

Evaluierung von Erscheinungsmodellen für die Segmentierung mit statistischen Formmodellen

Sascha Münzing, Tobias Heimann, Ivo Wolf und Hans-Peter Meinzer

Abteilung für Medizinische und Biologische Informatik, DKFZ Heidelberg
Email: s.muenzing@dkfz-heidelberg.de

Zusammenfassung. Zur Segmentierung mit Formmodellen werden Erscheinungsmodelle benötigt, die eine exakte Anpassung an die Bilddaten ermöglichen. Wir haben Standardverfahren und nichtlineare Modellierungsmethoden für die 3D-Segmentierung erweitert und auf klinischen Leberdaten evaluiert. Mit nichtlinearen Erscheinungsmodellen wurden insgesamt deutlich bessere Ergebnisse erzielt. Dabei erbrachte die Erweiterung um einen generischen Ansatz zur Erhöhung der Datendichte eine zusätzliche signifikante Verbesserung.

1 Einleitung

Die Segmentierung ist ein wichtiger Schritt in vielen Bereichen der medizinischen Bildverarbeitung, wie z.B. in der Operationsplanung, Bestrahlungsplanung oder bildbasierter Diagnostik. Da medizinische Bilder häufig durch Rauschen und andere Artefakte gestört sind, ist zur automatische Erkennung und Extraktion von relevanten Gewebsstrukturen (z.B. der Leber oder dem Herz) oftmals Vorwissen nötig, d.h. ein internes Modell der erwarteten Form und/oder Intensitäten.

Bei Statistischen Formmodellen, auch als Active Shape Models [1] bekannt, wird Vorwissen aus Trainingsbildern gelernt. Dabei wird jedes Trainingsbild durch eine Menge von Punkten auf der Objektoberfläche repräsentiert, so genannten Landmarken. Aus den korrespondierenden Positionsdaten aller Trainingsbilder wird eine mittlere Form berechnet und anhand einer Hauptkomponentenanalyse die zugehörigen Varianzen ermittelt. Jedes Trainingsbild kann somit aus der mittleren Form zuzüglich einer Linearkombination von Verschiebungsvektoren beschrieben werden.

Um das generierte Formmodell zur Segmentierung verwenden zu können, müssen sich die Landmarken des Formmodells an die Bilddaten anpassen. Es wird, vergleichbar zum Formmodell, Vorwissen über die Landmarken benötigt, um das Aussehen an einer bestimmten Position, mit dem erwartenden Aussehen für einen gegebenen Landmarkenpunkt zu vergleichen. Hierzu werden so genannte Erscheinungsmodelle erstellt, die die Grauwertumgebung einer Landmarke anhand bestimmter Eigenschaften abbilden. Man spricht auch von lokalen Erscheinungsmodellen, da nicht die globale Erscheinung des ganzen Formmodells, sondern jeweils der lokale Bereich um eine Landmarke abgebildet wird.

2 Stand der Forschung und Fortschritt durch den Beitrag

In der ersten Version der Active Shape Models wurden die Landmarken mit einem simplen Gradientenprofil gesucht. Später folgten Statistische Modelle [2], bei denen das Aussehen einer Landmarke durch ein mittleres Profil aus allen Trainingsbildern, sowie die zugehörigen Kovarianzen modelliert wird. Normalverteilte Grauwerte werden damit korrekt abgebildet, um aber beliebige Verteilungen modellieren zu können, werden in neuerer Zeit nichtlineare Modelle verwendet. In [3] wird ein k NN-Klassifikator mit Grauwertprofilen verwendet, um sowohl die Objektgrenze, als auch Bereiche in der Landmarkenumgebung zu modellieren. In [4] wird ein Multiskalenklassifikator zur 2D-Segmentierung eingesetzt, bei dem anhand des Multi Local Jet Verfahrens differenzierte Bildmerkmale aus sogenannten locally orderless images berechnet werden. Diese Verfahren wurden für die 3D-Segmentierung erweitert und auf klinischen Daten der Leber evaluiert. Durch die einheitliche Datenbasis ist ein direkter Vergleich der Methoden gegeben.

3 Methoden

Die evaluierten Erscheinungsmodelle wurden zur Verwendung innerhalb eines Multiresolution-Frameworks entwickelt. Dabei wird für jedes CT Volumen eine Bildpyramide aus fünf verschiedenen Auflösungen R_0 bis R_4 erstellt. Auf der Originalauflösung R_0 beträgt die Schrittweite 1mm und auf den nachfolgenden Stufen wird diese jeweils verdoppelt. Für jede Auflösungsstufe werden separate Erscheinungsmodelle erstellt.

3.1 Standardverfahren

Bei der einfachsten Methode werden Landmarken anhand der stärksten Kante gesucht. Entlang der Oberflächennormalen wird ein normalisiertes Gradientenprofil der Länge $k = 2n + 1$ berechnet, mit jeweils n Punkten zu jeder Seite der Oberfläche. Hierbei besteht das Vorwissen lediglich aus der Annahme, dass der Grauwertgradient an der Position der Objektkante maximal sei.

Bei den Statistischen Modellen werden Grauwerte durch Profile der Länge k - analog zum obigen Modell - für jede Landmarke und aus allen Trainingsbildern extrahiert. Aus den gesammelten Profilen $\mathbf{g}_i = (g_1, g_2, \dots, g_k)^T$ wird zu jeder Landmarke ein mittleres Profil $\bar{\mathbf{g}}$ und eine Kovarianzmatrix \mathbf{S}_g berechnet. Zur Anpassung einer Landmarke wird ein neues Profil \mathbf{g}_s mit den Profilen aus der Verteilung verglichen, indem die quadrierte Mahalanobisdistanz $f(\mathbf{g}_s)$ berechnet wird

$$f(\mathbf{g}_s) = (\mathbf{g}_s - \bar{\mathbf{g}})^T \mathbf{S}_g^{-1} (\mathbf{g}_s - \bar{\mathbf{g}}) \quad (1)$$

Wenn das Profil \mathbf{g}_s auf der optimalen Position liegt, ist die Mahalanobisdistanz minimal. Das Verfahren wurde mit einfachen Grauwertprofilen und normalisierten Gradientenprofilen evaluiert, da diese in [2] die besten Ergebnisse erzielten.

3.2 Nichtlineares Grauwertmodell

Beim nichtlinearen Grauwertmodell wird das Aussehen einer Landmarke anhand eines k NN-Klassifikators verteilungsfrei modelliert. Es werden Grauwertprofile der Länge $k = 2n + 1$ entlang der Oberflächennormalen abgetastet, aber im Gegensatz zum Statistischen Modell werden nicht nur Profile direkt auf der Objektgrenze (Kantenprofile) verwendet, sondern zusätzlich $2n$ Profile aus der Landmarkenumgebung (Umgebungsprofile), jeweils n Profile nach außen und n in Richtung zum Objektinneren. Die Menge der Trainingsprofile besteht also aus zwei Klassen. Die Wahrscheinlichkeit, dass ein Profil g_s aus einem neuen Bild auf dem Objektrand liegt, ist durch die Posteriori Wahrscheinlichkeit des k NN-Klassifikators gegeben

$$P(\text{Kantenpunkt}|g_s) = \frac{n_{\text{true}}}{k_{\text{nn}}} \quad (2)$$

wobei n_{true} die Anzahl der Kantenprofile aus den k_{nn} nächsten Nachbarn ist. Bei der NN-Suche werden die Trainingsprofile in einem KD-Tree gespeichert, wobei die Grauwerte eines Profils die Koordinaten eines k -dimensionalen Punktes darstellen. Gewöhnlich ist die resultierende Datendichte gering, deshalb wurde wie in [5] ein K-Means-Clustering durchgeführt. Durch dieses Verfahren kann man die Leber in ca. 20 Cluster einteilen und somit Daten aus durchschnittlich 128 Landmarken in einem Erscheinungsmodell zusammenfassen.

3.3 MultiLocalJet-Merkmale

Bei diesem nichtlinearen Erscheinungsmodell werden Landmarken nicht durch Grauwertprofile sondern durch differenzierte Merkmale aus Bildpunkten des Objekts und der angrenzenden Umgebung modelliert. Für ein Bild werden einzelne Voxel anhand des Localjet-Verfahrens als ein Taylorpolynom interpretiert. Die benötigten Taylorkoeffizienten sind die Ableitungen an einem Bildpunkt bis zum entsprechenden Grad des Polynoms. Auf den Skalenraum des Bildes angewendet, resultiert eine große Anzahl an Merkmalskombinationen. Wählt man fünf Auflösungsstufen und alle Ableitungen bis zum zweiten Grad (auch gemischte), erhält man für den 3D-Fall insgesamt 45 Merkmale. Um aus dieser Menge die besten Merkmale zu selektieren, wird zunächst eine klassifikatorabhängige Sequential Feature Forward Selection durchgeführt, wobei zu jeder Landmarke maximal 10 Merkmale ausgewählt werden. Zusammen mit einem k NN-Klassifikator ($k_{\text{nn}}=10$, $e^{-\text{Distanz}}$) ergibt sich ein Multiskalenklassifikator aus den besten Merkmalen für eine entsprechende Anwendung und einen bestimmten Bildbereich. Um in einem neuen Bild eine Landmarke zu lokalisieren, werden zunächst $k = 2n + 1$ Bildpunkte entlang der Oberflächennormalen klassifiziert

$$f(g) = \sum_{i=-n}^{-1} g_i + \sum_{i=0}^{+n} (1 - g_i) \quad (3)$$

Anhand der Kostenfunktion $f(g)$ wird die optimale Landmarkenposition als die Stelle definiert, an der die Wahrscheinlichkeit am größten ist, dass die Profilpunkte g_i innerhalb der Kontur $i = 0 \dots +n$ tatsächlich zum Objekt gehören und die Punkte außerhalb $i = -n \dots -1$ zur Umgebung.

3.4 Evaluationsverfahren

Die Güte eines Erscheinungsmodells wird unabhängig von einem Suchverfahren für alle Trainingsbilder bestimmt, indem die Anpassung der Modelle an der Originallandmarkenposition und an drei Positionen auf jeder Seite der Oberfläche untersucht wird. Um die Bedingungen während einer Bildsuche zu simulieren, erfolgt eine Randomisierung der Landmarkenposition mit einer Varianz von 1 mm in der Originalauflösung entlang der Oberfläche (in den nachfolgenden Auflösungen jeweils verdoppelt). Gleichzeitig wird die Richtung des Normalenvektors mit einer Varianz von ungefähr zehn Grad randomisiert. Auf diese Art werden in jedem Trainingsbild zu jeder Landmarke 20 Stichproben entnommen. Der Index der Stelle mit der besten Anpassung (ein Bereich zwischen -3 und 3) wird gespeichert und zur Erzeugung eines Histogramms verwendet. Aus den Häufigkeiten der auftretenden Verschiebungen kann als Gütemaß der mittlere Quadratfehler berechnet werden [6].

4 Ergebnisse

Das Bildmaterial bestand aus 32 CT-Aufnahmen des Abdomens, mit einer Auflösung von 512x512 Voxeln in der Ebene und einer variierenden Anzahl von 60 bis 130 Schichten. Der Voxelabstand variiert in der Ebene zwischen 0,55mm und 0,8mm und der Schichtabstand beträgt meistens 3mm, in einigen Fällen 5mm.

Zu jedem Erscheinungsmodell sind in Abbildung 1 die mittleren quadratischen Fehler und in Abbildung 2 die entsprechenden Histogramme für alle fünf Auflösungsstufen dargestellt.

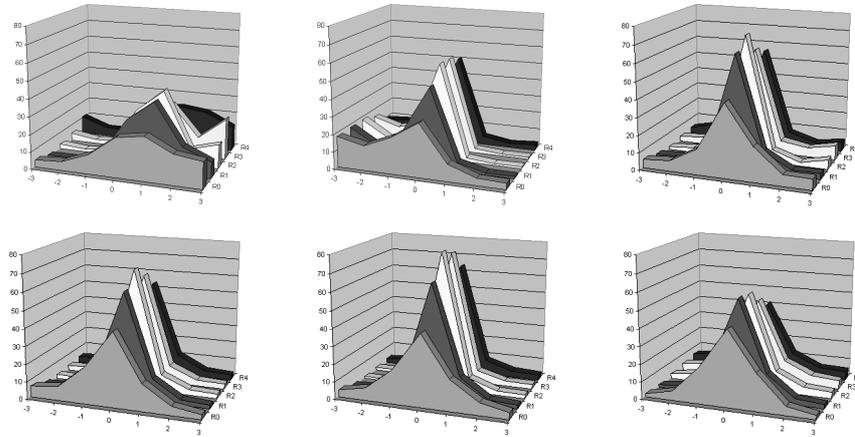
5 Diskussion

Das erreichbare Maximum wird durch die anisotropen Auflösungen der verwendeten Trainingsbilder und insbesondere in der Originalauflösung durch die Un-

Aufl.- stufe	ohne Training	Statistische Modelle		Nichtlineare Modelle		
	Gradient	Intensität	Gradient	ungeclustert	geclustert	LocalJet
R0	2,97	2,92	1,85	1,67	1,63	1,22
R1	2,10	2,38	0,99	0,95	0,77	0,94
R2	2,36	2,22	0,99	0,86	0,57	1,27
R3	3,38	2,09	1,59	1,06	0,75	1,60
R4	3,56	1,84	1,48	1,26	1,02	1,84

Abb. 1. Tabelle der mittleren quadratischen Fehler der Erscheinungsmodelle

Abb. 2. Standardverfahren (oben): Stärkste Kante, Statistische Modelle mit Intensitätsprofil und normalisierten Gradientenprofil; Nichtlineare Modelle (unten): ungeclusterte und geclusterte Grauwertprofile, LocalJet-Merkmale



genauigkeiten bei den manuellen Segmentierungen (klinischer Goldstandard) begrenzt. Diese haben ab der zweiten Stufe R_1 durch die geringere Auflösung und größere Schrittweite immer weniger Einfluss. Die schlechteren Ergebnisse ab R_3 resultieren vermutlich aus den ab dieser Stufe relativ großen Profilen und der damit einhergehenden großen Bildvariationen, teilweise ragen die Profile sogar über den Bildrand hinaus.

Mit den besten nichtlinearen Modellen werden Landmarken im Vergleich zum besten Standardverfahren um durchschnittlich 30% besser abgebildet. Besonders das geclusterte Modell liefert insgesamt bessere Ergebnisse, darum soll in Zukunft auch das LocalJet-Modell um diesen Ansatz erweitert werden.

Literaturverzeichnis

1. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models: Their training and application. *Comput Vis Image Underst* 1995;61(1):38–59.
2. Cootes T, Taylor C. Active shape model search using local grey-level models: A quantitative evaluation. In: *Procs BMVC*. BMVA Press; 1993. 639–648.
3. de Bruijne M, van Ginneken B, Niessen WJ, et al. Active shape model segmentation using a non-linear appearance model: Application to 3D AAA segmentation. *Inst. of Inform. and Comp. Sciences, Utrecht University*; 2003.
4. van Ginneken B, Frangi AF, Staal JJ, et al. Active shape model segmentation with optimal features. *IEEE Trans Med Imaging* 2002;21(8):924–933.
5. Bacher MG, Pekar V, Kaus MR. Model-based segmentation of anatomical structures in MR images of the head and neck area. *Procs BVM* 2005.
6. Heimann T, Wolf I, Meinzer HP. Active shape models for a fully automated 3D segmentation of the liver: An evaluation on clinical data. *Procs MICCAI* 2006; 41–48.