

# Ein Verfahren zur objektiven Quantifizierung der Genauigkeit von dreidimensionalen Fusionsalgorithmen

## Ein Optimierungs- und Bewertungswerkzeug

Falk Uhlemann<sup>1</sup>, Ute Morgenstern<sup>2</sup> und Ralf Steinmeier<sup>3</sup>

<sup>1</sup>Lehrstuhl Erkennende Systeme und Bildverarbeitung, Institut für Künstliche Intelligenz, Fakultät Informatik, Technische Universität Dresden, 01307 Dresden

<sup>2</sup>Institut für Biomedizinische Technik, Fakultät Elektrotechnik und Informationstechnik, Technische Universität Dresden, 01062 Dresden

<sup>3</sup>Klinik für Neurochirurgie, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden  
Email: uhlemann@ifwt.et.tu-dresden.de

**Zusammenfassung.** Für die Bewertung von Ergebnissen dreidimensionaler Bildverarbeitungsalgorithmen in der Medizin wurde ein Verfahren für die numerische Ergebnisbewertung entwickelt, welches in diesem Beitrag vorgestellt wird. Sowohl für die Optimierung während der Entwicklung, als auch für den Vergleich verschiedener Algorithmen stellt diese objektive quantitative Beurteilung ein wertvolles Werkzeug dar. Die implementierte Software erlaubt die automatische Generierung und Transformation von simulierten realistischen Datensätzen sowie die systematische Berechnung der entsprechenden Verarbeitungsergebnisse und Gütemaße.

## 1 Einleitung

In den letzten Jahren erfuhr die Entwicklung und der klinische Einsatz von dreidimensionalen Bildverarbeitungsalgorithmen für die Medizin einen enormen Aufschwung.

Ein Anwendungsgebiet dreidimensionaler Bildverarbeitungsalgorithmen ist die Überlagerung multimodaler Daten. Dabei besteht die Aufgabe im Finden von Korrespondenzen in Aufnahmen des gleichen Objektes von verschiedenen Bildgewinnungsverfahren oder zu unterschiedlichen Zeitpunkten.

Für diese Fusionsaufgabe wurden durch zahlreiche Forschergruppen leistungsfähige Verfahren entwickelt, die auch bei vollautomatischer Berechnung bereits befriedigende Ergebnisse liefern. Eine Einschätzung der erreichten Fusionsergebnisse erfolgt dabei meist visuell durch den Entwickler bzw. die klinischen Nutzer.

Für eine erste Abschätzung der Fusionsgüte ist die visuelle Begutachtung meist ausreichend, für eine quantitative, objektive Evaluation der Fusionsgüte oder einen Vergleich mit anderen Verfahren jedoch ungenügend. Deshalb wurde in dieser Arbeit die Möglichkeit einer objektiven Bewertung unter Benutzung simulierter realistischer Daten untersucht und eine Methode entwickelt, die die

**Tabelle 1.** Vergleich von Verfahren zur Erzeugung von Referenzdaten

Nr.	Verfahren	Klinische Relevanz	Validität	Genauigkeit	Kosten/Aufwand
1	Aufnahme realer klinischer Datensätze mit Einschraubmarkern [1]	Hoch	Hoch	Mittel	Hoch/Hoch
2	Aufnahme realer Daten von physikalischen Phantomen mit Markern [2]	Gering	Hoch	Niedrig	Hoch/Mittel
3	Simulation realistischer Daten durch Modellierung des Aufnahmeprozesses für einfache Phantomgeometrien [3]	Gering	Mittel	Hoch	Niedrig/Niedrig
4	Simulation realistischer Daten durch Modellierung des Aufnahmeprozesses und entsprechende Modifikation segmentierter klinischer Daten [4,5,6]	Hoch	Mittel	Hoch	Niedrig/Mittel bis Hoch

automatisierbare Analyse von Fusionsalgorithmen hinsichtlich ihrer Qualität anhand numerischer Gütekennzahlen erlaubt.

## 2 Methode

Gegenwärtig existieren verschiedene Ansätze für die Erzeugung von Referenzdatensätzen, d.h. Aufnahmen bei denen die ideale Überlagerung der korrespondierenden Koordinatensysteme bekannt ist, die im folgenden näher vorgestellt werden sollen.

### 2.1 Methodenvergleich

In 2.3 sind die gegenwärtig gebräuchlichsten Verfahren für die Generierung dreidimensionaler medizinischer Daten mit einigen Vor- und Nachteilen aufgeführt. Dabei bezieht sich die „Klinische Relevanz“ auf die Häufigkeit der abgebildeten Objekte in der klinischen Routine und die Validität auf die Sicherheit der Übereinstimmung der jeweiligen generierten Daten mit realen Aufnahmen.

Bei diesem Vergleich ist erkennbar, dass die Auswahl nur eines Verfahren immer einen Kompromiss zwischen den aufgeführten Kriterien darstellt. Deshalb sollten bei der Entwicklung bzw. dem Vergleich von Algorithmen mehrere Methoden der Bewertung zum Einsatz kommen. Leider ist die Berechnung objektiver quantitativer Gütekennzahlen anhand von Referenzdaten gegenwärtig noch nicht gängige Praxis bei der Entwicklung von Bildverarbeitungsalgorithmen.

### 2.2 Gütemaße und Datenbasis

Neben dem Einsatz von Daten mit bekannten Transformationsparametern (Referenzdaten) gilt es, numerische Ausdrücke zu finden, die eine Aussage über die

Qualität eines Algorithmus' ermöglichen. Dabei sollen diese Gütemaße einerseits einfach interpretierbar sein, andererseits müssen die verschiedensten Einflüsse, die durch die Unterschiede im bearbeiteten Datenmaterial bedingt sind, berücksichtigt werden.

Gegenwärtig werden zur Ergebnisbewertung von Bildverarbeitungsalgorithmen für die Fusion von dreidimensionalen Daten häufig die mittlere Abweichung bzw. die *Figure-of-Merit* (*FOM*) des Abstandes für eine Auswahl von definierten Punkten im Volumen von den Referenzkoordinaten oder die Abweichung der berechneten Transformationsparameter gegenüber den Referenzparametern eingesetzt. Dabei ist zu beachten, dass die Güte der erreichten Ergebnisse in entscheidendem Maße von der Qualität des bearbeiteten Datenmaterials abhängt. Deshalb sind Vergleiche von Ergebnissen bei Benutzung verschiedener Daten meist nicht sehr aussagekräftig.

Das „Retrospective Registration Evaluation Project“ an der *Vanderbilt University – School of Engineering* [1] versucht dieses Problem durch eine öffentlich zugängliche Bilddatenbank zu lösen. Dennoch bleibt bei dieser Datenbank, selbst bei Vernachlässigung der begrenzten Genauigkeit der Referenztransformation (ca. 1 mm), das Problem der eingeschränkten Datenvariabilität. So sind Störgrößen wie Rauschen und Geometrieinhomogenitäten nicht nur vom aufgenommenen Objekt, sondern auch vom Aufnahmengerät abhängig. Wollte man all diese Parameter in angemessener Weise berücksichtigen, so würde die Anzahl von Datensätzen dramatisch steigen.

Deshalb wurde während der Entwicklungsphase von Fusionsalgorithmen auf die Simulation realistischer Datensätze zurückgegriffen [3,4]. Dies erlaubt die freie Parameterwahl für die Transformation und das Hinzufügen von Artefakten.

### 2.3 Modellierung des Bildgewinnungsprozesses

Heutige Bildgewinnungssysteme stellen hoch entwickelte Geräte dar, deren Modellierung ein schwieriges Problem darstellt. Hinzu kommt, dass die physikalischen Vorgänge der Wechselwirkungen in biologischem Gewebe bei der Bilderzeugung (Emission, Absorption, Reflexion, Streuung...) äußerst komplex sind. Für einen ersten Ansatz der softwaremäßigen Simulation erscheint es demnach sinnvoll, geeignete Vereinfachungen vorzunehmen.

Als Ausgangsbasis für die Verfahren 3 und 4 dienen segmentierte und entsprechend dem Gewebetyp mit einer Nummer versehene (gelabelte) anatomische Daten. Anschließend werden sowohl algorithmische (z.B. Berechnung der Schichtdaten aus den Projektionen) als auch physikalische Vorgänge (z.B. Helligkeit und Verteilung der Grauwerte in Abhängigkeit des Labels, Addition von Rauschen...) softwaretechnisch nachgebildet. Dabei werden nur die wichtigsten Schritte und Einflussgrößen berücksichtigt (siehe [3,4,5,6]).

### 2.4 Implementierung des Verfahrens

Während der Fusionsalgorithmenentwicklung und für erste Voruntersuchungen wurde eine Simulationssoftware entsprechend [3] eingesetzt, die eine schnelle

Berechnung einfacher Phantomgeometrien unter Hinzufügen von Störungen ermöglicht. Da für den Vergleich verschiedener Fusionsalgorithmen jedoch die klinische Relevanz der Bilddaten eine entscheidende Rolle spielt, wurde dafür auf die Daten aus [4] zurückgegriffen (aus segmentierten MRT- Daten simulierte PET-Aufnahmen mit und ohne Rauschen). Der Einsatz beliebiger anderer Ausgangsdaten ist jedoch problemlos möglich.

Diese Daten werden im nächsten Schritt wahlweise mit systematisch oder randomisiert variierten Parametern transformiert (Festkörpertransformation). Ein entsprechendes Auswertemodul gestattet das automatische Einlesen, Vorverarbeiten und (Rück-) Transformieren der Datensätze. Dabei ist es möglich, verschiedenste Matching-Funktionen einzubinden und hinsichtlich ihrer Genauigkeit zu untersuchen. Die Ergebnisberechnung erfolgt mit Hilfe von zusätzlich in die jeweiligen DICOM- Datei- Header eingebrachten Informationen über die Referenztransformation. Dieses Werkzeug erlaubt eine automatisierte Analyse von Algorithmen über eine große Anzahl von Datensätzen.

## 2.5 Ein Beispiel für die Anwendung des Verfahrens

Im folgenden wird der Einsatz des beschriebenen Verfahren am Beispiel einer intra- bzw. intermodalen Überlagerung (PET-PET bzw. MRT-PET) von dreidimensionalen Datensätzen veranschaulicht. Dabei wurden die Aufnahmen sowohl mit äquidistanten als auch mit randomisierten Parametern transformiert.

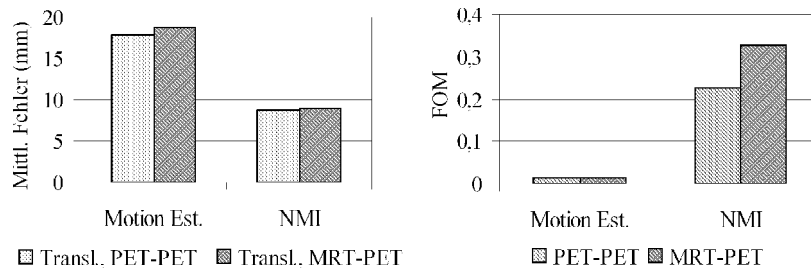
Für die Berechnung der Überlagerung wurden Algorithmen mit verschiedenen Optimierungskriterien für eine Festkörpertransformation mit 6 Freiheitsgraden herangezogen. Diese waren einerseits eine Bewegungsabschätzung (*Motion Estimation*) und andererseits die relative Entropie (*Normalised Mutual Information - NMI*).

Einige Ergebnisse der Untersuchung sind in Abb. 1 gezeigt. Dafür wurden die simulierten PET- Datensätze um bis zu 40 mm verschoben bzw. bis zu 45 Grad rotiert. Die relativ hohen Abweichungen gegenüber der idealen Überlagerung (bei Translation von MRT- PET bis zu 18 mm) sind hauptsächlich auf das, im Sinne einer *worst- case-* Abschätzung starke, hinzugefügte Rauschen in den PET- Aufnahmen zurückzuführen. Insgesamt konnte eindeutig festgestellt werden, dass die *NMI* genauer und robuster als der eingesetzte *Motion Estimation* Algorithmus ist.

## 3 Ergebnisse

Mit dem dargestellten Verfahren konnten die Algorithmen hinsichtlich verschiedenster Transformationsaufgaben untersucht und die Ergebnisse genau quantifiziert werden.

Die Analyse von Bildverarbeitungsalgorithmen durch eine numerische Bewertung erlaubt eine gezielte Optimierung von Verfahren während der Entwicklung, sowie einen objektiven Vergleich von Bildverarbeitungsalgorithmen. Durch entsprechende Erweiterungen dieses Ansatzes ist eine Schaffung von Standards für

**Abb. 1.** Gütemaße für verschiedene Fusionsalgorithmen.

die Validierung bzw. die Qualitätssicherung von Methoden in der Bildverarbeitung möglich.

#### 4 Ausblick

Künftige Untersuchungen werden weitere Algorithmen, Datensätze und Gütemaßzahlen einbeziehen, um ein universelles Werkzeug für die Entwicklung, Optimierung und den Vergleich von Bildverarbeitungsalgorithmen bereitstellen zu können. Diese Arbeit stieß bereits auf internationales Interesse, so dass eine Fortführung der Untersuchungen in einer internationalen Arbeitsgruppe geplant ist.

#### Literaturverzeichnis

1. West J, Fitzpatrick JM, Wang MY et al.: Comparison and Evaluation of Retrospective Intermodality Brain Image Registration Techniques. SPIE Medim '96 conference, Newport Beach, CA, February 10-15, 1996. vol. proc. SPIE 2710:332-347, 1996
2. Sobottka S, Uhlemann F, Beuthin- Baumann B, et al.: Genauigkeit der digitalen Bildfusion von PET und MRT/CT eines automatischen Matching- Algorithmus. Duffner, F.; Grote, E.H., 1. Jahrestagung der Sektion Neurochirurgie und Neuronavigation der Deutschen Gesellschaft für Neurochirurgie 05. und 06. Oktober 2001, Universitätsklinikum Tübingen - Klinik für Neurochirurgie:60, 2001
3. Uhlemann F: Segmentierung, Volumenbestimmung und Visualisierung medizinischer Daten verschiedener Modalitäten. Institut für Biomedizinische Technik, Fakultät Elektrotechnik, Technische Universität Dresden, Diplomarbeit, 2000.
4. Chodkowski B: A Simulation of Clinically Realistic PET Data and its Use in the Evaluation of Image Registration Algorithms. Master of Science, School of Engineering and Applied Science of George Washington University, 1996.
5. Kwan RKS, Evans AC, Pike GB: An Extensible MRI Simulator for Post-Processing Evaluation, Visualization in Biomedical Computing (VBC'96), Lecture Notes in Computer Science, vol. 1131, Springer,:135-140, 1996.
6. Grova C, Biraben A, Scarabin JM, et al.: A methodology to validate MRI/ SPECT registration methods using realistic simulated SPECT data. In W.J. Niessen and M.A. Viergever, editors, MICCAI 2001, Utrecht (The Netherlands), Lecture Notes in Computer Science, volume 2208:275-282, Springer, 2001.