In: Paulus, Dietrich und Wagner, Thomas (Herausg.):
Tagungsband 3. Workshop „Farbbildverarbeitung", Fraunhofer-Institut für Integrierte Schaltungen A, Seiten 67-72 & 101, 25./26. 9. 97, Erlangen

Ein neuronales Verfahren zur Segmentierung hautfarbener Bildregionen

Ulf-Dietrich Braumann, Anja Brakensiek,
Hans-Joachim Böhme, Horst-Michael Groß

Technische Universität Ilmenau
Fachgebiet Neuroinformatik

Zusammenfassung

Der vorliegende Beitrag stellt die Anwendung eines inkrementellen neuronalen Netzes zur Klassifikation hautfarbener Pixel als Grundlage zur Segmentierung hautfarbener Bildregionen vor. Es werden die Probleme erläutert, die beim praktischen Einsatz in Real-World-Umgebungen ohne Einhaltung von Referenzbedingungen auftreten. In Abwandlung eines Verfahrens von Pomierski (siehe [2]) wird ein einfaches Adaptationsverfahren zur Farbumstimmung mit impliziter Referenz vorgestellt und diskutiert.

1 Einleitung

Die Lokalisierung von Personen in Real-World-Szenen bildet eine wichtige Voraussetzung zur visuell basierten Kommunikation von Nutzern mit einem mobilen Robotersystem. Ein Weg dafür ist die Klassifizierung der Bildfarbwerte in haut- und nicht-hautfarbene Teilmengen, um damit *ein spezielles Auffälligkeitsmaß* für Bildregionen zu erhalten, in denen Gesichter und ggf. Hände enthalten sind. Dieses Auffälligkeitsmaß stellt eine Voraussetzung für eine erhebliche Datenreduktion für die (wissensbasierte) effektive Szenenanalyse mittels *selektiver visueller Aufmerksamkeit* durch den Roboter dar.

Im Kontext der Roboternavigation ist generell ein erhebliches Maß an Invarianzleistung für die Merkmalsextraktion aus der Umwelt gefordert. Aus unserer Sicht erscheint es für die visuelle Merkmalsextraktion dabei am aussichtsreichsten, verschiedene Informationsmodalitäten zu verknüpfen (siehe Abbildung 1). Neben (a) der Bildung von Bewegtheitmaßen aus Abschnitten von zeitlichen Bildfolgen (Detektion von Winkbewegungen), (b) der Gestaltsanalyse anhand spezieller Konturverläufe (Detektion von Kopf-Schulter-Partien) und (c) der strukturbasierten Gesichtsdetektion (Hauptkomponentenanalyse (PCA)), wird durch die Konzentration auf sichtbare Hautpartien durch (d) Verwendung eines für typische Hautfarben ausgelegten Klassifikators die *Lokalisierbarkeit von potentiellen Nutzern* für ein mobiles Robotersystem entscheidend verbessert.

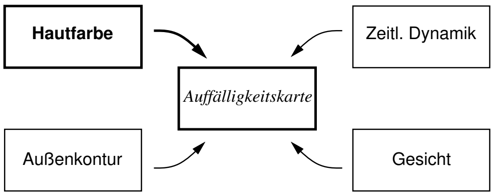

Abbildung 1: Die Beiträge zur Bildung einer Auffälligkeitskarte

2 Growing Neural Gas

Unser Ansatz zur Klassifikation hautfarbener Bildregionen verwendet einen adaptiven inkrementellen Vektorquantisierer, das Growing Neural Gas (GNG, siehe [1]).

2.1 Eigenschaften

Das GNG adaptiert während der Trainingsphase nicht nur die Parameter seiner Neuronen (Gewichte, Radien), sondern auch die Anzahl der Neuronen selbst. Im Laufe der Trainingsphase kann das Netz sich optimal an die Statistik der Daten anpassen. Die Neuronen stehen über ihre Kanten nicht in einem festen Schema (etwa aufgespannte Hypertetraeder) mit ihren Nachbarn in Verbindung, sondern die Kanten können sich so aufbauen, wie die Stimuli das erfordern, so daß beispielsweise unzusammenhängende Netze entstehen können.

2.2 Funktionsweise

Initial existieren zwei zufällig im Merkmalsraum positionierte Neuronen, die mit einer Kante verbunden sind. Das Hinzufügen von Neuronen zur besseren Abdeckung des Merkmalsraums geschieht aufgrund eines Fehlermaßes, das für jedes Neuron gebildet wird. Das Fehlermaß wird akkumuliert, indem für das aktuell am stärksten aktivierte Neuron (Best-Matching-Neuron) die Abweichung vom Stimulus ermittelt wird und dem Fehlermaß dieses Neurons zugeschlagen wird. Immer nach Verstreichen einer sog. Einfügeperiode wird das Neuron mit dem *größten akkumulierten Fehler insgesamt* und ein über eine Kante benachbartes Neuron mit dem *größten akkumulierten Fehler in der Nachbarschaft* ermittelt und genau zwischen beiden Neuronen ein neues Neuron eingefügt. Die bestehende Kante wird entfernt und durch zwei neue Kanten ersetzt. Die Fehlermaße aller drei beteiligter Knoten werden auf Null rückgesetzt.

Kanten können entfernt werden, wenn sie „zu alt“ sind. Immer, wenn ein Best-Matching-Neuron bestimmt wurde, werden alle angrenzenden Kanten „älter“. Sie können aber ein maximales Alter nicht überschreiten, so daß sie genau dann entfernt werden müssen. Sofern es eine Kante vom Best-Matching-Knoten zu einem Knoten gibt, der keine weitere Kante hat, und sofern diese Kante das Maximalalter erreicht, wird neben der Kante auch der betreffende Knoten gelöscht.

Das Einfügen von Knoten kann solange fortgeführt werden, wie ein bestimmter Maximalfehler nicht unterschritten wird. In unserer Implementation wurde allerdings der Weg gegangen, eine bestimmte Anzahl von Trainingsdaten (ermittelter Hautfarbdatensatz) zu präsentieren.

2.3 Gewinnung der Trainingsdaten

Generell wurden die Datensätze aus manuell segmentierten Farbbildern gewonnen, bei deren Aufnahme (RGB-Einchip-CCD-Kamera) Referenzbedingungen geherrscht haben. Das betrifft insbesondere die Vornahme des Weißabgleichs entsprechend der vorherrschenden Beleuchtung. In einen Datensatz sind typischerweise hautfarbene Regionen mehrerer Personen aufgenommen worden, um die gesamte Vielfalt der auftretenden Hauttypen berücksichtigen zu können.

3 Verwendeter Farbraum

3.1 Modellbildung

In [2] wird ein Modell zur Farbadaptation mit dem Ziel der *Realisierung stabiler Farbempfindungen* vorgestellt. Es enthält im Kern einen *neurophysiologisch motivierten Elementarfarbraum* mit den zueinander orthogonalen Achsen Rot-Grün (RG), Blau-Yellow (BY) und Weiß-Schwarz (WS), also zwei jeweils opponenten Farbachsen und einer Unbuntachse.

Die Modellierung erstreckt sich über mehrere Stufen des parvozellulären (kleinzelligen) Pfades des visuellen Systems des Menschen, in welchem die primäre Farbverarbeitung konzentriert ist. Es werden berücksichtigt

- die *spektralen Absorptionskurven* der lang-, mittel- und kurzwellig (LMS) sensitiven Zapfen (Rezeptoren des Tagsehens) der Retina (Netzhaut),
- das *zahlenmäßige Verhältnis der Zapfen* im Bereich der Fovea (Bereich höchster örtlicher Auflösung der Retina),
- die *Art und Stärke der Gewichtung bei der Verschaltung der L-, M- und S-Aktivierungen* innerhalb des parvozellulären Teils des Lateralen Kniehöckers (LGN, erste Stufe der visuellen Verarbeitung nach der Retina).

Als Ergebnis der Verarbeitung liegen im LGN *bezüglich der Retina topographisch organisierte (kein Verlust der Ortsinformation) Aktivierungskarten RG, BY und WS* vor. Nach einer weiteren Verarbeitungsstufe (Areal V 1 des visuellen Cortex) liegt im Areal V 4 eine hinsichtlich der Retina nicht-topographische Repräsentation, der Elementarfarbraum vor.

Mit dem Ziel, ständig ein *mittleres Empfindungsniveau für die Gesamtszene* für eine *optimale visuelle Informationsextraktion* einzunehmen, kann in diesem Elementarfarbraum eine *Farbadaptation* ausgeführt werden. Ein mittleres Empfidungsniveau besteht dann, wenn der mittlere Farbwert einer Farbszene *unbunt* ist. Der Autor charakterisiert diese Adaptation als *referenzfreie Farbumstimmung* (s. Abb. 2, 3. Spalte).

3.2 Umsetzung des Modells

Als Voraussetzung für den Einsatz dieser Umstimmung wird von einer *farblich ausgewogen zusammengesetzten Szene* ausgegangen, was bei natürlichen Szenen im Mittel vorausgesetzt wird.

Unter Berücksichtigung der *spektralen Empfindlichkeitskurven der R-, G- und B-Sensoren* der Kamera (indirekt gegeben über die spektralen Emissionskurven von Monitoren nach U. S.-amerikanischer FCC-Norm) wird eine lineare Transformationsvorschrift angegeben, die die unmittelbare Überführung von RGB nach RG/BY/WS erlaubt.

Im einzelnen werden folgende Schritte zur *empfindungsgemäßen Farbumstimmung* ausgeführt:

- Bestimmung des eigenwertgrößten Eigenvektors (Hauptachse) aller Abbildungspunkte (Punktwolke) im Farbraum,
- Scherung dieser Punktwolke derart, daß ihre Hauptachse in Deckung mit der Unbuntachse gelangt,
- Reskalierung der Punktwolke entlang der Unbuntachse, so daß der maximale Dynamikumfang entlang der Unbuntachse ausgenutzt wird.

Das Verfahren kann immer dann, wenn keine farblich ausgewogen zusammengesetzte Szene vorliegt, zu problematischen, fehlerhaften Umstimmungen führen, da lediglich der Bildinhalt selbst das Kriterium für die Umstimmung liefert und nicht unterschieden werden kann, was die Ursache für eine Auslenkung der Unbuntachse ist: die Beleuchtung der Szene oder die unausgewogene farbliche Zusammensetzung der Objekte innerhalb der Szene. In den für die Roboternavigation relevanten Umgebungen ist die farblich ausgewogene Zusammensetzung häufig nicht erfüllt, beispielsweise durch (bildbezogen) anteilsmäßig große einzelne einfarbige Flächen wie Fußböden und Wände oder Kombinationen farblich ähnlicher Objekte. Im Ergebnis sind dann beispielsweise solche farblichen Flächen nach der Farbadaptation entsättigt und/oder bei von den farblich dominanten Flächen abweichenden anteilsmäßig kleinen Objekten mehr oder weniger starke Farbverfälschungen zu beobachten. Das Ziel dieser Farbumstimmung ist in erster Linie, den Gesamteindruck eines Bildes zu verbessern, während für eine sichere Hautklassifikation vor allem die Lage der Hautfarbwolke entscheidend ist und die Adaption des Hintergrundes nur eine untergeordnete Rolle spielt.

4 Farbumstimmung mit impliziter Referenz

Die vorliegende Arbeit führt eine *objektspezifische Farbrepräsentation* mit dem Ziel ein, eine *sicherere Lokalisation von hautfarbenen Bildregionen* zu erzielen und unter Hinzunahme von Strukturinformation (grauwertbasiert) zur Lokalisation von Gesichtern von Personen die jeweilige im Bereich des Gesichts vorherrschende (Haut-)Farbe als Referenz zur Farbkorrektur bezogen auf ein vorhandenes Hautfarbmodell zu verwenden. Während die Lokalisation frontaler Gesichter grauwertbasiert nach dem Stand gegenwärtiger Untersuchungen beherrschbar erscheint (z. B. mittels Filtersätzen, die aus den resultierenden Eigenvektoren einer PCA gewonnen wurden), ist die grauwertbasierte Lokalisation von

Händen (und Armen) aufgrund der Vielfalt der möglichen Ausrichtungen und Haltungen wesentlich schwieriger.

Somit kann eine farbliche Referenz gebildet werden, mit deren Hilfe eine Farbumstimmung hin zu einem Hautfarbmodell erfolgen kann. Dieses Hautfarbmodell stellt das trainierte GNG dar.

Die implizite Adaption erfolgt über die Lage der Schwerpunkte unseres Hautfarbmodells und der aktuellen Referenzhautwolke. Dazu wird das ursprüngliche Adaptionsverfahren von [2] so verändert, daß in einem ersten Schritt eine Scherung (Verschiebung der Farbpunkte innerhalb einer Ebene in Abhängigkeit des WS-Anteils) in Richtung des Hautfarbmodells durchgeführt wird, wobei der Buntgrad der Farbwolke vorläufig erhalten bleibt. Anschließend erfolgt eine Buntgradkorrektur in RG- und BY-Richtung auf den Schwerpunkt des Modells durch eine Skalierung und eine Streckung in WS-Richtung.

5 Ergebnisse

Die Hautfarbklassifikation eines Pixels erfolgt nach dessen Transformation in den beschriebenen Elementarfarbraum und durchgeführter Farbadaption mit Hilfe eines 3-dim. GNG. Das Klassifikationsmaß entspricht dabei der Aktivierung des Best-Matching-Neurons.

Im folgenden werden sowohl farbadaptierte Bilder und ihre Punktwolken im Farbraum (gelb: Hautfarbe, blau: Hintergrund) als auch die zugehörigen Klassifikationsergebnisse (schwarz: höchste Hautwahrscheinlichkeit) dargestellt (s. Abb. 2).

5.1 Hautfarbklassifikation

Die linke Spalte der im Anhang dargestellten Abbildungen (Abb. 2) zeigt ein Bild, welches unter Trainingsbedingungen aufgenommen wurde. Entsprechend gut erfolgt die Hautfarbklassifikation, da das gelernte GNG sich mit der unten dargestellten Hautfarbwolke deckt. Man kann jedoch erkennen, daß auch unter diesen Referenzbedingungen (Halogenbeleuchtung) hautfarbene Regionen des Hintergrundes ebenfalls klassifiziert werden. Bei einer entsprechenden Aufnahme bei Leuchtstoffröhrenbeleuchtung (2. Spalte) stimmt die aktuelle Hautfarbe nur noch zu einem geringen Teil mit dem Hautfarbmodell überein. Nach einer referenzfreien Adaption (3. Spalte) verbessert sich zwar die Hautklassifikation, die vollständigen Hautpartien werden jedoch erst nach der Farbadaption mit impliziter Referenz (4. Spalte) ermittelt.

Nebeneffekt der impliziten Umstimmung ist im Durchschnitt auch eine korrektere Darstellung des Hintergrundes. Ungenauigkeiten bei der Hintergrundfarbe treten auf, weil zum einen die Hautfarbe sehr stark variieren kann (Winter/Sommer) und zum anderen die tatsächliche Abbildungsverschiebung von Objekten im Farbraum bei Beleuchtungsänderung nicht allein eine Funktion der jeweiligen Farbraumkoordinate ist. Anzumerken ist, daß es kein universelles Farbadaptionsverfahren gibt, das alle Farben eines Bildes geschlossen korrigieren kann, da die Zuordnung von physikalischem Spektrum und Farbraumdarstellung (insbesondere bei unterschiedlicher, unbekannter Beleuchtung) nicht eindeutig ist.

5.2 Regionenorientierte Segmentierung

Das bisher vorgestellte Verfahren stellt kein regionenorientiertes, sondern ein pixelbezogenes Segmentierungsverfahren dar. Da kein Zusammenhangstest erfolgt, ist es leicht möglich, daß verschiedene örtlich eng begrenzte Störeinflüsse zu losgelösten, falsch positiven Segmentierungsergebnissen führen, die nicht eliminiert werden. Um dennoch isolierte positive Bildpunkte eliminieren zu können, wendeten wir zunächst folgende einfachen Schritte an: Herabskalierung des Bildes (aliasingfrei) um den Faktor 4, anschließende Binarisierung des Ergebnisses anhand einer geeigneten Schwelle.

Eine anderes, *neuronales* Verfahren für eine regionenbasierte Segmentierung wurde in [3] vorgestellt, das darin anhand eines beispielhaften Ergebnisses unserer Hautfarbklassifikation demonstriert wird. Der Segmentierungseffekt beruht auf der inneren Dynamik eines Feldes aus Columnen dynamischer Neuronen. Im Laufe der Iterationsschritte bilden sich die einzelnen Hautfarbsegmente sequentiell aus und eignen sich daher insbesondere als Vorstufe für eine selektive Bildanalyse. Der generelle Einsatz dieses Verfahrens wird gegenwärtig vorbereitet.

6 Diskussion

Unter der Zielstellung, ein aus mehreren Komponenten bestehendes Auffälligkeitsmaß zu bilden, für das *ein* Beitrag aus der Segmentierung hautfarbener Bildregionen stammt, konnten gute qualitative Ergebnisse sowohl bei der Klassifikation unter Referenzbedingungen als auch unter abweichenden Bedingungen erzielt werden. Ein referenzfreies Farbadaptationsverfahren (siehe [2]) bildete den Ausgangspunkt für ein Verfahren, welches strukturbasiert ermittelte Regionen im Bild als implizite Referenz zur Farbumstimmung benutzt.

Literatur

[1] Fritzke, Bernd. A Growing Neural Gas Network Learns Topologies. In Tesauro, Gerald, Touretzky, David S., and Leen, Todd K., editors, *Advances in Neural Information Processing Systems 7 (NIPS'94)*, pages 625–632. The MIT Press, 1995.

[2] Pomierski, Torsten and Gross, Horst-Michael. Neural Architecture for Chromatic Adaptation of Single Images and Picture Sequences. In *Tagungsband 2. Nationaler Workshop „Farbbildverarbeitung“*, Schriftenreihe des Zentrums für Bild- und Signalverarbeitung e. V. Ilmenau, pages 29–34, 1996.

[3] Stephan, Volker and Gross, Horst-Michael. Formerhaltende sequentielle visuelle Aufmerksamkeit in columnar organisierten neuronalen Feldern. In Paulus, Erwin and Wahl, Friedrich M., editors, *Mustererkennung 1997 – 19. DAGM-Symposium (Braunschweig)*, Informatik aktuell. Springer-Verlag, 1997. Noch nicht erschienen.

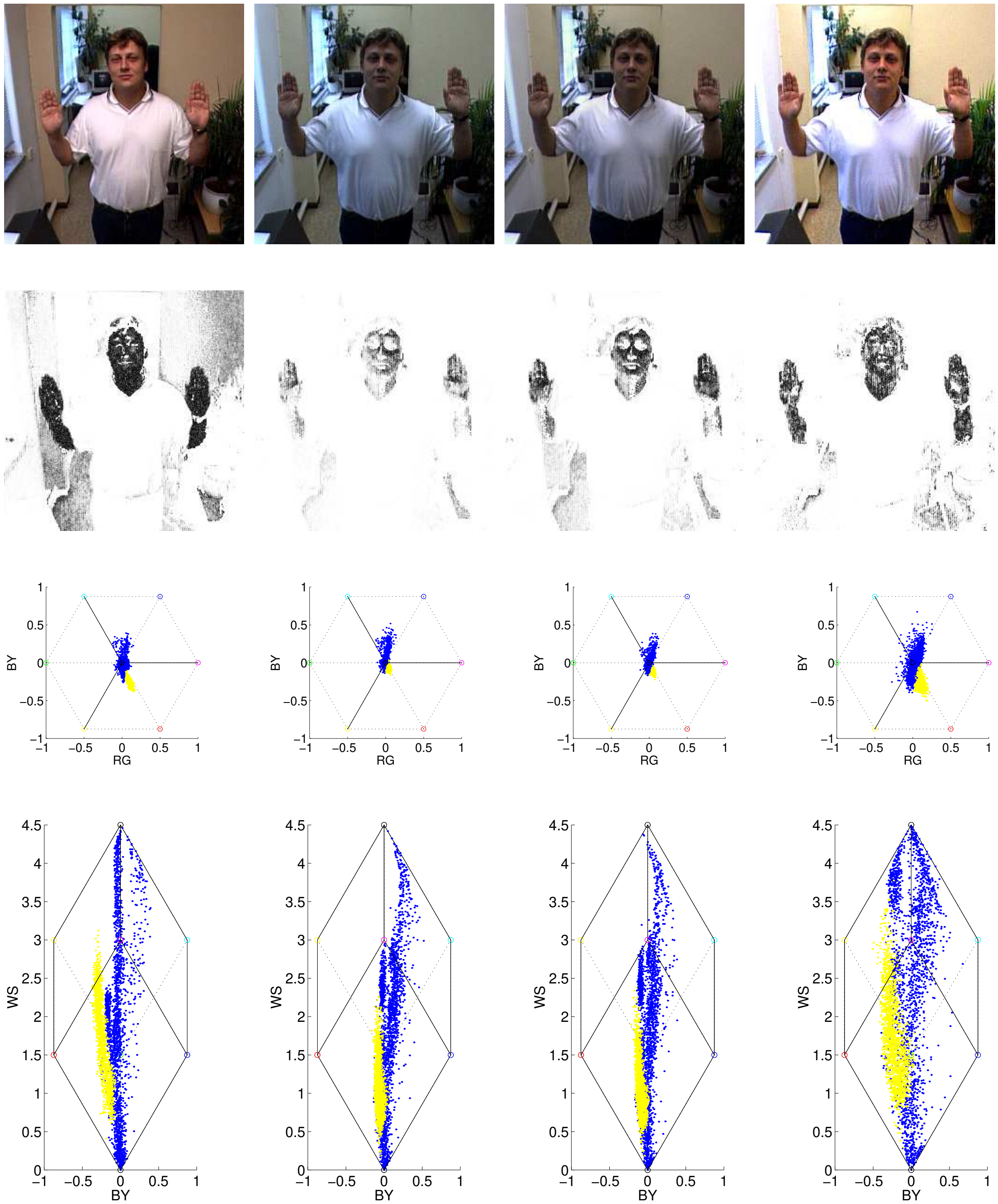

Abbildung 2: Hautfarbklassifikation eines Trainingsbildes (links), eines Testbildes bei veränderter Beleuchtung (2.v.l.), des referenzfrei adaptierten Testbildes (2.v.r.) und des mit impliziter Referenz adaptierten Testbildes (rechts)