Hierarchische Wasserscheiden-Transformation zur Lippensegmentierung in Farbbildsequenzen

Christoph Palm, Benedikt Fischer,
Thomas Lehmann und Klaus Spitzer

Institut für Medizinische Informatik
Rheinisch-Westfälische Technische Hochschule (RWTH), 52057 Aachen
Email: cpalm@mi.rwth-aachen.de

Zusammenfassung. Zur Lösung komplexer Segmentierungsprobleme wird eine hierarchische und farbbasierte Wasserscheidentransformation vorgestellt. Geringe Modifikationen bezüglich Startpunktwahl und Flutungsprozess resultieren in signifikanten Verbesserungen der Segmentierung. Das Verfahren wurde zur Lippendetektion in Farbbildsequenzen eingesetzt, die zur quantitativen Beschreibung von Sprechbewegungsabläufen automatisch ausgewertet werden. Die Experimente mit 245 Bildern aus 6 Sequenzen zeigten eine Fehlerrate von 13%.

Schlüsselwörter: Hierarchische Wasserscheiden-Transformation, Segmentierung der Lippen, Bewegungsanalyse, Farbbildverarbeitung

1 Einleitung

Physiologische und pathologische Lippenbewegungsabläufe [1] werden mit Hilfe der Farbvideotechnik dokumentiert und analysiert. Zur Bestimmung quantitativer Parameter, die solche Abläufe beschreiben, ist eine automatisierte Segmentierung der Lippen in digitalisierten Farbbildsequenzen erforderlich. In der Bildverarbeitung werden mit kantenorientierten und regionenbasierten Segmentierungsverfahren im wesentlichen zwei Ansätze unterschieden. Während z.B. die häufig erfolgreich eingesetzten Snake-Verfahren meist anhand des Gradienten die Grenzen homogener Gebiete kantenorientiert detektieren, fassen z.B. Split-and-Merge-Methoden diese Gebiete mit Hilfe von Ähnlichkeitskriterien regionenorientiert zusammen. Nachdem für die Lippensegmentierung ein aktives Konturmodell [1] zu häufig an den geringen Kontrasten scheiterte, wird hier die Wasserscheiden-Transformation [2] eingesetzt, die Eigenschaften regionen- und kantenorienterter Segmentierung verbindet. So stellen ausgeprägte Gradientenmaxima ein Hindernis für die Ausbreitung einer Region dar, während in kontrastarmen Gebieten durch die gleichzeitige Ausbreitung mehrerer Regionen dennoch eine Trennung erreicht werden kann. Nachteil dieser Methode ist die bekannte Neigung zur Übersegmentierung, da jedem initialen lokalen Minimum eine Region zugeordnet wird. Desweiteren ist der Einsatz der Farbinformation bislang nur indirekt über das Gradientenbild möglich [3].

In diesem Beitrag werden diese Nachteile durch geringe Modifikationen aufgehoben. Die neue Transformation führt zu einer deutlichen Verbesserung der Segmentierungsergebnisse und ist auch auf andere Fragestellungen übertragbar.

2 Methode

Die Wasserscheiden-Transformation basiert auf der Vorstellung eines Grauwertgebirges, bei der die Helligkeit eines Pixels mit der Höhe des Berges an entsprechender Stelle korrespondiert. Aus den Tälern (lokale Minima) steigt Wasser derart nach oben, daß der Wasserstand in allen Tälern gleich ist. Tiefe Täler werden also zuerst geflutet. Es werden immer dann Dämme (Wasserscheiden) errichtet, wenn zwei Seen im nächsten Schritt verschmelzen würden. Die Dämme stellen die Regionengrenzen dar. Grundlage des *Gebirges* ist meist das Gradientenbild, das an Objektgrenzen hohe Werte zeigt. Im folgenden werden Modifikationen bei Startpunktwahl sowie Fluten vorgestellt, die zu einem hierarchischen Algorithmus genutzt werden.

2.1 Startpunkte

Bei der Wahl der Startpunkte fließt Vorwissen über Art und Anzahl der zu segmentierenden Objekte ein. Die Punkte können z.B. interaktiv in der Region plaziert werden. Entscheidend ist, daß zu jedem Startpunkt das entprechende Objekt a-priori bekannt ist. Zur Steigerung der Stabilität werden in jedem Objekt mehrere Startpunkte markiert, die jeweils das gleiche Label besitzen. Beim Zusammenstoß der entsprechenden Regionen kommt es automatisch zu einer Verschmelzung unter Erhaltung der eindeutigen Objekt-Label Zuordnung. Das ist Vorrausetzung für das nachfolgende objektbezogene Fluten.

2.2 Fluten

Herkömmliches Fluten wird durch eine Sortierung der Flutpunkte gemäß ihrer Höhe und die jeweilige Betrachtung der Nachbarschaft realisiert. Liegt der Nachbar unterhalb eines Flutpunktes und ist noch keiner anderen Region zugeordnet, so wird er mit dem Label des aktuellen Flutpunktes versehen und selbst in die Liste der Flutpunkte eingetragen. Dabei werden alle Flutpunkte gleich behandelt. In einem zweiten Schritt werden die gefundenen Regionen anhand von global gültigen Ähnlichkeitskriterien zusammengefasst. Wissen über die Objekte kann dabei nicht eingebracht werden (Abb. 1).

Dieses Konzept wird hier durch bedingtes und richtungsbasiertes Fluten in Abhängigkeit vom Label des Flutpunktes erweitert. So kann das Fluten eines Nachbarn davon abhängig gemacht werden, in welcher Richtung er liegt und ob er labelspezifische Bedingungen erfüllt. Dabei ist beispielsweise die Farbe als Bedingung in Form eines geringen Farbabstandes zur Flutpunkt- oder zur mittleren Regionenfarbe in direkter Form zu integrieren. Im Gegensatz zu den klassischen Merging-Strategien sind die Flutungsbedingungen sowohl richtungsals auch objektspezifisch. Im Fall der Oberlippensegmentierung stellen z.B. helle, weissliche Pixel nur in Richtung Mundöffnung ein Hindernis dar (Zähne), nicht aber in perioraler Richtung (umflossenes Glanzlicht). So werden Glanzlichter

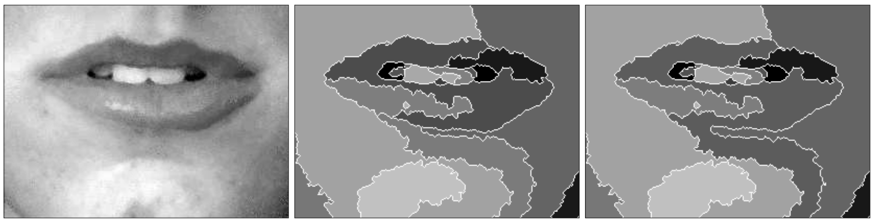

Abb. 1. Zusammenfassung von Regionen nach Farbähnlichkeit. Die Bilder in der Mitte und rechts stellen zwei aufeinanderfolgende Merging-Schritte zur Segmentierung des Originals (links) dar. Während im übersegmentierten Bild (Mitte) die Lippenkonturen vollständig vorhanden sind, werden beim Merging (rechts) beleuchtungsbedingt Lippen- und Hintergrundregion miteinander verschmolzen.

trotz ihrer geringen Ähnlichkeit in die Oberlippenregion integiert. Weitere Konsequenz des bedingten Flutens ist die möglicherweise unvollständige Segmentierung des Bildes. Solche Pixel, die die Bedingung nicht erfüllen bleiben fraglich und erhalten kein Label. Andererseits kann bei einer scharfen Formulierung der Bedingungen sichergestellt werden, daß jedes vom Verfahren gelabelte Pixel sicher der gewünschten Region zugeordnet wurde. Die nicht gelabelten Pixel, aber auch inhomogene Objekte können mit Hilfe einer hierarchischen Wasserscheiden-Transformation in nachfolgenden Schritten weiterverarbeitet werden.

2.3 Hierarchische Wasserscheiden-Transformation

Die Definition von Bedingungen für das Fluten gestaltet sich dann schwierig, wenn gleichzeitig viele unterschiedliche Regionen zu trennen sind. Zur Lösung bietet sich die schrittweise Verfeinerung von Segmenten an, die durch eine hierarchische Wasserscheiden-Transformation realisiert wird. Dabei werden sukzessive Regionen ermittelt, die sicher von anderen abgegrenzt werden können. Die erneute Anwendung der Wasserscheiden-Transformation bezogen auf eine spezifische Region kann zu einer Veränderung der Flutungsbedingungen genutzt und eine weitere Aufteilung in feinere Strukturen herbeigeführt werden (Abb. 2).

Die Wahrung der in vorherigen Segmentierungsstufen detektierten Außengrenzen wird durch ein binäres Maskenbild realisiert. Hier werden die Gebiete

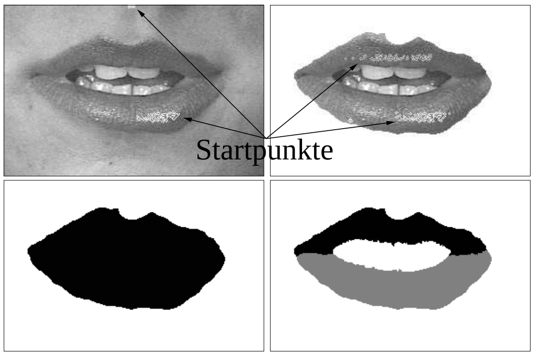

Abb. 2. Links: Hierarchiestufe 1, rechts: Hierarchiestufe 2.

festgelegt, in denen Fluten zulässig ist. Die Berücksichtigung des Maskenbildes kann in das Konzept des bedingten Flutens als weitere Bedingung integriert werden. Die hierarchische Wasserscheiden-Transformation ermöglicht auf diese Weise kaskadierte Flutungsbedingungen zur Abgrenzung von Objektdetails.

3 Anwendung

Die vorgestellten Erweiterungen der Wasserscheiden-Transformation werden zur Segmentierung von Lippen in Farbbildsequenzen eingesetzt. Dazu sprechen Patienten Laute wie *pill, mull, aba*, ... Die entsprechenden Lippenbewegungen werden auf Farbvideos gespeichert. Derzeit wird untersucht, inwieweit Bewegungsstörungen durch quantitative Parameter wie Symmetrie, Mundschluß,... manifestiert werden können.

Zur automatischen Auswertung wird ein zweistufiges Verfahren verwendet. Im ersten Schritt werden die Startpunkte des Hintergrundlabels am Bildrand plaziert, die der Mundregion dort, wo sich die zu den Lippen im vorherigen Bild ähnlichsten Farben befinden. Das erste Bild der Sequenz wird manuell vorsegmentiert. Das Ergebnis der Stufe 1 ist ein Maskenbild, das die Mundregion segmentiert (Abb. 2 links). Im zweiten Schritt bleibt die Hintergrundregion fixiert, während verschiedene Startpunkte für Ober- und Unterlippe aufgrund ihrer Farbähnlichkeit bestimmt werden (Abb. 2 rechts). Das Fluten basiert auf dem Farbgradienten des kontrastverstärkten Bildes [1]. Flutungsbedingung in Richtung Mundöffnung ist der Farbabstand des Nachbarpixels vom mittleren Farbwert der aktuellen Region. Das Fluten nach außen erfolgt dann ohne Prüfung der Farbähnlichkeit, wenn es sich um eine helle, ungesättigte Farbe handelt. Durch die Richtungsabhängigkeit lassen sich Glanzlichter in die Lippenregionen integrieren, obwohl deren Farbabstand sehr groß ist.

4 Ergebnisse und Diskussion

Bislang wurden 245 Bilder aus 6 Farbbildsequenzen mit Hilfe der modifizierten Wasserscheiden-Transformation segmentiert. Nach subjektiver Beurteilung wurden davon 212 (87%) korrekt segmentiert. Abb. 3 zeigt Beispielsequenzen für verschiedene Lautfolgen. Im letzten Bild der unteren Sequenz ist eine Fehlsegmentierung zu sehen, die durch zu geringe Kontraste zwischen Zunge und Lippe zu erklären ist. Das Vordringen der Lippenregion in die Mundöffnung kann aber meist durch die Farbabstandsbedingung auch bei einem unvollständigen Gradientengrat verhindert werden. Allerdings ist die Wahl des maximalen Farbabstandes ein kritischer Parameter, der vom Benutzer vorzugeben ist. Er kann jedoch über eine Bildsequenz konstant gehalten werden. Die Informationspropagation durch eine Bildsequenz erfolgt durch Weitergabe der mittleren Lippenfarbe und -position. Die Position wird im Folgebild zur Einschränkung des Suchraumes verwendet, in dem die ähnlichsten Farben bestimmt werden. Dennoch ist das Verfahren auch bei starken Bewegungen stabil. Im Vergleich zu snake-basierten

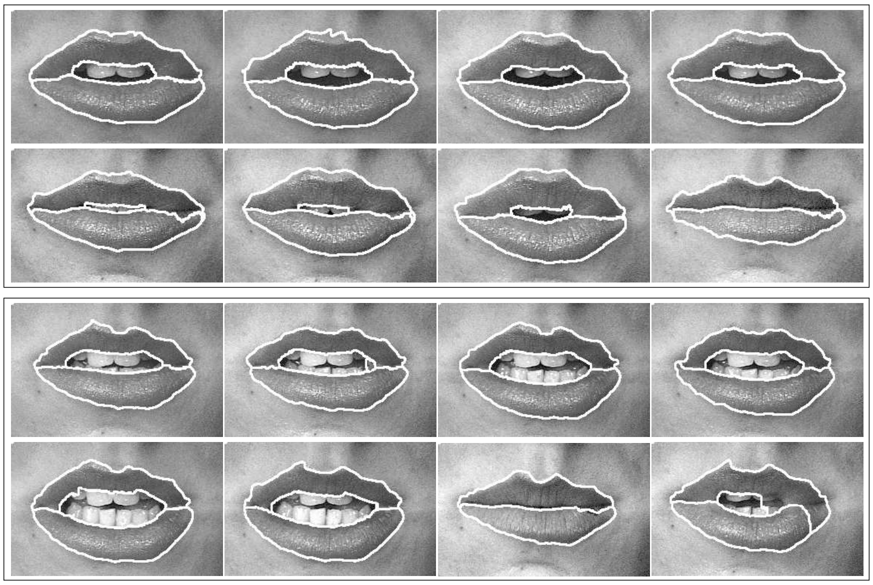

Abb. 3. Ausschnitte aus den Sequenzen für die Lautfolgen *owo* (oben) und *ibi* (unten) (jedes zweite Bild).

Verfahren führen Segmentierungsfehler nicht zwangsläufig zu Folgefehlern in der Restsequenz, da in der ersten Hierarchiestufe die Startpunkte beliebig innerhalb der Mundregion plazierbar sind. Folgefehler können durch eine Kombination von Farb- und Ortskriterien vermieden werden.

Probleme ergeben sich durch inhomogene Beleuchtung, die in den vorliegenden Bildern durch zwei seitlich vom Probanden angebrachte Punktlichtquellen entstand. Dadurch kommt es in einigen Bildern zu unvollständigen Lippensegmenten, die Mundöffnung bleibt davon allerdings unberührt.

Die Experimente zeigen, daß die vergleichsweise geringen Modifikationen der Wasserscheiden-Transformation (vgl. Abb. 1) zu einer deutlichen Verbesserung der Segmentierungsergebnisse führen. Nach einer vollständigen Automatisierung des Verfahrens sollen die Ergebnisse in der klinischen Anwendung zur quantitativen und qualitativen Bewertung von Bewegungsstörungen genutzt werden.

Literatur

1. Palm C, Neuschaefer-Rube C, Lehmann T, Spitzer K: Wissensbasierte Bewegungskompensation in aktiven Konturmodellen. In: Evers H, Glombitza G, Lehmann T, Meinzer H-P: Bildverarbeitung für die Medizin 1999, Springer Verlag, Berlin, 8-12, 1999.
2. Vincent L, Soille P: Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. IEEE Trans. PAMI, 13(6): 583-598, 1991.
3. Shafarenko L, Petrou M, Kittler J: Automatic Watershed Segmentation of Randomly Textured Color Images. IEEE Trans. IP, 6(11): 1530-1544, 1997.