Augenlokalisation und -analyse in klinischen Applikationen

Bernhard Fröba, Christian Münzenmayer, Sandra Stecher und Thomas Wittenberg

Fraunhofer Institut für Intergrierte Schaltungen
Am Weichselgarten 3, D-91058 Erlangen
Email: {bdf,muenzecn,stechesa,wbg}@iis.fhg.de

Zusammenfassung. Ein System zur automatischen Echtzeit-Detektion von Gesichtern und Segmentierung der Augen wird vorgestellt. Für die die beiden Schritte Detektion und Segmentierung werden spezielle Verfahren, basierend auf *Edge Orientation Matching* und *Deformable Template Matching* entwickelt und eingesetzt. Das Verfahren zur Augensegmentierung wird auf einer Stichprobe von 192 Einzelbildern ausgewertet und ein Vergleich von zwei Varianten präsentiert.

1 Einleitung

Für viele Anwendungen in der Medizin ist eine notwendige Basis-Technologie die automatische Detektion von Gesichtern und damit meist direkt verbunden die Lokalisation und Segmentierung der Augen. Ein Beispiel für eine solche Aufgabenstellung ist die rechnerunterstützte Diagnose und Therapieverlaufskontrolle von Gesichtslähmungen (Paresen) mit Hilfe von Kameras [1,2]. In diesem Szenario muss das Gesicht des Patienten vor der Kamera automatisch gefunden und verfolgt werden. Bei einer Fokussierung von Diagnose und Therapie auf die Augenpartie — z.B. für Untersuchungen des Lidschlages — müssen zudem beide Augen in Echtzeit (mehr als 10 Bilder/Sekunde) detektiert und segmentiert werden.

In diesem Beitrag wird eine Methode zur schnellen Detektion von Gesichtern vor beliebigem Hintergrund mit anschließender Segmentierung und Analyse der Augen vorgestellt, bei der zur Gesichtssegmentierung lediglich Strukturinformation im Grauwertbereich verwendet wird. Ausgehend von der bekannten Methode der *Deformable Templates* [3] wurde zur Segmentierung und Analyse der Augenpartien ein neues robustes Verfahren entwickelt, dass deutlich bessere Ergebnisse liefert als der ursprünglich vorgeschlagene Ansatz. Das Gesamtverfahren - Gesichtssegmentierung mit anschließender Analyse der Augen - ist echtzeitfähig, d.h. es kann auf Standardhardware (PC 700MHz, DirectShow Grabber) mit mehr als 10 Bilden pro Sekunde betrieben werden. Die Anzahl der gleichzeitig analysierbaren Gesichter ist dabei nicht beschränkt.

2 Gesichtsdetektion

Zur Gesichtsdetektion wird das in einer früheren Arbeit vorgestellte EOM (Edge Orientation Matching) [4] verwendet. Für das EOM wird aus einer Trainingsdatenbank ein Strukturmodell des Gesichtes erzeugt, dass die dominanten Kanten anhand ihrer Vorzugsrichtungen beschreibt. Das EOM liefert Kandidatenpositionen die mit einem ansichtsbasierten Klassifikator [5] verifiziert werden. Die gesamte Detektionskette lässt sich sehr effizient realisieren, so dass z.B. die Analyse eines 320×240 großen Bildes weniger als 40ms auf einem PC 700 MHz benötigt. Dabei werden Gesichter ab einer Größe von 32×40 Bildpunkten gefunden. Die Detektionsgenauigkeit ist im Mittel besser als 25% Abweichung der ermittelten Gesichtsgröße und Position, bezogen auf den tatsächlichen Augenabstand. Die Gesichtsdetektion findet somit eine gute initiale Schätzung der Augenposition für die nachfolgend beschriebene Augensegmentierung.

3 Augensegmentierung

3.1 Geometrisches Augenmodell

Zur Augensegmentierung wird in dieser Arbeit ein geometrisches Augenmodell verwendet wie es in [3] vorgeschlagen wird. Dies entspricht einem wissensbasierten Ansatz insofern, als das Kenntnis über das zu analysierende Objekt — hier das Auge — im Bild in den Algorithmus einfließt. Das geometrische Modell (Abb. 1 a), im folgenden auch *Augentemplate* genannt, beschreibt das Suchobjekt Auge durch einen Parametervektor $\boldsymbol{p} = \{\boldsymbol{x}_c, \boldsymbol{x}_e, r, a, b, c, \Theta\}$. Ein Kreis mit Radius r um das Iriszentrum $\boldsymbol{x}_c$ trennt die Iris vom Weißen des Auges. Zwei parabelförmige Bögen mit Zentrum $\boldsymbol{x}_e$ und Höhe a bzw. c beschreiben den oberen und unteren Lidrand. Mit dem Parameter b wird die Breite des Auges festgelegt und der Winkel Θ beschreibt die Neigung gegenüber der Horizontalen.

Die Passform einer Instanz dieses Parametervektors $\boldsymbol{p}$ im Bild bewertet das Verfahren anhand mehrerer Energiefunktionen, die zur Gesamtenergie $E_c := E_c(\boldsymbol{p})$ kombiniert werden:

$$\begin{aligned} E_c = & c_{IrisInt} \cdot E_{IrisIntensity} + c_{IrisEdge} \cdot E_{IrisEdge} + \\ & c_{ScleraInt} \cdot E_{ScleraIntensity} + c_{ScleraEdge} \cdot E_{ScleraEdge} + \\ & c_{ScleraEdgeOrient} \cdot E_{ScleraEdgeOrientation} + c_{EyeForm} \cdot E_{EyeForm} \end{aligned} \tag{1}$$

Die Teilenergiefunktionen $E_{IrisIntensity}$ und $E_{ScleraIntensity}$ ermitteln den mittleren Grauwert im Iriskreis bzw. in der Region des Weißen im Auge. $E_{IrisEdge}$ und $E_{ScleraEdge}$ bewerten die Kantenstärke entlang des Irisrandes und der Lidränder. Um fehlerhaften Deformationen des Templates entgegenzuwirken, wird die Energie $E_{EyeForm}$ eingeführt, die die geometrische Konfiguration in sinnvollen Proportionen zueinander halten soll. Im Rahmen der Arbeit wurde die weitere

Energie $E_{ScleraEdgeOrientation}$ eingeführt, die die Korrelation zwischen der Kantenorientierung und der Steigung des Kurvenverlaufs auf den Lidrändern bewertet. Die Koeffizienten c erlauben eine Gewichtung der Energien und können aus einer Lernstichprobe ermittelt werden.

Mittels eines Optimierungsverfahrens wird der Parametervektor $\boldsymbol{p}$ so bestimmt, dass sich das initiale Modell in gewissem Rahmen selbstständig in Größe, Form und Lage an das Suchobjekt anpasst und so die Bewertungsfunktion minimiert.

3.2 Optimierungsverfahren

Das ursprüngliche Optimierungsverfahren verwendete ein Gradientenabstiegsverfahren kombiniert mit der *Epochenmethode* aus [3]. Unter dem Begriff *Epochenmethode* versteht man hier, dass man die Teilenergiefunktionen nicht alle auf einmal optimiert, sondern dass man sie in mehreren Epochen, nacheinander mit unterschiedlichen Gewichtungsfaktoren auswertet. Ein Starttemplate soll sich so schrittweise an das Auge adaptieren. Der Nachteil dieses Verfahrens besteht darin, dass die Irisposition in der ersten Epoche nur anhand von Intensitätsinformation bestimmt wird und dass in den darauffolgenden Optimierungsschritten, der Feinanpassung, eine eventuelle Irisfehldetektion nicht mehr behoben werden kann. Ist die Startposition des Starttemplates vom Auge zu weit entfernt setzt sich das Template an den Brauen, den Haaren oder Nasenlöchern fest.

Wesentlich seltener ist dies zu beobachten, wenn man alle Teilenergiefunktionen in einem einzigen Schritt optimiert. Dazu wurde das Downhill Simplex Verfahren implementiert [6]. Es handelt sich dabei um ein regelbasiertes Verfahren, dass in einem geeignet beschränkten N-dimensionalen Parameterraum selbstständig zumindest ein lokales Minimum finden kann. Der Parametersuchraum wird durch eine Startmatrix begrenzt. Aus einer Trainingsstichprobe werden dazu die Mittelwerte der Templateparameter bestimmt und anhand derer die Startmatrix initialisiert.

Weitere Vorteile des Downhill Simplex Verfahrens sind, dass es einfacher zu verwenden ist, da wesentlich weniger Algorithmusparameter eingestellt werden müssen, dass die Augendrehung und die Augengröße besser angepasst werden und das grobe Fehldetektionen seltener sind als mit dem Gradientenabstiegsverfahren.

4 Versuche und Ergebnisse

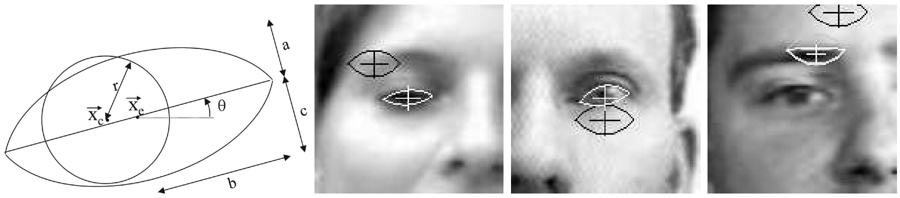
Bei den Stichproben die zur Verfügung standen, handelte es sich um jeweils 192 Aufnahmen der rechten bzw. linken Gesichtshälfte mit dem entsprechenden Auge. Die Auflösung für alle Bilder betrug 128×128 Pixel.

Zur Messung der Erkennungsraten werden die zwei Kriterien korrekte Irisposition bzw. korrektes Auge definiert. Damit ein Ergebnis der korrekten Irisposition zugeordnet werden kann, muss der Abstand des gefundenen Irismittelpunkts vom tatsächlichen einen Schwellwert von 6 Pixeln unterschreiten. Bei dieser Bildgröße und Stichprobenqualität entspricht dies der Genauigkeit, die bei der Handsegmentierung erreicht werden kann. Für eine korrekte Augenposition müssen alle Templateparameter einen Schwellwert von 6 Pixeln bzw. 10 Grad unterschreiten.

Die Erkennungsraten in Tab. 1 belegen die oben bereits erwähnten Vorteile des Downhill Simplex Verfahrens. Auffallend ist hier die weit bessere Anpassung des Gesamttemplates und des Drehwinkels Θ. Die bessere Funktionsweise des Downhill Simplex Verfahrens wird auch in einigen repräsentativen Ergebnisbildern (Abb. 1) sichtbar.

	Gradientenabstiegsverfahren		Downhill Simplex	
	Stichprobe 1	Stichprobe 2	Stichprobe 1	Stichprobe 2
Korrekte Irisposition	84	80	96	95
Korrektes Auge	53	48	81	84

Tabelle 1. Vergleich der Erkennungsraten in % für Gradientenabstiegs- und Downhill Simplex Verfahren



(a) (b) (c) (d)

Abb. 1. (a)Geometrisches Augenmodell mit den Parametern $a, b, c, r, \boldsymbol{x}_c$, $\boldsymbol{x}_e$ und θ. Beispiele für die Augensegmentierung mittels Downhill Simplex Optimierung, (b)und (c) zeigen erfolgreiche Adaptionen, (d) konnte nicht angepasst werden. Das Starttemplate ist schwarz, das Ergebnis weiß dargestellt.

Literatur

1. Wittenberg Th, Frischholz R et al: *Automatische Verfolgung von Augenlidbewegungen und Korrelation mit EMG-Daten.* In: Bildverarbeitung in der Medizin 2000, Springer Verlag, Berlin, 43-47.
2. Gebhard A, Paulus D et al: *Automatische Graduierung von Gesichtsparesen.* In: Bildverarbeitung in der Medizin 2001, Springer Verlag, Berlin, 352-356.
3. Yuille AL; Hallinan PW; Cohen DS:*Feature Extraction from Faces Using Deformable Templates.* In: International Journal of Computer Vision 8:99-111, 1992.
4. Fröba B, Küblbeck Ch: *Real-Time Face Detection using Edge-Orientation Matching.* In: Audio- and Video-based Biometric Person Authentication (AVBPA'2001) pp. 78-83, 2001.
5. Yang MH, Roth D, Ahuja N: *A SNoW-Based Face Detector.* In: Advances in Neural Information Processing Systems 12 (NIPS 12), pp 855-861, 1999.
6. Press W. H. et al.: *Numerical Recipes in C* The Press Syndicate of the University of Cambridge, 1999.