

# Gesichtsdetektion mithilfe von Farb- und Formmodellen

Christian Küblbeck / Bernhard Fröba  
Fraunhofergesellschaft, Institut für Integrierte Schaltungen  
Am Weichselgarten 3  
D-91058 Erlangen

Tel. +49 (0)9131 776 549 / +49 (0)9131 776 535  
Fax +49 (0)9131 776 588

Email: [C.Kueblbeck@iis.fhg.de](mailto:C.Kueblbeck@iis.fhg.de) / [bdf@iis.fhg.de](mailto:bdf@iis.fhg.de)  
Internet: <http://www.iis.fhg.de>

STICHWÖRTER: Gesichtslokalisierung Farbe Form Gauß-Mixtur Segmentierung

## Übersicht

In diesem Artikel wird ein Ansatz beschrieben, der auf farbbasierten Bildern Gesichter lokalisiert. Hierfür wird zunächst anhand einer Lernstichprobe ein Hautfarbenmodell unter Verwendung von Histogrammen oder von Gauß-Mixturen berechnet. Dieses wird dann dazu verwendet, in Testbildern hautfarbene Pixel zu lokalisieren.

Anschließend werden zusammengehörige Pixel mithilfe von Wachstumsverfahren zu Regionen zusammengefasst (segmentiert).

Schließlich wird anhand der Form der Regionen eine Klassifikation durchgeführt. Dazu werden parametrische Formmodelle, wie z.B. Ellipsen, wie auch Kantenmodelle von Gesichtern verwendet.

In den Ergebnissen werden die verschiedenen Ansätze mit einander verglichen.

## 1 Einleitung

Die Gesichtsdetektion in komplexen Szenen ist eine Aufgabe aus der Mustererkennung, für die in der Literatur viele verschiedene Ansätze existieren. Steht als Eingabequelle eine Farbkamera zur Verfügung, so liegt es nahe, die Informationen über die Hautfarbe bei der Gesichtsdetektion mit zu verwenden. Zum einen wird dieser Ansatz dadurch erschwert, dass die Menschen verschiedener Kontinente von vornherein aufgrund ihrer unterschiedlichen Pigmentierung verschiedene Hautfarben haben. Zum anderen ergibt sich aus unterschiedlichen Beleuchtungssituationen (Aufnahme im Kunst- oder Naturlicht, bei Helligkeit oder Dunkelheit, oder zum Beispiel ein Mensch der am Rechner vor einem „blue screen“ sitzt) eine sehr unterschiedliche Farbdarstellung physikalisch gleicher Gesichter.

Verwendet man (aus einer breiten Stichprobe gewonnene) sehr allgemeine Farbmodelle, so können zwar mit hoher Wahrscheinlichkeit Gesichter beliebiger Herkunft und unter variablen Beleuchtungen sehr zuverlässig gefunden werden, allerdings ist die Rate an Fehldetektionen relativ hoch.

Deshalb wird in dem hier beschriebenen Ansatz die Analyse (Verifikation) der gefundenen hautfarbenen Region mithilfe von Formmodellen durchgeführt. Auf diese Weise kann nicht nur die Zahl von Fehldektionen eingeschränkt werden, es ist außerdem möglich, andere hautfarbene Körperteile wie die Hände aufgrund ihrer Form auszuschließen.

## 2 Lokalisierung der hautfarbenen Regionen

In diesem Abschnitt wird das Vorgehen beschrieben, mit dem aus einem gegebenem Farbbild über Vorverarbeitung und Vergleich mit einem aus einer Lernstichprobe gewonnenen Modell eine Klassifikation einzelner Pixel durchgeführt werden kann. Hierbei wird insbesondere auf zwei verschiedene Arten der Modellbildung eingegangen, die Modell-schätzung mithilfe von Histogrammen oder mithilfe von Gauß-Mixturen.

### 2.1 Wahl des Farbraums

Die Aufnahme und Darstellung der Farbbilder wird grundsätzlich im RGB-Farbraum durchgeführt. Allerdings weisen die RGB-Werte zum einen hohe Redundanz, zum anderen eine hohe Anfälligkeit gegenüber Intensitätsveränderungen im Bild auf. Deshalb wurden hier der normalisierte rgb-Farbraum und der HSV-Farbraum verwendet.

Der rgb-Farbraum ergibt sich durch folgende Gleichung:

$$r = \frac{R}{R+G+B}; g = \frac{G}{R+G+B}; b = \frac{B}{R+G+B} \Rightarrow r + g + b = 1 \quad (1)$$

Aufgrund der Normierung geht die Information über die Intensität des Pixels verloren. Damit ist die blaue Farbe redundant, die Berechnung von Farbverteilungen wird vereinfacht.

Der HSV-Farbraum basiert auf den drei Parametern Farbton (**hue**), Sättigung (**saturation**) und Helligkeit (**value**) und wird hier nach folgender Vorschrift berechnet:

$$\begin{aligned} V &= \text{MAX}(R, G, B) \\ S &= \frac{V - \text{MIN}(R, G, B)}{V} \\ H &= \begin{cases} \frac{60(G - B)}{V - \text{MIN}(R, G, B)}, & \text{falls } R = \text{MAX}(R, G, B) \\ 120 + \frac{60(B - R)}{V - \text{MIN}(R, G, B)}, & \text{falls } G = \text{MAX}(R, G, B) \\ 240 + \frac{60(R - G)}{V - \text{MIN}(R, G, B)}, & \text{falls } B = \text{MAX}(R, G, B) \end{cases} \end{aligned} \quad (2)$$

Die Sättigung beschreibt hierbei die „Reinheit“ des Farbtons und wird für einander angleichende Farbwerte  $R, G, B$  singulär.

Ein Farbmodell kann nun entweder nur aus dem Farbtonwert oder aus der Kombination aus Farbton und Sättigung gebildet werden. Eine Hinzunahme der Intensität macht hingegen wenig Sinn.

## 2.2 Modellbildung mithilfe von Histogrammen

Histogramme lassen eine nichtparametrische Beschreibung einer Lernstichprobe im Farbraum zu. Hierbei lässt sich für jeden Farbwert  $\underline{p}$  eine Häufigkeit seines Auftretens

$$h(\underline{p}) \equiv \frac{n(\underline{p})}{N} \text{ mit } N \equiv \sum_{\text{alle } \underline{p}} n(\underline{p}) \quad (3)$$

angeben. Je nach Verwendung des Farbmodells wird ein eindimensionales (nur Verwendung des Farbtons) oder ein zweidimensionales (Verwendung von Farbton und Sättigung oder von normiertem Rot- und Grünwert) Histogramm verwendet. Die Dimension der Histogramme beträgt in unseren Versuchen jeweils 256.

## 2.3 Modellbildung mithilfe von Gauß-Mixturen

Anstelle von Histogrammen kann auch eine parametrische Modellierung der Hautfarbe mithilfe der Normalverteilung durchgeführt [1, 2].

Die Parameter der einfachen Gaußverteilung sind der Mittelwertvektor  $\underline{\mu}$  und die Kovarianzmatrix  $\underline{C}$ , berechnet aus der Lernstichprobe.

Da mithilfe von einfachen Normalverteilungen die Charakteristik der Hautfarbe nur annähernd beschrieben werden kann, bietet sich der Einsatz von Gaußmixturen zur Beschreibung der Modelle an [3, 4]. Hierfür beschreibt man mit

$$p(\underline{x}) = \sum_{i=1}^m \pi_i p_i(\underline{x}) = \sum_{i=1}^m \pi_i e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)^T (\underline{C}_i)^{-1} (\underline{x}-\underline{\mu}_i)}; \sum_{i=1}^m \pi_i = 1 \quad (4)$$

die Häufigkeitsverteilung des Pixels bei Modellierung mithilfe einer gewichteten Summe von einzelnen Gaußverteilungen.

Die Parameter können mit dem EM-Algorithmus<sup>1</sup> geschätzt werden. Bei dieser iterativen Methode werden zunächst die Gewichte  $\pi_i$  der einzelnen Verteilungen gleich gesetzt und die Mittelwerte und Kovarianzmatrizen mithilfe der Momentenmethode geschätzt. Anschließend werden gemäß [4] die Parameter so lange angepasst, bis der Schätzfehler eine vorgegebene Schranke unterschreitet.

Aus wie vielen Einzelverteilungen man nun die Mischverteilung zusammensetzt, hängt grob davon ab, wie viele Cluster in den Farbhistogrammen geschätzt werden können.

## 2.4 Farbklassifikation

Nach Erstellung der Modelle anhand einer Lernstichprobe wird nun zunächst pixelweise klassifiziert, ob ein bestimmter Punkt zu einer hautfarbenen Region gehört oder nicht.

---

<sup>1</sup> von expectation maximation

Bei Verwendung eines histogrammbasierten Modells verwendet man das Histogramm als „lookup“-Tabelle und erhält somit direkt für jeden Pixel eine Wahrscheinlichkeit, mit der dieser Pixel hautfarben ist.

Verwendet man das Gauß-Mixtur-Modell, so kann man ebenfalls pixelweise die Wahrscheinlichkeiten mithilfe von Gleichung 4 berechnen. Anschließend werden die gewonnenen Wahrscheinlichkeiten binarisiert: Die Schwelle  $S$  wird dynamisch gemäß [5] bestimmt. Der Ablauf der gesamten Kette ist in Abbildung 2 demonstriert.

Um Störungen aus den Bildern zu eliminieren, wird das Bild mithilfe des kantenerhaltenden Median-Filters bearbeitet.

### **3 Segmentierung der hautfarbenen Regionen**

Ziel der Segmentierung ist es, aus dem gewonnenen Binärbild räumlich benachbarte und zusammenhängende Regionen zu bestimmen und zusammenzufassen.

#### **3.1 Regionenwachstum**

Bei diesem Verfahren werden Regionen pixelweise aufgebaut. Dabei wird ausgehend von Keimpunkten eine Indizierung der einzelnen Pixel durchgeführt (siehe auch [6]). Hierbei werden benachbarte Pixel mit gleichen Indizes versehen. Hier wurde eine einfache Zweierachbarschaft verwendet, die eine schnellere Berechnung bei gleichbleibend guten Ergebnissen zulässt.

Idealerweise enthält das Binärbild eines Farbbildes mit einer einzigen Person genau eine zusammenhängende Region, nämlich die des Gesichts. Jedoch wird zum einen oft mehr als eine Region detektiert, zum anderen wird das Gesicht durch mehr als eine Region dargestellt. Abbildung 3 verdeutlicht diese Problematik.

#### **3.2 Meanshift-Algorithmus**

Das Meanshiftverfahren [7] basiert auf einem nicht parametrischen Verfahren zur Minimierung der Gradienten der Verteilungsdichte. Ziel ist hierbei die Detektion von Regionen mit hoher Dichte (*Cluster*). Für die Lokalisation werden Suchfenster verwendet, deren Größe und Form in Abhängigkeit von vorliegenden Wahrscheinlichkeitsdistribution verschoben und optimiert werden.

Hauptproblem dieses Verfahrens ist die Initialisierung der Bildfenster. Sie wurde durch den Einsatz einer Bildpyramide gelöst. In dem Bild der kleinsten Auflösungsstufe wird das Regionenwachstumsverfahren angewandt, um die initialen Zentren und Größen der Regionen zu bestimmen. Dieses Vorgehen bietet neben seiner Schnelligkeit den Vorteil, dass kleine, unbedeutende Regionen nicht berücksichtigt werden müssen.

### **4 Extraktion von Gesichtsmerkmalen**

Nach Auffinden und Segmentierung der hautfarbenen Bereiche ist der nächste Schritt die Verifikation der gefundenen Regionen. Dieser Schritt ist nötig, um die bei diesem Verfahren gegebene relativ hohe Falschdetektionsrate zu vermindern. Wir stellen hier kurz zwei verschiedenartige Verfahren vor, die vergleichsweise untersucht wurden.

## 4.1 Formbasierte Verfahren

Die Form des Gesichts kann durch eine Ellipse angenähert werden. Überprüft man also, ob eine segmentierte Region  $\mathfrak{R}$  von einer Ellipse gut umschrieben werden kann, so lässt sich auf die Anwesenheit eines Gesichtes schließen. Die Lage und die Orientierung der Ellipse lässt sich aus der segmentierten Region durch Bildung der ersten und zweiten Momente berechnen. Als Ähnlichkeitsmaß  $\xi$  zwischen Region und Ellipse lässt sich die Zahl falsch liegender Pixel geteilt durch die Fläche der Ellipse angeben. Als zusätzliche Konfidenzmaße lassen sich das Verhältnis zwischen großer und kleiner Halbachse  $b$  und  $a$  der Ellipse angeben, sowie die Lage  $\theta$  der Ellipse (siehe Abbildung 4).

Folgende Entscheidungsregel hat sich als vorteilhaft erwiesen:

$$\text{falls } \left( \xi \leq 0.7 \wedge 1 \leq \frac{b}{a} \leq 2.3 \wedge -\frac{\pi}{3} \leq \theta \leq \frac{\pi}{3} \right), \text{ dann ist } \mathfrak{R} \text{ ein Gesicht.}$$

## 4.2 Heuristisches Verfahren

Bei dem zweiten untersuchten Verfahren werden mehrere Gesichtsm Merkmale extrahiert. Die wichtigsten Merkmale sind hierbei die Augen, Augenbrauen, Nase und Nasenlöcher und der Mund. Diese können durch Kantenfilterung des Binärbildes gefunden werden (siehe Abbildung 5). Im Rahmen unserer Untersuchungen hat sich herausgestellt, dass insbesondere die Augenbrauen und der Mund detektiert werden können. Wie eine genaue Zuordnung der Merkmalskandidaten zu den Merkmalen gemacht werden kann, ist genauer in [8] beschrieben. Anschließend kann mithilfe antropometrischer Annahmen das Verhältnis überprüft und ein Konfidenzmaß ermittelt werden.

Die Verfahren 4.1 und 4.2 lassen sich prinzipiell auch kombinieren, da ersteres Verfahren auf die äußere Form und letzteres auf die innere Struktur des Gesichtes abzielt.

## 5 Versuche und Ergebnisse

Zur Erstellung der Hautfarbenmodelle wurde eine Lernstichprobe erstellt, die ein großes Spektrum sowohl verschiedenartiger Haut als auch verschiedenartiger Lichtquellen berücksichtigt (siehe Abbildung 1). Um die Algorithmen zu testen, wurde eine Teststichprobe aus 218 Farbbildern zusammengestellt. Hierbei wurde weniger Wert auf die Größe, sondern vielmehr auf den Variantenreichtum der Stichprobe Wert gelegt. Sie besteht aus 236 weißen, elf asiatischen und zwölf dunkelfarbigem Gesichtern (einige Bilder enthalten mehrere Gesichter).

Zunächst werden die Verfahren zur Modellbildung miteinander verglichen. Hierbei bedeutet die FRR-Rate die Zahl fälschlich als nicht-hautfarben klassifizierten Pixel, während die FAR-Rate die Zahl fälschlich als hautfarben klassifizierten Pixel bedeutet. Die Berechnung erfolgte im HSV-Raum.

Verfahren	1-D-Histogramm	2-D-Histogramm	1D-Gauß- verteilung	2D-Gauß- verteilung	2-D-Gaußmixture
FRR-Rate	0.16	0.13	0.03	0.09	0.13
FAR-Rate	0.33	0.43	0.73	0.62	0.51

Es lässt sich erkennen, dass mit den 1-D und 2-D-Histogrammen und mit dem Gaußmixture-Verfahren die beste Erkennungsrate erreichen lässt. Desweiteren wird überprüft, in wie weit die Wahl des Farbraums eine Rolle spielt.

Verfahren	HSV, 2-D-Histogramm	Norm. Rgb, 2-D-Histogramm	HSV, 2-D-Gaußmixture	Norm. Rgb, 2-D-Gaußmixture
FRR-Rate	0.13	0.26	0.13	0.17
FAR-Rate	0.43	0.36	0.51	0.42

Insgesamt lässt sich feststellen, dass der HSV-Raum etwas bessere Ergebnisse liefert, insbesondere unter schwierigen Beleuchtungsbedingungen.

Die letzte Tabelle zeigt nun die FAR- und FRR-Raten unter Hinzunahme der Merkmalsextraktionsverfahren. Die Ergebnisse beruhen auf der regionenbasierten Segmentierung des mit dem 2-D-HSV-Modell histogrammbasiert berechneten Binärbild.

Verfahren	Ellipsen-Fitting	Heuristisches Verfahren	Kombination aus beiden Verfahren
FRR-Rate	0.16	0.09	0.15
FAR-Rate	0.33	0.39	0.15

Es lässt sich erkennen, dass insbesondere durch eine Kombination der beiden Analyse-Verfahren eine deutlich Reduktion der Fehlerrate erreichen lässt.

## 6 Danksagung

Die hier beschriebene Arbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Projektes EMBASSI unterstützt.

## 7 Literatur

1. Yang, J., W. Lu, and A. Waibel, *Skin-color Modelling and Adaption*, in ACCV'98. 1998.
2. Yang, M.-H. and N. Ahuja, *Detecting Human Faces in Color Images*. Proceedings of the International conference on Image Processing, 1998. 1: p. 127-130.
3. Raja, Y., S.J. McKenna, and S. Gong, *Segmentation and Tracking using colour Mixture Models*, in *Third Asian Conference on Computer Vision*. 1998.
4. Yang, M.-H., *Gaussian Mixture Model for Human Skin Color and its Applications in Image and Video Databases*. 1999.
5. Sonka, M., V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. 1996: International Thomson Computer Press.
6. Hering, F., *Lagrangesche Untersuchungen des Strömungsfeldes unterhalb der wellenbewegten Wasseroberfläche mittels Bildfolgenanalyse*. 1996, Ruprecht - Karls - Universität: Heidelberg. p. 154.
7. Fukunaga, K., *Introduction to Statistical Pattern Recognition*. 1990: Academic Press.
8. Jeng, S.-H., et al., *Facial Feature Detection Using Geometrical Face Model: An Efficient Approach*. Pattern Recognition, 1998. 31.



Abbildung 1 Ausschnitt aus der Lernstichprobe zur Bildung eines hautfarbenen Modells. Deutlich zu erkennen ist hierbei, dass die Stichprobe sowohl verschiedenartige Haut als auch verschiedenartige Lichtquellen berücksichtigt.

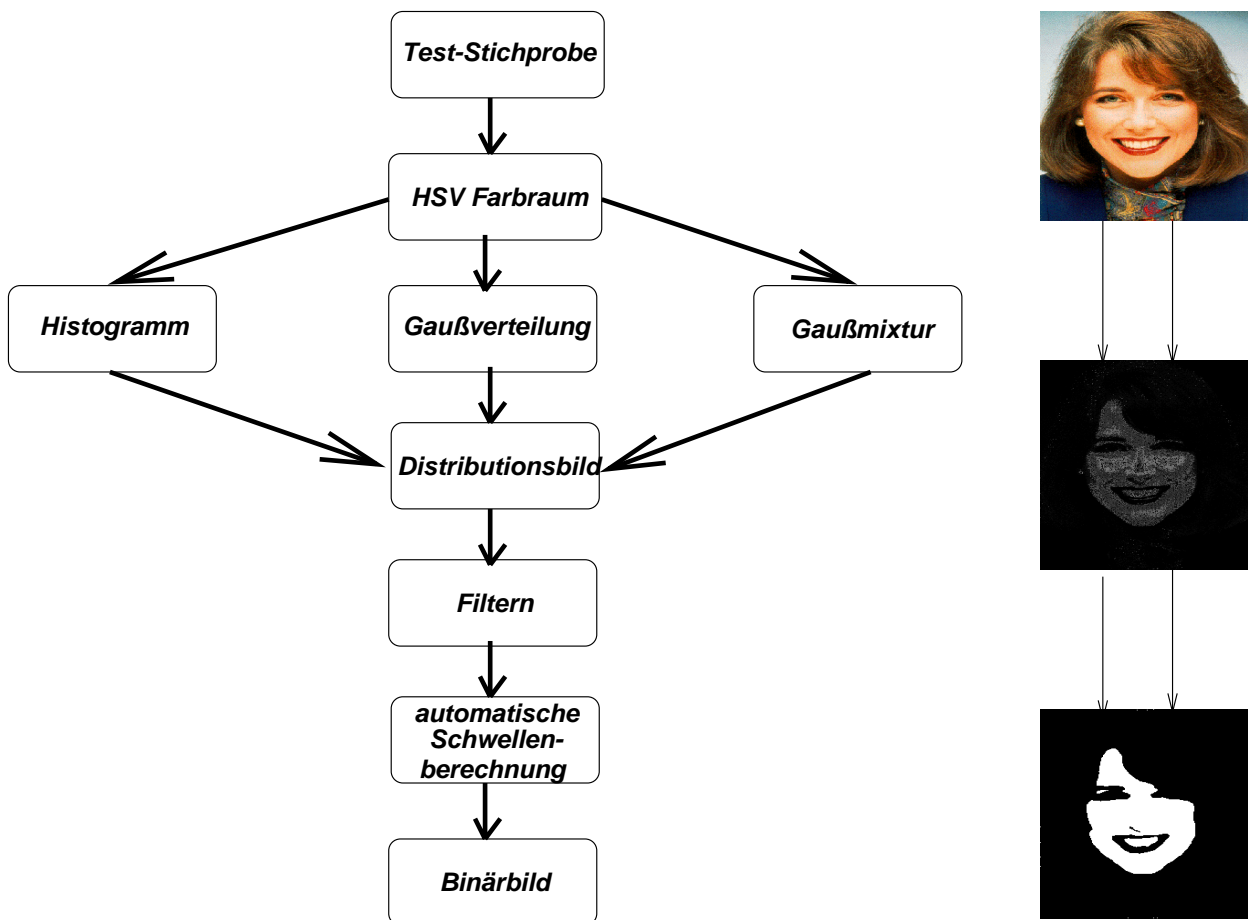


Abbildung 2 Blockdiagramm der Binärbildbildung aus einem gegebenem Testbild über ein Distributionsbild

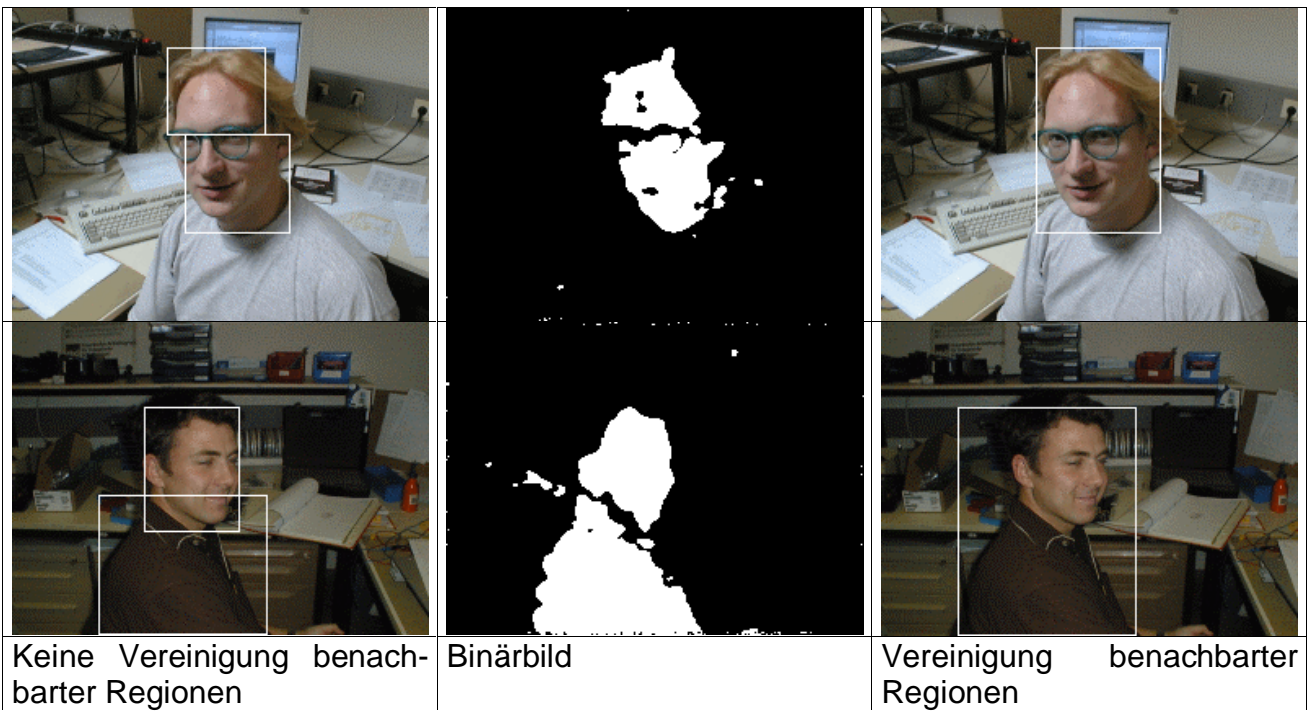


Abbildung 3 Beispiele zur regionenwachstumssegmentierung: Im oberen Fall macht eine Vereinigung benachbarter, hautfarbener Gebiete Sinn, das Gesicht wird vollständig erkannt. Unten jedoch tritt durch die Vereinigung einer Fehler auf, da das Gesicht schon zuvor vollständig erkannt ist.

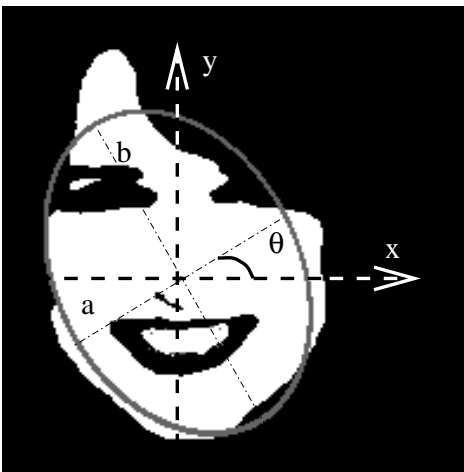


Abbildung 4 Überprüfung der segmentierten Region mithilfe einer Ellipse.



Abbildung 5 Berechnung zunächst des farbsegmentierten Bildes und anschließend des des Kantenbildes als Ausgang für die heuristische Formanalyse.