

Detektion und Verifikation von charakteristischen Bildpunkten in Gesichtsbildern*

R. Herpers¹ L. Witte², M. Michaelis¹, J. Bruske³ und G. Sommer³

¹ GSF-Institut für Medizinische Informatik und Systemforschung, Neuherberg, 85764 Oberschleißheim, E-mail: (herpers,michaeli)@gsf.de

^{2,3} Lehrstuhl für Mensch-Maschine-Kommunikation, TU-München, 80290 München

³ Institut für Informatik, Christian-Albrechts-Universität, 24105 Kiel

Zusammenfassung In diesem Beitrag wird ein Verfahren vorgestellt, welches automatisch charakteristische Punkte in Bildausschnitten aus Gesichtsbildern detektiert. Die Lokalisierung der Punkte basiert auf einer modellgesteuerten Detektion und Verfolgung der vorhandenen Linien und Kanten. Zur robusten Verfolgung der Kanten wird ein steuerbares Filterschema eingesetzt. Eine anschließende Verifikation der detektierten Bildpunkte verringert zusätzlich die Wahrscheinlichkeit einer Falschdetektion. Die Extraktion der benötigten Informationen und Merkmale basiert sowohl für die Detektion als auch für die anschließende Verifikation auf demselben Satz von Filtern. Zur Verifikation der gefundenen Bildpositionen wird eine Dynamische Zellstruktur (DCS-Netzwerk) verwendet, die durch ein überwachtetes Lernverfahren trainiert wird.

1 Einleitung

Charakteristische Bildpunkte in Gesichtsbildern, wie Augen- und Mundecken, stellen bedeutende Merkmale für eine Vielzahl von Anwendungen im Bereich der automatischen Gesichtserkennung dar [11]. Das grundlegende Problem bei einer automatischen Detektion ist, daß solche charakteristischen Punkte eher durch ihre semantische Bedeutung (Ecke des Mundes oder Auges) als durch eine einheitliche lokale Struktur in den Bilddaten definiert sind. Insbesondere für eine präzise Lokalisierung der Punkte stellt die große Variabilität der zugrundeliegenden Struktur in den Bilddaten aufgrund verschiedenster kleiner Falten, Glanzlichter, Schatten und anderer nicht vorhersehbarer Merkmale ein nicht triviales Problem dar (Abb. 1). Deshalb lassen sich diese markanten Punkte auch nicht mit Verfahren zuverlässig detektieren, die lediglich die Bilddaten in der lokalen Nachbarschaft auswerten, ohne das vorhandene Kontextwissen mit einzubeziehen.

In diesem Beitrag wird ein kombiniertes Verfahren vorgestellt, welches automatisch charakteristische Punkte in Gesichtsbildausschnitten detektiert und verifiziert. Die Bildregionen werden bereits vorab durch eine attentive Regionensuche extrahiert [3, 4], um diese anschließend gezielt nach markanten Punkten zu untersuchen. Die Detektion dieser Punkte basiert dabei auf einer sequentiellen

* Diese Arbeit wurde von der DFG unter den Az: So 320\1-1 und Ei 322\1-1 gefördert.

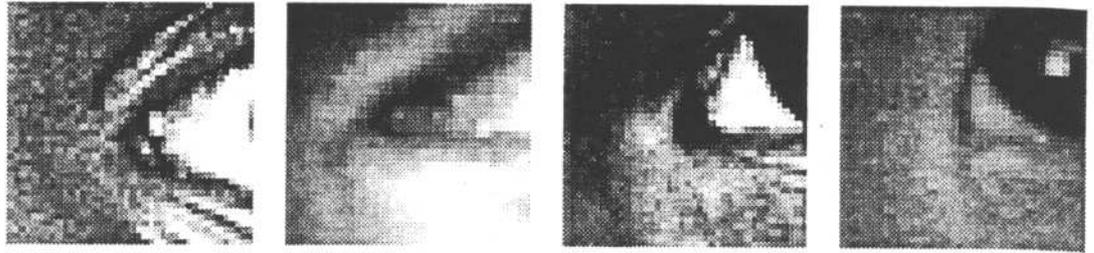


Abbildung 1. Vier Beispiele von realen inneren Augenecken. Die starke Variabilität der Bilddaten erlaubt keine rein datengetriebene Lokalisation.

Auswertung der Kanten- und Linieninformation. Wesentlich dabei ist, daß der Kantenverfolgung ein Modell der untersuchten Bildregion zugrundeliegt.

Zur Reduktion von Fehldetektionen schließt sich eine Verifikation der gefundenen Punkte an. Dazu wird an den detektierten Punkten aus der lokalen Bildinformation ein Merkmalvektor generiert und durch ein DCS-Netzwerk klassifiziert. Eine direkte Suche der charakteristischen Bildpunkte allein mit Hilfe eines DCS-Netzwerkes aufgrund der lokalen Information ist dagegen unzuverlässig. Eine Verifikation der bereits gefundenen Bildpositionen ist jedoch mit hoher Zuverlässigkeit möglich. Ordnet das DCS-Netzwerk den Merkmalvektor des gefundenen Bildpunktes der Klasse des gesuchten charakteristischen Punktes zu, so gilt der Punkt als detektiert.

Die Grundlage der Merkmalsextraktion für die Detektion als auch für die Verifikation bildet ein steuerbares Filterschema, welches Kanten- und Linienfilter einer hohen Orientierungsselektivität in beliebigen Skalen und Orientierungen auf das Bild anwendet. Dazu wird nur ein kleiner Satz sogenannter Basisfunktionen mit dem Bild gefaltet, mit dem aber alle benötigten Merkmale während der Suche der Punkte bei Bedarf abgeleitet werden können. Die strukturelle Vielfalt dieser Basisfunktionen ermöglicht außerdem die sensible Unterscheidung verschiedener lokaler Bildstrukturen für den sich anschließenden Verifikationsschritt.

Das steuerbare Filterschema, welches die Grundlage der hier vorgestellten Verfahren darstellt, wird in Kapitel 2 kurz eingeführt. Anschließend wird in Kapitel 3 die Detektion der markanten Punkte durch die modellgesteuerte Kantenverfolgung am Beispiel der Augenregion behandelt. In Kapitel 4 wird die zur Verifikation verwendete dynamische Zellstruktur und die Verknüpfung mit der Detektion über den Satz von Bayes vorgestellt.

2 Das Filterschema

2.1 Steuerbare Filter

Die Grundlage für das im folgenden vorgestellte Filterschema bilden die in Abbildung 2 dargestellten orientierungsselektiven Kanten- und Linienfilter. Diese Filter $F(\mathbf{x})$ werden in Orientierung und Skala (ϑ, σ) gesteuert. Der Begriff der Steuerung bezieht sich dabei auf die lineare Rekonstruktion der Filter aus einem festen und kleinen Satz von sogenannten Basisfunktionen $A_k(\mathbf{x})$ (siehe Formel und Abb. 2e-h).

$$F_{\vartheta, \sigma}(\mathbf{x}) \approx \sum_{k=1}^N b_k(\vartheta, \sigma) A_k(\mathbf{x})$$

Alle Faltungen und Projektionen des Bildes werden ausschließlich mit den Basisfunktionen ausgeführt. Die Kanten- und Liniendetektoren selbst und deren Antworten erhält man durch Superpositionen der Basisfunktionen bzw. deren Antworten. Die vom Deformationsparameter abhängigen Superpositionskoeffizienten $b_k(\vartheta, \sigma)$ werden Interpolationsfunktionen genannt. Die Basisfunktionen und die Interpolationsfunktionen werden durch ein auf der Singulärwertzerlegung basierendes Verfahren berechnet, das auf Perona zurückgeht [10]. Allerdings werden hier Orientierung und Skala simultan gesteuert. Diese Methode wird detailliert in [9] beschrieben.

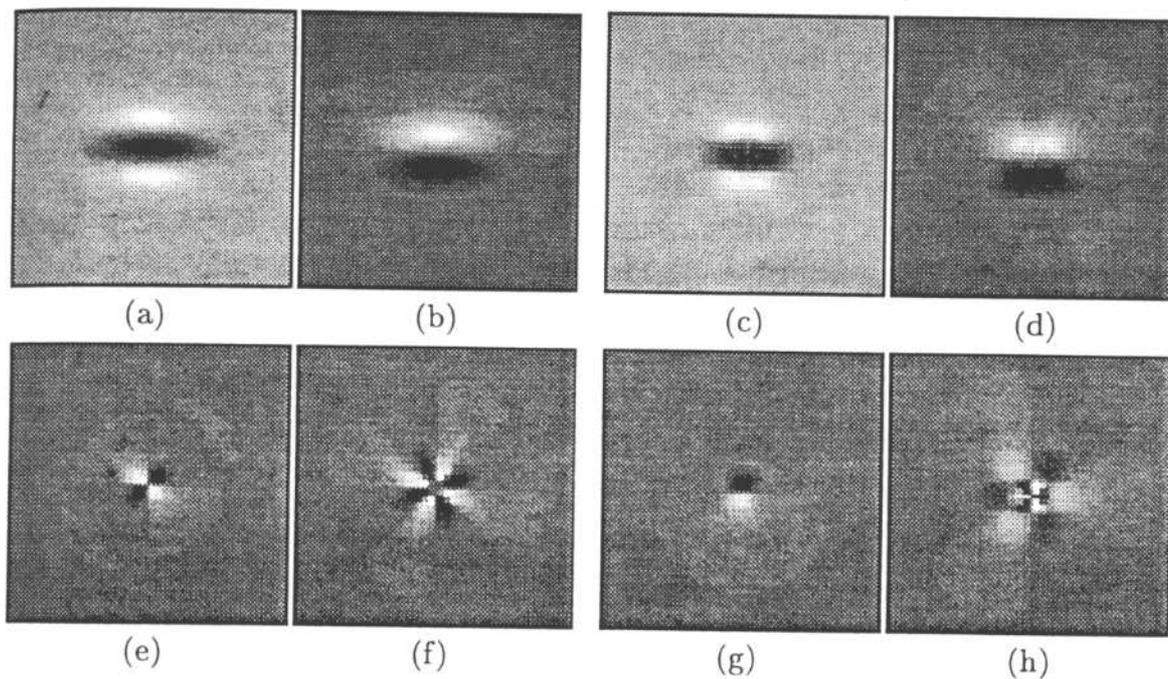


Abbildung 2. Linien- und Kantendetektionsfilter (a,b). Es handelt sich dabei um die erste (b) und zweite (a) Ableitung einer Gaußfunktion mit einem Seitenverhältnis von 2:1. Der Faltungskern hat eine Größe von 27x27 Pixeln. In (c) und (d) sind Approximationen dieser Filter unter Verwendung von 13 bzw. 10 Basisfunktionen abgebildet. Beispiele für Basisfunktionen (e-h) zur Steuerung der Filter von (a,b) in Orientierung und Skala.

2.2 Eigenschaften der Basisfunktionen

Die hier verwendeten Basisfunktionen erlauben eine einfache Anpassung der Approximationsqualität und der Geschwindigkeit. Dies beruht auf folgenden Eigenschaften:

- Die Basisfunktionen sind orthogonal, d.h. für eine bessere Approximationsqualität können Basisfunktionen hinzugenommen werden, ohne die Interpolationsfunktionen ändern zu müssen. Die neuen Funktionen bzw. Projektionskoeffizienten werden einfach zu den alten addiert.
- Eine beliebige Anzahl von Basisfunktionen rekonstruiert alle deformierten Filter. Lediglich die Approximationsqualität ist abhängig von der Anzahl der verwendeten Basisfunktionen. Deshalb kann die Geschwindigkeit der Merkmalsgewinnung gesteigert werden, indem anfangs grobe, aber schnelle Approximationen berechnet werden und erst bei Bedarf genauere.

3 Detektion markanter Punkte

3.1 Filtergrundoperationen

Die Detektion der markanten Punkte basiert auf der modellbasierten Suche und Verfolgung von Kanten- und Linienstrukturen in den Bildregionen. Die Kanten- und Liniendetektion wird durch die folgenden drei Grundoperationen realisiert, welche die eingeführte Steuerbarkeitseigenschaft der verwendeten Filter nutzen.

- Die erste Filtergrundoperation (BFO1) sucht eine Kante oder Linie bei vorgegebener Orientierung und Skala in einer definierten Region (Abb. 3a).
- Die zweite Filtergrundoperation (BFO2) bestimmt die Orientierung einer Linie oder Kante an einer vorgegebenen Position durch die maximale Antwort eines rotierten Filters, der bereits bzgl. der Skala an die gesuchte Struktur angepaßt ist. (Abb. 3b).
- Die dritte Filtergrundoperation (BFO3) verfolgt eine Linie oder Kante für einen Schritt. Dazu wird der Filter etwas in die Richtung des Verlaufs der bereits detektierten Struktur verschoben und anschließend senkrecht zu der neuen Position nach einer Kante oder Linie gesucht (Abb. 3c).

Diese Filtergrundoperationen werden in verschiedenen Orientierungen und Skalen verwendet und zu komplexeren Operationen zusammengesetzt.

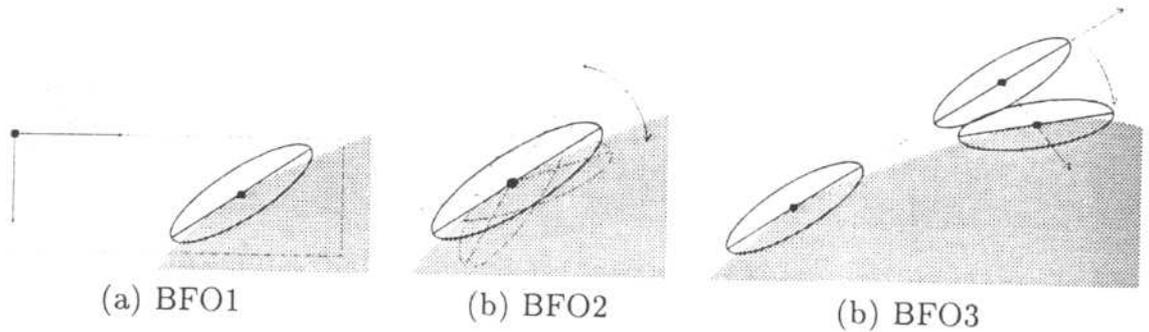


Abbildung 3. Filtergrundoperationen. BFO1: Detektion einer Linie oder Kante (a). BFO2: Bestimmung der Orientierung einer Linie oder Kante (b). BFO3: Stufenweise Verfolgung einer Linie oder Kante (c).

3.2 Definition des Modellwissens

Die Anwendung unserer Verfahren zielt auf die Detektion von Dysmorphiezeichen in Gesichtsbildern zur Unterstützung einer medizinischen Diagnose [12, 13]. Die Anwendung von Modellwissen bezieht sich daher auf die durch die konkrete Aufgabenstellung gegebene, bestimmte Klasse von Bildern. Zulässige Bilder zeigen frontale Gesichtsaufnahmen hoher Auflösung (512^2), frontale diffuse Beleuchtung, keine Brillen, Bärte oder Verdeckungen.

Das Augenmodell besteht aus einem Kanten- und Linienmodell der für eine robuste Kantenverfolgung notwendigen Kanten und Linien der Augenregion (Abb. 4). So ist etwa der Irisrand ein prominentes, gekrümmtes weiß-schwarzes Kantensegment vertikaler Orientierung, welches durch die horizontalen Kanten der Augenlieder begrenzt ist. Das Modell beinhaltet dabei auch einige Details, die zu einer genauen Lokalisierung der anatomisch definierten Punkte notwendig sind. Eine vollständige Beschreibung aller zugrundeliegenden Detailannahmen befindet sich in [7].

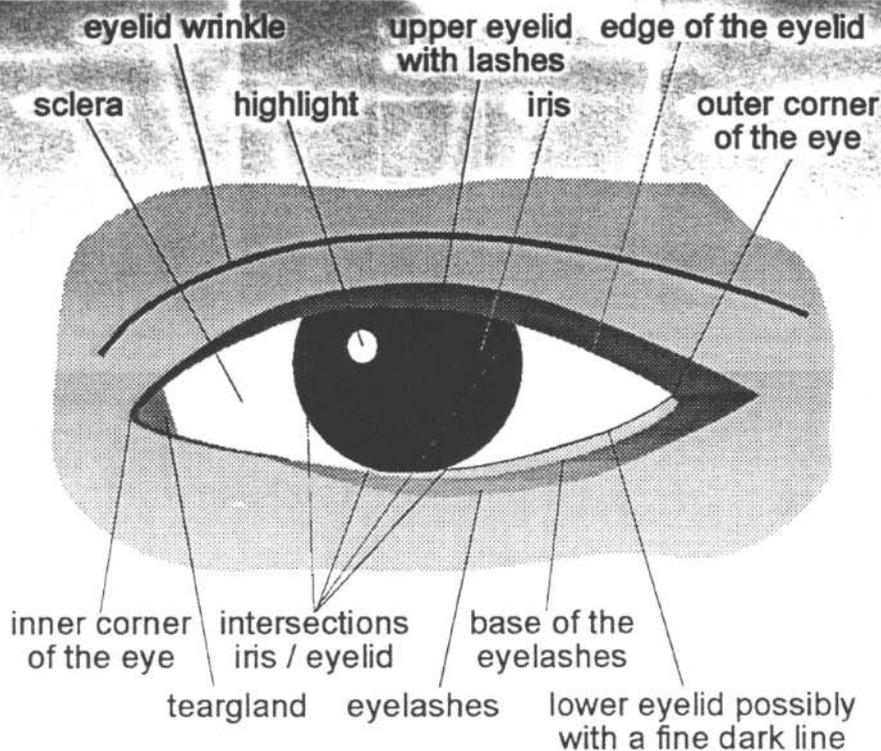


Abbildung 4. Modell des linken Auges. Die dargestellten Details werden bei der sequentiellen Suche und bei Konsistenzüberprüfungen verwendet.

3.3 Sequentielles Suchschema

Zur Verdeutlichung des Prinzips der sequentiellen Suchstrategie wird im folgenden die Detektion der Iris genauer erklärt (siehe Abb. 5). Jeder Schritt beinhaltet dabei verschiedene Anwendungen der bereits vorgestellten Filtergrundoperationen (Kap. 3.1) wobei deren Auswahl und Parametrisierung hinsichtlich Filtertyp (Kante, Linie, lokale Energie), Skala und Orientierung bzw. Richtungskorridor von den Ergebnissen vorhergehender Detektionsschritte zusammen mit dem Modellwissen festgelegt wird.

Die sequentielle Suchstrategie startet mit der Detektion des linken Kantensegments der Iris (Abb. 5a). Dazu wird nach einer markanten Hell/Dunkel-Kante mit nahezu vertikaler Orientierung mit der Filtergrundoperation BFO1 gesucht. In diesem ersten Verarbeitungsschritt wird die ganze Augenregion betrachtet, während sich alle nachfolgenden Schritte lediglich auf kleine begrenzte Suchregionen in Abhängigkeit bereits bekannter Informationen beschränken. Um nun ein Kantensegment eines Glanzlichtes auf der Iris von einer Kante des Irisrandes unterscheiden zu können, wird das Kantensegment in beide Richtungen mit der Filtergrundoperation BFO3 verfolgt. Das Kantensegment wird solange verfolgt, bis ein signifikanter Wechsel der Filterantwort das Ende der Kante signalisiert. An der Krümmung des detektierten und verfolgten Kantensegments kann nun zwischen einem Kantensegment eines Glanzlichtes und dem des Irisrandes unterschieden werden. In einem ersten Bewertungsschritt wird der Radius und das Zentrum der Iris initial geschätzt. Die bereits abgeleiteten Information wird nun zur Definition der Suchregion für das zweite, rechte Kantensegment der Iris verwendet (Abb. 5b). Die Detektion des Kantensegments erfolgt wiederum durch Einsatz der Filtergrundoperation BFO1 und durch Kantenverfolgung mit Hilfe von Filtergrundoperation BFO3 (Abb. 5c). Anschließend wird überprüft, ob bei-

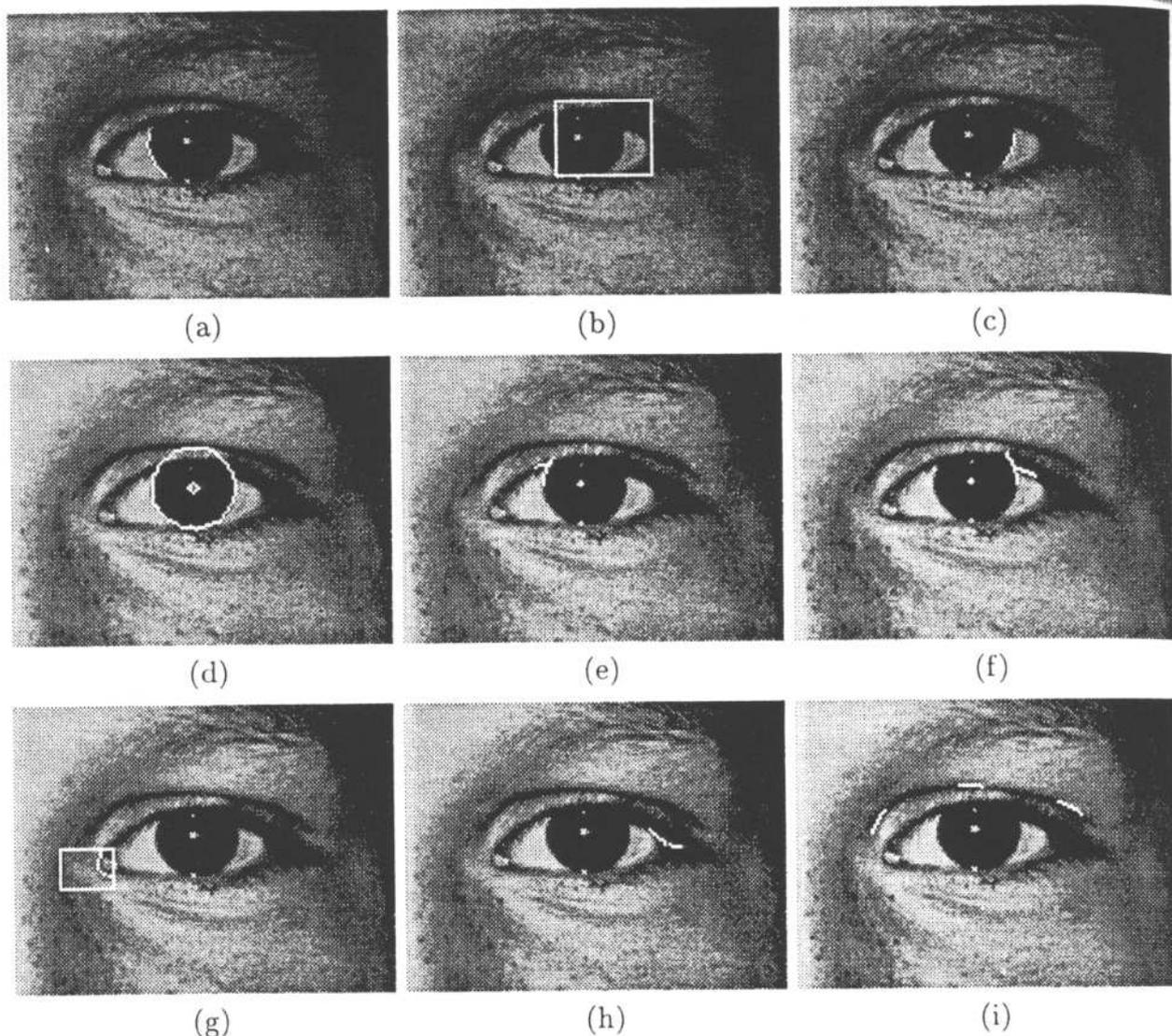


Abbildung 5. Sequentielle Suchstrategie, dargestellt an einem linken Auge. Die während der Laufzeit abgeleiteten Informationen über Skala, Ort und Orientierung einzelner Kantensegmente werden zur Suche weiterer Kantensegmente wiederverwendet. (a) Detektion der linken Iriskante. (b) Suchbereich für das dazugehörige rechte Kantensegment. (c) Detektion der rechten Iriskante. (d) Aus den detektierten Kantensegmenten berechneter Irisradius und Iriszentrum. (e,f) Detektion der Schnittpunkte mit den Augenlidern. (g) Suchbereich und Segmentation des inneren Augenwinkels. (h) Verfolgung der unteren Oberlidkante, um die äußere Augenecke zu detektieren. (i) Bestimmung der Orientierungen der Lidfalte.

de Kantensegmente zu ein und demselben Kreis gehören, wodurch eine genaue Schätzung des Radius und des Zentrums der Iris ermöglicht wird. Die Vorgehensweise zur Detektion der anderen Strukturen und markanten Punkte in der Augenregion (Abb. 5) orientiert sich in entsprechender Art und Weise an dem repräsentieren Modellwissen und setzt abhängig von der aktuellen Datenlage gezielt die Filtergrundoperationen zu deren Detektion und Verfolgung ein.

3.4 Detektionsergebnisse

Das vorgestellte Verfahren wurde an 98 Augenregionen von verschiedenen Gesichtsbildern getestet, die alle die oben aufgeführten allgemeinen Voraussetzungen erfüllten (siehe Kap. 3.2). Der gesamte Detektionsfehler für die vollständige Detektion der Iris betrug 2.35%. In 94.4% der verbleibenden Bildausschnitte

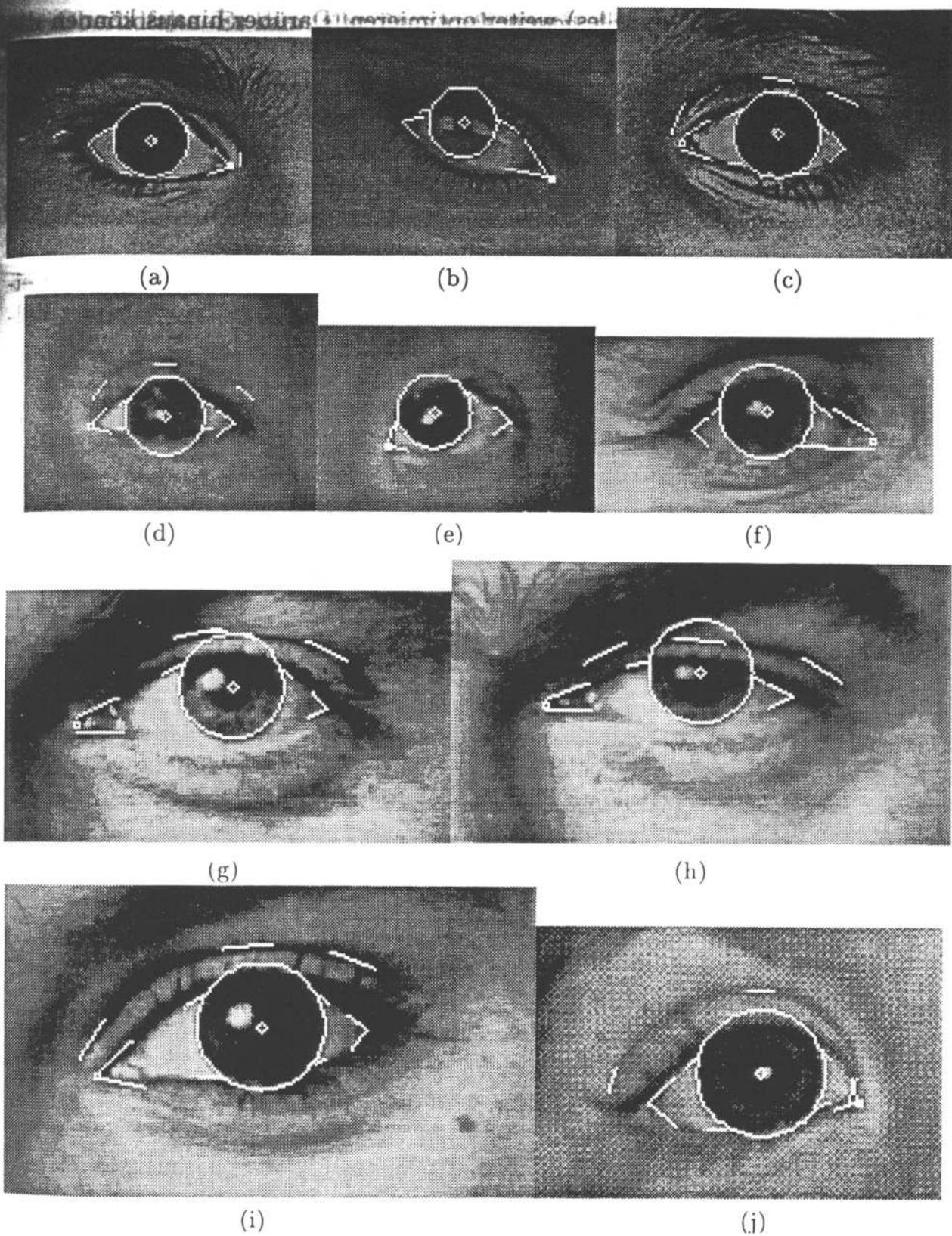


Abbildung 6. 10 Beispiele von erfolgreich untersuchten Augenregionen. Alle Bilder besitzen den gleichen Maßstab.

konnte die äußere Augenecke erfolgreich detektiert werden, während die innere Augenecke in 91.6% der Fälle richtig gefunden werden konnte. Der höhere Detektionsfehler bei der inneren Augenecke erklärt sich durch ihre komplexere Struktur und durch die höhere Variabilität der Daten, mit verursacht durch die Störungen, die von den kleinen Glanzlichtern auf dem Tränensack ausgehen (Abb. 1). Das sequentielle Suchschema läßt sich aber durch weitere Konsistenzüberprüfun-

gen (Verfeinerung des Modelles) weiter optimieren. Darüber hinaus können die Detektionsfehler durch einen sich anschließenden Verifikationsschritt der gefundenen Bildposition reduziert werden (siehe Kap. 4).

Die große Variabilität der erfolgreich untersuchten Augenregionen ist an 10 repräsentativen Beispielen in Abb. 6 dargestellt. Alle abgebildeten Augenregionen wurden mit derselben Parametereinstellung des sequentiellen Suchschemas analysiert. Eine Beispielregion, die sich nicht in horizontaler Orientierung befindet, wird in Abbildung 6b gezeigt. Der Rotationswinkel beträgt über 20° .

Das Verfahren ist unempfindlich gegen eine Variation der Augengröße in einem weiten Bereich. Augenregionen unterschiedlicher Größe sind in den Abbildungen 6i (Irisdurchmesser = 53 Pixel) und 6e (Irisdurchmesser = 28 Pixel) dargestellt. Der Irisdurchmesser des größten mit unserem Verfahren erfolgreich untersuchten Auges beträgt 104 Pixel (nicht abgebildet).

4 Verifikation markanter Bildpunkte

Zur Verifikation der markanten Bildpunkte, die mit Hilfe des sequentiellen Suchschemas gefunden wurden, wird ein neuronaler Klassifikator eingesetzt [5]. Dieser basiert auf einer dynamischen Zellstruktur, *Dynamic Cell Structures* (DCS-Netzwerk), die durch überwachtetes Lernen trainiert wird [1].

4.1 Grundlagen der DCS-Netzwerke

Die in [1] eingeführten *Dynamic Cell Structures* (DCS) gehören zu der Klasse der RBF-basierten Approximationsverfahren, von denen sie sich dadurch unterscheiden, daß sie zusätzlich die Topologie der Eingabe-Mannigfaltigkeit in Form von *Perfekt Topologie-erhaltenden Merkmalskarten* (PTM) erlernen [8]. Diese nutzen sie für verbesserte Approximationseigenschaften und Lernverfahren. DCS-Netzwerke weisen folgende Architektur auf:

- eine verdeckte Schicht von RBF Knoten (evtl. dynamisch wachsend oder schrumpfend),
- eine sich dynamisch an die Topologie anpassende laterale Verbindungsstruktur zwischen den Knoten und
- eine Schicht meist linearer Ausgabeeinheiten.

Die in diesem Artikel verwendeten *Growing Dynamic Cell Structures* (GD-CS), die ähnlich zu den von Fritzke [2] durch fehlerbasiertes Einfügen neuer RBF Knoten wachsen, vermeiden damit die Unzulänglichkeiten der klassischen Kohonen-Karten: Die starre Topologie des Netzes (2- oder 3-dimensional), die vorgegebene Anzahl von Knoten und die rein von der Eingabewahrscheinlichkeitsdichte abhängige Verteilung der Knoten. Ihre Eleganz, Einfachheit und problemabhängige Struktur gaben den Ausschlag für die Verwendung des DCS-Netzwerkes als Verifikator, zumal sie sich in einer Reihe von Klassifikations-Benchmarks herkömmlichen Klassifikatoren überlegen erwiesen [1].

4.2 Aufbau und Training des DCS-Netzwerkes

In unserer Anwendung unterscheiden wir 9 Klassen charakteristischer Bildpunkte: Innerer und äußerer Augenwinkel (jeweils für das rechte und linke Auge getrennt), 4 Schnittpunkte des Irisrandes mit den Lidern und einen Punkt auf

der Oberlidfalte. Es kann allerdings vorkommen, daß die Schnittpunkte der Iris mit dem Unterlid nicht in jeder Augenregion vorhanden und somit nicht zu detektieren und zu verifizieren sind. Obwohl die meisten Pixel in einer Augenregion keine charakteristischen Punkte sind, wird eine Restklasse nicht explizit berücksichtigt. Damit wird verhindert, daß das DCS-Netzwerk nicht die meisten Neurone zur Repräsentation einer Klasse verwendet, an der wir nicht interessiert sind. Stattdessen wird die Restklasse implizit durch einen relativen, klassenspezifischen Schwellwert definiert. Liegt die Netzaktivierung einer Bildposition bei der Klassifikation unterhalb dieses Schwellwertes, wird sie der Restklasse zugeordnet. Entsprechend werden nur Merkmalvektoren von vorselektierten charakteristischen Bildpunkten für das Training des DCS-Netzwerks verwendet.

Als Merkmalvektor für das Training des DCS-Netzwerks und auch für die Klassifikation eines unbekanntes Bildpunktes dienen die ersten 70 Projektionskoeffizienten, die durch Projektion des charakteristischen Bildpunktes und seiner lokalen Umgebung auf die Basisfunktionen des steuerbaren Filterschemas berechnet werden (vgl. Kap. 2). In den Projektionskoeffizienten ist die wesentliche Kanten- und Linieninformationen der lokalen Bildstruktur, die auch für die Verifikation wichtig ist, enthalten. Da diese Koeffizienten ohnehin für die Detektion der Punkte berechnet wurden (vgl. Kap. 3) entsteht kein zusätzlicher Berechnungsaufwand. Die Verwendung der Projektionskoeffizienten anstelle der direkten Bildinformation (originale Grauwertverteilung in der 27×27 Umgebung) ermöglicht darüber hinaus eine erhebliche Dimensionsreduzierung des Merkmalraumes, ohne auf die wesentliche Information der lokalen Bildstruktur verzichten zu müssen.

668 Merkmalvektoren wurden aus insgesamt 98 Augenregionen von rechten und linken Augen abgeleitet, davon bilden 527 Merkmalvektoren die Trainingsmenge und 141 Merkmalvektoren die Testmenge.

Der Klassifikationsfehler des Netzwerkes (gemittelt über alle Klassen) wird sowohl bezüglich der Trainingsmenge als auch bezüglich der Testmenge in Abhängigkeit zur Anzahl der Neurone, die während der Trainingsphase in das Netz eingefügt wurden, in Tabelle 1 dargestellt. Das Einfügen zusätzlicher Neurone wird beendet, wenn der Klassifikationsfehler für die Trainingsmenge den Wert 0 erreicht und der Klassifikationsfehler für die Testmenge nicht weiter sinkt. Durch weiteres Einfügen von Neuronen würde die Generalisierungsfähigkeiten des Netzwerkes leiden.

4.3 Der Bayes'sche Verifikationsansatz

Die Detektion und die Verifikation eines charakteristischen Punktes der Klasse C_i können als hintereinander ausgeführte, unabhängige Zufallsexperimente betrachtet werden. Im Falle der sequentiellen Suche besteht das Zufallsexperiment aus dem Finden der Punktkoordinaten des markanten Punktes, der gerade aufgesucht werden soll. Bei der Verifikation wird die Klasse des gerade betrachteten Bildpunktes bestimmt und der Zufallsvariablen Y zugewiesen. Die Wahrscheinlichkeit, daß die Bildposition, an der die sequentielle Suche terminiert, der Klasse C_i angehört, ist somit die a priori Wahrscheinlichkeit $P(X = C_i)$. Die Fehlerwahrscheinlichkeit, daß das sequentielle Suchschema einen Punkt detektiert,

Netzwerkgröße						
Anzahl Neurone	21	34	60	90	120	142
Trainingsmenge (527 Vektoren):						
Anzahl Fehler	52	41	31	19	7	0
Fehler (%)	9.8	7.7	5.8	3.6	1.3	0
Testmenge (141 Vektoren):						
Anzahl Fehler	11	10	10	9	8	8
Fehler (%)	7.8	7.0	7.0	6.3	5.6	5.6

Tabelle 1. Klassifikationsfehler bezüglich der Größe des DCS-Netzwerkes.

der nicht zur Klasse C_i gehört, ist somit $P(X \neq C_i)$. Wenn nun Y die Klasse bezeichnet, die der betrachteten Bildposition durch die Verifikation zugeordnet wird, dann ist $P(Y = C_i | X \neq C_i)$ die bedingte Wahrscheinlichkeit für eine falsch positive Klassifikation und $P(Y \neq C_i | X = C_i)$ die bedingte Wahrscheinlichkeit für eine falsch negative Klassifikation. Nach dem Satz von Bayes läßt sich nun die a-posteriori-Fehlerwahrscheinlichkeit, daß eine detektierte Bildposition nicht der gesuchten Klasse angehört, obwohl dies die Verifikation nicht bestätigt hat, berechnen als:

$$\begin{aligned}
 P(X \neq C_i | Y = C_i) &= \frac{P(Y = C_i | X \neq C_i) P(X \neq C_i)}{P(Y = C_i)} \\
 &= \frac{P(Y = C_i | X \neq C_i) P(X \neq C_i)}{P(Y = C_i | X \neq C_i) P(X \neq C_i) + P(Y = C_i | X = C_i) P(X = C_i)}
 \end{aligned} \quad (2)$$

Eine Verifikation ist nur dann sinnvoll, wenn die a posteriori Fehlerwahrscheinlichkeit $P(X \neq C_i | Y = C_i)$ kleiner ist, als die a priori Fehlerwahrscheinlichkeit $P(X \neq C_i)$. Tabelle 2 zeigt, daß die Fehlerrate durch die zusätzliche Anwendung des Verifikationssystems erheblich verkleinert werden konnte. Dies gilt sowohl für die klassenspezifischen Fehlerwahrscheinlichkeiten als auch für die gesamte Fehlerwahrscheinlichkeit, die von 17% auf unter 1% gesunken ist. Die in Tabelle 2 angegebenen empirischen Fehlerwahrscheinlichkeiten beziehen sich jeweils auf ein von der ursprünglichen Testmenge (Tab. 1) abweichendes Kollektiv. Insbesondere die a priori Fehlerwahrscheinlichkeiten der sequentiellen Suche beziehen sich abweichend zu den in Kap. 3.4 angegebenen Ergebnissen nur auf die Augenregionen, die im Testkollektiv des DCS-Netzwerkes berücksichtigt wurden. Eine direkte Detektion der markanten Punkte mit Hilfe des DCS-Netzwerkes ist zwar möglich, kommt dagegen über eine durchschnittliche Erfolgsquote von ca. 50% nicht hinaus [5].

5 Zusammenfassung und Diskussion

Die in dieser Arbeit vorgestellte Vorgehensweise ermöglicht eine robuste und präzise Lokalisierung von markanten Punkten in Gesichtsbildern. Am Beispiel der Augenregion wurde die Leistungsfähigkeit der entwickelten Verfahren gezeigt. Dieselbe Methodik wurde darüberhinaus auch zur Detektion der Mund-ecken und anderer markanter Punkte im Gesichtsbild angewendet [6].

char. Bildpos.	$\hat{P}(X \neq C_i)$	$\hat{P}(Y = C_i X \neq C_i)$	$\hat{P}(Y \neq C_i X = C_i)$	$\hat{P}(X \neq C_i Y = C_i)$
Augenecke	19.0	1.5	0.0	0.0
Augenecke	12.5	1.9	4.8	0.0
Schnittp. Iris/Lid	19.4	0.7	5.4	1.4
Augenlidfalte	13.3	1.1	0.0	0.0
gesamt	17.0	1.1	3.8	0.8

Tabelle 2. Empirische Fehlerwahrscheinlichkeiten in %.

Das Hauptproblem der Detektion markanter Punkte in realen Bilddaten besteht in der großen Variabilität der vorhandenen Bildstrukturen, welche den gesuchten Punkten bei verschiedenen Individuen und unterschiedlichen Aufnahmebedingungen zugrundeliegt. Eine rein daten-getriebene Detektion ist deshalb ausgeschlossen. Eine bekannte Möglichkeit zur Berücksichtigung des Bildkontextes sind Deformable Templates [14]. Für eine sehr präzise Lokalisierung der Augenecken besitzen diese jedoch eine zu geringe Flexibilität verglichen mit der interindividuellen Variabilität der Augenregionen.

Die relative hohe Detektionsfehlerrate der sequentiellen Suche läßt sich durch dieses hohe Maß an Variabilität der Bildstrukturen in Real-Weltbildern erklären. Der Detektionsfehler kann aber durch den Einsatz eines zusätzlichen Verifikationsschrittes drastisch minimiert werden, wie hier gezeigt werden konnte. Der Einsatz eines neuronalen Klassifikators auf der Grundlage einer dynamischen Netzstruktur ermöglicht eine optimale Diskriminierung des hochdimensionalen Entscheidungsraumes, wobei auf bereits vorliegende Merkmaldaten zurückgegriffen werden kann.

Das hier vorgeschlagene sequentielle Suchschema zur Detektion der markanten Bildpunkte berücksichtigt einerseits globale Zusammenhänge durch die modellgestützte Kantenverfolgung, erlaubt aber andererseits eine genügende Anpassungsfähigkeit an die lokalen Verhältnisse der Suchumgebung zur präzisen Lokalisierung der markanten Punkte. Durch die Verwendung von Filtern mit hoher Orientierungsselektivität ist die Kantenverfolgung in einer vorgegebenen Richtung robuster als mit entsprechenden isotropen Filtern. Isotrope Kantenfilter verlieren bei durch Falten, Wimpern oder Schatten verursachten Verzweigungspunkten leicht die vorgegebene Richtung und folgen dann nicht mehr der gewünschten Kante.

Die Adaptivität des verwendeten Filterschemas stellt eine wesentliche Innovation gegenüber einer auf 'einfachen' Kantendetektoren basierenden Kantenverfolgung dar. Durch die Berücksichtigung von klassenspezifischem Modellwissen ermöglicht sie gleichzeitig eine hohe Leistungsfähigkeit und Präzision. Das verwendete Filterschema trägt dabei wesentlich zur Robustheit und Effizienz der Vorgehensweise bei. Die Ergebnisse der Projektionen mit den Basisfunktionen werden nicht nur explizit für die Detektion der markanten Punkte verwendet sondern zusätzlich zur Generierung der Verifikationsresultate eingesetzt. Durch die Wahl des verwendeten Filterschemas werden nur die für die Detektion und für die anschließende Verifikation notwendigen Strukturinformationen abgeleitet und repräsentiert. Dies stellt im Vergleich zur originalen Grauwertverteilung einer 27×27 -Umgebung eine hohe Datenreduktion dar.

Darüber hinaus läßt sich die Idee der hybriden Vorgehensweise der modellbasierten, sequentiellen Suche, gefolgt von einer daten-getriebenen Verifikation, auf der Grundlage eines neuronalen Netzes auf andere Fragestellungen übertragen, in denen lokale Strukturen präzise zu detektieren sind.

Danksagung Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft unter den Aktenzeichen: So 320\1-1 and Ei 322\1-1 gefördert. Wir möchten uns bei Frau Prof. S. Stengel-Rutkowski und Herrn Dr. C. Apacik vom Kinderzentrum München für die bereitwillige Unterstützung der Arbeiten an dem Projekt bedanken.

References

1. J. Bruske and G. Sommer, *Dynamic cell structure learns perfectly topology preserving map*, Neural Computation, Vol. 7, no. 4, pp. 845-865, 1995.
2. B. Fritzsche, *Growing cell structures - a self organizing network for unsupervised and supervised training*, ICSI Berkeley, Tech.-Rep., tr-93-026, 1993.
3. R. Herpers, H. Kattner, and G. Sommer, *GAZE: Eine attentive Verarbeitungsstrategie zum Erkennen und Lernen der visuell auffälligen Gesichtsregionen*, in: Mustererkennung 1994, W.G. Kropatsch and H. Bischof (Eds.), 341-349, 1994.
4. R. Herpers, H. Kattner, H. Rodax, and G. Sommer, *GAZE: An attentive processing strategy to detect and analyze the prominent facial regions*, Proc. Int. Works. on Auto. Face- and Gesture-Rec., Zurich, Switzerland, 214-220, 1995.
5. R. Herpers, L. Witta, J. Bruske and G. Sommer, *Evaluation of local image structures applying a DCS network*, in: Solving Engineering Problems with Neural networks, Proc. 2nd int. Conf. EANN'96, A.B. Bulsari et al. (Eds.), 305-312, 1996.
6. R. Herpers, M. Michaelis, K.H. Lichtenauer and G. Sommer, *Edge and keypoint detection in facial regions*, eingereicht zur 2. int. Conf. on Automatic Face and Gesture Recognition, 14.-16.10.96, Killington, Vermont.
7. R. Herpers, M. Michaelis, L. Witta, and G. Sommer, *Context Based Detection of Keypoints and Features in Eye Regions* angenommen auf der 13. ICPR'96, 25.-29.08.96, Wien, Österreich.
8. T. Martinetz, *Competitive Hebbian learning rule forms perfectly topology preserving maps*, Proc. ICANN 93, pp. 426-438, 1993.
9. M. Michaelis, *Low level image processing using steerable filters*, PhD thesis, Christian-Albrechts-Universität, D-24105 Kiel, Germany, 1995.
10. P. Perona, *Steerable-scalable kernels for edge detection and junction analysis*, EC-CV'92, G. Sandini (Ed.), LNCS 588, Springer-Verlag, 3-18, 1992.
11. A. Samal and P.A. Iyengar, *Automatic recognition and analysis of human faces and facial expressions: A survey*, Pattern Recognition, Vol. 25, 65-77, 1992.
12. P. Schimanek, *Anthropologische und anthropometrische Definition von Dysmorphiezeichen als Grundlage der Diagnostik von Dysmorphiesyndromen*, PhD thesis, Universität München, Germany, 1988.
13. S. Stengel-Rutkowski, P. Schimanek, and A. Wernheimer, *Anthropometric definitions of dysmorphic facial signs*, Hum. Genet, vol.67, 272-295, 1984.
14. A.L. Yuille, D.S. Cohen, and P.W. Hallinan, *Feature extraction from faces using deformable templates*, Proc. of IEEE Conf. CVPR '89, 104-109, 1989.

und anderer markanter Punkte im Gesichtsbild.