

3D Gesichtserkennung - Merkmalsdetektion in 3D-Scans und merkmalsbasierter Vergleich von Gesichtern

Diplomarbeit

Vorgelegt von

Bettina Otten

Matrikelnr.: 200210302



Institut für Computervisualistik
Arbeitsgruppe Computergraphik

Betreuer: Dipl.-Inf. Frank Seibert (Fraunhofer IGD)

Prüfer: Prof. Dr.-Ing. Stefan Müller

März 2006

Ehrenwörtliche Erklärung

Hiermit versichere ich, die vorliegende Diplomarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Koblenz, 1. März 2006

Bettina Otten

Danksagung

An dieser Stelle möchte ich all jenen danken, die durch ihre fachliche und persönliche Unterstützung zum Gelingen dieser Diplomarbeit beigetragen haben. Besonderer Dank gebührt meiner Familie, auf deren Unterstützung ich mich jederzeit verlassen konnte. Ein großes Dankeschön aussprechen möchte ich außerdem meinem guten Freund Daniel Slawinski, der mir für die Dauer der Diplomarbeit sein Notebook zur Verfügung gestellt hat. Weiter bedanke ich mich bei Herrn Prof. Dr.-Ing. Stefan Müller sowie bei Herrn Dipl.-Inf. Frank Seiber für die Betreuung meiner Diplomarbeit am Fraunhofer IGD und die zahlreichen Ratschläge und Denkanstöße, die mir immer sehr geholfen haben.

Inhaltsverzeichnis

1	Einleitung	6
2	Biometrie	8
2.1	Grundlagen	8
2.2	Anforderungen an biometrische Merkmale	11
2.3	Verfahren	12
2.3.1	Physiologische Verfahren	12
2.3.2	Verhaltensbasierte Verfahren	14
2.4	Probleme biometrischer Verfahren	15
2.4.1	Toleranzschwelle	15
2.4.2	Mimik	16
2.4.3	Alterung	16
2.5	Benutzerakzeptanz	17
2.5.1	Transparenz	18
2.5.2	Handhabung	18
2.6	Rechtlicher Rahmen	19
2.7	Anwendungsbereiche	21
3	Gesichtserkennung	23
3.1	Motivation	23
3.2	3D Scans	25
3.2.1	Technologie	25
3.2.2	Eigenschaften	27
3.3	Normalisierung	29
3.4	Matchingverfahren	30
3.4.1	Holistische Verfahren	31
3.4.2	Merkmalsbasierte Verfahren	33
3.4.3	Elastic Bunch Graph Matching	34
3.4.4	Weitere Verfahren	35
4	Merkmalsbasierte Erkennung	37
4.1	Merkmale	37
4.2	Merkmalsdetektion	39
4.2.1	Curvatures	40
4.2.2	Wavelets	48
4.2.3	Interestoperatoren	51
4.2.4	Histogramme	54
4.3	Merkmalsvergleich	55
4.3.1	Ähnlichkeitsmaße	56
4.3.2	Registrierung	58
4.3.3	Elastische Graphenanpassung	59
4.3.4	Hauptkomponentenanalyse	66

4.3.5	Unabhängige Komponentenanalyse	68
4.3.6	Diskriminanten	70
4.3.7	Support Vector Machines	73
5	Auswertung	78
5.1	Teststandards	78
5.1.1	FAR und FRR	78
5.1.2	FER	79
5.1.3	ROC-Kurve	80
5.1.4	DET-Kurve	81
5.1.5	CMC-Kurve	81
5.2	Ergebnisse	82
6	Fazit	88
A	Das Gesicht	90
B	Differentielle Geometrie	92
B.1	Normalkrümmung einer Kurve	93
B.2	Hauptkrümmungen und -richtungen	94

Abbildungsverzeichnis

1	Löcher in Tiefenbildern	28
2	Spikes in Tiefenbildern	29
3	Formklassen	41
4	Segmentierungsergebnisse	42
5	Symmetrie bei Merkmalen	47
6	Symmetrien bei starker Glättung	48
7	Gaborfilter	50
8	Gabor Wavelets	50
9	Jet	51
10	“künstlerischer“ Bündelgraph	62
11	Bündelgraph	63
12	PCA vs. ICA	68
13	Hyperebenen	74
14	Maximum Margin Hyperplane	75
15	Verwendete Gesichtsgraphen	84
16	CSU Gesichtsgraphen	84
17	CMC-Kurve	85
18	DET-Kurve	86
19	ROC-Kurve	87
20	Gesichtmerkmale	90
21	Abmessungen der Augenregionen	91
22	Normalkrümmung	93
23	Shape Index Formklassen	95

1 Einleitung

Der Mensch ist in der Lage, ein Gesicht innerhalb von Bruchteilen einer Sekunde zu erkennen, selbst wenn das letzte Zusammentreffen mit der zugehörigen Person bereits Jahre zurückliegt. In biometrischen Gesichtserkennungsverfahren wird versucht, diese Leistung nachzuahmen.

Die Gesichtserkennung gewinnt neben anderen etablierten biometrischen Identifikations- und Verifikationsverfahren wie der Fingerabdruckerkennung immer mehr an Beliebtheit. Verschiedene Ansätze zur Gesichtserkennung wurden entwickelt, die das Problem des Wiedererkennens von Gesichtern auf unterschiedlichen Wegen lösen sollen. Die merkmalsbasierten Verfahren zeichnen sich dadurch aus, dass markante Merkmale aus einem Gesicht extrahiert werden und nur diese zum Vergleich herangezogen werden. Der Einsatz von zweidimensionalen Intensitätsbildern erforderte Kanten- und Eckendetektoren. Der Nachteil bei der Verwendung von Intensitätsbildern liegt darin, dass sie Unterschieden in der Beleuchtung unterworfen sind und Objekte bzw. Personen nur schwer in einer vergleichbaren Position aufnehmen können. Bei der 3D-Gesichtserkennung sind die üblichen Kanten- und Eckendetektoren nicht mehr ausreichend, da zum Beispiel Tiefenbilder nur wenige ausgeprägte Strukturen enthalten, die sich als Ecken oder Kanten identifizieren lassen. Die Merkmalsdetektion in Tiefenbildern ist ein zentraler Inhalt dieser Arbeit. Zusätzlich zu der Untersuchung von Oberflächeneigenschaften, die zur Merkmalsfindung beitragen, wurde im Rahmen dieser Arbeit ein Graph Matching Verfahren auf Tiefenbilder angewandt und getestet.

Diese Arbeit enthält im Wesentlichen vier Kapitel. In Kapitel 2 werden die Grundlagen der Biometrie erläutert und es wird eine Übersicht

über die gängigsten biometrischen Verfahren gegeben. Das Kapitel geht auf die verschiedenen Einsatzmöglichkeiten dieser Verfahren ein und beleuchtet die damit verbundenen rechtlichen Hintergründe. Kapitel 3 befasst sich mit der Gesichtserkennung, und der Entwicklung von 2D-Verfahren hin zu den heute bekannten 3D-Gesichtserkennungsverfahren. Die einzelnen Verarbeitungsschritte von der Bildaufnahme über die Weiterverarbeitung der Daten bis zur eigentlichen Erkennung werden aufgezeigt. Dieses Kapitel enthält außerdem einen Überblick über die aktuell verwendeten 3D-Gesichtserkennungsverfahren. In dem darauffolgenden Kapitel geht es vor allem um die merkmalsbasierten Verfahren. Verschiedene Möglichkeiten zur Merkmalsdetektion und zum Vergleich von Merkmalen werden vorgestellt. Die Ergebnisse des im Rahmen dieser Diplomarbeit verwendeten Verfahrens werden in Kapitel 5 präsentiert und ausgewertet.

2 Biometrie

Biometrie bezeichnet die Wissenschaft, die sich mit dem Vermessen von Merkmalen der Körperteile von Lebewesen beschäftigt. Die Bezeichnung „Biometrie“ setzt sich aus den zwei Worten „bios“ (Leben) und „metron“ (Maß) zusammen. Insbesondere werden in der Biometrie Merkmale von Menschen verwendet, um aus ihnen auf eine bestimmte Person zu schließen.

2.1 Grundlagen

Die Biometrie findet im Wesentlichen Einsatz in zwei Bereichen: der Verifizierung und der Identifizierung.

Unter Verifizierung versteht man die Überprüfung der vorgegebenen Identität einer Person. Dieser Beweis der Identität wird zum Beispiel häufig vor dem Zutritt zu bestimmten Gebäuden, vor finanziellen Transaktionen oder vor dem Passieren von Grenzen gefordert.

Die Identifizierung soll die Antwort auf die Frage „Wer ist die Person?“ liefern. Anhand von verschiedenen Hinweisen und Merkmalen wird nach der Identität einer unbekannt Person geforscht. Dies geschieht vor allem im Bereich der Strafverfolgung und Forensik. Meist stehen hier nur kleinste Spuren zur Verfügung, die auf die gesuchte Person hinweisen.

Das persönliche Wiedererkennen ist längst nicht mehr ausreichend als Authentifizierung und wurde ergänzt durch die Prinzipien von Besitz und Wissen. Allerdings sind sowohl Besitz als auch Wissen lediglich personenbezogen, d. h. sie können mutwillig an andere Personen übergeben werden oder durch Diebstahl entwendet werden.

Ein Schlüssel oder die heute weitverbreiteten Chipkarten bieten zwar die Berechtigung für den Zutritt zu einem Gebäude oder die Verwendung eines Geräts, aber es gibt keine Garantie, dass der Besitzer des Schlüssels tatsächlich auch der Berechtigte ist, dem ursprünglich der Schlüssel ausgehändigt wurde.

Ähnlich verhält es sich mit dem Wissen als Legitimationsmethode. Passwörter und Losungen ermöglichen es, den Missbrauch durch Diebstahl zu verringern. Die Sicherheit dieser Methode wird allerdings durch die Unachtsamkeit der Passwort-Besitzer reduziert, wenn diese ihr Passwort bzw. ihre PIN (Persönliche Identifikationsnummer) zu Erinnerung notieren. Somit ist das Passwort nicht mehr vor Zugriff von außen geschützt.

Eine Kombination von Wissen und Besitz wiederum erhöht die Sicherheit von Authentifizierungsverfahren, kann einen Betrug dennoch nicht vollständig ausschließen.

Durch die Verwendung von biometrischen Verfahren kann dieses Sicherheitsproblem weitestgehend gelöst werden. Biometrische Merkmale, die zur Authentifizierung verwendet werden, sind personengebunden, das heißt, eine Person kann diese Merkmale nicht an andere weitergeben und verlieren. Sie zeichnen sich dadurch aus, dass sie für jede Person individuelle Formen und Aussehen besitzen, so dass sie für eine zweifelsfreie Identifikation bzw. Verifikation herangezogen werden können.

Verifikation und Identifikation unterscheiden sich vor allem darin, dass sie unterschiedlich großen Aufwand erfordern. Für beide Verfahren wird im Vorfeld des Vergleichs ein Merkmalssatz aus den biometrischen Daten extrahiert. Bei der Verifikation wird dann ein Referenzmuster, das zum Beispiel auf einer vom Benutzer mitgeführten Speicherkarte zu finden ist, gelesen und

für den direkten Vergleich verwendet. Eine Alternative zur Speicherkarte ist die Verwendung einer ID, unter der das entsprechende Referenzmuster abgelegt ist. Bei diesem Prüfvorgang ist es nicht nötig, Informationen zu speichern oder weiterzuleiten, da der Vergleich direkt stattfindet. Man spricht hier auch von einem One-to-one-Matching. Die Verifikation ist einfach, und daher auch wesentlich schneller als die Identifikation. Im Gegensatz zur Verifikation kann die Identifikation sehr aufwendig sein. Bei diesem Verfahren wird zu dem vorhandenen Merkmalsatz eine Datenbank nach einem passenden Muster durchsucht. Es wird also mit zunehmender Größe der Datenbank eine höhere Rechenleistung beansprucht, da alle gespeicherten Muster geprüft werden müssen.

Für alle biometrischen Verfahren gilt die gleiche Voraussetzung: für eine Identifikation oder Verifikation muss die Person bzw. deren Identität dem System im Vorfeld bekannt gemacht werden. Während der Registrierung bzw. dem Enrollment wird ein Template erzeugt, das die Messdaten bzw. die berechneten Merkmale der zu registrierenden Person in einer Datenbank oder auf einer Chipkarte speichert. Um zu verhindern, dass die biometrischen Merkmale zusammen mit einer falschen Identität gespeichert werden, ist es nötig, eine genaue Identitätsprüfung vorzunehmen, die zum Beispiel auf der Prüfung des Personalausweises basiert.

Die Größe, bzw. der Speicherbedarf des Templates bzw. des Referenzmusters ist in der Regel wesentlich kleiner, als die vom Scanner aufgenommenen Daten. Die ursprünglich aufgenommene Datenmenge kann abhängig vom verwendeten biometrischen Verfahrens stark reduziert werden.

2.2 Anforderungen an biometrische Merkmale

Es existiert eine Vielzahl an biometrischen Merkmalen, die zur Authentifizierung verwendet werden können. Um gute Ergebnisse erzielen zu können, sollten die Merkmale möglichst die folgenden Eigenschaften erfüllen.

- Universalität

Ein biometrisches System ist nur dann sinnvoll, wenn es vom Großteil aller Menschen benutzt werden kann. Daher sollten die zu betrachtenden Merkmale bei allen Personen zu finden sein. Eine Ausnahme bilden jene Menschen, die diese Merkmale zum Beispiel aufgrund von Erbkrankheiten nicht ausbilden konnten oder deren Erscheinungsbild sich durch einen Unfall drastisch verändert hat, so dass entsprechende Merkmale nicht eindeutig gefunden werden können.

- Einzigartigkeit

Biometrisch relevante Merkmale sollten sich durch ihre Einzigartigkeit auszeichnen. Es sollten also vor allem die Merkmale für ein biometrisches Verfahren gewählt werden, die bei jedem Menschen unterschiedlich ausgeprägt sind. Die Einzigartigkeit der in einem System verwendeten Merkmale steht im direkten Verhältnis zur Sicherheit des Verfahrens. Je mehr Menschen es gibt, die für ein bestimmtes Merkmal die selben Eigenschaften aufweisen, desto weniger eignet sich dieses Merkmal für die biometrische Authentifizierung.

- Permanenz

Zeitliche Invarianz ist ein wichtiger Faktor bei der Auswahl von Merk-

malen. Für biometrische Systeme sind Merkmale, die sich im Laufe der Zeit leicht ändern oder verändern lassen, ungeeignet.

- Fälschungssicherheit

Um die Sicherheit, die ein biometrisches System zum Beispiel bei der Zutrittskontrolle von Gebäuden bieten soll, zu gewährleisten, dürfen nur Merkmale verwendet werden, die nicht durch Fälschungen ersetzt werden können. Wichtig ist die Erkennung des Merkmals am realen Menschen. Das System muss erkennen können, ob es sich lediglich um eine Kopie des Merkmals handelt.

- Erfassbarkeit

Das ausgewählte Merkmal muss messbar sein, damit eine elektronische Verarbeitung möglich ist. Auch sollte das Merkmal in relativ kurzer Zeit erfassbar sein.

2.3 Verfahren

Die verschiedenen biometrischen Verfahren lassen sich in zwei Kategorien einteilen: die physiologischen Verfahren und die verhaltensbasierten Verfahren.

2.3.1 Physiologische Verfahren

Zu den physiologischen Verfahren zählen jene, die sich mit Körpermerkmalen befassen. Dies können zum Beispiel die Gesichtsform, der Fingerabdruck oder die DNS eines Menschen sein. Bekannte physiologische Verfahren sind

der Iris-Scan, die Analyse des Fingerabdrucks, die DNS-Analyse, die Analyse der Finger- oder Handgeometrie oder die Gesichtserkennung.

Fingerabdruck Jeder Mensch zeichnet sich durch einen individuellen Fingerabdruck aus. Fingerabdrücke sind schon bei Kindern voll ausgeprägt und verändern sich im Laufe eines Lebens höchstens noch durch Fremdeinwirkungen wie Verletzungen oder chirurgische Eingriffe. Selbst eineiige Zwillinge unterscheiden sich durch ihre Fingerabdrücke. Fingerabdrücke entstehen aus nebeneinander verlaufenden Erhebungen der Haut. Man nennt diese auch Papillarleisten. Ihr Struktur bildet drei verschiedene Grundmuster aus: Schleifen, Bögen und Wirbel. Innerhalb dieser Muster lassen sich unterschiedliche Merkmale wiederfinden, die zur Identifikation bzw. Verifikation von Personen herangezogen werden können. Hierzu gehören beispielsweise Kreuzungen von Linien, Gabelungen, Linienenden, sehr kurze Linien, Delta oder Löcher innerhalb von Linien. Um eine Person anhand ihres Fingerabdrucks zu identifizieren, werden aus einem digitalen Bild diese spezifischen Merkmale der Fingerkuppe extrahiert und mit einem Referenzmuster verglichen. Die Verwendung von Fingerabdrücken in biometrischen Systemen ist heute weit verbreitet. Die dazu verwendeten Systeme sind klein, kostengünstig und einfach zu bedienen.

Iris-Scan Die menschliche Iris ist wie der Fingerabdruck ein sehr zuverlässiges biometrisches Merkmal, da sie sich in ihrer Struktur nach der Geburt nicht mehr verändert und zusätzlich durch ein sehr komplexes Muster charakterisiert wird. So ist die Wahrscheinlichkeit, dass zwei Personen eine identische Irisstruktur aufweisen, extrem gering. Die Iris, oder auch Regenbogenhaut, ist der Bereich des Auges, der die Pupille umschließt und

der dem menschlichen Auge seine Farbe verleiht. Dabei ist ihre Farbe die einzige Eigenschaft, in der die Iris sich noch verändern kann. Interessant für die Biometrie ist die Struktur, die in der Iris zu finden ist. Verschiedene Merkmale in unterschiedlichen Ausprägungen (Form, Länge, Größe, usw.) wie Streifen, Ringe, Punkte und Flecken lassen sich identifizieren.

2.3.2 Verhaltensbasierte Verfahren

Bei den verhaltensbasierten Verfahren handelt es sich um aktive Verfahren. Hierbei werden bestimmte Verhaltensmerkmale untersucht. Als Beispiel sind hier die Lippenbewegung oder die Stimme beim Sprechen, Unterschriftsdynamik oder auch der Bewegungsablauf beim Gehen zu nennen.

Unterschrifterkennung Die Unterschrift einer Person genießt auf der ganzen Welt große Akzeptanz im Bereich der Authentifizierung. Eine Unterschrift selbst aber auch der Vorgang des Unterschreibens weisen verschiedene Eigenschaften auf, die individuell unterschiedlich und personentypisch sind. Daher eignet sie sich zur biometrischen Erkennung. Bei dem Vergleich von Unterschriften wird nicht nur das optische Bild einer Unterschrift, sondern auch eine Vielzahl von zuvor extrahierten Einzelmerkmalen herangezogen. Statische Merkmale wie Aufstriche, Dichten, Kreuzungen, Abzweigungen, Schleifen- und Bogenformen sind ebenso von Interesse wie dynamische Merkmale wie Schreibdruck, Schreibdauer oder Schreibgeschwindigkeit. Mit Hilfe der Analyse von Linienstärke und Farbablagerungen, den sogenannten Schreibpastenablagerungen, lässt sich sogar sehr genau feststellen, ob es sich bei einer Unterschrift um eine Fälschung handelt.

2.4 Probleme biometrischer Verfahren

Bildbasierte biometrische Verfahren treffen immer wieder auf zwei grundlegende Schwierigkeiten: Variationen in der Beleuchtung und Verdeckungen. Verdeckungen entstehen, wenn eine Person sich einem Sensor nicht in der optimalen Position präsentiert. Das Resultat ist, dass Teile eines biometrischen Merkmals (zum Beispiel des Gesichts) nicht oder nur teilweise erfasst werden. Bei einem Gesicht lässt sich dies beispielsweise auf eine leichte Kopffrotation zurückführen. Dabei wird vor allem von der Nase ein Teil des Gesichts verdeckt.

Bei biometrischen Systemen, die in alltäglichen Umgebungen Einsatz finden, ist die Qualität der vom System getätigten Aufnahmen dem Wandel des Tageslichts unterworfen. Dieses Problem lässt sich nur lösen, in dem der erforderliche Scanvorgang unter festen Beleuchtungsverhältnissen durchgeführt wird. Dies wäre im Idealfall ein Raum ohne Fenster mit fest installierten Lichtquellen, die das aufzunehmende Objekt optimal ausleuchten.

Verdeckungen und Beleuchtungsvariationen vermindern die Qualität und damit den Wert des Bildes für das biometrische Verfahren, da so nicht sichergestellt werden kann, dass alle Bilder den zum Vergleich nötigen gleichen Informationsgehalt besitzen.

2.4.1 Toleranzschwelle

Ein weiteres Problem biometrischer Systeme ist die Neigung des Menschen zur Individualität. Menschen neigen dazu, sich äußerlich zu verändern, um ihrer Persönlichkeit Ausdruck zu verleihen. Dazu gehört unter anderem das Verändern der Frisur und das Tragen von Hüten und Brillen. Hinzu kommen Veränderungen bedingt durch Krankheit oder Verletzungen sowie durchs Al-

tern. Ein biometrisches System muss mit solchen Veränderungen umgehen können und entsprechende Abweichungen des Merkmals erlauben. Die Toleranzschwelle legt also fest, wie hoch der Grad an Veränderung höchstens sein darf, damit die geprüfte Person trotzdem akzeptiert wird.

2.4.2 Mimik

Probleme, die speziell die Effizienz von Gesichtserkennungssystemen beeinflussen, begründen sich in der Natur des Gesichts. Es besteht aus einer großen Menge von Muskeln und kann daher eine Vielzahl von Gesichtsausdrücken annehmen. Viele Gesichtserkennungsverfahren können mit unterschiedlichen Gesichtsausdrücken nicht umgehen und so wird meistens ein „neutraler“ Gesichtsausdruck während der Aufnahme gefordert. Aber selbst ein neutraler Ausdruck variiert von Aufnahme zu Aufnahme. Die jeweilige Gemütsstimmung spiegelt sich in einem Gesicht wieder, selbst wenn die betreffende Person sich dessen nicht bewusst ist. Leicht gehobene Mundwinkel oder Augenbrauen machen diese Variation zum Beispiel aus.

2.4.3 Alterung

Das zunehmende Alter macht sich vor allem im Gesicht bemerkbar. Daher sollten Referenzmuster für die Gesichtserkennung in regelmäßigen Abständen (ca. alle 10 Jahre) erneuert werden.

Das Alter äußert sich im Gesicht durch eine erschlaffende Gesichtshaut und eine zunehmende Faltenbildung. Die einzelnen Komponenten, die das Gesicht zusammenstellen, nämlich Haut, Unterhaut-Fettgewebe, Muskulatur und Knochenstruktur altern unabhängig voneinander. Beim jungen Men-

schen und noch im mittleren Lebensalter stehen alle in einem gewissen Gleichgewicht. Doch mit fortschreitendem Alter droht ein Ungleichgewicht zwischen Knochen, Muskeln, Fett und Haut. In einigen Bezirken des Gesichtes kommt es zu einer Zunahme von Fett, während in anderen Bereichen eine Reduktion erfolgt. Das führt dazu, dass das so genannte Weichgewebe der Schwerkraft folgend herabsinkt.

Bei Heranwachsenden liegen zum Beispiel die Augenbrauen noch oberhalb des oberen Rands der knöchernen Augenhöhle. Bei einem 50-Jährigen aber sind sie bereits über diesen Rand nach unten gewandert und bei einem 70-Jährigen befinden sie sich deutlich unterhalb der früheren Position.

2.5 Benutzerakzeptanz

Im Vergleich zur Verwendung von PINs oder Chipkarten ist die Verwendung eines biometrischen Verfahrens viel persönlicher und intimer, da sich die zu indentifizierende Person einer Maschine mit einem für das Verfahren verwendeten Körperteil präsentieren muss. Der Benutzer muss sich zum Beispiel von einem Gesichts- oder Iris-Scanner abtasten lassen oder seine Hand bzw. seine Fingerspitze auf ein entsprechendes Sensorfeld legen. Dies ist für die Erfassung der Daten notwendig. Daher ist es wichtig, dass das biometrische Verfahren eine hohe Akzeptanz beim Benutzer findet und er sich nicht gestört oder eingeschränkt fühlt. Die Benutzerakzeptanz hängt von den zwei Faktoren Informationstransparenz und Handhabung ab.

2.5.1 Transparenz

Zweifel an dem biometrischen Verfahren lassen sich am besten ausräumen, indem man den Benutzer mit umfassenden Informationen versorgt. Nur wenn er versteht, welche Risiken und Chance ein solches System bietet und wie genau der Vorgang des biometrischen Erkennens vor sich geht, kann mit einer hohen Akzeptanz seinerseits gerechnet werden. Wichtige Punkte bei der Aufklärung des Benutzers sind

- Umgang mit der Datenspeicherung
- Maßnahmen zur Verhinderung von Missbrauch
- Zugriffsrechte beim Betreiber
- Einstellung der Toleranzschwellen

2.5.2 Handhabung

Um dem Benutzer den Vorgang der biometrischen Identifizierung bzw. Verifizierung so angenehm wie möglich zu gestalten, sollten folgende drei Anforderungen erfüllt werden: Einfachheit, Schnelligkeit, Ergonomie.

- Einfachheit und Bequemlichkeit

Es sollte für den Benutzer möglichst auf den ersten Blick ersichtlich sein, wie er mit dem Bedienelement umzugehen hat. Dabei ist es sinnvoll, wenn der Benutzer zum Beispiel ganz entspannt vor einem Scanner stehen kann, ohne sich in eine unbequeme Körperhaltung bringen zu müssen. Unnötige Knöpfe und ähnliches würden den Benutzer nur verwirren. Dagegen könnte ein eindeutiges Feedback zum Beispiel in

Form einer Anzeige oder eines akustischen Signals für Klarheit sorgen.

- **Schnelligkeit**

Die wenigsten Menschen haben die Zeit und Geduld, lange auf eine biometrische Entscheidung zu warten. Daher sollte das System zumindest so schnell reagieren, dass kaum ein Unterschied zu herkömmlichen Verfahren wie der Prüfung von PINs besteht.

- **Ergonomie der Endgeräte**

Schon die unterschiedlichen Körpergrößen verschiedener Personen erzwingt die flexible Gestaltung und Anbringung der Anwendungsgeräte. Höhenverstellbarkeit ist aber nur ein Aspekt der Ergonomie. So sollte zum Beispiel auch auf eine behindertengerechte Gestaltung der Endgeräte geachtet werden.

2.6 Rechtlicher Rahmen

Die Verwendung von biometrischen Merkmalen einer Person kann je nach Verfahren und je nach Anwendungsgebiet mehr oder weniger eine Beeinträchtigung der Rechte dieser Person darstellen. Vor allem betrifft dies das Recht auf informationelle Selbstbestimmung¹. Dieses Recht garantiert jedem Einzelnen, dass er selbst darüber bestimmen kann, in welchem Umfang

¹Das Bundesverfassungsgericht hat im Volkszählungsurteil vom 15.12.1983 das Grundrecht auf informationelle Selbstbestimmung anerkannt:

„Das Grundrecht gewährleistet insoweit die Befugnis des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen. Einschränkungen dieses Rechts auf „informationelle Selbstbestimmung“ sind nur im überwiegenden Allgemeininteresse zulässig.“

Es besteht demnach ein „Schutz des Einzelnen gegen unbegrenzte Erhebung, Speicherung, Verwendung und Weitergabe seiner persönlichen Daten“. Das Grundrecht auf informationelle Selbstbestimmung wird verfassungsrechtlich aus Art. 2 Abs. 1 (sog. allgemeine Handlungsfreiheit) in Verbindung mit Art. 1 Abs. 1 GG hergeleitet.

und zu welchen Zwecken er persönliche Daten preisgibt. Zu persönlichen Daten² zählen Angaben über persönliche und sachliche Verhältnisse einer Person.

Der Datenschutz ist im Bereich der Biometrie dafür zuständig, die individuellen Daten von Benutzern zu kontrollieren und einen zuverlässigen Umgang mit ihnen sicherzustellen. Um den Ansprüchen des Datenschutzes zu genügen, sollte ein biometrisches Verfahren folgende Punkte berücksichtigen:

- Die Datenerhebung darf nicht erfolgen, wenn der Benutzer keine Kenntnis darüber besitzt. Jeder Nutzer muss in jedem Fall darüber informiert sein, dass seine persönlichen Daten aufgenommen und zu welchem Zweck sie verwendet werden sollen.
- Das biometrische System sollte sich in der Verwendung persönlicher Daten auf so wenige wie möglich beschränken und keine für das Verfahren überflüssige Daten erheben.
- Die Einschränkungen, die ein biometrisches System aufgrund von Datenschutzgesetzen erfährt, verlieren ihren Bestand, sobald die Daten einer Person soweit vom System verändert wurden, dass keine Identifizierung der zugehörigen Person aufgrund dieser Daten mehr möglich ist. Der Nutzer hat dann keine Entscheidungsbefugnis mehr über die anonymisierten Daten.

Für die Erzeugung anonymisierter Daten ist aber im Vorfeld eine Einwilligung zur Verarbeitung der zunächst personenbezogenen Informationen des Betroffenen nötig.

²Angaben über Betroffene, deren Identität bestimmt oder bestimmbar ist (Datenschutzgesetz, Art. 2, § 4)

Grundsätzlich ist die Erhebung biometrischer Daten also rechtlich unbedenklich, wenn dies freiwillig geschieht und dem Nutzer bekannt ist, zu welchem Zweck die Informationen weiterverarbeitet werden und wofür die Daten möglicherweise noch verwendet werden. Der Vorgang des Erkennens sollte nicht ohne den Willen oder das Wissen des Benutzers durchgeführt werden können, wie es zum Beispiel bei berührungslosen Systemen zur Gesichts- oder Iriserkennung möglich wäre.

2.7 Anwendungsbereiche

Biometrische Systeme finden heute bereits in verschiedenen Bereichen Anwendung. Diese Bereiche lassen sich wie folgt unterteilen:

- **Physische Zugangskontrolle**

Das Biometrische System steuert den Zugang zu Gebäuden und Räumlichkeiten, wie zum Beispiel zu Büroräumen oder Hochsicherheitstrakten.

- **Logische Zugangssysteme**

Anmeldeverfahren, die Zugriff auf PCs, Mobiltelefone oder Geldautomaten ermöglichen, verwenden normalerweise eine Kombination von Benutzername und Passwort. Diese lassen sich durch biometrische Verfahren ersetzen.

- **Legitimation**

Die Verwendung von biometrischen Systemen kann wichtige Dokumente wie Ausweise oder Führerscheine ersetzen.

- Überwachung und Kontrolle

Die Überwachung öffentlicher Plätze wird durch die Biometrie vereinfacht, weil Personen automatisch identifiziert werden können, ohne dass eine Interaktion nötig ist.

- Forensik

Die Forensik beschäftigt sich mit der Überführung von Straftätern anhand biometrischer Merkmale. Dabei kann schon die kleinste Spur, vom Haar oder Blutstropfen bis zum Fingerabdruck, die Identität des Straftäters offenbaren.

3 Gesichtserkennung

3.1 Motivation

Das Erkennen von Personen anhand ihrer Gesichter ist für den Menschen eine alltägliche Erfahrung. Dabei ist allerdings nicht eindeutig zu sagen, ob das menschliche Gehirn Gesichter als Einheit betrachtet, oder ob nur bestimmte Merkmale registriert und im Gehirn weiterverarbeitet werden. Die automatische Gesichtserkennung durch ein biometrisches System versucht, die natürliche Fähigkeit des menschlichen Gehirns, Gesichter zu erkennen und zu unterscheiden, nachzubilden.

Es existiert bereits eine Vielzahl von Ansätzen zur Gesichtserkennung im 2D-Bereich. Die Gesichtserkennung basierend auf Grauwert- oder Farbbildern ist sehr zuverlässig, allerdings ist es schwierig, für die verwendeten Bilder immer gleiche Ausgangssituationen zu schaffen, wenn die Gesichtserkennung im Alltag angewandt werden soll. Vor allem zwei Probleme treten immer wieder auf, denen nur durch kontrollierte Verhältnisse während der Bildaufnahme beizukommen ist: Variationen in der Beleuchtung und unterschiedliche Körper- bzw. Kopfhaltungen.

Das Entwickeln von Gesichtserkennungsalgorithmen, die gegenüber diesen Einflussfaktoren invariant sind, ist eine der Hauptaufgaben der aktuellen Forschung im Bereich der Gesichtserkennung. Es gibt zwei verschiedene grundlegende Ansätze. Einer von ihnen zielt darauf, Merkmale im Bild zu entdecken, die gegenüber Änderungen von Blickpunkt und Beleuchtung unempfindlich sind. Dieser Ansatz wird auch merkmalsbasierte Erkennung genannt. Leider ist es gerade wegen der Unterschiede in Beleuchtung und Blickwinkel schwierig, in 2D-Bildern verlässliche Merkmalspunkte zu finden, was dazu führt, dass die entsprechenden Algorithmen mit sehr wenigen Punkten auskommen müssen. Die Unterscheidung zwischen Gesichtern ist dadurch

nicht zuverlässig genug. Ein ähnlicher Ansatz, die template-basierte Erkennung, stößt an seine Grenzen, wenn während eines Vergleichs unterschiedliche Gesichtsausdrücke oder Kopfhaltungen auftreten.

In holistischen Gesichtserkennungsverfahren werden die Gesichtsbilder im Ganzen als multidimensionale Vektoren aufgefasst. Durch das Verringern der Dimension des Vektorraums ist es möglich, die gesichtsspezifischen Informationen zu extrahieren und den folgenden Vergleich auf diese Informationen zu konzentrieren. Dieses Verfahren funktioniert sehr gut, solange alle Bilder, sowohl gallery- als auch Testbilder, unter gleichen Bedingungen erzeugt wurden. Ansonsten erfordert es eine genaue Registrierung der Bilder, die wiederum das Auffinden von robusten Merkmalspunkten nötig macht.

2D-Gesichtserkennungsverfahren berücksichtigen nur eingeschränkt, dass das Gesicht ein dreidimensionales Objekt ist und dass gerade die daraus resultierende geometrische Beschaffenheit des Gesichts die Menschen grundlegend voneinander unterscheidet. Dem entsprechend lassen sich 2D-Verfahren auch einfacher täuschen. Zum Beispiel ist es für Maskenbildner aus Theater und Film ein leichtes, einer Person ein beliebiges anderes Aussehen zu verleihen, mit dem sie ein entsprechendes System überlisten kann.

Ein Grund für die Popularität von 2D-Gesichtserkennungsverfahren liegt sicherlich darin, dass 2D-Grauwertbilder bzw. Farbbilder einfach und günstig zu erzeugen und leicht in ihrer Handhabung sind. Aber auch in der 3D-Sensortechnik hat es in den letzten Jahren Fortschritte gegeben, so dass die Verwendung von 3D-Daten für die Gesichtserkennung immer attraktiver wird. Die oben erwähnten Einschränkungen, die mit der Benutzung von 2D-Bildern einhergehen, können mit Hilfe von 3D-Scans überwunden werden. Das Ergebnis eines 3D-Scans ist entweder ein 3D-Mesh oder ein Tiefenbild. Diese bieten strukturelle Informationen über die Oberfläche ei-

nes Objekts und sind unempfindlich gegen Beleuchtungsänderungen. Beide Repräsentationen eignen sich zur Gesichtserkennung, allerdings sind 3D-Meshes schwieriger zu handhaben. Innerhalb eines Tiefenbildes repräsentiert jeder Punkt einen Wert, der den Abstand des Punktes zur Bildebene beschreibt. Dieser Tiefenwert ist unabhängig von der Beleuchtung. Im Gegensatz zu 3D-Meshes sind die 3D-Informationen eines Tiefenbildes leichter zu verarbeiten, da die einzelnen Punkte des Tiefenbildes regulär angeordnet sind. Es finden sich also einfachere Nachbarschaftsbeziehungen.

Diese Eigenschaften haben dazu geführt, dass hauptsächlich Tiefenbilder in der 3D-Gesichtserkennung benutzt werden. Aktuelle Verfahren in der 3D-Gesichtserkennung werden in Kapitel 3.4 vorgestellt.

Durch die Verwendung von Tiefeninformationen eines Gesichts haben Unterschiede in der Beleuchtung während des Aufnahmeprozesses einen wesentlich geringeren Einfluss auf die Qualität der Aufnahme und auf das Gesichtserkennungsverfahren. Das Problem, dass Personen sich nie in exakt der gleichen Weise bzw. Position einem Scanner oder einer Kamera präsentieren, wird durch die Verarbeitung der Scans vor dem eigentlichen Erkennungsprozess gelöst. Ein Normalisierungsschritt sorgt dafür, dass sämtliche zu vergleichende Gesichter in gleicher Weise ausgerichtet werden und so eine einheitliche Basis den darauffolgenden Schritten im Erkennungsverfahren zugrunde liegt.

3.2 3D Scans

3.2.1 Technologie

3D Scanner bieten die Möglichkeit, ein beliebiges dreidimensionales Objekt oder eine Umgebung zu erfassen und ihre Forminformationen in verwert-

bare Daten umzusetzen. Oft werden außerdem noch Farbinformationen mit aufgenommen. Mit Hilfe der so erhaltenen Daten lassen sich digitale, dreidimensionale Modelle des gescannten Objektes erstellen. 3D Scanner finden heute Einsatz unter anderem in der Industrie als auch in der Unterhaltungsbranche.

Im Gegensatz zu 2D-Aufnahmeverfahren, in denen lediglich die Farbinformationen einer Oberfläche erfasst werden, werden in 3D-Verfahren die Distanzen des betrachteten Objektes gemessen und die resultierenden Tiefeninformationen aufgenommen. Die so erzeugten Bilder beschreiben also in jedem Punkt den Abstand zu der gescannten Oberfläche.

In vielen Fällen werden mehrere Aufnahmen eines Objektes benötigt, um eine Rundum-Ansicht eines Objektes zu erhalten. Ein 3D-Scanner erfasst nur „sichtbare“ Oberflächen und so muss dem Fehlen von Teilbereichen, die bei einer bestimmten Ansicht aufgrund von Verdeckungen nicht aufgenommen werden konnten, Rechnung getragen werden durch das wiederholte Aufnehmen des Objektes aus verschiedenen Blickwinkeln. Um ein vollständiges Objekt zu erhalten, müssen die Teilaufnahmen anschließend verbunden werden.

Die große Vielfalt an 3D Scannern lässt sich grob unterteilen in passive und aktive Verfahren. Eine weitere Einteilung ist möglich in Geräte, bei denen für eine Aufnahme Körperkontakt hergestellt werden muss (zum Beispiel Fingerabdruckscanner), und jene, bei denen auf einen direkten Kontakt verzichtet werden kann. Im Folgenden werden kurz einige Scanverfahren vorgestellt.

Der Structured Light 3D Scanner kommt ohne Körperkontakt aus. Er erzeugt ein Muster aus Licht, das auf das zu scannende Objekt projiziert wird. Die Form des Objektes lässt sich dann daraus ableiten, an welchen

Stellen und wie stark sich das Lichtmuster verformt. Eine Kamera erfasst diese Verformung, wobei die Distanz zu jedem einzelnen Punkt innerhalb des Lichtmusters berechnet wird. Die Verwendung von Structured Light Scannern hat den Vorteil, dass das Verfahren sehr schnell ist. Es wird nicht nur ein Punkt pro Zeiteinheit erfasst, sondern mehrere oder sogar das gesamte Blickfeld. Dadurch wird zusätzlich das Problem der Verzerrung durch Bewegung vermindert.

Stereoskopische Systeme basieren auf zwei Kameras, die mit einem geringfügigen Abstand voneinander fest montiert sind. Ein zu scannendes Objekt wird von beiden Kameras gleichzeitig aufgenommen. Durch die Analyse der geringen Unterschiede in den Aufnahmen lassen sich die Distanzen jedes einzelnen Punktes bestimmen. Das stereoskopische Scannen ist ein passives Verfahren, für das lediglich zwei Kameras benötigt werden. Weil in der Regel keine zusätzliche Hardware nötig ist, sind passive Verfahren oft günstiger als jene, die Lichtmuster verwenden. Bei dieser Aufnahmetechnik ist es allerdings wichtig, neben fest installierten Kameras auch fest installierte Lichtquellen zu verwenden, um eine optimale Ausleuchtung zu gewährleisten.

Die Daten, die von einem 3D Scanner erzeugt werden, bedürfen in der Regel einer Vorverarbeitung, bevor sie in einem speziellen Kontext verwendet werden können. Für Applikationen ist es meist erforderlich, die Daten in polygonale 3D-Modelle oder Tiefenbilder umzurechnen.

3.2.2 Eigenschaften

3D-Daten, die von einem 3D-Scanner erzeugt wurden, bieten selten ideale Voraussetzungen für ihre Verwendung. In gescannten Gesichtern treten

im Grunde immer irgendwelche Formen von Artefakten auf. Diese sind zurückzuführen auf Flächen, die der Scanner nicht erfassen konnte, wie zum Beispiel die glänzenden Oberflächen der Augen, und auf Regionen wie Augenbrauen und Bärte.

Es gibt grundlegend zwei Arten von Artefakten, die in 3D-Scans auftreten können: Löcher und Spikes. Löcher sind Bereiche, für die keine Daten vorhanden sind, weil der Scanner an diesen Stellen keine diffuse Oberfläche aufgefunden hat (vgl. Abbildung 1). Sie treten außerdem dann auf, wenn während des Scanvorgangs ein Objekt an einer Stelle von einem anderen Objekt verdeckt wurde.

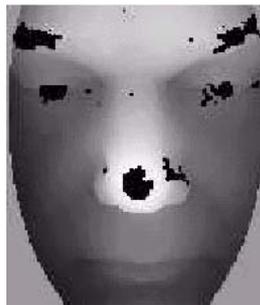


Abbildung 1: Löcher: 3D-Daten gerendert als Tiefenbild, die schwarzen Bereiche entsprechen fehlenden Daten, [21]

Spikes sind Ausreißer in den Daten, die zum Beispiel durch Reflexionen entstehen oder durch Korrespondenzprobleme im Stereoverfahren (vgl. Abbildung 2).

Löcher in den Daten lassen sich am leichtesten durch Interpolation unter Berücksichtigung von Werten am Rand der Löcher schließen. Für große Löcher ist dies allerdings nicht die optimale Lösung, da die Interpolation sehr ungenau werden kann und so die rekonstruierte Oberfläche stark vom Original abweicht.

Das Entfernen von Spikes kann beispielsweise durch Glätten der Oberfläche



Abbildung 2: Spikes: Ausreißer erscheinen als „Stachel“ mit großem Abstand zur eigentlichen Gesichtsoberfläche [21]

erfolgen. Die Glättung mit einem Median-Filter ist zum Beispiel sinnvoll, weil dadurch die korrekt erfasste Oberfläche im Umfeld des Spikes wenig beeinflusst wird.

3.3 Normalisierung

Wie bereits in Kapitel 3.1 erwähnt, ist es nötig, die erzeugten Tiefenbilder bzw. Meshes zu normalisieren, um eine gemeinsame Grundlage für einen erfolgreichen Vergleich zu schaffen. Mit Normalisierung bezeichnet man das Berechnen einer Transformation, die ein Gesicht auf eine definierte Position ausrichtet. Diese Transformation beinhaltet in der Regel

1. eine Translation - Das Gesicht wird an eine neue Position verschoben. Oft wird dabei die Nasenspitze in den Bildmittelpunkt bewegt.
2. eine Rotation - Um die rotationsmäßige Ausrichtung zu erreichen, werden mindestens zwei Punkte aus dem Gesicht benötigt. Häufig wird hierfür die Nasenspitze in Verbindung mit der Nasenwurzel verwendet.

Für die Normalisierung anhand von Merkmalspunkten werden markante Punkte eines Gesichts benötigt, die in jedem Gesicht einfach und sicher

wiederzufinden sind. In verschiedenen Ansätzen werden hierfür verschiedene Merkmale des Gesichts vorgeschlagen. Vor allem die Nasenspitze ist eine sinnvolle Wahl, weil sie leicht zu detektieren ist, da sie der am weitesten aus dem Gesicht herausragende Punkt ist. Des Weiteren ist die Position der Nase unabhängig vom Gesichtsausdruck. Die Mundwinkel sind aus diesem Grund zum Beispiel nicht besonders gut zur Normalisierung geeignet, da ein leichtes Lächeln bereits ganz andere Mundwinkelpositionen hervorbringt als ein neutraler Mund. Die Pupillen bzw. die Mittelpunkte der Augen sind im Prinzip ein zuverlässiges Mittel zur Normalisierung, da auch ihre Positionen sich wie die der Nase nicht verändern. Allerdings weisen die meisten Tiefenbilder von Gesichtern an Stelle der Pupillen Löcher auf, da diese andere Reflektionseigenschaften als die restliche Gesichtshaut besitzen und vom 3D-Scanner nicht immer korrekt erfasst werden.

Die ungefähre Position der Nase in einem Gesicht lässt sich zum Beispiel über die Berechnung der konvexen Hülle desselben bestimmen. Die Nase ist das Organ, das am weitesten aus dem Gesicht herausragt und sie befindet sich ungefähr auf der Symmetrieachse des Gesichts. Durch die Betrachtung der Verteilung der Tiefenwerte oder der Krümmungseigenschaften (vgl. Kapitel 4.2.1) des Gesichts entlang dieser Achse kann man dann die genaue Position der Nasenspitze sowie der Nasenwurzel lokalisieren.

3.4 Matchingverfahren

Gesichtserkennungsverfahren lassen sich grundlegend in zwei unterschiedliche Kategorien einteilen, die holistischen und die merkmalsbasierten Verfahren. Erstere behandeln die Gesichter oder Gesichtsbereiche im Ganzen als Vektor, der zum Vergleich mit anderen herangezogen wird, während in

einem merkmalsbasierten Ansatz zuerst eine Segmentierung des Gesichts vorgenommen wird, und die so extrahierten Merkmale den Merkmalsvektor bilden. Bei dieser Art der Gesichtserkennung geht ein großer Teil der vorhandenen Informationen nach der Merkmalsextraktion verloren, weshalb es besonders wichtig ist, ein verlässliches und robustes Verfahren für diesen Vorgang zu benutzen.

Die folgenden Kapitel geben einen kurzen Überblick über die aktuell verwendeten Gesichtserkennungsmethoden.

3.4.1 Holistische Verfahren

Ein beliebtes Verfahren in der Gesichtserkennung ist die Verwendung der Hauptkomponentenanalyse (PCA, principal component analysis). Mit ihrer Hilfe lässt sich die Dimension des Merkmalsraums verringern, so dass sich der anschließende Vergleich nur auf die wichtigsten Merkmale konzentriert. Die Gesichtserkennung mit PCA wird auch Eigengesicht-Methode genannt. Die Erkennung mittels Eigengesichtern wurde bereits 1991 von Turk und Pentland [4] vorgestellt. Ihre Anwendung auf Tiefenbilder wird in [5] beschrieben.

Die Eigengesicht-Methode lässt sich in zwei Hauptaufgaben unterteilen, der Lernphase und der Erkennungsphase. Im ersten Schritt werden die Gesichtsbilder mit Hilfe der Hauptkomponentenanalyse in den Gesichtsraum (face space) projiziert, wo sie durch n -dimensionale Vektoren repräsentiert werden. Dabei entspricht n der Größe des Ausgangsbildes, also *breite* \times *hoehe*. Alle Gesichter haben grundsätzlich eine ähnliche Struktur. Die aus den Gesichtsbildern gebildeten Vektoren besitzen also ebenfalls eine gewisse Ähnlichkeit. Das bedeutet, dass alle Vektoren im Merkmalsraum innerhalb

eines kleinen Bereiches auftreten. Sie bilden sozusagen eine Wolke von Datenpunkten. Ein Großteil der vorhandenen Dimensionen ist also gar nicht nötig und es macht Sinn, eine Verringerung der Dimension des Merkmalsraumes durchzuführen, damit im späteren Vergleich nur gesichtsrelevante Informationen verwendet werden. Unwichtige oder unerwünschte Informationen wie Hintergrund oder Rauschen können auf einfache Weise mittels PCA aussortiert werden.

In der zweiten Phase findet die eigentliche Gesichtserkennung statt. Für die Identifikation neuer Gesichter werden deren Vektoren mit Hilfe der Eigengesichter genauso reduziert wie die der Trainingsmenge, und können dann zum Vergleich verwendet werden. Zum Vergleich dienen unterschiedliche Distanzmaße, wie zum Beispiel die euklidische Distanz, oder Klassifizierungsverfahren wie zum Beispiel die Support Vector Machines.

Die Methode der Eigengesichter ist nur dann sinnvoll, wenn sowohl die Bilder der Trainingsmenge als auch die der Testmenge unter exakt den gleichen Umständen erzeugt wurden, da sie sehr empfindlich bezüglich Beleuchtungsänderungen, unterschiedlichen Posen und Gesichtsausdrücken ist.

In [5] wurde eine Datenbank aus 222 Trainingsbildern von 37 unterschiedlichen Personen verwendet. Zu jeder Person enthielt die Datenbank sechs Bilder mit unterschiedlichen Gesichtsausdrücken. Eine Erkennungsrate von bis zu 100% wurde von dem Verfahren erreicht, wenn die Gesichter der Testbilder bereits in der Trainingsmenge enthalten waren. Ohne Überschneidung der zwei Bildmengen liegt die Erkennungsrate bei bis zu 94%.

Eine weitere Möglichkeit, Gesichter zu kodieren, ist die lineare Diskriminanzanalyse (linear discriminant analysis, LDA). Erstmals fand die lineare

Diskriminanzanalyse Anwendung in der Gesichtserkennung 1997 [6]. Dabei zeigte sich, dass dieses Verfahren eine wesentliche Verbesserung gegenüber der Hauptkomponentenanalyse darstellte. Vor allem im Bezug auf Beleuchtungsänderungen wies die LDA-Methode eine geringere Fehlerrate auf, als die PCA-Methode. Die LDA-Methode wird auch als Fisherfaces-Methode bezeichnet.

Im Vergleich von PCA und LDA wurde schnell klar, dass die LDA nicht in allen Punkten erfolgreicher war als die PCA. Dabei spielte die Größe der Trainingsmenge eine besondere Rolle, denn bei besonders kleinen Trainingsmengen zeigte sich, dass die PCA wesentlich robuster reagierte.

3.4.2 Merkmalsbasierte Verfahren

Oberflächen lassen sich aufgrund ihrer Krümmung in Regionen unterschiedlicher Eigenschaften unterteilen: konvexe und konkave Regionen sowie Sattelpunkte. Diese Informationen über eine Oberfläche eignen sich sehr gut zur Objekterkennung.

In vielen verschiedenen Verfahren werden die Krümmungseigenschaften der Gesichtsoberfläche verwendet, um das Gesicht zu segmentieren und interessante Merkmale zu finden. In [9] verwendet Gordon die Mittlere und die Gausskrümmung in Verbindung mit Tiefenbildern, um die Regionen der Augen und der Nase zu extrahieren. Beim Vergleich dieser Regionen wurde eine Erkennungsrate von 97% erzielt. Die verwendete Datenbank enthielt Bilder von 24 verschiedenen Personen.

Moreno et al. benutzen ebenfalls die Eigenschaften der Mittleren und der Gausskrümmung, um Gesichter zu segmentieren. Aus den entstehenden Regionen und Linien werden 86 verschiedene Merkmale extrahiert, die un-

terschiedlich stark diskriminierende Eigenschaften besitzen. Mit Hilfe der Fisher-Koeffizienten (4.3.6), [13], wird berechnet, für welche Merkmale diese besonders ausgeprägt sind. Dies führt zu einem Merkmalsvektor, der die 35 wichtigsten Merkmale beinhaltet. Die Ähnlichkeit zwischen Merkmalsvektoren wird mit Hilfe der Euklidischen Distanz festgestellt. Das von Moreno vorgestellte Verfahren lieferte eine Erkennungsrate von bis zu 92%. Beste Ergebnisse wurden bei neutralen Gesichtsausdrücken erreicht.

3.4.3 Elastic Bunch Graph Matching

Die Gesichtserkennung mittels Elastic Bunch Graph Matching wurde von L. Wiskott et al. entwickelt. In [8] zeigte Wiskott, wie Face Bunch Graphen in der Lage sind, Gesichter zu beschreiben. Die Verwendung dieses Verfahrens im Rahmen der Gesichtserkennung auf der Basis von Gabor Wavelets wird in [7] erläutert. Das Elastic Bunch Graph Matching ist eine Methode zur Objekterkennung, die auf lokalen und geometrischen Merkmalen basiert. Objekte werden mit objekt-adaptierten Graphen und Gabor Wavelets modelliert. Für jedes Bild einer Trainingsmenge wird ein Graph erstellt und alle so erzeugten Graphen werden zu einem Bunch Graph zusammengefasst. Dieser stellt bezüglich der geometrischen Merkmalspositionen der Trainingsgraphen einen Durchschnittsgraph dar. Zusätzlich zu diesen Landmarken enthält er noch die Wavelet-Antworten zu jeder Landmarke aus jedem Bild der Trainingsmenge. Somit deckt der Bunch Graph sämtliche Variationen im Trainingsmaterial ab, also zum Beispiel unterschiedliche Hautfarben, Gesichtsausdrücke, Gesichtsformen, Bärte, Brillen oder Variationen durch Geschlecht oder Alter. Der Bunch Graph wird verwendet, um für neue Gesichter automatisch einen Graphen zu generieren. Die eigentliche Erkennung erfolgt

dann durch den Vergleich der Graphen der verschiedenen Gesichter. Laut [7] liegt die Erkennungsrate für Frontalaufnahmen bei 98%, für Profilaufnahmen wird eine Rate von 84% erreicht. Der Test umfasste eine Untermenge der FERET-Datenbank, die aus 250 verschiedenen Gesichtern bestand.

3.4.4 Weitere Verfahren

Das Gesichtserkennungsverfahren nach Lu, Colbry und Jain [15], [16] verwendet eine hybride Variante des Iterative Closest Point Algorithmus zur Registrierung von 3D-Oberflächen. Diese vereint die große Geschwindigkeit des Algorithmus von Besl und McKay [17] mit der Genauigkeit des Verfahrens nach Chen und Medioni [18]. Der bei der Registrierung zweier Oberflächen entstehende mittlere Fehler dient als Vergleichswert. Bei einer Datenbank mit 18 3D Modellen verschiedener Personen wurde eine Erkennungsrate von 79,3% [19] erreicht, wenn lediglich der mittlere Fehler berücksichtigt wurde. Zur Verbesserung des Verfahrens wurden zusätzliche Informationen wie Texturinformationen oder Shape Index-Werte für den Vergleich herangezogen.

Wang et al. [20] verwenden zur Gesichtserkennung sogenannte Punktsignaturen (point signatures) im 3D in Kombination mit Gaborfiltern in 2D. Die Merkmale aus 2D und 3D bilden zusammen einen Merkmalsvektor. Die Klassifikation der Merkmalsvektoren wurde mit Hilfe einer Support Vector Machine durchgeführt und erzielte Erkennungsraten von über 90% [21].

Ein weiterer, auf der Segmentierung mit Hilfe von Krümmungen basierender Ansatz verwendet Extended Gaussian Images (EGI) zur Repräsentation eines Gesichts. Lee und Milios [22], [21] finden unter Verwendung der

Vorzeichen der Mittleren und der Gaussskrümmung konvexe Regionen in Tiefenbildern und erzeugen für jede dieser Regionen ein EGI. Der Vergleich von Regionen erfolgt dann durch den Vergleich der EGIs. Die Ähnlichkeit eines Gesichts im Ganzen zu einem anderen wird mit Hilfe eines Graph Matching Algorithmus' festgestellt. In [23] werden ebenfalls EGIs verwendet. Das vorgestellte Verfahren erreichte bei 37 Testbildern eine 100%ige Erkennung [21].

4 Merkmalsbasierte Erkennung

4.1 Merkmale

Merkmale sind interessante Punkte oder Bereiche innerhalb eines Bildes. Sie weisen Eigenschaften auf, die sie von uninteressanten Punkten abgrenzen. In wieweit ein Punkt als interessant eingestuft werden kann, hängt allein von der Problemstellung ab, für deren Lösung die interessanten Punkte herangezogen werden sollen.

Es existiert eine große Anzahl an Merkmalen, die für ein (Tiefen-)Bild festgestellt werden können, z. B.

- 2D-/3D-Punkte
- Regionen
- Kanten
- durchschnittliche Grau- bzw Farbwerte
- Kontrast
- usw.

Innerhalb der merkmalsbasierten Gesichtserkennung sind vor allem jene Merkmale interessant, die allgemein hin als Landmarken bezeichnet werden: Punkte, die markante Bereiche im Bild bezeichnen. Weitere für den Vergleich von Gesichtern geeignete Merkmale lassen sich aus einer Regionenorientierten Segmentierung des Ausgangsbildes extrahieren. Regionen besitzen eine große Vielfalt an Eigenschaften, die für einen Vergleich herangezogen werden können:

- Fläche

- Schwerpunkt
- Formfaktor (Verhältnis von Umfang zu Fläche)
- Exzentrizität (das Verhältnis der längsten Sehne innerhalb einer Region zur Sehne quer dazu (Hauptachse zu Nebenachse))
- Länge der Kontur
- Ferret Box (kleinstes achsenparalleles Rechteck, welches die Region vollständig enthält)
- Aspect Ratio (das Verhältnis von Breite:Höhe der Ferret Box)
- kleinstes umgebendes Rechteck (MBR, minimum bounding rectangle - kleinstes, meist nicht achsenparalleles Rechteck, dass die Region vollständig enthält)
- Ausdehnung der Region (Verhältnis der Kantenlängen des MBR)

Merkmale, die zur Gesichtserkennung verwendet werden sollen, müssen, um gute Ergebnisse erzielen zu können, folgende Eigenschaften aufweisen:

- Eindeutigkeit

Merkmale sollten Eigenschaften besitzen, die sie unterscheidbar von anderen Merkmalen der Nachbarschaft machen. Ein Merkmalspunkt hebt sich von anderen Punkten ab, weil er Informationen repräsentiert, die besser für eine Problemlösung geeignet sind, als die der benachbarten Punkte.

- Stabilität

Die Eigenschaften der Merkmale sollten sich bei leichten Variationen der Szene nicht ändern. Merkmale sollten invariant gegenüber Translation, Skalierung und Rotation sein.

- Ortstreue

Die Lage des Merkmals sollte mit der Position der zugrundeliegenden Bildstruktur eng verknüpft sein

- Robustheit

Unempfindlichkeit gegenüber Qualitätsunterschieden in den verwendeten Bildern. Ein Merkmal muss zum Beispiel auch in stark veräuschten Bildern auffindbar sein, um für die Gesichtserkennung bzw. Objekterkennung geeignet zu sein.

- Einfachheit

Es sollten nur wenige und leicht bestimmbare Parameter vorkommen. Die Berechnung von Merkmalen darf den Prozess der Erkennung nicht unnötig in die Länge ziehen.

- Dichte

Merkmale sollten möglichst in allen zu bearbeitenden Szenen in hinreichender Dichte auftreten. Es macht keinen Sinn, Merkmale für Erkennungsverfahren zu verwenden, die sich nicht eindeutig für jedes Objekt bestimmen lassen.

4.2 Merkmalsdetektion

Bei der merkmalsbasierten Gesichtserkennung liegt ein besonderer Schwerpunkt auf der Detektion und Extraktion von Merkmalen aus einem Gesichtsbild. Die Verwendung von Merkmalen beinhaltet eine starke Reduktion der Ausgangsdaten. Da die Gesichtserkennung lediglich auf den extrahierten Merkmalen erfolgt und die restlichen Informationen über das Gesicht ignoriert werden, ist eine möglichst genau Lokalisierung der Merkmale unbedingt

nötig.

Die folgenden Kapitel geben einen kurzen Überblick über Verfahren zur Merkmalsdetektion bzw. -extraktion.

4.2.1 Curvatures

Innerhalb eines Tiefenbildes lassen sich markante Punkte nicht wie bei einem einfachen Grauwert- oder Farbbild durch starke Helligkeits- bzw. Farbbunterschiede lokalisieren. Zur Merkmalsdetektion in 3D-Tiefenbildern ist es daher sinnvoll, die Form der Oberfläche des betrachteten Objekts zu berücksichtigen.

Zur Segmentierung von Tiefenbildern sind besonders gut Curvatures geeignet. Mit ihrer Hilfe lassen sich die Eigenschaften von gekrümmten Flächen sehr genau beschreiben. Die Verwendung von Curvatures ist vor allem sinnvoll, weil sie unabhängig vom Blickpunkt sind.

Eine beliebige Oberfläche lässt sich mit Hilfe der HK-Segmentierung in Bereiche mit gleicher Oberflächenbeschaffenheit unterteilen. HK steht für die Verwendung der Mittleren Krümmung H sowie der Gausskrümmung K . Oberflächen können aus sechs verschiedenen Formen bestehen, die anhand der Vorzeichen der mittleren und der Gausskrümmung klassifiziert werden: konkav elliptisch, konvex elliptisch, hyperbolisch, konkav zylindrisch, konvex zylindrisch oder planar (vgl. Abbildung 3 und Tabelle 1).

Die Mittlere Krümmung sowie die Gausskrümmung werden berechnet aus der maximalen und der minimalen Hauptkrümmung (κ_1, κ_2):

$$H = -\frac{\kappa_1 + \kappa_2}{2}$$

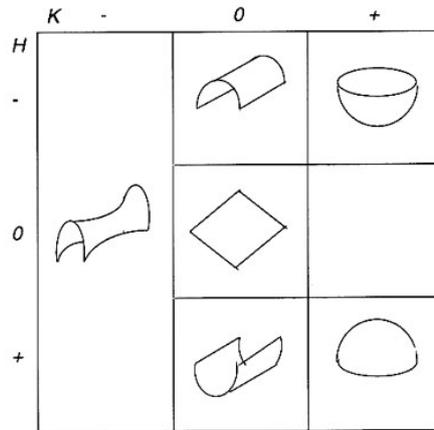


Abbildung 3: Formklassen für 3D-Oberflächen, [1]

K	H	lokale Formklasse
0	0	planar
0	+	konkav zylindrisch
0	-	konvex zylindrisch
+	+	konkav elliptisch
+	-	konvex elliptisch
-	beliebig	hyperbolisch

Tabelle 1: Klassifizierung von Obeflächenformen anhand der Vorzeichen der HK-Krümmung

$$K = \kappa_1 \kappa_2$$

Informationen zu den Hauptkrümmungen finden sich in (Anhang B).

Innerhalb eines Tiefenbildes h erfolgt die HK-Segmentierung an einem Punkt

P mit

$$K = \frac{h_{xx}h_{yy} - h_{xy}^2}{(1 + h_x^2 + h_y^2)^2} \quad (1)$$

und

$$2H = \frac{(1 + h_x^2)h_{yy} - 2h_x h_y h_{xy} + (1 + h_y^2)h_{xx}}{(1 + h_x^2 + h_y^2)^{3/2}} \quad (2)$$

wobei die Indizes partiellen Ableitungen entsprechen.

Für die Einordnung einer lokalen Region einer Oberfläche in eine der

oben genannten Formklassen genügt also die Berechnung obiger Formeln und das anschließende Zuordnen der Form entsprechend der Vorzeichen der Ergebnisse.

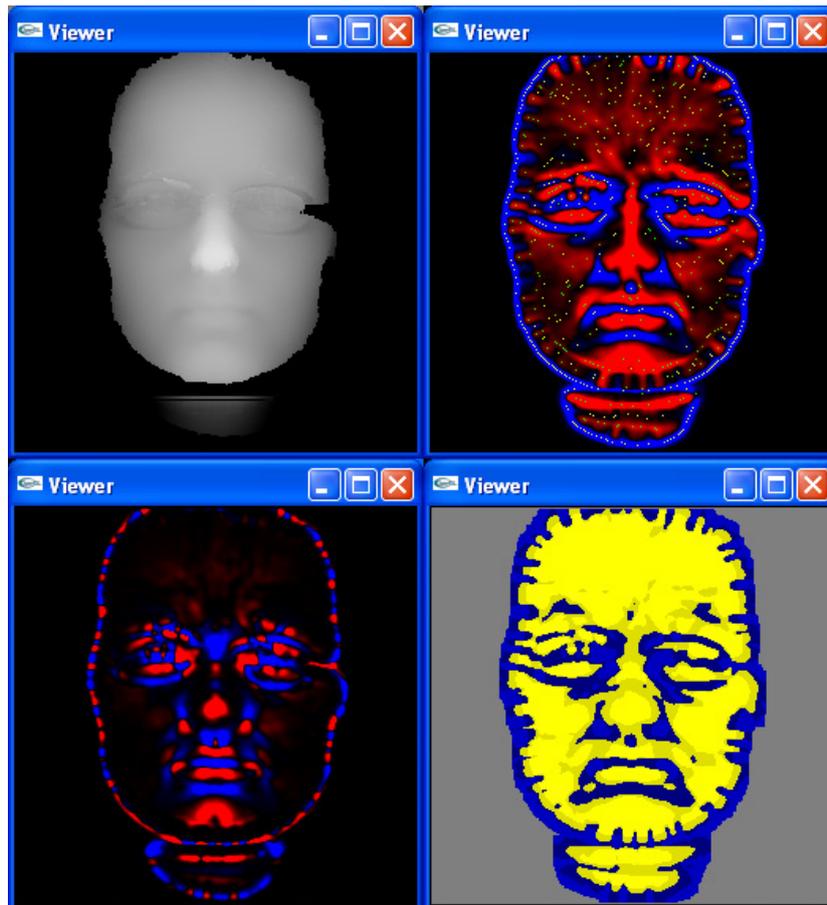


Abbildung 4: Verschiedene Segmentierungsergebnisse basierend auf Krümmungseigenschaften (links oben: Tiefenbild, rechts oben: Segmentierung anhand Mittlerer Krümmung, rot = konvex, blau = konkav, links unten: Segmentierung durch Gauss-Krümmung, rot = konvex, blau = konkav, rechts unten: HK-Segmentierung, blau = konkav, gelb = konvex)

Die Berechnung der Krümmungseigenschaften kann durch das Auftreten von Rauschen beeinflusst werden. Daher ist es sinnvoll, im Vorfeld der

Segmentierung eine Glättung durchzuführen, um möglichst gute Ergebnisse zu erzielen. Bei Filtern, die mit Durchschnittswerten arbeiten, wie der Gauss-Filter, ist allerdings zu beachten, dass diese dazu neigen, hohe und möglicherweise wichtige Krümmungswerte stark abzuschwächen, während an anderer Stelle schwache und unwichtige Krümmungen stärker hervorgehoben werden.

Ein weiteres zu beachtendes Problem ist die Klassifizierung planarer Oberflächen. Für planare Flächenbereiche gilt $H = K = 0$. Dieses Ergebnis jedoch kann mit den oben genannten numerischen Berechnungen nicht erreicht werden, so dass es nötig ist, einen Schwellwert für H und K einzurichten, um festzulegen, wann ein Ergebnis nah genug an Null ist, um die dazugehörige Fläche als planar einzustufen.

Nach der erfolgten Segmentierung eines Tiefenbildes in konvexe und konkave Regionen werden diese verwendet, um wichtige Merkmale, die sich zum Vergleich heranziehen lassen, in dem jeweiligen Gesicht zu finden.

Verschiedene Eigenschaften helfen, uninteressante Merkmals-Kandidaten auszuschließen [13]:

- die Größe einer Region auf der Oberfläche
- der Abstand einer Region oder eines Punktes von der Symmetrieachse, wobei diese Achse durch die Position und Lage des Nasenrückens definiert wird
- die Symmetrieeigenschaften einiger Merkmale (das Vorkommen eines Merkmals in beiden Gesichtshälften)
- die Nähe eines Punktes oder einer Region zu einer bestimmten Position

auf der Oberfläche

- das Ausmaß, in dem eine Region oder ein Punkt aus seiner umgebenden Oberfläche herausragt

Am einfachsten zu detektieren ist mit großer Wahrscheinlichkeit die Position der Nase, bzw. der Nasenspitze. Sie besitzt bei jedem Menschen die Eigenschaft, dass sie am weitesten aus dem Gesicht herausragt. Um die Suche nach der Nasenspitze einzuschränken, empfiehlt sich also die Betrachtung der Tiefenwerte. Für jede Reihe des Tiefenbildes lässt sich der maximale Tiefenwert bestimmen. Vor allem im Bereich der Nasenspitze und entlang des Nasenrückens sind die entsprechenden Punkte entlang der Mittelachse des Bildes angesiedelt (ausgehend von einer Frontalansicht). Durch Eintragen dieser Werte in ein Histogramm findet sich die Position dieser Mittellinie. Die weitere Suche nach der Nasenspitze kann anschließend auf eine Suche entlang dieser Linie reduziert werden [37]. Die Betrachtung des vertikalen z -Profils liefert mehrere Kandidaten, zum Beispiel für die Lippen, die Stirn und auch die Nasenspitze. Für die finale Auswahl der Nasenspitze werden die zeilenweisen z -Profile für die Kandidaten herangezogen. Der Punkt, in dessen Umgebung die z -Werte am stärksten variieren bzw. abfallen, wird als Nasenkandidat identifiziert. Die Nasenspitze ist dann der Punkt innerhalb der konvexen Region aus der lokalen Umgebung, der den größten z -Wert aufweist.

Die so gefundene Nasenspitze bildet den Ausgangspunkt für die Detektion der Nasenwurzel. Die Nasenwurzel, also der Bereich oberhalb der Nase, etwa auf Höhe der Augen, lässt sich leicht detektieren. Es erfordert lediglich die Suche nach einer konkaven Region oberhalb der Nasenspitze entlang der Mittellinie. Die Nasenspitze und -wurzel bilden zwei Punkte auf einer Linie, die als ungefähre Symmetrieachse des Gesichts betrachtet werden kann. Die

Existenz einer exakten Symmetrie in einem Gesicht bezüglich der Nase ist nahezu auszuschließen.

Wie bereits in Kapitel 3.3 erwähnt, ist die Nase ein sehr zuverlässiges Merkmal im menschlichen Gesicht, da sie nicht durch Gesichtsausdrücke in ihrer Position veränderlich ist. Aus diesem Grund lassen sich die Nasenspitze und die Nasenwurzel sowie die durch sie beschriebene Symmetrieachse dazu verwenden, die Suche nach weiteren Merkmalen zu vereinfachen.

Die durch die HK-Segmentierung gebildeten Regionen eignen sich nur dann zum Vergleich, wenn sie konsistent in allen Gesichtern gefunden werden können. Zu diesen gehören vor allem jene, die als symmetrische Paare auftreten, wie zum Beispiel die konkaven Bereiche an den äußeren und inneren Augenwinkeln. Konvexe Regionen, die in beiden Gesichtshälften auftreten, sind zum Beispiel die Augäpfel, bzw. die Augenlider (bei geschlossenen Augen). Allerdings ist die Segmentierung der Augäpfel nicht sehr verlässlich, da an Stelle der Augen in Tiefenbildern oft Löcher auftreten, die die Segmentierungsergebnisse beeinflussen.

Die Suche nach einem bestimmten Merkmal kann erheblich vereinfacht werden, wenn es sich in einer charakteristischen Lage bezüglich der Symmetrieachse bzw. Symmetrieebene befindet, oder wenn es für dieses Merkmal ein Gegenstück auf der anderen Gesichtshälfte gibt.

Für das Auffinden zweier symmetrischer Merkmale, zum Beispiel der inneren Augenwinkel, werden zuerst alle Regionen, die diesem Oberflächentyp (hier: konkav) entsprechen, betrachtet. Die Suchmenge lässt sich mit verschiedenen Einschränkungen verkleinern. Wenn Angaben für die typischen Maße einer bestimmten Region vorhanden sind, kann man aufgrund dessen all jene Regionen ausschließen, die die Grenzwerte für die Größe überschreiten. Eine weitere Möglichkeit zur Reduktion von Kandidaten ist die Lage der Re-

gionen. Für ein Merkmal aus dem Bereich der Augen macht es wenig Sinn, die Regionen unterhalb der Nasenspitze zu berücksichtigen.

Die einzelnen Regionen werden am einfachsten durch ihren jeweiligen Schwerpunkt repräsentiert. Für jeden noch in der Suchmenge vorhandenen Punkt werden zwei Maße berechnet: der Abstand d_S zur Symmetrieachse und die Projektion p_s des Punktes auf die Symmetrieachse.

Das eigentliche Auffinden von symmetrischen Regionen- bzw. Punkt-Paaren durchläuft mehrere Schritte. Zuerst werden alle Regionen eines Typs (konkav oder konvex) mit allen anderen des gleichen Typs verglichen. Dabei wird geprüft, ob ein gewisser Grad an Symmetrie zwischen den jeweils zwei Regionen existiert. Dies erfolgt durch den Vergleich von d_S und p_S . Liegen diese Werte innerhalb eines bestimmten Rahmens, wird die Region der einen Gesichtshälfte auf die andere Gesichtshälfte projiziert und der Abstand zwischen der Projektion und der zu prüfenden Region berechnet. Die geringste Distanz lässt auf ein gefundenes Paar schließen. Regionen, für die keine Symmetrien festgestellt werden können, werden verworfen und nicht länger als Kandidaten für symmetrische Paarungen betrachtet.

Da Gesichter in der Regel ein gewisses Maß an Asymmetrie aufweisen, können nicht alle Regionen symmetrische Gegenstücke besitzen. Es kommt daher vor, dass zu einigen Regionen mehrere Gegenstücke gefunden werden. In diesem Fall ist die Paarung nicht für jede Region wechselseitig korrekt. Eine Region, für die ein symmetrisches Gegenstück gefunden wurde, für das allerdings eine andere, bessere Paarung besteht, wird in eine erneute Suche nach Paaren aufgenommen, in der nur solche Regionen betrachtet werden, für die wechselseitig in der ersten Suche keine Übereinstimmung gefunden wurde.

Das Ergebnis ist eine Liste, die jeweils alle Paarungen konkaver oder kon-

vexer Regionen enthält. Um diese Paarungen bestimmten Merkmalen des Gesichts zuzuordnen, wird ihre jeweilige Lage zur Symmetrieachse, zur Nase oder auch zu anderen bereits zugeordneten Merkmalen betrachtet. Die Abbildungen 5 und 6 zeigen Ergebnisse der Merkmalsdetektion auf der Basis von Symmetrieeinformationen. Die Tiefenbilder wurden mit der Gauss Krümmung in konvexe und konkave Regionen segmentiert. Die starke Glättung in Abbildung 6 bewirkt eine Reduktion der auffindbaren Paarungen.

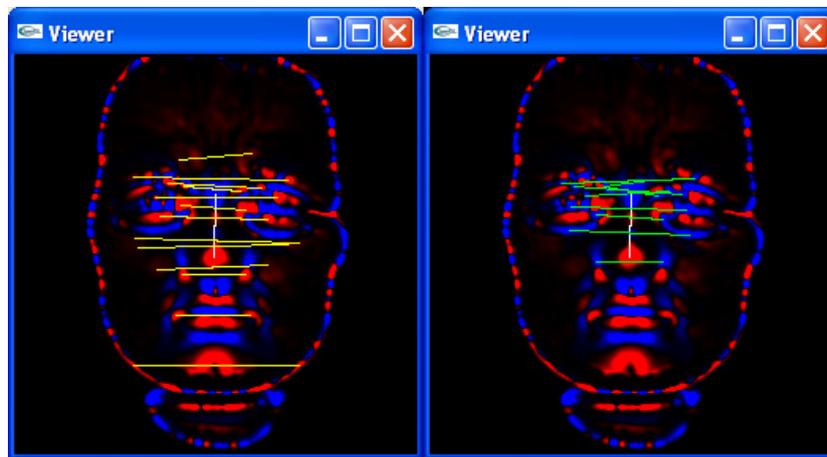


Abbildung 5: Auf Basis von Gesichtssymmetrien detektierte Merkmalspaarungen in einem mit Gausser Krümmung segmentierten Tiefenbild (weiß = Symmetrieachse, gelb = Paarungen konvexer Regionen, grün = Paarungen konkaver Regionen)

Regionen bieten eine Vielzahl an Vergleichsmöglichkeiten, die bereits in Kapitel 4.1 beschrieben wurden. Zusammen mit den jeweiligen Distanzen zwischen ihnen lassen sich umfangreiche Merkmalsvektoren erstellen, anhand derer unterschiedliche Gesichter verglichen werden können. Die Schwerpunkte der Regionen lassen sich direkt als Landmarken auffassen und können zum Beispiel in einem Graph Matching Verfahren (vgl. Kapitel 4.3.3) verwendet werden.

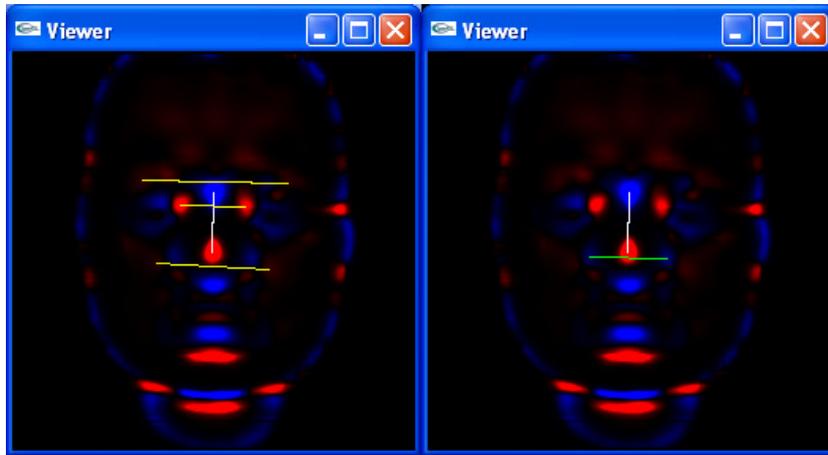


Abbildung 6: Auf Basis von Gesichtssymmetrien detektierte Merkmalspaarungen in einem mit Gaußscher Krümmung segmentierten, stark geglätteten Tiefenbild (weiß = Symmetrieachse, gelb = Paarungen konvexer Regionen, grün = Paarungen konkaver Regionen)

4.2.2 Wavelets

Die Verwendung von Wavelets für die Merkmalsextraktion ist vor allem biologisch zu begründen. Wavelets, darunter zum Beispiel Gabor Wavelets, modellieren das menschliche Sehen. Gaborfilter zeigen ähnliche Reaktionen wie die Rezeptivfelder des visuellen Cortex im menschlichen Gehirn.

Wie Wavelets „sehen“, wird in [34] beschrieben: im Wavelet-Modell des Sehvorgangs ist das Bild vor unseren Augen das Signal. Die Wavelets übernehmen die Rolle der Rezeptivfelder, auf die die Neuronen der Sehrinde ansprechen. Diese haben unterschiedliche Größen und Neigungen: verschiedene Neuronen sprechen auf unterschiedliche Rezeptivfelder an. In der Signalverarbeitung werden Signale nach verschiedenen großen Wavelets zerlegt. Ähnlich wird das, was das menschliche Auge sieht, nach unterschiedlichen Rezeptivfeldern zerlegt.

Der Wavelet-Koeffizient entspricht der Reaktion eines Neurons auf das Signal. Er wird sehr groß, wenn ein Neuron mehrfach und stark reagiert,

während bei einer schwachen Reaktion des Neurons der Koeffizient klein ist. Wie bei der Wavelet-Transformation kodieren kleine Rezeptivfelder hohe Frequenzen und große Rezeptivfelder niedrige Frequenzen.

Gabor Wavelets eignen sich gut zur Merkmalsextraktion, da man mit ihrer Hilfe Informationen über die lokale Richtung und Frequenz einer Bildstruktur aus Bildern gewinnen kann. Ein besonderer Vorteil von Wavelets liegt darin, dass sie relativ unempfindlich auf konstante Grauwertverschiebungen reagieren.

Das Gabor Wavelet ist eine Welle, die von einer Gauss-Fensterfunktion begrenzt wird. Das Gabor-Mutterwavelet hat folgende Form (vgl. [30]):

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right) \left[\exp(i\vec{k}_j \vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (3)$$

Aus dieser Darstellung lassen sich alle Wavelets für die verschiedenen Orientierungen und Frequenzen ableiten. Dies lässt sich über \vec{k} steuern.

$$k_{\nu,\mu}^{\vec{}} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_\nu \cos \phi_\mu \\ k_\nu \sin \phi_\mu \end{pmatrix}$$

ν indiziert die verschiedenen verwendeten Frequenzen, während μ die unterschiedlichen Orientierungen anzeigt. Entsprechend ist k_ν die Frequenz und ϕ_μ die Orientierung bzw. der zugehörige Winkel. Die Breite σ/k der Gaussfunktion wird mit σ kontrolliert. In der Objekterkennung hat sich die Verwendung von 5 Frequenzen ($\nu = 0, \dots, 4$) und 8 Orientierungen ($\mu = 0, \dots, 7$) als ausreichend erwiesen. Daher gilt $j = \mu + 8\nu$, $k_\nu = 2^{\frac{\nu+2}{2}} \pi$ und $\phi_\mu = \mu \frac{\pi}{8}$.

Ein komplexer Gabor-Filter besteht aus einem Realteil (cos-Anteil) und einem Imaginärteil (sin-Anteil) (vgl. Abbildung 7). Der Realteil zeigt die Re-

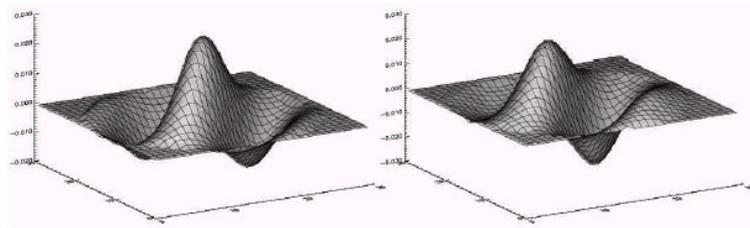


Abbildung 7: Gaborfilter: Realteil (links) und Imaginärteil (rechts), [29]

aktion auf Bildstrukturen mit gerader Symmetrie, während am Imaginärteil die Reaktion auf ungerade Symmetrien ablesbar ist. Auch innerhalb des Sehprozesses gibt es diese Zweiteilung. Zwei verschiedene Arten von Zellen reagieren entweder auf gerade oder auf ungerade Symmetrien. Dieses Verhalten wird durch die \cos - und die \sin -Funktionen imitiert.

Durch das Filtern eines Bildes mit einem Gabor-Filter wird dieses in eine Darstellung aus komplexen Komponenten überführt, die sich als Punkte im Polarkoordinatensystem auffassen lassen [31]. Der Betrag der Komponenten liefert die Amplituden und der Winkel zwischen beiden die Phasen. Die Amplitudendarstellung ist eine glatte Funktion und ist relativ robust gegenüber Translation und Rotation. Die Phase dagegen zeigt schon bei geringen Positionsänderungen starke Schwankungen (vgl. Abbildung 8).

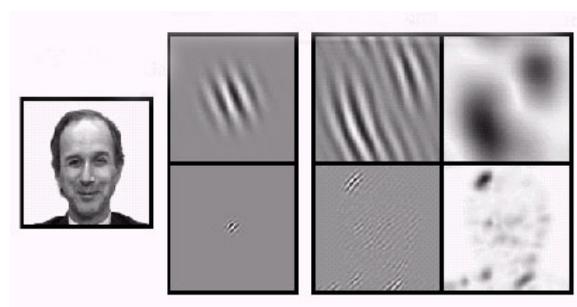


Abbildung 8: Gabor Wavelets: Originalbild (links), Gabor Wavelets (mitte), Ergebnisse der Faltung (links: Imaginärteil, rechts: Realteil)(rechts) [30]

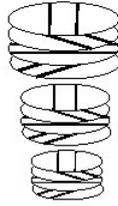


Abbildung 9: Jet (drei Frequenzen, vier Orientierungen) [31]

Die komplexen Antworten des Gabor-Filters lassen sich zu einem 40-dimensionalen Vektor zusammenfassen (bei zuvor gewählten 5 Frequenzen und 8 Orientierungen). Dieser wird auch als Jet bezeichnet. In Polardarstellung haben Jets die Form

$$J_j = a_j \exp(i\phi_j)$$

Jets beschreiben die lokalen Frequenzinformationen um eine Landmarke herum. Sie werden durch Faltung eines Bildes an der Position der Landmarke mit den Wavelets erzeugt. Abbildung 9 zeigt die symbolische Darstellung eines Jets, der aus Waveletes mit drei Frequenzen und vier Orientierungen entsteht.

Die Jets wirken als Repräsentanten ausgesuchter (Gesichts-)Merkmale auf mehreren Auflösungssebenen. Das bedeutet, dass sie eine große Menge an Informationen zu einem einzigen Punkt bieten. Dies macht sie für die Objekt- bzw. Gesichtserkennung so interessant.

4.2.3 Interestoperatoren

In der Merkmalsextraktion in 2D-Bildern finden oft die sogenannten Interestoperatoren Einsatz. Da diese allerdings Punkte auf Basis der Intensität der Pixel eines Bildes liefern, sind die 2D-Interestoperatoren als solche für die

Anwendung auf Tiefenbilder denkbar ungeeignet, weil Tiefenbilder in der Regel keine oder kaum signifikante Intensitätswechsel aufweisen, die auf Ecken oder Kanten hindeuten könnten. Dennoch können 2D-Interestoperatoren verwendet werden, um die Qualität der Merkmalsextraktion aus Tiefenbildern zu verbessern. Ein Beispiel für einen 2D-Interestoperator ist der Harris-Corner-Detektor.

Bei der Helligkeits- bzw. Intensitäts-basierten Detektion werden Veränderungen der Grauwerte um einen Punkt durch Verschieben eines Analysefensters in verschiedene Richtungen um diesen Punkt ermittelt [43]. Ob an einer Stelle eine Ecke vorliegt, hängt von der Nachbarschaft der Stelle ab. Weist der betrachtete Punkt einen Wert auf, der sich über einen Schwellwert hinaus von den Werten der Punkte seiner Umgebung abhebt, kann dieser Punkt als Ecke bewertet werden.

Der Harris-Detektor ist eine Erweiterung des Ansatzes von Moravec, [43]. Für jeden Bildpunkt wird dessen Umgebung betrachtet und die Intensitätsänderung bei allen möglichen Drehungen über dem Ausschnitt berechnet. Bei geringer Änderung liegt kein besonderes Merkmal wie eine Ecke oder eine Kante vor. Ein große Änderung in einer Richtung weist auf eine Kante hin, während bei einer Ecke das Minimum aller Änderungen groß ist. Als zusätzliche Ecken-Indikatoren werden auch die Richtung und die Variation der Änderung in die Berechnung mit einbezogen.

Ob eine Ecke vorliegt, lässt sich ermitteln mit Hilfe der Kovarianzmatrix. Diese hat folgende Form:

$$M = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

I_x und I_y sind die Gradienten in x- und y-Richtung. Die Eigenwerte dieser Matrix geben Aufschluss darüber, inwiefern ein Punkt als Ecke oder Kante eingestuft werden kann. Sind beide Eigenwerte groß, kann von einer Ecke

ausgegangen werden, während nur ein großer Eigenwert auf die Existenz einer Kante hindeutet. Die Eigenwerte sind also ein Maß für das Vorliegen von Kanten in zwei orthogonal zu einander stehenden Richtungen.

Die Anwendung des Spektralsatzes [43] erlaubt die Rotation der Matrix und somit die Prüfung auf „Eckigkeit“ in allen Richtungen. Die Bewertung der „Eckigkeit“ erfolgt über

$$e = a \cdot b - k(a + b)$$

bzw.

$$e_{Harris} = (I_x^2 \cdot I_y^2 - (I_x I_y)^2) - k \cdot (I_x^2 + I_y^2)^2$$

wobei a und b die Eigenvektoren sind. $a \cdot b$ entspricht der Determinanten von M und $a + b$ der Spur von M .

Wie bereits erwähnt, können 2D-Interestoperatoren helfen, geeignete Merkmale in Tiefenbildern zu finden. Dies wird möglich durch die zusätzliche Verwendung des dem Tiefenbild entsprechenden Intensitätsbildes. Die Gesichtserkennungsverfahren, die sich sowohl der 2D- als auch der 3D-Merkmalsextraktion bedienen, werden auch als multimodal bezeichnet. In [20] werden zum Beispiel die sogenannte Point Signatures mit den Antworten eines Gabor-Filters kombiniert. Die Verbindung von ShapeIndex und 2D-Interestoperatoren findet Einsatz in [37].

Die Ergebnisse der einzelnen Detektions- bzw. Segmentierungsverfahren in 2D und 3D lassen sich zum Beispiel durch die Summen-Regel (sum rule) zusammenfassen ([37])

$$F(p) = (1 - S'(p)) + C'(p)$$

Dabei ist S' das normalisierte Ergebnis des ersten verwendeten Verfahrens und C' das des zweiten. Normalisiert bedeutet in diesem Fall, dass die Wer-

te in Abhängigkeit ihrer lokalen Umgebung, der Region, in der nach dem Merkmal gesucht wurde, dargestellt werden:

$$S'(p) = \frac{S(p) - \min\{S_i\}}{\max\{S_i\} - \min\{S_i\}}$$

Der Punkt mit dem höchsten Ergebnis für $F(p)$ innerhalb einer Suchregion kann als Merkmalspunkt akzeptiert werden.

4.2.4 Histogramme

Für die Detektion von Merkmalen in Tiefenbildern ist die Betrachtung der lokalen Verteilung von Tiefenwerten interessant. Mit Hilfe von Histogrammen lassen sich diese bezüglich der Spalten oder Zeilen eines Bildes darstellen.

Die vertikale Projektion bildet die durchschnittlichen Tiefenwerte eines betrachteten Bildbereichs pro Spalte ab, während die horizontale Projektion das Bild zeilenweise abtastet. Maxima bzw. Minima in den Histogrammdarstellungen weisen auf entsprechende Strukturen im Tiefenbild hin. Für die Nasenspitze ist beispielsweise in beiden Histogrammen ein ausgeprägtes Maximum zu erwarten.

Die zeilen- bzw. spaltenweisen Projektionen werden berechnet durch

$$H(k) = \frac{1}{J} \sum_{j=1}^J I(j, k)$$

und

$$V(j) = \frac{1}{K} \sum_{k=1}^K I(j, k)$$

Wie bereits erwähnt, ist die Nasenspitze relativ einfach mit Hilfe von Histogrammen zu lokalisieren. Für andere Merkmale wie die Augenwinkel oder Mundwinkel ist es allerdings schwierig, die zugehörigen Minima oder Maxima in den Histogrammen zu finden. Daher macht es Sinn, den Suchbereich

innerhalb der Histogramme auf Basis bereits gefundener Merkmale einzuschränken, da die Lage der einzelnen Merkmale zueinander in der Regel bekannt ist.

Die Genauigkeit der mit Histogrammen gefundenen Merkmale hängt von der Qualität des verwendeten Tiefenbildes ab. Rauschen beeinflusst die Lage von lokalen Extremstellen im Histogramm und kann das Detektionsergebnis verfälschen. Die histogrammbasierte Merkmalsextraktion ist nicht für alle Merkmale gleich gut geeignet. Bei Brillenträgern zum Beispiel werden die Histogramme der Augenregionen kaum ausgeprägte Maxima oder Minima aufweisen. An dieser Stelle sind andere Verfahren zur Merkmalsdetektion robuster.

Bei der Betrachtung der Tiefenwerte ist das Histogrammverfahren nicht zuverlässig genug. Die gefundenen Extrema lassen sich allerdings als Ausgangspunkt für andere Verfahren verwenden, in dem auf der Basis der lokalen Extrema relativ kleine lokale Bereiche festgelegt werden, in denen mit anderen, möglicherweise aufwändigeren, Verfahren die Merkmalsdetektion fortgesetzt bzw. ergänzt und verfeinert werden kann.

Histogramme eignen sich nicht nur zur Darstellung der durchschnittlichen Tiefenwerte. Auch die Verteilung von Krümmungswerte in einem HK-segmentierten Tiefenbild lässt sich mit ihnen abbilden. Die Analyse der Histogramme besteht dann aus der Suchen nach Extremwerten in den Krümmungen.

4.3 Merkmalsvergleich

Das Resultat der Merkmalsdetektion und -extraktion ist ein Merkmalsvektor, der sämtliche gefundenen Merkmale vereint. Je nach Verfahren enthält

dieser Vektor unterschiedlich viele Komponenten. Um den Grad der Ähnlichkeit zwischen zwei Merkmalsvektoren festzustellen, müssen die einzelnen Komponenten verglichen werden. Ein direkter Vergleich der einzelnen Merkmale ist allerdings nicht immer sinnvoll, da nicht alle im Merkmalsvektor enthaltenen Merkmale von gleicher Wichtigkeit sind. Ein Verfahren, das die Aussagekraft eines Merkmals überprüft, kann dieses Problem eingrenzen. Oftmals ist ein Merkmalsvektor auch viel zu groß, das heißt er enthält viel zu viele Daten für einen direkten Vergleich. In diesem Fall empfiehlt sich die Reduktion der Dimension des Vektors.

Ein Überblick über die Möglichkeiten des Vergleichs von Merkmalsvektoren liefern die folgenden Kapitel.

4.3.1 Ähnlichkeitsmaße

Ähnlichkeitsmaße oder Distanzfunktionen beschreiben den Grad der Übereinstimmung von Vektoren, die eine Folge von Messwerten darstellen. Als Distanzfunktionen lassen sich verschiedene Metriken verwenden. Dieses Kapitel gibt einen kurzen Überblick über die gängigsten Ähnlichkeitsmaße.

- Manhattan-Distanz

Die Manhattan-Distanz ist definiert durch

$$d(x, y) = \sum_{i=1}^N |x_i - y_i| \quad x, y \in \mathfrak{R}^N$$

- Euklidischer Abstand

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad x, y \in \mathfrak{R}^N$$

Der Euklidische Abstand ist die mathematische Definition des normalen Abstands. Er liefert den Abstand für zwei Vektoren oder Punkte durch Berechnung der euklidischen Norm des Differenzvektors zwischen den beiden Punkten. Für zwei Punkte $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ gilt also

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

Die Euklidische Distanz ist translationsinvariant:

$$d(x, y) = d(x + \nu, y + \nu)$$

- Mahalanobis Distanz

$$d(x, y) = (x - y)^T S^{-1} (x - y)$$

Die Mahalanobis Distanz berechnet die Distanz zwischen Punkten im mehrdimensionalen Vektorraum. Dabei wird die Korrelationsstruktur der Vektoren berücksichtigt. Sie gibt den Abstand eines Beobachtungswertes einer mehrdimensionalen Normalverteilung vom Mittelwertvektor dieser Verteilung an. Vor allem in der Statistik und in multivariaten Analyseverfahren findet die Mahalanobis Distanz Anwendung. Sie wird zum Beispiel in der Diskriminanzanalyse dazu verwendet, einen Punkt zu einer bestimmten Population zuzuordnen.

In der oben angegebenen Formel entspricht S der Kovarianzmatrix der beiden Punkte bzw. Vektoren x und y . Die Mahalanobis Distanz ist skaleninvariant und translationsinvariant. Ein Nachteil der Mahalanobis Distanz ist, dass die Kovarianzmatrix durch Ausreißer verzerrt sein kann.

4.3.2 Registrierung

Bei dem Vergleich von Landmarken aus zwei verschiedenen Bildern muss davon ausgegangen werden, dass die Gesichter während der Extraktion der Merkmale nicht exakt gleich ausgerichtet waren, so dass eine bestimmte Transformation (Rotation und Translation) die resultierenden Merkmalsmengen voneinander unterscheidet. Um genaue Vergleichswerte zu erhalten, ist es sinnvoll, diese Transformation im Vorfeld zu berechnen. Ein beliebtes Verfahren dafür ist die Iterative Closest Point Methode (ICP).

Das ICP-Verfahren setzt sich aus vier Schritten zusammen [24]:

1. Finden von Punktkorrespondenzen
2. Berechnen der Rotation und der Translation
3. Anwenden der Transformation auf die entsprechende Punktmenge
4. Prüfen des Abbruchkriteriums, falls nicht erfüllt weiter mit neuen Punktkorrespondenzen bei Schritt 1

Die eigentliche Berechnung der Transformation erfolgt in Schritt 2 mit Hilfe der Singulärwertzerlegung. Zur Vereinfachung werden die einzelnen 3D-Punkte p_i bzw. y_i (Modellmenge und Testmenge) in ihr jeweiliges Zentrum verschoben. Auf diesem Weg wird die Translation aus der Berechnung entfernt und vorerst nur die Rotation berechnet.

$$p_{c,i} = p_i - \bar{p}$$

$$\bar{p} = 1/N_p \sum_{i=1}^{N_p} p_i$$

Entsprechendes gilt für die Testmenge Y .

Es ergibt sich also folgendes Problem:

$$\min_R \sum_{i=1}^N \|y_{c,i} - Rp_{c,i}\|^2 \quad (4)$$

das sich durch Maximieren der $Spur(RK)$ lösen lässt. K ist definiert als

$$K = \sum_{i=1}^{N_p} y_{c,i} p_{c,i} 1^T$$

Die Singulärwertzerlegung von K ergibt $K = VDU^T$. Die optimale Rotationsmatrix, die die Spur maximiert, ist dann $R = VU^T$

Abschließend muss noch die Translation berechnet werden. Dies erfolgt durch

$$t = \bar{t} - R\bar{P}$$

Das ICP-Verfahren zeigt gute Eigenschaften, wenn in beiden zu vergleichenden Punktmengen für alle Punkte Punktkorrespondenzen vorhanden sind. In Bezug auf Ausreißer reagiert diese Methode allerdings sensibel.

Ein robusteres Verfahren ist der sogenannte RICP Algorithmus [24]. RICP steht für robustes ICP. Es besteht ebenfalls aus 4 Schritten, wobei sich der Ablauf im Vergleich zum ICP nur in Schritt 2 unterscheidet. Zuerst wird für die zuvor berechneten Punktkorrespondenzen eine initiale Rotation berechnet, was zum Beispiel mit Hilfe der QR-Zerlegung geschieht. Die entstehenden Residuen werden zur Berechnung von Gewichten für die nachfolgende Least Squares Regression benötigt. Jeder 3D-Punkt erhält ein Gewicht entsprechend der Gewichte seiner einzelnen Koordinaten. Ist eines davon null, dann beträgt auch die Gesamtgewichtung für diesen Punkt null und der zugehörige Punkt wird als Ausreißer betrachtet. Die gesuchte Rotationsmatrix wird anschließend mit den gewichteten Punkten wie im ICP-Verfahren berechnet.

4.3.3 Elastische Graphenanpassung

Die Elastische Graphenanpassung ist eine aus der Dynamic Link Architecture (DLA) [32] motivierte Methode zur Objekterkennung. Sie basiert auf

lokalen und geometrischen Merkmalen. Das Verfahren ist theoretisch für die Erkennung aller Objekttypen, die sich hinreichend durch geometrische oder textuelle Merkmale unterscheiden, anwendbar. Die Elastische Graphen Anpassung hat gegenüber dem Dynamischen Link Matching den Vorteil, mit einem geringeren Rechenaufwand auszukommen [31].

Die Dynamische Link Architektur versucht, bestimmte Gehirnfunktionen nachzubilden. Sie besteht aus mehreren Graphen, die jeweils aus vielen Knoten aufgebaut sind. Dies sind über Verknüpfungen (Links) miteinander verbunden. Der Grundgedanke der DLA ist, dass Verknüpfungen zwischen Knoten, die zusammengehören, also gleiche Merkmale repräsentieren, verstärkt werden, während die Verbindung von Knoten mit zusammenhangslosem Inhalt abgeschwächt oder aufgehoben werden. Wenn zwei Knoten stark korreliert sind, führt dies zu einer Stärkung der Verbindung zwischen ihnen. Dies wiederum hat die Erhöhung der Korrelation zwischen ihnen zur Folge. Das bedeutet, dass starke Verbindungen sich noch weiter verstärken. Analog schwächen sich schwache Verbindungen weiter ab.

In der Gesichtserkennung mittels DLA wird jedes Gesicht als DLA-Graph repräsentiert, wobei die Knoten gleichmäßig verteilten Bildpunkten des Gesichts entsprechen und Gesichtsmerkmale aus der Umgebung des jeweiligen Knoten enthalten. Der Bildgraph der zu erkennenden Person wird mit denen aus der Datenbank verglichen, wobei starke Verknüpfungen auf große Ähnlichkeit hindeuten.

Für das Elastic Bunch Graph Matching werden Objekte mit beschrifteten Graphen modelliert. Die Graphen sind von der Form des jeweiligen Objektes abgeleitet. Daher werden sie auch als objekt-adaptiert bezeichnet.

Für die Modellierung von Objekten in der Gesichtserkennung werden Ge-

sichtsgraphen (Face Graph) benutzt. Ein Graph enthält Knoten, die die besonderen Merkmale des Gesichts repräsentieren. Es können zum Beispiel Knoten für die Augen, die Nasenspitze oder die Mundwinkel existieren. Jeder Knoten wird mit einem Jet (vgl. 4.2.2) beschriftet, also mit den lokalen Merkmalen des Punktes, auf dem der Knoten liegt. Die Verbindung zwischen zwei Knoten liefert eine Kante, die mit dem Abstand der beiden Knoten beschriftet ist.

Der so entstandene Graph ist in der Regel nicht vollständig, da es nicht für jedes benachbarte Knotenpaar Sinn macht, die Distanz zu prüfen.

Die Gesichtsgraphen einer Klasse lassen sich zu einem Bündelgraphen zusammenfassen. Dabei ist wichtig, dass die Gesichtsgraphen alle die gleiche Struktur besitzen, also für die selben Punkte im Gesicht Knoten und eine identische Auswahl an Kanten enthalten. Der resultierende Bündelgraph besitzt ebenfalls diese Struktur. Die Knoten sind allerdings nicht mit einzelnen Jets beschriftet, sondern mit Bündeln (Bunches) der Jets der Gesichtsgraphen für jeden Knoten (vgl. Abbildung 11). Jedes Bündel enthält eine Menge von Jets, die ganz unterschiedliche Ausprägungen eines Merkmals repräsentieren können, also zum Beispiel aus Gesichtern verschiedenen Alters, unterschiedlichen Geschlechts, verschiedene Formen der Gesichtsbehaarung, Brillen, geöffnete oder geschlossene Augen, usw. So wird eine große Bandbreite von möglichen Variationen im Gesicht abgedeckt. Abbildung 10 verdeutlicht dieses Prinzip anschaulich. Die Kanten sind mit der durchschnittlichen Distanz der Kanten der Gesichtsgraphen beschriftet.

Die eigentliche Gesichtserkennung beinhaltet das Vergleichen von Gesichtsgraphen. Um diesen Vergleich zu ermöglichen, werden verschiedene Funktionen benötigt, die eine Metrik für die Distanz zwischen den Graphen

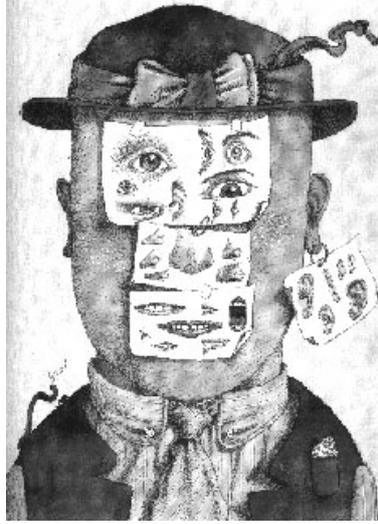


Abbildung 10: Bündelgraph aus einem künstlerischen Blickpunkt (“Unfinished Portrait“ von Tullio Pericoli (1985))

definieren. Die Ähnlichkeit der Graphen lässt sich auf Basis der Jets, aber auch auf Basis der geometrischen Struktur überprüfen. In [30] werden für die Jets zwei verschiedene Vergleichsfunktionen definiert.

$$S_a(J, J') = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \quad (5)$$

und

$$S_\phi(J, J') = \frac{1}{2} \left(1 + \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j - \vec{d}\vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \right) \quad (6)$$

wobei in 5 die Phase nicht berücksichtigt wird. Die Amplituden a_j ändern sich mit der Position nur sehr langsam, während die Phasen ϕ_j starke Variationen aufweisen können. Aus diesem Grund haben Jets, die nur wenige Pixel auseinander liegen, stark unterschiedliche Koeffizienten, auch wenn sie annähernd das selbe Merkmal im Gesicht repräsentieren. Die erste Formel

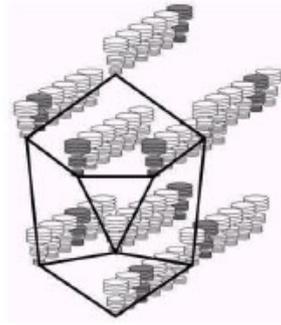


Abbildung 11: Bunch Graph

(5) dient daher vor allem dazu, eine grobe Positionierung der Jets zu erreichen.

Die Benutzung der Phaseninformationen bietet die Möglichkeit, zwei Muster mit ähnlicher oder gleicher Amplitude differenzierter zu unterscheiden und kann daher dazu verwendet werden, Jets genauer auszurichten. Eine Besonderheit in Formel 6 ist die Verwendung eines Verschiebungsvektors \vec{d} . Wenn zwei Jets J und J' sich auf den selben Ort beziehen, aber sich durch einen geringen Abstand \vec{d} unterscheiden, dann kann die daraus resultierende Phasenverschiebung näherungsweise durch $\vec{d}\vec{k}_j$ ausgeglichen werden. \vec{d} wird durch die Maximierung der Taylorentwicklung von S_ϕ berechnet. Die berechnete Verschiebung kann dazu benutzt werden, ein Jet exakter zu positionieren. Für ein Jet J in der Nähe eines Punktes x_0 wird J mit $J(x_0)$ verglichen und die Verschiebung $\vec{d}_0(J, J(x_0))$ berechnet. Darauf folgt der Vergleich von J mit $J(x_0 + d_0)$ und eine erneute Berechnung der Verschiebung. Dies ist ein iterativer Prozess, der solange fortgesetzt wird, bis J genau genug positioniert wurde.

Die Ähnlichkeit zwischen einem Gesichtsgraph und dem entsprechenden

Bündelgraph wird mit

$$S_B(G^I, B) = \frac{1}{N} \sum_n \max(S_\phi(J_n^I, J_n^{Bm})) - \frac{\lambda}{E} \sum_e \frac{(\Delta \vec{x}_e^I - \Delta \vec{x}_e^B)^2}{(\Delta \vec{x}_e^B)^2} \quad (7)$$

bestimmt. Dabei ist G^I ein Gesichtsgraph mit $n = 1, \dots, N$ Knoten und $e = 1, \dots, E$ Kanten und B der Bündelgraph, der sich zusammensetzt aus $m = 1, \dots, M$ Graphen der Trainingsmenge. λ ist ein Faktor zur Gewichtung der geometrischen Ähnlichkeit. Mit J_n werden die Jets am Knoten n bezeichnet. $\Delta \vec{x}_e$ sind die Abstände an den Kanten e .

Mit Hilfe der oben genannten Ähnlichkeitsfunktionen lassen sich die Gesichtsgraphen für neue Bilder berechnen. Die Positionierung der Knoten verläuft in mehreren Schritten, die im Folgenden kurz erklärt werden. Bei den unterschiedlichen existierenden Umsetzungen des Elastic Bunch Graph Matching variieren diese Schritte, der grundlegende Ablauf ist aber immer gleich.

Zuerst muss der Graph grob positioniert werden. In diesem Schritt wird der Graph noch nicht verformt, sondern nur in einem groben Raster über das Bild bewegt. Es wird lediglich S_a als Vergleichsfunktion verwendet. Es folgt die Anpassung der Größe und die genauere Positionierung. Der Graph wird skaliert. Als letzter Schritt kommt die lokale Verformung mit Hilfe der Phaseninformationen S_ϕ zum Einsatz.

Da der Bündelgraph für jeden Knoten mehrere Jets enthält, wird zu jedem Jet im neuen Gesichtsgraph der am besten passende aus dem Bündelgraph ausgewählt. Diese werden auch als lokale Experten (local experts) für das neue Gesicht bezeichnet. In jedem Schritt wird für die Position der Knoten der Abstand zum lokalen Experten aus dem Bündelgraphen berechnet. Die Position mit der höchsten Ähnlichkeit bzw. dem geringsten Abstand

wird für den nächsten Schritt verwendet.

Die Erkennung eines Gesichts mit der Elastischen Graphenanpassung ist sehr einfach. Zwei Gesichtsgraphen lassen sich auf Basis ihrer Jets oder auf Basis ihrer geometrischen Struktur vergleichen. Die Ähnlichkeitsmaße lauten

$$L_{Jet}G, G' = \frac{1}{N} \sum_{i=0}^N S(J_i, J'_i) \quad (8)$$

und

$$L_{Geo}(G_1, G_2) = \sqrt{\sum_{i=0}^M (x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2} \quad (9)$$

[33], also die durchschnittliche Ähnlichkeit zwischen korrespondierenden Jets, bzw. die durchschnittliche Ähnlichkeit zwischen den Kanten.

In [39] wird ein Verfahren vorgestellt, das die klassische Elastische Graphenanpassung um die Anwendung der Hauptkomponentenanalyse (4.3.4) erweitert und die Qualität der Merkmalsfindung verbessert. Da sich mittels der Hauptkomponentenanalyse hochdimensionale Merkmalsvektoren optimal darstellen lassen, wird diese auf die einzelnen Jets angewandt, die als 40- bzw. 80-dimensionale (bei Aufspaltung in Real- und Imaginärteil) Merkmalsvektoren aufgefasst werden können. Das Ergebnis sind sogenannte PCA-Jets, die die 80 Hauptkomponenten und 80 Varianzen enthalten. Ähnlich der Graphenanpassung mit einem Bündelgraph werden auch beim PCA-Matching (vgl. [39]) die einzelnen Jets zu einem Graphen mit denselben Verbindungen zwischen den Landmarken zusammengefasst. Für den Vergleich eines Jets, der aus einem Bild extrahiert wurde, mit einem PCA-Jet wird die Projektion des Jets auf die Hauptkomponenten berechnet.

Die Verwendung der Hauptkomponentenanalyse bewirkt eine Verbesserung der Merkmalspositionen im Vergleich zur klassischen Graphenanpassung, allerdings macht dieses Verfahren erst Sinn, wenn mehr als 80 Gesichter für die Erzeugung des Bündelgraphen herangezogen werden. Wenn die Zahl

der Merkmalsvektoren nämlich kleiner ist, als die Dimensionalität (hier: 80), dann liefert die Hauptkomponentenanalyse auch nur so viele linear unabhängige Hauptkomponenten wie Merkmalsvektoren existieren. Für Ergebnisse der PCA-basierten Graphenanpassung sei auf [39] verwiesen.

4.3.4 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (Principal Component Analysis, PCA) ist eine Methode, Variablen in einem hochdimensionalen Raum durch eine kleine Anzahl unkorrelierter Variablen zu ersetzen, die die gesamte Varianz der Ursprungsdaten erklären. Auf diese Weise ist es möglich, die effektive Dimension der Daten zu reduzieren.

Bei der PCA wird versucht, eine Datenmenge durch eine orthogonale Basis zu repräsentieren. Diese ergibt sich aus den Eigenvektoren der Kovarianzmatrix der Daten. Die entsprechenden Eigenwerte geben an, welche der Eigenvektoren besonders relevant sind. Geometrisch betrachtet wird die Basis gewählt, bei der die Summe der Abstände zu den Datenpunkten minimal wird.

Die Hauptkomponentenanalyse basiert auf der Annahme, dass die Eingabedaten normalverteilt sind.

Ausgangspunkt ist eine beliebig-dimensionale Datenmenge. Zuerst muss für jede Dimension das Mittel der Daten berechnet und subtrahiert werden. Das bedeutet zum Beispiel für eine Menge von zweidimensionalen Punktkoordinaten, dass von allen x -Werten der Durchschnitt aller vorhandenen x -Werte abgezogen wird. Entsprechendes gilt auch für die y -Werte. Das Ergebnis ist eine gemittelte Datenmenge, deren Durchschnitt null ist.

Als nächstes wird aus dieser Datenmenge die Kovarianzmatrix berechnet.

Die Größe dieser Matrix ist $n \times n$ bei einer n -dimensionalen Datenmenge. Da die Kovarianzmatrix quadratisch ist, lassen sich ihre Eigenvektoren und Eigenwerte berechnen. Die Eigenvektoren charakterisieren die verwendete Datenmenge. Jene, die zu den höchsten Eigenwerten gehören, geben die Richtungen an, entlang derer sich die Datenpunkte anordnen. Es macht also Sinn, die Eigenvektoren entsprechend der Größe ihrer Eigenwerte zu sortieren. Bei sehr kleinen Eigenwerten können die dazugehörigen Eigenvektoren ignoriert werden. So geht zwar ein Teil der Information aus den Daten verloren, aber dieser Anteil ist sehr gering. Das Weglassen einiger Eigenvektoren hat die Reduktion der Dimension der Datenmenge zur Folge. Nach der Auswahl der weiter zu verwendenden Eigenvektoren werden diese in einer Matrix zusammengefasst. Das linksseitige Multiplizieren der gemittelten Ausgangsdaten mit dieser Matrix entspricht dann einer Projektion der Daten in einen Merkmalsraum, der von den Eigenvektoren aufgespannt wird. Mit Hilfe der Hauptkomponentenanalyse werden Daten so transformiert, dass eine genaue Beschreibung ihrer Beziehung untereinander möglich wird. Die Daten werden durch die Ähnlichkeiten und Unterschiede zwischen ihnen beschrieben.

Um die Daten, die für die Berechnung der Eigenvektoren verwendet wurden, mit neuen Daten vergleichen zu können, müssen die neuen Daten auf die gleiche Weise mit den zuvor benutzten Eigenvektoren in den selben Merkmalsraum transformiert werden. Der eigentliche Vergleich erfolgt dann zum Beispiel mit der Euklidischen Distanz.

4.3.5 Unabhängige Komponentenanalyse

Die Unabhängige Komponentenanalyse (Independent Component Analysis, ICA) ist eine Weiterentwicklung der Hauptkomponentenanalyse. Ihr Ziel im Gegensatz zur PCA ist es, nicht nur die Korrelationen aus den Eingabedaten zu entfernen, sondern diese auch statistisch so unabhängig wie möglich zu machen. Die ICA versucht, nicht-orthogonale Basisvektoren zu finden, die zur Repräsentation der Daten verwendet werden können. Nicht-orthogonale Komponenten sind wesentlich besser in der Lage, die Struktur von Daten zu erfassen (vgl. Abbildung 12) [26].

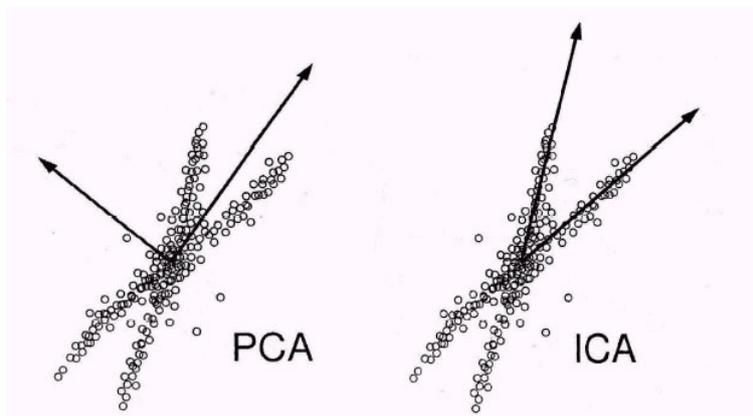


Abbildung 12: Vergleich der Ergebnisse von PCA und ICA

Ein wesentlicher Unterschied zur Hauptkomponentenanalyse liegt in der Grundvoraussetzung, dass die Daten nicht normalverteilt sein dürfen.

Ziel der ICA ist es, das Gleichungssystem

$$x = As$$

zu lösen. Dabei enthält x die Eingabedaten, während A , die sogenannte Mischungsmatrix, und s , die zu rekonstruierenden Ursprungssignale, unbekannt

sind.

Vor der Anwendung der eigentlichen ICA müssen die Eingabedaten x wie in der PCA durch Subtraktion des Mittelwertes m zentriert werden. Dadurch wird auch der gesuchte Vektor s zentriert. Ein weiterer sinnvoller Vorverarbeitungsschritt ist das sogenannte Whitening. Dabei wird auf den zentrierten Eingabevektor x eine lineare Transformation angewandt, die einen Vektor \tilde{x} zum Ergebnis hat, dessen Komponenten unkorreliert und deren Varianzen auf den Wert 1 normiert sind. Die Kovarianzmatrix von \tilde{x} ist also die Einheitsmatrix. Die Transformation lässt sich zum Beispiel mit einer Eigenwertzerlegung der Kovarianzmatrix $C = EDE^T$ berechnen. Die Matrix E enthält die Eigenvektoren und D die Eigenwerte. Die Berechnung erfolgt dann mit

$$\tilde{x} = ED^{-1/2}E^T x$$

Neben \tilde{x} liefert das Whitening außerdem eine neue Mischungsmatrix \tilde{A} . Diese ist orthogonal. Das Whitening ermöglicht eine Reduzierung der später mit ICA zu berechnenden Elemente der Mischmatrix von n^2 auf $n(n-1)/2$. Durch das Verwerfen von Eigenvektoren mit geringen Eigenwerten lässt sich außerdem die Dimension erheblich reduzieren.

Um die größtmögliche statistische Unabhängigkeit der Daten zu erreichen, muss die Ähnlichkeit der Verteilung der Daten zur Normalverteilung minimiert werden, bzw. die Nicht-Gaußähnlichkeit maximiert werden. Hierfür bieten sich verschiedene Verfahren an. Ein klassisches Maß für die Gaußähnlichkeit ist zum Beispiel die Kurtosis. Diese ist schnell zu berechnen, allerdings reagiert diese Methode sehr anfällig auf Ausreißer in den Daten. Die Negentropie lässt sich ebenfalls zur Bestimmung der Gaußähnlichkeit verwenden. Dieses Maß ist jedoch schwer zu berechnen, weil explizit die

Wahrscheinlichkeitsdichtefunktion vorausgesetzt wird [26]. Die Kurtosis und die Negentropie sind nur Beispiele aus einer Reihe von Möglichkeiten, die zur Maximierung der Nicht-Gaußähnlichkeit herangezogen werden können. Weitere Informationen hierzu finden sich in [27].

Die gesuchten Basisvektoren finden sich nach erfolgreicher Anwendung der ICA in den Spalten der resultierenden Mischungsmatrix A .

4.3.6 Diskriminanten

Die Diskriminanzanalyse ist laut [25] „ein multivariates Verfahren zur Analyse von Gruppenunterschieden. Sie ermöglicht es, die Unterschiedlichkeit von zwei oder mehreren Gruppen hinsichtlich einer Mehrzahl von Variablen zu untersuchen, um Fragen folgender Art zu beantworten:

- „Unterscheiden sich die Gruppen signifikant voneinander hinsichtlich der Variablen?“
- „Welche Variablen sind zur Unterscheidung zwischen den Gruppen geeignet bzw. ungeeignet?““

Ein Anwendungsgebiet der Diskriminanzanalyse ist die Bestimmung der Gruppenzugehörigkeit von Elementen, also die Klärung der Frage „In welche Gruppe ist ein „neues“ Element, dessen Gruppenzugehörigkeit nicht bekannt ist, aufgrund seiner Merkmalsausprägungen einzuordnen? [25]“.

Sowohl PCA als auch ICA verwenden die Trainingsdaten im ganzen. Die Lineare Diskriminanzanalyse (LDA) dagegen berücksichtigt die unterschiedlichen Gesichtsklassen. Dies ermöglicht es, individuelle Gesichter zu unterscheiden aber Gesichter von ein und der selben Person wiederzuerkennen.

Die Bilder aus der Trainingsmenge müssen im Vorfeld der Diskriminanzanalyse in Klassen eingeteilt werden, also pro Person eine Klasse.

Die Unterscheidung zwischen zwei Gruppen bzw. Klassen ist zum Beispiel möglich durch die Distanz ihrer Centroide. Allerdings ist dieser Vergleich ungenau, wenn die einzelnen Gruppen stark streuen. Ein besseres Maß für die Unterschiedlichkeit erhält man, wenn auch die Streuung der Gruppen berücksichtigt wird. Bei der Betrachtung mehrerer Gruppen sind verschiedene Arten der Streuung zu beachten: die Streuung zwischen den Gruppen S_B und die Streuung in den Gruppen S_W :

$$S_B = \sum_{i=1}^c N_i \cdot (x_i - \mu) \cdot (x_i - \mu)^T$$

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i) \cdot (x_k - \mu_i)^T$$

wobei N_i der Anzahl der Trainingsbilder in Klasse i entspricht. c ist die Anzahl der Klassen, μ_i ist das Mittel der Klasse i und X_i repräsentiert die Menge der Bilder, die zur Klasse i gehören.

Die Streuung zwischen den Gruppen wird durch die quadrierten Abweichungen der Gruppencentroide vom Gesamtmittel gemessen und kann so für beliebig viele Gruppen erfolgen. Um unterschiedliche Gruppengrößen zu berücksichtigen, werden die Abweichungen jeweils mit der Gruppengröße N_i multipliziert.

Die Streuung in den Gruppen wird durch die quadrierten Abweichungen der Gruppenelemente vom jeweiligen Gruppencentroid gemessen.

Ziel der linearen Diskriminanzanalyse ist es, das Diskriminanzkriterium

$$\Gamma = \frac{S_B}{S_W} \quad (10)$$

zu maximieren, also die Maximierung von S_B und die Minimierung von S_W . Dies wird erreicht, wenn die Spaltenvektoren der Projektionsmatrix W_{LDA} die Eigenvektoren von $S_W^{-1} \cdot S_B$ bilden.

$$W_{LDA} = \underset{W}{\operatorname{argmax}} \frac{W^T \cdot S_B \cdot W}{W^T \cdot S_W \cdot W}$$

Um Schwierigkeiten, wie der Singularität von S_W , entgegenzuwirken sowie zur Reduktion der Dimension, empfiehlt sich die Verwendung der PCA im Vorfeld.

Die Lineare Diskriminanzanalyse kann zur Definition von globalen, aber auch von lokalen Merkmalen verwendet werden [28]. Ein Bild besteht aus $w \times h$ Pixeln, die unterschiedlich wichtig für die Klassifizierung sind. Wenn man sich ausschließlich auf die diskriminanten Pixel konzentrieren möchte, lässt sich die Berechnungszeit reduzieren und möglicherweise sogar ein besseres Ergebnis erzielen. Mit Hilfe der Basisvektoren der Projektionsmatrix W aus der Linearen Diskriminanzanalyse lassen sich genau diese Pixel finden.

Jedes Pixel eines Bildes entspricht im hochdimensionalen Raum einer Dimension bzw. einer Achse. Wenn also die Dimension i stark diskriminant ist, so gilt dies auch für das Pixel i .

Ausgehend von C unterschiedlichen Klassen führt die LDA zu $C - 1$ diskriminanten Vektoren w_i . Zuerst muss ein neuer Vektor w' definiert werden durch

$$w' = \sum_{i=1}^m |w_i|$$

Anschließend werden die Elemente von w' absteigend sortiert in den Vektor v eingetragen. Entsprechend eines vordefinierten Schwellwertes t wird der Index r bestimmt, für den gilt $r/d \approx t$. d ist dabei die Dimension der

Basisvektoren. v_r liefert dann den Schwellwert, mit dem die Elemente von w' binarisiert werden. Der resultierende Ergebnisvektor w^{mask} enthält dann für diskriminante Pixel den Wert 1 und für weniger interessante Pixel den Wert 0. Die Visualisierung des Maskenvektors weist valide (diskriminante) Pixel an den Stellen auf, an denen sich im Gesicht die Augen, die Nase und der Mund befinden, während ein Großteil des restlichen Gesichts schwarz gefärbt ist, was auf uninteressante Bereiche hinweist [28]. Auf diese Weise lässt sich die weitere Verarbeitung auf diskriminante Bereiche im Gesicht beschränken.

Die diskriminierenden Eigenschaften von Merkmalen lassen sich auch direkt unter Verwendung der sogenannten Fisher Diskriminanzfunktion 4.3.6 ermitteln:

$$\frac{\sum_{i=1}^c (m_i - m)^2}{\sum_{i=1}^c \frac{1}{n_i} \sum_{x \in \Phi_i} (x - m_i)^2}$$

c ist die Anzahl an Klassen bzw. Personen, Φ_i ist die Menge an extrahierten Merkmalen für eine Klasse i , n_i entspricht der Größe von Φ_i , m_i ist das Mittel von Φ_i und m ist das globale Mittel eines Merkmals über alle Klassen. Entsprechend der Werte, die der Fisher Koeffizient liefert, kann man den einzelnen Merkmalen eines Merkmalsvektors eine Wichtigkeit bezüglich ihrer diskriminierenden Eigenschaft zuordnen.

4.3.7 Support Vector Machines

Support Vector Machines ist wie die lineare Diskriminanzanalyse ein Verfahren zur Klassifizierung von Objekten. Dabei werden die Daten in einen höherdimensionalen Merkmalsraum transformiert, in dem sie einfacher zu

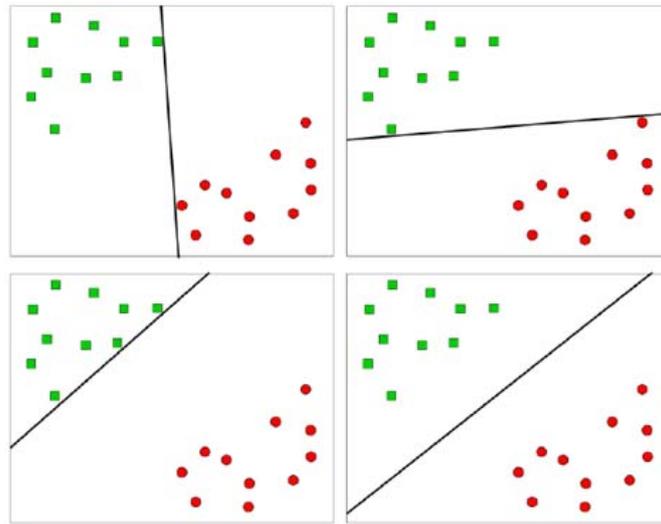


Abbildung 13: Separationsmöglichkeiten durch Hyperebenen im Zwei-Klassen-Fall, [41]

separieren sind. Die Unterscheidung der Objekte erfolgt durch das Erzeugen von sogenannten Hyperebenen im Merkmalsraum, die die verschiedenen Objektklassen voneinander trennen.

Die Klassifizierung von Daten mit niedriger Dimension ist oft schwierig, weil sie sich oft nicht linear trennen lassen. Daher ist es sinnvoll, die Daten in einen anderen, höherdimensionalen Merkmalsraum zu transformieren, in dem eine lineare Trennung wesentlich einfacher umzusetzen ist.

Die lineare Separation erfolgt dann (im Zwei-Klassen-Fall) mit einer Hyperebene, bzw. im zweidimensionalen Raum durch eine einfache Linie. Datenmengen lassen sich durch beliebige Hyperebenen trennen, allerdings ist die erzielte Separation nur unter bestimmten Umständen auch eine optimale Trennung. Betrachtet man die Beispiele in Abbildung 13, so würde man intuitiv die Hyperebene aus dem Beispiel rechts unten als optimal bezeichnen.

Der Grund hierfür ist, dass diese Ebene einen bestimmten Mindestabstand zu den Datenpunkten der beiden Klassen aufweist. Sie teilt den größten

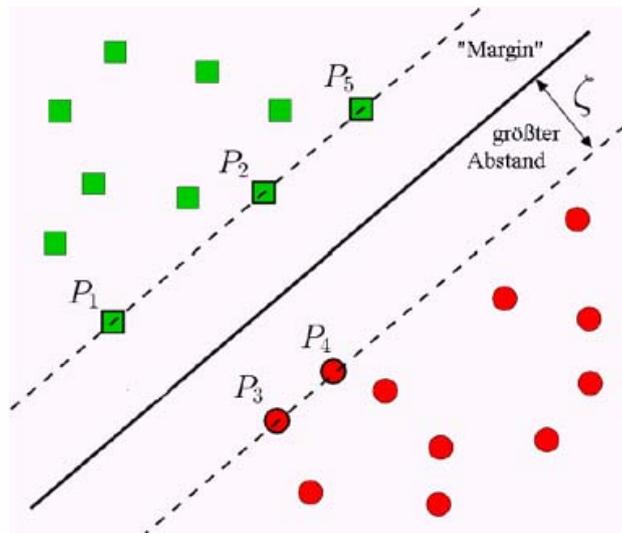


Abbildung 14: Hyperebene mit maximalem Rand im Zwei-Klassen-Fall, [41]

Spalt zwischen ihnen. Eine Hyperebene, die diese Eigenschaft aufweist, wird auch als Hyperebene mit maximalem Rand oder maximaler Trennspanne (Maximum Margin Hyperplane) bezeichnet (vgl. Abbildung 14). Die Punkte P_1 bis P_5 bestimmen die optimale Lage der Hyperebene. Sie sind die Punkte, für die die „Breite“ der Hyperebene genau der geringste Abstand von der Ebene ist. Diese Punkte werden auch als Support Vectors bezeichnet.

Neue Datenpunkte werden bei dieser Separation mit großer Wahrscheinlichkeit der richtigen Klasse zugeordnet. Je größer der Spalt zwischen den Klassen, desto bessere Ergebnisse liefert das Verfahren.

Eine Hyperebene $(w \cdot x) + b = 0$ erfüllt (im Zwei-Klassen-Fall) die Bedingung, dass die Daten einer Klasse oberhalb der Hyperebene zu finden sind, während die Daten der anderen Klasse sich unterhalb der Hyperebene befinden:

$$(w \cdot x_i) + b > 0 \text{ falls } y_i = 1$$

$$(w \cdot x_i) + b < 0 \text{ falls } y_i = -1$$

bzw.

$$y_i [(w \cdot x_i) + b] \geq 1 \quad i = 1, \dots, l$$

Um sicherzustellen, dass es sich bei einer Hyperebene um die optimale Hyperebene handelt, genügt diese Forderung allerdings nicht. Zusätzlich zur oben genannten Bedingung muss eine optimale Hyperebene den minimalen Abstand zu jedem Datenpunkt maximieren.

$$\rho(w, b) = \min_{x_i | y_i = 1} \frac{w \cdot x_i + b}{|w|} - \max_{x_j | y_j = -1} \frac{w \cdot x_j + b}{|w|}$$

ist die Summe der Abstände der zwei am nächsten zur Hyperebene gelegenen Punkte, welche zu maximieren ist. Dieses Optimierungsproblem wird gelöst durch Minimierung von

$$\Phi(w) = \frac{1}{2} w \cdot w$$

Die Lösung dieses Problems erfolgt zum Beispiel über die Lagrangefunktion.

$$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_{i=1}^l \alpha_i \{[(x_i \cdot w) + b] y_i - 1\} \quad (11)$$

Daraus ergeben sich folgende Eigenschaften für die optimale Hyperebene:

$$\sum_{i=1}^d \alpha_i y_i = 0 \quad (12)$$

und

$$\vec{w}_0 = \sum_{i=1}^d \alpha_i y_i x_i \quad (13)$$

α_i sind die Lagrangemultiplikatoren. w_0 ist die Linearkombination der Vektoren der Trainingsmenge. Nur die sogenannten Support Vectors besitzen Koeffizienten α_i ungleich Null:

$$\vec{w}_0 = \sum_{\text{Support Vectors}} \alpha_i y_i x_i, \quad \alpha_i > 0$$

Durch Einsetzen von 12 und 13 in 11 erhält man

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

was zu maximieren ist. Dies liefert den Vektor $\alpha = (\alpha_1, \dots, \alpha_l)$ der Lagrange-Multiplikatoren, die in der finalen Indikatorfunktion verwendet werden:

$$f(x) = \text{sign} \left(\sum_{\text{Support Vectors}} y_i \alpha_i (x_i \cdot x) - b_0 \right)$$

Dabei sind x_i die Support Vektoren und b_0 ist der Schwellwert

$$b_0 = \frac{1}{2} [(w_0 \cdot x^*(1)) + (w_0 \cdot x^*(-1))]$$

$x^*(1)$ ist ein beliebiger Support Vektor der ersten Klasse und $x^*(-1)$ einer der zweiten Klasse.

Support Vector Machines eignen sich gut zur Objekterkennung, da sie eine hohe Präzision mit einer geringen Fehlerwahrscheinlichkeit vereinen. Sie arbeiten gut auf Daten mit vielen Merkmalen und erzielen eine gute Performanz (vgl. [41]). Ein Nachteil der Support Vector Machines liegt in der Entscheidung, welche Transformation verwendet werden soll, um die Daten in einen hochdimensionalen Raum zu transformieren.

5 Auswertung

5.1 Teststandards

Um die Güte eines biometrischen Systems festzustellen, gibt es verschiedene statistische Größen, die Aussagen über das System erlauben. Die Verwendung dieser Größen ermöglicht es, unterschiedliche biometrische Verfahren mit einander zu vergleichen.

5.1.1 FAR und FRR

Die Falschakzeptanzrate (FAR) gibt die Häufigkeit an, mit der nichtberechtigte Personen als berechtigt akzeptiert werden. Die FAR ist wichtig für die Sicherheit eines Systems, da das fälschliche Akzeptieren von nichtberechtigten Personen nicht erwünscht ist und in der Regel zu Schäden führt.

Die Falschrückweisungsrate (FRR) sagt aus, wie oft Personen, die eine Berechtigung besitzen, vom System zurückgewiesen werden. Falsche Rückweisungen sind nicht wie die FAR sicherheitsrelevant. Dennoch sollte die FRR möglichst klein sein, da unnötiges Wiederholen von Identifizierungsversuchen lästig ist und die Benutzerakzeptanz beeinträchtigen kann.

Die FAR und FRR eines biometrischen Systems bedingen sich gegenseitig. Dies bedeutet, dass es eine Akzeptanzschwelle gibt, deren Erhöhung die FAR verkleinert, während gleichzeitig die FRR erhöht wird. Die Angabe von lediglich einer dieser zwei Wahrscheinlichkeiten ist daher nicht sinnvoll.

Für ein biometrisches Erkennungssystem ist es erstrebenswert, sowohl die FAR als auch die FRR zu minimieren. Um dieses Ziel zu erreichen, gibt es eine Reihe von Optimierungsmöglichkeiten. Eine positive Beeinflussung der FRR erreicht man zum Beispiel indem man sich nur mit der bestmöglichen Bildqualität zufrieden gibt und entsprechend den zu verwendenden Sensor

auswählt. Des Weiteren sollte die biometrische Anwendung so benutzerfreundlich wie möglich gestaltet sein, um Fehlbedienungen durch den Benutzer zu vermeiden. Auch die Art der biometrischen Merkmale spielen eine Rolle. So hat die Eindeutigkeit eines Merkmals einen wesentlichen Einfluss auf die FAR, während sich die Messbarkeit und Konstanz des Merkmals vor allem auf die FRR auswirken.

Die Wahrscheinlichkeitsdichtekurven von FAR und FRR lassen sich durch Aufsummieren der Ähnlichkeitswerte aus Messreihen für Berechtigte und Nichtberechtigte berechnen. Durch Integration dieser Kurven ergeben sich die FAR- und FRR-Kurven. Dieses Kurvenpaar eignet sich besonders gut dafür, einen optimalen Schwellwert EER (equal error rate) für das biometrische System festzulegen.

Die False Match Rate (FMR) ist das Äquivalent zur Falschakzeptanzrate. Sie gibt den Anteil der im Vergleich fälschlicherweise akzeptierten Identitäten an. Im Gegensatz zur FAR werden in der FMR allerdings die abgewiesenen Versuche, die auf schlechte (Bild-)Qualität zurückzuführen sind, nicht berücksichtigt. Entsprechendes gilt für die False Non-Match Rate (FNMR) bezüglich der FRR. Sie gibt den Anteil der fälschlicherweise nicht erkannten Identitäten an.

5.1.2 FER

Es kann vorkommen, dass für eine Person das Enrolment immer wieder fehlschlägt und diese Person folglich nicht enrollt werden kann. Die Nutzer-
ausfallrate (FER) eines Systems gibt den Anteil an Personen an, für die dies der Fall ist. Tritt ein ähnliches Problem bei der Verifikation auf, also dass eine bereits enrollte Person auch nach mehreren Verifikationsversuchen nicht

akzeptiert wird, spricht man von FTA-Raten (failure to acquire).

5.1.3 ROC-Kurve

Die FAR- und FRR-Kurven können beliebigen Skalierungen und Transformationen unterworfen sein. Dies ist vor allem auf die Interpretation des Schwellwertes bzw. des Ähnlichkeitsmaßes zurückzuführen. Prinzipiell ist so das Aussehen der beiden Kurven veränderbar, allerdings bleibt das Verhältnis zwischen einem Wertepaar FAR/FRR immer gleich.

Eine skalierungsunabhängige Darstellung verbessert die Vergleichbarkeit biometrischer Systeme. Eine solche Darstellung ist die Receiver Operating Characteristic (ROC). Die FAR-Werte werden gegen die FRR-Werte aufgetragen. Das Ergebnis ist unabhängig von irgendwelchen Schwellwerten.

Die Performanz eines biometrischen Systems kann durch die drei statistischen Größen FAR, FER und FRR bestimmt werden. Diese beeinflussen sich jedoch gegenseitig. Vor allem wenn die Akzeptanzschwelle verändert wird, ist ein Vergleich verschiedener Systeme schwierig. Ein objektiver Vergleich einer der oben genannten Größen erfordert, dass die anderen beiden Größen gleich sind. Beispielsweise muss sichergestellt sein, dass bei der Betrachtung der FARs zweier Systeme die zugehörigen FRRs beider Systeme gleich sind sowie die zugehörigen FERs.

Die ROC-Kurve ist eine traditionelle Methode zum Vergleich der Performanz verschiedener Systeme unter gleichen Bedingungen oder zum Vergleich verschiedener Konfigurationen eines Systems. Die Performanz eines Systems wird dargestellt durch die Abbildung der FMR auf der x-Achse und $1 - FNMR$ auf der y-Achse.

5.1.4 DET-Kurve

Die DET-Kurve (DET = detection error trade-off) ist wie die ROC-Kurve eine Möglichkeit, die Performanz eines Systems darzustellen. In der DET-Kurve werden die Fehlerraten FMR (auf der x-Achse) und die FNMR (auf der y-Achse) gegeneinander abgebildet. Bei der DET-Kurve handelt es sich um eine Variante der ROC-Kurve. Da beide Darstellungen auf den gleichen Daten basieren, reicht eine von ihnen aus, um ein biometrisches System zu charakterisieren. In der Regel werden bei der Darstellung von ROC- und DET-Kurven die Achsen des Koordinatensystems logarithmisch skaliert, um Unterschiede zwischen Systemen klarer herausstellen zu können.

5.1.5 CMC-Kurve

Die Cumulative Match Characteristic Kurve ist eine Methode, die Ergebnisse eines biometrischen Systems abzubilden. Eine CMC Kurve trägt den Rang gegen die Erkennungsrate auf.

Die Kurve gibt in Abhängigkeit von der Erkennungsrate an, wie oft eine korrekte Identität bei einem Test gegen alle Datenbanksätze unter den k ersten Rängen zu finden ist. k wird dabei auf der x-Achse aufgetragen. Auf der y-Achse sind die Matchingscores von 0 bis 1 aufgetragen.

Die Kurve ist in der Regel exponentiell ansteigend. Je stärker sie sich dem Wert 1 nähert, desto besser ist das geprüfte Verfahren. Im Idealfall bildet die CMC-Kurve eine Horizontale.

5.2 Ergebnisse

Im Rahmen dieser Diplomarbeit wurde das Elastic Bunch Graph Matching ([30]) auf Tiefenbilder angewandt. Ausgangspunkt hierfür war das System zur Evaluation von Gesichtserkennungsalgorithmen der Universität Colorado ([36]). Innerhalb dieses Systems wurden verschiedene grundlegende Verfahren zur Gesichtserkennung umgesetzt: PCA, eine Kombination von PCA und LDA, ein Bayes Klassifikator sowie das Elastic Bunch Graph Matching. Die Erweiterung des Verfahrens zur Anwendung auf Tiefenbilder beinhaltete unter anderem das Einführen neuer Ähnlichkeitsmaße.

Neben den bereits vorhandenen Ähnlichkeitsmaßen für den Vergleich von Jets sowie den Vergleich der geometrischen Struktur der Gesichtsgraphen auf 2D-Basis wurde zusätzlich ein Ähnlichkeitsmaß für den Vergleich der aufgrund der Tiefenwerte vorhandenen 3D-Struktur der Gesichtsgraphen verwendet.

$$L_{Geo3D}(G_1, G_2) = \sqrt{\sum_{i=0}^M (x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2 + (z_{i1} - z_{i2})^2}$$

Zur genaueren Ausrichtung der Graphen vor dem Vergleich wurden die in Kapitel 4.3.2 vorgestellten Verfahren ICP und RICP verwendet. Es zeigte sich, dass beide gleiche Ergebnisse liefern, da bei den Gesichtsgraphen feste Punktkorrespondenzen vorgegeben sind. Daher wurde für die Tests das 3D-Ähnlichkeitsmaß $Geo3D_{ICP}$ verwendet. Für den Vergleich der Jets wurde das in [33] definierte Maß $FGPredictiveIter$ benutzt. Der Vergleich der geometrischen Struktur auf 2D-Basis erfolgte mit dem Ähnlichkeitsmaß $GeoLeastSquaresPI$ (vgl. [33]).

Darüber hinaus wurden zwei weitere Ähnlichkeitsmaße definiert, die die Ergebnisse von Jet-Ähnlichkeit und geometrischer Ähnlichkeit kombinieren. Diese verwenden die Ergebnisse der jeweils zuerst berechneten Ähnlichkeit

zwischen zwei Knoten der Gesichtsgraphen als Gewichtung für das jeweils zweite Maß.

Für das Testen des Verfahrens stand eine Datenbank aus 4223 Tiefenbildern zur Verfügung. Sie enthielt Aufnahmen von insgesamt 89 Personen, die in einem Zeitraum von sechs Wochen (2 Aufnahmen pro Tag) entstanden sind. Die Personen zeichneten sich durch unterschiedliches Alter und Geschlecht aus. Sowohl Brillenträger, Personen mit Bartwuchs und verschiedenen Frisuren waren in der Datenbank enthalten. Als Galerie-Menge dienten die Bilder aus dem Enrollment der 89 Personen. Zum Testen wurden drei verschiedene Testmengen, bestehend aus 2856, 951 und 915 Bildern von jeweils 87, 28 und 28 Personen aus der Datenbank verwendet. Darin waren die Bilder aus dem Enrollment nicht enthalten.

Der verwendete Bündelgraph setzte sich zusammen aus 56 Bildern aus der Galerie-Menge. Die grundlegende Struktur der Gesichts- und Bündelgraphen wurde im Vergleich zum Ausgangsverfahren [36] verändert (vgl. Abbildungen 15 und 16). Dies beinhaltete eine Reduktion der Knotenanzahl im Gesichtsgraphen von 25 Knoten auf 16 Knoten. Entsprechend wurde auch die Anzahl Links zwischen den Knoten verringert (von 55 auf 38), was wiederum die Reduktion der Anzahl der zu berechnenden Jets an interpolierten Punkten zur Folge hatte. Für den Bündelgraphen wurden sämtliche Landmarken aus den 56 Gesichtern manuell ausgewählt.

Insgesamt bestand die verwendete Gesichtgraphstruktur also aus 16 Knoten, 38 Kanten und 38 möglichen interpolierten Punkten für den Jetvergleich. Die Änderung der Struktur begründet sich damit, dass in dem Ausgangsverfahren sehr viele Merkmalspunkte am Rand der betrachteten Gesichter extrahiert wurden. Für Tiefenbilder ist die genaue Lokalisation dieser Merkmale schwieriger und damit nicht genau genug. Weitere Knoten, die keine

Berücksichtigung fanden, befanden sich an Positionen, an denen in Tiefenbildern häufig Löcher auftreten, zum Beispiel an den Pupillen.

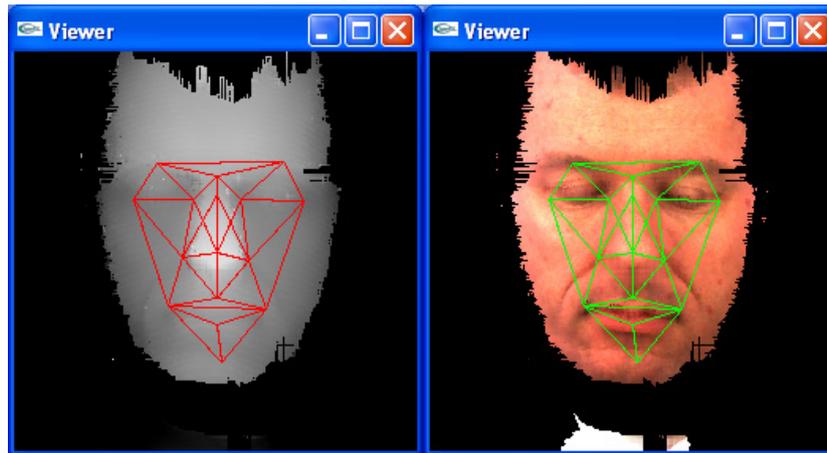


Abbildung 15: Struktur der im Test verwendeten Gesichtsgraphen

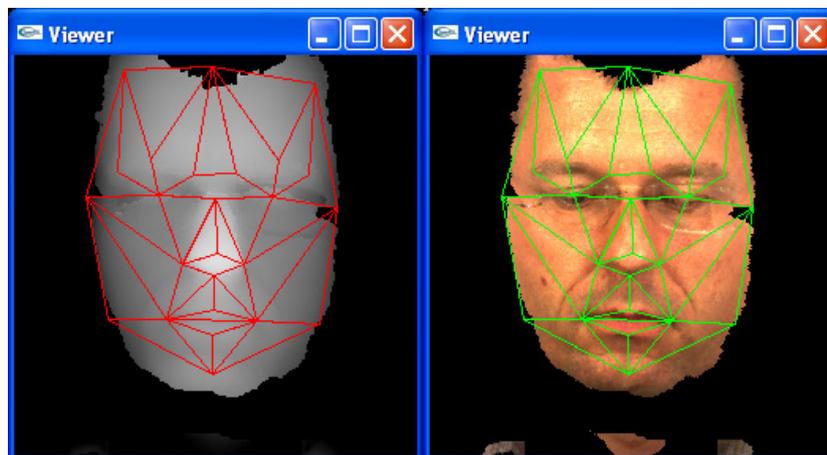


Abbildung 16: Struktur der in [36] verwendeten Gesichtsgraphen

In den Abbildungen 17, 18 und 19 sind die Ergebnisse aus dem Vergleich der Galerie-Bilder mit der ersten Testmenge (2856 Bilder) dargestellt.

Die CMC-Kurve (Abbildung 17) setzt sehr weit unten an und verläuft relativ steil. Laut dieser Darstellung wird eine maximale Erkennungsrate von lediglich 39% erreicht. Die Werte für die durchschnittliche Jet-Ähnlichkeit sowie eine Kombination von geometrischer Ähnlichkeit mit der Jet-Ähnlichkeit fallen besonders gering aus. Auch die DET-Kurve (Abbildung 18) und die

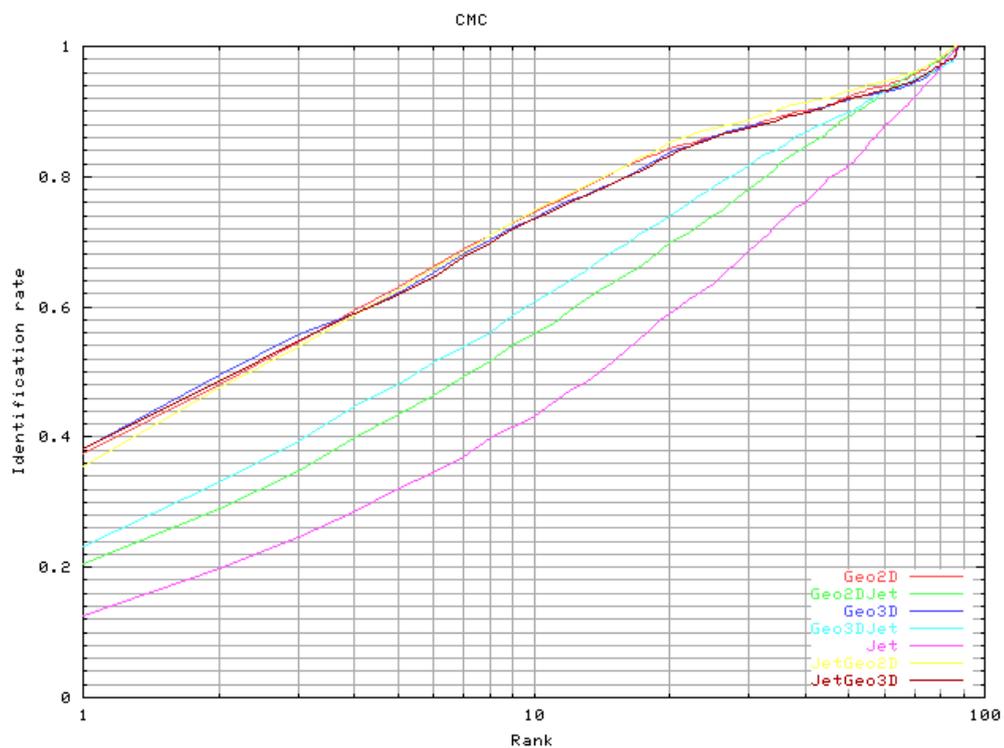


Abbildung 17: CMC-Kurve

ROC-Kurve (Abbildung 19) zeigen, dass für die Jet-Ähnlichkeit und die Kombination mit ihr schlechtere Ergebnisse erzielt werden. Die Falschrückweisungsrate ist insgesamt sehr hoch. Das Verfahren weist also zu viele berechnete Identitäten zurück.

Die relativ schlechten Ergebnisse sind vermutlich auf die Verwendung der Jets zurückzuführen. Die Antworten der Gabor-Wavelets fallen für Tiefenbilder nicht so ausgeprägt aus, wie für Intensitätsbilder. Wie bereits in Kapitel

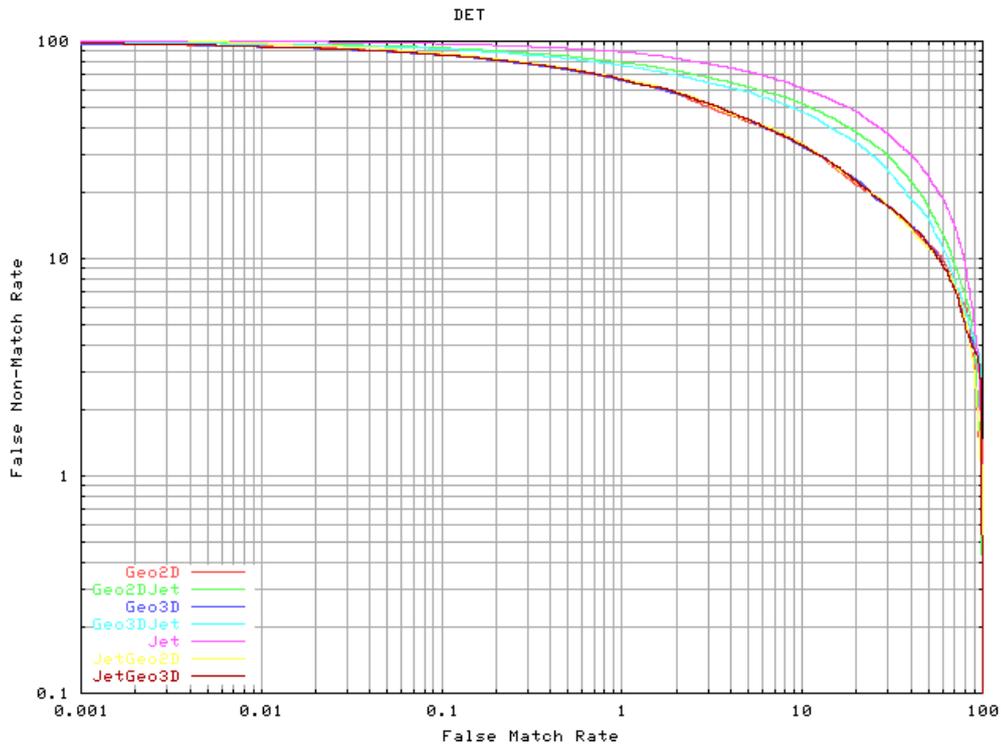


Abbildung 18: DET-Kurve

4.2.2 beschrieben, werden Wavelets in der Objekterkennung als Merkmalsdetektoren verwendet. Sie reagieren auf Kanten und Ecken stärker als auf einförmige Bereiche. In Tiefenbildern existieren kaum oder gar keine Strukturen, die mit Kanten und Ecken vergleichbar sind. Aus diesem Grund sind die Antworten von Wavelets für die Merkmalsdetektion in Tiefenbildern offensichtlich nicht ausreichend, um allein gute Ergebnisse in der Gesichtserkennung zu erzielen.

In [33] wird für die geometrische Ähnlichkeit zweidimensionaler Gesichtsgraphen eine Erkennungsrate von bis zu 43% angegeben (Ähnlichkeitsmaß *GeoLeastSquaresPI*). Die beim Vergleich von aus Tiefenbildern erzeugten Gesichtsgraphen erreichten 39% lassen darauf schließen, dass dennoch eine gute lokale Merkmalsdetektion mit Hilfe der Gabor-Wavelets erfolgt.

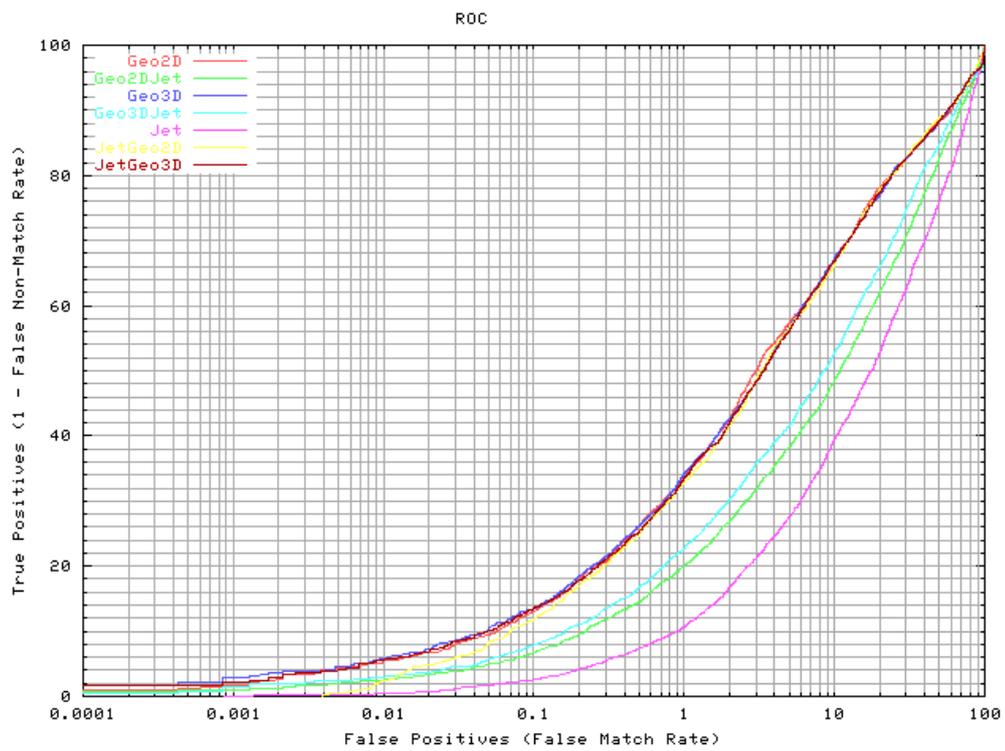


Abbildung 19: ROC-Kurve

6 Fazit

Die Ergebnisse in Kapitel 5 zeigen, dass das Elastic Bunch Graph Matching nach Wiskott et al. ([30]) für die Gesichtserkennung auf Tiefenbildern nicht in ausreichendem Maße zuverlässig ist. Das Graph Matching bedarf bei der Verwendung von Tiefenbildern einer anderen Methode der Merkmalsextraktion, die die Knoten des Graphen genauer positioniert. Außerdem sind die Wavelet-Antworten bei der Filterung von Tiefenbildern nicht hoch genug, da keine ausgeprägten Kanten oder Ecken existieren. Zur Verbesserung der Ergebnisse wäre zum Beispiel eine Kombination von Krümmungsinformationen und Waveletantworten, wobei die Jets aus den entsprechenden Intensitätsbildern extrahiert würden.

Die Ergebnisse aus [33] haben bereits gezeigt, dass der ausschließliche Vergleich der 3D-Koordinaten der Knoten von Gesichtsgraphen keine optimale Erkennung erzielt. Ein Vergleich auf der Basis der Oberflächenkrümmung verbunden mit den Jet-Antworten aus den Intensitätsbildern könnte erheblich zur Verbesserung der Ergebnisse beitragen.

Die merkmalsbasierte Erkennung ist im hohen Maße abhängig von einer möglichst exakten Detektion der Merkmale, die im Vergleich verwendet werden sollen. Tiefenbilder bieten die Möglichkeit, zusätzliche Tiefeninformationen in die Merkmalsuche mit einzubeziehen. Die Berechnung der Krümmung von Oberflächen ist zum Beispiel eine Methode, die Eigenschaften einer Oberfläche zu berücksichtigen. Das Auffinden von Merkmalspunkten, die den Ecken in Intensitätsbildern entsprechen würden, ist allerdings nur bedingt möglich. Die Krümmungsinformationen eignen sich vor allem zur Segmentierung von Tiefenbildern in konvexe und konkave Regionen, sowie zum Finden von Strukturen, die man als „Kämme“ (ridge lines) bezeichnen kann. Die Segmentierungsergebnisse lassen sich anschließend für die eigent-

liche Merkmalsdetektion verwenden.

Multimodale Systeme, in denen sowohl 2D- als auch 3D-Informationen Einsatz finden, zeigen bessere Erkennungsraten als die jeweiligen einfachen 2D- bzw. 3D-Systeme (vgl. [21]). Auch das Graph Matching könnte von einer Entwicklung zu einem multimodalen Verfahren profitieren.

A Das Gesicht

Das menschliche Gesicht weist eine Vielzahl an Merkmalen auf. Abbildung 20 gibt einen Überblick über die anthropologischen Merkmale, die bei einer Frontalansicht im Gesicht auffindbar sind.

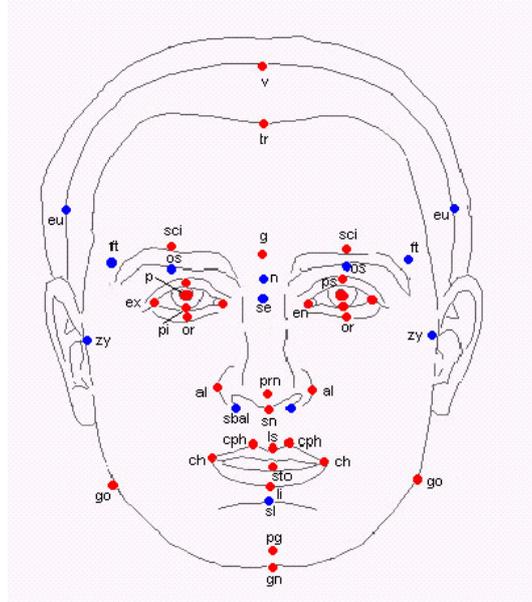


Abbildung 20: Gesichtsmerkmale

Die Tabelle in Abbildung 21 zeigt die durchschnittlichen Abmessungen der Augenregionen.

Feature	Geometric constraints			Symmetry constraints	
	area	width	height	dist. limits d_s	asymm. $(\Delta d_s, \Delta p_s)$
eyeballs (eb)	0 - 350	0 - 40	0 - 17	35 - 70	12, 15
inside corner eye cavity (iec)	150 - 900	0 - 300	0 - 300	10 - 25	12, 15
outside corner eye cavity (oec)	0 - 900	0 - 300	0 - 300	40 - 70	12, 15
Feature	Proximity Target		Verification Conditions		
eyeballs (eb)	$ p_s - p_s(\text{nose center}) < 40$		<ul style="list-style-type: none"> • p_s above p_s (nose base) 		
inside corner eye cavity (iec)	$ p_s - p_s(\text{eb}) < 14$ or $ p_s - p_s(\text{nose center}) < 40$		<ul style="list-style-type: none"> • p_s above p_s (nose base) • d_s (inside corner) $< d_s$ (eyeball) 		
outside corner eye cavity (oec)	$ p_s - p_s(\text{eb}) < 14$ or $ p_s - p_s(\text{iec}) < 17$ or $ p_s - p_s(\text{nose center}) < 30$		<ul style="list-style-type: none"> • p_s above p_s (nose base) • d_s (outside corner) $> d_s$ (eyeball) • d_s (inside corner) $> d_s$ (eyeball) 		

Abbildung 21: Durchschnittliche Maße für Augenregionen, [13]

B Differentielle Geometrie

Eine Fläche im Raum kann in ihrer parametrischen Form wie folgt ausgedrückt werden:

$$S = S(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \\ z(u, v) \end{pmatrix}$$

Die Komponenten x , y und z sind in einem Punkt P mindestens zweimal nach u und v partiell differenzierbar. Ohne diese Forderung können die Krümmungen nicht eindeutig ermittelt werden. Die beiden Tangentenvektoren an dem Kurvenpunkt P in Richtung der Parameter u und v sind gegeben durch die partiellen Ableitungen von $S(u, v)$ nach u bzw. v

$$S_u = \frac{\delta S}{\delta u}$$

und

$$S_v = \frac{\delta S}{\delta v}$$

Der Einheitsnormalenvektor der Tangentialebene von S am Punkt P ergibt sich aus dem normierten Kreuzprodukt

$$n_S(P) = \frac{S_u \times S_v}{\|S_u \times S_v\|}$$

Für Tiefenbilder mit der Parametrisierung

$$S(x, y) = \begin{pmatrix} x \\ y \\ h(x, y) \end{pmatrix}$$

ist $n_S(P)$

$$n_S(P) = \frac{(-h_x, -h_y, 1)}{\sqrt{1 + h_x^2 + h_y^2}}$$

mit $h_i = \delta h / \delta i$.

B.1 Normalkrümmung einer Kurve

Der Punkt P , der Flächennormalenvektor n_S und ein beliebiger Tangentenvektor t an P bestimmen eine Ebene $E_S = P + b_1 \cdot n_S + b_2 \cdot t$, deren Schnitt mit S eine Kurve C liefert (vgl. Abbildung 22). Der Kehrwert des Radius des Schmiegekreises an C im Punkt P ergibt den Betrag der Normalkrümmung κ_n der Fläche F im Punkt P bezüglich der Richtung des Tangentenvektors t . Die Normalkrümmung einer Kurve C auf S durch P ist definiert durch

$$\kappa_n = \kappa \cos \theta$$

wobei k die Krümmung von C im Punkt P ist und θ der Winkel zwischen der Oberflächennormalen und der Normalen der Kurve. Die so berechnete Krümmung ist allerdings noch nicht vorzeichenbehaftet. Die Normalkrümmung ist negativ, falls das Skalarprodukt von n mit dem Hauptnormalenvektor der Kurve negativ ist, anderenfalls ist sie positiv.

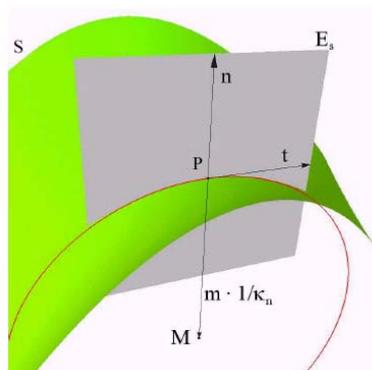


Abbildung 22: Darstellung der Normalkrümmung. Die Schnittebene E_S (grau) schneidet die Fläche S (grün). Die durch den Schnitt gebildete Kurve C wird durch den Schmiegekreis (rot) mit Mittelpunkt M im Punkt P optimal angenähert, [35]

B.2 Hauptkrümmungen und -richtungen

Variiert man den Tangentialvektor t in der Schnittebenengleichung, indem man ihn um 0° bis 180° um n rotiert, erhält man alle möglichen Normalkrümmungen κ_n . Die größte und die kleinste Normalkrümmung (κ_1 und κ_2) werden als die Hauptkrümmungen von S in P bezeichnet. Die Tangentialvektoren, die zu den beiden Hauptkrümmungen gehören, sind entsprechend die Hauptrichtungen.

Die lokale Form einer Oberfläche lässt sich mit Hilfe der Hauptkrümmungen und -richtungen eindeutig und vollständig beschreiben.

Zur Klassifizierung einer lokalen Region einer Oberfläche lassen sich aus den Hauptkrümmungen die Mittlere (H) und die Gausskrümmung (K) berechnen:

$$H = -\frac{\kappa_1 + \kappa_2}{2}$$

$$K = \kappa_1 \kappa_2$$

Eine weitere interessante Größe ist der Krümmungswert I , der oft auch als Shape Index bezeichnet wird:

$$I = \frac{2}{\pi} \arctan \frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}$$

Abbildung 23 zeigt die zugehörigen Formklassen.

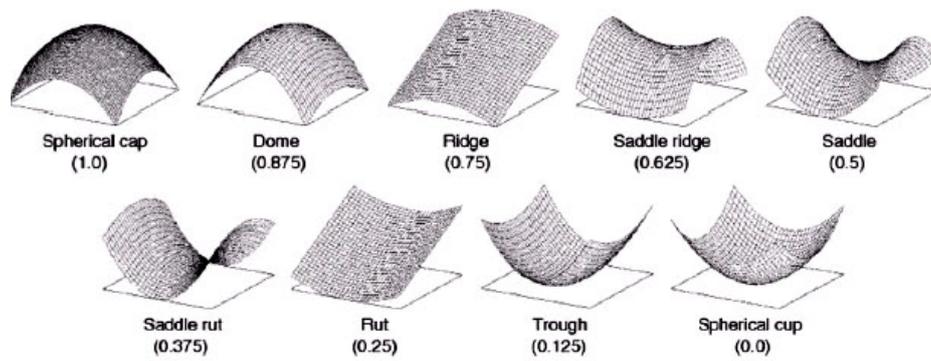


Abbildung 23: Shape Index Formklassen, [37]

Literatur

- [1] E. Trucco, A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998
- [2] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992
- [3] M. Amberg, S. Fischer, J. Röbler, *Biometrische Verfahren - Studie zum state of the art*, 2003
- [4] M. Turk, A. Pentland, *Eigenfaces for Recognition*, Journal of Cognitive Neuroscience Vol. 1, S. 71-86, 1991
- [5] C. Heshner, A. Srivastava, G. Erlebacher, *Principal Component Analysis of Range Images for Facial Recognition*,
- [6] P. N. Belhumeur, J. P. Hespana, D. J. Kriegmann, *Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection*, IEEE Transactions on Pattern Analysis and Machine Intelligence 19, S. 711-720, 1997
- [7] L. Wiskott, J.-M. Fellous, N. Krüger, C. v.d.Malsburg, *Face Recognition by Elastic Bunch Graph Matching*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, S. 775-779, 1997
- [8] L. Wiskott, *Phantom faces for face analysis*, IEEE International Conference on Image Processing, S. 308-311, 1997
- [9] G. Gordon, *Face recognition based on depth maps and surface curvature*, Geometric Methods in Computer Vision, Vol. 1570 of SPIE Proceedings, S. 234-247, 1991

- [10] G. Gordon, L. Vincent, *Application of Morphology to Feature Extraction for Face Recognition*,
- [11] X. Lu, A. Jain, *Multimodal Facial Feature Extraction for Automatic 3D Face Recognition*,
- [12] A. B. Moreno, Á. Sánchez, J. F. Vélez, F. J. Díaz, *Face recognition using 3D surface-extracted descriptors*, Irish Machine Vision and Image Processing Conference (IMVIP 2003), 2003
- [13] P. Hallinan, G. Gordon, A. Yuille, P. Giblin, D. Mumford, *Two and Three-dimensional patterns of the face*, A. K. Peters Ltd., 1999
- [14] T. K. Kim, S. C. Kee, S. R. Kim, *Real-Time Normalization and Feature Extraction of 3D Face Data Using Curvature Characteristics*,
- [15] X. Lu, D. Colbry, A. K. Jain, *Matching 2.5D Scans for Face Recognition*, ICBA, LNCS 3072, S. 30-36, 2004
- [16] X. Lu, D. Colbry, A. K. Jain, *Three Dimensional Model-Based Face Recognition*, Proc. 17th International Conference on Pattern Recognition, 2004, S. 362-366
- [17] P. Besl, N. McKay, *A Method for Registration of 3D Shapes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.14, S. 239-256
- [18] Y. Chen, G. Medioni, *Object Modeling by Registration of Multiple Range Images*, Image and Vision Computing, Vol. 10, S. 145-155, 1992
- [19] A. Rajwade, *Facial Pose Estimation and Face Recognition from Three-Dimensional Data*, 2004

- [20] Y. Wang, C. Chua, Y. Ho, *Facial Feature Detection and Face Recognition from 2D and 3D Images*,
- [21] K. W. Bowyer, K. Chang, P. Flynn, *A Survey Of 3D and Multi-Modal 3D+2D Face Recognition*, International Conference on Pattern Recognition 2004,
- [22] J. C. Lee, E. Milios, *Matching range images of human faces*, International Conference on Computer Vision, 1990, S. 722-726
- [23] H. T. Tanaka, M. Ikeda, H. Chiaki, *Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition*, International Conference on Automated Face and Gesture Recognition, S. 372-377, 1998
- [24] E. Trucco, A. Fusiello, V. Roberto, *Robust Motion and Correspondence of Noisy 3-D Point Sets with Missing Data*, Pattern Recognition Letters 20, S. 889-898, 1999
- [25] K. Backhaus, B. Erichson, W. Plinke, R. Weiber, *Multivariate Analysemethoden*, Springer Verlag Berlin, 2003
- [26] C. Bergmair, *Independent Component Analysis*
- [27] A. Hyvärinen, E. Oja, *Independent Component Analysis: Algorithms and Applications*, Neural Networks, Vol. 13, S. 411-430, 2000
- [28] Y. Pang, L. Zhang, M. Li, Z. Liu, W. Ma, *A Novel Gabor-LDA Based Face Recognition Method*, 5th Pacific Rim Conference on Multimedia, S. 352-358, 2004
- [29] D. Rybach, *Elastic Bunch Graph Matching für die Objekterkennung*, 2003

- [30] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, *Face Recognition by Elastic Bunch Graph Matching*, Proceedings of the 7th International Conference on Computer Analysis of Images and Patterns, S. 456–463, 1997
- [31] M. Braun, M. Kirschmann, *Gesichtserkennung in der Dynamic Link Architektur*,
- [32] L. Wiskott, C. von der Malsburg, *Face Recognition by Dynamic Link Matching*, 2003
- [33] D. S. Bolme, *Elastic Bunch Graph Matching*, 2003
- [34] B. Burke Hubbart, *Wavelets - Die Mathematik der kleinen Wellen*, Birkhäuser Verlag, 1997
- [35] A. Helbig, *Methoden zur Erkennung charakteristischer Eigenschaften von archäologischen Keramiken*, 2005
- [37] X. Lu, A. K. Jain, *Multimodal Facial Feature Extraction for Automatic 3D Face Recognition*
- [38] <http://de.wikipedia.org/wiki/Distanzfunktion>
- [39] M. Müller, *Finden von Punktkorrespondenzen zwischen Gesichtsbildern*, 2004
- [40] A. Colombo, C. Cusano, R. Schettini *Tri-dimensional face detection and localization*
- [41] K. Kunze, *Hauptseminar Machine Learning: Support Vector Machines, Kernels*, 2004
- [42] M. O. Stitson, J. A. Weston, A. Gammerman, V. Vovk, V. Vapnik, *Theory of Support Vector Machines*, 1996

- [43] J. Hartmann, S. Stesny, C. Sünderkamp, *Comparison of Interest Point Detectors*, 2005
- [44] *Bioidentifikation*, <http://www.bromba.com/faq/biofaqd.htm>
- [36] *The CSU Face Identification Evaluation System*, <http://www.cs.colostate.edu/evalfacerec/>