Auf dem Weg zur Gesichtserkennung unter Verwendung von Waveletnetzen

Diplomarbeit
am Lehrstuhl für Kognitive Systeme
des Instituts für Informatik und Praktische Mathematik
der Christian-Albrechts-Universität zu Kiel
Prof. Dr. Gerald Sommer

vorgelegt von

Alexander Happe

Betreuer: Dipl.-Inform. Volker Krüger

April 2000

Inhaltsverzeichnis

1 Einführung

Die rechnergestützte Erkennung von Gesichtern gehört zu den interessantesten und auch schwierigsten Problemstellungen der Bildverarbeitung. Die Einsatzmöglichkeiten automatischer Gesichtserkennungssysteme sind vielfältig, als Beispiele seien hier die Verifikation der Identität beispielsweise zur Zugangskontrolle, die Suche in Dateien mit Bildern von Straftätern oder das gezielte Suchen bestimmter Personen in den Aufnahmen von Überwachungskameras genannt.

Für die computergestützte Gesichtserkennung wird im allgemeinen zunächst eine Datenbank mit Gesichtsbildern akquiriert und für die spätere Suche geeignet kodiert. Ein Testgesicht wird dann mit jedem kodierten Gesicht in der Datenbank verglichen. Die **Erkennung** des Gesichts erfolgt, wenn eine signifikante Übereinstimmung des Testgesichts mit einem Gesicht aus der Datenbank gefunden wird. Die Herausforderung für ein solches System besteht darin, die Erkennung unter verschiedenen Variationen der Bedingungen zu leisten. Zu den Variationen zählen Beleuchtungsänderungen, verschiedene Gesichtsposen und Gesichtsausdrücke sowie die Skala des dargebotenen Testgesichts. Dazu wird eine geeignete Repräsentation (Kodierung) von Gesichtern benötigt, sowie ein Verfahren, die Ähnlichkeit zweier Repräsentationen feststellen zu können. Ein solches Maß für die Ähnlichkeit ist die Grundlage für die Erkennung. Ein Testgesicht wird identifiziert, indem es dem ähnlichsten gespeicherten Gesicht aus der Datenbank zugeordnet wird.

In dieser Diplomarbeit untersuchen wir die Verwendung von Waveletnetzen zur Erkennung von Gesichtern. Die in [21] für die Repräsentation von Grauwertbildern vorgestellten Waveletnetze stellen ein Bindeglied zwischen künstlichen neuronalen Netzen und der Wavelettheorie dar. Anstatt der üblichen Nichtlinearitäten neuronaler Netze werden bei Waveletnetzen zur Repräsentation von Grauwertbildern Gaborwavelets als Knoten eingesetzt. Wir verwenden Waveletnetze mit Gaborwavelets als Repräsentation von Gesichtern. Für jedes Gesicht der Datenbank wird eine solche Repräsentation erzeugt und gespeichert. Mit Gaborwavelets ist eine sehr kompakte Repräsentation von Gesichtern möglich. Wir verwenden Waveletnetze mit 42 und 64 Wavelets, die jeweils durch 6 Parameter charakterisiert sind. Damit benötigt jede Gesichtsrepräsentation lediglich einen Speicherplatz von 1 bzw. 1.5 Kilobyte.

Die Erkennung erfolgt in drei Schritten: Im ersten Schritt setzen wir die gespeicherten Repräsentationen ein, um in einem Testbild die Lage und Größe des zu erkennenden Gesichts exakt zu bestimmen. Die Gesichtsrepräsentation wird entsprechend dieser gewonnenen Information transformiert. Hierbei wird ausgenutzt, daß sich ein gesamtes Waveletnetz, eingesetzt als sogenanntes **Superwavelet**, kontinuierlich affin transformieren läßt.

In einem zweiten Schritt wird mit der transformierten Repräsentation das Testgesicht rekonstruiert. Die Rekonstruktion erfolgt durch die Projektion des Testgesichts in den durch die Wavelets eines Waveletnetzes aufgespannten Raum. Die Wavelets sind als Basisfunktionen spezifisch für das durch das Waveletnetz repräsentierte Individuum. Die Rekonstruktion liefert ein Waveletnetz als Repräsentation des Testgesichts. Für jedes Individuum der Datenbank stehen damit sowohl seine gespeicherte als auch die aus dem Testgesicht gewonnene Repräsentation zur Verfügung.

Im dritten Schritt der Erkennung wird diejenige Person, für die beide Waveletnetz-Repräsentationen am ähnlichsten sind, dem Testgesicht als Identität zugeordnet.

Im Abschnitt 1.1 skizzieren wir kurz einige wichtige Verfahren zur Gesichtserkennung. Im Abschnitt 1.2 geben wir einen Überblick über unsere Ansätze der Gesichtserkennung mit Hilfe der Waveletnetz-Repräsentation. In Abschnitt 1.3 wird die Gliederung der weiteren Arbeit beschrieben.

1.1 Einige Verfahren zur Gesichtserkennung

Einen aktuellen Überblick über die wichtigsten Verfahren zur Erkennung von Gesichtern findet man in [3]. Die Verfahren lassen sich grob in zwei Zweige unterteilen.

Lokale, merkmalsbasierte Verfahren lokalisieren zunächst markante Gesichtsmerkmale wie Augen, Nase und Mund. Aus diesen Merkmalen werden verschiedene Merkmalsvektoren gewonnen, die als Eingaben für neuronale Netze oder klassische Mustererkennungsverfahren dienen. Erfolgt die Erkennung auf der Basis der Geometrie der Gesichtsmerkmale, so spricht man von geometrischen, merkmalsbasierten Verfahren. Ein Beispiel hierfür ist in [1] zu finden.

Lades et al. stellen in [13] ein Verfahren vor, das sehr gute Erkennungsraten liefert. Das Verfahren verwendet *Dynamic-Link-Architecture* genannte Graphen, deren Knoten die Ergebnisse lokaler Gabortransformationen speichern. Die Kanten der Graphen sind mit den geometrischen Distanzen der Knoten beschriftet. Die lokalen Gabortransformationen werden pro Knoten mit einer Filterbank (*jet*) bestehend aus 40 Gaborwavelets berechnet. Für jede Person wird ein eigener Graph gespeichert. Die Knoten des Graphen sind in einem regelmäßigen Gitter angeordnet. In einem elastischen Graph-

Matching Verfahren werden die gespeicherten Graphen an das Testgesicht angepaßt und so der am besten passende Graph und damit die Identität des Testgesichts ermittelt. Wiskott et al. erweitern in [29] das Dynamic-Link-Architecture-Modell, so daß die Knoten des Graphen auf ausgesuchten Landmarken im Gesicht positioniert werden, wodurch eine größere Unabhängigkeit von der Gesichtspose erreicht wird. Zusätzlich wird das Graph-Modell zu sogenannten *Bunch Graphs* erweitert, in deren Knoten die Informationen der Jets mehrerer Personen gespeichert werden, um nicht länger für jede Person einen eigenen Graphen verwenden zu müssen.

Globale Verfahren fassen ein Gesicht als Ganzes auf, ohne explizit Gesichtsmerkmale zu finden.

In [14] dient ein Gesichtsbild als Eingabe einer Kombination neuronaler Netze zur Gesichtserkennung, bestehend aus rezeptiven Feldern, einer selbstorganisierenden Karte sowie einem vollständig verbundenen Multi-Layer-Perceptron.

Die bekanntesten globalen Verfahren basieren auf der Hauptkomponentenanalyse, auch als Karhunen-Loéve-Transformation bezeichnet [11, 28, 20]. Dazu wird das zu erkennende Gesicht in den durch die Hauptkomponenten (*eigenfaces*) aufgespannten Eigenraum projiziert und anhand von Abstandsmaßen in diesem Raum klassifiziert. Diese Verfahren sind sehr sensitiv gegenüber Änderungen in Pose, Skala und Beleuchtung, so daß entweder normierte Bedingungen vorausgesetzt werden müssen bzw. das Eingabebild vor der Erkennung bezüglich Orientierung und Skala zu normalisieren ist. Moghaddam und Pentland verwenden in [20] nicht nur den Eigenraum für Gesichter, sondern zusätzlich Eigenräume für Augen, Mund und Nase. Dieses Verfahren stellt damit eine Kombination zwischen den merkmalsbasierten und globalen Verfahren dar.

Edwards, Cootes und Taylor verwenden in [7] ihr *Active-Appearance-Model* [4] zur Gesichtserkennung. Dieser Ansatz ist modellbasiert und setzt sich aus einer Beschreibung der Gesichtsform sowie einer Parametrisierung der Grauwertvariationen zusammen. Die Repräsentation der Gesichtsform wird aus 400 Gesichtsbildern gewonnen, die jeweils mit 122 Landmarken versehen werden. Mit Hilfe einer Hauptkomponentenanalyse der Landmarken wird eine auf 23 Parametern beruhende Beschreibung der Gesichtsgeometrie gewonnen. Mit diesem Modell ist eine Formnormalisierung der Gesichtsbilder auf Basis der Landmarken möglich. Aus den normalisierten Gesichtsbildern wird wieder über die Hauptkomponentenanalyse ein Modell für die Grauwertvariationen der Gesichtsbilder entwickelt. Beide Modelle werden zu einer Gesichtsrepräsentation kombiniert, die über 80 Parameter die verarbeiteten Gesichtsbilder photorealistisch synthetisieren können. Die Erkennung erfolgt in zwei Schritten. Zunächst wird in einem Matchingverfahren das Gesichtsmodell an das zu erkennende Gesicht angepaßt. Der dabei erhaltene Parametersatz dient anschließend zur Identifikation des Gesichts. Das Active-Appearance-Model stellt eine Erweiterung zu Verfahren dar, die auf der reinen Hauptkomponentenanalyse der Grauwertverteilungen basieren, indem zusätzlich die aus den Landmarken gewonnenen geometrischen Merkmale zur Modellierung herangezogen werden.

Andere Verfahren arbeiten mit Templates. Nach geeigneten Vorverarbeitungsschritten zur Normalisierung von Skala und Orientierung wird das Testbild mit gespeicherten Templates verglichen. Das Template, mit dem die größte Korrelation erzielt wird, bestimmt die Identität des Testgesichts. Brunelli und Poggio verwenden in [2] zusätzlich zu Templates für das gesamte Gesicht Templates für die Augen-, Nasen- und Mundregion. Yuille verwendet in [31] deformierbare Templates, die rotiert, verschoben und skaliert werden können, um sie bestmöglich an das Testbild anzupassen.

1.2 Übersicht über Waveletnetze zur Gesichtserkennung

In diesem Kapitel soll nun ein Überblick über die Funktionsweise unseres Verfahrens zur Gesichtserkennung gegeben werden. Die in der Einleitung vorgestellte Grundidee wird nun ausführlicher erläutert.

Die Abbildung 1.1 zeigt schematisch unser System zur Erkennung von Gesichtern unter der Verwendung von Waveletnetzen. Im Rahmen dieser Diplomarbeit beschäftigen wir uns mit der Gesichtserkennung in einem **geschlossenen Universum**. Darunter verstehen wir, daß das Gesichtserkennungssystem in der Lage sein soll, eine feste Menge von Personen voneinander zu unterscheiden. Für jede dieser Personen steht ein **Referenzbild** in Form einer Frontalaufnahme zur Verfügung. Die Referenzbilder sind in einer **Referenzdatenbank** gespeichert. In einer weiteren Datenbank, der **Testdatenbank**, sind für jede Person der Referenzdatenbank mehrere, vom jeweiligen Referenzbild abweichende **Testbilder** gespeichert, die die Personen mit verschiedenen Gesichtsausdrücken sowie unter leicht variierenden Gesichtsposen zeigen.

Mit Hilfe von aus Gaborwavelets bestehenden Waveletnetzen werden Repräsentationen der Referenzgesichter erzeugt. Dabei stellen Waveletnetze Linearkombinationen von Basisfunktionen (**Wavelets**) dar, die aus einer Mutterfunktion durch Translation, Dilatation und Rotation abgeleitet werden. In unserem Fall verwenden wir ungerade Gaborfunktionen als Basisfunktionen. Die Basisfunktionen werden in einem Optimierungsprozeß an die Gesichtsmerkmale des Referenzgesichts angepaßt und so gewichtet, daß eine bestmögliche Approximation des Referenzgesichts durch die Linearkombination der Basisfunktionen erhalten wird. Die Güte der Approximation wird durch die Zahl der verwendeten Wavelets bestimmt. Mit einer geringen Anzahl von Wavelets erhält man eine Repräsentation auf einer groben Skala, verwendet man mehr Basisfunktionen, sind die Waveletnetze in der Lage, feinskaligere Details zu modellieren. Für jede Referenzperson wird ein individuelles Waveletnetz angepaßt und als Gesichtsrepräsentation gespeichert. Damit wird ein Gesicht durch einen spezifischen Satz von

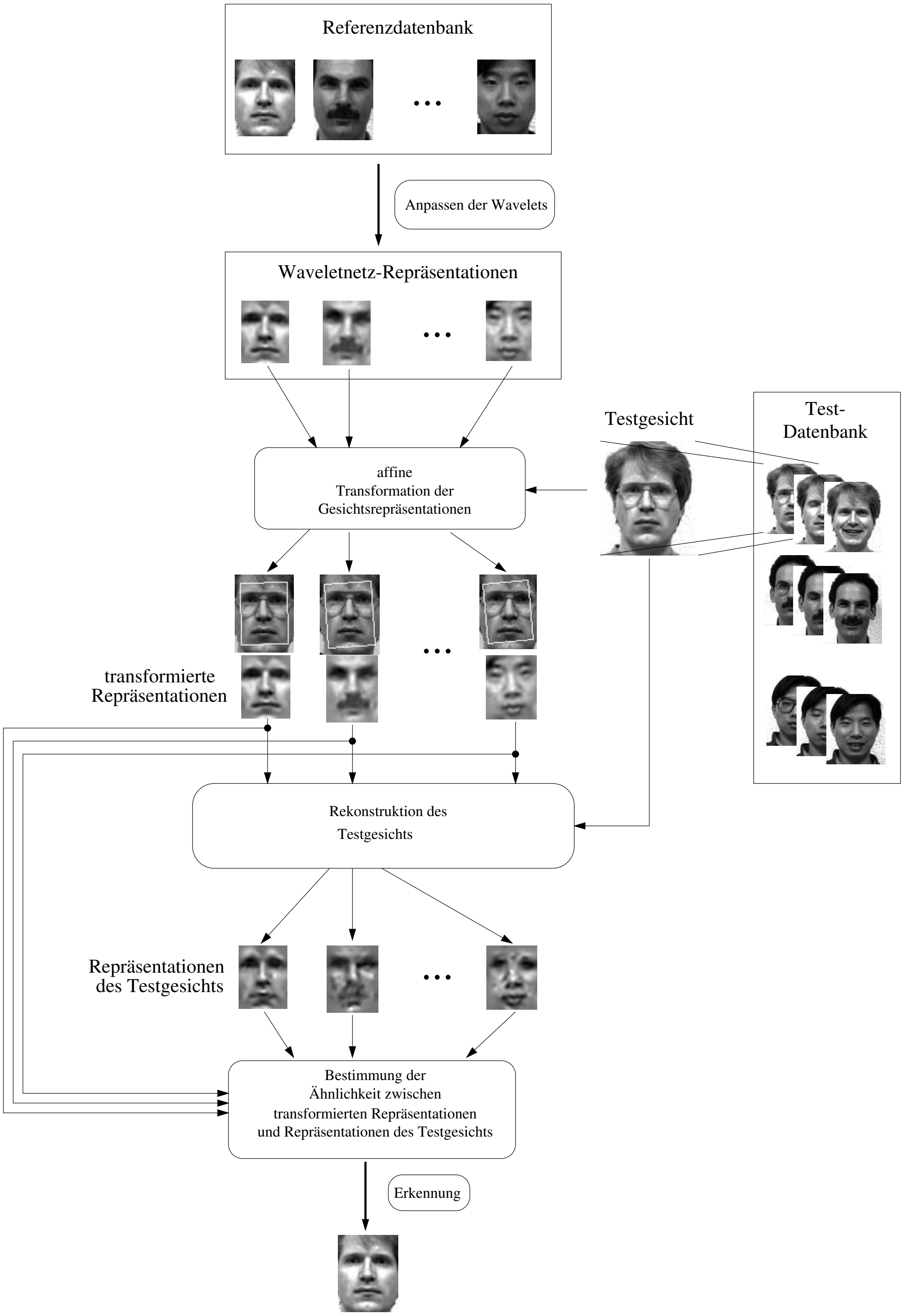


Abbildung 1.1: Übersichtsdiagramm der Gesichtserkennung mit Waveletnetzen.

Basisfunktionen in Kombination mit den zugehörigen Gewichten repräsentiert.

Als Eingabe dient dem System ein Bild aus der Testdatenbank. In dem ersten Schritt der Erkennung wird zunächst in diesem Testbild die Position, Orientierung und Größe des Gesichts festgestellt. Dazu verwenden wir die Waveletnetz-Repräsentation als kontinuierlich affin transformierbares Template, das in einem Minimierungsprozeß (**Matching**) mit dem Testgesicht in Deckung gebracht wird. Bei dem Matching-Verfahren wird das gesamte Waveletnetz als sogenanntes **Superwavelet** aufgefaßt, das wie ein Wavelet in einem kontinuierlichen Parameterraum reparametrisiert, d.h. verschoben, gedehnt und rotiert werden kann. Durch das Matching gelingt die Segmentierung des Testgesichts im Testbild. Dasselbe Matching-Verfahren kann auch zur Verfolgung von Gesichtern in Bildsequenzen verwendet werden, was wir in der Diplomarbeit ebenfalls kurz demonstrieren.

Die **Reparametrisierung** bezüglich des Testbildes wird für sämtliche gespeicherten Gesichtsrepräsentationen durchgeführt. Im zweiten Schritt des Erkennungsverfahrens wird mit den reparametrisierten Gesichtsrepräsentationen das Testgesicht **rekonstruiert**, d.h. eine neue Gewichtung bezüglich der transformierten Basisfunktionen bestimmt. Die Neubestimmung der Gewichte erfolgt durch die orthogonale Projektion des Testgesichts in den durch die reparametrisierten Basisfunktionen aufgespannten Unterraum. Damit erhalten wir für jede Person der Referenzdatenbank eine eigene Repräsentation des Testgesichts in Form eines Waveletnetzes mit transformierten und neu gewichteten Basisfunktionen.

Durch die spärliche Repräsentation mit einer geringen Anzahl von Basisfunktionen (in unseren Experimenten 42 bzw. 64) soll sichergestellt werden, daß ein individueller Satz von Basisfunktionen sehr spezifisch für das jeweilige Referenzgesicht ist. Damit ist sichergestellt, daß verschiedene Testgesichter einer Person mit den individuellen Basisfunktionen gut rekonstruiert werden können, da trotz unterschiedlicher Gesichtsausdrücke ein Großteil der Wavelets weiterhin mit den Gesichtsmerkmalen übereinstimmt. Bilder anderer Personen lassen sich im allgemeinen schlechter rekonstruieren, da eine große Zahl der Basisfunktionen schlecht zu den Merkmalen der anderen Personen passen. Diese Spezifität der individuellen Basisfunktionen ist die Grundlage für unsere Ansätze der Erkennung.

Für den dritten Schritt, die eigentliche Erkennung, benötigen wir ein Maß zur Bestimmung der **Ähnlichkeit** des Testgesichts mit den gespeicherten Referenzgesichtern. Wir untersuchen die Verwendbarkeit mehrerer **Ähnlichkeitsmaße**. Dabei unterscheiden wir Maße im Signalraum, in dem die durch Waveletnetze als Ausgabe gelieferten Grauwertbilder liegen, und Maße im Raum der Gewichte der Basisfunktionen. Diese Ähnlichkeitsmaße werden jeweils auf die bezüglich des Testbildes reparametrisierten und anschließend rekonstruierten individuellen Waveletnetze angewendet. Bei der eigentlichen Erkennung des Testgesichts bestimmt das rekonstruierte Waveletnetz, für das der höchste Wert des verwendeten Ähnlichkeitsmaßes erreicht wird, die Identität des Testbildes.

Über das bisher skizzierte Verfahren hinaus untersuchen wir eine Anwendung zur Gesichtserkennung, bei der die Repräsentation des Testgesichts nicht mit sämtlichen in einer Datenbank gespeicherten Repräsentationen verglichen wird. Das wird dann notwendig sein, wenn das Universum sehr groß ist, d.h. eine große Menge an Personen erkannt werden soll. Wir haben dabei das Szenario vor Augen, daß die Waveletnetz-Repräsentation einer Person z.B. auf einer Chipkarte gespeichert wird. Beispielsweise bei einem Zugangskontrollsystem soll verifiziert werden, ob es sich bei einem Testbild um die auf der Chipkarte gespeicherte Person handelt. Diese Verifikation soll lediglich auf der Grundlage des Vergleichs der gespeicherten Gesichtsrepräsentation und der aus dieser gewonnenen Repräsentation des Testgesichts erfolgen. Dazu führen wir ein weiteres Ähnlichkeitsmaß ein, das das rekonstruierte Testgesicht mit dem Testgesicht selbst im Signalraum vergleicht, d.h. die Korrelation zwischen der Ausgabe des rekonstruierten Waveletnetzes und dem Grauwertbild des Testgesichts wird bestimmt.

1.3 Aufbau der Arbeit

In Kapitel 2 werden Waveletnetze als Approximationsschemata vorgestellt, die Konzepte künstlicher neuronaler Netze mit Ergebnissen aus der Wavelettheorie verknüpfen. Nach einem knappen Überblick über künstliche neuronale Netze zur Funktionsapproximation in Abschnitt 2.1 und über die kontinuierliche und diskrete Wavelettransformation in Abschnitt 2.2 wird die Architektur der Waveletnetze in Abschnitt 2.3 zunächst für die Approximation eindimensionaler Funktionen vorgestellt. Der Zusammenhang von Waveletnetzen mit der Wavelettheorie und der Nutzen der Wavelettheorie für die Konstruktion von Waveletnetzen wird in Abschnitt 2.4 herausgearbeitet. Das Modell der Waveletnetze wird in Abschnitt 2.5 basierend auf den Ergebnissen der Arbeit von Alexander Pelc [21] für die Repräsentation von Grauwertbildern erweitert. In diesem Abschnitt stellen wir die gegenüber [21] verbesserte Initialisierung und Optimierung sowie ein neues Verfahren zur Bestimmung der Waveletgewichte vor. Abschließend wird als grundlegende Erweiterung gegenüber [21] das Konzept der Superwavelets beschrieben, das die elementare Grundlage für die Nutzung der Waveletnetze zur Gesichtserkennung ist. Dabei wird ein gesamtes Waveletnetz als (Super-)Wavelet aufgefaßt, das kontinuierlich verschoben, skaliert und rotiert werden kann. Damit ist ein Waveletnetz als affin veränderbares Template anwendbar, dessen Parameter kontinuierlich änderbar sind, ohne die bei herkömmlichen Templates auftretenden Diskretisierungsprobleme bei Rotationen, stufenloser Skalierung sowie subpixelgenauer Positionierung

In Kapitel 3 stellen wir unsere Ansätze zur Gesichtserkennung mit Waveletnetzen vor. In Abschnitt 3.1 wird die Erzeugung einer Gesichtsrepräsentation mittels Waveletnetzen beschrieben. In Abschnitt 3.2 beschreiben wir den ersten Schritt des Erkennungsverfahrens, die Verwendung der Waveletnetz-Repräsentationen in Form von Superwavelets zur Bestimmung von Größe, Orientierung und Position eines Gesichts in einem Testbild. In Abschnitt 3.3 zeigen wir, wie das in Abschnitt 3.2 vorgestellte Matching-Verfahren zur Verfolgung von Gesichtern in Videobildsequenzen eingesetzt werden kann. Den zweiten Schritt der Erkennung, die Rekonstruktion des Testgesichts durch die affin transformierte Gesichtsrepräsentation, erläutern wir in Abschnitt 3.4. In Abschnitt 3.5 beschreiben wir die im letzten Schritt der Erkennung benötigten Ähnlichkeitsmaße für Waveletnetz-Repräsentationen. Die eigentliche Gesichtserkennung wird in Abschnitt 3.6 erläutert.

Des weiteren stellen wir in Abschnitt 3.7 Ansätze vor, die Identität einer Person zu verifizieren, indem lediglich eine gespeicherte Repräsentation mit der daraus gewonnenen Repräsentation eines Testgesichts verglichen wird.

Die Ergebnisse von Experimenten mit den in Kapitel 3 eingeführten Verfahren stellen wir in Kapitel 4 vor.

Kapitel 5 beschließt die Arbeit mit einer Zusammenfassung und einem Ausblick.

2 Waveletnetze

Im Kontext dieser Diplomarbeit verwenden wir Waveletnetze im weiteren Sinne zur Approximation von Funktionen. Eine Möglichkeit der Funktionsapproximation ist die Verwendung künstlicher neuronaler Netze, z.B. Radialer-Basisfunktions-Netze (RBF-Netze). Waveletnetze verbinden das Konzept dieser RBF-Netze mit der Wavelettheorie, indem statt radialer Basisfunktionen Wavelets als Nichtlinearitäten der Knoten eines Netzes verwendet werden.

In Abschnitt 2.1 wird ein kurzer Abriß über die Verwendung neuronaler Netze zur Funktionsapproximation gegeben. In Abschnitt 2.2 werden die Begriffe *Wavelet* und *Wavelettransformation* eingeführt. Auf dieser Grundlage kann in Abschnitt 2.3 die Architektur der Waveletnetze erläutert werden, deren Verbindung zur Wavelettheorie in Abschnitt 2.4 dargestellt wird. In Abschnitt 2.5 beschreiben wir dann die Anwendung von Waveletnetzen zur Approximation von Grauwertbildern.

2.1 Neuronale Netze zur Funktionsapproximation

Mit künstlichen neuronalen Netzen werden biologische neuronale Netze als informationsverarbeitende Systeme nachgeahmt. Wir bezeichnen im folgenden künstliche neuronale Netze vereinfachend als neuronale Netze. In den vergangenen Jahren sind neuronale Netze für ein weites Spektrum an Aufgaben eingesetzt worden, z.B. für Probleme der Mustererkennung oder zum Erlernen nichtlinearer Abbildungen.

Wir betrachten hier vorwärtsgerichtete neuronale Netze, die aus einer oder mehreren Schichten von Berechnungselementen (Knoten, Neuronen) bestehen, wobei die Knoten jeweils nur mit Knoten der vorherigen oder nächsten Schicht verbunden sind. Die Verbindungen sind gerichtet, d.h. der Informationsfluß verläuft von den Knoten einer Eingabeschicht zu den sogenannten verdeckten Schichten und von dort zur Ausgabeschicht. Des weiteren sind die Verbindungen gewichtet, d.h. die Ausgänge der Knoten werden mit skalaren Werten multipliziert. Abbildung 2.1 zeigt ein vorwärtsgerichtetes Netz mit einer Eingabeschicht, einer verdeckten Schicht und einer aus einem Neuron bestehenden Ausgabeschicht. Neuronale Netze repräsentieren damit eine nichtlineare

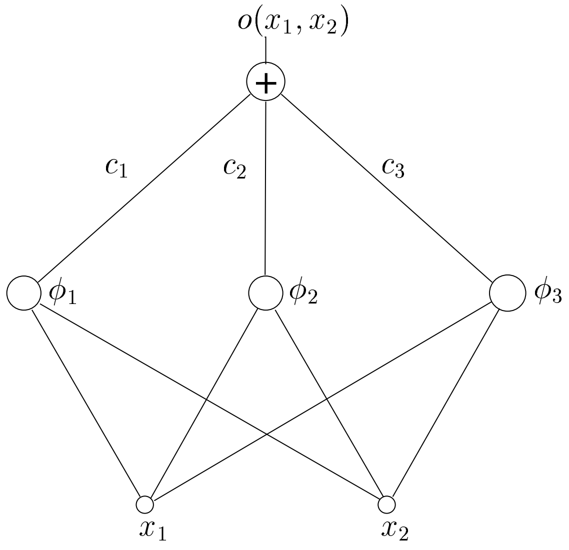


Abbildung 2.1: Vorwärtsgerichtetes neuronales Netz.

Abbildung von Eingangsvariablen (Eingänge der Knoten der Eingabeschicht) auf Ausgangsvariablen (Ausgänge der Knoten der Ausgabeschicht). Wir beschränken uns in der folgenden Darstellung auf vorwärtsgerichtete Netze mit d Knoten in der Eingangschicht, einer verdeckten Schicht mit m Knoten und einem Knoten in der Ausgabeschicht. Diese Netze lassen sich allgemein mathematisch darstellen als Abbildung

$$o : \mathbb{R}^d \to \mathbb{R},\ o(\boldsymbol{x}) = \sum_{i=1}^{m} c_i \phi_i(\boldsymbol{x}) \ , \tag{2.1}$$

mit ϕ_i als die mit dem Knoten i assoziierte Berechnungsfunktion. Haben die Funktionen ϕ_i die Form

$$\phi_i = \phi_i(\boldsymbol{x}; \boldsymbol{w}_i) = \sigma(\boldsymbol{w}_i^T \boldsymbol{x} + \theta_i)$$

mit σ als sigmoidale Aktivierungsfunktion, $\boldsymbol{w}_i$ als Vektor mit skalaren Gewichten und skalarem Schwellwert θ_i, so handelt es sich um ein *backpropagation*-Netz oder *multi-layer-perceptron*; bei

$$\phi_i = \phi(||\boldsymbol{x} - \boldsymbol{t}_i||)$$

handelt es sich um ein RBF-Netz mit radialen Basisfunktionen ϕ_i, deren Zentren durch $\boldsymbol{t}_i$ gegeben sind.

Netze der Gestalt (2.1) können unter dem Gesichtspunkt der Funktionsapproximation betrachtet werden. Eine (nichtlineare) Abbildung f soll durch ein neuronales

Netz mit Ausgabe o approximiert werden, so daß $f(\boldsymbol{x}) \approx o(\boldsymbol{x})$ gilt. In einem Lern- oder Trainingsprozeß sollen aus einer Menge $\mathcal{T} = \{(\boldsymbol{x}^{(1)}, y^{(1)}), \ldots, (\boldsymbol{x}^{(N)}, y^{(N)})\}$ von Ein-/Ausgabepaaren die Parameter der Funktionen ϕ_i so angepaßt werden, daß eine bestmögliche Approximation von f durch o erreicht wird.

2.2 Wavelettransformation

Die **Wavelettransformation** ist ein nützliches Hilfsmittel der Signalanalyse. Die Zielsetzung der Signalanalyse ist es, relevante Informationen aus einem Signal zu extrahieren, indem dieses transformiert wird. Ein Beispiel einer Transformation ist die Fouriertransformation eines eindimensionalen reell- oder komplexwertigen Signals $f \in L^2(\mathbb{C})$

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \;, \tag{2.2}$$

die das Signal in den Frequenzraum transformiert.

Damit ist die Fouriertransformation eine lineare Abbildung in den durch die Basisfunktionen

$$b_\omega(t) = e^{i\omega t} = \cos(\omega t) + i \sin(\omega t)$$

aufgespannten Vektorraum. Diese Basisfunktionen sind räumlich (bzw. zeitlich) unendlich ausgedehnt, also räumlich (zeitlich) nicht lokalisiert. Damit sind im Frequenzraum keine Informationen über die räumliche (zeitliche) Gestalt des Signals enthalten.

Im Hilbertraum $L^2(\mathbb{C})$ der quadratintegrierbaren Funktionen über $\mathbb{C}$ ist das innere Produkt oder Skalarprodukt zweier Funktionen g_1 und g_2 gegeben durch

$$<g_1, g_2> = \int_{-\infty}^{+\infty} g_1(t) g_2^*(t) \; dt \;. \tag{2.3}$$

Gleichung (2.2) besagt damit, daß sich die Fouriertransformierte als inneres Produkt des Signals f mit den Basisfunktionen b_ω ergibt. Das ist gleichbedeutend mit einer orthogonalen Projektion des Signals auf die Basisfunktionen.

Mit Hilfe der Fouriertransformierten $F(\omega)$ läßt sich f darstellen als

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega t} d\omega \;. \tag{2.4}$$

Die Wavelettransformation ist wie die Fouriertransformation (2.2) eine Integraltransformation in einen Raum, der durch die Waveletbasisfunktionen (**Wavelets**) aufgespannt wird. Die Basisfunktionen

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad a > 0,\ b \in \mathbb{R}, \tag{2.5}$$

entstehen durch Dilatation und Translation aus einem "**Mutterwavelet**" (*motherwavelet*) ψ. Damit ergibt sich die **kontinuierliche Wavelettransformation** als

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t)\psi^*_{a,b}(t)\ dt\ . \tag{2.6}$$

Damit das Integral in (2.6) konvergiert, muß ein Mutterwavelet ψ einer sogenannten **Zulässigkeitsbedingung** genügen [16]:

$$K_\psi = \int_{-\infty}^{+\infty} \frac{|\mathcal{F}\{\psi\}(t)|^2}{|t|}\ dt\ , \tag{2.7}$$

mit $\mathcal{F}\{\psi\}$ als Fouriertransformierten von ψ. Aus dieser Bedingung folgt, daß ψ als Impulsantwort eines Bandpaßfilters interpretiert werden kann und somit mittelwertfrei ist, d.h.

$$\int_{-\infty}^{+\infty} \psi(t)\ dt = 0\ . \tag{2.8}$$

Gleichung (2.6) besagt, daß die Wavelettransformierte $W_f(a,b)$ das Skalarprodukt von f und der Basisfunktion $\psi_{a,b}$ ist, also durch die orthogonale Projektion von f auf $\psi_{a,b}$ gewonnen wird. Durch $W_f(a,b)$ wird gemessen, wie ähnlich f dem Wavelet $\psi_{a,b}$ ist.

Die Funktion f läßt sich aus der Wavelettransformierten durch

$$f(t) = K_\psi^{-1} \int_{a>0} \int_{-\infty}^{+\infty} W_f(a,b)\psi_{a,b}\frac{dadb}{a^2} \tag{2.9}$$

zurückgewinnen. Zu beachten ist, daß gemäß Gleichung (2.7) K_ψ eine lediglich von ψ und nicht vom Signal abhängige Konstante ist.

Die Wavelettransformation kann diskretisiert werden, indem sowohl der Dilatationsparameter a als auch der Verschiebungsparameter b abgetastet werden. In [5] wird folgendes Abtastschema vorgeschlagen, auf das wir noch in Abschnitt 2.3 zurückkommen werden:

$$\begin{aligned} a &= a_0^j \qquad a_0 > 1 \\ b &= kb_0a_0^j \\ \psi_{j,k}(t) &= a_0^{-j/2}\psi(a_0^{-j}t - kb_0) \qquad j,k \in \mathbb{Z}\ , \end{aligned} \tag{2.10}$$

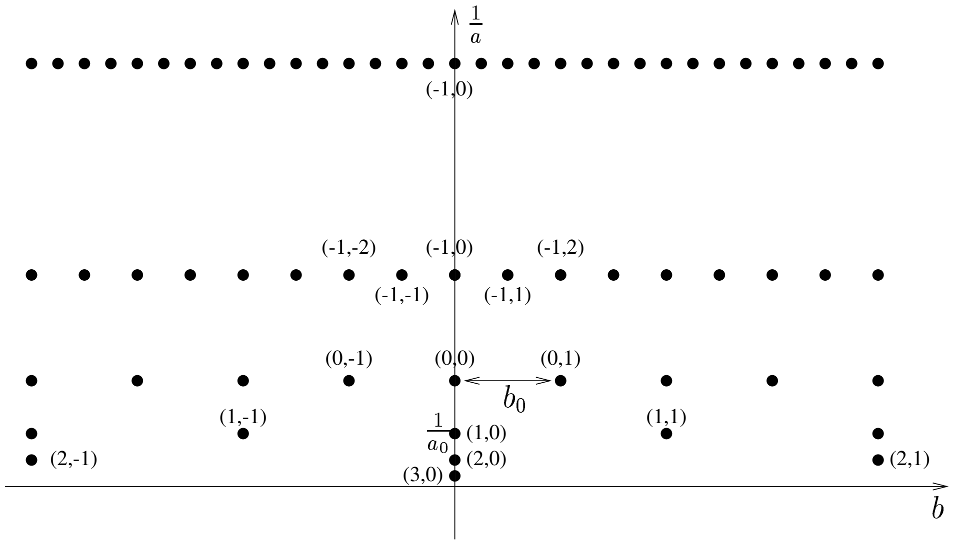


Abbildung 2.2: Diskretisierung der Wavelettransformation. Abtastschema gemäß Gleichung (2.10) für $a_0 = 2$. In Klammern sind jeweils die Werte für (j,k) angegeben.

mit a_0 und b_0 als Schrittweiten der Diskretisierung. Abbildung 2.2 illustriert das Abtastschema für den Fall $a_0 = 2$.

Damit ergeben sich die Koeffizienten der **diskreten Wavelettransformation** zu

$$c_{j,k} = \int_{-\infty}^{+\infty} f(t)\psi_{j,k}(t)\ dt. \tag{2.11}$$

Das Rekonstruktionsproblem besteht darin, a_0, b_0 und $\psi(t)$ so zu wählen, daß

$$f(t) \approx \sum_j \sum_k c_{j,k}\psi_{j,k}(t) \tag{2.12}$$

gilt. Im Abschnitt 2.4 gehen wir im Kontext von *frames* näher auf diesen Aspekt ein.

2.3 Architektur der Waveletnetze

Zhang und Benveniste führen in [32] **Waveletnetze** ein, die vorwärtsgerichtete neuronale Netze mit einer verdeckten Schicht sind und Konzepte der Theorie der neuronalen Netze und der Wavelettheorie vereinen, indem als Nichtlinearitäten Wavelets gewählt werden. Netze vergleichbarer Struktur wurden von Szu et al. in [27] und [26] vorgestellt.

Zunächst betrachten wir Waveletnetze mit eindimensionalen Eingaben $x \in \mathbb{R}$. Ein **Waveletnetz** $\mathcal{W}$ stellt die Abbildung

$$z(x) = \sum_{i=1}^{N} w_i \psi(s_i(x - t_i)) \quad s_i > 0,\ w_i,\ t_i \in \mathbb{R} \tag{2.13}$$

dar, wobei ψ das Mutterwavelet, N die Anzahl der Wavelets im Netz, s_i die Skalierungs- und t_i die Translationsparameter und w_i skalare Gewichte sind.

Ein Vergleich von (2.13) mit (2.1) zeigt sofort die Verbindung der Waveletnetze zu herkömmlichen neuronalen Netzen. Vergleicht man (2.13) mit Gleichung (2.12) erkennt man, wie Waveletnetze durch die diskrete Wavelettransformation motiviert sind.

Faßt man die Skalierungsparameter s_i und die Translationsparameter t_i zu Parametervektoren

$$\boldsymbol{a}_i = (s_i, t_i)^T$$

zusammen, läßt sich (2.13) schreiben als

$$z(x) = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}(x)\ . \tag{2.14}$$

Mit

$$\boldsymbol{\Psi} = \{\psi_{\boldsymbol{a}_1}, \dots, \psi_{\boldsymbol{a}_N}\}$$

als Familie der Basisfunktionen und dem Gewichtsvektor

$$\boldsymbol{w} = (w_1, \dots, w_N)^T$$

läßt sich ein Waveletnetz $\mathcal{W}$ als Tupel

$$\mathcal{W} = (\boldsymbol{\Psi}, \boldsymbol{w}) \tag{2.15}$$

schreiben.

Wie oben erwähnt, wenden wir Waveletnetze zur Approximation von Funktionen an. Um dieses Ziel zu erreichen, müssen in einem Trainingsprozeß die Parameter $\boldsymbol{a}_i$ der Wavelets und der Gewichtsvektor $\boldsymbol{w}$ so angepaßt werden, daß die zu approximierende Funktion $f : \mathbb{R} \to \mathbb{R}$ durch die Ausgabe z eines Netzes $\mathcal{W}$ mit möglichst geringem Approximationsfehler dargestellt wird. Da die einzelnen Wavelets $\psi_{\boldsymbol{a}_i}$ mittelwertfrei sind, kann der Gleichanteil $\bar{f}$ durch (2.14) nicht modelliert werden, so daß der Gleichanteil $\bar{f}$ explizit modelliert werden muß:

$$z(x; \boldsymbol{a}, \boldsymbol{w}) = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}(x) + m\ , \tag{2.16}$$

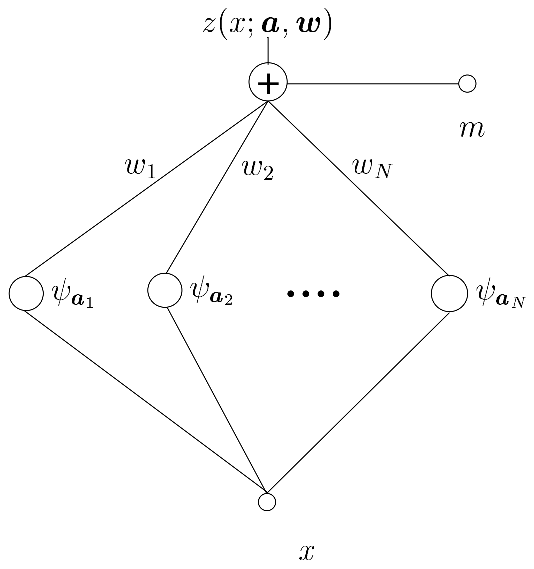


Abbildung 2.3: Waveletnetz mit eindimensionaler Eingabe.

mit $m = \bar{f}$ und $\boldsymbol{a} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_N)^T$ als Gesamtparametervektor des Waveletnetzes. Abbildung 2.3 zeigt die Struktur eines Waveletnetzes. Der Mittelwert m wird zusammen mit den Parametern $\boldsymbol{a}_i$ der Basisfunktionen und den Gewichten w_i gespeichert. Deshalb erweitern wir die Tupelschreibweise für Waveletnetze und erhalten

$$\mathcal{W} = (\Psi, \boldsymbol{w}, m) \ . \tag{2.17}$$

Zum Trainieren des Waveletnetzes wird die zu approximierende Funktion geeignet abgetastet, um eine Trainingsmenge bestehend aus Eingabewerten $\mathcal{X} = \{x^{(1)}, \dots, x^{(M)}\}$ und Sollwerten $\mathcal{Y} = \{y^{(1)}, \dots, y^{(M)}\}$ mit $y^{(i)} = f(x^{(i)})$ zu erhalten. Durch ein geeignetes Trainingsverfahren ist das Minimierungsproblem

$$(\boldsymbol{a}_{\text{opt}}, \boldsymbol{w}_{opt}) = \min_{(\boldsymbol{a}, \boldsymbol{w})} \sum_{j=1}^{M} \left(y^{(j)} - z(x^{(i)}; \boldsymbol{a}, w) \right)^2 \tag{2.18}$$

zu lösen.

Zhang und Benveniste verwenden in [32] ein stochastisches Gradientenverfahren für das Training von Waveletnetzen. In Kapitel 2.5 beschreiben wir das in dieser Diplomarbeit verwendete Optimierungsverfahren. Hervorzuheben ist, daß sowohl die Gewichte w_i als auch die Parameter der Wavelets (Dilatations- und Translationsparameter) kontinuierlich angepaßt werden.

2.4 Waveletnetze aus der Sicht der Wavelettheorie

In diesem Abschnitt soll der Bezug der Waveletnetze zur Theorie der diskreten und kontinuierlichen Wavelettransformation verdeutlicht werden. Im vorherigen Abschnitt wurde durch den Vergleich von Gleichung (2.13) und (2.12) auf die strukturelle Ähnlichkeit der Ausgabe eines Waveletnetzes und der Rekonstruktion aus der diskreten Wavelettransformation hingewiesen. Die Wavelettheorie liefert als Ergebnis, daß eine Familie

$$\Psi = \{\psi_{j,k} = a_0^{-j/2}\psi(a_0^{-j}t - kb_0)\ ,\ j,k \in \mathbb{Z}\} \tag{2.19}$$

von Wavelets bei geeigneter Wahl von ψ, a_0 und b_0 einen sogenannten *frame* bilden können [5]. In einem Hilbertraum $\mathcal{H}$ (z.B. $L^2(\mathbb{R})$, $L^2(\mathbb{C})$) wird eine Familie

$$\{\psi_{j,k}\ ; j,k \in \mathbb{Z}\}$$

frame genannt, wenn es Konstanten $A > 0$ und $B < \infty$ gibt, so daß für alle $f \in \mathcal{H}$ gilt

$$A||f||^2 \leq \sum_{j,k\in\mathbb{Z}} |<f,\psi_{j,k}>|^2 \leq B||f||^2 \tag{2.20}$$

wobei $||f|| = \int_{-\infty}^{+\infty} |f(x)|^2\ dx$ die L^2-Norm von f ist. Eine Folgerung der *frame*-Bedingung (2.20) ist, daß ein *frame* $\{\psi_{j,k}\}$ dicht in $\mathcal{H}$ liegt, also eine vollständige Repräsentation der Funktionen f mit Hilfe der Koeffizienten der diskreten Wavelettransformation mit den Wavelets aus $\psi_{j,k}$ möglich ist. Im Falle nichtorthogonaler Wavelets – wie wir sie für die Repräsentation von Grauwertbildern verwenden – ist dazu ein *dualer frame* $\tilde{\Psi}$ notwendig, der sich aus dem *frame* Ψ berechnen läßt [5]. Es gilt

$$f = \sum_{j,k} <f,\psi_{j,k}>\tilde{\psi}_{j,k} = \sum_{j,k} <f,\tilde{\psi}_{j,k}>\psi_{j,k}\ . \tag{2.21}$$

Wenn die Grenzen A und B gleich sind, dann liegt ein sogenannter *tight frame* vor. Dann sind die Funktionen $\tilde{\psi}_{j,k}$ des dualen *frames* bis auf eine Konstante identisch mit den Funktionen $\psi_{j,k}$ des *frames*:

$$\tilde{\psi}_{j,k} = A^{-1}\psi_{j,k} \qquad \text{falls } A = B\ . \tag{2.22}$$

Im Fall $A = B = 1$ bildet der *frame* eine orthonormale Basis und Ψ und $\tilde{\Psi}$ sind identisch.

Ein Waveletnetz ist nichts anderes als eine endliche Linearkombination von Wavelets, so daß die Ergebnisse der Wavelettheorie in die Wahl geeigneter Wavelets und die Konstruktion von Waveletnetzen einfließen können [32].

Durch die Anwendung des in Abschnitt 2.3 angesprochenen Optimierungsverfahrens stellen Waveletnetze im gewissen Sinne ein Bindeglied zwischen der kontinuierlichen und diskreten Wavelettransformation dar. Bei der diskreten Wavelettransformation wird der Parameterraum der Wavelets nach einem festen Schema (z.B. gemäß (2.19)) abgetastet. Die Anpassung der Waveletparameter der Wavelets erfolgt bei Waveletnetzen im **kontinuierlichen** Parameterraum. Die zu approximierende Funktion f wird nicht durch einen festen, universellen Satz von Basisfunktionen, sondern durch speziell an die Funktion f angepaßte Funktionen approximiert. Die Anpassung der Waveletparameter an f erlaubt eine Approximation mit einer erheblich geringeren Anzahl von Basisfunktionen gegenüber der Verwendung von *frames*.

2.5 Waveletnetze zur Repräsentation von Grauwertbildern

Bisher haben wir Waveletnetze mit eindimensionalen Eingaben vorgestellt. Um Grauwertbilder mit Waveletnetzen repräsentieren zu können, muß das Konzept der Waveletnetze auf zwei Dimensionen erweitert werden. Dies wird in Abschnitt 2.5.1 erläutert. In Abschnitt 2.5.2 werden die von uns verwendeten Gaborwavelets beschrieben. Nach der Beschreibung der Initialisierung eines Netzes in Abschnitt 2.5.3 wird das Optimierungsverfahren zum Anpassen der Wavelets in Abschnitt 2.5.4 vorgestellt. In Abschnitt 2.5.5 werden die Unterschiede der Repräsentation durch Waveletnetze zu einer vollständigen Waveletrepräsentation dargestellt. Unser entwickeltes Verfahren, zu gegebenen Wavelets einen optimalen Gewichtsvektor zu bestimmen, stellen wir in Abschnitt 2.5.6 vor. Abschnitt 2.5.7 beschreibt, wie ein Waveletnetz als Linearkombination von Wavelets seinerseits wieder als Mutterwavelet aufgefaßt werden kann, aus dem durch Dilatation, Rotation und Translation neue Wavelets erzeugt werden können. Für diese übernehmen wir den in [27] eingeführten Begriff des **Superwavelets**.

2.5.1 Parametrisierung der Wavelets

Ein Grauwertbild g mit $K \times L$ Pixeln stellt eine digitale Bildsignalfunktion dar, die aus einem kontinuierlichen Meßsignal durch Abtastung und Quantisierung (hier in 256 Grauwertstufen) entstanden ist:

$$g : [0, \dots, K-1] \times [0, \dots, L-1] \to [0, \dots, 255] \; .$$

Um die zweidimensionale Bildfunktion g zu approximieren, benötigen wir ein zweidimensionales Mutterwavelet $\psi : \mathbb{R}^2 \to \mathbb{R}$. Aus diesem Mutterwavelet werden analog zum eindimensionalen Fall Wavelets durch Dilatation und Verschiebung gewonnen, wobei wir nun im Gegensatz zum eindimensionalen Fall vier Freiheitsgrade haben:

- t_x Translation in x-Richtung,
- t_y Translation in y-Richtung,
- d_x Dilatation in x-Richtung,
- d_y Dilatation in y-Richtung.

Nicht rotationssymmetrische Mutterwavelets, wie wir sie verwenden werden, verfügen über einen weiteren Freiheitsgrad:

- α Rotation des Wavelets um das durch t_x und t_y gegebene Zentrum.

Fassen wir die Freiheitsgrade der Wavelets jeweils zu einem Parametervektor

$$\boldsymbol{a}_i = (t_{x_i}, t_{y_i}, d_{x_i}, d_{y_i}, \alpha_i)^T$$

zusammen, ergeben sich aus dem Mutterwavelet durch Translation, Dilatation und Rotation die Wavelets

$$\psi_{\boldsymbol{a}_i}(\boldsymbol{x}) = \psi\left(\boldsymbol{D}_i \boldsymbol{R}_i(\boldsymbol{x} - \boldsymbol{t}_i)\right) \tag{2.23}$$

mit $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\boldsymbol{t}_i = \begin{pmatrix} t_{x_i} \\ t_{y_i} \end{pmatrix}$, $\boldsymbol{D}_i = \begin{pmatrix} d_{x_i} & 0 \\ 0 & d_{y_i} \end{pmatrix}$ und $\boldsymbol{R}_i = \begin{pmatrix} \cos\alpha_i & -\sin\alpha_i \\ \sin\alpha_i & \cos\alpha_i \end{pmatrix}$.

Mit dem Gleichanteil $m = \bar{g}$, dem Gewichtsvektor $\boldsymbol{w}$ und Ψ als Familie von Wavelets $\psi_{\boldsymbol{a}_i}$ erhalten wir das Waveletnetz $\mathcal{W} = (\Psi, \boldsymbol{w}, m)$, durch dessen Ausgabe $z(\boldsymbol{x})$ das Grauwertbild g approximiert werden soll:

$$g(\boldsymbol{x}) \approx z(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}(\boldsymbol{x}) + m\ . \tag{2.24}$$

2.5.2 Gabor Wavelets

In [21] wurden Waveletnetze zur Repräsentation von Bildern vorgestellt. Dabei haben sich Gaborfunktionen als Wavelets bewährt. Gaborfunktionen sind mit einer Gaußfunktion modulierte komplexharmonische Funktionen der Form

$$h(x, y) = e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} e^{-i(\omega_x x + \omega_y y)}\ . \tag{2.25}$$

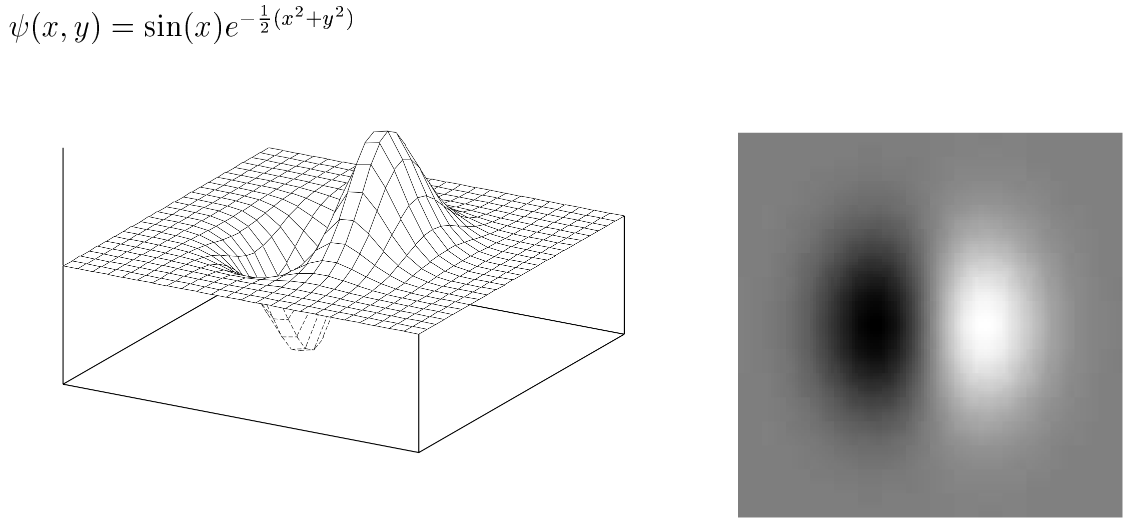
Gaborfunktionen sind in der Bildverarbeitung und Signaltheorie wohlbekannt. So sind sie die Basisfunktionen der gefensterten Fouriertransformation, bei der eine räumliche

Lokalisation durch die Wahl der Gaußfunktion als Fensterfunktion erzielt wird. Somit eignen sich auf Gaborfunktionen basierende Filter zur orientierungsselektiven spektralen Analyse. Das Gesichtserkennungssystem von Wiskott et al. verwendet als Eingabe die Filterantworten von Filterbänken bestehend aus Gaborfiltern [29]. Die Verwendung von Gaborfunktionen ist auch biologisch motivierbar, so sind im visuellen Kortex von Katzen Zellen zu finden, deren Antwortverhalten lokalen Frequenzmessungen entsprechen und gut durch Gaborfunktionen modellierbar sind [18, 10]. Daugman verwendet in [6] auf diesen Untersuchungen basierende Gaborfunktionen zur Repräsentation und Kompression von Bildern. Lee untersucht in [15] die Repräsentation von Bildern mit aus Gaborfunktionen gewonnenen Wavelets, deren Parametrisierung sich ebenfalls auf die physiologischen Beobachtungen z.B. in [10] stützt.

Wie in [21] verwenden wir für die Waveletnetze den ungeraden Imaginärteil der Gaborfunktion (2.25) als Mutterwavelet

$$\psi(x,y) = \sin(x)e^{-\frac{1}{2}(x^2+y^2)} \,. \tag{2.26}$$

Abbildung 2.4 zeigt dieses Mutterwavelet.



(a) (b)

Abbildung 2.4: Gabormutterwavelet als Oberflächenplot (a) und abgetastet als Grauwertbild (b).

2.5.3 Initialisierung

Die Wavelets eines Waveletnetzes werden durch ein Optimierungsverfahren an die zu approximierende Funktion angepaßt. Für die schnelle Konvergenz eines Optimierungsverfahrens ist eine gute Initialisierung als Ausgangspunkt für die Optimierung notwendig. Ein guter Startwert ist insbesondere dann wichtig, wenn das angewendete Optimierungsverfahren kein globales sondern lediglich ein lokales Minimum findet. Je dichter der Startpunkt für die Optimierung an einem globale Minimum liegt, desto größer ist die Chance, daß die Optimierung in einem lokalen Minimum nahe beim globalen Minimum konvergiert.

Die Initialisierung eines Waveletnetzes dient als Startwert für das Optimierungsverfahren zur Anpassung der Wavelets. Wir verwenden wie in [21] ein Initialisierungsschema, das an das in 2.2 vorgestellte Abtastschema bei der diskreten Wavelettransformation angelehnt ist. Die Zentren und Dilatationsparameter der Wavelets werden in Form einer Auflösungspyramide initialisiert, wie es in Abbildung 2.5 symbolisch dargestellt ist. Solche Pyramiden sind in der Bildverarbeitung als Mehrgitterdatenstrukturen bekannt und werden in Form von Gauß- und Laplacepyramiden eingesetzt. Sie erlauben eine Mehrskalenrepräsentation von Bildern. Näheres dazu ist z.B. in [9] zu finden.

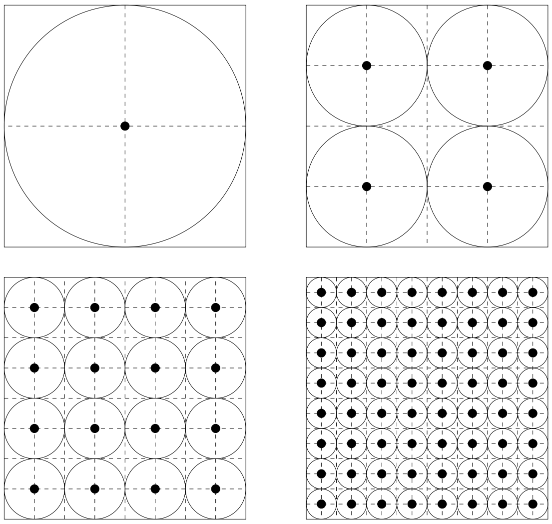


Abbildung 2.5: Initialisierung der ersten 4 Ebenen eines Waveletnetzes. Die Kreise geben den Abstand vom Zentrum der Wavelets an, bei dem die einhüllende Gaußfunktion auf die Hälfte ihrer Amplitude abgefallen ist.

Zur Realisierung der pyramidenartigen Initialisierung der Waveletnetze werden die Dilatationsparameter des Gaborwavelets auf der ersten, gröbsten Ebene der Pyramide so gewählt, daß die Amplitude der einhüllenden Gaußfunktion am Bildrand auf die

Hälfte ihres Wertes im Zentrum $\boldsymbol{t}$ abgefallen ist. Von Ebene zu Ebene verdoppeln sich jeweils die Dilatationsparameter d_x und d_y, die Abstände der Waveletzentren halbieren sich dementsprechend. Für jeden in Abbildung 2.5 dargestellten Aufpunkt wird lediglich ein Wavelet verwendet, im Unterschied zur vollständigen Waveletrepräsentation in [15], wo pro Abtastpunkt mehrere Orientierungen und Skalen verwendet werden.

Für die Initialisierung der Orientierung eines Wavelets bieten sich zwei Strategien an. Entweder wird wie in [21] das Wavelet mit einer zufälligen Orientierung initialisiert, oder es wird die Orientierung gewählt, bei der der Betrag der Projektion $<\psi_{\boldsymbol{a}_i}, g>$ des Wavelets auf das Bild maximal wird. Für die Experimente dieser Diplomarbeit haben wir letztere Vorgehensweise bei der Initialisierung der Orientierungen gewählt. Dazu werden pro Wavelet die Projektionskoeffizienten $<\psi_{\boldsymbol{a}_i}, g>$ für 16 äquidistant verteilte Orientierungen berechnet und die Orientierung gewählt, bei der $|<\psi_{\boldsymbol{a}_i}, g>|$ ein Maximum einnimmt. Wie in Abbildung 2.7 demonstriert, führt dieses Verfahren zu deutlich besseren Initialisierungen und daraus resultierend zu besseren Approximationen nach dem Optimierungsverfahren, als es bei einer zufälligen Initialisierung der Orientierungen der Fall ist. Abbildung 2.6 zeigt die Wavelets eines Waveletnetzes mit drei Ebenen.

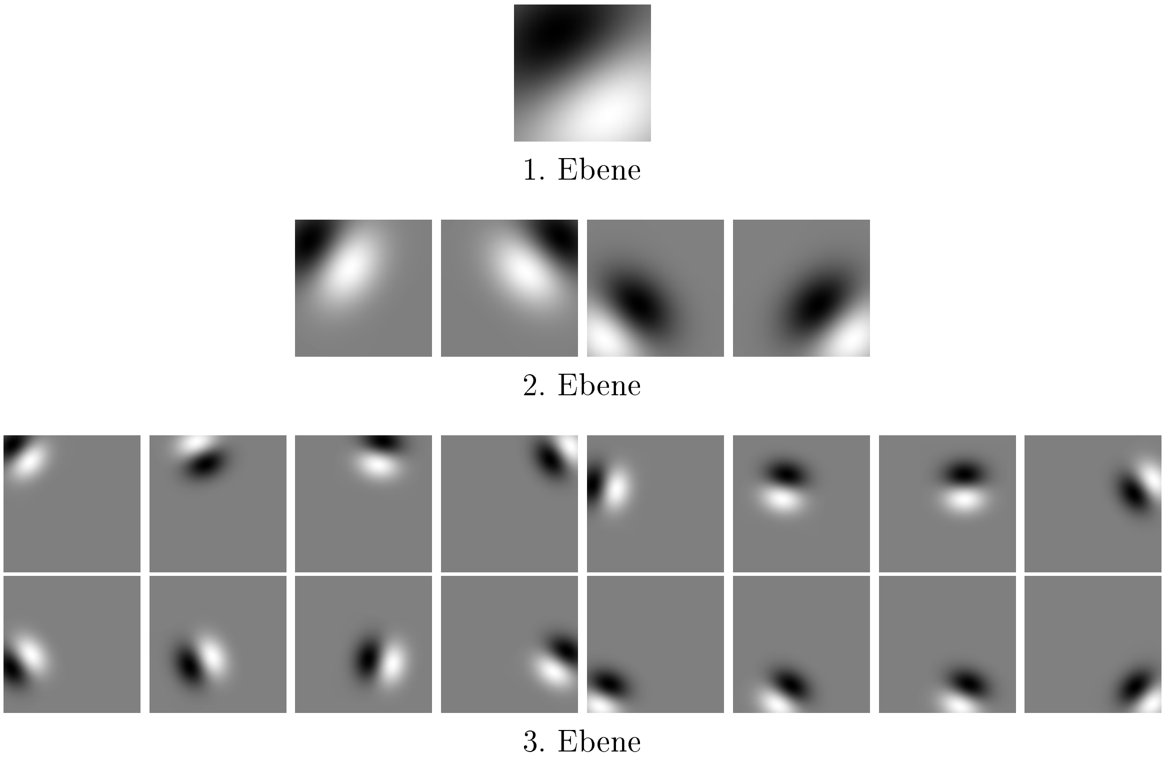
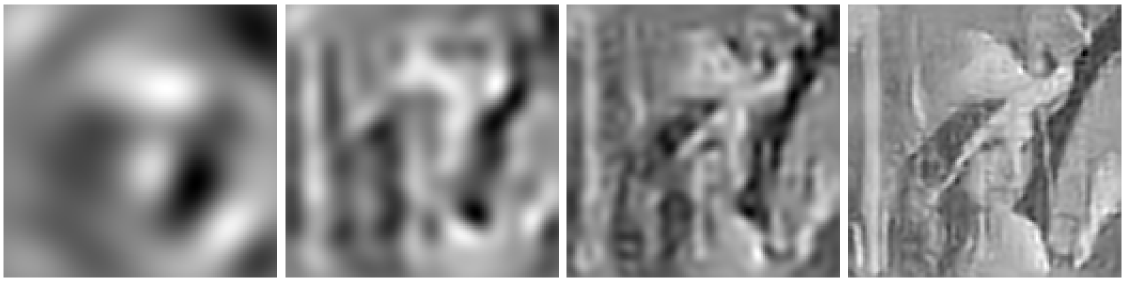


Abbildung 2.6: Die Wavelets der ersten drei Ebenen eines initialisierten Waveletnetzes.

Es verbleibt noch die Initialisierung der Waveletgewichte. In [21] wurde vorgeschlagen, die Projektionen der Wavelets auf das Bild als Initialisierung der Gewichte zu verwenden. Tatsächlich wäre dies nur dann die optimale Wahl, wenn orthogonale Wavelets verwendet würden. In Abschnitt 2.5.6 beschreiben wir das von uns implementierte Verfahren, die Nichtorthogonalität der Gaborwavelets bei der Bestimmung der optimalen Gewichte zu berücksichtigen. Abbildung 2.7 zeigt die Ergebnisse dieses Initialisierungsverfahrens am Beispiel des Bildes "Lenna", wobei jeweils die Initialisierung mit zufälligen Orientierungen der Initialisierung mit optimalen Orientierungen gegenüber gestellt werden. Man sieht deutlich, daß die Wahl der initialen Orientierung in Abhängigkeit vom Betrag der Projektionen eine bessere Initialisierung liefert. In unseren Experimenten führt diese bessere Initialisierung stets zu besseren Approximationen durch das Optimierungsverfahren.

(a)

(b)



(c)

Abbildung 2.7: Initialisierung eines Waveletnetzes mit drei, vier, fünf und sechs Ebenen am Beispiel des Bildes "Lenna" (a). Abbildung (b) zeigt die Ergebnisse mit zufälligen Orientierungen, (c) mit optimalen Orientierungen. Es wurde das in Abschnitt 2.5.6 eingeführte Verfahren zur optimalen Initialisierung der Gewichte verwendet. Man sieht bei den aus vier, fünf und sechs Ebenen bestehenden Netzen deutlich, daß die Wahl optimaler Orientierungen bessere Approximationen liefert.

2.5.4 Optimierungsverfahren

An die im vorigen Abschnitt beschriebene Initialisierung eines Waveletnetzes schließt sich die Optimierung der Waveletparameter $\boldsymbol{a}_i$ sowie der Gewichte w_i der Wavelets an. Die Anpassung der Parameter erfolgt mit dem Ziel, den quadratischen Fehler zwischen dem Grauwertbild g und der Ausgabe z eines Waveletnetzes zu minimieren. Faßt man die Parametervektoren $\boldsymbol{a}_i = (t_{x_i}, t_{y_i}, d_{x_i}, d_{y_i}, \alpha_i)^T$ und den Gewichtsvektor $\boldsymbol{w} = (w_1, \ldots, w_N)^T$ zu einem Vektor

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{6N})^T = (w_1, t_{x_1}, t_{y_1}, d_{x_1}, d_{y_1}, \alpha_1, \ldots, d_{y_N}, \alpha_N)^T$$

zusammen, dann ergibt sich der quadratische Fehler als

$$\chi^2(\boldsymbol{\theta}) = \frac{1}{2} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} \big(g(m,n) - z(m,n;\boldsymbol{\theta})\big)^2 = \frac{1}{2} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} \left(g(m,n) - \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}(m,n) \right)^2 . \tag{2.27}$$

Wie in [21] verwenden wir den Levenberg-Marquardt-Algorithmus [22], um den optimalen Parametervektor $\boldsymbol{\theta}$ zu bestimmen. Im nächsten Abschnitt soll nun der Levenberg-Marquardt-Algorithmus erläutert werden.

2.5.4.1 Levenberg-Marquardt-Verfahren

Ausgehend von einem Parametervektor $\boldsymbol{\theta}_{akt}$ läßt sich die Fehlerfunktion χ^2 durch eine Taylorreihe zweiter Ordnung approximieren:

$$\chi^2(\boldsymbol{\theta}) \approx \chi^2(\boldsymbol{\theta}_{\text{akt}}) + \nabla\chi^2\big|_{\boldsymbol{\theta}_{\text{akt}}} \cdot \boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\theta}^T \boldsymbol{H} \boldsymbol{\theta} \,, \tag{2.28}$$

mit der Hesse-Matrix $\boldsymbol{H} = \left(\left. \dfrac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j} \right|_{\boldsymbol{\theta}_{\text{akt}}} \right)_{ij}$. Ist $\boldsymbol{\theta}_{akt}$ in der Nähe eines Minimums von χ^2, so wird die Approximation (2.28) relativ gut sein. In diesem Fall wird durch

$$\boldsymbol{\theta}_{\text{neu}} = \boldsymbol{\theta}_{\text{akt}} + \boldsymbol{H}^{-1} \left(- \nabla\chi^2\big|_{\boldsymbol{\theta}_{\text{akt}}} \right) . \tag{2.29}$$

eine gute Näherung des Minimums von χ^2 erreicht. Die iterierte Anwendung von Gleichung (2.29) stellt das Newton-Verfahren dar.

Wenn (2.28) eine schlechte lokale Approximation der Fehlerfunktion χ^2 ist, führt ein Gradientenabstieg in Richtung des gesuchten Minimums:

$$\boldsymbol{\theta}_{\text{neu}} = \boldsymbol{\theta}_{\text{akt}} - c\, \nabla\chi^2\big|_{\boldsymbol{\theta}_{\text{akt}}} \,, \tag{2.30}$$

mit einer geeigneten Konstanten c.

Mit $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_{\text{neu}} - \boldsymbol{\theta}_{\text{akt}}$ werden (2.29) zu

$$\Delta\boldsymbol{\theta} = \boldsymbol{H}^{-1}(-\left.\nabla\chi^2\right|_{\boldsymbol{\theta}_{\text{akt}}}) \tag{2.31}$$

und (2.30) zu

$$\Delta\boldsymbol{\theta} = -c\left.\nabla\chi^2\right|_{\boldsymbol{\theta}_{\text{akt}}} . \tag{2.32}$$

Als Gradient von χ^2 bezüglich $\boldsymbol{\theta}$ erhält man

$$\frac{\partial\chi^2}{\theta_k} = -\sum_{m=0}^{K-1}\sum_{n=0}^{L-1}\left[(g(m,n) - z(m,n;\boldsymbol{\theta}))\,\frac{\partial z(m,n;\boldsymbol{\theta})}{\partial\theta_k}\right] . \tag{2.33}$$

Für die zweiten Ableitungen gilt

$$\frac{\partial^2\chi^2}{\partial\theta_k\partial\theta_l} = \sum_{m=0}^{K-1}\sum_{n=0}^{L-1}\left[\frac{\partial z(m,n;\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial z(m,n;\boldsymbol{\theta})}{\partial\theta_l} - (g(m,n) - z(m,n;\boldsymbol{\theta}))\,\frac{\partial^2 z(m,n;\boldsymbol{\theta})}{\partial\theta_l\partial\theta_k}\right] . \tag{2.34}$$

In der Praxis hat sich gezeigt, daß Optimierungsverfahren, die die Hesse-Matrix verwenden, stabiler sind, wenn in (2.34) der Term mit den zweiten Ableitungen von z vernachlässigt wird, so daß stattdessen die Matrix

$$\boldsymbol{H}' = \left(\sum_{m=0}^{K-1}\sum_{n=0}^{L-1}\left[\left.\frac{\partial z}{\partial\theta_k}\right|_{(m,n;\boldsymbol{\theta}_{\text{akt}})}\left.\frac{\partial z}{\partial\theta_l}\right|_{(m,n;\boldsymbol{\theta}_{\text{akt}})}\right]\right)_{kl} \tag{2.35}$$

anstelle der Hesse-Matrix verwendet wird.

Mit $\boldsymbol{b} = -\nabla\chi^2\ |_{\boldsymbol{\theta}_{\text{akt}}}$ kann Gleichung (2.31) geschrieben werden als

$$\boldsymbol{H}'\delta\boldsymbol{\theta} = \boldsymbol{b} . \tag{2.36}$$

Aus Gleichung (2.32) wird dann

$$\frac{1}{c}\delta\boldsymbol{\theta} = \boldsymbol{b} . \tag{2.37}$$

Levenberg und Marquardt haben eine Methode entwickelt, um kontinuierlich zwischen der Newton-Methode (2.36) und dem Gradientenabstieg (2.37) zu variieren. Das Ziel ist es, bei großer Distanz vom Minimum den Gradientenabstieg zu verwenden und in der Nähe des Minimums allmählich zum Newton-Verfahren überzugehen. Dieses Levenberg-Marquardt-Verfahren arbeitet in der Praxis sehr gut und ist zu einem Standardverfahren für nichtlineare Kleinste-Quadrate-Probleme geworden.

Die Variation zwischen dem Gradienten- und dem Newtonverfahren gelingt durch das modifizierte Gleichungssystem

$$\boldsymbol{H}'(1+\lambda\boldsymbol{I})\delta\boldsymbol{\theta} = \boldsymbol{b}, \tag{2.38}$$

mit der Einheitsmatrix $\boldsymbol{I}$ und einem skalaren Faktor λ.

Für kleine λ geht (2.38) in das Newton-Verfahren (2.36) über. Für große λ wird $\boldsymbol{H}'(1+\lambda\boldsymbol{I})$ diagonaldominant. Bezeichnen wir die Diagonalelemente von $\boldsymbol{H}'(1+\lambda\boldsymbol{I})$ mit α_i, dann geht (2.38) für große λ über in

$$\lambda\alpha_i\delta\theta_i = b_i \qquad i = 1\ldots 6N\ . \tag{2.39}$$

Das entspricht dem Gradientenabstieg (2.37), wobei für jede Komponente $\delta\theta_i$ eine eigene Konstante $c = \frac{1}{\lambda\alpha_i}$ verwendet wird.

Damit läßt sich das Levenberg-Marquardt-Verfahren als Algorithmus formulieren:

0. Setze $\boldsymbol{\theta}$ entsprechend der Initialisierung des Waveletnetzes, wähle ein λ (z.B. 10^{-4}) und einen Schwellwert k (z.B. 10^{-6})
1. Berechne $\chi^2(\boldsymbol{\theta})$
2. Löse das Gleichungssystem (2.38) für $\delta\boldsymbol{\theta}$ und berechne $\chi^2(\boldsymbol{\theta}+\delta\boldsymbol{\theta})$
3. Wenn $\frac{\chi^2(\boldsymbol{\theta}+\delta\boldsymbol{\theta})-\chi^2(\boldsymbol{\theta})}{\chi^2(\boldsymbol{\theta})} < k$, dann gehe zu 6
4. Wenn $\chi^2(\boldsymbol{\theta}+\delta\boldsymbol{\theta}) \leq \chi^2(\boldsymbol{\theta})$, dann erhöhe λ um den Faktor 10 und gehe zu 2
5. Wenn $\chi^2(\boldsymbol{\theta}+\delta\boldsymbol{\theta}) > \chi^2(\boldsymbol{\theta})$, dann erniedrige λ um den Faktor 10, aktualisiere $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}+\delta\boldsymbol{\theta}$ und gehe zu 2
6. Stoppe. Konvergenz des Verfahrens

2.5.4.2 Progressive Optimierung

Bei dem im vorigen Abschnitt beschriebenen Levenberg-Marquardt-Verfahren werden die Parameter sämtlicher Wavelets im Waveletnetz simultan adaptiert. Pro Ebene der Auflösungspyramide vervierfacht sich die Zahl der Wavelets gegenüber der vorherigen Ebene. Tabelle 2.1 zeigt die Anzahl der Wavelets in Abhängigkeit der Anzahl der Ebenen der Auflösungspyramide. Mit zunehmender Anzahl der Ebenen erhöhen sich entsprechend zum einem die Anzahl der Funktionsauswertungen zur Berechnung der

Anzahl der Ebenen	Anzahl der Wavelets im Netz
1	1
2	5
3	21
4	85
5	341
6	1365

Tabelle 2.1: Anzahl der Wavelets in Abhängigkeit der Ebenenzahl der Auflösungspyramide.

Wavelets und zum anderen die Dimension des Parametervektors. Beim Levenberg-Marquardt-Verfahren ist die Invertierung der Hesse-Matrix notwendig, deren Dimension durch die Dimension des Parametervektors gegeben ist. Die sich aus diesen Zusammenhängen ergebenden Rechenzeiten für die Anpassung der Wavelet-Parameter und Gewichte sind nicht akzeptabel.

Deshalb wird in [21] ein progressives Verfahren vorgeschlagen, bei dem die einzelnen Wavelets sukzessive adaptiert werden. Beginnend bei der gröbsten Ebene werden die Wavelets nacheinander zeilenweise einzeln optimiert, wobei die Parameter bereits optimierter Wavelets festgehalten werden. So dient die Optimierung jedes neu hinzugenommenen Wavelets der Reduzierung des verbleibenden Restfehlers. Die Vorgehensweise entspricht einer Grob-Fein-Strategie (*coarse to fine*); die grobskaligen Wavelets tragen die Bildinformation auf einer groben Skala, für die kleinskaligen Wavelets verbleibt die Approximierung der entsprechend feineren Bildstrukturen. Abbildung 2.8

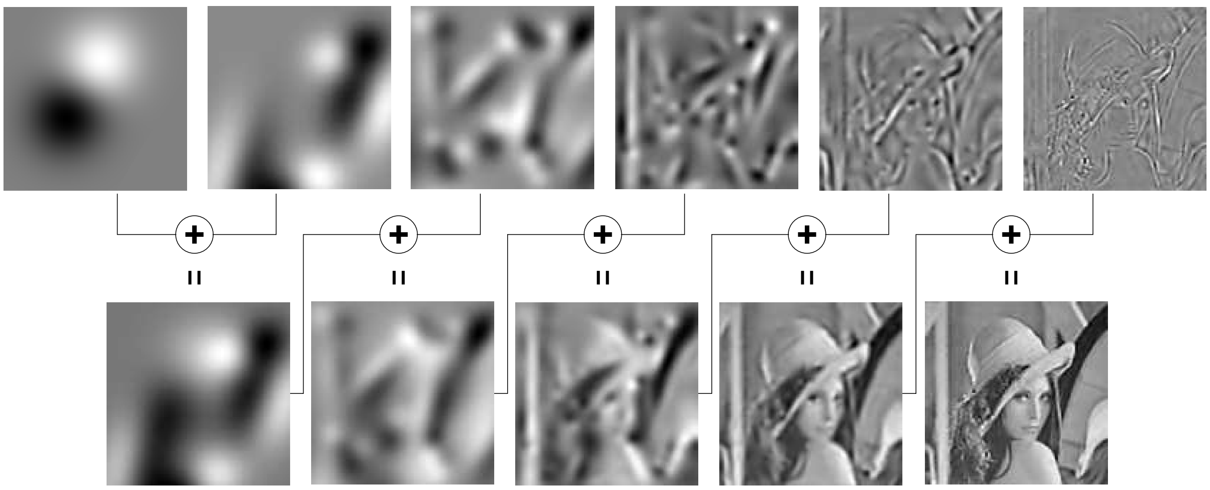


Abbildung 2.8: Grob-Fein-Strategie.

veranschaulicht dieses Prinzip. Die sechs Ebenen eines Waveletnetzes sowie die sukzessive Aufsummierung der Ebenen sind einzeln dargestellt. Das Ergebnis der letzten Summation ist die Ausgabe des gesamten Waveletnetzes.

2.5.4.3 Lokale Optimierung

Die in Abschnitt 2.5.3 vorgestellte Initialisierung eines Waveletnetzes (vgl. Abbildung 2.6) liefert bereits ein Waveletnetz, dessen Ausgabe starke Ähnlichkeiten zum zu approximierenden Bild aufweist und damit einen guten Ausgangspunkt für die Optimierung der Waveletparameter und Gewichte liefert.

Bei der im letzten Abschnitt vorgestellten progressiven Optimierung wird die Initialisierung nicht optimal genutzt. Die progressive Optimierung startet mit einem leeren Waveletnetz, die Wavelets des vorinitialisierten Netzes werden Ebene für Ebene einzeln hinzugenommen und angepaßt. Als Startparameter für die Optimierung werden für jedes Wavelet die Parameter der Initialisierung verwendet. Diese Startwerte sind optimal in Bezug auf die Gesamtheit aller Wavelets im initialisierten Netz. Damit sind sie aber nicht zwingend optimal im Zusammenspiel mit den bereits optimierten Wavelets, die gegenüber der Initialisierung andere Positionen, Skalen, Orientierungen und Gewichte erhalten haben.

Wir schlagen daher eine *lokale* Optimierung vor, bei der wie bei der progressiven Optimierung jeweils lediglich einzelne Wavelets angepaßt werden. Im Gegensatz zur progressiven Optimierung wird als Ausgangspunkt der Optimierung das gesamte initialisierte Netz genommen. Von Ebene zu Ebene werden sukzessive die Wavelets optimiert, wobei wie bei der progressiven Optimierung zeilenweise vorgegangen wird. Die übrigen Wavelets werden bei dieser Optimierung festgehalten und tragen so zur Reduzierung des Approximationsfehlers bei. Die Optimierung einer Ebene kann optional mehrfach durchlaufen werden.

In Abbildung 2.9 sind Zwischenergebnisse der lokalen Optimierung der letzten Ebene eines Netzes mit fünf Ebenen abgebildet. Die Abbildung verdeutlicht das Prinzip der lokalen Optimierung. Man sieht, wie die Wavelets einer Ebene nacheinander angepaßt werden.

Abbildung 2.10 stellt für das Bild "Lenna" die Ergebnisse der progressiven Optimierung denen der lokalen Optimierung mit drei Optimierungsdurchläufen pro Ebene gegenüber. In Tabelle 2.2 sind die entsprechenden Approximationsfehler ausgegeben. Durch die lokale Optimierung werden sowohl für den Betrachter als auch im Sinne der Minimierung des quadratischen Fehlers bessere Ergebnisse gegenüber der progressiven Optimierung erzielt. Diese Aussage ist auch für andere Bilder als "Lenna" gültig, so daß wir uns für die Verwendung der lokalen Optimierung entschieden haben.

(a) (b) (c)

(d) (e)

Abbildung 2.9: Zwischenergebnisse der lokalen Optimierung am Beispiel der letzten Ebene eines Waveletnetzes mit 5 Ebenen. (a) Netz nach Optimierung der Ebenen 1 bis 4, (b) nach Optimierung von 25 % der 5. Ebene, (c) nach Optimierung von 50 % der 5. Ebene, (d) nach Optimierung von 75 % der 5. Ebene; (e) Endresultat der Optimierung.

(a)

(b)

Abbildung 2.10: Resultate nach der Optimierung von Waveletnetzen mit drei, vier, fünf und sechs Ebenen. (a) Progressive Optimierung, (b) lokale Optimierung. Für diese Abbildung wurde die Optimierung jeder Ebene dreimal durchlaufen, wodurch die gegenüber Abbildung 2.9 besseren Approximationsergebnisse begründet sind.

	Anzahl der Ebenen			
χ^2	3	4	5	6
progressiv	$3.13 \cdot 10^7$	$1.78 \cdot 10^7$	$9.08 \cdot 10^6$	$3.78 \cdot 10^6$
lokal	$2.88 \cdot 10^7$	$1.63 \cdot 10^7$	$8.11 \cdot 10^6$	$3.22 \cdot 10^6$

Tabelle 2.2: Approximationsfehler bei progressiver und lokaler Optimierung am Beispiele des Bildes “Lenna”.

2.5.4.4 Regularisierung der Parameter

In [21] wird berichtet, daß bei der Optimierung eine Regularisierung der Parameter notwendig wird. Auch bei unseren Experimenten ist dies der Fall. Mit Hilfe der Regularisierung muß dafür gesorgt werden, daß im Verlauf des Optimierungsverfahrens die Zentren der Wavelets nicht aus dem Bild herauslaufen und die Skalen der Wavelets nicht entarten, d.h. sehr große oder kleine Werte annehmen. Das Herauswandern der

Wavelets bzw. das Entarten der Skalen ist eine Folge lokaler Minima der Fehlerfunktion, die durch den Algorithmus eingenommen werden. Bildlich gesprochen kann es im Sinne der Minimierung des quadratischen Fehlers günstiger sein, ein Wavelet aus dem Bild zu bewegen, als das Wavelet irgendwo im Bild zu positionieren. Ohne Regularisierung wird jedoch häufig die Hesse-Matrix im Levenberg-Marquardt-Algorithmus singulär, wodurch der Algorithmus abbricht.

Die Regularisierung der Parameter wird durch eine Modifikation der Fehlerfunktion χ^2 erreicht, indem ein zusätzlicher Bestrafungsterm φ zur Fehlerfunktion addiert wird:

$$\hat{\chi}^2(\boldsymbol{\theta}) = \chi^2(\boldsymbol{\theta}) + \varphi(\boldsymbol{\theta}) \tag{2.40}$$

mit $\varphi(\boldsymbol{\theta}) = \varphi(t_x) + \varphi(t_y) + \varphi(d_x) + \varphi(d_y)$. Die Bestrafungsterme haben die Form eines Potentials, das einsetzt, sobald ein Parameter $\theta \in \{t_x, t_y, d_x, d_y\}$ ein vorgegebenes Intervall $[\theta_u, \theta_o]$ verläßt. Das Bestrafungs- oder Regularisierungspotential hat die Form

$$\varphi(\theta) = \begin{cases} \varphi_0 \varphi(\theta_u - \theta)^n & \text{falls } \theta < \theta_u \\ \varphi_0 \varphi(\theta - \theta_o)^n & \text{falls } \theta > \theta_o \\ 0 & \text{sonst} \end{cases} \tag{2.41}$$

mit geeigneter Amplitude φ_0 und Exponenten n. Die Ober- und Untergrenzen θ_o und θ_u sind für die Translationsparameter durch die Bildgröße gegeben. Für die Dilatationsparameter werden diese Grenzen durch eine wählbare maximale prozentuale Abweichung vom Wert der Initialisierung vorgegeben.

2.5.5 Waveletnetze vs. vollständige Waveletrepräsentation

Lee hat in [15] den in Abschnitt 2.4 eingeführten Begriff der *frames* auf zwei Dimensionen erweitert und für komplexwertige Gaborwavelets die Anforderungen an ein Diskretisierungsschema untersucht, mit dem eine vollständige Repräsentation durch Gaborwavelets möglich ist, also ein *frame* gebildet wird. Um einen *frame* zu erhalten, müssen pro Abtastpunkt mehrere Orientierungen gewählt werden (z.B. 8 oder 16 Orientierungen) und sowohl die Skalen als auch die Orte dichter abgetastet werden, als dies bei der Initialisierung der Waveletnetze erfolgt. Das bedeutet, daß eine sehr hohe Anzahl von Wavelets für eine vollständige Repräsentation notwendig ist.

Wir wollen hier ganz explizit darauf hinweisen, daß durch unsere Initialisierung und anschließende Optimierung der Waveletnetze **nicht** das Ziel verfolgt wird, eine vollständige Repräsentation durch Gaborwavelets zu erhalten. Eher das Gegenteil ist der Fall. Durch das Optimierungsverfahren soll ein Satz von Wavelets erzeugt werden, der gut geeignet ist zur **effizienten** Approximation *einer* fest gegebenen Funktion. Die Möglichkeit, mit diesem speziellen Satz von Wavelets andere, von der gegebenen deutlich verschiedene Grauwertfunktionen approximieren zu können, ist nicht gefordert. Sie ist, wie wir in Kapitel 3 sehen werden, für die Verwendung der Wavelets zur Erkennung von Gesichtern sogar unerwünscht.

2.5.6 Bestimmung optimaler Gewichte

In diesem Abschnitt wollen wir uns mit der Frage beschäftigen, wie für ein Waveletnetz $\mathcal{W}$ mit gegebenen Wavelets $\Psi = \{\psi_{\boldsymbol{a}_1}, \ldots, \psi_{\boldsymbol{a}_N}\}$ ein Gewichtsvektor $\boldsymbol{w} = (w_1, \ldots, w_N)$ bestimmt werden kann, um ein Grauwertbild g bestmöglich durch eine Linearkombination

$$z(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i} \tag{2.42}$$

der Funktionen aus Ψ zu approximieren. Dieses Problem stellt sich z.B. bei der Initialisierung eines Waveletnetzes. Wir nehmen o.B.d.A. an, daß g mittelwertfrei ist, also daß $\bar{g} = 0$. Für Bilder mit $\bar{g} \neq 0$ muß der Mittelwert $\bar{g}$ vor der Bestimmung der Gewichte abgezogen und anschließend zur Linearkombination (2.42) addiert werden, entsprechend der Ausgabe z eines Waveletnetzes $\mathcal{W} = (\Psi, \boldsymbol{w}, m)$ (vgl. Gleichung (2.24)).

Eine Möglichkeit zur Bestimmung der Gewichte ist die Verwendung des Optimierungsverfahrens aus 2.5.4, wobei die Parameter der Wavelets festgehalten werden und nur die Gewichte adaptiert werden. In der Regel wird das Optimierungsverfahren lediglich ein lokales Minimum finden.

Optimale Gewichte, die dem globalen Minimum des Approximationsfehlers entsprechen, sind jedoch auch direkt berechenbar, was im folgenden dargestellt wird.

In [21] wird vorgeschlagen, die Projektionen $<\psi_{\boldsymbol{a}_i}, g>$ der Wavelets $\psi_{\boldsymbol{a}_i}$ auf das zu approximierende Bild g als Gewichte zu verwenden. Wie wir in Abschnitt 2.4 ausgeführt haben, liefert das nur im Falle orthogonaler Wavelets das gewünschte Ergebnis. Bei nichtorthogonalen Wavelets, wie es die Gaborwavelets sind, müssen zur Rekonstruktion des Bildes g aus den Projektionen $<\psi_{\boldsymbol{a}_i}, g>$ duale Basisfunktionen $\tilde{\psi}_{\boldsymbol{a}_i}$ verwendet werden. Darauf kommen wir später zurück. Wir wollen uns zunächst von einer anderen Seite dem Problem nähern.

Dazu fassen wir das Bild g als kontinuierliche Signalfunktion in einem Hilbertraum $\mathcal{H}$ auf. Durch Ψ wird ein Unterraum $\mathcal{M} = \mathrm{Span}\{\psi_{\boldsymbol{a}_1}, \ldots, \psi_{\boldsymbol{a}_N}\}$ aufgespannt. Ein Element $\hat{g} \in \mathcal{M}$ heißt genau dann **beste Approximation** von $g \in \mathcal{H}$, wenn

$$||g - \hat{g}|| \leq ||g - m||$$

für alle $m \in \mathcal{M}$ gilt, wobei $||\cdot||$ die durch das Skalarprodukt $<\cdot,\cdot>$ induzierte Norm ist. Es gilt, daß $\hat{g}$ genau dann die beste Approximation von g ist, wenn $\hat{g}$ die orthogonale

Projektion g_p von g auf $\mathcal{M}$ ist [19]. Dabei ist genau dann ein Element $g_p \in \mathcal{M}$ die orthogonale Projektion von g auf $\mathcal{M}$, wenn

$$<g - g_p, m> = 0$$

für alle $m \in \mathcal{M}$ [19].

Ein leicht zu beweisender Satz [19] besagt, daß die orthogonale Projektion von g auf $\mathcal{M}$ gegeben ist durch

$$g_p = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i} \ , \tag{2.43}$$

wobei $\boldsymbol{w} = (w_1, \ldots, w_N)$ die Lösung des linearen Gleichungssystems

$$\boldsymbol{\Psi w} = \boldsymbol{c} \tag{2.44}$$

ist, mit $\boldsymbol{c} = (c_1, \ldots, c_N)^T = (<\psi_{\boldsymbol{a}_1}, g>, \ldots, <\psi_{\boldsymbol{a}_N}, g>)^T$ und $\boldsymbol{\Psi}_{jk} = <\psi_{\boldsymbol{a}_j}, \psi_{\boldsymbol{a}_k}>$. Die Matrix $\boldsymbol{\Psi}$ mit den Skalarprodukten der Funktionen aus $\mathcal{M}$ wird Gramsche Matrix genannt.

Damit haben wir die Lösung für die optimalen Gewichte gefunden:

$$w_k = \sum_{i=1}^{N} \boldsymbol{\Psi}_{ki}^{-1} <\psi_{\boldsymbol{a}_i}, g> \ . \tag{2.45}$$

Unter Berücksichtigung der Symmetrie von $\boldsymbol{\Psi}^{-1}$ gilt

$$\begin{aligned} \hat{g} &= \sum_{k=1}^{N} w_k \psi_{\boldsymbol{a}_k} \\ &= \sum_{k=1}^{N} \sum_{i=1}^{N} \boldsymbol{\Psi}_{ik}^{-1} c_i \psi_{\boldsymbol{a}_k} \\ &= \sum_{i=1}^{N} c_i \sum_{k=1}^{N} \boldsymbol{\Psi}_{ik}^{-1} \psi_{\boldsymbol{a}_k} \ . \end{aligned} \tag{2.46}$$

Aus (2.46) lassen sich die oben bereits angesprochenen dualen Basisfunktionen

$$\tilde{\psi}_{\boldsymbol{a}_i} = \sum_{k=1}^{N} \boldsymbol{\Psi}_{ik}^{-1} \psi_{\boldsymbol{a}_k} \tag{2.47}$$

ablesen, mit denen die Rekonstruktion

$$\hat{g} = \sum_{i=1}^{N} <\psi_{\boldsymbol{a}_i}, g> \tilde{\psi}_{\boldsymbol{a}_i} \tag{2.48}$$

von g mittels der Projektionen $<\psi_{\boldsymbol{a}_i}, g>$ möglich ist.

Das Skalarprodukt der Basisfunktionen $\psi_{\boldsymbol{a}_i}$ und dem Grauwertbild g mit $K \times L$ Pixeln berechnen wir in der Form

$$<\psi_{\boldsymbol{a}_i}, g> = \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} g(m,n) \psi_{\boldsymbol{a}_i}(m,n) \ . \tag{2.49}$$

Analog berechnen wir die Elemente der Gramschen Matrix:

$$<\psi_{\boldsymbol{a}_i}, \psi_{\boldsymbol{a}_k}> = \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} \psi_{\boldsymbol{a}_i}(m,n) \psi_{\boldsymbol{a}_k}(m,n) \ . \tag{2.50}$$

Wir wollen nun noch eine alternative, der Berechnung über die Inverse der Gramschen Matrix äquivalente Berechnungsmöglichkeit angeben, aus der deutlich wird, warum wir uns bei der Berechnung der Skalarprodukte der Wavelets in Gleichung (2.50) auf den Definitionsbereich des Bildes g beschränken, obwohl die Träger von Wavelets am Bildrand über diesen herausragen können.

Die Bestimmung der optimalen Gewichte kann als das Problem interpretiert werden, eine Lösung eines überbestimmten Gleichungssystems zu finden. Seien $I_1, \dots, I_N$ $K \times L$-Grauwertbilder erzeugt durch Abtasten der Basisfunktionen, also

$$I_j(m,n) = \psi_{\boldsymbol{a}_j}(m,n) \qquad j \in \{1, \dots, N\}, (m,n) \in \{0, \dots, K-1\} \times \{0, \dots, L-1\} \ .$$

Seien $\boldsymbol{y}$ der Vektor, der die konkatenierten Zeilen des Bildes g enthält und $\boldsymbol{A}$ eine $K \cdot L \times N$ Matrix, deren Spalten durch Konkatenation der Zeilen der Waveletbilder I_j gebildet werden. Dann ergeben sich die optimalen Gewichte als Lösung des überbestimmten Gleichungssystems

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{w} \ . \tag{2.51}$$

Dieses Gleichungssystem läßt sich mit Hilfe der Pseudoinversen

$$\boldsymbol{A}^+ = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T$$

lösen. Dazu wird Gleichung (2.51) von links mit $\boldsymbol{A}^T$ multipliziert, man erhält

$$\boldsymbol{A}^T \boldsymbol{y} = \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{w} \tag{2.52}$$

und daraus die Lösung

$$\boldsymbol{w} = \boldsymbol{A}^+ \boldsymbol{y} \ . \tag{2.53}$$

Betrachten wir Gleichung (2.52) genauer: Die k.te Komponente des Vektors $\boldsymbol{A}^T\boldsymbol{y}$ lautet

$$\begin{aligned}(\boldsymbol{A}^T\boldsymbol{y})_k &= \sum_{i=1}^{K\cdot L} \boldsymbol{A}_{ik} y_i \\ &= \sum_{m=0}^{K-1}\sum_{n=0}^{L-1} \psi_{\boldsymbol{a}_g}(m,n) \ . \end{aligned} \tag{2.54}$$

Gleichung (2.54) ist also identisch mit (2.49), der Projektion von $\psi_{\boldsymbol{a}_k}$ auf g. Analog gilt für die Elemente von $\boldsymbol{A}^T\boldsymbol{A}$

$$\begin{aligned}(\boldsymbol{A}^T\boldsymbol{A})_{ik} &= \sum_{j=1}^{K\cdot L} \boldsymbol{A}_{ji}\boldsymbol{A}_{jk} \\ &= \sum_{m=0}^{K-1}\sum_{n=0}^{L-1} \psi_{\boldsymbol{a}_i}(m,n)\psi_{\boldsymbol{a}_k}(m,n) \ . \end{aligned} \tag{2.55}$$

Aus Gleichung (2.55) wird deutlich, daß $\boldsymbol{A}^T\boldsymbol{A}$ die Gramsche Matrix $\boldsymbol{\Psi}$ mit den Skalarprodukten der Wavelets ist, wenn die Skalarprodukte gemäß Gleichung (2.50) berechnet werden.

Damit ist das Gleichungssystem (2.52) identisch mit (2.44). Daraus folgt, daß die Bestimmung der Gewichte über die Pseudoinverse $\boldsymbol{A}^+$ äquivalent zur Bestimmung der Gewichte über die orthogonale Projektion von g in den durch die Wavelets aufgespannten Raum $\mathcal{M}$ ist, wenn die Skalarprodukte gemäß (2.49) und (2.50) berechnet werden. Die Rekonstruktion von g durch die orthogonale Projektion auf $\mathcal{M}$ ist wiederum analog zur Bestimmung der Gewichte aus den dualen Wavelets $\tilde{\psi}_{\boldsymbol{a}_i}$ gemäß (2.47) und (2.48).

In der Implementation der Waveletnetze wurde für die Berechnung der Waveletnetze der Weg über die Berechnung der inversen Gramschen Matrix gemäß Gleichung (2.45) gewählt. Bei der alternativen Berechnung über die Pseudoinverse müßten die Abbilder der Wavelets zur Berechnung der Gewichte im Speicher gehalten werden. Bei einer hohen Anzahl von Wavelets ist dies ineffizient bzw. nicht möglich. Die Verwendung der Gramschen Matrix hat einen weiteren Vorteil. Wie wir im Abschnitt 3.2 sehen werden, bleibt die Inverse der Gramschen Matrix bis auf eine Konstante invariant gegenüber affinen Transformationen der Basisfunktionen.

In diesem Abschnitt haben wir ein Verfahren entwickelt, mit dem zu der Familie Ψ aus Wavelets optimale Gewichte berechnet werden können, um ein Grauwertbild g bestmöglich durch diese Familie zu approximieren. Das Verfahren wird bei der Initialisierung von Waveletnetzen eingesetzt und wird sich in Kapitel 3 als essentiell für die Bestimmung der Ähnlichkeitsmaße für Gesichter erweisen.

2.5.7 Superwavelets

Die Ausgabe eines Waveletnetzes ist eine Linearkombination von Wavelets. Diese Linearkombination ihrerseits kann wiederum als Mutterwavelet aufgefaßt werden, aus dem durch Rotation, Translation und Dilatation Wavelets abgeleitet werden können. In [27] wird ein solches Mutterwavelet der Form

$$\Psi(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}(\boldsymbol{x}) \tag{2.56}$$

als *Superwavelet* bezeichnet, einer Kurzform für *superposition-wavelet*. Wir nennen das Mutterwavelet der Form (2.56) Mutter-Superwavelet und wollen die Bezeichnung Superwavelet für aus diesem durch Rotation, Translation und Dilatation gewonnene Wavelets verwenden.

Mit dem Parametervektor $\boldsymbol{n} = (t_x, t_y, d_x, d_y, d_{xy}, \alpha)$ wird aus dem Mutter-Superwavelet (2.56) das Superwavelet

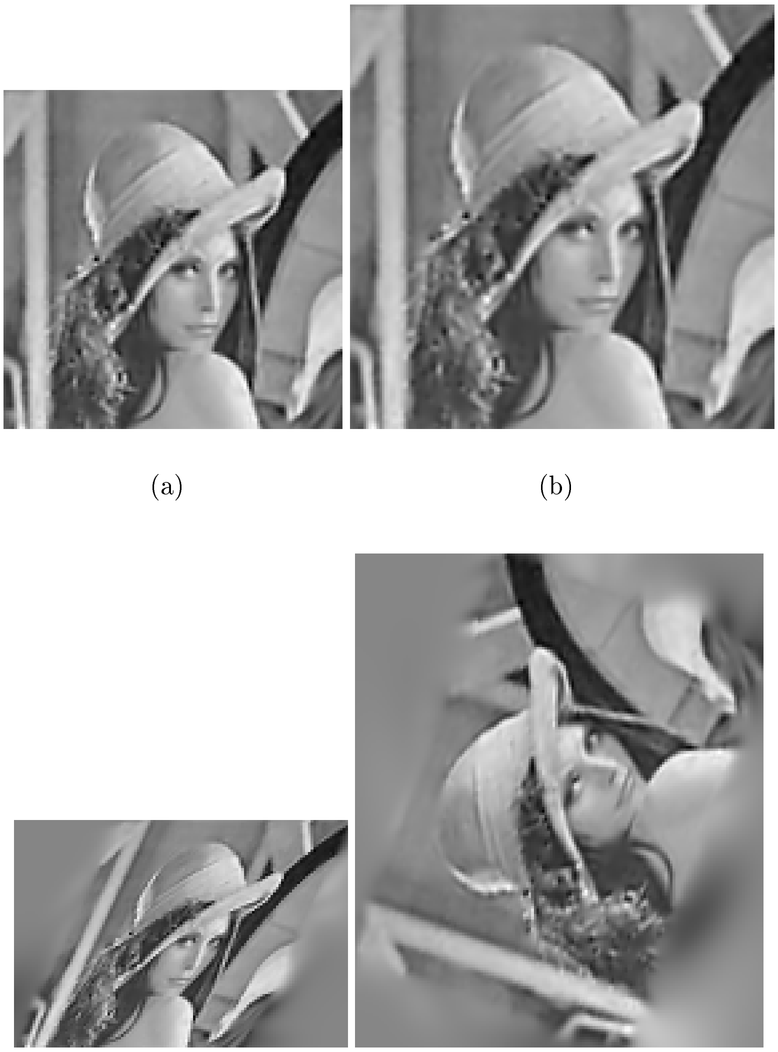
$$\Psi_{\boldsymbol{n}}(\boldsymbol{x}) = \Psi\left(\boldsymbol{D}\boldsymbol{R}[\boldsymbol{x} - \boldsymbol{t}]\right) \tag{2.57}$$

erzeugt, mit $\boldsymbol{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$, $\boldsymbol{D} = \begin{pmatrix} d_x & d_{xy} \\ 0 & d_y \end{pmatrix}$ und $\boldsymbol{R} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$. Mit den Notationen aus Abschnitt 2.5.1 wird (2.57) zu

$$\Psi_{\boldsymbol{n}}(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \psi\left(\boldsymbol{D}_i \boldsymbol{R}_i (\boldsymbol{D}\boldsymbol{R}[\boldsymbol{x} - \boldsymbol{t}] - \boldsymbol{t}_i)\right) \ . \tag{2.58}$$

Zusätzlich zur Parametrisierung der bisherigen Wavelets wird durch den zusätzlichen Dilatationsparameter d_{xy} eine Scherung von Superwavelets ermöglicht, womit das volle Spektrum affiner Transformationen für Superwavelets zur Verfügung steht. Damit stellt ein Superwavelet ein Template dar, das kontinuierlich skalierbar, scherbar, orientierbar sowie durch den Translationsvektor $\boldsymbol{t}$ in Subpixelgenauigkeit positionierbar ist. Abbildung 2.11 veranschaulicht, wie sich die Änderungen der Parameter eines Superwavelets auswirken.

Mit dem Konzept des Superwavelets als Template lassen sich die bei herkömmlichen Templates auftretenden Probleme bei der Rotation und Skalierung auf einem diskreten Gitter vermeiden.



(a) (b)

(c) (d)

Abbildung 2.11: Waveletnetz als Superwavelet. (a) Mutter-Superwavelet; (b) Superwavelet mit $\alpha = 0°$, $d_x = d_y = 0.8$, $d_{xy} = 0$; (c) $\alpha = 0°$, $d_x = d_y = 1.5$, $d_{xy} = 0.7$; (d) $d_x = 0.8$, $d_y = 1.2$, $d_{xy} = 0$, $\alpha = 70°$.

3 Waveletnetze zur Gesichtserkennung

In diesem Kapitel wollen wir unsere Ansätze vorstellen, Waveletnetze zur Erkennung von Gesichtern einzusetzen. Die Voraussetzungen für die Erkennung sind für diese Diplomarbeit dadurch gegeben, daß eine feste Menge zu identifizierender Personen vorgegeben ist. Für jede Person steht als **Referenzbild** eine Frontalaufnahme zur Verfügung, die die Person mit einem neutralen Gesichtsausdruck zeigt. Die Referenzbilder werden in einer **Referenzdatenbank** gespeichert. An jedes Referenzbild der Datenbank wird ein Waveletnetz angepaßt. Das angepaßte **individuelle Waveletnetz** dient als Repräsentation des Referenzbildes und wird gewissermaßen als Modell des speziellen Gesichts verwendet. Beispiele von Waveletnetzen, die Gesichter repräsentieren, zeigt Abbildung 3.1.

Abbildung 3.1: Repräsentation von Referenzgesichtern (obere Reihe) durch individuelle Waveletnetze bestehend aus 64 Gaborwavelets (untere Reihe).

In einer weiteren Datenbank stehen für jede Person der Referenzdatenbank mehrere Frontalaufnahmen zur Verfügung, die die Person mit verschiedenen Gesichtsausdrücken und ggf. leicht unterschiedlichen Kopfhaltungen zeigen. Aus dieser **Testda-**

tenbank werden dem Gesichtserkennungssystem Bilder als Eingabe gegeben. Diese Eingaben nennen wir im folgenden **Testbilder**. Das System hat die Aufgabe, auf der Basis der Gesichtsrepräsentation zu entscheiden, um welche Person aus der Referenzdatenbank es sich bei einem aus der Testdatenbank entnommenen Testbild handelt.

Abschnitt 3.1 beschreibt, wie die Gesichtsrepräsentationen in Form von individuellen Waveletnetzen gewonnen werden. Dazu erweitern wir in Abschnitt 3.1.1 die Waveletnetze derart, daß zusätzlich zum explizit modellierten Gleichanteil ein linearer Term eingeführt wird, um unterschiedliche Beleuchtungseinflüsse zu kompensieren. Abschnitt 3.1.2 beschreibt die gegenüber den bisher dargestellten Waveletnetzen geänderte Initialisierung, so werden Waveletnetze zur Repräsentation von Gesichtern z.B. auf einen elliptischen Bereich eingeschränkt. In Abschnitt 3.1.3 erläutern wir, wie die Wavelets eines initialisierten Waveletnetzes an ein Referenzgesicht angepaßt werden und so ein individuelles Waveletnetz erzeugt wird. Hierfür haben sich zusätzliche Einschränkungen gegenüber den allgemeinen Waveletnetzen zur Repräsentation von Grauwertbildern als nützlich erwiesen, insbesondere die Regularisierung der Dilatationsparameter haben wir modifiziert.

Die in Abschnitt 3.1 gewonnenen individuellen Waveletnetze werden in Abschnitt 3.2 als Superwavelets verwendet. Damit stehen kontinuierlich affin transformierbare und somit in Subpixelgenauigkeit positionierbare Templates zur Verfügung, mit denen durch ein Optimierungsverfahren (**Matching**) Skalen, Orientierungen und Positionen zu identifizierender Testgesichter bestimmt werden können.

Die Einsatzmöglichkeit des Matchingverfahrens zum Verfolgen von Gesichtern in Videosequenzen (*tracking*) wird in Abschnitt 3.3 aufgezeigt.

Durch das Matching wird ein Waveletnetz affin transformiert. Dadurch werden die Basisfunktionen (Wavelets) des Netzes entsprechend der Skala, Position und Orientierung des Testgesichts geändert (*warping*). Hierfür verwenden wir den Begriff der **Reparametrisierung**. In Abschnitt 3.4 beschreiben wir, wie mit diesen geänderten Basisfunktionen das Testgesicht rekonstruiert wird. Die Güte der Rekonstruktion wird später ein Kriterium zur Identifizierung des Testgesichts sein.

In Abschnitt 3.5 führen wir Maße zur Bestimmung der Ähnlichkeit von Referenzgesichtern und Testgesichtern ein. Dazu wird die bezüglich eines Testgesichts reparametrisierte Waveletnetz-Repräsentation des Referenzgesichts mit der aus der reparametrisierten Repräsentation gewonnen Rekonstruktion des Testgesichts verglichen. Die Ähnlichkeit kann entweder, wie in Abschnitt 3.5.1 beschrieben, im Signalraum oder im Raum der Waveletgewichte (Abschnitt 3.5.2) bestimmt werden. Über die reine Beschreibung der Maße hinaus arbeiten wir den Zusammenhang der Ähnlichkeitsmaße im Gewichtsraum und im Signalraum heraus.

Den Ähnlichkeitsmaßen liegt die Annahme zu Grunde, daß das bezüglich des Testbildes reparametrisierte individuelle Waveletnetz nach dem Matching Aufnahmen der

Referenzperson mit geringem Fehler rekonstruieren kann und hingegen bei der Rekonstruktion von Bildern anderer Personen größere Rekonstruktionsfehler entstehen.

Abschnitt 3.6 stellt die auf den Ähnlichkeitsmaßen basierenden Verfahren zur Identifikation einer Person bzw. zur Verifikation der Identität vor.

3.1 Gesichtsrepräsentation

In diesem Abschnitt beschreiben wir, wie wir, ausgehend von einem Referenzbild, ein individuelles Waveletnetz erzeugen, das als Repräsentation eines Gesichts dient. Bei einem Referenzbild handelt es sich um eine Frontalansicht eines Gesichts mit neutralem Gesichtsausdruck.

3.1.1 Zusätzliche lineare Komponente

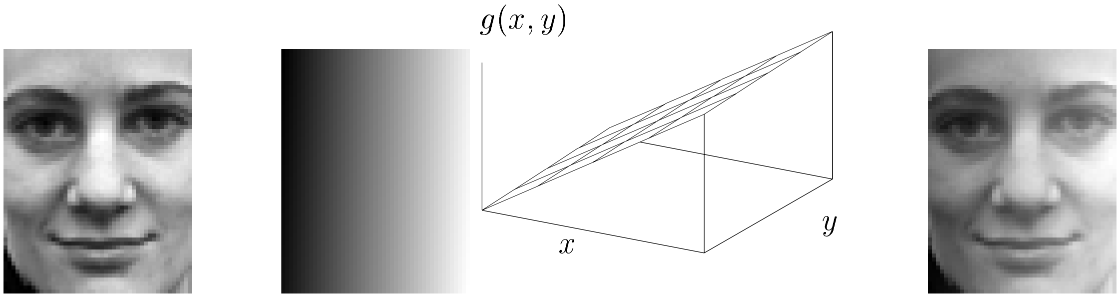
Die Linearkombination aus den Wavelets der in Kapitel 2 eingeführten Waveletnetze beschreibt die approximierte Bildstruktur bezüglich des Mittelwerts (Gleichanteils) des Bildes (vgl. Gleichung (2.16)). Abbildung 3.2(a) zeigt ein Gesicht, zu dem die lineare Grauwertfunktion (**Grauwertebene**) aus Abbildung 3.2(b) addiert wurde, um eine seitliche Beleuchtung zu simulieren (Abbildung 3.2(c)). Abbildung 3.3 zeigt die Ausgaben zweier Waveletnetze, die an Abbildung 3.2(c) angepaßt wurden. Dabei wurde das Netz in 3.3(b) zur besseren Modellierung linearer Grauwertverläufe zusätzlich zum Gleichanteil um einen lineareren Term erweitert. Der quadratische Fehler nach der Optimierung beträgt für Netz 3.3(a) $1.17 \cdot 10^6$ und für Netz (b) $4.87 \cdot 10^5$. Diese Verbesserung der Approximation durch die explizite Modellierung des linearen Verlaufs ist auch optisch in Abbildung 3.3 wahrnehmbar. In Abbildung 3.3(b) mußten die Wavelets nicht zusätzlich die überlagerte Grauwertebene approximieren, so daß sie besser die Strukturen abbilden konnten, was besonders bei der Nase und in der Augenregion deutlich wird. In Abbildung 3.3(b) ist das von links nach rechts ansteigende Grauwertniveau wesentlich besser wiedergegeben, als dies in Abbildung 3.3(a) ohne Modellierung eines linearen Anteils der Fall ist.

Ein um einen linearen Term ergänztes Waveletnetz liefert eine Ausgabe der Form

$$z(x, y; \boldsymbol{a}, \boldsymbol{w}) = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}(x, y) + l(x, y) + m \;, \tag{3.1}$$

mit

$$l(x, y) = \xi x + \eta y + \tau \tag{3.2}$$



(a) (b) (c)

Abbildung 3.2: Simulation einer einseitigen Beleuchtung. (a) Gesicht bei zentraler Beleuchtung, (b) linearer Grauwertverlauf als Grauwertbild und Flächenplot, (c) Addition der Grauwertebene zum Gesicht zur Simulation einer Beleuchtungsquelle rechts vom Gesicht.

(a) (b)

Abbildung 3.3: Approximation von Abbildung 3.2(c) durch Waveletnetze. (a) ohne zusätzlichen linearen Term, (b) mit zusätzlichem linearen Term.

zur Modellierung einer Grauwertebene. Die Parameter ξ, η und τ der Grauwertebene werden durch Lösen des linearen Ausgleichsproblems

$$(\xi, \eta, \tau) = \min_{\xi,\eta,\tau} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} \left(g(m,n) - l(m,n;\xi,\eta,\tau)\right)^2 \tag{3.3}$$

bestimmt.

In unserer Implementation der Waveletnetze zur Repräsentation von Gesichtern gehen wir einen etwas anderen Weg. Anstatt die bestmögliche lineare Approximation eines Bildes g zur Ausgabe eines Waveletnetzes zu addieren (Gleichung (3.1)), wird eine an g optimal angepaßte Grauwertebene der Form (3.2) von g abgezogen. Die Subtraktion der Grauwertebene dient damit dem Zweck, Einflüsse z.B. durch unterschiedliche Beleuchtungsrichtungen zu verringern, analog zum von Sung und Poggio vorgeschlagenen Verfahren [25]. Das auf diese Weise vorverarbeitete Bild wird durch ein Waveletnetz approximiert, um eine von Beleuchtungsbedingungen weitgehend unabhängige Gesichtsrepräsentation zu erhalten.

3.1.2 Initialisierung

Abbildung 3.4 zeigt einige typische Beispiele der Referenzbilder, aus denen wir eine Gesichtsrepräsentation erzeugen. Die in Kapitel 2 vorgestellten Waveletnetze liefern

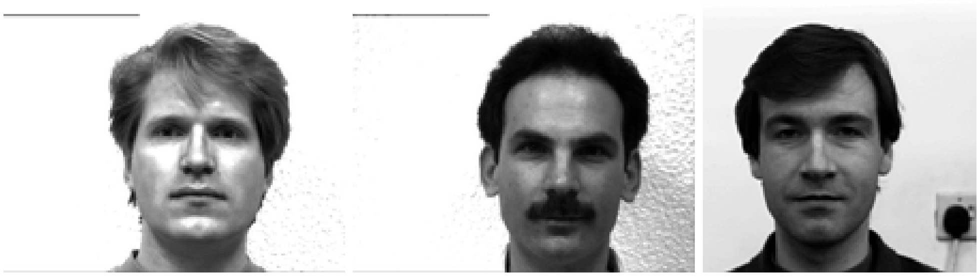


Abbildung 3.4: Beispiele von Referenzbildern, aus denen Gesichtsrepräsentationen erzeugt werden.

eine Repräsentation jeweils eines gesamten Bildes inklusive Hintergrund. Wir wollen hingegen lediglich eine Repräsentation der Gesichter entwickeln. Dazu ist es naheliegend, lediglich Wavelets in einem elliptischen Bereich zu positionieren, der das **innere Gesicht** (*inner face*) darstellt. In Abbildung 3.5 sind diese elliptischen Ausschnitte hervorgehoben.

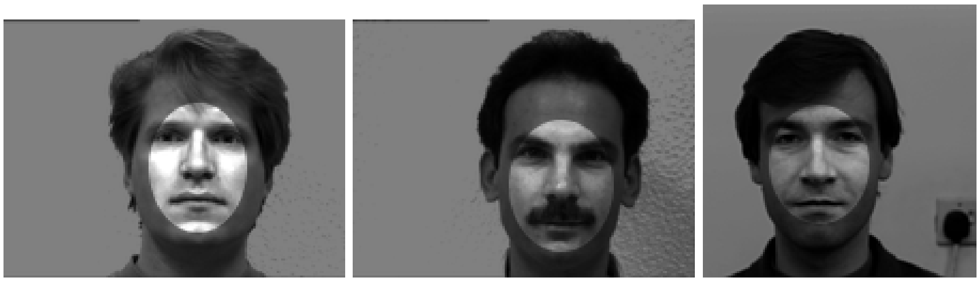
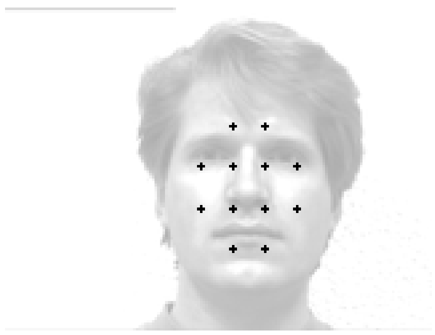
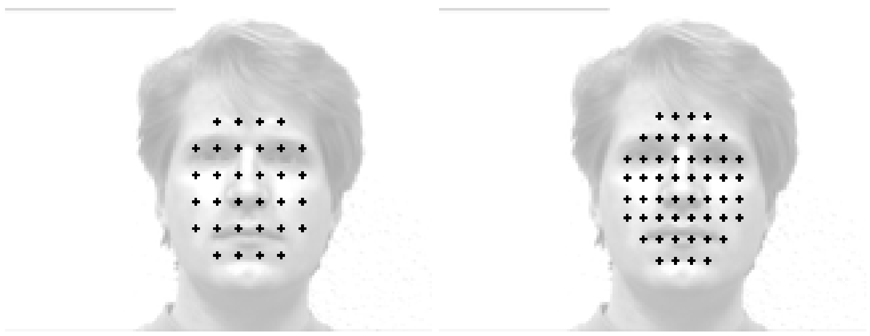


Abbildung 3.5: Elliptische Gesichtsregionen, in denen Wavelets positioniert werden.

Wir möchten eine effiziente Repräsentation erhalten, die mit möglichst wenigen Wavelets auskommt. Deshalb verwenden wir für die Gesichtsrepräsentation Waveletnetze, die nicht gemäß einer kompletten Skalenpyramide initialisiert werden, wie es in [21] und Abschnitt 2.5.3 beschrieben wird. Für unsere Experimente haben wir uns für Waveletnetze bestehend aus zwei Ebenen entschieden. Abbildung 3.6 zeigt die Waveletzentren der verwendeten Netzvarianten mit 42 bzw. 64 Wavelets. Die erste Ebene



(a)



(b)

Abbildung 3.6: Initialisierung der Waveletzentren. (a) 1. Ebene mit 12 Wavelets, (b) 2. Ebene mit 30 bzw. 52 Wavelets.

enthält bei beiden Varianten 12 Wavelets, die zweite 30 bzw. 52. Die Dilatationsparameter der Wavelets werden so initialisiert, daß wie zuvor auf der halben Distanz benachbarter Wavelets die einhüllende Gaußfunktion auf den Wert 0.5 abgefallen ist.

Die Initialisierung der Dilatationsparameter, Orientierungen und Gewichte unterscheidet sich nicht von den im Abschnitt 2.5.3 beschriebenen Verfahren. Abbildung 3.7 zeigt die initialisierten Waveletnetze der drei Referenzbilder aus Abbildung 3.4. Wie in Abschnitt 3.1 beschrieben, wird vor der Initialisierung vom Referenzbild eine an das innere

(a)

(b)

Abbildung 3.7: Initialisierte Waveletnetze. (a) 42 Wavelets, (b) 64 Wavelets.

Gesicht optimal angepaßte Grauwertebene abgezogen, um die Einflüsse verschiedener Beleuchtungsrichtungen zu verringern. Zur Lösung des Ausgleichsproblems werden damit nur die Pixel herangezogen, die im elliptischen Bildausschnitt des inneren Gesichts liegen.

3.1.3 Optimierung der Waveletnetze

Die Optimierung der Waveletnetze zur Repräsentation von Gesichtern erfolgt mit dem im Abschnitt 2.5.4 beschriebenen Levenberg-Marquardt-Algorithmus. In diesem Abschnitt beschreiben wir die gegenüber Waveletnetzen zur allgemeinen Grauwertbildrepräsentation vorgenommenen Modifikationen des Optimierungsverfahrens. Wir haben den Bereich, über dem der Approximationsfehler berechnet wird und die Regularisierung sowohl der Waveletzentren als auch der Dilatationsparameter verändert, um zu verhindern, daß einzelne Wavelets aus dem Gesichtsbereich herausragen.

Die Auswertung der Fehlerfunktion χ^2 eines Waveletnetzes $\mathcal{W}$ gemäß Gleichung (2.27) wird bei der Optimierung auf einen rechteckigen Bereich eingeschränkt. Diese im folgenden als **Auswertungsbereich** $A_{\mathcal{W}}$ bezeichnete Region ergibt sich als umschreibendes Rechteck der elliptischen Gesichtsregion aus dem vorigen Abschnitt. Abbildung 3.8 zeigt die Auswertungsbereiche anhand der drei Referenzbilder aus dem letzten Abschnitt. Durch die Einschränkung des Auswertungsbereiches entstehen bei der Opti-

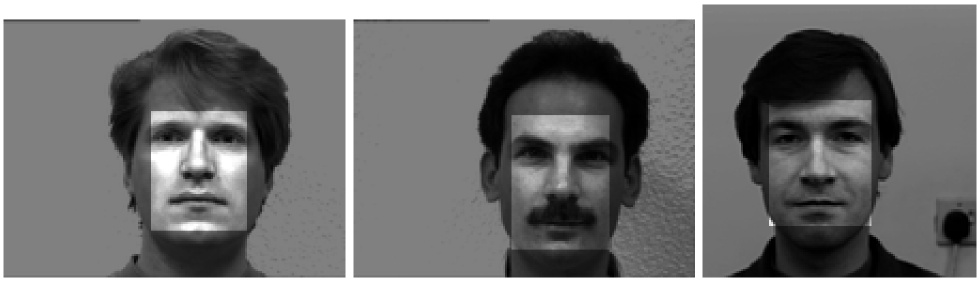
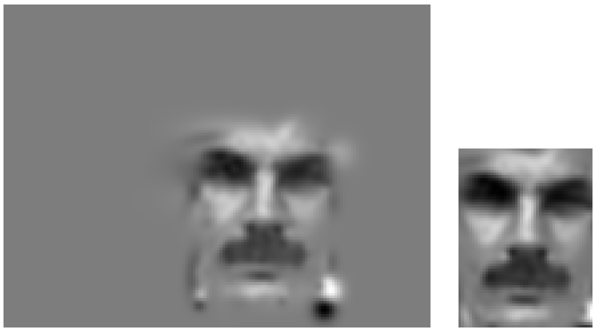


Abbildung 3.8: Auswertungsbereiche der Waveletnetze.

mierung unerwünschte **Randeffekte**, verursacht durch am Rand des Auswertungsbereiches positionierte Wavelets, die versuchen, Teile des Hintergrunds zu modellieren. Auch wenn die Zentren der Wavelets noch innerhalb des erlaubten Bereiches liegen, kann deren Träger aus diesem herausragen. Die Gewichte solcher Randwavelets werden im Optimierungsverfahren so angepaßt, daß der Anteil innerhalb des Auswertungsbereiches möglichst gut lokal zur Approximation beiträgt. Die Auswirkung des Waveletanteils außerhalb des betrachteten Bereiches bleibt dabei unberücksichtigt. Bei dem späteren Matching werden Waveletnetzes jedoch auch außerhalb des zur Optimierung festgelegten Auswertungsbereiches ausgewertet. Deshalb müssen Randeffekte, wie sie in Abbildung 3.9 gezeigt werden, vermieden werden. So dient beispielsweise in Abbildung 3.9(a) nur der obere Bereich des stark gewichteten Wavelets in der rechten unteren Ecke zur Modellierung des hellen Hintergrunds. Daß der dunkle, untere Bereich des Wavelets, der außerhalb des Auswertungsbereiches liegt, nicht zum Bild "paßt", ist für das Optimierungsverfahren nicht sichtbar, da nur der Auswertungsbereich zur Berechnung des Approximationsfehlers berücksichtigt wird.

Um die geschilderten Randeffekte zu vermeiden, vergrößern wir zunächst den Auswertungsbereich. Abbildung 3.10 zeigt das Ergebnis am Beispiel des Waveletnetzes aus Abbildung 3.9. Der Auswertungsbereich des Approximationsfehlers ist in diesem Bild um 20 % vergrößert. Trotz dieser Vergrößerung wird die Regularisierung der Waveletzentren bezüglich des ursprünglichen Bildausschnitts beibehalten, damit die Wavelets auf dem inneren Gesicht positioniert bleiben. In Abbildung 3.10 wird deutlich, daß die am Rand liegenden Wavelets nun komplett im Auswertungsbereich liegen und korrekt sowohl an den Gesichtsrand als auch an den Hintergrund angepaßt sind.

Da die Gesichtsrepräsentation weiter unten als affin transformierbares Template zum Matchen verwendet wird, soll möglichst vermieden werden, daß einzelne Wavelets ver-



(a) (b)

Abbildung 3.9: (a) Randeffekte, verursacht durch Wavelets, deren Träger nicht komplett im Auswertungsbereich des Waveletnetzes liegt. (b) Dasselbe Waveletnetz, eingeschränkt auf den Auswertungsbereich.

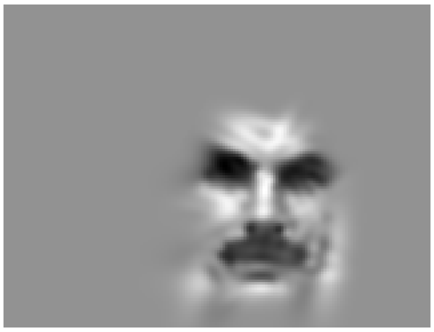


Abbildung 3.10: Ergebnis der Optimierung bei um 20 % vergrößertem Auswertungsbereich. Die Zentren der Wavelets bleiben weiterhin auf den ursprünglichen Bildausschnitt eingeschränkt.

suchen, den Hintergrund zu modellieren, wie es in Abbildung 3.10 immer noch der Fall ist. Deshalb regularisieren wir die Zentren der Wavelets nun zusätzlich so, daß das Verlassen des im vorigen Abschnitt beschriebenen elliptischen Bereiches stark bestraft wird. Für unser Beispielnetz zeigt Abbildung 3.11 das Ergebnis der Einschränkung der Waveletzentren auf den elliptischen Bereich.

Bei sämtlichen in diesem Abschnitt gezeigten Waveletnetzen haben wir stillschweigend eine weitere Ursache dafür beseitigt, daß die Träger von Wavelets den Auswertungsbereich überschreiten. Die von Pelc in [21] beschriebene Regularisierung der Dilatationsparameter verhindert nicht, daß Wavelets in einer Richtung stark gestaucht, in der anderen hingegen gestreckt werden und sich so insgesamt “in die Länge ziehen”

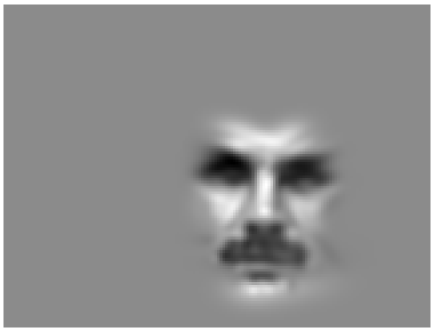
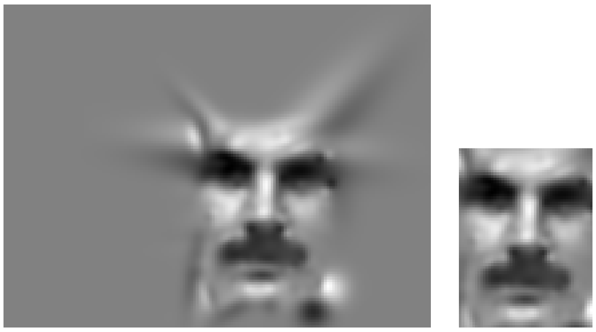


Abbildung 3.11: Regularisierung der Waveletzentren bezüglich des elliptischen Bereiches, der dem inneren Gesicht entspricht. Eine Approximation des Hintergrunds findet nicht mehr statt.

und dadurch aus dem Auswertungsbereich herausragen. Abbildung 3.12 zeigt dazu ein Beispiel. Der bei der Regularisierung der Dilatationsparameter in [21] erlaubte Para-



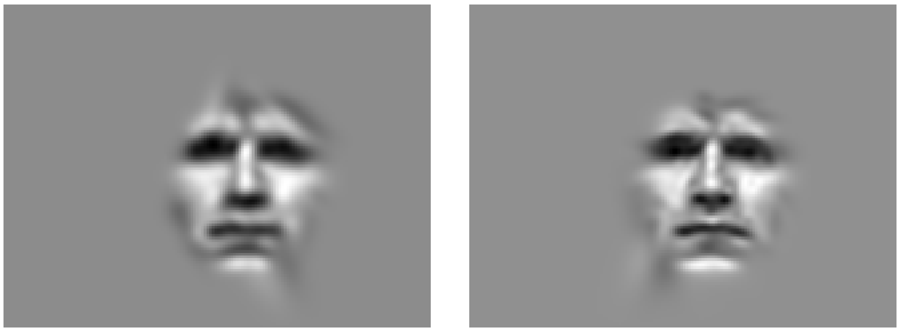
(a) (b)

Abbildung 3.12: (a) Langgezogene Wavelets ragen über den Auswertungsbereich. (b) Dasselbe Netz auf den Auswertungsbereich eingeschränkt.

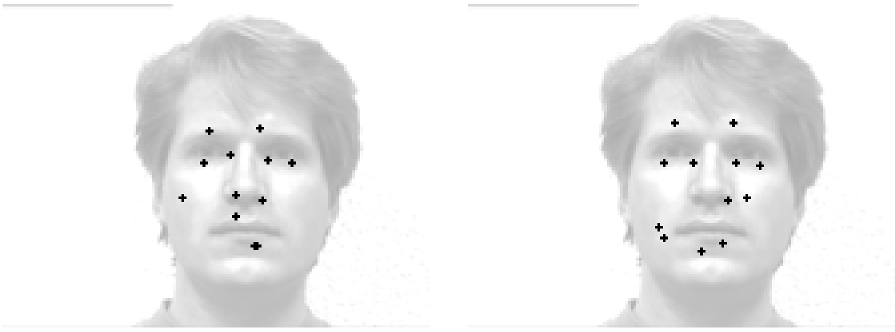
meterbereich für die Dilatationsparameter war durch die initialen Werte der gröbsten und der feinsten Pyramidenebenen gegeben. Demnach kann sich ein Wavelet in einer Richtung bis auf den Wert der feinsten Ebene zusammenziehen und in der anderen Richtung bis auf den Wert der gröbsten Ebene strecken, was in Abbildung 3.12 zu sehen ist. Um diese Effekte zu vermeiden, regularisieren wir den Dilatationsparameter d_x und sorgen durch Bestrafungsterme dafür, daß das **Verhältnis** von d_y und d_x in einem festlegbaren Intervall (z.B. 0.5 bis 1.5) liegt. Damit können die Skalen nicht länger unabhängig voneinander angepaßt werden, extrem langestreckte Wavelets treten nicht mehr auf. Es sei darauf hingewiesen, daß diese Strategie zur Gewinnung von Repräsentation anderer Objekte, die langgestreckte, gerade Kanten besitzen, nicht optimal ist, im Fall der Gesichtsrepräsentation jedoch zu guten Ergebnissen führt.

Die im letzten Absatz beschriebene Regularisierung der Dilatationsparameter läuft noch der Grob-Fein-Strategie zuwider. Zwar werden die Wavelets auf zwei Skalenstufen initialisiert, der zulässige Wertebereich für den Parameter d_x ist jedoch global über das gesamte Netz festgelegt. Dadurch können sich nach der Optimierung in der "groben" Ebene kleinere Wavelets befinden als in der "feinen". Deshalb regularisieren wir den Dilatationsparameter d_x unter Berücksichtigung seines Initialwertes. Abweichung von $\pm 10\%$ vom initialen Wert werden zugelassen. So wird sichergestellt, daß ein Waveletnetz zur Gesichtsrepräsentation auch nach der Optimierung über eine Ebene grobskaliger und eine Ebene feinskaliger Wavelets verfügt.

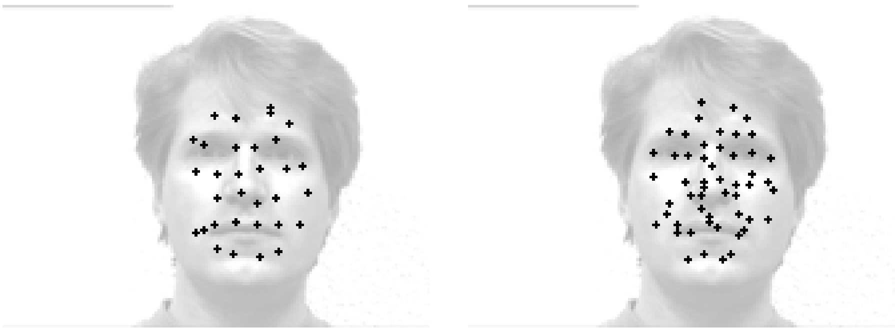
Abbildung 3.13 zeigt die Resultate des Optimierungsverfahrens unter Berücksichtigung der in diesem Abschnitt beschriebenen Modifikationen.



(a)



(b)



(c)

Abbildung 3.13: Waveletnetze zur Repräsentation von Gesichtern. Links: Waveletnetz mit 42 Wavelets, rechts Waveletnetz mit 64 Wavelets. (a) Ausgabe der Waveletnetze, (b) Positionen der Wavelets der ersten Ebene, (c) Waveletpositionen der 2. Ebene.

3.2 Matching mit Superwavelets

Im letzten Abschnitt haben wir beschrieben, wie wir unsere Gesichtsrepräsentation erzeugen. Für jede zu identifizierende Person wird ein individuelles Waveletnetz an das innere Gesicht des Referenzbildes angepaßt. In diesem Abschnitt verwenden wir die Waveletnetz-Repräsentation als affin transformierbares Superwavelet. Dieses Superwavelet dient als kontinuierlich affin transformierbares Template. Dieses Template wird verwendet, um im dem System als Eingabe gegebenen Testbild die genaue Position, Größe und Orientierung des zu identifizierenden Gesichts festzustellen. Durch diese Vorgehensweise entfällt die Notwendigkeit, Testbilder unter genau festgelegten Bedingungen zu akquirieren, um sicherzustellen, daß in jedem Testbild das zu erkennende Gesicht sich an einer bestimmten Position und Orientierung im jeweils gleichen Abstand zur Kamera befindet.

Beim im Abschnitt 2.5.7 vorgestellten Konzept der Superwavelets wird ein gesamtes Waveletnetz aufgefaßt als Wavelet, das kontinuierlich verschoben, gestreckt und gestaucht sowie rotiert werden kann. Zu einem Waveletnetz $\mathcal{W} = (\mathit{\Psi}, \boldsymbol{w}, m)$ hat das zugehörige Superwavelet die Form

$$\Psi_{\boldsymbol{n}}(\boldsymbol{x}) = \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}\left(\boldsymbol{D}\boldsymbol{R}[\boldsymbol{x} - \boldsymbol{t}]\right) + m \;, \tag{3.4}$$

mit $\boldsymbol{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$, $\boldsymbol{D} = \begin{pmatrix} d_x & d_{xy} \\ 0 & d_y \end{pmatrix}$, $\boldsymbol{R} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ und dem Parametervektor $\boldsymbol{n} = (t_x, t_y, d_x, d_y, d_{xy}, \alpha)^T$.

Das Matching mit dem Superwavelet erfolgt über die iterative Bestimmung eines Parametervektors $\boldsymbol{n}_{\text{opt}}$, so daß der quadratische Fehler zwischen Superwavelet $\Psi_{\boldsymbol{n}}$ und Testbild h

$$\chi^2(\boldsymbol{n}) = \frac{1}{2} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} \left[h(m,n) - \Psi_{\boldsymbol{n}}(m,n)\right]^2 \tag{3.5}$$

für $\boldsymbol{n}_{\text{opt}}$ minimal wird.

Dieses Minimierungsproblem lösen wir analog zur Anpassung der einzelnen Wavelets mit dem Levenberg-Marquardt-Algorithmus, wobei hier beim Matching nur der Parametervektor $\boldsymbol{n}$ des Superwavelets verändert wird, die Parameter $\boldsymbol{a}_i$ und Gewichte w_i der einzelnen Wavelets werden dabei natürlich festgehalten. Es ist hervorzuheben, daß zum Matchen **derselbe** Algorithmus verwendet wird, wie bei der Anpassung der einzelnen Wavelets zur Gewinnung der individuellen Gesichtsrepräsentation. Damit bewährt sich das Konzept der Superwavelets, da ein gesamtes Waveletnetz als einzelnes Wavelet mit entsprechenden Freiheitsgraden interpretiert wird.

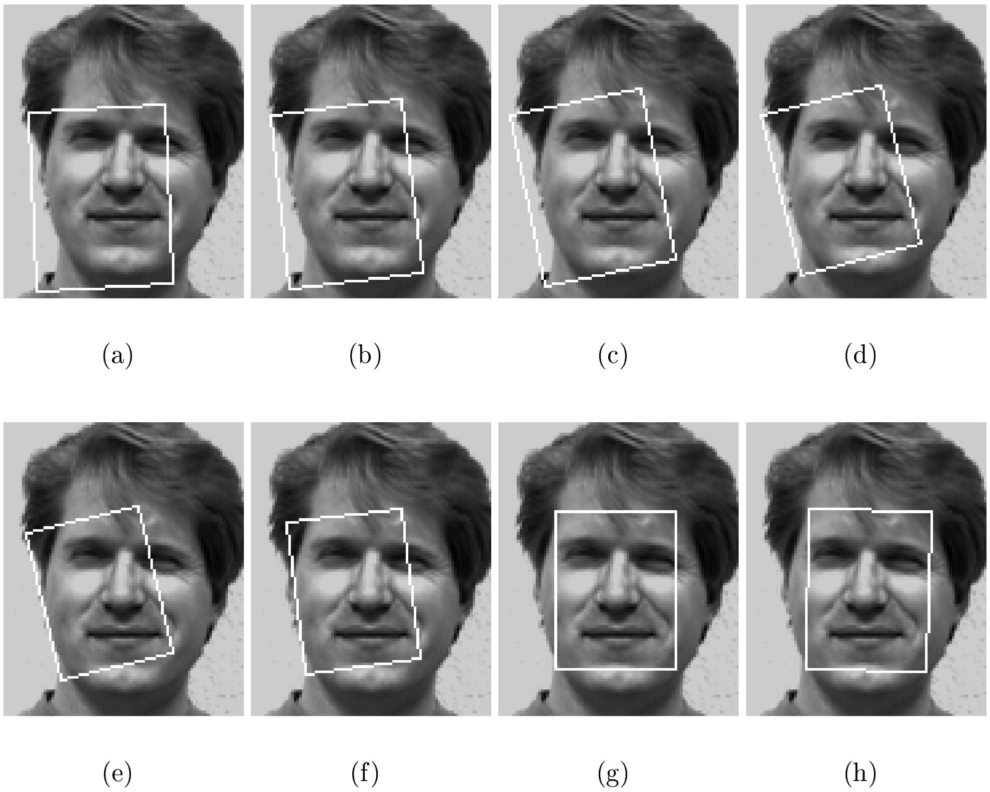


Abbildung 3.14: Matching mit dem Superwavelet. Es wird ein Waveletnetz mit 64 Wavelets verwendet, das an das Referenzgesicht aus Abbildung 3.4 angepaßt ist. Position, Orientierung und Skala werden durch den Rahmen veranschaulicht. Das in diesem Beispiel verwendete Bild unterscheidet sich vom Referenzgesicht durch den Gesichtsausdruck. (a) Ausgangsposition für den Optimierungsalgorithmus, in diesem Beispiel beträgt die Abweichung im Ort horizontal und vertikal jeweils 10 Pixel, als Startwert für die Orientierung wurde 3° gewählt. Das Superwavelet ist um 10% gegenüber dem ursprünglichen Waveletnetz vergrößert. (b) - (g) Superwavelets nach $5, 10, \ldots 30$ Iterationen, (h) Ergebnis nach Konvergenz des Algorithmus (32 Iterationen).

Die drei Dilatationsparameter d_x, d_y und d_{xy} erlauben eine unabhängige Skalierung des Superwavelets in $x-$ und $y-$Richtung sowie eine Scherung. Mit diesen Freiheitsgraden läßt sich die orthographische Projektion eines Gesichts in die Bildebene gut modellieren. Die Gesichtserkennung soll jedoch in Frontalaufnahmen von Gesichtern erfolgen, so daß lediglich eine gleichmäßige Skalierung in $x-$ und $y-$ Richtung erforderlich ist. Deshalb wird für das Matching der Parameter d_{xy} auf Null gesetzt und festgehalten. Der Dilatationsparameter d_x in $x-$Richtung wird im Optimierungsalgorithmus freigegeben, der Parameter d_y so angepaßt, daß ein konstantes Verhältnis

$\frac{d_x}{d_y} = 1$ gegeben ist. Damit erreichen wir eine gleichmäßige Skalierung des Superwavelets in $x-$ und $y-$Richtung. Im Abschnitt 3.3 skizzieren wir kurz, wie das Matching mit Superwavelets zum Verfolgen von Gesichtern in Bildsequenzen verwendet wird. Dabei ist es sinnvoll, sämtliche affinen Freiheitsgrade freizugeben.

Für den Levenberg-Marquardt-Algorithmus wird ein initialer Parametervektor als Startwert benötigt. In mehreren Versuchen haben wir festgestellt, daß die Startposition im allgemeinen bis zu ± 10 Pixel von der tatsächlichen Position des Gesichts entfernt liegen sollte. Andernfalls konvergiert das Verfahren häufig in lokalen Minima, die nicht der gesuchten Position entsprechen. Für die Orientierung erhalten wir entsprechend eine Toleranz von ca. $\pm 10°$. Die Dilatationsparameter sollten im Bereich von 0.9 bis 1.1 liegen, was einer Abweichung in der Skala von $\pm 10\%$ entspricht. Für unsere Experimente nutzen wir für die Bestimmung des Startwerts des Matchingverfahrens einen Gesichtsdetektor [8], der auf dem von Sung und Poggio in [25] veröffentlichten Verfahren basiert. Dieser Detektor liefert eine gute Schätzung der Position und Skala des Gesichts im Testbild. Der oben beschriebene Spielraum für den Startwert läßt jedoch auch einfachere Verfahren zur Bestimmung des Startwerts zu, z.B. einfache Segmentierungsverfahren basierend auf Grau- oder Farbwerten.

Abbildung 3.14 zeigt einige Iterationen des Levenberg-Marquardt-Algorithmus, die Position, Orientierung und Skala des Superwavelets wird durch einen weißen Rahmen illustriert.

Die oben angegebene Toleranz von ± 10 Pixeln entspricht einer relativen Toleranz gegenüber Verschiebungen von 17%, bezogen auf die Breite des inneren Gesichts. Edward, Cootes und Taylor verwenden ein vergleichbares Matchingverfahren, um ihr Active-Appearance-Model, das wir in Abschnitt 1.1 kurz vorgestellt haben, auf ein Testgesicht zu positionieren. In [4] sprechen sie von einer Toleranz von 7.5%. Damit erweist sich das Matching mit Waveletnetzen als deutlich robuster bezüglich der Wahl der Startposition.

Im Abschnitt 3.4, der die sich an das Matching anschließende Rekonstruktion des Testgesichts beschreibt, betonen wir die Wichtigkeit der Reparametrisierung eines Superwavelets im kontinuierlichen Parameterraum. Je genauer das durch das Superwavelet gegebene Gesichtstemplate auf das Testgesicht paßt, desto besser wird eine Rekonstruktion möglich sein.

3.3 Anwendung des Matchings: Verfolgen von Gesichtern

In unseren Experimenten haben wir festgestellt, daß der Optimierungsalgorithmus für das Matching mit Superwavelets in wenigen Iterationsschritten konvergiert, wenn der

Startwert für den Parametervektor $\boldsymbol{n}$ im Parameterraum dicht am optimalen Wert liegt, der den quadratischen Fehler zwischen Gesicht und Superwavelet minimiert.

In Videosequenzen mit Aufnahmen von Gesichtern z.B. bei Telekonferenz-Systemen sind die Abweichungen in Position, Orientierung und Skala zwischen aufeinanderfolgenden Bildern einer Sequenz verhältnismäßig gering. Dies legt nahe, Superwavelets mit dem im vorigen Abschnitt beschriebenen Matchingverfahren zur Verfolgung von Gesichtern in Videosequenzen einzusetzen. Das Ergebnis des Matchingverfahrens für Bild i der Sequenz wird als Startwert für das Matching in Bild $i+1$ verwendet usw. In der Regel konvergiert das Matchingverfahren wegen des jeweils guten Startwerts in wenigen Iterationsschritten. Damit ist ein Verfolgen von Gesichtern in Echtzeit möglich [12].

Abbildung 3.15 zeigt ein Beispiel für das Verfolgen von Gesichtern in einer Sequenz von Grauwertbildern.

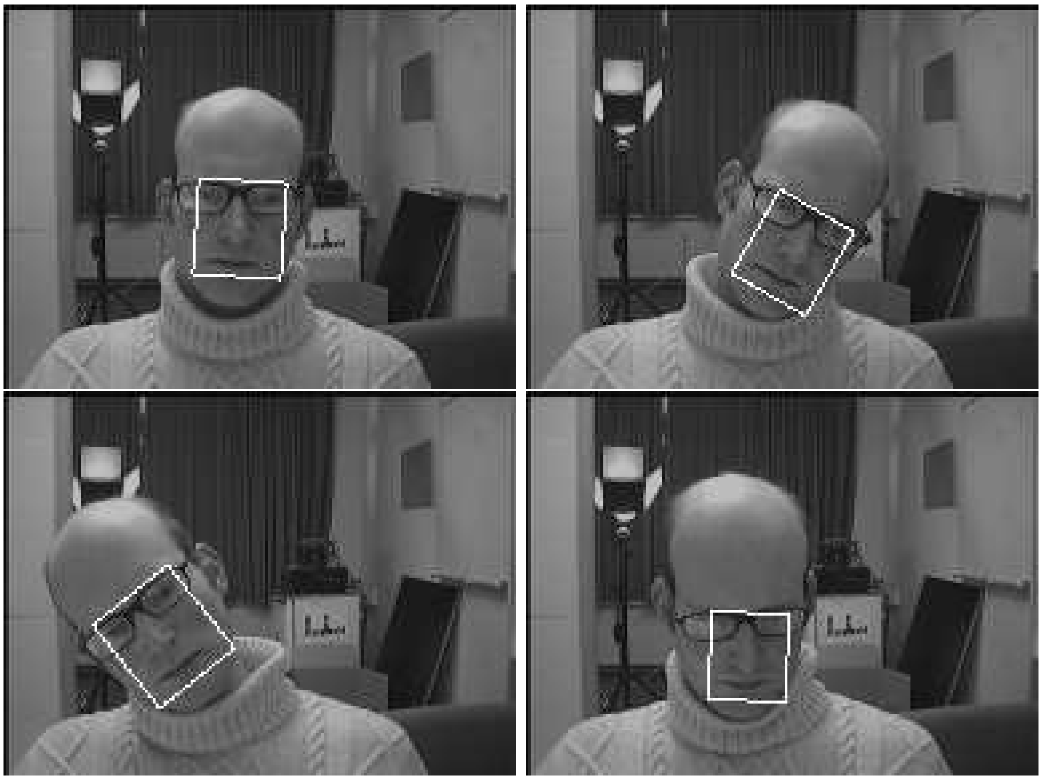


Abbildung 3.15: Verfolgen von Gesichtern in einer Grauwertbildsequenz.

3.4 Rekonstruktion mit Superwavelets

Durch das im Abschnitt 3.2 beschriebene Matchingverfahren sind wir in der Lage, mit Hilfe von Waveletnetzen die Größe, Position und Orientierung eines Gesichts in einem gegebenen Testbild zu bestimmen. Interpretiert als Superwavelet wird das Waveletnetz in dem Matchingverfahren so transformiert, daß bei festgehaltenen Gewichten ein lokales Minimum des quadratischen Fehlers zwischen Testgesicht und der Ausgabe des Waveletnetzes eingenommen wird. Diese Transformation bezeichnen wir im folgenden als **Reparametrisierung** bezüglich des Testgesichts.

Sei g das Referenzgesicht, auf das das Waveletnetz $\mathcal{W} = (\varPsi, \boldsymbol{w}, m)$ angepaßt wurde. Durch das Matchingverfahren, in dem das Waveletnetz $\mathcal{W}$ als Superwavelet $\Psi_{\boldsymbol{n}}(\boldsymbol{x})$ aufgefaßt wird, erhalten wir einen optimalen Parametervektor $\boldsymbol{n}_{\mathrm{opt}}$, der das Superwavelet so reparametrisiert, daß das Gesicht im Testbild h durch das Superwavelet bestmöglich angenähert wird.

Ein reparametrisiertes Superwavelet kann umgekehrt auch als Waveletnetz aufgefaßt werden. Mit den Notationen aus Abschnitt 2.5.7 erhalten wir:

$$\begin{aligned}\Psi_{\boldsymbol{n}_{\mathrm{opt}}}(\boldsymbol{x}) &= \sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i}(\boldsymbol{D}\boldsymbol{R}[\boldsymbol{x}-\boldsymbol{t}]) + m \\ &= \sum_{i=1}^{N} w_i \psi(\boldsymbol{D}_i\boldsymbol{R}_i\,[\boldsymbol{D}\boldsymbol{R}[\boldsymbol{x}-\boldsymbol{t}] - \boldsymbol{t}_i]) + m \\ &= \sum_{i=1}^{N} w_i \psi(\boldsymbol{D}'_i\boldsymbol{R}'_i(\boldsymbol{x}-\boldsymbol{t}'_i)) + m \\ &= \sum_{i=1}^{N} w_i \psi'_{\boldsymbol{a}_i} + m\ , \end{aligned} \tag{3.6}$$

mit

$$\psi'_{\boldsymbol{a}_i}(\boldsymbol{x}) = \psi(\boldsymbol{D}'_i\boldsymbol{R}'_i[\boldsymbol{x}-\boldsymbol{t}'_i]) \tag{3.7}$$

und

$$\boldsymbol{t}'_i = \boldsymbol{t} + \boldsymbol{R}^{-1}\boldsymbol{t}_i\ .$$

Die Matrizen $\boldsymbol{D}'_i$ und $\boldsymbol{R}'_i$ erhält man jeweils aus der Zerlegung der Matrix $\boldsymbol{D}_i\boldsymbol{R}_i\boldsymbol{D}\boldsymbol{R}$ in eine untere Dreiecksmatrix $\boldsymbol{D}'_i$ und eine orthogonale Matrix $\boldsymbol{R}'_i$.[1]

[1] Für eine nichtsinguläre Matrix $\boldsymbol{M}$ existiert die sogenannte QR-Zerlegung $\boldsymbol{M} = \boldsymbol{Q}\boldsymbol{R}$ in eine orthogonale Matrix $\boldsymbol{Q}$ und obere Dreiecksmatrix $\boldsymbol{R}$, die sich mit dem Householder-Verfahren berechnen läßt [24]. Daraus folgt, daß für $\boldsymbol{M}$ eine LQ-Zerlegung in eine untere Dreiecksmatrix $\boldsymbol{L}$ und eine orthogonale Matrix $\boldsymbol{Q}$ existiert. Seien $\boldsymbol{Q}$ und $\boldsymbol{R}$ die Matrizen der QR-Zerlegung von $\boldsymbol{M}^T$, also $\boldsymbol{M}^T = \boldsymbol{Q}\boldsymbol{R}$, dann folgt $\boldsymbol{M} = \boldsymbol{R}^T\boldsymbol{Q}^T$, wobei $\boldsymbol{R}^T$ eine untere Dreiecksmatrix und $\boldsymbol{Q}^T$ orthogonal sind.

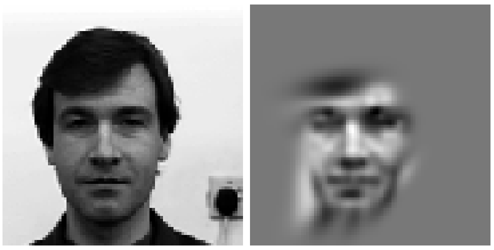
Gleichung (3.6) stellt die Ausgabe eines Waveletnetzes dar, dessen Wavelets durch Gleichung (3.7) gegeben sind. Damit haben wir gezeigt, daß durch das Matchingverfahren das Waveletnetz $\mathcal{W} = (\Psi, \boldsymbol{w}, m)$ in ein Waveletnetz $\mathcal{W}' = (\Psi', \boldsymbol{w}, m)$ überführt wird, indem die Basisfunktionen aus Ψ transformiert werden in $\Psi' = \{\psi'_{\boldsymbol{a}_1}, \ldots, \psi'_{\boldsymbol{a}_N}\}$ gemäß Gleichung (3.7).

Ein an eine Referenzperson individuell angepaßtes Waveletnetz beschreibt durch die Linearkombination der Basisfunktionen (Wavelets) die Merkmale des Gesichts der Person. Die Wavelets sind in ihrer Position, Skala und Orientierung den individuellen Gesichtsmerkmalen angepaßt. Trotz unterschiedlicher Gesichtsausdrücke bleiben wesentliche Merkmale des Gesichts weitgehend unverändert. Das bedeutet, daß ein Großteil der individuell angepaßten Wavelets nach der Reparametrisierung durch das Matching weiterhin gut zu den Merkmalen der Referenzperson "passen", also eine neue Gewichtung der modifizierten Basisfunktionen gefunden werden kann, die das veränderte Gesicht rekonstruieren. Die Gewichte bezüglich des transformierten Satzes an Basisfunktionen sind neu zu bestimmen. Abbildung 3.16 zeigt Beispiele für die Rekonstruktion mit einem Waveletnetz bestehend aus 42 Wavelets. Abbildung 3.16(a) zeigt das Referenzbild sowie das angepaßte Waveletnetz mit 42 Wavelets. Abbildung 3.16(b) zeigt Testbilder mit Aufnahmen der Referenzperson, auf die das Waveletnetz optimiert wurde. Die Rekonstruktionen dieser Bilder nach der Reparametrisierung sind in Abbildung 3.16(c) abgebildet. Abbildung 3.17 zeigt die Rekonstruktionsergebnisse von anderen, nicht zur Referenzperson gehörenden Testbildern. Es zeigt sich, daß das Superwavelet durch die spärliche Repräsentation mit einer geringen Zahl von Basisfunktion diese Bilder deutlich schlechter rekonstruieren kann. Viele der Wavelets, die die lokalen Merkmale des Referenzgesichts repräsentieren, weichen in ihrer Lage von den Merkmalen der Testgesichter ab. Es kann keine Gewichtung gefunden werden, die eine gute Rekonstruktion erlaubt. Diese Beobachtung bildet die Grundlage für unsere in den Abschnitten 3.5 und 3.6 vorgestellten Ansätze zur Erkennung von Gesichtern.

Die Rekonstruktion, d.h. die Berechnung neuer Gewichte, erfolgt gemäß Abschnitt 2.5.6, der die Bestimmung optimaler Gewichte bei einem gegebenen Satz von Basisfunktionen beschreibt. Dazu wird das Testgesicht in den durch die Wavelets $\psi'_{\boldsymbol{a}_i} \in \Psi'$ aufgespannten Unterraum projiziert. Zur Berechnung der orthogonalen Projektion verwenden wir den Ansatz über die Gramsche Matrix $\boldsymbol{\Psi}'$, deren Einträge durch die Skalarprodukte der Wavelets $\psi'_{\boldsymbol{a}_i}$ gegeben sind. Die Auswertung der Skalarprodukte der Wavelets sowie der Projektionen der Wavelets auf das Testgesicht beschränken wir auf den Auswertungsbereich $A_{\Psi'}$ des Waveletnetzes. Damit erhalten wir als Skalarprodukte

$$<\psi_{\boldsymbol{a}_i}, \psi_{\boldsymbol{a}_j}>_{A_{\Psi'}} = \sum_{(m,n)\in A_\Psi} \psi_{\boldsymbol{a}_i}(m,n)\psi_{\boldsymbol{a}_j}(m,n) \tag{3.8}$$

$$<\psi_{\boldsymbol{a}_i}, h>_{A_{\Psi'}} = \sum_{(m,n)\in A_\Psi} \psi_{\boldsymbol{a}_i}(m,n)h(m,n)\ . \tag{3.9}$$



(a)

(b)

(c)

Abbildung 3.16: Rekonstruktion mit Superwavelets. (a) Referenzperson und an diese angepaßtes Waveletnetz mit 42 Wavelets, (b) Testbilder mit Aufnahmen der Referenzperson, (c) Rekonstruktion der Testbilder nach erfolgter Reparametrisierung.

Mit diesen Skalarprodukten können die Gewichte gemäß Gleichung (2.45) berechnet werden.

Die Verwendung der Gramschen Matrix hat gegenüber der Berechnung über die Pseudoinverse den Vorteil, daß sich unter der Reparametrisierung die Matrix $\mathbf{\Psi}'$ aus $\mathbf{\Psi}$ durch eine einfache Multiplikation mit einer Konstanten ergibt. Damit unterscheiden sich die Inversen $\mathbf{\Psi}^{-1}$ und $\mathbf{\Psi}'^{-1}$ ebenfalls nur durch einen Faktor voneinander. Eine zeitaufwendige Neuberechnung der Gramschen Matrix und ihrer Inversen ent-

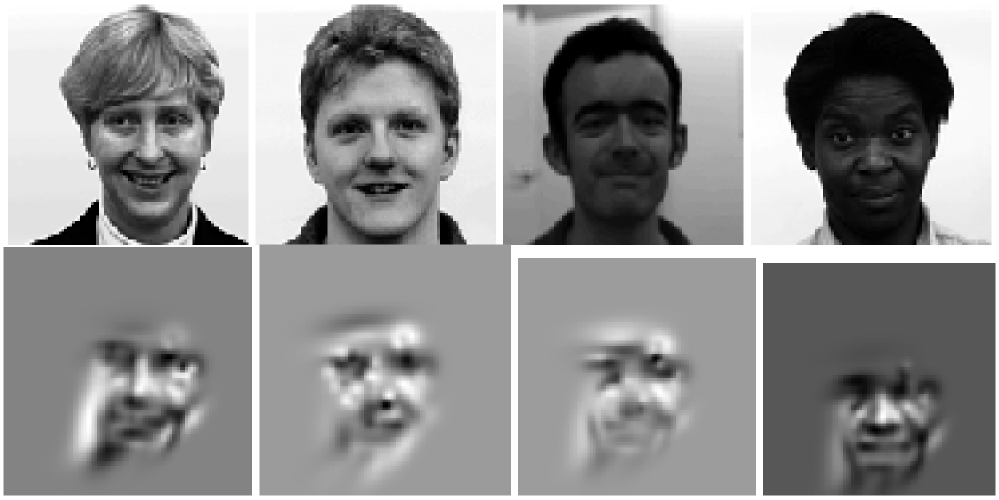


Abbildung 3.17: Rekonstruktion mit dem Waveletnetz aus Abbildung 3.16 bei Testbildern mit von der Referenzperson abweichenden Gesichtern.

fällt damit, wenn die Inverse der Gramschen Matrix zusammen mit dem Waveletnetz gespeichert und nach dem Matchen durch die Multiplikation mit dem der Reparametrisierung entsprechenden Faktor aktualisiert wird. Der Faktor für die Anpassung der Gramschen Matrix wird dabei wie folgt bestimmt:

Für ein reparametrisiertes Wavelet $\psi'_{\boldsymbol{a}_i}$ gilt

$$\psi'_{\boldsymbol{a}_i}(x,y) = \psi_{\boldsymbol{a}_j}(u(x,y), v(x,y)) \tag{3.10}$$

mit

$$\begin{aligned}\begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix} &= \boldsymbol{DR}\left(\begin{pmatrix} x \\ y \end{pmatrix} - \boldsymbol{t}\right) \\ &= \begin{pmatrix} d_x & d_{xy} \\ 0 & d_y \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x - t_x \\ y - t_y \end{pmatrix} \;,\end{aligned}$$

also

$$u(x,y) = (d_x \cos\alpha + d_{xy} \sin\alpha)(x - t_x) - (d_x \sin\alpha - d_{xy} \cos\alpha)(y - t_y) \tag{3.11}$$

$$v(x,y) = d_y \sin\alpha (x - t_x) + d_y \cos\alpha (y - t_y) \;. \tag{3.12}$$

Damit geht ein reparametrisiertes Wavelet $\psi'_{\boldsymbol{a}_i}$ durch eine Koordinatentransformation, gegeben durch $u(x,y)$ und $v(x,y)$, in das ursprüngliche Wavelet $\psi_{\boldsymbol{a}_i}$ über. Die Funktionaldeterminante der Koordinatentransformation ist gegeben durch

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = d_x d_y \;.$$

Mit dem Zusammenhang

$$\frac{\partial(x,y)}{\partial(u,v)} = \frac{1}{\frac{\partial(u,v)}{\partial(x,y)}}$$

erhalten wir als Skalarprodukt zweier reparametrisierter Wavelets $\psi'_{\boldsymbol{a}_i}$ und $\psi'_{\boldsymbol{a}_j}$

$$\begin{aligned} <\psi'_{\boldsymbol{a}_i}, \psi'_{\boldsymbol{a}_j}> &= \int_{-\infty}^{+\infty} \psi'_{\boldsymbol{a}_i}(x,y)\psi'_{\boldsymbol{a}_j}(x,y)\, dxdy \\ &= \int_{-\infty}^{+\infty} \psi_{\boldsymbol{a}_i}(u(x,y), v(x,y))\psi_{\boldsymbol{a}_j}(u(x,y), v(x,y))\, dxdy\ , \\ &= \int_{-\infty}^{+\infty} \psi_{\boldsymbol{a}_i}(u,v)\psi_{\boldsymbol{a}_j}(u,v)\frac{\partial(x,y)}{\partial(u,v)}\, dudv \\ &= \frac{1}{d_x d_y} <\psi_{\boldsymbol{a}_i}, \psi_{\boldsymbol{a}_j}>\ . \end{aligned}$$

Damit gilt

$$\boldsymbol{\Psi}' = \frac{1}{d_x d_y}\boldsymbol{\Psi}\ ,$$

also

$$\boldsymbol{\Psi}'^{-1} = d_x d_y \boldsymbol{\Psi}^{-1}\ . \tag{3.13}$$

Damit erhalten wir analog zur Gleichung (2.45) die Gewichte w'_i der Rekonstruktion:

$$w'_i = d_x d_y \sum_{k=1}^{N} \boldsymbol{\Psi}^{-1}_{ik} <\psi'_{\boldsymbol{a}_k}, h>\ . \tag{3.14}$$

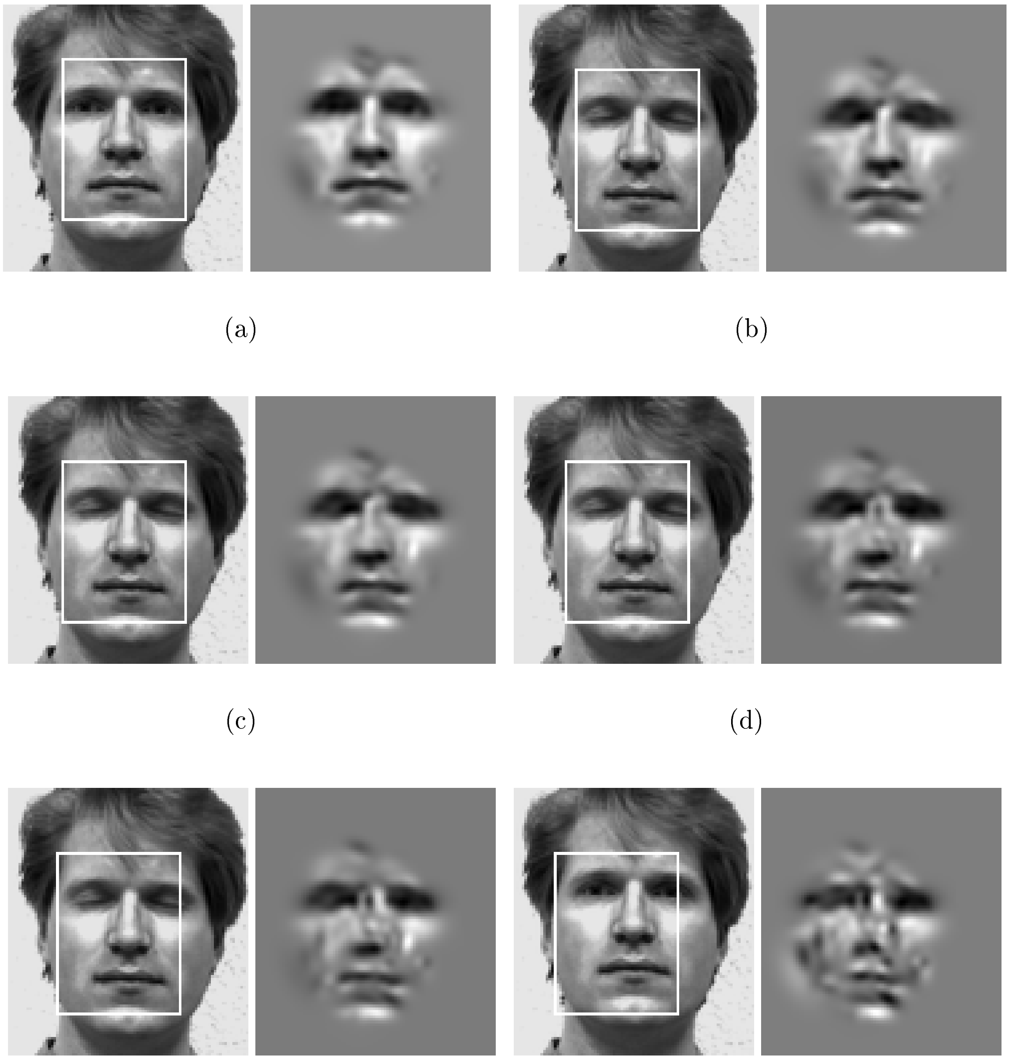
In Gleichung (3.14) werden die Projektionen $<\psi'_{\boldsymbol{a}_k}, h>$ der reparametrisierten Wavelets auf das Testbild im Auswertungsbereich des modifizierten Waveletnetzes ausgewertet. Der Auswertungsbereich wurde bei der Beschreibung der Anpassung des Waveletnetzes an die Referenzperson im Abschnitt 3.1.3 beschrieben.

Nach der Rekonstruktion erhalten wir ein Waveletnetz

$$\mathcal{W}'' = (\Psi', \boldsymbol{w}', m')\ ,$$

wobei m' der Mittelwert im Auswertungsbereich ist. In den Abbildungen 3.16 und 3.17 wurden bereits Beispiele für die Rekonstruktion mit Hilfe von Waveletnetzen gezeigt.

Nachdem wir erläutert haben, wie nach der Reparametrisierung die Gewichte der Wavelets zur Rekonstruktion des Testbildes berechnet werden, wollen wir zum Schluß



(a) (b) (c) (d) (e) (f)

Abbildung 3.18: Abhängigkeit der Qualität der Rekonstruktion von der Positionierung des Superwavelets. (a) Referenzperson und individuelles Waveletnetz. Der Auswertungsbereich des Netzes ist durch den Rahmen illustriert, (b) Referenzperson mit anderem Gesichtsausdruck und Rekonstruktion nach Reparametrisierung, (c) Rekonstruktion mit gegenüber (b) um 1 Pixel nach links verschobenen Superwavelet, (d) Verschiebung um 2 Pixel nach links, (e) Verschiebung um 3 Pixel nach links, (f) Verschiebung um 7 Pixel nach links.

dieses Abschnitts die Notwendigkeit einer genauen Reparametrisierung im kontinuierlichen Parameterraum aufzeigen. Die Abbildungen 3.18(b)-(f) zeigen die Ergebnisse der Rekonstruktion eines Testgesichts. In Abbildung 3.18(b) ist das Ergebnis der Rekonstruktion nach der Reparametrisierung mit dem Levenberg-Marquardt-Algorithmus abgebildet, die Abbildungen (c) - (f) zeigen die Rekonstruktionen mit gegenüber (b) nach links verschobenen Superwavelets. Je weiter die das Superwavelet gegenüber der durch das Matching gefundenen optimalen Position verschoben ist, desto schlechter wird die Rekonstruktion. Dies wird auch in Abbildung 3.19 deutlich, die die Abhängigkeit des mittleren Rekonstruktionsfehlers von der horizontalen Position zeigt.

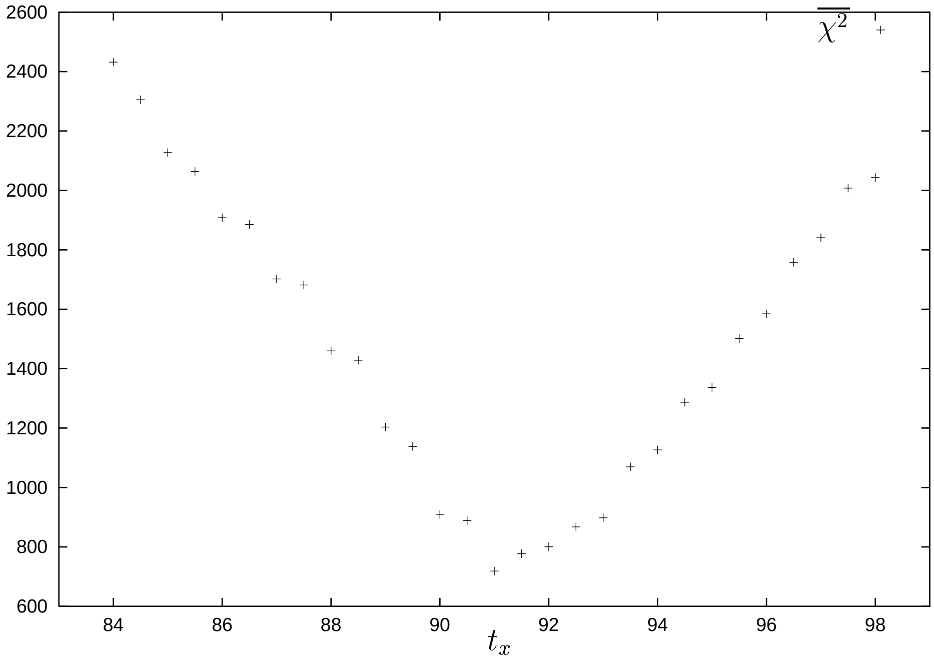
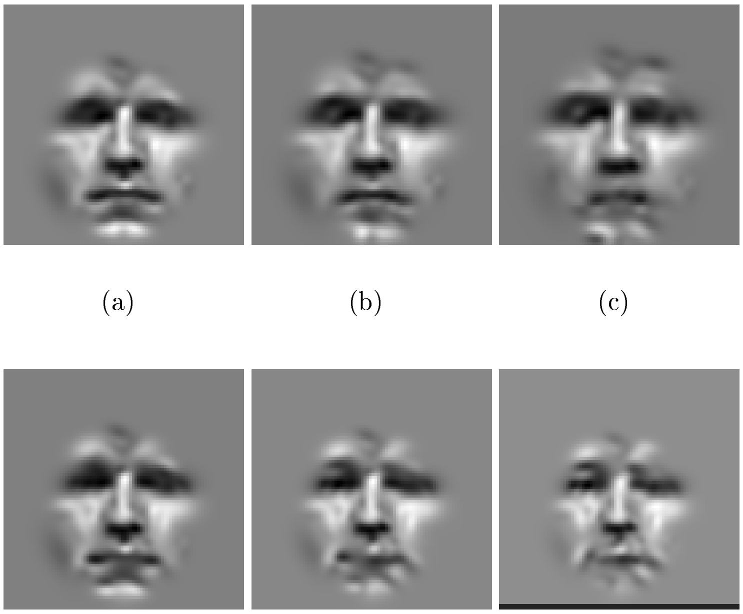


Abbildung 3.19: Rekonstruktionsfehler in Abhängigkeit der horizontalen Position t_x am Beispiel des Superwavelets und Testbildes aus Abbildung 3.18. Auf der y-Achse ist der mittlere quadratische Rekonstruktionsfehler $\overline{\chi^2}$ im Auswertungsbereich des jeweiligen Superwavelets abgetragen.

Die korrekte Skalierung durch das Matchingverfahren ist ebenso wichtig wie die Positionierung. Abbildung 3.20 zeigt Beispiele für die Rekonstruktion mit optimal positionierten, aber unterschiedlich skalierten Superwavelets.



(a) (b) (c)

(d) (e) (f)

Abbildung 3.20: Rekonstruktion in Abhängigkeit von der Skalierung des Superwavelets. (a) optimale Position und Skala, (b) Rekonstruktion mit um 5 % vergrößertem Superwavelet, (c) 10 % vergrößert, (d) 5 % verkleinert, (e) 10 % verkleinert, (f) 15 % verkleinert.

3.5 Ähnlichkeitsmaße

In diesem Abschnitt beschreiben wir die Ähnlichkeitsmaße zum Vergleich zweier Waveletnetz-Repräsentationen. Als Ausgangssituation haben wir zwei Waveletnetze $\mathcal{W}_1 = (\Psi, \boldsymbol{w}_1, m_1)$ und $\mathcal{W}_2 = (\Psi, \boldsymbol{w}_2, m_2)$ gegeben. Zu beachten ist, daß die beiden Netze $\mathcal{W}_1$ und $\mathcal{W}_2$ über **dieselben** Basisfunktionen Ψ verfügen und sich nur durch die Gewichte und ggf. die gespeicherten Mittelwerte unterscheiden. Genau diese Ausgangssituation liegt bei der Gesichtserkennung vor. Dort haben wir zum einem das aus einem Waveletnetz $\mathcal{W} = (\Psi, \boldsymbol{w}, m)$ durch die affine Transformation gemäß einem Testbild h entstandene Waveletnetz $\mathcal{W}' = (\Psi', \boldsymbol{w}, m)$. Zum anderen haben wir das aus der Rekonstruktion des Testbildes mit dem Netz $\mathcal{W}'$ entstandene Waveletnetz $\mathcal{W}'' = (\Psi', \boldsymbol{w}', m')$ mit neu berechnetem Gewichtsvektor $\boldsymbol{w}'$. Grundlage unserer Erkennungsverfahren ist die Bestimmung der Ähnlichkeit dieser beiden Netze $\mathcal{W}'$ und $\mathcal{W}''$.

Die Ausgabe $z_{\mathcal{W}}$ (vgl. Gleichung (2.16)) eines Waveletnetzes $\mathcal{W}$ kann als Grauwertbild $g_{\mathcal{W}}$ aufgefaßt werden, dessen Definitionsbereich durch den Auswertungsbereich

$A_\Psi \subset \mathbb{N} \times \mathbb{N}$ (vgl. Abschnitt 3.1.3) gegeben ist. Der Auswertungsbereich eines Netzes ist mit dem Satz der Basisfunktionen verknüpft, d.h. für alle Netze mit denselben Basisfunktionen Ψ ist der Auswertungsbereich jeweils durch A_Ψ gegeben. Das mit dem Waveletnetz $\mathcal{W}$ identifizierte Grauwertbild $g_\mathcal{W}$ erhält man durch die Abtastung der Ausgabe $z_\mathcal{W}$ innerhalb des Auswertungsbereiches:

$$g_\mathcal{W}(m,n) = z_\mathcal{W}(m,n) \quad \text{für alle } (m,n) \in A_\Psi \subset \mathbb{N} \times \mathbb{N} \,. \tag{3.15}$$

Die Menge aller möglichen Grauwertbilder mit dem Auswertungsbereich A_Ψ als Definitionsbereich bildet den **Signalraum** $\mathcal{S}_{A_\Psi}$. Die aus den Ausgaben von Waveletnetzen erzeugten Grauwertebilder sind Elemente dieses Signalraums. So ist es naheliegend, die Ähnlichkeit zweier zweier Waveletnetze $\mathcal{W}_1$ und $\mathcal{W}_2$ im Signalraum $\mathcal{S}_{A_\Psi}$ zu bestimmen, indem Ähnlichkeitsmaße für die Grauwertbilder $g_{\mathcal{W}_1}$ und $g_{\mathcal{W}_2}$ berechnet werden. Die beiden von uns im Signalraum berechneten Ähnlichkeitsmaße stellen wir in Abschnitt 3.5.1 vor.

Als **Gewichtsraum** bezeichnen wir den Raum, der durch die Gewichtsvektoren zu einem Satz Ψ von Basisfunktionen aufgespannt wird. In Abschnitt 3.5.2 stellen wir Ähnlichkeitsmaße vor, die im Gewichtsraum berechnet werden.

Zusätzlich führen wir ein weiteres Ähnlichkeitsmaß ein, das nicht zwei Waveletnetze miteinander vergleicht, sondern Aussagen über die Ähnlichkeit der Ausgabe eines Waveletnetzes und eines gegebenen Grauwertbildes erlaubt. Dieses Ähnlichkeitsmaße setzen wir ein, um die Güte der Rekonstruktion eines Testbildes durch ein Waveletnetz zu bestimmen. Damit wird die Fähigkeit eines Waveletnetzes gemessen, ein gegebenes Testgesicht zu approximieren. Unter der Annahme, daß durch die angepaßten Wavelets Bilder der Referenzperson gut und Bilder anderer Personen schlecht approximiert werden können, kann die Ähnlichkeit zwischen Testbild und rekonstruiertem Waveletnetz Hinweise auf die Identität des Testgesichts liefern.

3.5.1 Ähnlichkeitsmaße im Signalraum

Die in diesem Abschnitt vorgestellten Maße $\mathsf{S}_{||\cdot||_2}$ und S_C sind im Signalraum motiviert, da die mit Waveletnetzen identifizierten Grauwertbilder verglichen werden.

Die aus den Ausgaben $z_{\mathcal{W}_1}$ und $z_{\mathcal{W}_2}$ der Waveletnetze $\mathcal{W}_1 = (\Psi, \boldsymbol{w}_1, m_1)$ und $\mathcal{W}_2 = (\Psi, \boldsymbol{w}_2, m_2)$ gewonnenen Grauwertbilder $g_{\mathcal{W}_1}$ und $g_{\mathcal{W}_2}$ sind Vektoren des Signalraums $\mathcal{S}_{A_\Psi}$, den wir oben eingeführt haben. Als Abstandsmaß zweier Vektoren g_1 und g_2 im Signalraum kann deren euklidischer Abstand

$$||g_1 - g_2||_2 = \sqrt{\sum_{(m,n)\in A_\Psi} \left(g_1(m,n) - g_{g_2}(m,n)\right)^2} \tag{3.16}$$

dienen. Der euklidische Abstand (3.16) ist äquivalent zum quadratischen Fehler zwischen den beiden Grauwertbildern.

Unser erstes Ähnlichkeitsmaß $\mathsf{S}_{\|\cdot\|_2}$ ist der Kehrwert des euklidischen Abstands. Damit erhalten wir das folgende Ähnlichkeitsmaß

$$\mathsf{S}_{\|\cdot\|_2}(\mathcal{W}_1, \mathcal{W}_2) = \frac{1}{\|g_{\mathcal{W}_1}, g_{\mathcal{W}_2}\|_2} \ . \tag{3.17}$$

Um unterschiedliche Gleichanteile m_1 und m_2 zu kompensieren, setzen wir beide Werte vor der Berechnung des euklidischen Abstands auf Null. Der Wert für $\mathsf{S}_{\|\cdot\|_2}$ wird umso größer, je geringer der quadratische Fehler und damit der euklidische Abstand zwischen $g_{\mathcal{W}_1}$ und $g_{\mathcal{W}_2}$ wird.

Betrachten wir den euklidischen Abstand (3.16) näher. Mit

$$\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)^T = \boldsymbol{w}_1 - \boldsymbol{w}_2$$

gilt

$$\begin{aligned}
\|g_1 - g_2\|_2^2 &= \sum_{(m,n)\in A_\Psi} (z_{\mathcal{W}_1}(m,n) - z_{\mathcal{W}_2}(m,n))^2 \\
&= \sum_{(m,n)\in A_\Psi} \left(\sum_{i=1}^{N} \delta_i \psi_{\boldsymbol{a}_i} \right)^2 \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_i \delta_j \sum_{(m,n)\in A_\Psi} \psi_{\boldsymbol{a}_i}(m,n) \psi_{\boldsymbol{a}_j}(m,n) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_i \delta_j < \psi_{\boldsymbol{a}_i}, \psi_{\boldsymbol{a}_j} >_{A_\Psi} \\
&= \boldsymbol{\delta}^T \boldsymbol{\Psi} \boldsymbol{\delta} \ ,
\end{aligned} \tag{3.18}$$

wobei $\boldsymbol{\Psi}$ die Gramsche Matrix aus Gleichung (2.44) ist. Wie wir in Abschnitt 3.4 beschrieben haben, bleibt die Gramsche Matrix unter affinen Transformationen bis auf einen Faktor invariant, womit ihre Neuberechnung entfällt. Es ist also gemäß Gleichung (3.4) lediglich die Division der Matrixelemente durch das Produkt der Dilatationsparameter notwendig. Mit den Umformungen (3.18) wird gezeigt, daß sich $\mathsf{S}_{\|\cdot\|_2}$ alternativ zur Berechnung im Signalraum effizient im Gewichtsraum unter Zuhilfenahme der Gramschen Matrix $\boldsymbol{\Psi}$ berechnen läßt:

$$\mathsf{S}_{\|\cdot\|_2}(\mathcal{W}_1, \mathcal{W}_2) = \frac{1}{\sqrt{\boldsymbol{\delta}^T \boldsymbol{\Psi} \boldsymbol{\delta}}} \ . \tag{3.19}$$

Bei Template-Matching-Verfahren wird häufig die **normalisierte Kreuzkorrelation** verwendet, um die Ähnlichkeit zwischen einem Template und einem Bildausschnitt zu bestimmen [2]. Die normalisierte Kreuzkorrelation zweier Grauwertbilder g_1 und g_2 ist gegeben durch

$$\mathsf{C}(g_1, g_2) = \frac{\overline{g_1 \cdot g_2} - \overline{g}_1\overline{g}_2}{\sqrt{\mathrm{Var}(g_1)\,\mathrm{Var}(g_2)}} , \tag{3.20}$$

wobei $g_1 \cdot g_2$ für das pixelweise Produkt, $\bar{\cdot}$ für die Mittelwertbildung und $\mathrm{Var}(\cdot)$ für die Varianz stehen. Die Normalisierung in (3.20) transformiert die Energieverteilungen von g_1 und g_2 so, daß sie mittelwertfrei sind und ihre Varianzen den Wert 1 annehmen. Zusätzlich wird erreicht, daß

$$-1 \leq \mathsf{C}(g_1, g_2) \leq 1$$

gilt, was später in Abschnitt 3.7 hilfreich bei der Bestimmung von Schwellwerten ist.

Als zweites Ähnlichkeitsmaß im Signalraum neben $\mathsf{S}_{\|\cdot\|_2}$ schlagen wir die normalisierte Kreuzkorrelation der Grauwertbilder $g_{\mathcal{W}_1}$ und $g_{\mathcal{W}_2}$ vor:

$$\mathsf{S}_{\mathsf{C}}(\mathcal{W}_1, \mathcal{W}_2) = \mathsf{C}(g_{\mathcal{W}_1}, g_{\mathcal{W}_2}) . \tag{3.21}$$

Auch dieses Ähnlichkeitsmaß läßt sich ausschließlich durch die Gewichtsvektoren sowie die Gramsche Matrix ausdrücken. Man erhält

$$\mathsf{S}_{\mathsf{C}}(\mathcal{W}_1, \mathcal{W}_2) = \frac{\boldsymbol{w}_1^T \boldsymbol{\Psi} \boldsymbol{w}_2}{\sqrt{\boldsymbol{w}_1^T \boldsymbol{\Psi} \boldsymbol{w}_1}\sqrt{\boldsymbol{w}_2^T \boldsymbol{\Psi} \boldsymbol{w}_2}} . \tag{3.22}$$

3.5.2 Ähnlichkeitsmaße im Gewichtsraum

Die beiden betrachteten Waveletnetze $\mathcal{W}_1 = (\Psi, \boldsymbol{w}_1, m_1)$ und $\mathcal{W}_2 = (\Psi, \boldsymbol{w}_2, m_2)$ unterscheiden sich – abgesehen von den gespeicherten Gleichanteilen – nur durch die beiden Gewichtsvektoren $\boldsymbol{w}_1$ und $\boldsymbol{w}_2$. Also liegt es auf der Hand, Ähnlichkeitsmaße lediglich basierend auf den Gewichtsvektoren zu formulieren. Der Vektorraum, der durch alle möglichen Gewichtsvektoren zu Ψ gebildet wird, bezeichnen wir als **Gewichtsraum**. Für die beiden in diesem Abschnitt vorgestellten Maße $\mathsf{S}_{\boldsymbol{w}}$ und $\mathsf{S}_{\hat{\boldsymbol{w}}}$ werden bloß die Gewichtsvektoren $\boldsymbol{w}_1$ und $\boldsymbol{w}_2$ herangezogen, die Maße werden folglich im Gewichtsraum bestimmt.

Ein Möglichkeit zur Bestimmung der Ähnlichkeit zweier Vektoren eines Vektorraums ist der Kosinus des eingeschlossenen Winkels. Als erstes Maß im Gewichtsraum verwenden wir daher den Kosinus des durch die Gewichtsvektoren $\boldsymbol{w}_1$ und $\boldsymbol{w}_2$ eingeschlossenen Winkels:

$$\mathsf{S}_{\boldsymbol{w}}(\mathcal{W}_1, \mathcal{W}_2) = \frac{\boldsymbol{w}_1^T \boldsymbol{w}_2}{\|\boldsymbol{w}_1\|_2 \|\boldsymbol{w}_2\|_2} . \tag{3.23}$$

Es gilt

$$-1 \leq \mathsf{S}_{\boldsymbol{w}}(\mathcal{W}_1, \mathcal{W}_2) \leq 1 \;.$$

Bei diesem Maß tragen alle Basisfunktionen aus Ψ in gleicher Weise zum Ähnlichkeitsmaß $\mathsf{S}_{\boldsymbol{w}}$ bei, ungeachtet ihrer Skala und damit ihres Einflusses auf die Ausgabe des Waveletnetzes. Es erscheint daher offensichtlich, feinskalige Wavelets mit sehr lokaler Wirkung weniger stark zu berücksichtigen als grobskalige Wavelets mit weiterreichendem Einfluß. Dazu leiten wir aus einem Gewichtsvektor $\boldsymbol{w}$ einen Vektor $\hat{\boldsymbol{w}} = (\hat{w}_1, \ldots, \hat{w}_N)^T$ ab, indem jede Komponente von $\hat{\boldsymbol{w}}$ mit der L^2-Norm der entsprechenden Basisfunktion gewichtet wird:

$$\hat{w}_i = \|\psi_{\boldsymbol{a}_i}\|_2 w_i = \sqrt{<\psi_{\boldsymbol{a}_i}, \psi_{\boldsymbol{a}_i}>}\; w_i, \qquad 1 \leq i \leq N \;. \tag{3.24}$$

Damit erhalten wir als zweites Ähnlichkeitsmaß

$$\mathsf{S}_{\hat{\boldsymbol{w}}}(\mathcal{W}_1, \mathcal{W}_2) = \frac{\hat{\boldsymbol{w}}_1^T \hat{\boldsymbol{w}}_2}{\|\hat{\boldsymbol{w}}_1\|_2 \|\hat{\boldsymbol{w}}_2\|_2} \;. \tag{3.25}$$

Es sei darauf hingewiesen, daß $\hat{\boldsymbol{w}}$ der zu $\boldsymbol{w}$ äquivalente Gewichtsvektor bezüglich der gemäß der L^2-normierten Basisfunktionen $\hat{\Psi} = \left\{ \frac{\psi_{\boldsymbol{a}_1}}{\|\psi_{\boldsymbol{a}_1}\|_2}, \ldots, \frac{\psi_{\boldsymbol{a}_N}}{\|\psi_{\boldsymbol{a}_N}\|_2} \right\}$ ist, also

$$\sum_{i=1}^{N} w_i \psi_{\boldsymbol{a}_i} = \sum_{i=1}^{N} \hat{w}_i \frac{\psi_{\boldsymbol{a}_i}}{\|\psi_{\boldsymbol{a}_i}\|_2} \;.$$

Nehmen wir nun einmal an, daß wir orthogonale Basisfunktionen vorliegen hätten, also eine Diagonalmatrix als Gramsche Matrix. Im letzten Abschnitt hatten wir gezeigt, daß für das Maß S_{C} gilt

$$\mathsf{S}_{\mathsf{C}}(\mathcal{W}_1, \mathcal{W}_2) = \frac{\boldsymbol{w}_1^T \boldsymbol{\Psi} \boldsymbol{w}_2}{\sqrt{\boldsymbol{w}_1^T \boldsymbol{\Psi} \boldsymbol{w}_1} \sqrt{\boldsymbol{w}_2^T \boldsymbol{\Psi} \boldsymbol{w}_2}} \;. \tag{3.26}$$

Für den Zähler von (3.26) gilt unter der Annahme orthogonaler Basisfunktionen

$$\begin{aligned}
\boldsymbol{w}_1^T \boldsymbol{\Psi} \boldsymbol{w}_2 &= \sum_{i=1}^{N} w_{1_i} <\psi_{\boldsymbol{a}_i}, \psi_{\boldsymbol{a}_i}> w_{2_i} \\
&= \sum_{i=1}^{N} w_{1_i} \|\psi_{\boldsymbol{a}_i}\|_2^2 w_{2_i} \\
&= \sum_{i=1}^{N} \hat{w}_{1_i} \hat{w}_{2_i} \\
&= \hat{\boldsymbol{w}}_1^T \hat{\boldsymbol{w}}_2 \;.
\end{aligned}$$

Analog gilt für den Nenner

$$\begin{aligned}\sqrt{\boldsymbol{w}^T \boldsymbol{\Psi} \boldsymbol{w}} &= \sqrt{\hat{\boldsymbol{w}}^T \hat{\boldsymbol{w}}} \\ &= \|\hat{\boldsymbol{w}}\|_2 \,.\end{aligned}$$

Mit diesen Umformungen erhalten wir aus (3.26)

$$\mathsf{S}_{\mathsf{C}}(\mathcal{W}_1, \mathcal{W}_2) = \frac{\hat{\boldsymbol{w}}_1^T \hat{\boldsymbol{w}}_2}{\|\hat{\boldsymbol{w}}_1\|_2 \|\hat{\boldsymbol{w}}_2\|_2} = \mathsf{S}_{\hat{\boldsymbol{w}}}(\mathcal{W}_1, \mathcal{W}_2) \;, \tag{3.27}$$

wenn die Basisfunktionen orthogonal sind. Also liefern in diesem Fall die normalisierte Kreuzkorrelation (3.26) sowie der Kosinus des durch die Vektoren $\hat{\boldsymbol{w}}_1$ und $\hat{\boldsymbol{w}}_2$ eingeschlossenen Winkels dasselbe Ähnlichkeitsmaß. Bei nichtorthogonalen Basisfunktionen trägt die normalisierte Kreuzkorrelation (3.26) der Nichtorthogonalität Rechnung, das Maß $\mathsf{S}_{\hat{\boldsymbol{w}}}$ hingegen nicht. Wir vermuten, daß dieser Umstand die Begründung dafür ist, daß durch $\mathsf{S}_{\hat{\boldsymbol{w}}}$ schlechtere Erkennungsraten erzielt werden, als durch S_{C}, wie wir im Kapitel 4 zeigen werden.

Mit ähnlichen Rechnungen läßt sich zeigen, daß im Fall orthogonaler Basisfunktionen das Maß $\mathsf{S}_{\|\cdot\|_2}$, also der euklidische Abstand der Ausgaben der Waveletnetze, identisch ist mit dem euklidischen Abstand der Vektoren $\hat{\boldsymbol{w}}_1$ und $\hat{\boldsymbol{w}}_2$, also

$$\mathsf{S}_{\|\cdot\|_2}(\mathcal{W}_1, \mathcal{W}_2) = \|\hat{\boldsymbol{w}}_1 - \hat{\boldsymbol{w}}_2\|_2 \,. \tag{3.28}$$

3.6 Erkennung von Gesichtern

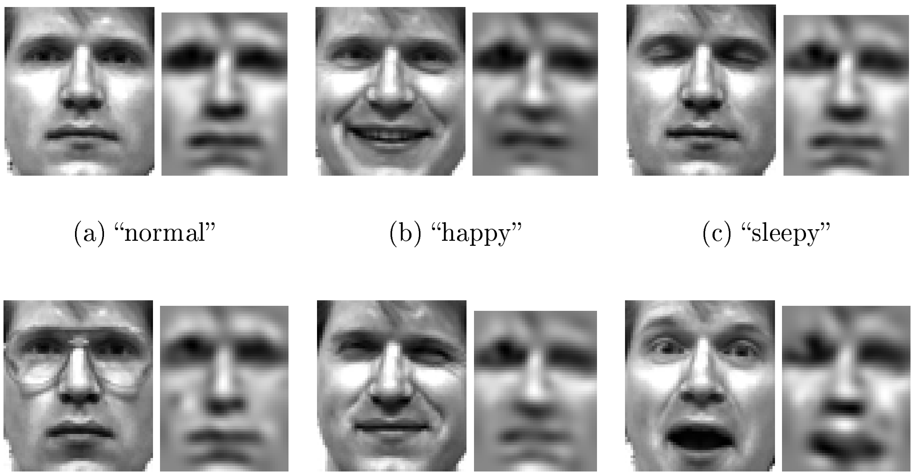
Die Abbildung 1.1 auf Seite 5 zeigt in Form eines Ablaufdiagrammes den Aufbau des Systems zur Erkennung von Gesichtern. In diesem Abschnitt beschäftigen wir uns mit dem letzten Schritt, der eigentlichen Erkennung. Die Ausgangssituation hierfür ist wie folgt: Wir verfügen über die in Abschnitt 3.1 beschriebenen Waveletnetz-Repräsentationen $\mathcal{W}_1$ bis $\mathcal{W}_M$ der M Referenzbilder der Referenzdatenbank. Als Eingabe ist ein Testbild h vorgegeben. Zunächst werden die M Waveletnetze $\mathcal{W}_i = (\Psi_i, \boldsymbol{w}_i, m_i)$ mit dem im Abschnitt 3.2 eingeführten Matchingverfahren so affin transformiert, daß sie in Position, Größe und Orientierung mit dem Gesicht im Testbild übereinstimmen. Die affine Transformation (Reparametrisierung) liefert die Waveletnetze $\mathcal{W}'_i = (\Psi'_i, \boldsymbol{w}_i, m_i), 1 \leq i \leq M$. Mit diesen reparametrisierten Waveletnetzen wird jeweils das Testbild h gemäß Abschnitt 3.4 rekonstruiert. Durch die Rekonstruktion entstehen die M Repräsentationen $\mathcal{W}''_1$ bis $\mathcal{W}''_M$ des Testbildes, mit $\mathcal{W}''_i = (\Psi'_i, \boldsymbol{w}'_i, m'_i)$. Mit den im Abschnitt 3.5 eingeführten Ähnlichkeitsmaßen bestimmen wir jeweils die Ähnlichkeit von $\mathcal{W}'_i$ und $\mathcal{W}''_i$ als Grundlage für die Zuordnung einer Identität.

Zunächst wollen wir in diesem Abschnitt an Beispielen motivieren, daß die Ähnlichkeitsmaße geeignet sind, die Basis für die Erkennung zu bilden. Die Beispiele sind aus der Yale-Datenbank [30] mit $M = 15$ verschiedenen Gesichtern entstanden. In der Yale-Datenbank sind Aufnahmen von als “subject01” bis ”subject15” bezeichneten Personen mit jeweils verschiedenen Gesichtsausdrücken, z.B. “normal”, “happy”, “sleepy”, “wink” und “surprised” enthalten. Im Kapitel 4 gehen wir näher auf die für diese Diplomarbeit eingesetzten Gesichtsdatenbanken ein. Die Aufnahmen “normal” mit jeweils neutralen Gesichtsausdrücken dienen uns als Referenzbilder, aus denen die Waveletnetz-Repräsentationen gewonnen werden.

Wir verwenden Gesichtsrepräsentationen mit 42 bzw. 64 Gaborwavelets. Diese spärliche Repräsentation mit einer geringen Anzahl von Wavelets soll sicherstellen, daß durch das Anpassen der Wavelets des Netzes $\mathcal{W}_i$ an das i.te Referenzgesicht ein an die Gesichtsmerkmale angepaßter, spezifischer Satz von Basisfunktionen gefunden wird. Mit Hilfe dieser spezifischen Wavelets erwarten wir, daß sich Gesichtsbilder der Person i gut rekonstruieren lassen. Von Bildern anderer Personen $j \neq i$ nehmen wir an, daß sie sich schlechter rekonstruieren lassen, da ein Großteil der Wavelets nicht mehr auf den Gesichtsmerkmalen liegt. Eine zu große Zahl von Wavelets würde es ermöglichen, durch das Rekonstruktionsverfahren beliebige Gesichtsbilder zu approximieren, mit zunehmender Waveletanzahl nähert man sich einer kompletten Wavelettransformation, die Repräsentation beliebiger Bilder wäre möglich.

Die Verwendung verhältnismäßig grobskaliger Wavelets soll die Robustheit gegenüber verschiedenen Gesichtsausdrücken sicherstellen. Dabei gehen wir davon aus, daß die Variationen bei verschiedenen Gesichtsausdrücken im Wesentlichen auf einer feinen Skala beschrieben werden, die groben Gesichtsmerkmale (Augen, Nase) bleiben verhältnismäßig invariant.

Ein Beispiel hierfür zeigt Abbildung 3.21: Es sind verschiedene Gesichtsausdrücke der Person “subject01” der Yale-Datenbank abgebildet. Abbildung 3.21(a) zeigt das Referenzgesicht der Person “subject01” sowie die Waveletnetz-Repräsentation mit 42 Wavelets. Die Abbildungen 3.21(b)-(f) zeigen vom Referenzgesicht abweichende Gesichtsausdrücke sowie die Rekonstruktionen mit dem individuellen Waveletnetz der Person “subject01”. Hervorzuheben ist das Bild “glasses”. Der dünne Rand der Brille ist für die grobskaligen Wavelets “nicht sichtbar” und hat keinen nennenswerten Einfluß auf die Rekonstruktion. Auch die Rekonstruktionen der Bilder “sleepy” und “wink” weichen im optischen Eindruck nur gering von der Repräsentation des Referenznetzes ab. Beim Gesichtsausdruck “happy” ist gut erkennbar, daß der gegenüber dem neutralen Gesichtsausdruck stark veränderte Mund verhältnismäßig schlecht rekonstruiert werden kann, doch die Augenpartie und die Nase bleiben im Bild “happy” weitgehend unverändert, diese Gesichtsregionen können gut rekonstruiert werden. Der **lokale Charakter** der Waveletnetz-Repräsentation wird bei diesen Bildern deutlich. Lokale Abweichungen im Gesicht wirken sich lediglich auf die in den entsprechenden Bereichen positionierten Wavelets aus, die übrigen Wavelets vermögen weiterhin die nur unwesentlich veränderten Regionen zu rekonstruieren. Der Gesichtsausdruck “surprised” zeigt die



(a) “normal” (b) “happy” (c) “sleepy”

(d) “glasses” (e) “wink” (f) “surprised”

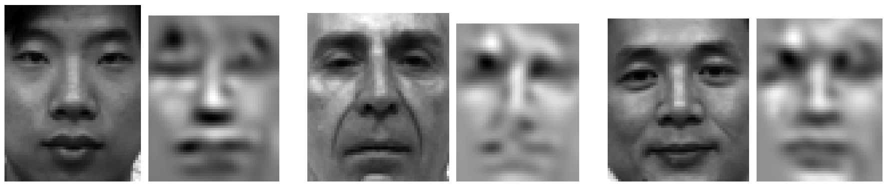
Abbildung 3.21: (a) Referenzgesicht der Person “subject01” und gespeicherte Waveletnetz-Repräsentation. (b) - (f) Unterschiedliche Gesichtsausdrücke und Ergebnisse der Rekonstruktion mit dem transformierten Waveletnetz der Person “subject01”.

Grenzen der Gesichtsrepräsentation auf. Die Änderung des überraschten Gesichts gegenüber dem neutralen Gesichtsausdruck sind eher globaler Natur, dementsprechend läßt die Qualität der Rekonstruktion sichtbar nach.

In Abbildung 3.22 sind die übrigen Referenzgesichter der Yale-Datenbank zusammen mit den Rekonstruktionen durch das Waveletnetz von “subject01” abgebildet. Optisch scheinen sich die Bilder und die Rekonstruktionen jeweils deutlich unähnlicher zu sein, als die Bild-/Rekonstruktionspaare in Abbildung 3.21. Wir erwarten, daß sich dieser optische Eindruck in den vorherigen Abschnitt eingeführten Ähnlichkeitsmaße widerspiegelt. Die Abbildungen 3.23 und 3.24 zeigen die entsprechenden Diagramme mit den Ähnlichkeitsmaßen. Mit Ausnahme des Gesichtsausdrucks “surprised” ergeben sich für die Gesichter von “subject01” höhere Werte als für die Bilder der Personen “subject02” bis “subject15”. Für das Testbild “subject01.surprised” ergeben sich im Signalraum Werte für die Ähnlichkeitsmaße $\mathsf{S}_{\|\cdot\|_2}$ und S_{C}, die deutlich unter denen der übrigen Testbilder von “subject01” liegen und nur noch knapp die Werten für die anderen Personen überschreiten. Dies entspricht in Abbildung 3.21 gewonnenen optischen Eindruck. Bei den beiden Maßen $\mathsf{S}_{\boldsymbol{w}}$ und $\mathsf{S}_{\hat{\boldsymbol{w}}}$ im Gewichtsraum ergeben sich für “subject01.surprised” sogar geringere Ähnlichkeitswerte als für einige der übrigen Personen. Die Werte der anderen Bilder von “subject01” liegen jedoch deutlich höher.

Damit scheinen die Ähnlichkeitsmaße insgesamt dem aus den Abbildungen 3.21 und

(a) subject01 (b) subject02 (c) subject03



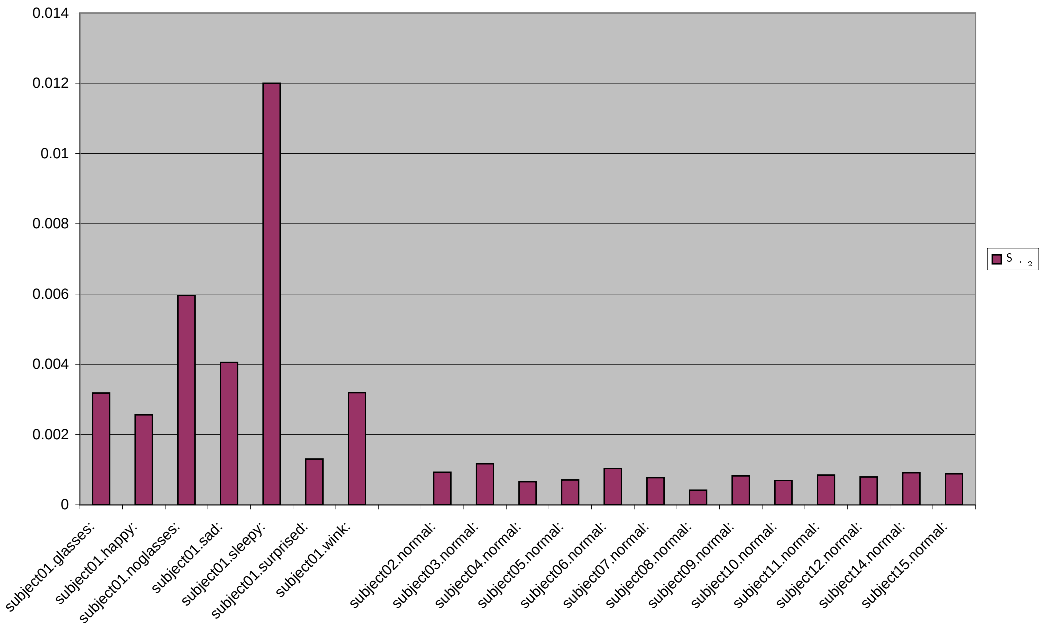
(d) subject04 (e) subject05 (f) subject06

(g) subject07 (h) subject08 (i) subject09

(j) subject10 (k) subject11 (l) subject12

(m) subject13 (n) subject14 (o) subject15

Abbildung 3.22: Gesichter der Personen “subject01” bis “subject15” als Testgesichter mit den Ergebnissen der Rekonstruktion durch das Waveletnetz von “subject01”. Die Rekonstruktion von “subject01” ist deutlich besser als die Rekonstruktionen der übrigen Gesichter.



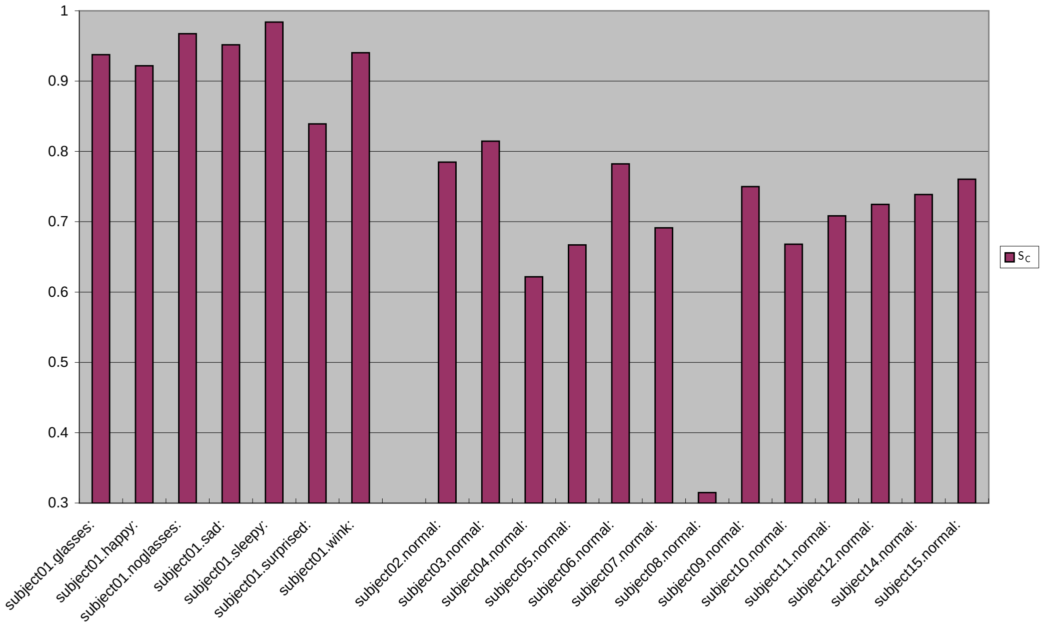


Abbildung 3.23: Ähnlichkeitsmaße $\mathsf{S}_{\|\cdot\|_2}$ und S_C im Signalraum. Mit der Waveletnetz-Repräsentation von “subject01” wurden die an der horizontalen Achse abgetragenen Testbilder rekonstruiert (vgl. Abbildungen 3.21 und 3.22) und die Werte der Ähnlichkeitsmaße im Signalraum bestimmt. Die Rekonstruktionen der Testbilder von ”subject01“ erreichen höhere Werte als die übrigen Testbilder, wobei ”subject01.surprised” nur knapp über den Werten der übrigen Personen liegt.

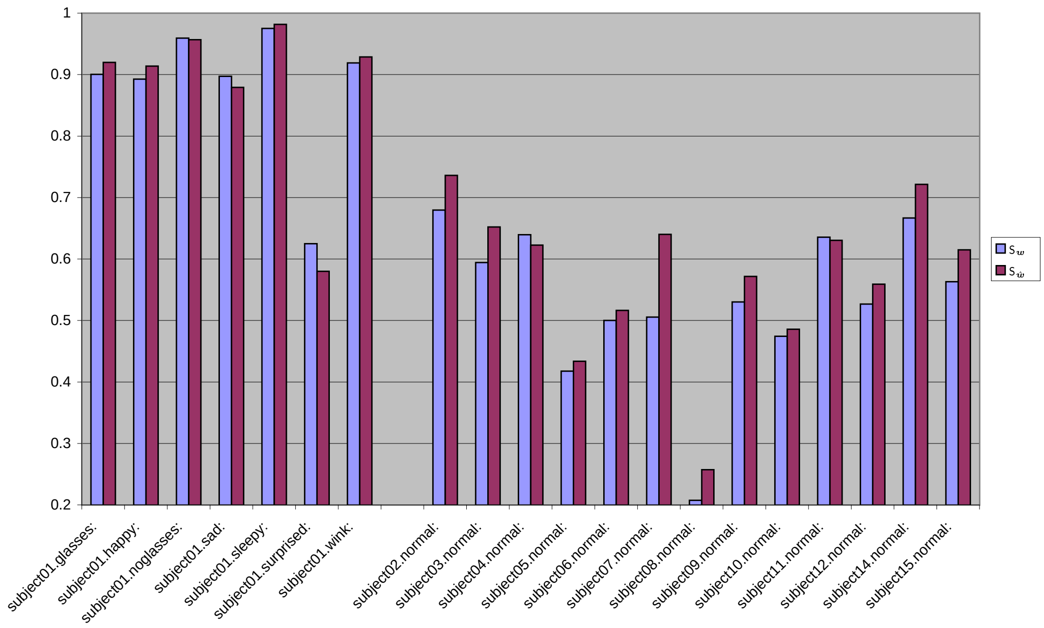


Abbildung 3.24: Ähnlichkeitsmaße im Gewichtsraum. Mit der Waveletnetz-Repräsentation von “subject01” wurden die an der horizontalen Achse abgetragenen Testbilder rekonstruiert (vgl. Abbildungen 3.21 und 3.22) und die Werte der Ähnlichkeitsmaße $\mathsf{S}_{\boldsymbol{w}}$ und $\mathsf{S}_{\hat{\boldsymbol{w}}}$ im Gewichtsraum bestimmt. Bis auf “subject01.surprised” erreichen die Rekonstruktionen von “subject01” höhere Werte als die Rekonstruktionen der übrigen Personen.

3.22 gewonnenen optischen Eindruck zu entsprechen und bestätigen unsere Vermutung, daß die Gesichtsrepräsentation von “subject01” im wesentlichen gut geeignet ist, Testbilder von ”subject01” zu rekonstruieren. Die Rekonstruktion der übrigen Personen ist erwartungsgemäß nur eingeschränkt möglich.

Nun wollen wir die Rekonstruktionen eines Testgesichts h durch sämtliche affin transformierten Waveletnetze $\mathcal{W}_i$ miteinander vergleichen. In Abbildung 3.25 zeigen wir die 15 Rekonstruktionen für das Testgesicht “subject01.happy”. Die Waveletnetz-

Abbildung 3.25: Rekonstruktion des Testgesichts “subject01.happy” mit den 15 Waveletnetz-Repräsentationen der Yale-Datenbank. Mit der Repräsentation von “subject01” ist die optisch beste Rekonstruktion möglich (Bild links oben).

Repräsentation von “subject01” liefert erwartungsgemäß die beste Rekonstruktion des Testbildes “subject01.happy”. Die Werte der Abstandsmaße für dieses Testbild sowie für die Testbilder “sad”, “sleepy”, “surprised” und “wink” sind den Abbildungen 3.26 und 3.27 zu entnehmen. Anhand der Diagramme wird deutlich, daß das Netz von “subject01” höhere Ähnlichkeitsmaße $\mathsf{S}_{\|\cdot\|_2}$ und S_C im Signalraum erzielt als die Netze der übrigen Personen. Allerdings liegen die Werte für das Bild “surprised” dicht bei den Werten einiger anderer Testbilder. Für die Ähnlichkeitsmaße $\mathsf{S}_{\boldsymbol{w}}$ und $\mathsf{S}_{\hat{\boldsymbol{w}}}$ im Gewichtsraum werden sogar in einigen Fällen für Bilder anderer Personen höhere Werte erreicht, als für das Bild “surprised” der Person “subject01”.

Für die Gesichtserkennung wollen wir uns in dieser Diplomarbeit jeweils auf ein einzelnes Abstandsmaß $\mathsf{S} \in \{\mathsf{S}_{\|\cdot\|_2}, \mathsf{S}_\mathsf{C}, \mathsf{S}_{\boldsymbol{w}}, \mathsf{S}_{\hat{\boldsymbol{w}}}\}$ beschränken. Die Diagramme der Abbildungen 3.26 und 3.27 legen folgende Strategie nahe: Dem Testbild wird eine Identität I gemäß

$$I = \arg \min_{1 \le i \le M} \mathsf{S}(\mathcal{W}_i', \mathcal{W}_i'') \tag{3.29}$$

zugeordnet.

Eine genaue Auswertung der Erkennungsraten in Abhängigkeit des gewählten Ähnlichkeitsmaßes S und der Anzahl der Wavelets pro Waveletnetz erfolgt in Kapitel 4.

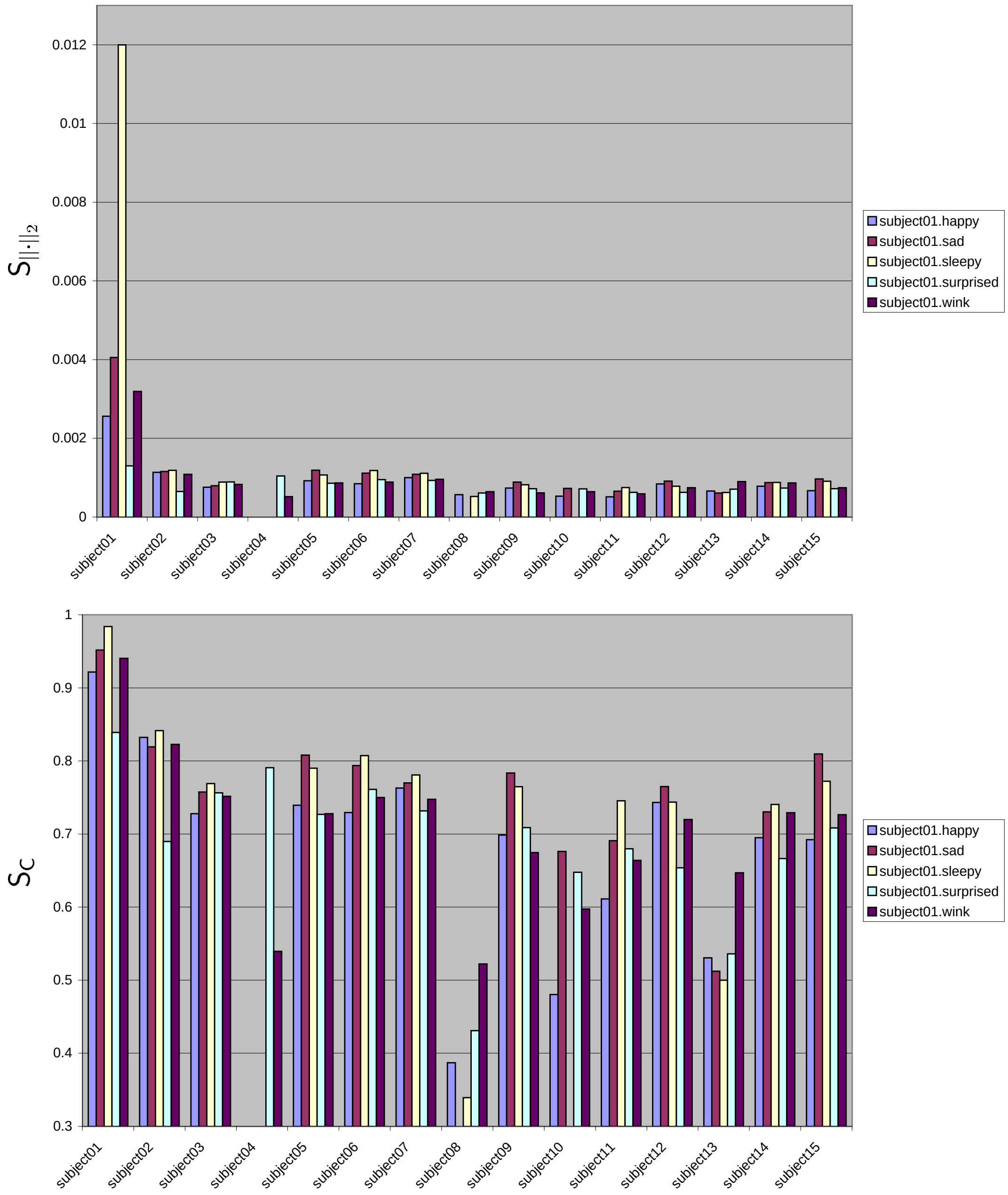


Abbildung 3.26: Ähnlichkeitsmaße im Signalraum für die Testbilder “happy”, “sleepy”, “sad”, “surprised” und “wink” der Person “subject01”. Es sind die Werte für die 15 Waveletnetze der Referenzpersonen “subject01” bis “subject15“ angegeben. Fehlende Werte bedeuten, daß das Matchingverfahren nicht sinnvoll konvergierte. Wird ein Maximum für “subject01” eingenommen, ist eine korrekte Identifizierung möglich. Dies ist für “subject01.surprised” nur sehr knapp der Fall. Für die übrigen Testbilder liegen die Ähnlichkeitsmaße für die Rekonstruktionen durch das Netz von “subject01” deutlich über denen der übrigen Waveletnetze.

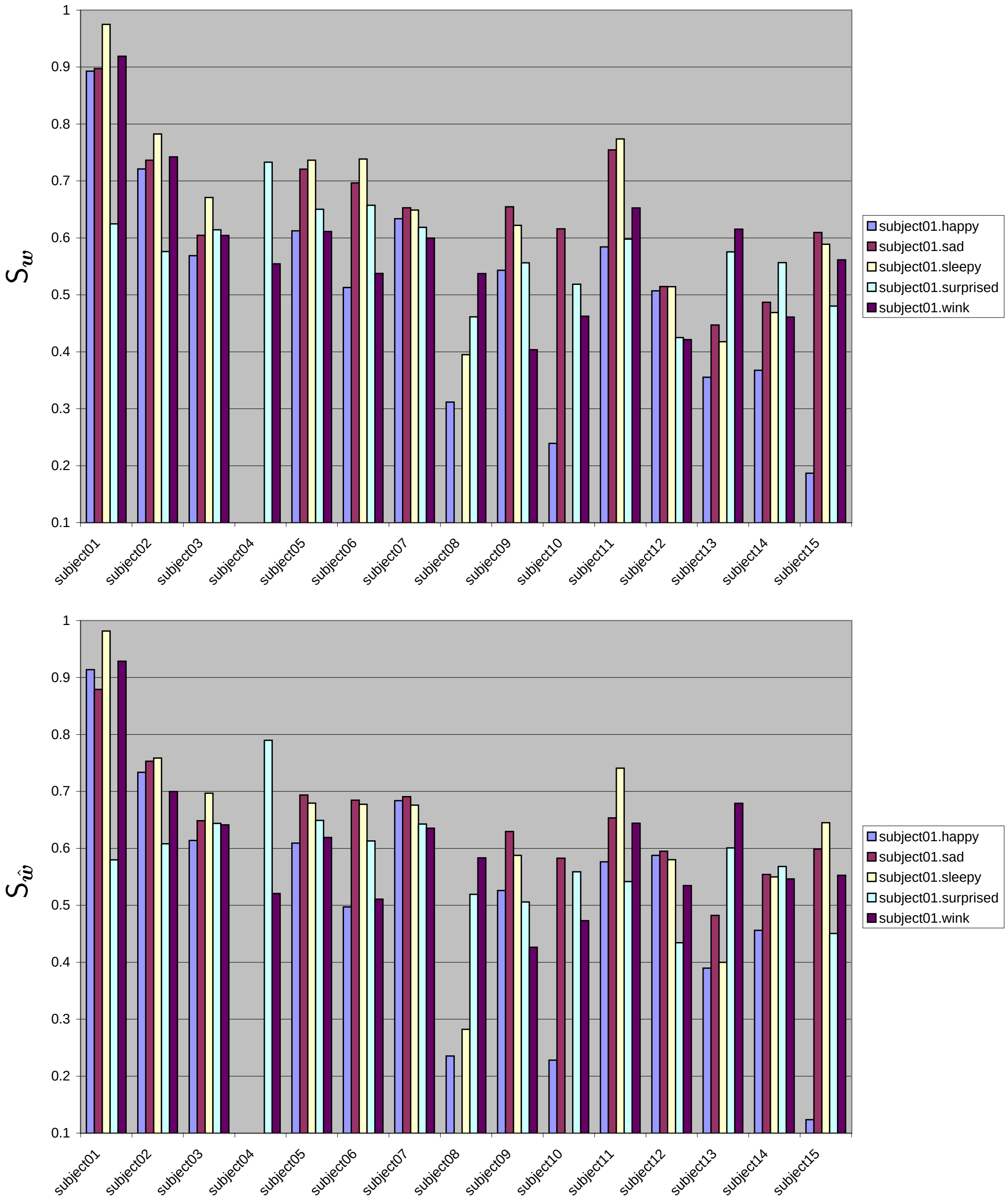


Abbildung 3.27: Ähnlichkeitsmaße im Gewichtsraum für die Testbilder "happy", "sleepy", "sad", "surprised" und "wink" der Person "subject01". Es sind die Werte für die 15 Waveletnetze der Referenzpersonen "subject01" bis "subject15" angegeben. Mit Ausnahme des Testbildes "surprised" erreichen die Rekonstruktionen mit dem Netz von "subject01" jeweils die höchsten Werte, die Testbilder würden korrekt erkannt. Das Testbild "surprised" würde fälschlicherweise als "subject04" erkannt.

3.7 Verifikation der Identität

Bei dem im letzten Abschnitt erläuterten Verfahren zur Gesichtserkennung wird für **jede** Referenzperson aus dem gespeicherten individuellen Waveletnetz eine eigene Repräsentation des zu erkennenden Testgesichts erzeugt. Sind viele Personen in der Referenzdatenbank gespeichert, wird diese Vorgehensweise inpraktikabel; die zur Erkennung benötigte Rechenzeit wächst mit der Anzahl der zu erzeugenden und zu überprüfenden Repräsentationen des Testgesichts.

Wir wollen in diesem Abschnitt mit der Situation beschäftigen, daß zum gegebenen Testbild h die Hypothese geprüft werden soll, ob es sich um eine bestimmte Person i handelt, die durch ein Waveletnetz W_i repräsentiert wird. Diese Situation tritt beispielsweise bei Zugangskontrollsystemen auf, bei denen autorisierte Personen eine Chipkarte besitzen, auf der ihr individuelles Waveletnetz gespeichert ist. Zur Überprüfung wird mit einer Kamera eine Aufnahme der Person als Testbild erzeugt. Das gespeicherte Waveletnetz $\mathcal{W}$ wird mit dem Matchingverfahren aus Abschnitt 3.2 entsprechend dem Gesicht im Testbild affin transformiert. Mit dem transformierten Waveletnetz $\mathcal{W}'$ wird das Testgesicht rekonstruiert (Abschnitt 3.4), wodurch das Waveletnetz $\mathcal{W}''$ entsteht. Mit einem der Ähnlichkeitsmaße aus Abschnitt 3.5 wird die Ähnlichkeit $\mathsf{S}(\mathcal{W}', \mathcal{W}'')$ von $\mathcal{W}'$ und $\mathcal{W}''$ bestimmt. Überschreitet die Ähnlichkeit einen bestimmten Schwellwert, dann wird davon ausgegangen, daß es sich bei dem Testbild um eine Aufnahme der auf der Chipkarte gespeicherten Person handelt. Die Identifikation soll also im Gegensatz zu Abschnitt 3.6 lediglich durch den Vergleich der gespeicherten Repräsentation mit **einer** Repräsentation des Testgesichts erfolgen.

Für diese Aufgabenstellung wollen wir ein weiteres Ähnlichkeitsmaß S_h überprüfen, das das Testgesicht direkt mit seiner Rekonstruktion vergleicht. Der Vergleich erfolgt über die normalisierte Kreuzkorrelation zwischen Testbild und dem aus der Ausgabe von $\mathcal{W}''$ gewonnenen Grauwertbild. Die normalisierte Kreuzkorrelation ist gegeben durch

$$\mathsf{S}_h(\mathcal{W}'', h) = \mathsf{C}(g_{\mathcal{W}''}, h) = \frac{\overline{g_{\mathcal{W}''} \cdot h} - \overline{g}_{\mathcal{W}''}\overline{h}}{\sqrt{\mathrm{Var}(g_{\mathcal{W}''})\,\mathrm{Var}(h)}} \,. \tag{3.30}$$

Die Wahl geeigneter Schwellwerte wollen wir an dem Beispiel aus dem letzten Abschnitt illustrieren. Dort sind in den Abbildungen 3.21 und 3.22 die Rekonstruktionsergebnisse des Waveletnetzes der Person "subject01" für Testbilder aller Personen der Yale-Datenbank abgebildet. Die Abbildungen 3.23 und 3.24 zeigen Diagramme mit den entsprechenden Werten der Ähnlichkeitsmaße aus Abschnitt 3.5. Für Sicherheitsanwendungen, wie z.B. Zugangskontrollen, müssen möglichst hohe Schwellwerte eingesetzt werden, um das Risiko zu minimieren, daß eine nicht autorisierte Person fälschlicherweise als die auf der Chipkarte gespeicherte Person identifiziert wird. Dabei muß in Kauf genommen werden, daß Aufnahmen der autorisierten Person nicht erkannt werden, wenn der Gesichtsausdruck deutlich vom Referenzbild abweicht, das durch das gespeicherte Waveletnetz repräsentiert wird. In den oben zitierten Beispielen ist dies bei dem Testbild "subject01.surprised" der Fall, dessen Ähnlichkeitsmaße nur knapp über und zum Teil sogar unter den Werten der übrigen Personen liegen. Die Diagramme in Abbildung 3.23 legen für das Maß $\mathsf{S}_{\|\cdot\|_2}$ beispielsweise eine Schwelle von 0.003 und für S_{C} einen Wert von 0.9 nahe. Für die Maße $\mathsf{S}_{\boldsymbol{w}}$ und $\mathsf{S}_{\hat{\boldsymbol{w}}}$ liest man aus den Diagrammen der Abbildung 3.24 einen Schwellwert von beispielsweise 0.85 ab. Wird eine autorisierte Person bei solchen Schwellen nicht erkannt (in unserem Beispiel die Person "subject01" mit Gesichtsausdruck "surprised"), dann kann ein Identifikationssystem den Nutzer auffordern, mit möglichst neutralem Gesichtsausdruck in die Kamera zu schauen.

In Abbildung 3.28 sind analog zu den Abbildungen 3.23 und 3.24 die Werte für das oben eingeführte Ähnlichkeitsmaß S_h dargestellt. Hier fällt auf, daß der Wert der nicht-

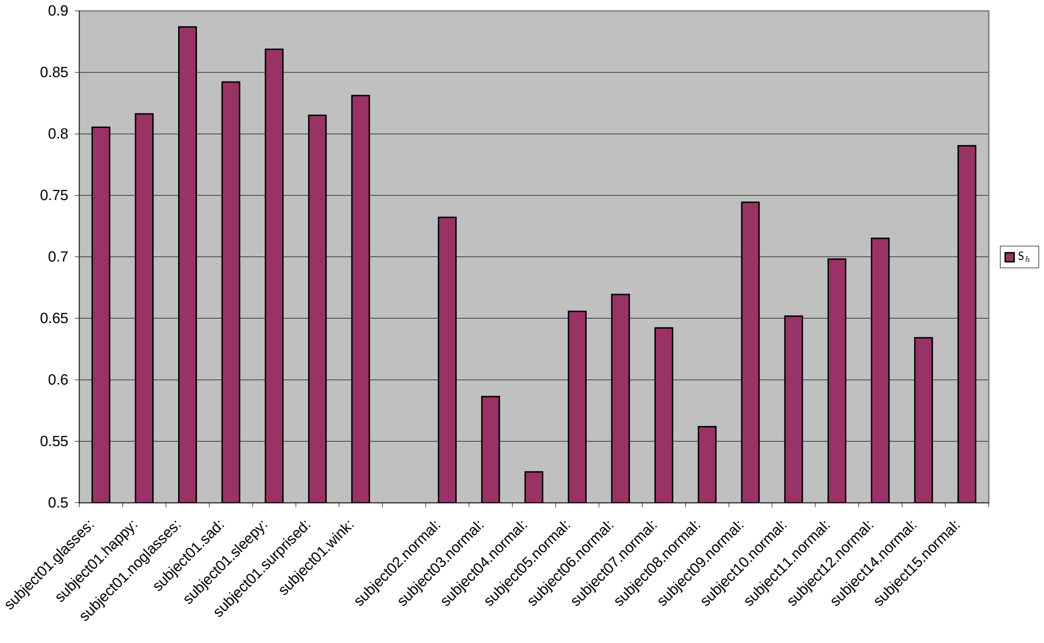


Abbildung 3.28: Ähnlichkeitsmaß S_h. Mit der Waveletnetz-Repräsentation von "subject01" wurden die an der horizontalen Achse abgetragenen Testbilder rekonstruiert (vgl. Abbildungen 3.21 und 3.22) und die Werte des Ähnlichkeitsmaßes S_h berechnet. Für "subject15" wird ein Wert erreicht, der dicht bei den Werten der Testbilder der autorisierten Person "subject01" liegt.

autorisierten Person "subject15" dicht bei den Werten von "subject01" liegt. Bei den in Abschnitt 3.5 eingeführten Ähnlichkeitsmaßen trat das Testbild "subject01.surprised" als einzelner Ausreißer auf. Beim Maß S_h liegen mehrere Testbilder von "subject01" auf und unter dem Niveau von “subject01.surprised”. In Diagramm 3.28 müßte eine sichere Schwelle für S_h z.B. bei einem Wert von 0.85 liegen. Damit würden 5 der sieben Testbilder von “subject01” nicht korrekt identifiziert. In Kapitel 4.2 werden wir näher untersuchen, ob sich das schlechtere Abschneiden von S_h verallgemeinern läßt.

Aufgrund der kompakten Gesichtsrepräsentation durch Waveletnetze scheinen diese prädestiniert für eine oben skizzierte Sicherheitsanwendung zu sein. Wir haben Waveletnetze mit 42 und 64 Gaborwavelets eingesetzt. Jedes Wavelet ist durch sechs Gleitkommaparameter vollständig beschrieben, damit ergibt sich für ein Netz mit 64 Wavelets ein Speicherbedarf von 384 Gleitkommawerten, was 1536 Bytes entspricht. In Abschnitt 4.2 untersuchen wir in Abhängigkeit von gegebenen Schwellwerten die Anzahl der Fehlidentifizierungen und damit die Einsetzbarkeit der verschiedenen Ähnlichkeitsmaße für Sicherheitsanwendungen.

4 Experimente und Auswertung

In diesem Kapitel untersuchen wir die Erkennungsraten der in Kapitel 3 vorgestellten Verfahren.

Die Experimente führen wir mit der Yale- [30] und der Manchester-Gesichtsdatenbank [17] durch. Abbildung 4.1 zeigt die verwendeten Testbilder der Yale-Datenbank. Diese Datenbank umfaßt Gesichtsbilder von 15 Personen. Pro Person stehen mehrere Gesichtsausdrücke zur Verfügung, wir nutzen die Bilder “happy”, “sleepy”, “sad”, “surprised” und “wink”. Als Referenzbilder zur Erzeugung der Waveletnetz-Repräsentationen verwenden wir jeweils das Bild mit neutralem Gesichtsausdruck (“normal”).

Abbildung 4.1: Testbilder der Yale-Datenbank [30]. Pro Person sind jeweils die Gesichtsausdrücke “happy”, “sleepy”, “sad”, “surprised” und “wink” abgebildet.

In Abbildung 4.2 sind die als Testbilder verwendeten Aufnahmen aus der Manchester-Datenbank aufgeführt. In dieser Datenbank sind je 10 Aufnahmen von 30 Personen gespeichert. Wir haben pro Person jeweils ein Bild als Referenzbild sowie drei weitere als Testbilder gewählt. Die Manchester-Datenbank ist im Gegensatz zur Yale-Datenbank nicht nach bestimmten Gesichtsausdrücken organisiert. Nichtsdestotrotz umfassen die von uns zufällig ausgesuchten Testbilder ein weites Spektrum von Gesichtsmimiken.

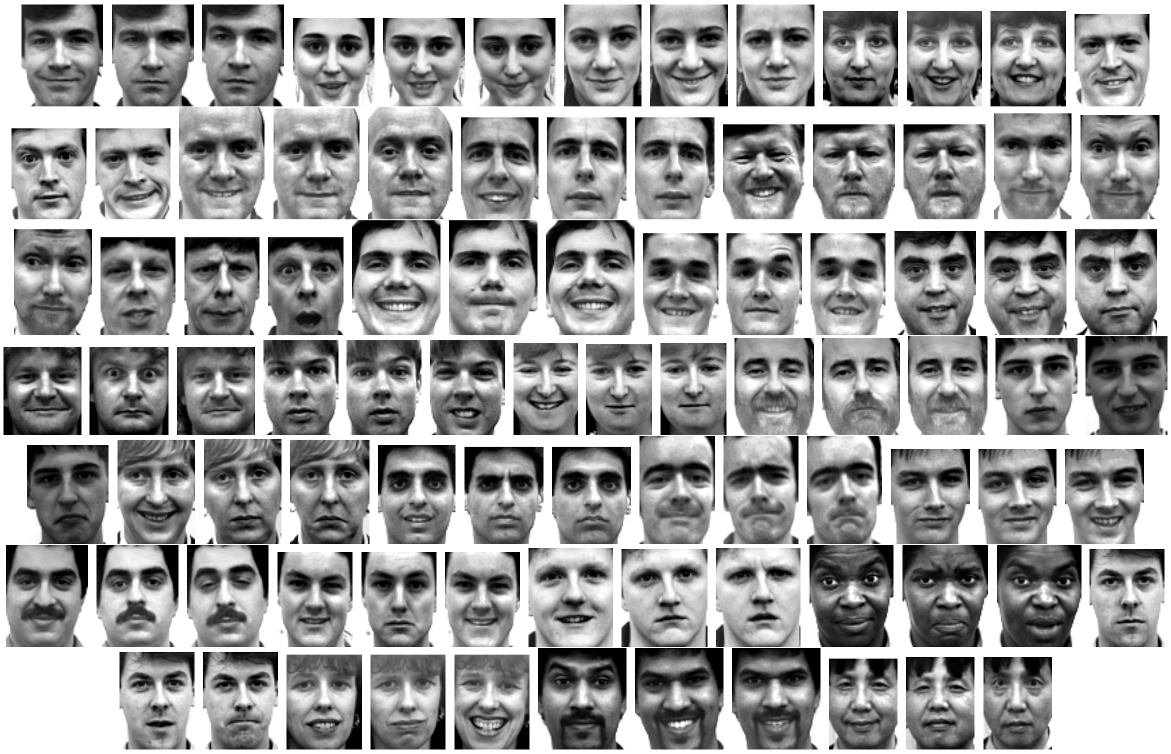


Abbildung 4.2: Testbilder aus der Manchester-Datenbank [17].

In Abschnitt 4.1 untersuchen wir die Erkennungsraten des im Abschnitt 3.6 vorgestellten Erkennungsverfahrens in Abhängigkeit vom verwendeten Ähnlichkeitsmaß sowie der Zahl der Wavelets pro Waveletnetz.

Eine entsprechende Auswertung der Verifikationsanwendung aus Abschnitt 3.7 erfolgt in Abschnitt 4.2.

4.1 Experimente zur Gesichtserkennung

Das im Abschnitt 3.6 vorgestellte Verfahren zur Gesichtserkennung wollen wir nun mit den oben angegebenen Gesichtsdatenbanken testen. Aus der Yale-Datenbank verwenden wir dazu 75 Testbilder, aus der Manchester-Datenbank 90. Wir wollen die

Abhängigkeit der Erkennungsrate, also dem Anteil der korrekt erkannten Gesichter, von der Anzahl der Wavelets pro Waveletnetz sowie vom eingesetzten Ähnlichkeitsmaß feststellen.

In Tabelle 4.1 sind die Erkennungsraten für Waveletnetze mit 42 und 64 Wavelets in Abhängigkeit der in Abschnitt 3.5 vorgestellten Ähnlichkeitsmaße aufgeführt. In Abbildung 4.3 sind die Werte graphisch dargestellt.

	Erkennungsrate [%]			
	Yale-Datenbank		Manchester-Datenbank	
S	42 Wavelets	64 Wavelets	42 Wavelets	64 Wavelets
$\mathsf{S}_{\|\cdot\|_2}$	94.7	92.0	93.3	91.1
S_{C}	96.0	88.0	92.2	90.0
$\mathsf{S}_{\boldsymbol{w}}$	89.3	85.3	87.8	83.3
$\mathsf{S}_{\hat{\boldsymbol{w}}}$	90.7	86.7	88.9	84.4

Tabelle 4.1: Erkennungsraten.

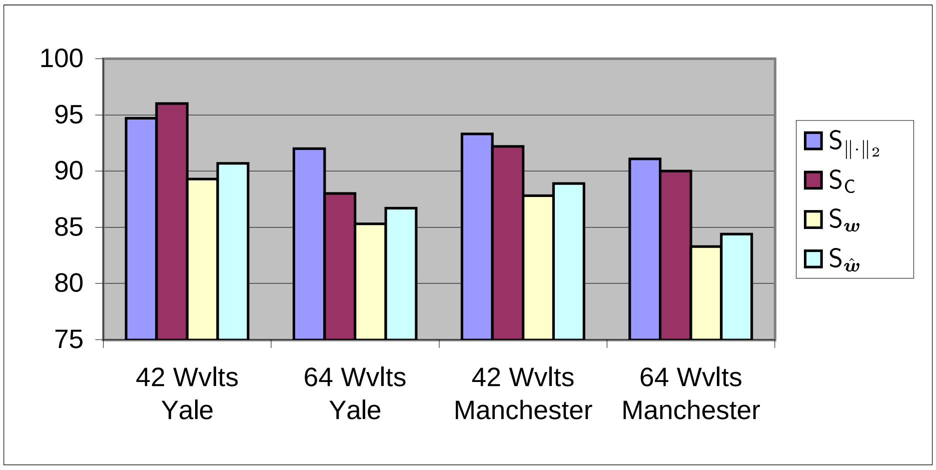


Abbildung 4.3: Erkennungsraten.

Es fällt auf, daß die Erkennungsleistung bei der Verwendung der Waveletnetze mit 64 Wavelets gegenüber Netzen mit 42 Wavelets geringer ist. Das ist dadurch zu begründen, daß mit steigender Zahl von Wavelets ein Waveletnetz der Person i Testbilder von Personen $j \neq i$ mit zunehmend geringerem Fehler rekonstruieren kann. Weiterhin wird ersichtlich, daß die im Signalraum motivierten Maße $\mathsf{S}_{\|\cdot\|_2}$ und S_{C} bessere Ergebnisse erzielen als die Ähnlichkeitsmaße $\mathsf{S}_{\boldsymbol{w}}$ und $\mathsf{S}_{\hat{\boldsymbol{w}}}$ im Gewichtsraum. Die Berücksichtigung der Größen der Wavelets bei dem Maß $\mathsf{S}_{\hat{\boldsymbol{w}}}$ durch die Verwendung gemäß Gleichung

(3.24) modifizierter Gewichtsvektoren $\hat{\boldsymbol{w}}$ liefert nicht die erhoffte Verbesserung gegenüber dem Maß $\mathsf{S}_{\boldsymbol{w}}$. Es wird zwar eine Verbesserung von durchschnittlich 1.3% erreicht, das Maß $\mathsf{S}_{\|\cdot\|_2}$ liefert jedoch durchschnittlich um 5.1% bessere Ergebnisse.

Insgesamt liefert das Maß $\mathsf{S}_{\|\cdot\|_2}$ mit Werten von 91.1% bis 94.7% die besten Erkennungsraten, auch wenn die maximale Erkennungsrate von 96% für die Yale-Datenbank mit dem Ähnlichkeitsmaß S_C erreicht wird.

Die Ergebnisse für die Manchester-Datenbank sind etwas schlechter als die Ergebnisse der Yale-Datenbank. Durch die Verdopplung der Anzahl zu erkennender Individuen steigt die Wahrscheinlichkeit einer Fehlerkennung.

Für die Yale-Datenbank analysieren wir in Tabelle 4.2, welche Gesichtsausdrücke in den Testbildern die meisten Fehler verursacht haben. Wie wir im Abschnitt 3.6 vermuteten, sind dies die Gesichtsausdrücke “surprised” gefolgt von “happy”, da bei diesen die stärksten Abweichungen zu den jeweiligen Referenzgesichtern auftreten.

	Anzahl der Fehler							
	42 Wavelets				64 Wavelets			
	$\mathsf{S}_{\|\cdot\|_2}$	S_C	$\mathsf{S}_{\boldsymbol{w}}$	$\mathsf{S}_{\hat{\boldsymbol{w}}}$	$\mathsf{S}_{\|\cdot\|_2}$	S_C	$\mathsf{S}_{\boldsymbol{w}}$	$\mathsf{S}_{\hat{\boldsymbol{w}}}$
“happy”	2	2	2	2	1	4	2	4
“sleepy”	0	0	0	0	0	0	0	0
“sad”	0	0	1	1	1	0	1	0
“surprised”	2	1	5	4	3	3	7	4
“wink”	0	0	0	0	1	2	1	2

Tabelle 4.2: Verteilung der Fehler bei der Yale-Datenbank auf die einzelnen Gesichtsausdrücke.

4.2 Experimente zur Verifikation der Identität

Wir wollen nun die Eignung der in Abschnitt 3.5 eingeführten Ähnlichkeitsmaße für die in Abschnitt 3.7 beschriebene Anwendung zur Verifikation der Identität untersuchen. Die z.B. auf einer Chipkarte gespeicherte Waveletnetz-Repräsentation einer autorisierten Person soll mit der aus dem gespeicherten Waveletnetz erzeugten Repräsentation eines Testgesichts verglichen werden. Dazu werden Schwellwerte für die Ähnlichkeitsmaße benötigt, deren Überschreitung zur Bestätigung der Identität und Unterschreitung zur Zurückweisung durch das Verifikationssystem führen.

Da wir eine Sicherheitsanwendung vor Augen haben, wollen wir nicht in Kauf nehmen, daß eine nicht autorisierte Person durch das System als autorisiert klassifiziert wird (*false positives*). Deshalb wählen wir grundsätzlich Schwellwerte, für die keine solchen falsch-positiven Klassifikationen auftreten und zählen, wie viele Testbilder der autorisierten Person fälschlicherweise zurückgewiesen werden (*misses*). Die Wahl der Schwellwerte erfolgt so, daß die Anzahl der falschen Zurückweisungen minimiert wird, jedoch keine falsch positiven Klassifikationen auftreten.

Unsere Experimente führen wir mit Bildern der im letzten Abschnitt beschriebenen Yale-Datenbank durch. Als Testbilder einer autorisierten Person verwenden wir jeweils die Bilder "sad", "sleepy", "happy", "surprised" und "wink". Für die übrigen, als nicht autorisiert zu klassifizierenden Personen, verwenden wir zusätzlich zu diesen Gesichtsausdrücken das jeweilige Referenzbild "normal". Diese Bildmenge wird jeweils als Eingabe für die Verifikation mit den Waveletnetz-Repräsentationen der 15 Referenzpersonen verwendet. Damit können pro Person von diesen 89 Bildern maximal 5 falsch zurückgewiesen werden.

Wir betrachten zwei Ansätze für Schwellwerte. Zunächst prüfen wir, ob zu einem Ähnlichkeitsmaß ein universeller Schwellwert gefunden werden kann, der für alle Waveletnetz-Repräsentationen gute Ergebnisse liefert. Anschließend untersuchen wir, ob bessere Ergebnisse erzielt werden können, wenn zu jedem gespeicherten Waveletnetz ein individuell bestimmter Schwellwert verwendet wird. Wir setzen für die Experimente in diesem Abschnitt die Ähnlichkeitsmaße ein, die auf das Intervall $[-1, 1]$ normiert sind. Dies sind die normalisierten Kreuzkorrelationen S_{C} und S_h sowie die Kosinusse $\mathsf{S}_{\boldsymbol{w}}$ und $\mathsf{S}_{\hat{\boldsymbol{w}}}$ der durch die Gewichtsvektoren eingeschlossenen Winkel.

In Tabelle 4.3 sind die jeweils verwendete universelle Schwelle sowie für jede Person der Datenbank und jedes Ähnlichkeitsmaß die Anzahl der falsch zurückgewiesenen Testbilder angegeben. Die universellen Schwellen sind so bestimmt, daß keine falsch positiven Klassifikationen auftreten und dabei eine minimale Anzahl falsch zurückgewiesener Testbilder erreicht wird.

		Anzahl falsch zurückgewiesener Testbilder														
Maß	θ	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
S_{C}	0.95	4	2	1	2	2	2	3	4	3	2	1	4	5	1	2
$\mathsf{S}_{\boldsymbol{w}}$	0.97	4	3	3	1	2	2	2	3	4	1	1	5	4	0	3
$\mathsf{S}_{\hat{\boldsymbol{w}}}$	0.94	4	3	2	0	2	2	3	3	3	1	1	4	3	1	3
S_h	0.87	5	5	5	5	1	5	5	5	5	5	4	5	5	5	4

Tabelle 4.3: Anzahl falsch zurückgewiesener Testbilder in Abhängigkeit vom verwendeten Ähnlichkeitsmaß. Für die 15 Waveletnetze der Personen "subject01" bis "subject15" wurde jeweils derselbe universelle Schwellwert θ verwendet, der in der Tabelle angegeben ist.

Es fällt die hohe Anzahl falsch zurückgewiesener Testbilder auf, in einigen Fällen werden sämtliche Testbilder einer autorisierten Person zurückgewiesen. Besonders schlecht bei der Verwendung eines einheitlichen Schwellwertes schneidet das Maß S_h ab. Dies ist nicht verwunderlich, da für verschiedene Referenzpersonen der Approximationsfehler der Waveletnetzrepräsentation bezogen auf das Referenzgesicht unterschiedlich hoch ist. Vom Referenzbild verschiedene Testbilder werden mit einem höheren Approximationsfehler rekonstruiert. Für jede Waveletnetzrepräsentation liegt demnach ein Grund-Approximationsfehler vor, gegeben durch den Approximationsfehler des Referenzbildes. Die Absolutwerte der Approximationsfehler der einzelnen Waveletnetze variieren von Person zu Person. Da das Maß S_h umgekehrt proportional zu diesen Approximationsfehlern ist, ist es verständlich, daß kaum ein universeller Schwellwert gefunden werden kann.

Es liegt auf der Hand, den jeder Waveletnetz-Repräsentation eigenen Approximationsfehler des Referenzgesichts zu berücksichtigen, indem für die einzelnen Waveletnetze individuelle Schwellwerte für S_h bestimmt werden. Analog zu den universellen Schwellwerten werden die individuellen Schwellwerte so festgelegt, daß keine falsch positiven Klassifikationen zugelassen werden und dabei die Anzahl falsch zurückgewiesener Testbilder minimiert wird. Den Ansatz mit individuellen Schwellwerten testen wir ebenfalls für die übrigen Ähnlichkeitsmaße.

In Tabelle 4.4 sind die ermittelten individuellen Schwellwerte für die verschiedenen Ähnlichkeitsmaße aufgeführt. Die daraus resultierenden falschen Zurückweisungen zeigt Tabelle 4.5. In Abbildung 4.4 sind die Ergebnisse mit universellen Schwellwerten denen mit der Verwendung individueller Schwellwerte gegenübergestellt. Bei allen Ähnlichkeitsmaßen wird durch den Einsatz individueller Schwellen eine geringere Anzahl falscher Zurückweisungen erreicht.

	Schwellwert θ														
Maß	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
S_C	0.83	0.93	0.9	0.88	0.95	0.94	0.87	0.88	0.94	0.89	0.87	0.96	0.84	0.87	0.95
$\mathsf{S}_{\boldsymbol{w}}$	0.71	0.87	0.88	0.91	0.9	0.91	0.78	0.89	0.89	0.86	0.84	0.93	0.81	0.88	0.89
$\mathsf{S}_{\hat{\boldsymbol{w}}}$	0.76	0.89	0.89	0.92	0.93	0.91	0.83	0.93	0.9	0.91	0.86	0.94	0.88	0.88	0.93
S_h	0.81	0.85	0.87	0.79	0.87	0.86	0.81	0.76	0.87	0.79	0.79	0.87	0.78	0.8	0.85

Tabelle 4.4: Individuelle Schwellwerte für die 15 Waveletnetze zur Yale-Datenbank in Abhängigkeit vom Ähnlichkeitsmaß.

Das Ähnlichkeitsmaß S_h erweist sich als gänzlich ungeeignet für unsere Verifikationsaufgabe. Trotz der Verwendung individueller Schwellwerte werden von sechs Personen sämtliche Testbilder zurückgewiesen. Die besten Resultate liefert das Ähnlichkeitsmaß S_C, wenn auch für die Person "subject12" vier der fünf Testbilder zurückgewiesen

Maß	Anzahl falsch zurückgewiesener Testbilder														
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
S_{C}	0	1	0	0	1	2	0	2	2	0	0	4	0	0	0
$\mathsf{S}_{\boldsymbol{w}}$	1	2	1	0	1	2	1	3	2	1	1	5	0	0	3
$\mathsf{S}_{\hat{\boldsymbol{w}}}$	1	2	1	0	1	1	0	3	1	1	1	4	1	0	3
S_{h}	0	4	5	2	1	5	5	5	5	1	0	5	5	5	3

Tabelle 4.5: Anzahl falsch zurückgewiesener Testbilder in Abhängigkeit vom Ähnlichkeitsmaß. Als Schwellwert θ wurde jeweils der individuelle Schwellwert aus Tabelle 4.4 verwendet.

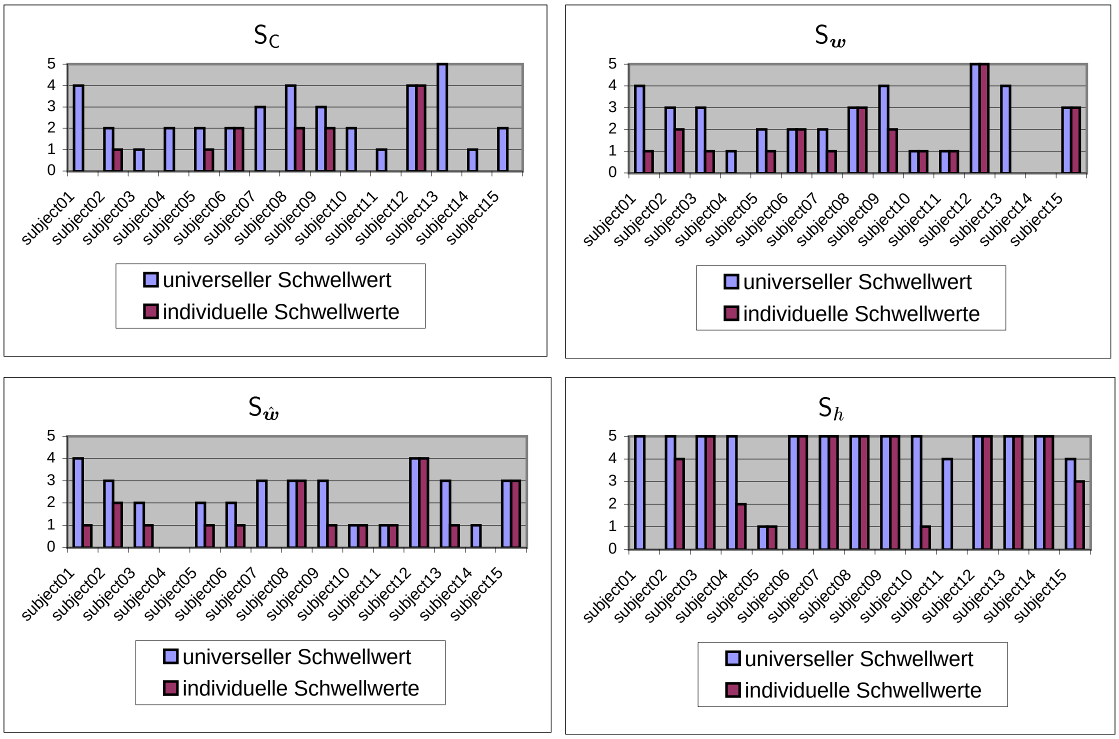


Abbildung 4.4: Vergleich der Verifikationsergebnisse bei der Verwendung von universellen und individuellen Schwellwerten. Angegeben sind jeweils die Anzahl falsch zurückgewiesener Testbilder.

werden.

In Tabelle 4.6 listen wir die falsch zurückgewiesenen Testbilder auf. Wie bereits in Abschnitt 3.7 vermutet, sind dies meist die Gesichtsausdrücke “happy”, “wink” und “surprised”, da sie die stärksten Abweichungen zum jeweiligen Referenzbild aufweisen.

Person	zurückgewiesene Testbilder
subject02	surprised
subject05	happy
subject06	surprised, wink
subject08	surprised, wink
subject09	happy, surprised
subject12	happy, sad, surprised, wink

Tabelle 4.6: Zurückgewiesene Testbilder bei der Verwendung des Ähnlichkeitsmaßes $\mathsf{S_C}$ und individueller Schwellwerte.

Mit Ausnahme von “subject12” werden jeweils die Bilder “sleepy” und “sad” korrekt als Gesichter der autorisierten Personen erkannt.

5 Zusammenfassung und Ausblick

In dieser Diplomarbeit haben wir den Einsatz von Gabor-Waveletnetzen zur Gesichtserkennung untersucht.

Die in [21] zur Repräsentation allgemeiner Grauwertbilder eingeführte Architektur der Waveletnetze mit Gaborwavelets haben wir angepaßt und erweitert zur Repräsentation von Gesichtsbildern. Dazu haben wir den Optimierungsprozeß modifiziert, der ausgehend von einer Initialisierung die Wavelets des Netzes an die Gesichtsmerkmale des zu repräsentierenden Gesichts anpaßt. Das Optimierungsverfahren liefert als Gesichtsrepräsentation einen individuellen Satz von Wavelets, der charakteristisch ist für das repräsentierte Gesicht. Die Gesichtsrepräsentation setzt sich aus diesen Wavelet-Basisfunktionen und einer entsprechenden Gewichtung zusammen.

Für die so gegebene Gesichtsrepräsentation haben wir ein Verfahren vorgestellt, die Basisfunktion affin zu transformieren, um die Gesichtsrepräsentation mit einem Testbild in Deckung zu bringen. Dieses Matchingverfahren setzen wir zur Bestimmung von Position, Skala und Orientierung zu identifizierender Gesichter ein. Die Anwendungsmöglichkeit des Matchingverfahrens zum Verfolgen von Gesichtern in Bildsequenzen haben wir ebenfalls demonstriert. Das Matching beruht auf dem Konzept der Superwavelets, bei dem ein gesamtes Waveletnetz als neues Wavelet behandelt wird, dessen Parameter mit den bereits zur Anpassung von Wavelets bei der Gewinnung der Gesichtsrepräsentation vorgestellten Algorithmen optimiert werden. Das Matching funktioniert stabil, gegenüber dem ebenfalls zum Matching eingesetzten Active-Appearence-Model von Cootes et al. [4] kann die Robustheit gegenüber der Wahl der Startposition nahezu verdoppelt werden.

Ganz wesentlich für die Erkennung ist unsere vorgestellte Methode, die Gewichte eines affin transformierten Waveletnetzes für ein Testbild neu zu berechnen, also die Rekonstruktion des Testbildes mit der Gesichtsrepräsentation. Die neuen Gewichte werden durch die orthogonale Projektion des Testbildes in den durch die Wavelets aufgespannten Raum berechnet. Dieses Verfahren setzen wir auch zur Initialisierung von Waveletnetzen ein.

Die Erkennung von Gesichtern erfolgt auf der Basis von Ähnlichkeitsmaßen, die es erlauben, zwei Waveletnetze miteinander zu vergleichen. Ein Ziel der Diplomarbeit ist die Untersuchung verschiedener Ähnlichkeitsmaße im Hinblick auf die Leistung bei der Gesichtserkennung. Hierbei stellt sich heraus, daß die im Signalraum motivierten Ähnlichkeitsmaße, bei denen die durch Waveletnetze gelieferten Grauwertbilder verglichen werden, am besten zur Erkennung geeignet sind. Der Vergleich von Waveletnetzen allein durch den paarweisen Vergleich der Gewichte der Basisfunktionen liefert schlechtere Ergebnisse.

Des weiteren habe wir den Einsatz der Ähnlichkeitsmaße zur Verifikation der Identität einer Person getestet. Die Verifikation soll dabei allein durch den Einsatz *eines* Waveletnetzes erfolgen. Mit diesem Waveletnetz wird das zu prüfende Gesicht rekonstruiert und ein ermittelter Ähnlichkeitswert mit einem Schwellwert verglichen. Liegt der Wert über dem Schwellwert, gilt die Identität als bestätigt. Um zu erreichen, daß keine falsch-positive Einstufung der Identität erfolgt, müssen die Schwellwerte so hoch gesetzt werden, daß im Gegenzug viele eigentlich als positiv zu erkennende Testbilder fälschlich zurückgewiesen werden. Nur wenn der Gesichtsausdruck auf dem Testbild dem durch das Waveletnetz repräsentierten Referenzbild nahezu entspricht, wird eine Person als autorisiert eingestuft. Damit ist ein sinnvoller Einsatz von Waveletnetzen der hier vorgestellten Form für Verifikationsaufgaben kaum möglich.

Mit Waveletnetzen ist eine sehr kompakte Repräsentation von Gesichtern möglich. Für unsere Experimente verwendeten wir Netze bestehend aus 42 und 64 Gaborwavelets, die jeweils durch 6 Parameter vollständig beschrieben werden. Zusätzlich zu diesen Parametern wird kein *code book* z.B. mit Hauptkomponenten wie bei dem Active-Appearence-Model [4] oder bei den auf der Hauptkomponentenanalyse basierenden Erkennungsverfahren benötigt. Die Eigenschaft der kompakten Repräsentation ist ideal für den Einsatz für oben beschriebene Verifikationsaufgaben, leider machen die für diesen Einsatzzweck wenig geeigneten Ähnlichkeitsmaße diesen Vorteil zunichte.

Die Waveletnetz-Repräsentation erhält man mit minimaler Nutzerinteraktion. Lediglich der zu repräsentierende Bildbereich (in unserem Fall das innere Gesicht) ist durch den Nutzer anzugeben. Ein aufwendiges Markieren einer großen Anzahl von Bildern mit Landmarken zur Gewinnung von Gesichtsrepräsentationen, wie es z.B. in [7] erforderlich ist, ist für Waveletnetze nicht notwendig. Eine Kombination mit einem Gesichtsdetektor könnte auch die Bestimmung des inneren Gesichts automatisieren.

Ein Nachteil der untersuchten Verfahren ist, daß die Gesichtsrepräsentation lediglich aus jeweils einem einzelnen Referenzbild gewonnen wird. Dadurch ist es schwierig, einen großen Spielraum bei Variationen der Gesichtsmimik zuzulassen. Um vom Referenzbild stark abweichende Gesichtsausdrücke sicher erkennen zu können, müßte die Waveletnetz-Repräsentation so erweitert werden, daß mit einem Netz verschiedene Gesichtsausdrücke modelliert werden können, vergleichbar mit dem Active-Appearance-Model von Edwards et al. [7]. Das Active-Appearance-Model ist in der Lage, mit einem Modell sämtliche zur Gewinnung des Modells gespeicherten Gesichtsbilder nahezu photorealisitisch zu synthetisieren. Bei dem Bunch-Graph-Verfahren von Wiskott et al. [29] wird durch das elastische Graph-Matching sichergestellt, daß die Gaborfilterbänke auch bei stärkeren Variationen des Gesichtsausdrucks auf den entsprechenden lokalen Gesichtsmerkmalen positioniert bleiben. Damit ist zu erwarten, daß der Bunch-Graph-Ansatz deutlich robuster gegenüber Variationen in der Gesichtsmimik ist als das hier vorgestellte Verfahren für Waveletnetze.

Ein weiterer Nachteil der Waveletnetze ist, daß bisher für jede zu erkennende Person eine eigenes Waveletnetz benötigt wird, das zeitaufwendig auf das Testgesicht gematcht und rekonstruiert werden muß. Beispielsweise bei der Hauptkomponentenanalyse bilden alle zu erkennenden Personen einen *gemeinsamen* Eigenraum, in den das Testbild projiziert werden muß. Anhand der Lage der Projektion im Eigenraum erfolgt die Identifizierung. Zur Erkennung mit den Waveletnetzen muß das Testbild in *jeden* der durch die Basisfunktionen der Waveletnetz-Repräsentationen aufgespannten Räume projiziert werden, wodurch das Verfahren langsam und bei einer größeren Zahl zu erkennender Personen inpraktikabel wird.

Im Rahmen der Diplomarbeit haben wir eine Implementierung der Waveletnetze in C++ unter dem Bildverarbeitungssystem Khoros 2 vogenommen. Die entstandenen Programme laufen unter Linux und Solaris. Die Implementation erfolgte streng objektorientiert unter dem Gesichtspunkt einer möglichst leichten Weiterverwendung der Quelltexte für nachfolgende Arbeiten. Der modulare Aufbau ermöglicht eine leichte Erweiterung beispielsweise um neue Optimierungsalgorithmen oder andere Wavelettypen. Großen Wert legten wir auf die Kapselung der Implementation der Waveletnetze gegenüber dem speziellen Bildverarbeitungssystem. Dies gelang durch geeignete Schnittstellen zu den Bildformaten von Khoros. Mit geringem Aufwand kann die Implementierung der Waveletnetze von Khoros abgekoppelt und an andere Bildformate angepaßt werden. Diese Abkoppelung ist bereits für eine aktuell laufende Diplomarbeit am Lehrstuhl Kognitive Systeme problemlos durchgeführt worden.

Eine naheliegende Erweiterung der Wavletnetze wäre die Repräsentation einzelner Gesichtsmerkmale wie Augen, Mund und Nase durch Teilnetze. Beim Matching könnten dann in Analogie zu Brunellis und Poggios templatebasierten Ansätzen [2] die Teilnetze einzeln gematcht werden und die aus den relativen Positionierungen gewonnenen Informationen zur Erkennung zusätzlich genutzt werden. Die durch Teilnetze repräsentierten Merkmale könnten zur Erkennung detailgetreuer repräsentiert werden, wobei zum robusten Matching grobskalige Merkmalsrepräsentationen eingesetzt werden sollten. Die softwareseitige Realisierung von Teilnetzen ist bereits implementiert.

Für den erfolgreichen Einsatz zur Gesichtserkennung müssen Wege gefunden werden, wie in einem Waveletnetz (gegebenenfalls mit zusätzlichen Unternetzen für Augen, Nase etc.) mehrere Individuen kodiert werden können, wie es beispielsweise bei den Bunch-Graphen der Fall ist.

Symbolverzeichnis

Übersicht der verwendeten mathematischen Notationen und wichtigsten Symbole.

$\Big\vert_{\boldsymbol{x}}$	"ausgewertet an der Stelle $\boldsymbol{x}$", z.B. $\frac{\partial \chi^2}{\partial \theta_i}\Big\vert_{\boldsymbol{x}}$
$< f, g >$	Skalarprodukt im $L^2(\mathbb{C})$: $< f, g >= \int_{-\infty}^{\infty} f(x) g^*(x)\, dx$
f^*	komplex Konjugierte von $f \in \mathbb{C}$
$\bar{f}$	Gleichanteil: $\bar{f} = \int_{-\infty}^{\infty} f(x)\, dx$
$\boldsymbol{X}^T$, $\boldsymbol{x}^T$	transponierte Matrix bzw. transponierter Vektor
$\boldsymbol{a}_i$	Parametervektor eines Wavelets
A_Ψ	Auswertungsbereich eines Waveletnetzes mit Basisfunktionen Ψ
C	normalisierte Kreuzkorrelation
$\boldsymbol{D}_i$	Dilatationsmatrix
d_{x_i}, d_{y_i}	Dilatationsparameter
g	Grauwertbild
$g_{\mathcal{W}}$	Grauwertbild der Ausgabe des Waveletnetzes $\mathcal{W}$
h	Testbild
$\boldsymbol{H}$	Hesse-Matrix
K	Anzahl der Zeilen eines Grauwertbilds
L	Anzahl der Spalten eines Grauwertbilds
$L^2(\mathbb{R})$	Raum der quadratintegrierbaren Funktionen über $\mathbb{R}$
$L^2(\mathbb{C})$	Raum der quadratintegrierbaren Funktionen über $\mathbb{C}$
$\boldsymbol{n}$	Parametervektor eines Superwavelets
N	Anzahl der Wavelets im Waveletnetz $\mathcal{W}$
M	Anzahl der Bilder in der Referenzdatenbank
φ	Regularisierungspotential
ψ	Mutterwavelet
$\psi_{a,b}$	Wavelet mit Dilatationsparameter a und Translationsparameter b
$\psi_{\boldsymbol{a}}$	Wavelet mit Parametervektor $\boldsymbol{a}$
Ψ	Familie von Wavelets
$\boldsymbol{\Psi}$	Gramsche Matrix

Ψ	Superwavelet
$\boldsymbol{R}_i$	Rotationsmatrix
$\mathsf{S}_{\Vert\cdot\Vert_2}$	Ähnlichkeitsmaß im Signalraum, Kehrwert des euklidischen Abstands
S_{C}	Ähnlichkeitsmaß im Signalraum, normalisierte Kreuzkorrelation
$\mathsf{S}_{\boldsymbol{w}}, \mathsf{S}_{\hat{\boldsymbol{w}}}$	Ähnlichkeitsmaße im Gewichtsraum
S_h	Normalisierte Kreuzkorrelation zwischen Testbild h und Ausgabe eines Waveletnetzes
$\boldsymbol{t}_i = (t_{x_i} t_{y_i})^T$	Translationsparameter
$\boldsymbol{\theta}$	Parametervektor des Levenberg-Marquardt-Verfahrens, $\boldsymbol{\theta} = (\theta_{1,\ldots,\theta_{6N}})^T$
$\nabla\chi^2$	Gradient bezüglich $\boldsymbol{\theta}$, $\nabla\chi^2 = (\frac{\partial\chi^2}{\partial\theta 1}, \ldots \frac{\partial\chi^2}{\partial\theta_{6N}})^T$
w_i	Gewicht des i.ten Wavelets eines Waveletnetzes
$\boldsymbol{w}$	Gewichtsvektor $\boldsymbol{w} = (w_1, \ldots, w_N)^T$
$\mathcal{W}$	Waveletnetz
χ^2	quadratischer Fehler
$z(\boldsymbol{x})$	Ausgabe eines Waveletnetzes $\mathcal{W}$ zur Eingabe $\boldsymbol{x}$

Literaturverzeichnis

[1] R. Brunelli & T. Poggio, 1992. Face Recognition through Geometrical Features. In *Proc. European Conference on Computer Vision*, S. 792–800.

[2] R. Brunelli & T. Poggio, 1993. Face Recognition: Features versus Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052.

[3] R. Chellappa, C.L. Wilson & S. Sirohey, 1995. Human and Machine Recognition of Faces: A Survey. *Proceedings of the IEEE*, 83(5):705–740.

[4] T.F. Cootes, G.J. Edwards & C.J. Taylor, 1998. Active Appearance Models. In *Proc. European Conference on Computer Vision*, Band 2, S. 484–498.

[5] I. Daubechis, 1990. The Wavelet Transform, Time Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005.

[6] J.G. Daugman, 1988. Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7):1169–1179.

[7] G.J. Edwards, T.F. Cootes & C.J. Taylor, 1998. Face Recognition Using Active Appearance Models. In *Proc. European Conference on Computer Vision*, Band 2, S. 581–595.

[8] A. Happe, 1998. Implementation eines Gesichtsdetektors. Studienarbeit am Lehrstuhl Kognitive Systeme, Christian-Albrechts-Universität zu Kiel.

[9] B. Jähne, 1997. *Digitale Bildverarbeitung.* Springer, 4. Auflage.

[10] J.P. Jones & L.A. Palmer, 1987. An Evaluation of the Two-dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex. *Journal of Neurophysiology*, 58(6):1233–1258.

[11] M. Kirby & L. Sirovich, 1990. Application of the Karhunen-Loève Procedure for the Characterization of Human Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108.

[12] V. Krüger & G. Sommer, 2000. Gabor Wavelet Networks for Object Representation. Technischer Bericht 2002, Institut für Informatik und Praktische Mathematik, Christian-Albrechts-Universität zu Kiel.

[13] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R.P. Würtz & W. Konen, 1993. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers*, 42(3):300–310.

[14] S. Lawrence, C.L Giles, A.C. Tsoi & A.D. Back, 1997. Face Recognition: A Convolutional Neural Network Approach. *IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition*, 8(1):98–113.

[15] T.S. Lee, 1996. Image Representation Using 2D Gabor Wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10).

[16] S.G. Mallat, 1989. Multifrequency Channel Decompositions of Images and Wavelet Models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2091–2110.

[17] University of Manchester. Manchester face database.
http://peipa.essex.ac.uk/ipa/pix/faces/manchester.

[18] S. Marčelja, 1980. Mathematical Description of the Response of Simple Cortical Cells. *Journal of Optical Society of America*, 70(11):1297–1300.

[19] L. Máté, 1989. *Hilbert Space Methods in Science and Engineering.* Adam Hilger.

[20] B. Moghaddam & A. Pentland, 1997. Probabilistic Visual Learning for Object Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710.

[21] A. Pelz, 1997. *Aufgabenbezogen progressive Repräsentation von Bildern durch Waveletnetze.* Diplomarbeit, Christian-Albrechts-Universität zu Kiel.

[22] W.H. Press, S.A. Teukolsky, W.T. Vetterling & B.P. Flannery, 1992. *Numerical Recipes in C.* Cambridge University Press, 2. Auflage.

[23] O. Rioul & M. Vertterli, 1991. Wavelets and Signal Processing. *IEEE SP Magazine*, S. 14–38.

[24] J. Stoer, 1989. *Numerische Mathematik 1.* Springer-Lehrbuch, 5. Auflage.

[25] K.-K. Sung & T. Poggio, 1994. Example-based Learning for View-based Human Face Detection. A.I. Memo No. 1521, Massachusetts Institute of Technology.

[26] H.H. Szu, B. Telfer & J. Garcia, 1996. Wavelet Transforms and Neural Networks for Compression and Recognition. *Neural Networks*, 9(4):695–708.

[27] H.H. Szu, B. Telfer & S. Kadambe, 1992. Neural Network Adaptive Wavelets for Signal Representation and Classification. *Optical Engineering*, 31(9):1907–1916.

[28] M. Turk & A. Pentland, 1991. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(7):71–86.

[29] L. Wiskott, J.-M. Fellous, N. Krüger & C. von der Malsburg, 1997. Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779.

[30] Yale University. Yale face database. http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[31] Alan L. Yuille, 1991. Deformable Templates for Face Recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70.

[32] Q. Zhang & A. Benveniste, 1992. Wavelet Networks. *IEEE Transactions on Neural Networks*, 3(6):889–898.

Danksagung

Mein Dank gilt Volker Krüger für die intensive Betreuung bei der Anfertigung der vorliegenden Diplomarbeit. Die vielen ausführlichen Diskussionen waren sehr hilfreich für die Entwicklung des "roten Fadens".
Dem Lehrstuhl Kognitive Systeme danke ich für die angenehme, kollegiale und hilfsbereite Atmosphäre, insbesondere die Diskussionen mit Norbert Krüger und Thomas Bülow haben sehr geholfen.
Doris Jentzsch danke ich besonders dafür, daß sie mir stets unterstützend zur Seite stand. Ihr sowie Sibylle Bahr und Lars Kleinschmidt danke ich für kritische Kommentare und Hinweise auf Fehler bei der Durchsicht der Arbeit.
Mein ganz besonderer Dank gilt meinen Eltern für ihre Unterstützung während meines gesamten Studiums. Ohne sie wäre das Studium und letztlich diese Arbeit nicht in dieser Form möglich gewesen.

Erklärung

Hiermit versichere ich, daß ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Kiel, den 30. April 2000

- -

Alexander Happe