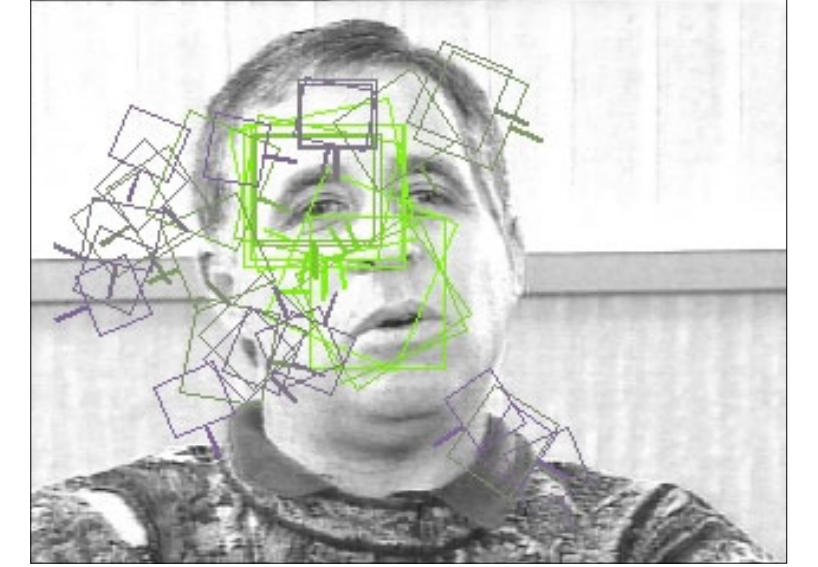


Entwurf, Implementierung und Evaluation eines Verfahrens zur Detektion von Gesichtsmerkmalen in Einzelbildern mittels künstlicher neuronaler Netze



Diplomarbeit, vorgelegt von Holger Ripp

Aufgabenstellung:

Um ein Gesicht in einem Bild automatisch zu verarbeiten (Identifikation, Verifikation usw.), ist es notwendig zu wissen, ob und wo sich ein Gesicht im Bild befindet. Dieser als Gesichtsdetektion bezeichnete Vorgang kann mit verschiedenen Ansätzen realisiert werden.

In dieser Arbeit sollte ein Verfahren entwickelt werden, welches Gesichtsmerkmale (Augen, Nase und Mund) in Einzelbildern detektieren kann. Das Gesicht sollte sich dabei in der Frontalansicht befinden und beliebig positioniert, skaliert und rotiert sein dürfen. Weiterhin sollte eine beliebige Anzahl desselben Gesichtsmerkmals (z.B. des rechten Auges) erlaubt und das Verfahren Beleuchtungsinvariant sein.

Um das gesamte Gesicht zu detektieren, müssen die einzelnen Gesichtsmerkmale anschließend kombiniert werden, was jedoch nicht Gegenstand dieser Arbeit war.

Problembeschreibung:

Es handelt sich um ein Mustererkennungsproblem mit sehr hoher Komplexität. Die automatische Verarbeitung von Gesichtern in Bildern wird durch die Variabilität der Abbildung des Gesichtes durch folgende Eigenschaften erschwert:

Eigenschaften des Gesichtes:

- verschiedene Gesichtsausdrücke
- Bart, Brille, Frisur
- Bekleidung (Hut, Mütze, Kapuze, Schal)
- Gesichtsfarbe
- Verdeckung

Eigenschaften des Gesichtes im Bild:

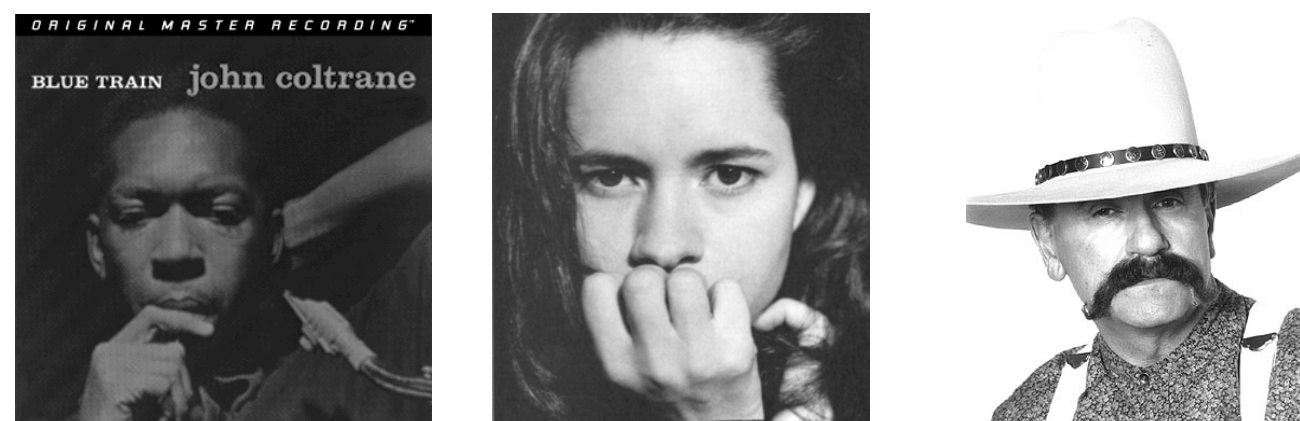
- Position im Bild
- Größe im Bild
- Rotation des Kopfes (3 Achsen)

Eigenschaften des Bildes:

- Komplexität des Hintergrundes
- verschiedene Beleuchtungssituationen
- mehrere Gesichter im Bild

Eigenschaften des Aufnahmesensors:

- Auflösung
- Farbfehler
- Rauschen



Beispiele für die Variabilität. (Quelle: CMU-MIT-Gesichtsdatenbank)

Verfahren zur automatischen Verarbeitung von Gesichtern in Bildern legen Einschränkungen fest, um die aus diesen Eigenschaften entstehende Variabilität und damit die Komplexität zu verringern. Nötig sind diese Einschränkungen, damit die Verarbeitung in, für den produktiven Einsatz, brauchbarer Zeit erfolgt bzw. die Erkennungsrate möglichst hoch ist.

Aktuelle Verfahren:

Aktuelle Verfahren zur Gesichtsdetektion lassen sich grob in vier Ansätzen unterteilen. Die Ansätze 1 und 2 beschreiben, welcher Teil des Gesichtes betrachtet wird und die Ansätze 3 und 4 welche Eigenschaften des jeweiligen Teils verwendet werden. Die meisten Verfahren nutzen somit zwei oder mehr Ansätze.

- 1.) Wissensbasierte Ansätze betrachten das ganze Gesicht.

2.) Merkmalsbasierte Ansätze betrachten einzelne Gesichtsteile oder -merkmale.

3.) Template Matching Ansätze nutzen geometrische Eigenschaften.

4.) Erscheinungsbasierte Ansätze nutzen statistische Eigenschaften.

Lösungsansatz:

Resultierend aus der Bewertung der aktuellen Verfahren, wurde sich für eine Kombination aus merkmalsbasiertem und erscheinungsbasiertem Ansatz entschieden. Verfahren nach dem merkmalsbasierten Ansatz sind robust gegenüber Verdeckung, Bekleidung und komplexen Hintergründen. Und Verfahren nach dem erscheinungsbasierten Ansatz besitzen die beste Generalisierungsfähigkeit, speziell wurde sich für künstliche neuronale Netze entschieden, da diese eine hohe Toleranz gegenüber variablen Eingaben aufweisen und somit eine hohe Generalisierungsfähigkeit erwarten lassen.

Es wird also nach Gesichtsmerkmalen (Augen, Nase und Mund) gesucht und diese werden von künstlichen neuronalen Netzen modelliert. Die Beleuchtungsinvarianz soll mit einer Kombination aus Beleuchtungskorrektur und Kantenfilter sichergestellt werden. Mit einer simplen Suchstrategie wird die geforderte Positions-, Skalierungs- und Rotationsinvarianz realisiert, indem ein Fenster mehrfach über das Bild bewegt und dabei skaliert und gedreht wird.

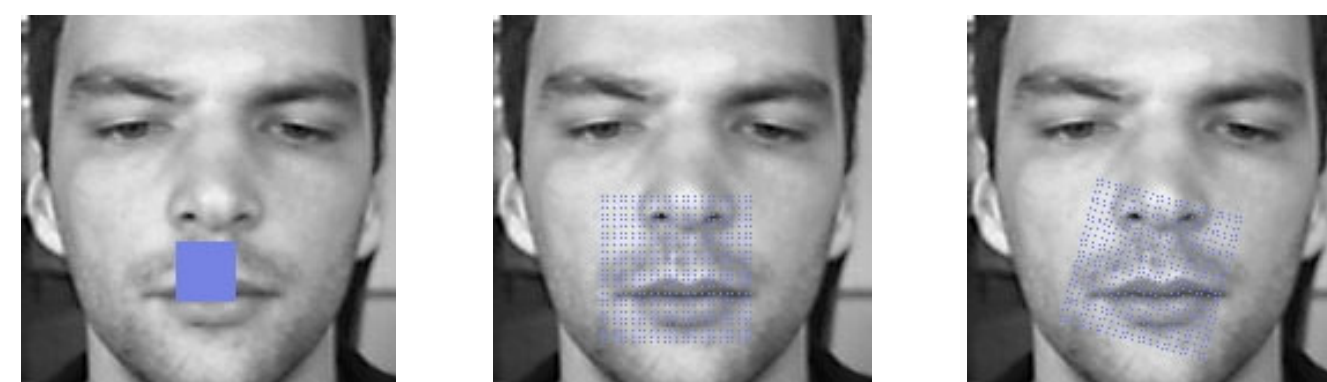
Folgende Einschränkungen werden festgelegt:

- Gesichter sind mindestens 60 x 60 Pixel groß
 - Gesichter befinden sich vollständig im Bild
 - Gesichter sind nicht stärker als 5° aus der Frontalansicht gedreht
 - keine Gesichter mit Brillen oder Bärten
- Außerdem soll die Verarbeitungszeit vernachlässigt werden.

Fenstertechnik für die Suche:

Das Fenster ist ein zweidimensionales quadratisches Feld von Punkten mit ungeraden Seitenlängen. Der Mittelpunkt hat die Koordinaten 0,0 (x,y) und alle anderen Punkte die Abstandskordinaten zum Mittelpunkt.

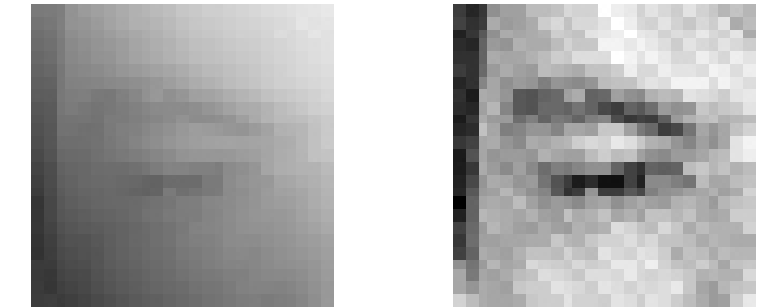
Soll ein Bildausschnitt extrahiert werden, wird ein Pixel im Bild festgelegt. Zu jedem Punkt des Fensters werden nun die Koordinaten dieses Pixels addiert. Die entstehenden Koordinaten entsprechen den Pixeln des Bildes, deren Farb- bzw. Grauwerte weiterverarbeitet werden.



Beispielhafte Darstellung der Fenstertechnik (mit den Fensterpunkten korrespondierende Bildpixel sind zur Verdeutlichung blau dargestellt)

Beleuchtungskorrektur:

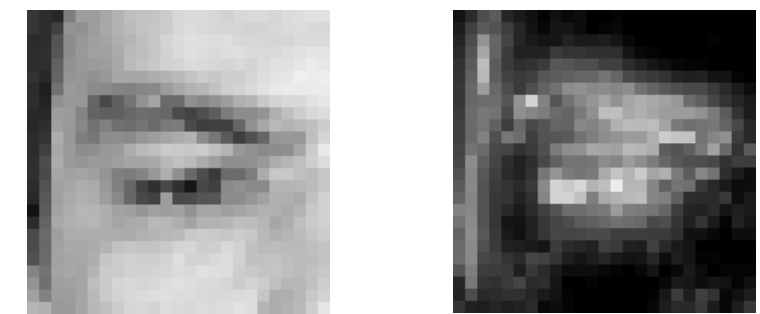
Es sollten mögliche lineare Helligkeitsverläufe im extrahierten Bildausschnitt erkannt und eliminiert werden. Dazu wurden in einem kreisförmigen Bereich innerhalb des Bildausschnitts in mehreren Winkeln Helligkeitsverläufe ermittelt. Wurde Linearität festgestellt und überschritt die Helligkeitsdifferenz einen bestimmten Schwellwert, wurde versucht in diesem Winkel den Helligkeitsverlauf zu eliminieren.



Beispiel der Beleuchtungskorrektur anhand eines Verlaufes von links unten (dunkel) nach rechts oben (hell) mit anschließender Kontrastverstärkung.

Kantenfilter:

Als Kantenfilter wurde der SUSAN-Kantenfilter angewendet. Er war jedoch so modifiziert, dass der Bildausschnitt nicht binarisiert, sondern das Filterergebnis auf Werte zwischen 0 und 255 abgebildet wurde.



Anwendung des modifizierten SUSAN-Kantenfilters.

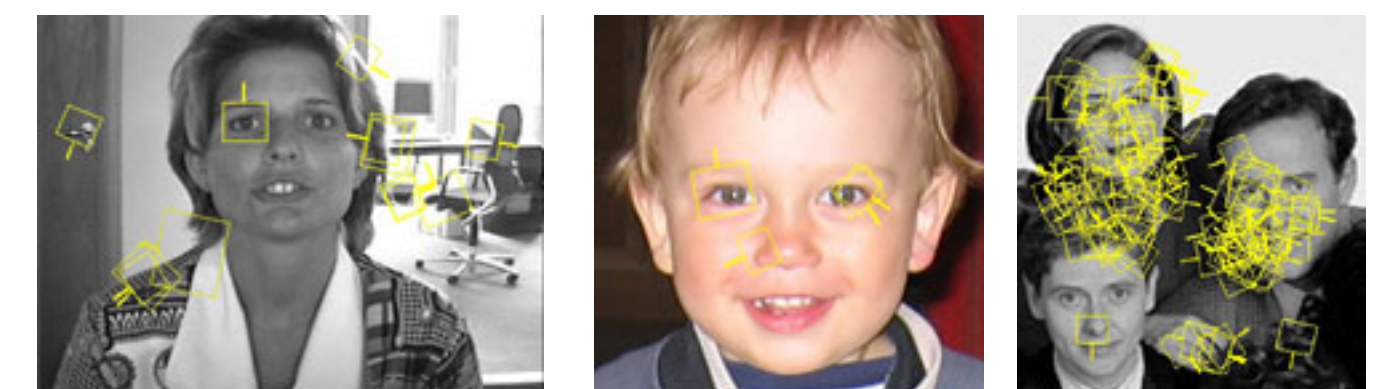
Künstliche neuronale Netze:

Es wurden einfache zweischichtige vorwärtsgerichtete Netze verwendet, d.h. sie besaßen eine Eingabe-, eine Hidden- und eine Ausgabeneuronenschicht. Diese wurden mit Hilfe des Backpropagationverfahrens trainiert. Das Training erfolgte mit Hilfe der BioID-Gesichtsdatenbank. Diese zeichnet sich aus durch:

- 1521 Grauwertbilder mit jeweils einem Gesicht
- 23 verschiedene Personen
- 20 Punktwerte zu jedem Bild in einer separaten Datei

Tests:

Stellvertretend für alle Gesichtsmerkmale erfolgte das Training der künstlichen neuronalen Netze, sowie die Suche ausschließlich mit bzw. nach dem rechten Auge. Die Klassifikations- und Generalisierungsfähigkeit wurde einerseits visuell beurteilt und andererseits wurden die Rückgabewerte der Netze um die rechten Augen ausgewertet, um eine Aussage über die Skalierungs- und Winkelgenauigkeit treffen zu können.



Visualisierte Suchergebnisse. Die abgebildeten Rechtecke entsprechen Bildausschnitten mit einem Reaktionswert des Netzes $\geq 0,99$. Das linke Bild stammt aus der BioID-Gesichtsdatenbank und diente der Beurteilung der Klassifikationsfähigkeit des Verfahrens. Das mittlere private und das rechte aus der CMU-MIT-Gesichtsdatenbank stammende Bild dienten der Beurteilung der Generalisierungsfähigkeit des Verfahrens.

Zusammenfassung und Ausblick:

Das entwickelte Verfahren ist aufgrund der verwendeten Suchstrategie Positions-, Skalierungs- und Rotationsinvariant. Die geforderte Beleuchtungsinvarianz konnte mit den gewählten Bildvorverarbeitungsalgorithmen realisiert werden. Durch zunehmend verbesserte Trainingsmusterdateien wurde die erwartete Generalisierungsfähigkeit der künstlichen neuronalen Netze erreicht.

Die Anzahl der falsch klassifizierten Bildausschnitte ist jedoch noch zu hoch und die für die spätere Kombination der Gesichtsmerkmale notwendige Skalierungs- und Winkelgenauigkeit ist nicht gewährleistet. Das Verfahren hat somit nur Experimentalcharakter und eignet sich nicht zur Detektion von Gesichtern. Weiterführende Arbeiten müssen die falsch klassifizierten Bildausschnitte eliminieren, die Skalierungs- und Winkelgenauigkeit erhöhen und die bisher vernachlässigte Verarbeitungszeit verringern.