Grobe Lokalisation und Verfolgung von Patientengesichtern und Gesichtsmerkmalen in Farbblidfolgen in Echtzeit

A. Gebhard, M. Dege, D. Paulus

Lehrstuhl für Mustererkennung (Informatik 5) (*LME*)
Friedrich–Alexander–Universität Erlangen–Nürnberg
Martensstraße 3, 91058 Erlangen
email:gebhard@informatik.uni-erlangen.de

6 Seiten
Erscheint in den Proceedings zum
4. Workshop Farbbildverarbeitung

Koblenz, 17.–18. September 1998

Grobe Lokalisation und Verfolgung von Patientengesichtern und Gesichtsmerkmalen in Farbblidfolgen in Echtzeit

A. Gebhard, M. Dege, D. Paulus

Lehrstuhl für Mustererkennung (Informatik 5) (*LME*)
Friedrich–Alexander–Universität Erlangen–Nürnberg
Martensstraße 3, 91058 Erlangen
email:gebhard@informatik.uni-erlangen.de

Zusammenfassung

Für die Echtzeit–Verfolgung von Gesichtern in Bildfolgen schlagen Wang und Chang ein Verfahren vor, das direkt auf den DCT-Koeffizienten von JPEG–komprimierten Bildern arbeitet. Dieses Verfahren wurde erweitert um eine automatische Schätzung von Gesichtsfarbe und der automatischen Verfolgung und groben Lokalisation von Gesichtsmerkmalen. Das System wird als Vorverarbeitungsstufe eines Systems eingesetzt, in dem Gesichtsbilder von Patienten in einer medizinischen Anwendung analysiert werden.

1 Einleitung

Bildanalyse nimmt zunehmend ihren Platz in medizinischen Anwendungen ein. Das im folgenden vorgestellte Modul ist Teil eines System, mit dem Bildsequenzen von Patienten, die vor einer Kamera gezielte mimische Bewegungen durchführen, in Echtzeit verarbeitet werden [1]. Die Patienten leiden dabei unter einer Teillähumg der Gesichtsnerven, wie sie beispielsweise nach einem Schlaganfall auftritt. Da das System in der Klinik wie auch im häuslichen Bereich eingesetzt werden soll, ist eine kontrollierte Beleuchtung nicht vorgesehen. Ein Überblick der medizinischen Problematik findet sich in [4, 3].

In der Anwendung ist es zum einen erforderlich, Merkmale in den Gesichtsbildern zu segmentieren, die sich an anatomischen Gegegebenheiten orientieren; dies erfordert die Segmentierung der Mundregion. Beispielsweise sind die Mundwinkel von großer Bedeutung. Zum anderen ist es vorgesehen, daß der Patient sich weitgehend frei vor der Kamera bewegen kann; daher muß das Gesicht in den Bildfolgen verfolgt werden. Zur Lösung beider Probleme wird Farbinformation eingesetzt.

Im folgenden wird der Ansatz zur Echtzeitverfolgung von Gesichtern von Wang und Chang [2] vorgestellt und um die automatische Schätzung von Gesichtsfarbe und die Verfolgung von Gesichtsmekrmalen (hier Augen und Mund) erweitert. Dabei ist das Ziel eine grobe Lokalisierung dieser Regionen.

Im anschließenden Abschnitt 2 werden die Grundlagen der JPEG-Komrimierung beschrieben. Abschnitt 3 beinhaltet die Schätzung der Farbverteilung eines Gesichtes. In Abschnitt 4 wird das Gesamtsystem erläutert. Ergebnisse werden in Abschnitt 5 präsentiert. Den Abschluß des Artikels bildet eine Zusammenfassung.

2 Grundlagen der JPEG–Komprimierung

Bei der JPEG–Kompression wird als erstes das gesamte Bild in 8×8 große Blöcke aufgeteilt und der Wertebereich der Bildpunkte auf $[-127, \ldots, 128]$ normiert. Im nächsten Schritt wird auf jeden Block die diskrete Kosinustransformation (DCT) angewendet. Als Ergebnis erhält man einen ebenfalls 8×8 großen Block von DCT–Koeffizienten, wobei ein Großteil der Energie des Ausgangsblockes in nur wenige niedrigfrequente Koeffizienten konzentriert wurde. Der Koeffizient $(0, 0)$ entspricht dabei — bis auf einen Faktor — dem Mittelwert der Grau– bzw. Farbwerte des Blockes.

Als nächstes erfolgt die Quantisierung, wobei jeder Eintrag des Koeffizientenblocks durch den entsprechenden in einer 8×8 großen Quantisierungstabelle geteilt und das Ergebnis auf eine ganze Zahl gerundet wird. Da die meisten Koeffizienten bereits einen sehr kleinen Wert besitzen, werden sie durch die Quantisierung auf Null gesetzt, so daß im Endeffekt hauptsächlich nur noch Koeffizienten im linken oberen Eck des Blockes von Null verschieden sind. Anschließend an die Quantisierung wird der Block in einem Zickzack–Muster — links oben beginnend — abgelaufen und die Werte in einen 64 Elemente langen Vektor übertragen. Dadurch wird erreicht, daß diejenigen Einträge, die von Null verschieden sind, vor allem an den Anfang des Vektors positioniert werden. Bei der darauffolgenden Entropiekodierung, dem letzten Schritt der JPEG–Kompression, wird dieser Vektor zuerst lauflängenkodiert, wobei nur die von Null verschiedenen Einträge gespeichert werden, jeweils mit einer Angabe, wieviele Nuller ihnen vorangegangen sind, und das daraus resultierende Ergebnis nach einem weiteren Zwischenschritt einer Huffmankodierung unterworfen.

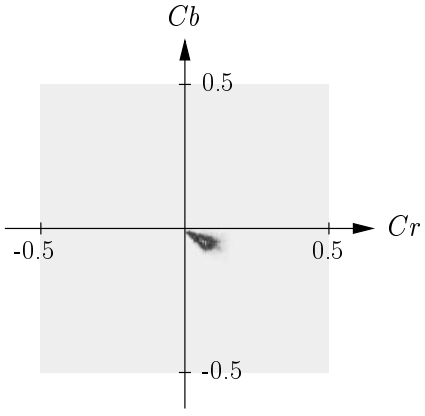

Bild 1: Farbverteilung eines Gesichts

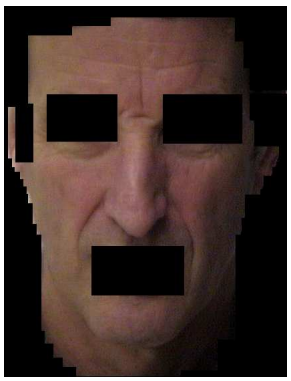

Bild 2: Handsegmentiertes Gesicht

Bei der JPEG–Kompression von Farbbildern wird in der Regel der YCrCb–Farbraum verwendet, der aus einem Helligkeits– (Y) und zwei Farbkanälen (Cr und Cb) besteht. Da das menschliche Auge empfindlicher gegenüber Helligkeit als gegenüber Farbe ist, liegen die Farbkanäle gröber abgtastet vor als der Helligkeitskanal (im Verhältnis 1:2 oder 1:4). Auch im Rahmen der Gesichtserkennung erweist sich dieser Farbraum als gut geeignet. Zum einen beschränkt sich die Hautfarbe auf einen relativ kleinen und kompakten Bereich der Cr–Cb–Ebene (vgl. Bild. 1), der aufgrund seiner ellipsenänhlichen Form durch eine Normalverteilung angenähert werden kann. Zum anderen bewirkt

die Trennung von Farb– und Helligkeitsinformation ein gewisses Maß an Robustheit gegenüber wechselnden Beleuchtungssitationen. Die Verteilung der Hautfarbe wurde anhand 51 handsegmentierter Gesichter trainiert (siehe Bild 2).

3 Schätzen der Farbverteilung

Ein Problem stellt allerdings die Gewinnung der Normalverteilung zur Farbsegmentierung dar. Da in die wahrgenommene Gesichtsfarbe nicht nur die Eigenfarbe der Haut sondern auch die Farbe der Beleuchtung mit hineinspielt, wurde in diesem Verfahren darauf verzichtet eine an Hand einer Stichprobe a priori geschätzte Verteilung zu verwenden. Statt dessen wird die Verteilung am Beginn (und unter Umständen auch während) des Programmablaufes auf die momentane Person adaptiert, wobei — wie bei der Gesichtssuche — ausschließlich auf den DC–Koeffizienten der Makroblöcke gearbeitet wird.

Um eine Verteilung überhaupt schätzen zu können, muß zuerst das Gesicht der Person im Bild gefunden werden. Deshalb wird nach einem rechteckigen Bereich gesucht, dessen Proportionen denen eines Gesichtes entsprechen, und dessen Farbe sich deutlich von seiner Umgebung unterscheidet. Zu diesem Zweck werden Schablonen verschiedener Größe über das Bild geschoben. Für jede Position wird zuerst zuerst aus den Blöcken, die innerhalb der Schablone liegen, Mittelwert und Kovarianzmatrix einer Normalverteilung geschätzt. Anschließend wird für den linken, rechten und oberen Randstreifen um die Schablone herum jeweils der Abstand von der vorher bestimmten Verteilung berechnet. Überschreitet dieser für alle drei Ränder einen Schwellwert θ, wird innerhalb der Schablone nach Augen und Mund gesucht. Ist diese Suche erfolgreich, so wurde ein Gesicht gefunden und gleichzeitig die dazugehörige Verteilung. Ein „Nebenprodukt“ dieses Vorgehens ist, daß der bei der Suche verwendete Schwellwert θ direkt bei der späteren Farbsegmentierung weiterverwendet werden kann.

Da diese Suche nach der Verteilung sehr rechenintensiv ist, da unter Umständen für sehr viele verschiedene Positionen und Schablonengrößen die Parameter geschätzt werden müssen, ist es wünschenswert, den Suchraum für das Gesicht einzuschränken. Bewegt die Person, die sich vor der Kamera befindet, während der Initialisierungsphase leicht ihren Kopf, z.B. durch Nicken, so kann eine Art Bewegungssegmentierung durchgeführt werden. Dabei werden aus zwei aufeinanderfolgenden semi–komprimierten Frames des Helligkeitskanals ein Differenzbild erstellt und diejenigen Blöcke markiert, die zwischen den beiden Aufnahmen ihren Wert signifikant geändert haben. Auf dem so entstandenen Binärbild kann wieder mittels eines Schablonenvergleiches nach einer Gesichtsregion gesucht werden, die anschließend als Suchraum für die Bestimmung der Verteilung verwendet wird.

4 Gesichtslokalisierung

4.1 Initiale Schätzung der Gesichtsposition

Zur Gesichtslokalisation wird auf den DCT–Koeffizienten eine Farbsegmentierung mit der gewonnenen Farbverteilung (siehe Abschnitt 3) durchgeführt und somit alle Blöcke im Bild markiert, die Gesichtsfarbe entsprechen. Auf dem dadurch entstandenen binaristierten Bild werden anhand von Schablonen (s. a. [2]), die über das Bild geschoben werden, Gesichter gesucht (siehe Bilder 3 u. 4).

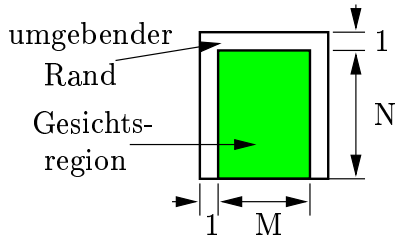

Bild 3: Gesichtsschablone

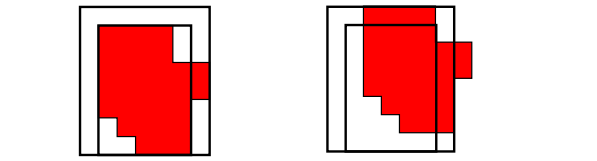

Bild 4: Segmentierungsbeispiele

Die Suche ist erfolgreich, wenn innerhalb der Maske viele Blöcke mit Gesichtsfarbe liegen und im umgebenden Rand möglichst keine.

4.2 Bestimmung der Lage der Gesichstmerkmale

Zur Lagebestimmung der Gesichtsmerkmale wird ausgenutzt, daß Augen und Mund deutliche Sprünge im Intensitätskanal in vertikaler Richtung verursachen. Aus diesen Sprüngen resultieren u. a. relativ große Werte in den ersten zwei Spalten der DCT–Koeffizientenblöcke (Bild 5).

Als vorbereitender Schritt zur Suche nach Mund und Augen wird für jeden Block innerhalb des Gesichtes die Energie über eine Auswahl von Koeffizienten der ersten und zeweiten Spalte berechnet (vgl. Bild 5). Die eigentliche Suche erfolgt auf der so erstellten „Energiekarte“ und es werden Masken für Augen und Mund verwendet, für deren Inneres jeweils die Gesamtenergie bestimmt wird. Für die Augen haben solche Masken bei halber Bildauflösung beispielsweise die Größ 4×3, für den Mund von 7×2 bis 4×4. Augen bzw. Mund befinden sich an der Position, an welcher die Maske den größten Ausschlag erzielt, wobei die Gesamtenergie zusätzlich über einer Schwelle liegen muß.

Die Suchbereiche für Augen und Mund sind durch die Anatomie des Gesichtes eingeschränkt. Um sie zu erhalten, wird das Gesicht in fünf horizontale Streifen unterteilt, wobei die Augen in den Streifen 2 und 3 zu finden sind und der Mund in 4 und 5 (siehe 6). Für die Augen gilt weiterhin, daß sie einen bestimmten Abstand zueinander haben müssen (ca. $\frac{1}{2}$ Gesichtsbreite) und daß sie nur zwei Blöcke in der Höhe differieren dürfen. Der Mund wird jeweils in Abhängigkeit von der aktuellen Augenposition gesucht, wobei sein Abstand von den Augen in y–Richtung relativ genau dem Abstand der Augenmittelpunkte zueinander entspricht ($\pm$ einen Block).

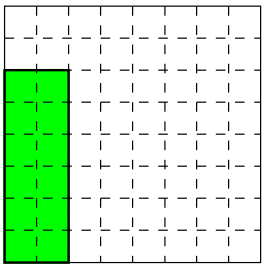

Bild 5: Suchbereich für vertikale Energien

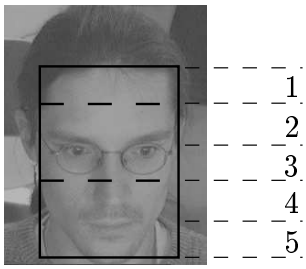

Bild 6: Einteilung des Gesichts

	Mittel	Min	Max
Zeit DCT	5,943ms	5,830ms	6,045ms
Zeit Farbverteilung	200,63ms	102,0ms	291,0ms
Zeit Verfolgen	24,76ms	23,37	26,75ms
Länge Bildfolge	28,87	14,21	38,07
Verfolgungsrate			
Gesicht	0,985	0.895	1,00
Augen	0,956	0,887	1,00
Mund	0,998	0,991	1,00

Tabelle 1: Ergebnisse

4.3 Verfolgung des Gesichtes und der Gesichtsmerkmale

Da die Lokalisation auf jedem Videobild durchgeführt wird, gehen wir davon aus, daß sich die Position des Gesichts innerhalb der Frabkanäle um höchstens um einen Block geändert hat. Die Gesichtsmerkmale sollen sich ebenfalls um höchstens einen Block ändern, und die anatomischen Lagebeziehungen müssen weiterhin erfüllt bleiben.

Wenn sich in einer festen Anzahl von aufeinanderfolgenden Aufnahmen kein Gesicht lokalisieren läßt, so wird die Suche abgebrochen und zunächst eine neue Farbverteilung geschätzt. Danach wird die Gesichtssuche bzw. Gesichtsmerkmalssuche wieder begonnen.

5 Ergebnisse

Das vorgeschlagene Verfahren wurde auf einer SGI O2 (R10000, 195 MHz, 128 MB) implementiert und getestet. Aufgrund des Hardware–JPEG–Kompressors der O2 erfolgt die JPEG–Kompression der Bilddaten in wenigen Millisekunden. Die Teil–Dekompression (bis zu den DCT-Koeffizienten erfolgt per Software.

In Tabelle 1 sind Ergebnisse der Messungen dargestellt. In der ersten Zeile werden die Zeiten für Hardware–Kompression und Software–Dekompression der Bilddaten gezeigt. Die zweite und Zeile zeigen die Zeiten für die Berechnung der Gesichtsfarbverteilung bzw. für die Verfolgung eines gefundenen Gesichts. In Zeile 4 ist die Länge der Bildfolgen (Anzahl der Bilder in der Folge) angegeben, in denen ein Gesicht er-

folgreich verfolgt werden konnte. Die weiteren Zeilen enthalten die Wiedererkennungswahrscheinlichkeiten des Gesichts bzw. der Gesichtsmerkmale.

Die Bilder 7 – 12 (6 Farbbilder) zeigen Testsituationen, Ergebnisse und Probleme des Systems.

6 Zusammenfassung

In diesem Beitrag wird ein Ansatz zur Echtzeitverfolgung von Gesichtern in Farbbildfolgen beschrieben. Das angestrebte Ziel ist die grobe Lokalisierung von Gesichtern verschiedener Farben sowie die grobe Lokalisierung von Gesichtsmerkmalen.

Das vorgeschlagene Verfahren arbeitet auf 8×8 DCT–Koeffizientenblöcken von JPEG–kompromierten Bildern. Daher werden als Ergebnis nur relativ grobe Rechtecke für Gesicht und Gesichtsmerkmale angegeben.

Dieses Verfahren wird zur Lokalisation von Patientengesichtern mit Gesichtslähmungen eingesetzt. Es ermittelt einen Bereich in der Szene beziehungsweise im Gesicht, auf den anschließend eine Kamera zoomen kann. Die Nahaufnahmen werden dann zur Analyse von Gesichtsmerkmalen und Gesichtsparesen benutzt.

Literaturverzeichnis

1. U. Ahlrichs, D. Paulus, S. Wolf: *Objektivierung der Beurteilung von Gesichtsasymmetrien durch Bildanalyse*, in T. Lehmann, I. Scholl, K. Spitzer (Hrsg.): *Workshop Bildverarbeitung für die Medizin*, Aachen, 1996, S. 125–130.
2. H. Wang, S.-F. Chang: *A Highly Efficient System for Automatic Face Detection in MPEG video*, *IEEE Transactions on Circuits and Systems for Video Technology*, Bd. 7, Nr. 4, 1997, S. 477–461.
3. S. R. Wolf, W. Schneider: *Läßt die transkranielle Magnetstimulation eine verbesserte Prognoseeinschätzung der idiopathischen Fazialisparese zu?*, *HNO*, Bd. 42, 1994, S. 559–564.
4. S. Wolf, U. Heininger, W. Schneider, D. Wenzel: *Aktuelle Aspekte in der Diagnostik und Therapie der kindlichen Faszialisparese*, *HNO*, Bd. 42, 1994, S. 624–628.

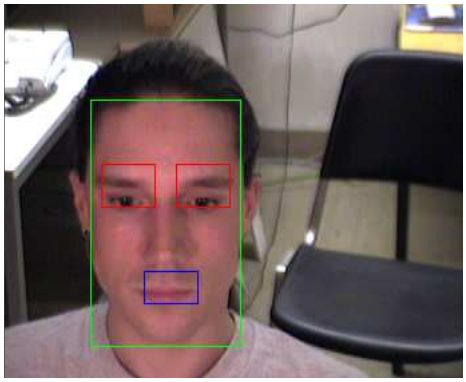

Bild 7: Gesicht gefunden

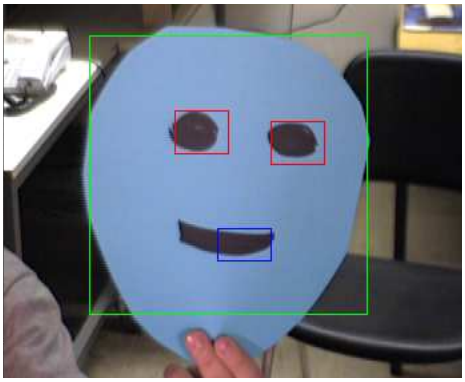

Bild 8: Änderung der Gesichtsfarbe

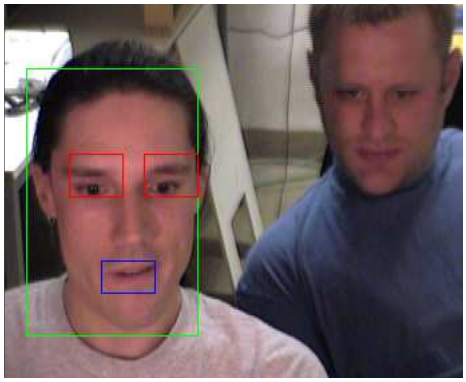

Bild 9: Zwei Gesichter

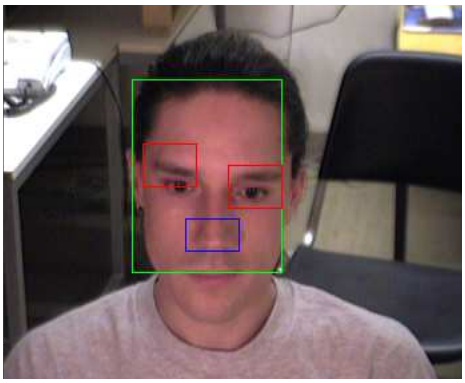

Bild 10: Erkennung n. i. O.

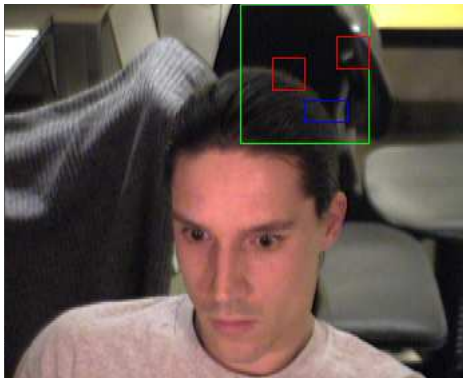

Bild 11: Fehler

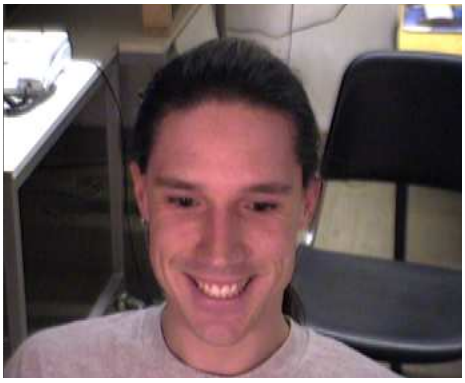

Bild 12: Fehler