

# Modellbasierte Echtzeit-Bewegungsschätzung in der Fluoreszenzendoskopie

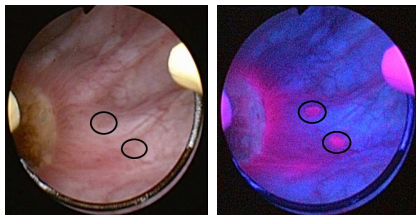
Thomas Stehle, Jonas Wulff, Alexander Behrens, Sebastian Gross, Til Aach

Lehrstuhl für Bildverarbeitung, RWTH Aachen University, 52056 Aachen  
thomas.stehle@lfb.rwth-aachen.de

**Kurzfassung.** In diesem Beitrag wird eine Methode zur echtzeitfähigen Block-Matching-basierten Bewegungsschätzung beschrieben, die auch in der Präsenz von starkem Rauschen zuverlässige Ergebnisse liefert. Zu diesem Zweck werden zunächst Blöcke kreisförmig im Bild verteilt. Mit Hilfe des Struktur-Tensors wird entschieden, ob ein Block genügend Strukturinformation aufweist, um zur Bewegungsschätzung herangezogen zu werden. Die ausgewählten Blöcke werden verwendet, um die Parameter eines globalen Bewegungsmodells unter Verwendung des RANSAC-Algorithmus robust zu bestimmen.

## 1 Einleitung

In Deutschland werden pro Jahr laut Robert Koch Institut etwa 28.750 neue Fälle von Blasenkrebs diagnostiziert. Diese Krebsart ist mit normaler Weißlicht-Endoskopie nur schwer zu festzustellen. Aus diesem Grund wird häufig die photodynamische Diagnose (PDD) angewandt, bei der dem Patienten eine Marker-substanz verabreicht wird, die sich vorwiegend an bösartigem Gewebe anlagert. Bei Verwendung einer schmalbandigen blauen Lichtquelle beginnen diese Areale zu fluoreszieren und der Tumor wird deutlich sichtbar. Ein Beispiel für Weißlicht- und PDD-Endoskopie wird in Abb. 1 gegeben. Die Tumore sind in beiden Fällen mit schwarzen Kreisen markiert. Unter Weißlicht sind die Tumore kaum zu erkennen, während sie sich bei der PDD deutlich von der gesunden Blasenwand unterscheiden.



**Abb. 1.** Blasentumore unter Weißlicht- (links) und PDD-Endoskopie (rechts). Unter Weißlicht sind die Tumore (schwarze Markierung) schwer zu erkennen, während sie in der PDD deutlich sichtbar sind.

Durch die schmalbandige Beleuchtung und die lichtschwache Fluoreszenz trifft nur eine geringe Lichtmenge auf dem Videosensor des Endoskops auf. Aus

diesem Grund ist in solchen Videosequenzen meist ein starkes Rauschen präsent. Stehle et al. haben in [1] ein Verfahren vorgestellt, das eine echtzeitfähige Rauschreduktion in der PDD-Endoskopie ermöglicht. Dieses Verfahren setzt allerdings ein vorgegebenes Bewegungsvektorfeld voraus. Bewegungsschätzer, die zur Videokodierung eingesetzt werden, erzeugen nicht notwendigerweise ein glattes Bewegungsvektorfeld. Die Verwendung solcher Felder maximieren zwar die Prädiktionsleistung des Videocoders, bei der zeitlichen Filterung führen derartige Felder jedoch zu störenden Artefakten. Behrens hat in [2] ein Verfahren vorgestellt, das zur robusten modellbasierten Bewegungsschätzung zwischen zwei PDD-Bildern geeignet ist. Das Verfahren verwendet allerdings nur eine zeitlich unterabgetastete Folge der Eingangsbilder und ist trotzdem nicht echtzeitfähig.

Im Folgenden wird ein Verfahren zur modellbasierten Bewegungsschätzung vorgestellt. Das Verfahren ist geeignet, die benötigten Bewegungsvektoren in Echtzeit zu berechnen, so dass eine zeitlich rekursive Rauschfilterung in Echtzeit ermöglicht wird.

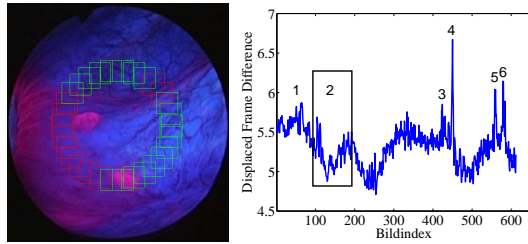
## 2 Material und Methoden

Das Videomaterial des Endoskopiesystems weist eine Auflösung von  $768 \times 576$  Bildpunkten auf. Jedes Pixel besteht aus den drei Farbwerten rot, grün und blau und das System hat eine Aktualisierungsrate von 50 Hz. Das bedeutet, dass für eine Bewegungsschätzung (gefolgt von einer bewegungskompensierten zeitlich rekursiven Filterung) maximal 20 ms an Rechenzeit pro Bild zur Verfügung stehen, wenn eine echtzeitfähige Verarbeitung erwünscht ist.

Es wurde ein Verfahren entworfen, das ein globales Bewegungsmodell mit Hilfe von Punktkorrespondenzen bestimmt. Diese werden mit Hilfe eines Block-Matching-Verfahrens ermittelt. Die Verwendung eines solchen globalen Bewegungsmodells wird dadurch gerechtfertigt, dass sich das Endoskop meist sehr nah an der Blasenwand befindet, so dass diese in sehr guter Näherung als Ebene beschrieben werden kann.

### 2.1 Bestimmung der Punktkorrespondenzen

Eine Anzahl von Blöcken wird kreisförmig im Bild angeordnet und die entsprechenden Regionen werden im vorangegangenen Bild gesucht. Die Zuordnung der Mittelpunkte dieser Regionen sind die gesuchten Punktkorrespondenzen  $[(x_i, y_i), (x'_i, y'_i)]$ . Als Ähnlichkeitsmaß wird die Summe der absoluten Differenzen verwendet, da sich diese mit Hilfe der Befehlssätze heutiger Prozessoren sehr effizient berechnen lässt. Da Block-Matching mit Blockgrößen von  $8 \times 8$  oder  $16 \times 16$  Pixeln, wie sie in der Videokodierung angewandt werden, aufgrund des starken Rauschens kein robustes Ergebnis liefern, mussten in diesem Fall wesentlich größere Blöcke gewählt werden. Für diese Anwendung haben sich Blöcke der Größe  $40 \times 40$  als stabil erwiesen. Auf der linken Seite von Abb. 2 ist das Block-Matching-Setup zu sehen.



**Abb. 2.** Links: Anordnung der Such-Blöcke im Endoskopiebild. Die roten Blöcke wurden mit Hilfe des Struktur-Tensors von der Bewegungsschätzung ausgeschlossen. Rechts: Darstellung der DFD für jedes Bild der Testsequenz. Signifikante Ereignisse wurden mit Ziffern gekennzeichnet.

Um die Zuverlässigkeit potentieller Korrespondenzen abzuschätzen, wird vor dem eigentlichen Block-Matching der Struktur-Tensor [3] jedes Blocks bestimmt. Die Eigenwerte des Struktur-Tensors geben ein Maß dafür an, wie viel Struktur in einem Block enthalten ist. Zwei kleine Eigenwerte deuten auf einen strukturlosen Block hin, der ggf. von Rauschen dominiert wird. Ein großer und ein kleiner Eigenwert lassen auf eine einfach gerichtete Struktur wie einen geraden Kantenzug schließen. Auch in diesem Fall ist aufgrund des Aperturproblems keine eindeutige Bewegungsschätzung möglich. Erst Blöcke mit zwei betragsmäßig großen Eigenwerten weisen auf Strukturen wie beispielsweise Ecken hin, die eine zuverlässige Bestimmung der Verschiebung erlauben.

## 2.2 Globale Bewegungsschätzung

Das globale Bewegungsmodell weist vier Freiheitsgrade auf: Translation in  $x$ - und  $y$ -Richtung ( $t_x$  und  $t_y$ ), Rotation um den Winkel  $\alpha$  und Skalierung um den Faktor  $s$ . Mit Hilfe dieser Freiheitsgrade kann ein Großteil der in der PDD-Endoskopie auftretenden Bildbewegungen modelliert werden. Durch die hohe zeitliche Abtastung ist der Einfluss anderer Bewegungen wie z. B. Scherung oder perspektivische Verzerrung zwischen zwei zeitlich benachbarten Bildern gering. Die entsprechende Abbildungsgleichung lautet

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} s \cdot \cos(\alpha) & -s \cdot \sin(\alpha) & t_x \\ s \cdot \sin(\alpha) & s \cdot \cos(\alpha) & t_y \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (1)$$

Mit Hilfe der Punktkorrespondenzen lässt sich nun folgendes lineares Gleichungssystem aufstellen, das eine Least-Squares-Lösung der Parameter des Bewegungsmodells liefert.

$$\begin{pmatrix} x'_0 \\ y'_0 \\ \vdots \\ x'_n \\ y'_n \end{pmatrix} = \begin{pmatrix} x_0 & y_0 & 1 & 0 \\ y_0 & -x_0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 \\ y_n & -x_n & 0 & 1 \end{pmatrix} \begin{pmatrix} s \cdot \sin(\alpha) \\ s \cdot \cos(\alpha) \\ t_x \\ t_y \end{pmatrix} \quad (2)$$

Da sich trotz der Selektion der Blöcke noch Ausreißer in den Punktkorrespondenzen befinden können, wird der RANSAC-Algorithmus [4] zur robusteren Least-Squares-Schätzung der Modellparameter angewandt.

### 2.3 Experimente und Ergebnisse

Die Auswertung des vorgestellten Verfahrens wurde auf einer Endoskopiesequenz durchgeführt, bei der ein Arzt das Endoskop vom Ostium zur linken Blasenwand schwenkt. In dieser Sequenz sind neben Translation, Rotation und Skalierung aufgrund der Freihandbewegung ebenfalls perspektivische Bewegungskomponenten enthalten. Während der kompletten Sequenz sind bewegten sich kleine losgelöste Gewebepartikel durch das Bild, die nicht dem globalen Bewegungsmodell folgen und somit die Bewegungsschätzung erschweren.

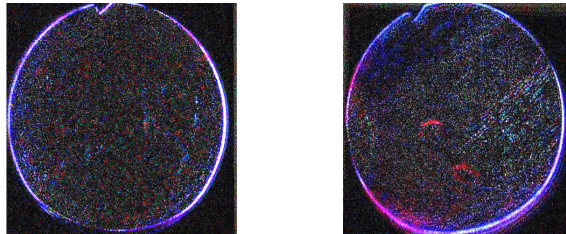
Als Bewertungsmetrik wurden die absoluten Differenzen zwischen einem zentralen Bereich des aktuellen Bilds und des entsprechend Gl. (1) bewegungskompensierten Bilds des vorherigen Zeitpunkts berechnet (displaced frame difference, DFD). Die DFD wird auf die Größe des zentralen Bereichs und auf die Anzahl der Farbkanäle normiert. Die Verwendung dieser Fehlerfunktion ist durch die zeitlich rekursive Filterung motiviert, für die das Verfahren entwickelt wurde. Ein niedriger Wert spricht für eine gute Bewegungskompensation und führt später zu einer starken Filterwirkung.

Der Verlauf der DFD während der Sequenz ist auf der rechten Seite von Abb. 2 zu sehen. Während der kompletten Sequenz findet eine kontinuierliche Bewegung statt. Verschiedene Ereignisse wurden dabei mit Ziffern gekennzeichnet. An der Stelle '1' wurde das Endoskop rotiert und eine Bewegungsunschärfe tritt auf. Das Verfahren erkennt die Bewegung korrekt und die DFD steigt nicht. Im Bereich '2' ist rechts im Bild eine Falte in der Blasenwand zu sehen, so dass die Hypothese einer einzigen globalen Bewegung verletzt ist. Die Inkonsistenz wird vom RANSAC korrekt erkannt und ausgeschlossen. Die DFD steigt hier nicht, da der Randbereich, in dem die Falte auftaucht, nicht zur DFD-Berechnung verwendet wird. Bei '3' findet eine ruckartige Rotation mit Bewegungsunschärfe statt, die das Verfahren nicht komplett erfasst. An den Stellen '4'-'6' treten sehr schnelle Zooms auf, die zu relativ großen Fehlern führen. Beispiele der DFD für eine erfolgreiche und eine ungenaue Bewegungsschätzung finden sich in Abb. 3. Das Verfahren benötigt auf einem 4 Kern Intel XEON Rechner mit 2,3 GHz ca. 18 ms zur Bewegungsschätzung, so dass für eine nachfolgende Filterung nur noch 2 ms zur Verfügung stehen. Da die Filterung nochmals etwa 12-15 ms benötigt, wird so keine Echtzeitfähigkeit erreicht. Durch die Anwendung eines Greedy-Optimierungsverfahrens [5] lässt sich die Rechenzeit auf 4 ms reduzieren, so dass eine Echtzeitverarbeitung ermöglicht wird. Die DFD nimmt dadurch im Mittel um etwa 1,2% zu.

## 3 Diskussion

Es wurde ein Verfahren zur echtzeitfähigen, modellbasierten Bewegungsschätzung in PDD-Endoskopie vorgestellt. Entsprechend der Anforderungen an das

**Abb. 3.** Links: DFD bei erfolgreicher Bewegungsschätzung. Bis auf die Randbereiche des Sichtfelds besteht die DFD vorwiegend aus Rauschen. Rechts: DFD bei ungenauer Bewegungsschätzung (Stelle '4' aus Abb. 2 rechts). Zusätzlich zum Bildrauschen sind Strukturinformationen sichtbar (2 Tumore und Blutgefäße der Blase, Abb. 2 links).



stark verrauschte Bildmaterial wurde ein Block-Matching-Verfahren angewandt, das aufgrund großer Blöcke robust gegenüber Rauschen ist. Die Blöcke wurden mit Hilfe des Struktur-Tensors auf Eignung zum Block-Matching hin untersucht und ggf. davon ausgeschlossen. Um die Robustheit gegenüber Ausreißern zu erhöhen, wurde der RANSAC-Algorithmus zur Bestimmung der Parameter des Bewegungsmodells verwendet.

Das Verfahren liefert gute Ergebnisse während langsamer Bewegungen und geringfügige Verletzungen der Annahme einer globalen Bewegung werden toleriert. Schnelle Zooms führen jedoch zu Fehlern in der Bewegungsschätzung.

Die Rechenzeit der vollständigen Suche konnte unter Verwendung eines Greedy Optimierungsverfahrens von 18 ms auf 4 ms reduziert werden, so dass noch ausreichend Rechenzeit für eine echtzeitfähige Filterung der Videosequenz zur Verfügung steht. Durch diese Maßnahme steigt die DFD im Mittel um 1,2% an und die Qualität der Bewegungsschätzung nimmt geringfügig ab.

## Literaturverzeichnis

1. Stehle T, Wulff J, Behrens A, et al. Denoising fluorescence endoscopy: a motion-compensated temporal recursive video filter with an optimal minimum mean square error parametrization. In: Proc IEEE ISBI; 2009. p. 314–7.
2. Behrens A. Creating panoramic images for bladder fluorescence endoscopy. Acta Polytechn J Adv Eng. 2008;48(3):50–4.
3. Bigun J, Granlund GH. Optimal orientation detection of linear symmetry. In: Proc IEEE ICCV; 1987. p. 433–8.
4. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM. 1981;24(6):381–95.
5. Williams DJ, Shah M. A fast algorithm for active contours and curvature estimation. Comput Vis Image Underst. 1992;55(1):14–26.