

Gesichtserkennung mit handelsüblicher Hard- und Software

Wettbewerb "Jugend Forscht" 2008

Daniel Rocholz (19 Jahre)

Arbeitsgemeinschaft "Jugend Forscht"
des Christian-Gymnasiums Hermannsburg
Leitung: StD Thomas Biedermann

Inhaltsverzeichnis

Idee	3
Definition	3
Andere Begriffe	3
Erkennungsmethoden	4
Anwenden von Definitionen	4
Vergleichen mit Trainingsbildern	4
Kombination der Methoden	4
Anwendungen	4
Etablierte Einsatzgebiete	4
Eigene Ideen	5
Eigene Umsetzung	6
Bilderfassung	6
Bildaufbereitung	6
Geschickt ignorieren	6
Die Suche nach Gesichts-Kandidaten	6
Auffinden von hautfarbigen Flächen	7
Untersuchung von Kandidaten	8
Fazit	8
Die Zukunft von Gesichtserkennung	9
Danksagung	9
Quellen	10
Anhang	11

Idee

Die Idee zu dieser Arbeit kam mit dem Interesse an Informationstechnologie im Allgemeinen und den öffentlichen Diskussionen um die Vereinbarkeit von Überwachungsmaßnahmen zur Terror-Bekämpfung mit den Bürgerrechten. Wie auch bei der Online-Anmeldung in der Kurzfassung angegeben, beabsichtigte ich zunächst, mich mit den Mechanismen zur Gesichtserkennung in Kamera-Überwachungssystemen zu beschäftigen. Die Vielfalt an Forschungsprojekten und privaten Veröffentlichungen der Entwicklergemeinschaft animierte mich jedoch im Laufe des Projektes, mein Augenmerk auf die nicht-kommerziellen und nicht politisch motivierten Entwicklungen zu legen. Zwei eigene Ideen (siehe "Anwendungen") waren der Grund, das Projekt auf herkömmliche Computerhardware und Software zu beschränken. Ziel des Projektes ist es somit, einen Einblick in existierende Methoden zur Gesichtserkennung zu gewinnen und mit dem erworbenen Wissen eine eigene Anwendung zu entwickeln, welche das Videobild einer gewöhnlichen Webcam nach Gesichtern durchsucht.

Definition

"Gesichtserkennung" wird von verschiedenen Quellen auf unterschiedlichste Art definiert. Es ist zu unterscheiden zwischen

- dem Auffinden einer bekannten Anzahl von Gesichtern,
- dem Auffinden einer unbekannt Anzahl von eventuell vorhandenen Gesichtern und
- der Identifizierung eines Gesichtes anhand eines gespeicherten Abbildes.

Eine gute Unterscheidung für diese Aufgaben bieten die englischen Begriffe

- face localisation,
- face recognition und
- face authentication.

Face recognition ist somit eine spezialisierte Form von face localisation, face authentication wiederum setzt auf das lokalisieren von Gesichtern, um anschließend einen Vergleich mit einer gespeicherten Aufnahme anzustellen.

In dieser Arbeit konzentriere ich mich auf "face recognition" und benutze nur für diesen Vorgang die Übersetzung "Gesichtserkennung".

Andere Begriffe

Die Leistungsfähigkeit eines Gesichtserkennungs-Algorithmus wird in der Regel anhand einer Datenbank von Bildern bestimmt, deren Inhalt, also die Anzahl und Position von Gesichtern, bekannt ist. Die *Erkennungsrate* wird errechnet aus dem Anteil der erkannten Gesichter an der Anzahl aller in der Datenbank vorhandenen Gesichter. Diese Angabe lässt allerdings noch keine Aussage über die Qualität des Algorithmus zu, da eine hohe Erkennungsrate auch eine hohe Fehleranzahl beim Ausschluss von Bereichen bedingt, die kein Gesicht enthalten (sogenannte *non-faces*). Diese Art von Fehlerkennung wird allgemein als "falsch positiv" (*false positive*) bezeichnet. Sie wird als Absolutwert angegeben und lässt in Relation zur Anzahl der Testbilder einen Rückschluss zu.

Erkennungsmethoden

Grundsätzlich lassen sich die existierenden Herangehensweisen zum Erkennen von Gesichtern in Bildmaterial in zwei Kategorien unterteilen:

Anwenden von Definitionen

Für einige Merkmale eines Gesichtes gelingt es, aus Beobachtungen klare Definitionen zu formulieren, die auf das Bildmaterial angewendet werden. Dazu gehören Proportionen und Abstände der Gesichtsmarkmale, Farben, Texturen und Formen.

Ein Beispiel für das Gewinnen einer solchen Definition gibt das Unterkapitel "Farbe" im praktischen Teil der Arbeit.

Vergleichen mit Trainingsbildern

Auch ohne feste Definitionen können durch statistische Vergleiche der Pixelmuster mit Datenbanken von Gesichtern gute Ergebnisse erzielt werden. Der Erfolg ist dabei abhängig von der Vielfalt und Größe der Datenbank. Die derzeitigen Fortschritte bei der Simulation von Neuronalen Netzwerken und die ständig steigenden Rechenkapazitäten machen es möglich, die analysierten Daten selbst als Trainingsmaterial zu verwenden. Die neuesten Erkennungsmethoden verwenden lernfähige Algorithmen und erhöhen somit im Idealfall ihre Erkennungsrate über die Laufzeit.

Kombination der Methoden

Gewöhnlich werden beide Ansätze kombiniert, um eine hohe Erkennungsrate und gleichzeitig wenige falsch-positiv Ergebnisse zu erreichen. Je nach Priorisierung der Methode wird unterschieden:

- Die *definitionsbasierte Top-Down-Methode* ("knowledge based Top-Down method") zeichnet sich durch eine schnelle Verarbeitung der Bilddaten aus, da die zu untersuchenden Bereiche zunächst durch klare Definitionen stark eingegrenzt werden. Die Schwierigkeit dieses Ansatzes ist es, das menschliche Wissen in klar definierte Programmparameter zu "übersetzen", welche dennoch auf verschiedene Bildsituationen, Posen und Lichtverhältnisse ansprechen.
- Die *merkmalbasierte Bottom-Up-Methode* ("Bottom-Up feature-based method") verfährt anders herum: es wird zunächst nach allgemeingültigen (unabhängig von Beleuchtung, Pose, Blickwinkel) Merkmalen gesucht, deren Verhältnisse statistisch ausgewertet werden.

Anwendungen

Durch die neuesten politischen Entwicklungen um die Terror-Bekämpfung wird der Begriff "Gesichtserkennung" vermehrt mit Überwachungsmechanismen assoziiert. Entsprechend der Idee dieser Jugend-forscht Arbeit soll auch die Nützlichkeit der Technologie für den Privatanwender aufgezeigt werden:

Etablierte Einsatzgebiete

- Moderne Unterhaltungselektronik (Spielekonsolen, Mobiltelefone) setzt Gesichts- und Gestenerkennung als neue Methode der Mensch-Maschine-Interaktion ein.
- Bei Internet-Videotelefonie und Videokonferenzen wird Gesichtserkennung eingesetzt, um das Gesicht des Kommunizierenden auf den Bildausschnitt zu vergrößern - eine einfachere Ausrichtung der Kamera auf der einen und ein besseres Bild auf der anderen Seite sind das Ziel.
- Die neueste Generation von Foto-Verwaltungssoftware nutzt Gesichtserkennung (und -wiedererkennung) um Fotosammlungen anhand der abgebildeten Personen zu katalogisieren.
- Der Suchmaschinenendienst Google bietet (in der Englischen Version) seit 2007 eine Filterung der Suche nach Gesichtern an, um das Finden von Portraits zu erleichtern.

- Bei Diensten wie Google StreetView werden mitunter große Mengen an Bilddaten veröffentlicht, die unter Umständen die Privatsphäre von Unbeteiligten verletzen. Seit Mitte 2008 wird im Fall der Google Straßenfotografien Gesichtserkennung eingesetzt, um Gesichter vollautomatisch unkenntlich zu machen.

Eigene Ideen

Zwei eigene Ideen, die ich mit dem erworbenen Wissen versuchen werde umzusetzen:

- Eine auf dem PC-Monitor installierte Webcam kann eingesetzt werden, um Strom zu sparen: Die Bildausgabe (Grafikkarte und Monitor) kann in den Standby versetzt werden, wenn der Nutzer den Arbeitsplatz verlässt. Außerdem kann ein mitunter lästiges Problem umgangen werden: Bei rein passiver Nutzung des Computers (z.B. Fernsehen, Lesen) kann das ungewollte Aktivieren des Bildschirmschoners verhindert werden. Der Bildschirmschoner kann, durch die Webcam gesteuert, immer nur dann aktiviert werden, wenn der Monitor "nicht in Gebrauch" ist, unabhängig davon, ob der Nutzer die Maus in der Hand hält. Die dazu notwendige Gesichtserkennung kann auf einem sehr ressourcenschonenden Niveau betrieben werden, da nur ein einzelnes, groß abgebildetes Gesicht detektiert werden muss.
- Wie durch den Student Johnny Chung Lee der Carnegie Mellon Universität in Pittsburgh demonstriert [1], kann ein normaler Computermonitor als dreidimensionales Display genutzt werden, wenn der Standpunkt des Betrachters bekannt ist. Während sein Ansatz den Kopf des Betrachters mithilfe der Infrarotkamera einer Wii Spielekonsole und "Positionslampen" verfolgt, könnte eine normale Webcam ebenso eingesetzt werden, um die genaue Position des Betrachters per Gesichtserkennung festzustellen.



Abb. 1: Johnny Lee mit Infrarotsender und Fernbedienung

Mit der Weiterentwicklung von Techniken wie "face authentication", Posen-, Gesten- und Mimik- sowie Spracherkennung werden zahllose andere Anwendungen denkbar, z.B. das Ersetzen von Passwörtern durch die eigenen Biometriedaten [2] und die Bedienung des Computers mit Gesten und Sprachkommandos.

Eigene Umsetzung

Um nach der Entwicklung der Gesichtserkennung die beiden Ideen umsetzen zu können, bemühe ich mich um eine klare Strukturierung und leichte Adaptierbarkeit des Programmcodes für andere Anwendungen. Der Einsatz von Windows Standardschnittstellen und der Visual Basic Programmierumgebung gewährleistet, dass das Programm auf allen neueren Windows-Betriebssystemen lauffähig ist.

Bilderfassung

Eine normale Webcam wird per USB an den Rechner angeschlossen. Über die Treiberschnittstelle WIA ("Windows Image Acquisition") und die Windows API ("Application Programming Interface") kann das Videobild in Form von Einzelbildern abgefragt werden. Jedes Einzelbild ("Frame") wird durch das Programm in seine Pixel und deren Farbinformationen zerlegt, welche dann die durch den Benutzer aktivierten Routinen durchlaufen. Die bildverarbeitenden Mechanismen wurden modular aufgebaut, um eine spätere Ergänzung von neuen Methoden zu erleichtern.

Zum einfacheren Testen des Programmes an unterschiedlichen Bildsituationen wurde, außer der Anbindung an die Webcam, auch das Verarbeiten von Bild- und Videodateien vorgesehen. Durch das Testen der nachfolgenden Verarbeitungsschritte an jeweils dem gleichen Bildmaterial lassen sich auf einfache Weise Vergleiche zwischen verschiedenen Konfigurationen anstellen.

Bildaufbereitung

Da günstige Webcams häufig ein stark verrauschtes und unter- oder überbelichtetes Bild liefern, sollte das Bild vor der Verarbeitung bestmöglich aufbereitet werden. Da falsch belichtete Bereiche ohnehin keine Bildinformationen enthalten, kann durch deren Ignorierung Rechenarbeit gespart werden und eine Verfälschung der nachfolgenden Schritte verhindert werden. Die Bildschirmfotos in diesem Abschnitt der Arbeit markieren ignorierte Bereiche rot. Eine Methode zum Entrauschen des Bildes wurde bis zum Fertigstellen der Arbeit noch nicht implementiert; zur Verbesserung der Erkennungsrate wird vermutlich in der Weiterentwicklung des Projektes eine freie Sammlung von Grafikmethoden (z.B. ein Windows-Kompilat der ursprünglich von Linux-Entwicklern programmierten "Imagemagick" Bibliothek [3]) verwendet.

Geschickt ignorieren

Nachdem unverwertbare Pixel bereits ignoriert werden, lässt sich die Menge der zu untersuchenden Pixel weiter reduzieren: Basierend auf der Annahme, dass die Webcam fest installiert und der Benutzer des Computers ständig leicht in Bewegung ist, können statische Bereiche des Bildes ignoriert werden. Die umgekehrte Anwendung, die Erkennung von Bewegungen in Videomaterial nennt man "motion-detection".

Die Suche nach Gesichts-Kandidaten

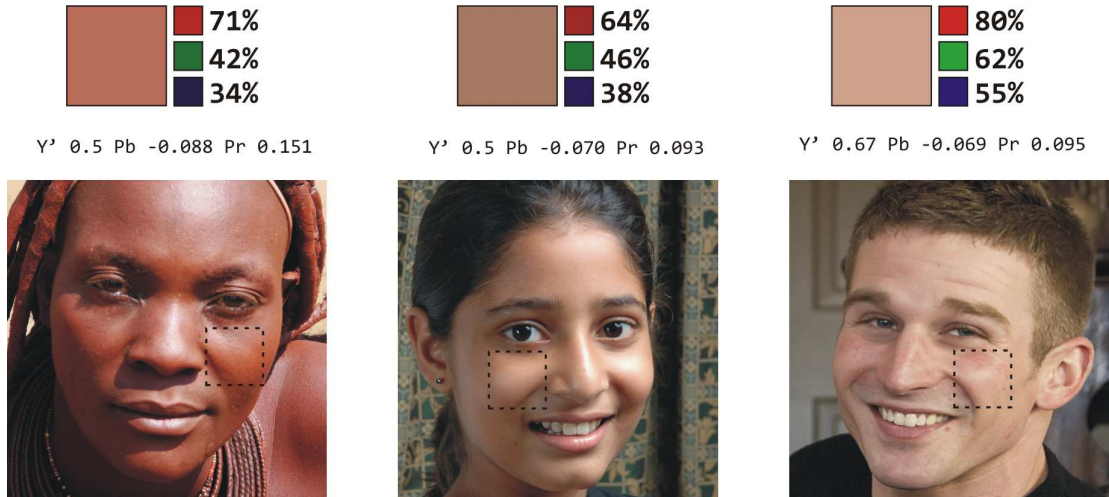
Um effizient nach Gesichtern zu suchen, wird die Top-Down-Methode (Siehe "Erkennungsmethoden") angewandt: Das Bildmaterial durchläuft eine Hierarchie von Untersuchungen.

Fundstellen von möglichen Kandidaten werden in Form eines Rechtecks beschrieben und zur jeweils nächsten Stufe weitergeleitet, dort mit höherem Rechenaufwand genauer untersucht und eventuell verworfen. Bei dieser Verfeinerungsmethode nimmt die false-positive-Rate von oben nach unten ab, der Rechenaufwand jedoch zu.

Diese Hierarchie bedingt, dass die Filterregeln in den obersten Regeln nicht zu strikt ausfallen dürfen, um eine hohe Erkennungsrate zu erreichen. Im Gegensatz dazu steigt mit Lockerung der Regeln die Anzahl von false-positives und der Rechenaufwand - ein guter Mittelweg will gefunden werden.

Auffinden von hautfarbigen Flächen

Bilddaten von Digitalkameras und Webcams liegen auf dem Computer meistens im RGB Format und einer Farbauflösung von 8 Bit vor. Jedem Pixel wird dabei eine Intensität zwischen 0 und 255 für die drei Farben Rot, Grün und Blau zugeordnet. Einem solchen Bild können somit schon ohne eine aufwändige Verarbeitung der Daten Informationen entnommen werden, die auf das Vorhandensein eines Gesichtes hindeuten: die Hautfarbe.



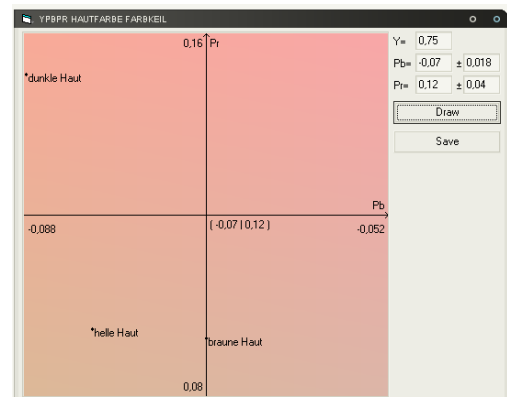
In Abb. 1 wurden die RGB-Werte von drei verschiedenen Hautfarben über mehrere Hundert Pixel (gestrichelte Markierung) gemittelt und im oberen großen Feld dargestellt, zusätzlich dazu wurde die Zusammensetzung der Farbe ermittelt.

Bei der Verarbeitung der RGB-Daten ergibt sich ein Problem: ein- und dieselbe Farbe erhält unter unterschiedlichen Lichtbedingungen deutlich andere RGB-Werte, während die Relationen der Farbanteile untereinander konstant bleiben.

Eine einfache Möglichkeit, die Farbdarstellung von der Helligkeit zu trennen, ist die Umwandlung in ein anderes Farbsystem. Zur Demonstration sind die Hautfarben in Abb. 1 zusätzlich im $Y'P_bP_r$ -System angegeben, wobei Y' die "Luminanz" wiedergibt, welche der Farbe eine dem menschlichen Sehen nachempfundenen Intensitätswert zwischen 0 und 1 zuordnet. Die Farbinformation ist in den beiden Komponenten P_b und P_r enthalten, welche Werte von -0.5 bis 0.5 annehmen. [4]

Bei einem Vergleich der drei Farben in $Y'P_bP_r$ -Darstellung fällt auf, dass alle drei Farbzusammensetzungen nur 1,9% bzw. 5,8% des Farbraums einnehmen, während die Helligkeit stärker auseinander liegt.

Abb. 2 stellt eine Farbtafel dar, welche mittels eines einfachen Testprogramms aus den gewonnenen Daten erzeugt wurde. Um eine zweidimensionale Darstellung zu ermöglichen, wurde dem Diagramm ein Intensitäts- (Y' -) Wert von 75% zugrundegelegt. Im Koordinatenursprung liegt der aus den drei Hautfarben errechnete Mittelwert, auf den Achsen wurde eine Abweichung um 1,8% bzw. 4% jeweils in positive und negative Richtung aufgetragen.



Da in der Praxis je nach Aufnahmemethode leichte Farbverfälschungen auftreten, sollte der Definitionsbereich für "Hautfarbe" etwas größer gewählt werden. Sofern möglich liefert eine Kalibrierung des Bildes auf eine bestimmte Hautfarbe, ähnlich einem Weißabgleich bei Digitalkameras, zusammen mit strikteren Filterparametern die besten Ergebnisse, jedoch wird die Erkennung von unterschiedlichen Hauttönen dadurch negativ beeinträchtigt.

Abb. 3 zeigt eine Streuung von 5% bzw. 10% um den gleichen Mittelwert wie in Abb. 2, diese Werte lassen eine zuverlässigere Erkennung von Hautfarbe in Bilddaten mit unterschiedlicher Farbtemperatur zu. Abschließend wurden die Parameter aus Abb. 1 in das Programm implementiert und auf eine Reihe von Testbildern angewandt. Abb. 4 zeigt das Ergebnis eines Tests - als Hautfarbe erkannte Bereiche sind grün markiert.

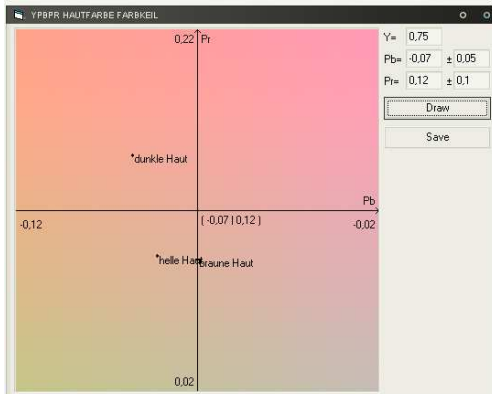


Abb. 5: Farbkeil mit stärkerer Streuung

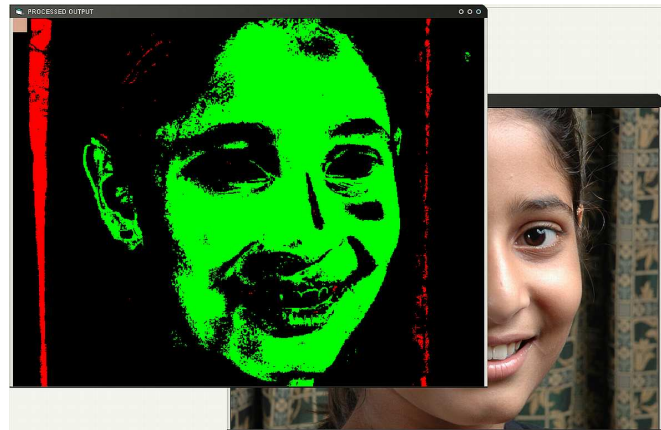


Abb. 4: Auffinden von Hautfarbe

Untersuchung von Kandidaten

Bei Fertigstellung des Programmes wird der Ablauf wahrscheinlich folgender sein:

1. Auffinden von hautfarbigen Flächen
2. Aussortieren anhand der Größe und der Seitenverhältnisse
3. Auffinden von Gesichtsmerkmalen anhand der Helligkeitsverteilung
4. Extraktion von Proportionen ("Biometriedaten")
5. Template Matching - Je nach Genauigkeit wird eine Identifizierung von bestimmten Gesichtern möglich.

Fazit

Die Ergebnisse der Untersuchung auf Hautfarbe mithilfe des Y'P_bP_r-Farbmodells sind ermutigend. Sie liefert bei gewöhnlichen Webcambildern sogar unter den ungünstigsten Bedingungen eine gute Grundlage (= gute Kandidaten) für die Entwicklung eines eigenen face-detection-Algorithmus. Ich bedauere, bis zur Abgabe dieser Arbeit mit der Entwicklung nicht so weit fortgeschritten zu sein.

Ich bemühe mich, meine Ideen und Erkenntnisse bis zum Regionalwettbewerb in Celle umzusetzen.

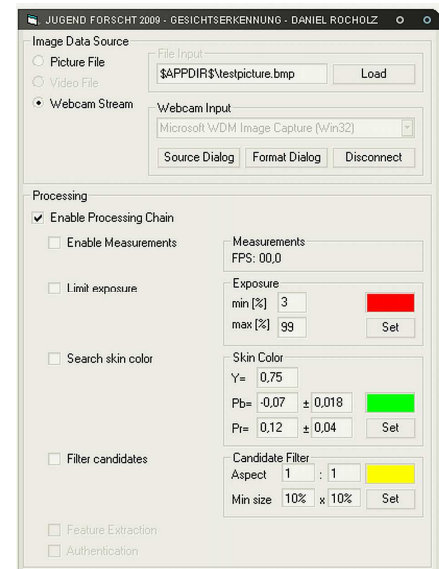


Abb. 6: Programmoberfläche

Die Zukunft von Gesichtserkennung

Durch die Verbreitung von 3D-Kameras und der rasanten Entwicklung der Rechenkapazitäten werden in der nahen Zukunft deutlich akkuratere Lösungen wachsen, die in allen Bereichen des täglichen Lebens Einzug halten. Denkbar ist personalisierte Werbung, Interaktion mit dem Computer über Gesten und Mimik, natürliche Kommunikation mit künstlicher Intelligenz sowie neue Überwachungstechnologien, die die Verfolgung von Personen geradezu automatisieren. Manche dieser Zukunftsvisionen lassen das Herz der Techno-Generation höher schlagen, andere wecken eher Erinnerungen an dystopische Geschichten à la 1984 von George Orwell. Meine persönliche Meinung ist, dass die junge (meine) Generation nur durch Aufklärung und Wissen über die Technologie und deren Anwendungsmöglichkeiten vor der erneuten Ausbildung einer Überwachungsgesellschaft, wie sie vor nunmehr 20 Jahren aufgelöst wurde, bewahrt werden kann.

"Technological progress has merely provided us with more efficient means for going backwards."

~Aldous Huxley

Danksagung

Wie bei jedem Projekt danke ich vielmals meinem Betreuer Thomas Biedermann für seine Unterstützung und die lehrreichen Gespräche. Mein Dank geht außerdem an die Entwicklergemeinde des Forums ActiveVB.de, welche mir beim Erlernen der Programmiersprache Visual Basic zu jeder Zeit eine gute Hilfe war und auch in einem konkreten Fall bei der Lösung eines Problems geholfen hat (Siehe Quellen).

Quellen

- [1] Head Tracking for Desktop VR Displays using the Wii Remote
<http://www.cs.cmu.edu/~johnny/projects/wii/>

- [2] Tilo Gockel, Pedram Azad, Rüdiger Dillmann: Gesichtskontrolle - Biometrie per Webcam selbstgemacht (aus cT 13/2008, S. 169, erschienen im Heise Verlag)

- [1] Lizenzfreie Bilder von StockExchange (www.sxc.hu)
[http://www.sxc.hu/photo/1099451,](http://www.sxc.hu/photo/1099451)
[http://www.sxc.hu/photo/950674,](http://www.sxc.hu/photo/950674)
<http://www.sxc.hu/photo/949466b>

- [3] Webseite der freien Grafikbibliothek ImageMagick
<http://www.imagemagick.org/script/index.php>

- [4] Charles Poynton - Frequently Asked Questions about Color, S. 14 (Abs. 28)
http://www.poynton.com/notes/colour_and_gamma/ColorFAQ.html

- MSDN-Artikel zur WIA-Schnittstelle
<http://msdn.microsoft.com/en-us/library/ms630368.aspx>

- ActiveVB Forum: capCreateCaptureWindow ohne WS_CHILD
<http://foren.activevb.de/cgi-bin/foren/view.pl?forum=4&id=e3de9da127f1706f7a2bf14e13b1667b&msg=372243&root=372243>

- IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, January 2002
Detecting Faces in Images: A Survey
Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, S. 34-58
Vorliegend als PDF der IEEE Computer society (www2.computer.org), zur Zeit online nicht erreichbar