

Aus der Klinik und Poliklinik für Nuklearmedizin  
der Universität München

Vorstand: Prof. Dr. med. K. Hahn

Bildfusion von MRT und ECD-SPECT Daten  
des menschlichen Gehirns –  
Entwicklung eines Modells zur Fehlerberechnung und  
Vergleich manueller und automatischer Methoden

Dissertation  
zum Erwerb des Doktorgrades der Medizin  
an der Medizinischen Fakultät  
der Ludwig-Maximilians-Universität  
zu München

Vorgelegt von Christian Vollmar aus München

2005

Mit Genehmigung der Medizinischen Fakultät  
Der Universität München

Berichterstatter: Priv. Doz. Dr. Th. Pfluger

Mitberichterstatter: Priv. Doz. Dr. R. Brüning  
Prof. Dr. K.J. Pfeifer

Dekan: Prof. Dr. med. Dr. h.c. K. Peter

Tag der mündlichen Prüfung: 13. Januar 2005

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung .....</b>	<b>6</b>
<b>2</b>	<b>Grundlagen .....</b>	<b>9</b>
2.1	<b>Theoretische Grundlagen der Untersuchungstechniken .....</b>	<b>9</b>
2.1.1	Magnetresonanztomographie .....	9
2.1.2	Single-Photonenemissions-Computertomographie .....	13
2.1.2.1	Radionuklide .....	13
2.1.2.2	Radiopharmaka .....	13
2.1.2.3	Gammakamera .....	14
2.2	<b>Bildfusion - Prinzipien und Methodenübersicht .....</b>	<b>16</b>
2.2.1	Manuelle Methoden .....	17
2.2.2	Automatische Methoden .....	18
2.2.2.1	Prinzipien der iterativen Bildfusion .....	18
2.2.2.2	Landmarkenbasierte Methoden .....	21
2.2.2.3	externe Marker .....	22
2.2.2.4	Surface matching nach Pelizzari .....	24
2.2.2.5	Pixel uniformity matching nach Woods .....	25
2.2.2.6	Mutual Information .....	26
2.2.2.7	Principal Axes .....	28
2.3	<b>Ziele der Arbeit .....</b>	<b>29</b>
<b>3</b>	<b>Material und Methoden .....</b>	<b>30</b>
3.1	<b>Patienten .....</b>	<b>30</b>
3.2	<b>Magnetresonanztomographie .....</b>	<b>30</b>
3.3	<b>ECD-SPECT .....</b>	<b>31</b>
3.4	<b>Bildfusion .....</b>	<b>31</b>
3.4.1	Systeme .....	31

3.4.1.1	Picker VoxelQ.....	31
3.4.1.2	Nuclear Diagnostics Hermes .....	32
3.4.2	Manuelle Bildfusion.....	32
3.4.3	Surface matching nach Pelizzari .....	34
3.4.4	Pixel uniformity matching nach Woods .....	36
<b>3.5</b>	<b>Fehleranalyse .....</b>	<b>36</b>
3.5.1	Unterschiede der verwendeten Fusionssysteme .....	36
3.5.2	Definition der Referenzposition.....	40
3.5.3	Berechnung der Abweichung von der Referenzposition.....	41
3.5.4	Intermethodenvergleich .....	42
<b>4</b>	<b>Ergebnisse.....</b>	<b>43</b>
4.1	<b>Konsolidierung des Fehlerwertes.....</b>	<b>43</b>
4.2	<b>Fusionsgenauigkeit.....</b>	<b>45</b>
4.2.1	Manuelle Bildfusion.....	46
4.2.1.1	Intraobservervariabilität .....	46
4.2.1.2	Interobservervariabilität .....	47
4.2.2	Surface matching nach Pelizzari .....	48
4.2.2.1	Intraobservervariabilität .....	48
4.2.2.2	Interobservervariabilität .....	48
4.2.3	Pixel uniformity matching nach Woods .....	49
4.2.4	Intermethodenvergleich .....	50
4.3	<b>Geometrische Verteilung des Fusionsfehlers .....</b>	<b>51</b>
4.4	<b>Zeitbedarf für die Fusion .....</b>	<b>51</b>
<b>5</b>	<b>Diskussion.....</b>	<b>52</b>
5.1	<b>Alternative Verfahren zur Bildfusion.....</b>	<b>52</b>
5.2	<b>Fehleranalyse .....</b>	<b>53</b>
5.2.1	Definition des Goldstandards .....	53

5.2.1.1	Externe Marker .....	53
5.2.1.2	Simulierte Studien .....	54
5.2.1.3	Phantomstudie .....	54
5.2.1.4	Statistische Auswertung .....	55
5.2.2	Art der Fehlerberechnung .....	57
5.2.2.1	Parameterangabe, 2D Fehler .....	57
5.2.2.2	3D Fehler-Angabe.....	57
5.2.2.3	Trefferquote .....	59
<b>5.3</b>	<b>Ergebnisse der Fehlerberechnung .....</b>	<b>59</b>
<b>5.4</b>	<b>Zeitbedarf.....</b>	<b>61</b>
<b>5.5</b>	<b>Gesamtbewertung der Ergebnisse .....</b>	<b>62</b>
<b>6</b>	<b>Zusammenfassung .....</b>	<b>64</b>
<b>7</b>	<b>Literaturverzeichnis .....</b>	<b>67</b>
<b>8</b>	<b>Danksagung .....</b>	<b>73</b>
<b>9</b>	<b>Lebenslauf.....</b>	<b>74</b>

# 1 Einführung

Die Diagnostik in vielen Bereichen der Medizin stützt sich seit der Einführung moderner Schnittbildverfahren in zunehmendem Maße auf bildgebende diagnostische Verfahren.

Das Spektrum der zur Verfügung stehenden diagnostischen Methoden ist weit: Die Computertomographie (CT) oder Magnetresonanztomographie (MRT) liefern genaue morphologisch-anatomische Informationen, während nuklearmedizinische Untersuchungen, wie die Single-Photonenemissions Computertomographie (SPECT) oder Positronenemissionstomographie (PET), funktionelle oder metabolische Informationen beisteuern können.

Für entsprechende klinische Fragestellungen ist es sinnvoll diese komplementären Informationen verschiedener diagnostischer Verfahren zu kombinieren. Für viele Fragestellungen kann mit diesem multimodalen Ansatz sowohl die Sensitivität als auch die Spezifität der Diagnostik erhöht werden (Hutton, Braun et al. 2002). Der diagnostische Gewinn beruht auf der Möglichkeit, funktionelle Informationen genauer zu lokalisieren und definierten anatomischen Strukturen zuzuordnen (Thurfjell, Pagani et al. 2000).

In der Epilepsiediagnostik erlaubt die Bildfusion eine exaktere Lokalisation der Anfallsursprungszone durch eine Perfusions-SPECT (O'Brien, O'Connor et al. 1998; Valenti, Froelich et al. 2002) oder die Korrelation des hippocampalen Hypometabolismus in der PET mit einer Volumenreduktion in der MRT (Knowlton, Laxer et al. 2001). In einer weiteren Studie konnten in Kenntnis der Lokalisation von pathologischen Befunden aus SPECT und Magnetencephalographie in der MRT diskrete kortikale Dysplasien diagnostiziert werden, die histologisch gesichert waren, aber in der Primärbefundung der MRT nicht als pathologisch befundet wurden (Zhang, Simos et al. 2003). Oft ist für eine weitergehende Untersuchung bei epilepsiechirurgischen Kandidaten eine invasive Untersuchung mit subdural implantierten EEG-Elektroden notwendig. Hier ermöglicht die Bildfusion von MRT und CT eine genauere Lokalisation der implantierten Elektroden als es mit nur einer bildgebenden Methode alleine möglich wäre (Winkler, Vollmar et al. 2000). Gleiches gilt für die postoperative Lokalisation von stereotaktisch implantierten Tiefenelektroden (Ferroli, Franzini et al. 2004). Für alle Bildfusionen ist schließlich die Integration der fusionierten Daten in ein Navigationssystem möglich, um die

verschiedenen Befunde intraoperativ zu lokalisieren. Auch wenn für eine Befundlokalisation nicht zwingend ein Navigationssystem notwendig ist, wird dessen Einsatz in einigen Zentren für ein resektives operatives Vorgehen in komplexen Fällen favorisiert (Murphy, O'Brien et al. 2004). In der Rezidivdiagnostik von Hirntumoren kann die Bildfusion von MRT und PET die Differenzierung zwischen Tumorrezidiv und unspezifischen posttherapeutischen Veränderungen oder Nekrosen verbessern (Nelson, Day et al. 1997). Für die Strahlentherapie erlaubt eine solche fusionierte "anatomometabolische" Darstellung des Gehirns die Identifizierung vitaler Tumoranteile, auf die eine Bestahlungsplanung fokussiert werden kann - bei gleichzeitiger Dosisreduktion in anderen Gewebeanteilen (Rajasekar, Datta et al. 2003). Bei der Untersuchung von Demenzerkrankungen erlaubt die Bildfusion eine Korrelation von Hypoperfusion und Atrophie und konnte so zeigen, dass die Perfusionsreduktion der strukturell nachweisbaren Atrophie vorausgeht und über deren Maß hinausgeht (Imran, Kawashima et al. 1999). Die Möglichkeit, regions of interest (ROI) anhand einer MRT zu definieren und diese dann für die Auswertung auf eine SPECT oder PET zu übertragen erlaubt eine verbesserte quantitative Auswertung nuklearmedizinischer Studien. Die genauere anatomische Informationen für die Definition der ROIs hilft dabei, z.B. Partialvolumeneffekte zu reduzieren (Evans, Marrett et al. 1991; Quarantelli, Berkouk et al. 2004).

Außerhalb des Gehirns liegt die häufigste Anwendung der Bildfusion in der onkologischen Diagnostik. Hier ergibt sich der diagnostische Gewinn ebenfalls aus der Möglichkeit, insbesondere fokale Glucosemetabolismussteigerungen einer FDG-PET-Untersuchung beispielsweise einem Lymphknoten zuzuordnen und so von physiologischer Aktivitätsanreicherung z.B. in den ableitenden Harnwegen oder in Muskulatur abzugrenzen (Wahl, Quint et al. 1993). Auch in der Strahlentherapie findet die Bildfusion hier Anwendung. Die Fusion einer CT mit einer PET erlaubt – wie bei Hirntumoren - eine Fokussierung der CT-gesteuerten Dosisverteilung auf vitale Tumoranteile (Munley, Marks et al. 1999), oder die zusätzliche Detektion befallener Lymphknoten, die so in die Bestrahlung eingeschlossen werden können (Scarfone, Lavelly et al. 2004).

Der wesentliche erste Schritt jeder multimodalen Bildgebung ist die Bildfusion, also der Vorgang, bei dem die dreidimensionalen Datensätze der verschiedenen Untersuchungen geometrisch exakt zur Deckung gebracht werden, so dass beide

Untersuchungen in jedem Bildpunkt die gleiche anatomische Struktur darstellen. Für diesen Prozess steht eine Vielzahl an manuellen oder automatischen Methoden zur Verfügung (Hawkes 1998; Myers 2002). Jede klinische Anwendung einer Bildfusion erfordert eine Qualitätskontrolle, um die Verlässlichkeit des verwendeten Verfahrens einschätzen zu können. Dies gilt insbesondere, wenn therapeutische Konsequenzen vom diagnostischen Ergebnis der Bildfusion anhängig sind. Die Daten zur Genauigkeit von Bildfusionen sind in der Literatur jedoch sehr inhomogen wiedergegeben, häufig in Form einzelner Parameter des Fusionsprozesses. Hieraus lässt sich vom Arzt nur schwer eine konkrete klinisch relevante Abschätzung der erzielten Genauigkeit ableiten. Auch wenn Fehlerabschätzungen in Form einer konkreten Distanz in Millimetern angegeben sind, sind diese zwischen unterschiedlichen Studien nicht direkt vergleichbar, da sie auf sehr unterschiedlichen Referenzsystemen und Berechnungsmethoden basieren.

Ein Ziel der vorliegenden Arbeit war daher die Entwicklung eines universell anwendbaren Fehlermodells zur Beurteilung der Qualität verschiedener Bildfusionsmethoden in Form einer absoluten Angabe des Fusionsfehlers in Millimetern.

Bei der Auswahl nach der geeigneten Methode für die Bildfusion wird man mit einer Vielzahl methodisch oft sehr komplexer automatischer Algorithmen konfrontiert, deren exakte Vorgehensweise einem Nichtmathematiker oft nur schwer zu erschließen ist. Die vorliegenden Arbeiten zum Vergleich von Bildfusionsmethoden beschränken sich meist auf den Vergleich verschiedener *automatischer* Verfahren (Strother, Anderson et al. 1994; West, Fitzpatrick et al. 1997), es fehlen Daten zum direkten Vergleich automatischer und *manueller*, also interaktiv vom Benutzer geführter Methoden. Zweites Ziel der Arbeit war daher die Anwendung des erarbeiteten Fehlerberechnungsmodells zum direkten Vergleich der am besten etablierten automatischen Bildfusionsmethoden mit einem manuellen Ansatz für die Bildfusion von MRT und SPECT Aufnahmen des menschlichen Gehirns.

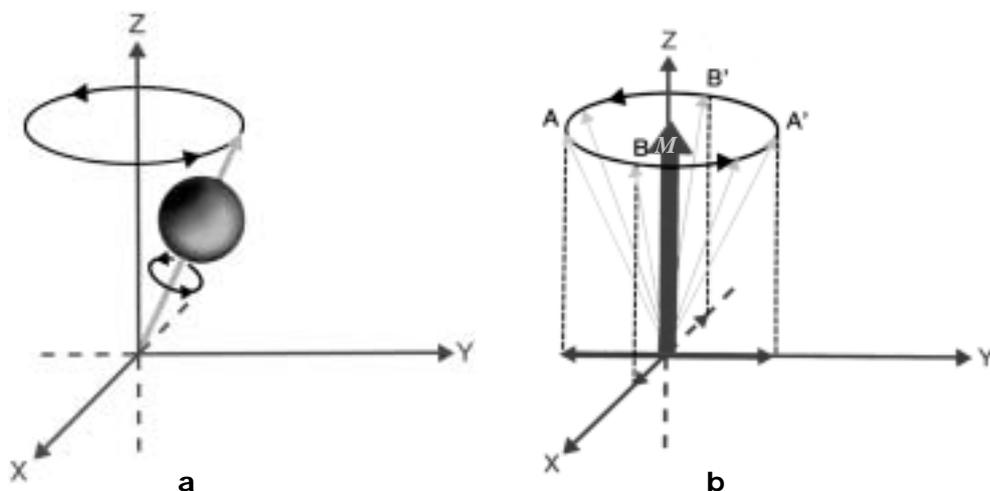
## 2 Grundlagen

### 2.1 Theoretische Grundlagen der Untersuchungstechniken

#### 2.1.1 Magnetresonanztomographie

Grundlage der Magnetresonanztomographie oder Kernspintomographie ist der als Kernspin bezeichnete Eigendrehimpuls von Wasserstoffatomkernen. Der Wasserstoffatomkern besteht aus nur einem Proton mit positiver Ladung, die Rotation dieses geladenen Kernteilchens erzeugt ein kleines, atomares Magnetfeld, ein magnetisches Dipolmoment. Das System ist mit einem um seine Längsachse rotierenden Stabmagneten vergleichbar (Abbildung 1a).

Ohne äußere Einflüsse sind die magnetischen Momente der Protonen in der Materie zufällig ausgerichtet, so dass sich ihre Gesamtwirkung nach außen neutralisiert. Wirkt ein äußeres Magnetfeld auf die Kerne ein, richten sich die magnetischen Momente der Protonen parallel oder antiparallel zum externen Magnetfeld aus. Da die Ausrichtung parallel zum externen Magnetfeld energetisch günstiger ist, liegt ein leichter Überschuss an parallel ausgerichteten Protonen vor. Die unterschiedlichen Querkomponenten der Magnetisierung sind zufällig verteilt und neutralisieren sich daher gegenseitig (Abbildung 1b: A und A', B und B'). Für die Gesamtheit aller Spins ergibt sich damit eine makroskopische Kernmagnetisierung  $M$  parallel zum externen Magnetfeld.



**Abbildung 1:** Darstellung eines einzelnen Protons (a) sowie des Gesamtmagnetisierungsvektors  $M$  unter dem Einfluss eines externen Magnetfeldes  $B_0$  parallel zur Z-Achse (b) (Schild 1990).

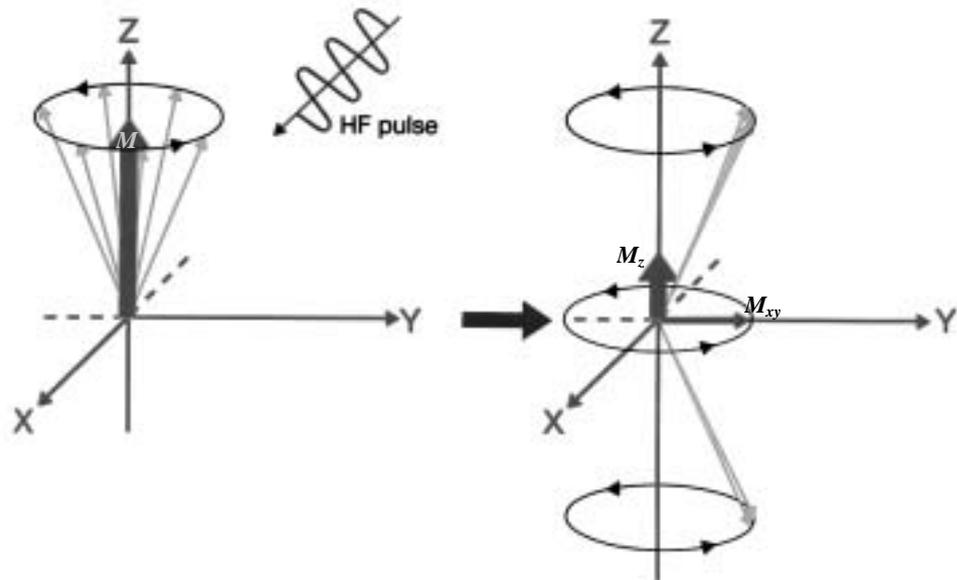
Durch Aufnahme oder Abgabe eines bestimmten Energiequants sind jeweils Übergänge zwischen der parallelen und der antiparallelen Ausrichtung möglich. Ein zusätzliches elektromagnetisches Feld kann also durch Energieaustausch mit den Kernspins in Wechselwirkung treten. Eine solche Wechselwirkung ist jedoch nur möglich, wenn das elektromagnetische Feld mit einer definierten Frequenz auf die Kerne einwirkt. Diese Frequenz  $\omega_L$  wird Resonanzfrequenz oder Larmorfrequenz genannt und ist von spezifischen Eigenschaften der Atomkerne und von der Stärke des angelegten externen Magnetfeldes abhängig:

$$\omega_L = \gamma \cdot B_0$$

Hierbei steht  $\gamma$  für die charakteristische gyromagnetische Konstante des Atomkerns und  $B_0$  für die Stärke des externen statischen Magnetfeldes. Für den Wasserstoffkern beträgt die gyromagnetische Konstante ca. 42,6 MHz/T. In einem Magnetfeld der Stärke von 1,5 Tesla ergibt sich damit eine Larmorfrequenz von ca. 63,9 MHz bei der eine energetische Wechselwirkung mit den Kernen möglich ist.

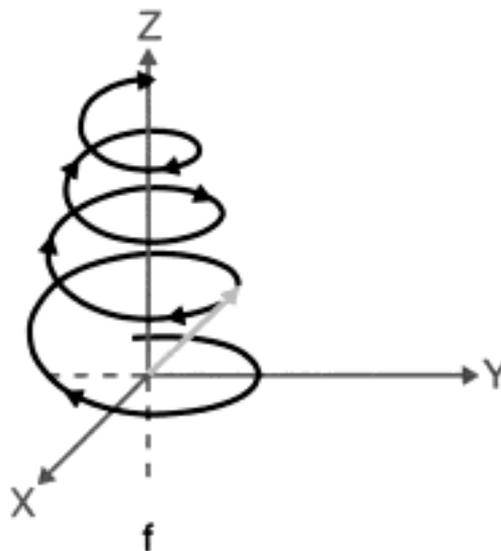
Durch das kurzzeitige Anlegen eines hochfrequenten elektromagnetischen Feldes  $B_1$  mit der Frequenz  $\omega_L$  - einen so genannten HF-Puls - der senkrecht zum statischen Magnetfeld  $B_0$  ausgerichtet ist kann die makroskopische Kernmagnetisierung aus dem Gleichgewicht gebracht werden. Der HF-Puls bewirkt eine Auslenkung der Gesamtkernmagnetisierung  $M$  gegenüber dem statischen Magnetfeld  $B_0$  um einen Winkel  $\alpha$ , der als Flipwinkel bezeichnet wird. Durch Wahl entsprechender Parameter für den HF-Puls ist der Flipwinkel von wenigen Grad bis zu 180 Grad, also einer vollständigen Umkehr der Magnetisierungsrichtung steuerbar.

Der Richtungsvektor der Kernmagnetisierung  $M$  kann nach der Auslenkung durch den HF-Puls in zwei Komponenten aufgeteilt werden: die jetzt reduzierte Längsmagnetisierung  $M_z$  parallel zu  $B_0$  und die neu entstandene Quermagnetisierung  $M_{xy}$  senkrecht zu  $B_0$  (Abbildung 2). Durch den Eigendrehimpuls der Kernmagnetisierung ändert die Quermagnetisierung  $M_{xy}$  ständig ihre Orientierung, sie rotiert mit der Larmorfrequenz  $\omega_L$  um die Richtung des Magnetfeldes  $B_0$ .



**Abbildung 2:** Die Auslenkung der Gesamtmagnetisierung  $M$  durch einen HF-Puls erzeugt eine Quermagnetisierungskomponente  $M_{xy}$ , senkrecht zur Richtung des externen Magnetfeldes  $B_0$  (= Richtung der Z-Achse), die mit der Larmorfrequenz  $\omega_L$  um die Z-Achse rotiert (Schild 1990).

Gleichzeitig bewirkt das weiterhin bestehende Feld  $B_0$  ein Drehmoment auf die ausgelenkte Kernmagnetisierung zurück in die parallele Ausrichtung. Es kommt damit zu einer spiralförmigen, abnehmenden Präzession von  $M$  (Abbildung 3).



**Abbildung 3:** Spiralförmige Präzession der Hauptmagnetisierungsrichtung zurück in Richtung der Z-Achse unter dem Einfluss des weiter bestehenden Magnetfeldes  $B_0$  (Schild 1990).

Die Präzession der Quermagnetisierung  $M_{xy}$  stellt physikalisch eine sich bewegende, elektrische Ladung in einem Magnetfeld dar und induziert damit ein mit der Frequenz  $\omega_L$  oszillierendes, elektrisches Feld, dessen Amplitude sich mit Abnahme der Präzession verringert. Dieses durch die Präzessionsbewegung induzierte, elektrische Feld ist das eigentliche Signal, das von den Empfangsspulen gemessen wird und zur Bildrekonstruktion der Magnetresonanztomographie verwendet wird.

Die Rückbildung der Quermagnetisierung  $M_{xy}$  und die Wiederherstellung der Längsmagnetisierung  $M_z$  in Richtung des Feldes  $B_0$  wird als Relaxation bezeichnet. Das Relaxationsverhalten, also die Geschwindigkeit, mit der sich  $M_{xy}$  und  $M_z$  unter Einfluss des statischen Feldes  $B_0$  wieder zurückbilden, ist von der molekularen Umgebung der Protonen abhängig und unterscheidet sich beispielsweise für Protonen je nachdem ob sie Bestandteil eines Wassermoleküls, einer Fettsäure oder von Hämoglobin sind.

Das Phänomen der Kernspinresonanz wurde ursprünglich in der chemischen Diagnostik zur Charakterisierung organischer Moleküle eingesetzt, erst später entstand die Idee, das Verfahren auch bildgebend einzusetzen. Nach ersten Anwendungen durch Paul Lauterbur und Peter Mansfield in den 70er Jahren (die dafür 2003 mit dem Nobelpreis für Medizin ausgezeichnet wurden) kommt die MRT seit Anfang der 80er Jahre für die bildgebende Diagnostik beim Menschen zum Einsatz.

Einer der wesentlichen Vorteile der MRT für den klinischen Einsatz ist die Nebenwirkungsarmut. Da keinerlei ionisierende Strahlung zum Einsatz kommt, kann das Verfahren zum Beispiel auch bei Kindern, bei Bedarf auch wiederholt, eingesetzt werden. Nach kontinuierlicher Weiterentwicklung der Aufnahme- und Nachverarbeitungstechnik bietet die MRT heute einen sehr guten Weichteilkontrast. Im Gegensatz zur Computertomographie, die primär nur axiale Schnittbilder akquirieren kann, ermöglicht die MRT auch die Aufnahme primär koronar oder sagittal orientierter Schnittbilder, die für entsprechende Fragestellungen diagnostisch besser verwertbar sind. Wie für alle digital vorliegenden Bilddaten besteht auch die Möglichkeit, die Daten *nach* der Akquisition in anderen Ebenen zu rekonstruieren.

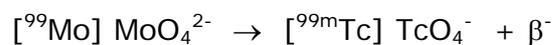
## 2.1.2 Single-Photonenemissions-Computertomographie

Grundlage aller nuklearmedizinischer Untersuchungen ist die Verwendung von radioaktiven Atomen, die in bestimmte physiologische Prozesse „eingeschleust“ werden. Hierzu wird dem Patienten ein untersuchungsspezifisches Radiopharmakon injiziert. Wie der Name Radiopharmakon bereits andeutet, besteht dieses aus einem organischen Molekül, dem „Pharmakon“ und einem zusätzlichen radioaktiven Liganden, der an das organische Molekül gekoppelt ist.

### 2.1.2.1 Radionuklide

Für die nuklearmedizinische Diagnostik werden niedrigenergetische  $\gamma$ -Strahler mit kurzer Halbwertszeit als Radionuklide bevorzugt. Die  $\gamma$ -Strahlung bewirkt im Vergleich zu korpuskulärer  $\alpha$ - oder  $\beta$ -Strahlung deutlich weniger Nebenwirkungen im Gewebe. Eine kurze Halbwertszeit, am besten im Bereich von Stunden, ist wichtig, um die Strahlenbelastung des Patienten nach der Untersuchung so gering wie möglich zu halten.

Für die Mehrzahl an Untersuchungen wird heute [ $^{99m}\text{Tc}$ ] Technetium eingesetzt, ein metastabiles Nuklid, das mit einer physikalischen Halbwertszeit von 6 Stunden unter Aussendung von Gammastrahlung mit einer Energie von 140 keV in seinen Grundzustand [ $^{99}\text{Tc}$ ] Technetium übergeht. [ $^{99m}\text{Tc}$ ] Technetium wird in der Regel in einem Generator durch  $\beta$ -Zerfall des Mutternuklids [ $^{99}\text{Mo}$ ] Molybdän gewonnen. Das Molybdän ist im Generator als negativ geladenes Molybdat  $\text{MoO}_4^{2-}$  an Aluminiumoxid gebunden. Das durch den  $\beta$ -Zerfall



entstehende Pertechnetat [ $^{99m}\text{Tc}$ ]  $\text{TcO}_4^-$  wird nicht gebunden und kann bei Bedarf mit Kochsalzlösung aus dem Generator eluiert werden (Brandau, Becker et al. 2001).

### 2.1.2.2 Radiopharmaka

Das so gewonnene [ $^{99m}\text{Tc}$ ] Technetium muss im nächsten Arbeitsschritt an einen passenden, untersuchungsspezifischen organischen Träger gebunden werden, was meist durch Komplexbildung erfolgt. So kann das Technetium beispielsweise an das rasch tubulär sezernierte Mercaptoacetyltriglycin (MAG3) gebunden

werden, um die Nierenfunktion darzustellen oder an patienteneigene Erythrozyten, die nach Markierung reinjiziert werden, wenn eine Blutungsquelle gesucht wird.

In dieser Arbeit wurden Untersuchungen mit [ $^{99m}\text{Tc}$ ] Ethylcysteinatdimer (ECD) als Radiopharmakon verwendet. [ $^{99m}\text{Tc}$ ] ECD dient zur Darstellung der regionalen Hirnperfusion zum Zeitpunkt der Injektion. Wegen seiner guten Diffusionsfähigkeit in das Hirnparenchym weist das lipophile ECD eine hohe Extraktionsrate aus dem Blut von circa 50% bei der ersten Kapillarpassage auf. Nach Diffusion in das Hirnparenchym erfolgt eine enzymatische Trennung des Dimers. Die entstandenen Monomere sind polar, was eine Rückdiffusion durch die lipophile Zellmembran weitgehend verhindert. So kommt es zu einer intrazellulären Akkumulation und damit zu einer ausreichend langen stabilen Retention des Radiopharmakons im Hirnparenchym. Dadurch ist es möglich, bis zu einige Stunden nach der Injektion szintigraphische Aufnahmen anzufertigen, deren Aktivitätsverteilung die Hirnperfusion zum Injektionszeitpunkt widerspiegelt. Nicht extrahierte Anteile des injizierten ECD werden relativ rasch renal aus dem Blutpool eliminiert, was die Hintergrundaktivität bei der Aufnahme reduziert und den Kontrast verbessert (O'Brien, Brinkmann et al. 1999).

### 2.1.2.3 Gammakamera

Nach Injektion und entsprechender Verteilung des Radiopharmakons wird mittels einer Gammakamera von außen die zuvor injizierte Radioaktivität lokalisiert.

Eine Gammakamera zur Detektion der Strahlung besteht aus folgenden Komponenten: die vorderste Lage bildet der Kollimator, der aus dünnen Bleilamellen besteht und als Filter für unerwünschte Streustrahlung dient: Strahlung die schräg auftrifft wird an den Bleilamellen absorbiert, nur Strahlung die senkrecht von vorne auf den Kollimator trifft, kann diesen ungehindert passieren. Nach der Passage des Kollimators trifft die Strahlung auf einen Szintillationskristall, meist auf einen bis zu 40 cm großen Natriumiodid-Kristall. In diesem löst die einfallende Gammastrahlung einen sichtbaren Lichtblitz aus. Durch flächendeckenden Kontakt des Kristalls zu einer Matrix aus elektronischen Photomultipliern wird ein solcher Lichtblitz registriert, lokalisiert und an den Auswertrechner übertragen.

Eine über einen längeren Zeitraum aufgenommene Szintigraphie zeigt dann ein Summationsbild aller senkrecht zur Detektorebene eingefallenen Gammastrahlen. Für viele Anwendungen, wie die Schilddrüsenszintigraphie oder Skelettszintigraphie ist eine solche Projektionsaufnahme eine geeignete Darstellung. Um statt einer Gesamtprojektion einzelne Schichten aus dem Körper darzustellen, ist eine Tomographie notwendig. Hierfür werden mehrere Aufnahmen aus verschiedenen Projektionsrichtungen angefertigt, meist unter Verwendung von zwei oder drei Kameraköpfen, die ringförmig montiert, in definierten Winkelschritten um den Patienten rotieren. Aus diesem Satz an Projektionsdaten kann dann durch Rückprojektion ein Volumendatensatz generiert werden, der sich als Schnittbildserie in allen Raumebenen rekonstruieren lässt.

## 2.2 Bildfusion - Prinzipien und Methodenübersicht

Mit Hilfe der Bildfusion können die komplementären Informationen verschiedener bildgebender Verfahren – hier der Magnetresonanztomographie und der ECD-SPECT kombiniert werden, nachdem beide Datensätze geometrisch zur Deckung gebracht wurden (van den Elsen, Evert et al. 1993; Hutton, Braun et al. 2002).

Der mathematische Vorgang der Abbildung eines dreidimensionalen Datensatzes auf einen zweiten Datensatz taucht in der Literatur unter verschiedenen Begriffen wie „Bildfusion“, „Bildüberlagerung“ oder „Bildregistrierung“ beziehungsweise in der internationalen Literatur als „Image Fusion“, „Image Registration“, „Image Matching“ oder „Geometric Correlation“ auf. Im Folgenden soll zur Beschreibung dieses Vorganges der Terminus „Bildfusion“ verwendet werden.

In der konkreten Anwendung wird für die Bildfusion immer ein Datensatz als Primärdatensatz definiert und unverändert belassen, während ein zweiter Sekundärdatensatz entsprechend im Raum bewegt und an die Position des Primärdatensatzes angepasst wird.

Bei der Bildfusion gibt es verschiedene Freiheitsgrade zu berücksichtigen. Der einfachste Fall einer Fusion zweier dreidimensionaler Datensätze ist die sog. „starre Transformation“, d.h. der bewegte Sekundärdatensatz wird als starrer Körper behandelt. Um einen starren, dreidimensionalen Körper auf einen zweiten abzubilden, sind sechs Parameter notwendig: drei Verschiebungen (im folgenden Translation genannt) und drei Rotationen, jeweils entlang der orthogonalen Achsen eines Koordinatensystems. Mit diesen sechs Parametern kann jede Position eines starren Körpers im Raum definiert werden.

Voraussetzung für die Anwendung einer starren Transformation ist, dass die Pixelgrößen beider Datensätze bekannt und richtig sind, so dass keine Skalierung zur Größenkorrektur notwendig ist. Dies ist bei medizinischen Bilddaten durch regelmäßige Qualitätskontrollen und Kalibrierung der Aufnahmegereäte gewährleistet. Weiterhin wird angenommen, dass keine korrekturbedürftigen geometrischen Verzerrungen innerhalb eines Datensatzes vorkommen, so dass die beiden Datensätze nach dieser Transformation optimal aufeinander passen. Unter Zugrundelegung dieser Annahmen lassen sich verschiedene Untersuchungen *eines* Patienten mit gleichen anatomischen Verhältnissen mittels der

starren Transformation zur Deckung bringen (van den Elsen, Evert et al. 1993; Hawkes 1998; Woods, Grafton et al. 1998).

Werden die Daten verschiedener Patienten fusioniert, oder haben sich die anatomischen Verhältnisse eines Patienten deutlich verändert, sind zusätzlich zu den sechs Freiheitsgraden der starren Transformation weitere Freiheitsgrade notwendig, um die individuellen anatomischen Abweichungen auszugleichen: es kommen zusätzliche Skalierungen, Stauchungen oder Verzerrungen des Datensatzes zur Anwendung. Im klinischen Einsatz wird eine solche nichtstarre, „elastische“ Transformation z.B. verwendet, um individuelle Patientendaten mit einem Normkollektiv oder einem Atlas zu vergleichen (Collins, Neelin et al. 1994; Kostelec, Weaver et al. 1998; Guimond, Roche et al. 2001; Van Laere, Warwick et al. 2002), oder wenn sich die anatomischen Verhältnisse, z.B. durch raumfordernde Prozesse, deutlich verschoben haben.

Aufgrund der methodischen Komplexität der nichtstarren Transformation wird im weiteren Verlauf nur der Fall der *starren Transformation* zur Fusion intraindividuell-eller Daten betrachtet.

### **2.2.1 Manuelle Methoden**

Manuelle Verfahren zur Bildfusion stützen sich - wie das häufig gebrauchte Synonym "interaktive Methoden" zeigt - auf die Interaktion mit dem Benutzer, der die geometrische Transformation des sekundären Datensatzes manuell durchführt. Dieser Vorgang basiert vollständig auf seiner visuellen Kontrolle und Einschätzung der Qualität der erzielten Fusion, es finden keinerlei automatische Prozesse statt. In einem entsprechenden Userinterface wird dem Benutzer in der Regel eine farbcodierte, kombinierte Ansicht beider Datensätze in mehreren Ebenen angeboten, häufig wird auch der Sekundärdatensatz durch Konturlinien oder „Isolinien“ entlang bestimmter Pixelwerte dargestellt. (Pietrzyk, Herholz et al. 1990). Die eigentliche Transformation des sekundären Datensatzes erfolgt entweder durch direktes anklicken und verschieben bzw. drehen mit der Maus, oder über ein separates Bedienfeld mit schrittweiser Ausführung kleiner Bewegungen oder Rotationen. Auch die direkte Eingabe numerischer Transformationsparameter ist möglich.

## 2.2.2 Automatische Methoden

Die Methoden zur automatischen Bildfusion stützen sich auf eine software-basierte, automatisierte Durchführung bzw. Optimierung der Bildfusion. Die einzelnen Methoden berücksichtigen dabei unterschiedliche Bildmerkmale, die zur Abschätzung der Qualität einer Bildfusion herangezogen werden. Die Mehrzahl der Programme erfordern zunächst einige Bearbeitungsschritte vom Benutzer und werden deshalb auch als „halbautomatische“ Methoden bezeichnet. Hier liegen allerdings fließende Übergänge vor, eine eindeutige Zuordnung von Methoden gibt es nur an den beiden Extremen des Spektrums: die rein manuell durchgeführte Bildfusion und die vollautomatische Bildfusion ohne jegliche Vorverarbeitung. Dazwischen liegen die halbautomatischen Methoden, die je nach Intensität der notwendigen oder möglichen Benutzerinteraktion näher an den manuellen oder näher an den vollautomatischen Methoden liegen. Im Folgenden werden daher alle Methoden bei denen letztendlich ein automatisierter Optimierungsprozess stattfindet, als "Automatische Methoden" zusammengefasst, unabhängig vom Umfang einer eventuellen vorherigen Benutzerinteraktion.

### 2.2.2.1 Prinzipien der iterativen Bildfusion

Die meisten automatischen Algorithmen zur Bildfusion arbeiten nach dem iterativen Prinzip, das aus folgenden drei Schritten besteht:

1. Merkmalsextraktion
2. Fehlerdefinition
3. Fehlerminimierung

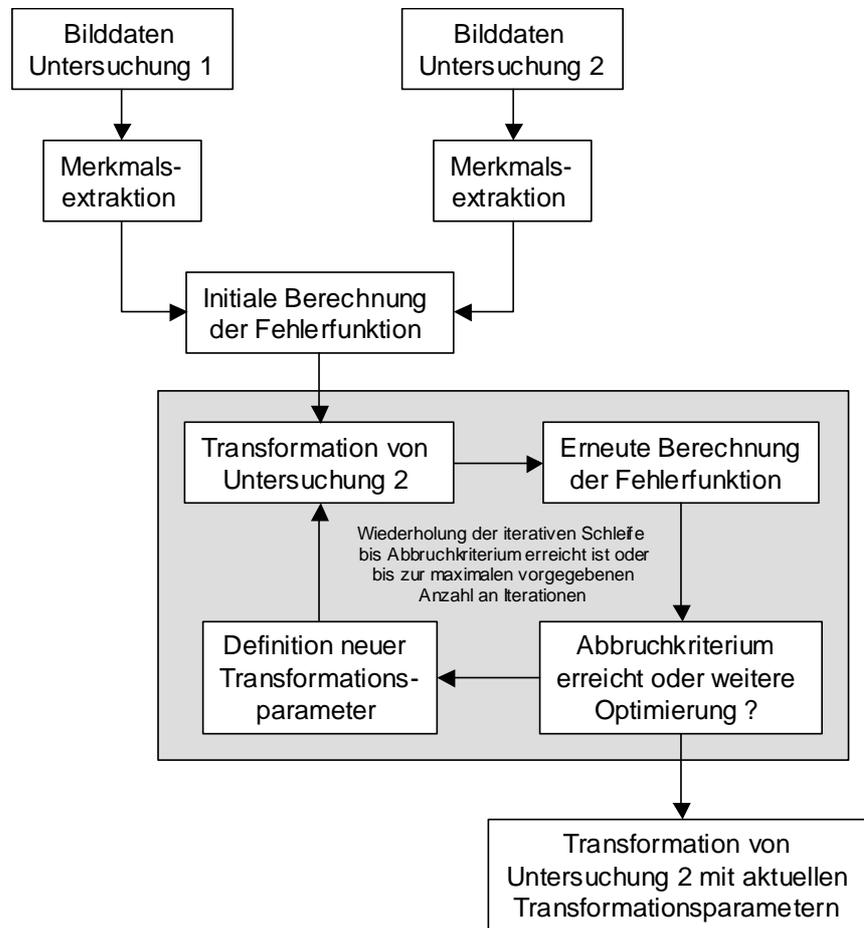
die in Abbildung 4 schematisch zusammengefasst sind.

Zuerst findet eine Form von Mustererkennung in beiden Studien statt. Hierfür kommen sowohl eine vom Benutzer durchgeführte Segmentierung oder Punktauswahl, als auch eine automatische Erkennung bestimmter Datenbereiche in Frage. Häufig verwendet werden hier die Oberflächen eines oder mehrerer Organe, oder eine Einteilung in bestimmte Graustufenbereiche. Unabhängig von der Komplexität dieses Schrittes wird letztendlich eine Teilinformation aus dem Datensatz als fusionsrelevantes Merkmal extrahiert.

Für die Beurteilung eines einzelnen Fusionsversuchs, wird nun ein bestimmtes Fehlermaß definiert, das möglichst gut die "Differenz" zwischen den beiden Datensätzen beschreibt. Das Ergebnis ist eine Fehlerfunktion, die von den sechs Transformationsparametern abhängt und ein Maß für die Qualität der derzeitigen Fusion darstellen soll. Werte, die dafür in Frage kommen, sind beispielsweise die Entfernung zwischen korrespondierenden Punkten, die Summe der Entfernungen zwischen Punkten auf Organoberflächen oder die Homogenität der Graustufenverteilung innerhalb definierter Gewebeklassen. Es wird postuliert, dass eine Reduktion dieses errechneten Fehlermaßes wirklich eine Verbesserung der Bildfusion widerspiegelt.

Der letzte Schritt ist die iterative Minimierung dieser Fehlerfunktion d.h. die Suche nach den Transformationsparametern, bei denen der definierte Fehler am kleinsten ist. Iterative Verfahren ermitteln nicht mit einer einzigen Berechnung die optimalen Parameter, sondern ändern nach einem bestimmten Prinzip die einzelnen Parameter, rekonstruieren eine Studie nach diesen Parametern und bestimmen für diese Transformation erneut das vorher festgelegte Fehlermaß. Nach einigen initialen „Blindversuchen“ erkennen die Algorithmen die Transformationen, die zu einer Verbesserung des Fehlermaßes geführt haben und können die Parameter gezielt „in der gleichen Richtung“ verändern. In der Regel sind bis zu mehrerer hundert solcher „Versuche“ nötig, um die optimale Transformation zu finden. Die durchschnittliche Rechenzeit für eine solche Iteration mit Berechnung des aktuellen Fehlerwertes liegt auf aktuellen Rechnern meist deutlich unter einer Sekunde. Die Minimierung der Fehlerfunktion erfolgt nach verschiedenen Verfahren der numerischen Mathematik, wobei sich die Verfahren Powell (Powell 1964), Simplex (Nelder und Mead 1965), Mark-1, Jakob, Newton-Raphson u.a. bewährt haben. Das Abbruchkriterium für die Suche nach dem optimalen Parametersatz kann variabel definiert werden, z.B. wenn mehrere aufeinander folgende Iterationen keine weitere oder nur eine geringe Verbesserung des Fehlers gebracht haben. Eine Fehlerquelle für diesen Minimierungsprozess stellen dabei lokale Minima der Fehlerfunktion dar, in denen ein solcher Suchalgorithmus fälschlicherweise das definierte Abbruchkriterium erreicht, obwohl das globale, „wahre“ Minimum der Fehlerfunktion noch nicht erreicht ist. Um diese Gefahr zu verringern, muss das Abbruchkriterium der Minimierung entsprechend definiert sein, was durch eine steigende Zahl

notwendiger Iterationen und damit längerer Rechenzeit erkauft wird. In jedem Fall ist eine abschließende, visuelle Kontrolle der Transformation notwendig.

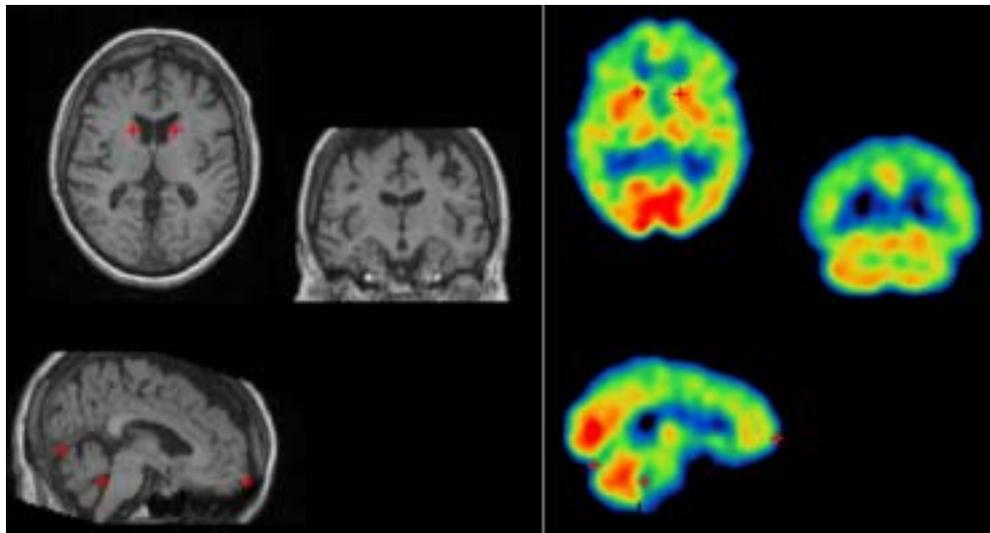


**Abbildung 4:** Flussdiagramm einer iterativen Bildüberlagerung. Nach Extraktion der für die Fehlerfunktion relevanten Bildmerkmale wird die Fehlerfunktion initial errechnet. Der Algorithmus versucht durch zunächst zufällige, später dann gezielte Veränderung der Transformationsparameter die Fehlerfunktion zu minimieren.

Die Weiterentwicklung von automatischen, iterativen Verfahren kann an allen drei Schritten einzeln ansetzen: die Merkmalsextraktion, die Definition der Fehlerfunktion sowie die Minimierungsmethode können unabhängig voneinander modifiziert und optimiert werden.

### 2.2.2.2 Landmarkenbasierte Methoden

Eine der am längsten etablierten Methoden für die Bildfusion ist die Verwendung von anatomischen Landmarken. Dabei markiert der Anwender manuell eine möglichst große Anzahl von anatomisch klar definierten Landmarken, die in beiden zu überlagernden Studien sicher erkennbar sind (Abbildung 5). Nach Identifikation von mindestens drei solcher korrespondierenden Landmarken wird durch eine Funktion zur Fehlerminimierung die Transformation bestimmt, welche die definierten Landmarken mit der geringsten Abweichung zur Deckung bringt. Erwartungsgemäß steigt bei dieser Methode die erzielbare Genauigkeit zunächst mit der Anzahl der verwendeten Punktepaare, da Ungenauigkeiten in der Position einzelner Landmarken so weniger ins Gewicht fallen.



**Abbildung 5:** Beispiel für mögliche anatomische Landmarken, die in beiden Untersuchungsmodalitäten (MRT und ECD-SPECT) sichtbar sind: der Kopf des Nucleus caudatus, das Tentorium, der vierte Ventrikel und der Frontalpol

Es sind auch Programme beschrieben, die nach jedem markierten Punktepaar die Transformation anpassen, die Ansicht der fusionierten Daten aktualisieren, und so dem Benutzer eine schrittweise, interaktive, immer genauer werdende Optimierung der Fusion erlauben (Habboush, Mitchell et al. 1996; Treves, Mitchell et al. 1998).

Eine Variante dieser Technik für die Fusion von Gehirnstudien wurde zuerst von Kapouleas (Kapouleas, Alavi et al. 1991) beschrieben und von Ge (Ge, Fitzpatrick et al. 1994) und Wang (Wang, Pahl et al. 1994) aufgegriffen: zuerst wird in

beiden Studien in jeder Schicht manuell der Interhemisphärenspalt durch zwei Punkte markiert, woraus die sagittalen Mittelschichten der beiden Studien bestimmt werden. Nachdem diese automatisch zur Deckung gebracht wurden, ist nur noch ein Verschieben und Rotieren innerhalb der medianen Sagittalebene nötig, wobei als Orientierungspunkte die beiden Kommissuren oder die äußeren Hirnkonturen verwendet werden können.

Für alle landmarkenbasierten Verfahren gilt, dass der relative Anteil der Benutzerinteraktion am Gesamtergebnis der Bildfusion relativ hoch ist.

### 2.2.2.3 externe Marker

Bei der Fusion mittels externer Marker werden am Patienten fixierte Markierungen verwendet. Die Marker müssen in allen geplanten Untersuchungen sichtbar sein, was häufig nicht mit einem einzigen Material möglich ist. Deshalb basieren viele Marker auf Halterungen, die für unterschiedliche Untersuchungsmodalitäten verschiedene Einsätze aufnehmen können: z.B. Kunststoff- oder Metallkugeln für MR- und CT-Untersuchungen und mit Radionukliden gefüllte Hohlkugeln für nuklearmedizinische Untersuchungen. Für die Positionierung der Marker kommen verschiedene Möglichkeiten in Frage: Die einfachste Form sind oberflächliche, auf der Haut des Patienten aufgeklebte Markierungspunkte (Turkington, Jaszczak et al. 1993; Elsen und Viergever 1994; Turkington, Hoffman et al. 1995). Eine zuverlässigere aber auch invasivere Methode sind in den Schädelknochen eingedrehte Schrauben, an denen verschiedene Markierungen (Black, Videen et al. 1996; Maurer, Fitzpatrick et al. 1997) oder ein stereotaktischer Rahmen (Woods, Mazziotta et al. 1993; Ge, Fitzpatrick et al. 1994; West, Fitzpatrick et al. 1997) befestigt werden können. Außerdem existieren eine Vielzahl von Halterungen, die die eigentlichen Marker aufnehmen und auch eine spätere Wiederpositionierung des Markers ermöglichen. Die Palette reicht hier von individuell anmodellierten Gesichtsmasken aus thermoplastischem Kunststoff bis zu zwischen den Zähnen oder im äußeren Gehörgang fixierten Halterungen (Kaiser, Sabri et al. 1994; Bale, Vogele et al. 1997; Bale, Burtscher et al. 2000).

Mit steigender Anzahl der verwendeten Marker lässt sich die Genauigkeit bis zu einem gewissen Grad erhöhen (Sipila, Nikkinen et al. 1997). West et al. konnten

zeigen, dass die Positionierung externer Hautmarker einen wesentlichen Einfluss auf die erzielbare Genauigkeit hat und empfehlen eine Positionierung die gleichmäßig über beide Körperseiten verteilt ist und eine lineare Anordnung mehrerer Marker vermeidet. Der Schwerpunkt der verwendeten Marker sollte möglichst nahe der Zielregion liegen (West, Fitzpatrick et al. 2001).

Für die Fusion zweier Studien werden diese externen Markerpunkte in beiden Studien manuell markiert oder automatisch erkannt (Krishnan, Hermann et al. 2003) und die Software bestimmt analog zum Vorgehen bei internen anatomischen Landmarken mit verschiedenen mathematischen Ansätzen die Transformation, mit der die Punktepaare optimal zur Deckung gebracht werden. Die erzielbare Genauigkeit der Bildfusion mittels externer Marker wird teilweise mit 1,4 mm angegeben (Maurer, Fitzpatrick et al. 1997), kann jedoch bei ungünstiger Platzierung der Marker auch über 10 mm liegen (West, Fitzpatrick et al. 2001). Strother et al. kamen zu dem Ergebnis, dass alle anderen untersuchten Verfahren, einschließlich der Methoden nach Pelizzari und Woods, den externen Markern überlegen sind (Strother, Anderson et al. 1994).

Ein prinzipieller Nachteil aller auf externen Markern basierenden Verfahren ist der prospektive Charakter. Häufig ergibt sich jedoch erst nach Akquisition der ersten bildgebenden Untersuchung die Notwendigkeit einer multimodalen Diagnostik und einer Bildfusion. In der ersten durchgeführten Untersuchung stehen in solchen Fällen keine externen Marker zur Verfügung. Auch für die hier durchgeführten Untersuchungen schied die Methode daher aus.

#### 2.2.2.4 Surface matching nach Pelizzari

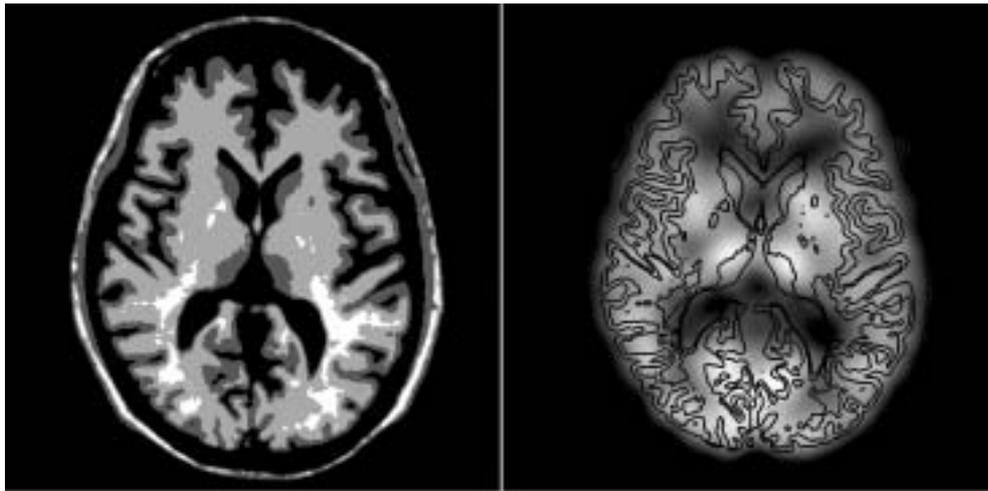
Pelizzari et al. haben 1989 mit dem „surface matching“ Algorithmus einen neuen, routinetauglichen Algorithmus zur Bildfusion vorgestellt (Pelizzari, Chen et al. 1989). Auch hierbei handelt es sich um einen iterativen Algorithmus. Als extrahiertes Bildmerkmal wird die äußere Kontur eines Organs verwendet, d.h. für die Anwendung am Gehirn muss in der MRT zunächst die Oberfläche des Gehirns vom Benutzer manuell segmentiert werden. Teilautomatisierte Verfahren zur Hirnsegmentierung existieren zwar, sind aber häufig fehleranfällig, insbesondere im Bereich der Schädelbasis. Für eine ECD-SPECT Untersuchung ist diese Segmentierung deutlich einfacher: da sich das injizierte Radiopharmakon ohnehin nur im Hirnparenchym anreichert, kann meist ein einfaches Schwellenwertverfahren angewendet werden, das alle zusammenhängenden Bildpunkte oberhalb eines bestimmten Pixelwertes auswählt und daraus die Organoberfläche definiert. Der Pelizzari-Algorithmus bestimmt nun iterativ die Transformation, die die Hirnoberflächen aus beiden Untersuchungen mit der geringsten Differenz zur Deckung bringt. Fehlermaß ist hier die Summe der Quadrate der Abstände zwischen den Oberflächen. Die Methode wird auch als „head-hat“ Verfahren bezeichnet, da gewissermaßen versucht wird, der einen „Kopfstudie“ die Oberfläche der anderen Studie wie einen „Hut“ anzupassen und aufzusetzen. Die mit dieser Methode erzielbare Qualität der Fusion hängt von der Komplexität der Oberfläche bzw. von der Anzahl und räumlichen Anordnung der segmentierten Organe ab. Für ein einzelnes, weitgehend kugelförmiges Organ besteht natürlich bei Betrachtung der reinen Oberflächenkontur ein großes Risiko für Rotationsfehler. Neben der Fusion verschiedener bildgebender Daten wird das surface matching auch für die Kalibrierung von Neuronavigationssystemen eingesetzt. Hierbei wird die Kopfoberfläche des Patienten im OP-Saal mittels eines 3D Laserscanners erfasst und wird dann mit der Kopfoberfläche des für die Navigation vorgesehenen Datensatzes fusioniert (Miga, Sinha et al. 2003).

#### 2.2.2.5 Pixel uniformity matching nach Woods

Der zur Zeit wohl am häufigsten angewendete und am meisten variierte Algorithmus zur automatischen Bildfusion wurde 1993 von Woods et al vorgestellt (Woods, Mazziotta et al. 1993). Die zugrunde liegende Überlegung war, dass verschiedene Untersuchungsmodalitäten zwar nie absolut gleiche Pixelwerte für ein Organ aufweisen, dass aber einzelne Organe in beiden Modalitäten jeweils konstante Pixelwerte haben werden, d.h. die Struktur von Organen sollte in beiden Modalitäten anhand homogener Pixelwerte zu erkennen sein.

Aufgrund dieser Annahme wird die MRT als primärer Datensatz in bestimmte Graustufenbereiche aufgeteilt, von denen postuliert wird, dass sie eine bestimmtes "Organ" darstellen. Dieses Organ sollte also auch in der nuklearmedizinischen Untersuchung konstante Pixelwerte aufweisen. Als Maß für die Qualität der Fusion wird daher die Homogenität der Pixelwerte ("pixel uniformity") der nuklearmedizinischen Studie innerhalb jeder - aus MR Daten definierten - „Struktur“ verwendet (Abbildung 6). Dieses Fehlermaß wird nun iterativ minimiert, bis die optimale Transformation bestimmt wurde.

Die Methode taucht in der Literatur unter verschiedenen Namen auf: z.B. als „corresponding PET intensity variation“ bei (Studholme, Hill et al. 1997), als „correlation of intensity values“ bei (West, Fitzpatrick et al. 1997) oder „ratio image uniformity“ bei (Strother, Anderson et al. 1994) (das „ratio image“ ist hier ein Bild, in dem der Quotient der Pixelwerte beider Studien dargestellt ist). Bei verschiedenen Implementierungen dieser Methode ist eine vorherige manuelle Segmentierung der zu berücksichtigenden Bereiche eines Datensatzes notwendig.



**Abbildung 6:** Der Algorithmus nach Woods unterteilt den primären Datensatz in Bereiche mit gleichen Pixelwerten, hier symbolisch durch vier Graustufen in der MRT dargestellt. Die Konturen dieser Einteilung werden auf den zweiten Datensatz übertragen und es wird gefordert, dass auch hier möglichst homogene Pixelwerte innerhalb der "Organkonturen" vorliegen. (Woods, Mazziotta et al. 1993).

Modifiziert wurde das Verfahren u.a. von Ardekani et al. (Ardekani, Braun et al. 1995): statt einer einfachen, graustufenbasierten Segmentierung der MRT wurde hier ein neuronales k-means clustering Verfahren angewendet, das den Datensatz in sog. „connected components“ mit gleichen Pixelwerten aufteilt. Die Berücksichtigung der topographischen Lage der Pixel entspricht dabei eher einer realistischen „Organsegmentierung“.

#### 2.2.2.6 Mutual Information

Inzwischen hat sich als weiterer automatischer Algorithmus die Bildfusion nach dem Prinzip der mutual information etabliert, die – ähnlich wie die Methode nach Woods – auf der Ähnlichkeit der Pixelwerte zweier Studien basiert (Studholme, Hill et al. 1997; Zhu 2002; Chen und Varshney 2003).

Die Methode orientiert sich dabei an der statistischen Abhängigkeit der Pixelwerte von korrespondierenden Bildpunkten. D.h. die „Vorhersagbarkeit“ des Pixelwertes in einer Studie in Abhängigkeit vom Pixelwert der anderen Studie wird betrachtet. Anders formuliert wird untersucht, wie viel Information aus dem einen Datensatz über den anderen Datensatz gewonnen werden kann, was den

englischen Begriff „wechselseitige Information“ erklärt. Diese ist bei korrekt fusionierten Bildern am höchsten, da nur dann gleiche anatomische Strukturen zur Deckung kommen und die Information über den Pixelwert einer Studie eine Prognose über den Wert in der anderen Untersuchung zulässt.

In der Informationstheorie wird als Maß für den Informationsgehalt auch der Begriff der Entropie verwendet (Shannon 1948). Dies schlägt sich in der Bezeichnung der Algorithmen nach dem Prinzip der mutual information nieder, die in der Literatur auch unter Termini wie „entropy-based registration“, „cross-entropy optimization“ oder „joint entropy registration“ auftauchen.

Für die Genauigkeit der Bildfusion mittels mutual information liegen heterogene Literaturdaten vor, die Mehrzahl der Studien konnte jedoch keine Überlegenheit gegenüber dem pixel-uniformity-Algorithmus nach Woods belegen (He und Narayana 2002). In der Studie von West schnitt die Fusion mittels mutual information mit einem Fehler von 3,6 mm schlechter ab, als der Algorithmus nach Woods mit 2,3 mm Fehler (West, Fitzpatrick et al. 1997).

Thurfjell et al. untersuchten den Einfluß einer reduzierten Auflösung der Datensätze und von Filterung mit einem Gaußschen Weichzeichner zur Erhöhung der Verarbeitungsgeschwindigkeit bei der Fusion von MRT und Perfusions-SPECT mittels Mutual Information. Neben simulierten Studien wurden hier auch reale Bilddaten von Probanden untersucht, es lagen jeweils zwei SPECT Untersuchungen vor, die unabhängig voneinander mit der MRT fusioniert wurden. Bei Verwendung der vollen Auflösung ergaben sich hier Abweichungen von durchschnittlich 1,2 mm Translation und 1,7° Rotation (Thurfjell, Lau et al. 2000).

Fiedler et al. untersuchten die Fusion von ECD-SPECT Daten mit verschiedenen MRT-Untersuchungstechniken. Bei Verwendung einer dünn-schichtigen 3D MRT betrug die Dauer für die automatische Fusion mittels Mutual Information im Mittel knapp 4 Minuten. Die automatische Fusion führte hier nur in 20% der Fälle zu einem visuell korrekten Ergebnis, in 80% musste das Ergebnis manuell nachbearbeitet werden. In Fällen mit schlechtem Ergebnis der automatischen Bildfusion betrug die Dauer für diese manuelle Nachbearbeitung im Mittel 5,5 Minuten (Fiedler, Platsch et al. 2003). Der gesamte Zeitaufwand belief sich damit auf fast 10 Minuten, was gegenüber einer vollständig manuell durchgeführten Bildfusion kaum noch eine Zeitersparnis bedeutet.

### 2.2.2.7 Principal Axes

Ein vollautomatisches analytisches Verfahren zur Bildfusion ist die Hauptkomponentenanalyse (engl. principal axes analysis) (Alpert, Bradshaw et al. 1990; Rusinek, Tsui et al. 1993). Dabei werden beide Datensätze analysiert und es werden drei orthogonalen Achsen eines Koordinatensystems, die sog. Hauptachsen bestimmt. Der Ursprung dieses Koordinatensystems liegt im Massenschwerpunkt eines Datensatzes. Die Achsen werden entsprechend der Massenverteilung so angeordnet, dass sie möglichst symmetrisch durch den Datensatz laufen. Im einem rechteckigen Quader würden sie beispielsweise parallel zu den Außenkanten durch den Schwerpunkt laufen, in einem unregelmäßig geformten Organ parallel zu dessen Hauptlängs- und -querrichtung. Für die Fusion zweier Datensätze müssen nun nur die beiden Massenschwerpunkte aufeinandergelegt und die Hauptachsen aufeinanderrotiert werden. Mit diesem analytischen Lösungsansatz, der in einem Schritt die Hauptkomponentenanalyse und die daraus resultierende Transformation vornimmt, unterscheidet sich das principal axis Verfahren von den übrigen automatischen Ansätzen die iterativ, also in mehreren wiederholten Schritten, arbeiten. Im Vergleich zum surface matching zeigte sich in der Untersuchung von Rusinek (Rusinek, Tsui et al. 1993) ein größerer Fehler für die Hauptkomponentenanalyse. Dhawan et al. haben einen principal axes Algorithmus speziell für die Fusion von Datensätzen mit unterschiedlichem field of view vorgestellt, der iterativ das betrachtete field of view verändert und so bei nur teilweise überlappenden Volumina bessere Ergebnisse liefert (Dhawan und Arata 1992).

## 2.3 Ziele der Arbeit

Wie jeder Schritt in der Nachverarbeitung digitaler medizinischer Bilddaten muss auch die Bildfusion Gegenstand von Qualitätskontrollen sein. Hierfür sind zwei Punkte zu klären:

Zum einen muss ein Goldstandard definiert werden, also die „richtige“ Position des transformierten Datensatzes im Raum, mit der das Ergebnis einer Bildfusion verglichen werden kann. Insbesondere bei der retrospektiven Bildfusion von Datensätzen die im Rahmen der klinischen Routine erhoben wurden ist dies problematisch, da hier oft kein verbindlicher Goldstandard vorliegt. Insbesondere werden in der Regel keine externen Referenzsysteme verwendet, auf die man sich später beziehen könnte.

Für den Vergleich mit dem Goldstandard muss weiterhin definiert werden, in welcher Form der Unterschied zweier Positionen eines dreidimensionalen Datensatzes im Raum angegeben werden soll. Eine Angabe absoluter Entfernungen zwischen korrespondierenden Punkten in Millimetern erlaubt hier am schnellsten eine klinisch relevante Einschätzung der Genauigkeit.

Die Entwicklung eines universell anwendbaren Fehlermodells unter Berücksichtigung dieser beiden Punkte ist Gegenstand der vorliegenden Arbeit.

Das entwickelte Fehlermodell soll anschließend für einen Vergleich manueller und automatischer Bildfusionsmethoden auf verschiedenen Systemplattformen für die Fusion von MRT und ECD-SPECT Daten des menschlichen Gehirns angewendet werden.

## **3 Material und Methoden**

### **3.1 Patienten**

Für die Studie wurden die MRT und ECD-SPECT Untersuchungen von 42 Patienten der psychiatrischen Klinik der Universität München retrospektiv ausgewertet. Die Untersuchungen wurden im Rahmen klinischer Routinefragestellungen durchgeführt. Häufigste Indikation für die bildgebenden Untersuchungen waren unklare dementielle Syndrome sowie deren differentialdiagnostische Abgrenzung zu depressiven Störungen beziehungsweise der Ausschluss struktureller Läsionen. Es wurden nur Patienten in die Studie aufgenommen, bei denen sowohl die MRT als auch die ECD-SPECT Untersuchung als altersentsprechend unauffällig befundet wurden. Weitere Voraussetzungen waren die Akquisition einer hochauflösenden T1-gewichteten MRT Sequenz sowie ein maximaler zeitlicher Abstand von zwei Wochen zwischen MRT und ECD-SPECT Untersuchung. Bei 15 Patienten (6 weiblich, 9 männlich) waren alle Voraussetzungen erfüllt, das Alter der eingeschlossenen Patienten reichte von 25 bis 74 Jahren, mit einem Durchschnittsalter von 58 Jahren.

### **3.2 Magnetresonanztomographie**

Die Kernspintomographien wurden auf einem Siemens Magnetom Vision mit 1,5 Tesla Feldstärke akquiriert (Siemens Medizintechnik, Erlangen). Die verwendete mprage-Sequenz (magnetisation prepared rapid gradient echo) liefert T1 gewichtete Bilder (TR: 11,6 ms, TE: 4,9 ms) mit einer Bildmatrix von 256 x 256 Pixel und einer Auflösung von 1 x 1 mm. Die Sequenz zeichnet 128 sagittale Schichten mit einer Schichtdicke von 1,2 mm auf, die Akquisitionszeit beträgt 5:30 Minuten. In der klinischen Routinediagnostik hat die Sequenz einen festen Stellenwert, da die hohe Auflösung eine Rekonstruktion der Bilddaten in beliebigen Raumebenen praktisch ohne Qualitätsverlust erlaubt. Dies prädestiniert die Sequenz auch als Primärdatensatz für eine multimodale Bildfusion sowie für alle Formen der 3D Rekonstruktion. Die Bilddaten wurden digital auf Magneto-optischen Disks (MOD) gespeichert und via Netzwerk im Dicom Format auf die Workstations für die Bildfusion übertragen.

### 3.3 ECD-SPECT

Die SPECT Untersuchungen wurden auf einer Picker Prism 3000 Gammakamera durchgeführt (Picker international, Cleveland OH, USA, jetzt Philips Medizintechnik). Die Kamera verfügt über drei Detektorköpfe, so dass mit einer Rotation von 120 Grad vollständige Projektionen aus 360 Grad aufgezeichnet werden konnten. Die Bilder wurden mit einer Matrix von 128 x 128 Pixel aufgezeichnet, es wurden transversale Schnittbilder mit einer Pixelgröße und Schichtdicke von 2,1 mm rekonstruiert. Für die Aufnahmen wurden konisch fokussierende HR-Fan-Beam Kollimatoren verwendet, das Maximum des Energiefensters lag bei 140 keV. Nach Injektion von durchschnittlich 350 MBq [ $^{99m}\text{Tc}$ ]-ECD wurden Aufnahmen über 30 Minuten angefertigt und nach gefilterter Rückprojektion Schnittbilddatensätze in allen Ebenen rekonstruiert und archiviert. Für die Bildfusion wurden die transversal rekonstruierten Schnittbilder verwendet, die im interfile-Format auf die entsprechenden Arbeitsstationen übertragen wurden.

### 3.4 Bildfusion

#### 3.4.1 Systeme

Die Bildfusion wurde auf zwei verschiedenen Systemen durchgeführt, die zum Zeitpunkt der Studie in der Klinik für Nuklearmedizin der LMU zur Verfügung standen. Die verfügbaren Methoden zur Bildfusion die im Folgenden miteinander verglichen wurden waren:

- Manuelle, interaktive Bildfusion
- Surface matching nach Pelizzari
- Pixel uniformity matching nach Woods

##### 3.4.1.1 Picker VoxelQ

Die interaktive Bildfusion sowie das automatische surface matching nach Pelizzari wurden auf einer Picker VoxelQ Workstation durchgeführt. Das System basiert auf einer Sun Ultra Sparc mit speziellem 3D Rendering Modul, unter dem

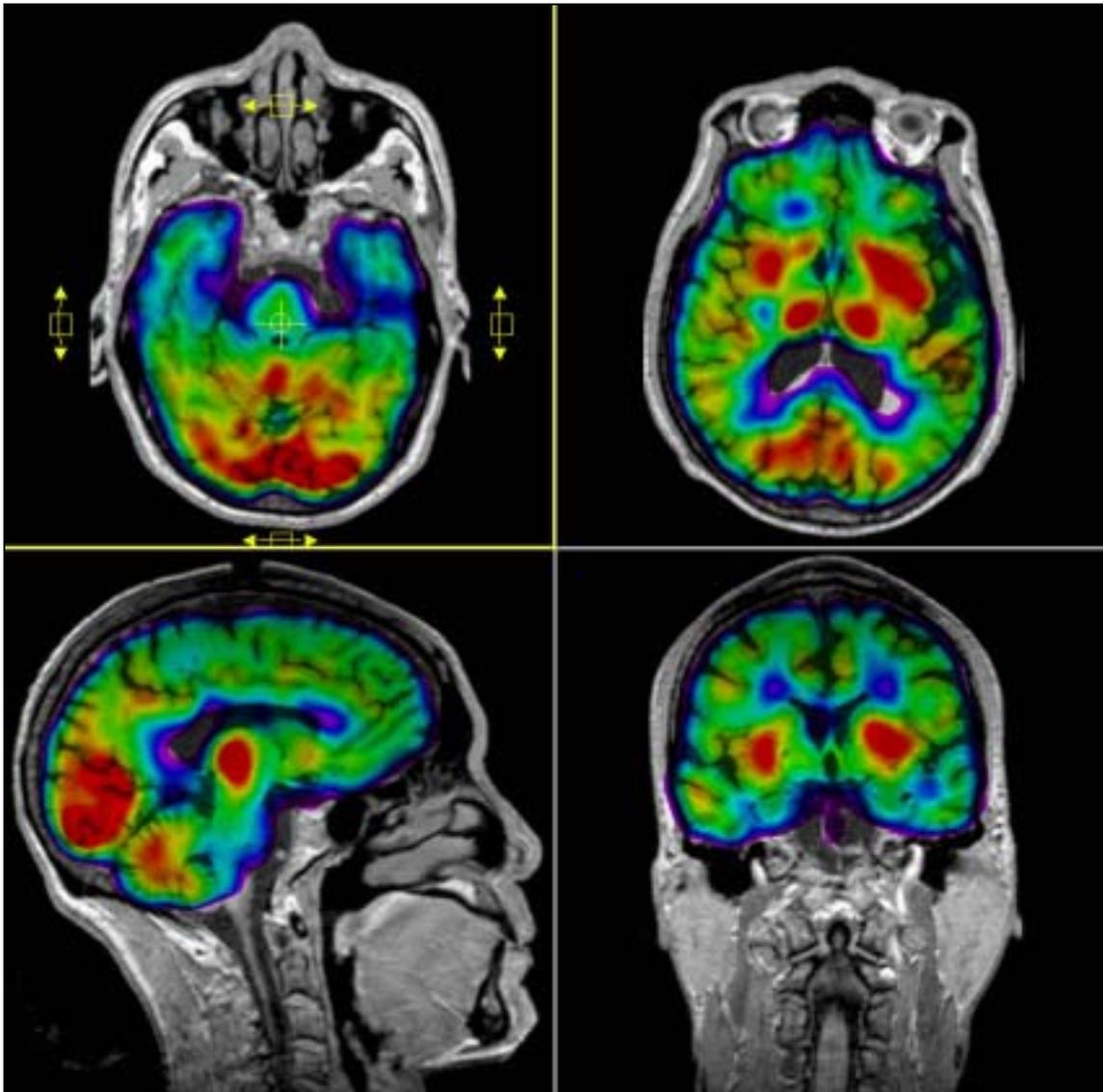
Betriebssystem Solaris 5.1. Haupteinsatzgebiet der Workstation sind 3D Rekonstruktionen von CT und MRT Daten.

#### 3.4.1.2 Nuclear Diagnostics Hermes

Die automatische Bildfusion nach dem Algorithmus von Woods wurde auf einer Hermes Workstation der Fa. Nuclear Diagnostics durchgeführt. Das System basiert auf Standard-PC Hardware (Intel CPU) mit Solaris als Betriebssystem. In der Klinik für Nuklearmedizin werden Sie zur Befundung aller Routineuntersuchungen verwendet.

#### **3.4.2 Manuelle Bildfusion**

Abbildung 7 zeigt das in der Studie verwendete Benutzerinterface für die interaktive Bildfusion. Hier werden dem Benutzer auf dem Bildschirm vier überlagerte Bilder in verschiedenen Orientierungen angezeigt. Im dargestellten Beispiel ist im Hintergrund als Primärdatensatz die Kernspintomographie des Kopfes als Graustufenbild sichtbar. Farbcodiert überlagert ist eine SPECT Untersuchung sichtbar, die die regionale Hirnperfusion anzeigt. Die verwendete Farbtabelle orientiert sich an einem prismatischen Farbspektrum und stellt die regionale Hirnperfusion in ansteigender Intensität von violett über blau, grün, gelb und rot dar. Die Helligkeits- und Kontrasteinstellungen für den Primärdatensatz sowie Farbintensität und -spektrum für den Sekundärdatensatz können unabhängig voneinander angepasst werden, um dem Benutzer eine möglichst optimale Bildqualität anzubieten. Auch Veränderungen während des Fusionsprozesses sind noch möglich. Der SPECT Datensatz kann mit der Maus gegenüber der MRT im Hintergrund interaktiv verschoben und rotiert werden. Eine Manipulation in einer Bildebene bewirkt eine sofortige Aktualisierung der Darstellung in den anderen Ebenen. Der Benutzer wählt initial für jede dargestellte Orientierung eine Bildebene, die eine möglichst detaillierte, anatomische Korrelation zwischen den beiden Datensätzen erlaubt (z.B. Darstellung der Basalganglien in der transversalen Ansicht, Darstellung der Medianebene in sagittaler Richtung). Die interaktive Transformation des SPECT Datensatzes wird vom Benutzer so lange angepasst, bis visuell eine optimale Deckung der beiden Studien in allen drei Bildebenen erreicht worden ist.



**Abbildung 7:** Benutzerinterface für die manuelle Bildfusion. Der transparent farbcodiert dargestellte SPECT Datensatz wird gegenüber der statischen MRT im Hintergrund interaktiv verschoben und rotiert, bis in allen Bildebenen eine gute Übereinstimmung vorliegt.

Nach erfolgter Transformation wird der komplette Datensatz in allen Ebenen durchgescrollt um noch verbliebene Ungenauigkeiten, die in anderen Bildebenen sichtbar sind, zu erkennen und gegebenenfalls zu korrigieren. Nach Abschluss der Transformation werden die aktuellen Transformationsparameter gespeichert.

Zur Überprüfung der Intraobservervariabilität wurde die interaktive Bildfusion für alle 15 Patienten jeweils fünfmal wiederholt, zusätzlich wurde die benötigte Zeit für die einzelnen Fusionsversuche ermittelt. Zur Untersuchung der Interobservervariabilität, und damit der Subjektivität der durchgeführten Transformation

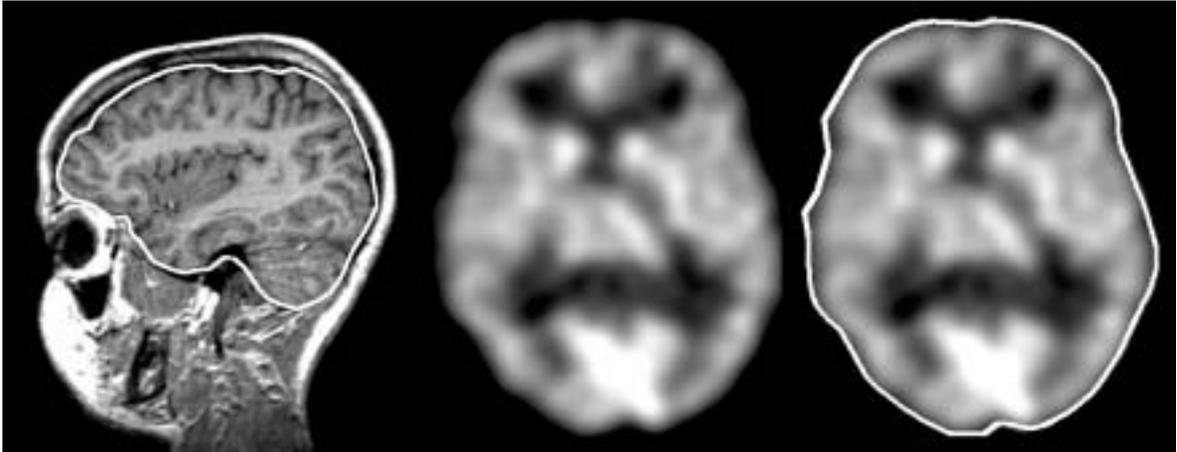
wurden die Daten von fünf zufällig ausgewählten Patienten zusätzlich von fünf anderen Untersuchern fusioniert. Vor jeder interaktiven Bildfusion wurde die Position des Sekundärdatensatzes auf die unveränderte Ausgangsposition zurückgesetzt. Die Transformationsparameter dieser wiederholten Bildfusionen wurden jeweils als ascii-Datei abgelegt und für die weitere statistische Auswertung und Fehleranalyse exportiert (Abbildung 8).

-2.1088000	-7.1156000	-3.4136000	0.0608000	0.0267000	0.0129000
-2.2006000	-6.8741000	-3.1480000	0.0608000	0.0320000	0.0185000
-2.0577000	-6.2299000	-3.1328000	0.0724000	0.0204000	0.0229000
-2.9376000	-7.2157000	-2.7243000	0.0562000	0.0340000	0.0166000
-2.6514000	-7.4796000	-2.3246000	0.0178000	0.0205000	0.0112000

**Abbildung 8:** Beispiel für die Transformationsparameter von fünf wiederholten manuellen Bildfusionen. Die ersten drei Spalten geben die Verschiebung entlang der X-, Y- und Z-Achse in Millimetern an, die letzten drei Spalten die entsprechenden Rotationen um die Achsen im Bogenmaß

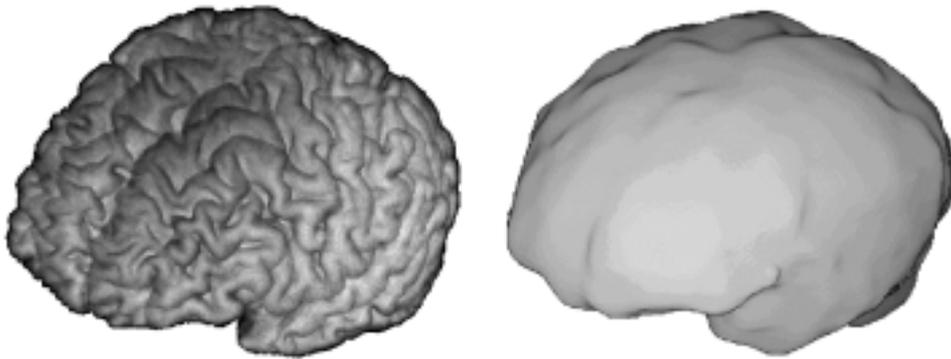
### 3.4.3 Surface matching nach Pelizzari

Die Bildfusion mittels des surface matching Algorithmus nach Pelizzari (Pelizzari, Chen et al. 1989) wurde ebenfalls auf der Picker VoxelQ Workstation durchgeführt. Als Vorverarbeitung war die Definition der Hirnoberfläche in beiden Untersuchungen notwendig. Für die MRT erfolgte dies durch manuelle Segmentierung aller 128 sagittalen Bildschichten, für die SPECT Untersuchung wurde eine schwellenwertbasierte, automatische Segmentierung durchgeführt (Abbildung 9).



**Abbildung 9:** manuelle Segmentierung der Hirnoberfläche in der MRT und automatische Segmentierung in der SPECT.

Der Algorithmus nach Pelizzari minimierte nun iterativ die Summe der Quadrate der Abstände zwischen den beiden Oberflächen (Abbildung 10).



**Abbildung 10:** korrespondierende Organoberflächen nach Segmentierung in MRT und SPECT

Um den Einfluss der manuellen Segmentierung der MRT-Daten zu untersuchen, wurde diese für alle Fälle fünfmal wiederholt und die Reproduzierbarkeit der folgenden automatischen Bildfusion ausgewertet. Analog zum Vorgehen bei der manuellen Bildfusion wurden auch hier die fünf ausgewählten Datensätze zusätzlich von fünf anderen Untersuchern segmentiert und die Interobservervariabilität ermittelt.

#### **3.4.4 Pixel uniformity matching nach Woods**

Die automatische Bildfusion mittels des Algorithmus nach Woods (Woods, Mazziotta et al. 1993) wurden auf Hermes Workstations der Firma Nuclear Diagnostics durchgeführt.

Die Bildfusion wurde ohne weitere Datenvorverarbeitung, insbesondere ohne vorherige manuelle Segmentierung des Gehirns in der MRT durchgeführt. Um einen systematische Beeinflussung des Algorithmus durch diese fehlende Vorverarbeitung auszuschließen, wurden fünf Datensätze zusätzlich manuell segmentiert und die Transformationsergebnisse mit denen der unsegmentierten Daten verglichen.

Auch für die Bildfusion nach Woods wurde die Transformation jeweils fünfmal pro Datensatz wiederholt. Hierzu wurde manuell die Position des SPECT Datensatzes in einem Bereich von  $\pm 20^\circ$  Rotation und  $\pm 20$  mm Translation verändert, bevor der automatische Algorithmus mit der Bildfusion begann. In einer abschließenden visuellen Kontrolle wurde das Fusionsergebnis auf grobe Fehler überprüft.

Die erzeugten Transformationsparameter wurden ebenfalls als Textdatei exportiert.

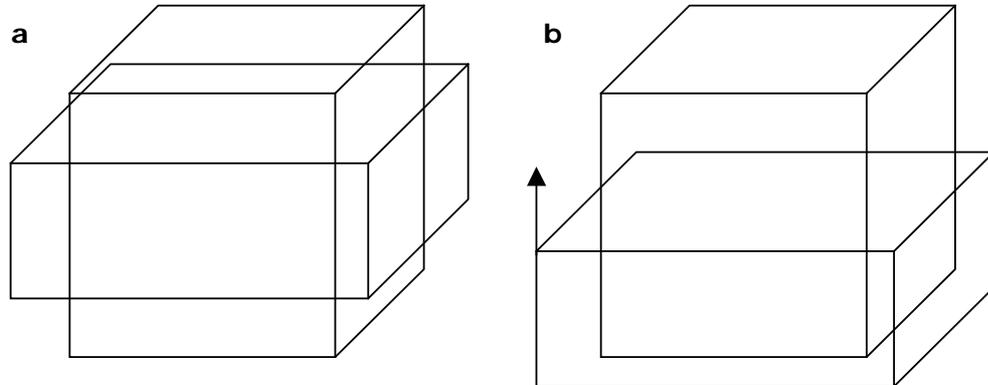
### **3.5 Fehleranalyse**

#### **3.5.1 Unterschiede der verwendeten Fusionssysteme**

Ergebnis der durchgeführten manuellen und automatischen Bildfusionen war jeweils ein Satz der Transformationsparameter, die die Position des sekundären Datensatzes im Raum definieren. Vor einer Weiterverarbeitung der Parameter waren einige Unterschiede in der zugrunde liegenden Bildgeometrie der beiden Systeme zu berücksichtigen:

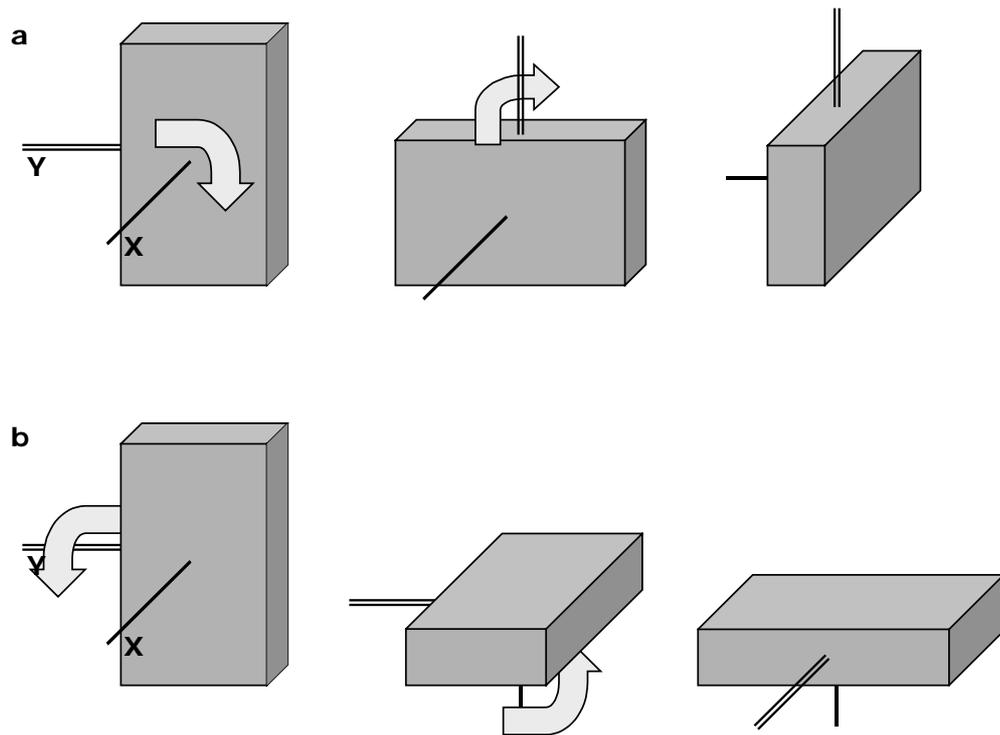
Ein Unterschied bestand in der initialen Positionierung der Datensätze, also der Ausgangsposition vor Durchführung der Bildfusion. Beide Systeme zentrierten die Daten in der Transversalebene, wo eine Korrektur somit nicht notwendig war. Die Höhenpositionierung unterschied sich jedoch: Die VoxelQ zentriert auch in lateraler Ansicht beide Datenvolumina auf ihren geometrischen Mittelpunkt

(Abbildung 11a), die Hermes Workstation legte dagegen beide Datenvolumina mit ihrer jeweils untersten transversalen Schicht aufeinander (Abbildung 11b). Hier war eine entsprechende Korrektur des Verschiebungsparameters in Z-Richtung notwendig (Pfeil)



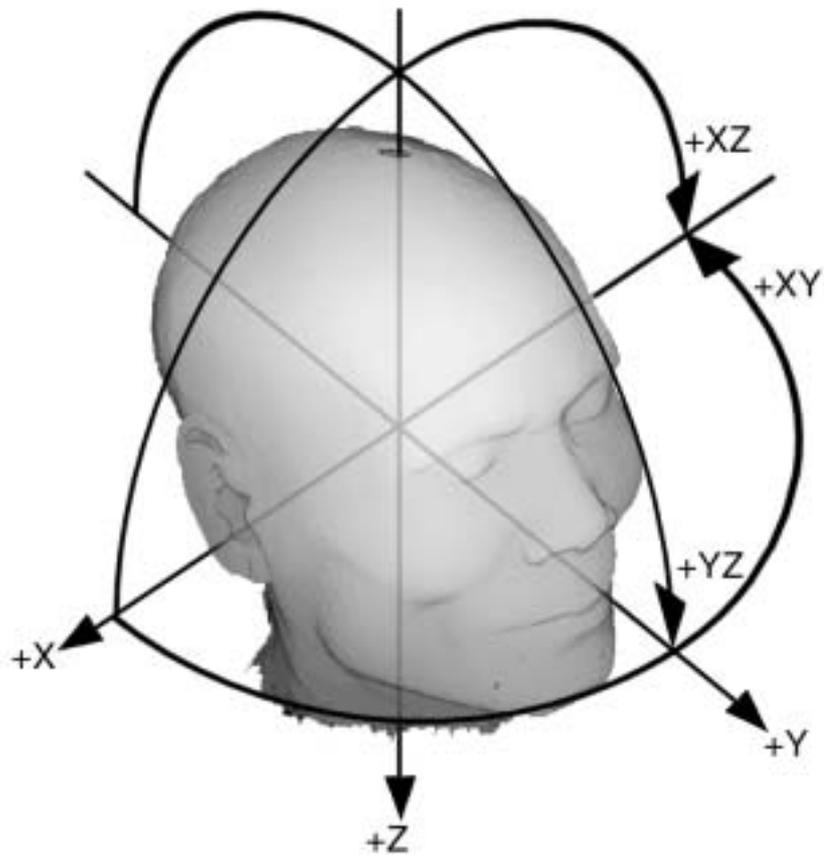
**Abbildung 11:** Unterschiedliche Startpositionen der Datenvolumina auf der VoxelQ (a) bzw. Hermes (b) Workstation. Der Würfel symbolisiert den MRT Datensatz, der flachere Quader den SPECT Datensatz.

Ein weiterer Unterschied bestand in der Reihenfolge der Drehungen, die als Grundlage für die angegebenen Rotationsparameter verwendet wurden. Die Reihenfolge von Drehungen ist nicht kommutativ, eine veränderte Reihenfolge von gleichen Drehungen führt zu unterschiedlichen Ergebnissen. Abbildung 12 illustriert dies anhand eines Quaders, der zweimal um je  $90^\circ$  um zwei Achsen gedreht wird (im Uhrzeigersinn bei Blick auf die Achse). In Abbildung 12a erfolgt die Drehung zuerst um die X-Achse, danach um die Y-Achse, in Abbildung 12b umgekehrt. Obwohl beide Male die gleichen Drehungen von je  $90^\circ$  durchgeführt wurden unterscheiden sich die resultierenden Positionen des Quaders.



**Abbildung 12:** Illustration der Abhängigkeit der Endposition von der Reihenfolge zweier identischer Rotationen um zwei Achsen (jeweils Rechtsdrehung um  $90^\circ$  um die X- und Y-Achse).

Auch die hier verwendeten Plattformen für die Bildfusion unterschieden sich bezüglich der Reihenfolge von Rotationen: an der VoxelQ Workstation mit der Abfolge X-, Y- und zuletzt Z-Achse, die Hermes Workstations begannen dagegen mit der Rotation um die Z-Achse gefolgt von X- und Y-Achse (Abbildung 13). Vor einem Vergleich der Rotationsparameter mussten diese entsprechend konvertiert werden, konkret wurden die Hermes-Parameter in die Konvention der VoxelQ-Parameter übertragen.



**Abbildung 13:** Konvention des Koordinatensystems und der Achsenbezeichnung für beide verwendeten Systeme.

Der Mittelpunkt der Rotationen war auf beiden Systemen der geometrische Mittelpunkt des SPECT-Datensatzes.

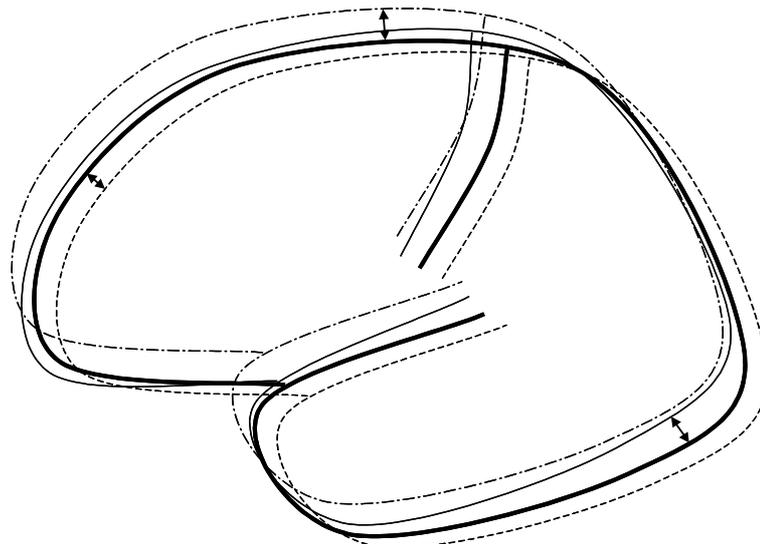
Die Umrechnung der Rotationsparameter wurde mit der Programmiersprache IDL (Interactive Data Language, Research Systems Inc., Kanada) auf einer Sun Ultra-Sparc Workstation durchgeführt.

### 3.5.2 Definition der Referenzposition

Für die quantitative Analyse der Bildfusionsgenauigkeit wurde jeweils die Reproduzierbarkeit jeder der drei angewandten Bildfusionsmethoden bei fünf Wiederholungen überprüft. Jede einzelne Bildfusion liefert als Ergebnis eine Raumposition des überlagerten Datensatzes, der durch die sechs Transformationsparameter für Translation und Rotation definiert ist. Da ein verbindlicher Goldstandard für eine Fehlerbestimmung fehlt, wurde postuliert, dass der Mittelwert dieser fünf einzelnen Fusionsversuche der "richtigen" Position am nächsten kommt, und diese mittlere Position wurde als Referenzposition definiert. Anschließend wurde die Abweichung jedes einzelnen Fusionsversuches von dieser Referenzposition bestimmt. Die Abweichung von der Referenzposition stellt ein Maß für die Reproduzierbarkeit einer Bildfusionsmethode dar und wird im Weiteren als "Fehler" bezeichnet.

Die Berechnung der mittleren Position des Datensatzes nach fünf einzelnen Transformationen erfolgt als Bestimmung des Schwerpunktes im sechsdimensionalen Parameterraum, d.h. durch Mittelwertbildung für jeden einzelnen Transformationsparameter, ebenfalls unter Verwendung von IDL.

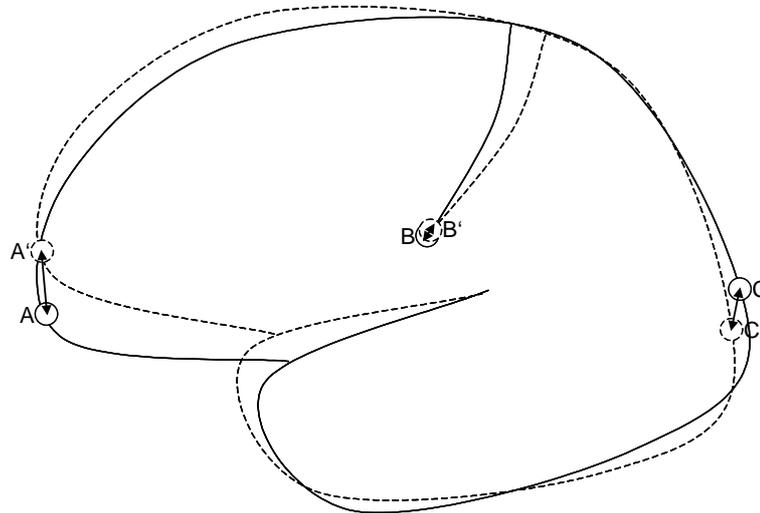
Abbildung 14 illustriert das Prinzip anhand von drei einzelnen Raumpositionen, deren Abweichung von der Referenzposition ermittelt wird.



**Abbildung 14:** Bestimmung der mittleren Position (dicke Konturlinie) aus drei einzelnen Fusionsprozessen (dünne und gestrichelte Konturen) als Referenzposition. Anschließend wird die Abweichung der einzelnen Fusionsergebnisse von dieser Referenzposition ermittelt (Pfeile)

### 3.5.3 Berechnung der Abweichung von der Referenzposition

Je nach Position und Rotationsstellung kann diese Abweichung natürlich für jeden einzelnen Punkt innerhalb eines Datensatzes sehr unterschiedlich ausfallen, wie die korrespondierenden Punktepaare A und A', B und B' sowie C und C' in Abbildung 15 zeigen.



**Abbildung 15:** Die Abweichung einzelner Punkte A', B' und C' eines Fusionsergebnisses von den korrespondierenden Punkten A, B und C in der Referenzposition ist in verschiedenen Regionen des Datensatzes unterschiedlich stark.

Die Konvention, für welche Punkte innerhalb eines Datensatzes der Fehler bestimmt wird, spielt somit eine entscheidende Rolle für eine Fehlerangabe.

Hier wurde zunächst für jeden der 15 Fälle durch ein Schwellenwertverfahren (Countrate > 12000) eine binäre Maske aus dem SPECT Datensatz erstellt, analog zu dem unter 0 beschriebenen Vorgehen für die automatische Oberflächensegmentierung. Die Berechnung des Fehlers wurde auf Punkte innerhalb dieser Maske, also innerhalb des Hirnparenchyms, beschränkt.

Über eine Monte Carlo Simulation (Williamson 1987) wurden für jeden einzelnen Fusionsvorgang zufällige Pixelkoordinaten generiert. Lag der entsprechende Bildpunkt außerhalb der Maske wurden die Koordinaten verworfen und ein neuer Punkt generiert. Lag der Punkt innerhalb der Maske wurde die Abweichung von der Referenzposition in Millimetern für diesen Punkt bestimmt. Dies wurde sooft wiederholt bis für jeden einzelnen Fusionsvorgang der Fehler an 5000 Punkten berechnet wurde. Der Mittelwert dieser 5000 Fehlerberechnungen wurde als Maß

für die Qualität eines einzelnen Fusionsversuches verwendet. Weiterhin wurde der Mittelwert aller fünf Fusionsversuche pro Fall und Methode bestimmt, und für jede Fusionsmethode wurde der mittlere Fehler über alle 15 Fälle bestimmt. Auch der maximale aufgetretene Fehler für jeden einzelnen Fusionsvorgang wurde dokumentiert.

Insgesamt wurden bei 15 Fällen, drei Methoden mit je fünf bzw. zweimal fünf Wiederholungen (Intra- und Interobservervariabilität) die räumliche Abweichung für 1.375.000 Punkte berechnet.

Um zu überprüfen, wie viele Punkte in die Fehlerberechnung eingehen müssen, um ein repräsentatives Maß für den gesamten Datensatz zu erhalten, wurde vor der Festlegung der letztendlich verwendeten Anzahl von 5000 Punkten pro Position die Abhängigkeit des ermittelten Fehlers von der Anzahl der berechneten Punkte bestimmt.

Die Berechnung der Referenzposition, die Erstellung der Datenmaske, die zufallsgesteuerte Auswahl der 5000 Punkte sowie die Abstandsberechnung zwischen den korrespondierenden Punkten in der Referenzposition wurden ebenfalls mit der Software IDL durchgeführt. Die Ausgabe des Fehlerberechnungsprogramms unter IDL wurde jeweils als ascii Textdatei mitprotokolliert und gespeichert. Die Mittelwertbildungen und weitere statistische Auswertungen wurden mit Microsoft Excel durchgeführt.

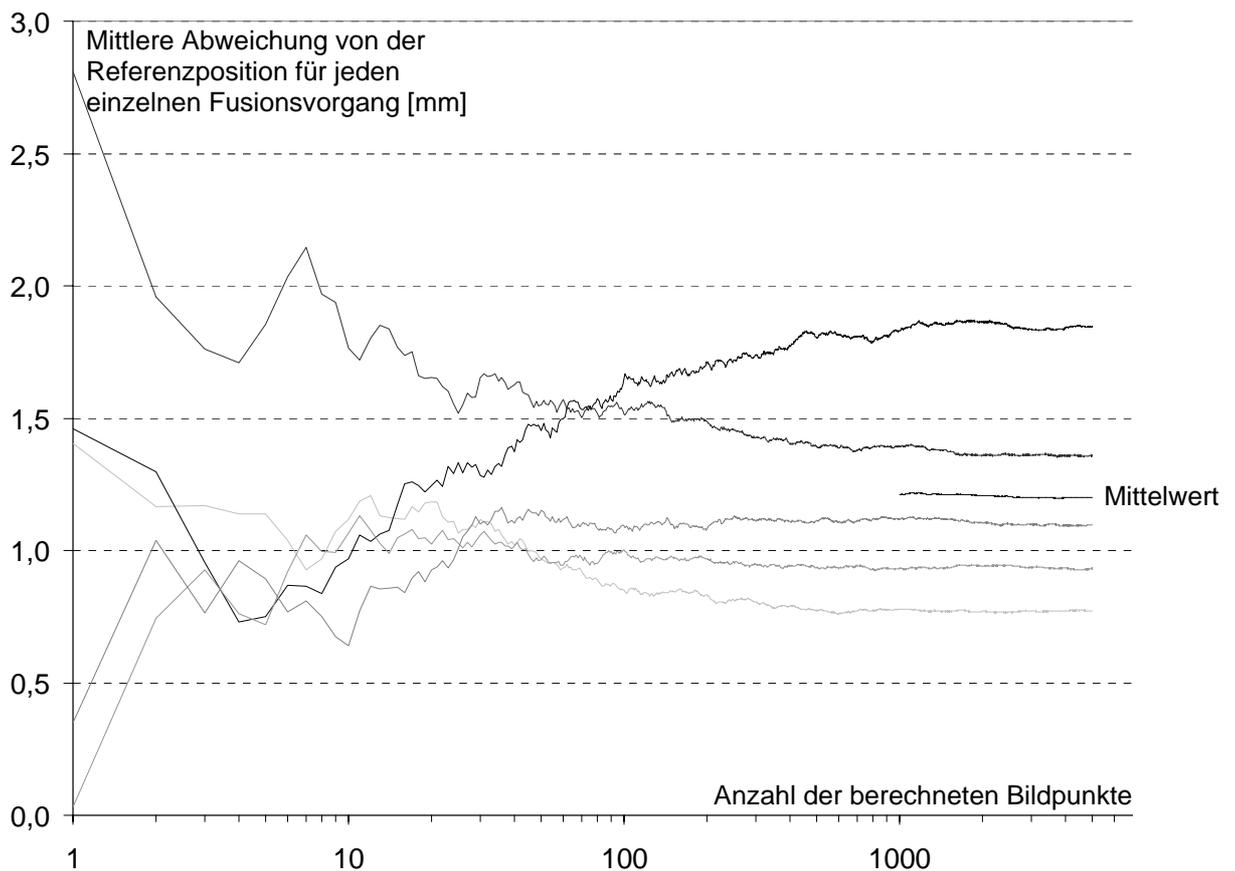
#### **3.5.4 Intermethodenvergleich**

Nach Bestimmung des mittleren Fusionsfehlers für die einzelnen Verfahren wurde zusätzlich ein direkter Vergleich der Referenzpositionen der verschiedenen Verfahren durchgeführt. Analog zu dem vorher beschriebenen Vorgehen wurde eine neue Referenzposition aus den Ergebnissen der verschiedenen Verfahren ermittelt und wiederum die Abweichung der einzelnen Methoden von dieser neuen Referenzposition bestimmt.

## 4 Ergebnisse

### 4.1 Konsolidierung des Fehlerwertes

Mittels der in 3.5.3 beschriebenen Simulation wurde der räumliche Abstand eines Bildpunktes von seiner Position in der Referenztransformation für eine definierte Anzahl von zufällig ausgewählten Bildpunkten innerhalb der individuellen Maske für jeden Fusionsvorgang berechnet. Abbildung 16 zeigt exemplarisch den Verlauf der Fehlerkonsolidierung für fünf einzelne manuelle Bildfusionen des ersten Datensatzpaares. An der logarithmisch skalierten X-Achse, ist die Anzahl an berechneten Bildpunkten aufgetragen. Die Y-Achse zeigte die Entfernung vom korrespondierenden Punkt in der Referenzposition in Millimetern, dargestellt ist jeweils der Mittelwert aus allen bis dahin berechneten Bildpunkten.



**Abbildung 16:** Konsolidierung der gemittelten Fehlerberechnung mit zunehmender Anzahl an Bildpunkten deren Abweichung berechnet wurde.

Es zeigt sich, dass der ermittelte Fehler bei Berechnung für die ersten Bildpunkte noch stark schwankt, aber bereits ab circa 100 berechneten Bildpunkten zunehmend zum endgültigen Wert konvergiert. Ab einer Anzahl von 1000 berechneten Bildpunkten lagen die Schwankungen für jeden einzelnen Fusionsvorgang unter 0,1 mm. Der ab 1000 Bildpunkten zusätzlich dargestellte Mittelwert aus den fünf einzelnen Fusionsfehlern war damit stabil zu ermitteln. Bei 5000 berechneten Bildpunkten lag die Schwankung des Fehlermittelwertes in allen Fällen unter 0,02 mm. Die Berechnung stellt somit ein hinreichend genaues, repräsentatives Maß für die Fusionsgenauigkeit innerhalb der verwendeten Maske dar.

Die starken Schwankungen innerhalb der ersten Berechnungen zeigen gleichzeitig, wie sehr der ermittelte Fehler von der zufällig bestimmten Position weniger Bildpunkte abhängt.

In der Konsolidierung der Fehlerberechnungen ergab sich kein Unterschied für die unterschiedlichen Fusionsmethoden.

## 4.2 Fusionsgenauigkeit

Tabelle 1 zeigt eine Übersicht der Fusionsgenauigkeiten für die drei verwendeten Methoden bei den 15 Patientendatensätzen. Der angegebene Fehlerwert entspricht der mittleren Abweichung von je fünf Wiederholungen der Bildfusion von der als Goldstandard definierten Referenzposition des Datensatzes. Es sind jeweils der Mittelwert aller ausgewerteten Datensatzpaare sowie die Standardabweichung angegeben.

Das angegebene Maximum bezieht sich auf die Fehlerberechnung eines *einzelnen* Bildpunktes, nicht auf den Gesamtfehler innerhalb eines Datensatzes.

Fehler im mm	Mittelwert	Standard- abweichung	Maximum
Manuelle Bildfusion Intraobservervariabilität	1,5	0,3	5,6
Manuelle Bildfusion Interobservervariabilität	1,6	0,3	5,8
Surface matching nach Pelizzari Intraobservervariabilität	2,9	1,1	10,5
Surface matching nach Pelizzari Interobservervariabilität	3,8	1,0	11,7
Pixel uniformity matching nach Woods Reproduzierbarkeit	2,2	0,8	8,2
Intermethodendifferenz	2,3	0,8	8,4

**Tabelle 1:** Ergebnisübersicht des mittleren Fusionsfehlers sowie des maximalen aufgetretenen Einzelfehlers

### 4.2.1 Manuelle Bildfusion

Tabelle 2 zeigt eine Übersicht der Fusionsfehler für die manuelle Bildfusion.

Fehler [mm]	Intraobserver- variabilität	Interobserver- variabilität
Fall 1	1,16	1,57
Fall 2	1,84	1,52
Fall 3	0,99	1,07
Fall 4	1,77	1,49
Fall 5	1,29	1,28
Fall 6	1,72	1,85
Fall 7	1,71	2,04
Fall 8	1,44	1,30
Fall 9	1,79	1,78
Fall 10	1,60	1,82
Fall 11	1,92	1,83
Fall 12	1,90	1,03
Fall 13	1,55	2,25
Fall 14	0,96	1,83
Fall 15	1,18	1,49
<b>Mittelwert</b>	<b>1,52</b>	<b>1,61</b>
<b>Standardabweichung</b>	<b>0,33</b>	<b>0,35</b>
<b>Maximum</b>	<b>1,92</b>	<b>2,25</b>
<b>Minimum</b>	<b>0,96</b>	<b>1,03</b>

**Tabelle 2:** Mittelwert der Abweichung von der Referenzposition aus fünf Wiederholungen der manuellen Bildfusion, durchgeführt von einem (Intraobserver) oder mehreren (Interobserver) Benutzern.

#### 4.2.1.1 Intraobservervariabilität

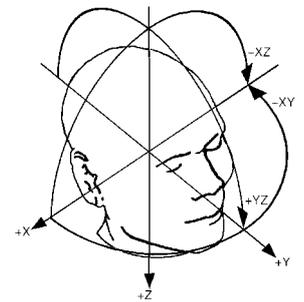
Für die Intraobservervariabilität der manuellen, interaktiven Bildfusion ergab sich ein mittlerer Gesamtfehler von 1,5 mm.

Der maximale Fehler pro Fall, also der schlechteste Mittelwert aus 15 Fällen, lag bei 1,9 mm, der geringste Fehler bei 1,0 mm. Die Standardabweichung betrug 0,3 mm.

Der maximale Fehler pro einzeltem Pixel, also der Fehler für den schlechtesten von 5000 Pixeln im schlechtesten von fünf Fusionsversuchen beim schlechtesten von 15 Fällen betrug 5,6 mm. Die Angabe eines minimalen Fehlers für einzelne Pixel erübrigt sich, da in fast jeder Transformation einzelne Pixel vorkommen, deren Position sich kaum verändert hat. Entsprechend gab es auch in allen Fällen Pixel mit einem berechneten Fehler von unter 0,1 mm, was jedoch keinen Rückschluss auf die Genauigkeit einer Bildfusion erlaubt.

Bei der Betrachtung der einzelnen Transformationsparameter (Tabelle 3) zeigte sich hier die größte Variabilität bei der Translation von anterior nach posterior (y) und mit einer Standardabweichung von 1,2 Grad bei der Rotation in der Sagittalebene (YZ).

	Translation [mm]			Rotation [Grad]		
	x	y	z	YZ	XZ	XY
Fall 1	0,43	0,68	0,70	0,54	0,36	0,30
Fall 2	0,67	0,98	0,85	1,77	0,79	0,71
Fall 3	0,53	0,43	0,39	1,31	1,40	0,62
Fall 4	1,15	0,89	0,67	0,94	0,96	0,41
Fall 5	0,29	0,81	0,63	1,18	0,54	0,84
Fall 6	1,23	0,78	0,53	1,23	0,99	1,08
Fall 7	0,13	1,47	0,58	0,98	0,21	1,02
Fall 8	0,41	0,59	0,95	1,42	0,62	0,44
Fall 9	0,76	1,04	1,14	1,28	0,42	0,57
Fall 10	0,70	1,16	0,78	1,17	0,73	0,27
Fall 11	1,04	1,60	1,07	1,07	0,64	0,46
Fall 12	0,30	1,77	0,74	2,49	0,15	0,33
Fall 13	0,70	1,26	0,60	1,20	1,27	0,28
Fall 14	0,47	0,68	0,70	0,76	0,53	0,49
Fall 15	0,54	0,53	0,52	0,96	1,29	0,27
<b>Mittelwert</b>	<b>0,62</b>	<b>0,98</b>	<b>0,72</b>	<b>1,22</b>	<b>0,73</b>	<b>0,54</b>



**Tabelle 3:** Standardabweichungen der einzelnen Transformationsparameter bei je fünf Wiederholungen der manuellen Bildfusion für jeden Fall

Insgesamt ergaben sich für die Standardabweichungen der einzelnen Transformationsparameter jedoch keine eindeutige Benachteiligung eines Parameters und kein signifikanter Unterschied zwischen den einzelnen Bildfusionsverfahren. Daher wird diese Standardabweichung hier nur exemplarisch für die Intra-observervariabilität der manuellen Bildfusion angegeben, bei den anderen Methoden wird auf die Darstellung verzichtet.

#### 4.2.1.2 Interobservervariabilität

Für die Interobservervariabilität der interaktiven Bildfusion ergab sich ein mittlerer Gesamtfehler von 1,6 mm.

Pro Fall betrug der Fehler maximal 2,2 mm, minimal 1,0 mm, bei einer Standardabweichung von 0,3 mm.

Der maximale Fehler für einen einzelnen Pixel lag bei 5,8mm.

## 4.2.2 Surface matching nach Pelizzari

Fehler [mm]	Intraobserver- variabilität	Interobserver- variabilität
Fall 1	4,20	3,88
Fall 2	3,47	3,63
Fall 3	2,21	3,09
Fall 4	2,05	3,80
Fall 5	2,42	4,03
Fall 6	2,80	4,29
Fall 7	3,63	3,82
Fall 8	1,30	2,79
Fall 9	5,34	6,85
Fall 10	3,29	4,36
Fall 11	2,63	3,32
Fall 12	3,31	4,03
Fall 13	2,27	2,71
Fall 14	1,52	2,37
Fall 15	3,64	3,92
<b>Mittelwert</b>	<b>2,94</b>	<b>3,79</b>
<b>Standardabweichung</b>	<b>1,06</b>	<b>1,03</b>
<b>Maximum</b>	<b>5,34</b>	<b>6,85</b>
<b>Minimum</b>	<b>1,30</b>	<b>2,37</b>

**Tabelle 4:** Übersicht über den Fusionsfehler für den surface matching Algorithmus nach Pelizzari in Abhängigkeit von der vorhergehenden Segmentierung der Hirnoberfläche.

### 4.2.2.1 Intraobservervariabilität

Für die Intraobservervariabilität der automatischen Bildfusion nach Pelizzari, also die Abhängigkeit des Fusionsergebnisses von der vorherigen manuellen Segmentierung der Hirnoberfläche, ergab sich ein mittlerer Gesamtfehler von 2,9 mm. Der größte Fehler pro Fall lag bei 5,3 mm, der geringste Fehler bei 1,3 mm. Die Standardabweichung betrug 1,1 mm (Tabelle 4). Der maximale Fehler für einen einzelnen Pixel lag bei 10,5 mm.

### 4.2.2.2 Interobservervariabilität

Für die Interobservervariabilität der lag der mittlere Gesamtfehler bei 3,8 mm mit einem Minimum von 2,4 mm, einem Maximum von 6,7 mm und einer Standardabweichung von 1,0 mm. Der maximale Fehler für einen einzelnen Pixel betrug 11,7 mm.

### 4.2.3 Pixel uniformity matching nach Woods

In vier der 15 Fälle zeigte sich für die automatische Bildfusion nach Woods eine massive Fehlüberlagerung mit Rotationsfehlern, insbesondere in der Sagittalebene, von über 30°. Diese waren in der visuellen Kontrolle der automatischen Fusion eindeutig als schwere systematische Fehler erkennbar und wurden nicht in die Berechnung des mittleren Fusionsfehlers für diese Methode einbezogen. In allen vier Fällen war diese grobe Fehlüberlagerung für alle fünf verschiedenen initialen Startpositionen reproduzierbar. Auch nach einer initialen manuellen Bildfusion, d.h. aus einer korrekten Startposition heraus, wiederholte der Algorithmus jeweils den gleichen Fehler und beendete die automatische Fusion in einer vergleichbaren Fehlstellung.

Fehler [mm]	Reproduzierbarkeit
Fall 1	1,94
Fall 2	-
Fall 3	0,67
Fall 4	2,55
Fall 5	2,21
Fall 6	1,90
Fall 7	2,39
Fall 8	-
Fall 9	2,39
Fall 10	2,74
Fall 11	2,60
Fall 12	1,59
Fall 13	-
Fall 14	-
Fall 15	3,72
<b>Mittelwert</b>	<b>2,24</b>
<b>Standardabweichung</b>	<b>0,76</b>
<b>Maximum</b>	<b>3,72</b>
<b>Minimum</b>	<b>0,67</b>

**Tabelle 5:** Fusionsfehler bei der automatischen Bildfusion nach Woods. Vier Fälle mit offensichtlichen groben Fehlern in der Bildfusion wurden nach visueller Kontrolle ausgeschlossen.

Die Untersuchung der Reproduzierbarkeit der übrigen elf Fälle (Tabelle 5) ergab einen mittleren Gesamtfehler von 2,2 mm. Bei einer Standardabweichung von 0,8 mm lag das Minimum pro Fall bei 0,7 mm, das Maximum bei 3,7 mm. Für einen einzelnen Pixel lag der maximale Fehler bei 8,2 mm.

#### 4.2.4 Intermethodenvergleich

In den Fällen für die der Woods-Algorithmus ein verwertbares Ergebnis lieferte, wurden die Referenzpositionen der einzelnen Bildfusionsverfahren wiederum miteinander verglichen und der Abstand der Referenzpositionen der einzelnen Verfahren ermittelt (Tabelle 6).

Fehler [mm]	
Fall 1	2,35
Fall 2	
Fall 3	1,53
Fall 4	1,42
Fall 5	2,03
Fall 6	1,16
Fall 7	2,62
Fall 8	
Fall 9	3,03
Fall 10	2,92
Fall 11	3,18
Fall 12	1,69
Fall 13	
Fall 14	
Fall 15	3,43
<b>Mittelwert</b>	<b>2,30</b>
<b>Standardabweichung</b>	<b>0,79</b>
<b>Maximum</b>	<b>3,43</b>
<b>Minimum</b>	<b>1,16</b>

**Tabelle 6:** Abweichung der Referenzpositionen der einzelnen Fusionsverfahren

Hier ergab sich eine mittlere Abweichung von 2,3 mm. Pro Fall betrug die Abweichung minimal 1,2 mm, maximal 3,4 mm, bei einer Standardabweichung von 0,8 mm. Die maximale Abweichung für einen einzelnen Pixel lag bei 8,4 mm.

### 4.3 Geometrische Verteilung des Fusionsfehlers

Für die interaktive Bildfusion, die insgesamt zu den besten Ergebnissen führte, wurde zusätzlich die räumliche Verteilung des Fusionsfehlers untersucht. Für jeden Fusionsvorgang der Intraobserverwiederholungen wurden die fünf Punkte mit dem größten Fehler innerhalb der für die Fehlerberechnung verwendeten Maske bestimmt. Von diesen 375 besonders fehleranfälligen Punkten lagen 34 % am Frontalpol, 19 % in der Zentralregion, 25 % occipital und 22 % im Kleinhirn.

### 4.4 Zeitbedarf für die Fusion

Tabelle 7 zeigt zusammengefasst jeweils die mittlere, die maximale und die minimale benötigte Zeit für die Bildfusion. Für die interaktive Bildfusion entspricht dies vollständig interaktiver Zeit, während der der Benutzer aktiv den Fusionsprozess durchführt. Für die automatischen Algorithmen ist jeweils die Gesamtzeit angegeben, in Klammern der jeweilige Anteil an aktiver Tätigkeit des Benutzers und der anschließenden Rechenzeit des Computers zur Bestimmung der optimalen Transformationsparameter.

Zeitbedarf für einen Fusionsvorgang [Minuten] (Interaktion + Rechenzeit)	Mittelwert	Standardabweichung	Minimum	Maximum
Interaktives Matching	11	5	7	14
Surface Matching	26 (22 + 4)	12	21	48
Woods Algorithmus	9 (4 + 5)	3	5	12

**Tabelle 7:** Vergleich des Zeitbedarfs für einen einzelnen Fusionsvorgang

Den höchsten Zeitaufwand erforderte die Fusion mit der surface matching Methode nach Pelizzari, hier insbesondere die manuelle Segmentierung der Hirnoberfläche in der MRT.

## 5 Diskussion

### 5.1 Alternative Verfahren zur Bildfusion

Die Auswahl der hier verglichenen Methoden zur Bildfusion war zum Zeitpunkt der Studie durch deren Verfügbarkeit an Routinearbeitsstationen in der nuklearmedizinischen Klinik der LMU München definiert. Selbstverständlich stellt dies, insbesondere für die automatischen Verfahren, nur einen Ausschnitt aus dem gesamten Spektrum der bekannten Verfahren für die Bildfusion dar.

Aus verschiedenen Gründen können die hier verwendeten automatischen Methoden nach Woods und Pelizzari jedoch als repräsentativ bezeichnet werden: Verschiedene Studien zum Vergleich automatischer Bildfusionsverfahren zeigten eine Überlegenheit oder zumindest ein gleichwertiges Abschneiden für pixel uniformity basierte Verfahren nach der Methode von Woods (Strother, Anderson et al. 1994; West, Fitzpatrick et al. 1999). Insbesondere gegenüber der neueren Methode der Bildfusion mittels Mutual Information (2.2.2.6) konnten sich die pixel-uniformity basierten Verfahren behaupten (West, Fitzpatrick et al. 1997; He und Narayana 2002). Diese können somit insgesamt als die mit am besten etablierten automatischen Verfahren bezeichnet werden. Berichte zu anderen Methoden beziehen sich oft auf experimentelle Arbeitsumgebungen, wobei die Methoden nicht als kommerziell erhältliche Produkte für die klinische Routine zur Verfügung stehen.

Für den surface matching Algorithmus spricht vor allem die lange Verfügbarkeit der Methode, die bereits eine ausführliche Validierung der Vorgehensweise an Phantommessungen (Rusinek, Tsui et al. 1993) und realen Datensätzen (Levin, Pelizzari et al. 1988; Mangin, Frouin et al. 1994; Turkington, Hoffman et al. 1995; de Munck, Verster et al. 1998) erlaubte.

Beim Vergleich von Fehlerangaben für andere automatische Verfahren muss darüber hinaus stets die genaue Berechnungsweise des Fehlers betrachtet werden.

## 5.2 Fehleranalyse

Jedes neu entwickelte Verfahren für die Bildfusion muss seinen Stellenwert durch Daten zur erzielbaren Genauigkeit bzw. dem auftretenden Fehler belegen. Die genaue Vorgehensweise bei der Fehlerberechnung spielt hier eine entscheidende Rolle für den ermittelten Fehler eines Bildfusionsverfahrens. Unterschiedliche Berechnungsmethoden erschweren hierbei oft einen direkten Vergleich verschiedenener Studien, ungünstige Formen der Fehlerangabe erschweren oft eine schnelle Abschätzung des tatsächlich auftretenden klinisch relevanten Fehlers. Wegen dieser großen Bedeutung sollen daher im Folgenden die bisher verwendeten Vorgehensweisen der Validierung und Fehleranalyse für Bildfusionsverfahren dargestellt und kritisch diskutiert werden.

### 5.2.1 Definition des Goldstandards

Für die Beurteilung der Qualität einer Bildfusion ist es notwendig, die Abweichung einer Transformation von der tatsächlich richtigen Transformation, die beide Studien exakt aufeinander abbildet, zu untersuchen. Dabei stellt sich wie unter 2.3 erwähnt das grundsätzliche Problem, dass die genauen "richtigen" Transformationsparameter für klinische Datensätze letztlich niemals genau bekannt sind, und es damit keinen verbindlichen Goldstandard gibt, an dem man die eigenen Parameter messen könnte. Um dieses Problem zu umgehen gibt es eine Reihe von Möglichkeiten zur Fehlerberechnung, die trotzdem ein Maß für die zu erwartende Genauigkeit der Fusion bieten.

#### 5.2.1.1 Externe Marker

Eine Möglichkeit besteht darin, durch externe Markerpunkte ein Referenzsystem zu schaffen, das nicht wie unter 2.2.2.3 beschrieben primär für die Bildfusion verwendet wird, sondern nach erfolgter Bildfusion durch eine andere Methode lediglich zur Validierung des Ergebnisses dient. Doch die mit diesen Referenzsystemen erzielbare Genauigkeit liegt meist in der gleichen Größenordnung wie der Fusionsfehler, der damit gemessen werden soll. Selbst die invasive Variante des am Knochen verschraubten stereotaktischen Rahmens verhindert hierbei nicht, dass sich das Gehirn innerhalb des knöchernen Schädels bewegt. So berichten West et al. (West, Fitzpatrick et al. 1997) einen Fehler von 1,65 mm für die Lokalisation externer Marker an einem stereotaktischen Rahmen bei der

PET-MRT Fusion von Hirnstudien. Die Fehler der untersuchten Methoden lagen dabei zwischen 2,0 und 4,9 mm. Bei auf die Haut geklebten Obeflächenmarkern muss von noch höheren Fehlern bei der Markerlokalisierung ausgegangen werden, da hier gegenüber einem verschraubten stereotaktischen Rahmen noch die Hautverschieblichkeit als zusätzliche Fehlerquelle hinzukommt. Für die Beurteilung geringer Fehler bei der Bildfusion erscheinen externe Marker daher als Referenzsystem ungeeignet.

#### 5.2.1.2 Simulierte Studien

Eine weitere Möglichkeit zur Validierung von Fusionsverfahren ist das Erzeugen von simulierten Datensätzen. Für die Fusion gleicher Modalitäten, z.B. MR - MR oder PET - PET wird ein Originaldatensatz zunächst um bekannte Werte verschoben und rotiert und als neuer Datensatz gespeichert. Dieser transformierte Datensatz wird anschließend mit dem unveränderten Original fusioniert. Eine erfolgreiche Bildfusion muss also genau die negativen ursprünglichen Transformationsparameter ergeben. Eine noch verbleibende Differenz reflektiert einen entsprechenden Fehler der Bildfusion (Lin, Huang et al. 1996; Zuk und Atkins 1996).

Aber auch die Fusion unterschiedlicher Modalitäten ist mittels Simulationen möglich. Meist wird hierbei aus der anatomischen Modalität, z.B. einer MRT durch Segmentierung, Grauwertveränderungen, Filterungen und Interpolationen ein simulierter nuklearmedizinischer Datensatz erzeugt. Dieser wird wieder mit bekannten Parametern transformiert und mit dem Original MRT in der Ausgangsposition fusioniert (Strother, Anderson et al. 1994). Allerdings ist ein derart simulierter Datensatz nicht direkt mit realen nuklearmedizinischen Daten vergleichbar, wo schon physiologischerweise und erst recht bei pathologischen Befunden größere Diskrepanzen zwischen den Datensätzen unterschiedlicher Modalitäten bestehen, als dies bei einem simulierten Datensatz der Fall ist. Ein auf der Basis simulierter Studien bestimmter Fehler für eine Bildfusion unterschiedlicher Untersuchungsmodalitäten liegt somit eher unter dem Fehler, der bei Anwendung realer klinischer Daten zu erwarten wäre.

#### 5.2.1.3 Phantomstudie

Auch die Verwendung von Phantomen kommt in Frage um ein Bezugssystem zu schaffen. Es können Kunststoffphantome untersucht werden (Sipila, Nikkinen et

al. 1997), die allerdings in einigen Modalitäten deutlichere Konturen aufweisen als reale klinische Datensätze und damit ebenso wie simulierte Studien unrepräsentative Ergebnisse mit zu geringen Fehlern liefern können. Eine weitere Alternative sind Kadaverstudien, bei denen z.B. in einem Hirnpräparat interne Marker mit Kontrastmittel platziert werden, die bei der Fusion zur Deckung gebracht werden, wie bei Hemler et al (Hemler, Napel et al. 1995), die damit die Genauigkeit der Fusion von Schädel-CT und MRT prüften. Für nuklear-medizinische Untersuchungen wie die Perfusions-SPECT sind anatomische Präparate allerdings nicht anwendbar, da keine physiologischen Parameter wie Perfusion oder Stoffwechsel abgebildet werden können.

#### 5.2.1.4 Statistische Auswertung

Bei den beschriebenen Problemen der anderen Verfahren zur Definition eines Goldstandards bietet sich als weitere Lösungsmöglichkeit eine statistische Untersuchung der Transformationsparameter an.

So kann, wie in dieser Arbeit, aus den Transformationsparametern mehrerer Fusionsversuche ein Mittelwert gebildet werden, der als Referenzposition definiert wird und die Abweichung jedes einzelnen Fusionsversuches von dieser Referenzposition kann berechnet werden. Auch wenn systematische Fehler mit diesem Ansatz nicht erkannt werden, liefert die Streuung der einzelnen Transformationen um die Referenzposition innerhalb einer Methode dennoch ein Maß für die Reproduzierbarkeit und damit die Stabilität einer Fusionsmethode.

Bei der Untersuchung von interaktiven Methoden könnten die einzelnen Transformationen für die Mittelwertbildung jeweils von mehreren Fusionen verschiedener Anwendern stammen (Pietrzyk, Herholz et al. 1994; Perault, Schwartz et al. 1997; Woods, Grafton et al. 1998). Ein Vergleich zwischen Intra- und Interobservervariabilität erlaubt hierbei eine Untersuchung der Subjektivität eines Verfahrens.

Kapouleas et al. (Kapouleas, Alavi et al. 1991) betrachteten bei ihrem auf internen Landmarken basierenden Verfahren nicht nur die Differenz zwischen den Ergebnissen verschiedener Anwender, sondern auch die Veränderung des Ergebnisses durch den Ausfall jeweils einzelner von insgesamt ca. 35 identifizierten Landmarken.

Bei automatischen Methoden kann die Reproduzierbarkeit von Fusionen mit jeweils verschiedenen initialen Startpositionen oder mit unterschiedlichen methodischen Parametern des verwendeten Algorithmus bewertet werden.

Ardekani et al. untersuchten die Reproduzierbarkeit ihrer Ergebnisse bei der Fusion von acht nacheinander aufgenommenen PET Datensätzen mit immer dem gleichen MR Datensatz (Ardekani, Braun et al. 1995).

Da bei der reinen Betrachtung der Reproduzierbarkeit von Fusionsparametern immer noch systematische Fehler möglich wären, ist eine zusätzliche visuelle Plausibilitätsprüfung der erzielten Fusionsergebnisse notwendig.

Auch eine systemübergreifende Mittelwertbildung aus den Ergebnissen verschiedener Fusionsmethoden ist möglich, um systematische Fehler auszuschließen.

Andersson et al. (Andersson, Sundin et al. 1995) verwendeten als Fehlermaß den direkten Vergleich zwischen den Transformationsparametern eines interaktiven Verfahrens und denen des zu untersuchenden automatischen Algorithmus. Wahl et al (Wahl, Quint et al. 1993) berechneten den Fehler aus der Diskrepanz zwischen nachträglicher Identifizierung interner Landmarken und deren durch die aktuelle Fusion vorgegebenen Position.

Insgesamt kann die statistische Auswertung wiederholter Fusionsvorgänge als etabliertes Verfahren betrachtet werden. Die damit untersuchbare Reproduzierbarkeit einer Bildfusion bietet sich als Fehlermaß bei der Fusion klinischer Datensätze an.

## 5.2.2 Art der Fehlerberechnung

Hat man einen Goldstandard definiert, gibt es mehrere Möglichkeiten die Abweichung einer Datensatzposition von dieser Referenzposition anzugeben:

### 5.2.2.1 Parameterangabe, 2D Fehler

Die einfachste Möglichkeit ist hierbei die Angabe der Transformationsparameter selbst, bzw. deren Streuung, z.B. in Millimetern Translation und Grad Rotation für jede Achse des Koordinatensystems (Zuk und Atkins 1996; Vaarkamp 2001).

Diese Form ist jedoch sehr unanschaulich, da es schwer ist, aus der Kombination dieser einzelnen Parametern ein Gesamtmaß für den Fehler abzuschätzen.

Insbesondere der Rotationswinkel kann mit zunehmender Entfernung vom Drehpunkt einen sehr großen Einfluss haben, der aus der Angabe einzelner Parameter praktisch nicht abzuschätzen ist.

Auch die Messung von Abständen innerhalb korrespondierender Schichten wird als Fehlermaß verwendet: Holman et al. (Holman, Zimmerman et al. 1991) verwenden den in Millimetern angegebenen Abstand zwischen den manuell segmentierten Thalami in beiden Modalitäten nach erfolgter Fusion von Hirn-SPECT und MRT. Analog dazu haben Scott et al. (Scott, Macapinlac et al. 1994) den Abstand zwischen lokalisierten Tumoren in abdominalen SPECT und CT/MR Studien gemessen. Steinmetz et al. geben für jede orthogonale Projektionsrichtung den Abstand zwischen Organkonturen an (Steinmetz, Huang et al. 1992) was bereits eine bessere Abschätzung der klinischen Relevanz erlaubt als die bloße Angabe der Transformationsparameter.

### 5.2.2.2 3D Fehler-Angabe

Ein klinisch verwertbares Fehlermaß sollte auf einen Blick eine quantitative Beurteilung der zu erwartenden Differenz zwischen der Referenzposition und dem aktuellen Fusionsergebnis erlauben. Am anschaulichsten erfolgt dies wie in der hier durchgeführten Untersuchung durch Berechnung von absoluten räumlichen Abständen zwischen korrespondierenden Punkten aus den Transformationsparametern, einschließlich der Rotationen.

Die Größe dieses Fehlers ist wie unter 3.5.3 beschrieben nicht gleichmäßig über den ganzen Datensatz verteilt, insbesondere nimmt der Fehler mit zunehmender Entfernung vom geometrischen Mittelpunkt des Datensatzes zu, da hier die

Rotationskomponente der Transformationen einen größeren Einfluss bekommt. West et al. konnten hier zeigen, dass diese zunehmende Rotationskomponente des Fusionsfehlers für "oberflächenbasierte" Methoden wie den Pelizzari-Algorithmus größer war, als für "volumenbasierte" Methoden wie nach Woods (West, Fitzpatrick et al. 1999).

Bei der Angabe eines Fehlers für die Bildfusion muss daher genau spezifiziert werden, auf welche Positionen im Datensatz sich der ermittelte Fehler bezieht.

Bei Verwendung externer Marker ist es möglich, den Fusionsfehler für die Markerposition zu ermitteln (Elsen und Viergever 1994), wenngleich diese oberflächlichen Marker oft sehr weit von den eigentlich interessierenden Strukturen entfernt sind.

Idealerweise wird eine individuelle region of interest (ROI) gewählt, um zu gewährleisten, dass nur Punkte innerhalb der tatsächlich interessierenden Regionen für die Fehlerberechnung herangezogen werden. Für Hirnuntersuchungen lässt sich dazu, wie in dieser Arbeit, eine Maske erstellen, die das Gehirn markiert. Für die Fehlerberechnung werden jetzt beispielsweise nur Punkte innerhalb dieser Maske verwendet, außerhalb liegende Punkte bleiben unberücksichtigt (Strother, Anderson et al. 1994). West et al. haben als ROI acht Punkte definiert, die jeweils den Mittelpunkt einer „chirurgisch interessanten“ Region darstellen. Der Fehler wurde als Mittelwert dieser acht Punkte angegeben (West, Fitzpatrick et al. 1997). Pietrzyk et al. berechnen den Fehler als Mittelwert aus fünf definierten Punkten, im Zentrum und an den vier äußeren Polen des Hirns (Pietrzyk, Herholz et al. 1994).

Speziell die geometrische Verteilung der Fusionsfehlern findet in klinischen Anwendungen oft nur wenig Beachtung. Bei Neuronavigationssystemen ist es beispielsweise üblich, nach erfolgter Kalibrierung des Systems die erzielte Genauigkeit im geometrischen Mittelpunkt des Datensatzes anzugeben. Genau hier ist jedoch die Rotationskomponente einer fehlerhaften Transformation am geringsten und es entsteht der Eindruck einer unrepräsentativ hohe Genauigkeit. Sinnvoller wäre auch hier die Berechnung des Fehlers innerhalb einer vom Operateur definierten, auf den individuell geplanten Eingriff zugeschnittenen ROI. Für Untersuchungen bei denen z.B. mangels scharfer Organkonturen die Definition einer ROI schwierig ist, kann man sich der Organkontur durch künstlich

erstellte Masken anzunähern. Die dafür meist am besten geeignete Möglichkeit ist die Verwendung kugelförmiger Masken in der Größe des zu untersuchenden Organs. Dem Gehirn kann man sich z.B. durch eine Kugel mit 15 cm Durchmesser relativ gut annähern, und damit für die Fehlerberechnung einen annähernd repräsentativen Ausschnitt aus dem Datensatz markieren.

Die eigentlich ideale Vorgehensweise, für jede Bildfusion eine individuelle ROI zu definieren, erfordert andererseits einen zusätzlichen Zeitaufwand, so dass hier realisierbare Kompromisslösungen für den klinischen Alltag gefragt sind.

### 5.2.2.3 Trefferquote

Anstelle der direkten Angabe des absoluten Fehlers ist auch die Definition eines „Trefferbereiches“ und die Angabe einer „Trefferquote“ möglich, die angibt, wie viele Matchingversuche eines Verfahrens innerhalb des vorher definierten Toleranzintervalls liegen (van Herk und Kooy 1994; Studholme, Hill et al. 1997). Dieses Toleranzintervall ist natürlich beliebig definierbar und stellt sicherlich eine vermeidbare Vergrößerung der Genauigkeitsbestimmung dar.

## 5.3 Ergebnisse der Fehlerberechnung

In der vorliegenden Arbeit wurden die besten Ergebnisse durch die manuelle Bildfusion erzielt, mit einer Intraobservervariabilität von 1,5 mm. Die nur minimal höhere Interobservervariabilität von 1,6 mm belegt, dass die manuelle Bildfusion keine starke Benutzerabhängigkeit zeigt, sondern auch von verschiedenen Benutzern reproduzierbar zu vergleichbaren Ergebnissen führt.

Frühere Studien zu manuellen Fusionsmethoden erbrachten teilweise größere Fehler: Pietrzyk et al. untersuchten die Reproduzierbarkeit der einzelnen Transformationsparameter für die Fusion von MRT und PET und berichteten eine Standardabweichung von maximal 2,2 mm für die Translation und maximal 1,7 Grad für die Rotation, ein dreidimensional ermittelter Abstand zwischen Punkten wurde hier nicht berechnet (Pietrzyk, Herholz et al. 1994). Beide Standardabweichungen liegen über unseren Werten von max. 0,7 mm für die Translation bzw. 1,2 Grad für die Rotation. Teilweise ist dieser größere Fehler durch die von Pietrzyk et al. verwendeten MRT Daten mit einer Schickdicke von 2 mm (hier 1,2 mm) erklärbar. Ein weiterer wichtiger Faktor ist sicherlich das Benutzerinterface

für die manuelle Bildfusion. Anders als in der Studie von Pietrzyk et al. ist bei der hier verwendeten VoxelQ Workstation eine beliebig vergrößerte Anzeige der zu fusionierenden Daten möglich, was eine Beurteilung und Manipulationen im Pixel- und Subpixelbereich erleichtert. Eine Studie von Habboush et al. (Habboush, Mitchell et al. 1996) untersuchte die interaktive manuelle Bildfusion von MRT und SPECT Daten, wobei auch hier MRT Daten mit größerer Schichtdicke (3 mm) verwendet wurden, die zudem auf die Pixelgeometrie der SPECT Daten interpoliert wurden (Auflösung 128 x 128 pixel à 2,45 mm). Bei der Fusion von Patientendaten wurde auch hier die Reproduzierbarkeit der interaktiven Bildfusion durch mehrere Benutzer als Fehlermaß herangezogen. Für die Position dreier Testpunkte im Datensatz ergaben sich Abweichungen bis zu 5,9 mm mit einem Mittelwert von 2,9 mm. Bemerkenswert ist, dass auch in der damaligen Studie der Fehler für die interaktive Bildfusion mit 2,9 mm knapp oberhalb der Auflösung der MRT lag. In der vorliegenden Arbeit betrug der Fehler 1,5 mm bei einer MRT-Auflösung von 1,2 mm.

Für die automatische Bildfusion mittels surface matching nach Pelizzari ermittelten wir einen Fehler von 2,9 bzw. 3,8 mm. Dieses Ergebnis ist relativ konsistent mit früheren Studien: So ermittelten West et al. im Rahmen einer Multicenter-Studie bei der Fusion vom MRT und PET Daten einen Fehler von 2,9 mm für die Fusion durch die Arbeitsgruppe um Pelizzari, sowie Fehler von 2,8 , 4,0 und 4,6 mm für andere Implementationen von oberflächenbasierten Bildfusionsmethoden (West, Fitzpatrick et al. 1997). In anderen Studien wurden Fehler bis 4,6 mm Translation (Turkington, Hoffman et al. 1995) bzw. eine räumliche Distanz von 2,4 mm zwischen korrespondierende Punkten (Strother, Anderson et al. 1994) angegeben, ermittelt jeweils für die Fusion von MRT und PET. Obwohl die Mehrzahl der früheren Studien mit Daten geringerer Auflösung durchgeführt wurden, zeigen sich hier vergleichbare Ergebnisse. Oder um es anders zu formulieren: Die Bildfusion mittels surface matching scheint nicht im gleichen Maße von einer höheren Auflösung der Bilddaten zu profitieren, wie die manuelle Bildfusion. Dies erscheint plausibel, da eine höhere Auflösung lediglich zu einer Glättung der Gesamtkontur der Hirnoberfläche führt, die Form der Kontur, die für die Fusion entscheidend ist, aber nicht wesentlich verändert wird.

Für die pixel-uniformity basierte Bildfusion nach Woods ergab sich in unserer Studie eine Reproduzierbarkeit mit einem Fehler von 2,2 mm. Auch diese Ergebnisse stimmen gut mit Literaturdaten überein. Die West-Studie (West, Fitzpatrick et al. 1997) berichtete einen Fehler von 2,3 mm für die Fusion durch Wood's Alorithmus. Eine manuelle Maskierung fokaler pathologischer Befunde vor der Fusion konnte dort die Ergebnisse nicht verbessern (Fehler nach Maskierung: 3,1 mm). Dies scheint eine gewisse Robustheit der Methode gegenüber einzelnen, lokalen Inkonsistenzen zu belegen. Offensichtlich schöpft hier die pixelbasierte Methode Vorteile aus dem Prinzip, Korrelationen aus den gesamten Volumen des Datensatzes zu berücksichtigen. Ein Überwiegen des "passendes" Volumen kann damit ausreichen, um lokale "Fehler" im Datensatz zu kompensieren.

Auch Black et al. ermittelten einen Fehler von 2,2 mm für die Fusion nach Woods. (Black, Videen et al. 1996). Das deutlich bessere Ergebnis für die Fusion nach Woods in der Studie von Strother et al. mit einem Fehler von nur 1,3 mm ist möglicherweise dadurch zu begründen, dass dort simulierte PET Daten durch Nachbearbeitung und Filterung einer MRT erzeugt wurden, die dann mit der Original-MRT fusioniert wurden. Diese Vorgehensweise erklärt natürlich ein sehr hohes Maß an Korrelation zwischen beiden Datensätzen, die wohl nicht repräsentativ für klinische Daten ist (s. auch 5.2.1.2). Eine pixelbasierte Methode wie der Woods-Algorithmus profitiert von dieser Verfälschung offensichtlich stärker als beispielsweise das surface matching, wo der von Strother bestimmte Fehler nur etwas geringer war, als der von uns ermittelte.

Das Versagen der Methode nach Woods in vier von 15 Fällen belegt, dass auch ein komplexer Algorithmus immer gefährdet ist, einem systematischen Fehler zu unterliegen. Meistens ist hierfür wohl die Definition der Fehlerfunktion verantwortlich, deren Minimierung offensichtlich in einigen Fällen nicht mit einer Verbesserung des Fusionsergebnisses einhergeht.

## **5.4 Zeitbedarf**

Der ermittelte Zeitbedarf für die Bildfusion war für das pixel uniformity matching nach Woods mit 9 Minuten am geringsten. Die Fehleranfälligkeit dieser Methode zeigt allerdings, dass jede automatisch durchgeführte Bildfusion einer visuellen

Plausibilitätskontrolle des Fusionsergebnisses durch einen erfahrenen Auswerter bedarf. Wong et al. konnten zeigen, dass Dislokationen von 2-3 Millimetern Translation und 2-4° Rotation zuverlässig von jedem Untersucher detektiert werden konnten (Wong, Studholme et al. 1997). Um auch diskrete Fehler von wenigen Millimetern Dislokation zu entdecken, erfordert diese visuelle Kontrolle einen zusätzlichen Zeitaufwand von ca. 3-5 Minuten pro Fall. Vor diesem Hintergrund ist ein Vorteil für die interaktive Bildfusion zu sehen, da hier während des gesamten Fusionsprozesses ständig eine visuelle Beurteilung erfolgt und somit keine separate anschließende Kontrolle mehr notwendig ist. Die Gesamtbearbeitungszeit für den Fusionsvorgang und die Überprüfung wäre damit für die manuelle Bildfusion mit 11 Minuten an kürzesten.

Die Methode nach Pelizzari hat gezeigt, dass die Datenvorverarbeitung, hier insbesondere die Segmentierung der Hirnoberfläche einen z.T. erheblichen zeitlichen Aufwand darstellt, der bei Verwendung der manuellen Bildfusion eingespart werden kann.

## **5.5 Gesamtbewertung der Ergebnisse**

Die manuelle Bildfusion mit einem geeigneten interaktiven Benutzerinterface führte zu den besten Fusionsergebnissen. Dem Nachteil der zeitaufwendigen Interaktion steht der Vorteil gegenüber, dass bei manuellen Verfahren zum einen keine Vorverarbeitung der Daten notwendig ist und zum anderen keine separate Überprüfung der erfolgten Fusion mehr notwendig ist. Unter Berücksichtigung der Gesamtbearbeitungszeit ist somit ebenfalls die manuelle Bildfusion die Methode der ersten Wahl.

Die Leistung der hier verwendeten automatischen Verfahren soll jedoch nicht unterschätzt werden: Alle getesteten Methoden erzielten Genauigkeiten die unter der physikalischen Auflösung vieler nuklearmedizinischer Untersuchungen liegt. Somit sind im Prinzip alle drei Methoden für die Bildfusion von MRT und SPECT Untersuchungen des Gehirns einsetzbar.

Dennoch zeigt die manuelle Bildfusion hier eine deutliche Überlegenheit. Die Stärke der automatischen Verfahren ist daher aktuell eher bei der Fusion von Bilddaten gleicher Modalitäten zu sehen, also beispielsweise PET mit PET oder MRT mit MRT. Prinzipiell werden bei der Fusion gleicher Modalitäten immer

höhere Korrelationen der Daten vorliegen, was eine Erfassung durch ein mathematisches Modell einfacher und robuster macht. Dies ermöglicht es den Methoden, ihre Komplexität zu entfalten und hohe Genauigkeiten zu erzielen (Woods, Grafton et al. 1998; Zhilkin und Alexander 2004).

## 6 Zusammenfassung

Ein Ziel der Arbeit war die Entwicklung eines universell anwendbaren Fehlermodells, das eine repräsentative Angabe der Bildfusionsgenauigkeit ermöglicht.

Das vorgestellte Fehlermodell erlaubt eine solche Fehlerberechnung. Die Verwendung des Mittelwertes aus fünf wiederholten Fusionsvorgängen als Goldstandard erlaubt eine praktikable Reliabilitätsabschätzung einer Fusionsmethode. Alternative Referenzsysteme sind selbst mit zu großen Messungenauigkeiten behaftet oder sind für eine retrospektive Datenverarbeitung nicht anwendbar. Die Ergebnisse aus Simulationen oder Phantommessungen erlauben keinen Rückschluss auf die Leistungsfähigkeit einer Methode bei Anwendung auf reale klinische Daten. Insgesamt gibt es damit für klinische Datensätze praktisch keine Alternative zu dem hier verwendeten Mittelwert als definiertem Goldstandard. Die grundsätzlich mögliche Schwäche, hierbei systematische Fehler zu übersehen, kann durch visuelle Kontrolle der Fusionsergebnisse zuverlässig vermieden werden.

Die hier etablierte Bestimmung eines direkten räumlichen Abstandes zwischen korrespondierenden Punkten im dreidimensionalen Raum erlaubt eine anschauliche, für den klinischen Einsatz gut einzuordnende Abschätzung der Fusionsgenauigkeit. Die Ermittlung eines durchschnittlichen Fehlers für 5000 einzeln berechnete Bildpunkte erlaubt eine zuverlässige, repräsentative Darstellung des Fehlers. Die Verwendung individueller Masken innerhalb derer die Fehlerberechnung erfolgt, ermöglicht die Bestimmung des Fehlers speziell für die relevanten Bereiche eines Datensatzes und vermeidet eine Verfälschung des ermittelten Fehlers durch Messwerte aus klinisch nicht relevanten Bereichen des Datensatzes.

Das erarbeitete Fehlerberechnungsmodell wurde für den Vergleich von drei unterschiedlichen Bildfusionsverfahren angewendet. Mittels manueller Bildfusion und durch die beiden am besten etablierten automatischen Bildfusionsverfahren wurden 55 Bildfusionen von T1-gewichteter MRT und ECD-Perfusions-SPECT jeweils 5 mal wiederholt und der Fehler für jeden der 275 einzelnen Fusionsvorgänge bestimmt.

Der Methodenvergleich hat die Überlegenheit der manuellen Bildfusion gegenüber den automatischen Ansätzen in allen Bereichen belegt: Sowohl die erzielbare Genauigkeit mit einer Intraobservervariabilität von 1,5 Millimetern, als auch der Zeitaufwand von insgesamt durchschnittlich 11 Minuten sprechen für die interaktive Bildfusion. Die nur minimal höhere Interobservervariabilität von 1,6 Millimetern belegt, dass der Vorwurf der fehlenden Objektivität für die manuelle Bildfusion nicht haltbar ist. Das relativ gute Ergebnis für die manuelle Bildfusion im Vergleich zu Literaturdaten belegt zum einen, dass hier alle Vorteile einer höheren räumlichen Auflösung der Untersuchungsdaten ausgeschöpft werden können und zeigt gleichzeitig, dass ein optimiertes Benutzerinterface für die erzielbare Qualität einer manuellen Bildfusion entscheidend ist.

Das surface matching nach Pelizzari disqualifiziert sich für einen Routineeinsatz sowohl durch den hohen Zeitaufwand von durchschnittlich 26 Minuten, als auch durch den ermittelten Fehler von 2,9 bzw. 3,8 Millimeter Intra- bzw. Interobservervariabilität. Die hier ermittelten Ergebnisse für das surface matching, die nur unwesentlich besser waren als in früheren Studien mit schlechterer Auflösung der Bilddaten zeigen auch, dass diese Methode kaum von einer Verbesserung der Gerätetechnik, z.B. einer höheren Auflösung profitiert.

Für das pixel uniformity matching nach Woods lag der ermittelte Fehler bei durchschnittlich 2,2 Millimeter, was für eine klinische Anwendung prinzipiell ausreichend wäre. Das Auftreten von groben Fusionsfehlern mit reproduzierbaren Rotationsfehlstellungen von  $> 30$  Grad in vier der 15 Fälle zeigt aber die Anfälligkeit der Methode für systematische Fehler. Hiermit wird klar, dass eine visuelle Kontrolle der Ergebnisse zwingend notwendig ist, so dass sich der Zeitbedarf von 9 Minuten für die eigentliche Bildfusion um diese Kontrolle verlängert und der Gesamtaufwand damit den für eine manuelle Bildfusion übersteigt.

Prinzipielle Überlegungen lassen erwarten, dass auch zukünftige, verbesserte Verfahren zur automatischen Bildfusion Probleme aufweisen werden:

Letztendlich basieren alle automatischen Algorithmen auf der Annahme einer Analogie der zu fusionierenden Datensätze. Die verschiedenen klinischen Fragestellungen an bildgebende Untersuchungen ergeben sich aber gerade aus der unterschiedlichen Sensitivität einzelner Verfahren für die Detektion bestimmter Pathologien. Ein Befund in einer Modalität, der in einer anderen

Modalität kein Korrelat zeigt, ist daher ein elementares Prinzip der multimodalen Diagnostik und aus klinischer Sicht gerade erwünscht. Für einen automatischen Bildfusionsalgorithmus bedeutet gerade dies aber immer einen „Fehler“ der durch Änderung der Transformationsparameter zu korrigieren versucht wird. Diese systematische Schwäche aller automatischen Fusionsmethoden wird auch bei der Weiterentwicklung komplexerer automatischer Algorithmen weiter bestehen bleiben.

Der zusätzliche Aspekt der Akzeptanz für klinische Anwendungen ist nicht zu unterschätzen. Hier kann die manuelle Bildfusion durch einen erfahrenen Benutzer weitere erhebliche Vorteile für sich in Anspruch nehmen. Insbesondere bei vorliegen pathologischer Befunde oder bei der Erkennung von Artefakten kann ein geschulter Benutzer schnell und individuell reagieren.

Insgesamt hat sich damit eine eindeutige Überlegenheit der manuellen Bildfusion gegenüber den automatischen Verfahren nach Pelizzari und Woods gezeigt. In der prächirurgischen Diagnostik des Epilepsiezentrum München wird daher ausschließlich die interaktive Bildfusion eingesetzt und hat dort einen festen Stellenwert in der routinemäßig durchgeführten multimodalen Diagnostik, insbesondere bei der Befundintegration für die Planung resektiver neurochirurgischer Eingriffe.

Das erarbeitete Fehlerberechnungsmodell stellt eine Erweiterung der bisher in der Literatur beschriebenen Methoden dar und erlaubt eine zuverlässige, repräsentative und klinisch verwertbare Beurteilung der erzielbaren Fusionsgenauigkeit. Neben der hier beschriebenen Bildfusion von MRT und SPECT Daten wurde das etablierte Modell inzwischen auch für die Evaluation der Bildfusion anderer Modalitäten wie MRT – PET, MRT - CT oder MRT - MR-Angiographie eingesetzt.

## Literaturverzeichnis

- Alpert, N. M., Bradshaw, J. F., et al. (1990). "The principal axes transformation-- a method for image registration." J Nucl Med **31**(10): 1717-22.
- Andersson, J. L., Sundin, A., et al. (1995). "A method for coregistration of PET and MR brain images." J Nucl Med **36**(7): 1307-15.
- Ardekani, B. A., Braun, M., et al. (1995). "A fully automatic multimodality image registration algorithm." J Comput Assist Tomogr **19**(4): 615-23.
- Bale, R. J., Burtscher, J., et al. (2000). "Computer-assisted neurosurgery by using a noninvasive vacuum-affixed dental cast that acts as a reference base: another step toward a unified approach in the treatment of brain tumors." J Neurosurg **93**(2): 208-13.
- Bale, R. J., Vogele, M., et al. (1997). "VBH head holder to improve frameless stereotactic brachytherapy of cranial tumors." Comput Aided Surg **2**(5): 286-91.
- Black, K. J., Videen, T. O., et al. (1996). "A metric for testing the accuracy of cross-modality image registration: validation and application." J Comput Assist Tomogr **20**(5): 855-61.
- Brandau, W., Becker, W., et al. (2001). Radiochemie. Nuklearmedizin. U. Büll. Stuttgart, Georg Thieme Verlag: 79-112.
- Chen, H. M. und Varshney, P. K. (2003). "Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation." IEEE Trans Med Imaging **22**(9): 1111-9.
- Collins, D. L., Neelin, P., et al. (1994). "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space." J Comput Assist Tomogr **18**(2): 192-205.
- de Munck, J. C., Verster, F. C., et al. (1998). "Registration of MR and SPECT without using external fiducial markers." Phys Med Biol **43**(5): 1255-69.
- Dhawan, A. P. und Arata, L. (1992). "Knowledge-based multi-modality three-dimensional image analysis of the brain." Am J Physiol Imaging **7**(3-4): 210-9.
- Elsen, P. A. v. d. und Viergever, M. A. (1994). "Marker-guided multimodality matching of the brain." Eur. Radiol. **4**: 45-51.
- Evans, A. C., Marrett, S., et al. (1991). "MRI-PET correlation in three dimensions using a volume-of-interest (VOI) atlas." J Cereb Blood Flow Metab **11**(2): A69-78.
- Ferrolli, P., Franzini, A., et al. (2004). "A simple method to assess accuracy of deep brain stimulation electrode placement: pre-operative stereotactic CT + postoperative MR image fusion." Stereotact Funct Neurosurg **82**(1): 14-9.
- Fiedler, E., Platsch, G., et al. (2003). "[Time consumption and quality of an automated fusion tool for SPECT and MRI images of the brain]." Nuklearmedizin **42**(5): 215-9.

- Ge, Y., Fitzpatrick, J. M., et al. (1994). "Retrospective registration of PET and MR brain images: an algorithm and its stereotactic validation." J Comput Assist Tomogr **18**(5): 800-10.
- Guimond, A., Roche, A., et al. (2001). "Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections." IEEE Trans Med Imaging **20**(1): 58-69.
- Habboush, I. H., Mitchell, K. D., et al. (1996). "Registration and alignment of three-dimensional images: an interactive visual approach." Radiology **199**(2): 573-8.
- Hawkes, D. J. (1998). "Algorithms for radiological image registration and their clinical application." J Anat **193**(Pt 3): 347-61.
- He, R. und Narayana, P. A. (2002). "Global optimization of mutual information: application to three-dimensional retrospective registration of magnetic resonance images." Comput Med Imaging Graph **26**(4): 277-92.
- Hemler, P. F., Napel, S., et al. (1995). "Registration error quantification of a surface-based multimodality image fusion system." Med Phys **22**(7): 1049-56.
- Holman, B. L., Zimmerman, R. E., et al. (1991). "Computer-assisted superimposition of magnetic resonance and high- resolution technetium-99m-HMPAO and thallium-201 SPECT images of the brain." J Nucl Med **32**(8): 1478-84.
- Hutton, B. F., Braun, M., et al. (2002). "Image registration: an essential tool for nuclear medicine." Eur J Nucl Med Mol Imaging **29**(4): 559-77.
- Imran, M. B., Kawashima, R., et al. (1999). "Parametric mapping of cerebral blood flow deficits in Alzheimer's disease: a SPECT study using HMPAO and image standardization technique." J Nucl Med **40**(2): 244-9.
- Kaiser, H. J., Sabri, O., et al. (1994). "A method of correlating and merging cerebral morphology and function by a special head holder." Nuklearmedizin **33**(3): 123-6.
- Kapouleas, I., Alavi, A., et al. (1991). "Registration of three-dimensional MR and PET images of the human brain without markers." Radiology **181**(3): 731-9.
- Knowlton, R. C., Laxer, K. D., et al. (2001). "In vivo hippocampal glucose metabolism in mesial temporal lobe epilepsy." Neurology **57**(7): 1184-90.
- Kostelec, P. J., Weaver, J. B., et al. (1998). "Multiresolution elastic image registration." Med Phys **25**(9): 1593-604.
- Krishnan, R., Hermann, E., et al. (2003). "Automated fiducial marker detection for patient registration in image-guided neurosurgery." Comput Aided Surg **8**(1): 17-23.
- Levin, D. N., Pelizzari, C. A., et al. (1988). "Retrospective geometric correlation of MR, CT, and PET images." Radiology **169**(3): 817-23.
- Lin, K. P., Huang, S. C., et al. (1996). "Automated image registration for FDOPA PET studies." Phys Med Biol **41**(12): 2775-88.

- Mangin, J. F., Frouin, V., et al. (1994). "Fast nonsupervised 3D registration of PET and MR images of the brain." J Cereb Blood Flow Metab **14**(5): 749-62.
- Maurer, C. R., Jr., Fitzpatrick, J. M., et al. (1997). "Registration of head volume images using implantable fiducial markers." IEEE Trans Med Imaging **16**(4): 447-62.
- Miga, M. I., Sinha, T. K., et al. (2003). "Cortical surface registration for image-guided neurosurgery using laser-range scanning." IEEE Trans Med Imaging **22**(8): 973-85.
- Munley, M. T., Marks, L. B., et al. (1999). "Multimodality nuclear medicine imaging in three-dimensional radiation treatment planning for lung cancer: challenges and prospects." Lung Cancer **23**(2): 105-14.
- Murphy, M. A., O'Brien, T. J., et al. (2004). "Multimodality image-guided surgery for the treatment of medically refractory epilepsy." J Neurosurg **100**(3): 452-62.
- Myers, R. (2002). "The application of PET-MR image registration in the brain." Br J Radiol **75 Spec No**: S31-5.
- Nelder, J. A. und Mead, R. (1965). "The simplex method for function minimization." Comput. J. **7**: 308-313.
- Nelson, S. J., Day, M. R., et al. (1997). "Alignment of volume MR images and high resolution [18F]fluorodeoxyglucose PET images for the evaluation of patients with brain tumors." J Comput Assist Tomogr **21**(2): 183-91.
- O'Brien, T. J., Brinkmann, B. H., et al. (1999). "Comparative study of 99mTc-ECD and 99mTc-HMPAO for peri-ictal SPECT: qualitative and quantitative analysis." J Neurol Neurosurg Psychiatry **66**(3): 331-9.
- O'Brien, T. J., O'Connor, M. K., et al. (1998). "Subtraction ictal SPET co-registered to MRI in partial epilepsy: description and technical validation of the method with phantom and patient studies." Nucl Med Commun **19**(1): 31-45.
- Pelizzari, C. A., Chen, G. T., et al. (1989). "Accurate three-dimensional registration of CT, PET, and/or MR images of the brain." J Comput Assist Tomogr **13**(1): 20-6.
- Perault, C., Schwartz, C., et al. (1997). "Thoracic and abdominal SPECT-CT image fusion without external markers in endocrine carcinomas. The Group of Thyroid Tumoral Pathology of Champagne-Ardenne." J Nucl Med **38**(8): 1234-42.
- Pietrzyk, U., Herholz, K., et al. (1994). "An interactive technique for three-dimensional image registration: validation for PET, SPECT, MRI and CT brain studies." J Nucl Med **35**(12): 2011-8.
- Pietrzyk, U., Herholz, K., et al. (1990). "Three-dimensional alignment of functional and morphological tomograms." J Comput Assist Tomogr **14**(1): 51-9.

- Powell, M. J. D. (1964). "An efficient method for finding the minimum of a function of several variables without calculating derivatives." Comput. J. **7**: 155-163.
- Quarantelli, M., Berkouk, K., et al. (2004). "Integrated software for the analysis of brain PET/SPECT studies with partial-volume-effect correction." J Nucl Med **45**(2): 192-201.
- Rajasekar, D., Datta, N. R., et al. (2003). "Multimodality image fusion in dose escalation studies of brain tumors." J Appl Clin Med Phys **4**(1): 8-16.
- Rusinek, H., Tsui, W. H., et al. (1993). "Principal axes and surface fitting methods for three-dimensional image registration." J Nucl Med **34**(11): 2019-24.
- Scarfone, C., Lavelly, W. C., et al. (2004). "Prospective Feasibility Trial of Radiotherapy Target Definition for Head and Neck Cancer Using 3-Dimensional PET and CT Imaging." J Nucl Med **45**(4): 543-552.
- Schild, H. (1990). MRI made easy. Berlin, Schering AG.
- Scott, A. M., Macapinlac, H. A., et al. (1994). "Clinical validation of SPECT and CT/MRI image registration in radiolabeled monoclonal antibody studies of colorectal carcinoma." J Nucl Med **35**: 1976-84.
- Shannon, C. E. (1948). "A mathematical theory of communication." Bell System Tech J **27**: 379-423.
- Sipila, O., Nikkinen, P., et al. (1997). "Accuracy of a registration procedure for brain SPET and MRI: phantom and simulation studies." Nucl Med Commun **18**(6): 517-26.
- Steinmetz, H., Huang, Y., et al. (1992). "Individual integration of positron emission tomography and high- resolution magnetic resonance imaging." J Cereb Blood Flow Metab **12**(6): 919-26.
- Strother, S. C., Anderson, J. R., et al. (1994). "Quantitative comparisons of image registration techniques based on high- resolution MRI of the brain." J Comput Assist Tomogr **18**(6): 954-62.
- Studholme, C., Hill, D. L., et al. (1997). "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures." Med Phys **24**(1): 25-35.
- Thurfjell, L., Lau, Y. H., et al. (2000). "Improved efficiency for MRI-SPET registration based on mutual information." Eur J Nucl Med **27**(7): 847-56.
- Thurfjell, L., Pagani, M., et al. (2000). "Registration of neuroimaging data: implementation and clinical applications." J Neuroimaging **10**(1): 39-46.
- Treves, S. T., Mitchell, K. D., et al. (1998). "Three dimensional image alignment, registration and fusion." Q J Nucl Med **42**(2): 83-92.
- Turkington, T. G., Hoffman, J. M., et al. (1995). "Accuracy of surface fit registration for PET and MR brain images using full and incomplete brain surfaces." J Comput Assist Tomogr **19**(1): 117-24.

- Turkington, T. G., Jaszczak, R. J., et al. (1993). "Accuracy of registration of PET, SPECT and MR images of a brain phantom." J Nucl Med **34**(9): 1587-94.
- Vaarkamp, J. (2001). "Reproducibility of interactive registration of 3D CT and MR pediatric treatment planning head images." J Appl Clin Med Phys **2**(3): 131-7.
- Valenti, M. P., Froelich, S., et al. (2002). "Contribution of SISCOM imaging in the presurgical evaluation of temporal lobe epilepsy related to dysembryoplastic neuroepithelial tumors." Epilepsia **43**(3): 270-6.
- van den Elsen, P. A., Evert, J., et al. (1993). "Medical image matching --- a review with classification." IEEE Eng Med Biol **93**(3): 26-39.
- van Herk, M. und Kooy, H. M. (1994). "Automatic three-dimensional correlation of CT-CT, CT-MRI, and CT-SPECT using chamfer matching." Med Phys **21**(7): 1163-78.
- Van Laere, K. J., Warwick, J., et al. (2002). "Analysis of clinical brain SPECT data based on anatomic standardization and reference to normal data: an ROC-based comparison of visual, semiquantitative, and voxel-based methods." J Nucl Med **43**(4): 458-69.
- Wahl, R. L., Quint, L. E., et al. (1993). "'Anatomometabolic' tumor imaging: fusion of FDG PET with CT or MRI to localize foci of increased activity." J Nucl Med **34**(7): 1190-7.
- Wang, C., Pahl, J. J., et al. (1994). "A method for co-registering three-dimensional multi-modality brain images." Comput Methods Programs Biomed **44**(2): 131-40.
- West, J., Fitzpatrick, J. M., et al. (1999). "Retrospective intermodality registration techniques for images of the head: surface-based versus volume-based." IEEE Trans Med Imaging **18**(2): 144-50.
- West, J., Fitzpatrick, J. M., et al. (1997). "Comparison and evaluation of retrospective intermodality brain image registration techniques." J Comput Assist Tomogr **21**(4): 554-66.
- West, J. B., Fitzpatrick, J. M., et al. (2001). "Fiducial point placement and the accuracy of point-based, rigid body registration." Neurosurgery **48**(4): 810-6; discussion 816-7.
- Williamson, J. F. (1987). "Random selection of points distributed on curved surfaces." Phy Med Biol **32**: 1311-19.
- Winkler, P. A., Vollmar, C., et al. (2000). "Usefulness of 3-D reconstructed images of the human cerebral cortex for localization of subdural electrodes in epilepsy surgery." Epilepsy Res **41**(2): 169-78.
- Wong, J. C. H., Studholme, C., et al. (1997). "Evaluation of the limits of visual detection of image misregistration in a brain fluorine-18 fluorodeoxyglucose PET-MRI study." Eur J Nucl Med **24**(6): 642-50.
- Woods, R. P., Grafton, S. T., et al. (1998). "Automated image registration: I. General methods and intrasubject, intramodality validation." J Comput Assist Tomogr **22**(1): 139-52.

- Woods, R. P., Mazziotta, J. C., et al. (1993). "MRI-PET registration with automated algorithm." J Comput Assist Tomogr **17**(4): 536-46.
- Zhang, W., Simos, P. G., et al. (2003). "Multimodality neuroimaging evaluation improves the detection of subtle cortical dysplasia in seizure patients." Neurol Res **25**(1): 53-7.
- Zhilkin, P. und Alexander, M. E. (2004). "Affine registration: a comparison of several programs." Magn Reson Imaging **22**(1): 55-66.
- Zhu, Y. M. (2002). "Volume image registration by cross-entropy optimization." IEEE Trans Med Imaging **21**(2): 174-80.
- Zuk, T. D. und Atkins, M. S. (1996). "A Comparison of Manual and Automatic Methods for Registering Scans of the Head." IEEE Trans Med Imaging **15**: 732-44.

## 7 Danksagung

Mein Dank gilt meiner Mutter, die in mir das Interesse an der Medizin weckte und mir das Studium ermöglichte. Ihr und meinem Bruder Sebastian danke ich für das Verständnis und die Freiräume, die notwendig waren um diese Arbeit zu erstellen.

Herrn Prof Dr. Klaus Hahn danke ich für die Möglichkeit diese Arbeit in seiner Abteilung durchzuführen, die ein harmonisches Miteinander der diagnostischen Disziplinen erleben ließ und so mein Interesse an der bildgebenden Diagnostik weckte. PD Dr. Thomas Pfluger danke ich für die Überlassung des Themas und seine stets erreichbare und freundschaftliche Betreuung. Neben der allgemeinen Einführung in wissenschaftliches Arbeiten verdanke ich ihm viele konstruktive Hilfen bei der Entwicklung der Methodik sowie wertvolle Hinweise und Korrekturen für die Erstellung der Arbeit.

Dr. Axel Wismüller danke ich für die Unterstützung bei der Erstellung der IDL-Programme zur Fehlerberechnung, Herrn Dr. Dirk Heiss für die schnelle Hilfe bei EDV-Problemen. Ihnen beiden sowie Lars Brinkmann, Johannes Behrens und Oliver Lange sei außerdem gedankt für die Hilfe bei der Durchführung der manuellen Bildfusionen und Organsegmentierung für die Interobserver-Testung.

Den MTAs der Kernspintomographie sowie der Nuklearmedizin in der Ziemssenstraße danke ich für die stets freundliche Unterstützung bei technischen und logistischen Problemen.

Allen Mitarbeitern der Radiologie und Nuklearmedizin der Innenstadtkliniken danke ich schließlich für die angenehme Atmosphäre in den Abteilungen, in der ich mich als Doktorand stets wohlfühlte.

## 8 Lebenslauf

### Persönliche Informationen

Geburtsdatum 23.12.1972  
 Geburtsort: München  
 Eltern: Dr. Helga Vollmar, Ärztin  
 Hans-Peter Vollmar, Studienrat

### Ausbildung

1979 – 1983 Grundschule, Kirchheim b. München  
 1983 – 1992 Gymnasium Kirchheim, Kirchheim b. München  
 1992 – 2000 Studium der Humanmedizin an der  
 Ludwig Maximilians Universität München  
 1999 – 2000 Praktisches Jahr am Klinikum der Universität München:  
 1. Terial: Institut für Radiologische Diagnostik, Innenstadt  
 2. Terial: Kliniken für Chirurgie und Neurochirurgie,  
 Großhadern  
 3. Terial: Medizinische Klinik, Ziemssenstraße  
 19. 10. 2000 Bestehen der ärztlichen Prüfung

### Berufstätigkeit

2000 – 2002 Arzt im Praktikum am Institut für Radiologische Diagnostik  
 und in der Klinik für Nuklearmedizin der Universität München,  
 Innenstadt.  
 Assistenzarzt an der Klinik für Nuklearmedizin.  
 17. 07. 2002 Approbation als Arzt  
 ab 01. 08. 2002 Assistenzarzt an der Neurologischen Klinik und Poliklinik der  
 Universität München, Schwerpunkt Epileptologie