Wissensbasierte Bewegungskompensation in aktiven Konturmodellen

C. Palm, C. Neuschaefer-Rube*, T. Lehmann, K. Spitzer

Institut für Medizinische Informatik

*Klinik für Phoniatrie und Pädaudiologie
Rheinisch-Westfälische Technische Hochschule (RWTH), 52057 Aachen
Email: cpalm@imib.rwth-aachen.de

Zusammenfassung. Zur Analyse von Lippenbewegungsabläufen wird ein aktives Konturmodell eingesetzt. Probleme bereitet die hohe Sprechgeschwindigkeit, die in starken Objektverschiebungen resultiert und bislang nicht durch eine alleinige Konturanpassung kompensiert werden kann. In diesem Beitrag werden die klassischen aktiven Konturmodelle um eine Vorjustierung der Grobkonturen erweitert, die eine energiebasierte Konturanpassung erst möglich macht. Die Schätzung der Verschiebung zur Vorjustierung basiert auf dem Gradientenbild und einem prädikatenlogisch formulierten Regelwerk, das Annahmen und Nebenbedingungen als Wissensbasis enthält. Mit Hilfe dieser Erweiterungen ist eine automatisierte Konturverfolgung der Lippen möglich.

Schlüsselwörter: Aktives Konturmodell, Prädikatenlogik, Bewegungsschätzung

1 Einleitung

Die Farbvideotechnik wird zur Dokumentation von physiologischen und pathologischen Lippenbewegungsabläufen eingesetzt. Mit Hilfe von Bildverarbeitungsmethoden ist es möglich, durch eine automatische Lippensegmentierung eine Analyse dieser Bewegungsabläufe durchzuführen. Versuche zur Munddetektion- und verfolgung sind bislang vornehmlich in der Gesichtserkennung (z.B. in Sicherungs- oder Spracherkennungssystemen [1]) bekannt. Dabei werden deformierbare Templates verwendet, die der Lokalisation der Lippen im Gesicht und als Formbeschreibung dienen [2]. Medizinische Anwendung fand diese Technik bei der Untersuchung von Gesichtslähmungen (Fazialisparesen) [3].

Die Analyse von Lippenbewegungsabläufen erfordert andere Schwerpunkte, die weniger in der Formdeskription als in der pixelgenauen Konturerkennung liegen. Diese kann durch Einsatz von kantenbasierten aktiven Konturmodellen (Snakes) erreicht werden, die allerdings einen starken Gradienten und eine präzise Grobkontur voraussetzen. Bei langsamen Bewegungen kann die Ergebniskontur des vorherigen Bildes als aktuelle Grobkontur verwendet werden. Sie wird mit Hilfe von Energieminimierungsstrategien an die veränderte Lippenposition angepaßt.

Bei der Analyse schneller Sprechbewegungen ergeben sich hingegen Schwierigkeiten bei der Lippendetektion, da bereits in aufeinanderfolgenden Einzelbildern große Positionsänderungen der Lippen auftreten. Diese können mit den bisher eingesetzten Verfahren zur aktiven Konturanpassung nicht ausreichend erfaßt werden. In diesem Beitrag wird ein neuartiges Konzept vorgestellt, das Bild– und Bewegungsinformationen in Form einer prädikatenlogischen Wissensbank zur Vorjustierung der Grobkonturen einsetzt, um die Konturverfolgung den Erfordernissen schneller Bewegungsvorgänge anzupassen.

2 Aktives Konturmodell

Ein aktives Konturmodell [4] besteht aus Knoten und Kanten mit variabler Länge. Mit Hilfe einer iterativen Optimierung werden die Knoten einer groben Kontur so verschoben, daß ein Energieterm minimiert wird. Dieser Energieterm besteht aus einer gewichteten Summe: Die *interne Energie* dient der Glättung und nimmt bei einer Kreisform ihr Minimum an. Die *externe Energie* basiert auf dem (Farb-)Gradienten des Bildes, der dort hohe Werte zeigt, wo Grauwert– oder Farbsprünge auftreten. Da zwei Konturen, die sich schließen und wieder öffnen, quasiparallel verfolgt werden sollen, wird eine Abstoßenergie modelliert, die bei geringen Konturabständen zu einer Abstoßung führt, ohne ein echtes Schließen zu verhindern [5].

3 Bewegungsschätzung

Zur Bewegungsschätzung werden die Lippen orientierend lokalisiert, indem im aktuellen Gradientenbild die Zuordnung der Gradientenmaxima zu den Objekträndern erfaßt wird. Zur Vereinfachung wird das Problem auf den 1D-Fall zurückgeführt, indem die Werte des Gradienten entlang einer Scanlinie aufgetragen und dort lokale Maxima detektiert werden. Die Scanlinie verläuft durch den Schwerpunkt beider Konturen in Hauptbewegungsrichtung orthogonal zur Hauptachse. Dennoch ist die Maximazuordnung zu den Lippenkanten problematisch, weil durch Risse auf den Lippen, Glanzlichter usw., lokale Maxima entstehen, die nicht den Lippenrändern zuzuordnen sind. Desweiteren reduziert sich die Zahl der echten Randmaxima bei geschlossenen Lippen von vier auf drei (Abb. 1).

Zur Lösung wurde ein wissensbasiertes Regelwerk geschaffen, in das Annahmen über konstante Lippenfarbe und –breite während der Bewegung ebenso einfließen wie die Nebenbedingung der Überschneidungsfreiheit. Vorteil einer solchen Wissensbank ist die logische Nachvollziehbarkeit des einfließenden Wissens und die Anpassungsfähigkeit an andere Problemstellungen. Die Verzahnung erfolgt durch Anfragen der Wissensbank an die Bildverarbeitung, die in Form von Prädikaten antwortet. Durch logische Schlüsse und Variablenbindung erfolgt die gesuchte Zuordnung der Maxima zu den Objektkanten.

Zur Einführung der Prädikate gehen wir von diskreten Scanlinien S aus, die mit Hilfe der *Vorgänger*-Relation in eine Reihenfolge gebracht werden:

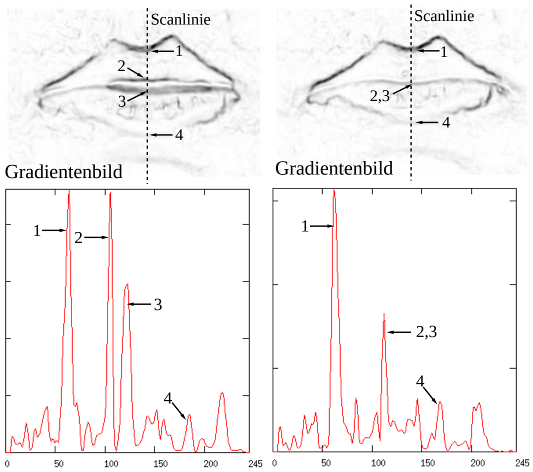

Abb. 1. Lokale Maxima auf einer Scanlinie bei offenem (links) und bei geschlossenem Mund (rechts), die den Lippenrändern zuzuordnen sind. Die Scanlinie verläuft orthogonal zur Hauptachse der vorherigen Kontur

- $iS(S)$ [*ist Scanlinie*]
- $iVS(S_1, S_2)$ [*ist Vorgänger-Scanlinie*]

Die detektierten lokalen Maxima werden mit der entsprechenden Scanlinie verbunden und über die *liegt-unter*-Relation zu Paaren verknüpft. Diese dienen letztlich als vermutete Ränder der Ober– bzw. Unterlippe:

- $ilMS(X, S)$ [*ist lokales Maximum auf Scanlinie*]
- $lu(X, Y)$ [*liegt unter*]
- $iPuM(X_1, X_2, S_2)$ [*ist Paar untereinanderliegender Maxima*]
- $vOL/UL(X_1, X_2, S_2)$ [*vermutete Ober(Unter-)lippe*]

Eine solche Vermutung ist unter folgenden Bedingungen gerechtfertigt:

1. Die mittlere Farbe der Lippen bleibt während der Bewegung konstant (Farbinvarianz). Dazu testet die Bildverarbeitung den Farbunterschied der gemittelten Farbe zwischen jeweils zwei Punkten auf zwei Scanlinien:
   - $gmFaD(X_1, X_2, S_2, Y_1, Y_2, S_1)$ [*geringe mittlere Farbdifferenz*]
   - $FaIg_{OL/UL}(X_1, Y_1, S_1, X_2, Y_2, S_2)$ [*Farbinvarianz Ober-/Unterlippe gilt*]
2. Die Vertikalausdehnung der Lippen bleibt während der Bewegung konstant (Forminvarianz). Die Differenz der paarweisen Abstände von Punkten auf der Scanlinie wird über ein Prädikat an die Wissensbank übergeben:
   - $gpAd(X_1, X_2, Y_1, Y_2)$ [*geringe paarweise Abstandsdifferenz*]
   - $FoIg_{OL/UL}(X_1, X_2, Y_1, Y_2)$ [*Forminvarianz Ober-/Unterlippe gilt*]
3. Die beiden Konturen überlappen sich nicht:
   - $kUe(X_1, X_2, X_1', X_2')$ [*kein Überlapp*]

Aus den verbleibenden Hypothesen werden nun diejenigen als faktische Lippenbegrenzungen ausgewählt, für die die erforderliche Verschiebung minimal wird:

- $hV(X_1, X_2, Y_1, Y_2, V)$ [*hat Verschiebung*]
- $fLb(X_1, X_2, X_1', X_2', S)$ [*faktische Lippenbegrenzungen*]

Die Grundlage der Inferenz bilden dann die Fakten und Regeln der Wissensbank:

$$iS(S_1); \qquad iS(S_2); \qquad iVS(S_1,S_2); \tag{1}$$

$$iPuM(X_1,X_2,S_2) \longleftarrow \\ ilMS(X_1,S_2) \ \wedge \ ilMS(X_2,S_2) \ \wedge \ lu(X_1,X_2) \tag{2}$$

$$FaIg_{OL}(X_1,X_2,S_2) \longleftarrow gmFaD(X_1,X_2,S_2,Y_1,Y_2,S_1) \\ \wedge \ fLb(Y_1,Y_2,Y_1',Y_2',S_1) \ \wedge \ iPuM(X_1,X_2,S_2) \tag{3}$$

$$FoIg_{OL}(X_1,X_2,S_2) \longleftarrow gpAd(X_1,X_2,Y_1,Y_2) \\ \wedge \ fLb(Y_1,Y_2,Y_1',Y_2',S_1) \ \wedge \ iPuM(X_1,X_2,S_2) \tag{4}$$

$$FoIg_{UL}(X_1',X_2',S_2), \qquad FaIg_{UL}(X_1',X_2',S_2) \qquad \text{dito.} \tag{5}$$

$$vOL(X_1,X_2,S_2) \longleftarrow FaIg_{OL}(X_1,X_2,S_2) \ \wedge \ FoIg_{OL}(X_1,X_2,S_2) \tag{6}$$

$$vUL(X_1',X_2',S_2) \longleftarrow FaIg_{UL}(X_1',X_2',S_2) \ \wedge \ FoIg_{UL}(X_1',X_2',S_2) \tag{7}$$

$$kUe(X_1,X_2,X_1',X_2') \longleftarrow \\ vOL(X_1,X_2,S_2) \ \wedge \ vUL(X_1',X_2',S_2) \ \wedge \ lu(X_2,X_1') \tag{8}$$

$$fLb(X_1,X_2,X_1',X_2',S_2) \longleftarrow kUe(X_1,X_2,X_1',X_2') \tag{9}$$
$$\wedge\big[\neg\,\exists\, X_1'',X_2''\,[hV(X_1'',X_2'',Y_1,Y_2,V_1) \ \wedge \ hV(X_1,X_2,Y_1,Y_2,V_2) \ \wedge \ V_1 < V_2]\big]$$
$$\wedge\big[\neg\,\exists\, X_1''',X_2'''[hV(X_1''',X_2''',Y_1',Y_2',V_3) \wedge hV(X_1',X_2',Y_1',Y_2',V_4) \wedge V_3 < V_4]\big]$$

Regel 9 liefert mit der Bindung von X_1, X_2, X_1' und X_2' die Lokalitäten der besten Hypothese für die Lippenkanten auf der aktuellen Scanlinie. Hierdurch kann die Kontur vorab in der Nähe der richtigen Gradientenpeaks positioniert werden.

4 Vorjustierung

Geht man von einer nicht pathologisch veränderten Lippenbewegung aus, so ist das Maximum der Bewegung, das zum Rand hin abnimmt, im Schwerpunkt beider Konturen zu finden. Die geschätzte Verschiebung auf der Scanlinie (Abschn. 3) stellt eine gute Näherung für dieses Maximum dar. Die seitlichen Ränder beider Lippen müssen sich immer berühren, so daß zum Rand hin die Bewegung gegen Null geht. Findet hier dennoch eine Verschiebung statt, so ist diese auf eine Kopfbewegung des Probanden zurückzuführen, die sich ebenso auf den Gesamtschwerpunkt auswirkt. Solche Kopfbewegungen können durch Überlagerung des Startkonturschwerpunktes und des durch Mittelung der Scanlinienmaxima geschätzten Schwerpunktes eliminiert werden.

Vereinfachend wird die Gesamtverschiebung der Lippen gemäß einer doppelten Exponentialfunktion modelliert, mit der sich die Verschiebung der Kontur zum Rand hin verringert. Sei $v_{x,y}^{\max}$ die mit Hilfe der Scanlinie bestimmte maximale Verschiebung, $(x_{\text{alt}}, y_{\text{alt}})$ der zu justierende Knoten der oberen bzw. unteren Kontur und a die Länge dessen Projektionsvektors auf die gemeinsame Hauptachse beider Konturen. Dann ergeben sich neue Koordinaten $(x_{\text{neu}}, y_{\text{neu}})$ aus

$$x_{\text{neu}} = \eta(a) \cdot v_x^{\max} \pm x_{\text{alt}} \qquad \text{bzw.} \qquad y_{\text{neu}} = \eta(a) \cdot v_y^{\max} \pm y_{\text{alt}} \tag{10}$$

Abb. 2. Konturverfolgung aufeinanderfolgender Bilder einer Sequenz.

$$\eta(a) = \exp\left(- \left| \frac{\ln(M)}{HL} \cdot a \right| \right) \tag{11}$$

bezeichnet den relativen Verschiebungsanteil, HL die Länge der Hauptachse und M die minimale Verschiebung an beiden äußeren Rändern der Konturen. Da M ebenfalls die Stärke des Abfalls der Verschiebungskurve entscheidend beeinflußt, muß ein heuristischer Kompromißwert nahe Null gefunden werden. Dessen Spannbreite ist aber groß, da die justierte Kontur nur eine grobe Näherung darstellt und anschließend mit Hilfe des aktiven Konturmodells verfeinert wird.

5 Ergebnisse und Ausblick

Die vorgestellte Methode wird eingesetzt, um physiologische und pathologische Bewegungsabläufe der Lippen zu studieren. Dazu werden Patienten zu sehr schnellen Lippenbewegungen aufgefordert, die mit einer Standardfarbkamera aufgenommen und digitalisiert werden. Um den Farbkontrast zu verbessern, wird zur Zeit noch Lippenstift aufgetragen. Zur Analyse werden jeweils Sequenzen von ca. 50 Bildern ausgewertet (Abb. 2).

Die Lippenkonturen können mit $M \in [0.01; 0.1]$ automatisch verfolgt werden. Auch wenn die in der Wissensbank formulierten Annahmen über konstante Form und Farbe nur eingeschränkt zutreffen, ist durch ihre Kombination eine Zuordnung der Maxima möglich. Allerdings zeigt die Unterlippenkontur Ungenauigkeiten am unteren Rand, die auf Blurring-Effekte zurückzuführen sind.

Zukünftig wird die Verbesserung des Gradientenbildes angestrebt, um auch wenig kontrastreiche Bilder auswerten zu können. Desweiteren werden diagnostische Parameter zur Beschreibung pathologischer Bewegungsabläufe entwickelt.

Literatur

1. Mirhosseini AR, Yan H, Lam K-M: Adaptive deformable model for mouth boundary detection. Optical Engeneering 37(3): 869-875, 1998.
2. Yuille AL, Hallinan PW, Cohen DS: Feature Extraction from Faces Using Deformable Tamplates. Internat. Journal of Computer Vision 8(2): 99-111, 1992.
3. Ahlrichs U, Paulus D, Wolf S: Objektivierung der Beurteilung von Gesichtsasymmetrien durch Bildanalyse. In: Lehmann et al.(Hrsg.): Bildverarbeitung für die Medizin, Augustinus Verlag, Aachen, 125-130, 1996.
4. Kass M, Witkin A, Terzopoulos D: Snakes: Active Contour Models. Internat. Journal of Computer Vision 1(4): 321-331, 1988.
5. Scholl I, Sovakar A, Lehmann T, Spitzer K: Motion Analysis of Vocal Folds using adaptive Snakes. In: Wittenberg et al.(eds.): Advances in Quantitative Laryngoscopy, Friedrich-Alexander-University Erlangen-Nuremberg, 29-38,1997.