Verwendung von attentiven Verarbeitungsstrategien zur lokalen und globalen Bildanalyse

R. Herpers[1], H. Kattner[1], H. Rodax[1] und G. Sommer[2]

1 GSF-Institut für medizinische Informatik und Systemforschung, Medis, Postfach 1129
D-85758 Oberschleißheim, Email: herpers@gsf.de

2 Institut für Informatik, Lehrstuhl für Kognitive Systeme, Christian-Albrechts-Universität Kiel, Preußerstr. 1-9 D-24105 Kiel

In diesem Beitrag wird ein attentives Verarbeitungssystem vorgestellt, welches durch die Simulation physiologischer Blickbewegungsstrategien die visuell auffälligen Bildregionen von Gesichtsbildern lokalisiert und identifiziert. Dazu werden aus den Orginalbilddaten visuell relevante Kantenmerkmale abgeleitet und diese in einer Skalenhierarchie repräsentiert. Auf der Grundlage der entwickelten Attentionsstrategie werden sequentiell die interessanten Bildbereiche lokalisiert. Die ausgewählten Bildausschnitte werden normalisiert und anschließend auf der Basis der erzeugten Repräsentation einer topologischen Merkmalskarte klassifiziert. Das durch die Klassifikation abgeleitete Wissen wird zur Steuerung der weiteren Attentionsstrategie verwendet.
Der vorgestellte Attentionsmechanismus ist ein wesentlicher Bestandteil eines umfassenden Bildverarbeitungssystems, welches Gesichtsbilder mit dysmorphen Abnormalitäten analysieren und typisieren soll. Es wird gezeigt, daß auf der Grundlage einer attentiven Verarbeitungsstrategie unter Integration von verschiedenen Invarianzeigenschaften eine Objektsuche effizient durchgeführt werden kann. Das vorgestellte Verfahren kann darüber hinaus universell für eine Lokalisation von unterschiedlichen attentiven Regionen in beliebigen Bilddaten eingesetzt werden.

1. Einleitung

Der Aufwand der Objektsuche in Realwelt-Bildern hat im Falle einer vollständigen Suche, bei der die gesamte Bildmatrix unter Variation der Parameter der Freiheitsgrade sequentiell durchsucht werden muß, eine NP-vollständige Zeitkomplexität [Tsotsos89]. Eine aufgabenorientierte Suche auf der Grundlage selektiver visueller Aufmerksamkeit benötigt dagegen lediglich linearen Aufwand, der von der Anzahl der enthaltenen Objekte in den ausgewählten Bildausschnitten abhängig ist [Tsotsos90]. Die Verwendung von attentiven Verarbeitungsstrategien ermöglicht daher eine sequentielle Verarbeitung und Analyse lokal begrenzter Bildbereiche und eröffnet zusätzliche Informationen über die in dem Bildbereich enthaltenen Objekte [Sandon89].
Motiviert durch die Blicksteuerung biologisch visueller Systeme werten künstliche Attentionssysteme einfache Merkmale auf einer groben Auflösungsstufe aus und sind somit in der Lage, aus den extrahierten Informationen die Blickbewegung auf eng begrenzte, für die aktuelle Fragestellung interessante Bildausschnitte zu steuern [Koch85, Califano90, Culhane92, Olshausen92]. Irrelevante Bildregionen können von einer weiteren Bearbeitung ausgeschlossen werden und erfordern somit keinen weiteren Verarbeitungsaufwand.
In biologisch visuellen Systemen werden während einer frühen präattentiven Verarbeitungsstufe elementare visuelle Merkmale wie Bewegung, Farbe, Orientierung, Kanten etc. abgeleitet. Diese Merkmale erzeugen eine präattentive Repräsentation, die die Grundlage für die weitere Steuerung der visuellen Aufmerksamkeit darstellt [Treisman85]. Ein attentiver Verarbeitungsmechanismus lokalisiert auf dieser Grundlage verschiedene, lokal begrenzte Bildbereiche und erzeugt somit sowohl eine sinnvolle Vorauswahl als auch eine sequentielle Ordnung der Bildausschnitte, deren weitere Betrachtung lohnenswert erscheint.
In dem hier vorliegenden Beitrag wird eine technische Realisierung eines Attentionsmechanismus vorgestellt, welche durch die Simulation physiologischer Blickbewegungsstrategien die visuell auffälligen Bildregionen von Gesichtsbildern (Augen, Nase, Mund und Ohren) lokalisiert. Anschließend wird in dem Verarbeitungssystem versucht semantische Teilinterpretationen dieser

extrahierten Bildbereiche durchzuführen. Dazu werden aus den Orginalbilddaten die visuell relevanten Kantenmerkmale abgeleitet, diese auf unterschiedlichen Skalen repräsentiert, dort mit Hilfe einer Attentionsstrategie ausgewertet und anschließend die ausgewählten Bildbereiche klassifiziert. Zur Klassifikationen der einzelnen Bildausschnitte wird ein assoziatives Verfahren eingesetzt. Durch ein neu entwickeltes, selbstüberwachtes Lernverfahren wird eine Merkmalrepräsentation der einzelnen Gesichtsregionen wie Augen-, Nasen-, Mund- und Ohrenregion auf der Merkmalkarte erzeugt. Auf der Basis der dadurch möglichen ersten Interpretationen wird in einem weiteren Schritt eine lokale Analyse von besonders charakteristischen Bildpunkten (keypoints) wieder unter Verwendung einer attentiven Strategie durchgeführt. Diese sowohl datengetriebene als auch durch Top-Down-Wissen gesteuerte, selektive Aufmerksamkeit ist in der Lage diese Schlüsselpunkte automatisch zu lokalisieren und zu identifizieren. Diese Vorgehensweise stellt somit eine Verknüpfung von attentiven Verarbeitungsmechanismen mit dynamisch generiertem semantischen Wissen dar, d. h. es wird nach einer unwillkürlichen Lokalisierung lokal begrenzter Bildregionen die Aufmerksamkeit auf visuell relevante Bildbereiche zur weiteren effektiven Merkmalsanalyse und Bildinterpretation konzentriert.
Der vorgestellte Ansatz ist motiviert durch Untersuchungen der Arbeitsgruppe von J. Tsotsos, der einen Prototyp für einen datengetriebenen visuellen Attentionsmechanismus entwickelt hat [Tsotsos91, Culhane92]. Die Betonung des hier vorgestellten Ansatzes liegt auf einer effizienten Realisierung einer Attentionsstrategie, die die prominenten Bildbereiche in Gesichtsbildern lokalisiert und interpretiert. Das beschriebene Attentionssystem stellt dabei einen wesentlichen Bestandteil eines umfassenden Bildanalysesystems dar, welches Gesichtsbilder mit dysmorphen Abnormalitäten (Dysmorphien) identifizieren und typisieren soll [Stengel85].

2. Adaptive Regionensuche in Verarbeitungshierarchien

Auf der Basis frontaler Gesichtsaufnahmen werden aufgabenbezogene Bildmerkmale abgeleitet und auf einer Merkmalkarte kodiert. Die Bilddaten liegen als 8-bit Grauwertbilddaten der Größe 256x256 oder 512x512 Pixel vor. Auf der Grundlage eigener Untersuchungen als auch der anderer Arbeitsgruppen [Brunelli92-93] stellt die Kanteninformation ein ausreichendes Merkmal für die Lokalisierung der visuell auffälligen Gesichtsregionen wie Auge, Nase, Mund und Ohren dar.

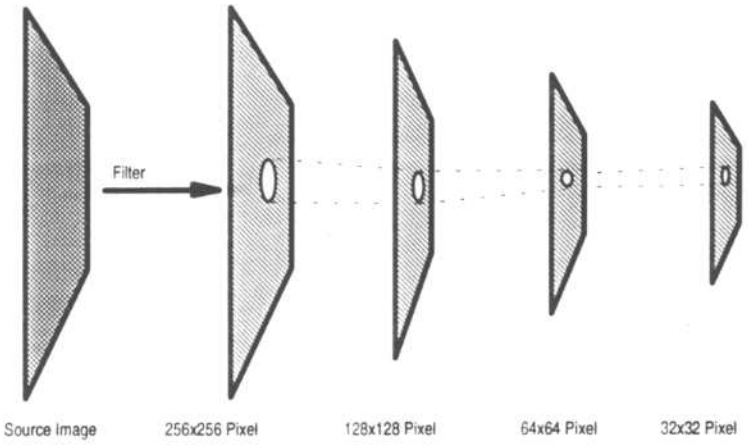


Abb.: 2.1 Modell der attentiven Suche einer Bildregion innerhalb einer oktaven Multiskalenrepräsentation. Die linke Skizze stellt das Originalbild dar, gefolgt von der Repräsentation der abgeleiteten Merkmale auf der Aufmerksamkeitskarte, die anschließend auf 3 zusätzlichen Skalenstufen repräsentiert wird. Die niedrigste Auflösungsstufe (rechts) stellt die Basisrepräsentation bzw. minimale Bildskizze der Aufmerksamkeitskarte dar, die als Ausgangspunkt für die anschließende attentive Suche verwendet wird. Die Propagierung eines lokalisierten und angepaßten Bildausschnittes ist in den einzelnen Skalen eingetragen.

In dem vorgestellten System werden Kanteninformationen auf der Basis von unterschiedlich skalierbaren Filtern (z. B 5x5 Sobel, 7x7 Blockkanten in X- und Y-Richtung) abgeleitet. Die abgeleiteten Merkmale können sowohl einzeln als auch in verknüpfter Form auf der sogenannten Aufmerksamkeitskarte repräsentiert werden. Aus dieser Merkmalkarte werden Repräsentationen unterschiedlicher Auflösungsstufen, sogenannte Skalenpyramiden, unter Verwendung von

Gaussfilterungen erzeugt. Die Repräsentation der Merkmale auf der niedrigsten Auflösungsstufe stellt eine Art Basisrepräsentation oder auch minimale Merkmalskizze dar, die als Ausgangspunkt für den sich anschließenden Attentionsmechanismus verwendet wird. Eine solche Repräsentation ist vergleichbar mit der der präattentiven visuellen Stimuli in biologischen Systemen, bei der die Reizsignale bereits in reduzierter Form repräsentiert sind. Als Ausgangspunkt für eine effektive Analyse einer Verarbeitungshierarchie von Kantenrepräsentationen wurden 4-stufige Oktavpyramiden bei 256x256 Gesichtsbildern und 5-stufige bei 512x512 gewählt (siehe Abb. 2.1).
Allein auf der Grundlage der Operatorantworten eines 5x5 Sobelfilters (ohne Orientierungsselektivität) und einer anschließenden 5-stufigen Repräsentation in einer Gausspyramide (512, 256, 128, 64, 32) läßt sich eine sinnvolle selektive Attentionsstrategie zur Lokalisation der markanten Gesichtsregionen erzeugen [Herpers93].
Die Auswahl einer attentiven Region beginnt mit einer Maximumsuche auf der minimalen Merkmalskizze der Aufmerksamkeitskarte. Anschließend wird die zu lokalisierende Bildregion ausgehend von dem angesteuerten maximalen Kartenelement soweit expandiert, daß die Größe der Bildregion der Ausdehnung der zugrundeliegenden Merkmale entspricht. Das bedeutet, daß die Größe der attentiven Region möglichst an die Ausdehnung der Repräsentation der Reizsignale des zugrundeliegenden Objektes angepaßt wird, so daß eine vollständige Überdeckung entsteht. Dieser so gewonnene Bildausschnitt wird anschließend auf die nächst höhere Auflösungsstufe propagiert, während gleichzeitig die Bildregion wieder an die Ausdehnung des zugrundeliegenden Objektes angepaßt wird (siehe Abb. 2.1).
Dies kann sowohl eine Vergrößerung als auch eine Verkleinerung des projezierten Bildausschnittes beinhalten. In dem hier vorgestellten Verfahren werden z. Z. lediglich horizontale und vertikale, rechteckige Bildausschnitte selektiert. Die Rechtecke werden während der Regionenanpassung jeweils an den Rändern abhängig von einem lokalen Gewicht erweitert oder verkleinert. Dieses Gewicht ist abhängig von der Fläche der Bildregion F(A) und seiner mittleren Merkmalstärke M(A). Die Expansion der Fläche wird über die Funktion F(A) gesteuert bzw. begrenzt.

$$f(A) \quad = \quad 1 \quad + \quad \frac{A-1}{a \quad (A-1) \quad + \quad b}; \qquad a, b > 0; \qquad (1)$$

mit A := Fläche des Bildausschnittes und a, b sind Parameter zur Erzielung eines asymptotischen Verhaltens der Wachstumseigenschaft.
Bei der Berechnung der Regionenanpassung wird eine Maximierung des Produktes der Funktionen F(A) und M(A) angestrebt. Dieses Produkt stellt somit ein gewichtetes Maß für die mittlere Merkmalstärke eines Bildausschnittes einer bestimmten Größe dar. Vergrößert sich dieses Maß bei der Erweiterung der Bildregion an einem Rand des berechneten Rechteckes so wird der Bildausschnitt expandiert, verkleinert sich das Maß so wird der Bildausschnitt reduziert.
Es müssen zwei globale Bedingungen bei der Berechnung und auch bei der Anpassung der rechteckigen Bildausschnitte erfüllt sein, um das Regionenwachstum (bzw. die Regionenverkleinerung) zu beschränken. Einerseits ist die minimale Kantenlänge auf 8-10 Pixel und die maximale Kantenlänge auf 1/4 der Bildgröße beschränkt. Andererseits darf das Kantenverhältnis eines Rechteckes einen vorgegebenen Schwellwert nicht übersteigen. Es besteht dabei die Möglichkeit die Kantenverhältnisse zwischen 1:2 und 1:6 zu wählen, wobei sich aber ein Verhältnis von 1:3 oder 1:4 als sinnvoll erwiesen hat.
Die Propagierung des lokalisierten Bildausschnittes auf die nächst höhere Skala und die Regionenanpassung wird für die gesamte Skalenhierarchie iteriert und somit für alle vorhandenen Auflösungsstufen bis zur maximalen Auflösung des Orginalbildes durchgeführt. Durch diese Vorgehensweise wird eine effiziente Lokalisierung visuell auffälliger Bildregionen ermöglicht (siehe Abb. 2.2a).
Neben den Merkmalrepräsentationen auf oktaven Gausspyramiden wurden Auswertungen mit Skalenhierarchien, die nicht in Oktavstufen aufgebaut sind, durchgeführt. Dadurch ließ sich einerseits die Detektionsgenauigkeit der zu lokalisierenden Bildregionen und andererseits der Rechenaufwand beeinflussen. Bei individueller Wahl von nicht oktaven Auflösungshierachien konnte die Anpassung der zu selektierenden Bildregionen an die zugrundeliegende Objektgröße wesentlich verbessert werden.

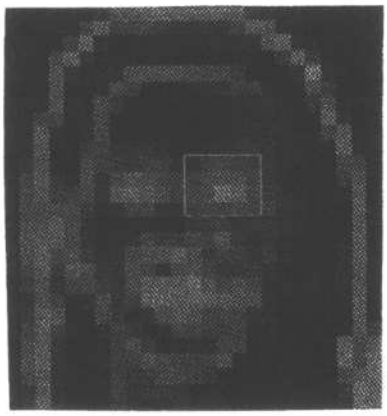


Abb. 2.2a Minimale Bildskizze mit der 1. attentiven Bildregion (Auge). Nach der Detektion des Maximums wurde der Bildausschnitt auf die Ausdehnung der Repräsentation des zugrundeliegenden Objektes expandiert.

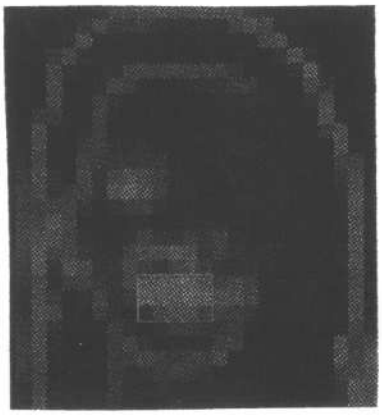


Abb. 2.2b Detektion der 2. attentiven Gesichtsregion (Mund) auf der minimalen Bildskizze, bei der die Ausblendung der zuvor lokalisierten Bildregion (Auge) bereits integriert ist. Der Bildausschnitt ist noch nicht optimal an das zugrundeliegende Objekt angepaßt.

Abb. 2.2c Aufmerksamkeitskarte mit der 2. attentiven Bildregion (Mund) in höchster Auflösung. Das zugrundeliegende Objekt wurde durch die Regionenanpassungen optimal überdeckt. An den Ausblendungen ist die sanfte abgestufte Ausblendungstechnik zu erkennen, bei der graduell die Kanteninformation reduziert wird Die Merkmale der Lippen sind lediglich abgeschwächt, während die der Zähne fast vollständig gedämpft sind.

Zur Berechnung zusätzlicher, attentiver Bildausschnitte werden die bereits gefundenen von einer weiteren Verarbeitung ausgeschlossen, indem die gefundenen Regionen auf der Merkmalkarte ausgeblendet werden (Abb. 2.2b). Diese Ausblendung wird durch einen Eintrag in eine Gewichtsmatrix realisiert, die nach jeder erfolgreichen Suche einer attentiven Region aktualisiert und anschließend mit der Merkmalkarte multipliziert wird. In dem hier vorgestellen Verfahren wird eine "weiche" Ausblendung des berechneten Rechteckausschnittes verwendet, die auf der Basis einer 2-dimensionalen Gaussfunktion dimensioniert nach der Ausdehnung des Rechtecks in X- und Y-Richtung berechnet wird. Dadurch läßt sich eine sanft abgestufte Ausblendung der detektierten Bildregionen realisieren. Es entstehen somit keine statischen Ausblendungen (geschwärzte Bildregionen) mit scharfen Übergängen, sondern es wird eine graduelle Reduktion der Merkmalstärke beginnend im Schwerpunkt der bereits selektierten Region vorgenommen. Dies entspricht auch näherungsweise der Reduktion der Stimulusstärke in biologisch visuellen Systemen. Durch diese Vorgehensweise können Artefakte bei anschließenden Lokalisationsprozessen vermieden werden, die sonst durch eine unvollständige Überdeckungen von Objekten an den Rändern (Übergängen) der ausgeblendeten Rechtecke entstehen würden (siehe Abb. 2.2c).
Eine zusätzliche Komponente der entwickelten Ausblendungstechnik bereits selektierter Bildausschnitte stellt die zeitliche Abhängigkeit der Ausblendung einer Region dar, d.h. die Einträge in der Ausblendungsmatrix unterliegen einer zeitlichen Veränderung. Es lassen sich somit Ausblendungen nach einer vorgegebenen Latenzzeit zurücknehmen und bereits lokalisierte Regionen erneut ansteuern. Dadurch entsteht eine Dynamik in der Aufmerksamkeitssteuerung, die der des biologischen Vorbildes ähnelt [Walker77] (siehe Abb. 2.5). Darüber hinaus wird durch die wiederholte Ansteuerung einzelner Bildregionen ein eventuell fehlgeschlagener oder unvollständiger, lokaler Analyseprozess mit leicht veränderten Daten der Bildregion erneut ermöglicht. Dies entspricht auch wieder den Blickbewegungsstrategien biologisch visueller Systeme, die nach einer bestimmten Latenzzeit erneut ihren Blick auf besonders markante Bildregionen (Augen) richten [Yarbus67].

Abb. 2.3 Detektion der markanten Gesichtsregionen

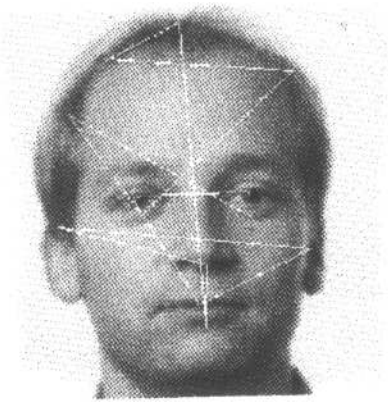


Abb. 2.4 Berechneter Attentionspfad mit statischer Ausblendung

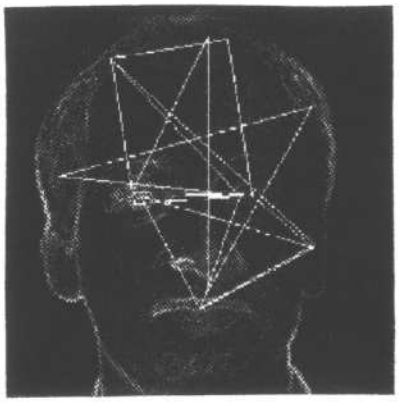
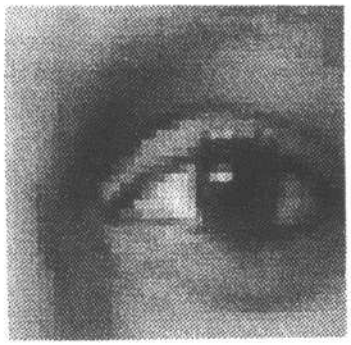


Abb. 2.5 Attentionspfad mit dynamischer Ausblendung

In den Abb. 2.3 und 2.4 werden zwei Ergebnisse des entwickelten Attentionsmechanismus vorgestellt. Die visuell markanten Bildregionen wurden detektiert und die selektierten, rechteckigen Bildausschnitte eingezeichnet. Neben der Berechnung von rechteckigen Bildausschnitten wurden erste Untersuchungen mit elliptischen Bildausschnitten durchgeführt, wobei sich aber Diskretisierungsprobleme ergeben haben.
In der Abb. 2.5 wird ein Attentionspfad mit zeitlich abhängiger Reduktion der ausgeblendeten Bildregionen vorgestellt. Dadurch entsteht eine wiederholte Ansteuerung bestimmter Bildregionen, was dem Verhalten biologisch visueller Syteme ähnelt.

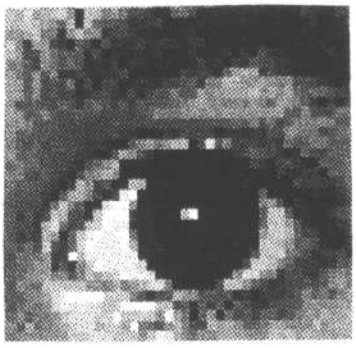
3. Klassifikation der selektierten Bildregionen

Die Klassifikation der lokalisierten Bildregionen wird durch eine Repräsentation der Merkmaleigenschaften der einzelnen Bildausschnitte auf einer Kohonen-Karte der Größe 6x6 oder 8x8 realisiert [Kohonen82]. Als Eingabegrößen für das Netzwerk können sowohl die Grauwerte der normalisierten Bildausschnitte als auch die Repräsentation der zugehörigen, abgeleiteten Merkmale verwendet werden, wobei jeweils ein 256-dimensionaler Inputvektor vorausgesetzt wurde. Für die Erzeugung einer standardisierten Eingabe in die Kohonen-Karte müssen deshalb die lokalisierten Bildausschnitte auf eine fest vorgegebene Größe von 16x16 Pixel normalisiert werden (siehe Abb. 3.1). Bei der Standardisierung der lokalisierten Bildausschnitte treten Diskretisierungsprobleme selbst bei ausreichender Interpolation auf, die nicht vollständig vermieden werden können. Weiterführende Entwicklungen werden sich mit Ansätzen beschäftigen, die eine solche Normalisierung vermeiden bzw. umgehen. Vorstellbar wäre bereits eine Auswahl nur von quadratischen Bildausschnitten oder die Verwendung von skalierten Repräsentationsformen der einzelnen attentiven Bildregionen, die keine fest definierte Bildausschnittsgrenze voraussetzen, so wie sie von Burt vorgeschlagen wurden [Burt89].



a) linkes Auge aus Abb. 4.1a

b) rechtes Auge aus Abb. 4.1b



c) linkes Auge aus Abb. 3.2

Abb. 3.1 Die drei Beispiele der extrahierten und normalisierten Bildregionen konnten von dem Klassifikationssystem als Augen identifiziert werden. An den vorgestellten Bildausschnitten lassen sich die Invarianzeigenschaften des Klassifikationsprozesses bezüglich Skalierung, Position, Rotation und Hell-Dunkel Kontrast erkennen.

Die extrahierten Bildregionen werden auf der Basis von a priori Wissen unter Anwendung eines selbstüberwachten Lernverfahrens trainiert, welches eine Vorauswahl der zu berücksichtigenden Kartenelemente trifft. Nach einer initialen Trainingsphase werden neu zu lernende Bildregionen nur den Kartenpositionen angeboten, die der selben Objektklasse angehören. Diese Vorgehensweise bewirkt bei der Anpassung der Kohonen-Karte an neue Eingabevektoren eine sinnvolle, selektive Auswahl, so daß eine zusätzliche Repräsentation von neuen Bildausschnitten auf der Karte realisiert werden kann. Dies entspricht einer gezielten, dynamischen Erweiterung der repräsentierten Information auf der Merkmalkarte und vermeidet Konfliktzuweisungen zwischen Eingabevektoren und Kartenvektoren, die unter Umständen unterschiedlichen Objektklassen angehören können. Durch die Vorauswahl der Kartenelemente einer Objektklasse läßt sich der Rechenaufwand des Anpassungsprozesses reduzieren, da nur noch die Distanzen und Kartenadaptationen bezüglich des Inputvektors und der vorausgewählten Kartenvektoren berechnet werden müssen.
Das a priori Wissen besteht aus relativen Wahrscheinlichkeiten einer Gesichtsregion bezüglich der Ortskoordinaten der Bildmatrix. Die hier verwendete Ortsinformation wird aber lediglich für den selbstüberwachten Lernvorgang verwendet und nicht für die anschließende Klassifikation, denn die Ortsinformation gehört nicht zu den Eingabegrößen während der Lernphase des Netzwerkes.
Zur Segmentation der gelernten topologischen Merkmalkarte wurde eine von Kohonen vorgeschlagene Rückzuweisungsmethode verwendet [Kohonen89], wobei aber bei der Zuweisung zusätzlich eine Ähnlichkeitsbedingung (euklidische Distanz) eingeführt wurde. Wird ein diesbezüglich definiertes Ähnlichkeitsmaß nicht erfüllt, so wird das Kartenelement einer Rückweisungsklasse zugeordnet. Dadurch konnten Fehlklassifikationen vermieden werden, die durch eine Rückzuweisung von Verbindungselementen der Karte, die zwischen zwei Repräsentationen von zwei Objektklassen liegen, entstehen würden. Aufgrund der geringen Klassenanzahl relativ zur Kartengröße (4 Klassen plus Rückweisungsklasse) und aufgrund relativ ausgewogenem und ausreichendem Stichprobenumfang für alle Gesichtsregionenklassen entstanden vergleichbar ausgeprägte Kartenrepräsentationen der einzelnen Objektklassen, ohne daß ein Gesichtsbereich unterrepräsentiert wurde.

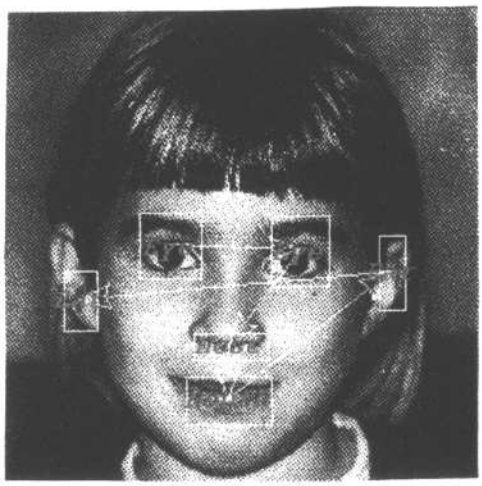


Abb. 3.2 Beispiel eines Gesichtsbildes, in dem die lokalisierten Bildregionen richtig klassifiziert wurden.

Die auf der Grundlage der genannten Repräsentation an über 50 Gesichtsbildern durchgeführten Klassifikationen erreichten für die Zuordnungen von Bildausschnitten der Augen-, Nasen-, Mund- und Ohrenregion eine Sensitivität von über 90% bei einer Spezifität von ca. 80%. Voraussetzung waren ausreichend große Bildausschnitte der einzelnen Bildregionen (siehe Abb. 3.2). Zu kleine Teilausschnitte, die nur einen Bruchteil der entsprechenden Bildregion beinhalten, wurden von einer Klassifikation ausgeschlossen. Die Augenregion wurde nahezu zu 100% richtig identifiziert, da die zugehörigen Bildausschnitte durch den verwendeten Attentionsmechanismus recht homogen bezüglich Größe und Positionierung ausgewählt werden konnten (siehe auch Abb. 3.1). Die Ohrenregion konnte im Vergleich dazu unsicherer klassifiziert werden, da diese Ausschnitte erheblich heterogener selektiert wurden. Die Qualität der Klassifikation ist also entscheidend von einer standardisierten Wahl der betreffenden Bildausschnitte abhängig. Je standardisierter die

Lokalisierung der einzelnen Bildausschnitte in Größe und Positionierung berechnet werden kann, desto besser kann sich eine gute Repräsentation auf der topologischen Merkmalkarte ausbilden und um so besser und sicherer können anschließende Klassifikationen vorgenommen werden. Eine Optimierung der Klassifikationen erfordert also sowohl eine Verbesserung der Lokalisierung der Bildausschnitte als auch eine Verbesserung der Repräsentation auf der topologischen Merkmalkarte.

4. Diskussion und Ausblick

In dieser Arbeit wurde an dem Beispiel der Gesichtserkennung gezeigt, daß motiviert durch das biologische Vorbild attentive Verarbeitungsstrategien sinnvoll und effizient in komplexen Bildverarbeitungssystemen eingesetzt werden können. Dazu gehört einerseits eine effiziente Methode zu einer aufgabenorientierten Blicksteuerung als auch Klassifikationsverfahren mit selbstlernenden Eigenschaften zur robusten Identifikation extrahierter Bildbereiche.
Auf der Basis attentiver Verarbeitungsmechanismen läßt sich die Realisation einer Objektlokalisation in Realweltbildern soweit reduzieren, daß sie in linearem Zeitaufwand durchgeführt werden kann. Diese Vorgehensweise zeichnet sich insbesondere durch eine Robustheit gegenüber unterschiedlichen Rahmenbedingungen aus. Neben der Eigenschaft der Skaleninvarianz, die durch die Auswertung der Repräsentation der Aufmerksamkeitskarte in Skalenhierarchien errreicht wird, konnten Ortsinvarianzeigenschaften, wie Verschiebungs- und Rotationsinvarianzen, durch das attentive Vorgehen integriert werden (Abb. 4.1). Das Verfahren ist darüber hinaus robust gegenüber unterschiedlichen Beleuchtungsbedingungen und Kontrasten der Bilddaten (siehe Abb. 3.1).

Abb. 4.1 a) - c) Dreistufige und vierstufige Attentionspfade mit Darstellung der lokalisierten und richtig klassifizierten Bildregionen. Der Zeitpunkt der Detektion der Augen und der Mundregion liegt während der ersten 3 bis 4 Attentionsschritte. Darüber hinaus ist eine vergleichbare Ausdehnung der selektierten Bildregionen insbesondere der Augenausschnitte zu erkennen.

Neben den bereits angesprochenen Untersuchungen ist beabsichtigt, die Informationen aus bereits lokalisierten und interpretierten Bildregionen für die weitere Attentionssteuerung zu nutzen, so daß die Aufmerksamkeit nur auf die bezüglich einer bestimmten Fragestellung relevanten Bildregionen gelenkt wird. Dazu soll kontinuierlich das bereits vorhandendene semantische Wissen über die erkannten Bildobjekte gesammelt und ausgewertet werden. Auf dieser Basis ließe sich eine gezielte Aufmerksamkeitssteuerung bezüglich einer bestimmten Fragestellung realisieren und die Navigation innerhalb der vorliegenden Bildmatrix optimieren. Ziel ist dabei eine weitere Verbesserung der Skalen- und Ortsinvarianzeigenschaften, so daß die z. Z. vorhandenen standardisierenden Voraussetzungen schrittweise abgebaut werden können. Darüber hinaus soll dadurch die Voraussetzung geschaffen werden, die Aufmerksamkeit bezüglich einer bestimmten Fragestellung zu konzentrieren. Als Beispiel wäre die Suche und möglichst genaue Lokalisierung der Augenwinkel zur Bestimmung der verschiedenen Augenabstände denkbar. Dazu sind Verfahren zu entwickeln, die eine Unterteilung der globalen und lokalen Verarbeitung steuern und somit eine Lokalisation von besonders charakteristischen Bildpunkten (keypoints) ermöglichen. Erste Untersuchungen diesbezüglich haben gezeigt, daß die Augenecken durch Kombination von standard Filterverfahren zu detektieren sind. Eine Klassifikation verschiedener solcher charakteristischen Schlüsselpunkte wurde bereits durch spezifische Verfahren erreicht [Michaelis94].

Danksagung

Für die Unterstützung der vorliegenden Arbeit und für die Bereitstellung der Routinen des Bildverarbeitungssystems HORUS möchten wir uns bei Herrn Dr. W. Eckstein, Institut für Informatik, Lehrstuhl für Informatik IX der Technischen Universität München bedanken. Desweiteren gilt unser Dank Frau Prof. Dr. S. Stengel-Rutkowski, Kinderzentrum München der Universität München für die Bereitstellung des Bildmaterials, welches aber aus Datenschutzgründen in der vorliegenden Arbeit nicht verwendet wurde.

Literatur

[Ahmad91] Ahmad S., *VISIT: An efficient computational model of human visual attention*, PhD thesis, University of Illinois at Urbana-Champaign, September 1991
also TR-91-049, International Computer Science Institute, Berkeley, CA

[Brunelli92] Brunelli R. and T. Poggio, *Face recognition through geometrical features*, in G. Sandini (edt.) Computer Vision, Proc. of the ECCV'92, S. 792-800 (1992)

[Brunelli93] Brunelli R. and T. Poggio, *Face recognition: features versus templares*, IEEE Trans. Patt. Anal. Machine Intell., Vol. 15, Nr. 10, S. 1042-1052 (1993)

[Burt89] Burt P. J., *Multiresolution techniques for image representation, analysis and smart transmission*, SPIE Conf. 1199, S. 1-14 (1989)

[Califano90] Califano A., R. Kjeldsen and R. M. Bolle, *Data and model driven foveation*, Proc. 10th. ICPR 90, S. 1-7 (1990)

[Culhane92a] Culhane S. M. and J. K. Tsotsos, *An attentional prototype for early vision*, In G. Sandini (ed.) Computer Vision ECCV'92, Springer-Verlag, pp. 551-560 (1992)

[Culhane92b] Culhane S. M. and J. K. Tsotsos, *A prototype for data-driven visual attention*, Proc. of the 11th IAPR, S. 36-40 (1992)

[Herpers93] Herpers R., H. Kattner und G. Sommer, *GAZE: An attentional mechanism for the extraction of prominent facial features*, in Artificial Intelligence in Medicine, S. Andreassen et al. (eds.) IOS Press Amsterdam, S. 159-163 (1993)

[Koch85] Koch C. and S. Ullman, *Shifts in selective visual attention: towards the underlying neural circuity*, Human Neurobiol, Vol. 4, S. 219-227 (1985)

[Kohonen82] Kohonen T., *Clustering, taxonomy and topological maps of patterns*, Proc. of the 6th. Int. Conf. on Patt. Rec., Computer Society Press, Silver Spring, S. 114-128 (1982)

[Kohonen89] Kohonen T., *Self-organization and assoziative memory*, Springer Verlag Berlin (1989)

[Michaelis94] Michaelis M. and G. Sommer, *Junction classification by multipleorientation detection*, accepted by ECCV'94, Stockholm (1994)

[Olshausen92] Olshausen B., C. Anderson and D. v. Essen, *A neural model of visual attention and invariant pattern recognition*, California Institute of Technology, Tec. Report CNS Memo 18, September 1992

[Sandon89] Sandon P. A., *Simulating visual attention*, J. of cog. Neuroscience, Vol. 2 (3), S. 213-231 (1989)

[Stengel85] Stengel-Rutkowski S. und P. Schimanek, *Chromosomale und nicht-chromosomale Dysmorphiesyndrome*, Enke Verlag Stuttgart (1985)

[Treisman85] Treisman A., *Preattentive processing in vision*, CVGIP Vol. 31, pp. 156-177 (1985)

[Tsotsos89] Tsotsos J. K., *The complexity of perceptual search task*, In: Proc. of the IJCAI, Detroit, pp. 1571-1577 (1989)

[Tsotsos90] Tsotsos J. K., *Analyzing vision at the complexity level*, Behavioral and Brain Sciences, Vol.13 S. 423-469 (1990)

[Tsotsos91] Tsotsos J. K., *Localizing stimuli in a sensory field using an inhibitory attentional beam*, Tec. Report of the University of Toronto RBCV-TR-91-37, October 1991

[Walker77] Walker-Smith G. J., A. G. Gale and J. M. Findlay, *Eye movement strategies involved in face perception*, Perception, Vol. 6, S. 313-326 (1977)

[Yarbus67] Yarbus A. L., *Eye movements and vision*, New York: Plenum Press (1967)