Verfolgung und Analyse von Stimmlippenkonturen in stroboskopischen Videosequenzen

A. Sovakar, I. Scholl, Ch. Neuschaefer-Rube, Th. Lehmann und K. Spitzer

Institut für Medizinische Informatik und Biometrie,
Klinik für Phoniatrie und Pädaudiologie, RWTH Aachen
abi@vaire.imib.rwth-aachen.de
scholl@imib.rwth-aachen.de

Zur Diagnostik funktioneller Stimmstörungen wird das Schwingungsverhalten der Stimmlippen lupenlaryngostroboskopisch untersucht und videodokumentiert. Die Beurteilung der einzelnen Schwingungsparameter basiert derzeit auf der subjektiven Beobachtung des jeweiligen Untersuchers. Unter der Zielvorgabe die Stimmlippenschwingungen stimmgestörter Patienten untersucherunabhängig quantifizieren zu können, wird hier ein digitales Bildverarbeitungsverfahren vorgestellt, das die Konturen der Stimmlippen mit einem erweiterten aktiven Konturmodell aus den laryngoskopischen Bildsequenzen extrahiert und verfolgt. Basierend auf den detektierten Konturen werden anschließend Parameter für die Schwingungsanalyse extrahiert.

Keywords: Aktive Konturmodelle, Konturverfolgung, funktionelle Stimmstörung, Stimmlippen, Bildsequenzen

1 Einleitung

Die Kommunikationsfähigkeit des Meschen kann durch eine Stimmstörung erheblich beeinträchtigt werden. Dies ist bei Personen in Sprechberufen, wie LehrerInnen oder SängerInnen von besonderer Bedeutung. Beim Auftreten von Mißempfindungen im Kehlkopf und bei Veränderungen des Stimmklangs bis zur völligen Stimmlosigkeit werden Untersuchungen des Kehlkopfes durchgeführt, um Veränderungen der Stimmlippenbewegung und -schwingung festzustellen

Zur endoskopischen Beurteilung der Kehlkopfstrukturen und zur Beobachtung der respiratorischen Öffnungs- sowie der phonatorischen Schließbewegungen der Stimmlippen werden seit zwei Jahrzehnten starre Optiken (Lupenlaryngoskope) und zunehmend auch flexible Fiberglasoptiken eingesetzt. Zur Schwingungsbeobachtung werden die Endoskope mit einer Stroboskoplichtquelle gekoppelt. Mittels einer CCD- Kamera wird die Stimmlippenbewegung- und schwingung videodokumentiert und anschließend in PAL-Auflösung digitalisiert.

Während die endoskopische Untersuchung weitgehend standardisiert durchgeführt wird, erfolgt die Auswertung der Befunde in der klinischen Routine auf der Grundlage überwiegend subjektiver Beschreibungsparameter. Befunde, die von unterschiedlichen Untersuchern oder zu verschiedenen Zeitpunkten erhoben werden, sind nur eingeschränkt vergleichbar. In dem hier zugrundeliegendem Forschungsprojekt *Quantitative Digitale Laryngoskopie* wird ein Bildanalyseverfahren entwickelt, das eine quantitative Beschreibung der Stimmlippenbeweglichkeit, ihres Schwingungsverhaltens und der morphologischen Beschaffenheit des Endolarynx ermöglichen.

In diesem Beitrag wird zur Analyse des Schwingungsverhaltens ein aktives Konturmodell zur Konturverfolgung der linken und rechten Stimmlippe eingesetzt, das im folgenden Abschnitt 2 vorgestellt wird. Dieses Modell wurde durch eine adaptive Gewichtung und durch eine Abstoßkraft erweitert. Ergebnisse der Konturanpassung und -verfolgung werden im Abschnitt 3 beschrieben. Eine abschließende Diskussion und ein Ausblick des vorgestellten Verfahrens werden in Abschnitt 4 gegeben.

2 Material und Methode

2.1 Konturverfolgung

Zur Untersuchung einer stroboskopischen Schwingungssequenz wird das Videobild mit der weitesten Glottisöffnung als Startbild definiert. In diese Bild werden interaktiv grobe Startkonturen zu der linken und rechten Stimmlippe vorgegeben werden (Abb. 2 links). Im nächsten Schritt werden die Startkonturen iterativ mit einem aktiven Konturmodell (Snakes) verfeinert und den realen Stimmlippenrändern angepaßt (z.B. Abb. 2c). Die angepaßten Konturen werden als grobe Startkonturen für das nächste Einzelbild der Sequenz verwendet.

2.2 Konturanpassung

Die bei der Konturanpassung stattfindende Bewegung der einzelnen Konturknoten basiert auf einem Energieminimierungsprozeß nach einem *aktiven Konturmodell* [4, 5]. Die Kontur besteht aus Knoten, die über Kanten miteinander verbundenen sind und deren Positionen einen Zustand definieren. Aus unterschiedlichen Zuständen (andere Form der Kontur) lassen sich Energieunterschiede ∇E ableiten, aus denen wiederum Kräfte ermittelt werden können, deren Richtung und Betrag die Bewegung des Knotens beeinflussen und somit die Form der Kontur verändern. Das Ziel ist es, einen Zustand zu finden, in dem die Energie minimal wird. Die Kraft setzt sich aus vier Kompenenten zusammen, welche unterschiedliche Eigenschaften der Kontur modellieren und verschiedenen Energien zugeordnet sind.

$$\mathbf{f} = \mathbf{w}_{\text{int}}\, \mathbf{f}_{\text{int}} + \mathbf{w}_{\text{ext}}\, (\mathbf{w}_{\text{img}}\, \mathbf{f}_{\text{img}} + \mathbf{w}_{\text{usr}}\, \mathbf{f}_{\text{usr}}) + \mathbf{f}_{\text{damp}} \tag{1}$$

1. Die interne Kraft $\mathbf{f}_{\text{int}}$ minimiert die lokale Krümmung der Kontur und glättet damit die Kontur. Je glatter die Kontur ist, desto geringer ist die *interne Energie*.

2. Eine Anpassung an Bildeigenschaften, wie Kanten oder Linien wird durch die Bildkraft $\mathbf{f}_{\text{img}}$ erreicht. Dazu werden Energiebilder I_E verwendet. Bei der Stimmlippenverfolgung werden Kanten gesucht, so daß hier Gradientenoperatoren benutzt werden. Folgt die Kontur genau der Kante, so wird die *Bildenergie* minimal. Dabei entsprechen lokale Maximas im Gradientbild lokalen Minimas in der Energieverteilung, welche durch das Energiebild bestimmt ist: $I_{\text{grad}} = -I_E$.

3. Die User-Kraft $\mathbf{f}_{\text{usr}}$ ist eine frei definierbare Komponente, die es ermöglicht, zusätzliche Maßnahmen für die Konturanpassung einzuleiten. Für die Anwendung der Stimmlippenverfolgung wurde eine lokale Abstoßkraft definiert, welche durch den Abstand eines Knotens zur gegenüberliegenden Stimmlippenkontur bestimmt wird.

4. Die Geschwindigkeit der einzelnen Knoten während der Konturanpassung wird durch eine Dämpfungskraft $\mathbf{f}_{\text{damp}}$ gebremst. Dies bewirkt eine Stabilisierung des Systems und beschleunigt die Konvergenz.

Die Bildenergie und die Userenergie sind "externe" Eigenschaften, während die interne Energie aus der Form der Kontur selbst ermittelt wird. Deshalb werden die ersteren zusammengefaßt zur externen Energie, welche durch w_{ext} eine eigene Gewichtung erhält (siehe Gl. 1).

Die Gewichtung der einzelnen Kräfte ($w_{\text{int}}, w_{\text{img}}, w_{\text{usr}} \in [0,1]$ und $w_{\text{damp}} \in [-1,0]$) betont die verschiedenen Eigenschaften. Die Parameterwahl ist kritisch für das Gelingen der Konturverfolgung. Eine Vereinfachung wurde durch eine adaptive Gewichtung erreicht, die nach zwei Prinzipien gesetzt wird. a) Bereiche im Gradientenbild I_{grad} mit starken Gradienten tragen eine verläßliche Information über die Konturform, so daß in diesem Fall das Gewicht zugunsten der externen Kraft verlagert wird: $w_{\text{int}} = 1 - w_{\text{ext}}$. b) Eine hohe Abstoßkraft bekommt höhere Priorität über die interne Kraft und die Bildkraft. Diese Prinzipien werden durch eine lineare Funktion modelliert, die nur noch von w_{usr} und dem normierten Gradienten x abhängt:

$$w_{\text{img}} = 1 - w_{\text{usr}} \tag{2}$$

$$w_{\text{ext},\lambda}(x, w_{\text{usr}}) = (\lambda - w_{\text{usr}}\lambda)\, x + (w_{\text{usr}} - \frac{1}{2})\lambda + \frac{1}{2} \tag{3}$$

$$\text{mit}\, x, w_{\text{usr}}, \lambda \in [0,1]\, \text{und}\, \lambda\, \text{konstant}$$

Die Abstoßkraft, welche anhand der Distanz *dist* bestimmt wird, setzt sich aus zwei Teilen zusammen, wovon der eine für die Wichtung w_{usr} verwendet wird:

$$w_{\text{usr}}(dist) = l_w/(l_w + dist^2) \tag{4}$$

$$\mathbf{f}_{\text{usr}}(dist) = k \cdot l_f/(l_f + dist) \tag{5}$$

Der Wert $l_w > 0$ bestimmt dabei den Einfluß der Abstoßkraft auf die Gewichtung, während $k, l_f > 0$ die Stärke der Abstoßkraft modulieren. Durch l_f wird gewährleistet, daß $\mathbf{f}_{\text{usr}}(0) = k$ ist.

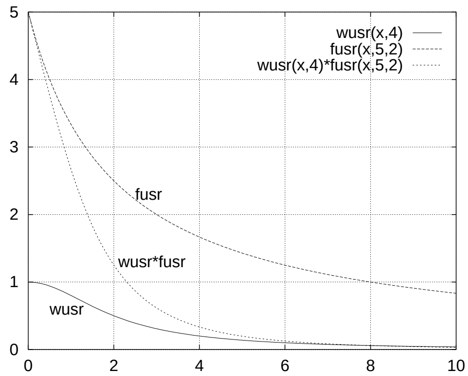

Abb. 1: Die Abstoßkraft besteht aus den Komponenten w_{usr} und $\mathbf{f}_{\text{usr}}$, welche miteinander multipiziert werden (1).

3 Ergebnisse

Als Testsequenz wurde eine stroboskopische Videolaryngoskopiesequenz digitalisiert. Sie umfaßt eine Dauer von 4,4 Sekunden und besteht insgesamt aus 111 Bildern. Zur Ausrichtung der Einzelbilder mit der Kepstrum–Technik [3] wurden quadratische Bildausschnitte

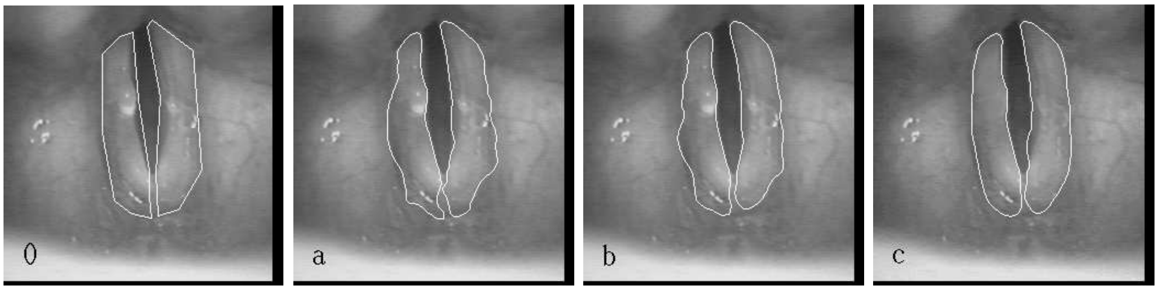

Abbildung 2: (0) Initialkonturen im Startbild; Konturanpassung: (a), (b) ohne adaptive Gewichtung mit (a) $w_{\text{int}} = w_{\text{ext}} = 0.5$, (b) $w_{\text{int}} = 0.7$, $w_{\text{ext}} = 0.3$; (c) mit adaptiver Gewichtung ($\lambda = 0.9$).

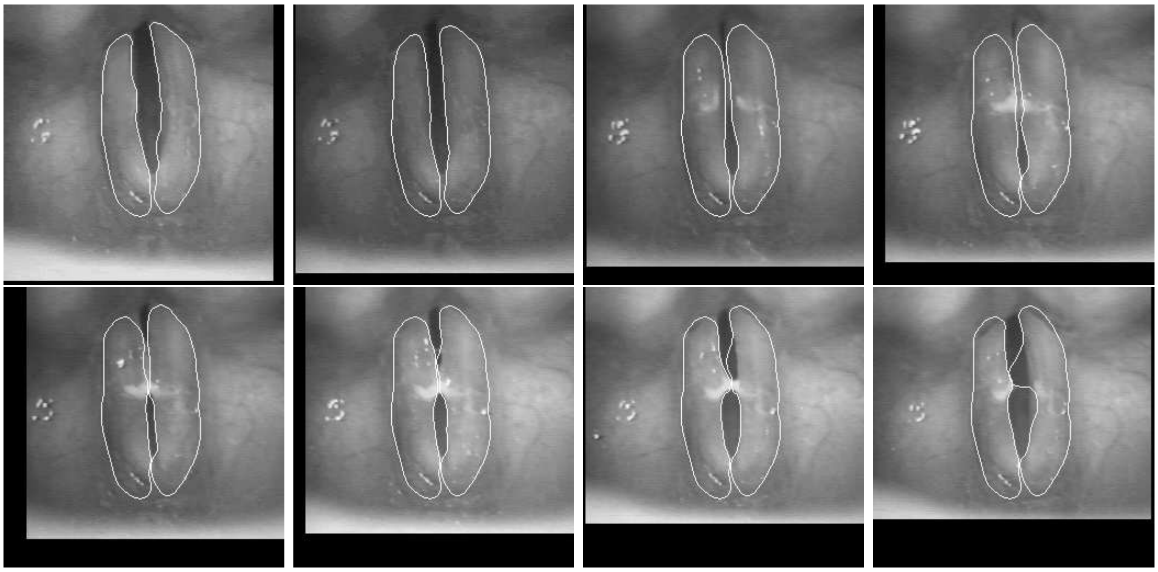

Abbildung 3: Konturverfolgung mit adaptiver Gewichtung

der Größe 256×256 aus der Originalsequenz extrahiert und in Grauwertbilder umgewandelt.

Im folgenden werden die Ergebnisse der adaptiven Gewichtung und der Abstoßkraft dargestellt. Abbildung 2 zeigt die Auswirkung der adaptiven Gewichtung für die Konturanpassung im ersten Bild der Testsequenz. Die Konturen in 2(a) und (b) wurden ohne und die in (c) mit adaptiver Gewichtung angepaßt. Mit steigender interner Kraft werden die Konturen glatter, die Form der Kontur wird dadurch runder und verläßt die *wahren* Konturen. Die adaptive Gewichtung bewirkt einen Kompromiß zwischen diesen Kräften und sorgt dafür, daß am lateralen Stimmlippenrand die interne Kraft einen höheren Einfluß hat und am freien Stimmlippenrand, wo starke Gradienten vorhanden sind, das Gewicht auf Seiten der externen Kraft liegt. Dadurch bleibt die Kontur am lateralen Rand stabil während der freie Rand seine Bewegungsfreiheit beibehält.

Die Abbildungen 4 und 3 zeigen die Auswirkungen zweier Konturverfolgungen mit und

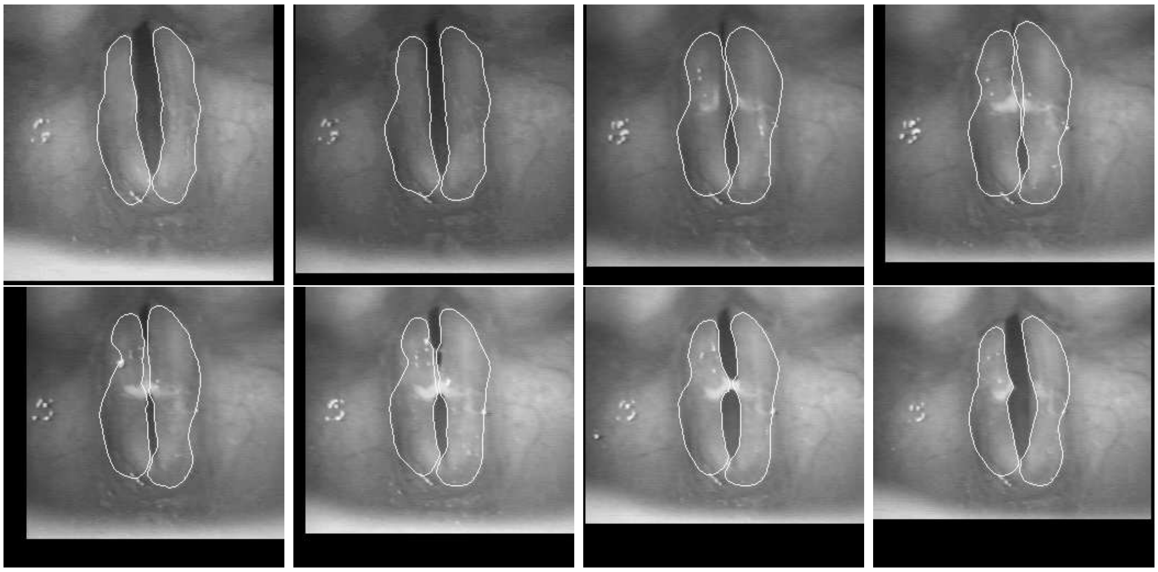

Abbildung 4: Konturverfolgung mit adaptiven Gewichtung und Abstoßkraft

ohne der Abstoßkraft. Die Konturanpassung wurde dabei für beide Sequenzen mit adaptiver Gewichtung berechnet. In der Sequenz 3 zeigt sich während der Öffnungsphase der Stimmlippenschwingung keine vollständige Trennung der linken von der rechten Kontur, die durch die Abstoßkraft erzielt werden kann (Abb. 4).

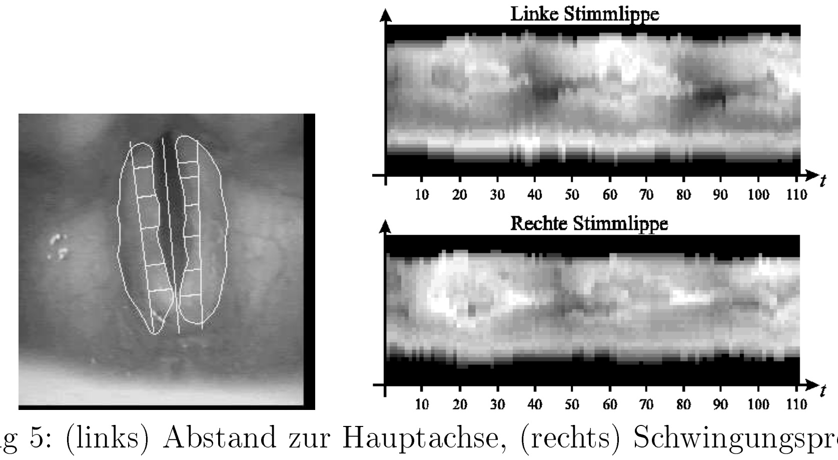

Abbildung 5: (links) Abstand zur Hauptachse, (rechts) Schwingungsprofil-Bilder.

Auf Grund der extrahierten Stimmlippenkonturen lassen sich Parameter zur Beurteilung der Stimmlippenschwingung berechnen. Dazu wurden im Startbild die Hauptachsen der Konturflächen mit der Karhunen-Loève-Transformation [6] berechnet (Abb. 5(links)). Als Parameter wurden entlang den Hauptachsen die Abstände zu den freien Stimmlippenrändern (Konturen des Glottisspaltes) bestimmt und als Grauwerte dreidimensional dargestellt (Abb. 5 (rechts)).

Kleine Abstände wurden auf einen dunklen, große Abstände auf einen hellen Grauwert abgebildet. Dunkle Bereiche lassen auf einen geöffneten Zustand der Stimmlippe schließen.

Desweiteren wurden Parameter wie Umfang, Fläche, Schwerpunkt der Konturen und mittlerer euklidischer Abstand zwischen den freien Stimmlippenrändern und der Hauptachse der linken bzw. rechten Stimmlippenkontur berechnet (Abb. 11). Anhand dieser Parameter kann eine nicht symmetrische Stimmlippenschwingung beobachtet werden, die für eine funktionelle Stimmstörung symptomatisch ist.

4 Zusammenfassung und Ausblick

Es zeigt sich, daß Aktive Konturmodelle für die Konturverfolgung der Stimmlippen in stroboskopischen Videosequenzen grundsätzlich geeignet sind. Mittels adaptiver Gewichtung wird eine höhere "Anpassungsfähigkeit" der Konturen erreicht und die zusätzliche definierte Abstoßkraft sorgt für eine korrekte Verfolgung. Adaptive Gewichtung und Abstoßkraft tragen zur Robustheit der Verfahrens bei. Auf Basis der ermittelten Konturen können verschiedene Parameter extrahiert werden, die globale und lokale Aussagen über die Stimmlippenschwingungen ermöglichen. Insbesondere bietet das Schwingungsprofil-Bild neue Möglichkeiten. Es erlaubt erstmalig einen Überblick über die Schwingung über eine Videosequenz und kann als Basis für weitere Parameter dienen, welche für die Diagnostik funktioneller Stimmstörungen sehr wertvoll sind.

Für einen routinemäßigen Einsatz muß das Verfahren allerdings noch weiter verbessert werden. Aufgrund mangelnder Qualität der Grandientenbilder schrumpft die Kontur am lateralen Stimmlippenrand (Rand, der nicht im Glottisbereich ist). Aus diesen Gründen werden zukünftig verstärkt weitere Verfahren für die Gewinnung der Energiebilder untersucht, welche das Vefahren besser stützten. Bei der adaptiven Gewichtung wird die Dämpfung, die für die Robustheit des Verfahrens entscheidend beiträgt, noch nicht beachtet, so daß hier noch Forschungbedarf herrscht. Ein weiterer Forschungsschwerpunkt bildet die Verbesserung des Schwingungsprofil-Bildes und weitergehende Parameterextraktion aus dem selben.

Literatur

[1] Böhme, G.: Klinik der Sprach-, Sprech- und Stimmstörungen, Stuttgart: G. Fischer Verlag 1983.

[2] Wirth, G.: Stimmstörungen. Lehrbuch für Ärzte, Logopäden, Sprachheilpädagogen und Sprecherzieher. Deutscher Ärzte-Verlag GmbH, Köln, 3. Auflage, 1991

[3] Lehmann , T.; Goerke, C.; Schmitt, W.; Repges, R.: Rotations- und Translationsbestimmung durch eine erweiterte Kepstrum-Technik. In: Sagerer, G.; Posch, S.; Kummert, K.(Hrsg.): Mustererkennung 1995, 17. DAGM- Symposium, Berlin: Springer, 395-402, 1995.

[4] Lobregt, S.; Viergever, M.A.: A discrete dynamic contour model. IEEE Trans. on Medical Imaging, Vol. 14, no. 1, pp. 12-24, 1995.

[5] Scholl, I.; Sovakar, A.; Spitzer, K.: Konturverfolgung der Stimmlippen in laryngoskopischen Bildsequenzen. In 41. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) e.V., 1996

[6] Gonzales, R.G.; Woods, R.E.: Digital Image Processing, Addison-Wesley 1992.