# Bewertung der Ergebnisse von Segmentierungslösungen in radiologischen Bilddaten am Beispiel des Einsatzes von aktiven Konturen

Regina Pohle, Markus Grohmann, Klaus Tönnies

AG Bildverarbeitung / Bildverstehen
Institut für Simulation und Graphik, Fakultät für Informatik
Otto-von-Guericke-Universität 39016 Magdeburg, PSF 4120
Email: regina@isg.cs.uni-magdeburg.de

**Zusammenfassung.** Im Bereich der medizinischen Bildverarbeitung werden jedes Jahr zahlreiche neue Segmentierungsverfahren entwickelt. Dadurch wird es für einen Anwender immer schwieriger, sich anhand objektiver Kriterien, die aus der Aufgabenstellung und aus dem Bildmaterial ableitbar sind, einen für sein Problem optimalen Algorithmus auszuwählen. Abhilfe könnte hier eine Charakterisierung der Leistungsfähigkeit der einzelnen Ansätze schaffen. In unserem Beitrag werden Möglichkeiten zu einer objektiven Bewertung von Segmentierungsergebnissen diskutiert und erste Erfahrungen bei der Untersuchung der Verhaltensweise aktiver Konturen dargelegt.

**Schlüsselwörter:** Segmentierung, aktive Konturen, Evaluation

## 1 Einleitung

Der Einsatz von Methoden der Bildverarbeitung in der Medizin hat aufgrund der weiteren Verbreitung digitaler Aufnahmetechniken in den letzten Jahren stark zugenommen. Gleichzeitig damit stieg die Anzahl der jährlich entwickelten und speziell an die verschiedenen Aufgabenstellungen angepaßten Segmentierungsalgorithmen. Die Begründung der Auswahl des Verfahrens und ein Vergleich mit Standardalgorithmen erfolgt in den meisten Veröffentlichungen nur auf Basis weniger praktischer Bilder, die eine sichere Aussage über die tatsächliche Leistungsfähigkeit des Verfahrens schwer ermöglichen. Unser Ziel ist es deshalb, anhand der Segmentierung mit aktiven Konturen exemplarisch zu untersuchen, welche Möglichkeiten zur objektiven Bewertung der erzielten Ergebnisse bestehen und wo die Grenzen der einzelnen Bewertungsmethoden liegen.

## 2 Aktive Konturen in der medizinischen Bildverarbeitung

Eine Motivation für unsere Untersuchung der Segmentierung mit aktiven Konturen war deren häufiger Einsatz zur Erkennung von Regionengrenzen in stark verrauschten medizinischen Bildern in den letzten Jahren [1]. Die hierbei verwendete Kombination aus eingesetztem High-level- und Low-level-Wissen ermöglicht zwar zum einen eine

sehr flexible Gestaltung des Verfahrens, zum anderen aber wird die Vorhersage der Qualität des Ergebnisses erschwert.

Die aktive Kontur wurde als Energie minimierende B-Spline-Kurve realisiert. Die parametrische Beschreibung der Kontur erfolgt über v(s)=(x(s), y(s)), wobei x(s) und y(s) die Koordinaten entlang der Kontur sind und s∈[0, 1]. Die Energiefunktion der aktiven Kontur ergibt sich aus einer gewichteten Kombination interner und externer Kräfte.

$$E_{Kontur} = \int_0^1 E_{\text{int}}(v(s)) + E_{ext}(v(s))\, ds \tag{1}$$

Die interne Energie der Spline-Kurve kann über

$$E_{\text{int}} = \alpha(s) \left|\frac{dv}{ds}\right|^2 + \beta(s) \left|\frac{d^2 v}{ds^2}\right|^2 \tag{2}$$

beschrieben werden, wobei α(s) und β(s) die Elastizität und die Steifheit der Kontur spezifizieren. Ein großer Wert für α(s) bewirkt ein Zusammenziehen der Kurve und ein großer Wert für β(s) das Streben hin zu einer kreisförmigen Gestalt. Beide Werte müssen abhängig von der Form des zu segmentierenden Objekts interaktiv eingestellt werden. Den internen Kräften entgegengesetzt wirken die externen Kräfte, die sich in der genutzten Implementation wie folgt zusammensetzen:

$$E_{ext} = w_1 f(x, y) - w_2 \left| grad \quad f(x, y) \right|^2 \tag{3}$$

mit $w_1$ und $w_2$ als Wichtungsfaktoren für den Einfluß des Grauwertes und des Gradienten.

## 3 Evaluationsmethoden

Zur Bewertung von Segmentierungsalgorithmen können sowohl analytische als auch empirische Evaluationsmethoden verwendet werden [2]. Bei den analytischen Methoden erfolgt eine direkte Bewertung des Algorithmus, z.B. seiner Verarbeitungsstrategie, Komplexität und Effizienz ohne Betrachtung einer realen Applikation. Analytische Studien haben zwar den Vorteil, daß sie keine konkrete Implementierung zur Bewertung benötigen. Nachteilig ist jedoch, daß zumeist nur qualitative Aussagen ermittelt werden können.

Die empirischen Methoden bewerten die Qualität der Segmentierungsergebnisse indirekt anhand der Ergebnisse, die für Testbilder erzielt wurden. Sie lassen sich in die Gruppe der Gütemethoden und die Gruppe der Diskrepanzmethoden unterteilen. Bei den Gütemethoden wird die Qualität des Segmentierungsalgorithmus anhand der Erfüllung wünschenswerter Eigenschaften der segmentierten Regionen über Güteparameter gemessen. Gütemethoden haben den Vorteil, daß sie zur On-Line Bewertung genutzt werden können, da keine Referenzbilder benötigt werden. Nachteilig ist je-

doch, daß durch die Wahl der Art der Gütemerkmale dieselben Kriterien, die für die Segmentierung benutzt werden, auch zur Qualitätsbeurteilung herangezogen werden. Die aufwendigsten Verfahren zur Leistungsbewertung sind die Diskrepanzmethoden. Sie gehen davon aus, daß Fehler bei der Segmentierung unvermeidbar sind und beurteilen die Qualität der Ergebnisse anhand der Abweichung zwischen segmentiertem Bild und Referenzbild (Goldstandard). Da mit den Diskrepanzmethoden eine objektive und quantitative Bewertung des Segmentierungsverfahrens mit engem Bezug zur konkreten Anwendung am ehesten gegeben ist, wurden sie für die Untersuchung der Segmentierung mit aktiven Konturen favoritisiert. Probleme bei der praktischen Anwendung ergeben sich vor allem durch das Fehlen eines definierten Goldstandards und bei der Definition einer Metrik zur Fehlermessung.

### 3.1 Goldstandard

Der für die objektive Leistungsbewertung des Verfahrens notwendige Goldstandard ist für reale Probleme nur über eine aufwendige manuelle Festlegung der Regionen durch mehrere geschulte Benutzer zu erzielen. Um zu testen, ob die Güte eines Verfahrens auch einfacher charakterisiert werden kann, verwendeten wir für unsere Untersuchungen einfache künstliche Testbilder. Konkret untersuchten wir auf der Verfahrensseite den Einfluß von Parametrisierung und Initialisierung der Kontur. Auf der Bildseite ermittelten wir den Einfluß von Signal-Rausch-Verhältnis, Kontrast, Shading, Kantensteilheit und Kompaktheit der Objektform. Während für die ersten vier Einflußgrößen sehr leicht Bildserien für die Variation eines Parameters erzeugt werden können, mußte für die Objektform eine derartige Beschreibung erst entwickelt werden. Dabei berechnen sich die Polarkoordinaten der Objektkontur über

$$r(\varphi) = w_{radius} + sin(\varphi \cdot w_{Frequenz}) \cdot w_{Amplitude} + w_{oval} \cdot sin(\varphi)^2 \qquad (4)$$

wobei die vier Werte für w jeweils variiert werden können, um unterschiedliche Aspekte des Einflusses der Objektform auf das Segmentierungsergebnis zu untersuchen und eine möglichst realistische Nachbildung natürlicher Objekte zu erreichen.

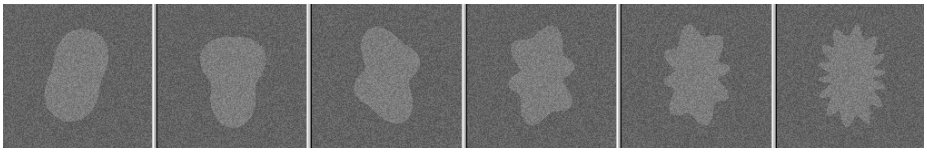

**Abb. 1:** Testobjekte mit konstanten Werten für $w_{Radius}$ von100, $w_{Amplitude}$ von 20, $w_{oval}$ von 50 sowie variablem $w_{Frequenz}$ von 2, 3, 4, 6, 8 und 14

### 3.2 Metriken zur Fehlerbewertung

Für die Bewertung der Qualität der Segmentierungsergebnisse benutzten wir die folgenden in der Literatur [2, 3] gefundenen Metriken, die eine möglichst von Objekt und Segmentierungsverfahren unabhängige Charakterisierung des Fehlermaßes erlauben:

1. mittlere Abweichung der Konturpixel von der Kontur im Goldstandard [2],

2. Hausdorf-Abstand, berechnet nach [3] über

$$d_{Hausdorf}\ (A, B) = \max\left( \max_i \{d(a_i . B)\}, \max_j \{d(b_j . A)\} \right) \tag{5}$$

mit A und B als Konturen vom segmentierten Bild und vom Goldstandard und

$$d(a_i . B) = \min_i \left\| b_j - a_i \right\| \tag{6}$$

3. Fehlerwahrscheinlichkeit basierend auf der Zahl falsch segmentierter Pixel [2],
4. relative Meßsicherheit basierend auf Merkmalswerten der segmentierten Objekte, exemplarisch betrachtet für den Flächeninhalt [2]

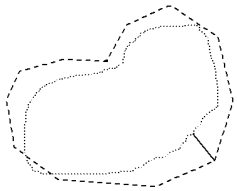

**Abb. 2:** Beispiel für die Berechnung des Hausdorf-Abstands (durchgehend) zwischen Kurve des Goldstandards (gestrichelt) und Ergebnis der Segmentierung (gepunktet)

## 4 Ergebnisse

Bei der Untersuchung des Einflusses der Initialisierung auf das Ergebnis anhand der Testbilder war festzustellen, daß bei kleiner werdendem Stützpunktabstand der die aktive Kontur bildenden Splinekurve die Abhängigkeit von der Form und Lage der Initialisierung abnahm. Bei einem Stützpunktabstand von weniger als 5 Pixeln war das Ergebnis reproduzierbar. Für reales Bildmaterial (Ultraschallbilder) traf diese Aussage jedoch aufgrund von möglichen Binnenstrukturen nicht immer zu. Ein anderer Grund für ein unterschiedliches Verhalten der aktiven Kontur in Test- und realen Bildern liegt in der verwendeten sehr einfachen Modellierung des Rauschens bei den Testbildern.

Auch bei der Wahl der Parameter für Steifheit und Elastizität konnte ein Unterschied zwischen der Anwendung auf Testbilder und reale Bilder festgestellt werden. Im Gegensatz zu den Testbildern mit quadratischen Objekten, bei denen die mittlere Abweichung der Konturpixel unabhängig von den eingestellten Parametern zwischen einem und zwei Pixeln betrug, kam es bei realen Bildern trotzt ähnlich kompakter Objektform bei einer Veränderung der Parameter auch zu unterschiedlichen Ergebnissen. Dies liegt daran, daß in realen Bildern der Kontrast und die Steilheit des Übergangs an den Regionengrenzen, z.B. in Ultraschallbildern in Form von Schallschatten, variiert. Aufgrund der daraus resultierenden schwankenden externen Kräfte gewinnen die internen Kräfte an Bedeutung für das endgültige Ergebnis.

Bei der Untersuchung des Einflusses der Bildparameter zeigte sich sowohl für ein Signal-Rausch-Verhältnis unterhalb von 2:1 als auch für eine Kontrastreduzierung unterhalb von 0.2 eine Verschlechterung des Segmentierungsergebnisses. In beiden Fällen nahm gleichzeitig die Abhängigkeit von der Initialisierung zu. Weiterhin konnte festgestellt werden, daß Shading-Effekte das Ergebnis der Segmentierung

kaum beeinflußten. Die Kantensteilheit spielte dagegen eine wichtige Rolle. So führte ein Graustufenübergang in Form einer idealen Rampenkante mit der Breite von 6 Pixeln im Vergleich zu einer idealen Stufenkante zu einer signifikanten Erhöhung des gemessenen Fehlers für alle ermittelten Fehlermaße. Die wie erwartet festgestellte Abhängigkeit der Ergebnisse von Kontrast, Signal-Rausch-Verhältnis und Kantensteilheit ist dadurch bedingt, daß alle diese Bildparameter den Gradienten beeinflussen, so daß sich jeweils die externen Kräfte verändern, die einen starken Einfluß auf die aktive Kontur ausüben. Bei den unterschiedlichen Diskrepanzmaßen erwies sich der Hausdorf-Abstand als besonders aussagekräftig (Abb. 3).

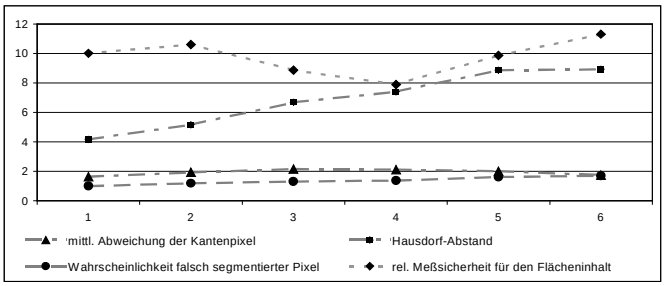


**Abb. 3:** Ergebnisse der Diskrepanzmessung gemittelt pro Meßpunkt über fünf Durchläufe für die Segmentierung von Testbildern bei Variation der Form durch Erhöhung des Parameters $w_{Frequenz}$

## 5 Diskussion

Bei der Überprüfung der Übertragbarkeit der Aussagen von den Testbildern auf die realen Bilder zeigte sich, daß diese nur für bestimmte Aufgabenklassen gelten, bei denen lokal nur am Objekt hohe Gradienten vorliegen (z.B. Abb. 4a). Bei Bildern mit mehreren Objekten bzw. anderen Strukturen mit hohen Gradienten (z.B. Abb. 4b) beeinflußten entfernter liegende Gradienten das Verhalten der aktiven Kontur. Für Untersuchungen nicht rein lokal arbeitender Segmentierungsverfahren müssen deshalb Testszenarien verwendet werden, die neben der interessierenden Region weitere Regionen mit hohen Gradienten enthalten.

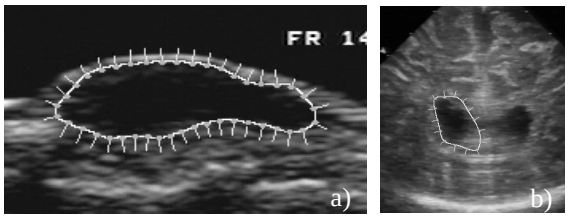


**Abb. 4:** Ultraschallbilder mit segmentiertem Hauttumor (a) und linken Ventrikel (b) durch Verwendung des aktiven Kontur-Algorithmus

## 6 Literatur


1. Singh A, Goldof D, Terzopoulos D: Deformable Models in Medical Image Analysis,. IEEE Computer Society, Los Alamitos, 1998.
2. Zhang YJ: A Survey on Evaluation Methods for Image Segmentation. Pattern Recognition, 29(8):1335-1346, 1996.
3. Chalana V, Kim Y: A Methodology for Evaluation of Bounary Detection Algorithms on Medical Images. IEEE Trans. on Medical Imaging. 16(5):642-652, 1997.