Automatische Gesichtserkennung: Methoden und Anwendungen

Dominikus Baur
baurd@ifi.lmu.de

Universität München
Amalienstrasse 17, 80333 München, Deutschland

Zusammenfassung Automatische Gesichtserkennung bietet eine Vielzahl von vorstellbaren Anwendungen und eine schier unüberschaubare Flut von Publikationen. Die vorliegende Arbeit versucht eine Einführung in das Thema zu bieten: Zuerst werden die geschichtlichen Hintergründe der Technik aufgezeigt, daraufhin ausgewählte Verfahren aus den Bereichen musterbasierte und holistische Gesichtserkennung vorgestellt und Möglichkeiten, deren Effizienz zu vergleichen, beschrieben. Aus der Fülle von Einsatzfeldern werden abschließend Multimedia-Metadaten mit MPEG-7 und Identifikationsverfahren mit dem ePass als beispielhafte Vertreter präsentiert.

1 Einleitung

Die computergestützte Erkennung von Personen unter Zuhilfenahme biometrischer Daten ist ein Forschungsbereich, der in den letzten Jahren ein rapides Wachstum erfahren hat. Aufgrund der starken Zunahme des Flugverkehrs und der scheinbar immer akuter werdenden Terrorismusgefahr stehen viele Länder vor einer Überwachungsaufgabe, die sich ohne maschinelle Unterstützung nur noch schwer lösen lässt. Die Forcierung einer Ausweitung des öffentlichen Kameranetzes, das von Polizei und anderen Sicherheitsbehörden genutzt wird, wird meist mit der Verfolgung und Bekämpfung von kriminellen Vorgängen und Personen begründet. Auch die Ausstattung der Bürger mit leicht maschinenlesbaren Identifikationsdokumenten wie dem neuen ePass in Deutschland ist ein Schritt in diese Richtung und steht ebenfalls im Zeichen der Verbrechensbekämpfung.
Ohne biometrische Technologien, die aus der Datenflut, die sich daraus ergibt Nutzen ziehen können, entbehrt jedoch das gesamte Programm der allgegenwärtigen Überwachung einer Daseinsberechtigung. Weil Fingerabdruck- und Iris-Scanner weitaus aufwändiger und teurer als Kameras sind, kommt gerade der Gesichtserkennung dabei eine Schlüsselrolle zu.
Doch die Biometrie lässt sich nicht nur für die Bekämpfung der Kriminalität einsetzen. Ihre Verfahren ermöglichen es, ganz neue Wege in der Nutzung von visuellen und akustischen Daten zu gehen. Durch den neuen 'Multimedia Content Description Standard', kurz MPEG-7, der eine gemeinsame Grundlage für die Beschreibung von multimedialer Information bietet, lassen sich verschiedenste Anwendungen zur Suche in und nach Bildern und Tönen entwickeln. Ein weiterer integrierter Punkt ist die Beschreibung von menschlichen Gesichtern, die ebenfalls zur Grundlage für Suchanfragen, statische Auswertung o.ä. gemacht werden können.
Diese Vielfältigkeit der Applikationen macht gerade die Gesichtserkennung zu

einem, wenn nicht dem wichtigsten biometrischen Verfahren. Im Folgenden soll auf die (vor allem mathematischen) Grundlagen der Technik und die beiden oben geschilderten Anwendungsgebiete im Detail eingangen werden.

2 Verwandte Arbeiten

Gerade der Bereich der Gesichtserkennung, der offenbar keine bestmögliche Lösung sondern nur Annäherungen an diese unter bestimmten Bedingungen bietet, eröffnet ein weites Feld für Methoden und Verbesserungen.
Ebenso vielfältig wie das Forschungsfeld ist auch das Angebot an Arbeiten darüber. Der Grossteil besteht dabei aus Vorstellungen von neuen Verfahren beziehungsweise Verbesserungen bestehender. In solchen Arbeiten wird die Effizienz meist auch im Zusammenhang mit ähnlichen Wegen vorgestellt, was immer auch einen Vergleich mit eben diesen impliziert.
Arbeiten, die explizit nur analysieren ohne etwas Neues einzubringen konzentrieren sich meistens auf einen Effizienzvergleich verschiedener Methoden (z.B. Feature- versus Template-basierte Ansätze [1]) oder deren Schwächen (z.B. Einfluss von Beleuchtung und Körperhaltung auf verschiedene Verfahren [2]).
Umfassendere Vergleiche der Methoden und deren Hintergründe bieten [3] und [4] (siehe 3.6). Die Anwendungen der Gesichtserkennung werden meist nur in der Einleitung erwähnt oder als bekannt vorausgesetzt. Einen Überblick über die Anwendungen UND die dahinterstehenden Verfahren bietet [5], bezieht sich jedoch nur auf die Jahre 1975 bis 1995.

3 Automatische Gesichtserkennung

Die Fähigkeit, Gesichter zu erkennen und unterscheiden zu können, ist für den Menschen als soziales Lebewesen geradezu essentiell.
Schon im Kleinkindalter und auch bei Menschen mit schweren geistigen Einschränkungen ist dieses Vermögen immer noch gegeben. Neueste Forschungen haben sogar nachgewiesen, dass das Gehirn Gesichter anders wahrnimmt als andere visuelle Reize und die Zuordnung in einem eigenen Teil des Gehirns erfolgt [6].
Trotz dieser zentralen Bedeutung des Erkennens von Gesichtern für Menschen stecken technische Lösungen für dieses Problem noch in den Kinderschuhen. Der folgende Abschnitt behandelt das Thema der maschinellen Erkennung von Gesichtern, deren Geschichte, verschiedene Ansätze und deren Effizienz.

3.1 Geschichte der maschinellen Gesichtserkennung

Gesichtserkennung war ursprünglich ein Thema mit dem sich vor allem Psychologen und Neurophysiologen beschäftigten. Die Frage war, wie genau die Erkennung im menschlichen Gehirn abläuft und welche Mechanismen dahinterstecken. Die automatisierte Erkennung von Gesichtern lag aufgrund der archaischen Hardware noch in weiter Ferne.
Erst als Mitte der 70er Jahre die Technik fortgeschritten genug war, begannen

sich Computergrafiker mit dem Thema zu beschäftigen und allgemeine Bilderkennungsalgorithmen auf die Gesichtserkennung anzuwenden. Die Erkennung von geometrischen Objekten hatte bereits solche Fortschritte gemacht, dass die Erkennung von Gesichtern den Forschern nur als ein Teilbereich des Gesamtproblems erschien. Gegen Ende der 80er Jahre waren die vorgegebenen Verfahren soweit verfeinert worden, dass eine Erkennung auch unter wechselnden Bedingungen mit akzeptabler Trefferquote möglich wurde. Dank des Preissturzes der Computerhardware wurden nun auch kommerzielle Anwendungen von Gesichtserkennung vorstell- und bald auch erwerbbar ([3]).
Das folgende Engagement der Privatwirtschaft ließ das Feld während der 90er Jahre boomen. Eine Vielzahl von Publikationen erschien und regelmäßig wurden konkrete Systeme Vergleichstests unterzogen (z.B.: [7]), wenn auch mit ernüchternden Ergebnissen (siehe 3.6). Mit der staatlichen Einführung von Biometrie in Sicherheitsprozesse ist nun auch eine konstante Nachfrage vorhanden, die den Anreiz, hochqualitative Systeme zu entwickeln, stetig erhöht.

3.2 Ablauf und Voraussetzungen

Der Ablauf einer generischen Gesichtserkennung lässt sich prinzipiell in zwei Phasen aufteilen:

1. Die Erstellung einer Datenbank, die bekannte Gesichter und deren Identitätszuordnung enthält
2. Die Überprüfung, ob in einem gegebenen (Kamera-)Bild bekannte Gesichter erkennbar sind

Dabei wird Phase 1 normalerweise vor Inbetriebnahme des Systems durchgeführt. Sie lässt sich aber auch überspringen und in Phase 2 eingliedern. Dann wird, sobald ein unbekanntes Gesicht im Bild erscheint, eine (manuelle) Identifzierung vorgenommen und das nun zugeordnete Gesicht in die Datenbank übernommen. Weiterhin können enorme Steigerungen der Erkennungsraten dadurch erreicht werden, dass nicht nur ein Bild, sondern verschiedene Bilder mit Variationen in Kopfposition, Beleuchtung, Mimik etc. für eine Person gespeichert werden.
Phase 2 läuft im Detail so ab (siehe [8] und Abbildung 1):

1. Finden des Gesichts im Bild (*face detection*)
2. Extrahieren der Merkmale des Gesichts (*feature extraction*)
3. Ermittlung der Identität (*face recognition*)

Diese Ermittlung der Identität erfolgt meist mit einem statistischen Ansatz: Der Datenbankeintrag, der dem Testbild (in seiner Struktur) am ähnlichsten ist, wird als Identität angenommen. Natürlich treten dadurch auch falsche Zuordnungen (*false positives*) auf, oder das Verfahren erkennt die Person nicht, obwohl sie sich in der Datenbank befindet (*false negatives*).
Ein Punkt in diesen Mechanismen, der für sich genommen schon eine enorme Herausforderung darstellt, ist die maschinelle Lokalisierung eines Gesichts in einem gegebenen Bild (*face detection*).

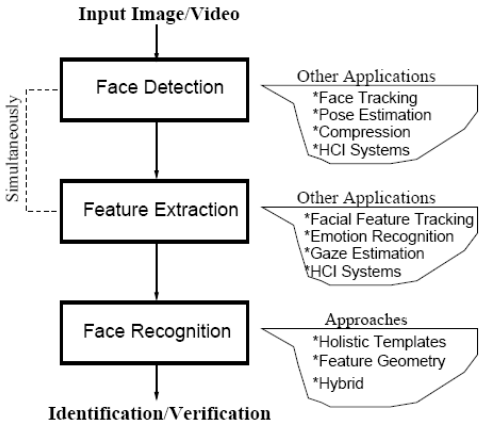

Abbildung 1. Ablauf der Gesichtserkennung aus [8]

Gesichtsdetektion Das Forschungsfeld der Gesichtsdetektion ist beinahe genauso umfangreich [9] wie das der Gesichtserkennung, auch wenn ihm erst in den letzten Jahren erhöhte Aufmerksamkeit gewidmet wurde.
Die Trefferquote eines Gesichtserkennungsalgorithmus lässt sich meist immens dadurch steigern, dass zur Erkennung ein standardisiertes (bezüglich Größe, Rotation, Farbe etc.) Bild eines menschlichen Gesichts ohne Hintergrund übergeben wird. Doch die Extraktion solcher Bilder aus einer gegebenen alltäglichen Vorlage (wie einem Fernsehprogramm oder einem Photo in einem Magazin) ist nicht eben trivial.
Die Verfahren teilen sich in die Kategorien musterbasierte Ansätze (*Feature-based approaches*) und bildbasierte Ansätze (*Image-based approaches*).
Bei musterbasierten Methoden wird von bestimmten allgemeinen Merkmalen des menschlichen Gesichts wie des Verhältnisses von Breite zu Höhe, der Position der Augen, des Abstands einzelner Gesichtsmerkmale zueinander, der Hell-Dunkel- oder Farb-Verteilung oder der typischen Bewegung im Vergleich zum restlichen Körper (in Videobildern) ausgegangen.
Die bildbasierten Verfahren gehen von der Gesichtsdetektion als Musterkennungsproblem aus und basieren auf der *Principal Component Analysis* (siehe 3.4), auf neuronalen Netzen oder statistischen Informationen [9].
Musterbasierte Ansätze zeigen zwar bei simplen Vorlagen wie Passphotos hohe Trefferquoten, versagen jedoch bei anspruchsvolleren (mehrere Gesichter in einem Bild, hohes Rauschen im Bildhintergrund) Problemen.
Bildbasierte Verfahren erzielen meist auch hier gute Ergebnisse und lassen sich durch die mathematische Verwandschaft mit den holistischen Gesichtserkennungsverfahren (siehe 3.4) oft mit diesen kombinieren und dadurch effizienter in ein System integrieren.

3.3 Musterbasierte Ansätze - *Feature-based Approaches*

Musterbasierte Ansätze, die zur Gesichtserkennung eingesetzt werden, nutzen menschliches Domänenwissen, um das Problem zu lösen.

Bestimmte Merkmale aller (oder zumindest nahezu aller) menschlichen Gesichter, die durch Forschungen und Erfahrenswerte in Medizin und Kunst gesammelt wurden, werden ausgenutzt, um die individuellen Aspekte eines bestimmten Gesichts zu messen und selbiges dadurch (für die Maschine) erkennbar zu machen. Brunelli und Poggio stellen in ihrem Aufsatz "Face Recognition: Features versus Templates" [1] zwei Verfahren dar, die beide in diese Kategorie fallen (siehe Abbildung 2).
Die Detektion des Gesichts wird dabei durch das Bild (*Template*) einer Augenpartie ermöglicht, das im Testbild gesucht wird. Dadurch kann die Position der Augen im Bild ermittelt werden. Danach setzen die Autoren auf einen geometrischen und einen weiterhin Template-basierten Ansatz.
Im geometrischen Verfahren werden durch Ausnutzung der Symmetrie und des festen Schemas eines Gesichts die einzelnen Merkmale (Augen, Nase, Mund, Gesichtsumriss, etc.) durch Kontrastunterschiede und statistische Schätzungen ermittelt, um schließlich ein geometrisches Modell des Ausgangsgesichts zu erstellen. Durch Anwendung desselben Verfahrens auf ein Testbild kann die Ähnlichkeit des Gesichts mit einem bereits gespeicherten Gesicht ermittelt und die Zuordnung damit vorgenommen werden.
Bei dem Template-basierten Ansatz wird bei den Bildern der Testpersonen in der Datenbank jeweils die Augenpartie, die Nase und der Mundbereich markiert. Der Algorithmus überprüft das Testbild auf Ähnlichkeiten zu gegebenen Mustern in den zu erwartenden Bereichen (also Nase in der Mitte des Gesichts, Mund darunter etc.) und ordnet das Testbild dem Bild in der Datenbank mit der höchsten Ähnlichkeit zu. Diese recht intuitiven Ansätze liefern leider nur

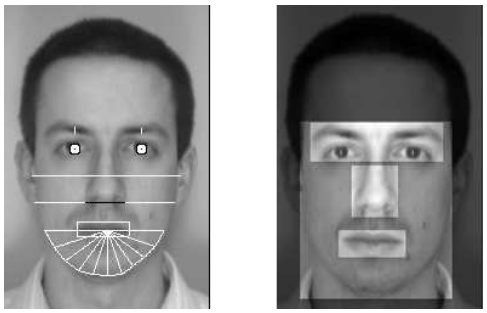

Abbildung 2. Geometrischer und Template-basierter Ansatz aus [1]

unter günstigen Bedingungen akzeptable Ergebnisse.
Ein Verfahren, das höhere Trefferquoten aufweist, ist das sogenannte *Elastic Bunch Graph Matching*, das von Wiskott et al. in ihrem Aufsatz "Face Recognition by Elastic Bunch Graph Matching" [10] vorgestellt wurde.
Es basiert darauf, jedes Gesicht in ein festgelegtes Raster aus Knoten und Kanten zu legen. Die Knoten, auch *Fiducial Points* genannt, liegen dabei auf bestimmten Punkten wie der Nasenspitze oder der Mitte der Pupillen. Bei jedem Knoten wird nun mithilfe sogenannter *Jets*, einem Bündel von bestimmten Wavelets, das den Knoten umgebende (Farb-)Muster gespeichert. Diese Jets ma-

chen das Verfahren unempfindlich gegen Schwankungen in Bildhelligkeit und -kontrast.
Jeder der möglichen Jets wird einem Fiducial Point zugeordnet und in einer Datenstruktur gespeichert, die die Entwickler *Face Bunch Graph* (FBG) nannten (siehe Abbildung 3).

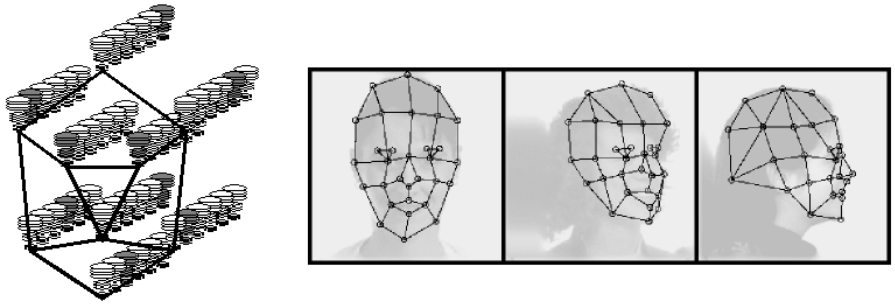

Abbildung 3. *Face Bunch Graph* und gerasterte Gesichter aus [10]

Der FBG stellt eine generelle Repräsentation von Gesichtern dar, weil die unterschiedlichen gespeicherten Jets aus einzelnen Testgesichtern beliebig kombiniert werden können.
Die Erkennung eines Gesichts läuft in drei Phasen ab:

1. Zuerst werden durch Bildung des Durchschnitts der Jets an einem Knoten des FBG und Vergleich dessen mit dem Testbild die Fiducial Points ermittelt (das macht das Verfahren robust gegen Veränderung der Pose der Testperson).
2. Das Bild wird dem Raster entsprechend normalisiert. Daraufhin werden die Jets der ermittelten Fiducial Points berechnet und gespeichert.
3. Abschließend werden die erzeugten Jets mit den Modellgraphen der Datenbank verglichen, um das Gesicht einer Person zuzuordnen.

Bunch Graph Matching ist der ausgereifteste Vertreter der Familie der musterbasierten Ansätze und zeigt gute Erkennungsraten. Ähnliche Effizienz wird sonst nur noch von gänzlich andersartigen Verfahren erbracht.

3.4 Holistische Ansätze - *Holistic Approaches*

Weil die Erkennung von *Gesichtern* in Bildern nur eine Untermenge der Erkennung von *Objekten* in Bildern darstellt, lassen sich Algorithmen, die für eben dieses Problem entwickelt wurden, auch auf den Bereich der Gesichtserkennung anwenden.
Diese holistischen Ansätze "wissen" (im Gegensatz zu musterbasierten Methoden) nichts über das Gebiet auf das sie angewandt werden. Sie lassen sich zwar auf die Eigenheiten von menschlichen Gesichtern optimieren, könnten aber in prinzipiell der gleichen Form auch Autos oder Bäume erkennen.

Die beiden vorzustellenden Algorithmen basieren auf der Interpretation von Bildern als Vektoren und deren Projektion in einen Unterraum:
Die Eingabebilder lassen sich als Vektoren von Grauwerten interpretieren. Ein Bild mit einer Auflösung von $x * y$ Pixeln würde damit eine Matrix A der Größe $x * y$ bilden.
Diese Matrix lässt sich nun als Vektor v darstellen, mit

$$v = \begin{pmatrix} A_{11} \\ A_{12} \\ \dots \\ A_{1x} \\ A_{21} \\ \dots \\ A_{yx} \end{pmatrix} \tag{1}$$

d.h. einer Verkettung aller Zeilen. Dieser Vektor hat die Dimension $x * y$.
Ein $x * y$-dimensionaler Vektorraum R enthält alle möglichen Bilder dieser Größe.
Die Ähnlichkeit zweier Bilder lässt sich (naiv) über die Euklidische Distanz zwischen ihren zugehörigen Vektoren ermitteln - je geringer die Distanz, desto ähnlicher die Bilder.
Aufgrund der hohen Anzahl der Dimensionen in R können dadurch aber auch völlig unterschiedliche Bilder als ähnlich gedeutet werden, wenn Symmetrieeffekte etc. zum Tragen kommen. Auch laufen Algorithmen, die in solch extrem hochdimensionalen Vektorräumen arbeiten, meist sehr ineffizient und langsam. Außerdem sind sich alle menschlichen Gesichter prinzipiell sehr ähnlich, so dass die Bildern von diesen auch im Vektorraum R sehr nahe beeinanderliegen und damit die Unterscheidung erschweren.
Die folgenden Algorithmen reduzieren durch Projektion in niedrigdimensionale Unterräume den Berechnungsaufwand und arbeiten die Unterschiede zwischen verschiedenen Gesichtern deutlicher heraus, um die Unterscheidung zu vereinfachen.

Diskriminanzanalyse - *Linear Discriminant Analysis* Die Diskriminanzanalyse ist ein statistisches Verfahren, dass den Zusammenhang zwischen verschiedenen Variablen darstellen kann und damit eine Klassifizierung von Testbildern ermöglicht. Es läuft folgendermaßen ab:
Für C Testpersonen werden jeweils eines oder mehrere Bilder aufgenommen und in der Datenbank gespeichert. Alle Bilder, die die gleiche Person zeigen, gehören zu einer *Klasse*. Für ein zu untersuchendes Testbild muß nur noch festgestellt werden, zu welcher Klasse es gehört, um es einer Person zuzuordnen.
Um die Unterscheidung zu erleichtern, werden alle Bilder in einen speziellen Unterraum, den sogenannten Fisher-Unterraum F, des Vektorraums R projiziert.
Die Basisvektoren des Unterraums F sind die Eigenvektoren V zu den Eigenwerten Λ der folgenden Gleichung:

$$S_B V = \Lambda S_W V \tag{2}$$

Dabei ist
S_B die Matrix der Streuung zwischen den Klassen (erklärte Streuung) und
S_W die Matrix der Streuung innerhalb der Klassen (nichterklärte Streuung).
Das Diskriminanzkriterium

$$\Gamma = \frac{erklärte\ Streuung}{nichterklärte\ Streuung} \tag{3}$$

gibt dabei die Qualität der Unterteilung in Klassen an; je höher der Wert von Γ, desto klarer die Trennung der einzelnen Klassen. Nutzt man die Eigenvektoren als Basis, wird das Diskriminanzkriterium maximiert.
Die beiden entscheidenden Streuungsmatrizen werden folgendermaßen berechnet:
Für die Bilder X jeder Klasse i ergibt sich eine spezifische Streuungsmatrix

$$S_i = \sum_{x \in X_i} (x - m_i)(x - m_i)^T \tag{4}$$

wobei m_i jeweils das Durchschnittsbild der Klasse i ist.
Die nichterklärte Streuungsmatrix ist nun einfach die Summe der spezifischen Streuungsmatrizen für jede Klasse und damit

$$S_W = \sum_{i=1}^{C} S_i \tag{5}$$

Die Streuung zwischen den Bildern ergibt sich aus dieser Formel

$$S_B = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T \tag{6}$$

wobei

n_i die Anzahl der Bilder in Klasse i
m_i das Durchschnittsbild der Klasse
m der Durchschnitt aller Bilder ist.

Nachdem sie berechnet wurden, werden die Eigenvektoren nach der Größe der zugehörigen Eigenwerte geordnet. Die ersten $C - 1$ Eigenvektoren bilden den Fisher-Unterraum F, in den alle Bilder Φ_i nach der Formel

$$\overline{\Phi_i} = V^T \Phi_i \tag{7}$$

projiziert werden ([11] [3]).
Je größer ein Eigenwert ist, desto relevanter ist die Koordinate eines Bildvektors im entsprechenden Eigen- bzw. Basisvektor. Die ersten Eigenvektoren tragen meistens schon einen Großteil der "Besonderheit" eines Bildes. Daher werden bei der Hauptkomponentenanalyse (siehe nächster Abschnitt) auch nur diese Eigenvektoren zum Aufspannen des Unterraums genutzt.
Das Verfahren optimiert das Koordinatensystem des Vektorraums so, dass die Mitglieder unterschiedlicher Klassen möglichst weit voneinander entfernt liegen. Testbilder werden in den Fisher-Unterraum projiziert und können dann mit einem Abstandsalgorithmus (z.B.: Mahalanobis-Distanz - ein Distanzmaß für höherdimensionale Vektoren) einer Klasse zugeordnet werden.

Hauptkomponentenanalyse - *Principal Component Analysis* Die Hauptkomponentenanalyse (*Principal Component Analysis - PCA*), auch Karhunen-Loève-Transformation, versucht ähnlich wie die Diskriminanzanalyse die Besonderheiten eines Gesichts im Vergleich mit anderen Gesichtern durch Projektion in einen Unterraum, diesmal *Face Space* genannt, herauszuarbeiten. Alle Bilder in der Datenbank werden dabei jedoch gleichwertig behandelt und nicht in Klassen eingeteilt.
Das Verfahren wurde erstmals von Sirovich & Kirby 1987 ([12]) auf Gesichtserkennung angewendet und von Turk & Pentland in "Eigenfaces for Recognition" ([13]) entscheidend verfeinert.
Grob skizziert umfasst das Verfahren die folgenden Schritte:

1. Wiederum wird von den Bildern als Vektoren $\Gamma_1, \Gamma_2, \ldots, \Gamma_M$ der Dimension N (=Höhe $*$ Breite in Pixeln) ausgegangen. Diese müssen zuerst normalisiert werden, indem zuerst ein Durchschnittsgesicht Ψ nach dem Schema $\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n$ berechnet und dieses von allen Bildern abgezogen wird. Daraus ergeben sich die normalisierten Bilder $\Phi_1, \Phi_2, \ldots, \Phi_M$.
2. Es wird eine die $N * M$-Matrix X aus den normalisierten Bildern erzeugt:

$$X = [\, \Phi_1 \; \Phi_2 \; \ldots \; \Phi_M \,] \tag{8}$$

3. Wir berechnen daraus die Kovarianzmatrix

$$C = \overline{XX}^T \tag{9}$$

4. und bilden nach der folgenden Formel die Eigenvektoren V und -werte Λ:

$$CV = \Lambda V \tag{10}$$

5. Die Eigenvektoren V werden wieder nach der Größe der Eigenwerte Λ sortiert.
6. Wie gezeigt werden kann, sind nur die ersten $M-1$ Eigenvektoren ungleich 0 (für $M \ll N$), daher spannen diese den Unterraum (*Face Space*) auf.
7. Jeder Bildvektor Φ_i wird mit der Formel 7 in den Unterraum projiziert. Diese projizierten Bilder, die wie geisterhafte Gesichter aussehen, nennen die Autoren *Eigenfaces*, was dem Verfahren seinen Namen gab.

Für eine ausführlichere Beschreibung siehe [11], [3] und [13].
Um ein neues Bild zu erkennen wird es nach 7 in den *Face Space* projiziert. Damit nur Gesichter (und keine beliebigen Objekte fälschlicherweise) als Gesichter erkannt werden, berechnen Turk & Pentland ([13]) noch den Abstand des (normalisierten (siehe Schritt 1)) Testbildes Φ_f vom Face Space:

$$\varepsilon^2 = \|\Phi - \Phi_f\|^2 \tag{11}$$

Übersteigt dieser Abstand einen bestimmten Wert ist klar, dass es sich um kein Gesicht handeln kann und das Verfahren wird abgebrochen. Sonst wird nach einem beliebigen Abstandsverfahren das ähnlichste Bild berechnet.
Die Literatur bietet eine Vielzahl von Optimierungen des Algorithmus (u.a. Einbeziehung der Nachbarschaft eines Pixels: [14], Optimierung für bewegte Bilder: [15]).

3.5 Weitere Ansätze

Nicht alle Verfahren lassen sich in die beiden Schemata musterbasiert und holistisch pressen - oft werden auch Hybridsysteme benutzt, die sich die Vorteile beider Methoden zunutze machen, oder es werden völlig andere Wege gegangen. Ein vielversprechender neuer Ansatz, der erst in den letzten Jahren seinen entscheidenden Durchbruch erlebt hat, ist die Betrachtung von Gesichtern als Texturen von 3D-Körpern.
Chang et al. nutzen in ihrer Arbeit ([16]) Höheninformationen in Gesichtern, um ein weiteres eindeutiges Identifizierungsmerkmal zu haben. Sie bilden ein Hybridverfahren aus PCA-basierter 2D-Bildanalyse und zusätzlichem Vergleich der Höheninformationen und können dadurch die Erkennungsgenauigkeit ihres Systems von knapp unter 95% (mit einem reinen 3D-Verfahren) auf 98,8% steigern.
Bronstein et al. stellen in ihrer Arbeit "Expression-Invariant 3D Face Recognition" ([17]) ein ähnliches Verfahren vor. Sie gehen jedoch zur Erkennung nicht von aufgenommenen 3D-Daten aus, sondern präsentieren eine Methode, mit der sich die sogenannte "kanonische Form" des Gesichts berechnen lässt und nutzen selbige. Das gleiche Gesicht wird unabhängig von Gesichtsausdruck, Kopfhaltung oder Lichtverhältnissen immer derselben kanonischen Form zugeordnet und ermöglicht damit eine Erkennung unabhängig von der Situation. Für das Testbild wird das gleiche Verfahren genutzt, um auch eine kanonische Form zu erzeugen und diese dann mit den gespeicherten Formen der Datenbank zu vergleichen. Das System weist eine beeindruckende Erkennungsquote auf und schafft es sogar zwei der Autoren, die eineiige Zwillinge sind, auseinander zu halten.

3.6 Effizienz der aufgezeigten Verfahren

Bei der Vielzahl von Verfahren ist es natürlich wichtig, regelmäßig die Effizienz verschiedener Ansätze miteinander zu vergleichen, um vielversprechende Kandidaten auszumachen und weiterentwickeln zu können.
Zu diesem Zweck wurden standardisierte Datenbanken angelegt, die Testumgebungen für Algorithmen bieten und damit eine Vergleichbarkeit erst ermöglichen.
Wichtige Datenbanken sind die *FERET*-Datenbank des amerikanischen Verteidigungsministeriums (`http://www.itl.nist.gov/iad/humanid/feret/feret_master.html`) oder die *PIE*-Datenbank der Carnegie Mellon University (`http://www.ri.cmu.edu/projects/project_418.html`).
Unglücklicherweise ist dadurch noch nicht die Vergleichbarkeit gesichert, weil 1. viele Forscher weiterhin eigene Daten nutzen (manche wahrscheinlich speziell auf ihren Algorithmus zugeschnittene) und 2. eine Vielzahl (`http://www.face-rec.org` listet 22) von konkurrierenden Datenbanken existieren.
Doch selbst wenn sich alle Forschergruppen auf eine Standard-Datenbank einigen könnten, kommt noch dazu, dass, wie Phillips & Newton, die 2002 eine Metaanalyse über 24 Arbeiten durchgeführt haben ([18]), bemerken, die Datenbanken aus so simplen Beispielen bestehen, dass eine Vergleichbarkeit der Algorithmen aufgrund der saturierten Erfolgsrate von über 90% nicht mehr möglich

ist. Ein einheitliches Testverfahren für Algorithmen dieser Art mit "anspruchsvollen" Problemen ist nötiger denn je.
Im kommerziellen Bereich existiert eine solch homogene Testumgebung mit dem "Face Recognition Vendor Test", der zuletzt 2002 durchgeführt wurde ([7]). Doch auch dieser konnte die teils euphorischen Botschaften der Firmen und Forscher nicht bestätigen: Das beste getestete System brachte eine Erkennungsrate von 90% (in Innenräumen; bei Außenaufnahmen fiel die Erkennungsgenauigkeit auf 50% ab) und die Erkennung nahm stetig ab, sobald die Liste, die überwacht werden sollte, wuchs.
Die Effizienz von Gesichtserkennungsverfahren lässt also noch sehr zu wünschen übrig und ob kommerzielle oder staatliche Anwendungen, die sich auf hohe Erkennungsraten verlassen, in naher Zukunft sinnvoll sind, ist fragwürdig (siehe 4.2).

4 Anwendungen

Der Reiz der Entwicklung von Gesichtserkennungssystemen liegt in der Vielzahl der vorstellbaren Anwendungen. Zhao et al. ([8]) nennen vier Gebiete, die davon profitieren könnten:

- *Entertainment* (Computerspiele, Mensch-Maschine-Interaktion)
- *Smart Cards* (Führerscheine, Pässe (siehe unten))
- *Information Security* (Passwort- und PIN-Ersatz)
- *Law Enforcement and Surveillance* (Automatische Kameraüberwachung, Verfolgung von Verdächtigen in Kameranetzen)

Gerade die Strafverfolgung könnte sich als sehr ergiebig dafür erweisen - Chellappa et al. ([5]) gehen hier auf spezielle Bereiche ein und nennen nicht nur Überwachung, sondern auch Manipulation von Gesichtsdaten, um beispielsweise maskierte Personen zu identifizieren oder den Alterungsprozess zu simulieren. Realweltliche Anwendungen sind trotz dieses breiten Spektrums noch rar gesät - im Folgenden werden der Bereich der Multimedia-Metadaten (anhand MPEG-7) und Identifikation (anhand des ePass) beschrieben.

4.1 Multimedia-Metadaten

MPEG-7 ist ein Multimedia-Metadaten Standard, der eine gemeinsame Grundlage für die Annotation und Analyse von Multimedia-Informationen bieten soll. MPEG-7 bietet folgende Hilfsmittel zur Strukturierung:

- *Descriptors* - konkrete Beschreibungen eines multimedialen Attributs
- *Description Schemes* - die Vorgaben zur Erstellung eines bestimmten Descriptors
- *Description Definition Language (DDL)* - eine Sprache um Description Schemes zu erstellen

Die Descriptors können in XML oder einem binären Format abgelegt werden, um die Verarbeitung zu beschleunigen. Die Transformation zwischen beiden Formaten ist verlustfrei.

Zusammen mit diesen Vorgaben hat die Moving Pictures Expert Group noch einige Beispieldescriptoren veröffentlicht, die sich auf die Kategorien Audio und Video aufteilen. Durch die Flexibilität des MPEG-7-Systems lassen sich beliebige weitere Descriptoren in DDL erstellen oder auch die vorgegebenen für eine bestimmte Anwendung einschränken.
Standardisierte Audio-Descriptoren existieren u.a. für Sprache, Melodien und Musikstücke.
Video-Descriptoren gibt es für einzelne Bilder (Farbe, Formen, Position, Texturen), Segmente (Farbe, Kamerabewegung, Bewegungsaktivität) und bewegliche Bereiche (Farbe, Flugbahn).
Für Stand- und Videobilder wurde auch ein Descriptor zur Erkennung von Gesichtern, der sogenannte *Face Descriptor*, vorgegeben ([19]).

Der Face Descriptor Zur Beschreibung eines Gesichts bietet MPEG-7 48 Vektoren.
Die Erkennung und Speicherung erfolgt über eine Variante von PCA (siehe 3.4), der Ablauf ist dabei folgendermaßen:

1. Das Gesicht wird normalisiert und in ein Raster von 56 Zeilen mit je 46 Luminanzwerten gelegt. Dabei sollen sich die Augen in der 24sten Zeile und der 16ten und 31sten Spalte befinden.
2. Nun wird der eindimensionale Gesichtsvektor Φ erstellt, indem man spaltenweise, von links-oben beginnend die Luminanzwerte kombiniert.
3. Φ wird in den durch den Standard vorgegebenen Unterraum projiziert. Nach der Normalisierung bleibt ein 48-dimensionaler *feature vector* aus 1-Byte-Werten W, der gespeichert werden kann.

Die Erkennung eines Gesichts wird über die Berechnung des euklidischen Abstands realisiert ([19]).

Mögliche und vorhandene Anwendungen Die Möglichkeiten, die sich durch das MPEG-7 Framework bieten, sind beeindruckend: Sobald ganze Filmdatenbanken eingespeist und mit Metainformationen versorgt wurden, lassen sich Szenen nach ganz anderen als den gewohnten Kriterien finden. Eine Suche nach Gesichtern, Farbgebungen, Hintergrundgeräuschen oder gesprochenen Worten ist möglich.
Dennoch sind die konkreten Implementierungen, die auch Gebrauch vom Face Descriptor machen, noch rar gesät (vermutlich wegen der Komplexität der Technik und der Neuheit des Standards).
Jae-Ho Lee stellt in seinem Artikel [20] ein Video-Indexierungssystem vor, das ein Video nach Gesichtern durchsucht, automatisch Metainformationen erstellt und die Suche und Editierung derselben ermöglicht.
Das Fraunhofer-Insititut hat mit dem *iFinder* ein Tool entwickelt, das die ganze Bandbreite von MPEG-7 Deskriptoren nutzt und das automatische Hinzufügen von Metainformationen zu einer multimedialen Quelle ermöglicht (`http://www.imk.fraunhofer.de/de/iFinder`). In dem System sind sowohl Mechanismen zur Erkennung als auch zur Speicherung und Suche vorhanden.

Dies sind nur die ersten Vorboten kommender MPEG-7 Anwendungen. Ob sich der Standard jedoch etablieren kann und das volle Potential ausgeschöpft wird, kann nur die Zukunft zeigen.

4.2 Identifikation

Der Vorteil der Gesichtserkennung gegenüber anderen biometrischen Verfahren ist die geringe Aufdringlichkeit. Das Gesicht als prägnantestes Körpermerkmal eines Menschen ist normalerweise immer (wenigstens teilweise) sichtbar und kann durch Kameras aufgezeichnet werden. Andere Verfahren wie Fingerabdruck- oder Iriserkennung erfordern immer die Kooperation der Testperson, sind deutlich aufwändiger zu realisieren und erfordern normalerweise spezielle Hardware.
Auch lässt sich ein Gesichtserkennungssystem mit hoher Erkennungsgenauigkeit nur sehr schwer überlisten, während beispielsweise Fingerabdrücke relativ leicht mit einfachen Mitteln wie Klebeband nachzumachen sind.
Daher ist die Gesichtserkennung besonders für staatliche Organe wie Polizei und Grenzschutz eine langerwartete technische Unterstützung. Um jedoch eine Zuordnung von Gesichtern zu Identitäten zu ermöglichen ist zuerst eine Datenbasis nötig. Daher wurde mit der EU-Verordnung Nr. 2252/2004 eine für alle europäischen Länder gültige Vorgabe geschaffen, wie in Zukunft biometrische Daten in Reisedokumente aufgenommen werden sollen ([21]). Als eines der ersten Länder der Europäischen Union hat Deutschland diese Verordnung mit dem sogenannten *ePass* umgesetzt, der seit dem 1. November 2005 der einzig erhältliche Reisepass für deutsche Bürgerinnen und Bürger ist.

Datenumfang und -speicherung Der ePass wird im Augenblick nur für die Speicherung des Namens, des Geburtsdatums, des Geschlechts und eines Gesichtsfotos des Ausweisinhabers genutzt, die Ausweise sollen jedoch in Zukunft auch die Fingerabdrücke (ab März 2007) enthalten.
Die Möglichkeit zur Speicherung und maschinellen Auswertung der Daten wird durch einen *RF*-Chip, der auf dem ePass vorhanden ist, gegeben. Die Daten werden mithilfe des ECDSA-Crypto-Algorithmus verschlüsselt und bieten somit eine zusätzliche elektronische Fälschungssicherheit, da gültige Signaturen nur von dem Ausstellerland selbst erzeugt werden können.
Um ein Auslesen der per Funk übertragenen Daten durch Unberechtigte zu verhindern, benötigen Lesegeräte einen Schlüssel, der im maschinenlesbaren Bereich des Ausweises vorhanden ist und nur durch optischen Kontakt erhalten werden kann. Nach der Authentifizierung wird die Kommunikation zwischen Lesegerät und RF-Chip mit 112-Bit-Triple-DES verschlüsselt, um diese abhörsicher zu machen ([22]).

Datennutzung Die neuen Daten, die im ePass gehalten werden, sollen wie bei bisherigen Pässen zur Identifikation von Personen an Grenzen oder Flughäfen genutzt werden. Dazu wurden verschiedene Lesegeräte entwickelt, die nach obenstehendem Protokoll den Ausweis auslesen können und die erfassten Daten

dem jeweiligen Beamten zugänglich machen.
Das Bundesamt für Sicherheit in der Informationstechnik hat dazu eine Softwareapplikation (das sogenannte *Golden Reader Tool*) entwickelt, das die grundlegende Kommunikation zwischen verschiedenen Lesegeräten und Ausweisen ermöglicht und die gelesenen Informationen an einem Terminal anzeigen kann ([23]).

Datenschutz und Kritik Die gesammelte elektronische Speicherung von so sensiblen Daten wie Namen, Geburtsdatum, Foto und Fingerabdrücken auf nur einem Dokument sowie die Übertragung dieser Daten per Funk bieten genug Anlass zur Kritik von Datenschützern. Besonders die Hackervereinigung *Chaos Computer Club* verurteilt die neue Technologie scharf ([24]). Die Kritikpunkte im Einzelnen sind:

- Die Funkchips ermöglichten unbemerkte Überwachung und Verfolgung einzelner Personen.
- Die genaue Nutzung und Speicherung der ausgelesenen Daten an Grenzen ist unklar.
- Die Erhöhung der Fälschungssicherheit sei nur ein Scheinargument, da diese auch ohne Speicherung von personenbezogenen Daten realisiert werden könne.
- Der hohe finanzielle Aufwand zur Erstellung der neuen Ausweise werde größtenteils vom Bund getragen.
- Der Hauptkritikpunkt ist jedoch die Unausgereiftheit der Technik. Eine Studie des BSI ([25]), die die Anwendbarkeit von biometrischen Verfahren im Alltag zeigen sollte, wies genau das Gegenteil nach: Eine Abweisungsrate von 3 bis 23 Prozent würde ein enormes Chaos an realen Flughäfen und Grenzen bedeuten und die gesonderte Untersuchung der zurückgewiesenen Personen einen untragbaren personellen Mehraufwand mit sich bringen.

Aufgrund dieser Probleme bleibt der ePass ein Problemkind und es ist fraglich, wie lange es noch dauert, bis Personenkontrollen an Grenzen tatsächlich mit biometrischer Unterstützung stattfinden werden und ob die scheinbar überstürzte Einführung Sinn gemacht hat.

5 Zusammenfassung

In der vorliegenden Arbeit wurde eine kurze Einführung in das Thema der automatischen Gesichtserkennung gegeben. Verschiedene Verfahren, die sich in die Bereiche musterbasierte und holistische Ansätze aufteilen, sowie deren Geschichte und Effizienz wurden vorgestellt. Außerdem wurde auf die Anwendungsgebiete solcher Technologien und im Detail auf Multimedia-Metadaten mit MPEG-7 und biometrische Informationen in Ausweisdokumenten mit dem ePass eingegangen.

Literatur

1. Brunelli, R., Poggio, T.: Face Recognition: Features versus Templates. Pattern Analysis and Machine Intelligence, IEEE Transactions on (1993) 1042 – 1052

2. Gross, R., Shi, J., Cohn, J.: Quo Vadis Face Recognition? In: Third Workshop on Empirical Evaluation Methods in Computer Vision. (2001)
3. Tseng, S.: Comparison of holistic and feature based approaches to face recognition. Master thesis (2003)
4. Hofmann, M.: Grundsätzliche Untersuchung von Bildverarbeitungsalgorithmen zur Gesichtererkennung. Diploma thesis (2002)
5. Chellappa, R., Sirohey, S., Wilson, C., Barnes, C.: Human and Machine Recognition of Faces: A Survey. In: Proceedings of the IEEE. Volume Volume 83. (1995) 705 – 741
6. Rotshtein, P., Henson, R., Treves, A., Driver, J., Dolan, R.: Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nature Neuroscience **8** (2005) 107 – 113
7. Phillips, P.J., Grother, P., Micheals, R., Blackburn, D.M., Tabassi, E., Bone, M.: Face Recognition Vendor Test 2002. In: AMFG '03: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Washington, DC, USA, IEEE Computer Society (2003) 44
8. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. **35** (2003) 399 – 458
9. Hjelmas, B., Low, B.: Face detection: A survey. Computer Vision and Image Understanding (2001) 236 – 274
10. Wiskott, L., Fellous, J.M., Krueger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. In: Intelligent Biometric Techniques in Fingerprint and Face Recognition. (1999) 355 – 396
11. Yambor, W.: Analysis of PCA-based and Fisher Discriminant-based Image Recognition Algorithms. M.S. Thesis (2000)
12. Sirovich, L., Kirby, M.: Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A: Optics, Image Science, and Vision **Volume 4, Issue 3** (1987) 519 – 524
13. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **Volume 3, Number 1** (1991)
14. Yang, M., Ahuja, N., Kriegman, D.: Face recognition using kernel eigenfaces (2000)
15. Lorente, L., Torres, L.: Face recognition of video sequences in a MPEG-7 context using a global eigen approach. In: ICIP 99: Proceedings. 1999 International Conference on Image Processing. Volume Volume 4., Kobe, Japan (1999) 187 – 191
16. Chang, K., Bowyer, K., Flynn, P.: Face recognition using 2D and 3D facial data. Workshop in Multimodal User Authentication (2003) 25 – 32
17. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Expression-invariant 3D face recognition. Lecture Notes in Computer Science **Volume 2688** (2003) 62 – 69
18. Phillips, P., Newton, E.: Meta-Analysis of Face Recognition Algorithms. In: Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FRG'02). Volume Volume 64/1., Washington, DC, USA (2002) 235 – 241
19. Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley & Sons, Inc., New York, NY, USA (2002)
20. Lee, J.: Automatic Video Management System Using Face Recognition and MPEG-7 Visual Descriptors. ETRI Journal **27** (2005) 806 – 809
21. Europäische Union: Verordnung (EG) Nr. 2252/2004 des Rates vom 13. Dezember 2004 über Normen für Sicherheitsmerkmale und biometrische Daten in von den Mitgliedsstaaten ausgestellten Pässen und Reisedokumenten (2004)
22. Bundesamt für Sicherheit in der Informationstechnik BSI: Digitale Sicherheitsmerkmale im elektronischen Reisepass. `http://www.bsi.de/fachthem/epass/Sicherheitsmerkmale.pdf` (2005)
23. Bundesamt für Sicherheit in der Informationstechnik BSI: Golden Reader Tool - Faltblatt. `http://www.bsi.bund.de/literat/faltbl/F25GRT.pdf` (2005)
24. Chaos Computer Club e.V.: Pressemitteilung des Chaos Computer Club e.V. zur Einführung biometrischer Reisepässe. `http://www.ccc.de/epass/CCC20051004` (2005)
25. Bundesamt für Sicherheit in der Informationstechnik BSI: Öffentlicher Abschlussbericht Bioface I & II. `http://www.bsi.de/literat/studien/BioFace/BioFaceIIBericht.pdf` (2003)