

Silberstandards aus Fourier-basierter Textursynthese zur Evaluierung von Segmentierungsalgorithmen

Thomas M. Lehmann, Jörg Bredno und Klaus Spitzer

Institut für Medizinische Informatik
Universitätsklinikum der RWTH Aachen, Pauwelsstr. 30, D – 52057 Aachen
Email: lehmann@computer.org

Zusammenfassung. Segmentierung ist eine wesentliche Aufgabe der medizinischen Bildverarbeitung, aber es fehlen einheitliche Methoden zu ihrer Evaluierung. Goldstandards stehen nur selten zur Verfügung und eine subjektive visuelle Inspektion ergibt keine reproduzierbaren Ergebnisse. Mit der hier vorgestellten Methode werden realistische Silberstandardbilder erzeugt, bei denen das richtige Ergebnis einer Segmentierung a-priori bekannt ist. Beispieltexuren aus realen Bildern werden gesammelt und in einer Fourier-Repräsentation gespeichert. Durch Neukombination der Amplituden- und Phasenspektren sowie der Mittelwerte werden synthetische Texturen gebildet. Diese werden dann in ein Bild mit vorgegebener Referenzkontur eingefügt. Mit der Methode wurde ein Ballon-Modell zur Segmentierung auf 2D Mikroskopien, 3D CT- und 4D MR-Bilddaten evaluiert. Dabei wurden auch systematische Abweichungen quantifiziert, die durch die Generierung der Silberstandards induziert werden. Sie liegen für alle Bilddimensionen im Subpixelbereich.

1 Einleitung

Obwohl die Segmentierung ein fundamentaler Schritt bei der automatischen Bildanalyse ist, existieren noch keine einheitlichen Methoden zur Evaluierung dieser Verfahren [1, 2]. Die theoretisch-analytische Betrachtung des Algorithmus ohne dessen Realisierung ist nicht praxisgerecht. Die empirische Bewertung durch den subjektiv-visuellen Eindruck eines Betrachters unterliegt einer inter- und intra-Beobachter-Variabilität. Zusätzlich ist die exakte Lage von Gewebegrenzen (Goldstandard) auf realem Bildmaterial meistens nicht bekannt [3]. Für die Segmentierung medizinischen Bildmaterials liegt bislang keine ausreichende Sammlung an Goldstandards vor [2]. Auf synthetischem Bildmaterial können zwar objektive Übereinstimmungsmaße bestimmt werden, diese sind aber nur bedingt verallgemeinerbar und kaum auf reale Probleme zu übertragen [4]. Oftmals werden zur Bildsynthese Modellannahmen getroffen, die auch der Segmentierung zugrunde liegen. Die Ergebnisse solcher Untersuchungen sind dann wertlos [5].

Der hier vorgestellte konsistente Ansatz erzeugt synthetische aber realistische Bilder, auf denen die Objektgrenzen a-priori bekannt sind. Diese Silberstandards sind zur Evaluierung unterschiedlichster Segmentierungsalgorithmen geeignet und werden hier exemplarisch für drei Modalitäten untersucht.

2 Methode

Ein Silberstandard wird aus drei synthetisierten Regionen im Bild zusammengesetzt: Dem Inneren und Äußeren eines dargestellten Objektes sowie einer optionalen Konturzone [6].

2.1 Referenztexturen

Zu allen Komponenten werden zunächst exemplarische Texturen aus realen Bildern extrahiert, die das charakteristische Erscheinungsbild des Gewebes mit allen auftretenden Artefakten widerspiegeln. Eine Texturausschnitt wird in beide Raumrichtungen gespiegelt. Das diskrete und stetige Textursignal $t(x, y)$ wird danach durch Subtraktion von μ_t in das mittelwertfreie Signal $t_0(x, y)$ überführt und Fourier-transformiert:

$$t(x, y) - \mu_t = t_0(x, y) \quad \circ \bullet \quad T_0(u, v) = r_t(u, v) \cdot \exp(-j \varphi_t(u, v)) \quad (1)$$

Zu jedem Texturbeispiel werden Amplitude $r_t(u, v)$ und Phase $\varphi_t(u, v)$ zusammen mit dem Mittelwert μ_t in einer Datenbank gespeichert.

Exemplarische Texturen einer Konturzone basieren auf einer manuellen oder automatischen Segmentierung ausgewählter Bilder. Entlang der bekannten Kontur wird ein schmales Band vorgegebener Dicke extrahiert und durch lineare Interpolation [7] in ein rechteckiges Bild z' umgewandelt. Dieses wird der Länge nach gespiegelt. Die so erhaltene Konturtextur $z(x, y)$ wird mit (1) nach Amplitude $r_z(u, v)$, Phase $\varphi_z(u, v)$ und Mittelwert μ_z zerlegt.

2.2 Textursynthese

Für eine flexible Generierung der Silberstandardbilder $s(x, y) = m_t(x, y) + m_g(x, y)$ wird eine mittelwertfreie Texturkarte m_t und eine Grauwertkarte m_g erstellt. Zur Bestimmung von m_t und m_g wird eine automatisch segmentierte oder manuell generierte Objektkontur vorgegeben. Die Segmentierung $m_s(x, y)$ enthält den Wert 1 für alle Punkte im Innern der Kontur und 0 außerhalb.

Für Inneres und Äußeres der Kontur wird je ein zufällig aus der Datenbank gewähltes Amplitudensignal $\hat{r}_t(u, v)$ mit einem zufällig gewählten $\hat{\varphi}_t(u, v)$ zum darzustellenden Gewebe kombiniert. Zusätzlich wird $\hat{\varphi}_t(u, v)$ mit additivem Rauschen $n_{0,\sigma}$ mit $\sigma = \frac{1}{10}\pi$ überlagert. Eine synthetische Textur $\tilde{t}(x, y)$ wird durch inverse Fourier-Transformation gebildet:

$$\tilde{t}(x, y) \quad \bullet \circ \quad \hat{r}_t(u, v) \cdot \exp\left(-j(\hat{\varphi}_t(u, v) + n_{0,\sigma}(u, v))\right) \quad (2)$$

Texturkarte. Die synthetische Textur \tilde{t} ist nicht mehr symmetrisch aber mittelwertfrei und kann stetig aneinandergesetzt werden. Die Texturkarte m_t wird jeweils mit der in m_s codierten Textur besetzt. Hierbei werden große Bildbereiche durch Wiederholung von \tilde{t} gefüllt (Abb. 1a).

Entlang der Konturzone kann eine weitere synthetische Textur $\tilde{z}(x, y)$ in das Bild eingefügt werden. Diese linearisierte Konturzone entsteht analog durch Rücktransformation zufällig ausgewählter Amplituden- und Phasenspektren, wobei die Phase jedoch nicht verrauscht wird. Die synthetische Konturzone \tilde{z} wird dann durch geometrische Transformationen entlang der Kontur des Silberstandards in die Texturkarte m_t eingefügt (Abb. 1c).

Grauwertkarte. Die Grauwertkarte m_g wird gebildet, indem zunächst zu den zu synthetisierenden Texturen passende Mittelwerte $\hat{\mu}_{\text{in}}$ und $\hat{\mu}_{\text{out}}$ zufällig aus der Datenbank gelesen werden. Für alle Pixel in m_g wird dann die kartesische Distanz zur Kontur selbst ermittelt. Wenn diese Distanz größer als ein vorgegebener Wert w ist, wird das Pixel gemäß dem Label aus m_s auf $\hat{\mu}_{\text{in}}$ oder $\hat{\mu}_{\text{out}}$ gesetzt. Für Punkte in der Nähe zur Kontur wird ein linearer Übergang der Grauwerte in einer Breite $2w$ über die Kontur erzeugt, damit im Silberstandard keine unrealistisch starken Gradienten auftreten (Abb. 1b). Wird eine eigene Konturzone synthetisiert, wird zusätzlich ein zufälliger Mittelwert $\hat{\mu}_{\text{zone}}$ aus der Datenbank gelesen. Der Bereich um die Kontur wird mit $\hat{\mu}_{\text{zone}}$ gefüllt. Dann wird jeweils ein linearer Übergang der Breite $2w$ von $\hat{\mu}_{\text{in}}$ nach $\hat{\mu}_{\text{zone}}$ sowie von $\hat{\mu}_{\text{zone}}$ nach $\hat{\mu}_{\text{out}}$ in m_g angelegt (Abb. 1d).

2.3 Höherdimensionale Silberstandards

Die Methode wurde auf drei- und vierdimensionales Bildmaterial übertragen. Dabei kann für alle Schichten eines Silberstandards eine gemeinsame, oder für jede Schicht eine individuelle Auswahl von \tilde{t} und $\hat{\mu}$ erfolgen. Zusätzlich müssen Bildgebungseigenschaften wie Schichtabstand oder Bewegungsartefakte berücksichtigt werden. Für Farbbilder erfolgt die Generierung von Silberstandards unabhängig auf den einzelnen Farbkanälen, wenn in diesen eine Mittelwertbestimmung möglich ist, bspw. RGB-Bilder. Rotatorische Kanäle wie der H-Kanal in HSV-Bildern können derzeit nicht modelliert werden.

Das Ergebnis der Segmentierung eines Silberstandards kann immer als binäre Segmentierungskarte $m_r(x, y)$ dargestellt werden. Da die Kontur a-priori bekannt ist, können quantitative Übereinstimmungs- und Distanzmaße bestimmt werden, bspw. $O = |m_r \cap m_s| / |m_r \cup m_s| \cdot 100\%$. Liegt die Segmentierung als Kontur mit Stützstellen vor, kann zusätzlich eine mittlere Distanz \bar{d} und die Hausdorff-Distanz (maximale Distanz) H der Knoten zur Referenzkontur ermittelt werden [6].

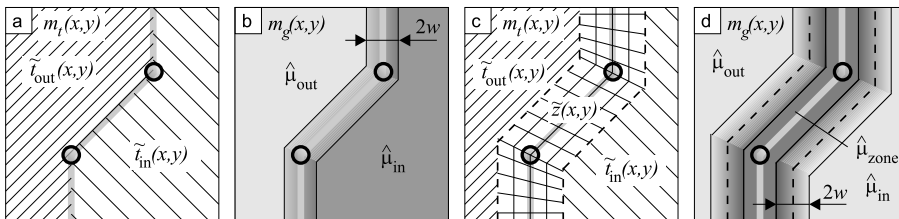


Abb. 1. Textur- und Grauwertkarten ohne (a,b) und mit (c,d) einer Konturzone.

Tabelle 1. Ergebnisse der Segmentierung von Silberstandardbildern.

Bild- material	Referenzkontur				Silberstandard		
	#	\bar{d}	H	O	#	\bar{d}'	Δ_d
2D	197	1,67	8,06	97,55 %	193	1,50	-0,17
3D	30	2,90	21,59	85,99 %	27	2,17	-0,73
4D	10	1,16	6,29	79,98 %	9	0,47	-0,74

3 Anwendung und Ergebnisse

Die Methode wurde zur kontextfreien Evaluierung eines Ballon-Modells [8] verwendet, das medizinisches Bildmaterial in 2, 3 und 4 Dimensionen segmentiert (Tab. 1, links). Die Tests erfolgten auf histochemisch gefärbten Dünnschnittmikroskopien von Neuronen (je 50 Silberstandards zu vier Vorlagen, Abb. 2), CT-Schichtbildern der Wirbelsäule (je 10 Standards zu drei Vorlagen) und MR-Zeitfolgen des linken Ventrikels über einen Herzzyklus (10 Silberstandards).

Im 2D-Fall war die Variabilität der erzeugten Silberstandards so hoch, daß die Ballon-Segmentierung bei drei Bildern scheiterte. Dies entspricht der Erkennungsrate von 96% auf realen Daten [9]. Die hohe Hausdorff-Distanz H für die CTs der Wirbelsäule zeigt, daß es dem Modell nicht fehlerfrei gelingt, die komplexe Geometrie abzubilden. Die geringe mittlere Distanz \bar{d} zeigt aber gleichzeitig, daß solche Fehler nur vereinzelt auftreten. Zu beachten ist, daß das Übereinstimmungsmaß O aus Normierungsgründen über die Bilddimensionen hinweg nicht vergleichbar ist.

Um durch Silberstandards induzierte, systematische Abweichungen zu bestimmen, wurden die Ergebniskonturen auch mit je einer zufällig ausgewählten Silberstandard-Segmentierung verglichen (Tab. 1, rechts). Der Unterschied Δ_d zur mittleren Distanz dieser Kontur \bar{d}' betrug im 2D-Fall nur 0.17 Pixel. Bei den 3D- und 4D-Silberstandards ist die methodisch bedingte Abweichung etwas größer, aber immer noch kleiner als ein Voxel bzw. Stixel.

4 Diskussion

Goldstandards stehen in der medizinischen Bildverarbeitung nur selten zur Verfügung, da sie häufig invasive Präparationen oder die Extraktion des dargestell-

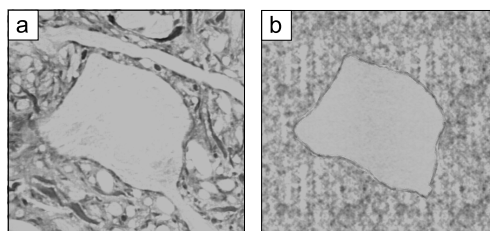


Abb. 2. Histochemisch gefärbtes Motoneuron (a) und ein Silberstandardbild (b).

ten Gewebes erfordern [1]. Fourier-Repräsentationen zur Textursynthese haben bereits beim Training von Klassifikatoren gute Ergebnisse gezeigt [10]. Unsere Fourier-basierte Methode zur Generierung von Silberstandards ist vielseitig, einfach zu handhaben und zur Evaluierung unterschiedlichster Segmentierungsmethoden geeignet. Die Variabilität der Silberstandardbilder erreicht oder überschreitet sogar die real vorkommende Variabilität heterogenen Bildmaterials, da Eigenschaften von jeweils drei Beispielt Texturen kombinatorisch vereint werden. Bereits mit je 3 Texturmustern können 256 verschiedene Silberstandards generiert werden. Durch das Verrauschen der Phase wird diese Zahl beliebig gesteigert. Damit findet die Methode ein breites Anwendungsfeld, das lediglich Fourier-basierte Segmentierungsverfahren ausschließt.

Nach wie vor muß eine akzeptable Segmentierung auch dem subjektiven Eindruck eines Experten auf realen Bilddaten entsprechen. Um diese Forderung zu erfüllen, war bisher intensive Interaktion nötig. Mit unseren Silberstandards wird dieser Aufwand erheblich verringert, ohne daß unzulässige Vereinfachungen der Evaluierung auftreten. Die Ergebnisse solcher Evaluationen sind unabdingbar für die klinischen Nutzer [3].

Mit den Silberstandards wird es möglich, die Eigenschaften einer Referenzkontur gemäß einer vorgegebenen Verteilung zu variieren. Damit kann die von HARALICK geforderte Evaluierung auf Basis der statistischen Verteilung von Werten vorgenommen werden [4]. Zusätzlich kann ein fruchtbarer Methodewettbewerb stattfinden, da mithilfe der Silberstandardbilder standardisierte Evaluationen möglich sind [2].

Literatur

1. Zhang YJ: A survey on evaluation methods for image segmentation. *Pattern Recognition* 29(8):1335–1346, 1996.
2. Gee JC: Performance evaluation of medical image processing algorithms. *Procs. SPIE* 3979:19–27, 2000.
3. Haynor DR: Performance evaluation of image processing algorithms in medicine: A clinical perspective. *Procs. SPIE* 3979:18, 2000.
4. Haralick RM: Validating image processing algorithms. *Procs. SPIE* 3979:2–16, 2000.
5. Chalana V, Kim Y: A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Trans. MI* 16(5):642–652, 1997.
6. Lehmann TM, Bredno J, Spitzer K: Silver standards obtained from Fourier-based texture synthesis to evaluate segmentation procedures. *Procs. SPIE* 4322, in press.
7. Lehmann TM, Gönner C, Spitzer K: Survey: Interpolation methods in medical image processing. *IEEE Trans. MI* 18(11):1049–1075, 1999.
8. Bredno J, Lehmann TM, Spitzer K: A general finite-element model for segmentation in 2, 3, and 4 dimensions. *Procs. SPIE* 3979:1174–1184, 2000.
9. Lehmann TM, Bredno J, Metzler V, Brook G, Nacimiento W: Computer-assisted quantification of axo-somatic boutons at the cell membrane of motoneurons. *IEEE Trans. BME*, accepted.
10. Egmont-Petersen M, Schreiner U, Tromp SC, Lehmann TM, Slaaf DW, Arts T: Detection of leukocytes in contact with the vessel wall from in-vivo microscope recordings using a neural network. *IEEE Trans. BME* 47(7):941–951, 2000.