Zungenverfolgung in medizinischen Röntgenbildsequenzen

Frank Höwing*†, Laurence Dooley* und Diederich Wermser†

*University of Glamorgan, School of Electronics,
Pontypridd, Mid Glamorgan, CF37 1DL. UK
Email: lsdooley@glamorgan.ac.uk
†Fachhochschule Braunschweig/Wolfenbüttel, FB E
Salzdahlumer Str. 46/48, 38302 Wolfenbüttel
Email: {f.hoewing|d.wermser}@fh-wolfenbuettel.de

Zusammenfassung. Dieser Beitrag stellt ein System zur Verfolgung von verformbaren Strukturen in zweidimensionalen Bildsequenzen vor. Es wird angewendet auf Röntgenbildsequenzen des menschlichen Vokaltraktes. Ziel dieser Anwendung ist das Vermessen artikulatorischer Organe. Die Messungen werden vorgenommen, um die komplexen dynamischen Vorgänge bei der menschlichen Spracherzeugung zu untersuchen. Von besonderem Interesse sind dabei verformbare Organe, insbesondere die Zunge.
Um dieses nicht eindeutige Erkennungsproblem auch unter dem Einfluß transparenter Überlagerungen sowie variierender Objekttexturen lösen zu können, wird ein zweistufiges System vorgestellt. Auf der unteren Ebene kombiniert der Ansatz kanten-, regionen- sowie bewegungsbasierte Bildverarbeitungsoperatoren, um deren jeweilige Vorteile nutzen und gleichzeitig Mängel kompensieren zu können. Auf einer höheren Ebene ermöglicht eine nicht-iterative Komponente auf der Basis aktiver Konturen die Nutzung von geometrischem Vorwissen.

Schlüsselwörter: Vokaltrakt, aktive Konturen, dynamische Programmierung

1 Merkmalsextraktion

Um die Vorteile der vielen existierenden Bildverarbeitungsoperatoren auszunutzen und gleichzeitig ihre jeweiligen Schwächen zu kompensieren, wird eine Struktur eingeführt, die es erlaubt einige kanten-, regionen- sowie bewegungsbasierten low-level Operatoren auf effiziente Weise zu kombinieren.

Um eine höhere Genauigkeit bei der Segmentierung zu erzielen, werden die Ergebnisse der Operatoren nicht als größere Elemente wie z.B. Liniensegmente repräsentiert, sondern bleiben pixelorientierte Bildbereichshinweise oder Evidenzen.

Experimente zeigen, daß jeder Operator für sich genommen fälschlicherweise niedrige Evidenzen für Konturpixel liefern und/oder hohe Werte für Pixel ausgeben kann, die nicht zu der gesuchten Kontur gehören. Die gewichtete, normalisierte Summe (Gl. (1)) aller Evidenzen für einen Pixel kann diesen Effekt kompensieren und zu einer Verstärkung der Gesamtevidenz e_{image} der meisten Konturpixel führen.

$$e_{image}(\mathbf{c}) = w_{edge}||e_{edge}(\mathbf{c})|| + w_{region}||e_{region}(\mathbf{c})|| + \ w_{motion}||e_{motion}(\mathbf{c})|| \quad (1)$$

Die tatsächlichen Evidenzfunktionen sind stark von der Anwendung abhängig. Die folgenden Definitionen in den Gleichungen (2) bis (4) wurden erfolgreich in der hier vorgestellten Anwendung eingesetzt. Abb. 1 sowie Abb. 2 demonstrieren die jeweiligen Teilkomponenten von e_{image}, welche im folgenden beschrieben werden.

Kantenbasierte Hinweise ergeben sich aus

$$e_{edge}(\mathbf{c}) = \nabla_{\perp}(\mathbf{c})\delta \quad (2)$$

wobei $\nabla_{\perp}$ den Gradienten senkrecht zu einer initialen Konturhypothese darstellt. Der Faktor δ bezeichnet dabei entweder eine steigende ($\delta = 1$) oder fallende ($\delta = -1$) Kante.

Die regionenbasierten Evidenzen aus Gl. (3) schwächen den Einfluß verdeckender Objekte, wie z.B. Füllungen, deren Grauwert einen Schwellwert θ unterschreitet. Pixel innerhalb dieser Objekte werden dabei nicht beeinflußt.

$$e_{region}(\mathbf{c}) = -|\nabla(\Theta, \mathbf{c})| \qquad \Theta(x_i, y_j) = \begin{cases} 1 & : \quad R(x_i, y_j) \leq \theta \\ 0 & : \quad R(x_i, y_j) > \theta \end{cases} \quad (3)$$

$R(x_i, y_j)$ bezeichnet hier einen Pixel innerhalb einer Region $\mathbf{R}$ um die Konturhypothese (vgl. Abb. 1). θ hängt von Vorwissen ab.

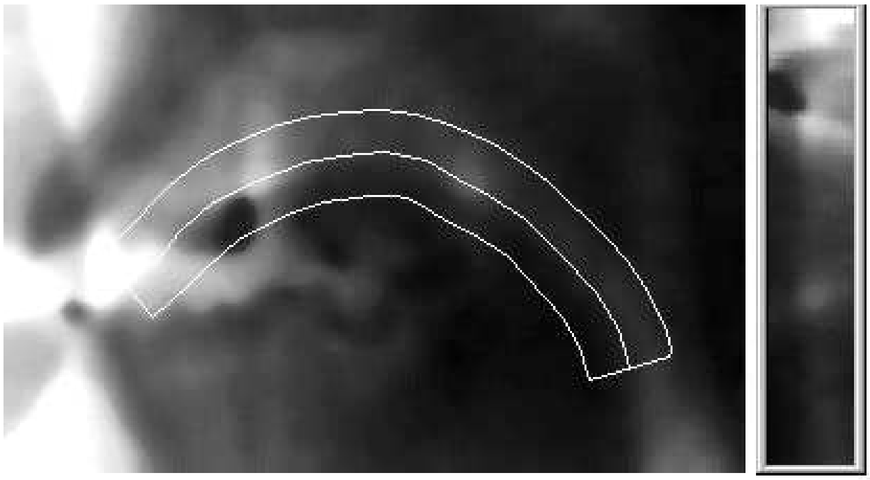

Abb. 1. Sagittale Röntgenaufnahme des Vokaltraktes (Ausschnitt). Der Suchraum $\mathbf{R}$ wird aus einer Konturhypothese (Mittellinie) abgeleitet.

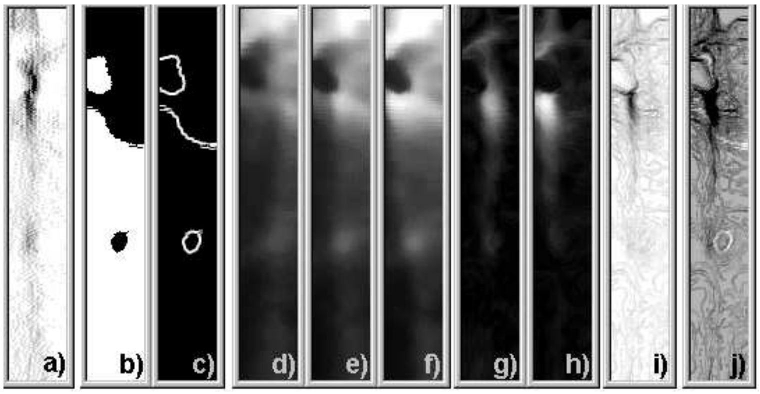

Abb. 2. Teilkomponenten von e_{image}. a) e_{edge}, b) Θ ($\theta = 80$), c) e_{region}, d) $\mathbf{R}_{t-d_1}$ ($d_1 = 5$), e) $\mathbf{R}_t$, f) $\mathbf{R}_{t+d_2}$ ($d_2 = 5$), g) $|\mathbf{R}_{t-d_1} - \mathbf{R}_t|$, h) $|\mathbf{R}_t - \mathbf{R}_{t+d_2}|$, i) e_{motion}, j) e_{image} ($w_{edge} = 0.1, w_{region} = 0.1, w_{motion} = 0.8$)

Um Bewegungsinformationen nutzen zu können wird die dritte Evidenzfuktion in Gl. 4 angewendet.

$$e_{motion}(\mathbf{c}) = \varepsilon_{med}(\mathbf{c}) = \sqrt{\nabla\left(|\mathbf{R}_{t-d_1} - \mathbf{R}_t|\right)}\sqrt{\nabla\left(|\mathbf{R}_t - \mathbf{R}_{t+d_2}|\right)} \qquad (4)$$

Dieser Operator liefert hohe Evidenzwerte für bewegte Kanten (*moving edge detector, med*), durch Multiplikation der Gradienten zweier Differenzbilder. Die Berechnung wird auf den Regionen $\mathbf{R}_i$ durchgeführt, die aus der Bildsequenz zu den Zeitpunkten $i = t$, $i = t - d_1$ und $i = t + d_2$ extrahiert werden. d_1 und d_2 sind dabei benutzerdefinierte Konstanten, die von der Bewegungsgeschwindigkeit des zu segmentierenden Objektes abhängen.

2 Kontursegmentation und -verfolgung

Ein Ansatz zur Segmentierung von Konturen, der nicht nur eine Zwischensymbolische Objektrepräsentation verwendet, sondern diese Repräsentation auch mit Bildmerkmalen und Vorwissen verbindet, sind die aktiven Konturen oder *Snakes*. Aktive Konturen wurden von Kass et al. [1] eingeführt und können als energieminimierende Splines oder Polygone bezeichnet werden. Für jedes Bild benötigt der Algorithmus ein initiales Polygon $\mathbf{P} = (\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{N-1})$, welches aus N Stützstellen $\mathbf{p}_i = (x_i, y_i)$ besteht, wobei x_i und y_i die Räumlichen Koordinaten von $\mathbf{p}_i$ sind. Die erkannte Kontur wird durch das Polygon $\mathbf{Q} = (\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_{N-1})$ mit $\mathbf{q}_i = (x_i, y_i)$ repräsentiert. Jede $\mathbf{q}_i$ wird aus einem Satz von Kandidaten $\mathbf{C}_i = (\mathbf{c}_{i,0}, \mathbf{c}_{i,1}, \ldots, \mathbf{c}_{i,M-1})$ ausgewählt, wobei die Kandidaten $\mathbf{c}_{i,j} = (x_j, y_j)$ in unserer Anwendung entlang einer Suchlinie abgetastet werden, die senkrecht zum initialen Polygon steht und $\mathbf{p}_i$ schneidet. Das Verfolgen der Kontur über eine Sequenz erfolgt Bild für Bild, wobei das jeweils resultierende $\mathbf{Q}(t)$ als initiale Konturhypothese $\mathbf{P}(t + 1)$ für das nächste Bild verwendet wird.

Tabelle 1. Energiminimierung der aktiven Kontur durch dynamische Programmierung.

1. **for all** m
2. $\quad S(0,m) = w_{image} E_{image}(\mathbf{c}_{0,m})$
3. **for all** m
4. $\quad S(1,m) = \min\limits_{k} \Big[w_{image} E_{image}(\mathbf{c}_{1,m}) + w_{int} E_{int} + S(0,k) \Big]$
5. $\quad T(1,m) = k^{\min}$
6. **for** $n = 2 \ldots N-1$
7. $\quad$ **for** $m = 0 \ldots M-1$
8. $\qquad S(n,m) = \min\limits_{k} \Big[w_{image} E_{image}(\mathbf{c}_{n,m}) + w_{int} E_{int} + w_{con} E_{con} + S(n-1,k) \Big]$
9. $\qquad T(n,m) = k^{\min}$

Die Energiefunktion der aktiven Kontur wird minimiert mittels dynamischer Programmierung. Tab.1 zeigt den Algorithmus, wobei $S(n,m)$ den minimalen Energiewert repräsentiert, der für die Stützstellen $0,\ldots,n$ erreicht werden kann, falls die $n$te Stützstelle der Kandidat $\mathbf{c}_{n,m}$ wäre. $T(n,m)$ speichert den Index k ($k = 0 \ldots M-1$), durch den der Ausdruck in Zeile 8 minimiert wird und zeigt so auf den optimalen Vorgänger des Kandidaten $\mathbf{c}_{n,m}$. Nachdem alle Stützstellen durchlaufen wurden, ergibt sich die neue Kontur durch Zurückverfolgen der Zeiger, wobei mit dem Kandidaten begonnen wird, der die minimale Gesamtenergie $S(N-1,m)$ aufweist. Die Gewichtungsfaktoren w_{image}, w_{int} und w_{con} in Tab. 1 steuern den relativen Einfluß der einzelnen Komponenten der Energiefunktion.

In der vorgestellten Anwendung wird die Energie der Bildmerkmale aus den oben beschriebenen Evidenzen abgeleitet:

$$E_{image} = 1 - e_{image} \tag{5}$$

Die interne Energie der aktiven Kontur ergibt sich aus[2]

$$E_{int} = ||e^{|k-m|}||, \tag{6}$$

wobei Kandidaten bevorzugt werden, die eine ähnliche lokale Krümmung wie die Anfangshypothese aufweisen. Dadurch existiert zwischen den Stützstellen eine Kraft, die der externen Kraft entgegenwirkt und so die Kontur glättet. Die Funktion $||\,.\,||$ normalisiert E_{int} auf den Bereich $[0,1]$, wobei ein niedriger Wert eine große Übereinstimmung bedeutet.

Eine geometrische Bedingung, die in [3] vorgeschlagen wurde, integriert Vorwissen über die erwartete Form des Objektes in die dynamische Programmierung. Gl. (7) bevorzugt konvexe oder konkave Konturen, durch Gewichtung eines Winkels γ der in dem Polygon gemessen wird, das aus den Stützstellen $\mathbf{c}_{n,m}$ (aktueller Kandidat), $\mathbf{c}_{n-1,k}$ (sein möglicher Vorgänger) sowie $\mathbf{c}_{n-2,T(n-1,k)}$ (der optimale Vorgänger von $\mathbf{c}_{n-1,k}$) besteht. Die resultierende Kontur kann in eine bestimmte Richtung gezwungen werden, indem γ immer in derselben Richtung gemessen, der erlaubte Bereich $[\gamma^{\min}, \gamma^{\max}]$ sowie der erwartete Winkel $\bar{\gamma}$ und die Standardabweichung σ eingestellt wird.

$$E_{con} = ||\nu(\gamma)|| \qquad \nu(\gamma) = \begin{cases} 0 & : \gamma^{\min} > \gamma > \gamma^{\max} \\ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\gamma - \bar{\gamma}}{\sigma}\right)^2} & : \gamma^{\min} \leq \gamma \leq \gamma^{\max} \end{cases} \tag{7}$$

3 Ergebnisse

Abb. 3 zeigt einige Resultate des vorgestellten Systems bei der Segmentierung und Verfolgung der Zunge. Jedes zehnte Bild einer Sequenz ist dargestellt (von oben nach unten). Die überlagerte Kontur ist in diesem Beispiel nicht geglättet.

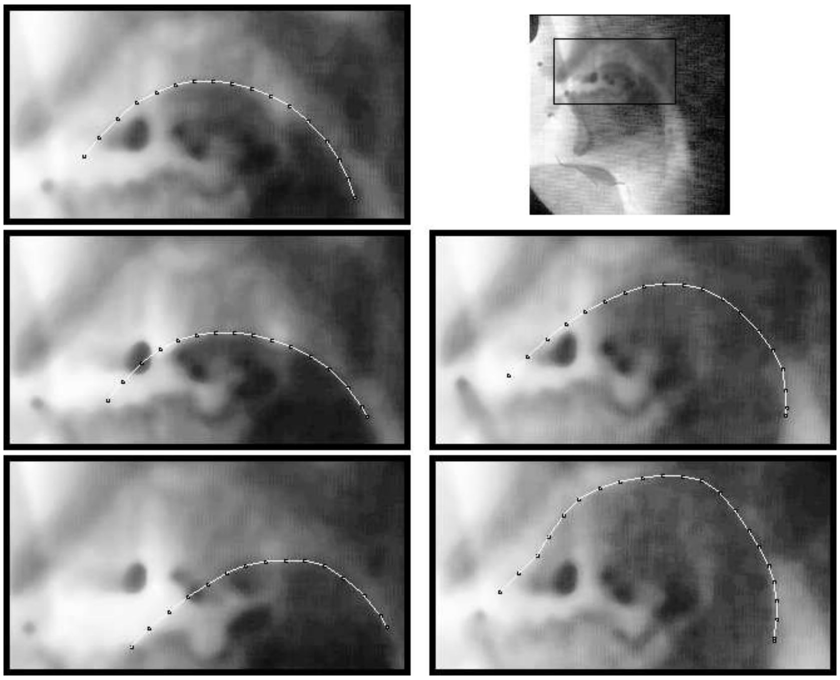

Abb. 3. Resultate des vorgestellten Systems.

Literatur

1. Kass M, Witkin A, Terzopoulos D: Snakes: Active contour models. Int. Conf. on Computer Vision, London, 259–268, June 1987
2. Olstad B: Automatic wall motion detection in the left ventricle using ultrasonic images. Electronic Imaging, San Jose, 1991
3. Höwing F, Wermser D, Dooley L: Recognition and tracking of articulatory organs in X-ray image sequences. IEE Electronics Letters, (32):444–445, 1996