

# Optimierte endoskopische Time-of-Flight Oberflächenrekonstruktion durch Integration eines Struktur-durch-Bewegung Ansatzes

A. Groch<sup>1</sup>, S. Haase<sup>2</sup>, M. Wagner<sup>3</sup>, T. Kilgus<sup>1</sup>, H. Kenngott<sup>3</sup>,  
H.-P. Schlemmer<sup>4</sup>, J. Hornegger<sup>2</sup>, H.-P. Meinzer<sup>1</sup>, L. Maier-Hein<sup>1</sup>

<sup>1</sup>Abt. Medizinische und Biologische Informatik, Deutsches Krebsforschungszentrum  
(DKFZ)

<sup>2</sup>Lehrstuhl für Mustererkennung, Friedrich-Alexander-Universität Erlangen-Nürnberg

<sup>3</sup>Klinik für Allgemein-, Viszeral- und Transplantationschirurgie, Universitätsklinikum  
Heidelberg

<sup>4</sup>Abt. Radiologie, Deutsches Krebsforschungszentrum (DKFZ)  
[a.groch@dkfz-heidelberg.de](mailto:a.groch@dkfz-heidelberg.de)

**Kurzfassung.** Eine der größten Herausforderungen im Kontext von computergestützten Assistenzsystemen für laparoskopische Eingriffe stellt die intraoperative akkurate und schnelle Rekonstruktion der Organoberfläche dar. Während Rekonstruktionstechniken basierend auf *Multiple View Methoden*, beispielsweise *Stereo*-Rekonstruktion, schon länger Gegenstand der Forschung sind, wurde erst kürzlich das weltweit erste *Time-of-Flight (ToF)* Endoskop vorgestellt. Die Vorteile gegenüber Stereo liegen in der hohen Aktualisierungsrate und dem dichten Tiefenbild unabhängig von der betrachteten Szene. Demgegenüber stehen allerdings Nachteile wie schlechte Genauigkeit bedingt durch hohes Rauschen und systematische Fehler. Um die Vorteile beider Verfahren zu vereinen, wurde in diesem Beitrag<sup>1</sup> ein Konzept entwickelt, welches die ToF-Endoskopie-Technik mit einem stereoähnlichen Multiple-View-Ansatz (*Struktur durch Bewegung*) fusioniert. Der Ansatz benötigt keine zusätzliche Bildgebungsmodalität wie z.B. ein Stereoskop, sondern nutzt die ohnehin akquirierten (Mono-) Farbdaten des ToF-Endoskops. Erste Ergebnisse zeigen, dass die Genauigkeit der Oberflächenrekonstruktion mit diesem Ansatz verbessert werden kann.

## 1 Einleitung

Während der letzten Jahre erfahren computergestützte Assistenzsysteme für laparoskopische Eingriffe zunehmend Aufmerksamkeit in der Wissenschaft. Solche Assistenzsysteme arbeiten in der Regel mit einem präoperativ erstellten Patientenmodell, welches während der Operation auf die aktuell vorliegende Patientenanatomie übertragen wird [1]. Endoskopische Oberflächenrekonstruktion ist

---

<sup>1</sup> Die vorliegende Arbeit wurde im Rahmen des von der Deutschen Forschungsgemeinschaft unterstützten Graduiertenkollegs 1126: "Intelligente Chirurgie" durchgeführt.

in diesem Kontext eine Schlüsseltechnik, da sie markerlos und ohne zusätzliche Bildgebungsmodalität und eventuelle Strahlenbelastung die Registrierung der präoperativ akquirierten Daten in Echtzeit ermöglicht. Häufig angewandte Verfahren zur Oberflächenrekonstruktion basieren auf sogenannten *Multiple View Methoden*. Dabei wird eine Korrespondenzanalyse auf zwei oder mehreren Bildern aus verschiedenen Kamerapositionen durchgeführt, was mittels Triangulation eine 3D-Rekonstruktion der abgebildeten Szene erlaubt. Werden zwei fest zueinander definierte Kameras benutzt, spricht man von *Stereo*-Rekonstruktion. Im Kontext der Endoskopie wird dies mit Stereoskopen als Bildgebungsmodalität realisiert [2]. Bei der Korrespondenzanalyse auf mehr als zwei Bildern werden die Bilder häufig sequentiell über die Bewegung der Kamera akquiriert (*Struktur-durch-Bewegung*, engl. *Structure-from-Motion (SfM)*) [3]. Im endoskopischen Kontext werden hierfür gewöhnliche (Mono-) Endoskope benutzt. Andere Ansätze zur Oberflächenrekonstruktion nutzen z.B. *Strukturiertes Licht* [4], welches auf einer aktiven Triangulation beruht.

Erst kürzlich wurde das erste Endoskop, das auf der neuen *Time-of-Flight (ToF)* Technik basiert, vorgestellt [5]. Das Prinzip von ToF-Sensoren besteht darin, dass die Flugdauer eines intensitätsmodulierten Lichtsignals gemessen wird. Wegen der Möglichkeit, dichte Tiefendaten mit einer videoähnlichen Aktualisierungsrate unabhängig von der betrachteten Szene zu generieren, gewinnt die ToF-Technik immer mehr Aufmerksamkeit als Alternative zu konventionellen Distanzmessungen. Nicht zu vernachlässigende Nachteile der ToF-Endoskopie sind dagegen eine schlechte Genauigkeit der rekonstruierten Oberflächen. Dies ist vor allem auf den starken Signalverlust und das damit einhergehende niedrige Signal-zu-Rausch-Verhältnis zurückzuführen, aber auch auf die fehlenden Forschungserfahrungen im Bereich der Kalibrierung von systematischen Fehlern, sowie die endoskopischen Anforderungen an die Genauigkeit im Submillimeterbereich.

Eine Möglichkeit, die niedrige Genauigkeit der rekonstruierten Oberfläche zu kompensieren, besteht darin, weitere Oberflächenrekonstruktionsverfahren in den Rekonstruktionsprozess einzubinden. Es liegt nahe, dies mit einem Structure-from-Motion Ansatz zu realisieren, da in der ToF-Endoskopie normale (Mono-) Endoskop-Farbbilder ohnehin zusätzlich zu den ToF-Tiefendaten akquiriert werden. Desweiteren sind die Eigenschaften von ToF und Structure-from-Motion komplementär. Während ToF ein dichtes, aber ungenaues Tiefenbild erzeugt, ist Structure-from-Motion an wenigen stabilen Merkmalen sehr genau. Zusätzlich erzielt ToF durch intensitätsbasierte Fehler die besten Ergebnisse bei homogenen Objekten, Structure-from-Motion wegen der Korrespondenzsuche hingegen bei texturierten Objekten. Forschungsarbeiten zur Fusion von ToF mit Multiple-View-Methoden gibt es bereits für herkömmliche, nicht-endoskopische Kameras, hauptsächlich für die Fusion von ToF- mit Stereodaten [6][7]. Die meisten Arbeiten basieren auf probabilistischen Ansätzen, in denen die unterschiedlichen Fehlerwahrscheinlichkeiten der Bildgebungsmodalitäten modelliert und dann fusioniert werden.

In diesem Beitrag stellen wir ein Konzept zur Verbesserung von ToF-Oberflächen durch Integration eines Structure-from-Motion Ansatzes im Kontext der Endoskopie vor.

## 2 Material und Methoden

In diesem Abschnitt werden die benutzte Hardware (2.1), das Fusionskonzept von ToF und Structure-from-Motion (2.2) und die Evaluation des Konzepts (2.3) näher erläutert.

### 2.1 Hardware

In dieser Arbeit wurde ein erster ToF-Endoskop-Prototyp der Firma Richard Wolf GmbH<sup>2</sup> benutzt. Er erzeugt ca 30 Distanzbilder pro Sekunde sowie dazu synchronisierte gewöhnliche endoskopische Farbdaten (Standard Definition (SD)). Die Aufnahme der Farbdaten durch dieselbe Optik ist eine Erweiterung zu dem von Penne vorgestellten ToF-Endoskop [5]. Für die vorliegende Arbeit wurde das Endoskop optisch getrackt, um seine Lage bestimmen zu können. Die intrinsischen Parameter der Farb- und ToF-Kamera wurden mit gängigen Kalibrierungsroutinen bestimmt, die Transformation des Trackingtools zu den Kameras mit einer gewöhnlichen Hand-Auge-Kalibrierung und die Transformation von den Kameras zueinander (ToF- zu Farbkamera) durch eine Stereo-Kalibrierung.

### 2.2 Fusionsansatz

In jedem Zeitschritt soll mit Hilfe der hier vorgestellten Fusionsmethode eine Oberfläche rekonstruiert werden. Dazu dienen als Input in jedem Zeitschritt (1) eine aus den ToF-Tiefendaten, unter Nutzung der intrinsischen Kameraparameter generierte, dichte Oberfläche sowie (2) eine 3D-Punktwolke aus dem Structure-from-Motion Ansatz. Die 3D-Punktwolke wird auf Basis der vorangegangenen Bildersequenz berechnet. Der Ansatz berücksichtigt nur die „besten“ Merkmale in den Farbbildern. Gute Merkmale zeichnen sich dadurch aus, dass sie in jedem Bild der Sequenz eindeutig detektiert und ihren korrespondierenden Merkmalen in allen Vorgängerbildern eindeutig zugeordnet werden können. Von diesen guten Merkmalen werden nur diejenigen  $n$  mit der kleinsten Deskriptordifferenz zu ihren korrespondierenden Vorgänger-Merkmalen berücksichtigt. Die Deskriptordifferenz ist ein Maß dafür, wie ähnlich sich die Merkmale sind.

Da für die ToF-Endoskopie noch keine etablierte Distanzkalibrierung existiert und die ToF-Distanzdaten deswegen mit einem Offset in Richtung des Sehstrahls vorliegen können, wird eine rigide Vorregistrierung mit Hilfe des kürzlich vorgestellten *anisotropen ICP* [8], eine Variante des *Iterative Closest Point (ICP)*

<sup>2</sup> Vielen Dank an die Firma Richard Wolf GmbH für die Leihgabe des ToF-Endoskop-Prototypen.

Algorithmus, durchgeführt. Dieser eignet sich besonders gut für rigide Registrierungen von ToF-Oberflächen, da anisotrope Fehlerwahrscheinlichkeitsverteilungen, wie sie bei ToF vorliegen, berücksichtigt werden können.

Ein probabilistischer Ansatz, ähnlich wie unter anderem in [6] für die Fusion mit Stereo beschrieben, wird nun für die weitere Fehlerminimierung der Oberflächen realisiert, indem ToF-Endoskop- und Structure-from-Motion-Daten fusioniert werden. Hierfür wird wie z.B. in [9] ein dreidimensionales *Belegungsnetz* (engl. *occupancy grid*) aufgestellt, welches das Volumen der zu fusionierenden Oberfläche bzw. Punktwolke umschließt und mit gitterförmig angeordneten Knoten hoher Auflösung (höher als die ToF-Oberfläche) gefüllt ist. Innerhalb dieses quaderförmigen Netzes liegt die gesuchte Oberfläche und setzt sich aus einer Teilmenge der Knoten des Belegungsnetzes zusammen. Jeder Knoten besitzt einen Wert abhängig davon, wie hoch die Wahrscheinlichkeit ist, dass genau dieser Knoten zur gesuchten Oberfläche gehört.

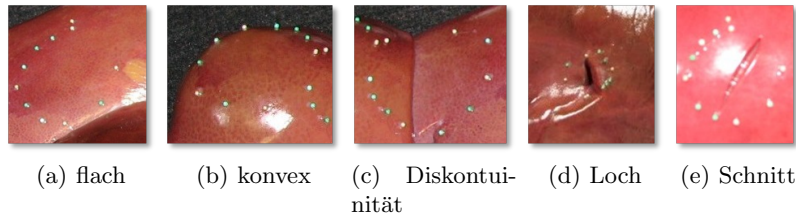
Die Wahrscheinlichkeit jedes Knotens  $v$  setzt sich zusammen aus  $p(v|O_{\text{ToF}}, \Sigma_{\text{ToF}}, O_{\text{SfM}}, \Sigma_{\text{SfM}})$ .  $O_{\text{ToF}}$  ist die Menge aller Knoten des Belegungsnetzes, welche zur ToF-Oberfläche gehören, und  $O_{\text{SfM}}$  die Menge derjenigen, die zur Structure-from-Motion Punktwolke gehören.  $\Sigma_{\text{ToF}}$  bzw.  $\Sigma_{\text{SfM}}$  beschreiben die Fehlerwahrscheinlichkeitsverteilung in alle drei Dimensionen für jeden Knoten  $o_{\text{ToF}} \in O_{\text{ToF}}$  bzw.  $o_{\text{SfM}} \in O_{\text{SfM}}$ . In der jetzigen Implementierung sind die Unsicherheitsverteilungen so gewählt, dass sie für ToF wesentlich höher als für SfM sind und generell in Sehstrahl-Richtung höher als in die Richtungen orthogonal dazu (anisotrope Gaussverteilung mit Erwartungswert 0). Der Ansatz läßt sich leicht erweitern, so dass auch die einzelnen Knoten unterschiedliche Verteilungen erhalten können (s. Abschnitt 4).

Nach Aufstellung des Belegungsnetzes wird eine Oberfläche in dem dreidimensionalen Netz gesucht, die (1) sowohl durch die Knoten mit möglichst hohen Wahrscheinlichkeiten geht, als auch (2) möglichst glatt ist. Außerdem muss berücksichtigt werden, dass es (3) nur jeweils einen Oberflächenpunkt in Richtung des Sehstrahls geben kann. Dies wird mit einem *Graph-Cut* basierten Verfahren [10] umgesetzt, welches mit dem *Min-Cut-Max-Flow*-Theorem einen Schnitt mit minimalen Kosten (= maximaler Wahrscheinlichkeit) bei maximaler Glätte sucht. Dieses Verfahren gewährleistet zusätzlich, dass der Schnitt das Netz senkrecht zur Sehstrahlrichtung in zwei Teile zerschneidet und somit jeweils genau ein Oberflächenknoten in Sehstrahlrichtung existiert.

### 2.3 Machbarkeitsstudie

Um die Machbarkeit des vorgestellten Fusionsansatzes zu überprüfen, wurde eine erste Evaluation an *in-vitro* an Schweineorganen durchgeführt. Hierfür wurden verschiedene Organformen (s. Abb. 1) und verschiedene Organtexturen aufgenommen. Die Genauigkeit der Oberfläche aus dem Fusionsansatz wurde mit der nur aus den ToF-Endoskopdaten erstellten Oberfläche verglichen. Als Goldstandard dienten CT Aufnahmen aller Objekte. Über Marker, die außerhalb der zu evaluierenden Oberfläche angebracht waren, wurden die rekonstruierten Oberflächen und die Goldstandard-Daten zueinander registriert. Damit konnte die mitt-

lere Distanz aller Knoten der rekonstruierten Oberfläche zu den Goldstandard-Oberflächen berechnet werden. Außerdem wurde verglichen, wie sich die Genauigkeit abhängig von der Anzahl der benutzten Structure-from-Motion Punkte verbessert.



**Abb. 1.** Verschiedene Organformen

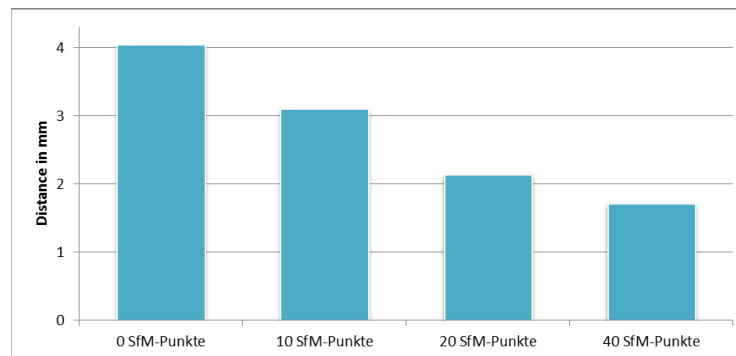
### 3 Ergebnisse

Um einen ersten quantitativen Eindruck des Fusionskonzept vor der Gesamtimplementierung zu erhalten, konzentrierte man sich bei dieser Evaluierung auf den probabilistischen Ansatz durch das Belegungsnetz. Dazu wurden ToF-Oberfläche und SfM-Punktwolke mit Hilfe der angebrachten Marker auf die CT-Referenz-Oberfläche registriert und dann fusioniert. Dies verschafft einen Eindruck, wie sehr die relative Genauigkeit der Oberfläche erhöht werden kann, nachdem die Vorpositionierung schon stattgefunden hat.

Ergebnisse an endoskopischen Bilddaten von Leber und Lunge mit verschiedenen Oberflächenformen und -texturen zeigen, dass sich die mittlere Genauigkeit aller akquirierten Organoberflächen bei einer Berücksichtigung schon von ca. 20 Structure-from-Motion Punkte verdoppelt (s. Abb. 2).

### 4 Diskussion

In dieser Arbeit haben wir ein Konzept zur Fusion von endoskopischer, ToF-basierter Oberflächenrekonstruktion mit einem Structure-from-Motion Ansatz vorgestellt. Dieses kombiniert die Vorteile beider Verfahren, so dass nach ersten Ergebnissen Fehler der noch sehr neuen ToF-Endoskopie reduziert werden. Es ist noch abzuwarten, wie gut die Ergebnisse nach Einsatz des A-ICP sein werden. Außerdem sollen bei der Aufstellung der Unsicherheiten im Belegungsnetz weitere Annahmen getroffen werden, wie z.B. dass Knoten, die aus ToF-Pixeln mit hoher Intensität entstanden sind, eine höhere Wahrscheinlichkeit zugeordnet werden, da helle Objektpixel bei ToF weniger rauschbehaftet sind als dunkle. Ähnliches kann für Structure-from-Motion angenommen werden. Hier sind die Knoten aus Merkmalen mit kleiner Deskriptordifferenz zuverlässiger und könnten deswegen eine höhere Wahrscheinlichkeit erhalten. Trotz des im Moment noch einfachen Ansatzes konnte in diesem Beitrag gezeigt werden, dass eine Fusionierung von ToF mit Structure from Motion die Genauigkeit der Oberflächenrekonstruktion erhöht.



**Abb. 2.** Mittlerer Fehler aller rekonstruierten Knoten aller akquirierten Oberflächen zur Goldstandard-Oberfläche abhängig von der Anzahl der benutzten Structure-from-Motion (SfM) Punkte. Null SfM-Punkte entspricht der nur ToF-basierten Rekonstruktion.

## Literaturverzeichnis

1. Baumhauer M, et al. Navigation in endoscopic soft tissue surgery: perspectives and limitations. *J Endourol.* 2008 Apr;22(4):751–766.
2. Stoyanov D, Scarzanella MV, Pratt P, Yang GZ. Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: MICCAI. Berlin, Heidelberg: Springer-Verlag; 2010. p. 275–282.
3. Wengert C, et al. Markerless endoscopic registration and referencing. In: MICCAI. LNCS 4190; 2006. p. 816–823.
4. Clancy NT, Stoyanov D, Yang GZ, Elson DS. An endoscopic structured lighting probe using spectral encoding. vol. 8090; 2011. (in press).
5. Penne J, et al. Time-of-Flight 3-D Endoscopy. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI. LNCS 5761; 2009. p. 467–474.
6. Zhu J, Wang L, Yang R, Davis JE, Pan Z. Reliability Fusion of Time-of-Flight Depth and Stereo Geometry for High Quality Depth Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2011;33:1400–1414.
7. Mutto CD, Zanuttigh P, Cortelazzo GM. A Probabilistic Approach to ToF and Stereo Data Fusion. In: 3DPVT. Paris, France; 2010. .
8. Maier-Hein L, Franz AM, dos Santos TR, Schmidt M, Meinzer HP, Fitzpatrick JM. Convergent Iterative Closest Point Algorithm to Account for Anisotropic Inhomogenous Localization Error. *IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI).* 2012 (in press);.
9. Guan L, Franco JS, Pollefeys M. 3D Object Reconstruction with Heterogeneous Sensor Data. In: International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT); 2008. .
10. Li K, Wu X, Chen DZ, Sonka M. Optimal surface segmentation in volumetric images – a graph-theoretic approach. *IEEE Trans Pattern Anal Machine Intell.* 2006;28(1):119–134.