

Segmentierung von Risikostrukturen der Otobasis für die minimal-invasive multi-port Chirurgie

M. Becker, M. Kirschner, G. Sakas

TU Darmstadt, Graphisch-Interaktive Systeme, Darmstadt

Kontakt: meike.becker@gris.tu-darmstadt.de

Abstract:

Das Forschungsprojekt MUKNO untersucht einen minimal-invasiven multi-port Ansatz für Operationen an der Otobasis. Eine der größten Herausforderungen dabei besteht darin, die verschiedenen Risikostrukturen wie Gesichtsnerv oder Blutgefäße nicht zu verletzen. Daher ist die Segmentierung dieser anatomischen Strukturen in Bilddaten für die Operationsplanung ein essentieller Schritt. Hierbei sind die geringe Größe der Strukturen, deren hohe Variabilität in Form und Intensität, sowie deren geringer Kontrast eine Herausforderung. Um das „Auslaufen“ der Segmentierung in benachbarte Strukturen zu verhindern und die fehlende Bildinformation zu kompensieren, wird Vorwissen eingesetzt. In dieser Arbeit stellen wir einen Ansatz zur Segmentierung der Risikostrukturen Cochlea, Bogengänge, innere Halsschlagader und Gesichtsnerv in Computertomographiedaten durch ein Probabilistic Active Shape Model vor. Wir testen unseren Ansatz durch Kreuzvalidierung auf 20 bis 42 Computertomographiedatensätzen (je nach Struktur). Die Ergebnisse zeigen eine gute Umrandung der oben genannten Strukturen.

Schlüsselworte: Active Shape Model, Segmentierung, minimal-invasive Chirurgie, Otobasis

1 Problemstellung

Bei Eingriffen an der Otobasis ist die Schonung von Risikostrukturen wie Gesichtsnerv oder Blutgefäße eine der größten Herausforderungen. Bisher ist der Eingriff stark traumatisch und es werden alle Risikostrukturen freigelegt, um ihre Unversehrtheit sicherzustellen. Das Forschungsprojekt MUKNO untersucht einen minimal-invasiven multi-port Ansatz für Eingriffe an der Otobasis. Hierbei sollen nur noch bis zu drei dünne Bohrkanäle von der Schädeloberfläche bis zum Operationsgebiet gebohrt werden mit dem Ziel, die Traumatisierung zu minimieren. In diesem Fall kann der Arzt die Risikostrukturen nicht mehr zur Orientierung verwenden, weshalb eine Planung der Bohrkanäle basierend auf Computertomographiedaten (CT-Daten) und eine entsprechende Navigation während des Bohrvorgangs nötig ist. Der erste Schritt hierbei ist die Segmentierung der Risikostrukturen, womit wir uns in dieser Arbeit beschäftigen.

Dabei gibt es folgende Herausforderungen: Neben der geringen Größe haben einige anatomische Strukturen eine hohe Variabilität in Form und Bildintensität sowie geringen Kontrast zur Umgebung. Beispielsweise hat die innere Halsschlagader teilweise keine sichtbare Abgrenzung zu benachbarten Strukturen in CT-Daten. Deshalb sind einfache pixelbasierte Verfahren wie Thresholding oder Regiongrowing nicht einsetzbar, wenn man keine hohe Fehlerquote oder aufwendige manuelle Korrektur in Kauf nehmen möchte. Bessere Ergebnisse versprechen wir uns durch modellbasierte Ansätze.

Im Vergleich zur Viszeralchirurgie, wo man eine Fülle von Segmentierungspublikationen finden kann, gibt es nur wenige modellbasierte Forschungsarbeiten zur Segmentierung von Strukturen der Otobasis. Voormolen et al. [1] präsentieren einen Ansatz für den Gesichtsnerv mit manueller Initialisierung durch Vorgabe des Start- und Endpunktes. Sie segmentieren die Mittellinie des Gesichtsnervs mit Active Appearance Models und legen anschließend eine tubuläre Sicherheitszone um die Mittellinie mit einem benutzerdefinierten Radius. Noble et al. haben verschiedene Ansätze entwickelt: Sie präsentieren einen atlas-basierten Ansatz für Cochlea, Bogengänge, Ossikel und äußeren Gehörgang [2]. Um speziell die Skala tympani und Skala vestibuli der Cochlea zu extrahieren, verwenden sie ein Active Shape Model, welches auf Micro CT-Daten trainiert wird [3]. Schließlich haben sie einen graph-basierten Ansatz für tubuläre Strukturen entwickelt, mit dem sie den Gesichtsnerv, die Chorda sowie die innere Halsschlagader segmentieren [4].

Das Active Shape Model [5] ist ein schnelles und genaues Segmentierungsverfahren, welches häufig und erfolgreich für medizinische Bilddaten eingesetzt wird. Das Probabilistic Active Shape Model (PASM) [6] ist eine Erweiterung dieses Ansatzes, welche eine flexiblere und spezifischere Segmentierung erlaubt. In dieser Arbeit stellen wir unseren Ansatz zur Segmentierung der Risikostrukturen Cochlea, Bogengänge, innere Halsschlagader und Gesichtsnerv in CT-Daten mit Hilfe des PASM vor. Genauer gesagt wird dabei nicht die Struktur selbst, sondern der in den CT-Daten sichtbare

knöcherne Kanal segmentiert. Wir testen unseren Ansatz durch Kreuzvalidierung auf 20 bis 42 CT-Datensätzen (je nach Struktur).

2 Material und Methoden

Das Active Shape Model [5] verwendet Vorwissen in Form eines statistischen Formmodells (SSM), welches die durchschnittliche Form einer Struktur sowie ihre Variabilität kodiert. Durch dieses Vorwissen wird das Auslaufen in benachbarte Strukturen reduziert bzw. verhindert. Für die Segmentierung wird zunächst die Durchschnittsform in das Bild platziert. Dann wird diese entsprechend der Features im Bild (wie z. B. Intensität oder Gradient) deformiert. Anschließend wird die deformierte Form wieder auf eine Form beschränkt, die plausibel im Sinne des gelernten SSM ist. Der Schritt der Deformation und der Beschränkung auf plausible Formen wird iteriert. Das Probabilistic Active Shape Model (PASM) [6] ist eine Erweiterung des Active Shape Models, welche eine flexiblere und spezifischere Segmentierung erlaubt. Im Folgenden werden die einzelnen Schritte des PASM kurz erläutert, für Details verweisen wir auf [5] und [6].

Das SSM wird anhand einer Menge T von Trainingsformen gelernt. Dabei wird jede Form i , d. h. jede anatomische Struktur, durch einen Vektor x_i dargestellt, der die aneinander gehängten 3D-Koordinaten der Punkte auf der Oberfläche dieser Struktur enthält. In einem ersten Schritt muss die Korrespondenz zwischen den $|T|$ Vektoren erstellt werden, d. h. der k -te Eintrag (= k -ter 3D Punkt) in x_i muss jeweils dem gleichen anatomischen Merkmal über alle Trainingsformen entsprechen. Dazu wird ein nicht-rigider Meshregistrierungsalgorithmus verwendet. Nach der Korrespondenzerstellung werden die Eigenvektoren der Trainingsmenge berechnet, sodass jede Form durch die Durchschnittsform plus einer Linearkombination der Eigenvektoren approximiert werden kann.

Der Algorithmus wird initialisiert, indem die Durchschnittsform des SSM in das zu segmentierende Bild platziert wird. Da das PASM ein lokales Suchverfahren ist, muss die Initialisierung ausreichend gut sein. Für die Mehrheit der Strukturen initialisieren wir das PASM durch Bestimmung der minimalen Boundingbox der Struktur. Dies kann durch manuelles Anklicken von sechs Punkten in den CT-Schichten erfolgen, die jeweils die minimale und maximale Ausdehnung in jede der drei Raumrichtungen markieren. Aufgrund der tubularen und gekrümmten Form des Gesichtsnervs, ist die Boundingbox hier als Initialisierung nicht ausreichend. Daher gibt der Arzt stattdessen drei Punkte durch Klicken in die Schicht des CT-Bildes vor: am Beginn des mastoidalen Segments, sowie am äußeren und am inneren Facialisknie. Die Punkte werden jeweils auf der Mittellinie des Gesichtsnervs gewählt.

Nach der Initialisierung wird die Durchschnittsform möglichst gut an das Bild angepasst. Zunächst werden für jeden Punkt auf der Kontur der aktuellen Form (d. h. im ersten Schritt die Durchschnittsform) mehrere Punkte entlang der Oberflächennormalen abgetastet. Die Samples werden mit einem KNN-Klassifizierer, der aus der Trainingsmenge gelernt wurde, in Hintergrund und Vordergrund klassifiziert. Daraus wird ein Fitnesswert für die Samples berechnet und die aktuelle Form entsprechend deformiert.

Für die anschließende Beschränkung der Formen verwendet das PASM (ebenso wie andere Ansätze) einen Energie-minimierungsansatz. Die Energie ist dabei wie folgt definiert: $E(x) = \alpha (E_{\text{Bild}}(x) + E_{\text{Lokal}}(x)) + E_{\text{Form}}(x)$, wobei der Parameter α die Energien gewichtet. Die Bildenergie E_{Bild} forciert eine gute Anpassung der Kontur an das Bild und die Energie $E_{\text{Lokal}}(x)$ sorgt lokal für eine glatte Kontur, da man annehmen kann, dass Organe glatt sind. Die Formenergie E_{Form} bestraft schließlich die Abweichungen von dem Formmodell.

Als Fehlermaße verwenden wir die auf der euklidischen Distanz basierenden Maße Average Symmetric Surface Distance d_{ASSD} sowie die maximale Distanz d_{max} . Sei M die Menge der Punkte, die auf der Oberfläche der manuellen Segmentierung liegen und S die Menge der Punkte auf der Oberfläche der Segmentierung des PASM. Wir definieren:

$$d_{\text{ASSD}} = \frac{1}{|M| + |S|} \left(\sum_{m \in M} \min_{s_1 \in S} \|s_1 - m\| + \sum_{s \in S} \min_{m_1 \in M} \|m_1 - s\| \right) \quad d_{\text{max}} = \max_{s \in S, m \in M} \left(\min_{s_1 \in S} \|s_1 - m\|, \min_{m_1 \in M} \|m_1 - s\| \right).$$

3 Ergebnisse

Die 42 CT-Daten, die für die Evaluation zur Verfügung standen, haben eine durchschnittliche Voxelgröße (VG) von $0.18 \times 0.18 \times 0.38 \text{ mm}^3$. Für jeden Datensatz haben wir von unseren klinischen Partnern manuell segmentierte Strukturen. Diese n Datensätze verwenden wir zum einen zum Trainieren des SSM und des lokalen Appearance Models und zum anderen als Ground Truth zur Evaluierung des PASM. Zur Validierung verwenden wir die einfache Kreuzvalidierung, d. h. wir trainieren das Modell auf $n-1$ Daten und testen den Algorithmus auf dem verbleibenden einen Datensatz. Dies wird $n-1$ mal wiederholt, sodass auf jedem Datensatz einmal getestet wurde.

Wir haben eine unterschiedliche Anzahl von Datensätzen für die verschiedenen Strukturen verwendet: Grund dafür ist, dass die Daten von zwei unterschiedlichen Personen segmentiert wurden und dadurch die manuellen Segmentierungen nicht immer exakt konsistent sind. Bei der inneren Halsschlagader haben wir nur 22 Daten zum Training und Testen verwendet, da bei der Hälfte der Daten nur der Strukturteil mit gutem Kontrast segmentiert wurde und damit nur die Hälfte der Struktur vorlag. Dies führt bei unserem Trainingsalgorithmus jedoch zu Problemen, da dadurch die Struktu-

ren zu unterschiedlich sind, um eine gute Korrespondenz zu erstellen: die Hälfte der Halsschlagader passt eben nicht zu der gesamten Halsschlagader. Wir haben das Problem pragmatisch gelöst, indem wir die 22 Datensätze verwendet haben, bei denen die gesamte innere Halsschlagader vorlag. Ähnlich wurde bei der Cochlea der Übergang zu den Bogengängen, wo es keinen Kontrast gibt, nicht immer einheitlich eingezeichnet, weshalb wir hier ebenfalls nur 20 Datensätze verwendet haben. Wir haben verschiedene Parametersettings mit der einfachen Kreuzvalidierung getestet. Die besten quantitativen Ergebnisse sind in Tabelle 1 abgebildet. Es wurde auf einem Standard Desktop PC (64-bit Windows 7, Intel Core i7 3770k, 3.5 GHz, 4 Kerne, 32 GB RAM) getestet.

	d_{ASSD} (in mm)		d_{max} (in mm)		Anzahl Daten	Zeit (in s)
	Mittelwert	SD	Mittelwert	SD		
Cochlea	0.10	0.03	1.08	0.22	20	8
Bogengänge	0.10	0.03	1.17	0.44	42	104
Gesichtsnerf	0.24	0.10	2.07	0.82	42	122
Innere Halsschlagader	0.29	0.09	2.42	0.77	22	166

Tabelle 1: Quantitative Ergebnisse des Probabilistic Active Shape Model im Vergleich zur manuellen Segmentierung. Mittelwert und Standardabweichung (SD) der Average Symmetric Surface Distance d_{ASSD} und des maximalen Fehlers d_{max} über die getesteten Daten sind angegeben.

In Abbildung 1 sind Beispiele für die qualitativen Ergebnisse der Cochlea und des Gesichtsnervs abgebildet.

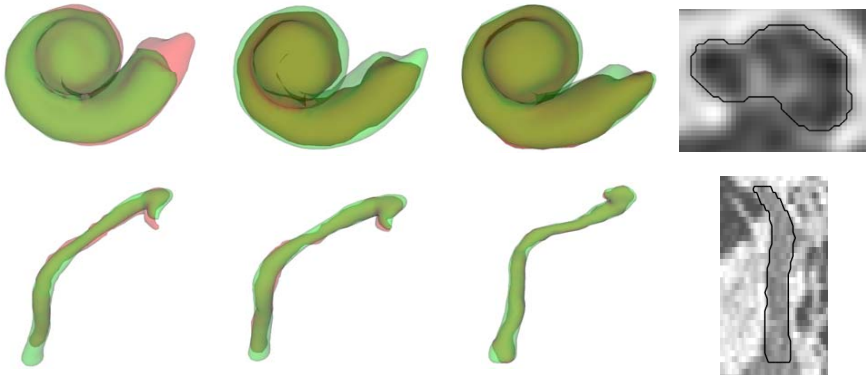


Abbildung 1: Beispiele für die Segmentierung der Cochlea sowie des Gesichtsnervs als 3D-Modell. Von links nach rechts ist ein Beispiel mit größtem, mittlerem und kleinstem maximalen Fehler abgebildet. Die manuelle Segmentierung ist in rot und das Ergebnis des PASM in grün dargestellt. Ganz rechts ist jeweils eine Schicht der CT-Daten der Segmentierung des PASM mit mittlerem maximalen Fehler abgebildet.

4 Diskussion

In Abschnitt 3 haben wir die Ergebnisse des Probabilistic Active Shape Model präsentiert. Es fällt dabei auf, dass die quantitativen Ergebnisse eine gute Average Symmetric Surface Distance aufweisen, jedoch der maximale Fehler vergleichsweise groß ist. Bei der Cochlea entsteht der maximale Fehlerwert am Übergang zu den Bogengängen, wo es keinen Kontrast gibt (s. Abb. 1). Hier weichen die Kontur von PASM und manueller Segmentierung voneinander ab, da die Kontur schwer eindeutig zu definieren ist. Der maximale Fehler liegt folglich für unseren Anwendungsfall in keinem kritischen Bereich. Ebenso liegt bei der inneren Halsschlagader der maximale Fehler an Start und Ende vor, wo ebenfalls kein Kontrast vorhanden ist. Der Gesichtsnerv ist durch seine hohe Variabilität in Form und Intensität, geringem Kontrast sowie dem umgebenden schwammartigen Knochen, die am schwierigsten zu segmentierende Struktur. Der größte Fehler entsteht am Beginn des Gesichtsnervs (im mastoidalen Segment). In den CT-Daten ist nicht erkennbar, wo der Gesichtsnerv unten beginnt und hier wird vom PASM in der Regel deutlich weiter segmentiert (s. Abb. 1). Am äußeren Facialisknie weicht die Segmentierung auch in einigen Fällen von der manuellen Segmentierung ab, da es auch hier

wenig Kontrast gibt. Es bleibt zu untersuchen, wie kritisch diese Abweichung ist. Bewertet man die Ergebnisse visuell, so sieht man im Allgemeinen eine gute Segmentierung der Risikostrukturen.

Ein direkter Vergleich der Ergebnisse mit den Resultaten von Noble et al. oder Voormolen et al. ist schwierig, da wir nicht auf den gleichen Daten trainiert und getestet haben. Vergleicht man die quantitativen Werte dennoch, so erzielen wir bei Cochlea und Bogengängen einen besseren d_{ASSD} -Wert und einen ähnlichen globalen maximalen Fehler wie Noble et al. [2], sie haben jedoch schlechter aufgelöste Daten (Noble: $d_{ASSD} = 0.42$ mm, $\max(d_{max}) = 2.62$ mm, $VG = 0.35^2 \times 0.4$ mm³). Bei der inneren Halsschlagader haben wir einen kleineren d_{ASSD} -Wert und einen ähnlichen mittleren maximalen Fehler, wobei wir deutlich kleinere Voxelgrößen haben (Noble: $d_{ASSD} \approx 0.75$ mm, $d_{max} \approx 2.6$ mm, $VG = 1^2 \times 1.5$ mm³) [4]. Wir haben hier keine exakten Angaben, da die Werte von einem Graph abgelesen sind. Mögliche Gründe für den relativ großen maximalen Fehler bei unseren Segmentierungen sind oben erläutert. Für den Gesichtsnerv haben Voormolen et al. [1] deutlich bessere Werte ($d_{max} = 0.42$ mm, $VG = 0.31^2 \times 1$ mm³), segmentieren allerdings auch nur die Mittellinie des Gesichtsnervs. Im Vergleich zu Noble et al. [4] erreichen wir bei dem Gesichtsnerv einen ähnlichen d_{ASSD} -Wert, erhalten jedoch einen größeren maximalen Fehler (Noble: $d_{ASSD} \approx 0.2$ mm, $d_{max} \approx 0.75$ mm, $VG = 0.3^2 \times 0.4$ mm³). Dies könnte wie oben beschrieben zum einen daran liegen, dass das PASM in vielen Fällen am Anfang des Gesichtsnervs deutlich weiter segmentiert als die manuelle Segmentierung (s. Abb. 1). Zum anderen betrachten wir auch den Teil hinter dem äußeren Facialisknie, was eine weitere Herausforderung darstellt.

Momentan beschäftigen wir uns mit der Segmentierung der Chorda tympani, der Ossikel sowie des inneren und äußeren Gehörgangs. Ferner ist es interessant, die bisherigen Segmentierungen intensiver zu untersuchen, z. B. eine gemeinsame Fehlerbestimmung für Cochlea und Bogengänge durchzuführen und eine Untersuchung der Abhängigkeit der Gesichtsnervsegmentierung von der Initialisierung. Um den Algorithmus weiter zu automatisieren, ist eine robuste Detektion der Boundingbox der Strukturen wünschenswert.

5 Zusammenfassung

Das Forschungsprojekt MUKNO untersucht einen minimal-invasiven multi-port Ansatz für Operationen an der Otobasis. Ein wichtiger Schritt dabei ist die Segmentierung der Risikostrukturen. In dieser Arbeit haben wir einen Ansatz zur Segmentierung der Risikostrukturen Cochlea, Bogengänge, innere Halsschlagader und Gesichtsnerv in CT-Daten mit dem Probabilistic Active Shape Model präsentiert. Die Kreuzvalidierung auf bis zu 42 Daten zeigt eine gute Umrandung der Risikostrukturen.

6 Danksagungen

Wir danken unseren klinischen Partnern der HNO-Klinik des Universitätsklinikums Düsseldorf für die manuelle Segmentierung der Datensätze. Das Projekt wurde gefördert von der Deutschen Forschungsgemeinschaft: FOR 1585; FE 431/13-1.

7 Referenzen

- [1] E. H. J. Voormolen et al., *Determination of a facial nerve safety zone for navigated temporal bone surgery*, *Neurosurgery*, **70(1 Suppl Operative)** 50-60 (2012).
- [2] J. H. Noble, R. F. Labadie, O. Majdani, B. M. Dawant, *Automatic Identification and 3D Rendering of Temporal Bone Anatomy*, *Otology & Neurotology*, **30(4)** 436-442 (2009).
- [3] J. H. Noble, B. M. Dawant, F. M. Warren, R. F. Labadie, *Automatic segmentation of intracochlear anatomy in conventional CT*, *IEEE Transactions on Biomedical Engineering*, **58(9)** 2625-2632 (2011).
- [4] J. H. Noble, B. M. Dawant, *A new approach for tubular structure modeling and segmentation using graph-based techniques*, *MICCAI 2011*, LNCS 6893, 305-312, Springer, Berlin Heidelberg, (2011)
- [5] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, *Active shape models - their training and Application*, *Comput. Vis. Image Underst.*, **61(1)** 38-59 (1995)
- [6] M. Kirschner, S. Wesarg, *Active Shape Models Unleashed*, *Proceedings of SPIE*, **7962(11)** 1-9 (2011)