing for Cinematographic Camera Motion Control'. In: *Proceedings of The 16th IFIP TC.13 International Conference on Human-Computer Interaction*. INTERACT '17. Mumbai, India, 2017, pp. 523–541. DOI: 10.1007/978-3-319-67744-6_33.

[123] **Axel Hoesl**, Phuong Anh Vu, Christina Rosenmöller, Florian Lehmann, and Andreas Butz. 'Towards an Evaluation Framework: Implicit Evaluation of Sense of Agency in a Creative Continuous Control Task'. In: *Proceedings of the 35th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. CHI LBW '17. Denver, Colorado, USA: ACM, 2017, pp. 1686–1693. DOI: 10.1145/3027063.3053072.

[124] **Axel Hoesl**, Julie Wagner, and Andreas Butz. 'Delegation Impossible? Towards Novel Interfaces for Camera Motion'. In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '15. ACM, 2015, pp. 1729–1734. DOI: 10.1145/2702613.2732904.

[125] Berthold K.P. Horn and Brian G. Schunck. 'Determining Optical Flow'. In: *Artificial Intelligence* 17.1 (1981), pp. 185–203. ISSN: 0004-3702. DOI: 10.1016/0004-3702(81)90024-2.

[126] Dries Hulens, Toon Goedemé, and Tom Rumes. 'Autonomous Lecture Recording with a PTZ Camera While Complying with Cinematographic Rules'. In: *Proceedings of the 2014 Canadian Conference on Computer and Robot Vision*. CRV'14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 371–377. ISBN: 978-1-4799-4337-1. DOI: 10.1109/CRV.2014.57.

[127] Toshiyuki Inagaki. 'Adaptive Automation: Sharing and Trading of Control'. In: *Handbook of Cognitive Task Design* 8 (2003), pp. 147–169.

[128] Wolfgang Iser. *Das Fiktive Und Das Imaginäre: Perspektiven Literarischer Anthropologie*. 6th ed. (in German). Suhrkamp, 1993. ISBN: 978-3-518-28701-9.

[129] Jacek Jankowski and Martin Hachet. 'A Survey of Interaction Techniques for Interactive 3D Environments'. In: *Eurographics 2013 - STAR*. Girona, Spain, 2013. URL: https://hal.inria.fr/hal-00789413 (visited on 06/12/2019).

[130] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 'Measuring and Defining the Experience of Immersion in Games'. In: *International Journal of Human-Computer Studies* 66.9 (2008), pp. 641–661. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2008.04.004.

[131] Ian Johnston. *The Mozi: A Complete Translation*. Translations from the Asian Classics. New York, NY, USA: Columbia University Press, 2010. 1032 pp. ISBN: 978-0-231-15240-2.

[132] Stephen D. Jones, Claus Andresen, and James L. Crowley. 'Appearance Based Process for Visual Navigation'. In: *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE/RSJ International Conference on Intelligent Robots and Systems. Vol. 2. IROS '97. Grenoble, France: IEEE, 1997, pp. 551–557. DOI: 10.1109/IROS.1997.655066.

[133] Bhautik Joshi, Kristen Stewart, and David Shapiro. 'Bringing Impressionism to Life with Neural Style Transfer in Come Swim'. In: *Proceedings of the ACM SIGGRAPH Digital Production Symposium*. Los Angeles, California: ACM, 2017, pp. 1–5. ISBN: 978-1-4503-5102-7.

[134] Simon Julier, Marco Lanzagorta, Yohan Baillot, Lawrence Rosenblum, Steven Feiner, Tobias Hollerer, and Sabrina Sestito. 'Information Filtering for Mobile Augmented Reality'. In: *Pro-*

ceedings IEEE and ACM International Symposium on Augmented Reality. ISAR '00. Munich, Germany: IEEE, 2000, pp. 3–11. ISBN: 0-7695-0846-4. DOI: 10.1109/ISAR.2000.880917.

[135] David B. Kaber. 'Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation'. In: *Journal of Cognitive Engineering and Decision Making* 12.1 (2017), pp. 7–24. ISSN: 1555-3434. DOI: 10.1177/1555343417737203.

[136] David B. Kaber and Mica R. Endsley. 'Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety'. In: *Process Safety Progress* 16.3 (1997), pp. 126–131. ISSN: 1547-5913. DOI: 10.1002/prs.680160304.

[137] David B. Kaber and Mica R. Endsley. 'The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task'. In: *Theoretical Issues in Ergonomics Science* 5.2 (2004), pp. 113–153. DOI: 10.1080/1463922021000054335.

[138] David B. Kaber and Jennifer M. Riley. 'Adaptive Automation of a Dynamic Control Task Based on Workload Assessment Through a Secondary Monitoring Task'. In: *Automation Technology and Human Performance: Current Research and Trends* (1999), pp. 129–133.

[139] Maria Karam. *A Framework for Research and Design of Gesture-Based Human-Computer Interactions*. Doctoral. University of Southampton, 2006. URL: `https://eprints.soton.ac.uk/263149` (visited on 06/12/2019).

[140] Daiichiro Kato, Tetsuo Katsuura, and Hideo Koyama. 'Automatic Control of a Robot Camera for Broadcasting Based on Cameramen's Techniques and Subjective Evaluation and Analysis of Reproduced Images'. In: *Journal of Physiological Anthropology and Applied Human Science* 19.2 (2000), pp. 61–71. DOI: 10.2114/jpa.19.61.

[141] Daishiro Kato, Mitsuho Yamada, K. Abe, A. Ishikawa, K. Ishiyama, and M. Obata. 'Analysis of the Camerawork of Broadcasting Cameramen'. In: *SMPTE Journal* 106.2 (1997), pp. 108–116. ISSN: 0036-1682. DOI: 10.5594/J15810.

[142] Mohamed Khamis, **Axel Hoesl**, Andreas Bulling, and Florian Alt. 'EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays'. In: *Proceedings of 30th ACM User Interface Software and Technology Symposium*. UIST '17. Québec City, Québec, Canada: ACM, 2017, pp. 155–166. DOI: 10.1145/3126594.3126630.

[143] Nabeel Y. Khan, Brendan McCane, and Geoff Wyvill. 'SIFT and SURF Performance Evaluation against Various Image Deformations on Benchmark Dataset'. In: *2011 International Conference on Digital Image Computing: Techniques and Applications*. 2011 International

Conference on Digital Image Computing: Techniques and Applications. Noosa, QLD, Australia: IEEE, 2011, pp. 501–506. ISBN: 978-1-4577-2006-2. DOI: 10.1109/DICTA.2011.90.

[144] Jan J. Koenderink and Andrea J. van Doorn. 'Two-Plus-One-Dimensional Differential Geometry'. In: *Volume Image Processing*. VIP '93 15.5 (1994), pp. 439–443. ISSN: 0167-8655. DOI: 10.1016/0167-8655(94)90134-1.

[145] Devpriya Kumar and Narayanan Srinivasan. 'Hierarchical Control and Sense of Agency: Differential Effects of Control on Implicit and Explicit Measures of Agency'. In: *Proceedings of 35th Annual Meeting of the Cognitive Science Society*. 2013.

[146] Nikolaos Kyriazis and Antonis A. Argyros. 'Scalable 3D Tracking of Multiple Interacting Objects'. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. CVPR '14. IEEE, 2014, pp. 3430–3437. DOI: 10.1109/CVPR.2014.438.

[147] Daniel Lakens. 'Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for T-Tests and Anovas'. In: *Frontiers in Psychology* 4 (2013), p. 863. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00863.

[148] Gierad Laput, Robert Xiao, and Chris Harrison. 'ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers'. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST '16. Tokyo, Japan: ACM, 2016, pp. 321–333. ISBN: 978-1-4503-4189-9. DOI: 10.1145/2984511.2984582.

[149] Anatole Lécuyer, Jean-Marie Burkhardt, Jean-Marie Henaff, and Stéphane Donikian. 'Camera Motions Improve the Sensation of Walking in Virtual Environments'. In: *IEEE Virtual Reality Conference*. IEEE Virtual Reality Conference. VR '06. Alexandria, VA, USA: IEEE, 2006, pp. 11–18. ISBN: 1-4244-0224-7. DOI: 10.1109/VR.2006.31.

[150] Ahreum Lee, Kiburm Song, Hokyoung Blake Ryu, Jieun Kim, and Gyuhyun Kwon. 'Fingerstroke Time Estimates for Touchscreen-Based Mobile Gaming Interaction'. In: *Human Movement Science* 44 (2015), pp. 211–224. ISSN: 0167-9457. DOI: 10.1016/j.humov.2015.09.003.

[151] G. Julian Lepinski, Tovi Grossman, and George Fitzmaurice. 'The Design and Evaluation of Multitouch Marking Menus'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA: ACM, 2010, pp. 2233–2242. ISBN: 978-1-60558-929-9. DOI: 10.1145/1753326.1753663.

[152] Tsai-Yen Li and Chung-Chiang Cheng. 'Real-Time Camera Planning for Navigation in Virtual Environments'. In: *Smart Graphics*. Ed. by Andreas Butz, Brian Fisher, Antonio Krüger, Patrick Olivier, and Marc Christie. Springer Berlin Heidelberg, 2008, pp. 118–129. ISBN: 978-3-540-85412-8. DOI: 10.1007/978-3-540-85412-8_11.

[153] Bojian Liang and Nick Pears. 'Visual Navigation Using Planar Homographies'. In: *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*. Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292). Vol. 1. Washington, DC, USA: IEEE, 2002, 205–210 vol.1. ISBN: 0-7803-7272-7. DOI: 10.1109/ROBOT.2002.1013362.

[154] Benjamin Libet. 'Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action'. In: *Behavioral and Brain Sciences* 8.4 (1985), pp. 529–539. ISSN: 0140-525X. DOI: 10.1017/S0140525X00044903.

[155] Benjamin Libet, Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl. 'Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act'. In: *Brain* 106.3 (1983), p. 623. DOI: 10.1093/brain/106.3.623.

[156] Hannah Limerick, David Coyle, and James W. Moore. 'The Experience of Agency in Human-Computer Interactions: A Review'. In: *Frontiers in Human Neuroscience* 8 (2014), p. 643. DOI: 10.3389/fnhum.2014.00643.

[157] Hannah Limerick, James W. Moore, and David Coyle. 'Empirical Evidence for a Diminished Sense of Agency in Speech Interfaces'. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, 2015, pp. 3967–3970. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702379.

[158] David C. Lindberg. *Theories of Vision from Al-Kindi to Kepler*. Chicago History of Science and Medicine. University of Chicago Press, 1976. 331 pp. ISBN: 0-226-48235-9.

[159] Tony Lindeberg. 'Scale-Space'. In: *Wiley Encyclopedia of Computer Science and Engineering*. American Cancer Society, 2008, pp. 2495–2504. ISBN: 978-0-470-05011-8. DOI: 10.1002/9780470050118.ecse609.

[160] Tony Lindeberg. 'Spatio-Temporal Scale Selection in Video Data'. In: *Journal of Mathematical Imaging and Vision* 60.4 (2018), pp. 525–562. ISSN: 1573-7683. DOI: 10.1007/s10851-017-0766-9.

[161] Christophe Lino, Marc Christie, Roberto Ranon, and William Bares. 'The Director's Lens: An Intelligent Assistant for Virtual Cinematography'. In: *Proceedings of the 19th ACM International Conference on Multimedia*. MM '11. Scottsdale, AZ, USA: ACM, 2011, pp. 323–332. ISBN: 978-1-4503-0616-4. DOI: 10.1145/2072298.2072341.

[162] Qiong Liu, Yong Rui, Anoop Gupta, and J. J. Cadiz. 'Automating Camera Management for Lecture Room Environments'. In: *Proceedings of the SIGCHI Conference on Human Factors*

in Computing Systems. CHI '01. Seattle, WA, USA: ACM, 2001, pp. 442–449. ISBN: 1-58113-327-8. DOI: 10.1145/365024.365310.

[163] Shuaicheng Liu, Lu Yuan, Ping Tan, and Jian Sun. 'SteadyFlow: Spatially Smooth Optical Flow for Video Stabilization'. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014 IEEE Conference on Computer Vision and Pattern Recognition. CVPR '14. Columbus, OH, USA: IEEE, 2014, pp. 4209–4216. ISBN: 978-1-4799-5118-5. DOI: 10.1109/CVPR.2014.536.

[164] Matthew M. Loper, Naureen Mahmood, and Michael J. Black. 'MoSh: Motion and Shape Capture from Sparse Markers'. In: *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)* 33.6 (2014), 220:1–220:13. DOI: 10.1145/2661229.2661273.

[165] David G. Lowe. 'Object Recognition from Local Scale-Invariant Features'. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Proceedings of the Seventh IEEE International Conference on Computer Vision. Vol. 2. Kerkyra, Greece: IEEE, 1999, pp. 1150–1157. ISBN: 0-7695-0164-8. DOI: 10.1109/ICCV.1999.790410.

[166] Bruce D. Lucas and Takeo Kanade. 'An Iterative Image Registration Technique with an Application to Stereo Vision'. In: *Proceedings of the 7th Joint Conference on Aritficial Intelligence*. IJCAI '81 (1981), pp. 674–679.

[167] Wendy E. Mackay and Anne-Laure Fayard. 'HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines'. In: *Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. ACM, 1997, pp. 223–234. DOI: 10.1145/263552.263612.

[168] Jock D. Mackinlay, Stuart K. Card, and George G. Robertson. 'Rapid Controlled Movement Through a Virtual 3D Workspace'. In: *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '90. Dallas, TX, USA: ACM, 1990, pp. 171–176. ISBN: 0-89791-344-2. DOI: 10.1145/97879.97898.

[169] Norman H. Mackworth and H. W. Sinaiko. 'Researches on the Measurement of Human Performance'. In: *Selected Papers on Human Factors in the Design and Use of Control Systems* (1950). publisher=Dover Publications;place=New York, NY, USA, pp. 174–331.

[170] Jost J. Marchesi. *Handbuch Der Fotografie*. 1st ed. Vol. 1. 3 vols. (in German). Schaffhausen, Switzerland: Verlag Photographie AG, 1993. ISBN: 3-7231-0024-4.

[171] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. 'Rethinking 'Multi-User': An In-the-Wild Study of How Groups Approach a Walk-up-and-Use Tabletop Interface'. In: *Proceedings of the SIGCHI Conference on Human Factors*

in Computing Systems. CHI '11. New York, NY, USA: ACM, 2011, pp. 3033–3042. DOI: 10.1145/1978942.1979392.

[172] Yoshio Matsumoto, Masayuki Inaba, and Hirochika Inoue. 'Visual Navigation Using View-Sequenced Route Representation'. In: *Proceedings of IEEE International Conference on Robotics and Automation*. Proceedings of IEEE International Conference on Robotics and Automation. Vol. 1. Minneapolis, MN, USA: IEEE, 1996, pp. 83–88. ISBN: 0-7803-2988-0. DOI: 10.1109/ROBOT.1996.503577.

[173] Aditya Mavlankar, Piyush Agrawal, Derek Pang, Sherif Halawa, Ngai-Man Cheung, and Bernd Girod. 'An Interactive Region-of-Interest Video Streaming System for Online Lecture Viewing'. In: *2010 18th International Packet Video Workshop*. 2010 18th International Packet Video Workshop. Hong Kong, China: IEEE, 2010, pp. 64–71. ISBN: 978-1-4244-9522-1. DOI: 10.1109/PV.2010.5706821.

[174] Thomas B. McCord and James A. Westphal. 'Two-Dimensional Silicon Vidicon Astronomical Photometer'. In: *Appl. Opt.* 11.3 (1972), pp. 522–526. DOI: 10.1364/AO.11.000522.

[175] Jon McCormack. 'Creative Ecosystems'. In: *Computers and Creativity*. Ed. by Jon McCormack and Mark d'Inverno. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 39–60. ISBN: 978-3-642-31727-9. DOI: 10.1007/978-3-642-31727-9_2.

[176] James C. McGuire, John A. Zich, Richard T. Goins, Jeffrey B. Erickson, John P. Dwyer, William J. Cody, and William B. Rouse. *Functional Decomposition of the Commercial Flight Domain for Function Allocation*. NASA Contractor Report 4374. Langley, VA, USA, 1990.

[177] Gustavo Mercado. *The Filmmaker's Eye: Learning (and Breaking) the Rules of Cinematic Composition*. Burlington, MA, USA: Taylor & Francis Ltd., 2010. 192 pp. ISBN: 978-0-240-81217-5.

[178] F. Mertes and L. Jenney. 'Automation Applications in an Advanced Air Traffic Management System'. In: *Methodology for Man-Machine Task Allocation*. Vol. 3. (Rep. No. DOT-TSC-OST-74-14-III). McLean, VA, USA, 1974, p. 220.

[179] Christopher A. Miller and Raja Parasuraman. 'Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49.1 (2007), pp. 57–75. DOI: 10.1518/001872007779598037.

[180] Christopher A. Miller, Michael Pelican, and Robert Goldman. 'Tasking" Interfaces to Keep the Operator in Control'. In: *Proceedings of the Fifth Annual Symposium on Human Interaction with Complex Systems*. 2000, pp. 87–91.

[181] Thomas B. Moeslund and Erik Granum. 'A Survey of Computer Vision-Based Human Motion Capture'. In: *Computer Vision and Image Understanding* 81.3 (2001), pp. 231–268. ISSN: 1077-3142. DOI: 10.1006/cviu.2000.0897.

[182] Valiallah M. Monajjemi, Jens Wawerla, Richard Vaughan, and Greg Mori. 'HRI in the Sky: Creating and Commanding Teams of Uavs with a Vision-Mediated Gestural Interface'. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS '13. 2013, pp. 617–623. ISBN: 978-1-4673-6358-7. DOI: 10.1109/IROS.2013.6696415.

[183] James W. Moore and Paul C. Fletcher. 'Sense of Agency in Health and Disease: A Review of Cue Integration Approaches'. In: *Beyond the Comparator Model* 21.1 (2012), pp. 59–68. ISSN: 1053-8100. DOI: 10.1016/j.concog.2011.08.010.

[184] James W. Moore and Patrick Haggard. 'Awareness of Action: Inference and Prediction'. In: *Consciousness and Cognition* 17.1 (2008), pp. 136–144. ISSN: 1053-8100. DOI: 10.1016/j.concog.2006.12.004.

[185] James W. Moore and Sukhvinder S. Obhi. 'Intentional Binding and the Sense of Agency: A Review'. In: *Beyond the Comparator Model* 21.1 (2012), pp. 546–561. ISSN: 1053-8100. DOI: 10.1016/j.concog.2011.12.002.

[186] James W. Moore, Daniel M. Wegner, and Patrick Haggard. 'Modulating the Sense of Agency with External Cues'. In: *Consciousness and Cognition* 18.4 (2009), pp. 1056–1064. ISSN: 1053-8100. DOI: 10.1016/j.concog.2009.05.004.

[187] Neville Moray, Toshiyuki Inagaki, and Makoto Itoh. 'Adaptive Automation, Trust, and Self-Confidence in Fault Management of Time-Critical Tasks'. In: *Journal of Experimental Psychology: Applied* 6.1 (2000), pp. 44–58. DOI: 10.1037//0278-7393.6.1.44.

[188] Masahiro Mori. 'The Uncanny Valley'. In: *Energy* 7.4 (1970), pp. 33–35.

[189] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 'The Uncanny Valley [From the Field]'. In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100. ISSN: 1070-9932. DOI: 10.1109/MRA.2012.2192811.

[190] William F. Moroney, David W. Biers, F. Thomas Eggemeier, and Jennifer A. Mitchell. 'A Comparison of Two Scoring Procedures with the NASA Task Load Index in a Simulated Flight Task'. In: *Proceedings of the IEEE 1992 National Aerospace and Electronics Conference*. Proceedings of the IEEE 1992 National Aerospace and Electronics Conference. NAECON '92. Dayton, OH, USA: IEEE, 1992, 734–740 vol.2. DOI: 10.1109/NAECON.1992.220513.

[191] Jawad Nagi, Alessandro Giusti, Gianni A. Di Caro, and Luca M. Gambardella. 'Human Control of UAVs Using Face Pose Estimates and Hand Gestures'. In: *Proceedings of the*

2014 ACM/IEEE International Conference on Human-Robot Interaction. Bielefeld, Germany: ACM, 2014, pp. 252–253. ISBN: 978-1-4503-2658-2.

[192] Peter Nemenyi. *Distribution-Free Multiple Comparisons*. Ph.D. thesis. Princeton, NJ, USA: Princeton University, 1963. 254 pp.

[193] Jakob Nielsen. 'Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier'. In: *Cost-Justifying Usability*. Ed. by R. G. Bias and D. J. Mayhew. Orlando, FL, USA: Academic Press, 1994, pp. 245–272. ISBN: 0-12-095810-4.

[194] Jakob I. Nielsen. *Camera Movement in Narrative Cinema: Towards a Taxonomy of Functions*. Århus: Århus Universitet, 2007. ISBN: 87-92052-25-8.

[195] Annu-Maaria Nivala and Tiina L. Sarjakoski. 'User Aspects of Adaptive Visualization for Mobile Maps'. In: *Cartography and Geographic Information Science* 34.4 (2007), pp. 275–284. ISSN: 1523-0406. DOI: 10.1559/152304007782382954.

[196] Donald A. Norman. 'Cognitive Engineering'. In: *User Centered System Design* (1986). Ed. by D. A. Norman and S. W. Draper, pp. 31–61.

[197] Donald A. Norman. *The Design of Everyday Things: Revised and Expanded Edition*. 2nd ed. New York, NY, USA: Basic Books, 2013. 368 pp. ISBN: 978-0-465-05065-9.

[198] Donald A. Norman. *The Psychology of Everyday Things*. United States, 1988. ISBN: 978-0-465-06710-7.

[199] Ian Oakley and Jun-Seok Park. 'Designing Eyes-Free Interaction'. In: *Haptic and Audio Interaction Design*. Ed. by Ian Oakley and Stephen Brewster. HAID '07. Springer Berlin Heidelberg, 2007, pp. 121–132. ISBN: 978-3-540-76702-2. DOI: 10.1007/978-3-540-76702-2_13.

[200] Mohammad Obaid, Felix Kistler, Gabrielė Kasparavǐciūtė, Asim Evren Yantąc, and Morten Fjeld. 'How Would You Gesture Navigate a Drone?: A User-Centered Approach to Control a Drone'. In: *Proceedings of the 20th International Academic Mindtrek Conference*. AcademicMindtrek '16. Tampere, Finland: ACM, 2016, pp. 113–121. ISBN: 978-1-4503-4367-1. DOI: 10.1145/2994310.2994348.

[201] Aníbal Ollero, Joaquín Ferruz, Fernando Caballero, Sebastián Hurtado, and Luis Merino. 'Motion Compensation and Object Detection for Autonomous Helicopter Visual Navigation in the COMETS System'. In: *IEEE International Conference on Robotics and Automation*. IEEE International Conference on Robotics and Automation. Vol. 1. ICRA '04. New Orleans, LA, USA: IEEE, 2004, 19–24 Vol.1. ISBN: 0-7803-8232-3. DOI: 10.1109/ROBOT.2004.1307123.

[202] Raja Parasuraman. 'Designing Automation for Human Use: Empirical Studies and Quantitative Models'. In: *Ergonomics* 43.7 (2000), pp. 931–951. DOI: 10.1080/001401300409125.

[203] Raja Parasuraman. ‘Effects of Adaptive Function Allocation on Human Performance’. In: *Human Factors and Advanced Aviation Technologies* (1993), pp. 147–157.

[204] Raja Parasuraman, Toufik Bahri, John E. Deaton, Jeffrey G. Morrison, and Michael Barnes. *Theory and Design of Adaptive Automation in Aviation Systems*. NAWCADWAR-92033-60. Washington, DC, USA: Cognitive Science Laboratory, The Catholic University of America, 1992.

[205] Raja Parasuraman and Mustapha Mouloua, eds. *Automation and Human Performance: Theory and Applications*. Lawrence Erlbaum Associates, 1996. 536 pp. ISBN: 978-0-8058-1616-7.

[206] Raja Parasuraman, Mustapha Mouloua, and Robert Molloy. ‘Effects of Adaptive Task Allocation on Monitoring of Automated Systems’. In: *Human Factors* 38.4 (1996), pp. 665–679. ISSN: 0018-7208. DOI: 10.1518/001872096778827279.

[207] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. ‘A Model for Types and Levels of Human Interaction with Automation’. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 30.3 (2000), pp. 286–297. ISSN: 1083-4427. DOI: 10.1109/3468.844354.

[208] Pekka Parhi, Amy K. Karlson, and Benjamin B. Bederson. ‘Target Size Study for One-Handed Thumb Use on Small Touchscreen Devices’. In: *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI ’06. Helsinki, Finland: ACM, 2006, pp. 203–210. ISBN: 1-59593-390-5. DOI: 10.1145/1152215.1152260.

[209] Nick Pears and Bojian Liang. ‘Ground Plane Segmentation for Mobile Robot Visual Navigation’. In: *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180)*. Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180). Vol. 3. Maui, HI, USA: IEEE, 2001, 1513–1518 vol.3. ISBN: 0-7803-6612-3. DOI: 10.1109/IROS.2001.977194.

[210] Carlo Pedretti. *The Codex Atlanticus of Leonardo Da Vinci: A Catalogue of Its Newly Restored Sheets*. New York, NY, USA: Johnson Reprint: Harcourt, Brace, Jovanovich, 1979.

[211] David N. Perkins. ‘Creativity: Beyond the Darwinian Paradigm’. In: *Dimensions of Creativity* (1994). Ed. by Margaret A. Boden, pp. 119–142.

[212] Ekaterina Peshkova, Martin Hitz, and David Ahlström. ‘Exploring User-Defined Gestures and Voice Commands to Control an Unmanned Aerial Vehicle’. In: *Intelligent Technologies for Interactive Entertainment*. Ed. by Ronald Poppe, John-Jules Meyer, Remco Veltkamp, and

Mehdi Dastani. INTETAIN '16. Utrecht, The Netherlands: Springer International Publishing, 2017, pp. 47–62. ISBN: 978-3-319-49616-0.

[213] Bastian Pfleging, Drea K. Fekety, Albrecht Schmidt, and Andrew L. Kun. 'A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions'. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, CA, USA: ACM, 2016, pp. 5776–5788. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858117.

[214] Claudio S. Pinhanez and Pentland, Alex P. *Intelligent Studios: Using Computer Vision to Control TV Cameras*. 324. Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology, 1995.

[215] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J. Black. 'ClothCap: Seamless 4D Clothing Capture and Retargeting'. In: *ACM Transactions on Graphics, (Proc. SIGGRAPH)* 36.4 (2017). DOI: 10.1145/3072959.3073711.

[216] Alan T. Pope, Edward H. Bogart, and Debbie S. Bartolome. 'Biocybernetic System Evaluates Indices of Operator Engagement in Automated Task'. In: *Biological Psychology* 40.1–2 (1995), pp. 187–195.

[217] Lawrence J. Prinzel, Frederick G. Freeman, Mark W. Scerbo, Peter J. Mikulka, and Alan T. Pope. 'Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component'. In: *Human Factors* 45.4 (2003), pp. 601–614. ISSN: 0018-7208. DOI: 10.1518/hfes.45.4.601.27092.

[218] Siddharth S. Rautaray and Anupam Agrawal. 'Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey'. In: *Artificial Intelligence Review* 43.1 (2015), pp. 1–54. ISSN: 1573-7462. DOI: 10.1007/s10462-012-9356-9.

[219] Jason L. Reisman, Philip L. Davidson, and Jefferson Y. Han. 'A Screen-Space Formulation for 2D and 3D Direct Manipulation'. In: *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*. UIST '09. Victoria, BC, Canada: ACM, 2009, pp. 69–78. ISBN: 978-1-60558-745-5. DOI: 10.1145/1622176.1622190.

[220] Eric Renner. *Pinhole Photography: From Historic Technique to Digital Application*. 4th ed. Alternative Process Photography. Taylor & Francis, 2012. ISBN: 978-1-136-09533-7.

[221] Mitchel Resnick, Brad Myers, Kumiyo Nakakoji, Ben Shneiderman, Randy Pausch, Ted Selker, and Mike Eisenberg. 'Design Principles for Tools to Support Creative Thinking'. In: *NSF Workshop Report on Creativity Support Tools*. Washington, DC, USA, 2005.

[222] Beverly F. Ronalds. *The Beginnings of Continuous Scientific Recording Using Photography: Sir Francis Ronalds' Contribution*. Vienna, Austria, 2016. URL: `http://www.eshph.org/wp-content/uploads/2016/05/ronalds_camera.pdf` (visited on 06/12/2019).

[223] Ruth Rosenholtz, Nathaniel R. Twarog, Nadja Schinkel-Bielefeld, and Martin Wattenberg. 'An Intuitive Model of Perceptual Grouping for HCI Design'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Boston, MA, USA: ACM, 2009, pp. 1331–1340. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518903.

[224] Edward Rosten and Tom Drummond. 'Machine Learning for High-Speed Corner Detection'. In: *Computer Vision – ECCV 2006*. Ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. Springer Berlin Heidelberg, 2006, pp. 430–443. ISBN: 978-3-540-33833-8. DOI: 10.1007/11744023_34.

[225] William B. Rouse. 'Adaptive Aiding for Human/Computer Control'. In: *Human Factors* 30.4 (1988), pp. 431–443. ISSN: 0018-7208. DOI: 10.1177/001872088803000405.

[226] William B. Rouse. 'Adaptive Allocation of Decision Making Responsibility between Supervisor and Computer'. In: *Monitoring Behavior and Supervisory Control*. Ed. by Thomas B. Sheridan and Gunnar Johannsen. Boston, MA, USA: Springer US, 1976, pp. 295–306. ISBN: 978-1-4684-2523-9. DOI: 10.1007/978-1-4684-2523-9_24.

[227] William B. Rouse, Norman D. Geddes, and Renwick E. Curry. 'An Architecture for Intelligent Interfaces: Outline of an Approach to Supporting Operators of Complex Systems'. In: *Human–Computer Interaction* 3.2 (1987), pp. 87–122. ISSN: 0737-0024. DOI: 10.1207/s15327051hci0302_1.

[228] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 'ORB: An Efficient Alternative to SIFT or SURF'. In: *2011 International Conference on Computer Vision*. 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011, pp. 2564–2571. ISBN: 978-1-4577-1101-5. DOI: 10.1109/ICCV.2011.6126544.

[229] Abdelhamid I. Sabra. *The Optics of Ibn Al-Haytham, Books I–II–III: On Direct Vision*. Vol. 40. Studies of the Warburg Institute. Leeds, Great Britain: The Warburg Institute, 1989. ISBN: 0-85481-072-2.

[230] Abdelhamid I. Sabra. 'The Optics of Ibn Al-Haytham. Edition of the Arabic Text of Books IV–V: On Reflection and Images Seen by Reflection'. In: (2002).

[231] Lothar Sachs. *Angewandte Statistik: Anwendung Statistischer Methoden*. 8th ed. DOI=10.1007/978-3-662-05746-9. Springer Berlin Heidelberg, 1997. 885 pp. ISBN: 978-3-662-05746-9.

[232] Cunka B. Sanokho, Clément Desoche, Billal Merabti, Tsai-Yen Li, and Marc Christie. 'Camera Motion Graphs'. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '14. Copenhagen, Denmark: Eurographics Association, 2014, pp. 177–188. URL: `http://dl.acm.org/citation.cfm?id=2849517.2849546` (visited on 06/12/2019).

[233] Mark W. Scerbo. ‘Theoretical Perspectives on Adaptive Automation’. In: *Automation and Human Performance*. Ed. by Mustapha Mouloua. 1st ed. New York, NY, USA: Routledge, 1996, pp. 57–84. ISBN: 978-1-351-46505-2.

[234] Paul Schatzkin. *The Boy Who Invented Television: A Story of Inspiration, Persistence, and Quiet Passion*. Silver Spring, MD, USA: TeamCom Books, 2002. ISBN: 1-928791-30-1.

[235] C. James Scheirer, William S. Ray, and Nathan Hare. ‘The Analysis of Ranked Data Derived from Completely Randomized Factorial Designs’. In: *Biometrics* 32.2 (1976), pp. 429–434. ISSN: 0006341X, 15410420. DOI: 10.2307/2529511.

[236] Mario J. G. Schuivens. *Die Historische Entwicklung Der Cockpit-Instrumentierungen von Verkehrsflugzeugen*. (in German). Ph.D. thesis. Technische Universität München, 2015. URL: `https://mediatum.ub.tum.de/doc/1245318/file.pdf` (visited on 06/12/2019).

[237] Paul Schutte, Kenneth Goodrich, and Ralph Williams. ‘Synergistic Allocation of Flight Expertise on the Flight Deck (SAFEdeck): A Design Concept to Combat Mode Confusion, Complacency, and Skill Loss in the Flight Deck’. In: *Advances in Human Aspects of Transportation*. Ed. by Neville A. Stanton, Steven Landry, Giuseppe Di Bucchianico, and Andrea Vallicelli. Springer International Publishing, 2017, pp. 899–911. ISBN: 978-3-319-41682-3. DOI: 10.1007/978-3-319-41682-3_74.

[238] Stephen Se, David G. Lowe, and James J. Little. ‘Vision-Based Global Localization and Mapping for Mobile Robots’. In: *IEEE Transactions on Robotics* 21.3 (2005), pp. 364–375. ISSN: 1552-3098. DOI: 10.1109/TRO.2004.839228.

[239] Doree Duncan Seligmann. *Interactive Intent-Based Illustration: A Visual Language for Three-Dimensional Worlds*. UMI Order No. GAX94-12841. New York, NY, USA: Columbia University, 1993.

[240] Rajeev Sharma and Seth Hutchinson. ‘On the Observability of Robot Motion Under Active Camera Control’. In: *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*. Proceedings of the 1994 IEEE International Conference on Robotics and Automation. Vol. 1. San Diego, CA, USA: IEEE, 1994, pp. 162–167. DOI: 10.1109/ROBOT.1994.350994.

[241] Thomas B. Sheridan. *Humans and Automation: System Design and Research Issues*. Vol. 3. System Engineering and Management: HFES Issues in Human Factors and Ergonomics Series. Hoboken, NJ, USA: Wiley-Interscience, 2002. 280 pp. ISBN: 978-0-471-23428-9.

[242] Thomas B. Sheridan. ‘Supervisory Control’. In: *Handbook of Human Factors*. Ed. by G. Salvendy. New York, NY, USA: Wiley, 1987, pp. 1244–1268.

[243] Thomas B. Sheridan and William L. Verplank. *Human and Computer Control of Undersea Teleoperators*. DTIC Document, 1978.

[244] Jianbo Shi and Carlo Tomasi. 'Good Features to Track'. In: *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. CVPR '94. Seattle, WA, USA: IEEE, 1994, pp. 593–600. ISBN: 0-8186-5825-8. DOI: 10.1109/CVPR.1994.323794.

[245] Ben Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing, 1986. ISBN: 0-201-16505-8.

[246] Blake Snyder. *Save the Cat: The Last Book on Screenwriting You'll Ever Need*. Saline, MI, USA: Michael Wiese Productions, 2005. 195 pp. ISBN: 978-1-932907-00-1.

[247] Mandyam V. Srinivasan, Saul Thurrowgood, and Dean Soccol. 'An Optical System for Guidance of Terrain Following in UAVs'. In: *2006 IEEE International Conference on Video and Signal Based Surveillance*. 2006 IEEE International Conference on Video and Signal Based Surveillance. AVSS '06. Sydney, Australia: IEEE, 2006, pp. 51–51. ISBN: 0-7695-2688-8. DOI: 10.1109/AVSS.2006.23.

[248] Mandyam V. Srinivasan, Shaowu W. Zhang, Javaan S. Chahl, Gert Stange, and Matt Garratt. 'An Overview of Insect-Inspired Guidance for Application in Ground and Airborne Platforms'. In: *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 218.6 (2004), pp. 375–388. ISSN: 0954-4100. DOI: 10.1243/0954410042794966.

[249] Rares Stanciu and Paul Y. Oh. 'Designing Visually Servoed Tracking to Augment Camera Teleoperators'. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vol. 1. IRDS '02. Lausanne, Switzerland: IEEE, 2002, pp. 342–347. DOI: 10.1109/IRDS.2002.1041412.

[250] Rares Stanciu and Paul Y. Oh. 'Feedforward-Output Tracking Regulation Control for Human-in-the-Loop Camera Systems'. In: *Proceedings of the 2005, American Control Conference*. Vol. 5. ACC '05. Portland, OR, USA: IEEE, 2005, pp. 3676–3681. DOI: 10.1109/ACC.2005.1470546.

[251] Giota Stratou and Louis-Philippe Morency. 'MultiSense: Context-Aware Nonverbal Behavior Analysis Framework: A Psychological Distress Use Case'. In: *IEEE Transactions on Affective Computing* 8.2 (2017), pp. 190–203. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2016.2614300.

[252] Anselm Strauss and Juliet M. Corbin. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Thousand Oaks, CA, USA: Sage Publications, Inc, 1990.

[253] Alan D. Swain and Henry E. Guttmann. *Handbook of Human-Reliability Analysis with Emphasis on Nuclear Power Plant Applications*. Final Report. Albuquerque, NM, USA: Sandia National Labs, 1983.

[254] Matthis Synofzik, Gottfried Vosgerau, and Albert Newen. 'Beyond the Comparator Model: A Multifactorial Two-Step Account of Agency'. In: *Consciousness and Cognition* 17.1 (2008), pp. 219–239. DOI: 10.1016/j.concog.2007.03.010.

[255] Desney S. Tan, George G. Robertson, and Mary Czerwinski. 'Exploring 3D Navigation: Combining Speed-Coupled Flying with Orbiting'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '01. ACM. Seattle, WA, USA, 2001, pp. 418–425. DOI: 10.1145/365024.365307.

[256] Sarah Tausch, Doris Hausen, Ismail Kosan, Andrey Raltchev, and Heinrich Hussmann. 'Groupgarden: Supporting Brainstorming Through a Metaphorical Group Mirror on Table or Wall'. In: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. NordiCHI '15. Helsinki, Finland: ACM, 2014, pp. 541–550. ISBN: 978-1-4503-2542-4.

[257] Richard M. Taylor. 'Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design'. In: *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations*. Seuilly-sur Seine, France, 1990, pp. 3/1–3/17.

[258] Michael Terry and Elizabeth D. Mynatt. 'Side Views: Persistent, On-Demand Previews for Open-Ended Tasks'. In: *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*. UIST '02. Paris, France: ACM, 2002, pp. 71–80. DOI: 10.1145/571985.571996.

[259] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Niessner. 'Face2Face: Real-Time Face Capture and Reenactment of RGB Videos'. In: *The IEEE Conference on Computer Vision and Pattern Recognition*. CVPR '16. Las Vegas, NV, USA: IEEE, 2016, pp. 2387–2395.

[260] Jenifer Tidwell. *Designing Interfaces*. Sebastopol, CA, USA: O'Reilly Media, 2005. 352 pp. ISBN: 978-0-596-00803-1.

[261] Bruce Tognazzini. 'The "Starfire" Video Prototype Project: A Case History'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'94. ACM, 1994, pp. 99–105. DOI: 10.1145/191666.191712.

[262] Emanuele Trucco and Konstantinos Plakas. 'Video Tracking: A Concise Survey'. In: *IEEE Journal of Oceanic Engineering* 31.2 (2006), pp. 520–529. ISSN: 0364-9059. DOI: 10.1109/JOE.2004.839933.

[263] Karl F. Van Orden, Wendy Limbert, Scott Makeig, and Tzyy-Ping Jung. 'Eye Activity Correlates of Workload during a Visuospatial Memory Task'. In: *Human Factors* 43.1 (2001), pp. 111–121. DOI: 10.1518/001872001775992570.

[264] Radu-Daniel Vatavu and Jacob O. Wobbrock. 'Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit'. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, 2015, pp. 1325–1334. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702223.

[265] Barbara Ventarola. 'Fiktionen Als Medien Möglicher Kommunikationen'. In: *Fiktion Im Vergleich Der Künste Und Medien*. Ed. by Anne von Enderwitz and Irina O. Rajewsky. WeltLiteraturen / World Literatures 13. (in German). Leck, Germany: De Gruyter, 2016, pp. 63–96. ISBN: 978-3-11-049864-6.

[266] Joris C. Verster and Thomas Roth. 'Standard Operation Procedures for Conducting the on-the-Road Driving Test, and Measurement of the Standard Deviation of Lateral Position (SDLP)'. In: *International Journal of General Medicine* 4.4 (2011), pp. 359–371. DOI: 10.2147/IJGM.S19639.

[267] Paul Viola and Michael J. Jones. 'Robust Real-Time Face Detection'. In: *International Journal of Computer Vision* 57.2 (2004), pp. 137–154. ISSN: 1573-1405. DOI: 10.1023/B:VISI.0000013087.49260.fb.

[268] Jinjun Wang, Changsheng Xu, Engsiong Chng, Hanqing Lu, and Qi Tian. 'Automatic Composition of Broadcast Sports Video'. In: *Multimedia Systems* 14.4 (2008), pp. 179–193. ISSN: 1432-1882. DOI: 10.1007/s00530-008-0112-6.

[269] Daniel M. Wegner and Thalia Wheatley. 'Apparent Mental Causation: Sources of the Experience of Will'. In: *American Psychologist* 54.7 (1999), p. 480. DOI: 10.1037/0003-066X.54.7.480.

[270] Wen Wen, Atsushi Yamashita, and Hajime Asama. 'The Sense of Agency During Continuous Action: Performance Is More Important Than Action-Feedback Association'. In: *PLoS One* 10.4 (2015), e0125226. DOI: 10.1371/journal.pone.0125226.

[271] Michael J. Wenger. 'On the Rhetorical Contract in Human—Computer Interaction'. In: *Computers in Human Behavior* 7.4 (1991). Locus of Control questionnaire in Appendix D, pp. 245–262. ISSN: 0747-5632. DOI: http://dx.doi.org/10.1016/0747-5632(91)90013-Q.

[272] Tom White and Ian Loh. 'Generating Animations by Sketching in Conceptual Space'. In: *Proceedings of the 8th International Computational Creativity Conference*. ICCC '17. Atlanta, GE, USA: ACC, 2017. ISBN: 978-0-692-89564-1.

[273] 'The Future of Air Traffic Control: Human Operators and Automation'. In: Panel on Human Factors in Air Traffic Control Automation, National Research Council (1998). Ed. by Christopher D. Wickens, Anne S. Mavor, Raja Parasuraman, and James P. McGee, p. 336.

[274] Frank Wilcoxon. 'Individual Comparisons by Ranking Methods'. In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83. ISSN: 00994987. DOI: 10.2307/3001968. JSTOR: 3001968.

[275] Geert Willems, Tinne Tuytelaars, and Luc Van Gool. 'An Efficient Dense and Scale-Invariant Spatio-Temporal Interest Point Detector'. In: *Computer Vision – ECCV 2008*. Ed. by David Forsyth, Philip Torr, and Andrew Zisserman. Springer Berlin Heidelberg, 2008, pp. 650–663. ISBN: 978-3-540-88688-4.

[276] Glenn F. Wilson. 'An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures'. In: *The International Journal of Aviation Psychology* 12.1 (2002), pp. 3–18. ISSN: 1050-8414. DOI: 10.1207/S15327108IJAP1201_2.

[277] Jacob O. Wobbrock and Julie A. Kientz. 'Research Contributions in Human-Computer Interaction'. In: *Interactions* 23.3 (2016), pp. 38–44.

[278] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 'User-Defined Gestures for Surface Computing'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Boston, MA, USA: ACM, 2009, pp. 1083–1092. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518866.

[279] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. 'Eulerian Video Magnification for Revealing Subtle Changes in the World'. In: *ACM Transactions on Graphics (TOG)* 31.4 (2012), 65:1–65:8. ISSN: 0730-0301. DOI: 10.1145/2185520.2185561.

[280] Benjamin Wulff and Alexander Fecke. 'LectureSight - An Open Source System for Automatic Camera Control in Lecture Recordings'. In: *Proceedings of the 2012 IEEE International Symposium on Multimedia*. ISM '12. Irvine, CA, USA: IEEE, 2012, pp. 461–466. DOI: 10.1109/ISM.2012.94.

[281] Benjamin Wulff, Alexander Fecke, Lisa Rupp, and Kai-Christoph Hamborg. 'LectureSight: An Open Source System for Automatic Camera Control for Lecture Recordings'. In: *Interactive Technology and Smart Education* 11.3 (2014), pp. 184–200. ISSN: 1741-5659. DOI: 10.1108/ITSE-07-2014-0019.

[282] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. 'Detecting Faces in Images: A Survey'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.1 (2002). publisher=IEEE, pp. 34–58. ISSN: 0162-8828. DOI: 10.1109/34.982883.

[283] Georgios N. Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 'Mixed-Initiative Co-Creativity'. In: *FDG*. 2014.

[284] Takao Yokoi and Hironobu Fujiyoshi. 'Virtual Camerawork for Generating Lecture Video from High Resolution Images'. In: *2005 IEEE International Conference on Multimedia and Expo*. 2005 IEEE International Conference on Multimedia and Expo. Amsterdam, Netherlands: IEEE, 2005. ISBN: 0-7803-9331-7. DOI: 10.1109/ICME.2005.1521532.

[285] Wenyi Zhao, Rama Chellappa, P. Jonathon Phillips, and Azriel Rosenfeld. 'Face Recognition: A Literature Survey'. In: *ACM Computing Surveys* 35.4 (2003), pp. 399–458. ISSN: 0360-0300. DOI: 10.1145/954339.954342.

[286] Silvia Zuffi, Javier Romero, Cordelia Schmid, and Michael J. Black. 'Estimating Human Pose with Flowing Puppets'. In: *2013 IEEE International Conference on Computer Vision*. ICCV '13. IEEE, 2013, pp. 3312–3319.

Supervision

Supervised student work including theses and reports on practical courses and seminars

Master Theses

[287] Sarah Aragon Bartsch. *Through-The-Lens Camera Motion Control: An Exploration of Interactions*. Master Thesis. LMU Munich, 2016.

[288] Thomas Burghart. *User Interface Concepts for Motion Control*. Master Thesis. LMU Munich, 2014.

[289] John-Louis Gao. *Image-Based Gimbal Control*. Master Thesis. LMU Munich, 2016.

[290] Alexander Klimczak. *EyeScout: Active Gaze-Interaction on Large Displays*. Co-supervised with Mohamed Khamis. Master Thesis. LMU Munich, 2016.

[291] Martin Reiss. *EyeScout: Active Gaze-Interaction on Large Displays*. Co-supervised with Mohamed Khamis. Master Thesis. LMU Munich, 2016.

Bachelor Theses

[292] Mujo Alic. *Exploring the Effect of Progressive Reduction on Perceived Control in a Cinematographic User Interface*. Bachelor Thesis. LMU Munich, 2017.

[293] Korbinian Blanz. *Development and Exploration of Tools for Evaluating the Workload/Unpredictability Trade-Off*. Bachelor Thesis. LMU Munich, 2017.

[294] Philipp Burgdorf. *Take-Over Strategies and Their Effect on Control and Recording of Automated Camera Tracking Shots*. Bachelor Thesis. LMU Munich, 2016.

[295] Elisabeth Dreßler. *Kontrolle in Automatisierten Steuerungen: Design-Konzepte Für Kamerasysteme*. Bachelor Thesis. LMU Munich, 2016.

[296] Andreas Ellwanger. *Axis plus Content-Based Control of Camera Motion: Design and Evaluation of User Interface Concepts*. Bachelor Thesis. LMU Munich, 2017.

[297] Filip Hristov. *Elicitation of Gestures and Mappings for Gesture-Based Computer Supported Camera Motion Control*. Bachelor Thesis. LMU Munich, 2017.

[298] Patrick Mörwald. *Development of a Remote Controlled Camera Slider*. with object druckerei. Bachelor Thesis. LMU Munich, 2014.

[299] Theresa Mücke. *On Camera Motion - Interviews, Opportunities and Concepts*. In conjuction with Stephan Vorbrugg Bildgestaltung, co-supervised with Julie Wagner. Bachelor Thesis. LMU Munich, 2014.

[300] Michael Puriss. *Examining and Adressing Visual Overload in Cinematographic User Interfaces in the Wild*. In conjuction with eyeslovetosee. Bachelor Thesis. LMU Munich, 2017.

Practical Course Reports

[301] Florian Lehmann, Christina Rosenmöller, and Phuong Anh Vu. *Exploring the Perception of Control in Cinematographic Motion Control*. Practical Course Report. Project in Advanced Topics in HCI. LMU Munich, 2017.

[302] Ludwig Trotter, Andrea Attwenger, and Maximilian Körner. *Image-Based Gimbal Control*. Practical Course Report. LMU Munich, 2017.

Seminar Reports

[303] Katharina Sachmann. *Strengthening the Evaluation of Creativity Support Tools with Implicit Measures*. Seminar Report. 2016.

[304] Daniel Seliger. *Evaluating Perceived Control and Authorship in Virtual Reality*. Seminar Report. 2017.

Electronic Sources

Websites

[305] AccelStepper. *AccelStepper: AccelStepper Library for Arduino*. 2018. URL: http://www.airspayce.com/mikem/arduino/AccelStepper (visited on 06/12/2019).

[306] ACR. *The Link | ACR Systems*. Video: https://vimeo.com/162514507. 2018. URL: http://acr-sys.com/products/the-link (visited on 06/12/2019).

[307] Adafruit. *Bluefruit LE - Bluetooth Low Energy (BLE 4.0)*. 2018. URL: https://www.adafruit.com/product/1697 (visited on 06/12/2019).

[308] Aimetis. *Aimetis: Home*. 2017. URL: http://www.aimetis.com (visited on 06/12/2019).

[309] Arduino. *Arduino Mega 2560 Rev3*. 2018. URL: https://store.arduino.cc/usa/arduino-mega-2560-rev3 (visited on 06/12/2019).

[310] ARRI. *ARRI Group: AMIRA*. 2018. URL: https://www.arri.com/en/camera-systems/cameras/amira (visited on 06/12/2019).

[311] ArriMotion. *ArriMotion Operator's Homepage*. 2013. URL: http://www.arrimotion.com (visited on 06/12/2019).

[312] ASC. *The American Society of Cinematographers*. 2018. URL: https://www.theasc.com (visited on 06/12/2019).

[313] BaseCam Electronics. *BaseCam SimpleBGC 32-Bit : BaseCam Electronics*. 2018. URL: https://www.basecamelectronics.com/simplebgc32bit (visited on 06/12/2019).

[314] BlackMagic. *Blackmagic URSA Mini Pro | Blackmagic Design*. 2018. URL: https://www.blackmagicdesign.com/products/blackmagicursaminipro (visited on 06/12/2019).

[315] Bolex Collector. *Bolex Collector | Cameras | H-16 Leader*. 2013. URL: http://www.bolexcollector.com/cameras/h16leader.html (visited on 06/12/2019).

[316] Tom Butts. *TV Technology Announces STAR & Mario Awards for NAB2007 | TvTechnology*. 2007. URL: https://www.tvtechnology.com/news/tv-technology-announces-star-mario-awards-for-nab2007 (visited on 06/12/2019).

[317] BVK. *Berufsverband Kinematografie - German Society Of Cinematographers*. 2018. URL: https://www.kinematografie.org/english/bvk/index.php (visited on 06/12/2019).

[318] campilots. *Turtlecam - Campilots - Fast and Furious Low Shots - Germany*. 2016. URL: http://www.campilots.com/lowrider.html (visited on 06/12/2019).

[319] Chapman. *Hustler IV | Chapman-Leonard.Com*. 2015. URL: http://www.chapman-leonard.com/product/620 (visited on 06/12/2019).

[320] Chicagology. *Bell & Howell 2709 35mm Camera*. 2018. URL: https://chicagology.com/silentmovies/bellhowell2709 (visited on 06/12/2019).

[321] C-mocos. *C-MOCOS*. 2018. URL: http://www.cmocos.com (visited on 06/12/2019).

[322] David S. Cohen and Dave McNary. *Alfonso Cuaron Returns to Bigscreen After Seven Years With 'Gravity' – Variety*. 2013. URL: http://variety.com/2013/film/news/alfonso-cuaron-returns-to-the-bigscreen-after-seven-years-with-gravity-1200596518 (visited on 06/12/2019).

[323] damperzen. *DAMPERZEN | Advanced Gimbal Damper System*. 2018. URL: http://www.damperzen.com (visited on 06/12/2019).

[324] Dan Birman. *Progressive Reduction: Evolving the Experience for Your Most Frequent Users*. 2016. URL: http://www.dtelepathy.com/blog/design/progressive-reduction-evolving-the-experience-for-your-most-frequent-users (visited on 06/12/2019).

[325] Disney. *The Evolution of the Acrobat Robot: Disney Imagineers Unveil Stuntronics - The Walt Disney Company*. 2018. URL: https://www.thewaltdisneycompany.com/the-evolution-of-the-acrobat-robot-disney-imagineers-unveil-stuntronics (visited on 06/12/2019).

[326] DJI. *DJI GO - Capture and Share Beautiful Content Using This New App*. 2018. URL: https://www.dji.com/uk/goapp (visited on 06/12/2019).

[327] f4. *F4transkript - Faster Transcription of Interviews and Recordings | Audiotranscription.De*. 2018. URL: https://www.audiotranskription.de/english/f4 (visited on 06/12/2019).

[328] FFmpeg. *FFmpeg*. 2018. URL: https://www.ffmpeg.org (visited on 06/12/2019).

[329] Firefly. *Firefly*. 2017. URL: http://en.t-firefly.com (visited on 06/12/2019).

[330] framer. *Framer - Interactive Design and Prototyping Tool*. 2018. URL: https://www.framer.com (visited on 06/12/2019).

[331] GitHub. *GitHub - AmroShohoud/Myo-Drone*. 2015. URL: https://github.com/AmroShohoud/myo-drone (visited on 06/12/2019).

[332] Allan Grinshtein. *Progressive Reduction - LayerVault Blog*. 2013. URL: http://layervault.tumblr.com/post/42361566927/progressive-reduction (visited on 06/12/2019).

[333] David Hancock. *The Global Digital Conversion of Cinemas Is Almost over - IHS Technology*. 2016. URL: https://technology.ihs.com/577835/the-global-digital-conversion-of-cinemas-is-almost-over (visited on 06/12/2019).

[334] Stephen Herbert. *Who's Who of Victorian Cinema - Wordsworth Donisthorpe*. URL: http://www.victorian-cinema.net/donisthorpe (visited on 06/12/2019).

[335] Wolfgang Hock. *Wolfgang Hock: Erste Photo-Kameras*. URL: http://www.wolfganghock.com/images%20F%20ist%20tot/Camera%20obscura%20von%20Niepce%201825%20kl.jpg (visited on 06/12/2019).

[336] **Axel Hoesl**. *Slider · Master · Lmu08360 / CameraMotion · GitLab*. 2017. URL: https://gitlab.lrz.de/lmu08360/CameraMotion/tree/master/Slider (visited on 06/12/2019).

[337] **Axel Hoesl**. *Tracker · Master · Lmu08360 / CameraMotion · GitLab*. 2017. URL: https://gitlab.lrz.de/lmu08360/CameraMotion/tree/master/Tracker (visited on 06/12/2019).

[338] Hurricane Wheels. *Hurricane Wheels Balanced Handwheel f. Gimbal Interface - Kids of All Ages - Motion Control*. 2018. URL: https://web.archive.org/web/20171021071729/http://kids-of-all-ages.de/hersteller/hurricane-wheels/hurricane-wheels-balanced-handwheel-f-gimbal-interface (visited on 06/12/2019).

[339] igus. *Igus® Drylin® Stepper Motor NEMA 23*. 2018. URL: https://www.igus.com/wpck/10753/DryLin_NEMA_23 (visited on 06/12/2019).

[340] igus. *Linear Profile Guides Drylin® by Igus®*. 2018. URL: https://www.igus.com/drylin/profile-rail-guide (visited on 06/12/2019).

[341] IMDb. *Come Swim (2017) - IMDb*. URL: https://www.imdb.com/title/tt5968238 (visited on 06/12/2019).

[342] IMDb. *Julia and Julia (1987) - IMDb*. URL: http://www.imdb.com/title/tt0093092/technical?ref_=tt_dt_spec (visited on 06/12/2019).

[343] IMDb. *Sunspring (2016) - IMDb*. URL: https://www.imdb.com/title/tt5794766 (visited on 06/12/2019).

[344] IMDb. *Vidocq (2001) - Technical Specifications - IMDb*. 2017. URL: http://www.imdb.com/title/tt0164961/technical?ref_=tt_dt_spec (visited on 06/12/2019).

[345] JCUDA. *Jcuda.Org - Java Bindings for CUDA*. 2018. URL: http://www.jcuda.org (visited on 06/12/2019).

[346] Keemotion. *Keemotion*. 2018. URL: http://www.keemotion.com (visited on 06/12/2019).

[347] Nokia Bell Labs. *Charge Coupled Device - Bell Labs*. 2017. URL: https://www.bell-labs.com/about/history-bell-labs/stories-changed-world/charge-coupled-device (visited on 06/12/2019).

[348] Leadshine. *DM556 Stepper Drive - Leadshine Technology Co., Ltd*. 2018. URL: http://www.leadshine.com/productdetail.aspx?type=products&category=stepper-products&producttype=stepper-drives&series=DM&model=DM556 (visited on 06/12/2019).

[349] Leap Motion. *Controlling a Parrot AR.Drone with Leap Motion + Cylon.Js*. 2015. URL: http://blog.leapmotion.com/controlling-parrot-ar-drone-leap-motion-cylon-js (visited on 06/12/2019).

[350] Andrew Leonard. *How Netflix Is Turning Viewers into Puppets | Salon.Com*. 2013. URL: https://www.salon.com/2013/02/01/how_netflix_is_turning_viewers_into_puppets/?123 (visited on 06/12/2019).

[351] *Library of Mixed-Initiative Creative Interfaces*. 2018. URL: http://mici.codingconduct.cc (visited on 06/12/2019).

[352] Maedler. *Maedler - Beam Coupling*. 2018. URL: http://www.maedler.de/Article/60271600 (visited on 06/12/2019).

[353] Maedler. *Maedler - Neoprene Timing Belt HTD 3M Width 9mm 3m-09 Open Length*. 2018. URL: https://www.maedler.de/product/1643/1616/970/zahnriemen-profil-htd-3m-breite-9-mm (visited on 06/12/2019).

[354] Maedler. *Maedler - SKF Deep Groove Ball Bearing Single Row*. 2018. URL: https://www.maedler.de/Article/626-2Z-SKF (visited on 06/12/2019).

[355] Maedler. *Maedler - Timing Belt Pulley HTD 3M Material Alu- Minium 26 Teeth for Belt Width 9mm*. 2018. URL: http://www.maedler.de/Article/17022600 (visited on 06/12/2019).

[356] Magic Lantern. *Magic Lantern*. 2014. URL: http://www.magiclantern.fm (visited on 06/12/2019).

[357] Material. *Accessibility - Material Design*. 2018. URL: https://material.io/design/usability/accessibility.html#layout-typography (visited on 06/12/2019).

[358] Mean Well. *Mean Well SP-320 Series*. 2018. URL: https://www.meanwell.com/webapp/product/search.aspx?prod=sp-320 (visited on 06/12/2019).

[359] Microsoft. *Targeting - UWP App Developer | Microsoft Docs*. 2017. URL: https://docs.microsoft.com/en-us/windows/uwp/input-and-devices/guidelines-for-targeting (visited on 06/12/2019).

[360] Myo. *Myo | Flying Drones with MyoPilot*. 2015. URL: https://developerblog.myo.com/flying-drones-with-myopilot (visited on 06/12/2019).

[361] NielsenNorman Group. *Progressive Disclosure*. 2006. URL: https://www.nngroup.com/articles/progressive-disclosure (visited on 06/12/2019).

[362] Oscars. *19 Scientific And Technical Achievements To Be Honored With Academy Awards® | Oscars.Org | Academy of Motion Picture Arts and Sciences*. 2014. URL: http://www.oscars.org/news/19-scientific-and-technical-achievements-be-honored-academy-awardsr (visited on 06/12/2019).

[363] Oscars. *2009 | Oscars.Org | Academy of Motion Picture Arts and Sciences*. URL: https://www.oscars.org/oscars/ceremonies/2009 (visited on 06/12/2019).

[364] Panther. *Classic Plus Kamera Dolly - Panther*. 2018. URL: http://www.panther.tv/product/classic-plus (visited on 06/12/2019).

[365] Parrot. *FreeFlight Pro | Parrot Official*. 2018. URL: https://www.parrot.com/global/freeflight-pro#pilot-your-parrot-drone-with-freeflight-pro (visited on 06/12/2019).

[366] Raspberry Pi. *Raspberry Pi 3 Model B+ - Raspberry Pi*. 2018. URL: https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus (visited on 06/12/2019).

[367] Raytrix. *Technology – Raytrix*. 2019. URL: https://raytrix.de/technology (visited on 06/12/2019).

[368] RED. *RED ONE | The Revolutionary 4K Digital Cinema Camera*. 2018. URL: http://www.red.com/products/red-one (visited on 04/12/2018).

[369] RED. *RED WEAPON*. 2018. URL: http://www.red.com/products/weapon (visited on 04/12/2018).

[370] SAE. *J3016B: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International*. 2018. URL: https://www.sae.org/standards/content/j3016_201806 (visited on 06/12/2019).

[371] ScriptBook. *ScriptBook - Hard Science. Better Box Office*. 2019. URL: https://www.scriptbook.io (visited on 06/12/2019).

[372] Sony. *IMX219PQ | Sony Semiconductor Solutions*. 2014. URL: https://www.sony-semicon.co.jp/products_en/new_pro/april_2014/imx219_e.html (visited on 06/12/2019).

[373] Spidercam. *Spidercam - Home*. 2017. URL: http://www.spidercam.tv (visited on 06/12/2019).

[374] StarWars. *Mythic Discovery: Revisiting the Meeting between George Lucas and Joseph Campbell*. 2015. URL: https://www.starwars.com/news/mythic-discovery-within-the-inner-reaches-of-outer-space-joseph-campbell-meets-george-lucas-part-i (visited on 06/12/2019).

[375] Statista. *Leading Countries Based on Number of Monthly Active YouTube Users as of 1st Quarter 2016 (in Millions)*. 2017. URL: https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users (visited on 06/12/2019).

[376] Statista. *Number of Vimeo Subscribers Worldwide from 1st Quarter 2015 to 3rd Quarter 2017 (in Millions)*. 2017. URL: https://www.statista.com/statistics/705598/vimeo-subscribers-worldwide (visited on 06/12/2019).

[377] StypeGrip. *STYPE KIT - 3D Virtual Studio and Augmented Reality System Stype Grip*. 2018. URL: http://www.stypegrip.com/stype-kit (visited on 06/12/2019).

[378] Supertechno. *Telescopic Camera Cranes | Robotic Cranes by Supertechno.Com*. 2018. URL: https://www.supertechno.com/telescopic-camera-cranes (visited on 06/12/2019).

[379] Syrp. *Genie | Syrp*. 2018. URL: https://syrp.co/eu/product/genie (visited on 06/12/2019).

[380] The Mill. *The Mill*. 2019. URL: http://www.themill.com/portfolio/3002/the-blackbird%5C%00ae (visited on 06/12/2019).

[381] The R Foundation. *R: The R Project for Statistical Computing*. 2018. URL: https://www.r-project.org (visited on 06/12/2019).

[382] Tornado. *Tornado Web Server - Tornado 5.1 Documentation*. 2018. URL: http://www.tornadoweb.org/en/stable (visited on 06/12/2019).

[383] TUM. *GitHub - InstituteOfErgonomics/ArduinoDRT: Arduino Detection Response Task*. 2017. URL: https://github.com/InstituteOfErgonomics/ArduinoDRT (visited on 06/12/2019).

[384] Unity. *Unity at SIGGRAPH 2018*. 2018. URL: https://www.unity3d.com/siggraph2018 (visited on 06/12/2019).

[385] Viscoda. *Voodoo Camera Tracker*. 2018. URL: https://www.viscoda.com/index.php?option=com_content&view=article&id=104&Itemid=548&lang=en (visited on 06/12/2019).

[386] WayBack Machine. *Arduino Multi Kamera IR Control Library for Nikon, Canon, & More | Sebastian.Setz.Name*. 2017. URL: https://web.archive.org/web/20170131204041/http://sebastian.setz.name/arduino/my-libraries/multi-camera-ir-control (visited on 06/12/2019).

[387] YouTube. *Introducing the All-New DJI GO Mobile App - YouTube*. 2015. URL: https://www.youtube.com/watch?v=_t0MI-I_yKU (visited on 06/12/2019).

[388] YouTube. *Parrot Bebop 2 - Tutorial #2 - Piloting - YouTube*. 2016. URL: https://www.youtube.com/watch?v=XDV35wuCayk (visited on 06/12/2019).

[389] YouTube. *Yuneec BREEZE 4K - Pilot Mode - YouTube*. 2016. URL: https://www.youtube.com/watch?v=JwhrgClARV8 (visited on 06/12/2019).

Figures

[390] Robert Couto. *Robert Couto — Bot & Dolly - a New Generation of Camera Movement*. 2015. URL: https://www.robertcouto.com/#/botdolly (visited on 06/12/2019).

[391] Flickr. *162 Die Entfesselte Kamera Auf Der Gleitbahn_Cigaretten Bi... | Flickr*. 2011. URL: https://www.flickr.com/photos/42399206@N03/6433506737 (visited on 06/12/2019).

[392] Flickr. *First Digital Camera | In 1975 Steven Sasson Created the Fir... | Flickr*. 2016. URL: https://www.flickr.com/photos/socsci/31380508405 (visited on 06/12/2019).

[393] IMDb. *Garrett Brown on IMDb: Movies, TV, Celebs, and More... - Photo Gallery - IMDb*. URL: http://www.imdb.com/name/nm0113593/mediaviewer/rm2150773248 (visited on 06/12/2019).

[394] IMDb. *Jean-Luc Godard on IMDb: Movies, TV, Celebs, and More... - Photo Gallery - IMDb*. URL: http://www.imdb.com/name/nm0000419/mediaviewer/rm2965681920 (visited on 06/12/2019).

[395] Plos One. *The Sense of Agency during Continuous Action: Performance Is More Important than Action-Feedback Association*. 2015. URL: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0125226 (visited on 06/12/2019).

[396] Rishi Sanyal. *Lytro Poised to Forever Change Filmmaking: Debuts Cinema Prototype and Short Film at NAB*. 2016. URL: https://www.dpreview.com/files/p/articles/6720444400/IMG_4306.jpeg (visited on 06/12/2019).

[397] Rotten Tomatoes. *5 Technical Breakthroughs in Star Wars That Changed Movies Forever << Rotten Tomatoes – Movie and TV News*. 2017. URL: https://editorial.rottentomatoes.com/article/5-technical-breakthroughs-in-star-wars-that-changed-movies-forever/ (visited on 06/12/2019).

[398] Wikimedia. *File:Arri Alexa Camera.Jpg - Wikimedia Commons*. 2014. URL: https://commons.wikimedia.org/wiki/File:Arri_Alexa_camera.jpg (visited on 06/12/2019).

[399] Wikimedia. *File:Boston, as the Eagle and the Wild Goose See It.Jpg - Wikimedia Commons*. 2013. URL: https://commons.wikimedia.org/wiki/File:Boston,_as_the_Eagle_and_the_Wild_Goose_See_It.jpg (visited on 06/12/2019).

[400] Wikimedia. *File:Camera Obscura 1.Jpg - Wikimedia Commons*. 2008. URL: https://commons.wikimedia.org/wiki/File:Camera_obscura_1.jpg (visited on 06/12/2019).

[401] Wikimedia. *File:Camera Obscura box18thCentury.Jpg - Wikimedia Commons*. 2011. URL: https://commons.wikimedia.org/wiki/File:Camera_Obscura_box18thCentury.jpg (visited on 06/12/2019).

[402] Wikimedia. *File:Courtesy Edison National Historic Site, West Orange, NJ.Jpg - Wikimedia Commons*. 2010. URL: https://commons.wikimedia.org/wiki/File:Courtesy_Edison_National_Historic_Site,_West_Orange,_NJ.jpg (visited on 06/12/2019).

[403] Wikimedia. *File:DJI OSMO Camera and Gimbal 2.Jpg - Wikimedia Commons*. 2015. URL: https://commons.wikimedia.org/wiki/File:DJI_OSMO_camera_and_gimbal_2.jpg (visited on 06/12/2019).

[404] Wikimedia. *File:Flying DJI Mavic Pro 3.Jpg - Wikimedia Commons*. 2017. URL: https://commons.wikimedia.org/wiki/File:Flying_DJI_Mavic_Pro_3.jpg (visited on 06/12/2019).

[405] Wikimedia. *File:Institut Lumière - CINEMATOGRAPHE Camera.Jpg - Wikimedia Commons*. 2014. URL: https://commons.wikimedia.org/wiki/File:Institut_Lumi%C3%A8re_-_CINEMATOGRAPHE_Camera.jpg (visited on 06/12/2019).

[406] Wikimedia. *File:Kodak Nr 1.Jpg - Wikimedia Commons*. (converted to black and white). 2012. URL: https://commons.wikimedia.org/wiki/File:Kodak_nr_1.jpg (visited on 06/12/2019).

[407] Wikimedia. *File:Mori Uncanny Valley.Svg - Wikimedia Commons*. 2007. URL: https://commons.wikimedia.org/wiki/File:Mori_Uncanny_Valley.svg (visited on 06/12/2019).

[408] Wikimedia. *File:My Nikon - D90 (13536773553).Jpg - Wikimedia Commons*. 2014. URL: https://commons.wikimedia.org/wiki/File:My_Nikon_-_D90_(13536773553).jpg (visited on 06/12/2019).

[409] Wikimedia. *File:Optic Projection Fig 89.Jpg - Wikimedia Commons*. 2008. URL: https://commons.wikimedia.org/wiki/File:Optic_Projection_fig_89.jpg (visited on 06/12/2019).

[410] Wikimedia. *File:PFD.Png - Wikimedia Commons*. 2016. URL: https://commons.wikimedia.org/wiki/File:PFD.png (visited on 06/12/2019).

[411] Wikimedia. *File:Redonecamera.JPG - Wikimedia Commons*. 2008. URL: https://commons.wikimedia.org/wiki/File:Redonecamera.JPG (visited on 06/12/2019).

[412] Wikimedia. *File:Six Flight Instruments.JPG - Wikimedia Commons*. 2008. URL: https://commons.wikimedia.org/wiki/File:Six_flight_instruments.JPG (visited on 06/12/2019).

[413] Wikimedia. *File:SliderPLUS PRO Fs700.Jpg - Wikimedia Commons*. 2014. URL: https://commons.wikimedia.org/wiki/File:SliderPLUS_PRO_Fs700.jpg (visited on 06/12/2019).

[414] Wikimedia. *File:Sony Cinealta.Jpg - Wikimedia Commons*. 2006. URL: https://commons.wikimedia.org/wiki/File:Sony_Cinealta.jpg (visited on 06/12/2019).

[415] Wikimedia. *File:View from the Window at Le Gras, Joseph Nicéphore Niépce.Jpg - Wikimedia Commons*. 2005. URL: https://commons.wikimedia.org/wiki/File:View_from_the_Window_at_Le_Gras,_Joseph_Nic%C3%A9phore_Ni%C3%A9pce.jpg (visited on 06/12/2019).

[416] Wikipedia. *File:FairchildF-1AerialCamera.Jpg - Wikipedia*. 2010. URL: https://en.wikipedia.org/wiki/File:FairchildF-1AerialCamera.jpg (visited on 06/12/2019).

[417] Wikipedia. *File:Le-Prince-Cameraprojector-Type1-Mark2-1888.Png - Wikipedia*. 2011. URL: https://en.wikipedia.org/wiki/File:Le-prince-cameraprojector-type1-mark2-1888.png (visited on 06/12/2019).

Videos

[418] CNN. *Scriptbook Tries to Predict Box Office Takings - Video - Media*. 2017. URL: https://money.cnn.com/video/media/2017/01/04/script-book-box-office.cnnmoney (visited on 06/12/2019).

[419] TED. *Zeresenay Alemseged: The Search for Humanity's Roots | TED Talk*. 2007. URL: https://www.ted.com/talks/zeresenay_alemseged_looks_for_humanity_s_roots (visited on 06/12/2019).

[420] Vimeo. *2D and 3D Direct Manipulation*. 2012. URL: https://vimeo.com/39892336 (visited on 06/12/2019).

[421] Vimeo. *Automated Cinematography with Unmanned Aerial Vehicles*. 2016. URL: https://vimeo.com/157138672 (visited on 06/12/2019).

[422] Vimeo. *Colibri on Wheels on Vimeo*. 2014. URL: https://vimeo.com/87411305 (visited on 06/12/2019).

[423] Vimeo. *DitoGear OmniSlider SERVO + RED Epic - VFX Testing on Vimeo*. 2011. URL: http://vimeo.com/28306715 (visited on 06/12/2019).

[424] Vimeo. *Maximus 7 by SpaceCam Systems Inc. on Vimeo*. 2016. URL: https://vimeo.com/153473118 (visited on 06/12/2019).

[425] Vimeo. *Showreel DOP Stephan Vorbrugg on Vimeo*. 2014. URL: https://vimeo.com/108470180 (visited on 06/12/2019).

[426] Vimeo. *SteadyFlow: Spatially Smooth Optical Flow for Video Stabilization*. 2016. URL: https://vimeo.com/156073058 (visited on 06/12/2019).

[427] Vimeo. *The Mill BLACKBIRD on Vimeo*. 2016. URL: https://vimeo.com/171939943 (visited on 06/12/2019).

[428] Wikimedia. *File:Intolerance (1916).Ogv - Wikimedia Commons*. 2014. URL: https://commons.wikimedia.org/w/index.php?title=File%3AIntolerance_(1916).ogv (visited on 06/12/2019).

[429] Wikimedia. *File:Roundhay Garden Scene.Ogv - Wikimedia Commons*. 2007. URL: https://commons.wikimedia.org/w/index.php?title=File%3ARoundhay_Garden_Scene.ogv (visited on 06/12/2019).

[430] YouTube. *3 Technocrane Camera Moves! | Cinematography Composition 201 - YouTube*. 2017. URL: https://www.youtube.com/watch?v=1MGT_0QkmbA (visited on 06/12/2019).

[431] YouTube. *300 Hz Real-Time Optical Flow. City*. 2016. URL: https://www.youtube.com/watch?v=_Rpi7WS-HTw (visited on 06/12/2019).

[432] YouTube. *3D Virtual Studio/Augmented Reality System - the Stype Kit*. 2013. URL: https://www.youtube.com/watch?v=d8ktVGkHAes (visited on 06/12/2019).

[433] YouTube. *"A Moment of Silence" Behind The Scenes - Shot Using the Kessler CineDrive - YouTube*. 2013. URL: https://www.youtube.com/watch?v=y6xm7tBSs0g (visited on 01/04/2018).

[434] YouTube. *Accelerated KAZE Features*. 2013. URL: https://www.youtube.com/watch?v=lI50PGr2TEU (visited on 06/12/2019).

[435] YouTube. *Agency in Mid Air Interfaces*. 2017. URL: https://www.youtube.com/watch?v=ETIJsaEOhBI (visited on 06/12/2019).

[436] YouTube. *Aimcrane Kit Video - YouTube*. 2010. URL: https://www.youtube.com/watch?v=oGIHC8N9ZRc (visited on 06/12/2019).

[437] YouTube. *Aimetis Face Recognition - Accurate and Easy to Use*. 2017. URL: https://www.youtube.com/watch?v=12So0nUWG5Q (visited on 06/12/2019).

[438] YouTube. *Airways: Optimization-Based Planning of Quadrotor Trajectories According to High-Level User Goals - YouTube*. 2016. URL: https://www.youtube.com/watch?v=CHjtTAvnmno (visited on 06/12/2019).

[439] YouTube. *Appearance-Based Landmark Selection for Efficient Long-Term Visual Localization*. 2016. URL: https://www.youtube.com/watch?v=JL_5zMEQKYc (visited on 06/12/2019).

[440] YouTube. *Ausstellung — Stan Douglas: Mise En Scène - YouTube*. 2014. URL: https://www.youtube.com/watch?v=sFpMd656-Tc (visited on 06/12/2019).

[441] YouTube. *Ausstellungen September 2014 – Januar 2015 - YouTube*. 2014. URL: https://www.youtube.com/watch?v=8_mD-DeitEA (visited on 06/12/2019).

[442] YouTube. *Ausstellungsfilm — Kapsel 03 + 04 / Lynette Yiadom-Boakye Und Adele Röder - YouTube*. 2015. URL: https://www.youtube.com/watch?v=0kktDh7JKX8 (visited on 06/12/2019).

[443] YouTube. *Axis HDTV PTZ Dome Network Camera - YouTube*. 2013. URL: https://www.youtube.com/watch?v=Q45eOZ4NVsc (visited on 06/12/2019).

[444] YouTube. *AXIS Mobile Viewing App (Android) - View Recording*. 2016. URL: https://www.youtube.com/watch?v=-NM9C1WfMj0 (visited on 06/12/2019).

[445] YouTube. *Behaviour-Based Robotics - 3PI/Wixel Robot Controlled by Java*. 2012. URL: https://www.youtube.com/watch?v=HogA8Cot6_8 (visited on 06/12/2019).

[446] YouTube. *'Birdman': Filmed as One Long Shot - YouTube*. 2014. URL: https://www.youtube.com/watch?v=D4qKyNU5u6M (visited on 06/12/2019).

[447] YouTube. *Bolt On Track Showreel at Stiller Studios - YouTube*. 2016. URL: https://www.youtube.com/watch?v=sfOroSB2f64 (visited on 06/12/2019).

[448] YouTube. *Boogie Nights (HD) - Opening Steadicam Scene - YouTube*. 2012. URL: https://www.youtube.com/watch?v=iiXtFyZqvQQ (visited on 06/12/2019).

[449] YouTube. *Bot & Dolly's Iris, World's Most Advanced Robotic Motion Control Camera System - YouTube*. 2012. URL: https://www.youtube.com/watch?v=xWJCxxKuTOw (visited on 06/12/2019).

[450] YouTube. *BREAKING BAD - Motivated Camera Movement - YouTube*. 2014. URL: https://www.youtube.com/watch?v=hlE8H6RaduI (visited on 06/12/2019).

[451] YouTube. *Camera Dolly Shots | Cinematography Composition 201 - YouTube*. 2017. URL: https://www.youtube.com/watch?v=Zhvg-EG7Lfs (visited on 04/13/2018).

[452] YouTube. *Camera Jib Moves | Cinematography Composition 201 - YouTube*. 2017. URL: https://www.youtube.com/watch?v=6SPuR2Q5gY8 (visited on 06/12/2019).

[453] YouTube. *Campilots Showreel HoverCam 2013 - YouTube*. 2013. URL: https://www.youtube.com/watch?v=O4T9mTNMuQI (visited on 06/12/2019).

[454] YouTube. *Cine HoverCam Interview - YouTube*. 2014. URL: https://www.youtube.com/watch?v=HQw7KnmSfgs (visited on 06/12/2019).

[455] YouTube. *Citizen Kane: Camera Moving Through Objects (The Rancho) - YouTube*. 2011. URL: https://www.youtube.com/watch?v=Th8cuq9tzZk (visited on 06/12/2019).

[456] YouTube. *Come Swim Film Directed By Kristen Stewart | Shatterbox Anthology | Refinery29 - YouTube*. 2017. URL: https://www.youtube.com/watch?v=u37GTEjnQv4 (visited on 06/12/2019).

[457] YouTube. *Demo of OpenFace on Drone-Recorded Video*. 2017. URL: https://www.youtube.com/watch?v=j6l2vUUBDIw (visited on 06/12/2019).

[458] YouTube. *Die Neue Tagesschau - Das Neue Studio - YouTube*. 2014. URL: https://www.youtube.com/watch?v=jnbeR5eGRTk (visited on 06/12/2019).

[459] YouTube. *Disney's Stunt Robots Could Change How Hollywood Makes Action Movies - YouTube*. 2018. URL: https://www.youtube.com/watch?v=nZ950ywJy0M (visited on 06/12/2019).

[460] YouTube. *DJI - Introducing the Ronin 2 - YouTube*. 2017. URL: https://www.youtube.com/watch?v=W0PgZsP8yBQ (visited on 06/12/2019).

[461] YouTube. *DJI - Introducing the Ronin - YouTube*. 2014. URL: https://www.youtube.com/watch?v=bcjSdzYb9LY (visited on 06/12/2019).

[462] YouTube. *DJI Mavic Pro – Dual Remote Controller Mode - YouTube*. 2017. URL: https://www.youtube.com/watch?v=3UEFcd_O9jk (visited on 06/12/2019).

[463] YouTube. *DJI Tutorials - Mavic Pro - Standout Features - Gesture Mode*. 2016. URL: https://www.youtube.com/watch?v=8fdVpil2erI (visited on 06/12/2019).

[464] YouTube. *DJI Tutorials - Spark - Gesture Piloting - YouTube*. 2017. URL: https://www.youtube.com/watch?v=ZeTXGkkox_E (visited on 06/12/2019).

[465] YouTube. *Drone History - It Started with I Am Cuba's Long Shot - YouTube*. 2014. URL: https://www.youtube.com/watch?v=H7fRyrbmbjg (visited on 06/12/2019).

[466] YouTube. *Drone & Me: An Exploration Into Natural Human-Drone Interaction*. 2015. URL: https://www.youtube.com/watch?v=vrWF3t7a_HU (visited on 06/12/2019).

[467] YouTube. *Drone & Me: An Exploration Into Natural Human-Drone Interaction - YouTube*. 2015. URL: https://www.youtube.com/watch?v=vrWF3t7a_HU (visited on 06/12/2019).

[468] YouTube. *Early Photography: Making Daguerreotypes - YouTube*. 2012. URL: https://www.youtube.com/watch?v=N0Ambe4FwQk (visited on 06/12/2019).

[469] YouTube. *Empirical Evidence for a Diminished Sense of Agency in Speech Interfaces*. 2016. URL: https://www.youtube.com/watch?v=WfcmFfPDp6o (visited on 06/12/2019).

[470] YouTube. *Estimating Human Pose with Flowing Puppets (ICCV 2013)*. 2013. URL: https://www.youtube.com/watch?v=aUyZLtml9PM (visited on 06/12/2019).

[471] YouTube. *Étienne Œhmichen's Quad Copter - YouTube*. 2016. URL: https://www.youtube.com/watch?v=0DdcfT8P560 (visited on 06/12/2019).

[472] YouTube. *Exhibition – A History. Contemporary Art from the Centre Pompidou - YouTube*. 2016. URL: https://www.youtube.com/watch?v=Rp11QFEypLg (visited on 06/12/2019).

[473] YouTube. *Exhibition — Anri Sala: The Present Moment - YouTube*. 2015. URL: https://www.youtube.com/watch?v=ptcKqAtrABM (visited on 06/12/2019).

[474] YouTube. *EyeScout: Active Eye Tracking for Gaze Interaction with Large Public Displays (UIST'17) - YouTube*. 2017. URL: https://www.youtube.com/watch?v=J7_0iRqsmdM (visited on 06/12/2019).

[475] YouTube. *Face2Face: Real-Time Face Capture and Reenactment of RGB Videos (CVPR 2016 Oral)*. 2016. URL: https://www.youtube.com/watch?v=ohmajJTcpNk (visited on 06/12/2019).

[476] YouTube. *Firefly Face Recognition SDK Released*. 2017. URL: https://www.youtube.com/watch?v=Y3sgGpsctQ8 (visited on 06/12/2019).

[477] YouTube. *Flying Parrot AR.Drone Using Myo Armband - YouTube*. 2015. URL: https://www.youtube.com/watch?v=AvAdgFt22ms (visited on 06/12/2019).

[478] YouTube. *Gandalf Vs Balrog (First Scene) - The Lord of the Rings: The Two Towers - YouTube*. 2015. URL: https://www.youtube.com/watch?v=kXGfXW3OLAc (visited on 06/12/2019).

[479] YouTube. *GoPro HERO7 X Johnny FPV - DRIFT - YouTube*. 2018. URL: https://www.youtube.com/watch?v=SldJIisWFmE (visited on 06/12/2019).

[480] YouTube. *Gravity Behind-the-Scenes Featurette Takes Audiences From Script to Screen - YouTube*. 2013. URL: https://www.youtube.com/watch?v=QxHc8Ns5g1c (visited on 06/12/2019).

[481] YouTube. *Haus Der Kunst — Interventionen in Die Architektur - YouTube*. 2016. URL: https://www.youtube.com/watch?v=rLVu4VBhJ9Q (visited on 06/12/2019).

[482] YouTube. *HEXO+ : Your Autonomous Aerial Camera - YouTube*. 2014. URL: https://www.youtube.com/watch?v=sKy_Qa6lMU0 (visited on 06/12/2019).

[483] YouTube. *HRI in the Sky: Creating and Commanding Teams of UAVs with a Vision-Mediated Gestural Interface*. 2013. URL: https://www.youtube.com/watch?v=L3H-akDN5vs (visited on 06/14/2018).

[484] YouTube. *Hurricane Wheels Crowdfunding Video with English Subtitles - YouTube*. 2013. URL: https://www.youtube.com/watch?v=8HRaRp-00AQ (visited on 06/12/2019).

[485] YouTube. *IBC 2017: ARRI Camera Stabilizer Systems - YouTube*. 2017. URL: https://www.youtube.com/watch?v=ytS89N1l1-Y (visited on 06/12/2019).

[486] YouTube. *Il Carrello Di Cabiria (1914) - YouTube*. 2007. URL: https://www.youtube.com/watch?v=c2B3Et8dqLI (visited on 06/12/2019).

[487] YouTube. *Introducing the DJI Osmo - YouTube*. 2015. URL: https://www.youtube.com/watch?v=tJZzgNwLoNA (visited on 06/12/2019).

[488] YouTube. *Jean-Claude Van Damme Presents - The Moving Camera for Hard Target - YouTube*. 2014. URL: https://www.youtube.com/watch?v=fDLVjcmRgrE (visited on 06/12/2019).

[489] YouTube. *KAZE Planar Object Tracking*. 2013. URL: https://www.youtube.com/watch?v=t1epvNOYcyc (visited on 06/12/2019).

[490] YouTube. *Keanu Reeves. Bot & Dolly. Kungfu Proof of Concept - YouTube*. 2012. URL: https://www.youtube.com/watch?v=cuvYBph4YBg (visited on 06/12/2019).

[491] YouTube. *Kessler CineDrive kOS Overview - YouTube*. 2012. URL: https://www.youtube.com/watch?v=E2gCisHY99Q (visited on 06/12/2019).

[492] YouTube. *Laser-Cut Wolcott Camera - YouTube*. 2015. URL: https://www.youtube.com/watch?v=bAtvtKnbjUE (visited on 06/12/2019).

[493] YouTube. *Leap Motion Node Drone Flight in 2013 - YouTube*. 2013. URL: https://www.youtube.com/watch?v=hfq2SisPvCU (visited on 06/12/2019).

[494] YouTube. *Lecturer Tracking Medium Shot - YouTube*. 2014. URL: https://www.youtube.com/watch?v=C9qRWiFTKEo (visited on 06/12/2019).

[495] YouTube. *LectureSight in Action*. 2016. URL: https://www.youtube.com/watch?v=PibOj6Pt3kU (visited on 06/12/2019).

[496] YouTube. *Make a Move with TECHNODOLLY - YouTube*. 2013. URL: https://www.youtube.com/watch?v=2_RC6d-WjZ4 (visited on 06/12/2019).

[497] YouTube. *Martin Scorsese on the Importance of Visual Literacy - YouTube*. 2012. URL: https://www.youtube.com/watch?v=I90ZluYvHic (visited on 06/12/2019).

[498] YouTube. *Matrix Bullet Time Scene [HD] - YouTube*. 2012. URL: https://www.youtube.com/watch?v=3c8Dl2c1whM (visited on 06/12/2019).

[499] YouTube. *MIDWAY JOURNEY II - Behind the Scenes - Kessler Cineslider - YouTube*. 2010. URL: https://www.youtube.com/watch?v=DP_8sR6g9Yg (visited on 06/12/2019).

[500] YouTube. *Milo Motion Control Rig – Mark Roberts Motion Control - YouTube*. 2011. URL: https://www.youtube.com/watch?v=9e4wnJrnuqU (visited on 06/12/2019).

[501] YouTube. *MIT Computer Program Reveals Invisible Motion in Video | The New York Times*. 2013. URL: https://www.youtube.com/watch?v=3rWycBEHn3s (visited on 06/12/2019).

[502] YouTube. *Motion Kit for SliderPLUS - Captivating Camera Motion in Seconds! - YouTube*. 2017. URL: https://www.youtube.com/watch?v=QSKqmJs_CRA (visited on 06/12/2019).

[503] YouTube. *OK Go - "I Won't Let You Down" - Interview with Damian and Tim - YouTube*. 2014. URL: https://www.youtube.com/watch?v=WCwzaTPARtU (visited on 06/12/2019).

[504] YouTube. *OK Go - I Won't Let You Down - Official Video - YouTube*. 2014. URL: https://www.youtube.com/watch?v=u1ZB_rGFyeU (visited on 06/12/2019).

[505] YouTube. *Opencv Optical Flow Demo*. 2017. URL: https://www.youtube.com/watch?v=tWo1LEBkMjI (visited on 06/12/2019).

[506] YouTube. *Pan + Tilt Applied | Cinematography Composition 201 - YouTube*. 2016. URL: https://www.youtube.com/watch?v=1LsZFHg6-qo (visited on 06/12/2019).

[507] YouTube. *Panther | Tristar Dolly - YouTube*. 2012. URL: https://www.youtube.com/watch?v=MCt8zlul0TU (visited on 06/12/2019).

[508] YouTube. *Percepto: Computer Vision Apps for Drones - YouTube*. 2015. URL: https://www.youtube.com/watch?v=Eu7ldtSywC4 (visited on 06/12/2019).

[509] YouTube. *Restaurant Scene - Mission Impossible - YouTube*. 2015. URL: https://www.youtube.com/watch?v=yDs8li2_AO8 (visited on 06/12/2019).

[510] YouTube. *Revealing Invisible Changes In The World*. 2011. URL: https://www.youtube.com/watch?v=e9ASH8IBJ2U (visited on 06/12/2019).

[511] YouTube. *R/GA Keemotion NAIA Vid - YouTube*. 2017. URL: https://www.youtube.com/watch?v=CdTKs24bmS0 (visited on 06/12/2019).

[512] YouTube. *"Save the Cat" Examples*. 2009. URL: https://www.youtube.com/watch?v=4MRY6BP0EpE (visited on 06/12/2019).

[513] YouTube. *Save The Cat! Tribute Video: Back To The Future - YouTube*. 2010. URL: https://www.youtube.com/watch?v=HGA7ccCNrkE (visited on 06/12/2019).

[514] YouTube. *Save The Cat! Tribute Video: ET - YouTube*. 2010. URL: https://www.youtube.com/watch?v=Xeua51wHbHw (visited on 06/12/2019).

[515] YouTube. *Scalable 3D Tracking of Multiple Interacting Objects - CVPR 2014*. 2014. URL: https://www.youtube.com/watch?v=SCOtBdhDMKg (visited on 06/12/2019).

[516] YouTube. *Scorpio Telescopic Camera Crane: PAN BARS & MINI HEAD - YouTube*. 2011. URL: https://www.youtube.com/watch?v=i5O-y3-1AU4 (visited on 06/12/2019).

[517] YouTube. *Screen Design Starts with Framer*. 2017. URL: https://www.youtube.com/watch?v=4_Zy1V701qw (visited on 06/12/2019).

[518] YouTube. *ShowMotion: Camera Motion Based 3D Design Review*. 2015. URL: https://www.youtube.com/watch?v=pKGUDyPWHHo (visited on 06/12/2019).

[519] YouTube. *SHOWREEL 2014 __ 2015 | Stype Kit Camera Tracking System - YouTube*. 2015. URL: https://www.youtube.com/watch?v=sPW98MS1WFM (visited on 06/12/2019).

[520] YouTube. *SimSensei & MultiSense: Virtual Human and Multimodal Perception for Healthcare Support*. 2013. URL: https://www.youtube.com/watch?v=ejczMs6b1Q4 (visited on 06/12/2019).

[521] YouTube. *Skinput: Appropriating the Body as an Input Surface (CHI 2010)*. 2010. URL: https://www.youtube.com/watch?v=g3XPUdW9Ryg (visited on 06/12/2019).

[522] YouTube. *Spectre- Opening Tracking Shot in 1080p - YouTube*. 2016. URL: https://www.youtube.com/watch?v=cbqv1kbsNUY (visited on 06/12/2019).

[523] YouTube. *Spidercam Demoreel - YouTube*. 2013. URL: https://www.youtube.com/watch?v=t1EOCKLNBdw (visited on 06/12/2019).

[524] YouTube. *Star Wars Visual Effects, from AT-ATs to Tauntauns - YouTube*. 2014. URL: https://www.youtube.com/watch?v=mIlYk7KQe-s (visited on 06/12/2019).

[525] YouTube. *Steadicam: Then and Now with Garrett Brown (Rocky) - YouTube*. 2013. URL: https://www.youtube.com/watch?v=7GURHXOlGQ8 (visited on 06/12/2019).

[526] YouTube. *StyleCam Interactive Stylized 3D Navigation Using Integrated Spatial Temporal Controls Publicatio*. 2011. URL: https://www.youtube.com/watch?v=J7dmQPNi_3c (visited on 06/12/2019).

[527] YouTube. *Syrp Genie - Motion Control Timelapse Device - YouTube*. 2013. URL: https://www.youtube.com/watch?v=OHpPz7zW_gw (visited on 06/12/2019).

[528] YouTube. *The Blacksmiths (1895) - LOUIS LUMIERE - Les Forgerons - YouTube*. 2012. URL: https://www.youtube.com/watch?v=dlxKBmagjTc (visited on 06/12/2019).

[529] YouTube. *The Director's Lens (ACM Multimedia 2011): A Smart Assistant for Shooting Virtual Cinematography*. 2011. URL: https://www.youtube.com/watch?v=30yvCLjI9Sg (visited on 06/12/2019).

[530] YouTube. *The Evolution of the Zoom Dolly - YouTube*. 2016. URL: https://www.youtube.com/watch?v=WIpMtL68G8w (visited on 06/12/2019).

[531] YouTube. *The Good The Bad The Ugly Cemetery Scene(1080p) - YouTube*. 2012. URL: https://www.youtube.com/watch?v=_ZHEu7HusG4 (visited on 06/12/2019).

[532] YouTube. *The Sea (1895) - LOUIS LUMIERE - La Mer - YouTube*. 2012. URL: https://www.youtube.com/watch?v=CAbTn3KzkgY (visited on 06/12/2019).

[533] YouTube. *The Self-Programming Motion Kit for SliderPLUS - YouTube*. 2018. URL: https://www.youtube.com/watch?v=CteqtqCOfXk (visited on 06/12/2019).

[534] YouTube. *The Sprinkler Sprinkled (1895) - 1st Comedy Movie - LOUIS LUMIERE - L'Arroseur Arrose - YouTube*. 2011. URL: https://www.youtube.com/watch?v=IooPPilYzkM (visited on 06/12/2019).

[535] YouTube. *TheMarmalade Identity - Behind The Scenes - YouTube*. 2012. URL: https://www.youtube.com/watch?v=A2CLQdCU7O0 (visited on 06/12/2019).

[536] YouTube. *Through the Lens Camera Control (FullClip SVR 1992 Upconv) - YouTube*. 2017. URL: https://www.youtube.com/watch?v=GcWCklhmFOY (visited on 06/12/2019).

[537] YouTube. *TopoSketch Demo*. 2017. URL: https://www.youtube.com/watch?v=lNjXbKNhtGA (visited on 06/12/2019).

[538] YouTube. *Touch of Evil Opening Shot - YouTube*. 2008. URL: https://www.youtube.com/watch?v=Yg8MqjoFvy4 (visited on 06/12/2019).

[539] YouTube. *Trick Riding (1895) - LOUIS LUMIERE - La Voltige - YouTube*. 2012. URL: https://www.youtube.com/watch?v=iqbXxJEVYTk (visited on 06/12/2019).

[540] YouTube. *UI Canvas - Unity Official Tutorials - YouTube*. 2014. URL: https://www.youtube.com/watch?v=OD-p1eMsyrU&list=PLX2vGYjWbI0Qp0sD8_RKgbWul7z_eyNAv (visited on 06/12/2019).

[541] YouTube. *Unity for Film - Siggraph 2018 Trailer - YouTube*. 2018. URL: https://www.youtube.com/watch?v=dwVkOuvu-Nw (visited on 06/12/2019).

[542] YouTube. *Using Hand Gestures and Face Pose Estimates to Maneuver and Direct Airborne UAVs*. 2015. URL: https://www.youtube.com/watch?v=3ZYrZYMk314 (visited on 06/12/2019).

[543] YouTube. *Vertical Studio App - Demonstration - YouTube*. 2016. URL: https://www.youtube.com/watch?v=RqXnAREptlQ (visited on 06/12/2019).

[544] YouTube. *Vertical UI/UX Demo - YouTube*. 2015. URL: https://www.youtube.com/watch?v=fKsUeT8Iigs (visited on 06/12/2019).

[545] YouTube. *Vertigo - YouTube*. 2008. URL: https://www.youtube.com/watch?v=bDwPJ5IaJ2g (visited on 06/12/2019).

[546] YouTube. *ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers - YouTube*. 2016. URL: https://www.youtube.com/watch?v=Poi0MeASmuY (visited on 06/12/2019).

[547] YouTube. *Victoria | Press Conference Highlights | Berlinale 2015 - YouTube*. 2015. URL: https://www.youtube.com/watch?v=3wjh-CMj90U (visited on 06/12/2019).

[548] YouTube. *Who's in Control: Patrick Haggard at TEDxHelvetia*. 2012. URL: https://www.youtube.com/watch?v=knvGvWghRIE (visited on 06/12/2019).

[549] YouTube. *Wilbur Wright Und Seine Flugmaschine - 1909 - YouTube*. 2011. URL: https://www.youtube.com/watch?v=8osZHhkp-cM (visited on 06/12/2019).

[550] YouTube. *You've Got the Moves, We've Got the Motion -Understanding and Designing for Cinematographic .. - YouTube*. 2018. URL: https://www.youtube.com/watch?v=-b1k6PWJaOE (visited on 06/12/2019).

Documents

[551] EU Publications Office Top. *European Commission : CORDIS : Projects and Results : Apertus° eXtendable Integrated Open Modular Cinema Camera*. 2017. URL: http://cordis.europa.eu/project/rcn/194334_en.html (visited on 06/12/2019).

[552] Netflix. *Netflix Letter to Stockholders Q3 2017*. 2017. URL: https://s22.q4cdn.com/959853165/files/doc_financials/quarterly_reports/2017/q3/Q3_17_Shareholder_Letter_COMBINED.pdf (visited on 06/12/2019).

[553] Thorsten Pohlert. *Package 'PMCMR'*. 2018. URL: https://cran.r-project.org/web/packages/PMCMR/PMCMR.pdf (visited on 06/12/2019).

[554] Royal Swedisch Academy of Sciences. *Two Revolutionary Optical Technologies*. 2009. URL: https://www.nobelprize.org/nobel_prizes/physics/laureates/2009/advanced-physicsprize2009.pdf (visited on 11/29/2017).

[555] Spidercam. *SPIDERCAM® SC250 FIELD FEATURES AND SPECIFICATIONS*. 2015. URL: http://www.spidercam.tv/daten/uploads/Features_and_Specifications_SC250FIELD_en_V1.4.pdf (visited on 06/12/2019).

[556] Supertechno. *Technodolly Software Users Manual*. 2018. URL: https://www.supertechno.com/download.php?fid=1022 (visited on 06/12/2019).

Appendix

12.1 Online Survey Questionnaire

Online questionnaire used for the survey in 3.2

Expert user feedback

As a 2D slider remote UI will be designed for a certain user group of expert users, the needs they have are of major importance.

There is a need to identify user roles. How do you mostly use a slider? Who is the person in charge of motion control?
e.g. shooting alone or in a collaborative set with others

I also want to find out about user tasks. Can you name the most important tasks controlling a slider with a remote?

User requirements are vital for the development. Could you name affordances that an expert user requires when working with a remote-controlled slider?

In the following, some ideas for the functionality of the remote UI that have already come up in the development process are presented. Could you rate those ideas by the value of importance for an expert user?

Live image from the camera included in the UI

1	2	3	4	5
☐	☐	☐	☐	☐
Non-Relevant	Somewhat Non-Relevant	Neutral	Somewhat Relevant	Relevant

Direct motion control

1	2	3	4	5
☐	☐	☐	☐	☐
Non-Relevant	Somewhat Non-Relevant	Neutral	Somewhat Relevant	Relevant

Motion capturing and editing

1	2	3	4	5
☐	☐	☐	☐	☐
Non-Relevant	Somewhat Non-Relevant	Neutral	Somewhat Relevant	Relevant

Position feedback of the slider head

1	2	3	4	5
☐	☐	☐	☐	☐
Non-Relevant	Somewhat Non-Relevant	Neutral	Somewhat Relevant	Relevant

Review function of the recorded material

1	2	3	4	5
☐	☐	☐	☐	☐
Non-Relevant	Somewhat Non-Relevant	Neutral	Somewhat Relevant	Relevant

Collaborative/Mulit-User Environment

1	2	3	4	5
☐	☐	☐	☐	☐
Non-Relevant	Somewhat Non-Relevant	Neutral	Somewhat Relevant	Relevant

12.2 Expert Interview Questionnaire

Questionnaire for the expert interviews presented in 3.5. The interviews were conducted in German. To support the readability of this thesis the questionnaire was translated to English.

Demographics

Age:			
Sex:	☐ Male	☐ Female	
Occupation:	☐ Student ☐ Teacher	☐ Permanent Operator ☐ Freelance Operator	
Days of recording: *within past 12 month*	☐ < 10 ☐ 50 - 89	☐ 10 - 29 ☐ 90 - 200	☐ 30 - 49 ☐ > 200
Days of recording: *with support tool*	☐ < 10 ☐ 20 - 49	☐ 10 - 19 ☐ > 50	
Type of recording:	☐ Short Film ☐ Image Film ☐ Advertisement	☐ Feature Film ☐ Trailer	☐ Documentary ☐ Online Material
Last Projects:			

Warm-Up

Please think of the scene that took you the longest to shoot. Could you please describe the situation?

What is your definition of fun behind the camera?

Why did you choose camera operation as your occupation?

Can you think of a day of shooting where remarkably many takes were needed for a scene? Please tell me about it.

Best-Case Scenario

Do you remember a scene with *hardly any* complications? Can you describe the take?

- What was the content of the scene?
- At which location?
- At which time of the day?
- How was your motivation?
- Which camera did you use?
- How was the shot organised?
- Who was responsible for the shot?
- Who was responsible for lights and focus?
- Who was responsible for communicating ideas?
- How was the coordination between departments?
- How was the coordination within the camera department?
- What type of camera movement did you use?
- How many people were involved in the recording of the scene?
- How many people were on set?
- Did the team work together priorly or was it recently compiled?
- How many takes were necessary in total?
- How long did the recording take?
- How long should the recording take (according to schedule)?
- What was central that it worked out in the end?
- When you consider the work of others on set, what do you think causes unnecessary trouble in their practice?
- Are complex movements in general more difficult to shoot or are they better preplanned and hence work out easier?

Worst-Case Scenario

Do you remember a scene with *many* complications? Can you describe the take?

- What was the content of the scene?
- At which location?
- At which time of the day?
- How was your motivation?
- Which camera did you use?
- How was the shot organised?
- Who was responsible for the shot?
- Who was responsible for lights and focus?
- Who was responsible for communicating ideas?
- How was the coordination between departments?
- How was the coordination within the camera department?
- What type of camera movement did you use?
- How many people were involved in the recording of the scene?
- How many people were on set?
- Did the team work together priorly or was it recently compiled?
- How many takes were necessary in total?
- How long did the recording take?
- How long should the recording take (according to schedule)?
- What was central that it worked out in the end?
- When you consider the work of others on set, what do you think causes unnecessary trouble in their practice?
- Are complex movements in general more difficult to shoot or are they better preplanned and hence work out easier?

Preparation

How is a shot planned and carried out? Can you (cognitively) walk us through all the necessary steps from the ideation to the realisation?

- How do you ideate a scene?
- How do you translate your abstract idea to the camera?
- How do you shoot the scene?
- How do you evaluate and possibly retake the scene?

Communication

How do you communicate with colleagues on set? Can you expand on the details a bit?

- If any, what tools do you use?
- At which level of detail do you share an idea?
- Can you image a way to share your ideas more easily?
- Are their (obvious) pitfalls in communicating that you try to avoid?

Operation

Can you further dive into some details of your experience in camera operation?

- What are (your) indicators of a successful day of shooting?
- How important is the right timing for a successful shot?
- What roles do lags (in technical devices) play a role in camera operation?
- How much do you rely on *"blind operation"*? (focussing less on the viewfinder or screen and more on the physical aspect of operation)
- If at all, how are *"blind operation"* and getting the timing right related?

12.3 Non-Parametric Testing in R

The template of the R[1] [381] script we used for significance testing using the non-parametric Friedman's Test [90, 91] (for complete block designs) and Durbin's Test (for incomplete block designs). For post-hoc pairwise comparisons, Wilcoxon's Signed Rank Test [274] is sometimes suggested as a non-parametric test for pairwise comparison as by Fields and Hole [86]. However, its use as a posthoc test after Friedman's test is also controversial. Alternatives are recommended, for instance, by Sachs [231] such as Conover's [54] or Nemenyi's Test [192] and implemented in the PMCMR package [553]. We implemented all three mentioned tests in our template but commented out Wilcoxon's Signed Rank and Nemenyi's along with test for incomplete blocks designs.

```
# Parametric tests for significance testing
# Implemented omnibus tests: Friedman's, Durbin's
# Implemented posthoc tests: Wilcoxon Signed Rank, Conover's, Nemenyi's and Durbin's (and
    Conover's) post-hoc
# Expects a .csv file with colums named ID, User_Interface and Rating
# Date: 28-10-2018
# Author: Axel Hoesl

# (if necessary install and) load PMCMR package
if(!require('PMCMR')) { install.packages('PMCMR') }
library('PMCMR')

# location of this script (substitue PATH with your path)
working_dir <- 'PATH'

# set working directory
setwd(working_dir)

# Read a .csv file (substitute PATH/DATA.csv with your filepath and file)
the_dataset <- read.csv('PATH/DATA.csv', sep=';', fileEncoding='UTF-8', header=TRUE)

# print the dataset (if necessary)
# print(the_dataset)

# print the descriptives of the dataset
tapply(the_dataset$Rating, the_dataset$User_Interface, function(x){ summary(x)} )

# omnibus test for significant differences with Friedman's Test (for complete block
   design)
omnibus_result <- friedman.test(the_dataset$Rating,the_dataset$User_Interface,the_dataset
   $ID)

# omnibus test for significant differences with Durbin's Test (for incomplete block
   design)
```

[1] We used R version 3.4.0 (2017-04-21) running on a x86_64-apple-darwin15.6.0 (64-bit) platform

```
# omnibus_result <- durbin.test(Rating ~ User_Interface | Task, data=the_dataset)

# print the results of Friedman's Test
print(omnibus_result)

# set the alpha level at 0.05
alpha_level <- 0.05

# Post-hoc tests are conducted only if omnibus Friedman test p-value is less than the alpha level
if(omnibus_result$p.value < alpha_level) {

# Wilcoxon's Signed Rank Test is sometimes suggested as a non-parametric test for pairwise comparison as by Fields and Hole (How to Design and Report Experiments, Sage, 2002). However its use as a post-hoc test after Friedman's test is controversial. Alternatives are recommended, for instance, by Sachs (Angewandte Statistik: Anwendung Statistischer Methoden, Springer 1997) such as Conover's or Nemenyi's Test.

# Wilcoxon's Signed Rank Test for post-hoc pairwise comparisions (with Bonferroni correction)
# posthoc_wilcoxon <- pairwise.wilcox.test(the_dataset$Rating, the_dataset$User_Interface, paired=TRUE, p.adjust.method='bonferroni')

# print the results of Wilcoxon's Signed Rank Test as a matrix
# print(posthoc_wilcoxon)

# Conover's Test for post-hoc pairwise comparisions (with Bonferroni correction)
posthoc_result <- posthoc.friedman.conover.test(the_dataset$Rating,the_dataset$User_Interface,the_dataset$ID, p.adjust.method='bonferroni')

# Durbin and Conover's post-hoc test (for incomplete block design) (with Bonferroni correction)
#posthoc_result <- posthoc.durbin.test(the_dataset$Rating,the_dataset$User_Interface,the_dataset$ID, p.adjust.method='bonferroni')

# print the results of Conover's or Durbin's Test as a matrix
print(posthoc_result)

# print the results of Conover's or Durbin's Test as a table
summary(posthoc_result)

# Nemenyi's Test for post-hoc pairwise comparisons
# posthoc_nemenyi <- posthoc.friedman.nemenyi.test(the_dataset$Rating, the_dataset$User_Interface, the_dataset$ID)

# print the results of Nemenyi's Test as a matrix
# print(posthoc_nemenyi)

# print the results of Nemenyi's Test as a table
# summary(posthoc_nemenyi)
```

```
} else {
# print message that no significant difference was found
sprintf("No significant difference was found with the omnibus test (alpha: %s).
    Consequently, no post-hoc pairwise comparisons were conducted.", alpha_level)
}
```

12.4 Parametric Testing in R

The template of the R[2] [381] script we used for significance testing using parametric ANOVA (two way, two way repeated measures). For post-hoc pairwise comparisons of two way ANOVAs, Tukey's Test is used.

```
# Parametric tests for significance testing
# Implemented omnibus tests: ANOVA (two way, two way repeated measures)
# Implemented posthoc tests:
# Expects a .csv file with colums named ID, User_Interface and Rating
# Date: 28-10-2018
# Author: Axel Hoesl

# (if necessary install and) load nlme package
if(!require('nlme')) { install.packages('nlme') }
library('nlme')

# (if necessary install and) load multcomp package
if(!require('multcomp')) { install.packages('multcomp') }
library('multcomp')

# location of this script (substitue PATH with your path)
working_dir <- 'PATH'

# set working directory
setwd(working_dir)

# Read a .csv file (substitute PATH/DATA.csv with your filepath and file)
the_dataset <- read.csv('PATH/DATA.csv', sep=';', fileEncoding='UTF-8', header=TRUE)

# print the dataset
print(the_dataset)

# print the descriptives of the dataset
tapply(the_dataset$Rating, the_dataset$User_Interface, function(x){ summary(x)} )

# set the alpha level at 0.05
```

[2] We used R version 3.4.0 (2017-04-21) running on a x86_64-apple-darwin15.6.0 (64-bit) platform

```
alpha_level <- 0.05

# testing normality of distribution: Shapiro Wilk Test
distribution_result <- shapiro.test(the_dataset$Rating)

# print results of testing for normality of distribution
if(distribution_result$p.value > alpha_level) {
sprintf(”Normality of distribution can be assumed (alpha: %s).”, alpha_level)
} else {
sprintf(”Normality of distribution can NOT be assumed (alpha: %s).”, alpha_level)
}

# perform a two way ANOVA
# anova_results <- aov(Rating ~ User_Interface * Task + Error(ID / User_Interface), data=
    the_dataset)

# print the results of the two way ANOVA
# summary(anova_results)

# perform a two way repeated measures ANOVA
anova_results <- lme(Rating ~ User_Interface * Task, random=~1|ID, data=the_dataset)

# print the results of the two way repeated measures ANOVA
anova(anova_results)

# conduct post-hoc pairwise comparison with Tukey’s Test
the_dataset$UI_Task <- interaction(the_dataset$User_Interface, the_dataset$Task)
model <- lme(Rating ~ UI_Task, random = ~1 | ID, data=the_dataset)
posthoc_results <- glht(model, linfct=mcp(UI_Task=”Tukey”))

# print results of the post-hoc tests
summary(posthoc_results)
```

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.7.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 13.6.2019

Axel Hösl