

Multi-Cue-Ansatz für ein dynamisches Auffälligkeitssystem zur visuellen Personenlokalisierung

Dissertation

zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs
(Dr.-Ing.)

der Fakultät für Informatik und Automatisierung
der Technischen Universität Ilmenau

vorgelegt von

Dipl.-Ing. Ulf-Dietrich Braumann
geb. am 11. Juni 1967 in Leipzig

Berichterstatter:

Prof. Dr.-Ing. Horst-Michael Groß (TU Ilmenau)

Prof. Dr. rer. nat. Georg Hartmann (Univ. Paderborn)

PD Dr.-Ing. habil. Karl-Heinz Franke (TU Ilmenau)

Tag der Einreichung:

21. Februar 2001

Tag der wissenschaftlichen Aussprache: 4. September 2001

Zusammenfassung

Die Entwicklung von technischen Sehsystemen mit neuartigen Schnittstellen zur Mensch-Maschine-Interaktion (MMI) wird international vorangetrieben. Angestrebt werden dabei Kommunikations- und Steuerungsmöglichkeiten ohne hohe Präzisionsanforderungen über natürliche menschliche Ausdrucksformen, wie Sprache, Mimik oder Gestik. Dadurch sollen zugleich herkömmliche, speziell zugeschnittene Eingabegeräte, wie Tastatur, Maus oder Joystick, abgelöst werden. Begünstigt wird eine solche Entwicklung durch die rapide gestiegene und weiter steigende Leistungsfähigkeit digitaler Rechnersysteme (insbesondere Universal-Hardware), so daß sich damit die Möglichkeiten berechnungsintensiver neuartiger künstlicher Hör- und Sehsysteme potentiell deutlich erweitert haben.

Die Arbeit ordnet sich in ein Szenario eines teilautonomen mobilen Servicesystems ein. Darin soll ein Nutzer anhand eines Satzes an Gesten dem Servicesystem (Roboter) bestimmte elementare Befehle übermitteln, um es damit für Transportaufgaben geeignet anzuweisen und zu positionieren. Elementare Voraussetzung dafür ist, daß das Sehsystem des Roboters zunächst die jeweilige Person visuell richtig lokalisiert hat. Genau diese Problemstellung behandelt die vorliegende Arbeit, die dabei besonderes Gewicht auf die Realisierung dieser Lokalisation unter *Echtzeitbedingungen* legt.

Gegenstand dieser Dissertation sind daher *Konzeption und Umsetzung eines neuen Mehrkomponenten-(Multi-Cue)-Ansatzes für ein visuelles Auffälligkeitssystem zur Detektion und Lokalisation von Personen*. Der Ansatz basiert im einzelnen auf drei unterschiedlichen, allgemein personentypischen Merkmalen, die als Auffälligkeitskomponenten (Cues) für die visuelle Personenlokalisierung geeignet sind: der hautfarbene Gesichtsbereich, das Gesichtsmuster selbst und die Außenkontur der Kopf-Schulter-Partie. Für sich genommen sind diese drei Cues nur recht begrenzt einsetzbar. Erst ihr Zusammenwirken ermöglicht den Einsatz in unpräparierten Umgebungen im Innen- und Außenbereich.

Für den Cue *Hautfarbe* wird ein Normalverteilungsklassifikator benutzt, der anhand exemplarischer Daten trainiert wurde. Der Cue *Gesicht* nutzt als Ähnlichkeitsmaß die einfache Cosinusmetrik zwischen normierten Vektoren (Bildausschnitt und Durchschnittsgesicht). Der *Konturcue* verwendet ein Arrangement orientierter Filter (Template), das die Kopf-

Schulter-Partie approximiert. Um in Grenzen unabhängig von der Entfernung der Person sein zu können und zugleich eine Entfernungsschätzung in die Lokalisation einbeziehen zu können, werden alle drei Cues in mehreren Auflösungsstufen verarbeitet und damit durchweg *pyramidale Repräsentationen* verwendet.

Um alle drei Cues verschmelzen zu können, werden zwei Schritte ausgeführt: Zunächst erfolgt eine Fusion mit einem *speziellen Fuzzy-Operator* (Min-Max), der sowohl die Nachteile einfacher Superposition oder einfacher Maximumsauswahl vermeidet, jedoch nicht generell auf die Vorteile der Nutzung von Koinzidenzen verzichtet. Damit erhält der Fusionsschritt einen integrativen Charakter. Für den anschließenden Selektionsprozeß, dessen Zielstellung die eigentliche Lokalisation der Person ist, wurde ein *3D dynamisches neuronales Feld* vom AMARI-Typ eingesetzt. Es besteht aus gegenseitig wechselwirkenden formalen dynamischen Neuronen und kann als rückgekoppeltes nichtlineares dynamisches System beschrieben werden. Ziel des darin ablaufenden *Winner-Take-All-Prozesses* (WTA) ist die Ausbildung eines einzelnen lokalen Blobs aktiver Neuronen, die sich bei Annäherung an einen Gleichgewichtszustand gegenüber den übrigen behaupten können. Dies stellt ein zugleich datengetriebenes und zustandsbasiertes Verfahren mit einer hohen Robustheit der Selektion gegenüber Artefakten dar.

Zur *Funktionsweise und Leistungsfähigkeit* des Systems werden eine Reihe von Lokalisationsbeispielen von Personen gezeigt und diskutiert. Es werden Vorschläge für mögliche Modifikationen des Systems sowie für weiterführende Untersuchungen gegeben.

Verschiedene Anhänge vertiefen einzelne Aspekte der Cues oder stellen z.T. alternative Algorithmen vor. Ein umfangreiches Literaturverzeichnis enthält alle genutzten Quellen der Arbeit.

Danksagung

Diese Arbeit ist das Ergebnis meiner etwas mehr als fünfjährigen Doktorandenzeit – zunächst als Stipendiat der Landesgraduiertenförderung des Landes Thüringen, später als wissenschaftlicher Mitarbeiter – am Fachgebiet Neuroinformatik, Fakultät für Informatik und Automatisierung der Technischen Universität Ilmenau.

Mein herzlicher Dank gilt allen, die zum Entstehen dieser Arbeit beigetragen haben, sei es durch fachliche Diskussionen, Anregungen und experimentelle Unterstützung oder insbesondere durch Aufmunterung und moralischen Beistand. Besonders danke ich Prof. Horst-Michael Groß und Dr. Hans-Joachim Böhme, die mir als betreuender Hochschullehrer bzw. Forschungsprojektleiter über den gesamten Zeitraum meiner Arbeiten am FG Neuroinformatik mit fachlichem und persönlichem Rat zur Seite gestanden und Vertrauen in meine Arbeit und wissenschaftliche Entwicklung gesetzt haben.

Danken möchte ich aber auch allen weiteren ehemaligen und gegenwärtigen Mitarbeitern und Studenten des FG Neuroinformatik, mit denen ich während meiner Arbeiten zusammengearbeitet habe und die hier alle genannt seien: Anja Brakensiek, Christoph Brandstädt, Dr. Andrea Corradini, Dr. Klaus Debes, Michael Engel, Wolfram Fischer, Dr. Rüdiger Franke, Heike Groß, Sören Hader, Dr. Fred Hamker, Dr. Dietmar Heinke, Andrea Heinze, Torsten Hempel, Dr. Bärbel Herrnberger, Dirk Hohlbein, Jürgen Key, Thomas Kleemann, Sandy Koch, Gunther Köhler, Prof. Edgar Körner, Dr. Ursula Körner, Markus Krabbes, Susan Lau, Dr. Ralf Möller, Rolf Nestler, Peter Paschke, Vania Pentcheva, Dr. Torsten Pomierski, Carsten Rieck, Holger Salewski, Carsten Schauer, Sabine Schulz, Ute Schütz, Torsten Seiler, Volker Stephan, Dimitrij Surmeli, Thomas Vesper, Dirk Wandtke, Stefan Weber, Dag Wendt, Torsten Wilhelm, Sören Wohlfarth und Thomas Zahn.

Darüber hinaus danke ich Dr. Karl-Heinz Franke und Heiko Kempe vom Zentrum für Bild- und Signalverarbeitung Ilmenau, deren Rat wesentlich zur Effizienzsteigerung der Algorithmen des im Rahmen dieser Arbeit implementierten Systems beigetragen hat.

Nicht vergessen möchte ich meine ehemaligen Kommilitonen Thomas Neubert und Uta Wölflick, ohne die mein Studium wohl eher nicht so gut gelaufen wäre, daß ich hätte meine Arbeiten als Doktorand aufnehmen können.

Größter Dank gebührt meiner Partnerin Angelika Pfäfflin, ohne die ich diese Arbeit nicht begonnen hätte niederzuschreiben, und unserem Sohn Fabian Braumann. Beide haben deswegen an vielen Wochenenden auf mich verzichtet.

Abschließend danke ich meinen Eltern Heinrich (†) und Ursula Braumann, denen ich diese Arbeit widmen möchte.

Inhaltsverzeichnis

Zusammenfassung	v
Danksagung	vii
Inhaltsverzeichnis	ix
Abbildungsverzeichnis	xiii
Tabellenverzeichnis	xvii
Abkürzungsverzeichnis	xix
1 Motivation	1
1.1 Einleitung	1
1.2 Szenario für ein teilautonomes mobiles Servicesystem	4
1.3 Visuell-basierte Personenlokalisierung	5
1.4 Résumé	10
2 Pyramidale Repräsentationen als Basiskonzept	13
2.1 Einleitung	13
2.2 Abstufung und Glättung	14
2.3 Unterabtastung und Interpolation	17
2.4 Résumé	18
3 Hautfarbdetektion	21

3.1	Einleitung	21
3.1.1	Zielstellung	21
3.2	Lineare Farbraummodelle	23
3.2.1	Additive Farbmischung	23
3.2.2	Primärvalenzmodelle	24
3.2.3	Gegenfarbmodelle	26
3.3	Hautfarbesegmentierung	27
3.3.1	Vorverarbeitung	27
3.3.2	Gewinnung von Hautfarbdaten	28
3.3.3	Charakterisierung von Hautfarbe	29
3.3.4	Wahl des Klassifikationsverfahrens	29
3.4	Ergebnisse	33
3.5	Résumé	34
4	Gesichtsdetektion	37
4.1	Einleitung	37
4.1.1	Zielstellung	38
4.2	Gesichtsspezifische Filterung	39
4.2.1	Normierte gesichtsspezifische Operatoren	39
4.2.1.1	Datenmaterial	40
4.2.1.2	Normierte mittelwertfreie Skalarprodukte (NMS)	40
4.2.1.3	Erhöhung der Selektivität – Beseitigung von Störungen	42
4.3	Ergebnisse	44
4.4	Résumé	45
5	Konturdetektion	47
5.1	Einleitung	47
5.2	Bedeutung von Konturen	49

5.3	Konturmodellierung durch ein Arrangement orientierter Filter	51
5.4	Orientierungsspezifische Analyse	54
5.4.1	Generelle physiologische Bezüge und mathematische Aspekte orientierungsspezifischer Analyse	54
5.4.2	Nichtlineare lokale Strukturanalyse zur Orientierungsbestimmung . .	56
5.4.2.1	Prinzip und Umsetzung der 2D-Tensor-Methode	57
5.4.2.2	Güteabschätzung	61
5.5	Ergebnisse	61
5.6	Résumé	63
6	Personenlokalisierung	65
6.1	Einleitung	65
6.2	Bekannte methodische Ansätze	66
6.3	Fusion und dynamische Selektion	70
6.3.1	Fusion der Cues mittels Fuzzy-Operator	71
6.3.2	Selektion im dreidimensionalen dynamischen neuronalen Feld	74
6.3.2.1	Zielstellung und grundsätzliches Prinzip	74
6.3.2.2	Konkrete Umsetzung	77
6.3.2.3	Numerische Lösung	81
6.4	Demonstration des realisierten Systems	85
6.5	Résumé	107
7	Zusammenfassung und Ausblick	109
7.1	Zusammenfassung	109
7.2	Ausblick	111
A	Ansätze zur referenzfreien Farbkonstanz	113
A.1	Retinex-Verfahren	113
A.2	Elementarfarbraum-Verfahren	115

B Eigenfaces	119
B.1 Eigenfaces zur Gesichtsdetektion	121
B.1.1 Projektion in den Face-Space	122
B.1.2 Detektion von Gesichtern mit Eigenfaces	123
B.2 Bestimmung der Eigenfaces	126
B.3 Effektivisierte Bestimmung der Eigenfaces	128
C Implementierbarkeit der Eigenface-Methode auf Spezialrechentechnik	131
C.1 Problemstellung: Mehrfache Faltung	131
C.2 CNAPS	132
C.3 Implementation der Eigenface-Methode auf CNAPS	134
D Nichtlineare Strukturanalyse – Theoretische Grundlagen	137
D.1 Lokale Nachbarschaft	137
D.2 Vektorielle Beschreibung von lokaler Orientierung	139
D.3 Tensorielle Repräsentation zur Beschreibung lokaler Orientierung	139
D.4 Lösung des Eigenwertproblems	141
D.5 Interpretation im Ortsfrequenzraum – Trägheitstensoranalogie	142
E Weitere tabellarische Ergebnisübersichten	145
E.1 Skalierungsinvarianz der Gesichts- und Konturdetektion	145
E.2 Rotationsinvarianz der Personenlokalisierung	146
Literaturverzeichnis	149
Index	167

Abbildungsverzeichnis

1.1	Beispiele für typische Gesten bzw. Posen	5
1.2	Teilautonomes mobiles Servicesystem MILVA	6
1.3	Videokamera VC-C1	6
1.4	Kombination der Cues	9
1.5	Schema des Einbettung der verwendeten Auffälligkeitskomponenten in den Verarbeitungsablauf	11
2.1	Glättungstiefpaß zur spektralen Begrenzung für die Unterabtastung beim Py- ramidenaufbau	15
2.2	Auflösungsstufen der pyramidalen Repräsentation (1)	18
2.3	Auflösungsstufen der pyramidalen Repräsentation (2)	18
3.1	Hautfarb-Cue zur Fusion	21
3.2	Normspektralwertkurven	24
3.3	<i>RGB</i> -Spektralwertkurven nach EBU	25
3.4	Normfarbtafel	26
3.5	<i>RG/BY/WS</i> -Spektralwertkurven nach POMIERSKI /Pom1996/	27
3.6	Hautfarbdatensatz und korrespondierendes Klassifikationsergebnis	29
3.7	Remissionsfunktionen von Körperfarben	30
3.8	Hautfarbklassifikation mit einem Normalverteilungsklassifikator	33
3.9	Beispiele von Ergebnissen der Hautfarbklassifikation im Innen- und Außen- bereich	34

4.1	Gesichts-Cue zur Fusion	37
4.2	Schritte der gesichtsspezifischen Filterung	40
4.3	Beispiele der verwendeten Gesichtsprototypen	41
4.4	Stufen der Gesichtsdetektion (1)	44
4.5	Stufen der Gesichtsdetektion (2)	45
5.1	Kontur-Cue zur Fusion	47
5.2	Prinzip der Konturdetektion zur Lokalisation einer Person anhand des Um- risses ihrer frontalen Kopf-Schulter-Partie	48
5.3	Diskrete parametrische Repräsentation eines Gesichtes nach KOPECZ, J. /Kop1992/	50
5.4	Fitmaps von Haar Wavelet Templates nach OREN ET AL. /OPS ⁺ 1997/ . . .	51
5.5	Schema der prinzipiellen Anordnung der orientierten Filter	52
5.6	Prinzip der Gewinnung der Filterarrangements zur Detektion der Kopf-Schul- ter-Partie	52
5.7	Modellierung des rezeptiven Feldes einer einfachen Zelle mittels 2D-Gabor- funktion	55
5.8	Orientierungsbestimmung mittels Tensormethode (1)	62
5.9	Orientierungsbestimmung mittels Tensormethode (2)	62
5.10	Ergebnisse der Orientierungsbestimmung mittels der Tensormethode	63
5.11	Beispiele von Ergebnissen der konturbasierten Personenlokalisierung im Innen- bereich (1)	63
5.12	Beispiele von Ergebnissen der konturbasierten Lokalisation im Innenbereich (2)	63
5.13	Beispiele von Ergebnissen der konturbasierten Personenlokalisierung im Au- ßenbereich	64
5.14	Beispiele von Ergebnissen der konturbasierten Personenlokalisierung im Innen- bereich (3)	64
6.1	Fusion der Cues	66
6.2	Fusion der drei pyramidal vorliegenden Cues	74

6.3	Sequentialisierung der Relaxationsprozesse	77
6.4	Temporäre Umkodierung der pyramidalen in eine kubische Repräsentation .	78
6.5	Formales dynamisches Neuron in Integratordarstellung	79
6.6	Veranschaulichung der Zusammensetzung der lateralen Aktivierungsfunktion	80
6.7	Zeitlicher Verlauf eines WTA-Prozesses innerhalb eines kubischen 3D dyna- mischen neuronalen Feldes (1)	84
6.8	Zeitlicher Verlauf eines WTA-Prozesses innerhalb eines kubischen 3D dyna- mischen neuronalen Feldes (2)	85
6.9	Lokalisationsergebnis für Versuchsperson Sandy im Außenbereich auf ca. 0,75m	89
6.10	Lokalisationsergebnis für Versuchsperson Autor im Außenbereich auf ca. 1,0m	91
6.11	Lokalisationsergebnis für Versuchsperson Autor im Außenbereich auf ca. 1,0m	92
6.12	Lokalisationsergebnis für Versuchsperson Autor im Außenbereich auf ca. 1,5m	93
6.13	Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (1) auf ca. 1,5m	95
6.14	Lokalisationsergebnis für Versuchsperson Sandy im Innenbereich (1) auf ca. 0,75m	97
6.15	Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 0,5m	99
6.16	Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 1,0m	100
6.17	Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 1,5m	101
6.18	Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 2,0m	102
6.19	Lokalisationsergebnis für Versuchsperson Sandy im Innenbereich (2) auf ca. 1,0m	104
6.20	Lokalisationsergebnis für Versuchsperson Sandy im Innenbereich (2) auf ca. 1,5m	105
B.1	Eigenface-Filterung	126

C.1	CNAPS-Prinzipaufbau: Verschaltung des Arrays aus 128 Prozessoren	133
D.1	Lokale Nachbarschaft mit ideal orientierter Struktur	137
D.2	Vektordarstellung lokaler Orientierung	139

Tabellenverzeichnis

6.1	Lokalisationsergebnisse für Versuchsperson Sandy im Außenbereich	88
6.2	Lokalisationsergebnisse für Versuchsperson Autor im Außenbereich	90
6.3	Lokalisationsergebnisse für Versuchsperson Christoph im Innenbereich (1) . .	94
6.4	Lokalisationsergebnisse für Versuchsperson Sandy im Innenbereich (1)	96
6.5	Lokalisationsergebnisse für Versuchsperson Christoph im Innenbereich (2) . .	98
6.6	Lokalisationsergebnisse für Versuchsperson Sandy im Innenbereich (2)	103
E.1	Skalierungsinvarianz der Gesichtsdetektion für Versuchsperson Christoph . .	146
E.2	Skalierungsinvarianz der Konturdetektion für Versuchsperson Christoph . . .	146
E.3	Rotationsinvarianz der Personenlokalisierung für Versuchspersonen Christoph und Autor	147

Abkürzungsverzeichnis

CCD	C harge C oupled D evice <i>engl.</i> : ladungsgekoppelte integrierte Schaltungsanordnung; Einsatz im Kamerabereich als CCD-Festkörpurbildsensor, <i>siehe auch</i> VC-C1
CIE	C ommission I nternationale de l'Éclairage <i>frz.</i> : Internationale Beleuchtungskommission, <i>siehe auch</i> RGB, XYZ
CNAPS TM	C onnected N etwork of A daptive P rocessors <i>engl.</i> : Hardwarekonzept in SIMD-Technologie der Firma Adaptive Solutions Inc., Beaverton (Oregon), USA, <i>siehe auch</i> CPL, SIMD
CPL	C NAPS P rogramming L anguage <i>engl.</i> : Assemblersprache zur hardwarenahen CNAPS-Programmierung, <i>siehe auch</i> CNAPS, PN
DPR	D iskrete p arametrische R epräsentation: in Form verteilter Orientierungshistogramme, die auf einem regulären zweidimensionalen Raster angeordnet sind, erstellte Kodierung natürlicher Bilder nach KOPECZ, J. /Kop1992/
DSP	D igital S ignal P rocessor <i>engl.</i> : digitaler (spezialisierter) Sinalprozessor
EBU	E uropean B roadcasting U nion <i>engl.</i> : Union der Europäischen Rundfunkanstalten, <i>siehe auch</i> RGB
FCC	F ederal C ommunication C ommission <i>engl.</i> : Bundeskommunikationskommission der USA, <i>siehe auch</i> RGB
FFT	F ast F ourier T ransform <i>engl.</i> : schnelle Fouriertransformation, die aus der diskreten Fouriertransformation abgeleitet werden kann; wenn als Bedingung eingehalten wird, daß die Anzahl der Datenpunkte eine Zweierpotenz ist, ergeben sich Vereinfachungen, die die Komplexität des Berechnungsaufwands von $O(n^2)$ auf $O(n \lg n)$ verringern, <i>siehe auch</i> LOCHER /Loc1992/
GNG	G rowing N eural G as <i>engl.</i> : künstliches neuronales Netzwerk nach FRITZKE /Fri1995/ zur adaptiven inkrementellen Vektorquantisierung, <i>siehe auch</i> KNN

HMM	Hidden Markov Model <i>engl.</i> : endlicher, stochastischer und trainierbarer Automat mit Zuständen und Zustandsübergängen (<i>siehe z. B.</i> BAUM /Bau1972/); den Zuständen zugeordnet sind frei zugängliche Wahrscheinlichkeiten für Ausgabesymbole; Einsatz bisher meist in der Sprachverarbeitung (Phonem-, Wortmodelle).
KLT	KARHUNEN-LOÈVE-Transformation : ein nach den beteiligten Autoren KARHUNEN /Kar1984/ und Loève /Loè1977/ benanntes lineares Verfahren zur Dimensionsreduktion in hochdimensionalen Signalräumen mittels Bestimmung der Eigenvektoren der Kovarianzmatrix, anschließender Auswahl einer Untermenge der eigenwertgrößten Eigenvektoren und einzelner Skalarproduktbildung des Datensatzes mit jedem Eigenvektor aus der gebildeten Untermenge an Eigenvektoren, <i>siehe auch</i> PCA
KNN	Künstliches Neuronales Netz : Netzwerk aus formalen Neuronen (die in den Anfängen von MCCULLOCH und PITTS /MP1943/ entworfen wurden), deren adaptive Verbindungen mittels Lernregeln (wie z. B. bereits kurz darauf von HEBB /Heb1949/ vorgeschlagen) entsprechend der Netzwerkarchitektur und des verwendeten Lernparadigmas, typischerweise während einer Trainingsphase, modifiziert werden können. ROSENBLATTS <i>Perceptrons</i> /Ros1962a, Ros1962b/ brachten die ersten beachtenswerten Verarbeitungsleistungen von KNNs in der Musterklassifikation. Es existiert mittlerweile eine Fülle verschiedener Architekturen und Lernparadigmen, <i>siehe auch</i> MLP, GNG, LLM, RBF.
LLM	Local Linear Map <i>engl.</i> : lokal lineare Abbildung; künstliches neuronales Netz zur Vektorquantisierung, das auf RITTER, H. /Rit1991/ zurückgeht, <i>siehe auch</i> KNN
LUT	Look-Up Table <i>engl.</i> : Verweistabelle; Soft- oder Hardwarestruktur zur diskreten Funktionsapproximation
MILVA	Multisensorielles intelligentes lernfähiges Vehikel in neuronaler Architektur : autonome mobile Roboterplattform am FG Neuroinformatik der Fakultät für Informatik und Automatisierung an der TU Ilmenau
MLP	Multi Layer Perceptron <i>engl.</i> : vorwärtsgekoppeltes lernfähiges Mehrebenen-Netzwerk, typischerweise mittels Error-Backpropagation (BP) trainiert, das auf WERBOS /Wer1974/ zurückgeht, <i>siehe auch</i> KNN.
MMI	Mensch-Maschine-Interaktion : interdisziplinäres Forschungsthema zur Entwicklung neurartiger (ergonomischerer) Schnittstellen zwischen Anwendern

und technischen Systemen, das sich gegenwärtig mit großer Intensität entwickelt; wird insbesondere als Einsatz- und Bewährungsfeld lernfähiger Systeme angesehen

MMX	M ultimedia E xtension <i>engl.</i> : für multimediale Anwendungen bestimmter erweiterter Befehlssatz von Pentium [®] -Prozessoren der Firma Intel
NOA	N eighborhood O perational A ccelerator <i>engl.</i> : anwendungsspezifizierter integrierter Schaltkreis der Firma Matrox, zugeschnitten auf lokale Bildverarbeitungsroutinen, wie z. B. Faltungen
ORL	O livetti R esearch L aboratory <i>engl.</i> : Forschungsabteilung der Firma Olivetti in Cambridge (UK), die mittlerweile von AT&T geführt wird. Bekannt durch die umfangreiche Datenbank <i>ORL Database of Faces</i> , die unter http://www.cam-orl.co.uk/facedatabase.html zugänglich ist.
PCA	P rincipal C omponent A nalysis <i>engl.</i> : lineares Verfahren zur Bestimmung der eigenwertgrößten Eigenvektoren der Kovarianzmatrix, <i>siehe auch</i> KLT
PCI	P eripheral C omponent I nterconnect <i>engl.</i> : Standard der Firma Intel für eine direkte Adreß- und Datenverbindung, insbesondere zwischen Prozessor und Einsteckplätzen, nach dem sog. Local-Bus-Konzept
PN	P rocessing N ode <i>engl.</i> : Einzelprozessor innerhalb der CNAPS-Architektur, <i>siehe auch</i> CNAPS, SIMD, CPL
RBF	R adial B asis F unction <i>engl.</i> : spezielle „glockenförmige“ Ausgabefunktion formaler Neuronen (meist) in Hidden-Schichten vorwärtsgekoppelter KNN (RBF-Netze, <i>siehe z. B.</i> POGGIO und GIROSI /PG1990/); werden typischerweise als univariate Gaußfunktionen implementiert, <i>siehe auch</i> KNN.
RF	R ezeptives F eld: neurobiologischer Begriff, der den Bereich eines Sinnesepithels, von dem aus ein bestimmtes Neuron im zugehörigen afferenten (hinleitenden) System in seiner Aktivität beeinflusst werden kann, bezeichnet (<i>nach</i> WOLF /Wol1988/); Verwendung auch bei Modellen neuronaler Strukturen und bei künstlichen neuronalen Netzen, <i>siehe auch</i> KNN.
<i>RG/BY/WS</i>	R ot- G rün, B lau- Y ellow (Gelb), W eiß- S chwarz: Dimensionen des physiologischen Elementarfarbraumes im Gegenfarbprinzip nach POMIERSKI /Pom1996/
<i>RGB</i>	R ot, G rün, B lau: Dimensionen eines durch diese drei Primärvalenzen aufgespannten kartesischen Farbraumes; dabei ist grundsätzlich zu unterscheiden zwischen <i>RGB</i> nach CIE, <i>RGB</i> _{FCC} nach FCC und <i>RGB</i> _{EBU} nach EBU sowie ggf. weiteren Spezifikationen, <i>siehe auch</i> CIE, FCC, EBU

SIMD	Single Instruction, Multiple Data <i>engl.</i> : Einzelbefehl wirkt gleichzeitig auf die auf ein Prozessorfeld verteilten (zerlegten) Daten, <i>siehe auch</i> CNAPS
VC-C1	Video Conferencing Camera 1 <i>engl.</i> : $\frac{1}{3}$ -Zoll CCD-Videokamera der Firma Canon mit integrierter Schwenk-Neige-Einheit, <i>siehe auch</i> CCD
WTA	Winner-Take-All <i>engl.</i> : explizites oder implizites Selektionsverfahren mit dem Ziel, aus konkurrierenden Hypothesen die optimale auszuwählen.
XYZ	von der CIE als <i>die</i> umfassende und grundlegende Referenz (Normvalenzsystem) 1931 aufgestellter Farbraum, der zur Gruppe der Primärvalenzsysteme gehört; die zugrunde liegenden Messungen im Winkelbereich von 2° beziehen sich auf ein Testkollektiv an Personen, die Ergebnisse wurden durch Mittelung gebildet, so daß sie sich auf einen angenommenen 2°-Normalbeobachter beziehen, <i>siehe auch</i> CIE

Kapitel 1

Motivation

1.1 Einleitung

Die Entwicklung von technischen Systemen mit neuartigen Schnittstellen zur Mensch-Maschine-Interaktion (MMI) wird international vorangetrieben. Vor diesem Hintergrund behandelt die vorliegende Arbeit unter dem Titel „Multi-Cue¹-Ansatz für ein dynamisches Auffälligkeitssystem zur visuellen Personenlokalisierung“ ein Thema, das an einem Kreuzungspunkt verschiedener Disziplinen und Forschungszweige angesiedelt ist. Direkt einzubeziehen sind Bildverarbeitung und Neuroinformatik; indirekt spielen Mobile Robotik, Neurophysiologie und Arbeitswissenschaften eine Rolle. Konkreter Gegenstand dieser Dissertation sind *Konzeption* und *Umsetzung* eines neuen Mehrkomponentenansatzes für ein visuelles Auffälligkeitssystem zur *Detektion* und *Lokalisation* von Personen. Es bildet einen Baustein für ein Gesamtkonzept zur gestenbasierten Mensch-Maschine-Interaktion mit dem teilautonomen Servicerobotersystem MILVA, auf dem unter anderem neuartige robuste Schnittstellen zur MMI entwickelt werden. Neuartige Schnittstellen in der MMI zeichnet aus:

- Kommunikations- und Steuerungsmöglichkeiten ohne hohe Präzisionsanforderungen über natürliche menschliche Ausdrucksformen wie Sprache, Mimik, Gestik oder manuelle Eingriffe,
- umfassende Berücksichtigung ergonomischer Aspekte,
- Intuitivität der Benutzung bei Wegfall speziell zu erlernender zugeschnittener Kommandos.

¹Cue: *engl.* Anhaltspunkt, Stichwort; im Rahmen dieser Arbeit im Sinne von Auffälligkeitskomponente benutzt, die für das dynamische visuelle Auffälligkeitssystem verwendet wird.

Angestrebtes Ziel ist dabei, insgesamt eine neue Qualität der Mensch-Maschine-Interaktion zu erlangen. Dafür haben sich potentielle Anwendungsfelder der MMI neu entwickelt oder sind in Reichweite, die so in der Vergangenheit kaum eine Rolle gespielt haben und auf die sich viel Forschungsarbeit konzentriert. Neuere Arbeiten, z. B. vorgestellt von MAGGIONI /Mag1995/, KORTENKAMP ET AL. /KHB1996/ oder RAUTERBERG ET AL. /RBLM1997/ auf dem Gebiet der MMI streben in vielen Fällen die Ablösung herkömmlicher Eingabegeräte wie Tastatur, Maus und Datenhandschuh an. Andere Arbeiten haben sich völlig neue Schnittstellenarten zur intuitiven Kommunikation (verbal und non-verbal) von Nicht-Experten mit intelligenten Systemen zum Ziel gesetzt, in denen häufig gesten- oder posturbaasierte Interaktionsformen eine wesentliche Rolle spielen. Während darin geschultes Personal Semantik und Syntax eines speziellen Instruktionssatzes erlernt und somit *speziell zugeschnittene Bedienungsfertigkeiten* entwickelt hat, ist *ein* Entwicklungsziel der MMI, einen beschränkten, aber universellen Instruktionssatz zu realisieren, der sich aus Ausdrucksformen zusammensetzt, die in der Mensch-Mensch-Interaktion üblicherweise eine Rolle spielen. Eine Voraussetzung ist, daß es das jeweilige Anwendungsgebiet zuläßt, in Kontakt tretende Personen *intuitiv* an die Interaktionsaufgabe herangehen zu lassen. Es ist also erforderlich, daß die Interaktionsaufgabe a priori *ohne* vertiefte Systemkenntnisse (d. h. ohne beherrschtes Befehlsalphabet) lösbar ist, wodurch erst eine intuitive Herangehensweise sinnvoll möglich ist. PENTLAND /Pen1996a, Pen1996b/ stellte dazu ein umfassendes MMI-Konzept vor, dessen Kernstück die sog. *Smart Rooms* bilden. Ähnlich einzuordnen sind z. B. Ansätze von ZELINSKY und HEINZMANN /ZH1996/, BARTLETT ET AL. /BVS⁺1996/ oder HERPERS ET AL. /HKRS1995/, die mittels Verfahren zur Mimikierfassung und -bewertung ganz wesentlich über etablierte Wege in der MMI hinausgehen und damit ebenfalls neue Interaktionsformen anstreben sowie dafür Lösungswege vorschlagen.

Neben den genannten Prämissen für neuartige Schnittstellen der MMI gibt es vielfältige weitere Motivationen und Anstöße für diese Entwicklungen. Einerseits trägt die immer weiter fortschreitende Steigerung der Leistungsparameter digitaler Rechnersysteme (Spezial- und insbesondere Universal-Hardware) wesentlich dazu bei, daß zunehmend eine Verarbeitung der hohen anfallenden Datenmengen (z. B. Bildfolgen) direkt am Prozeß und mit zum Teil immer aufwendigeren Algorithmen möglich wird. Diese Leistungsentwicklung wird in den kommenden Jahren andauern und damit weiter die Möglichkeiten der MMI ausweiten, zumal der Preisfaktor durch Verbilligung der Hardware (z. B. bei Bildaufnahmetechnik) immer weiter in den Hintergrund tritt, so daß etwa in Kürze Miniaturkameras zur Standardperipherie von Heim-PCs zu zählen sein werden.

Weiterhin wird die MMI als Herausforderung z. B. für den Einsatz und die Bewährung lernfähiger Systeme gesehen. Damit wird das Themenfeld MMI erheblich erweitert. Während in der Vergangenheit unter dem Vorzeichen MMI Arbeitswissenschaftler etwa die Ergono-

mie herkömmlicher Verkaufsautomaten untersuchten, arbeitet heute ein breites Spektrum an Entwicklern verschiedener Disziplinen an experimentellen neuartigen Robotersystemen oder Nutzerschnittstellen für Computer. Im Mittelpunkt dieser Arbeiten steht, inwieweit die algorithmisch-technischen Voraussetzungen für den weiteren Verzicht auf speziell zugeschnittene Eingabegeräte für Nutzer des jeweiligen Systems gegeben sind und welche Wege mit welchen Methoden auf welchen Anwendungsfeldern dafür gegenwärtig und in Zukunft offenstehen.

Dafür verdient neben leistungsfähiger Hardware insbesondere die Verfügbarkeit und Entwicklung geeigneter Algorithmen eine besondere Beachtung, denn bei den allermeisten der zuletzt aufgezählten Arbeiten zur MMI, wie auch der vorliegenden, spielt die Verarbeitung optischer (visueller) Informationen mit entsprechendem hohen Datenanfall eine wichtige Rolle. Der Grund liegt schon allein darin, daß diese für die Analyse von gestural-habituellen Ausdrucksformen vor allen anderen Informationsmodalitäten geeignet sind. Methoden der digitalen Bildverarbeitung kommt es demzufolge zu, für die benötigten Auffälligkeitskomponenten die Basis zu schaffen. Maßstäbe für ihre Eignung sollten grundsätzlich Leistungsfähigkeit und Robustheit bei wechselnden Bedingungen ebenso wie das Verhalten unter den bei der Realisierung zu erwartenden Kompromissen (z. B. beschränkte Auflösung) sein. Dazu gehört außerdem die Effizienz der implementatorischen Umsetzbarkeit. Nur so läßt sich bereits im Entwicklungsstadium eine akzeptable Demonstrabilität erzielen, ohne die entsprechende experimentelle Untersuchungen auf der Zielplattform unmöglich sind. So erfüllt bspw. das in dieser Arbeit benutzte elegante Verfahren zur Orientierungsfilterung nach JÄHNE /Jäh1998/ (*siehe Abschnitt 5.4.2 Nichtlineare lokale Strukturanalyse zur Orientierungsbestimmung*) ganz besonders die genannten Eignungskriterien.

Hervorzuheben ist darüber hinaus, daß für einen späteren praktischen Einsatz die *Demonstrabilität unter Echtzeitbedingungen* einen hohen Stellenwert einnehmen muß. Dies muß Konsequenzen für die zulässige Berechnungskomplexität der Algorithmen und Implementationen auf der real verfügbaren Rechentechnik haben. Daher sind Wege (Cues, Algorithmen) zu suchen, die in der zur Verfügung stehenden Zeit bei zufriedenstellender Leistungsfähigkeit die benötigten Ergebnisse liefern können. Hier gilt es, eine Balance zwischen neuen, methodisch anspruchsvollen Ansätzen und tatsächlicher Realisierbarkeit² zu halten. Die vorliegende Arbeit versucht daher ganz bewußt, mit einer konsequent ingenieurtechnischen Herangehensweise unter Nutzung leistungsfähiger (z. T. bekannter) Konzepte diese durch Verknüpfung zu einer robusten Systemlösung zu führen.

²Daraus werden sich Konsequenzen für die erzielbare Genauigkeit der Lösung ergeben, so daß etwa der Einsatzbereich des Systems noch bestimmte Einschränkungen haben kann, was die Lösung selbst jedoch nicht in Frage stellt.

1.2 Szenario für ein teilautonomes mobiles Servicesystem

Mit einem mobilem Robotersystem, wie es in Form des Roboters MILVA als Experimentalplattform zur Verfügung steht (*siehe* Abb. 1.2), sind die geräteseitigen Voraussetzungen zur Entwicklung und prototypischen Realisierung eines teilautonomen mobilen Servicesystems gegeben. Entwicklungsziel dieser Experimentalplattform ist ein System mit Ansätzen zu intelligenten Verhaltensweisen in diesem Bereich. Gerade hier liegt ein interessantes künftiges Anwendungsfeld derartiger teilautonomer mobiler Systeme, durch welche verschiedene Aufgaben übernommen werden können. Neben Transportaufgaben bietet sich zugleich eine Funktion als mobiles Informationsterminal an, das zusätzlich eine Lotsenfunktion ausüben kann.

Als konkretes Einsatzfeld, in dem alle diese drei Aspekte direkt oder indirekt Berücksichtigung finden, kommt ein Einsatz in einem Baumarkt in Frage. Dort besteht sowohl Bedarf an Unterstützung bei Transporten als auch an Orientierungshilfe in den weitläufigen Regalbereichen, und nebenbei kann der Roboter auch als Werbeträger dienen. Nimmt man die Transportfunktion, so bieten sich gerade gestenbasierte Interaktionsformen an, indem die geeignete Positionierung des Roboters mittels eines elementaren Satzes an Befehlen gesteuert wird. Abb. 1.1 stellt beispielhaft eine Auswahl solcher einfachen Gesten vor, die hier in Form statischer Posen ausgedrückt werden. Unter der Vorgabe, keine Bewegungsabläufe einzubeziehen, können diese Posen als intuitiv angesehen werden. Es ist jedoch klar, daß dynamischen Gesten (Bewegungsabläufe) eine viel größere Ausdrucksvielfalt bieten. Sie stellen jedoch erheblich höhere Ansprüche an die Verarbeitung. Die Erkennung derartiger Gesten zum Ziel hat eine Arbeit von CORRADINI/Cor2000/, der mit einem hybriden Ansatz (HMMs in Verbindung mit RBF-Netzen oder rekursiven KNN) zwischen sechs³ dynamischen Gesten unterscheiden kann – sofern zuvor die betreffende Person erfolgreich lokalisiert worden ist.

Eine wesentliche Bedingung für die gestenbasierte Interaktion muß also die Fähigkeit des teilautonomen mobilen Systems zur robusten Lokalisation eines Nutzers und zum Kontakthalten mit diesem sein. Beim Kontakthalten, während z. B. ein Kunde zu einem bestimmten Marktbereich begleitet wird, ist insbesondere ein ständiges Lokalisieren notwendig. In dieser Weise ordnet sich die vorliegende Arbeit zu einem robusten Ansatz für ein visuell-basiertes Personenlokalisationsverfahren in ein Szenario für ein mobiles Servicesystem ein.

³mittlerweile erweitert auf 13



Abbildung 1.1: Beispiele für Szenen, in denen typische Gesten (Posen) von Personen zur Interaktion mit einem mobilen Roboter ausgeführt bzw. eingenommen werden (v. o. l. n. u. r.): Achtung/Hallo!, Halt!, Fahre nach vorn links!, Fahre nach vorn rechts!, Biege nach links ab!, Biege nach rechts ab!

1.3 Visuell-basierte Personenlokalisierung

Die *Lokalisation von Personen* bildet eine elementare Voraussetzung für jedwede Art der Interaktion von Personen und Servicerobotersystemen. Ganz besonders trifft dies für ein mobiles Robotersystem zu, das teilautonom agiert und mit Personen in Interaktion treten können soll (*siehe z. B. FEYRER und ZELL /FZ1999/*). Derartige Problemstellungen treten neben mobilen Robotersystemen aber ebenso bei anderen Real-World-Anwendungen von Mensch-Maschine-Schnittstellen auf, wie z. B. bei den bereits erwähnten „intelligenten Zimmern“ (*siehe PENTLAND /Pen1996b/*), wo die betreffende Person zunächst im Raum genau lokalisiert werden muß. Um in diesen und weiteren möglichen Szenarien immer weniger Einschränkungen (vereinfachte Randbedingungen u. ä.) in Kauf nehmen zu müssen, ist eine stärkere Konzentration auf das Thema „visuelle Personenlokalisierung“ von großer Bedeutung. Diese Arbeit unternimmt den Versuch, sich konsequent dieser Fragestellung zu stellen, da deren Lösung elementare Voraussetzung für die genannten Szenarien ist. Daher lautet das explizite Ziel dieser Arbeit, jeweils *diejenige* Bildkoordinate zu bestimmen, für die die größte Wahrscheinlichkeit besteht, an dieser richtigerweise eine Person zu lokalisie-



Abbildung 1.2: Das teilautonome mobile Servicesystem MILVA: der auf einem dreirädrigen Fahrwerk aufbauende Roboter mit elektrischem Vorderradantrieb besitzt eine umfangreiche Ausrüstung an Sensortechnik. Dazu zählen insbesondere die beiden oben montierten zoomfähigen S-Video-Kameras, die durch die Montage auf Schwenk-Neige-Einheiten zusammen sieben Freiheitsgrade besitzen. Beide Kameras werden gerade für die Interaktion mit Personen relevant. Die eine der Kameras dient zur Lokalisation von Personen, die andere zur Gestenanalyse. Eine dritte Kamera am Roboterbug dient zur visuell-basierten Navigation. Weitere für die vorliegenden Arbeit allerdings nicht relevante Sensorik ist in Form eines Laserscanners, zweier Infrarotsensoren, eines Ultraschallsensors, taktiler Sensorik sowie zusätzlich durch ein binaurales Hörsystem vorhanden.

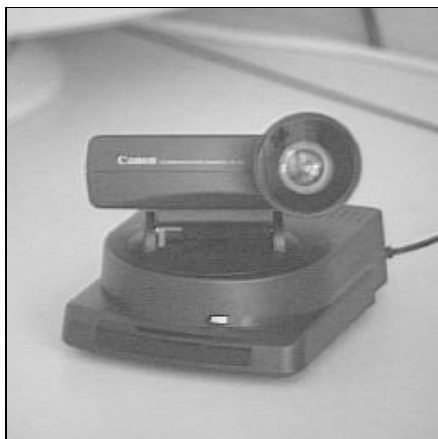


Abbildung 1.3: Um bei den Untersuchungen dieser Arbeit nicht direkt auf die Verfügbarkeit der Roboterplattform angewiesen zu sein, wurde als Vorstufe dazu ein flexibel einsetzbarer Computervision-Arbeitsplatz mit einer S-Video-Kamera VC-C1 (mit integrierter Schwenk-/Neigemechanik) eingesetzt. Prinzipielle Einschränkungen oder Vereinfachungen bringt diese Kamera nicht. Genau dieser Kameratyp wurde bereits von GOODRIDGE /Goo1997/ für dessen Untersuchungen zur visuo-akustischen Sensorfusion zur Mensch-Maschine-Interaktion ebenfalls mit Erfolg eingesetzt.

ren. Das potentielle Problem, zwischen alternativen Personen *definiert* zu entscheiden, um damit etwa nicht interaktionsberechtigte Personen auszuschließen, wird im Rahmen dieser Arbeit ausgeklammert.

Die Personenlokalisierung soll in Kamerabildern erfolgen, die das Umfeld des Roboters in

Übersichtsaufnahmen darstellen. Mittels sogenannter *pyramidaler Ansätze* aus der digitalen Bildsignalverarbeitung (*siehe* ROSENFELD und JOLION /Ros1984,JR1994/) kann aus einem solchen Bild eine Multiskalen-Repräsentation direkt im Ortsbereich erzeugt werden (*siehe* Kapitel 2 *Pyramidale Repräsentationen als Basiskonzept*). Diese erlaubt, die Auffälligkeitskomponenten auf verschiedenen Skalierungsstufen anzuwenden und damit ein bestimmtes Intervall eines Entfernungsbereichs einer Person abzudecken, wobei die Streuung der Größe von Personen hierbei als vernachlässigbar klein angesehen werden soll.

Aus Gründen der rechentechnischen Beherrschbarkeit wurde für diese Arbeit eine generelle Beschränkung der Bildauflösungen vorgenommen, jedoch aus Gründen des Platzbedarfs (insbesondere bei den Untersuchungen im Innenraum) zugleich die minimale Objektivbrennweite der Kamera eingestellt (ergibt maximale Öffnungswinkel von ca. 45° in Bildhöhe und 55° in Bildbreite). Der resultierende entfernungsmaßige Arbeitsbereich liegt daher nur zwischen 0,5m und 2,0m, was jedoch keine konzeptionelle Einschränkung darstellt. Für diesen Entfernungsbereich wurde eine pyramidale Abstufung im Halboktavabstand gewählt. Es ergeben sich daher fünf Auflösungsstufen. Auf diese Multiskalen-Repräsentation werden Algorithmen angewendet, die möglichst spezifisch auf Personenmerkmale eingestellt sind. Für jedes der visuellen Merkmale kann damit pro Bildkoordinate ein bestimmtes Auffälligkeitsmaß und damit eine eigene Auffälligkeitskarte gebildet werden.

Die Problemstellung dieser Arbeit ist insofern szenariospezifisch vereinfacht, als daß die Personen in den Bildern *aufrecht* und *frontal zur Kamera gewandt* angenommen werden. Die erste Annahme bzw. Einschränkung mag an sich plausibel sein. Die zweite kommt daher, daß davon ausgegangen werden kann, daß Personen ihre Interaktionsbereitschaft dadurch zum Ausdruck bringen, daß sie sich bei Interaktionsbedarf frontal zum Roboter gerichtet aufstellen und dabei das Gesicht ebenfalls frontal ausrichten. Diese Annahme erscheint zulässig, da auch Personen untereinander sich intuitiv frontal zu anderen Personen (Zielpersonen) ausrichten (und ggf. stehenbleiben oder auf die Zielperson zugehen), um leichter registriert zu werden. Deshalb stellen auch Oberkörper und davon speziell das Gesicht *die* Körperanteile dar, die am wahrscheinlichsten im Blickfeld sind, wenn es in der Mensch-Maschine-Interaktion um Personenlokalisierung geht. GOODRIDGE, der eine Arbeit /Goo1997/ zur audiovisuellen Sensorfusion in der Mensch-Maschine-Interaktion vorgelegt hat, stellt in ähnlicher Weise sowohl Oberkörper und Gesicht als die am meisten relevanten Körperpartien heraus, einfach weil das Blickfeld der Maschine prinzipiell dem von Personen ähnlich ist. Die frontale Ausrichtung bedeutet, daß praktisch nur nach Personen in *einer* typischen Ansicht gesucht werden muß. Personen in dieser frontalen Ansicht

- haben im Gesicht praktisch immer sichtbare Hautpartien; damit sind hautfarbbasierte Auffälligkeitsmechanismen einsetzbar,

- richten das Gesicht fast immer auch frontal aus, da der Blick überwiegend zum Roboter gerichtet ist; deshalb sind relativ einfache Gesichtsdetektionsverfahren einsetzbar,
- bilden insbesondere im Bereich ihres Oberkörpers eine typische Kontur gegenüber dem Hintergrund aus; die von Kopf und Schultern ausgebildeten Kontur bildet ebenso eine weitere Komponente für ein Auffälligkeitsmaß zur Lokalisation von Personen.

Aus der Vielzahl von Arbeiten, die zum Thema *visuelle Personenlokalisierung* verfügbar sind, soll auf eine Auswahl verwiesen werden, bei der deutliche Bezüge zum Gegenstand dieser Arbeit bestehen. WREN, AZARBAYEJANI, DARRELL und PENTLAND /WADP1997/ stellen ein System zur Personenlokalisierung vor, das sie als *Person Finder*, kurz *Pfinder*, bezeichnen. Dabei wird ein statistisches Mehrklassenmodell verwendet, mit dem personen-typische Farbe und Form für Personenbestandteile repräsentiert werden. Mit diesem wird versucht, *die* Bildkoordinaten zu finden, an denen das Vorhandensein einer Person am besten zu verifizieren ist. Das Modell besteht im verwendeten Merkmalsraum aus einer Menge an Hyperellipsen, die räumlich untereinander verbunden sind, so daß durch Resynthese mittels der Hyperellipsen die gesamte erfaßte Haltung der Person schließlich sichtbar gemacht werden kann, womit es klar über die Lokalisation hinausgeht. Nachteil ist, daß das System indirekt von einem deutlich weniger dynamischen Hintergrund ausgeht. DARRELL ET AL. /DGWH1998/ ebenso wie NASTAR und MITSCHKE /NM1998/ haben deshalb eine Anwendung vorgestellt, bei dem ein unbewegliches Sehsystem eingesetzt wird.

Andere Arbeiten stellen klar Gesichtsmuster als Basis zur Personenlokalisierung in den Mittelpunkt – ein Vorgehen, das gerechtfertigt ist, sofern gesichert ist, daß die erforderlichen Gesichtsdetails tatsächlich immer erfaßt werden können. Beispielsweise TERRILLON, DAVID und AKAMATSU /TDA1998/ wie auch STEFFENS, ELAGIN und NEVEN /SEN1998/ nutzen verschiedene leistungsfähige gesichtssensitive Verfahren zur Gesichtsdetektion und lokalisieren ausschließlich damit Personen.

Einen demgegenüber ganz anderen Zugang haben OREN, PAPAGEORGIOU und POGGIO /OPS⁺1997, POP1998/ gewählt. Sie beziehen sich bei der Lokalisation von Fußgängern auf den gesamten äußeren Umriß der Personen. Wohingegen die ersten beiden der genannten Wege schon als traditionell bezeichnet werden können, sind konturbasierte Ansätze bisher kaum zu finden gewesen. Diese Arbeiten sind auch deshalb interessant, weil unabhängig davon für die vorliegende Arbeit ebenfalls ein konturbasierter Ansatz entwickelt wurde, der in Kapitel 5 *Konturdetektion* dargestellt wird.

RAJA, MCKENNA und GONG /RMG1998/ andererseits nutzen derart genaue adaptive Farbmodelle von Personen (bes. Kleidung), daß diese unter bestimmten Bedingungen vollkommen ausreichen, bestimmte Personen zu lokalisieren. Bei dieser Herangehensweise bedarf es einer Adaptionphase, für die die jeweilige Person klar segmentiert vorliegen muß.

Weiterhin muß gesichert sein, daß angepaßt an die Genauigkeit der Farbmodelle auch die Beleuchtungsbedingungen sehr konstant gehalten werden können. Letzteres stellt eine erhebliche Einschränkung dar.

Die drei visuellen Auffälligkeitskomponenten (Hautfarbe, Gesicht und Kontur), die für diese Arbeit alle aufgegriffen werden (*siehe* Abb. 1.4), sind für sich genommen nur recht begrenzt erfolgreich mit robusten Ergebnissen einsetzbar. Erst im scheinbar besonders aufwendigen Zusammenwirken aller können Beschränkungen der einzelnen Komponenten überschritten werden, ohne daß der Be- und Verarbeitungsaufwand jeder Einzelkomponente besonders angehoben werden müßte. Die Idee dabei ist, daß ein Merkmalsverbund sich letztlich weniger aufwendig und zugleich leistungsfähiger erweisen kann, als ein einzelnes Merkmal. Für dieses eine Merkmal müßte eine sehr hohe Konstanzleistung erzielt werden, damit es unter allen vorkommenden Umgebungsbedingungen erfaßt werden könnte. Dadurch würde man u. U. erhebliche technische Anforderungen bekommen (etwa Kameratechnik mit erhöhtem Dynamikumfang) und trotzdem etwa bei extremen Kontrasten im Gegenlicht Probleme bekommen. Der Vorteil einer Fusion liegt also in einer Vereinfachung und Senkung des einzelnen algorithmischen Aufwandes, da der Bedarf an Konstanzleistung reduziert ist. Mit einem Mehrkomponentenansatz kann ein größerer Bereich unterschiedlicher Bedingungen a priori abgedeckt werden.

LEE und SEUNG /LS1998/ setzen zur Fusion personentypischer Farb- und Bewegungsmuster ein neuronales Netz ein, mit dem sie die Fusion als optimierte Klassifikation ausführen, ein Ansatz, der in ihrer Realisierung konkret davon profitiert, daß sie sehr große Datenmengen erzeugen können, da die Trainingsdaten online mit Hilfe eines auf die Stirn der Personen geklebten Infrarotreflektors erzeugt werden.

In der vorliegenden Arbeit wird dagegen mit einem dynamischen Fusionsansatz unter Nutzung dynamischer neuronaler Felder nach AMARI /Ama1977, KA1979/ der Effekt der dynamischen Selektion von Erregungsmustern ausgenutzt, mit dem räumliche Koinzidenzen in einem iterativen Prozeß zu einer globalen Selektionsentscheidung geführt werden können.

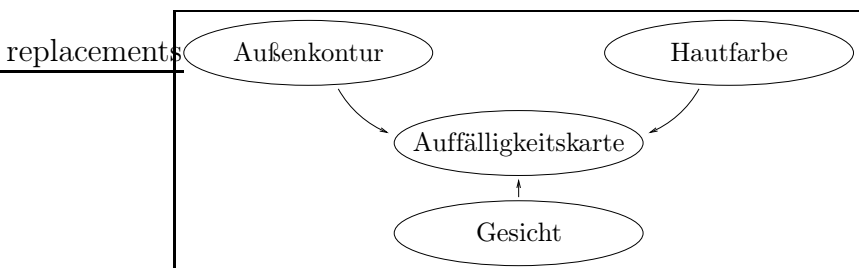


Abbildung 1.4: Schema der Kombination der in dieser Arbeit verwendeten Cues zu einer fusionierten Auffälligkeitskarte.

Es gehört zum Wesen der Lokalisationsproblematik, daß a priori keine Vorselektion im Bild einer visuellen Szene vorgenommen werden kann. Damit ist im Prinzip eine permanente *vollparallele Verarbeitung* nötig. Andererseits besteht kein permanenter Bedarf für eine

Personen- oder Nutzerlokalisierung durch das Robotersystem, vorausgesetzt, daß eine einmal lokalisierte Person tatsächlich mit dem Roboter in Interaktion tritt. Dadurch ist der Leistungsbedarf zur Lokalisation mindestens im Mittel *relativ beschränkt*, was allerdings eine Forderung nach möglichst effizienten Algorithmen keineswegs erübrigt.

1.4 Résumé

Als Ergebnis der Entwicklung eines neuen Mehrkomponentenansatzes für ein visuelles Auffälligkeitssystem zur Personenlokalisierung behandelt diese Arbeit sowohl die Bildung der Einzelkomponenten (Cues) als auch deren dynamische Fusion. In den Entwurf des Ansatzes gehen eine Reihe von Aspekten ein, die einmal die Einbettung der Arbeiten in das geschilderte Gesamtszenario sowie generelle, auf die konkrete Umsetzbarkeit ausgerichtete Überlegungen berücksichtigen. Das schlägt sich nieder in der Wahl der beteiligten Cues (Hautfarbe, Gesicht und Außenkontur) in einer erstmals so vorgenommenen Zusammenstellung. Gegenübergestellt werden andere Ansätze zur Personenlokalisierung aus der Literatur, deren Methoden zum Teil aufgegriffen und die mit Blick auf die vorhandene Hardware aufbereitet werden. Als Erweiterung des Methodenspektrums wird ein konturbasierter Cue genutzt, der unter den im Szenario gegebenen Randbedingungen besonders vorteilhaft einsetzbar ist. Darüber hinaus wird mit der dynamischen Fusion der Cues eine Lösung vorgestellt, bei der in einem kompetitiven Selektionsprozeß eine Lokalisationsentscheidung getroffen wird. Abb. 1.5 zeigt im Schema sowohl die Konzeption dieser Arbeit als auch technische Komponenten für deren Realisierung.

Die weiteren Kapitel der Arbeit stellen die Cues im einzelnen vor. Vorangestellt wird mit dem Kapitel 2 *Pyramidale Repräsentationen als Basiskonzept* eine kurze Darstellung der für alle eingesetzten Cues genutzten Methode der Vor- und Weiterverarbeitung des Bildmaterials. Darauf folgt Kapitel 3 *Hautfarbdetektion*, an das sich die Kapitel 4 und 5 *Gesichtsdetektion* und *Konturdetektion* anschließen. Die Ergebnisse dieser drei Kapitel werden danach im Kapitel 6 *Personenlokalisierung* aufgegriffen und darin zu einer Gesamtlösung in Form eines dynamischen Auffälligkeitssystems zur visuellen Personenlokalisierung zusammengeführt. Darin eingeschlossen sind eine Reihe von Lokalisierungsergebnissen in verschiedenen Beispielszenen. Kapitel 7 *Zusammenfassung und Ausblick* schließt die Arbeit ab. Verweise auf alle Anhänge finden sich in den jeweiligen Kapiteln.

placements

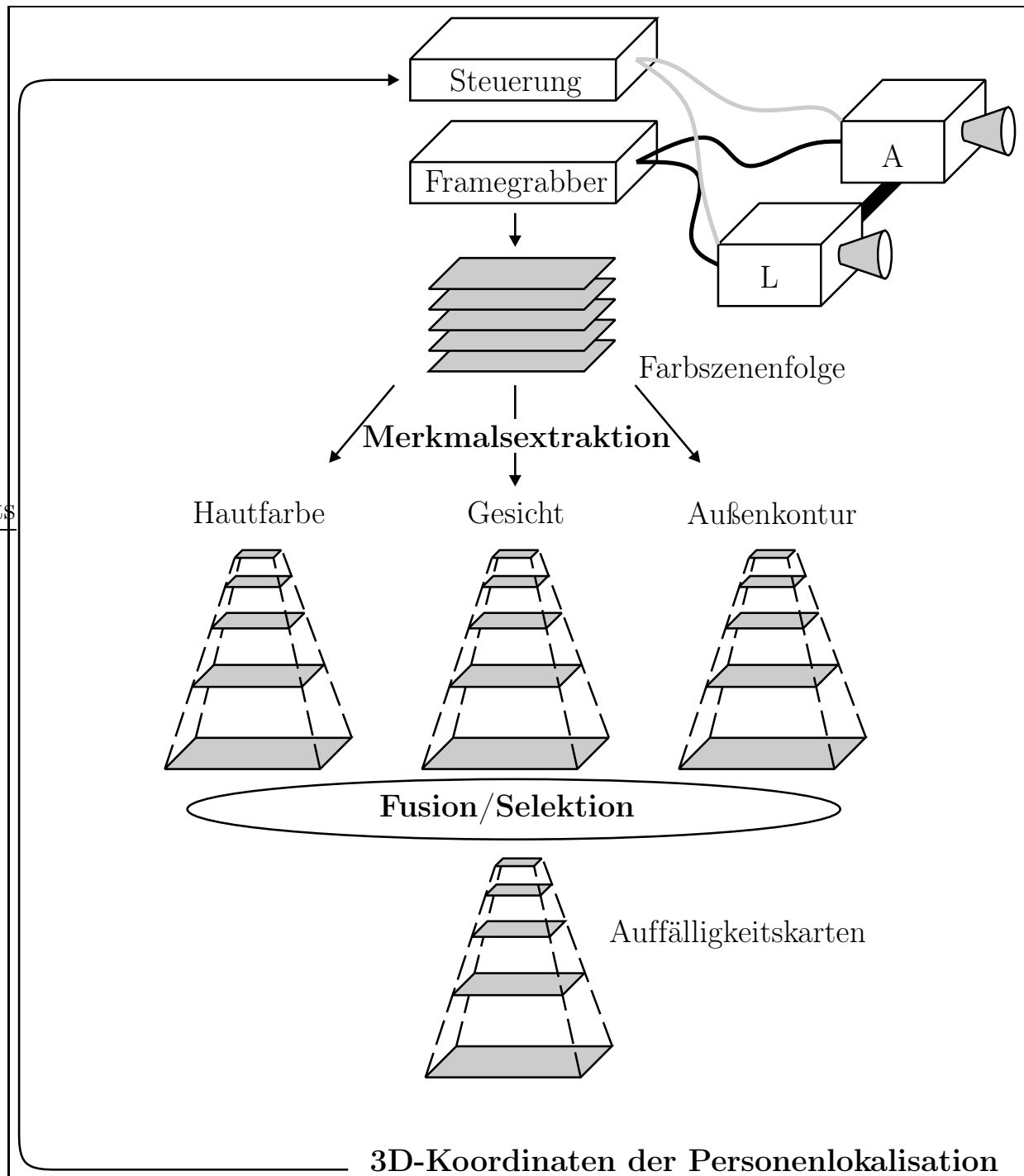


Abbildung 1.5: Schema der Einbettung der verwendeten Auffälligkeitskomponenten (Cues) in den Verarbeitungsablauf. Hervorzuheben ist die pyramidal angelegte Repräsentation aller drei Cues, die nach Verknüpfung den Input eines dreidimensionalen dynamischen neuronalen Feldes bilden, in dem eine Selektion erfolgt, so daß schließlich als Lokalisierungsergebnis die Koordinaten einer Person dreidimensional geschätzt werden können. Von den beiden Kameras dient die eine der Personenlokalisierung (L) in einer Übersichtsaufnahme der jeweiligen Szene und die andere der weiteren Analyse (A) der Person in einer Zoomaufnahme (nicht Gegenstand der vorliegenden Arbeit).

Kapitel 2

Pyramidale Repräsentationen als Basiskonzept

2.1 Einleitung

Pyramidale Repräsentationen sind eine seit langem in der Bildverarbeitung benutzte praktische und zugleich elegante Weise, bei Nutzung von Operatoren mit abgestuften bzw. gestaffelten spektralen Eigenschaften die Anzahl der auszuführenden Operationen auf ein erheblich vermindertes Maß zu reduzieren. ROSENFELD und JOLION /Ros1984, JR1994/ geben dazu einen umfassenden Überblick über das spezielle Gebiet *Auflösungspyramiden*¹ bzw. *Multiresolutional Image Processing*.

Die Relevanz für die vorliegende Arbeit liegt darin, daß die Lokalisation einer Person eine *Entfernungsschätzung* einschließen soll. Als mindestens erforderlich wird angesehen, in einem Entfernung- und damit Skalierungsintervall von zwei Oktaven diese Entfernungsschätzung vornehmen zu können, was ausgehend von einer festen minimal zulässigen Entfernung einen Spielraum bis zur Vervierfachung dieser Entfernung läßt. Daher rührt die Forderung, zumindestens einen Teil der zur Lokalisation notwendigen Operationen in aneinander anschließenden spektralen Intervallen auszuführen und somit indirekt über die spektrale Lage (sprich: Größe²) der Merkmale auf die Entfernung der Person zur Kamera schließen zu können.

Das führt zunächst dazu, einen Satz entsprechend abgestufter Operatoren zu entwerfen und die jeweiligen Operationen auszuführen. Dabei entsteht das Paradoxon, daß mit den

¹auch als Gaußpyramiden bezeichnet; wie noch erläutert wird, werden Gauß- oder Binomialfilter (als deren Approximation) unmittelbar zur Erstellung verwendet.

²Nur für solche Merkmale besteht eine Eignung, deren Größenvariation hinreichend klein bleibt.

aus diesem Satz größten und damit für die Rechnung aufwendigsten Operatoren nach den im Bild größten (und damit auch in der Abbildung am größten erscheinenden) enthaltenen jeweiligen Merkmalen gesucht wird.

Um dieses Paradoxon aufzulösen, besteht umgekehrt die Möglichkeit, auch ohne Informationsverlust das Bild zu verkleinern und *danach* die Operation auszuführen, sofern hierfür die durch die Verkleinerung wegfallenden höherfrequenten Bildbestandteile nicht relevant sind. Die Skalierung wird prinzipiell durch Unterabtastung einer Bildmatrix einer bestimmten Größe vorgenommen und resultiert in einer neuen Bildmatrix mit einer entsprechend verminderten Größe.

2.2 Abstufung und Glättung

Ein fehlerhaftes Vorgehen kann jedoch zu unerwünschten Ergebnissen führen, da zu beachten ist, daß hierbei das *Abtasttheorem* erfüllt bleibt. Dieses besagt, daß eine periodische Struktur nur dann aus Abtastwerten fehlerfrei rekonstruierbar bleibt, wenn die Abtastung dabei mindestens zweimal pro Wellenlänge erfolgt. Im Falle des genannten unterabzutastenden Bildes, das (bei der Bildaufnahme bereits abgetastete) periodische Strukturen bis zu einer minimalen Wellenlänge³ von $k_{\max}^{-1} = 2$ enthalten kann, muß also gesichert sein, daß *vor* der Unterabtastung idealerweise eine Beseitigung von spektralen Anteilen oberhalb $s \cdot k_{\max}$ erfolgt ist⁴. Andernfalls würden Strukturen enthalten bleiben, die nach erfolgter Unterabtastung eben *nicht* mehr hinreichend oft abgetastet wären. Damit ginge für diese Strukturen dem Theorem entsprechend die Rekonstruierbarkeit verloren, es käme zu falschen Rekonstruktionen. Der Effekt wird allgemein als *Aliasing* bezeichnet, womit gemeint ist, daß durch die falsche Rekonstruktion die Strukturen versetzt in einem falschen spektralen Bereich auftreten. Es kommt dabei zu einer Spiegelung von Strukturen, für die das Abtasttheorem verletzt ist, an der höchsten darstellbaren Frequenz in den darunterliegenden, richtig wiedergebbaren Frequenzbereich. Diese Überlagerung verfälscht das Bild.

In der Praxis sind spektrale Anteile oberhalb einer bestimmten Frequenz effektiv dämpfbar mit Tiefpaßfiltern, die im Ortsraum als Faltung (oder in erweiterter Form als Verknüpfung von Faltungen) mit bestimmten geeigneten Masken ausgeführt werden können. Die Wahl des Tiefpasses hängt jedoch vom Abstufungsintervall⁵ der Auflösungsstufen innerhalb der Pyramide ab, also vom Skalierungsfaktor s .

³Hier ist die unmittelbar im orthogonalen Raster darstellbare minimale Wellenlänge gemeint, die damit auch in allen Richtungen darstellbar ist. In Diagonalenrichtung ergibt sich eine um den Faktor $2^{-\frac{1}{2}}$ verminderte minimale Wellenlänge.

⁴ s symbolisiert dabei den gewählten Skalierungsfaktor im Intervall $(0, 1)$.

Ein typisches Abstufungsintervall von Auflösungspyramiden, die in der Literatur vorgestellt werden, ist der *Oktavabstand* mit einem Skalierungsfaktor von $s = 2^{-1}$. Dies entspricht also einer Halbierung der Auflösung von Ebene zu Ebene der Pyramide, wie es BURT und ADELSON /BA1983, Bur1984/ vorgeschlagen haben. Vorteil ist, daß der Zuwachs an Bild-daten (über die gesamte Pyramide gesehen) die theoretische Grenze⁶ von einem Drittel des Ursprungsdatenumfangs nicht überschreiten kann. Hinzu kommt der Wegfall des Bedarfs von Zwischenraster-Interpolationen. Als entscheidender Nachteil muß allerdings angesehen werden, daß der Oktavabstand sich leider für die Entfernungsbestimmung zu grob erweist, da z. B. die Merkmalsextraktionsmethoden für die Gesichtsdetektion eine geringere Skalierungsinvarianz als über eine Oktave aufweisen.

Dieser Umstand wurde von anderen Autoren (z. B. VIENNET und FOGELMAN SOULIÉ /VF1992/) ebenfalls erkannt, die sich stattdessen für den *Halboktavabstand* entschieden haben. Wenngleich damit Aufwand und Datenmenge der Pyramide steigen (der theoretisch maximal mögliche Datenzuwachs entspricht dem Umfang des Ursprungsbildes⁷), liegt damit die Genauigkeit der Entfernungsschätzung potentiell doppelt so hoch. Der Skalierungsfaktor s beträgt in diesem Falle $s = 2^{-\frac{1}{2}}$.

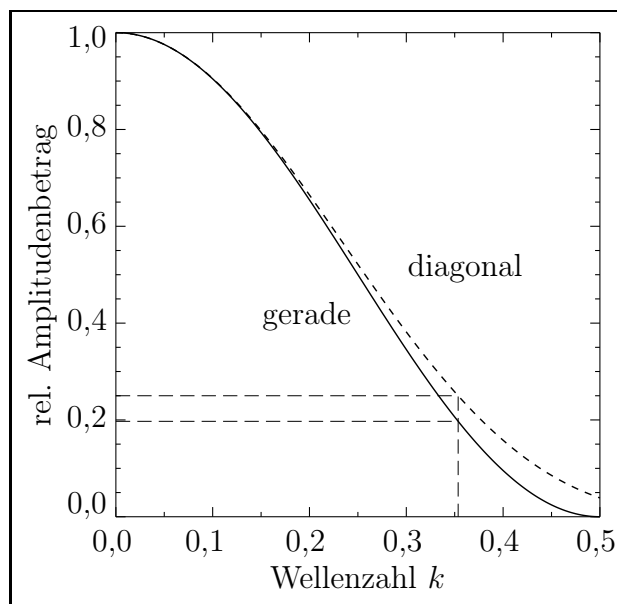


Abbildung 2.1: Glättungstiefpaß zur spektralen Begrenzung für die Unterabtastung beim Pyramidenaufbau. Es handelt sich um zwei einseitige Projektionen der Einhüllenden des reellen Amplitudenfrequenzspektrums des Tiefpasses ²**B** (siehe Gln. 2.2 und 2.3), um den Effekt der Anisotropie des Filters zu veranschaulichen. Die Projektion erfolgte entlang einer Koordinatenachse des Ortsfrequenzraumes (durchgezogene Kurve) und entlang einer Diagonalen (gestrichelte Kurve). Beide Kurven veranschaulichen die minimale und maximale Dämpfung, insbesondere bei der Grenzwellenzahl k_{grenz} .

Bei der Vorauswahl eines Tiefpasses wurde sich auf lineare Operatoren beschränkt. Als weiteres Auswahlkriterium kommt der Grad der Dämpfung oberhalb der Wellenzahl

⁵Da von vornherein keine Unterscheidung im entfernungsabhängigen Auflösungsbedarf der Entfernungsbestimmung gemacht werden soll, ist darüber hinaus eine *äquidistante* Abstufung der Pyramiden und damit der Einsatz *eines* allgemeinen Skalierungsfaktors sinnvoll.

$$^6 \lim_{n \rightarrow \infty} \sum_{i=0}^n \left(\frac{1}{2}\right)^{2n} = \frac{4}{3}$$

$$^7 \lim_{n \rightarrow \infty} \sum_{i=0}^n \left(\frac{1}{\sqrt{2}}\right)^{2n} = 2$$

$k_{\text{grenz}} = s \cdot k_{\text{max}}$ hinzu. Einen akzeptablen Kompromiß der Dämpfungseigenschaften weisen Binomialfilter auf. Es zeigt sich, daß bereits ein Binom 2. Ordnung zur Bildung eines geeigneten zweidimensionalen Binomialfilters ausreicht:

$${}^2\mathbf{B} = {}^2\mathbf{B}_1 * {}^2\mathbf{B}_2 \quad (2.1)$$

$$= \frac{1}{4} \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} * \frac{1}{4} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \quad (2.2)$$

Dessen Amplitudengang (*siehe* Abb. 2.1) wird bestimmt als:

$$\mathcal{F}\{{}^2\mathbf{B}\}(k_1, k_2) = \frac{1}{2}(1 + \cos(2\pi k_1)) \cdot \frac{1}{2}(1 + \cos(2\pi k_2)) \quad (2.3)$$

Von Interesse ist nun die erreichbare Dämpfung von Frequenzen oberhalb der Grenzwellenzahl

$$k_{\text{grenz}} = s \cdot k_{\text{max}} \quad (2.4)$$

$$= \frac{1}{\sqrt{2}} \cdot \frac{1}{2} \approx 0,35 \quad (2.5)$$

Diese Wellenzahl ist relativ bezogen auf eine betrachtete Auflösungsebene auf der nächstgrößeren Ebene noch darstellbar. Wie Abb. 2.1 veranschaulicht, ist ein kleiner Nachteil des gewählten Tiefpaßoperators, daß die Glättungs- und damit Dämpfungseigenschaften leicht *anisotrop* sind, so daß es eine leichte Abhängigkeit von der Richtung einer Struktur gibt. In einer Richtung gerade zum Raster beträgt die (max.) Dämpfung $\frac{1}{2}(1 + \cos(2\pi k_{\text{grenz}})) \approx 0,2$. In Diagonalenrichtung beträgt sie noch $\frac{1}{4}(1 + \cos(2\pi 2^{-\frac{1}{2}} k_{\text{grenz}}))^2 = 0,25$. Diese noch relativ hohen Werte werden jedoch für den Einsatz innerhalb des in dieser Arbeit vorgestellten Systems als noch vertretbar angesehen, da eine Verbesserung nur mit deutlich erhöhtem Aufwand mit Filtern höherer Ordnung erzielbar wäre. In Abb. 2.3 ist ein anschauliches Kriterium für die Güte der Filterung das Streifenmuster auf dem Hemd der Person. Auch auf der größten Auflösungsebene sind praktisch keine Aliasing-Artefakte erkennbar.

Die Glättung erfolgt nun auf jeder Farbauszugmatrix⁸ $\mathbf{I}_R$, $\mathbf{I}_G$ und $\mathbf{I}_B$ separat, was im Grunde der Vektorcharakteristik der Farbkodierung nicht voll gerecht wird, da hierbei kleinere Farbverfälschungen in Kauf genommen werden müssen. Vorschläge, diese Glättungsoperation besser in einem nichtlinearen oder in einem Differenzfarbraum nur bezogen auf zwei Buntheitsgrößen (komplexe Darstellung) auszuführen, wurden von FREY /Fre1988,Fre1990/ gemacht (*siehe auch* FRANKE /Fra1997/). Weil jedoch für diese Arbeit die Farbauszüge auch einzeln für die Auffälligkeitskomponenten verwendet werden, wurde eine Einzelglättung vorgenommen, natürlich nicht zuletzt auch wegen des geringeren Aufwandes.

⁸Zum weiteren Verständnis der Eigenschaften von Farbbildern sei hier bereits auf den Abschnitt 3.2 *Lineare Farbraummodelle* und auf Seite 32 in Abschnitt 3.3.4 *Wahl des Klassifikationsverfahrens* verwiesen.

2.3 Unterabtastung und Interpolation

Nachdem im vorangegangenen Abschnitt das Thema Glättung als Voraussetzung zur Erzeugung einer pyramidalen Repräsentation behandelt worden ist, soll nun mit Unterabtastung und Interpolation der eigentliche Schritt dazu dargestellt werden. Das Bild (einzeln Farbauszug) auf der n -ten Pyramidenebene $\mathbf{I}^{(n)}$ wird dafür tiefpaßgefiltert und dann mit einem Unterabtast- und Interpolationsoperator $\mathbf{R}$ behandelt, so daß sich die nächste Pyramidenebene formal folgendermaßen bilden läßt:

$$\mathbf{I}^{(n+1)} = \mathbf{R} \left({}^2\mathbf{B} * \mathbf{I}^{(n)} \right) \quad (2.6)$$

Was verbirgt sich hinter dem Unterabtast- und Interpolationsoperator $\mathbf{R}$? In einem Doppelschritt werden zunächst entsprechend eines Skalierungsfaktors (hier $s = \frac{1}{\sqrt{2}}$) die i. a. auf Zwischengitterplätzen liegenden Abtastkoordinaten bestimmt. Im einfachsten Fall könnten dafür die Werte des jeweiligen nächsten Nachbarn verwendet werden, wobei damit auf eine Interpolation im eigentlichen Sinne verzichtet würde.

Von den möglichen Interpolationsmethoden erscheint ein lineares Verfahren aufwandsmäßig am adäquatesten. Bei der sog. *bilinearen Interpolation* werden die vier Nachbarn eines solchen Zwischengitterplatzes zur Bestimmung eines Abtastwertes linear gewichtet einbezogen. Diese Interpolation erfolgt für die Koordinate k, l einer reduzierten Matrix an den i. a. auf Zwischengitterplätzen liegenden reellwertigen Koordinaten $s^{-1} \cdot k, s^{-1} \cdot l \in \mathbb{R}$, die daher jeweils zwischen zwei diskreten nichtnegativ ganzzahligen Koordinaten $p, q \in \mathbb{N}$ und $p+1, q+1 \in \mathbb{N}$ liegen, so daß gilt $p \leq s^{-1} \cdot k < p+1$ und $q \leq s^{-1} \cdot l < q+1$.

Sei $\tilde{i}^{(n)}(p, q)$ ein Bildpunkt eines (bereits geglätteten) Bildes ${}^2\mathbf{B} * \mathbf{I}^{(n)}$ (n -te Pyramidenebene), der zusammen mit den drei weiteren Bildpunkten $\tilde{i}^{(n)}(p+1, q)$, $\tilde{i}^{(n)}(p, q+1)$ und $\tilde{i}^{(n)}(p+1, q+1)$ die vier unmittelbar benachbarten Eckpunkte eines minimalen Matrixausschnitts bildet. Dann liegen die Koordinaten des abzutastenden Punktes, dessen Wert durch Interpolation gefunden werden muß, genau innerhalb dieses Ausschnitts. Die Vorschrift für die bilineare Interpolation dieses Abtastwertes für den Bildpunkt $i^{(n+1)}(k, l)$ der nächsten Pyramidenebene lautet:

$$\begin{aligned} i^{(n+1)}(k, l) = & \tilde{i}^{(n)}(p, q) \Delta p_2 \Delta q_2 + \\ & \tilde{i}^{(n)}(p, q+1) \Delta p_2 \Delta q_1 + \\ & \tilde{i}^{(n)}(p+1, q) \Delta p_1 \Delta q_2 + \\ & \tilde{i}^{(n)}(p+1, q+1) \Delta p_1 \Delta q_1 \end{aligned} \quad (2.7)$$

Darin sind $\Delta p_1 = s^{-1} \cdot k - p$ und $\Delta q_1 = s^{-1} \cdot l - q$ sowie $\Delta p_2 = 1 - \Delta p_1$ und $\Delta q_2 = 1 - \Delta q_1$ die jeweiligen Koordinatendistanzen zu den Koordinaten der Gitterpunkte.

Damit steht mit den Gln. 2.6 und 2.7 ein komplettes algorithmisches Instrumentarium für die Bildung der pyramidalen Repräsentation zur Verfügung. Die Abbn. 2.2 und 2.3 sollen diese abschließend für ein Beispiel auf zwei verschiedene Arten veranschaulichen. Es wird noch einmal ausdrücklich darauf hingewiesen, daß jeder der drei Farbauszüge einzeln auf diese Weise verarbeitet wird.

Abbildung 2.2: Auflösungsstufen der pyramidalen Repräsentation (1): Durch den gewählten Halboktavabstand bedingt, besträgt das Intervall zwischen der jeweils auf der obersten und der untersten Auflösungsebene maximal darstellbaren Wellenzahl genau zwei Oktaven bei fünf umfaßten Auflösungsebenen.

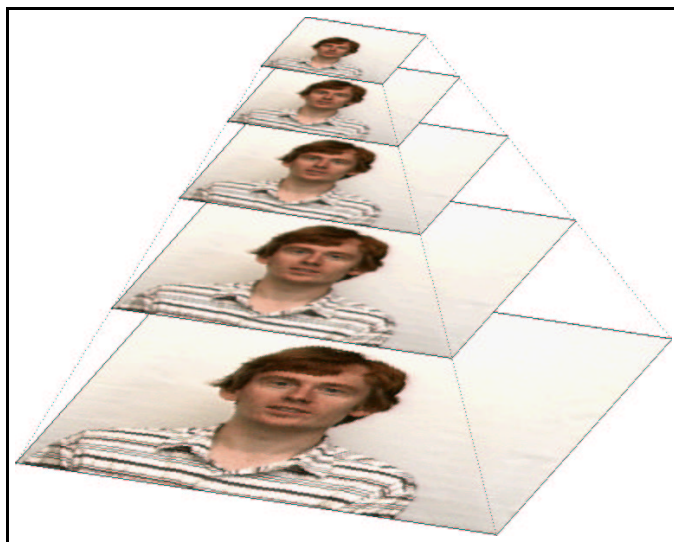


Abbildung 2.3: Auflösungsstufen der pyramidalen Repräsentation (2, v.l.n.r.): Testbild in Grundauflösung; erste bis vierte Unterabtastungsstufe, jeweils im Halboktavabstand. Die Skalierung der Darstellung erfolgte individuell derart, daß die Größe im Druck vereinheitlicht ist. In allen Darstellungen von Untersuchungsergebnissen innerhalb dieser Arbeit, insbesondere im Abschnitt 6.4 Demonstration des realisierten Systems, wurde derart vorgegangen.

2.4 Résumé

In diesem Kapitel wurde eine kurze Darstellung des Basiskonzepts *pyramidale Repräsentationen* gegeben. In den folgenden Kapiteln wird darauf häufig Bezug genommen werden, da es für alle Cues so konsequent wie möglich eingesetzt wird, um eine Entfernungsschätzung in die Lokalisation der Person einbeziehen zu können. Pyramidale Repräsentationen erlauben dabei eine optimale spektrale Anpassung der Operatoren, so daß nur ein Mindestmaß an Rechenoperationen ausgeführt werden muß. Die für diese Arbeit benutzte Konfiguration geht

von einem Arbeitsbereich aus, der sich über zwei Oktaven erstreckt. Die Abstufung erfolgt im Halboktavabstand, so daß sich genau fünf verschiedene Auflösungsstufen ergeben. Die dafür eingesetzte Glättungsoperation ist für die Problemstellung qualitativ zufriedenstellend. Ebenso ist die gewählte Unterabtastung mit bilinearer Interpolation für den Einsatz in einer solchen Auflösungspyramide geeignet.

Kapitel 3

Hautfarbdetektion

3.1 Einleitung

Vor dem Hintergrund der Entwicklung eines aufgabenspezifischen Auffälligkeitssystems zur Personenlokalisierung wird in diesem Kapitel neben der Einordnung eine Darstellung des Problemkreises *Farbe* mit einer Vertiefung speziell zu *Hautfarbe* gegeben, die mit der Vorstellung von Methoden zur Hautfarbdektion und mit Beispielergebnissen der für diese Arbeit ausgewählten Methode abschließt.

An einigen Stellen dieses Kapitels macht sich die Darstellung bekannter grundlegender Zusammenhänge erforderlich, um den relevanten theoretischen Hintergrund zu erfassen, der für eine Umsetzung des hautfarbbasierten Cues sowie für das Verständnis der verschiedenen Verarbeitungsstufen bedeutsam ist.

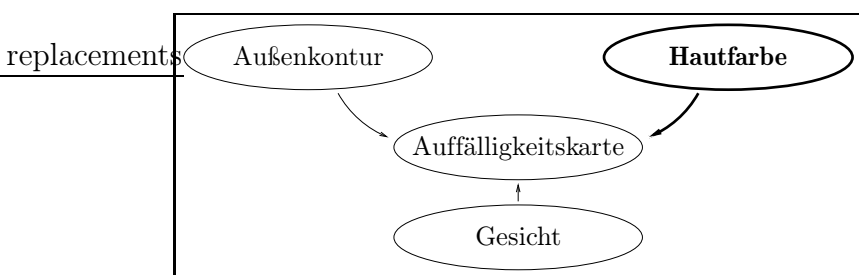


Abbildung 3.1: Der Hautfarb-Cue bildet die erste Komponente zur Kombination einer fusionierten Auffälligkeitskarte.

3.1.1 Zielstellung

Die Verwendung von Hautfarbe als einem der Cues für ein visuelles Auffälligkeitssystem ist zunächst der Versuch, ein für Personen typisches lokales Farbmerkmal¹ heranzuziehen

¹Es wird sich auf Gesichtspartien hellhäutiger Personen kaukasischen Typs beschränkt.

und entsprechende sichtbare Körperpartien von Personen zu segmentieren. Im Ergebnis soll ein Wahrscheinlichkeitsmaß für die Zugehörigkeit zu einer Klasse der hautfarbenen Pixel gebildet werden.

Das Merkmal *Hautfarbe* erweist sich allerdings bei genauerer Betrachtung weniger spezifisch, als das zunächst vermutet werden könnte. Selbst wenn die Reflexionseigenschaften von sichtbaren Hautpartien keine inter- und intra-individuellen Schwankungen aufweisen würden, verbliebe das nichttriviale Problem, Beleuchtungsschwankungen zu kompensieren. Die Lösung dafür wird als Farbkonstanzleistung bezeichnet (*siehe Anhang A Ansätze zur referenzfreien Farbkonstanz*).

Der visuelle Reiz *Hautfarbe* weist keine Einzigartigkeit innerhalb der Menge möglicher Farbreize auf, weder von den verschiedenen auftretenden Hauttypen mit inter- und intra-individuellen Schwankungen her gesehen, noch mit Blick auf die vielfältigen möglichen Beleuchtungsbedingungen.

TREISMAN /Tre1988/ stellt im Zusammenhang mit Untersuchungen zur visuellen Auffälligkeit fest, daß insbesondere von der vorherrschenden Farbe abweichende farbliche Objekte beschleunigt detektiert werden können.² TREISMANS Ergebnisse bilden die Basismotivation für die Nutzung von Farbinformation im Rahmen dieser Arbeit: durch Farblichkeit kann die visuelle Analyse vereinfacht und damit beschleunigt werden, weil unter ihrer Einbeziehung weniger Mehrdeutigkeiten verbleiben und damit eine aufwendigere Extraktion weiterer (verteilter) Merkmale häufig unterbleiben kann.

Technisch realisiert wird die Merkmalsextraktion von Farbmerkmalen durch Farbklassifikation/-segmentierung in Farbräumen. Im nachfolgenden Abschnitt wird eine Auswahl linearer Farbraummodelle³ vorgestellt, die für das im darauffolgenden Abschnitt vorgestellte Verfahren zur Hautfarbsegmentierung relevant sind.

²Analoge Effekte wurde von ihr schon bei rein orientierungs- oder größenbezogen voneinander abweichenden Objekten festgestellt /Tre1986/.

³Diese bilden in Form des Normvalenzsystems XYZ die Basis der Farbraummodelle, auf die sich alle anderen zurückführen lassen. Weiterhin wurden aus Gründen der Einfachheit lineare Modelle zunächst als ausreichend für das zu lösende Hautfarbsegmentierungsproblem angesehen.

3.2 Lineare Farbraummodelle

3.2.1 Additive Farbmischung

Farben sind aus Primärvalenzen (Grundfarben) mischbar, indem das räumliche oder zeitliche Auflösungsvermögen des Auges überschritten wird und der Beobachter eine Mischfarbe wahrnimmt. Mathematisch beschreiben läßt sich die additive Mischung, indem man die gewählten festen Primärvalenzen als Basisvektoren eines kartesischen Vektorraumes ausdrückt. Seien also die drei Primärvalenzen den Vektoren $\underline{\mathbf{R}} = (1, 0, 0)^\top$, $\underline{\mathbf{G}} = (0, 1, 0)^\top$ und $\underline{\mathbf{B}} = (0, 0, 1)^\top$ zugeordnet, dann läßt sich, wie bereits SCHRÖDINGER /Sch1920/ vorschlug, die resultierende Mischfarbvalenz $\underline{\mathbf{M}}$ als Vektoraddition entsprechend bestimmter Farbwerte R , G und B beschreiben:

$$\underline{\mathbf{M}} = R\underline{\mathbf{R}} + G\underline{\mathbf{G}} + B\underline{\mathbf{B}} \quad (3.1)$$

Der von diesen Basisvektoren aufgespannte Vektorraum eignet sich prinzipiell zur Repräsentation von Farben, da sowohl Mischungen als auch Distanzen von Farben dargestellt werden können. Um bestimmen zu können, welche Farbempfindung ein gegebener Farbreiz erzielen wird, wird eine Beziehung für die farbreizabhängige Bildung der Farbwerte (Koeffizienten in Gl. 3.1) benötigt. Jeder dieser Farbwerte wird als Integral über das Produkt von Farbreiz L (Funktion der Wellenlänge λ) mit der jeweiligen Spektralwertkurve⁴ (ebenfalls Funktion von λ) gebildet:

$$R = \int \bar{r}(\lambda) L(\lambda) d\lambda \quad (3.2)$$

$$G = \int \bar{g}(\lambda) L(\lambda) d\lambda \quad (3.3)$$

$$B = \int \bar{b}(\lambda) L(\lambda) d\lambda \quad (3.4)$$

Diese drei Integrale drücken eine immense Dimensionsreduktion aus. Das auf Sensoren auftreffende Strahlungsspektrum $L(\lambda)$ des Farbreizes mit theoretisch unendlicher Dimensionalität wird auf drei Dimensionen abgebildet und damit kompaktiert.

⁴Spektralwertkurve: zu jeder Primärvalenz korrespondierende Funktion der Wellenlänge, deren Funktionswert denjenigen wellenlängeabhängigen Farbwertanteil der Primärvalenz angibt, den diese zum Ermschen der Spektralfarbe beitragen muß; diese Kurven beziehen sich auf einen normalsichtigen Beobachter und wurden mit psychophysischen Messungen ermittelt; sie sind auf verschiedene Farbraummodelle umrechenbar.

3.2.2 Primärvalenzmodelle

replacements

Wichtigstes lineares Farbraummodell ist das Normvalenzsystem XYZ . Es wurde von der CIE 1931 eingeführt. Die Farbwerte des Normvalenzsystems werden analog Gln. 3.2 bis 3.4

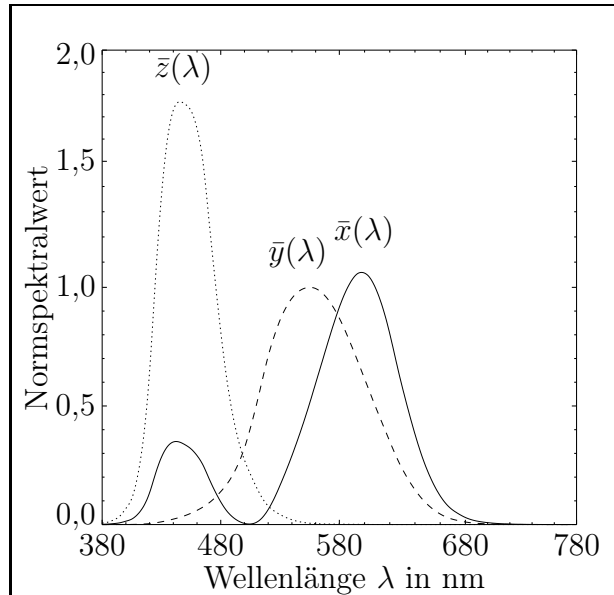


Abbildung 3.2: Normspektralwertkurven nach CIE; gemäß den von der CIE 1931 für den sog. farbmtrischen 2° -Beobachter im Bereich $\lambda \in 380\text{nm} \dots 780\text{nm}$ in 5nm-Schritten festgelegten Normfarbwertanteilen $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ und $\bar{z}(\lambda)$ (interpoliert).

bestimmt:

$$X = \int \bar{x}(\lambda) L(\lambda) d\lambda \quad (3.5)$$

$$Y = \int \bar{y}(\lambda) L(\lambda) d\lambda \quad (3.6)$$

$$Z = \int \bar{z}(\lambda) L(\lambda) d\lambda \quad (3.7)$$

Eine wichtige Darstellungsform von Farbvalenzen in Farbraummodellen mit Primärvalenzen sind Farbwertanteile. Im Falle des Normvalenzsystems bilden sich die Normfarbwertanteile zu:

$$x = \frac{X}{X + Y + Z} \quad y = \frac{Y}{X + Y + Z} \quad z = \frac{Z}{X + Y + Z} \quad (3.9-3.10)$$

Diese auf die Summe normierte Darstellung stellt im kartesischen Koordinatensystem eine Projektion auf eine Ebene dar, die durch die drei Normprimärvalenzen $\underline{X}$, $\underline{Y}$ und $\underline{Z}$ verläuft. Die Summe der Normfarbwertanteile selbst ist immer eins. Deshalb reicht es aus, für eine Darstellung von Farbwertanteilen immer nur zwei auszuwählen. Allerdings ist bei dieser Normierung eine Dimension zur Beschreibung verlorengegangen. Deshalb werden die typischerweise ausgewählten Anteile x und y als Farbart bezeichnet, bei der keine explizite Intensitätsinformation mehr enthalten ist. Jedoch hat gerade diese als CIE-Normfarbtafel bezeichnete Darstellungsform (siehe Abb. 3.4) für die Veranschaulichung und Gegenüberstellung von anderen Farbraummodellen eine sehr große Bedeutung erlangt.

Mit Hilfe des Normvalenzsystems lassen sich nun weitere auf Primärvalenzen basierende Farbraummodelle darstellen. Die für diese Arbeit eingesetzte Kamera VC-C1 (siehe Abb. 1.3) nutzt Primärvalenzen nach EBU-Norm. Für diese existiert das Farbraummodell RGB_{EBU} und ist mit folgenden Transformationsmatrizen zu bilden:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix}_{EBU} = \begin{pmatrix} 3,0651 & -1,3942 & -0,4761 \\ -0,9690 & 1,8755 & 0,0415 \\ 0,0679 & -0,2290 & 1,0698 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (3.11)$$

replacements

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0,4303 & 0,3416 & 0,1782 \\ 0,2219 & 0,7086 & 0,0713 \\ 0,0202 & 0,1296 & 0,9387 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}_{EBU} \quad (3.12)$$

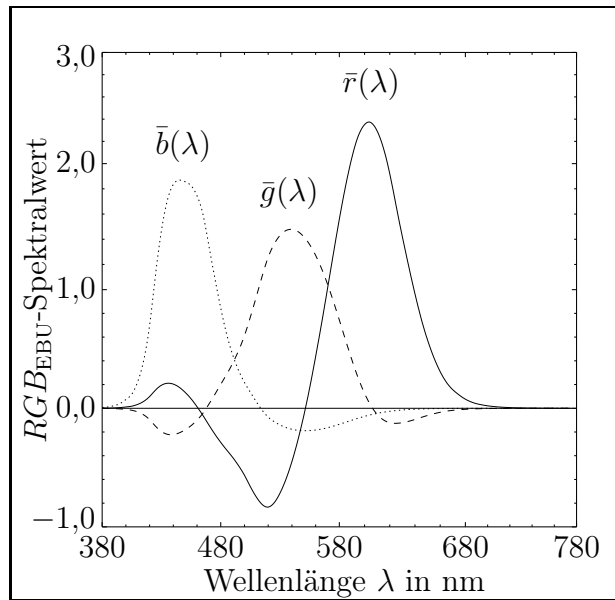


Abbildung 3.3: RGB-Spektralwertkurven nach EBU, die entsprechend der Transformationsmatrix aus Gl. 3.11 aus den Normspektralwertkurven in Abb. 3.2 gebildet werden. Im Unterschied zu diesen treten offensichtlich negative Anteile auf, ein Indiz dafür, daß die zugrunde liegenden Primärvalenzen realen Farbreizen entsprechen, so daß die Spektralfarben, die außerhalb des mit diesen Primärvalenzen eingeschlossenen Bereichs liegen, nur mit negativen Anteilen ermischt werden können.

Abb. 3.4 stellt anhand der Normfarbwertanteile x und y den möglichen Umfang der Farbwahrnehmung dar. Alle Farbörter⁵ von Farben, die ein Beobachter wahrnehmen kann, liegen innerhalb dieses Kurvenzuges. Dieser wird zwischen seinen spektralen Enden durch die sog. Purpurgerade geschlossen. Da die Normvalenzen, deren Lage für X bei $x = 1, y = 0$ bzw. für Y bei $y = 1, x = 0$ gegeben ist, außerhalb des Kurvenzuges liegen, wird von virtuellen Primärvalenzen gesprochen. Der abgegrenzte dreieckförmige Bereich stellt die Untermenge an Farbörtern dar, die aus technischer Sicht tatsächlich synthetisierbar sind. Die Abmaße des Bereichs werden unmittelbar durch die drei Primärvalenzen nach EBU bestimmt. Real außerhalb liegende Farbörter werden auf die Schnittpunkte von Verbindungsgeraden von ihnen zum Unbuntpunkt mit den Begrenzungen des Dreiecks projiziert. Dadurch können quasi Sättigungseffekte an den Dreiecksgrenzen auftreten.

⁵Farbörter: Projektionen in die Farbtafel; in der Farbmatrik wird, wie in der Geometrie auch, von ‚Örtern‘ als Plural von ‚Ort‘ gesprochen.

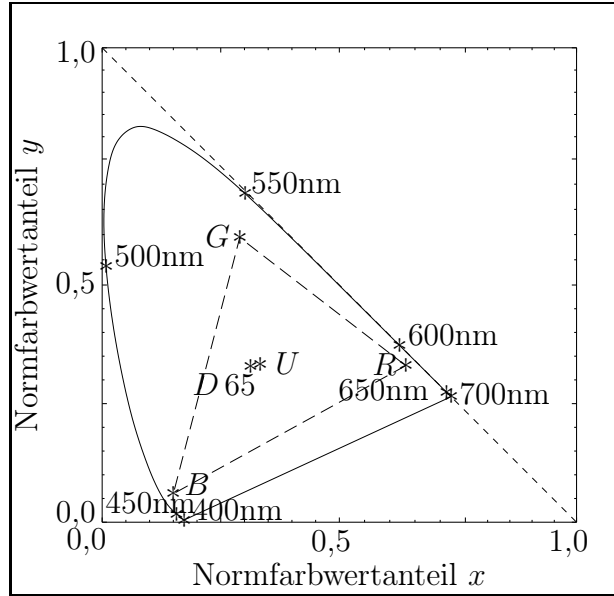


Abbildung 3.4: Normfarbtafel nach CIE; dargestellt ist die interpolierte Verbindung aller von der CIE 1931 für den sog. farbmimetrischen 2°-Beobachter im Bereich $\lambda \in 400\text{nm} \dots 700\text{nm}$ in 5nm-Schritten festgelegten Normfarbwertanteile x und y . Der darin abgegrenzte Bereich kann durch das Farbraummodell RGB_{EBU} maximal erfasst werden. Markiert ist außerdem der Unbunt-punkt U , an dem alle drei Normfarbwertanteile den Wert $\frac{1}{3}$ annehmen, sowie der Farbort von Normlichtart $D 65$ (repräsentiert mittleres Tageslicht).

3.2.3 Gegenfarbmodelle

Weitere Farbraummodelle beruhen auf Achsen entlang von Gegenfarben. Sie sind einerseits aufgrund von neurophysiologischen Erkenntnissen über die postrezeptorale Kodierung motivierbar. Interessanterweise sind darüber hinaus auch in den Übertragungsnormen von Fernsehsignalen Gegenfarbkodierungen vorgesehen worden. Für diese Arbeit ist davon das Farbraummodell von POMIERSKI /Pom1996/ von Interesse, der dieses unter der Bezeichnung *physiologisch motivierter Elementarfarbraum* $RG/BY/WS$ vorgeschlagen hat. Die dazugehörige lineare Transformationsvorschrift berücksichtigt die postrezeptorale Verschaltung im parvozellulären⁶ visuellen System des Menschen, in dem nach HERINGS /Her1905/ bestätigter Vermutung drei *opponente* Aktivierungen gebildet werden. Diese liegen im Modell von POMIERSKI als Aktivierungen entlang der Achsen RG , BY und WS vor. In Bezug auf das Normvalenzsystem ergeben sich folgende Transformationen:

$$\begin{pmatrix} RG \\ BY \\ WS \end{pmatrix} = \begin{pmatrix} 1,9690 & -2,3251 & 0,3325 \\ -1,6201 & 0,3624 & 1,0367 \\ 1,4744 & 2,0238 & 0,8702 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (3.13)$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0,1536 & -0,2322 & 0,2180 \\ -0,2531 & -0,1054 & 0,2222 \\ 0,3284 & 0,6385 & 0,2630 \end{pmatrix} \begin{pmatrix} RG \\ BY \\ WS \end{pmatrix} \quad (3.14)$$

Bedeutung hat dieses Gegenfarbmodell für diese Arbeit deshalb, weil es die Basis für ein referenzfreies Verfahren zur Farbkonstanz bildet, das ebenfalls von POMIERSKI /Pom1996/ vorgeschlagen wird. Im Anhang A.2 *Elementarfarbraum-Verfahren* wird das Verfahren in

⁶kleinzelligen

einer speziellen Anpassung vorgestellt.

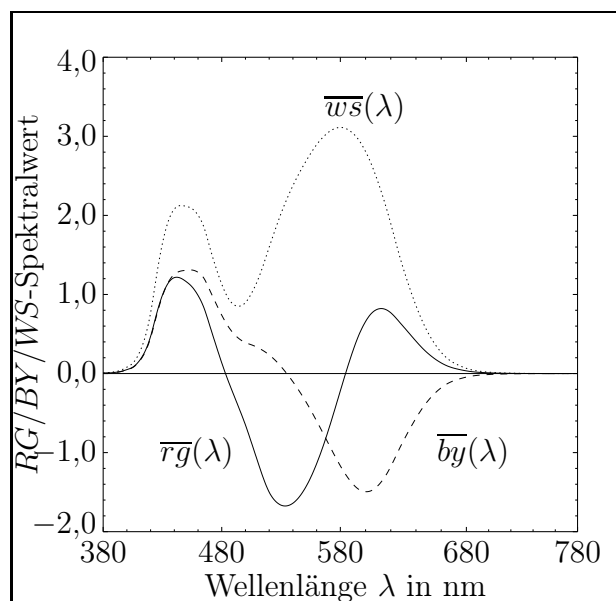


Abbildung 3.5: *RG/BY/WS-Spektralwertkurven nach POMIERSKI; entsprechend der Transformationsmatrix aus Gl. 3.13 aus den Normspektralwertkurven gebildet.*

3.3 Hautfarbsegmentierung

3.3.1 Vorverarbeitung

Gerade weil in den technischen Realisierungen letztlich sehr häufig Farbbilder mit *RGB*-Farbwerten (nach EBU wie für diese Arbeit, oder nach FCC) vorliegen bzw. gebildet werden, ist der einfachste Weg die Segmentierung im jeweiligen *RGB*-Farbraummodell. Deshalb haben verschiedene Autoren, die Arbeiten zur Hautfarbsegmentierung vorstellen (*siehe* YANG, WEIER und WAIBEL /YW1995,YW1996,YLW1997/ oder z. T. LITTMANN und RITTER, H. /LR1995/), gerade darauf zurückgegriffen. Jedoch bleiben dabei zwei prinzipiell verschiedene Möglichkeiten: die direkte Verwendung der Farbwerte oder die Bildung von Farbwertanteilen. Farbwertanteile, wie sie bereits prinzipiell für das Normvalenzsystem in Gl. 3.9–3.10 vorgestellt worden sind, lassen sich für alle anderen Primärfarbmodelle ebenso bilden.

$$r = \frac{R}{R + G + B} \quad g = \frac{G}{R + G + B} \quad b = \frac{B}{R + G + B} \quad (3.16\text{--}3.17)$$

Zwar stellen sie eine (weitere) Dimensionsreduktion $\mathbb{R}^3 \mapsto \mathbb{R}^2$ dar, da einer der drei Farbwertanteile redundant ist. Durch die Projektion erfolgt aber gleichzeitig eine Beseitigung der Abhängigkeit von der Intensität (Summe der Farbwerte) und damit von der Intensität der Beleuchtung⁷. Dieser Verarbeitungsschritt kann damit Konstanzleistung liefern, die jedoch nicht getrennt werden kann von einem gleichzeitigen Verlust an Diskriminationsvermögen.

⁷Intensität der Beleuchtung: ein Maß für den Fluß an elektromagnetischer Strahlung innerhalb des sichtbaren spektralen Intervalls.

Intensitätsunterschiede in der Beleuchtung in einer realen visuellen Szene entstehen insbesondere durch Inhomogenitäten der Ausleuchtung oder lokale Abschattungen. Diese Effekte werden durch die Nutzung von Farbwertanteilen weitgehend tolerierbar. Deshalb bilden Farbwertanteile für die Hautfarbsegmentierung innerhalb dieser Arbeit die Grundlage. In Abb. 3.6 wird der verwendete Hautfarbdatensatz auf Basis von Farbwertanteilen dargestellt.

Darüber hinaus kommt jedoch, wie bereits erwähnt, einer Farbkonstanzleistung (*siehe* Anhang A *Ansätze zur referenzfreien Farbkonstanz*) eine wichtige Bedeutung zu. Die Intensitätsnormierung stellt dazu keine Alternative dar.

3.3.2 Gewinnung von Hautfarbdaten

Zur Gewinnung eines Datensatzes von hautfarbenen Pixeln kommt nur die saubere manuelle Markierung entsprechender Regionen aus Gesichts- und Handbereichen in repräsentativen Beispielbildern in Frage. Wenn von einem repräsentativen Personenkollektiv⁸ ausgegangen wird, sind dabei folgende zwei Varianten möglich:

1. Schaffung von definierten Beleuchtungsbedingungen durch Einsatz einer Normlichtart nach CIE bei darauf abgestimmten Kameraeinstellungen
2. keine Schaffung von definierten Beleuchtungsbedingungen, möglichst breite Variation der Beleuchtungsbedingungen entsprechend der zu erwartenden möglichen Änderungen

Im ersten Fall ist die Streuung im Datensatz ausschließlich durch die tatsächliche Streuung der Remissionseigenschaften der erfaßten Hautpartien bedingt. Im zweiten Fall ist diese Streuung von den Fluktuationen der Beleuchtung überlagert.

Sofern die im Einsatz erzielbare Farbkonstanzleistung sehr gut ist, ist die erste Variante sinnvoller, mit der damit eine höhere Spezifität der Klassifikation erwartet werden kann. Die zweite Variante läßt demgegenüber eine schlechtere Spezifität erwarten, dafür aber durch die breitere Streuung im Datensatz eine höhere Sensitivität der Klassifikation. Höhere Sensitivität aber bringt bei schlechterer Spezifität auch mehr falsch positive Ergebnisse, was damit der Preis für eine etwaige mangelhafte Farbkonstanzleistung wäre.

Da nur von einer befriedigenden Farbkonstanzleistung ausgegangen werden kann, (*siehe* Anhang A *Ansätze zur referenzfreien Farbkonstanz*) sind die Daten deshalb etwa zu gleichen Teilen unter zwei verschiedenen Beleuchtungsbedingungen (Tageslicht verschiedener Stadien und Halogenlicht⁹) aufgenommen worden. Im oberen Teil der Abb. 3.6 sind alle erfaßten

⁸Hautfarbänderungen durch körperliche Belastungen o. ä. werden hier ausgeklammert.

hautfarbenen Pixel zusammen in einem „Bild“ dargestellt. Der Datensatz hat einen Umfang von 13300 Pixeln.¹⁰

Außerdem ist die Bedeutung der Qualität und Ausgewogenheit der manuellen Markierung der relevanten Hautpartien hervorzuheben. Sollten darin übersteuerte oder spiegelnde Bildabschnitte (z. B. im Stirnbereich) enthalten sein, fehlt einem Teil des Datensatzes die entsprechende Spezifität, ein Mangel, der sich auf das Klassifikationsergebnis und damit auf die erreichbare Segmentierungsleistung übertragen wird.

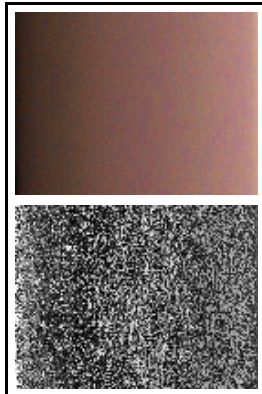


Abbildung 3.6: Hautfarbdatensatz (oberes Bild intensitätsbasiert sortiert, im Schwarzweißdruck nur als Intensitätsbild darstellbar), sowie die pixelweise korrespondierende Wahrscheinlichkeit für ein hautfarbendes Pixel (unteres Bild, schwarz: hohe Wahrscheinlichkeit). Durch die Intensitätsnormierung besteht nur noch eine geringe erkennbare Intensitätsabhängigkeit der Klassifikation. Ein Vergleich der durch lineare Regression bestimmten Anstiege der Intensitäten im intensitätsbasiert sortierten Datensatz und dem korrespondierenden Selbstklassifikationsergebnis ergab ein Verhältnis von ca. 130:1.

3.3.3 Charakterisierung von Hautfarbe

Diese Farbe lässt sich nicht so exakt definieren, wie das etwa bei Druckfarben möglich ist. Hautfarbe erweist sich als erstaunlich variabel, doch wird sie aus naiver Sicht eher als ganz charakteristisch angenommen, häufig gesättigter als tatsächlich beobachtbar. In Abbildung 3.7 sind zwei Spektren des Farbreizes von Hautpartien dargestellt. Beide unterscheiden sich nicht nur durch einen Skalierungsfaktor. Während die untere eher rotbetont ist, enthält die obere anteilig wesentlich mehr spektrale Grünanteile, so daß sie mehr zu einer bräunlich-gelblichen Erscheinung führt.

3.3.4 Wahl des Klassifikationsverfahrens

Offenbar handelt es sich bei Hautfarbsegmentierung zunächst um ein einfaches Klassifikationsproblem mit zwei Klassen: Hautfarbe oder sonstige Farbe. Sehr viele, praktisch alle,

⁹Auf Bedingungen explizit nur unter Leuchtstoffröhrenbeleuchtung wurde bei der Datenakquisition verzichtet, da diese in den vorhandenen Räumlichkeiten leider nicht blendfrei vorlag.

¹⁰Da es weniger darauf ankam, bei der manuellen Segmentierung große, eher homogen erscheinende Regionen zu segmentieren, wurden (mit einem entsprechenden Aufwand) mehr als 100 kleinere, nichtzusammenhängende Regionen einbezogen, die ca. 60 Aufnahmen entstammen, so daß der Umfang des Datensatzes insgesamt relativ klein bleibt. Die Datenbasis erscheint damit dennoch hinreichend groß genug.

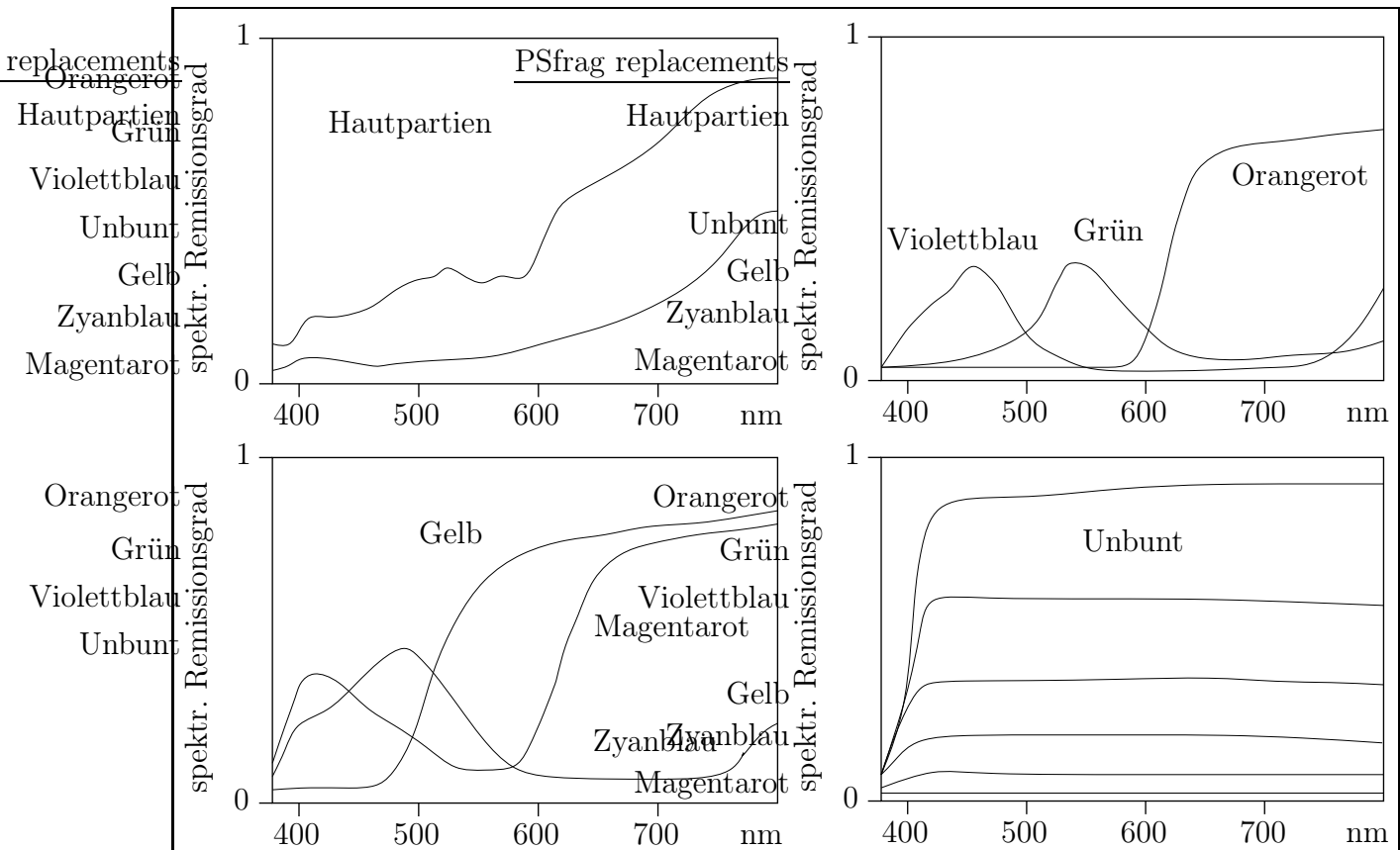


Abbildung 3.7: Remissionsfunktionen von Körperfarben (dargestellt sind prinzipielle Verläufe nach LANG /Lan1995/, S. 177). Während unbunte Körperfarben (rechts unten) durch weitgehend spektral ausgeglichene Remissionsfunktionen zustande kommen, sind die der gesättigten Farben (links unten und rechts oben) in vielen Bereichen relativ stark gekrümmt und weisen große Unterschiede auf. Dagegen unterscheiden sich die beiden links oben dargestellten Spektren von Hautfarben (leider keine näheren Angaben über die Aufnahmebedingungen) nicht qualitativ voneinander; bestimmte Unterschiede bestehen konkret in einem globalen Skalierungsfaktor (Intensität) als auch in der Gewichtung der spektralen Bereiche um 500nm und 650nm.

mehrdimensionalen Klassifikationsverfahren sind zunächst dafür prinzipiell geeignet. Die Auswahl sollte in Abhängigkeit der Komplexität des jeweiligen Referenzfarbdatensatzes, der Güteanforderungen und der zu erwartenden Einsatzbedingungen erfolgen.

Eigene Arbeiten (BRAUMANN, BRAKENSIEK ET AL. /BBBG1997/) untersuchen die Leistungsfähigkeit eines neuronalen Vektorquantisierers, der von FRITZKE /Fri1995/ unter der Bezeichnung *Growing Neural Gas Network* (GNG) entworfen wurde. Dieses GNG ist ein sehr leistungsfähiges Netzwerk und erzielt sehr genaue Klassifikationsergebnisse. Nachteilig kann sich jedoch der hohe Rechenaufwand beim GNG auswirken. Er ist bedingt durch eine z.T. erhebliche Anzahl von im Wachstumsprozeß entstehenden Neuronen. Hinzu kommt, daß sich der Vorteil des GNG, die Kopplung von datengetriebener Clusterbildung mit der Bewahrung von topologischen Relationen zwischen den Clusterzentren, für die Problemstellung Hautfarbsegmentierung als nicht entscheidend erwiesen hat.

LITTMANN und RITTER, H. /LR1995, LR1997/ führen eine Gegenüberstellung von statistischen und neuronalen Methoden gerade am Beispiel der Hautfarbsegmentierung durch, ohne jedoch entscheidende Unterschiede im Leistungsvermögen bei den von ihnen gewählten Klassifikatoren, des *multimodalen Normalverteilungsklassifikators* (multivariate Gaußfunktionen) vs. *lokal linearer Abbildungen* (LLM) festzustellen. In Arbeiten von RITTER, W. /Rit1992, Rit1997/ zur farbbasierten Verkehrszeichensegmentierung werden Polynomklassifikatoren nach SCHÜRMANN /Sch1977/, Bayes-Klassifikatoren sowie k -nächste-Nachbarn-Klassifikatoren verwendet, alles leistungsfähige Methoden der statistischen Mustererkennung.

HUNKE /Hun1994/ setzt normierte rg -Farbwertanteilhistogramme mit einer Größe von 2^{15} Einträgen zur Hautfarbklassifikation quasi in Form einer Verweistabelle (LUT) ein. Dieser Ansatz wird jedoch als zu einfach angesehen, da hier gesichert sein muß, daß das Histogramm auch mit einer hinreichenden Datenmenge besetzt ist¹¹, um nicht einen zusätzlichen Rauscheinfluß in den Klassifikationsprozeß hineinzutragen – ein klares Argument für ein kontinuierliches Klassifikatormodell.

Deswegen entscheiden sich YANG ET AL. /YW1995, YW1996, YLW1997/ für einen unimodalen bivariaten Normalverteilungsklassifikator, den sie zur Hautfarbsegmentierung einsetzen, also dem auch in dieser Arbeit gestellten Farbsegmentierungsproblem.

Unter den genannten Gesichtspunkten (Datensatzkomplexität, Güteanforderungen, Einsatzbedingungen) sowie zusätzlich bei Forderung eines minimierten Berechnungsaufwandes kommt für eine praktische Realisierung als Kompromiß der Ansatz von YANG ET AL. mit einem bivariaten Normalverteilungsklassifikator trotz (oder gerade wegen) seiner Einfachheit in Frage. Insbesondere erscheint es darüber hinaus kaum gerechtfertigt, die Genauigkeit aufwendigerer Algorithmen nutzen zu wollen, wenn keine ähnlich aufwendigen Verarbeitungsschritte zur Farbkonstanzleistung benutzt werden. Beides zusammen wird aber im Rahmen dieser Arbeit als zu aufwendig für eine Realisierung innerhalb einer in Echtzeit demonstrierbaren Implementierung aller Auffälligkeitskomponenten angesehen. Deshalb wurde der einfache bivariate Normalverteilungsklassifikator gewählt.

Dieser Normalverteilungsklassifikator (*siehe* z. B. BISHOP /Bis1995/) beruht auf einer einzigen geschlossenen Funktion, anhand derer eine parametrische Beschreibung einer Klassenzugehörigkeit gegeben werden kann. Im Fall der Hautfarbklassifikation verkörpert diese Beschreibung die Statistik der entsprechenden Daten mittels der Parameter einer mehrdimensionalen Gaußverteilung. Die Parameter bestehen in den *Koordinaten des Zentrums* dieser Glockenkurve (Mittelwertvektor $\underline{\mu}$) sowie in ihrer *multivariaten*¹² *Ausdehnung* (Kovarianzmatrix $\underline{C}$). Die Farbdaten, die als Farbwertanteile verarbeitet werden, liegen als N

¹¹Mindestens wäre eine Closing-Operation (Dilatation gefolgt von Erosion) erforderlich.

Vektoren $\underline{\xi}_i = (r_i, g_i)^\top$, $i = 1, \dots, N$ vor. Der Mittelwertvektor $\underline{\mu}$ enthält als Elemente die arithmetischen Mittel der korrespondierenden Elemente aller Datenvektoren:

$$\underline{\mu} = \frac{1}{N} \sum_{i=1}^N \underline{\xi}_i \quad (3.18)$$

Damit können die Daten zentriert werden, so daß die Kovarianzmatrix bestimmt werden kann als:

$$\underline{\mathbf{C}} = \frac{1}{N} \sum_{i=1}^N (\underline{\xi}_i - \underline{\mu})(\underline{\xi}_i - \underline{\mu})^\top \quad (3.19)$$

Das Distanzmaß, welches als Argument in die Exponentialfunktion (Gl. 3.21) eingeht, die die multivariate Gaußfunktion beschreibt, wird als Mahalanobis-Distanz m_i^2 zwischen $\underline{\mu}$ und $\underline{\xi}_i$ bezeichnet:

$$m_i^2 = (\underline{\xi}_i - \underline{\mu})^\top \underline{\mathbf{C}}^{-1} (\underline{\xi}_i - \underline{\mu}) \quad (3.20)$$

Damit kann die bivariate Gaußverteilung beschrieben werden als:

$$o(\underline{\xi}_i) = \exp\left(-\frac{1}{2}m_i^2\right) \quad \text{mit } o(\underline{\xi}_i) \in (0, 1] \quad (3.21)$$

Abb. 3.8 zeigt die im Datensatz erfaßte Verteilung von Hautfarbwerten und die gebildete Klassenzugehörigkeitsfunktion.

Seien $\underline{\mathbf{I}}_r$ und $\underline{\mathbf{I}}_b$ 2D-Farbwertanteilmatrizen aus dem $\mathbb{R}^1$, die analog zu Gl. 3.16–3.17 aus dem dreibändigen EBU-Farbbild $\underline{\mathbf{I}}_{RGB}$ (2D-Matrix aus Vektoren aus dem $\mathbb{R}^3$, zusammengesetzt aus den Farbauszugmatrizen $\underline{\mathbf{I}}_R$, $\underline{\mathbf{I}}_G$ und $\underline{\mathbf{I}}_B$) gebildet werden als:

$$\underline{\mathbf{I}}_r = \frac{\underline{\mathbf{I}}_R}{\underline{\mathbf{I}}_R + \underline{\mathbf{I}}_G + \underline{\mathbf{I}}_B} \quad \text{bzw.} \quad \underline{\mathbf{I}}_g = \frac{\underline{\mathbf{I}}_G}{\underline{\mathbf{I}}_R + \underline{\mathbf{I}}_G + \underline{\mathbf{I}}_B} \quad (3.23-3.23)$$

Aus beiden setzt sich die 2D-Matrix $\underline{\mathbf{I}}_{rg}$ zusammen, die damit Vektoren $\underline{\mathbf{i}}_{jk}$ aus dem $\mathbb{R}^2$ enthält. Somit kann unter analoger Nutzung von Gl. 3.20 und Gl. 3.21 das Element $p_{H_{jk}}$ der zum EBU-Farbbild korrespondierenden Matrix $\underline{\mathbf{P}}_H$ (Klassenzugehörigkeit) für das Vorliegen von hautfarbenen Pixeln gebildet werden:

$$p_{H_{jk}} = \exp\left(-\frac{1}{2a}(\underline{\mathbf{i}}_{jk} - \underline{\mu})^\top \underline{\mathbf{C}}^{-1} (\underline{\mathbf{i}}_{jk} - \underline{\mu})\right) \quad (3.24)$$

Beim Einsatz zur Klassifikation ist es zweckmäßig, die jeweilige Kovarianzmatrix mit einem zusätzlichen Parameter a skalieren zu können, um damit die effektive Streuung der Gaußfunktion einstellen und damit die Sensitivität der Klassifikation und damit auch die Spezifität modifizieren zu können. Mit einem Koeffizienten $a \geq 1$ ist der Klassifikator gröber und kann unter der Annahme, daß der Mittelwertvektor $\underline{\mu}$ weiterhin gültig bleibt, Fluktuationen tolerieren, die über die im Datensatz steckende Streuung hinausgehen.

¹²hier: bivariaten

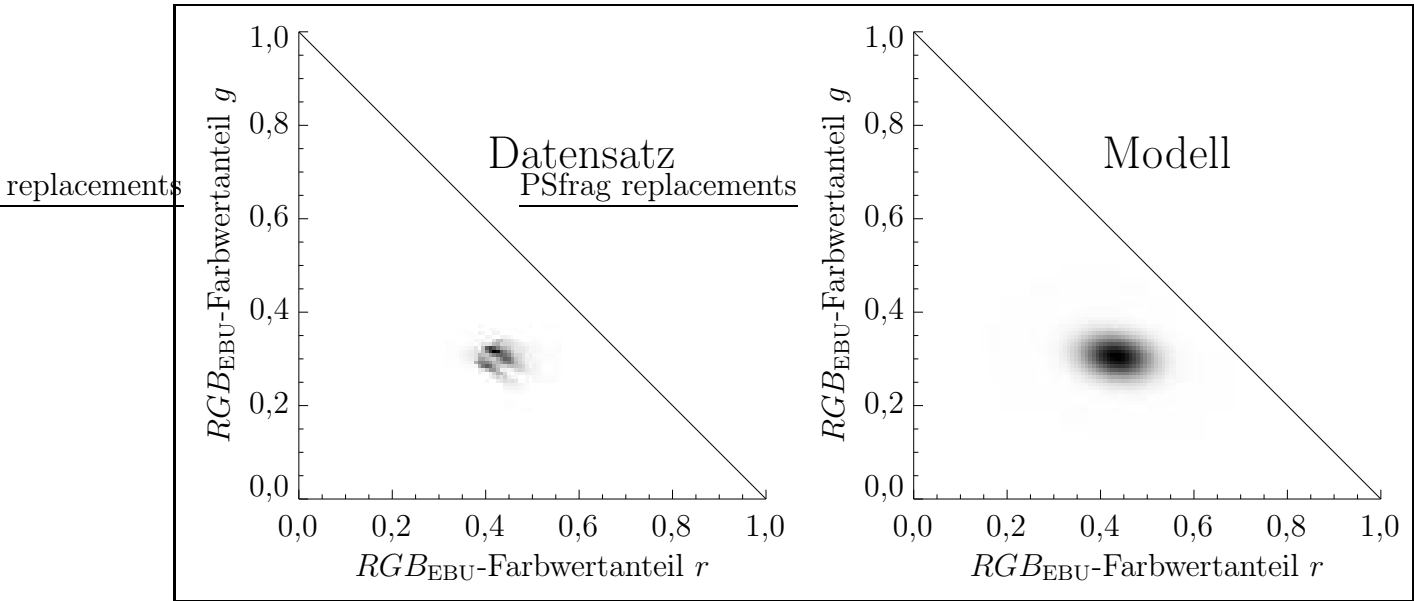


Abbildung 3.8: Hautfarbklassifikation mit einem Normalverteilungsklassifikator in der r - g -Farbwertanteilebene: links das Histogramm des Datensatzes, rechts das daraus gewonnene Verteilungsmodell. Die Farbdaten stammen aus vier unter Referenzbedingungen aufgenommenen Bildern und umfassen insgesamt 13300 RGB-Pixel, siehe auch Abb. 3.6. Die Berechnung des Mittelwertvektors für diesen Datensatz ergab $\underline{\mu} = (0,43180; 0,30078)^\top$. Die resultierende Kovarianzmatrix hat die Belegung $\underline{C} = ((0,0016; -0,00013)^\top; (-0,00013; 0,00046)^\top)$. Dargestellt ist bereits die Klassenzugehörigkeit basierend auf der um den Faktor $a = 2$ skalierten Kovarianzmatrix (siehe Gl. 3.24).

3.4 Ergebnisse

Ohne eine breit angelegte quantitative Bewertung vorzunehmen, sollen hier anhand dreier nicht für die Erstellung des Hautfarbdatensatzes verwendeter Beispielbilder die erzielten Ergebnisse qualitativ diskutiert werden. In Abb. 3.9 erscheint das Ergebnis des linken Beispiels einer Szene im Innenraum unter frontalem Halogenlicht nahezu perfekt. Da hier der Hintergrund außerdem keine Hautfarbähnlichkeit aufweist, sind (abgesehen von einer Applikation auf dem Pullover der rechten Person) keine falsch positiven Klassifikationsergebnisse zu verzeichnen.

Im mittleren und rechten Beispiel sind verschiedene Artefakte in Form sowohl leicht falsch positiver als auch falsch negativer Klassifikationsergebnisse zu finden. In diesen beiden Fällen wurde eine referenzfreie Farbumstimmung entsprechend Anhang A.2 *Elementarfarbraum-Verfahren* vorgenommen. Günstig für das Verfahren sind die ausgeprägten Unbuntbereiche in beiden Bildern. Die Artefakte treten dennoch auf, einmal weil die Korrektur nicht perfekt ist, weil z. T. deutliche Abschattungen auftreten und weil Mehrdeutigkeiten nicht beseitigt werden können. So sind im rechten Bild im rechten der Fenster im Hintergrund Spiegelungen einer anderen Hausfassade in einem dunklen gelblich-rötlichen Farbton vorhanden, die deshalb falsch positiv klassifiziert wird. Ebenso ist das violett-rötliche Hemd der Person im

Vordergrund zumindest teilweise falsch klassifiziert. Eine höhere Spezifität zur Vermeidung der Effekte würde aber nur bei gleichzeitigem Verlust an Sensitivität möglich sein – eine Wirkung, die nicht akzeptiert werden kann. Interessant am rechten Beispiel ist weiterhin, daß offenbar nur ein geringer Einfluß der Entfernung auf das Klassifikationsergebnis besteht.

Das Beispiel im mittleren Bild verkörpert einen typischen Fall: sowohl Tages- als auch Kunstlicht (Leuchtstoffröhrenbeleuchtung) treten als Mischbeleuchtung auf. Lediglich in einem kleinen Teil des Fußbodens treten Artefakte auf.

Die Beispielbilder zeigen, daß Artefakte mit dem verwendeten Methoden nicht komplett zu vermeiden sind. Die Güte wird aber vor dem Hintergrund der anderen Cues für diese Arbeit als völlig akzeptabel angesehen.



Abbildung 3.9: Beispiele von Ergebnissen der Hautfarbklassifikation im Innen- und Außenbereich. Die Beispiele zeigen, daß die Klassifikation offensichtlich trotz Anwendung eines referenzfreien Farbkonstanzverfahrens (mittleres und rechtes Beispiel) noch von den Umgebungsbedingungen abhängig bleibt. Weitere Erläuterungen werden im Text gegeben.

3.5 Résumé

Der in diesem Kapitel vorgenommene Entwurf eines einfachen Systems zur hautfarbbasierenden Klassifikation bezieht alle dafür wesentlichen Aspekte ein. Dabei erfolgt bewußt eine

Beschränkung nur auf einen Ausschnitt der prinzipiell sehr umfangreichen Problematik. Die gefundene einfache Lösung in Form einer Klassifikation mit einem unimodalen Normalverteilungsklassifikator auf Basis von Farbwertanteilen wird insbesondere durch den Kontext der Arbeit motiviert, weswegen ein einfaches und damit sehr schnelles Verfahren bevorzugt worden ist. Die erzielten guten Ergebnisse unterstreichen, daß die damit in Kauf genommenen Kompromisse klar zu rechtfertigen sind.

Potential für kommende Arbeiten besteht in folgenden Punkten:

- Hinzunahme nichtlinearer Farbraummodelle,
- vertieftere Berücksichtigung physiologischer und psychophysischer Aspekte,
- Optimierung der Bildgewinnung durch stärkere Berücksichtigung technischer Belange zur Konfiguration der Bildaufnahmetechnik.

Kapitel 4

Gesichtsdetektion

4.1 Einleitung

Neben dem im vorigen Kapitel vorgestellten Weg, Personen hautfarbbasiert zu lokalisieren, kommt in diesem Kapitel eine weitere personenspezifische Auffälligkeitskomponente für einen Multi-Cue-Ansatz hinzu. Im Mittelpunkt steht dabei die Lokalisation von Personen *anhand frontal zur Kamera ausgerichteter Gesichter*. Dafür werden spezielle Filter verwendet, die anhand von Bildern von *Gesichtsprototypen* trainiert werden. Dazu wird im folgenden ein spezieller Algorithmus vorgestellt, der im Anhang B *Eigenfaces* um eine Alternative bei Verwendung von massiv paralleler Spezialrechentechne (Neurocomputer CNAPS) ergänzt wird.

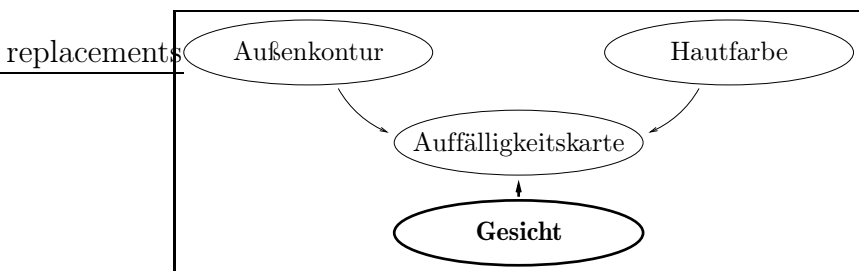


Abbildung 4.1: Der Gesichts-Cue bildet die zweite Komponente zur Kombination einer fusionierten Auffälligkeitskarte.

Eigenfaces eignen sich sowohl zur visuellen Lokalisation als auch zur Erkennung von Personen (*siehe dazu* die Arbeiten von TURK und PENTLAND /TP1991/ sowie von ROWLEY, BALUJA und KANADE /RBK1996/). Dieses Verfahren wird jedoch erst in einer Implementation auf CNAPS (s. o.) als hinreichend schnell angesehen. Eine abgewandelte Eigenface-Methode wurde in eigenen Arbeiten (BRAKENSIEK, BRAUMANN ET AL. /BBC⁺1998/) erfolgreich eingesetzt, bei der eine Kombination von Eigenfacefilterung und Klassifikation mittels neuronalem Netz (überwacht trainiertes GNG, *siehe* FRITZKE /Fri1995/) vorgenommen wurde. Auch hier sprach (neben theoretischen Überlegungen, *siehe* S. 120 im Anhang B

Eigenfaces) insbesondere die zusätzlich für das GNG benötigte hohe Rechenzeit (das GNG hätte im trainierten Zustand allenfalls mit einer Verweistabelle (LUT) ersetzt werden können) gegen den Einsatz im Rahmen dieser Arbeit.

Als gegenüber Eigenfacemethoden weniger berechnungsintensive Alternative wird hier anstelle dessen ein einfacher und leistungsfähiger Algorithmus vorgestellt, der als Methode der *normierten mittelwertfreien Skalarprodukte* bezeichnet wird. Hierbei werden zur gesichtsspezifischen Filterung andere algorithmische Schwerpunkte gesetzt, so daß die Lösung insgesamt flexibler und damit auch auf Universalrechentchnik hinreichend effizient implementierbar wird.

4.1.1 Zielstellung

Ziel der Gesichtsdetektion ist die Bildung einer Auffälligkeitskarte, in der die wahrscheinlichsten Bildkoordinaten für im Bild enthaltene frontale Gesichter bestimmter Größe markiert sind. Es wird davon ausgegangen, daß falsch positive Ergebnisse kaum zu vermeiden sind, will man den Aufwand begrenzt halten.

Psychophysische Experimente belegen die Bedeutung des visuellen Musters *Gesicht* für die menschliche visuelle Wahrnehmung insbesondere im frühkindlichen Entwicklungsstadium. BRUCE und GREEN /BG1993/ verweisen in einem Abriß zur Gesichtswahrnehmung auf Experimente von GOREN, SARY und WU /GSW1975/, die gezeigt haben, daß Neugeborene mit einem mittleren Alter von neun Minuten (!) schematische gesichtsähnliche Muster blickmäßig verfolgen. Jedoch geschieht das kaum bei Kontrollmustern, die zur Gegenüberstellung aus derselben Menge von Merkmalen bestehen, die aber räumlich umgeordnet wurden. Daraus kann geschlußfolgert werden, daß diese Neugeborenen mit einer groben Vorprägung darüber, wie Köpfe und Gesichter aussehen, zur Welt gekommen sind. Gerade dadurch können bereits frühzeitig selektiv und gezielt weitere Informationen über diejenigen Personen integriert werden, die um das Kind sorgen.

HANCOCK, BRUCE und BURTON /HBB1998/ führten eine experimentelle vergleichende Untersuchung von menschlicher und gegenwärtigem Stand maschineller Gesichtsanalyse durch. Sie bezogen dabei einerseits die Methode der *Eigenfaces* und andererseits Verfahren, die auf elastischem Graph-Matching beruhen (*siehe* V. D. MALSBURG ET AL. /WFKM1997, TM1998, Tri1999/), ein. Sie gelangen zu dem Ergebnis, daß sowohl das relativ starre Matching der *Eigenfaces* als auch das elastische Matching der Graphenstruktur für die Modellierung des menschlichen Sehens als wichtig anzusehen sind, um vergleichbares Leistungsvermögen zu erzielen. Darüber hinaus besitzen beide für sich eine biologische Plausibilität. Die Graph-Matching-Methode hat besondere Stärken gerade für Identifikati-

onsaufgaben und ist potentiell mächtiger als Eigenface-Methoden, da die Graphenstruktur auch Verzerrungen durch abweichende Ansichten kompensieren kann. Zur Merkmalsextraktion werden für das Verfahren 2D-Gaborfilter (*siehe* GÁBOR /Gáb1946/ für den 1D-Fall und DAUGMAN /Dau1980/ für den 2D-Fall) verschiedener Orientierung verwendet. Die hohe Flexibilität des Verfahrens ist jedoch mit besonders erhöhtem Berechnungsaufwand verbunden, so daß es für diese Arbeit nicht in Betracht gezogen wurde, obwohl es von STEFFENS, ELAGIN und NEVEN /SEN1998/ ebenfalls u. a. zur Personenlokalisierung benutzt wird.

Darüber hinaus stellen SIM ET AL. /SSMB1999/ ein gegenüber den genannten Verfahren erheblich vereinfachtes Prinzip zur Gesichtsanalyse vor, das zwar ausschließlich zur Gesichtserkennung eingesetzt wird, dessen Einfachheit bei hoher Leistungsfähigkeit jedoch stark überrascht. Als Distanzmaß fungiert lediglich die Summe der um eine Mindestschwelle zwischen Prototypbild und Testbild voneinander abweichenden Pixel. Diese prinzipielle Vorgehensweise konnte in dieser Arbeit nicht zur Gesichtsdetektion aufgegriffen werden, da die Veröffentlichung erst nach Abschluß der Untersuchungen erfolgte. Für künftige Arbeiten sollte das Prinzip dennoch unbedingt überprüft und ggf. aufgegriffen werden.

4.2 Gesichtsspezifische Filterung

4.2.1 Normierte gesichtsspezifische Operatoren

Obwohl die Methode der Eigenfaces nach TURK und PENTLAND /TP1991/ sowohl eine klare theoretische wie implementationsorientierte Grundlage besitzt, ist sie stark auf Hardware ausgerichtet, mit der 2D-Faltungen gut realisiert werden können, wozu gerade Additionen und Multiplikationen benötigt werden. Die Probleme der vorgenommenen CNAPS-Implementation (*siehe* Anhang C.3 *Implementation der Eigenface-Methode auf CNAPS*) führen zu Überlegungen, eine Operation zu finden, die auch auf Universalhardware akzeptabel realisierbar bleibt. Dabei sollte zur Aufwandssenkung auf mehrfache Faltungen verzichtet werden. Zusätzlich kommen auf Universalhardware Normierungen in Frage (die bei Eigenfaces allerdings nicht benötigt werden). Von Normierungen ist eine Steigerung der Tolerierung von Intensitätsabweichungen der Beleuchtung zu erwarten. Divisionen sind jedoch auf CNAPS (ebenso wie Radizierungen) sehr problematisch.¹ Auf Universalhardware sind dagegen zwar Faltungen eher aufwendig, aber Divisionen kein Problem. Deshalb wurde der folgende einfache Ansatz mit einem normierten gesichtsspezifischen Operator entworfen, der einen guten Kompromiß darstellt und im Ergebnis kaum hinter den Eigenfaces zurückbleibt².

¹Mit dem CNAPS-Rechner kann nur dividiert oder radiziert werden, wenn man das durch eigene (langsame und aufwendige) Emulationssoftware realisiert. Das allerdings ist vollkommen unpraktikabel.

Zum besseren Verständnis der nun folgenden Abschnitte gibt Abb. 4.2 einen schematischen Überblick über die Abfolge aller Schritte der gesichtsspezifischen Filterung.



Abbildung 4.2: Schritte der gesichtsspezifischen Filterung. Während in den ersten beiden Schritte für die eigentlichen normierten mittelwertfreien Skalarprodukte gebildet werden, erfolgt beginnend mit der Medianfilterung sowohl eine Störungsbeseitigung als auch die Aufbereitung für die spätere Fusion.

4.2.1.1 Datenmaterial

Das Datenmaterial an Gesichtsprototypen bestand aus 87 frontalen Gesichtern (siehe Abb. 4.3), die aus der frei zugänglichen *ORL Database of Faces*³ entnommen wurden. Die Gesichtsbilder wurden in das Format 15×15 Pixel gebracht, wobei sie oben durch die Stirn begrenzt sind, so daß kaum Haarfrisuranteile einbezogen werden. Um die Symmetrie zu sichern, ist der Datensatz durch Spiegelung aller Gesichter im Umfang verdoppelt worden.

Üblich ist es, Skalarbilder („Grauwertbilder“) aus Farbbildern durch Aufsummation aller drei Farbauszüge (ergibt Intensitätsbild) zu erzeugen. Für die vorliegende Arbeit wurde spezifischer vorgegangen. Experimente in Vorarbeiten haben ergeben, daß gerade die Gesichtsdetektion davon profitiert, *direkt* den Farbauszug $\mathbf{I}_R$ (siehe auch S. 32) als Skalarbild zu verarbeiten. Die Begründung liegt darin, daß (wie in Abb. 3.7 auf S. 30 dargestellt) die Remissionsspektren von hautfarbenen Oberflächen ein deutliches Gewicht im langwelligen („roten“) Bereich zeigen. Dadurch ergeben sich im Farbauszug $\mathbf{I}_R$ gerade in Gesichtern höhere Kontraste, was zu einer Anhebung des Signal-Rausch-Abstandes gerade auch vor dem Hintergrund wechselnder, unausgewogener und schlechter Beleuchtungsbedingungen führt. Diese Vorgehensweise steht nicht im Widerspruch zu der Tatsache, daß die Gesichtsbilder des Datensatzes (s. o.) mit sehr großer Wahrscheinlichkeit Intensitätsbilder sind (oder auch direkt mit einer SW-Kamera aufgenommen wurden), da diese unter Studiobedingungen mit optimierten Beleuchtungsbedingungen erfaßt wurden.

4.2.1.2 Normierte mittelwertfreie Skalarprodukte (NMS)

Die Idee hinter dieser Methode ist im Prinzip die einfache Cosinusmetrik. Ein betrachteter Bildausschnitt $\underline{\gamma}$ und ein gesichtsspezifischer Operand $\underline{\psi}$ seien dann identisch, wenn der

²Vgl. dazu auch Abb. 4.4 in diesem Kapitel auf S. 44 mit Abb. B.1 im Anhang B.1.2 *Detektion von Gesichtern mit Eigenfaces* auf S. 126.

³Auf diese Datenbank kann unter <http://www.cam-orl.co.uk/facedatabase.html> zugegriffen werden.

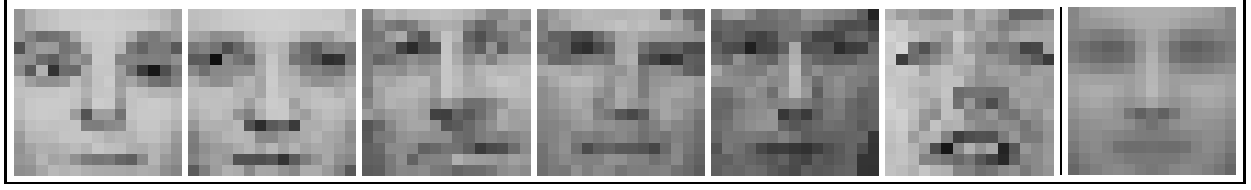


Abbildung 4.3: Beispiele der verwendeten Gesichtsprototypen aus der ORL Database of Faces. Ganz rechts ist das Durchschnittsgesicht aus allen Prototypen abgebildet. Es entspricht einem Ortsvektor zum Zentrum der Datenverteilung und damit dem Mittelwertvektor $\underline{\psi}$ (siehe Gl. B.24, S. 126).

Cosinus ihres eingeschlossenen Winkels $\cos \angle(\underline{\gamma}, \underline{\psi})$ den Wert eins annimmt. Dieser Cosinus wird bestimmt als

$$\cos \angle(\underline{\gamma}, \underline{\psi}) = \frac{\underline{\gamma}^\top}{\|\underline{\gamma}\|} \cdot \frac{\underline{\psi}}{\|\underline{\psi}\|} \quad (4.1)$$

Die euklidischen Normen der Vektoren in den Nennern stellen einen wesentlichen Punkt der Metrik dar. Ohne die Normierung ist kein klarer Wertebereich der Projektion des einen Vektors auf den anderen abgesteckt. Gleichzeitig werden die Vektoren *nur noch* anhand des eingeschlossenen Winkels verglichen, wodurch die bereits erwähnten Intensitätsabweichungen, verursacht durch geänderte Beleuchtung, stärker tolerabel werden können.

Als gesichtsspezifischer Operand $\underline{\psi}$ dient das Durchschnittsgesicht aus Gl. B.24 (siehe Abb. 4.3 rechts). Als Zentrum der Verteilung der Gesichtsprototypen ist es der geeignetste Referenzvektor für ein winkelbasiertes Distanzmaß.

Es erweist sich darüber hinaus als sinnvoll, die Operanden *mittelwertfrei* zu machen. Nur so ist eine Unabhängigkeit des gesichtsspezifischen Ähnlichkeitsmaßes von unterschiedlichen lokalen Gleichanteilen an Intensität im Bild zu erzielen. Stattdessen ist dann das Maß nur noch von (lokalen) Kontrasten im Bild abhängig. Seien die entsprechenden Mittelwerte⁴ für Bildausschnitt $\underline{\gamma}$ und für den gesichtsspezifischen Operanden $\underline{\psi}$ also wie folgt zu bestimmen:

$$\underline{\mu}_{\underline{\gamma}} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \frac{1}{N^2} \sum_{i=1}^{N^2} \gamma_i \quad (4.2)$$

$$\underline{\mu}_{\underline{\psi}} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \frac{1}{N^2} \sum_{i=1}^{N^2} \psi_i \quad (4.3)$$

⁴Die Darstellung der Mittelwerte als *Vektoren* (aus dem $\mathbb{R}^{N^2}$) erfolgt deshalb, weil die Subtraktion eines Skalars von einem Vektor (siehe Gl. 4.5) nicht definiert ist. Alle Elemente dieser Vektoren $\underline{\mu}_{\underline{\gamma}}$ bzw. $\underline{\mu}_{\underline{\psi}}$ sind identisch.

Dadurch ist gesichert, daß die Beziehungen $\sum_{i=1}^{N^2} \underline{\gamma} - \underline{\mu}_{\underline{\gamma}} = 0$ und $\sum_{i=1}^{N^2} \underline{\psi} - \underline{\mu}_{\underline{\psi}} = 0$ gelten. Dann wird damit der Cosinus des Winkels zweier somit mittelwertfreier Vektoren gebildet:

$$p_G^*(\underline{\gamma}) = \cos \angle(\underline{\gamma} - \underline{\mu}_{\underline{\gamma}}, \underline{\psi} - \underline{\mu}_{\underline{\psi}}) \quad (4.4)$$

$$= \frac{(\underline{\gamma} - \underline{\mu}_{\underline{\gamma}})^\top \cdot (\underline{\psi} - \underline{\mu}_{\underline{\psi}})}{\|\underline{\gamma} - \underline{\mu}_{\underline{\gamma}}\| \cdot \|\underline{\psi} - \underline{\mu}_{\underline{\psi}}\|} \quad (4.5)$$

Darin symbolisiert der hochgestellte Stern $\star$, daß es sich erst um ein Rohergebnis handelt; die tiefgestellte Vektorvariable weist auf den jeweiligen bezogenen Vektor hin, wobei $\underline{\gamma}$ als Bildausschnitt eben auch immer eine Abhängigkeit von einem Ortsvektor $\underline{r} = (x_1, x_2)^\top$ besitzt. Setzt man voraus, daß der gesichtsspezifische Operand bereits aufbereitet für alle $p_G^*(\underline{\gamma})$ vorliegt, dann bleiben noch ca. $4N^2$ Additionen und $2N^2$ Multiplikationen sowie eine Quadratwurzel für jedes $p_G^*(\underline{\gamma})$ zu berechnen⁵. Aus den $p_G^*(\underline{\gamma})$ als Elementen wird die zur Bildmatrix $\mathbf{I}_R$ korrespondierende Rohergebnismatrix $\mathbf{P}_G^*$ gebildet.

Mittelwertfreiheit bedeutet außerdem, daß *alle* Quadranten des Hyperraumes $\mathbb{R}^{N^2}$, in dem die Gesichtsmuster dargestellt werden, belegt sind, da auf allen Hyperraumachsen sowohl positive als auch negative Werte auftreten können. Damit werden gleichzeitig die auftretenden Winkel zwischen unterschiedlichen Mustern erheblich diversifiziert. Im Ergebnis wird eine klarere Unterscheidung von Mustern möglich.

4.2.1.3 Erhöhung der Selektivität – Beseitigung von Störungen

Bei der bisher dargestellten Vorgehensweise wird bei der Normierung (d. h. Projektion auf die Einheitshyperkugeloberfläche) der entsprechende Vektor entweder gestreckt oder gestaucht. Überlegungen zur Verbesserung der Störsicherheit des Verfahrens führen dazu, bei der Normierung eine Fallunterscheidung vorzunehmen. Anhand der Norm des Operanden $\underline{\psi} - \underline{\mu}_{\underline{\psi}}$ kann bewertet werden, wie sich die Norm des aktuellen Bildausschnitts $\underline{\gamma}(x, y) - \underline{\mu}_{\underline{\gamma}(x, y)}$ zu dieser verhält. Ist das Verhältnis $\frac{\underline{\psi} - \underline{\mu}_{\underline{\psi}}}{\underline{\gamma}(x, y) - \underline{\mu}_{\underline{\gamma}(x, y)}}$ größer eins, handelt es sich um einen Bildausschnitt mit einem geringeren Kontrast als dem des Operanden (Durchschnittsgesicht). Andernfalls ist der Kontrast von $\underline{\gamma}(x, y)$ größer (oder gleich) dem von $\underline{\psi}$. Letzterer Fall wird als unkritisch angesehen, da guter Kontrast auch entsprechende Beleuchtungsintensität voraussetzt und somit eine klare visuelle Information darstellt. Im anderen Fall aber kann gegenüber dem Operanden geringerer Kontrast immer auch als ein Hinweis auf wenig

⁵Vgl. die Angaben zur Eigenface-Methode auf S. 125 im Anhang B.1.2 *Detektion von Gesichtern mit Eigenfaces*: mit $M' = 5$ Eigenface-Filtern werden ca. $7N^2$ Multiplikationen und Additionen für jedes $p_G^*(\underline{\gamma})$ erforderlich.

Information und erhöhtes Rauschen angesehen werden.

Um davon herrührende Artefakte bei der Gesichtsdetektion zu vermindern, kann die Norm in solchen Fällen künstlich vergrößert werden, wodurch der betreffende Vektor des Bildausschnitts nicht bis auf die Einheitshyperkugeloberfläche gestreckt wird, was zur Folge hat, daß der Betrag des entsprechenden Skalarprodukts kleiner als der Cosinus des eingeschlossenen Winkels bleibt. Damit ist eine Verringerung von Störungen aus kontrastarmen Bildregionen möglich. Konkret wird vorgeschlagen, die künstliche Normvergrößerung so vorzunehmen, daß diese anschließend immer um genau *den* Betrag größer als die Operatornorm ist, um den sie zuvor kleiner war:

$$\|\underline{\gamma} - \underline{\mu}_{\underline{\gamma}}\|^* = \begin{cases} \|\underline{\gamma} - \underline{\mu}_{\underline{\gamma}}\|, & \text{wenn } \|\underline{\gamma} - \underline{\mu}_{\underline{\gamma}}\| \geq \|\underline{\psi} - \underline{\mu}_{\underline{\psi}}\| \\ 2\|\underline{\psi} - \underline{\mu}_{\underline{\psi}}\| - \|\underline{\gamma} - \underline{\mu}_{\underline{\gamma}}\|, & \text{sonst} \end{cases} \quad (4.6)$$

Diese modifizierte Norm ersetzt den entsprechenden Term in Gl. 4.5.

Ein weiterer wichtiger Nachbearbeitungsschritt zur Störungsbeseitigung zielt auf die Unterdrückung *derjenigen* lokalen Bereiche innerhalb $\underline{\mathbf{P}}_G^*$ ab, die falsch positive Ergebnisse darstellen. Gerade in den Abbn. 4.4 und 4.5 (jeweils 2. Bild v.l.) sind diese Artefakte leicht erkennbar, insbesondere entlang von Kanten mit hinreichendem Kontrast. Gleichzeitig fällt auf, daß das jeweilige Gesicht der Personen jeweils nur zu einem sehr punktuellen Eintrag in $\underline{\mathbf{P}}_G^*$ führt.

Zur Lösung dieses Problems wird zunächst ein Rangordnungsoperator (Medianfilter der Größe 3×3) eingesetzt. Mit diesem wird eine morphologische Operation ausgeführt, in deren Ergebnis gerade punktuelle Einträge eliminiert werden. Jedoch ist das Gegenteil das Ziel, d. h. es sollen gerade *punktuelle Einträge bestehen bleiben*, da diese typische Ergebnisse tatsächlich vorhandener Gesichter darstellen. Deshalb muß anschließend noch das Komplement durch Subtraktion gebildet werden. In einer zweiten Stufe wird dieses Komplement noch einer nichtlinearen Funktion unterworfen, um weiteres Rauschen unterdrücken zu können. Die hier vorgenommene Binarisierung stellt nur eine Möglichkeit dafür dar.

$$\underline{\mathbf{P}}_G^\diamond = f(\underline{\mathbf{P}}_G^* - \text{median}_{3 \times 3}(\underline{\mathbf{P}}_G^*)) \quad (4.7)$$

mit

$$f(x) = \begin{cases} 0, & x < \theta \\ 1, & x \geq \theta \end{cases} \quad \text{mit } \theta = 0,1 \quad (4.8)$$

Damit liegt nun das selektierte Ergebnis der Gesichtsdetektion als Matrix vor, die in jedem Element einen Wert enthält, der mit hinreichender Sicherheit im darum herum zentrierten Bildausschnitt einer Größe 15×15 Pixel ein Gesicht enthält (*siehe* 2. Bild v.r. in den Abbn. 4.4 und 4.5).

Mit Blick auf die weitere Verarbeitung der Cues mit dem Ziel der späteren Fusionierung erfolgt nun abschließend ein pragmatischer Schritt (*siehe* rechtes Bild in den Abbn. 4.4 und 4.5), indem der genannte zentrierte Bildausschnitt der Größe 15×15 um jedes Pixel dann auf den Wert dieses zentralen Pixels an der Stelle m, n gesetzt wird, sofern dieses zentrale Pixel einen Wert verschieden von Null hat. Anders ausgedrückt lautet das:

$$p_{G_{jk}} = \begin{cases} \max(p_{G_{mn}}^\diamond), \exists p_{G_{mn}}^\diamond \mid p_{G_{mn}}^\diamond > 0 & \text{mit } m = j - 3, \dots, j + 4 \text{ und} \\ & n = k - 3, \dots, k + 4 \\ 0, & \text{sonst} \end{cases} \quad (4.9)$$

Der sich daraus ergebende Vorteil liegt darin, daß eine Anpassung an die Charakteristik der bereits zuvor vorgestellten, flächenhaft ausfallenden Personendetektion durch Hautfarbsegmentierung erfolgt. Als Vorgriff auf das Kapitel 5 *Konturdetektion* sei hier bereits gesagt, daß auch der Cue *Kontur* (hervorgerufen durch eine bestimmte Translationsinvarianz) viel weniger lokal spezifisch arbeitet. Somit ist diese pragmatische Vorgehensweise zur dargestellten Aufbereitung des Gesichtsdetektionsergebnisses an dieser Stelle motivierbar.

Die sich aus den $p_{G_{jk}}$ zusammensetzende Ergebnismatrix wird nun mit $\underline{\mathbf{P}}_G$ bezeichnet und bildet das Endergebnis der Gesichtsdetektion, das so in die Fusion im Kapitel 6 *Personenlokalisierung* eingeht.

4.3 Ergebnisse

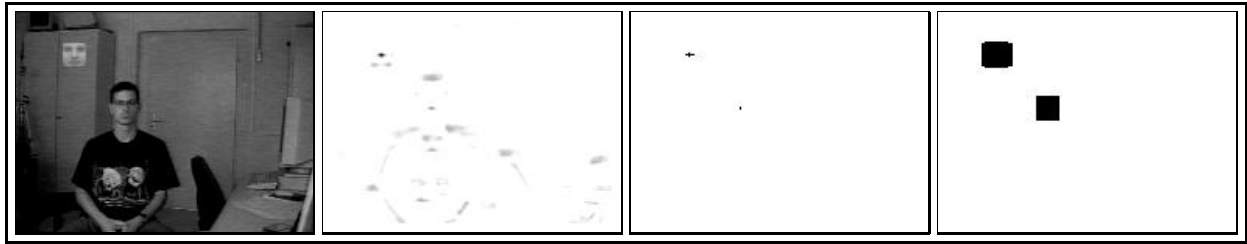


Abbildung 4.4: Stufen der Gesichtsdetektion (1): (v.l.n.r.) Skalarbild $\underline{\mathbf{I}}_R$ (Farbauszug „Rot“); Rohergebnis der Detektionsoperation $\underline{\mathbf{P}}_G^*$ (große Skalarprodukte sind dunkel, kleine hell dargestellt); Selektion punktueller Bereiche $\underline{\mathbf{P}}_G^\diamond$ (durch Medianfilterung und Komplementbildung); Markierung der dazu korrespondierenden potentiellen Gesichtsregionen $\underline{\mathbf{P}}_G$. Links oben ist mit dem „mittleren Gesicht“ der ideale Gesichtsinput als Referenz eingeblendet. Anschaulich sichtbar wird hier, daß erst die spezielle morphologische Operation zu einer akzeptablen Selektivität der Detektion führt.

Die beiden Beispiele, anhand derer die Ergebnisse der Gesichtsdetektion kurz qualitativ diskutiert werden sollen, sind jeweils unter Innenraumbedingungen aufgenommen worden. In beiden Fällen wurde in die linke obere Ecke als Referenz der gesichtsspezifische Operand $\underline{\psi}$ (als Matrix $\underline{\Psi}$) eingeblendet. Unterschiede bestehen vor allem in der Art und Richtung der Beleuchtung: In Abb. 4.4 besteht eine vorwiegend frontale Ausleuchtung der Szene, was

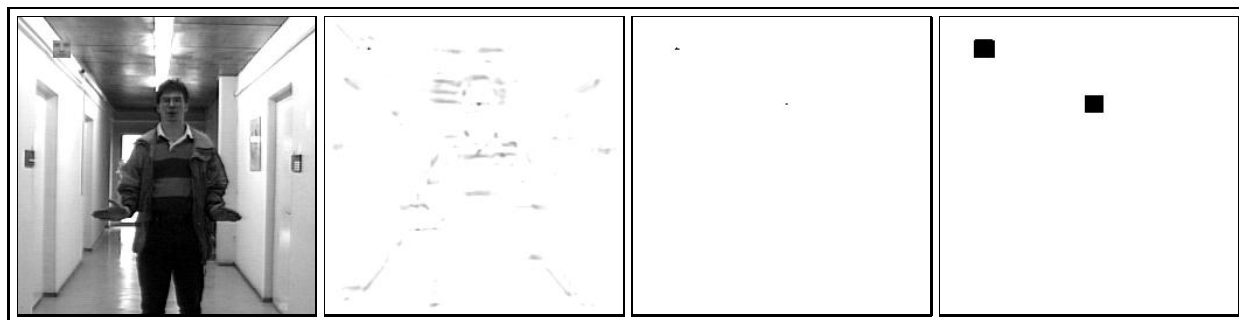


Abbildung 4.5: *Stufen der Gesichtsdetektion (2): weiteres Beispiel anhand einer gestikulierenden Person. Bedeutung der Bilder analog zu Abb. 4.4*

im Gesichtsbereich mit vermindertem Kontrast einhergeht. Dagegen weist das Beispiel in Abb. 4.5 lediglich Deckenbeleuchtung auf, was im Gesichtsbereich durch Schlagschatten den Kontrast wegen des Gesichtsreliefs anhebt. Interessant ist, daß offenbar gerade durch die zusätzlichen Verarbeitungsschritte zur Beseitigung von Störungen sehr effektiv Artefakte eliminiert werden können. Da die punktuellen Lokalisationen für die weitere Fusion nicht adäquat sind, werden diese entsprechend Gl. 4.9 genau zentriert um die Koordinaten der Gesichtslokalisation mit Quadraten in genau der Größe des gesichtsspezifischen Operanden markiert.

4.4 *Résumé*

Die in diesem Kapitel vorgenommene Entwicklung eines Detektionsverfahrens frontaler Gesichter orientierte sich zunächst an einem Algorithmus unter Verwendung sog. Eigenfaces. In den Anhängen B und C wird dieser im Detail vorgestellt. Zu große Kompromisse unter den ganz konkret vorhandenen Hardwaregegebenheiten führten zum Entwurf eines einfacheren Ansatzes mit *normierten gesichtsspezifischen Operatoren*. Die Ergebnisse belegen, daß das Verfahren trotz des bewußt einfachen Prinzips gute Ergebnisse liefert.

Für kommende Arbeiten besteht Verbesserungspotential bzw. Untersuchungsbedarf hinsichtlich folgender Aspekte:

- Überprüfung der Ergebnisse von SIM ET AL. /SSMB1999/
- Kompensation seitlicher Abschattungen aus seitlicher Beleuchtung, etwa mittels linearer Regression entsprechend des Vorschlags von ROWLEY ET AL. /RBK1996/
- Prüfung der Möglichkeiten zur weiteren Reduzierung der Auflösung, auf der dieser Cue berechnet wird.

Kapitel 5

Konturdetektion

5.1 Einleitung

In Ergänzung der beiden Auffälligkeitskomponenten, die in den beiden vorhergehenden Kapiteln vorgestellt wurden – einerseits hautfarbbasierte und andererseits gesichtsbasierte Methoden – folgt nun in diesem Kapitel mit der Vorstellung eines konturbasierten Ansatzes eine weitere Auffälligkeitskomponente. Die Idee, die hinter der Detektion einer bestimmten Kontur steht, soll Abb. 5.2 prinzipiell illustrieren. Ihr liegt eine entscheidende einschränkende Annahme zugrunde: es wird davon ausgegangen, daß eine Person, die mit einem mobilen Serviceroboter in Interaktion treten möchte, diese Absicht auch dadurch ausdrückt, daß sie sich frontal zum Roboter gerichtet aufstellt (*siehe dazu* die Erläuterungen auf S. 7 im Abschnitt 1.3 *Visuell-basierte Personenlokalisierung*).

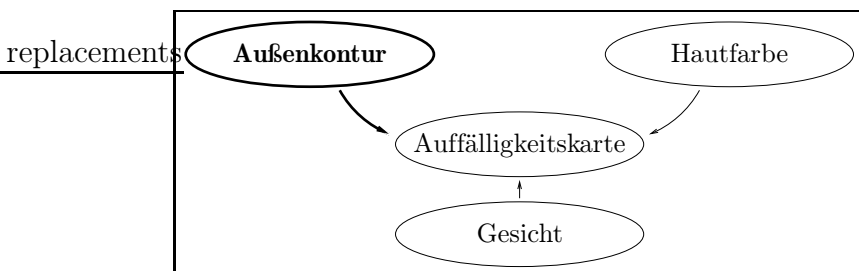


Abbildung 5.1: Der Kontur-Cue bildet die dritte Komponente zur Kombination einer fusionierten Auffälligkeitskarte.

Weiterhin hat sich gezeigt, daß unter einigen Umgebungsbedingungen durch Beleuchtungseinflüsse beispielsweise Schlagschatten entstehen, die die Anwendung der Methoden zur Gesichtsdetektion (*siehe* Kapitel 4) aber auch zur Hautfarbdetektion (*siehe* Kapitel 3) zur Personenlokalisierung anhand des Gesichts erheblich erschweren. Das Beispielbild in Abb. 5.2 (links) enthält zwar prinzipiell eine einfache Innenraumszene, allerdings wurde es offenbar relativ stark abgeblendet aufgenommen¹. Außerdem wird in der gegebenen Aufstellung der Person deren Gesicht kaum frontal beleuchtet, praktisch ein Resultat des Blendschutzes der

vorhandenen Leuchten, was zu relativ schlechten Kontrasten im Bereich des Gesichts der Person führt.

Viel makroskopischer als das Gesicht erscheint in diesem Beispiel dagegen die Außenkontur im Bereich der Kopf-Schulter-Partie. Diese hebt sich im gezeigten Beispiel zwar nicht durchgängig gut vom Hintergrund ab (etwa im Kopfbereich), bildet zumindest im Schulterbereich aber einen deutlichen Kontrast entlang einer bestimmten Kontur.



Abbildung 5.2: Prinzip der Konturdetektion zur Lokalisation einer Person anhand des Umrisses ihrer frontalen Kopf-Schulter-Partie: links ein Beispiel eines kontrastarmen Bildes einer Person, das aufgrund der konkreten Umgebungsbedingungen eines Ganges wenig intensitäts- und farbgesättigt ist; rechts die schematische Darstellung des Umrisses der Kopf-Schulter-Partie in korrespondierender Lage und Größe.

Welche Möglichkeiten bestehen nun, um diese Kontur als Ausschnitt der vorliegenden Silhouette zu erfassen und damit zur Lokalisation von Personen nutzbar zu machen? MOKHTARIAN /Mok1997/ stellt einen allgemeinen Ansatz zur silhouettenbasierten Objekterkennung vor. Dieser erfordert zunächst, die Konturlinie zu finden. Darauf aufbauend werden dann die darin lokalisierten Wendepunkte der Konturkrümmung für weitere Analysen genutzt. Eine derartige Methode erscheint jedoch nur dann sinnvoll, wenn die Konturlinie durchgängig sehr sicher und zuverlässig gefunden wird, was unter den vorliegenden Einsatzbedingungen definitiv nicht vorausgesetzt werden kann.

Ohne auf Konturlinien angewiesen sein zu müssen, wird es als praktikabler angesehen, direkt lokale Orientierungen im Bild auszuwerten. Dafür stehen Methoden der lokalen Orientierungsfilterung bzw. -messung zur Verfügung, von denen ein sehr effizientes Verfahren im Abschnitt 5.4 *Orientierungsspezifische Analyse* vorgestellt wird. Eine derartige Filterung

¹Ansonsten hätten die Leuchten die Kamera zu noch größerer Übersteuerung in einigen Bildbereichen geführt.

wird in einem folgenden Verarbeitungsschritt für eine stückweise Approximation der entsprechenden Kontur die Grundlage bilden. Der nachfolgende Abschnitt erläutert den Ansatz im Detail.

5.2 Bedeutung von Konturen

Konturen können für die visuelle Szenenanalyse ein wesentliches Merkmal von Objekten darstellen. Sie bilden die Umrisse von Silhouetten, die durch die Projektion dreidimensionaler Objekte auf die zweidimensionale Bildfläche (Netzhaut bzw. Bildwandler) entstehen. Genau genommen sind Konturen gerade eine Auswahl von linienhaft verlaufenden lokalen Intensitätsunterschieden (Kanten) zwischen Bildregionen. Eine Auswahl deshalb, weil ebenso innerhalb von Silhouetten von Objekten solche Kanten auftreten können, die aber aufgrund von *Gruppierungseffekten* (s. u.) keine weitere Rolle spielen. MARR /Mar1982/ verwendet Konturinformation als eine wichtige Komponente (neben Textur und Schattierung) in seiner Theorie vom $2\frac{1}{2}$ D-Sketch, mit der er die menschlichen visuellen Wahrnehmungseigenschaften modelliert. Konkret bezeichnet er diese Kontur als *Raw Primal Sketch*, die über differentielle Methoden bzw. Operationen gewinnbar ist. Damit liegt eine Urrepräsentation einer visuellen Szene anhand von Intensitätsunterschieden ihrer Abbildung vor, in die in erster Linie äußere Umrisse visueller Objekte eingehen. Genaue lokale Orientierungen sind dafür jedoch nicht notwendigerweise zu bestimmen.

Wie aus der sog. Gestalt-Psychologie bekannt ist (*siehe* z. B. KOFFKA /Kof1935/), nutzt die Wahrnehmung Mechanismen zur Kombination diskreter Empfindungen, um damit deren räumliche und/oder zeitliche Beziehungen bewerten zu können. Derartige Effekte werden als Gruppierungseffekte bezeichnet (*siehe* BRUCE /BG1993/). Die bevorzugte Wahrnehmung von Gestalt kann daher als elementare Eigenschaft des visuellen Systems angesehen werden. Die zugrundeliegenden Mechanismen werden auf frühen Stufen der visuellen Verarbeitung angenommen und speziellen visuellen Pfaden zugeordnet. Von einem abstrakten Standpunkt aus betrachtet treten Gestalt-Effekte gerade dort auf, wo bestimmte Qualitätskriterien bezogen auf eine Menge lokaler Merkmale erfüllt sind. Bei Konturen sind solche Merkmale orientierte Kantenstücke. Von Bedeutung ist dabei die korrekte Integration der Kantenstücke entsprechend der relevanten Kontur.

Zu den hierfür relevanten Gestaltgesetzen gehört das Gesetz der *guten Fortsetzung* für allgemeine Konturen. Hinzu kommt noch die *Symmetrie* als spezielle Eigenschaft der frontalen Kontur von Personen. Beide Gesetze erfordern die bereits erwähnten Gruppierungsmechanismen, die sogar soweit gehen können, daß im Falle von Verdeckungen eine Konturvervollständigung erfolgen kann, was damit implizit auf die Detektion von Gestalt hinweist.

Aus der neueren Literatur sollen nun kurz Beispiele von angewandter Modellierung von Konturen, z. T. aus dem Kopf-Schulter-Bereich, vorgestellt werden:

KOPECZ, J. /Kop1992/ stellt in Form *Diskreter Parametrischer Repräsentationen (DPR)* eine Darstellung natürlicher Bilder vor, mit der er Schritte der biologischen visuellen Verarbeitung in einem Modell mit einer sog. *Hyperkolumnenstruktur* nachempfunden. Die dadurch gewonnene Repräsentation sieht er als Zwischenstufe zur Analyse geometrisch dominanter Bilddetails an. Eine Modell-Hyperkolumne besteht aus einer Anzahl Modellneuronen, deren zweidimensionale rezeptive Felder allesamt eine spezielle Orientierungsselektivität aufweisen. Der Ansatz des Verfahrens wird jedoch als nicht effizient genug für einen Einsatz zur Personenlokalisierung angesehen, da für die Orientierungshistogramme (siehe Abb. 5.3) eine Vielzahl von Orientierungsfilterungen erforderlich ist.

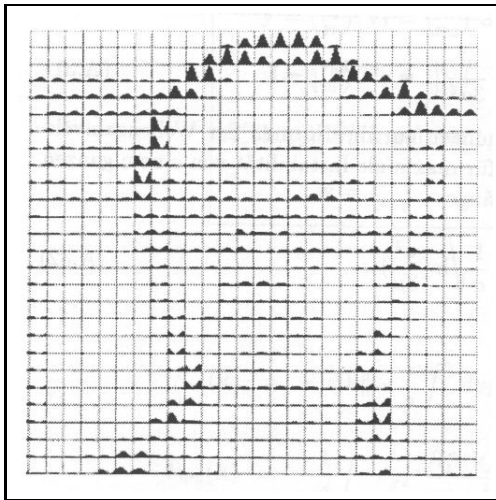


Abbildung 5.3: Diskrete parametrische Repräsentation (DPR) eines Gesichtes (entnommen aus /Kop1992/, Anhang B, S. 104).

In einem System zur visuellen Verfolgung von Personen setzen FEYERER und ZELL /FZ1999/ ein interessantes, unkonventionelles Verfahren zur Einbeziehung von Konturinformation ein, das von DENZLER /Den1997/ unter dem Begriff *Aktive Strahlen* vorgestellt wurde. Hierbei werden Grauwertverläufe *entlang* von Strahlen analysiert, die sternförmig von einem Referenzpunkt ausgehen. Ein Energieminimierungsschritt führt dann jeweils zur wahrscheinlichsten Kontur bzw. deren aktuellen Referenzpunkt. Das zweidimensionale Muster *Kontur* wird hier in viele eindimensionale Muster zerlegt, was allerdings mit entsprechendem Zerlegungs- aber auch Zusammensetzungsaufwand mittels Energieminimierung erkauft werden muß. Deshalb wurde das Verfahren für diese Arbeit nicht weiter betrachtet.

5.3 Konturmodellierung durch ein Arrangement orientierter Filter

Für die vorliegende Arbeit wurde eine sehr effiziente Methode zur Konturdetektion entwickelt, die unter der Bezeichnung *Arrangement orientierter Filter* nachfolgend vorgestellt wird. Dabei handelt es sich um ein einfaches Verfahren im Ortsraum, mit dem die in Abb. 5.2 (rechts) schematisch dargestellte Kontur approximiert wird (siehe Abb. 5.5).

Ähnlich wie bei KOPECZ' DPR's wird eine Analyse von lokaler Orientierung vorgenommen, jedoch auf die Erstellung von Orientierungshistogrammen (siehe Abb. 5.3) verzichtet. Dafür wird sich auf die Bestimmung des lokal dominanten Orientierungswinkels beschränkt. Entlang des zuvor ermittelten Konturmodells der Kopf-Schulter-Partie wird ein Fitmaß gebildet, das sich aus der Summe einzelner lokaler Fitmaße zusammensetzt. Benötigt wird dafür ein Verfahren zur Bestimmung des dominanten Orientierungswinkels in einer lokalen Umgebung².

Da das Konturmodell als Arrangement orientierter Filter gerade *Ortsrelationen* von lokalen Umgebungen verschiedener fester Orientierungen repräsentieren soll, ist die Verarbeitung im Ortsraum selbst am besten für die Bestimmung des Fitmaßes geeignet.

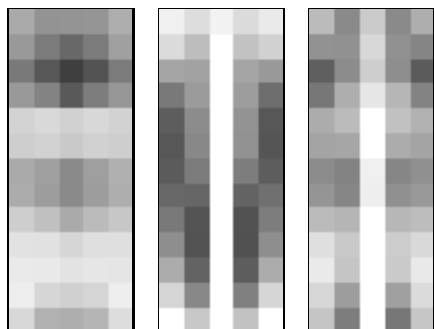


Abbildung 5.4: *Fitmaps von Haar Wavelet Templates nach OREN ET AL.: Ein derartiges Tripel an Templates aus Koeffizienten horizontaler (links), vertikaler (Mitte) und schräger (rechts) Haar-Wavelets nutzen OREN ET AL. (entnommen aus /OPS⁺1997/). Mit einiger Phantasie läßt sich insbesondere im mittleren Template erkennen, daß darin offenbar die vertikal verlaufenden Konturbereiche einer Personensilhouette repräsentiert werden. Auch die beiden anderen Templates (links: horizontal, rechts: schräg) können analog anschaulich interpretiert werden.*

Interessante Bezüge bestehen zu Arbeiten von OREN, PAPAGEORGIOU ET AL. /OPS⁺1997, POP1998/, die parallel ebenfalls einen Ansatz zur konturbasierten Personenlokalisierung entworfen haben. Hierbei wird anhand sog. *Haar Wavelet Templates* der komplette Umriss von Personen (Fußgänger) modelliert (siehe Abb. 5.4). Der gemeinsame Aspekt besteht darin, daß ebenfalls räumlich verteilte lokale orientierte Filter genutzt werden. Diese sind allerdings in Form von Haar-Wavelets äußerst einfach gehalten. Ebenfalls interessante Bezüge bestehen zu Arbeiten von MELLER /Mel1997b/, der u. a. mittels verschiedener git-

²Der dominante Orientierungswinkel entspricht qualitativ dem Winkel des Maximums in den Orientierungshistogrammen von Kopecz' DPR's.

terartig angeordneter Gaborfilter (s. u.) bestimmte Teilmerkmale von Personen extrahiert, deren räumliche Relationen zueinander zur Personenlokalisierung verwendet werden.

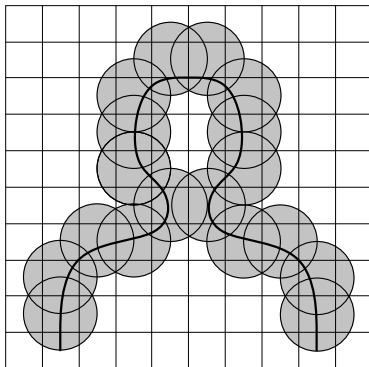


Abbildung 5.5: Schema der prinzipiellen Anordnung der orientierten Filter.

Der Einsatz des hier entwickelten Filterarrangement erfolgt wie für die anderen Cues mit fünf diskreten Abstufungen in einer Auflösungspyramide, so daß die Lokalisation auch hinsichtlich der Entfernung der Person erfolgen kann. Eine Eigenschaft einer idealen Konturkante im Bild ist, daß diese spektrale Anteile enthält, die umgekehrt proportional zur Wellenzahl k abfallen, so daß eine ideale Kante dementsprechend zunächst auf allen Ebenen der Pyramide abgebildet wird. Jedes orientierte Filter bringt eine lokale Unschärfe mit und hat gleichzeitig ein frequenz- und orientierungsmäßiges Bandpaßverhalten. Das führt für das gesamte Arrangement ebenfalls zu einer Art Unschärfe mit Bandpaßverhalten, so daß das Filterarrangement mindestens auf benachbarten Pyramidenebenen zu lokal koinzidenten Fitmaßen führt.

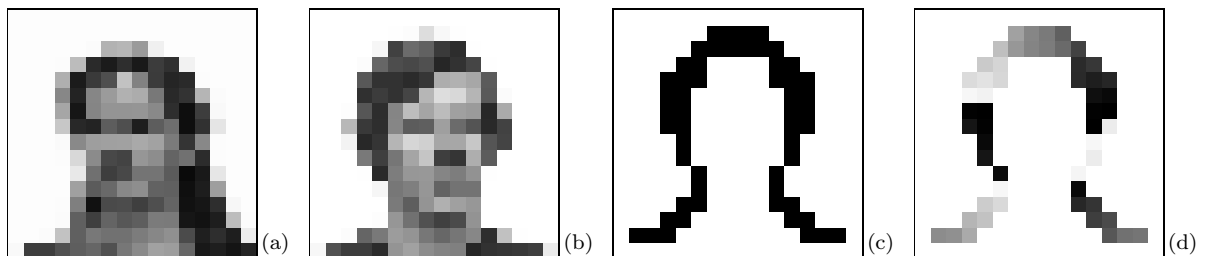


Abbildung 5.6: Prinzip der Gewinnung der Filterarrangements zur Detektion der Kopf-Schulter-Partie. Von links nach rechts: Zwei Beispielbilder (a) und (b) der Größe 16×16 aus dem Datensatz; (c) die daraus bestimmte äußere Kontur $\text{supp}(\underline{\mathbf{A}})$; (d) der Orientierungswinkel kodiert in Grauwerten $\underline{\mathbf{A}}$ (0° : Schwarz; 90° : Mittelgrau; 180° : Weiß). Man beachte, daß der Orientierungswinkel senkrecht zur Kante steht, so daß gerade im Scheitelbereich Winkel von ca. 90° auftreten. Im Bereich der Ohren treten dementsprechend Winkel von nahe 0° bzw. 180° auf, weswegen in der Darstellung in diesem Bereich sowohl schwarze als auch weiße Felder vorkommen. (vgl. auch mit Abb. 5.9 auf S. 62.) Die Kontur ist symmetrisch, da die Originalbilder jeweils zusätzlich gespiegelt vorlagen.

Das Datenmaterial zur Erstellung eines Kontur-Prototypen besteht aus unter identischen Bedingungen vor einem homogenen Hintergrund aufgenommenen Portraits von insgesamt

zehn Personen. Jede Person wurde direkt frontal sowie jeweils $(+10^\circ)$ und (-10°) davon abweichend aufgenommen. Um ein symmetrisches Konturmodell zu erhalten, wurde aus jedem dieser Bilder noch eine horizontal gespiegelte Kopie erzeugt, so daß insgesamt 60 Bilder zur Verfügung standen. Diese Skalarbilder im Format 256×256 wurden entsprechend tiefpaßgefiltert und auf eine Größe von 16×16 unterabgetastet. Anschließend wurden mit einem Sobel-Operator die Kanten dieser Bilder hervorgehoben und schließlich über alle diese Kantenbilder gemittelt. Mit einer einfachen Schwellenoperation erfolgte durch eine Binarisierung die Bestimmung eines repräsentativen Konturverlaufs, der im Mittel einer Kontur der Kopf-Schulter-Partie der Personen aus dem Datensatz entspricht. Die optimale Schwellenhöhe wurde bestimmt, indem sie in einem automatisierten Verfahren solange angehoben wurde, bis durch die Binarisierung nur noch *ein einzelnes, lückenlos zusammenhängendes, binäres Kontursegment* entstand.

Damit liegt der interessierende Konturverlauf in Form einer 16×16 -Binärmatrix $\text{supp}(\underline{\mathbf{A}})$ vor. Was noch offen ist, ist die lokale Orientierung der einzelnen Konturabschnitte. Diese wird bestimmt mittels orientierungsspezifischer Analyse (s. u.) der Binärmatrix $\underline{\mathbf{A}}$. Somit wird jedem Element $\lambda_{i,j}$ ein Wert zugewiesen, der dem in der Nachbarschaft von i, j dominanten Orientierungswinkel entspricht.

Im folgenden wird ein Maß für das Vorhandensein der *visuellen Auffälligkeitskomponente Kopf-Schulter-Partie*, die durch das Template $\underline{\mathbf{A}}$ repräsentiert wird, innerhalb der Matrix $\underline{\Phi}$ gebildet. Dafür wird ein Matchingverfahren genutzt, das auf einem *Ähnlichkeitsmaß* $p_{K_{jk}}$ basiert. Wegen der π -Periodizität des Ergebnisses der orientierungsspezifischen Analyse einerseits, und um andererseits geeignet die „Ähnlichkeit“ zweier Elemente von $\underline{\mathbf{A}}$ und $\underline{\Phi}$ (siehe Gl. 5.8) zu bestimmen, benötigt die Ähnlichkeitsfunktion dieselbe Periodizität.

$$p_{K_{jk}} = \frac{\sum_{l=0}^{L-1} \sum_{\substack{m=0 \\ \lambda_{lm} \neq 0}}^{M-1} \frac{1}{2} \left[\cos \left(2 \left| \lambda_{lm} - \phi_{(j+l-\frac{L}{2})(k+m-\frac{M}{2})} \right| \right) + 1 \right]}{\text{card}(\text{supp}(\underline{\mathbf{A}}))} \quad (5.1)$$

Hierin bezeichnet λ_{lm} ein Element aus $\underline{\mathbf{A}}$ mit den Indizes l und m sowie $\phi_{(j+l-\frac{L}{2})(k+m-\frac{M}{2})}$ ein Element aus $\underline{\Phi}$ an der Stelle $j+l-\frac{M}{2}, k+m-\frac{M}{2}$. $L = M = 16$ sind die Dimensionen der Matrix $\underline{\mathbf{A}}$. Die Normierung auf die Kardinalität des Supports³ von $\underline{\mathbf{A}}$ sichert $p_{K_{jk}} \in [0, 1]$ für die nachfolgenden Verarbeitungsschritte.

Aus $p_{K_{jk}}$ setzt sich nun die zum Skalarbild $\underline{\mathbf{I}}_G$ (siehe dazu die Diskussion auf S. 60 sowie die Darstellung zur Bilddimensionalität auf S. 32) korrespondierende Matrix $\underline{\mathbf{P}}_K$ zusammen, so daß dieser Cue darüber hinaus ebenso zu den in den vergangenen Kapiteln behandelten Cues korrespondiert. Damit bestehen die Voraussetzungen für eine erfolgreiche Fusion (siehe

³Der Support einer Matrix selektiert nur diejenigen Elemente daraus, die verschieden von Null sind.

Kapitel 6 *Personenlokalisierung*).

Der nachfolgende Abschnitt behandelt nun ausführlich die orientierungsspezifische Analyse von $\mathbf{I}_G$ und damit die Bildung der Matrix $\underline{\Phi}$, die Voraussetzung für Gl. 5.1 ist.

5.4 Orientierungsspezifische Analyse

Zunächst als Begriffsbestimmung: es ist zu trennen zwischen Richtung und Orientierung, die umgangssprachlich teilweise synonym verwendet werden. Richtung im zweidimensionalen ist ein Maß mit Werten aus dem Intervall $[0, 2\pi)$. Im diesem Sinne sind beispielsweise Gradienten *gerichtet*. Orientierung dagegen umfaßt nur Werte aus dem Intervall $[0, \pi)$. Kanten sind in diesem Sinne *orientiert*. Dabei steht der Orientierungsvektor einer Kante auf dieser senkrecht. In allen Darstellungen in den folgenden Abschnitten werden diese Begriffe gemäß dieser Vereinbarung verwendet.

5.4.1 Generelle physiologische Bezüge und mathematische Aspekte orientierungsspezifischer Analyse

Eindeutige Hinweise auf orientierungsspezifische Analyseelemente in natürlichen Sehsystemen sind lange bekannt. Deshalb soll in diesem Abschnitt zunächst auf dazu bekannte Fakten eingegangen werden, um die Sinnfälligkeit der Verwendung derartiger Methoden für die vorliegende Arbeit deutlicher hervorzuheben.

Bisher hat eine Vielzahl von Arbeiten gezeigt (z. B. von EDELMAN /Ede1995/ oder von OLZAK und THOMAS /OT1986/), daß das Profil der rezeptiven Felder (RF) von einfachen Zellen im primären visuellen Cortex von Säugern durch bestimmte zweidimensionale mathematische Funktionen beschrieben werden kann. *Gaborfunktionen*, mit denen bspw. JONES und PALMER /JP1987a, JP1987b/ (*siehe* Abb. 5.7) oder WEBSTER und DE VALOIS /WD1985/ arbeiten, sowie *Gaußfunktionen* (einschließlich Ableitungen niedriger Ordnung), die etwa KOENDERINK und VAN DOORN /KD1990/ oder YOUNG /You1986, You1987/ verwenden, erscheinen insbesondere zur Modellierung typischer Profile visueller rezeptiver Felder als gut geeignet.

Andere Arbeiten, wie die von OLSHAUSEN und FIELD /OF1996/, zeigen, wie durch die Auswertung der Statistik natürlicher Bilder lokale Operatoren derart adaptiert werden können, daß sie ebenfalls sehr ähnlich den Profilen der rezeptiven Felder einfacher Zellen werden.

KOENDERINK und VAN DOORN /KD1990/ nehmen eine Unterteilung von drei linearen

Modellfamilien rezeptiver Felder vor: Modelle in *kartesischen* Koordinaten (aus Hermite-schen Polynomen zusammengesetzt), Modelle mit Wavelets (einem Spezialfall der karte-sischen Modelle), bei denen die Hermite-schen Funktionen mit Linearkombinationen von goniometrischen Funktionen moduliert werden, sowie Modelle in *polaren* Koordinaten. Die Verwandtschaft aller dieser Modellfamilien ist jedoch sehr groß, da alle durch Lineartransfor-mationen ineinander überführt werden können. Der Vorteil dieser drei Familien ist, daß man dadurch die RFs mathematisch sehr bequem an verschiedene Bedürfnisse von RF-Charakte-ristiken, die durch ganz spezielle Stimuli entstehen können, anpassen kann, indem man die Beschreibungsweise nach Bedarf wechselt. Dieses Vorgehen ist deshalb gerechtfertigt, weil zu allen drei Familien Belege für deren prinzipielle Existenz in Form von Familien realer RFs angeführt werden.

Die lokale Ausrichtung der Zellen bezogen auf die Retina kann prinzipiell im gesamten Bereich $[0, 2\pi)$ liegen (*siehe* /JP1987a/), und es scheint Häufungen in bestimmten Rich-tungen zu geben, was nicht verwundert, da durch die natürliche und künstliche Umwelt be-stimmte Vorzugsrichtungen bzw. -orientierungen (vertikal, horizontal) existieren. Derartige lokale „Operatoren“ ermöglichen es, eine spektrale Zerlegung unter besonderer Berücksich-tigung der Orientierung vorzunehmen, indem verschiedene Frequenzbänder, Orientierungen und Phasenlagen erfaßt werden.

POLLEN und RONNER /PR1981/ stellten fest, daß benachbarte einfache Zellen häufig einen Phasenversatz von $\frac{\pi}{2}$ aufweisen. Diese Beobachtung ist dann plausibel, wenn man berücksichtigt, daß die Verbindung zweier benachbarter, um 90° verschobener Zellen prin-zipiell eine weitgehende Unabhängigkeit des lokalen Analyseergebnisses von der Phasenlage des Stimulus erlaubt (Quadraturfilter).

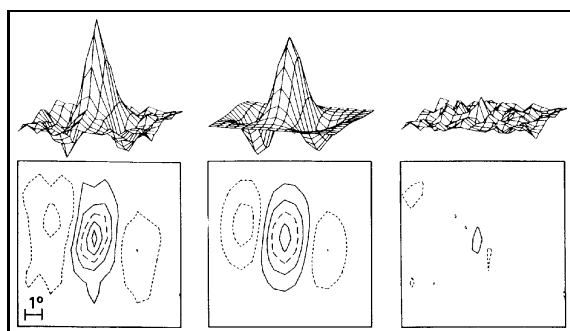


Abbildung 5.7: Modellierung des rezepti-ven Feldes einer einfachen Zelle im visuellen Cortex von Katzen mittels 2D-Gaborfunktion (v. l. n. r.): Empfindlichkeitsprofil der Zelle, Mo-dellierung mittels Gaborfunktion und Differenz zwischen Meßergebnis und Modell (Ergebnisse von JONES/PALMER /JP1987a/, entnommen aus /Pöt1994/, S. 15).

An o. g. Eigenschaften einfacher Zellen lassen sich wie folgt zusammenfassen:

- *Linearität* bezüglich der Stimuli
- *Mittelwertfreiheit*, d. h. kein Signal bei reinem Gleichanteil
- *Orientierungsspezifität*

- *Lokalisierbarkeit im Ortsraum*
- *Lokalisierbarkeit im Frequenzraum*

Alle diese Eigenschaften sind in einer technisch-algorithmischen Umsetzung von orientierungsspezifischer Filterung vor dem Hintergrund der Konturdetektion von Bedeutung. Für einige Punkte bestehen höhere Anforderungen, für andere geringere. Beispielsweise besteht für diese Arbeit kein Zwang für ein lineares Antwortverhalten bzgl. der Stimuli, wichtig ist lediglich Monotonie und eine bestimmte Signifikanz. Kompromisse bzgl. der genannten Eigenschaften einfacher Zellen werden immer dort als gerechtfertigt angesehen, wo bei praktisch immer vorhandenen Beschränkungen der verwendeten Hardware und insbesondere zugleich bei Ausnutzung ihrer jeweiligen Stärken Ergebnisse erzielt werden können, die gütemäßig akzeptiert werden können.

Sowohl Gaborfilter (*siehe* GÁBOR /Gáb1946/, insbesondere die zweidimensionale Erweiterung von GÁBORs Ansatz durch DAUGMAN /Dau1980/) als auch Steuerbare Filter (*siehe* FREEMAN /FA1991, Fre1992/) sind in den Vorarbeiten zu dieser Arbeit mit Erfolg zur Merkmalsextraktion für die Konturdetektion verwendet worden (*siehe* BRAUMANN und CORRADINI ET AL. /BBBG1998, CBB⁺1998, CBBG1999, BBBG1998, CBBG1998/). Beide Verfahren, die nach dem Quadraturfilterprinzip (Filterpaar) funktionieren, haben sich jedoch als zu ineffizient und damit unpraktikabel herausgestellt, da die jeweiligen Rechenzeiten zu groß geblieben sind und damit der Datendurchsatz selbst bei kleinen Bilddimensionen für Echtzeitanalysen nicht ausreichte. Mit dem nachfolgend vorgestellten Verfahren ist es demgegenüber viel besser gelungen, den Anforderungen bei hinreichend hoher Genauigkeit und gleichzeitig akzeptabler Rechenzeit zu genügen.

5.4.2 Nichtlineare lokale Strukturanalyse zur Orientierungsbestimmung

Ein schwerwiegender Nachteil der Quadraturfilter-Technik besteht zunächst im quasi doppelten Rechenaufwand wegen des zu verwendenden Filterpaars. Hinzu kommt, daß es sich Filterungen in diskreten Orientierungen handelt, so daß sich ein zweiter Schritt zur Bestimmung (Interpolation) einer effektiven Orientierung anschließen muß.

Ein modernes, alternatives Verfahren zur Orientierungsfilterung, was demgegenüber sehr viel weniger aufwendig ist, stellt die sog. 2D-Tensor-Methode dar. JÄHNE hat dieses Verfahren in verschiedenen Monographien/Jäh1993, Jäh1995, JMNS1996, Jäh1997/ vorgestellt und bezieht sich darin auf Arbeiten von GRANLUND /Gra1978/ und insbesondere von BIGÜN /BG1988/. Eine spezielle Darstellung der Methode gibt JÄHNE außerdem in /Jäh1998/.

Sie wird bisher noch nicht in größerem Umfang eingesetzt, was jedoch für die nähere Zukunft erwartet werden kann. Für die kompletten theoretischen Grundlagen sei auf Anhang D *Nichtlineare Strukturanalyse – Theoretische Grundlagen* verwiesen. Bevor im folgenden Abschnitt auf die Umsetzung der Methode eingegangen wird, soll zunächst ihr Prinzip erläutert werden.

5.4.2.1 Prinzip und Umsetzung der 2D-Tensor-Methode

Das Verfahren bietet zwei Betrachtungsmöglichkeiten: anschaulich läßt es sich im Fourierraum erläutern, realisierbar ist es aber besonders effektiv im Ortsraum selbst. Ausgegangen werden soll von einer lokalen Nachbarschaft (*siehe* Abschnitt D.1 *Lokale Nachbarschaft* im Anhang), die sich dadurch auszeichnet, daß darin eine exakte Kohärenz der Orientierung und damit eine ideale Orientierung vorliegt.

Fouriertransformiert man diese lokale Nachbarschaft, so erhält man ein exakt linienhaft durch den Koordinatenursprung verlaufendes Spektrum. Voraussetzung dafür ist, daß darin ein spektrales Gemisch vorliegt. Da angenommen wird, daß dieses Spektrum nach oben begrenzt ist, wird aus der Geraden eine Strecke. Der Winkel, den dieses Linienspektrum mit der k_1 -Achse des zweidimensionalen Fourierraumes einnimmt, entspricht aber genau dem Orientierungswinkel, den die lokale Nachbarschaft auch im Ortsraum aufweist.

Bei nicht exakt kohärenter Orientierung in der Nachbarschaft ergibt sich nun kein exakt linienhaftes Spektrum mehr. Der Ansatz der Tensormethode liegt einfach darin, durch Geradenfit die Hauptachse dieses nunmehr bandförmigen Spektrums und damit zugleich die dominierende Orientierung in dieser Nachbarschaft zu bestimmen.

Es besteht hierbei eine direkte Analogie zur Mechanik des starren Körpers, sofern man das Spektrum modellhaft als Fläche inhomogener Dichte auffaßt. Dieser spezielle Körper läßt sich gerade mit minimalem Trägheitsmoment um seine Hauptachse rotieren. Das allgemeine 2D-Trägheitsmoment eines flächenhaften Körpers wird bestimmt als:

$$J = \begin{pmatrix} k_1 & k_2 \end{pmatrix} \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} \quad (5.2)$$

Darin beschreibt der Vektor $\underline{k} = \begin{pmatrix} k_1 & k_2 \end{pmatrix}^\top$ genau die (zunächst beliebige) Rotationsachse. Die Matrix

$$\underline{\mathbf{J}} = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \quad (5.3)$$

stellt eine Beschreibung der Dichte (und darin eingeschlossen der Geometrie) des modellhaften Trägheitskörpers dar und wird als Trägheitstensor bezeichnet. Die einzelnen Elemente

setzen sich aus Integralen über k_1 und/oder k_2 zusammen, deren Integranden Produkte von k_1 bzw. k_2 mit der Dichtefunktion des Körpers bilden. Das Trägheitsmoment J ist genau dann minimal, wenn $\underline{k}$ die Rotationsachse minimaler Trägheit verkörpert.

Wird nun das Koordinatensystem so gedreht, daß die Rotationsachse $\underline{k}$ genau mit der k_1 -Achse zusammenfällt, benötigt man zur Beschreibung des Trägheitsmoments desselben Körpers einen Trägheitstensor $\underline{\mathbf{J}}'$, dessen Nebendiagonalelemente zu Null werden. Der Winkel ϕ , um den das Koordinatensystem zu drehen ist, beschreibt aber selbst genau die Lage der Rotationsachse minimalen Trägheitsmoments. Gesucht ist also gerade eine Transformationsmatrix

$$\underline{\mathbf{T}} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \quad (5.4)$$

zur Beschreibung der Rotation des Koordinatensystems. Nun kann prinzipiell nach ϕ aufgelöst werden.

Diese Betrachtung erfolgte bisher im Fourierraum. Das würde nun jedoch in der Praxis eine sehr große Zahl von nachbarschaftsweisen numerischen Fouriertransformationen erfordern, um dann die Komponenten des „Trägheitstensors“ der jeweiligen Nachbarschaft zu bestimmen. Dem kommt entgegen, daß das gesamte Problem – sogar noch einfacher – auch direkt im Ortsraum handhabbar ist. Die Integrale, aus denen sich der Tensor zusammensetzt, können auch im Ortsraum gelöst werden, da die Fouriertransformation eine unitäre Transformation ist, die die Norm einer Funktion dabei erhält. Zugleich werden (durch formale Fourierrücktransformation) aus den Produkten der k_1 und k_2 mit der Dichtefunktion partielle Ableitungen der Originalfunktion nach x_1 bzw. x_2 . Soweit die kurze Darstellung des Prinzips.

Als Lösung von JÄHNES Methode liegt folgende Vorschrift (Herleitung in Anhang D.4 *Lösung des Eigenwertproblems*) zur Bildung des Tangens des doppelten Orientierungswinkels vor (siehe auch Gl. D.18):

$$\tan(2\phi(\underline{\mathbf{x}})) = \frac{2J_{12}(\underline{\mathbf{x}})}{J_{22}(\underline{\mathbf{x}}) - J_{11}(\underline{\mathbf{x}})} \quad (5.5)$$

In diese prinzipielle Lösung müssen nun für eine Umsetzung des Ansatzes drei Tensorelemente eingesetzt werden. Sie werden prinzipiell entsprechend Gl. D.9 bestimmt. Die partiellen Ableitungen werden ausgeführt durch Faltungen mit diskreten Ableitungsoperatoren in Richtung von x_1 und x_2 (für den diskreten Fall nunmehr als p und q bezeichnet), und die Integration korrespondiert zu einer Faltung mit einem Glättungsoperator, der damit die Fensterfunktion darstellt. Zur Notation der einzelnen Schritte wird jetzt die Operator-schreibweise im Ortsraum verwendet. Ein vorverarbeitetes Bild $\underline{\mathbf{I}}^*$ (siehe Gl. 5.18 auf S. 61) wird gefaltet mit den diskreten Ableitungsoperatoren $\underline{\mathbf{D}}_p$ und $\underline{\mathbf{D}}_q$ und mit einer Binomialmaske $\underline{\mathbf{B}}$ geglättet. Daraus bildet man die zum (vorverarbeiteten) Bild korrespondierende

Tensorelementmatrix $\underline{\mathbf{J}}_{pq}$ nach:

$$\underline{\mathbf{J}}_{pq} = \underline{\mathbf{B}} * ((\underline{\mathbf{D}}_p * \underline{\mathbf{I}}^*) \cdot (\underline{\mathbf{D}}_q * \underline{\mathbf{I}}^*)) \quad (5.6)$$

Dabei ist zu beachten, daß die Notation $\cdot$ das sog. Schurprodukt symbolisiert, das als *elementweise* Multiplikation der jeweiligen Matrizen gebildet wird. Die Notation $*$ bezeichnet eine Faltungsoperation. Entsprechend Gl. 5.6 kann das zu $\underline{\mathbf{I}}^*$ korrespondierende Feld an Vektoren mit Richtung des doppelten Orientierungswinkels gebildet werden⁴:

$$\begin{pmatrix} \underline{\mathbf{B}} * ((\underline{\mathbf{D}}_q * \underline{\mathbf{I}}^*) \cdot (\underline{\mathbf{D}}_p * \underline{\mathbf{I}}^*) - (\underline{\mathbf{D}}_p * \underline{\mathbf{I}}^*) \cdot (\underline{\mathbf{D}}_q * \underline{\mathbf{I}}^*)) \\ 2\underline{\mathbf{B}} * ((\underline{\mathbf{D}}_p * \underline{\mathbf{I}}^*) \cdot (\underline{\mathbf{D}}_q * \underline{\mathbf{I}}^*)) \end{pmatrix} \quad (5.7)$$

Der Betrag eines solchen Vektors kann als Konfidenzmaß angesehen werden, das angibt, in welchem Maße in der lokalen Nachbarschaft enthaltene orientierte Strukturkomponenten ausgeprägt sind.

Entsprechend gelangt man zum Feld der Tangens' des doppelten Orientierungswinkels ϕ :

$$\tan(2\underline{\Phi}) = \left(2\underline{\mathbf{B}} * ((\underline{\mathbf{D}}_p * \underline{\mathbf{I}}^*) \cdot (\underline{\mathbf{D}}_q * \underline{\mathbf{I}}^*)) \right) \div \left(\underline{\mathbf{B}} * ((\underline{\mathbf{D}}_q * \underline{\mathbf{I}}^*) \cdot (\underline{\mathbf{D}}_p * \underline{\mathbf{I}}^*) - (\underline{\mathbf{D}}_p * \underline{\mathbf{I}}^*) \cdot (\underline{\mathbf{D}}_q * \underline{\mathbf{I}}^*)) \right) \quad (5.8)$$

Die Notation $\div$ bezeichnet hier eine elementweise Division. Zur Bestimmung eines Feldes $\underline{\Phi}$ an Orientierungswinkeln muß nun noch elementweise der halbe Arcustangens gebildet werden. Damit sind alle nötigen Operationen abgeschlossen. Es waren also insbesondere diskrete Ableitungen $\underline{\mathbf{D}}_p$ und $\underline{\mathbf{D}}_q$ zu bilden, diese paarweise zu multiplizieren, eine Differenz zu bilden, zwei Glättungen mit Binomialmasken auszuführen und anschließend der Arcustangens zu bilden und zu halbieren. Wählbar sind bei diesen Operationen folgende Parameter, die für die konkrete Umsetzung mit KEMPE /Kem1997/ diskutiert wurden:

Art der Ableitungsoperatoren: Es werden im ersten Schritt sehr einfache symmetrische Operatoren verwendet:

$$\underline{\mathbf{D}}_{p_1} = \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} \quad \text{und} \quad \underline{\mathbf{D}}_{q_1} = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad (5.10-5.10)$$

Diese Art von Ableitungsoperatoren erster Ordnung läßt sich als Differenz zweier Schiebeoperationen (um +1 und um -1) sehr einfach realisieren.

SCHARR ET AL. /SKJ1997/ weisen darauf hin, daß diese Ableitungsfilter durch Querglättung bezüglich ihrer Isotropie optimiert werden können. JÄHNE gibt in /Jäh1998/

⁴Die verwendete Glättungsoperation ist linear und kann deshalb ausgeklammert werden.

als konkreten Vorschlag folgende Glättungsoperanden an:

$$\underline{\mathbf{D}}_{p_2} = \frac{1}{16} \begin{pmatrix} 3 \\ 10 \\ 3 \end{pmatrix} \quad \text{und} \quad \underline{\mathbf{D}}_{q_2} = \frac{1}{16} \begin{pmatrix} 3 & 10 & 3 \end{pmatrix} \quad (5.12-5.12)$$

Damit ergeben sich die Ableitungsoperatoren aus den kartesisch separiert angewendeten Filtern $\underline{\mathbf{D}}_{p_1}$ und $\underline{\mathbf{D}}_{p_2}$ bzw. $\underline{\mathbf{D}}_{q_1}$ und $\underline{\mathbf{D}}_{q_2}$ zu

$$\underline{\mathbf{D}}_p = \underline{\mathbf{D}}_{p_1} * \underline{\mathbf{D}}_{p_2} \quad (5.13)$$

$$\underline{\mathbf{D}}_q = \underline{\mathbf{D}}_{q_1} * \underline{\mathbf{D}}_{q_2} \quad (5.14)$$

Auf den Einsatz von Ableitungsoperatoren höherer Ordnung kann damit verzichtet werden.

Ordnung des Binomialfilters zur Glättung: Das Filter ${}^4\mathbf{B}$ wird sinnvollerweise ebenfalls kartesisch separiert eingesetzt

$${}^4\mathbf{B} = {}^4\mathbf{B}_1 * {}^4\mathbf{B}_2 \quad (5.15)$$

und besteht entsprechend der Notation aus Binomen 4. Ordnung:

$${}^4\mathbf{B}_1 = \frac{1}{16} \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \end{pmatrix} \quad \text{und} \quad {}^4\mathbf{B}_2 = \frac{1}{16} \begin{pmatrix} 1 \\ 4 \\ 6 \\ 4 \\ 1 \end{pmatrix} \quad (5.17-5.17)$$

Wahl des Bandpasses zur Vorverarbeitung des Bildes: JÄHNE /Jäh1993/ weist darauf hin, daß eine initiale Glättung des Bildes erforderlich ist, um die Ableitungsberechnung zu verbessern. Implementiert werden kann dafür ein Bandpaß, der den Einsatz des genannten einfachen Operators zur Ableitungsberechnung auch hinsichtlich niederfrequenter Anteile verbessert, da diese dann eliminiert sind. Insgesamt gesehen wird durch die Bandbegrenzung die Filterung auf ein bestimmtes Ortsfrequenzintervall abgestimmt, was beim Einsatz in einer Multiresolutionspyramide gefordert ist.

Bereits im Abschnitt 4.2.1.1 *Datenmaterial* wurde kurz diskutiert, wie ein Skalarbild aus dem direkt verfügbaren EBU-Farbbild $\mathbf{I}_{RGB}$ gewonnen wird. Auch hierfür ist die Bildung eines Intensitätsbildes durch eine Aufsummation der drei Farbauszüge von sich aus sinnvoll. Folgende Überlegungen führten dazu, den Farbauszug $\mathbf{I}_G$ (*siehe auch* S. 32) *direkt* zu verwenden und als Skalarbild in die Orientierungsanalyse und damit in die Konturdetektion eingehen zu lassen:

- die Analyse soll empfindlich auf Helligkeitskontraste sein
- hierzu liegt als Grundlage die von der CIE 1924 als Standard herausgegebene Kurve $V(\lambda)$ zur Spezifizierung des spektralen Hellempfindungsgrades vor
- die Kurve besitzt eine gewisse Ähnlichkeit mit der grünen Farbmischkurve einer CCD-Farbfernsehkamera, beide Kurven weisen ein Maximum bei $\lambda \approx 555\text{nm}$ auf, wobei $V(\lambda)$ flacher zu den Rändern des sichtbaren Spektrums abfällt.
- $V(\lambda)$ stellt ein Ergebnis evolutionärer Entwicklung in natürlichen Umwelten dar; umgekehrt werden auch künstliche Umwelten meist angepaßt an die Fähigkeiten und Eigenschaften des menschlichen visuellen Systems entworfen

Ein vorverarbeitetes Bild $\underline{\mathbf{I}}^*$ wird also aus dem Farbauszug $\underline{\mathbf{I}}_G$ als Grauwert-Rohbild wie folgt gebildet:

$$\underline{\mathbf{I}}^* = 2,5 \cdot ({}^2\underline{\mathbf{B}} * \underline{\mathbf{I}}_G - {}^6\underline{\mathbf{B}} * \underline{\mathbf{I}}_G) \quad (5.18)$$

Dabei sollte ${}^6\underline{\mathbf{B}}$ aufgeteilt werden in ${}^6\underline{\mathbf{B}} = {}^4\underline{\mathbf{B}} * {}^2\underline{\mathbf{B}}$, wodurch das Zwischenergebnis ${}^2\underline{\mathbf{B}} * \underline{\mathbf{I}}^*$ doppelt genutzt werden kann. ${}^2\underline{\mathbf{B}}$ wird analog Gl. 5.15 gebildet mit:

$${}^2\underline{\mathbf{B}}_1 = \frac{1}{4} \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \quad \text{und} \quad {}^2\underline{\mathbf{B}}_2 = \frac{1}{4} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \quad (5.20-5.20)$$

Der Koeffizient in Gl. 5.18 skaliert den Bandpaß bei seiner Mittenfrequenz auf ≈ 1 , da die maximale Abweichung von ${}^2\underline{\mathbf{B}}$ und ${}^6\underline{\mathbf{B}}$ bei $\approx 0,4$ liegt.

5.4.2.2 Güteabschätzung

Anhand eines Testbildes soll eine Abschätzung der Güte der Orientierungsbestimmung mittels nichtlinearer Strukturanalyse gegeben werden. Die Abbn. 5.8–5.10 enthalten verschiedene Stadien der Verarbeitung.

Die Abbildungen 5.9 und 5.10 belegen, daß die Methode im relevanten Wellenzahlbereich zwischen ca. $\frac{\pi}{4}$ bis $\frac{3\pi}{4}$ eine sehr zufriedenstellende Genauigkeit liefert. Die Nutzung des Maximums des Konfidenzmaßes als Kriterium der Signifikanz führt zu einer Unterdrückung von homogenen Flächen. Für den Einsatz im Arrangement orientierter Filter auf Realwelt-Szenen bestehen damit gute Voraussetzungen.

5.5 Ergebnisse

Ähnlich wie in den beiden vorangegangenen Kapiteln werden nun mit den Abbn. 5.11–5.14 einige empirische Beispiele gegeben, die die prinzipielle Eignung des Ansatzes belegen sollen.

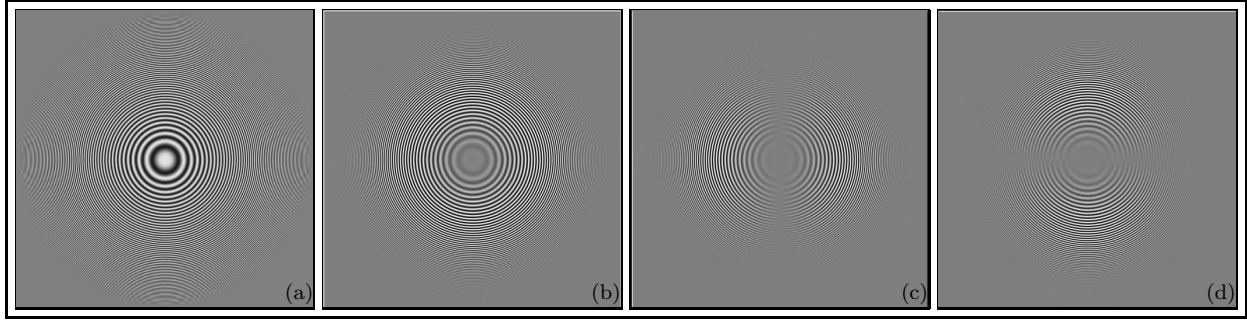


Abbildung 5.8: Orientierungsbestimmung mittels Tensormethode (1): (a) Testbild (lineare Erhöhung der Wellenzahl vom Zentrum zum Rand im Intervall $0 \dots \pi$, d. h. minimal hat eine Halbperiode eine L nänge von einem Pixel); (b) Bandpaßgefiltertes Testbild nach Gl. 5.18; (c) Ableitungsergebnis $\underline{D}_p * \underline{I}^*$; (d) Ableitungsergebnis $\underline{D}_q * \underline{I}^*$. Die sichtbaren Moiré-Effekte sind nahezu vollständig bedingt durch die begrenzte Auflösung des Drucks.

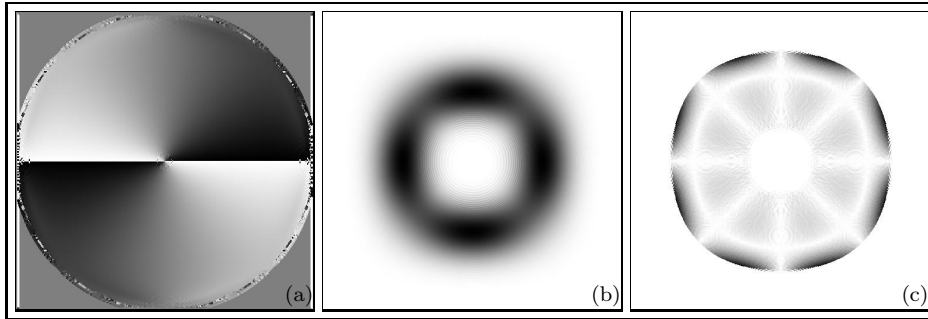


Abbildung 5.9: Orientierungsbestimmung mittels Tensormethode (2): (a) Orientierungswinkel kodiert in Grauwerten (0° : Schwarz, 90° : Mittelgrau, 180° : Weiß; wegen der zyklischen Eigenschaften des Orientierungswinkels kommt es in der Darstellung zu Weiß-Schwarz-Sprüngen); (b) Betrag des Vektors mit doppeltem Orientierungswinkel aus Gl. 5.7. Dieser Betrag bildet ein Konfidenzmaß für die lokale Orientierung. Ersichtlich wird, daß eine hohe Sicherheit gerade im mittleren Ortsfrequenzbereich vorliegt (indirekte Folge der initialen Bandpaßfilterung). Im rechten Abbildungsteil (c) ist der Fehler der Winkelbestimmung in dem Ortsfrequenzbereich dargestellt, in dem das Konfidenzmaß noch 10% des Maximums beträgt (siehe auch Abb. 5.10). Der maximale Winkelfehler in diesem Bereich liegt bei ca. $2,2^\circ$ und wird gerade an der äußeren Abgrenzung erreicht, wobei dieser darüber hinaus bis zu 20° erreicht. Der mittlere absolute Fehler liegt unter $0,2^\circ$, die Standardabweichung unter $0,3^\circ$.

Dabei sind besonders die Fälle von Interesse, in denen durch die Beleuchtungsverhältnisse die Person eher als dunkle (helle) Silhouette vor hellem (dunklem) Hintergrund erscheint und dabei gleichzeitig Gesichtsdetails nur sehr kontrastarm erscheinen.

Die guten Ergebnisse unterstreichen, daß das Prinzip das Potential hat, als Cue im Mehrkomponenten-Auffälligkeitssystem gerade dann einen wirksamen Beitrag zu guten Lokalisationsergebnissen zu leisten, wenn die beiden anderen Cues durch Beleuchtungseffekte an ihre Grenzen kommen.

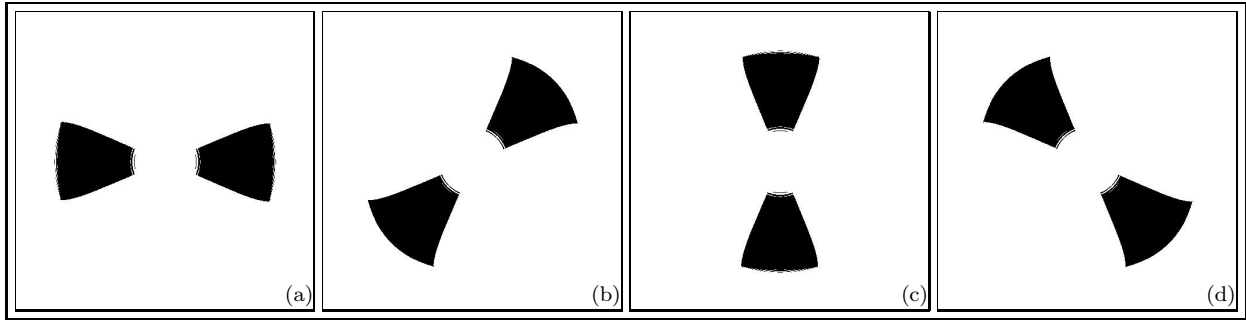


Abbildung 5.10: Ergebnisse der Orientierungsbestimmung mittels der Tensormethode nach Einteilung in Orientierungsintervalle (v.l.n.r.): $[-22,5^\circ \cdots +22,5^\circ]$; $[+22,5^\circ \cdots +67,5^\circ]$; $[+67,5^\circ \cdots +112,5^\circ]$; $[+112,5^\circ \cdots +157,5^\circ]$. Diese binäre Darstellung hebt genau diejenigen Bildpunkte schwarz hervor, an denen mindestens noch 10% des Bestimmtheitsmaßes vorlagen. Die Rotation einer Grauwertkante mit maximalem Kontrast in Gradschritten von 0 bis π ergab einen mittleren quadratischen Fehler der Winkelbestimmung kleiner $0,026^\circ$.

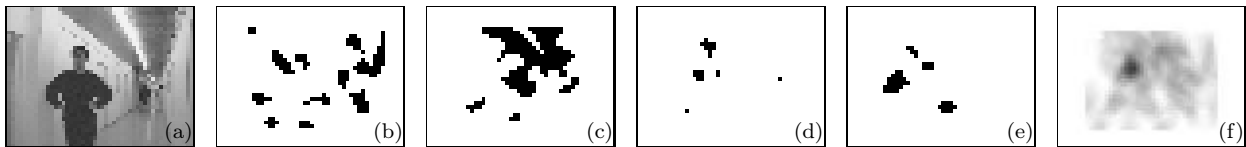


Abbildung 5.11: Beispiele von Ergebnissen der konturbasierten Personenlokalisierung im Innenbereich (1): Unter den Bedingungen des weitgehend homogen hell gestalteten Ganges und der dunkel gekleideten Person liegt hier mit dieser Szene (a) ein optimaler Fall vor. Ganz rechts (f) ist das Ähnlichkeitsmaß dargestellt, worin sich gerade im Kopfbereich die höchsten Werte stark häufen. In den mittleren Rahmen (b) bis (e) sind die in vier Intervalle eingeteilten Ergebnisse der Orientierungsbestimmung analog zu Abb. 5.10 wiedergegeben.

5.6 Résumé

Mit dem Entwurf eines einfachen und damit gut umsetzbaren konturbasierten Ansatzes zur Personenlokalisierung liegt ein weiterer Cue vor, der gerade unter Bedingungen, in denen die bereits vorgestellten Cues *Hautfarbe* und *Gesicht* z.T. starke Defizite aufweisen, gute Lokalisationsergebnisse liefert.

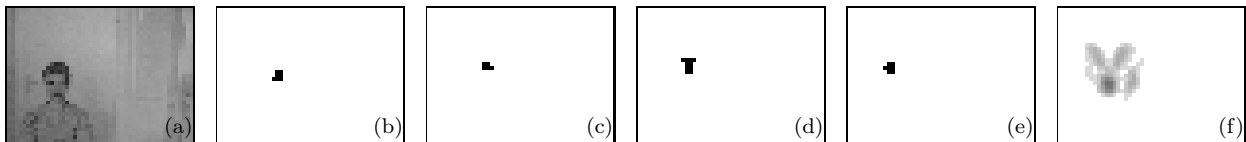


Abbildung 5.12: Beispiele von Ergebnissen der konturbasierten Lokalisation im Innenbereich (2): Zwar weist diese Szene (a) einen homogenen Hintergrund auf, jedoch ist die Person hier äußerst kontrastarm erfaßt worden, so daß hier etwa die kontrastmäßige Untergrenze erreicht ist. Wären noch sonstige kontrastreichere Gegenstände im Bild enthalten, hätte hier durchaus das Maximum des Ähnlichkeitsmaßes in (f) nicht im Gesichtsbereich liegen können. – Zur Bedeutung der Einzelbilder (a) bis (f) siehe Abb. 5.11.



Abbildung 5.13: Beispiele von Ergebnissen der konturbasierten Personenlokalisierung im Außenbereich: Hier ist in (a) die Homogenität des Hintergrundes gegenüber Abb. 5.11 deutlich herabgesetzt. Die Auflösung ist eine halbe Oktave feiner, die Distanz zur Kamera damit etwa eine halbe Oktave größer. Auch hier erfolgt im Ähnlichkeitsmaß (f) eine deutliche starke Häufung hoher Werte korrespondierend zum Kopf der Person. – Zur Bedeutung der Einzelbilder (a) bis (f) siehe Abb. 5.11.

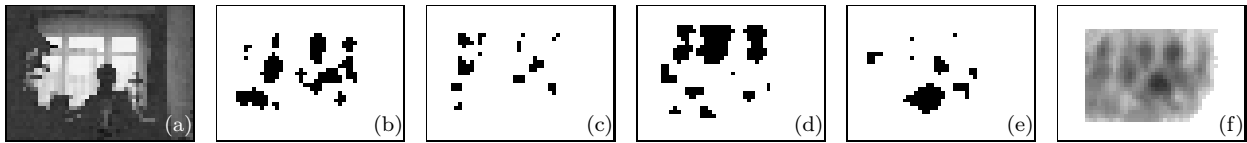


Abbildung 5.14: Beispiele von Ergebnissen der konturbasierten Personenlokalisierung im Innenbereich (3): Die Sicht in einen Büroraum könnte typisch sein für einen mobilen Roboter, der eine Suchaufgabe in einem Gebäude zu lösen hat. Es ist nicht überraschend, daß auch hier ein akzeptables Ergebnis zustande kommt. – Zur Bedeutung der Einzelbilder (a) bis (f) siehe Abb. 5.11.

Für weitere Arbeiten besteht folgendes Verbesserungspotential:

- Erweiterung von Gl. 5.1 um einen Term, der dem Ähnlichkeitsmaß einen Bonus erteilt, sofern zusammenhängende Konturabschnitte gut dem Modell entsprechen (Person z. B. teilverdeckt), so daß zufällige Matches durch Fits an auf der Kontur weit auseinanderliegenden Punkten besser unterdrückt werden. Diese Idee wurde zugunsten schneller Berechnung nicht umgesetzt.
- deutlichere Einbeziehung des Schulterbereiches in das Template.

Kapitel 6

Personenlokalisierung

6.1 Einleitung

Das Thema dieses Kapitels bildet zugleich den Kern der vorliegenden Arbeit. Nachdem bisher drei einzelne unterschiedliche personenspezifische Auffälligkeitskomponenten vorgestellt wurden – hautfarbbasierte, gesichtsbasierte und konturbasierte Methoden – geht es nun darum, diese zu *verschmelzen*. Zielstellung ist, als Ergebnis eines Selektionsprozesses *die* Bildkoordinate zu bestimmen, an der am wahrscheinlichsten eine Person richtigerweise zu lokalisieren ist.

Dafür wird ein zweischrittiges Vorgehen vorgeschlagen, was im Detail in den nachfolgenden Abschnitten vorgestellt wird. Zunächst soll darin die Fusion der drei Auffälligkeitskomponenten mittels einer einfachen Fuzzy-Operation erfolgen. Daran schließt sich ein Selektionsprozeß an, der innerhalb eines *dreidimensionalen dynamischen neuronalen Feldes* abläuft. Während dieses iterativen Prozesses bildet sich eine räumlich abgegrenzte Region aktiver Feldneuronen als sog. Aktivitätsblob heraus (Effekt der dynamischen Selektion von Erregungsmustern nach AMARI /Ama1977/), über dessen Lage innerhalb des Feldes die Lokalisierung erfolgen kann. Die wesentlichen Argumente für den Einsatz eines dynamischen neuronalen Feldes sind:

- kompetitive Ausbildung von Gewinnerregionen durch lokale Kooperation benachbarter Neuronen bei gleichzeitiger globaler Konkurrenz (wesentliche Bedingungen: Input mit hinreichender Amplitude, Ausdehnung und Zeitdauer)
- immanentes Hystereseverhalten (Gewinnerregion mit begrenzter räumlicher und zeitlicher Eigenstabilisierung; Verfolgen bewegter Inputs)
- geringe Anforderungen an numerisches Lösungsverfahren

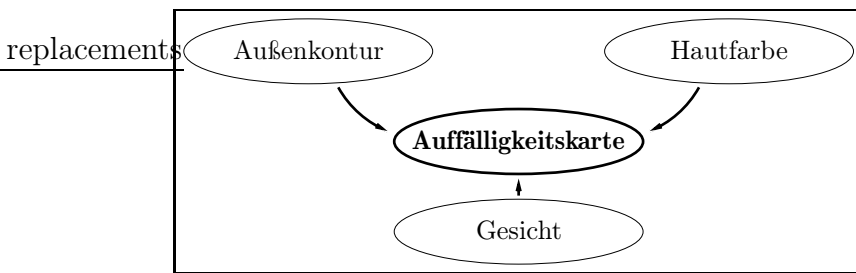


Abbildung 6.1: Die Auffälligkeitskarte ist das Ergebnis einer Fuzzy-Überlagerung der verschiedenen Cues und einer anschließenden dynamischen Selektion.

6.2 Bekannte methodische Ansätze

Im Zusammenhang mit der visuellen Personenlokalisierung werden in diesem Abschnitt Verweise auf eine Reihe von Arbeiten aus der Literatur gegeben, in denen ebenfalls eine Fusionsproblematik behandelt wird. Es erfolgt jeweils eine Skizzierung der Verfahren und eine kritische Wertung, wobei im Abschnitt 1.3 *Visuell-basierte Personenlokalisierung* auf einige der Arbeiten bereits verwiesen wurde.

LEE und SEUNG /LS1998/ nutzen zur Personenlokalisierung drei Farbauszüge und ein Framedifferenzbild und fusionieren mittels eines zweischichtigen Mehrebenenperzeptrons (MLP). In einer überwachten Trainingsphase verwenden sie einen Infrarotreflektor, der eine praktisch fehlerfreie Personenlokalisierung (als Referenz) erlaubt. Dadurch kann die Trainingsmenge des Bildmaterials *on the fly* erzeugt werden, wodurch das Training sehr realitätsbezogen gestaltet werden kann. In der Anwendung werden die Gewichtsmatrizen der Hidden-Elemente als Faltungskerne auf die Farbauszüge bzw. das Differenzbild angewendet und jedes Ergebnisbild der einzelnen Faltungen gewichtet aufaddiert. Das Ergebnisbild wird einer Maximumssuche unterzogen und die Koordinaten des Maximums werden direkt zur Lokalisation der Person eingesetzt. Vorteile des Ansatzes sind vor allem die gute Möglichkeit der Datenakquisition und, untrennbar damit verbunden, die Einbeziehung von Bewegungsinformation, da offenbar unterstellt wird, daß relevante Personen sich bewegen oder durch Eigenbewegung der Kamera im Bild bewegt erscheinen. Nachteil ist die simple Maximumssuche, die als sehr störanfällig vermutet werden muß, da einzelne lokale Fits unabhängig zu deren Umgebung betrachtet werden. Somit bleiben benachbarte Koinzidenzen unberücksichtigt, von denen ein wirksamer Beitrag zur Erhöhung der Robustheit zu erwarten wäre.

KUBISCH und RITTER, H. /KR1996/ stellen ein System zur Erkennung menschlicher Kopfhaltungen vor, bei dem die Lokalisation des Gesichts zwar nicht unmittelbar als Fusion verschiedener Cues gelöst wird, jedoch durch Kaskadierung implizit doch eine Fusion vorgenommen wird: Ein KNN (hier ein LLM-Netzwerk) führt eine Hautfarbsegmentierung auf dem in Blöcken gemittelten RGB_{EBU} -Bild durch, an die sich eine Regionensegmentierung anschließt, so daß danach die größte im Bild befindliche hautfarbene Region gefunden wird. Darauf folgt eine Analyse dieser Region mittels Gaborfilterung, die mittels eines weiteren LLM zur Bestimmung der besten Lage einer Ellipse innerhalb der segmentierten Region

genutzt wird. Diese Ellipse umschließt eine darin angenommene sichtbare Gesichtsfläche minimal. Damit ist die Lokalisation gelöst; es schließt sich hier eine weiterführende Analyse der Kopforientierung an. Vorteil dieses Ansatzes zur Gesichtslokalisation ist die Einheitlichkeit der Methoden (LLM) und die hohe Präzision, die durch die Nutzung geometrischer Information erzielt werden kann. Nachteil ist offensichtlich, daß das Verfahren grundsätzlich nur hautfarbbasiert ist.

FEYRER und ZELL /FZ1998/ stellen ein System zum visuell basierten Nachfolgen einer Person durch eine mobile Roboterplattform vor. Offensichtlich handelt es sich um eine zur vorliegenden Arbeit sehr verwandte Problemstellung – Detektion/Lokalisation und Tracking liegen eng beieinander. Die Autoren verwenden ganz explizit einen multimodalen Ansatz. Darin gehen im einzelnen – wie bei den beiden zuvor vorgestellten Vorgehensweisen – das Ergebnis einer Hautfarbsegmentierung und zusätzlich – wie bei LEES/SEUNGS Verfahren – Bewegungsinformation ein. Für letzteres besteht insbesondere der Anspruch, durch eine aufwendige Eigenbewegungskompensation Bildbereiche mit signifikanter aktiver Objektbewegung zu finden, wofür zusätzlich eine sehr genaue Entfernungsbestimmung durch ein Stereokamerasystem vorgenommen wird. Für die Hautfarbsegmentierung wurden alle manuell segmentierten prototypischen Daten in eine Verweistabelle (LUT) eingetragen. Beide genannten Komponenten bilden die Basis zur Personenlokalisierung, jedoch wird die Hautfarbe dominant verwendet (was die Frage nach dem Aufwand-Nutzen-Verhältnis für die Bewegungsdetektion aufwirft). Nachdem eine Regionensegmentierung erfolgte und daran anschließend gerade *die* Segmente ausgewählt wurden, bei denen ein für Gesichter plausibles Höhen-Breiten-Verhältnis eingehalten ist sowie eine Reihe weiterer spezieller Hautfarbparameter und Heuristiken (Fusion unmittelbar benachbarter Segmente anhand einer Adjazenzmatrix) erfüllt sind, modulieren nun die Ergebnisse der aufwendigen Bewegungsanalyse diese verbliebenen Hautfarbsegmente, indem solche mit hinreichender Bewegung in ihrer Auffälligkeit angehoben werden. Das Segment mit der höchsten Auffälligkeit wird selektiert.¹

TRIESCH /Tri1999/ behandelt das Problem der Fusion von bildbeschreibenden Merkmalskarten zunächst auf einfache Weise. Drei Merkmalskarten (lokale Orientierung auf versch. Skalen, mittlere lokale Farbe, lokale Orientierung auf einem hautfarbsegmentierten Bild) mit jeweiligen Auffälligkeitsmaßen werden durch gewichtete Überlagerung im Verhältnis 1 : 2 : 1 zu einer Verbundauffälligkeitskarte fusioniert. Teilziel ist dabei, sicher und

¹Leider bleibt unklar, in welcher Weise im zweiten Teil des Verfahrens die bereits im Abschnitt 5.2 *Bedeutung von Konturen* auf S. 50 kurz angerissene Detektion einer Person anhand einer charakteristischen Umrißlinie im Kopf-/Schulterbereich mittels des Verfahrens *Aktive Strahlen* einfließt. Die Vermutung, daß der Einfluß ähnlich gering wie der Bewegungsinformation ausfällt, liegt nahe, denn es wird alles in allem als Nachteil eingeräumt, daß das Tracking „sehr wesentlich von der Farbkomponente“ abhängt.

robust Fingerspitzen in Bildern zu finden, um eine gute Handgestenerkennung durchführen zu können – eine Problemstellung, die mit der hier behandelten Personenlokalisierung verwandt ist. TRIESCH weist ausdrücklich darauf hin, daß diese einfache Methode auch mit Blick auf konzeptionelle Einfachheit und gute Implementierbarkeit auf Computern gewählt wurde.

Darüber hinaus stellt TRIESCH einen Ansatz zur Objektverfolgung (realisiert am Beispiel von Gesichtern) vor, für den die Fusion der Merkmale *adaptiv* erfolgt. Hierzu werden die Gewichtungen der Cues als schnell modifizierbare, reversible und zugleich plastische Kurzzeit-Speicher nach V. D. MALSBERG /Mal1994/ ausgeführt. Realisiert wird dies von TRIESCH, indem für die aktuell gefundene Zielposition zum Zeitpunkt t ermittelt wird, in welchem Maße jeder beteiligte Cue *über* dem aktuellen globalen Durchschnittswert dieses Cues liegt. Für jeden der Cues wird daraus ein relativer Anteil an der aktuellen Lokalisation bestimmt, der als Störterm in eine lineare Differentialgleichung (DGL) 1. Ordnung eingeht. Mit einer derartigen DGL wird für jeden Cue ein sog. *Leaky Integrator* modelliert, der einen gleitenden Mittelwert für die Gewichtung des jeweiligen Cues bestimmt. Die dafür benötigte Zeitkonstante muß dabei eine hinreichende Größe aufweisen, um dieser Methode die erforderliche Trägheit zu verleihen. Zu schnelle Gewichtungsänderungen könnten sonst die Methode zu schnell in zu extreme Gewichtungskonstellationen führen. Ein derartiges Vorgehen wurde für die vorliegende Arbeit (ungeachtet der nicht in der Weise vorgenommenen linearen Überlagerung, s. u.) aus Gründen der Handhabbarkeit zunächst nicht in Betracht gezogen, sollte aber beim weiteren Ausbau unbedingt einfließen, um damit flexibler auf mögliche starke Änderungen der Umgebungsbedingungen reagieren zu können.

WREN, AZARBAYEJANI, DARRELL und PENTLAND /WADP1997, Pen1996b, Pen1996a, Pen1996c, Pen1997/ stellen mit dem sog. *Pfinder* ein sehr aufwendiges System vor (*siehe auch* Abschnitt 1.3 *Visuell-basierte Personenlokalisierung* auf S. 8). Sie verwenden für die Personenlokalisierung zwei Modi der Merkmalsextraktion, die parallel arbeiten: im *Blob-Modus* werden mit Labels versehene Hyperellipsen gebildet, die vor allem durch Farbsegmentierung (Farbe insbesondere der Kleidung) gefunden werden. Diese dienen als geometrische Modelle von Körperbestandteilen (Arme, Beine, Hände). Allerdings dürfen sie nur innerhalb einer Region liegen, die über eine morphologische 2D-Konturformanalyse zuvor gefunden wurde. Die Personenlokalisierung erfolgt anhand einer Schätzung der Position von Armen, Beinen, Händen etc. Sowohl Blobs als auch Konturanalyse werden fusioniert, um besonders im Falle des Ausfalls einer Komponente robust zu sein (bspw. Hand erscheint vor dem Hintergrund einer „hautfarbenen“ Wand). Dazu wird jeder Komponente eine lokale, heuristisch bestimmte Güte beigemessen, anhand derer eine lineare Superposition erfolgt. Der dazu parallel arbeitende *Vorhersage-Modus*, der das Verfolgen (Tracking) der Person zum Ziel hat, soll hier nicht weiter betrachtet werden. Zum Einsatz kommt Pfinder als Personenlokalisierung

für ein Gesichtserkennungssystem (*siehe* NASTAR und MITSCHKE /NM1998/). Der Ansatz besitzt eine Reihe von Abhängigkeiten zwischen morphologischer und Farbanalyse, so daß keine wirkliche Redundanz erwartet wird.

MILANESE, PUN und WECHSLER /MPW1993/ diskutieren die Fusionsproblematik visueller Information vor einem allgemeinen Hintergrund. Die Merkmalstypen, die berücksichtigt werden, sind Konturen und Regionen, anhand derer bestimmte Auffälligkeitskarten jeweils für jede verwendete Auffälligkeitskomponente gebildet werden – ganz analog zur Konstellation der vorliegenden Arbeit. Die Fusionsstrategie, die vorgeschlagen wird, ist hier grundsätzlich iterativ, wobei eine Energiefunktion mit einem nichtlinearen Relaxationsverfahren minimiert wird. Ziel ist, über alle Auffälligkeitskarten *Kohärenzen* zu finden. Damit ist zugleich implizit die verwendete Definition von Auffälligkeit vorgegeben. Vorteil des Verfahrens ist, daß keine starre Gewichtung der einzelnen Auffälligkeitskarten bei deren Überlagerung benötigt wird. Nachteil ist jedoch, daß alle beteiligten Merkmale tatsächlich kohärent auftreten müssen. Gerade diese Voraussetzung ist in typischen Real-World-Problemstellungen kaum einzuhalten, wenn der Aufwand für die Merkmalsextraktionsverfahren beherrschbar bleiben soll.

ITTI, KOCH und NIEBUR /IKN1998/ haben ein neuronal inspiriertes, allgemeines Modell zur selektiven visuellen Auffälligkeit vorgestellt. Auch hier ist das Vorgehen derart, daß multimodale Auffälligkeitskomponenten (repräsentiert in jeweils mehreren Farb-, Intensitäts- und Orientierungskarten) durch arithmetische Mittelung zu *einer* Gesamtauffälligkeitskarte fusioniert werden. Allerdings werden alle beitragenden Karten einer Normierungsoperation unterworfen, durch den Karten mit wenigen einzelnen Auffälligkeitsbeiträgen verstärkt und andere mit vielen gedämpft werden. Somit erfolgt keine reine Mittelung. Es schließt sich eine Selektion der auffälligsten Szeneregion mittels eines Winner-Take-All-Prozesses (WTA) an, was durch ein zweidimensionales Feld aus *Leaky Integrate-And-Fire-Neuronen* nach KOCH und ULLMAN /KU1985/ realisiert wird. Auch wenn sich dabei auf ein zweidimensionales Feld beschränkt wurde, erscheint eine derartige Modellierung eines WTA-Prozesses durch Selektionsvorgänge mit dynamischen Neuronenmodellen vor dem Hintergrund der vorliegenden Arbeit als sehr interessant.

MEL /Mel1997a/ stellt mit dem allgemeinen visuellen Objekterkennungssystem „SEEMORE“ einen multimodalen Ansatz vor, der durch seinen Umfang beeindruckt: 102 Merkmalskanäle werden verwendet, die in folgende fünf Gruppen eingeteilt werden: Farbkanäle, zwei Intensitätskanäle (Ecken, Häufungen), Konturkanäle und Texturkanäle. Die Fusion (und damit die Klassifikation) überläßt MEL einem Nächster-Nachbar-Klassifikator. Diese äußerst breit angelegte Merkmalsextraktion ist allerdings dementsprechend berechnungsintensiv – sowohl für die Vorverarbeitung als auch für die Fusion/Klassifikation. Allerdings kann nicht bestritten werden, daß sowohl Breite als auch Spektrum sehr interessant und von

der Herangehensweise plausibel sind.

Zusammenfassend kann zunächst festgestellt werden, daß in den vorgestellten konkreten Anwendungen immer die Segmentierung von Hautfarbe eine wichtige Rolle spielt. Darum herum ranken sich jeweils grauwertbasierte Verfahren, die zueinander wenig Gemeinsamkeiten aufweisen. Die Fusion der Auffälligkeitskomponenten ist meist darauf angelegt, daß die Hautfarbe die dominierende Rolle spielt. Mehrfach wird mit *gewichteter Superposition* gearbeitet – dem einfachsten und am leichtesten rechentechnisch zu implementierenden Verfahren – jedoch besteht hier immer die Forderung nach Kohärenz. Tatsächliche Redundanz, also die Ersetzbarkeit bestimmter Komponenten durch verfügbare andere, wird damit nicht erzielt. Andere Verfahren setzen *konditionale Cues* ein (meist die richtige Hautfarbsegmentierung als Bedingung für weitere Cues). Sofern diese Bedingung allerdings nicht erfüllt ist, bleiben andere Cues jedoch chancenlos. Dies ist kein wirklich plausibles Vorgehen.

Das in dieser Arbeit entwickelte Verfahren soll diese beiden Hauptkritikpunkte aufgreifen und in einer neuartigen Lösung berücksichtigen. Dazu sollte a priori der prinzipielle Vorschlag von MILANESE ET AL. aufgegriffen werden, ein iteratives Relaxationsverfahren einzusetzen. Gerade darin wird das Potential gesehen, eine deutlich verbesserte Robustheit zu erzielen, da in die Selektionslösung räumliche Kontextinformation, etwa nachbarschaftliche Koinzidenzen, eingehen können. Dafür wird es nicht erforderlich sein, ein bestimmtes aufwendiges Regelwerk aufzustellen, da einfache, uniform angewendete lokale Regeln in Form von Vorschriften für lokale Wechselwirkungen genügen.

6.3 Fusion und dynamische Selektion

Dieser Abschnitt behandelt die beiden Schritte des Systems, die zur Personenlokalisierung führen: Fusion und Selektion. Hierbei steht die Frage nach der sinnvollsten Reihenfolge. Möglich wäre, zunächst in jedem Cue eine Selektion vorzunehmen und daran anschließend die einzelnen Selektionsergebnisse zu fusionieren.

Es wird im Anschluß dargelegt, daß es sowohl für die Handhabbarkeit als auch für die Leistungsfähigkeit wichtig ist, umgekehrt zunächst die Fusion der Cues vorzunehmen und die Selektion für die Personenlokalisationsentscheidung in einem zweiten Schritt anzuschließen.

Im zweiten Teil des Abschnitts wird das Teilsystem, mit dem die angestrebte iterative Selektion ausgeführt wird, vorgestellt: das dreidimensionale dynamische neuronale Feld.

6.3.1 Fusion der Cues mittels Fuzzy-Operator

Die in dieser Arbeit eingesetzte Selektionsmethode war bereits in der Einleitung zu diesem Kapitel genannt worden: dynamische neuronale Felder. Da hierbei berechnungsintensive dreidimensionale Faltungen über eine Anzahl von Iterationsschritten auszuführen sind, ist wichtig, zunächst eine Abwägung von Aufwand und erwartetem Nutzen vorzunehmen. Insbesondere ist dies für eine Entscheidung über die Sinnfälligkeit einer separaten Selektion pro Cue erforderlich. Für separate Selektionen spricht, daß dabei die in ihren Lokalisationseigenschaften teilweise voneinander abweichenden² Cues stärker getrennt gehalten und als Zwischenergebnis zunächst drei spezifische Selektionsergebnisse liefern würden. Die Fusion selbst würde sich vereinfachen, da ein Endergebnis nur aus diesen drei Teilselektionsergebnissen zu bilden wäre.

Für den in der Abfolge umgekehrten Weg spricht, daß damit ein mehrfacher (für alle drei Cues also dreifacher) Selektionsaufwand vermieden wird. Damit geht eine erhebliche Beschleunigung und somit eine Verbesserung der Umsetzbarkeit des Verfahrens einher, da sich das Datenvolumen für die Selektion sehr deutlich auf ein Drittel reduziert. Damit fällt die Entscheidung zugunsten dieser Vorgehensweise.

Zu entscheiden wäre nun, welche Algorithmen für die zunächst anstehende Fusion in Frage kommen. Einfachste Methode wäre die lineare Superposition (*siehe* Abschnitt 6.2 *Bekannte methodische Ansätze*). Polynomiale Ansätze höherer Ordnung wären ebenfalls möglich. In beiden Fällen bestünden die Parameter aus einem Koeffizientensatz zur Gewichtung der Einzelbeiträge. Diese könnten mit einem Optimierungsverfahren bestimmt werden, wie etwa beim Polynomklassifikator. Ebenso in Frage kämen konnektionistische Ansätze, z. B. Mehrebenenperzeptrons. Zwar gibt es für alle genannten Möglichkeiten auch verschiedene Methoden der Vereinfachung (LUTs, ausgedünnte Netze). In Voruntersuchungen hat sich anstelle davon bereits der Einsatz eines Fuzzy-Operators als einfache und hinreichend robuste Alternative erwiesen – auch ohne besondere Parameteroptimierung.

Fusion im allgemeinen bedeutet Mittelung. Aus der Theorie über *Unscharfe Mengen* (Fuzzy Sets, *siehe dazu* ZIMMERMANN /Zim1991/) sind verschiedene sog. *Durchschnittsoperatoren* (Averaging Operators) bekannt. Ohne diese Theorie komplett umzusetzen, z. B. ausgehend von linguistischen Variablen über die Folge *Fuzzyifizierung* – *Inferenz* – *Defuzzyifizierung* einen bestimmten Ausgabewert zu bilden, sind die Operatoren *für sich genommen* für diese Arbeit interessant. Im Grunde wird bei den meisten Operatoren zur Verknüpfung unscharfer Mengen eine parametrische³ Variation vom reinen geometrischen oder arithme-

²besonders bzgl. der Möglichkeit/Genauigkeit der Entfernungsschätzung

³im Spezialfall auch nichtparametrische

tischen Mittel hin zu einer von zwei⁴ Grenzen ausgeführt: dem *Durchschnitt* (entspr. dem log. UND) oder der *Vereinigung* (entspr. dem log. ODER).

WERNERS /Wer1984, Wer1988/ stellte dazu konkret die Fuzzy-Operatoren μ_{UND} „Fuzzy And“ und μ_{ODER} „Fuzzy Or“ vor. Beide Operatoren sind parametrisch und erlauben einen Ausgleich (Kompensation, Mittelung) zwischen den *Zugehörigkeitsfunktionen*⁵ (Membership Values) der zu verschmelzenden Datensätze. Je nach Parametrisierung kann aus dem Bereich zwischen dem einfachen Minimum (bei μ_{UND}) bzw. Maximum (bei μ_{ODER}) und dem arithmetischen Mittel ($\gamma \in [0, 1], x \in \mathbb{R}^N$) gewählt werden:

$$\mu_{\text{UND}}(\mu_A, \mu_B) = \gamma \cdot \min(\mu_A(x), \mu_B(x)) + \frac{(1 - \gamma)(\mu_A(x) + \mu_B(x))}{2} \quad (6.1)$$

$$\mu_{\text{ODER}}(\mu_A, \mu_B) = \gamma \cdot \max(\mu_A(x), \mu_B(x)) + \frac{(1 - \gamma)(\mu_A(x) + \mu_B(x))}{2} \quad (6.2)$$

Jedoch findet sich in beiden Operatoren jeweils nur ein Teil der benötigten Eigenschaften. Erforderlich ist:

- keine einfache arithmetische Mittelung, da damit die hinreichende Präsenz *aller* Komponenten erforderlich wäre
- keine unmittelbare Selektion des Maximums, da damit immer eine der Komponenten nahezu perfekt arbeiten müßte
- deshalb Einbeziehung von Koinzidenzen über alle drei Komponenten, ohne dabei eine explizite Mittelung vorzunehmen

Zusammengefaßt zielen diese drei Forderungen auf einen *integrativen Charakter* der Fusion der Cues ab.

ZIMMERMANN und ZYSNO /ZZ1980, Zim1991/ haben auf der Suche nach Operatoren, die die menschliche Entscheidungsweise möglichst adäquat nachbilden, verschiedene Operortypen vorgestellt. Darunter findet sich folgender sog. Minimum-Maximum-(Min-Max)-Operator:

$$\mu_{\text{MINMAX}}(\mu_A, \mu_B) = \gamma \cdot \min(\mu_A(x), \mu_B(x)) + (1 - \gamma) \cdot \max(\mu_A(x), \mu_B(x)) \quad (6.3)$$

mit $\gamma \in [0, 1]$

Es handelt sich hier offenkundig um eine parametrische Zusammensetzung aus Operatoren, die die *logischen* UND- bzw. ODER-Operatoren modellieren (s. o.). Genauer gesagt, wird

⁴Die Negation als Entsprechung zum logischen NICHT soll hier nicht betrachtet werden.

⁵Mit Zugehörigkeitsfunktionen werden linguistische Variablen auf unscharfe Mengen abgebildet. Diese Vorgehensweise wird hier nicht direkt umgesetzt. Dafür werden unmittelbar die drei personenspezifischen Auffälligkeitskomponenten eingesetzt.

eine einfache lineare *Konvexkombination*⁶ aus Minimum- und Maximum-Operator vorgenommen. Hier liegt zugleich die besondere Eignung: einerseits wird keine explizite Mittelung vorgenommen, andererseits werden Koinzidenzen einbezogen. Die Konvexkombination kann so parametrisiert werden, daß der Maximumterm dominant ist. Die Folge ist, daß nur bei ebenfalls hinreichend großem Minimumterm die gewichtete Summe beider Terme den Maximumterm soweit aufstocken kann, daß ein bestimmtes Schwellenmaß überschritten werden kann.

Damit könnten die genannten Forderungen erfüllt werden, daß sowohl die Präsenz aller Komponenten als auch die optimale Präsenz nur einer der Komponenten *nicht* erforderlich sind. Somit wird eine gute Artefaktunterdrückung auf einfache Weise möglich; Koinzidenzen sind notwendig, aber nur bis zu einem gewissen Grade – einzelne starke Inputs sind ebenfalls notwendig, aber nicht notwendigerweise ausreichend für ein starkes Fusionsergebnis.

Um nun die in den vergangenen Kapiteln vorgestellten Cues $\underline{\mathbf{P}}_H$, $\underline{\mathbf{P}}_G$ und $\underline{\mathbf{P}}_K$ überhaupt fusionieren zu können, muß zunächst gesichert sein, daß diese drei Matrizen eine einheitliche Größe und damit zueinander korrespondierende Elemente aufweisen. Das hier verwendete Bildmaterial wurde generell nicht unmittelbar ausgehend von der Auflösung des Framegrabbers verarbeitet. Hautfarbcue und Konturcue arbeiten auf den Auflösungsstufen 109×81 , 77×57 , 55×41 , 39×29 und 27×20 .⁷ Dagegen nutzt der Gesichtscue eine Auflösung von 256×191 , 181×135 , 128×95 , 91×67 und 64×48 ,⁸ da dabei mehr Bilddetails erfaßt werden müssen. Deswegen muß also die Pyramide des Gesichtscues mittels Unterabtastung an die Auflösung der anderen beiden Pyramiden angepaßt werden. Weiterhin muß gesichert sein, daß einheitliche Wertebereiche verwendet werden. Bereits in den Gln. 3.24, 4.9 und 5.1 zur (elementweisen) Bestimmungen der Einzelcues $p_{H_{jk}}$, $p_{G_{jk}}$ und $p_{K_{jk}}$ wird deshalb ein Wertebereich von $[0, 1]$ gesichert.

Die Konkretisierung von Gl. 6.3 mit den in den vorangegangenen Kapiteln vorgestellten Cues lautet für ein Matrixelement an der Stelle j, k :

$$p_{jk}(p_{H_{jk}}, p_{G_{jk}}, p_{K_{jk}}) = \gamma \cdot \min(p_{H_{jk}}, p_{G_{jk}}, p_{K_{jk}}) + (1 - \gamma) \cdot \max(p_{H_{jk}}, p_{G_{jk}}, p_{K_{jk}}) \quad (6.4)$$

Dieses fuzzy-verknüpfte Auffälligkeitsmaß ist ein Zwischenergebnis und bildet die Grundlage für die abschließende dynamische Selektion, die im folgenden Abschnitt erläutert wird. Es wird für alle fünf Auflösungsstufen gebildet (*siehe* Kapitel 2 *Pyramidale Repräsentationen als Basiskonzept*), was Konsequenzen für die Topologie der Feldstruktur haben wird.

⁶eine Einbeziehung der wertemäßigen Ränder des Datensatzes

⁷ca. ein Siebtel bis ein Achtundzwanzigstel der Originalauflösung

⁸ca. ein Drittel bis ein Zwölftel der Originalauflösung

replacements

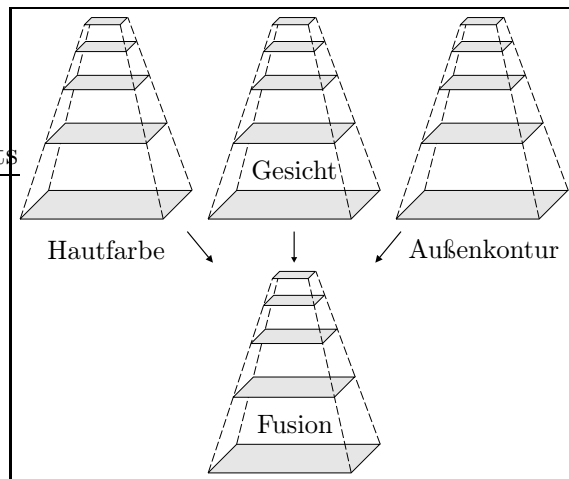


Abbildung 6.2: Fusion der drei pyramidal vorliegenden Cues. Es wird als erforderlich angesehen, daß die Fusion **integrativ** erfolgt. Daran schließt sich eine Selektion dabei entstehender Hypothesen an. Im Ergebnis werden die bis dahin getrennten Einzelcues in **eine** Personenlokalisationsentscheidung überführt. Die Lösung hat zu berücksichtigen, daß es sich durchweg um Multiskalen-Repräsentationen handelt.

6.3.2 Selektion im dreidimensionalen dynamischen neuronalen Feld

6.3.2.1 Zielstellung und grundsätzliches Prinzip

Um eine akzeptable Lokalisation von Personen erzielen zu können, reicht die Bildung des Auffälligkeitsmaßes entsprechend Gl. 6.4 allein nicht aus. Für eine definierte, sichere Lokalisationsentscheidung bedarf es darüber hinaus eines *Selektionsmechanismus*, der ausnutzt, daß eine Lokalisation umso sicherer erfolgen kann, je mehr es eine abgegrenzte Konzentration hoher Auffälligkeitswerte gibt.

Dynamische neuronale Felder eignen sich für diese Selektionsaufgabe besonders gut, weil darin kein starres *Pattern Matching* (wie bei den vorgestellten zugrundeliegenden Cues) erfolgt, sondern eine anpassungsfähige, auf impliziten Vorgaben beruhende dynamische Bildung eines Selektionsergebnisses.

Derartige Felder sind bereits von anderen Autoren in vergleichbaren Anwendungsgebieten mit Erfolg eingesetzt worden. KOPECZ, K. und SCHÖNER /KS1995,Kop1995,Kop1996/ stellen ein System zur Modellierung der aufmerksamkeitsbasierten Analyse visueller Szenen vor, in dem die sakkadische Szenenabtastung mit Feldern vom AMARI-Typ (s. u.) realisiert ist. Ebenso werden solche Felder von GIESE /Gie1997,Gie1998/ genutzt, der sie zur Segmentierung in speziellen Merkmalsräumen zur visuellen Bewegungsdetektion einsetzt und damit Effekte der Bewegungswahrnehmung modelliert.

HAMKER /HG1997,Ham1999/ geht ähnlich wie KOPECZ, K. und SCHÖNER vor, setzt jedoch die zu AMARIS Modell alternative Beschreibung dynamischer neuronaler Felder nach KASKI und KOHONEN /KK1994/ mit *nichtlinearer Verlustfunktion* ein. WILHELM /Wil1998/ hat in vergleichenden Untersuchungen allerdings festgestellt, daß dieser Algorithmus einen höheren Rechenaufwand (insbesondere eine kleinere Schrittweite zur numerischen Iteration und ein anspruchsvolleres numerisches Lösungsverfahren) erfordert.

Ähnlich wie HAMKER arbeitet auch STEPHAN /SG1997/, erweitert dabei jedoch AMARIS Ansatz strukturell und funktional derart, daß Aufmerksamkeitsregionen *formerhaltend* selektiert werden können. Diese zusätzliche Eigenschaft wird allerdings im Rahmen dieser Arbeit nicht benötigt. Ebenfalls wurden durch MÖLLER /Möl1996, MMTM1998/ eigene dynamische Neuronenmodelle als Grundelemente eines Modells der handlungsorientierten Wahrnehmung eingesetzt.

Favorisiert (und im weiteren detailliert vorgestellt) werden also Felder vom AMARI-Typ (*siehe* AMARI und KISHIMOTO /Ama1977, KA1979/), die sich gut für eine dynamische Selektion von Erregungsmustern (WTA) innerhalb iterativer Relaxationsprozesse eignen. Sie können mit einfachen homogenen feldinternen Wechselwirkungsmechanismen selektiv einzelne, eng abgegrenzte Gruppen aktiver Neuronen (sog. Blobs) herausbilden, deren Position im Feld als Lokalisationsergebnis genutzt werden kann. Als Besonderheit für den Einsatz im Rahmen dieser Arbeit kommt hinzu, daß die Relaxation auf einem kontinuierlichen Datenstrom erfolgen soll.

In AMARIS dynamischen Neuronenmodell werden nicht einzelne, sondern *Gruppen benachbarter Neuronen* modelliert. Deren Verschaltungsstruktur und deren darüber ablaufende Wechselwirkungen werden in *einem* Neuronenmodell zusammengefaßt. Darin werden Zeitverzögerungen von Wechselwirkungen allerdings vernachlässigt. Das dafür aufgestellte kontinuierliche nichtlineare DGL-System realisiert mit jeder Einzelgleichung eine Modellbeschreibung eines durchschnittlichen Membranpotentials einer Neuronengruppe. Es hat mit n Elementen die prinzipielle Form:

$$\tau \frac{dz_i}{dt} = -z_i + f(z_1, \dots, z_n, x_i) \quad \text{mit } i = 1, \dots, n \quad (6.5)$$

Diese *lineare Verlustfunktion* (Leaky Integrator) bewirkt, daß in gleichen Zeitintervallen gleiche Anteile an Aktivität abgebaut werden. Generell ist hierbei der Gradient der Aktivität immer proportional der Abweichung der aktuellen Aktivität von der bestehenden Erregung.

AMARI unterscheidet in /Ama1977/ folgende fünf Aktivitätsmodi⁹ eines solchen Systems:

- monostabiles Feld, in dem alle Aktivität schließlich verschwindet (ϕ -Lösung)
- monostabiles Feld, in dem schließlich alle Elemente aktiv sind (∞ -Lösung)
- bistabiles Feld, alle Elemente werden explosionsartig aktiv, wenn der gegebene Stimulus weit genug ausgedehnt ist, alle Elemente werden deaktiv, sofern der Stimulus diese Größe nicht erreicht.
- bistabiles Feld, in welchem initiale Erregung entweder zu (stabiler) lokaler Aktivität (Anregung) bestimmter Ausdehnung oder zu übergangsweise instabiler Aktivität

führt, die aber schließlich verschwindet. Im ersteren Fall bewegt sich die lokalisierte Aktivität solange umher, bis sie die Position des Maximums (Zentrums) des Input-Stimulus findet (WTA-Prozeß).

- räumlich-periodisches Aktivitätsmuster, dessen Dynamik-Typ mit dem mittleren Erregungsniveau zusammenhängt.

Für diese Arbeit ist genau der Fall des bistabilen Feldes mit stabiler lokaler Aktivität interessant. Dieser Zustand wird als α -Lösung bezeichnet, da ein Gleichgewicht ($\frac{dz}{dt} = 0$) mit einem Bereich $R[z] = \{\underline{x} | z(\underline{x}) > 0\}$ von konkret $R[z] = (\underline{a}_1, \underline{a}_2)$ der Diagonalenlänge $a = \|\underline{a}_2 - \underline{a}_1\|$ aktiver Neuronen vorliegt. Da das Feld als homogen angenommen wird, ist gesichert, daß eine Gleichgewichtslösung $z(\underline{r}, t)$ im Feld einhergeht mit weiteren Gleichgewichtslösungen $z(\underline{r} + \underline{c}, t)$ für beliebiges $\underline{c}$. VEIT /Vei1997/ hat das Verhalten dynamischer neuronaler Felder ergänzend zu AMARI für diesen Fall untersucht. Insbesondere hat er theoretisch gezeigt, daß bei einer kontinuierlichen monoton wachsenden Ausgabefunktion $y(z(\underline{r}, t))$ lokale Anregungen ebenfalls existieren und unter geringen Zusatzbedingungen auch ebenso stabil sein können.

KEY /Key1998/ stellt in einer vergleichenden Untersuchung eine Vielzahl bekannter WTA-Prozesse vor und systematisiert diese. Die Hauptunterteilung erfolgt in *starke* und *schwache* WTA-Prozesse. Stark ist ein solcher Prozeß dann, wenn dieser eine initiale inhomogene Feldaktivität der Neuronen toleriert.¹⁰ Stärke bedeutet hier also konkret, daß sich ein solches System jederzeit flexibel auf wechselnde Erregungsmuster einstellen kann. KEY bezeichnet AMARIS Modell einschränkend als *teilstark*, da nicht für alle Parameter-einstellungen jede beliebige initiale Aktivitätskonfiguration für beliebige Eingaben zu einem Selektionsverhalten führen. Dieser Umstand wurde in dieser Arbeit empirisch bei der Parameterwahl berücksichtigt. Es ist klar, daß ein schwacher WTA-Prozeß ungeeignet wäre, da im vorliegenden Fall raum-zeitliche inhomogene Eingaben¹¹ mit entsprechend vielfältigen Aktivitätskonfigurationen zu erwarten sind, insbesondere weil auf einem kontinuierlichen Datenstrom gearbeitet werden soll, wobei die Aktivität am Ende einer Relaxation in den nächsten WTA-Prozeß übernommen wird (*siehe* Abb. 6.3).

⁹Für alle fünf Typen gibt AMARI Stabilitätsbeweise. Vereinfachend nimmt er allerdings binäre Aktivitätszustände der Elemente an, so daß die Beweise streng genommen nur für diesen Fall gelten, erwartet jedoch für monoton wachsende Ausgabefunktionen kein anderes stabilitätsmäßiges Verhalten. VEIT /Vei1997/ bringt genau diese fehlenden Beweise.

¹⁰Schwach ist der Prozeß entsprechend bei geforderter homogener Aktivität, also identischen Ausgangsbedingungen der Neuronen.

¹¹teilweise auch nur räumlich inhomogene Eingaben

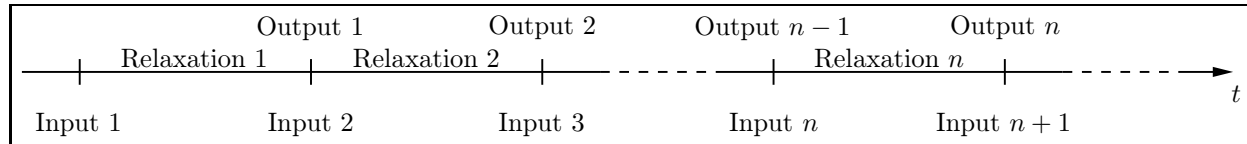


Abbildung 6.3: Sequentialisierung der Relaxationsprozesse. In jeder Relaxationsphase operiert das dynamische neuronale Feld auf einer Mikrozeitskala und bildet in einem WTA-Prozeß eine Lösung, die vom korrespondierenden Input und vom vorherigen Output abhängt. Die Lösung wird nach einem definiertem Abbruch der Relaxation als Output ausgegeben. Zugleich kann mit einem neuem Input der nächste Relaxations- und zugleich WTA-Prozeß gestartet werden.

6.3.2.2 Konkrete Umsetzung

Die erforderliche Topologie des dynamischen Feldes verdient besondere Beachtung. Im Abschnitt 1.3 *Visuell-basierte Personenlokalisierung* wurde bereits darauf hingewiesen, daß für die Problemstellung dieser Arbeit der Einsatz einer Multiskalen-Repräsentation erforderlich ist.¹² Erst damit ist es möglich, die Lokalisation auch hinsichtlich verschiedener Distanzen vorzunehmen.

Für die Lösung der Lokalisationsaufgabe erweist es sich jedoch als zweckmäßig, die pyramidale Repräsentation übergangsweise zu verlassen (*siehe* Abb. 6.4), um die Korrespondenzen zwischen den Auflösungsebenen der Repräsentation auf ein orthogonales Raster abzubilden. Der Grund ist, daß das wegen der dreidimensionalen Topologie der Daten erforderliche dreidimensionale Feld viel einfacher als *kubisches* Feld ausgeführt werden kann. Aus der Literatur ist eine solche formale und praktische kubische Erweiterung eines Ansatzes mit dynamischen neuronalen Feldern nicht bekannt.¹³ Nach dieser kubischen Umkodierung liegen entsprechend der Pyramidenabstufung fünf Ebenen vor (*siehe* Abb. 6.4), mit denen ein dreidimensionales Feld aufgespannt wird, innerhalb dessen die Selektion abläuft. Als Kompromiß für eine einheitliche Auflösung dieser fünf Ebenen würde sich die Auflösung der mittleren Pyramidenebene anbieten. Jedoch ist im weiteren zugunsten einer höheren Abarbeitungsgeschwindigkeit die Auflösung der obersten (größten) Pyramidenebene von Hautfarb- und Konturcues übernommen worden, so daß das kubische Feld eine Größe von $27 \times 20 \times 5$ dynamischen Neuronen aufweist.

Das verwendete dynamische neuronale Feld F vom AMARI-Typ besteht aus gegenseitig wechselwirkenden formalen dynamischen Neuronen (*siehe* Abb. 6.5) und kann als *rückgekoppeltes nichtlineares dynamisches System* beschrieben werden. Ziel des darin ablaufenden

¹²Im Anhang 2 *Pyramidale Repräsentationen als Basiskonzept* wird die Bestimmung derartiger Repräsentationen in Einzelheiten dargestellt.

¹³Zwar stellt z. B. AMARI Felder auch mit mehreren Ebenen vor, allerdings sind darin nur punktweise korrespondierende ebenenübergreifende Verbindungen vorgesehen. Insbesondere werden dabei in jeder Ebene unterschiedliche Neuronentypen (erregende/hemmende) eingesetzt.

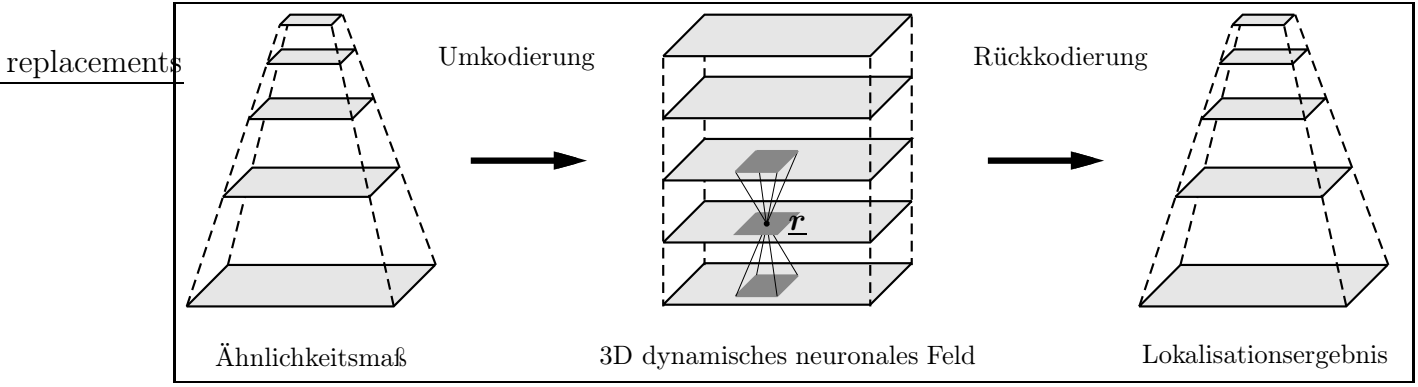


Abbildung 6.4: Temporäre Umkodierung der pyramidalen in eine kubische Repräsentation. Dieser Schritt ist erforderlich, um Korrespondenzen zwischen Ebenen für den Algorithmus des dreidimensionalen dynamischen neuronalen Feldes zu sichern. Die im Halboktavabstand angeordneten Ebenen erscheinen hier gemäß dem Logarithmus ihrer Kantenlängen (oder auch Abstände) gleichabständig. Alle Ebenen liegen im Kubus mit der Auflösung der mittleren Pyramidenstufe vor (der Kubus ist aus Anschauungsgründen vergrößert dargestellt). – Nach erfolgter Relaxation des Feldes schließt sich implizit eine Rückkodierung an, so daß der Bezug zwischen Kubusebene und Pyramidenstufe hergestellt ist und damit eine Entfernungsschätzung erfolgen kann. Die Skizze zeigt beispielhaft im Inneren des Kubus’ schematisch den Umgebungsbereich N , aus dem ein Element an der Stelle $\underline{r} = (x_1, x_2, x_3)^\top$ direkt aktiviert werden kann (siehe Gl. 6.6 ff.).

WTA-Prozesses ist die Ausbildung eines *einzelnen* lokalen Blobs aktiver Neuronen, die sich bei Annäherung an einen Gleichgewichtszustand gegenüber den übrigen behaupten können. Die folgende DGL beschreibt die nichtlineare Dynamik des Aktivierungszustandes $z(\underline{r}, t)$ eines Neurons (Elements) in kontinuierlicher Notation und konkretisiert die prinzipielle Darstellung des DGL-Systems aus Gl. 6.5. Die darin enthaltenen Koeffizienten c_h , c_l und c_i sind Parameter des Systems, anhand derer die externen und internen Erregungen (s. u.), die die Störterme in der DGL bilden, gewichtet werden.

$$\tau \frac{d}{dt} z(\underline{r}, t) = -z(\underline{r}, t) - c_h h(t) + c_i x(\underline{r}, t') + c_l \int_{N \subseteq \mathbb{R}^3} w(\underline{r} - \underline{r}') y(z(\underline{r}', t)) d^3 \underline{r}' \quad (6.6)$$

Hierbei steht $\underline{r}$ für die dreidimensionale Koordinate einer Neuronenposition im Feld.

$z(\underline{r}, t)$ ist die Aktivierung des Neurons $\underline{r}$ zum diskreten Zeitpunkt t . Korrespondierend dazu symbolisiert $y(z(\underline{r}, t))$ die Ausgabeaktivität dieses Neurons (Wertebereich $[0 \dots 1]$). Diese wird mittels der logistischen Sigmoidfunktion (Fermifunktion) bestimmt:

$$y(z(\underline{r}, t)) = \frac{1}{1 + \exp(-s \cdot z(\underline{r}, t))} \quad (6.7)$$

Hier liegt die Ursache für die Nichtlinearität der DGL in Gl. 6.6. Bezüglich des Neuronenmodells von AMARI stellt $y(z(\underline{r}, t))$ ein Maß für die Impulsemissionsrate an Feldkoordinate $\underline{r}$ zur Zeit t dar.

$x(\underline{r}, t')$ bezeichnet den lokalen externen Input, der dem umkodierten Auffälligkeitsmaß $p(p_H, p_G, p_K)$ (siehe Gl. 6.4) entspricht. Durch die Notation t' wird ausgedrückt, daß

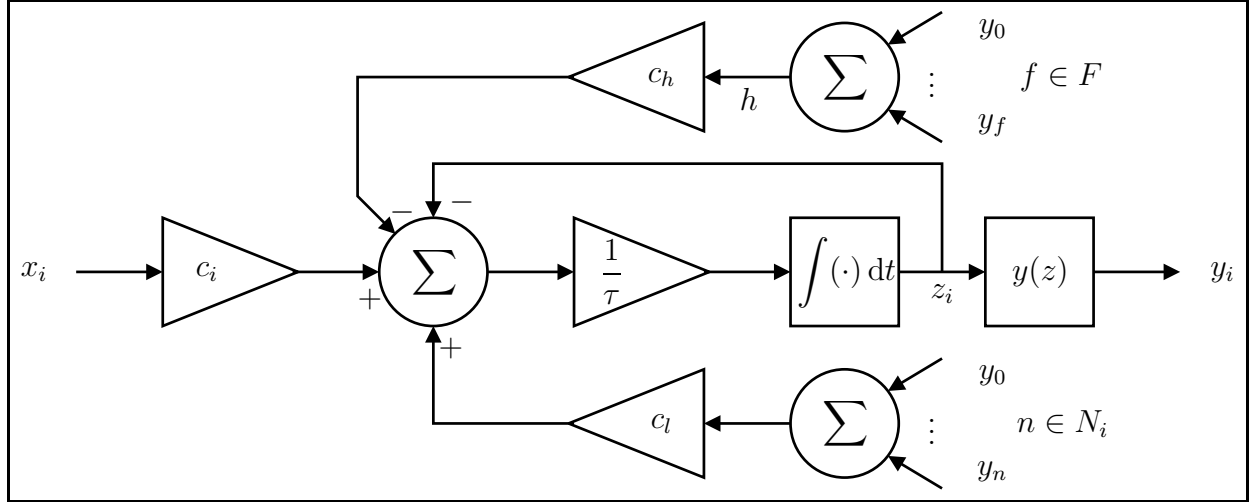


Abbildung 6.5: Formales dynamisches Neuron in Integratordarstellung (siehe Gln. 6.6 ff.). F symbolisiert die Menge aller Feldneuronen, N_i die Menge der Neuronen in der Nachbarschaft des i -ten Neurons.

dieser Stimulus ab der Startzeit t' für einen Iterationsvorgang bis zum Ende des Relaxationsprozesses konstant bleibt. $h(t)$ ist eine feldglobale Inhibition zum Zeitpunkt t , die über alle Neuronenaktivitäten $y(t - \Delta T)$ im gesamten Feld $F \subseteq \mathbb{R}^3$ aufsummiert wird:

$$h(t) = \int_{F \subseteq \mathbb{R}^3} y(\underline{r}', t - \Delta T) d^3 \underline{r}' \quad (6.8)$$

Die laterale Gewichtungsfunktion $w(\underline{r} - \underline{r}')$ ist vom Mexikanerhut-Typ und wurde folgendermaßen realisiert:

$$w(\underline{r} - \underline{r}') = 2 \exp \left(-(\underline{r} - \underline{r}')^\top \Sigma_1^{-1} (\underline{r} - \underline{r}') \right) - \exp \left(-(\underline{r} - \underline{r}')^\top \Sigma_2^{-1} (\underline{r} - \underline{r}') \right) \quad (6.9)$$

mit

$$\Sigma_2^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6.10)$$

$$\Sigma_1^{-1} = \left(\frac{2}{3} \Sigma_2 \right)^{-1} = \begin{pmatrix} \frac{4}{3} & 0 & 0 \\ 0 & \frac{4}{3} & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{3}{4} & 0 & 0 \\ 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{3}{2} \end{pmatrix} \quad (6.11)$$

Sie gewichtet als Modell eines rezeptiven Feldes (RF) die laterale Aktivierung eines Neurons $\underline{r}$ aus seiner Nachbarschaft $N \subseteq \mathbb{R}^3$, die in die DGL (siehe Gl. 6.6) eingeht. Für ein ausgewähltes $\underline{r}$ ist N in Abb. 6.4 dunkelgrau markiert. Die gewählte diskrete Ausdehnung von N beträgt $5 \times 5 \times 3$ Matrixelemente, d. h. es werden gerade drei umkodierte Ebenen umfaßt. Diese laterale Aktivierung erfolgt, wie die Parametrisierung von Σ_1 und Σ_2 und

auch die diskrete Ausdehnung zeigen, nicht isotrop. Das hat seine Ursache darin, daß der dreidimensionale Gitterabstand des Feldes nicht isotrop sein *kann*, da die Information über räumliche Tiefe, die über die fünf Ebenen kodiert wird, in einem erheblich größeren Raster vorliegt, als in den dazu orthogonal angeordneten und durchweg viel feiner aufgelösten, zu den Bildmatrizen korrespondierenden Matrizen (Ebenen). Dieser Umstand fand in der anisotropen Parametrisierung von w bzw. $\underline{\mathbf{W}}$ (siehe Gln. 6.12 ff.) eine grobe Berücksichtigung. Qualitativ vorgegeben ist, zentral eine positive Gewichtung zu verwenden, die nach außen hin in eine negative Gewichtung übergeht. Der Mexikanerhut-Typ ist keine ausdrückliche Vorgabe AMARIS. Hervorzuheben ist weiterhin, daß diese Funktion hier nicht von absoluten Feldkoordinaten abhängt (jedes Neuron hat das gleiche RF) und daß sie symmetrisch ist, weswegen F ein *homogenes* und *symmetrisches* dynamisches neuronales Feld darstellt. Ferner besteht die *zeitliche* Veränderlichkeit von w bzw. $\underline{\mathbf{W}}$.

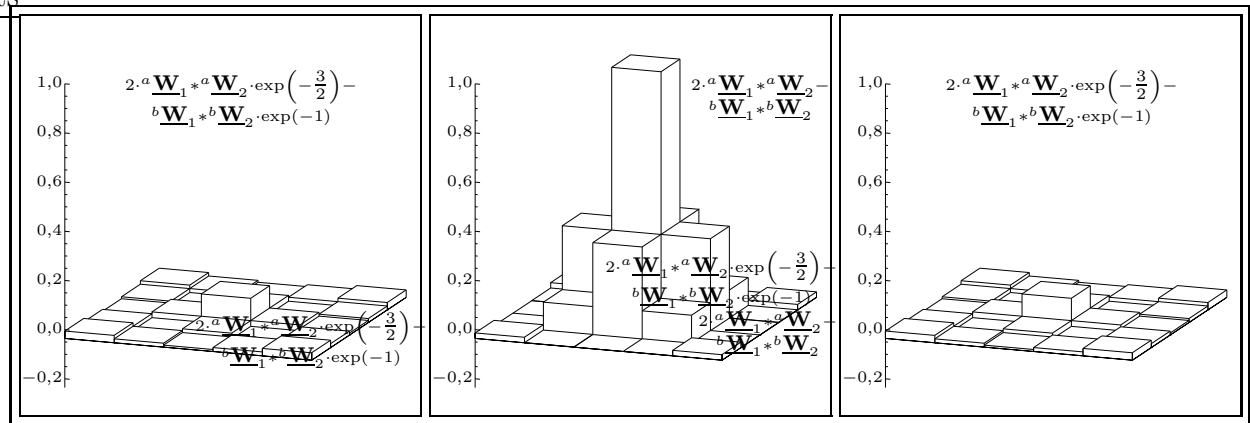


Abbildung 6.6: Veranschaulichung der Zusammensetzung der lateralen Aktivierungsfunktion $\underline{\mathbf{W}}$ als dreidimensionaler symmetrischer Faltungskern (Abmessungen $5 \times 5 \times 3$) anhand dreier Matrizen (siehe Gln. 6.12 ff.).

Da w sich aus einer Differenz von Funktionen vom Gauß-Typ zusammensetzt, besteht die Gelegenheit, implementationsseitig die Faltung mit den einzelnen Gaußfunktionen zunächst jeweils linear separiert zu berechnen und anschließend die Differenz der Faltungsergebnisse

zu bilden:

$$\underline{\mathbf{W}} = 2 \cdot {}^a\underline{\mathbf{W}}_1 * {}^a\underline{\mathbf{W}}_2 * {}^a\underline{\mathbf{W}}_3 - {}^b\underline{\mathbf{W}}_1 * {}^b\underline{\mathbf{W}}_2 * {}^b\underline{\mathbf{W}}_3 \quad (6.12)$$

mit

$${}^a\underline{\mathbf{W}}_1 = \begin{bmatrix} \exp(-3) & \exp\left(-\frac{3}{4}\right) & 1 & \exp\left(-\frac{3}{4}\right) & \exp(-3) \end{bmatrix} \quad (6.13)$$

und

$${}^b\underline{\mathbf{W}}_1 = \begin{bmatrix} \exp(-2) & \exp\left(-\frac{1}{2}\right) & 1 & \exp\left(-\frac{1}{2}\right) & \exp(-2) \end{bmatrix} \quad (6.14)$$

sowie

$${}^a\underline{\mathbf{W}}_2 = \begin{bmatrix} \exp(-3) \\ \exp\left(-\frac{3}{4}\right) \\ 1 \\ \exp\left(-\frac{3}{4}\right) \\ \exp(-3) \end{bmatrix} \quad (6.15) \quad \text{und} \quad {}^b\underline{\mathbf{W}}_2 = \begin{bmatrix} \exp(-2) \\ \exp\left(-\frac{1}{2}\right) \\ 1 \\ \exp\left(-\frac{1}{2}\right) \\ \exp(-2) \end{bmatrix} \quad (6.16)$$

In die dritte Dimension gerichtet sind die Vektoren

$${}^a\underline{\mathbf{W}}_3 = \begin{bmatrix} & & \exp\left(-\frac{3}{2}\right) \\ & & 1 \\ \exp\left(-\frac{3}{2}\right) & & \end{bmatrix} \quad (6.17)$$

und

$${}^b\underline{\mathbf{W}}_3 = \begin{bmatrix} & & \exp(-1) \\ & & 1 \\ \exp(-1) & & \end{bmatrix} \quad (6.18)$$

6.3.2.3 Numerische Lösung

Insbesondere der kubische Aufbau mit einer entsprechenden Erhöhung der Anzahl der Feldneuronen zwingt zu einer sehr effektiven Lösung zur numerischen Behandlung des DGL-Systems. Prinzipiell zur Verfügung stehen Kombinationen von

- ein- oder mehrschrittigen Verfahren,
- Verfahren verschiedener Ordnung,
- impliziten oder expliziten Verfahren und
- Verfahren mit fester oder variabler Iterationsschrittweite.

WILHELM /Wil1998/ hat in seinen Untersuchungen zu neuronalen Feldynamiken festgestellt, daß weder mehrschrittige Lösungsverfahren noch Verfahren höherer Ordnung für

Felder von AMARI-Typ erforderlich sind. Darüber hinaus genügt eine feste Schrittweite. Da implizite Verfahren nur bei steifen DGL-Systemen¹⁴ in Frage kommen, ist die Auswahl des numerischen Verfahrens getroffen: explizites einschrittiges Verfahren erster Ordnung mit fester Schrittweite, das nachfolgend kurz dargestellt wird.

Entsprechend Gl. 6.5 ist die prinzipielle Form einer DGL zur Beschreibung der Dynamik der Aktivierung eines Neurons

$$\tau \frac{dz}{dt} = -z + f \quad (6.19)$$

Die Differenzengleichung dafür lautet (abhängig vom diskreten Zeitpunkt k):

$$\tau \frac{z(k+1) - z(k)}{\Delta T} = -z(k) + f(k) \quad (6.20)$$

Nach wenigen Umstellungen ergibt sich

$$\begin{aligned} z(k+1) &= z(k) \left(1 - \frac{\Delta T}{\tau}\right) + \frac{\Delta T}{\tau} f(k) \\ &= z(k) + \frac{\Delta T}{\tau} (f - z(k)) \end{aligned} \quad (6.21)$$

Gl. 6.21 wird Polygonzug- oder Eulersche Methode genannt. Das Prinzip stellt generell die einfachste numerische Näherungsmethode zur Lösung von Differentialgleichungen dar. Die Schrittweite kann (*siehe* /Wil1998/) ohne Probleme bei etwa $\frac{\Delta T}{\tau} = 0,1$ (oder sogar darüber) liegen.

In den Abbn. 6.7 und 6.8 sind für zwei unterschiedliche synthetische Testfälle folgende Resultate erkennbar: Im WTA-Prozeß setzt sich nicht zwingend die ausdehnungsmäßig größere Region höherer Aktivierung durch, sondern die mit der lateralen Aktivierungsfunktion $w(\underline{\mathbf{r}} - \underline{\mathbf{r}}')$ der Neuronen besser im Einklang stehende. Daneben wird ersichtlich, daß die kompetitive Selektion zur Unterdrückung eines Bereiches und zum festen Aufbau des anderen Bereiches (Blob) führt.

Im Beispiel in Abb. 6.7 mit zwei auf getrennten Ebenen beschränkten Bereichen verstärkter Inputs setzt sich ein Blob nach und nach gegenüber dem anderen durch und repositioniert sich im weiteren leicht, so daß bei Abbruch der Iteration nach 75 Schritten seine Ausdehnung sich bis auf die benachbarte Ebene erstreckt. Grund dafür ist, daß im Zufallsmuster hinreichend starke Korrespondenzen auf der Nachbarebene liegen.

¹⁴Bedingung für steife Systeme sind stark unterschiedliche Zeitkonstanten; evtl. führen auch bestimmte spezielle Anfangsbedingungen zu steifem Verhalten. Implizite Verfahren machen sich Steifheit zunutze, indem sie eine Schätzung des Funktionswerts nachfolgender Iterationsschritte einbeziehen, die bei Steifheit mit guter Genauigkeit angegeben werden kann. Damit sind größere Schrittweiten möglich, so daß insgesamt in solchen Fällen Rechenzeit gespart werden kann.

Im Beispiel in Abb. 6.8 besteht der Unterschied zum ersten Beispiel darin, daß sich die Bereiche verstärkter Inputs noch auf die jeweiligen Nachbarebenen erstrecken. Es erfolgt die Selektion ähnlich wie zuvor, jedoch bildet sich im WTA-Prozeß ein Blob heraus, der frühzeitig nahezu gleichermaßen auf den Ebenen 1 und 2 (Spalten 2 und 3) verteilt ist. Grund ist hier, daß korrespondierend dazu im Input offenbar die konzentrierteste Häufung lokaler Inputs besteht. Die Gewichtung der globalen Inhibition h ist hier so bemessen, daß ca. 8 bis 10 dynamischen Neuronen einen Blob formieren können. Die globale Inhibition h und die laterale Gewichtungsfunktion w verhindern, daß der Blob sich mit derselben Stärke auf Ebene 0 (Spalte 1) etabliert, wo außerdem auch im verstärkten Zufallsmuster nur relativ schwache Inputs auftreten.

Das Entropiemaß¹⁵ erleichtert die Interpretation der in der Abbildung im Abstand von je fünf Iterationsschritten festgehaltenen Beispielergebnisse. Es könnte auch für ein Abbruchkriterium der numerischen Lösung des DGL-Systems verwendet werden. Da jedoch *repetitiv* auf einem Bilddatenstrom ohne explizite Begrenzung gearbeitet wird, aus dem jeweils nach Abschluß eines Lokalisationszyklusses erneut ein Bild entnommen wird und dazu korrespondierend nach kalkulierbarer Abarbeitungszeit ein Lokalisationsergebnis zur Verfügung gestellt werden soll, wurde als Abbruchkriterium eine feste verfügbare Iterationsschrittzahl eingesetzt. Damit wird zwar riskiert, daß eine Relaxation noch nicht akzeptabel abgeschlossen ist und eine entsprechend unsichere Lokalisationsentscheidung getroffen wird. Der Vorteil ist jedoch, daß die Lokalisation in einem festen Zeitraster erfolgt. Außerdem ist damit für den Ausnahmefall, daß durch *extreme* Erregungsmuster¹⁶ $x(\underline{r}, t')$ das Feld keinen stabilen Aktivitätsmodus mehr einnimmt und stattdessen etwa ein räumlich-periodisches Aktivitätsmuster zeigt, das Abbruchkriterium ebenfalls wirksam.

Ein Nebeneffekt der Umkodierung in eine kubische Struktur (*siehe* die Erläuterungen auf S. 77 und Abb. 6.4) ist folgender: je gröber die Auflösungsebene in der Pyramide wird, umso räumlich unschärfer wird das Ähnlichkeitsmaß auf der Kubusebene und desto ausgedehnter treten dementsprechend lokale Häufungen als Stimuli des dynamischen Feldes auf. Allerdings wurde dieser Effekt nicht gesondert kompensiert. Es war lediglich zu beobachten, daß in *den* Ebenen des Feldes, deren Input aus groben Auflösungsebenen entstammt, einige Schritte

¹⁵Die Entropie S läßt sich interpretieren als Maß für die Ungeordnetheit (vs. Kompaktheit) bzw. für den Informationsgehalt und wird für die Ausgabeaktivitäten des dynamischen neuronalen Feldes F formal gebildet als

$$S = - \sum_{i,j \in F} (y_{ij} \ln y_{ij}) \quad (6.22)$$

¹⁶Extreme Erregungsmuster für das Feld F sind z.B. gegeben wenn nur noch ein geringer Anteil der Feldneuronen keinen oder nur einen geringen direkten Stimulus x hat. Unter normalen Umständen des Szenarios zur Personenlokalisierung ist das wenig wahrscheinlich.

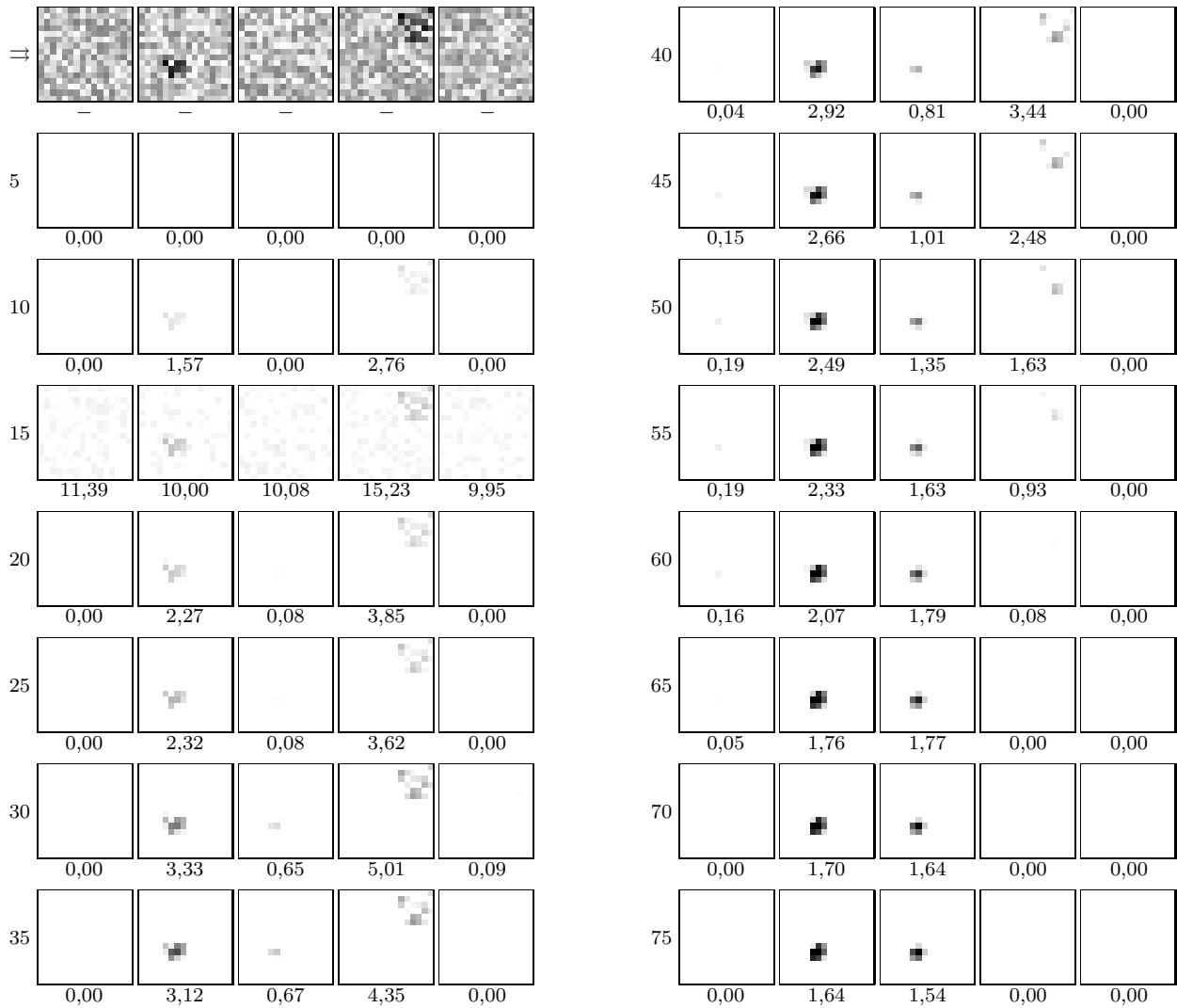


Abbildung 6.7: Zeitlicher Verlauf eines WTA-Prozesses innerhalb eines kubischen 3D dynamischen neuronalen Feldes (1): In diesem wie auch im zweiten Beispiel (Abb. 6.8) hat das Feld die Dimensionen $16 \times 16 \times 5$ (Darstellung in fünf Ebenen). In der Zeile oben links ($\Rightarrow$) ist der lokale externe Input $x(\underline{r}, t')$ dargestellt, der hier aus gleichverteilten Zufallszahlen aus dem Wertebereich $0 \dots 0,5$ besteht. Ausnahmen davon bilden zwei Bereiche, in denen der Zufallszahlenbereich $0 \dots 1$ beträgt, in der zweiten Ebene von der Größe 4×4 Elemente, in der vierten Ebene 6×6 Elemente. Alle anderen Zeilen der Darstellung enthalten die Aktivitäten der Feldneuronen $y(\underline{r}, t)$ nach bestimmten Iterationsschrittzahlen (in Fünferschritten). Zu jeder Ebene ist die jeweils vorherrschende Entropie angegeben (die jeweilige Gesamtentropie ergibt sich als Summe der Einzelentropien über alle fünf Ebenen; alle Angaben von 0,00 sind durch Abrunden sehr kleiner positiver Werte entstanden). Erkennbar ist hier, daß sich mehr und mehr ein (z. T. ebenenübergreifender) Blob aktiver Feldneuronen herausbilden kann, der auch Bereiche einschließt, an denen offenbar im Rauschmuster auch ohne explizite Verstärkung stärkere Inputs bestehen.

mehr zur Herausbildung eines Blobs bzw. Bestandteil eines Blobs benötigt werden.

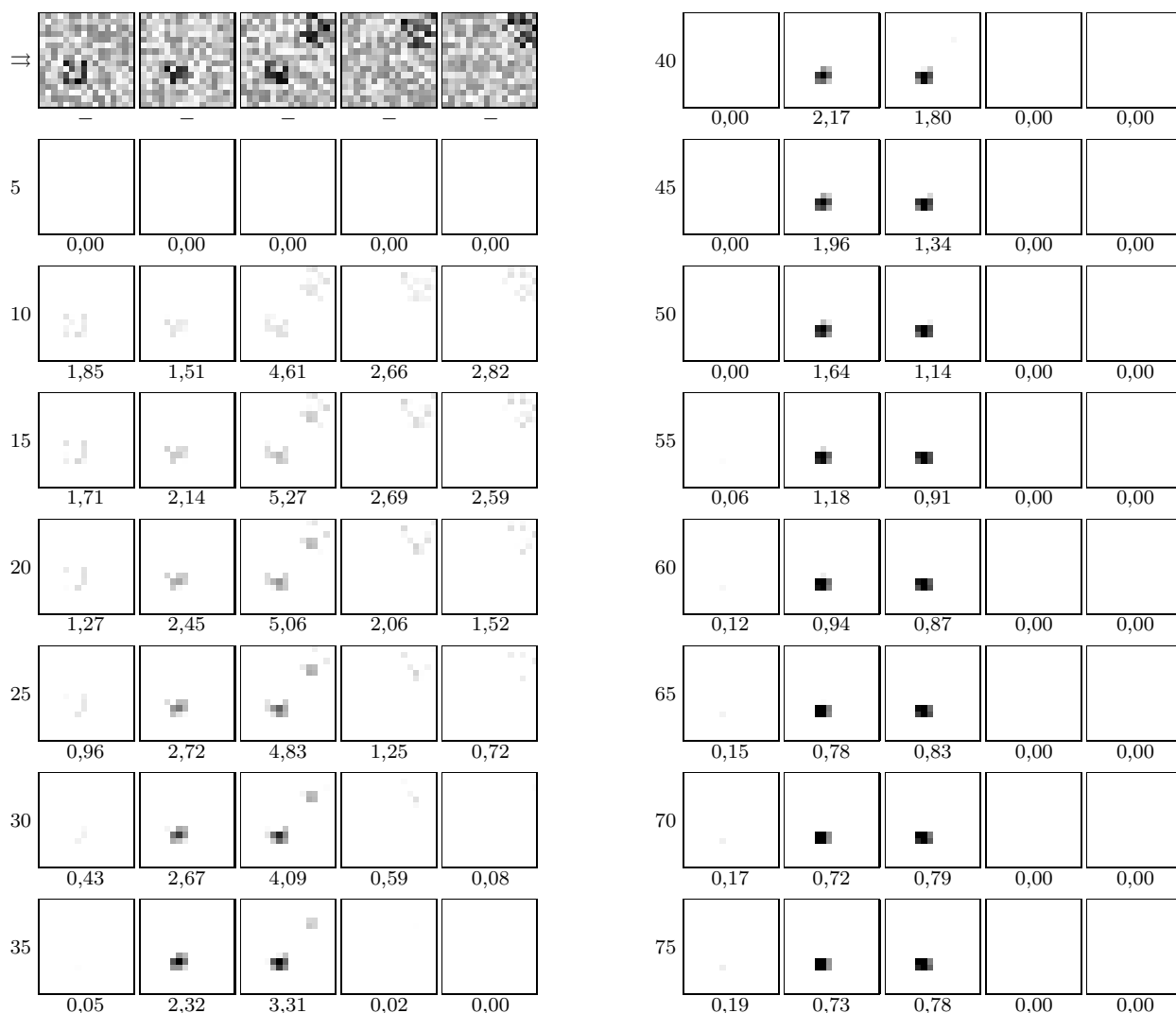


Abbildung 6.8: Zeitlicher Verlauf eines WTA-Prozesses innerhalb eines kubischen 3D dynamischen neuronalen Feldes (2): Dieses zweite Beispiel unterscheidet sich von dem in Abb. 6.7 dadurch, daß die beiden hervorgehobenen Aktivitätsbereiche sich jeweils noch über die benachbarten Ebenen erstrecken.

6.4 Demonstration des realisierten Systems

Zur Demonstration der Arbeitsweise und Leistungsfähigkeit des Gesamtsystems zur Personenlokalisierung wurde einer halbquantitativen Ergebnisdarstellung anhand typischer Fälle und Problemfälle hauptsächlich aus folgenden Gründen der Vorzug gegeben:

- eine Galerie an Szenen (als Voraussetzung für eine streng quantitative Bewertung) mit einem einfach handhabbaren Umfang wurde im Stadium der Entwicklung des Multi-Cue-Ansatzes als zu starr für Tests angesehen
- der Aufwand für eine Galerie mit genügend großem Umfang für quantitative Aussagen wurde unter den gegebenen Möglichkeiten als zu groß angesehen

- der verfügbare flexibel einsetzbare Computervision-Arbeitsplatz als Vorstufe zum mobilen Roboter erlaubte den Einsatz unter einer großen Variationsbreite an Umgebungsbedingungen zur Personenlokalisierung (Arrangement der Szenen, Beleuchtungsbedingungen, Personen usw.) im Innen- und Außenbereich

Nachteil ist jedoch, daß durch dieses Vorgehen keine genauen szenenbezogenen Referenzkoordinaten für die zu lokalisierende Person vorliegen, so daß damit eine genauere Gütebestimmung der Lokalisationsschätzung nur eingeschränkt möglich ist.

Genaue Koordinaten zur Bestimmung des Blobzentrums und somit der Personenlokalisierung können durch einfache gewichtete Mittelung der Ausgabeaktivität y des dynamischen Feldes (Kubus) F bestimmt werden.

$$\bar{\mathbf{r}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} = \frac{1}{\text{ord}(F)} \sum_{\mathbf{r} \in F} y(\mathbf{r}) \cdot \mathbf{r} \quad (6.23)$$

Mit einem somit ermittelten Ergebnis können nachgeordnete Prozesse nun die weitere Analyse der Person vornehmen. Je nach Realisierung kann als Referenzkoordinate ein bestimmtes Personenmerkmal, etwa die Nasenspitze, verwendet werden. Wesentlich sind die gestellten Genauigkeitsanforderungen. Im Grunde müssen hierfür Betrachtungen angestellt werden, die Abweichungen/Fehler in Raumwinkeln ausdrücken. Für eine sinnvolle Genauigkeitsbetrachtung sollte in zukünftigen Arbeiten mit einer Galerie an Szenen mit protokollierten Aufnahmebedingungen gearbeitet werden, so daß die Raumwinkel zwischen Kamerahauptachse und Ortsvektor zum Referenzpunkt auf der Person bekannt sind. Bei den hier zur Demonstration verwendeten Szenen war der Kamera-Personen-Abstand mittels Bandmaß bestimmt worden, so daß in Gegenüberstellung dazu in den Tabellen 6.1–6.6 für die Distanz eine (gerundete) Schätzung der Koordinate $\bar{x}_3$ angegeben werden konnte. Im Rahmen dieser Arbeit wurde es richtungsmäßig als völlig ausreichend angesehen, daß das Blobzentrum auf dem jeweiligen Gesicht liegt.

Im Rahmen dieses Ergebniskapitels soll hier anhand von insgesamt 24 Szenen (Abbn. 6.9–6.12 im Außen- und Abbn. 6.13–6.20 im Innenbereich bilden daraus eine Auswahl von zwölf variierten Szenen) eine Diskussion der erkennbaren Stärken und Schwächen gegeben werden. Das Versuchsprogramm ist derart angelegt, ein möglichst breites Spektrum an Konstellationen der einzelnen Cues abzudecken und damit Belege für das prinzipielle Leistungsvermögen des Multi-Cue-Ansatzes dieser Arbeit zu geben. Neben der Fallunterscheidung *Innen-/Außenbereich* mit entsprechenden *Beleuchtungsstärken-* und *-richtungsunterschieden* variieren insbesondere *Personenentfernungen* und *Hintergrund-/Vordergrundkomplexitäten*.

Zur Quantifizierung der Intensität der Beleuchtung wurde anhand eines Belichtungsmessers (unkalibriert, mit dual logarithmischer Skala) ein relatives Maß e_S für die Beleuch-

tungsstärke E_S [$\text{lx} = \text{lm}/\text{m}^2$] ermittelt. Alle weiteren Untersuchungen beziehen sich auf eine Referenzbeleuchtungsstärke, die bei den Untersuchungen zur Skalierungs- und Rotationsinvarianz (siehe Anhang E *Weitere tabellarische Ergebnisübersichten*) vorlag und unter allen weiteren vorgenommenen Untersuchungen das Minimum bildete.

Zur besseren Lesbarkeit soll hier eine kurze Übersicht der folgenden Seiten gegeben werden:

- Versuchsperson Sandy im Außenbereich: Tab. 6.1 & Abb. 6.9, S. 88–89
- Versuchsperson Autor im Außenbereich: Tab. 6.2 & Abbn. 6.10–6.12, S. 90–93
- Versuchsperson Christoph im Innenbereich (1): Tab. 6.3 & Abb. 6.13, S. 94–95
- Versuchsperson Sandy im Innenbereich (1): Tab. 6.4 & Abb. 6.14, S. 96–97
- Versuchsperson Christoph im Innenbereich (2): Tab. 6.5 & Abbn. 6.15–6.18, S. 98–102
- Versuchsperson Sandy in Innenbereich (2): Tab. 6.6 & Abbn. 6.19–6.20, S. 103–105
- verwendete Parameter: S. 106
- Rechenzeitbedarf: S. 106

Die konkrete Darstellung erfolgt sowohl graphisch als auch korrespondierend dazu tabellarisch. Somit werden Protokolle von Versuchsdurchführungen wiedergegeben. Alle Tests setzen sich dabei aus vier Einstellungen zusammen, von denen ein Teil mit Abbildungen erläutert und diskutiert wird.

Die im einzelnen jeweils in den Abbildungs- und Tabellentexten diskutierten Schwerpunkte sind:

- Welche besonderen Bedingungen liegen in der Szene vor?
- Wie tragen die einzelnen Cues zur Lokalisation bei?
- Wodurch sind Artefakte und/oder Defizite bedingt?

Die Angabe der Auflösungsebene in der Tabellenspalte *Person* ergibt sich durch ganzzahlige Rundung der reellwertigen Ebenenkoordinate der Lokalisationsschätzung. In den Spalten *Beleuchtungsrichtung* und *Hautfarbe* wird eine verbale Quantifizierung gegeben. Für die Spalten *Gesicht* und *Kontur* werden die Auflösungsebenen angegeben, auf denen sich die Person dem jeweiligen Cue entsprechend abzeichnet (halbquantitative visuelle Auswertung).

Tabelle 6.1: Lokalisationsergebnisse für Versuchsperson Sandy im Außenbereich: Hier ist die Hautfarbdetektion für alle Einstellungen schlecht ausgefallen. Das diffuse Tageslicht selbst stellt kein Problem dar, sondern der bemessene Dynamikumfang der Kamera. Die Beleuchtungsstärke lag um das ca. 16-fache über der bei den allermeisten Versuchen in Innenbereichen (vgl. Tabn. 6.3–6.6). Hier wirken sich Regelungseffekte¹⁸ aus, die im verwendeten Kamerasystem die Verstärkung beeinflusst und damit offenbar zu einer Verschiebung in der rg -Ebene geführt haben – ein Effekt, der bei reinen Intensitätsänderungen theoretisch nicht zu erwarten ist.

Die beiden anderen Cues arbeiten ungeachtet dessen zufriedenstellend (ausgenommen die 0,75m-Entfernung, dort ergibt die Schätzung ca. 1m), so daß auch die Selektion überall zu eindeutigen Ergebnissen führt.

In den Spalten „Gesicht“, „Kontur“ und „Person“ ist jeweils die Pyramidenebene angegeben, auf der die Einzelcues am besten detektiert bzw. die Person lokalisiert werden. Zusätzlich ist in der Spalte „Person“ in Klammern als Referenz die Pyramidenebene abgegeben, auf der entsprechend der jeweiligen Entfernung die Person richtigerweise zu lokalisieren wäre.

Dist.	ld(e_S)	Ort	Bel.-richt.	Hautf.	Gesicht	Kontur	Person	Abb.
0,75m	4,5	außen	rundum	schwach	2	2-3	2 (3)	6.9
1,0m	4,5	außen	rundum	schwach	2	1-2	2 (2)	–
1,5m	4,5	außen	rundum	schwach	1	1	1 (1)	–
2,0m	4,5	außen	rundum	schwach	0-1	0	0 (0)	–

¹⁸Diese Regelung ließ sich nicht abschalten, konnte aber gerade bei Kontrastwechseln als schwer zu quantifizierender Effekt beobachtet werden.

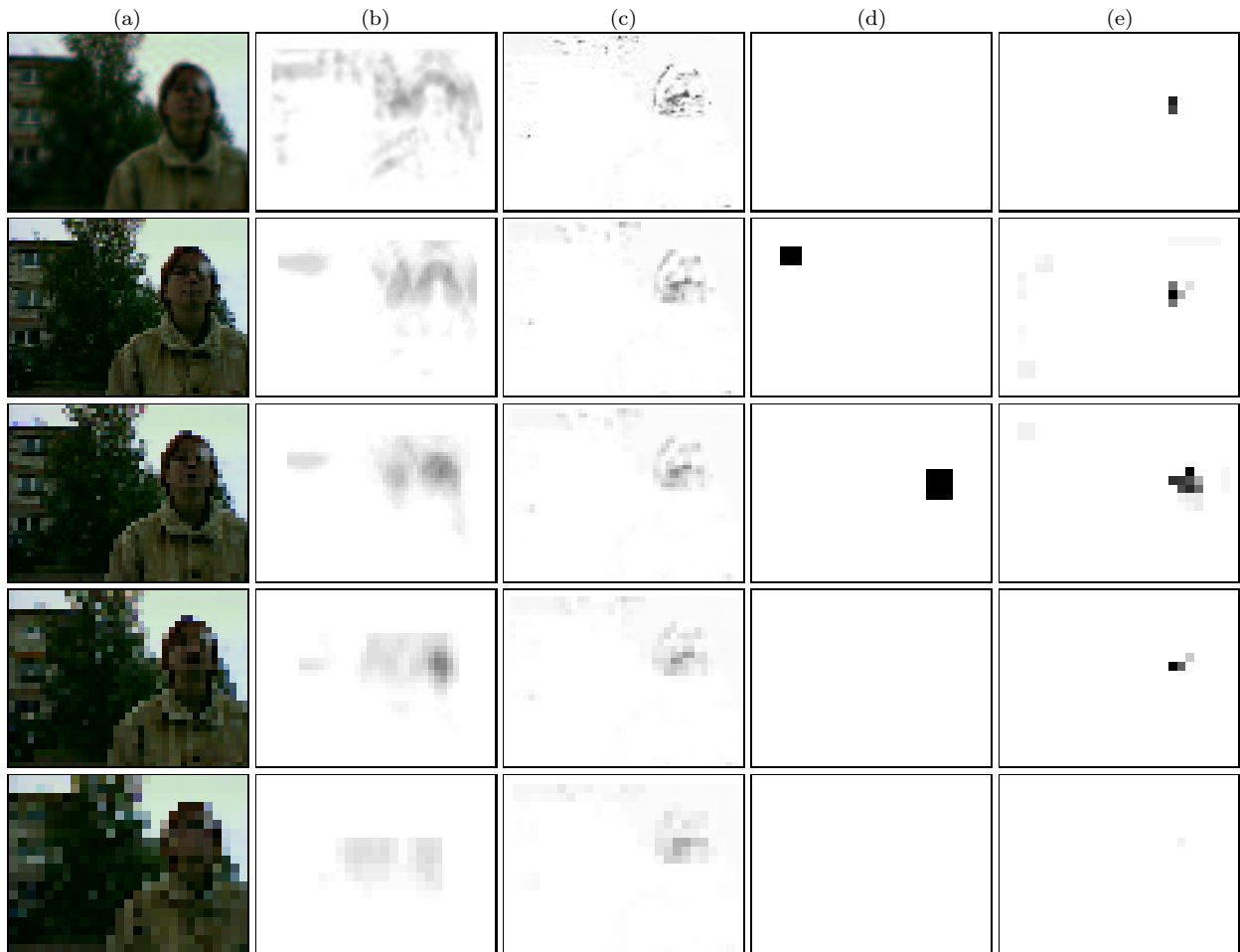


Abbildung 6.9: Lokalisierungsergebnis für Versuchsperson Sandy im Außenbereich auf ca. 0,75m: Kaum zum Tragen kommt hier die hautfarbbasierte Detektion in Spalte (c), erst die beiden anderen Cues, insbesondere die konturbasierte Detektion in Spalte (b) bringen den entscheidenden Beitrag zur Lokalisation, was sich bei der entsprechenden Blobausbildung mit größter Ausdehnung in der 3. Zeile als Ergebnis der dynamischen Selektion in Spalte (e) niederschlägt. Weitere Einzelheiten stellt Tab. 6.1 auf S. 88 dar.

Für dieses und alle folgenden Ergebnisbilder gilt folgende Anordnung der Zeilen und Spalten: In jeder **Zeile** sind alle Zwischenergebnisse und Ergebnisse für eine der fünf Auflösungsstufen (Abstand $1/\sqrt{2}$) dargestellt, in der 1. Zeile die 0. Ebene, in der 5. Zeile die 4. Ebene. Bei der hier gewählten minimalen Brennweiteinstellung führt ein Abstand Person – Kamera von 0,5m bei allen drei Merkmalsextraktionsverfahren zu einer optimalen Detektion auf der größten (anschaulich: obersten) Ebene der Auflösungspyramide (unterste Zeile). Entsprechend der Abstufung wird bei einem Abstand von 1,0m auf der mittleren Ebene und bei einem Abstand von 2,0m auf der feinsten Ebene (oberste Zeile) optimal detektiert.

Die fünf **Spalten** enthalten folgendes: Input in Spalte (a), das Ergebnis der Anwendung des Filterarrangements zur Detektion der Kopf-Schulter-Partie in Spalte (b), das Ergebnis der Hautfarbklassifikation in Spalte (c), die Gesichtsdetektion in Spalte (d), und schließlich das Ergebnis der Selektion innerhalb des dreidimensionalen Feldes dynamischer Neuronen in Spalte (e).

Tabelle 6.2: Lokalisationsergebnisse für Versuchsperson Autor im Außenbereich: Die Hautfarbdektion mit sehr guten Ergebnisse ebenso wie die beiden anderen Cues führen insgesamt zu guten Lokalisationsergebnissen.

Beide 1m-Einstellungen unterscheiden sich nur durch eine Seitwärtsverschiebung der Person. Am ersten Standort ist der Kontrast im Kopf-Schulter-Bereich geringer als am zweiten, was sich in der Genauigkeit dieses Cues niederschlägt, insgesamt jedoch keinen Einfluß auf das jeweilige Lokalisations hat.

Dist.	ld(e_S)	Ort	Bel.-richt.	Hautf.	Gesicht	Kontur	Person	Abb.
1,0m	5,0	außen	rundum	konz.	2	vert. 1-3	2 (2)	6.10
1,0m	5,0	außen	rundum	konz.	2	2	2 (2)	6.11
1,5m	5,0	außen	rundum	konz.	1	1	1 (1)	6.12
2,0m	5,0	außen	rundum	konz.	0	0	0 (0)	—

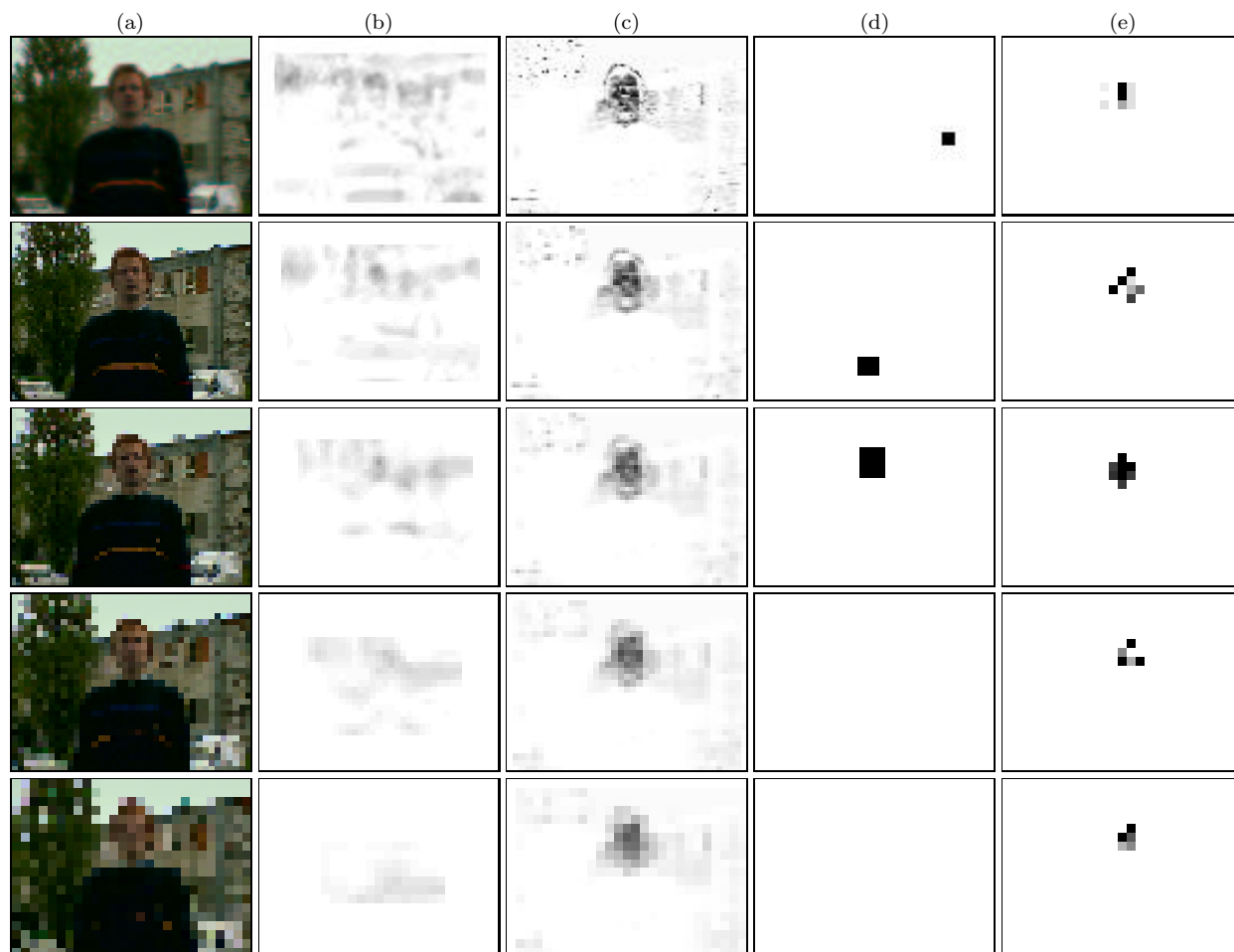


Abbildung 6.10: Lokalisationsergebnis für Versuchsperson Autor im Außenbereich auf ca. 1,0m: Personen- und Standortwechsel können eine drastische Änderung – hier eine sehr deutliche Verbesserung – der Hautfarbdetektion in Spalte (c) bringen. Allerdings tritt dadurch auch eine deutliche Verschmierung des im dynamischen WTA gebildeten dreidimensionalen Blobs in Spalte (e) über alle Ebenen auf – ein Problem, was gerade durch den lokalen Charakter des Farbmerkmals entsteht. Farbe bringt also keine besondere Information mit, in welcher Distanz die Person erscheint, nur die beiden anderen Komponenten sind sensitiv auf Entfernungsänderungen. Bei nur wenig ausgeprägter Konzentration der Konturdetektion in Spalte (b) ist hier schließlich die Gesichtsdetektion in Spalte (d) die Komponente, die das Zentrum der Blobbildung entscheidend beeinflusst. Weitere Einzelheiten stellt Tab. 6.2 auf S. 90 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

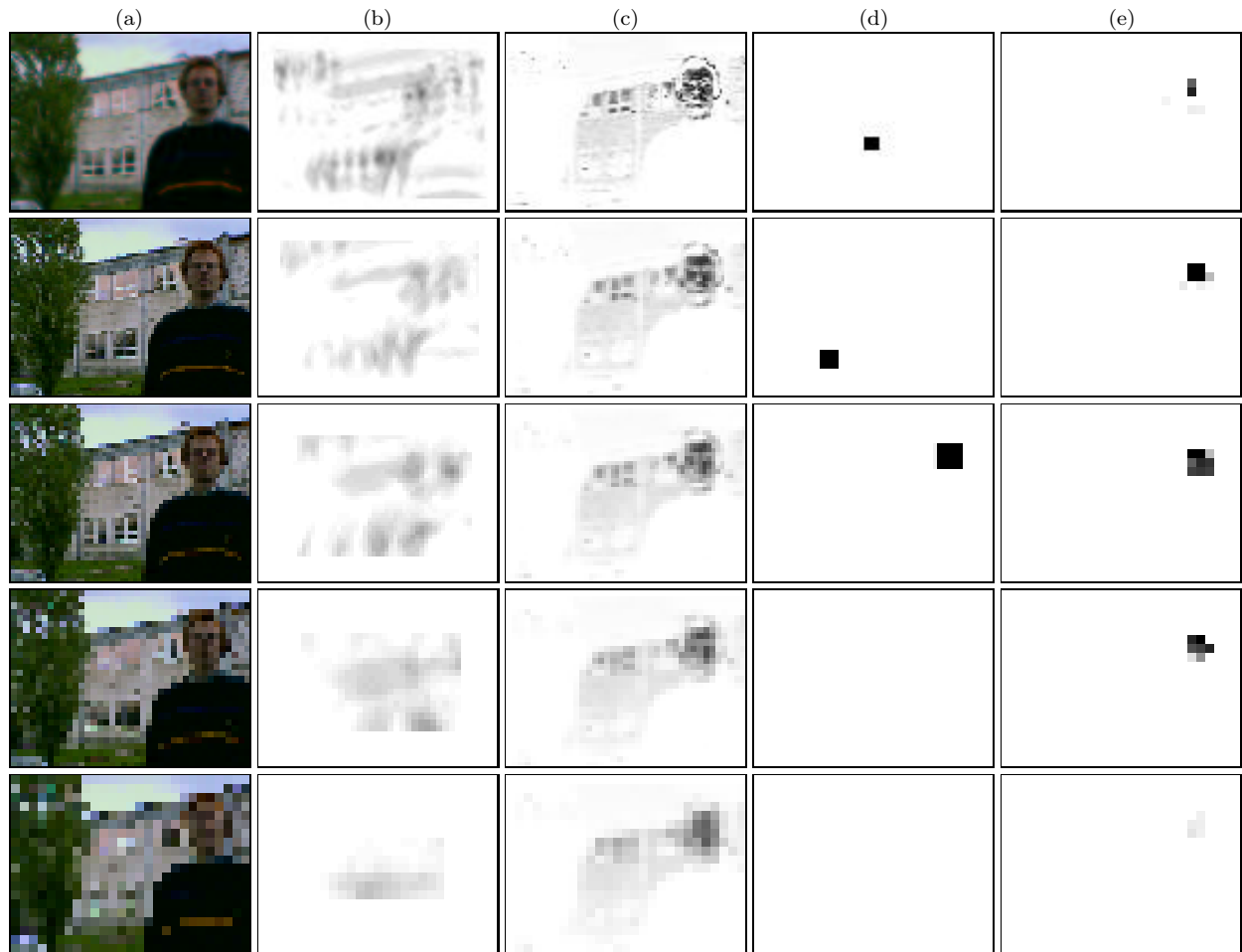


Abbildung 6.11: Lokalisierungsergebnis für Versuchsperson Autor im Außenbereich auf ca. 1,0m: Gegenüber Abb. 6.10 hat sich die Person nur zur Seite bewegt. Das geht offensichtlich einher mit einer insgesamt sich besser abzeichnenden Kontur der Kopf-Schulter-Partie, so daß sich dadurch der Blob etwas mehr konzentrieren kann. Da beide anderen Komponenten nahezu unverändert blieben, ist dadurch der Einfluß dieser kleinen Änderung auf das Gesamtergebnis gut nachvollziehbar. Weitere Einzelheiten stellt Tab. 6.2 auf S. 90 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.



Abbildung 6.12: Lokalisationsergebnis für Versuchsperson Autor im Außenbereich auf ca. 1,5m: Gegenüber den Abbn. 6.10 und 6.11 ist der Abstand der Person zur Kamera soweit vergrößert worden, daß das Gesicht auf der nächstfeineren Ebene detektiert wird, Spalte (d), 2. Zeile. Dadurch verschiebt sich das Blobzentrum in Spalte (e) hin zur 2. Zeile (1. Ebene). Die klaren Fehldetektionen von Gesichtern auf anderen Ebenen können wirksam unterdrückt werden. Anders als die beiden anderen Merkmale wird die Gesichtsdetektion generell nachbearbeitet, indem an die entsprechenden Koordinaten ein Rechteck genau der GesichtsfILTERgröße hineinprojiziert wird, da das Merkmal praktisch nicht translationsinvariant ist, wie das z. B. bei der Kopf-Schulter-Partie (in Grenzen) der Fall ist, so daß hier eine gewisse lokale Ausbreitung nicht inhärent ist. Die Rechtecke erscheinen entsprechend auf groben Auflösungsebenen größer als auf feinen. Weitere Einzelheiten stellt Tab. 6.2 auf S. 90 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

Tabelle 6.3: Lokalisationsergebnisse für Versuchsperson Christoph im Innenbereich (1): Die Anordnung der Person vor einem weitgehend homogenen Hintergrund im Innenraum bei frontaler Beleuchtung stellt eine deutliche komplexitätsmäßige Einschränkung dar. Dennoch sind die Ergebnisse nicht ideal. Dazu trägt hier der relativ kleine Kontrast zum Hintergrund im Kopfbereich bei. Obwohl dadurch die Kontur nur schwach detektiert wird, stellt das noch kein grundsätzliches Problem dar.

Mehr Probleme bringt die Häufung horizontaler Linien an der oberen Türeinfassung. Fehldetektionen liegen hier alle sehr nahe der eigentlichen Gesichtsposition und führen letztlich zur Streckung des Selektionsblobs über mehrere Ebenen, da die Artefakte immer auf mehreren Ebenen zugleich auftreten (siehe auch Abb. 6.13).

Dist.	ld(e_S)	Ort	Bel.-richt.	Hautf.	Gesicht	Kontur	Person	Abb.
0,5m	0,5	innen	frontal	konz.	4	3-4	4 (4)	–
1,0m	0,5	innen	frontal	konz.	2	2	2 (2)	–
1,5m	0,5	innen	frontal	konz.	1	0-1	1 (1)	6.13
2,0m	0,5	innen	frontal	konz.	0	0	0 (0)	–

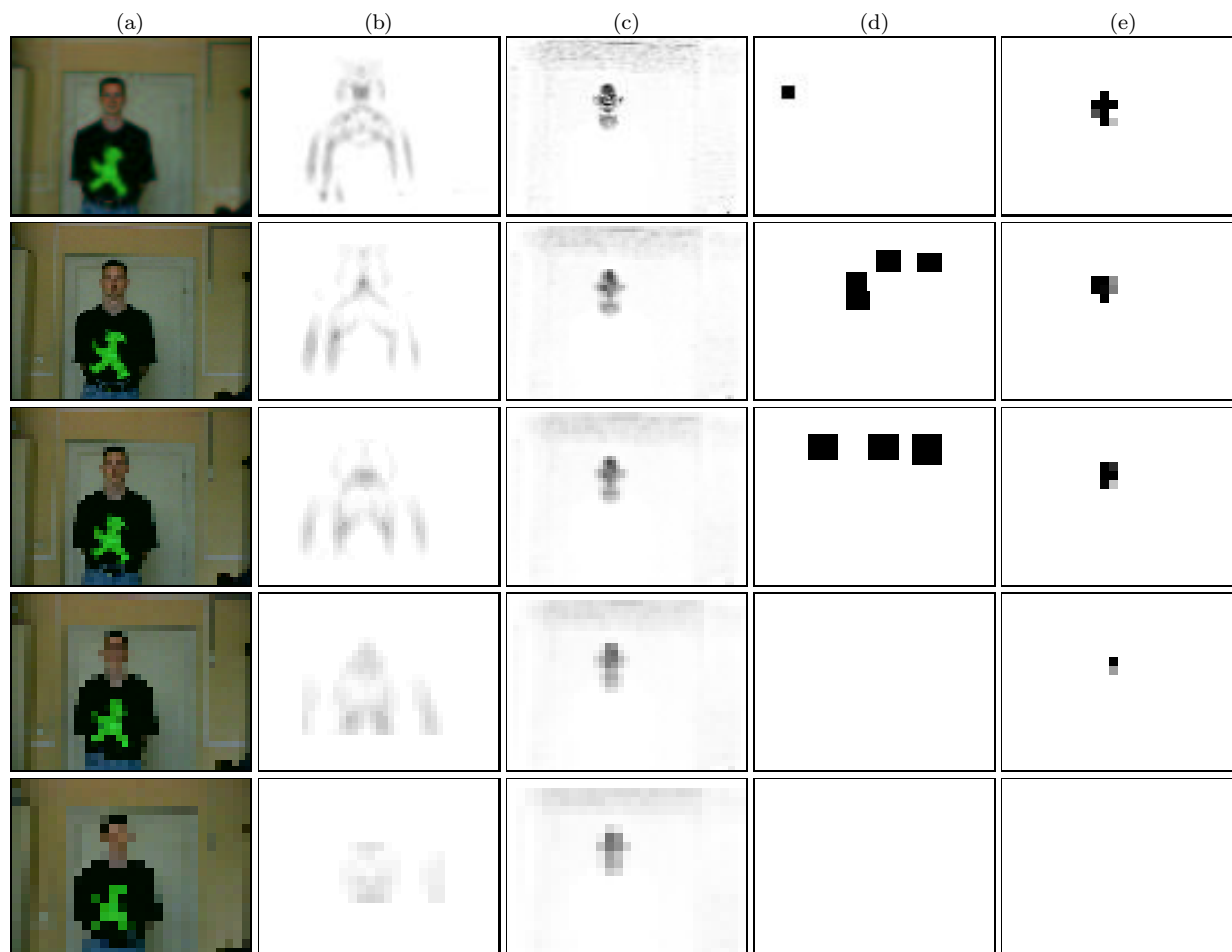


Abbildung 6.13: Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (1) auf ca. 1,5m: Unter bestimmten Bedingungen sind in Innenräumen sehr viele Einschränkungen der visuellen Szene möglich. Die Person vor nahezu homogenem Hintergrund stellt in diesem Fall ein Detektionsproblem vergleichsweise geringer Komplexität dar. Während etwa in Abb. 6.12 die konturbasierte Detektion kaum zu einer richtigen Lokalisation führt, ist diese in dieser Szene recht genau und lokal konzentriert. Dagegen erscheinen interessanterweise viele falsch positive gesichts-basierte Detektionen (Spalte (d)) im Bereich des oberen Türandes, die aber im dynamischen Selektionsprozeß wirksam unterdrückt werden. Insgesamt kann man erkennen, daß der entstandene Blob (Spalte (e)) seinen Schwerpunkt etwa im Bereich der ersten Ebene (2. Zeile) hat, so daß das Ergebnis auch hier entfernungsmäßig plausibel ist. Weitere Einzelheiten stellt Tab. 6.3 auf S. 94 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

Tabelle 6.4: Lokalisationsergebnisse für Versuchsperson Sandy im Innenbereich (1): Besonderheit in dieser Einstellung besteht in der unzureichenden Hautfarbklassifikation ausgelöst durch das Gegenlicht der im Blickfeld der Kamera liegenden Deckenleuchte infolge einer unzureichenden Abblendautomatik der Kamera.

Die Einstellung im Nahbereich von 0,75m führt zu einer falschen Entfernungsschätzung, da beide entfernungs sensitiven Cues gerade noch am unteren Rand ihres Toleranzbandes (z. T. darüber hinaus, siehe Tabn. E.1 und E.2). Keine weiteren Probleme treten dagegen bei den anderen getesteten Entfernungen auf, obwohl hier vom Kontur-Cue keine eindeutige Entfernungsschätzung einfließt.

Dist.	ld(e_S)	Ort	Bel.-richt.	Hautf.	Gesicht	Kontur	Person	Abb.
0,75m	0,75	innen	entgegen	–	2	1-2	2 (3)	6.14
1,0m	0,75	innen	entgegen	–	2	2	2 (2)	–
1,5m	0,75	innen	entgegen	–	1	0-1	1 (1)	–
2,0m	0,75	innen	entgegen	–	0	0-1	0 (0)	–

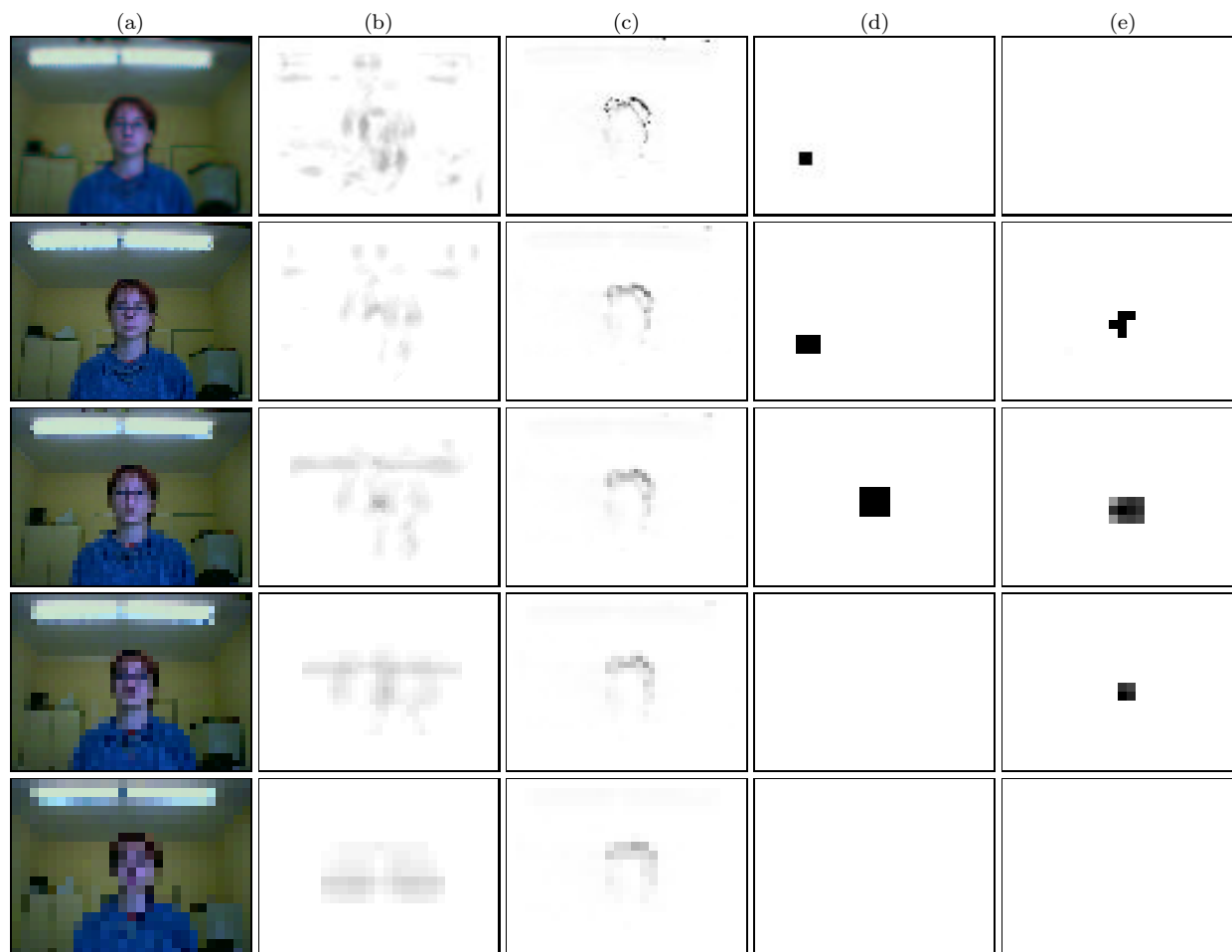


Abbildung 6.14: Lokalisationsergebnis für Versuchsperson Sandy im Innenbereich (1) auf ca. 0,75m: Hier ist durch den relativ geringen Kontrast zwischen Person und Hintergrund nur eine schwache konturbasierte Lokalisation möglich (Spalte (b)). Auch die hautfarbbasierte Lokalisation in Spalte (c) ist gestört. Das liegt daran, daß die Kamera z. T. direkt in Richtung Beleuchtungskörper gerichtet ist, wodurch kamerainterne elektronische Abblendmechanismen einsetzen, die allerdings offenbar zu einer ungleichgewichteten Intensitätsminderung führen, so daß die eigentlich sichtbaren Hautpartien nicht entsprechend klassifiziert werden können und stattdessen gerade der Haarbereich erfaßt wird. Da die Gesichtsdetektion in Spalte (d) als einziger der Cues ein deutliches und abgegrenztes Ergebnis bringt, bestimmt diese auch weitgehend das Lokalisationsergebnis. Jedoch ist die Entfernungsschätzung nicht exakt, beim eingestellten Abstand sollte der Ergebnisblob sein Zentrum eher in der 3. Ebene (4. Zeile) haben, die gerade einem Abstand von 0,71m entspricht. Weitere Einzelheiten stellt Tab. 6.4 auf S. 96 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

Tabelle 6.5: Lokalisationsergebnisse für Versuchsperson Christoph im Innenbereich (2): Es fällt hier auf, daß die Selektion fast immer über zwei Ebenen verteilt ausfällt, was hauptsächlich durch die Unschärfe des Konturcues bedingt ist.

Der Gesichts-Cue arbeitet einwandfrei, hingegen die Hautfarbdetektion erweist sich auch hier als abhängig von der Beleuchtungsrichtung (durch geringfügige Kameraneigung ist in der 0,5m-Einstellung der Beleuchtungskörper mit erfaßt worden).

Dist.	ld(e_S)	Ort	Bel.-richt.	Hautf.	Gesicht	Kontur	Person	Abb.
0,5m	0,5	innen	entgegen	verteilt	3	3-4	4 (4)	6.15
1,0m	0,25	innen	rundum	konz.	2	0-2	2 (2)	6.16
1,5m	0,5	innen	rundum	konz.	1	0-1	1 (1)	6.17
2,0m	0,5	innen	rundum	konz.	0	0-1	1 (0)	6.18

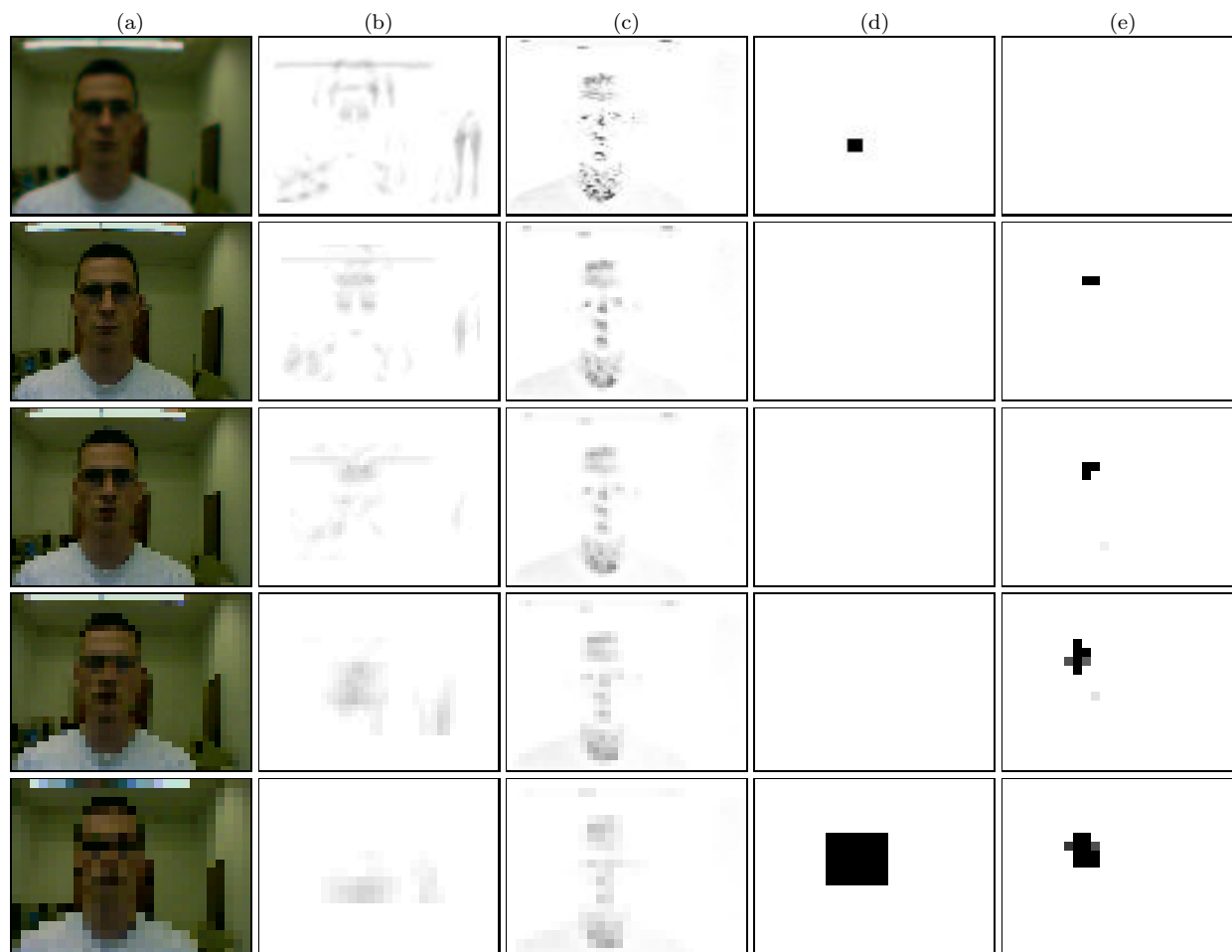


Abbildung 6.15: Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 0,5m: Hier wurde von der Person eine minimale Distanz eingenommen. Deutlich wird die präzise Gesichtsdetektion in Spalte (d), ähnlich auch die konturbasierte Detektion in Spalte (b), die gegenüber Abb. 6.14 von einem etwas besseren Kontrast Person/Hintergrund profitiert; gleichzeitig bleibt die hautfarbbasierte Detektion in Spalte (c) ähnlich dem Beispiel in Abb. 6.14 ungenau. Dennoch kann sich im dynamischen Selektionsprozeß ein Blob aktiver Neuronen so ausbilden, daß er am stärksten auf der Ebene ausfällt, auf der die Person größenrichtig abgebildet ist, (Spalte (e), 4. Zeile). Weitere Einzelheiten stellt Tab. 6.5 auf S. 98 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

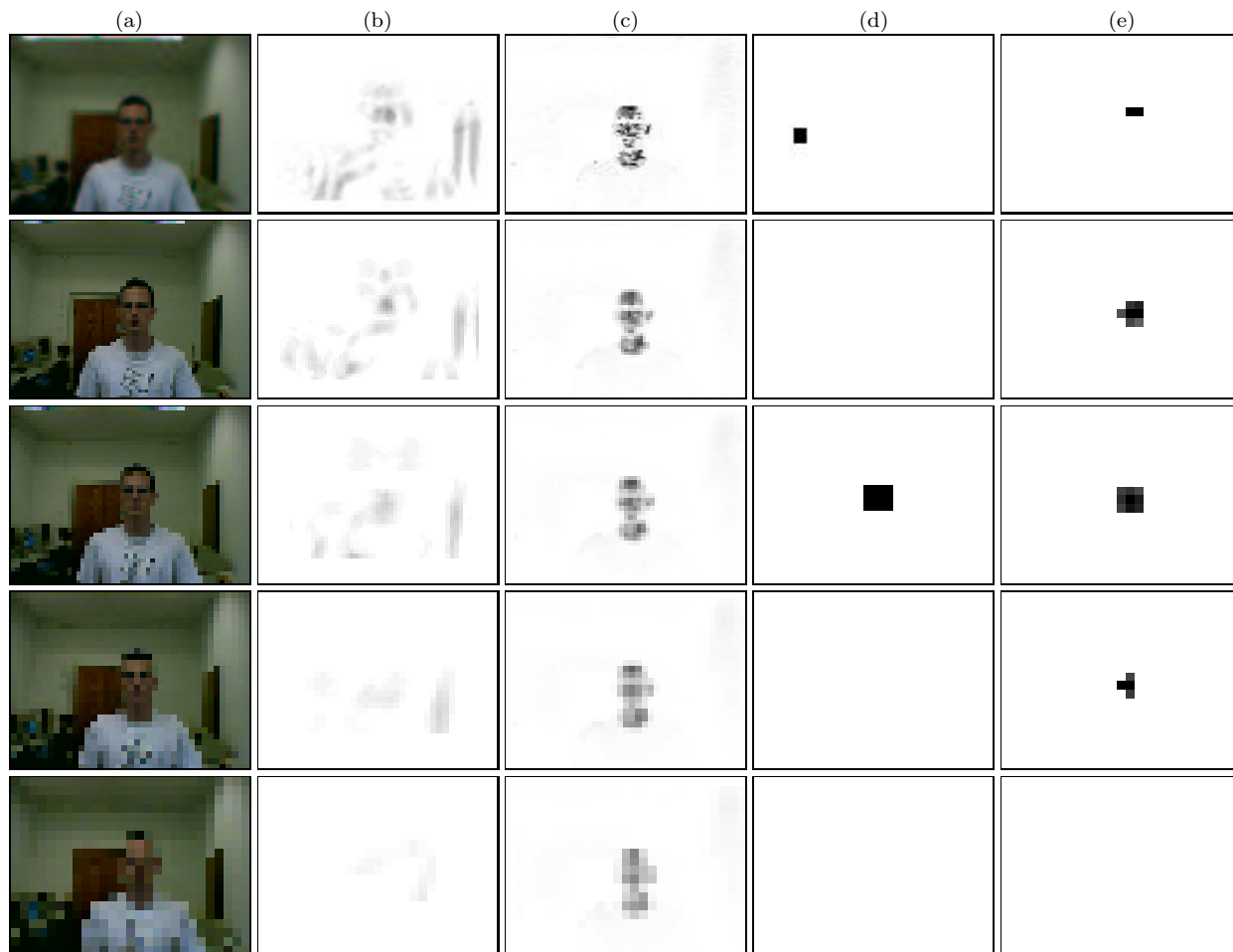


Abbildung 6.16: Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 1,0m: Dieselbe Person wie in Abb. 6.15 bei verdoppeltem Abstand. Auffällig ist die plötzliche Verbesserung der Hautfarbklassifikation in Spalte (c), die klar mit dem Verschwinden der Lampen aus dem Bild einhergeht (siehe den Fall in Abb. 6.14). Damit sind in diesem Beispiel alle drei Cues in gleichzeitig richtig präsent. Deutlich zu erkennen ist gegenüber Abb. 6.15 die Verlagerung des Lokalisationsblobs in Spalte (e) um zwei Ebenen nach oben, was bei einem $1/\sqrt{2}$ -Abstand der Ebenen auch genau der vorgenommenen Entfernungsänderung entspricht. Weitere Einzelheiten stellt Tab. 6.5 auf S. 98 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

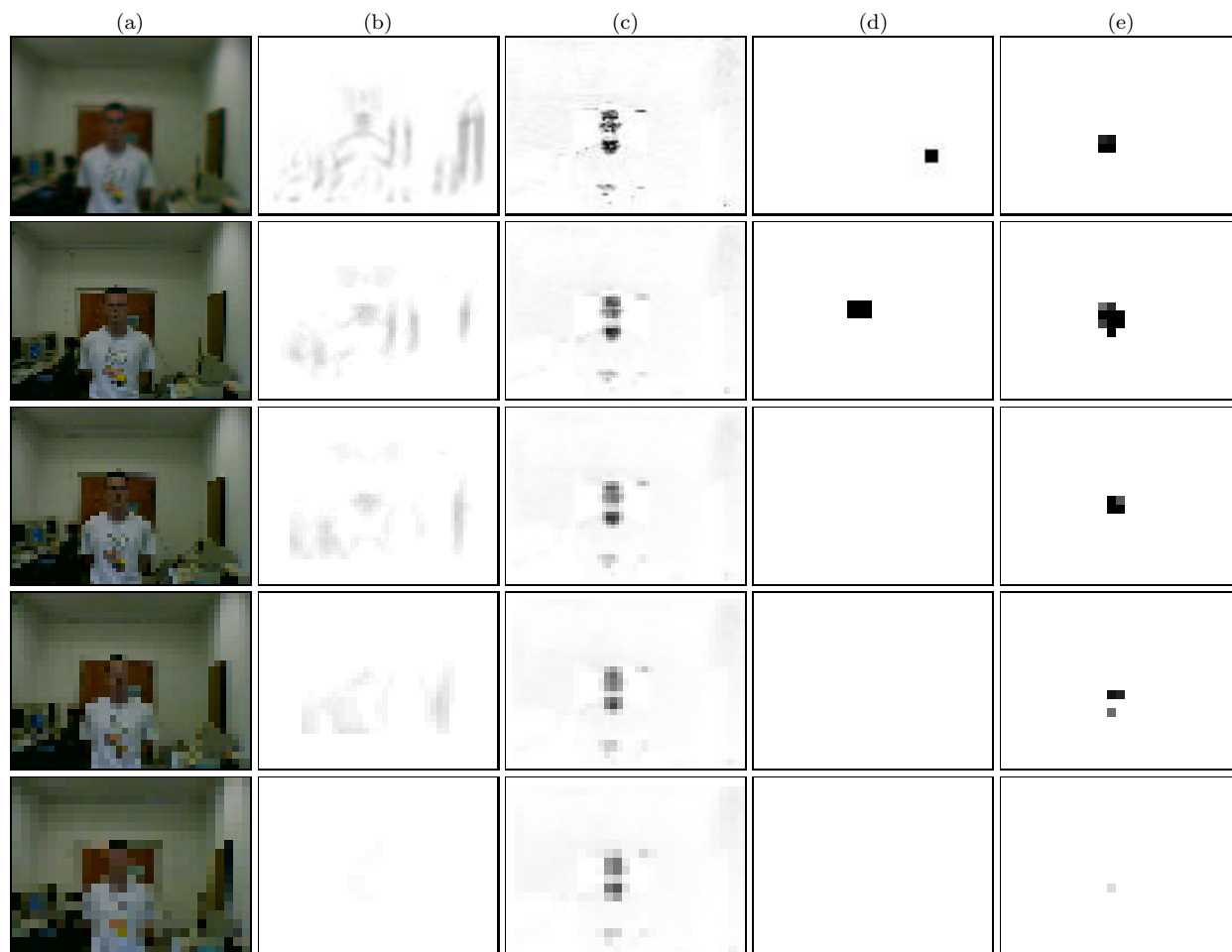


Abbildung 6.17: Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 1,5m: Eine weitere Distanzerhöhung führt auch hier zu einem klaren und plausiblen Ergebnis. Das Blobzentrum in Spalte (e) ist gegenüber Abb. 6.16 eine weitere Ebene nach oben ausgebildet. Dadurch, daß hier wieder ein relativ schwacher Kontrast zwischen Kopf und Hintergrundbereich (Tür) vorherrscht, ist auch das konturbasierte Lokalisationsergebnis in Spalte (b) nicht mehr so stark wie in Abb. 6.16. Trotzdem durch die pyramidale Struktur die räumliche Ausdehnung der Stimuli aus den Cues für das neuronale Feld immer mehr abnimmt, bleibt die Anzahl der am Blob beteiligten aktiven Elemente in etwa gleich (die Summe aller Aktivität für einen Gleichgewichtszustand ist konstant), was die Eignung der vorgenommenen Umkodierung hin zu einer kubischen Struktur des neuronalen Feldes belegt. Weitere Einzelheiten stellt Tab. 6.5 auf S. 98 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.



Abbildung 6.18: Lokalisationsergebnis für Versuchsperson Christoph im Innenbereich (2) auf ca. 2,0m: Hier steht die Person in maximaler Entfernung für eine Lokalisation unter den gegebenen technischen Randbedingungen. Was auffällt, ist die Verschlechterung des konturbasierten Ergebnisses in Spalte (b). Das liegt an der zufällig offenstehenden Seitentür (im Bild rechts), die mit einem hellen Plakat beklebt ist, was ganz automatisch auf bestimmten Auflösungsebenen vom konturbasierten Verfahren als ähnlich zu Bestandteilen der typischen Kopf-Schulter-Partie gewertet wird. Ähnlich verhält es sich beim sichtbaren vertikalen Spalt zwischen den beiden Türen, was sich in den großflächigen Fehldetektionen auf der 2. und 3. Ebene (3. und 4. Zeile) niederschlägt. Damit ist auch eine bestimmte Grenze für das Verfahren dieses Cues aufgezeigt. Interessant ist vielleicht noch, daß man hautfarbbasiert deutlich die Hände lokalisieren kann (Spalte (c)), ein Nebeneffekt, der durch das von der Person gehaltene Maßband mit ins Bild gekommen ist. Weitere Einzelheiten stellt Tab. 6.5 auf S. 98 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

Tabelle 6.6: Lokalisationsergebnisse für Versuchsperson Sandy im Innenbereich (2): Diese Versuchsreihe untersucht den Einfluß seitlicher Beleuchtung. Überraschend sind hier keine Ausfälle bei der Gesichtsdetektion zu verzeichnen. Auch die Konturdetektionen funktioniert, wenn auch überlappend.

Lediglich der Hautfarb-Cue funktioniert eingeschränkt, indem nur die direkt beleuchtete Gesichtshälfte richtig klassifiziert wird.

Dist.	ld(e_S)	Ort	Bel.-richt.	Hautf.	Gesicht	Kontur	Person	Abb.
0,75m	0,5	innen	seitlich	halb	3	3	3 (3)	–
1,0m	0,5	innen	seitlich	halb	2	2-3	2 (2)	6.19
1,5m	0,5	innen	seitlich	halb	1	1-2	1 (1)	6.20
2,0m	0,5	innen	seitlich	halb	0	0-1	0 (0)	–

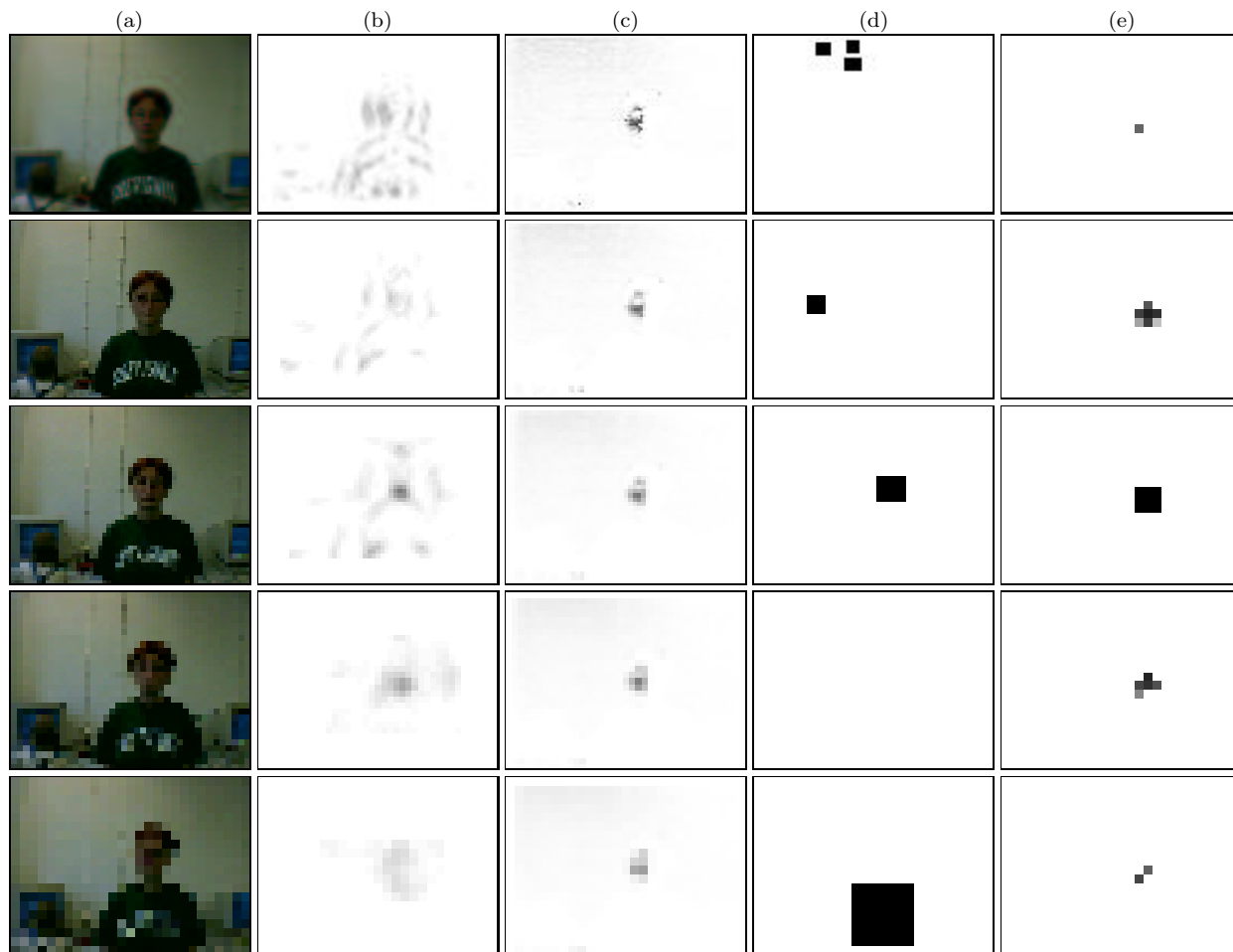


Abbildung 6.19: Lokalisationsergebnis für Versuchsperson Sandy im Innenbereich (2) auf ca. 1,0m: Mit diesem Beispiel, das unter relativ schlechten Beleuchtungsbedingungen aufgenommen wurde, wird noch einmal belegt, daß ohne solche direkten Gegenlichtverhältnisse wie in den Abbn. 6.14 und 6.15 auch unter schwacher Beleuchtung die richtige Hautfarbklassifikation möglich ist (Spalte (c)). Problemlos arbeitet das konturbasierte Verfahren in Spalte (b). Artefakte bei der Gesichtsdetektion in Spalte (d), die auf verschiedenen Ebenen auftreten, lassen sich nicht ganz ausschließen. Hervorzuheben ist aber, daß die Grundrichtung der direkten Beleuchtung seitlich ist (von links), dennoch auch das tatsächliche Gesicht sicher gefunden wird (mittlere Ebene). Das Blobzentrum in Spalte (e) liegt ebenfalls auf der mittleren Ebene, so daß das Ergebnis richtungs- und entfernungsmäßig plausibel ist. Weitere Einzelheiten stellt Tab. 6.6 auf S. 103 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

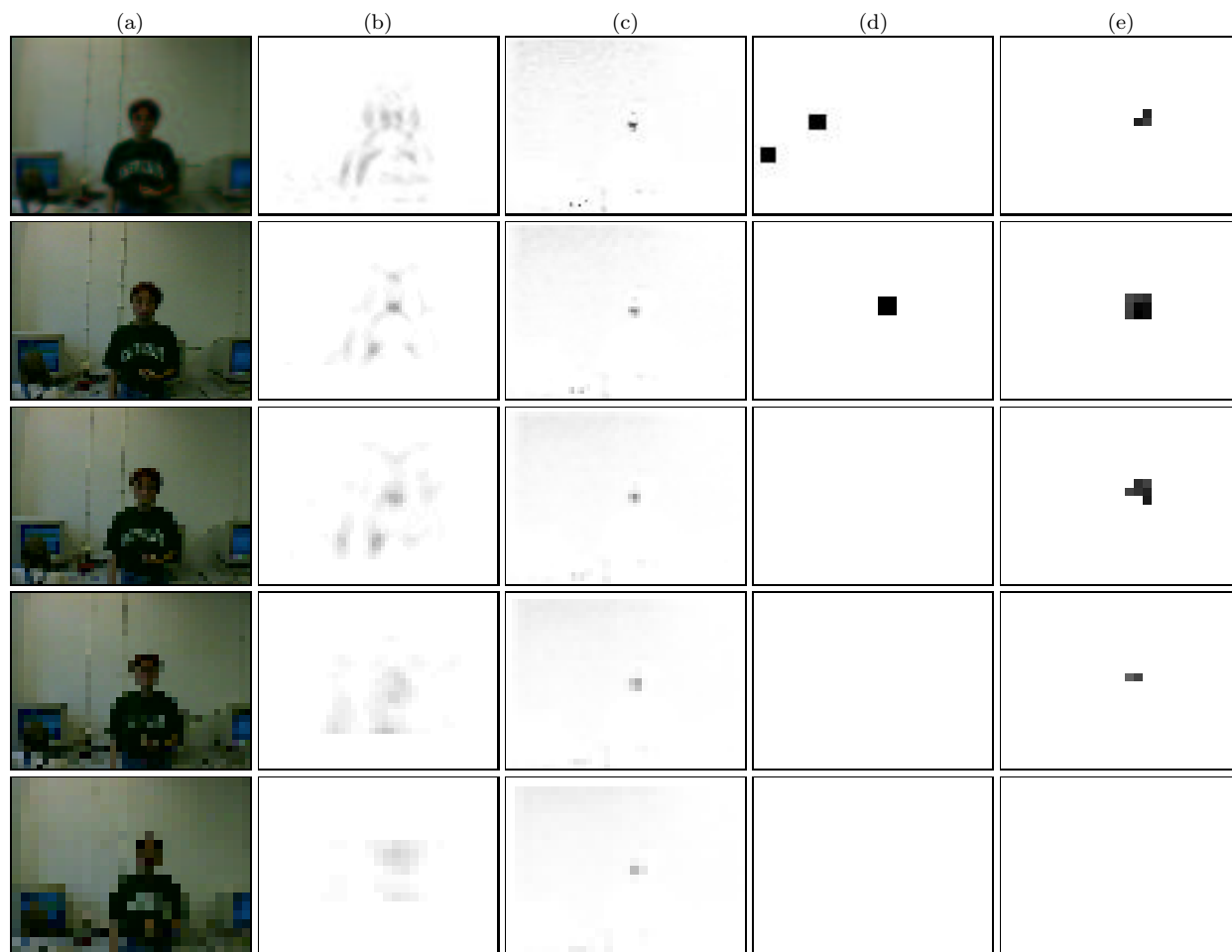


Abbildung 6.20: Lokalisationsergebnis für Versuchsperson Sandy im Innenbereich (2) auf ca. 1,5m: Auch dieses Beispiel mit einem gegenüber dem Beispiel in Abb. 6.19 um 0,5m größeren Abstand führt zu einer richtungs- und entfernungsmäßig plausiblen Lokalisation der Person. Die 2. Ebene, auf der sich der Blob konzentriert (3. Zeile in Spalte (e)), bildet idealerweise Personen in einer Entfernung von 1,41m größenrichtig ab. Allerdings ist der lokale Kontrast im Gesicht durch die Bedingungen schon so schwach, daß dieses Gesichtsdetektionsergebnis in Spalte (d) schon die äußerste Grenze des realisierbaren darstellt. Weitere Einzelheiten stellt Tab. 6.6 auf S. 103 dar.

Zur Bedeutung der Zeilen und Spalten siehe Abb. 6.9 auf S. 89.

Alle Ergebnisse wurden mit folgenden Parametern für die Gln. 6.4 und 6.6 zur Fusion bzw. Selektion erzielt:

- $\gamma = 0,1$ Wichtung des Min-Max-Operators (Gl. 6.4)
- $c_i = 1,0$ Wichtung der in $[0, 1]$ liegenden ext. Erregung (Fusionsergebnis, Gl. 6.6)
- $c_l = 0,1$ Wichtung der lateralen Erregung (Gl. 6.6)
- $c_h = 0,2$ Wichtung der feldglobalen Inhibition (Gl. 6.6)
- $s = 64,5$ Steilheit der logistischen Sigmoidfunktion (Gl. 6.7)
- $\frac{\Delta T}{\tau} = 0,1$ Iterationsschrittweite (Gl. 6.21)
- $k_{\text{STOP}} - k_{\text{START}} + 1 = 50$ Iterationsschrittzahl (Abbruchkriterium, S. 83)

Für den Rechenzeitbedarf wurden folgende typische Verarbeitungszeiten auf dem PC des genutzten Computervision-Arbeitsplatzes (CPU: Pentium I mit 200MHz) ermittelt:

0,2s	Bildakquisition (2%)
0,3s	Hautfarbdetektion (3%)
5,2s	Gesichtsdetektion (53%)
1,4s	Konturdetektion (14%)
+ 2,8s	Personenlokalisierung (Fusion/Selektion, 28%)
9,9s	Gesamtzeit zwischen Bildaufnahme und Lokalisation

Es fällt hier der relativ große anteilige Wert von 53% der Rechenzeit für die Gesichtsdetektion auf. Der Grund liegt insbesondere in der erforderlichen feineren Auflösung dieser Pyramide. Sie beträgt mit 256×191 , 181×135 , 128×95 , 91×67 und 64×48 ca. ein Drittel bis ein Zwölftel der Originalauflösung des Framegrabbers. Demgegenüber laufen Hautfarbdetektion und Konturdetektion auf Auflösungspyramiden mit Ebenen von 109×81 , 77×57 , 55×41 , 39×29 und 27×20 , also ca. ein Siebtel bis ein Achtundzwanzigstel der Originalauflösung. Das 3D dynamische neuronale Feld besitzt eine Auflösung von $27 \times 20 \times 5$. Für eine weitere Diskussion des Rechenzeitbedarfs wird auf den nachfolgenden Abschnitt 6.5 *Résumé* und das abschließende Kapitel 7 *Zusammenfassung und Ausblick* verwiesen.

6.5 Résumé

Dieses Kapitel stellte das Vorgehen bei der Nutzung der Auffälligkeitskomponenten zur Personenlokalisierung vor. Es wurde ein eigenes zweischrittiges Konzept entwickelt, das zunächst die Fusion der Cues und daran anschließend eine Selektion innerhalb des Fusionsergebnisses beinhaltet. Für die Fusion selbst wurde ein Min-Max-Fuzzy-Operator gewählt, der sowohl die Nachteile einfacher Superposition oder einfacher Maximumsauswahl vermeidet, jedoch nicht generell auf die Vorteile der Nutzung von Koinzidenzen verzichtet. Damit erhält der Fusionsschritt einen integrativen Charakter. Für den anschließenden Selektionsprozeß, dessen Zielstellung die eigentliche Lokalisation der Person ist, wurde ein dynamisches neuronales Feld eingesetzt. Darin wird iterativ ein einzelnes Selektionsergebnis gebildet.

Die gezeigten Ergebnisse belegen, daß das entworfene System zur Personenlokalisierung eines mobilen Robotersystems erfolgreich einsetzbar ist. Zu den Vorteilen des Ansatzes ist zu zählen, daß durch den expliziten Fusionsschritt eine Reduzierung des Aufwandes zur Selektion eintritt, so daß sich insgesamt auf ein einzelnes 3D dynamisches neuronales Feld beschränkt werden kann. Dadurch wird das System handhabbarer und überhaupt sinnvoll demonstrabel. Weiterhin erfolgt eine gute Störungsunterdrückung in der Fusion mittels Fuzzy-Operator, da Teilkoinzidenzen erforderlich sind. Die Selektion selbst erfolgt iterativ durch kompetitive lokale und feldglobale Wechselwirkungen innerhalb eines dynamischen neuronalen Feldes. Dies stellt ein zugleich datengetriebenes und zustandsbasiertes Verfahren dar, wodurch sich die Robustheit der Selektion gegenüber Artefakten erhöht.

Jedoch stellt die Wahl eines iterativen Selektionsverfahrens auch hohe Anforderungen an die Rechenleistung. Deshalb stellt dieser Verarbeitungsschritt einen sehr rechenzeitintensiven Teil im Gesamt Ablauf dar (*siehe* die Übersicht auf S. 106).¹⁹ Da eine geschlossene Lösung des nichtlinearen DGL-Systems nicht verfügbar ist, sollte für kommende Arbeiten geprüft werden, ob und ggf. mit welchen Nachteilen weitere Vereinfachungen des Lösungsverfahrens möglich sind.

Zur Ausführung der lateralen Gewichtungsfunktion w aus Gl. 6.6 legen weitere Überlegungen und Anregungen (KRABBES /Kra1997/) nahe, die Varianzen („Breiten“) der beiden Gaußglocken ebenenweise derart zu variieren, daß sie der Pyramidendivergenz angepaßt sind. Damit hätte dann etwa auf der obersten Ebene des Kubus’ (*siehe* Abb. 6.4) die dort jeweils verwendete Mexikanerhutfunktion eine vergleichsweise große Ausdehnung und auf der untersten Ebene dagegen eine kleinere. Die Motivation wäre, daß adäquate Stimuli auf der obersten Ebene relativ gesehen am ausgedehntesten erscheinen, wohingegen auf der un-

¹⁹Der Rechenzeitbedarf für die Gesichtsdetektion liegt zwar doppelt so hoch, hat aber seine Ursache im höheren Auflösungsbedarf dieses Cues.

tersten Ebene das umgekehrte der Fall ist. Jedoch wurde im Rahmen dieser Arbeit aus Gründen der Einfachheit und zugunsten einer effizienten Umsetzbarkeit (etwa linear separierbarer Faltungskern) auf eine Realisierung verzichtet und stattdessen ein symmetrischer uniformer Kern eingesetzt.

Die Problematik von realen Alternativen (z. B. zwei potentiell gleichermaßen relevante Personen in der Szene) wird nicht explizit behandelt, um das Szenario zunächst zu einfach zu halten. In einer konkurrenten Situation wird sich die erforderliche Zahl von Iterationsschritten zwangsläufig erhöhen, bevor sich die Selektion einer der Personen durchsetzt. In speziellen Fällen ist keine eindeutige Selektion möglich. Nähere Untersuchungen dazu müssen für einen erweiterten Einsatz des Systems vorgenommen werden.

Generell angestrebt werden sollte eine weitere Erhöhung der örtlichen und zeitlichen Auflösung, um eine deutlich erweiterte Arbeitsentfernung bzw. eine schnellere Selektion zu erzielen. Die Voraussetzungen dafür werden insbesondere auch durch die Fortentwicklung der rechentechnischen Möglichkeiten mittelfristig gegeben sein.

Kapitel 7

Zusammenfassung und Ausblick

7.1 Zusammenfassung

Die Arbeit hatte die Entwicklung eines Mehrkomponentenansatzes zur visuellen Personenlokalisierung bei der Mensch-Maschine-Interaktion (MMI) zwischen einem teilautonomen mobilen Servicesystem und gestenbasiert damit arbeitenden Nutzern zum Ziel. Zur Lösung dieser Problemstellung wurden Themenschwerpunkte bearbeitet, die sich im Aufbau dieser Arbeit unmittelbar niederschlagen. Sie beginnt zunächst mit einer Einleitung im Kapitel 1 *Motivation* zur Erläuterung von Problem- und Zielstellung. Darauf folgt ein Kapitel 2 *Pyramidale Repräsentationen als Basiskonzept*, in dem diese für diese Arbeit grundlegende Methodik detailliert dargestellt wird. Daran schließt an in den Kapiteln 3–5 die Darstellung der einzelnen verwendeten Cues *Hautfarbdetektion*, *Gesichtsdetektion* bzw. *Konturdetektion*. Diese drei Komponenten bilden die Basis der vorgeschlagenen Lösung, die dann im Kapitel 6 *Personenlokalisierung* vorgestellt wurde. Der verwendete Multi-Cue-Ansatz wurde aus folgenden Gründen favorisiert:

- einzelne Auffälligkeitskomponenten sind
 - (a) nur mit hohem algorithmischen Abstimmungsaufwand für eine hinreichende Konstanzleistung akzeptabel robust und
 - (b) a priori nur unter bestimmten Bedingungen einsetzbar;
- im Zusammenwirken können Beschränkungen einzelner Komponenten überwunden werden, ohne Aufwand und Perfektion der Einzelkomponenten besonders anheben zu müssen

Die erfolgreiche Funktionsweise des Ansatzes konnte an zahlreichen Beispielen belegt werden. Es kann damit auf einem mobilen Servicesystem für die visuelle Lokalisation eines damit

gestenbasiert interagierenden Nutzers zum Einsatz kommen. Die Arbeit liefert damit einen Beitrag für die Entwicklung und Etablierung einer neuen MMI-Schnittstelle.

Das System selbst ist unter Echtzeitbedingungen demonstrierbar. Die für den Computervision-Arbeitsplatz ermittelte Gesamtrechenzeit von ca. 10s zwischen Bildaufnahme und Personenlokalisierung wird noch als vertretbar angesehen, motiviert aber ungeachtet der mittlerweile deutlich erhöhten Rechenleistung schon von Standard-PCs zur weiteren Optimierung der Algorithmen und damit zur Senkung des Rechenzeitbedarfs. In der Gesichtsdetektion, die mehr als die Hälfte der Rechenzeit ausmacht, wird allerdings nur mit einer Senkung der Auflösung eine entscheidende Verbesserung möglich sein. Hier sind weitere gezielte Untersuchungen erforderlich, in denen auch der gesichtsspezifische Operand variiert werden sollte.

Der Ansatz dieser Arbeit unterscheidet sich von anderen in der Literatur vorgestellten zunächst in der Zusammensetzung der Auffälligkeitskomponenten. Neben der häufig anzutreffenden Hautfarb- und Gesichtsdetektion wird hier mit dem Kontur-Cue eine Komponente erarbeitet, die so bisher keine Verwendung gefunden hat. Es hat sich gezeigt, daß sich die Kopf-Schulter-Partie von frontalen Personensilhouetten gerade auch unter ungünstigen Beleuchtungsverhältnissen gut detektieren läßt. Die Methodik, zur Detektion ein Konturmodell mittels eines Arrangements orientierter Filter zu entwerfen, hat sich einfach umsetzen lassen. Er profitierte besonders von der Nutzung der 2D-Tensor-Methode.

Da die Lokalisierung auch eine Entfernungsschätzung der Person einbezieht, war eine durchgängige Behandlung der Einzelcues wie auch der Fusion und Selektion in pyramidalen bzw. kubischen Repräsentationen erforderlich. Um dafür eine akzeptable Genauigkeit ermöglichen zu können, wurde dabei mit einer Halboktav-Abstufung gearbeitet.

Für die Fusion der Cues und die anschließende Selektion zur Lokalisierung sind eigene Verfahren entwickelt worden, die aus einer Analyse der Schwachpunkte publizierter Verfahren resultieren. Die *Fusion* der Cues zu *einem* personenspezifischen Auffälligkeitsmaß erfolgt *integrativ* mittels einer Fuzzy-Min-Max-Operation, die auf einfache Weise die Nachteile von einfacher Superposition oder Maximumsauswahl vermeidet, jedoch ebenso Koinzidenzen berücksichtigt.

Die kompetitive *Selektion* wird durch einen WTA-Prozeß innerhalb eines auf 3D erweiterten dynamischen neuronalen Feldes erzielt. Mit einfachen internen Wechselwirkungsmechanismen (lokales Kooperations-, globale Konkurrenz) kann dieser Kohärenzen in der fusionierten dreidimensionalen Auffälligkeitskarte auffinden. Das Feld arbeitet sowohl datengetrieben als auch zustandsbasiert. Es besteht ein immanentes Hystereseverhalten, da die Gewinnerregion sich selbst räumlich und zeitlich stabilisieren kann. Bei den zu verarbeitenden Bildfolgen ist es unter Annahme einer gewissen Kontinuität somit möglich, Störungen

zu überbrücken. Die erforderliche iterative Lösung besitzt nur sehr geringe Anforderungen an das numerische Lösungsverfahren.

7.2 Ausblick

Diese Arbeit stellt ein Zwischenergebnis eines vielversprechenden Ansatzes für die MMI dar. Wie an verschiedenen Stellen der Arbeit bereits ausgeführt wurde, besteht für eine genauere quantitative Bewertung der Funktionsweise des vorgestellten Systems Bedarf an weiteren Tests und Untersuchungen.

Der Ansatz zur Personenlokalisierung für diese Arbeit bildete den ausdrücklichen Versuch, nur so genau wie nötig und damit aufwandsmäßig so gering wie möglich zu bleiben. Dafür wurde eine Reihe von Annahmen getroffen, die bis hin zur Auswahl der Cues führten. Im Rahmen von Arbeiten für den direkten Einsatz auf realen Robotersystemen sind mittlerweile weitere Erfahrungen gesammelt worden. Daraus resultieren Anregungen zu Veränderungen und Erweiterungen der Cues, die in diese Arbeit nicht eingegangen sind, da sie aus Gründen begrenzter Ressourcen bisher bewußt außer Acht gelassen worden sind. Hier besteht generell weiterer Entwicklungsbedarf. HEMPEL hat zwischenzeitlich den Ansatz zur Personenlokalisierung aus dieser Arbeit aufgegriffen und nach Tests auf einer mobilen Roboterplattform unter den realen Bedingungen eines Baumarktes modifiziert in ein System zur Personenverifikation und zum Personentracking integriert (*siehe* /Hem2000/).

Die Tests erfolgten bisher nur mit einzelnen Person. Hier werden jedoch keine prinzipiellen Einschränkungen gesehen. Es wird erwartet, daß bei mehreren Personen im Bild diejenige lokalisiert wird, für die die Fusion der Cues zu einer höheren Auffälligkeit führt. Für die Selektion dieser Person ist jedoch mit einer höheren erforderlichen Schrittzahl für die Iterationen im 3D dynamischen neuronalen Feld zu rechnen. Hierzu besteht für kommende Arbeiten weiterer Untersuchungs- und ggf. Entwicklungsbedarf.

Ausgeklammert bleiben konnten in dieser Arbeit andere als frontale Ansichten von Personen. Wie schon wenige Tests bestätigt haben (*siehe* Tab. E.3), sind die hier vorgeschlagenen Einzelcues dafür nur bedingt einsetzbar, da sie grundsätzlich ein starres Pattern Matching realisieren. Es wird erwartet, daß die Hinzunahme von Cues, die auf *elastischem Graph Matching* basieren, hier weitere Möglichkeiten eröffnen. Zugleich sollte künftig der nicht unerhebliche Aufwand getrieben werden, zur besseren Quantifizierbarkeit eine ausgewertete Testszenengalerie mit geeignetem Szenenspektrum zu erstellen.

Erforderlich ist, alle Genauigkeitsanforderungen auch an praktischen Kriterien zu messen, die aus Experimenten mit realen Systemen zur Gesten- oder Posturerkennung (*siehe*

CORRADINI/Cor2000/) zu ermitteln sind. Nur eine *effiziente, auf das wesentliche beschränkte Methodik* mit leistungsfähigen Algorithmen zur Personenlokalisierung wird einen sinnvollen Beitrag dafür leisten, sich weiter an die Zielstellung der MMI im Szenario eines teilautonomen mobilen Servicesystems anzunähern, einen *kontinuierlichen Aktions-Wahrnehmungs-Zyklus* in realen Umwelten aufzubauen. Mit dieser Arbeit ist ein weiterer Schritt in diese Richtung getan worden. Ihre Ergebnisse bilden eine tragfähige Basis für künftige Untersuchungen, für die eine Vielzahl von Anknüpfungspunkten bestehen, wie etwa die Erweiterung um nicht-frontale Ansichten, eine adaptive Gewichtung der Cues vor deren Fusion oder eine Verbindung zu akustischen Personenlokalisationsverfahren.

Anhang A

Ansätze zur referenzfreien Farbkonstanz

Unter Farbkonstanz versteht man die Repräsentation von Oberflächenfarben derart, daß eine Farberscheinung nicht von den spektralen Eigenschaften der Lichtquellen abhängt und damit von spektralen Veränderungen unbeeinflusst bleibt. Modelle, die die psychophysisch beobachtbare Farbkonstanz erklären und als Grundlage für Farbkonstanz in technischen Sehsystemen dienen können, lassen sich in zwei allgemeine Gruppen einteilen:

- referenzfreie Farbkonstanzmodelle
- referenzbezogene Farbkonstanzmodelle

Für die vorliegende Arbeit sind letztere nicht oder nur schlecht einsetzbar, da sie auf ein sicher im Bild enthaltenes definiertes Referenzobjekt angewiesen sind. Die für diese Arbeit in Frage kommenden *referenzfreien Modelle* sind vom Prinzip her schon sehr lange bekannt. Bereits anfangs des 20. Jahrhunderts entwarf VON KRIES /Kri1905/ ein Modell, in dem die Empfindlichkeiten der drei Sensoren unterschiedlich gewichtet werden können. HERING /Her1905/ vermutete eine Gegenfarbendarstellung in höheren (corticalen) visuellen Verarbeitungsstufen, da er anders sogenannte Nachbilder (Sukzessivkontrast) in Gegenfarben bei geschlossen Augen nicht erklären konnte. Entsprechend sollten Gewichtungsänderungen zur Farbumstimmung innerhalb dieser Gegenfarbendarstellung vorgenommen werden.

A.1 Retinex-Verfahren

Das bekannteste referenzfreie Modell, das allerdings ohne Gegenfarbendarstellung auskommt, stammt von LAND /LM1971/. Dessen Retinex-Theorie¹ zur Erklärung der von ihm selbst

sehr gut veranschaulichten Farbkonstanz-Phänomene drückt sich in einem zweiteiligen Algorithmus aus: Im ersten Teil wird zur Beseitigung von inhomogener Beleuchtung eine Trendeliminierung mit einem differentiellen Verfahren ausgeführt, womit psychophysisch ermittelte lokale Kontrastverstärkungsphänomene (sog. simultaner Farbkontrast) modelliert werden. Im zweiten Teil (Rekonstruktion der Farben) erfolgt eine bildbezogene Normierung aller Farbauszüge derart, daß Effekte aus wechselnder (abweichender) Beleuchtung ausbalanciert werden können. Das erfolgt einfach mittels Division durch den jeweiligen arithmetischen Mittelwert des Farbauszuges:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix}_{\text{EBU}_{\text{UR}}} = \begin{pmatrix} \frac{1}{R_s} & 0 & 0 \\ 0 & \frac{1}{G_s} & 0 \\ 0 & 0 & \frac{1}{B_s} \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}_{\text{EBU}} \quad (\text{A.1})$$

Praktischerweise schließt daran eine für alle drei Farbauszüge gleichartige Reskalierung an, so daß bspw. der durch die drei Mittelwerte gebildete Mittelwertvektor genau auf den Mittelpunkt des RGB-Farbraumes im benutzten Wertebereich fällt. D’ZMURA und IVERSON kommentieren die Methodik referenzfreier Modelle *insgesamt* einleitend in /DI1997,DI1998/ allerdings folgendermaßen:

“One problem with this sort of color constancy schemes is that if it works, one does not know why. If it does not work, one does not know why not.”

Wenngleich diese Aussage als überspitzt angesehen werden muß, ist es tatsächlich nicht unwahrscheinlich, daß das Verfahren fehlschlägt. MALLOT bemerkt in /Mal1998/ zu diesem Problem zusammenfassend:

„Das Retinex-Verfahren arbeitet befriedigend, wenn kleinere Schwankungen der Beleuchtung aus dem Bild eliminiert werden sollen. ... Dem liegt die Annahme zugrunde, daß die Welt im Mittel etwa grau ist (‘gray-world assumption’). Ist diese Annahme nicht erfüllt, z. B. bei Betrachtung einer Wiese, so macht das Retinex-Verfahren die erwarteten Fehler.“

Dennoch besitzen referenzfreie Modelle eine hohe Attraktivität wegen ihres Prinzips an sich, sowie wegen ihrer einfachen Umsetzbarkeit.

¹Retinex: von LAND /LM1971/ zusammengesetztes Kunstwort aus Retina und Cortex zur Veranschaulichung der übergreifenden Bezüge seines Farbkonstanzmodells

A.2 Elementarfarbraum-Verfahren

Ein weiteres referenzfreies Modell wird von POMIERSKI /Pom1996, PG1996a, PG1996b/ vorgestellt, für das dieser in seinen Arbeiten eindrucksvolle Ergebnisse zeigt. Für den von ihm aufgestellten Elementarfarbraum *RG/BY/WS* gibt er einen Algorithmus an, mittels dessen er eine Farbumstimmung ausführen kann, die er als *empfindungsgemäße Farbumstimmung* bezeichnet. Dabei handelt es sich um ein zweistufiges lineares Verfahren, bei dem zunächst, wie von LAND auch, der Mittelwertvektor aller Tristimulus-Vektoren des Bildes berechnet wird. Dessen euklidische Distanz zur Unbuntachse wird nun von POMIERSKI als Regelabweichung der ersten Stufe des Umstimmungsprozesses verwendet und diese mittels einer Scherung orthogonal zur Unbuntachse vollständig kompensiert. Im zweiten Schritt erfolgt eine Verstärkung in Richtung der Unbuntachse (Intensität), so daß der Dynamikumfang des Farbraumes in Richtung dieser Achse vollständig ausgeschöpft wird. Konkret wählt POMIERSKI eine Art der Verstärkung, die gleichzeitig zu einer Kontrastverstärkung führt, indem er eine nichtlineare Verstärkungskennlinie verwendet, die er aus zwei linearen Teilstücken zusammensetzt. Das erste Teilstück bis zu einer variablen Schwelle (Unbunt-Offset) hat den Anstieg 1, im zweiten Teilstück liegt der Anstieg entsprechend darüber. Dabei entspricht die variable Schwelle der Intensität des intensitätskleinsten Pixels im Bild. Damit wird auch für den zweiten Teil von POMIERSKIs Verfahren eine referenzfreie Korrektur vorgenommen, die davon ausgeht, daß im Bild die „schwärzesten“ Pixel als implizite Referenz für eine Kontrastverstärkung dienen können.

Die folgende formelmäßige Darstellung des Verfahrens erfolgt in homogenen Koordinaten und schließt beide Teilschritte in einer Hauptformel ein. Sofern die Multiplikation vierreihiger quadratischer Matrizen direkt auf spezieller Hardware oder mit speziell optimierter Software ausgeführt werden kann, kann die Implementierung unmittelbar in der angegebenen Weise erfolgen.

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix}_{\text{EBU}_{\text{UP}}} = \mathbf{T}^{-1} \cdot \mathbf{O}^{-1} \cdot \mathbf{K} \cdot \mathbf{O} \cdot \mathbf{T} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}_{\text{EBU}} \quad (\text{A.2})$$

Die darin verwendeten Transformationsmatrizen haben folgende Bedeutung und konkrete Belegung:

Transformation in den Elementarfarbraum²:

$$\mathbf{T} = \begin{pmatrix} 0,3380 & -0,9319 & 0,4972 & 0 \\ -0,5957 & -0,1622 & 0,7103 & 0 \\ 1,1011 & 2,0505 & 1,2238 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.3})$$

Subtraktion des Unbunt-Offsets:

$$\mathbf{O} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -WS_{\min} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.4})$$

Scherung im Elementarfarbraum:

$$\mathbf{K} = \begin{pmatrix} 1 & 0 & -\frac{RG_s}{WS_s} & 0 \\ 0 & 1 & -\frac{BY_s}{WS_s} & 0 \\ 0 & 0 & V & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.5})$$

Addition des Unbunt-Offsets:

$$\mathbf{O}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & WS_{\min} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.6})$$

Rücktransformation aus dem Elementarfarbraum²:

$$\mathbf{T}^{-1} = \begin{pmatrix} 0,6657 & -0,8688 & 0,2338 & 0 \\ -0,6078 & 0,0538 & 0,2157 & 0 \\ 0,4195 & 0,6915 & 0,2454 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.7})$$

Dabei werden die zuvor in den Gleichungen A.3 bis A.7 verwendeten Variablen wie folgt bestimmt:

$$WS_{\min} = \frac{3}{2} (R_{\min} + G_{\min} + B_{\min}) \quad (\text{A.8})$$

$$\text{mit } R_{\min} + G_{\min} + B_{\min} = \min_{(R,G,B)^\top} \quad (\text{A.9})$$

$$WS_{\max} = \frac{3}{2} (R_{\max} + G_{\max} + B_{\max}) \quad (\text{A.10})$$

$$\text{mit } R_{\max} + G_{\max} + B_{\max} = \max_{(R,G,B)^\top} \quad (\text{A.11})$$

Damit sind intensitätskleinster und intensitätsgrößter Unbuntwert bekannt. Der außerdem

²Diese Matrizen weichen von denen von POMIERSKI in /Pom1996/ angegebenen leicht ab, da diese dort für den RGB_{FCC} angegeben sind, hier jedoch linear transformiert für den RGB_{EBU} eingesetzt werden.

benötigte Schwerpunktvektor ergibt sich zu:

$$\begin{pmatrix} RG_s \\ BY_s \\ WS_s \end{pmatrix} = \mathbf{O} \cdot \mathbf{T} \cdot \begin{pmatrix} R_s \\ G_s \\ B_s \end{pmatrix}_{\text{EBU}} \quad (\text{A.12})$$

Dabei sind die R_s , G_s und B_s bzw. RG_s , BY_s und WS_s (wie schon in Gl. A.1 oder Gl. A.5) die arithmetischen Mittelwerte des jeweiligen Farbauszuges. Die Verstärkung V wird anhand des reduzierten verfügbaren Wertebereichsintervalls bestimmt:

$$V = \frac{\hat{WS} - WS_{\min}}{WS_{\max} - WS_{\min}} \quad (\text{A.13})$$

$$\text{mit} \quad \hat{WS} = \frac{3}{2}(\hat{R} + \hat{G} + \hat{B}) \quad (\text{A.14})$$

$$\text{und} \quad \hat{R} = \hat{G} = \hat{B} = 2^i - 1 \quad \text{sowie} \quad \check{R} = \check{G} = \check{B} = 0 \quad (\text{A.15--A.16})$$

Die Symbole $\hat{}$ und $\check{}$ bezeichnen darin die maximalen bzw. minimalen Wertebereichsgrenzen, die in der konkreten Darstellung durch den umfaßten Wertebereich und damit die Auflösung der verfügbaren Bilddigitalisierungseinheit (Framegrabber, hier mit $i = 8\text{bit}$) bestimmt sind.

POMIERSKIS Modell ist sowohl im Entwurf als auch in der Anwendung (Bestimmung der Transformationsmatrix) offensichtlich aufwendiger als LANDS, dessen Modell das denkbar einfachste ist. Es beruht in seiner Farbraumtopologie auf einer Modellierung von neurophysiologischen Erkenntnissen in der frühen visuellen Verarbeitung. Das beschriebene Umstimmungsverfahren entlang der Achsen des $RG/BY/WS$ -Farbraumes entspricht in seiner Form als Scherung einer unmittelbaren Nutzung der erzielten Umkodierung der sensorischen Repräsentation. Gegenüber anderen denkbaren möglichen Arten der Umstimmungstransformation ist diese innerhalb dieses Farbraumes die naheliegendste. Im Vergleich erscheint POMIERSKIS Verfahren, insbesondere in der Umsetzung direkt im technisch verfügbaren RGB_{EBU} -Raum, insgesamt plausibler, weil in ihm die Freiheitsgrade der Umstimmung durch Einsatz der Scherung eingeschränkt sind. In zwei Achsenrichtungen des Farbraumes erfolgen einfache Offsetkorrekturen.

Die o. g. Kritik von D'ZMURA und IVERSON kann nicht in vollem Maße mitgetragen werden. Die referenzfreien Modelle der Farbkonstanz sind genau dann immer gut anwendbar, wenn im Bild tatsächliche eine ausgewogene farbliche Zusammensetzung besteht. In diesem Falle gruppieren sich die Bildpixel als keulenförmig ausgebildete Punktwolke um die Unbuntachse, sofern eine weiße Beleuchtung vorherrscht. Bei deutlich davon abweichender Beleuchtung aber erscheint diese Punktwolke von dieser Lage ausgelenkt. Wenn also die Methode fehlschlägt, so kann das insbesondere daran liegen, daß die Ausgewogenheit der Bildfarben nicht gegeben ist. Das Fehlen einer Referenz muß gleichermaßen als Vor- wie

als Nachteil angesehen werden. Genauigkeitsdefizite gegenüber referenzbezogenen Modellen (*siehe* MALONEY und WANDELL /MW1986/) sind ohne Zweifel auch dann zu erwarten, wenn eine ausgewogene farbliche Zusammensetzung vorliegt. Die Umstimmung wird jedoch dadurch einfach und damit schnell und ist ad hoc einsetzbar. Für die vorliegende Arbeit sind referenzfreie Modelle daher besonders interessant.

Anhang B

Eigenfaces

Diese gesichtsspezifischen Filter stellen die eigenwertgrößten Eigenvektoren der mittleren Kovarianzmatrix aller Bilder an *Gesichtsprototypen* dar. Die Bestimmung entspricht der Ermittlung der Hauptachsen¹ der Verteilung der Gesichtsbildvektoren aller Gesichtsprototypen im Raum der Gesichtsprototypen. Die Anwendung der Eigenfaces als Filter kommt einer Projektion eines Bildausschnittes in einen (dimensionsreduzierten) Merkmalsraum gleich, der genau so angelegt ist, daß seine Achsen entlang der größten Varianzen der Datenverteilung der bekannten Gesichtsbilder (Prototypen) verlaufen. Die mit den Achsen dieses Merkmalsraumes aufeinanderfallenden Einheitsvektoren entsprechen den *Eigenfaces*.

Der Name *Eigenface* rührt daher, daß die Grauwertdarstellung der Eigenvektoren des Gesichtsdatenmaterials zu merkwürdigen „Geisterbildern“ führt, in denen gesichtsartige Strukturen zu erkennen sind. TURK und PENTLAND haben diese Eigenvektoren in /TP1991/ deshalb so bezeichnet. Wichtig ist, daß die Elemente der Eigenfacevektoren *dort* von einem mittleren Wert (Grauwert) abweichen, wo sich im Datensatz mehr oder weniger große Unterschiede zwischen den Prototypen befinden. Sie ergeben damit eine Art Karte davon, *wo* sich Unterschiede zwischen den Gesichtern befinden. Dabei ist die Stärke der Unterschiede unter der Orthogonalitätsbedingung und der Reihenfolge der Eigenvektoren zu bewerten.

Die Lokalisation eines Gesichts erfolgt dadurch, daß ein auf ein Gesicht hin zu prüfen-

¹Eng verwandt zu dieser Bestimmung der Hauptachsen ist die KARHUNEN-LOÈVE-Transformation (KLT) /Kar1984/, wobei deren Zielstellung vor allem die Informationsreduktion ist. In abgewandelter Weise geht es hier darum, genau *die* Information zu behalten, die als wesentlich für alle Gesichtsprototypen bestimmt wurde. Synonym wird die Bestimmung der eigenwertgrößten Eigenvektoren auch als Hauptkomponentenanalyse bzw. Principal Component Analysis (PCA) bezeichnet. Als Verfahren zur Bestimmung der Hauptachsen ist neben der geschlossenen Lösung (Bestimmung der Eigenwerte, Einsetzen der Eigenwerte, Bestimmung der Eigenvektoren durch Inversion der Kovarianzmatrix) alternativ ein lernendes rekursives Verfahren nach OJA /Oja1983/ und SANGER /San1989/ verwendbar (*siehe dazu* auch die Darstellungen von WENDT /Wen1993/ oder BRAUSE /Bra1995/).

der Bildausschnitt² in einen Merkmalsraum projiziert wird und anschließend sofort eine Rückprojektion in den Originalraum erfolgt. Zwischen originalem Bildausschnitt und dem Ergebnis der Rückprojektion wird ein Distanzmaß gebildet. Dieses Distanzmaß kann einem Schwellenkriterium unterworfen werden, um eine binäre Aussage zu treffen, ob im betreffenden Bildausschnitt ein Gesicht enthalten ist oder nicht.

In eigenen Arbeiten (BRAKENSIEK, BRAUMANN ET AL. /BBC⁺1998/) wurde getestet, auf die Rückprojektion in den Originalraum zu verzichten. Anstelle dessen wurde direkt im Merkmalsraum eine Modellierung der Gesichtsdatenverteilung mit einem überwacht trainierten *Growing Neural Gas Network* (GNG) nach FRITZKE /Fri1995/ und anhand dessen eine Klassifizierung vorgenommen. Die Methode wurde erfolgreich angewendet, allerdings wurde davon wieder abgegangen, da die Projektion in den Merkmalsraum nur anhand einer Auswahl an eigenwertgrößten Eigenfaces erfolgt. Dadurch sind Mehrdeutigkeiten auch mit im Originalraum stark gesichtsähnlichen Mustern nicht zu vermeiden³. Deshalb ist es nicht möglich, auf die Rückprojektion in den Originalraum zu verzichten und nur anhand eines Distanzmaßes innerhalb des Merkmalsraumes einen Bildausschnitt gesichtsmäßig zu bewerten. Es muß gesichert sein, daß die verwendeten Eigenvektoren insbesondere zur Beschreibung des betrachteten Bildausschnitts ausreichen. Die Projektion in den Merkmalsraum in die Nähe der Projektion der Prototypen genügt allein nicht.

Man wird leicht erkennen, daß das skizzierte Verfahren einen immensen Rechenaufwand erfordert. Dieser wird noch dadurch erhöht, daß eine gewisse Skalierungsinvarianz nur dadurch zu erzielen ist, wenn man dazu das Verfahren auf verschiedenen Skalierungsstufen des Bildes (Auflösungspyramide) anwendet. Der Berechnungsaufwand wird daher in Abhängigkeit der Anzahl der verwendeten Skalierungsstufen steigen, so daß ohne spezialisierte Hardware eine Verarbeitung von Bildern mit hoher Bildrate (ideal wäre die sog. Framerate⁴ von 25s^{-1}) weiter erschwert wird.

Bei den Bildausschnitten, die durch das Suchfenster zur Lokalisation erfaßt werden,

²Von Bildausschnitt wird im folgenden deshalb gesprochen, da das Verfahren nicht einfach lokal angewendet werden muß und die Suche zur Lokalisation ein Scannen eines Suchfensters über dem gesamten Bild erforderlich macht, wobei das Suchfenster jeweils einen Bildausschnitt erfaßt.

³Zur Gesichtserkennung wäre nach erfolgreicher Gesichtslokalisation eine Klassifikation der unterschiedlichen Gesichter direkt im Merkmalsraum adäquat. – Die o. g. guten Ergebnisse sind insbesondere darauf zurückzuführen, daß beim überwachten Zwei-Klassen-Training sowohl positive als auch negative Beispiele trainiert worden sind, was die Mehrdeutigkeiten deutlich einschränkt. Somit hängt das Verfahren jedoch stark von der Trainingsmenge an Negativbeispielen ab und ist damit nur eingeschränkt einsetzbar.

⁴Diese harte Anforderung steht realistischerweise für den Roboter MILVA weder für die Personenlokalisierung noch für weitere Verarbeitungsschritte, einfach weil die gegenwärtigen Hardware-Gegebenheiten das nicht erlauben.

handelt es sich um Ausschnitte der Größe $N \times N$ mit $N = 15$. Von eben dieser Größe sind die M verschiedenen Gesichtsprototypen. Zugunsten einer einfacheren Notation werden die Bildausschnitte im folgenden als Vektoren beschrieben, die damit eine Dimensionalität von N^2 besitzen.

Die Prototypen sind innerhalb des Originalraumes nicht wahllos verteilt. Genau diese Verteilung ist entscheidend für die Bildung der Eigenfaces. Sie ist das Kriterium, einen Merkmalsraum zu bilden, ohne daß die Art der Merkmale explizit vorgegeben wird. Gesucht sind die Hauptkomponenten der Verteilung der Bildausschnittvektoren im Originalraum bzw. die Eigenvektoren der Kovarianzmatrix der Menge aller Prototypen.

Der eigenwertgrößte Eigenvektor ist derart gerichtet, daß entlang dessen Richtung die Verteilung des Datensatzes (Prototypen) die größte Streuung hat. Der eigenwertgrößtmäßig folgende Eigenvektor ist so gerichtet, daß er unter der Bedingung, orthogonal zum eigenwertgrößten zu sein, in Richtung der zweitgrößten Varianz der Datenverteilung zeigt, usw. Jeder Eigenvektor ist zu jedem anderen Eigenvektor orthogonal, und alle Eigenvektoren werden gleichzeitig normiert bestimmt, so daß ein *Orthonormalsystem* durch die Menge aller Eigenvektoren aufgespannt wird.

B.1 Eigenfaces zur Gesichtsdetektion

Die Eigenfaces sollen die Basis zur Beschreibung von Gesichtsbildern bilden. SIROVICH und KIRBY /SK1987/ bewerteten ein Ensemble von $M = 115$ Prototypen (männliche Gesichter, kaukasischer Typ, 128×128), die sehr sorgfältig digitalisiert wurden. Sie stellten fest, daß bei Verwendung der $M' = 40$ eigenwertgrößten Eigenfaces ein im Datensatz enthaltenes Bild sehr gut rekonstruiert wird, so daß ein euklidischer Abstand von 2% bezogen auf das jeweilige Originalbild eingehalten werden konnte. Eigenfaces können somit als adäquat zur Beschreibung von Gesichtsbildern angenommen werden. Die weitergehende Fragestellung lautet nun, *inwieweit* man Eigenfaces als Mittel zur Gesichtslokalisation nutzen kann. Dabei sollte ein kleineres M' ausreichen, da es hier eben *nicht* um die Rekonstruktion (Kompression) von Gesichtsbildern, sondern um eine Mustererkennungsproblem geht, für das ein Kompromiß zwischen Aufwand, erzielbarer Genauigkeit und Leistungsfähigkeit zu finden ist.

Durch die Eigenfaces wird ein M' -dimensionaler Unterraum (Face-Space) des originalen N^2 -dimensionalen Datenraumes (Originalraum) aufgespannt. M' Eigenvektoren bilden den *Satz an Eigenfaces*. Klar ist, daß mit sinkendem M' der Berechnungsaufwand ebenfalls sinken wird. Für die Lokalisationsproblematik wurde hier ein Satz von $M' = 5$ Eigenfaces benutzt⁵.

B.1.1 Projektion in den Face-Space

Ein betrachteter Bildausschnitt $\underline{\gamma}$ (aus dem $\mathbb{R}^{N^2}$) wird in seine k -te Eigenface-Komponente transformiert durch folgende einfache Operation:

$$\omega_k = \underline{\mathbf{u}}_k^\top (\underline{\gamma} - \underline{\psi}) \quad (\text{B.1})$$

Die ω_k sind die Elemente (Fitwerte) eines Vektors $\underline{\omega}$ (Fitwertvektor), der entsprechend der Anzahl der Eigenfaces aus insgesamt M' Elementen besteht. $\underline{\psi}$ ist der aus den Prototypen $\underline{\gamma}_i$ gebildete Mittelwertvektor (*siehe* Gl. B.24). Die Projektion in den Face-Space entspricht also der Bestimmung des Fitwertvektors $\underline{\omega}^\top = (\omega_1, \omega_2, \dots, \omega_{M'})$ für einen betrachteten Bildausschnitt $\underline{\gamma}$:

$$\underline{\omega} = \underline{\mathbf{U}}'^\top (\underline{\gamma} - \underline{\psi}) \quad (\text{B.2})$$

Die Matrix $\underline{\mathbf{U}}'^\top$ enthält die Eigenfaces und ist aus dem $\mathbb{R}^{M' \times N^2}$.

Eine Rekonstruktion eines Bildausschnittes aus den M' Fitwerten ω_k erlaubt die Visualisierung der Projektion in den Face-Space im Originalraum:

$$\underline{\phi}_f = \underline{\mathbf{U}}' \underline{\omega} \quad (\text{B.3})$$

Die Rekonstruktion *kann* nicht den originalen Bildausschnitt wiederherstellen, sondern basiert nur auf Linearkombinationen der M' Eigenfaces. Wie später noch gezeigt werden wird, ist dieser Gesichtspunkt wesentlich für das Verfahren zur Personenlokalisierung. Der Fitwertvektor $\underline{\omega}$ selbst bildet den Ausgangspunkt für einen Mustererkennungsalgorithmus, dessen Ziel die Detektion und damit die Lokalisation von Gesichtern ist.

Eine mögliche Vorgehensweise zur Detektion könnte die Bestimmung eines Abstandsmaßes im Face-Space selbst sein. Dazu bietet sich zunächst ein euklidisches Abstandsmaß an. Bezeichnet als ζ^* zwischen einem Fitwertvektor $\underline{\omega}$ eines Bildausschnittes und dem typischen Fitwertvektor $\underline{\omega}^*$ aller Gesichtsprototypen lautet es:

$$\zeta^* = \|\underline{\omega} - \underline{\omega}^*\|_2 \quad (\text{siehe Fußnote}^6) \quad (\text{B.4})$$

$$= \sqrt{\sum_{i=1}^{M'} (\omega_i - \omega_i^*)^2} \quad (\text{B.5})$$

Die $\underline{\omega}^*$ werden dann derart gebildet, daß man alle F Fitwertvektoren aller Prototypen mittelt:

$$\underline{\omega}^* = \frac{1}{F_j} \sum_{k=1}^N \left(\underline{\mathbf{U}}'^\top (\underline{\gamma}_k - \underline{\psi}) \right) \quad (\text{B.6})$$

⁵TURK und PENTLAND verwendeten für die Untersuchungen in /TP1991/ $M' = 7$ Eigenfaces.

⁶Die tiefgestellte 2 symbolisiert, wie auch Gl. B.5 leicht klarmacht, die *quadratische* Norm.

Es wurde bereits darauf hingewiesen, daß dieses Maß jedoch allein nicht als Kriterium zur Gesichtsdetektion ausreicht. Der Grund ist, daß die Projektion in den niedrigdimensionalen Face-Space *allein* keine Trennung von Gesicht/Nichtgesicht (Hintergrund) erlauben *kann*. Im hochdimensionalen Originalraum sind unendlich viele Bildausschnitte denkbar, die mittels Projektion über die $\underline{\mathbf{U}}'^\top$ nur deshalb einem der Prototyp-Fitwertvektoren sehr nahe liegen, weil im Bildausschnitt *auch* Komponenten enthalten sind, die ebenfalls bei einem Gesicht auftreten.

Deshalb ist ein zweites Kriterium nötig, die Bestimmung des euklidischen Abstands ζ zwischen $\underline{\phi}$ und $\underline{\phi}_f$:

$$\zeta = \|\underline{\phi} - \underline{\phi}_f\|_2 \quad (\text{B.7})$$

$$= \sqrt{(\underline{\phi} - \underline{\phi}_f)^\top (\underline{\phi} - \underline{\phi}_f)} \quad (\text{B.8})$$

B.1.2 Detektion von Gesichtern mit Eigenfaces

Die Idee hinter Gl. B.7 ist, den Abstand ζ des betrachteten Bildausschnitts zum Face-Space zu bestimmen. Zunächst ist die Projektion eines (zentrierten⁷) Bildausschnitts in den Face-Space und wieder zurück:

$$\underline{\phi}_f = \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi} \quad (\text{B.9})$$

Die Gl. B.3 bzw. B.9 in Gl. B.7 eingesetzt, kann dieser euklidische Abstand auch so geschrieben werden:

$$\begin{aligned} \zeta &= \|\underline{\phi} - \underline{\mathbf{U}}' \underline{\omega}\|_2 \\ &= \|\underline{\phi} - \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi}\|_2 \\ &= \|\left(\underline{\mathbf{E}} - \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top\right) \underline{\phi}\|_2 \end{aligned} \quad (\text{B.10})$$

Dafür wären ca. jeweils $(2M' + 2)N^2$ Multiplikationen und Additionen nötig⁸. Nachfolgend soll ein effektiverer Weg zur Bildung des ζ gezeigt werden, so daß nur noch ca. jeweils $(M' + 2)N^2$ Multiplikationen und Additionen zu berechnen sind. Dafür wird zur Vereinfachung der

⁷Zentriert bedeutet, daß von jedem zu verarbeitenden Bildausschnitt der Mittelwertvektor $\underline{\psi}$ abgezogen ist, der anhand der Prototypen gebildet wurde.

⁸ N^2 Add. (Differenz), N^2 Add. (Zentrierung), $N^2 \cdot M'$ Mult./Add. ($\underline{\mathbf{v}}_{\text{temp}} = \underline{\mathbf{U}}'^\top \underline{\phi}$), $M' \cdot N^2$ Mult./Add. ($\underline{\mathbf{U}}' \underline{\mathbf{v}}_{\text{temp}}$), N^2 Mult./Add. (Skalarprodukt für Norm). Damit sind insgesamt $(2M' + 3)N^2$ Additionen und $(2M' + 1)N^2$ Multiplikationen auszuführen. Das entspricht je Bildausschnitt ca. $(2M' + 2)N^2$ Additionen und Multiplikationen.

Notation das Quadrat von ζ betrachtet:

$$\zeta^2 = \|\underline{\phi} - \underline{\phi}_f\|_2^2 \quad (\text{B.11})$$

$$= (\underline{\phi} - \underline{\phi}_f)^\top (\underline{\phi} - \underline{\phi}_f) \quad (\text{B.12})$$

$$= \underline{\phi}^\top \underline{\phi} - \underline{\phi}^\top \underline{\phi}_f - \underline{\phi}_f^\top (\underline{\phi} - \underline{\phi}_f) \quad (\text{B.13})$$

Wird aber die Projektion auf die transformierten Hauptachsen gebildet, so daß der Differenzvektor $(\underline{\phi} - \underline{\phi}_f)$ immer senkrecht auf $\underline{\phi}_f$ steht, kann daher der letzte Term von Gl. B.13 wegfallen⁹.

Wenn aber Gl. B.14 gilt, dann gilt auch:

$$\underline{\phi}_f^\top \underline{\phi} = \underline{\phi}_f^\top \underline{\phi}_f \quad , \quad (\text{B.15})$$

so daß Gl. B.13 sich unter Berücksichtigung von $\underline{\phi}_f^\top \underline{\phi} = \underline{\phi}_f^\top \underline{\phi}_f$ weiter vereinfacht:

$$\zeta^2 = \underline{\phi}^\top \underline{\phi} - \underline{\phi}_f^\top \underline{\phi}_f \quad , \quad (\text{B.16})$$

was letztlich mit $\underline{\phi}^\top \underline{\phi}$ als Hypothenuse dem Satz von Pythagoras entspricht.

Verglichen mit Gl. B.12 ist das scheinbar kaum eine Vereinfachung. Wesentlich vereinfachend wirkt sich aber aus, daß $\underline{\phi}_f^\top \underline{\phi}_f$ *nicht* über Gl. B.9 bestimmt werden muß, sondern

⁹Das kann folgendermaßen gezeigt werden:

$$\begin{aligned} \underline{\phi}_f^\top (\underline{\phi} - \underline{\phi}_f) &= (\underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi})^\top (\underline{\phi} - \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi}) \\ &= \underline{\phi}^\top (\underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top)^\top (\underline{\phi} - \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi}) \\ &= \underline{\phi}^\top \underline{\mathbf{U}}'^{\top\top} \underline{\mathbf{U}}'^\top (\underline{\phi} - \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi}) \\ &= \underline{\phi}^\top \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top (\underline{\phi} - \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi}) \\ &= \underline{\phi}^\top \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi} - \underbrace{\underline{\phi}^\top \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top}_{\underline{\mathbf{E}}} \underline{\phi} \\ &= \underline{\phi}^\top \underline{\mathbf{U}}' \underline{\mathbf{U}}'^\top \underline{\phi} (\underline{\mathbf{E}} - \underline{\mathbf{E}}) \\ &= 0 \end{aligned} \quad (\text{B.14})$$

Das Matrixprodukt $\underline{\mathbf{U}}'^\top \underline{\mathbf{U}}'$ ergibt die Einheitsmatrix $\underline{\mathbf{E}}$ (aus dem $\mathbb{R}^{M' \times M'}$), da $\underline{\mathbf{U}}'$ aus orthonormalen Spaltenvektoren besteht.

einfach als Summe der Quadrate der Fitwerte bestimmt werden kann:

$$\begin{aligned}\underline{\phi}_f^\top \underline{\phi}_f &= (\underline{\mathbf{U}}' \underline{\omega})^\top \underline{\mathbf{U}}' \underline{\omega} \\ &= \underline{\omega}^\top \underbrace{\underline{\mathbf{U}}'^\top \underline{\mathbf{U}}'}_{\underline{\mathbf{E}}} \underline{\omega} \\ &= \underline{\omega}^\top \underline{\omega}\end{aligned}\tag{B.17}$$

$$= \sum_{i=1}^{M'} \omega_i^2\tag{B.18}$$

Damit wird der Abstand ζ^2 durch folgenden Ausdruck bestimmt:

$$\zeta^2 = \underline{\phi}^\top \underline{\phi} - \underline{\omega}^\top \underline{\omega}\tag{B.19}$$

Dieses Maß wird über jedem Element eines Skalarbildes $\underline{\mathbf{I}}$ bestimmt, wobei $\underline{\phi}(x, y)$ einen Bildausschnitt der Größe $N^2 \times N^2$ symbolisiert, der über dem Bildelement $i(x, y)$ zentriert ist, so daß ζ^2 von den Koordinaten (x, y) abhängt:

$$\zeta^2(x, y) = \underline{\phi}^\top(x, y) \underline{\phi}(x, y) - \underline{\omega}^\top(x, y) \underline{\omega}(x, y)\tag{B.20}$$

$$= \underline{\phi}^\top(x, y) \underline{\phi}(x, y) - \left(\underline{\mathbf{U}}'^\top \underline{\phi}(x, y) \right)^\top \underline{\mathbf{U}}' \underline{\phi}(x, y)\tag{B.21}$$

Bei der Bestimmung von $\underline{\phi}^\top(x, y) \underline{\phi}(x, y)$ ist zu beachten, daß hier jeweils der in den Koordinatenursprung des Originalraumes zentrierte Bildausschnitt gemeint ist, so daß mit $\underline{\psi}$ aus Gl. B.24 gilt:

$$\underline{\phi}(x, y) = \underline{\gamma}(x, y) - \underline{\psi}\tag{B.22}$$

Als Endergebnis für die Abstandsbestimmung liegt nun vor:

$$\zeta(x, y) = \sqrt{\underline{\phi}^\top(x, y) \underline{\phi}(x, y) - \left(\underline{\mathbf{U}}'^\top \underline{\phi}(x, y) \right)^\top \underline{\mathbf{U}}' \underline{\phi}(x, y)}\tag{B.23}$$

Die dafür nötige Zahl von Operationen beträgt ca. jeweils $(M' + 2)N^2$ Multiplikationen und Additionen¹⁰. Damit fallen ca. $M' \cdot N^2$ Multiplikationen und Additionen weniger an als ohne die dargestellte Effektivierung.

Mit diesem Abstand $\zeta(x, y)$, der einer Schwellwertfunktion mit der bereits genannten Schwelle ϑ_ζ unterworfen werden kann, liegt nun die Grundlage eines gesichtsspezifischen Auffälligkeitsmaßes vor. Mittels $\zeta(x, y)$ ist das zur diskreten Bildmatrix $\underline{\mathbf{I}}$ korrespondierende Feld $\underline{\mathbf{Z}}$ einer Eigenface-basierten Auffälligkeit bestimmt.

¹⁰ N^2 Add. für Bildung von $\underline{\phi}$, N^2 Mult./Add. für Skalarprodukt der $\underline{\phi}$, $N^2 \cdot M'$ Mult./Add. für $\underline{\mathbf{v}}_{\text{temp}} = \underline{\mathbf{U}}' \underline{\phi}$, M' Mult./Add. für Skalarprodukt der $\underline{\mathbf{v}}_{\text{temp}}$. Damit sind insgesamt ca. $(M' + 1)N^2 + M'$ Multiplikationen und ca. $(M' + 2)N^2 + M'$ Additionen erforderlich. Das entspricht insgesamt ca. je $(M' + 2)N^2$ Multiplikationen und Additionen.

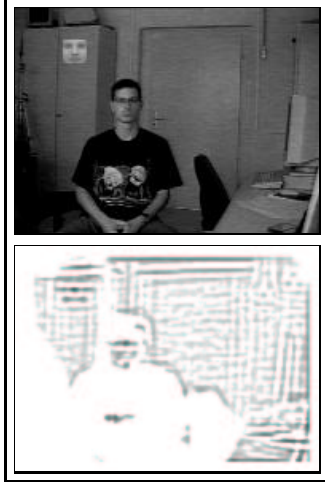


Abbildung B.1: Einsatz der Eigenface-Methode, Ergebnis bei Filterung mit zwei Eigenfaces. Kleine Distanzen sind dunkel, große hell dargestellt. Zur Nachbearbeitung dieses Ergebnisses siehe Abschnitt 4.2.1.3 Erhöhung der Selektivität – Beseitigung von Störungen.

B.2 Bestimmung der Eigenfaces

Sei $\{\underline{\gamma}_1, \underline{\gamma}_2, \dots, \underline{\gamma}_M\}$ die Menge aller Prototypen. Dann ist der Mittelwertvektor (geschätzter Erwartungswertvektor) aus allen Prototypen:

$$E(\underline{\gamma}) = \underline{\psi} = \frac{1}{M} \sum_{n=1}^M \underline{\gamma}_n \quad (\text{B.24})$$

Subtrahiert man diesen Vektor $\underline{\psi}$ von den $\underline{\gamma}_i$, erhält man die Menge der zentrierten Prototypen bzw. man verschiebt den Mittelwert der Datenverteilung der Prototypen in den Koordinatenursprung:

$$\underline{\phi}_i = \underline{\gamma}_i - \underline{\psi} \quad \forall i = 1, 2 \dots M \quad (\text{B.25})$$

Damit ist der Mittelwertvektor aus allen $\underline{\phi}_i$ der Nullvektor.

Mittels Hauptkomponentenanalyse sollen M orthonormale Vektoren (zueinander senkrecht stehende Einheitsvektoren) $\underline{u}_n$ gebildet werden, die die Datenverteilung der $\underline{\phi}_i$ optimal beschreiben. Für diese Vektoren gilt damit:

$$\underline{u}_l^\top \underline{u}_k = \delta_{lk} = \begin{cases} 1 & \text{wenn } l = k \\ 0 & \text{sonst} \end{cases} \quad (\text{B.26})$$

δ_{lk} bezeichnet darin das bekannte Kronecker-Delta.

Dafür muß man folgendes lineare Gleichungssystem (Eigenwertaufgabe) lösen:

$$\underline{\mathbf{C}} \underline{u} = \lambda \underline{u} \quad , \quad (\text{B.27})$$

d. h. die N^2 Eigenwerte $\lambda_k, k = 1 \dots N^2$ und die dazugehörigen Eigenvektoren $\underline{u}_k$ bestimmen. Eine Lösung existiert genau dann, wenn der Rang der Matrix $(\underline{\mathbf{C}} - \underline{\mathbf{E}}\lambda)$ kleiner als die Anzahl der Zeilen/Spalten dieser Matrix ist, d. h. wenn die Determinante dieser Matrix gleich Null

ist. Die Matrix $\underline{\mathbf{C}}$ ist die Kovarianzmatrix der Prototypen, die folgendermaßen bestimmt wird:

$$\begin{aligned}\underline{\mathbf{C}} &= \frac{1}{M} \sum_{n=1}^M \left(\underline{\gamma}_n - \underline{\psi} \right) \left(\underline{\gamma}_n - \underline{\psi} \right)^\top \quad (\text{siehe Fußnote}^{11}) \\ &= \frac{1}{M} \sum_{n=1}^M \underline{\phi}_n \underline{\phi}_n^\top\end{aligned}\tag{B.28}$$

Praktisch gesehen wird also der Durchschnitt der dyadischen Produkte der Prototypen gebildet.

Zur Lösung der Eigenwertaufgabe muß $\underline{\mathbf{C}}$ diagonalisiert (durch Hauptachsentransformation in die metrische Normalform $\underline{\mathbf{C}}_u$ überführt) werden, so daß nur noch die Hauptdiagonale (mit den Eigenwerten λ von $\underline{\mathbf{C}}$) besetzt ist. Dadurch ändert sich die Kovarianzmatrix:

$$\begin{aligned}\underline{\mathbf{C}}_u &= \frac{1}{M} \sum_{n=1}^M \underline{\mathbf{U}}^\top \underline{\phi}_n \left(\underline{\mathbf{U}}^\top \underline{\phi}_n \right)^\top \\ &= \frac{1}{M} \sum_{n=1}^M \underline{\mathbf{U}}^\top \underline{\phi}_n \underline{\phi}_n^\top \underline{\mathbf{U}}\end{aligned}\tag{B.29}$$

Transformiert wird mit einer Matrix $\underline{\mathbf{U}}$ vom Typ $N^2 \times N^2$, die sich aus den Eigenvektoren $\underline{\mathbf{u}}$ zusammensetzt. Die Matrix $\underline{\mathbf{U}}$, mit der die Hauptachsentransformation der Prototypen ausgeführt wird, kann vor bzw. hinter die Summe gezogen werden:

$$\underline{\mathbf{C}}_u = \underline{\mathbf{U}}^\top \left(\frac{1}{M} \sum_{n=1}^M \underline{\phi}_n \underline{\phi}_n^\top \right) \underline{\mathbf{U}}\tag{B.30}$$

Der Klammerausdruck entspricht Gl. B.28, weswegen er durch $\underline{\mathbf{C}}$ ersetzt werden kann:

$$\underline{\mathbf{C}}_u = \underline{\mathbf{U}}^\top \underline{\mathbf{C}} \underline{\mathbf{U}}\tag{B.31}$$

Das bedeutet, daß die Kovarianzmatrix der transformierten Verteilung durch lineare Operationen direkt aus der ursprünglichen Kovarianzmatrix gebildet werden kann. $\underline{\mathbf{C}}_u$ ist zu $\underline{\mathbf{C}}$ ähnlich. $\underline{\mathbf{U}}$ selbst kann $\underline{\mathbf{u}}_i$ -weise bestimmt werden über den Ansatz

$$(\underline{\mathbf{C}} - \lambda_i \underline{\mathbf{E}}) \underline{\mathbf{u}}_i = 0\tag{B.32}$$

Die $\underline{\mathbf{u}}_i$ sind dabei auf Einheitslänge zu normieren.

Die Hauptachsentransformation entspricht der Bestimmung *aller* Eigenvektoren von $\underline{\mathbf{C}}$. Daher ist $\underline{\mathbf{U}}^\top$ eine Matrix aus dem $\mathbb{R}^{N^2 \times N^2}$. Um die Dimensionalität zu reduzieren (*siehe auch* Karhunen-Loève-Transformation /Kar1984, Loè1977/, PCA) und nur mittels der

¹¹Der Koeffizient $\frac{1}{M}$ kann praktisch auch weggelassen werden, er beeinflußt die Größe der Eigenwerte, nicht aber deren Größenrelationen zueinander. Auf die Richtung der normierten $\underline{\mathbf{u}}$ hat er keinen Einfluß.

Hauptkomponenten eine Projektion auf einen geringerdimensionalen Raum auszuführen, soll $\underline{\mathbf{U}}^\top$ aus dem $\mathbb{R}^{M \times N^2}$ kommen. $\underline{\mathbf{U}}^{\top\top} = \underline{\mathbf{U}}$ besteht dann aus M Spaltenvektoren der Dimension N^2 , wobei N^2 die Dimension der Prototypen und somit der (vektorierten) Bildausschnitte ist, während M die verbleibende Dimension nach der Projektion ist. Somit stellt die Transformation mit $\underline{\mathbf{U}}^\top$ aus $\mathbb{R}^{M \times N^2}$ eine Projektion aus dem $\mathbb{R}^{N^2}$ in den $\mathbb{R}^M$ dar.

B.3 Effektivierte Bestimmung der Eigenfaces

Die Kovarianzmatrix $\underline{\mathbf{C}}$ ist aus dem $\mathbb{R}^{N^2 \times N^2}$. Sofern die Größe $N^2 \times N^2$ der Bildausschnitte relativ kleine Maße deutlich überschreitet, erweist sich nicht nur die Berechnung der Kovarianzmatrix als aufwendig, sondern es sind insbesondere viele Eigenwerte zu berechnen, anhand derer dann die Eigenvektoren zu bestimmen sind.

Hier stellt sich die Frage, ob es dafür einen rechentechnisch überhaupt gangbaren oder für die anderen Fälle zumindest wesentlich effektivierten Weg gibt. Dieser Weg existiert, wenn folgende Voraussetzung erfüllt ist: *Die Anzahl der vorhandenen Prototypen aus dem Originalraum ist kleiner als die Dimension dieses Raumes, d. h. $M < N^2$.* Ist diese Bedingung erfüllt, gibt es nur M anstelle N^2 signifikante Eigenvektoren der Dimension N^2 . (Den restlichen Eigenvektoren wird ein Eigenwert von 0 zugeordnet.)

Konkret kann man das Problem für die N^2 -dimensionalen Eigenvektoren lösen, indem man zunächst die M Eigenwerte (und deren M -dimensionale Eigenvektoren) einer $M \times M$ -Matrix bestimmt. Die tatsächlich benötigten Eigenvektoren bestimmt man anschließend durch geeignete Linearkombinationen der zentrierten Prototypen $\underline{\phi}_n$.

Die Kovarianzmatrix $\underline{\mathbf{C}}$ ist äquivalent zu Gl. B.28 darstellbar als folgendes Produkt:

$$\begin{aligned} \underline{\mathbf{C}} &= \frac{1}{M} \sum_{n=1}^M \underline{\phi}_n \underline{\phi}_n^\top \\ &= \frac{1}{M} \left[\underline{\phi}_1, \underline{\phi}_2, \dots, \underline{\phi}_M \right] \left[\underline{\phi}_1, \underline{\phi}_2, \dots, \underline{\phi}_M \right]^\top \end{aligned} \quad (\text{B.33})$$

$$= \underline{\mathbf{A}} \underline{\mathbf{A}}^\top \quad (\text{B.34})$$

Die Matrix $\underline{\mathbf{A}}$ ergibt sich infolgedessen aus Spaltenvektoren der $\underline{\phi}_n$ zu:

$$\underline{\mathbf{A}} = \frac{1}{\sqrt{M}} \left[\underline{\phi}_1, \underline{\phi}_2, \dots, \underline{\phi}_M \right] \quad (\text{B.35})$$

und ist aus dem $\mathbb{R}^{N^2 \times M}$. Betrachtet wird nun die Matrix

$$\underline{\mathbf{L}} = \underline{\mathbf{A}}^\top \underline{\mathbf{A}} \quad . \quad (\text{B.36})$$

Sie ist aus dem $\mathbb{R}^{M \times M}$. Deren Eigenwertaufgabe lautet:

$$\underline{\mathbf{L}} \underline{\mathbf{v}} = \mu \underline{\mathbf{v}} \quad (\text{B.37})$$

Linksseitiges Multiplizieren mit $\underline{\mathbf{A}}$ ergibt:

$$\begin{aligned} \underline{\mathbf{A}} \underline{\mathbf{L}} \underline{\mathbf{v}} &= \mu \underline{\mathbf{A}} \underline{\mathbf{v}} \\ \underline{\mathbf{A}} \underline{\mathbf{A}}^\top \underline{\mathbf{A}} \underline{\mathbf{v}} &= \mu \underline{\mathbf{A}} \underline{\mathbf{v}} \\ \underline{\mathbf{C}} \underline{\mathbf{A}} \underline{\mathbf{v}} &= \mu \underline{\mathbf{A}} \underline{\mathbf{v}} \end{aligned} \quad (\text{B.38})$$

Vergleicht man nun mit der Eigenwertaufgabe aus Gl. B.27, wird deutlich, daß eine Teilmenge der Eigenvektoren $\underline{\mathbf{u}}$ durch Lineartransformation mit $\underline{\mathbf{A}}$ aus den Eigenvektoren $\underline{\mathbf{v}}$ gebildet werden kann:

$$\underline{\mathbf{u}} = \underline{\mathbf{A}} \underline{\mathbf{v}} \quad (\text{B.39})$$

Derart können nur M Eigenvektoren von $\underline{\mathbf{C}}$ gebildet werden. Es sind genau die durch die Anzahl M der Prototypen $\underline{\phi}_i$ signifikant bestimmbar! Das heißt auch, daß die Menge der M Eigenwerte μ genau die Teilmenge der M größten Eigenwerte λ bildet.

Aus den zentrierten Prototypen $\underline{\phi}_i$ wird die Matrix $\underline{\mathbf{L}}$ entsprechend der Schreibweise von Gl. B.33 so gebildet:

$$\begin{aligned} \underline{\mathbf{L}} &= \underline{\mathbf{A}}^\top \underline{\mathbf{A}} \\ &= \frac{1}{M} \left[\underline{\phi}_1 \ \underline{\phi}_2 \ \cdots \ \underline{\phi}_M \right]^\top \left[\underline{\phi}_1 \ \underline{\phi}_2 \ \cdots \ \underline{\phi}_M \right] \end{aligned} \quad (\text{B.40})$$

Nachdem alle Eigenwerte λ_i und deren korrespondierende Eigenvektoren $\underline{\mathbf{u}}_i$ bestimmt worden sind, ist es sinnvoll, diese Paare anhand der Größe der Eigenwerte so zu sortieren, daß gilt:

$$\lambda_{i+1} \geq \lambda_i \quad \forall i = 1, 2, \dots, M \quad (\text{B.41})$$

Damit ist gesichert, daß jede Auswahl aus indexkleinsten Eigenvektoren immer eine Teilmenge der eigenwertgrößten Eigenvektoren ergibt.

Anhang C

Implementierbarkeit der Eigenface-Methode auf Spezialrechentechnik

C.1 Problemstellung: Mehrfache Faltung

Die Anwendung der Gl. B.23 (*siehe* S. 125) aus Abschnitt B.1.2 *Detektion von Gesichtern mit Eigenfaces* erfordert ein mehrfaches Falten jedes Bildes mit einem Satz von M' Operanden (Eigenfaces). Das betrifft den Term $\underline{\mathbf{U}}'^T \underline{\phi}(x, y)$, der Gl. B.2 zur Bestimmung des Fitwertvektors $\underline{\omega}$ entspricht. In Operatorschreibweise lautet die 2D-Faltung zur Bestimmung des Feldes des k -ten Fitwertes $\underline{\Omega}_k$ aus dem Bild $\underline{\mathbf{I}}$, wobei aus Handhabbarkeitsgründen wieder auf eine Matrixdarstellung der Eigenfaces ($\underline{\mathbf{U}}_k$: Eigenfacematrix aus dem $\mathbb{R}^{N \times N}$ gebildet aus dem Eigen(face)vektor $\underline{u}_k$ aus dem $\mathbb{R}^{N^2}$) übergegangen wird:

$$\underline{\Omega}_k = \underline{\mathbf{U}}_k * \underline{\mathbf{I}} \quad (\text{C.1})$$

Als Grundoperationen treten nur Additionen und Multiplikationen auf. Der Umfang ist allerdings erheblich und auch von sehr leistungsfähiger Universalrechentechnik unter bestimmten zeitkritischen Randbedingungen gegenwärtig noch nicht verarbeitet werden kann.

Ein prinzipiell mögliches Vorgehen bei Faltungen wäre, jedes Bild (und einmal alle M' Operanden) mittels schneller Fouriertransformation (FFT) in den Ortsfrequenzraum zu transformieren, das transformierte Bild mit allen Operanden punktweise zu multiplizieren und die M' Ergebnisse durch inverse schnelle Fouriertransformation zurück in den Ortsraum¹ zu bringen. Allerdings ist es für diesen Weg Voraussetzung, daß die Anzahl der Bildpunkte durch möglichst wenige Primfaktoren zerlegbar ist, da die Rechenzeit zur Transformation sich ungefähr proportional zur Summe der Primfaktoren multipliziert mit der Anzahl der

¹Dabei wird vorausgesetzt, daß das Ergebnis der Faltung tatsächlich im Ortsraum benötigt wird.

Bildpunkte verhält. Es werden deshalb Bildgrößen bevorzugt, die sich ausschließlich als Zweierpotenzen darstellen lassen, da dadurch das beste Datenvolumen-Rechenzeit-Verhältnis zu bilden ist.

Da die in dieser Arbeit benutzten Auffälligkeitskomponenten allesamt *ohne* Formateinschränkung pyramidal eingesetzt werden, ist die genannte Problematik der starken Abhängigkeit der FFT-Rechenzeit von der Bildpunktanzahl *ein* Grund, die Berechnung nicht über den FFT-Weg zu vollziehen. Noch wichtiger allerdings erscheint, daß quasi der doppelte FFT-Aufwand anfällt, da alle Teilergebnisse einzeln in den Ortsraum rücktransformiert werden müßten, um darin das Distanzmaß nach Gl. B.23 bestimmen zu können.

Andererseits stand jedoch für eine Faltung im Ortsraum keine spezielle Rechentechnik, wie etwa ein *Genesis Framegrabber mit NOA*² der Firma Matrox oder die *Convolver/Arithmetic Logic Unit* der Firma Imaging Technologies, zur Verfügung.

C.2 CNAPS

Verfügbar dagegen war in Form eines CNAPS-Neurocomputers (in SIMD-Technologie) ein spezielles DSP-Multiprozessorsystem als PCI-Einsteckkarte mit 128 Prozessoren (CNAPS/PCI-DLX), auf dem ein entsprechend hoher Parallelisierungsgrad und damit eine wesentliche Erhöhung der Rechengeschwindigkeit erzielbar ist. Auf dieser Hardware sind alle *solche* Operationen gut implementierbar, bei denen große Datenmengen mit uniformen Operationen bearbeitet werden müssen. Das ist gerade bei Faltungen der Fall. Deshalb braucht für den Einsatz des in Abschnitt B.1.2 *Detektion von Gesichtern mit Eigenfaces* vorgestellten Algorithmus' *nicht* der Weg über den Ortsfrequenzbereich gegangen werden, sondern die Berechnung kann direkt im Ortsraum ausgeführt werden.

Als Neurocomputer wird die CNAPS deshalb bezeichnet, weil vollverschaltete künstliche neuronale Netze (KNN), wie MLP's, sich gut auf die Architektur abbilden lassen. Jeder PN berechnet dabei Teilergebnisse einer Lineartransformation, die bei der Abbildung von internen oder externen Stimulusvektoren auf eine darüberliegende Netzsicht über eine Gewichtungsmatrix ausgeführt wird. Prinzipiell ist es sogar möglich, innerhalb eines Algorithmus' sowohl KNNs als auch (z. B. für die Vorverarbeitung der KNN-Inputs) Bildverarbeitungs-algorithmen auf der CNAPS-Karte gleichzeitig laufen zu lassen. Der Vorteil ist der Wegfall von zusätzlichen Datentransfers.

Es lassen sich auch erfolgreich spärlich verschaltete neuronale Netzwerke auf CNAPS

²NOA bezeichnet eine Spezialhardware, die auf lokale Bildverarbeitungs-routinen (etwa Faltungen) zugeschnitten ist.

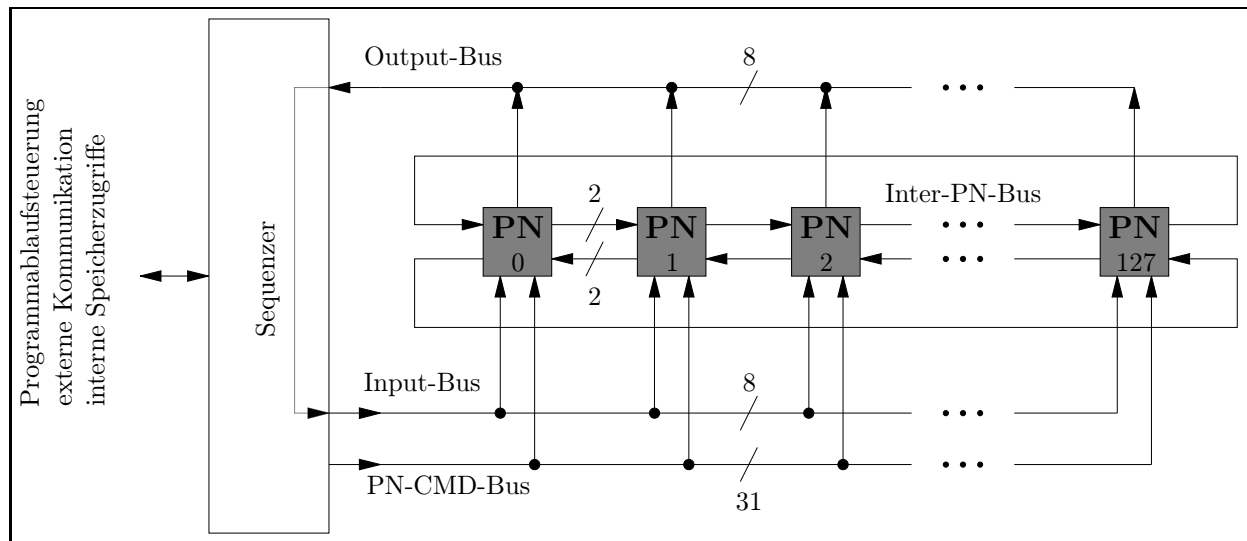


Abbildung C.1: CNAPS-Prinzipaufbau: Verschaltung des Arrays aus 128 Prozessoren (PN) verteilt auf 2 Chips. Der Sequenzer (Steereinheit) sendet einen Befehlsstrom (single instruction stream) im Broadcast-Verfahren an alle Prozessorknoten, die ihn synchron, aber jeweils auf ihren lokalen Daten (multiple data stream), abarbeiten. HAMMERSTROM erläutert in /Ham1995/ die Entwurfsentscheidungen der CNAPS-Hardware und schlägt Verbesserungen der Architektur vor, z. B. 16-Bit-Datenbusse, 16×16 -Bit-Multiplikation in einem Takt, vergrößerter lokaler Datenspeicher und höhere Taktfrequenzen als 20MHz.

implementieren, wobei die Algorithmen und Datenstrukturen darauf zugeschnitten werden müssen (siehe dazu Arbeiten von PASCHKE und MÖLLER /PM1997, PM1998/).

Jeder der CNAPS-Prozessoren innerhalb der Multiprozessorarchitektur stellt ein komplettes Rechenwerk für Festkomma-Arithmetik dar, was für die erforderliche schnelle Faltung gut nutzbar ist. Allerdings ist die Speichergröße pro Prozessor mit 4KByte recht eng bemessen, so daß dementsprechend eine feine Zerlegung der Bilddaten-Matrix erforderlich wird, was wiederum durch die Gesamtzahl der Prozessoren begrenzt wird. Die Leistungsfähigkeit dieses Rechners ist insbesondere darin begründet, daß in jedem Takt ein CNAPS-Prozessor eine Multiplikation (8- oder 16-Bit mit Festkomma) und eine Addition (32-Bit) *zugleich* ausführen kann. Für eine detaillierte Darstellung der CNAPS-Architektur sei auf einen Übersichtsartikel /Ham1995/ von HAMMERSTROM verwiesen, der selbst zu den Entwicklern des genutzten Chips³ CNAPS-1064 Ende der achtziger Jahre gehörte.

³Hervorzuheben sind einige Kenndaten des CNAPS-Prozessor-Chips: die Kantenlänge mißt ca. 26mm. Bei der verwendeten $0,8\mu$ -CMOS-Technologie können darauf mehr als 14 Millionen Transistoren implementiert werden. Damit handelt es sich bis heute um einen der größten jemals hergestellten Prozessorchips. Demgegenüber hatte ein populärer Prozessorchip aus ähnlicher Zeit, der für den 80486 DX (1989) der Firma Intel, erst ca. 1,2 Millionen Transistoren und war in 1μ -Technologie ausgeführt. Der Nachfolgeprozessor Pentium[®] (1993) wartete (zunächst ohne MMX) mit ca. 3,1 Millionen Transistoren in $0,8\mu$ -Technologie auf, was sich bei MMX-Einführung auf ca. 4,5 Transistoren bei mehr als halbiertem CMOS-Technologie steigerte. Der Pentium[®] Pro (1995) hat ca. 5,5 Millionen Transistoren. Der Chip des Prozessors Pentium[®] II

Allerdings besteht auf dem CNAPS-System generell die Forderung, die Bilddaten entsprechend der Prozessorzahl derart aufzuteilen (zu parzellieren), daß keine oder nur geringe Kommunikation zwischen den Prozessoren während der Abarbeitung anfällt, da der bidirektionale 2Bit-Bus, über den benachbarte Prozessorelemente miteinander verbunden sind, nur äußerst begrenzte Transferraten erlaubt. Das erfordert konkret, die Parzellierung mit entsprechenden „Überlapps“ vorzunehmen, wodurch Kommunikation zwischen den PN vermieden werden kann. An dieser Stelle liegt ein weiteres Problem der Architektur: die CNAPS-interne Beschickung der Prozessoren mit Bilddaten selbst ist auch zeitaufwendig, da sie über den 8Bit-Input-Bus in CNAPS-Taktfrequenz erfolgt. Sofern allerdings die Zerlegung extern (Host-PC) ausgeführt werden und die Daten bereits in prozessorweise zuordenbaren Blöcken geladen und auf analoge Weise zurückgeladen werden können, ist das Problem weniger kritisch. Allerdings bleibt die Zeit zum Datentransfer zur CNAPS und zurück im hier behandelten Fall verglichen mit der effektiven Rechenzeit hoch, da sie größenordnungsmäßig im selben Bereich liegt.

C.3 Implementation der Eigenface-Methode auf CNAPS

Von HEMPEL (*siehe* /Hem1998/) wurde eine Implementierung auf der genannten PCI-Einsteckkarte in CPL vorgenommen. Mit dieser Assemblersprache kann optimal an die CNAPS-Hardware angepaßter Programmcode erzeugt werden. Auf einem 768×572 Pixel⁴ großen Skalarbild wurde ohne Ladezeiten mit $M' = 4$ Faltungskernen (Eigenfaces) eine Taktzahl von ca. 8,5 Millionen benötigt, was einer Rechenzeit von etwas mehr als 0,4s entspricht.

Dabei muß das Bild allerdings aufgeteilt in 3 Teilbilder verarbeitet werden, wobei jedes Teilbild in 128 Stücken parzelliert wird, weil pro Prozessor nur 4KByte lokaler PN-Speicher zur Verfügung stehen.

Praktisch schwierig bleibt für die CNAPS das Ziehen der Quadratwurzel in Gl. B.23, die jedoch nicht essentiell für den Algorithmus erscheint. Das quadrierte Abstandsmaß $\zeta^2(x, y)$ ist ebenfalls ein gut nutzbares Kriterium, so daß auf die Wurzel verzichtet wurde.

Die Verwendung der CNAPS für die Faltungsaufgabe ist möglich, erfordert jedoch zum Teil deutliche Kompromisse. Als Nachteile treten dabei insbesondere der nötige Datenaufbereitungsaufwand und generell die geringe Flexibilität der Implementationen, insbesondere (1997) besitzt mit rund 7,5 Millionen Transistoren gut die Hälfte des CNAPS-Chips, momentan bereits in einer 0,25 μ -Technologie ausgeführt. Für den Pentium[®] III (1999) gab es noch einmal eine Erweiterung der Transistorzahl auf ca. 9,5 Millionen – aber die Zahl des CNAPS-Chips ist immer noch nicht erreicht!

⁴Das entspricht der maximalen Auflösung des verwendeten Framegrabbers ITI IC-PCI & AM-CLR. – Da normalerweise Breiten- zu Höhen-Verhältnisse von 4 : 3 üblich sind, ist unklar, warum die Bildhöhe nicht 576 Pixel beträgt.

auch im Hinblick auf den pyramidalen Einsatz des Algorithmus', in Erscheinung.

Anhang D

Nichtlineare Strukturanalyse – Theoretische Grundlagen

D.1 Lokale Nachbarschaft

Eine ideale lokale Orientierung sei dadurch gekennzeichnet, daß innerhalb einer *lokalen Nachbarschaft* eine bestimmte Richtung existiert, in der ein gerader Weg mit beliebigem Startpunkt in dieser Nachbarschaft verlaufen kann, *ohne* daß eine Grauwertänderung auftritt. Orthogonal dazu verläuft dementsprechend an jedem Punkt dieser Nachbarschaft die Richtung des maximalen Gradienten im Punkt $\underline{x}$. Die Richtung des maximalen Gradienten ist also für alle $\underline{x}$ mit $\underline{x}^\top = (x_1, x_2)$ in dieser lokalen Nachbarschaft gleich, nur der Richtungssinn kann wechseln. Diese Richtung entspricht der Richtung der lokalen Orientierung (siehe Abb. D.1). Sie wird mit dem Einheitsvektor $\underline{k}_0$ bezeichnet, der senkrecht zu den Linien konstanten Grauwertes steht.

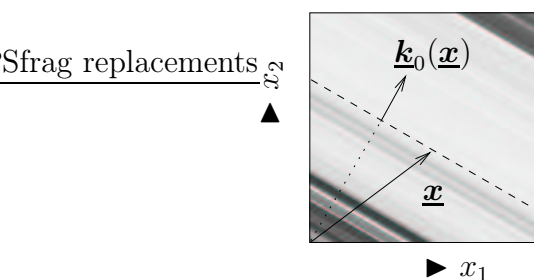


Abbildung D.1: Lokale Nachbarschaft mit ideal orientierter Struktur

Durch eine Projektion $\underline{x}^\top \underline{k}_0$ kann diese zweidimensionale Grauwertfunktion $g(\underline{x})$ als eindimensionale Funktion $g(\underline{x}^\top \underline{k}_0)$ entlang $\underline{k}_0$ dargestellt werden¹.

Transformiert man diese lokale Nachbarschaft in den Fourierraum, so erhält man eine

¹Daß der Gradientenvektor der Funktion im erläuterten Fall tatsächlich entlang $\underline{k}_0$ verläuft, zeigt folgende

spektrale Verteilung entlang einer geraden Linie durch den Koordinatenursprung:

$$\mathcal{F} \{g(\underline{x}^\top \underline{k}_0)\} = G(k\underline{k}_0) \delta(\underline{k} - \underline{k}_0(\underline{k}^\top \underline{k}_0)) \quad (\text{D.2})$$

Darin bezeichnet k die eindimensionale Koordinate in Richtung von $\underline{k}_0$. Das Argument der δ -Funktion² wird nur dann $\mathbf{0}$ (und damit der Funktionswert 1), wenn $\underline{k} \uparrow \downarrow \underline{k}_0$ oder $\underline{k} \uparrow \uparrow \underline{k}_0$ zueinander stehen.

Eine Frequenzanalyse entlang $\underline{k}_0$ wird im allgemeinen ein spektrales Gemisch ergeben, was jedoch, beginnend im Koordinatenursprung des Ortsfrequenzraumes, nur in einer speziellen Richtung auftritt, nämlich der von $\underline{k}_0$. Sofern die Nachbarschaft ein kontinuierliches Spektrum enthält, erhält man tatsächlich genau eine durch den Koordinatenursprung des Ortsfrequenzraums führende Linie aus Belegungspunkten.

Bisher wurde allerdings unausgesprochen angenommen, daß eine lokale Nachbarschaft unendlich ausgedehnt wäre, was jedoch logischerweise nicht der Fall ist. Vielmehr entspricht die Begrenzung einer Anwendung einer Fensterfunktion $w(\underline{x} - \underline{x}_0)$, so daß unter Anwendung des Verschiebungssatzes gilt:

$$\mathcal{F} \{w(\underline{x} - \underline{x}_0) \cdot g(\underline{x}^\top \underline{k}_0)\} = \exp(-i\underline{k}^\top \underline{x}_0) W(\underline{k}) * G(k\underline{k}_0) \delta(\underline{k} - \underline{k}_0(\underline{k}^\top \underline{k}_0)) \quad (\text{D.6})$$

Durch die Faltung wird, anschaulich betrachtet, die geradlinige Darstellung im Frequenzraum verschmiert. Die Breite dieser Verschmierung ist umgekehrt proportional zur Ausdehnung der lokalen Nachbarschaft. Davon kann man ableiten, daß die Genauigkeit der Orientierungsbestimmung proportional zum Verhältnis von Fenstergröße zur Wellenlänge der kleinsten Strukturen im Fenster ist.

kurze Herleitung:

$$\nabla g(\underline{x}^\top \underline{k}_0) = \begin{bmatrix} \frac{\partial g(\underline{x}^\top \underline{k}_0)}{\partial x_1} \\ \frac{\partial g(\underline{x}^\top \underline{k}_0)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} k_{0x1} g'(\underline{x}^\top \underline{k}_0) \\ k_{0x2} g'(\underline{x}^\top \underline{k}_0) \end{bmatrix} = \underline{k}_0 g'(\underline{x}^\top \underline{k}_0) \quad (\text{D.1})$$

Darin steht g' für die Ableitung von g nach $\underline{x}^\top \underline{k}_0$.

²Die Funktion $\delta(\underline{k})$ wird als Dirac-Delta bezeichnet und ist im $\mathbb{R}^2$ definiert als:

$$\delta(\underline{k}) = \prod_{i=1}^2 \delta(k_i) \quad (\text{D.3})$$

$$\text{mit} \quad \delta(k_i) = 0 \quad \text{falls} \quad k_i \neq 0 \quad (\text{D.4})$$

$$\text{und} \quad \int_{-\infty}^{\infty} \delta(k_i) dk_i = 1 \quad (\text{D.5})$$

D.2 Vektorielle Beschreibung von lokaler Orientierung

Eine Beschreibung einer lokalen Nachbarschaft durch eine skalare Größe ist ungeeignet, da es sich bei Orientierung um ein *zyklisches* Maß im Bereich $[0, \pi)$ handelt. Deshalb ist es sinnvoll, eine vektorielle Beschreibung vorzunehmen. Die Länge dieses Vektors soll als Konfidenzmaß interpretiert werden, welches ausdrückt, wie kontrastreich die orientierte Struktur in der lokalen Nachbarschaft ausgeprägt ist. Die Vektorrichtung wird durch den *doppelten*

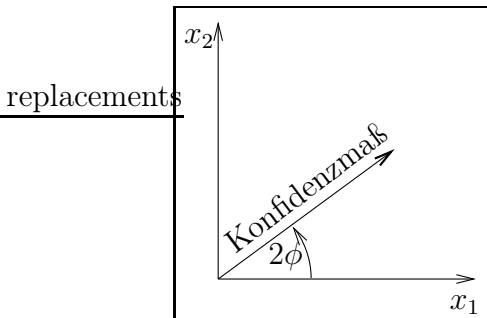


Abbildung D.2: Vektordarstellung lokaler Orientierung. Die Länge dieses Vektors entspricht einem Konfidenzmaß (siehe auch Gl. 5.7 und Abb. 5.9 in Abschnitt 5.4.2.1 Prinzip und Umsetzung der 2D-Tensor-Methode).

Orientierungswinkel bestimmt. Grund dafür ist, daß das Maß *Orientierung* gegenüber dem Maß *Richtung* nur den halben Wertebereich aufweist.

Betrachtet man nun im weiteren mehrere benachbarte lokale Nachbarschaften, und möchte dafür ein mittleres (effektives) Orientierungsmaß bilden, erweist sich die Vektorrepräsentation als geeignet: die Zusammenfassung zur Bildung eines effektiven Orientierungsmaßes könnte in einem ersten Ansatz einfach als Summation der Vektoren erfolgen. Es ist einsichtig, daß sich ein skalares Winkelmaß wegen der immanenten Diskontinuität nicht sinnvoll mitteln ließe.

Beispielsweise würde eine exakte Kohärenz in der Orientierung nur genau dann gegeben sein, wenn der Betrag des Summenvektors übereinstimmt mit der Summe der Beträge der einzelnen Vektoren. Paarweise orthogonale Orientierung schlägt sich in einer betragsmäßigen Subtraktion der Vektoren nieder. Miteinander unkorrelierte lokale Nachbarschaften stellen sich als Menge von Vektoren kurzer Länge dar, deren Summenvektor jedoch durch eine große Streuung der Richtungen klein bleibt.

D.3 Tensorielle Repräsentation zur Beschreibung lokaler Orientierung

Allerdings tritt bei der Methode der Vektoraddition das Problem auf, daß nicht unterschieden werden kann zwischen Nachbarschaften mit verschiedenen unkorrelierten lokalen Orientierungen und solchen ohne lokale Orientierung (konstanter Grauwert). Diese Unter-

scheidung kann nur vorgenommen werden, indem nicht der Summenvektor, sondern ein anderes Kriterium benutzt wird. Geeignet ist dafür, das quadrierte Skalarprodukt aus Gradient $\nabla g(\underline{\mathbf{x}})$ und Orientierungsvektor $\underline{\mathbf{k}}_0(\underline{\mathbf{x}})$ über der betrachteten lokalen Nachbarschaft (referenziert mit dem Ortsvektor $\underline{\mathbf{x}}$) zu integrieren und dieses Integral zu maximieren, was zur Bedingung hat, daß beide Vektoren parallel oder antiparallel gerichtet sind:

$$\begin{aligned} \overline{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} &= \arg \max_{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(\underline{\mathbf{x}} - \underline{\mathbf{x}}') (\nabla^\top g(\underline{\mathbf{x}}') \underline{\mathbf{k}}_0(\underline{\mathbf{x}}))^2 dx'_1 dx'_2 \\ &= \arg \max_{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(\underline{\mathbf{x}} - \underline{\mathbf{x}}') \underline{\mathbf{k}}_0^\top(\underline{\mathbf{x}}) \nabla g(\underline{\mathbf{x}}') \nabla^\top g(\underline{\mathbf{x}}') \underline{\mathbf{k}}_0(\underline{\mathbf{x}}) dx'_1 dx'_2 \\ &= \arg \max_{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} \left(\underline{\mathbf{k}}_0^\top(\underline{\mathbf{x}}) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(\underline{\mathbf{x}} - \underline{\mathbf{x}}') \nabla g(\underline{\mathbf{x}}') \nabla^\top g(\underline{\mathbf{x}}') dx'_1 dx'_2 \underline{\mathbf{k}}_0(\underline{\mathbf{x}}) \right) \quad (\text{D.7}) \end{aligned}$$

Das Doppelintegral in Gleichung D.7 stellt einen Tensor erster Ordnung dar (*siehe dazu* z.B. die Monographie von SCHMIDT /Sch1953/), der sich als geeignete mathematische Beschreibungsweise von lokaler Orientierung erweisen wird:

$$\underline{\mathbf{J}}(\underline{\mathbf{x}}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(\underline{\mathbf{x}} - \underline{\mathbf{x}}') \nabla g(\underline{\mathbf{x}}') \nabla g(\underline{\mathbf{x}}')^\top dx'_1 dx'_2 \quad (\text{D.8})$$

Dieser Tensor ist wegen des dyadischen Produkts von $\nabla g(\underline{\mathbf{x}})$ mit sich selbst immer symmetrisch ($J_{12} = J_{21}$). Seine Komponenten J_{pq} werden bestimmt als

$$J_{pq}(\underline{\mathbf{x}}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(\underline{\mathbf{x}} - \underline{\mathbf{x}}') \left(\frac{\partial g(\underline{\mathbf{x}}')}{\partial x'_p} \frac{\partial g(\underline{\mathbf{x}}')}{\partial x'_q} \right) dx'_p dx'_q \quad (\text{D.9})$$

Mit $\underline{\mathbf{J}}$ kann man Gleichung D.7 viel kompakter schreiben:

$$\overline{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} = \arg \max_{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} (\underline{\mathbf{k}}_0^\top(\underline{\mathbf{x}}) \underline{\mathbf{J}}(\underline{\mathbf{x}}) \underline{\mathbf{k}}_0(\underline{\mathbf{x}})) \quad (\text{D.10})$$

Diese Art von Problemstellung wird als Eigenwertproblem (eine sehr instruktive Darstellung geben GELLERT ET AL. in /GKN1990/) bezeichnet. Die Lösung, die hier zunächst nur angedeutet werden soll, erfolgt prinzipiell derart, daß $\underline{\mathbf{J}}$ durch Rotation des Koordinatensystems in eine Diagonalf orm $\underline{\mathbf{J}}'$ überführt wird. Auf der Hauptachse stehen dann die Eigenwerte von $\underline{\mathbf{J}}'$. Als Lösung für $\overline{\underline{\mathbf{k}}_0}$ kommt nur ein Vektor in Betracht, dessen Richtung mit einem der Eigenvektoren von $\underline{\mathbf{J}}$ übereinstimmt. Im hier betrachteten Fall der Maximierung von $\underline{\mathbf{J}}$ ist das der eigenwertgrößte Eigenvektor. Damit fällt $\overline{\underline{\mathbf{k}}_0}$ zugleich auf eine der Achsen des rotierten Koordinatensystems.

Diese Vorgehensweise führt zu einer Unterscheidbarkeit von Nachbarschaften mit konstantem Grauwert von solchen mit unkorrelierten Orientierungen. Sie ist damit leistungsfähiger als die Bildung eines Summenvektors über der lokalen Nachbarschaft. Das unterstreicht folgende Betrachtung:

Seien J'_1 und J'_2 (mit $J'_1 \geq J'_2$) die Eigenwerte des Tensors $\underline{\mathbf{J}}'$. Dann kann man folgende Fälle unterscheiden:

- $J'_1 = J'_2 = 0$

Da beide Eigenwerte Null sind, liegt keine lokale Orientierung vor, sondern es handelt sich bei der Struktur in der Nachbarschaft um einen *konstanten Grauwertbereich*.

- $J'_1 > 0, J'_2 = 0$

Da einer der beiden Eigenwerte Null geblieben ist, liegt eine lokale Nachbarschaft (siehe Abschnitt D.1 *Lokale Nachbarschaft*) vor, innerhalb derer eine *ideal orientierte Struktur* liegt.

- $J'_1 > 0, J'_2 > 0$

Es liegt keine lokale Nachbarschaft vor. Die Grauwerte ändern sich nicht nur entlang einer Richtung. Im Spezialfall $J'_1 = J'_2$ wird von einer isotropen Struktur gesprochen, es gibt *keine ausgezeichnete Orientierung*.

D.4 Lösung des Eigenwertproblems

Das Eigenwertproblem soll nun zwecks Bestimmung der Orientierung im Ortsraum gelöst werden, wozu der Tensor $\underline{\mathbf{J}}(\underline{\mathbf{x}})$ aus Gl. D.8 auf Hauptachsen gebracht werden muß. Dazu wird eine spezielle orthonormale Transformationsmatrix $\underline{\mathbf{T}}(\underline{\mathbf{x}})$ als Funktion eines Orientierungswinkels $\phi(\underline{\mathbf{x}})$ verwendet:

$$\underline{\mathbf{T}}(\underline{\mathbf{x}}) = \begin{bmatrix} \cos(\phi(\underline{\mathbf{x}})) & \sin(\phi(\underline{\mathbf{x}})) \\ -\sin(\phi(\underline{\mathbf{x}})) & \cos(\phi(\underline{\mathbf{x}})) \end{bmatrix} \quad (\text{D.11})$$

Damit wird aus $\underline{\mathbf{J}}(\underline{\mathbf{x}})$ eine Diagonalform $\underline{\mathbf{J}}'(\underline{\mathbf{x}})$ gebildet, die die Eigenwerte von $\underline{\mathbf{J}}(\underline{\mathbf{x}})$ auf der Hauptachse enthält:

$$\underline{\mathbf{J}}'(\underline{\mathbf{x}}) = \begin{bmatrix} J'_1(\underline{\mathbf{x}}) & 0 \\ 0 & J'_2(\underline{\mathbf{x}}) \end{bmatrix} \quad (\text{D.12})$$

$$= \underline{\mathbf{T}}^\top(\underline{\mathbf{x}}) \underline{\mathbf{J}}(\underline{\mathbf{x}}) \underline{\mathbf{T}}(\underline{\mathbf{x}}) \quad (\text{D.13})$$

Die Weiterführung des Ansatzes führt (unter Weglassung der Abhängigkeit von $\underline{\mathbf{x}}$ bei der Notation) zu:

$$\begin{bmatrix} J'_1 & 0 \\ 0 & J'_2 \end{bmatrix} = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{bmatrix} \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} \quad (\text{D.14})$$

$$= \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} J_{11} \cos \phi - J_{12} \sin \phi & J_{11} \sin \phi + J_{12} \cos \phi \\ J_{12} \cos \phi - J_{22} \sin \phi & J_{12} \sin \phi + J_{22} \cos \phi \end{bmatrix} \quad (\text{D.15})$$

Unter Anwendung goniometrischer Beziehungen für doppelte Winkel ($\sin 2\phi = 2 \sin \phi \cos \phi$ und $\cos 2\phi = \cos^2 \phi - \sin^2 \phi$) kann geschrieben werden:

$$\begin{bmatrix} J'_1 & 0 \\ 0 & J'_2 \end{bmatrix} = \begin{bmatrix} J_{11} \cos^2 \phi + J_{22} \sin^2 \phi - J_{12} \sin 2\phi & \frac{1}{2}(J_{11} - J_{22}) \sin 2\phi + J_{12} \cos 2\phi \\ \frac{1}{2}(J_{11} - J_{22}) \sin 2\phi + J_{12} \cos 2\phi & J_{22} \cos^2 \phi + J_{11} \sin^2 \phi + J_{12} \sin 2\phi \end{bmatrix} \quad (\text{D.16})$$

Dieses nichtlineare Gleichungssystem kann nun für ϕ gelöst werden, indem man ein Nebendiagonalelement von rechter und linker Seite gleichsetzt:

$$0 = \frac{1}{2}(J_{11} - J_{22}) \sin 2\phi + J_{12} \cos 2\phi \quad (\text{D.17})$$

Hieraus kann man direkt den Tangens des doppelten Orientierungswinkels bestimmen (Notation wieder als Funktion des Ortes):

$$\tan(2\phi(\underline{\mathbf{x}})) = \frac{2J_{12}(\underline{\mathbf{x}})}{J_{22}(\underline{\mathbf{x}}) - J_{11}(\underline{\mathbf{x}})} \quad (\text{D.18})$$

D.5 Interpretation im Ortsfrequenzraum – Trägheitstensoranalogie

Die bisher gegebene Darstellung zur Tensorrepräsentation erfolgte im Ortsraum. Der Tensor $\underline{\mathbf{J}}$ in Gl. D.8 basiert auf Gradienten ∇g . Im Ortsfrequenzraum ist eine andere Sicht möglich. Die Frequenzanalyse innerhalb einer Nachbarschaft ergibt eine spektrale Dichteverteilung $G(\underline{\mathbf{k}})$. Zur besseren Veranschaulichung der Tensormethode wird nachfolgend eine Darstellung im Ortsfrequenzraum gegeben.

Unter Verzicht auf die Gewichtungsfunktion $w(\underline{\mathbf{x}} - \underline{\mathbf{x}}')$ in Gl. D.7 zur Vereinfachung (die Darstellung würde sonst lediglich zu unhandlich) bleibt vom Doppelintegral I als Maximierung:

$$I_{\max}(\underline{\mathbf{x}}) = \max_{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\nabla^\top g(\underline{\mathbf{x}}') \underline{\mathbf{k}}_0(\underline{\mathbf{x}}))^2 dx'_1 dx'_2 \quad (\text{D.19})$$

Ohne Verletzung der Allgemeinheit kann darin der Integrand $(\cdot)^2$ als Quadrat eines Betrags $|\cdot|^2$ geschrieben werden:

$$I_{\max}(\underline{\mathbf{x}}) = \max_{\underline{\mathbf{k}}_0(\underline{\mathbf{x}})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\nabla^\top g(\underline{\mathbf{x}}') \underline{\mathbf{k}}_0(\underline{\mathbf{x}})|^2 dx'_1 dx'_2 \quad (\text{D.20})$$

Um nun den Schritt in den Ortsfrequenzraum tun zu können, wird die Eigenschaft der Fouriertransformation $\mathcal{F}$ ausgenutzt, daß die Norm einer Funktion $f(\underline{\mathbf{x}})$ erhalten bleibt:

$$\begin{aligned} \int_{-\infty}^{+\infty} |f(\underline{\mathbf{x}})|^2 d^n x &= \int_{-\infty}^{+\infty} |\mathcal{F}(f(\underline{\mathbf{x}}))|^2 d^n k \\ &= (2\pi)^n \int_{-\infty}^{+\infty} |F(\underline{\mathbf{k}})|^2 d^n k \end{aligned} \quad (\text{D.21})$$

Damit kann die Gl. D.20 im Ortsfrequenzraum geschrieben werden als:

$$I_{\max}(\underline{x}) = \max_{\underline{k}_0(\underline{x})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| \mathcal{F}(\nabla^\top g(\underline{x}') \underline{k}_0(\underline{x})) \right|^2 dk'_1 dk'_2 \quad (\text{D.22})$$

Die Fouriertransformation von ∇g erfolgt unter Zuhilfenahme des Differentiationssatzes $\mathcal{F}(\nabla f(\underline{x})) = i \underline{k} F(\underline{k})$. Entsprechend eingesetzt führt das zu:

$$\begin{aligned} I_{\max}(\underline{x}) &= \max_{\underline{k}_0(\underline{x})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| i \underline{k}'^\top G(\underline{k}') \underline{k}_0(\underline{x}) \right|^2 dk'_1 dk'_2 \\ &= \max_{\underline{k}_0(\underline{x})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left| \underline{k}'^\top \underline{k}_0(\underline{x}) \right|^2 \cdot |G(\underline{k}')|^2 dk'_1 dk'_2 \end{aligned} \quad (\text{D.23})$$

Betrachtet man den ersten Teil des Integranden einfach als Distanzmaß zwischen zwei Vektoren, so läßt sich das Skalarprodukt in Gl. D.23 prinzipiell auch durch die Summe der Abstandsquadrate ersetzen. Skalarprodukt und Abstandsquadratsumme sind jedoch gegenläufig, weswegen bei Ersetzung des Skalarprodukts anstelle der Maximierung des Integrals I eine Minimierung des neuen Integrals J vorgenommen werden muß:

$$J_{\min}(\underline{x}) = \min_{\underline{k}_0(\underline{x})} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\underline{k}' - \underline{k}_0(\underline{x}))^\top (\underline{k}' - \underline{k}_0(\underline{x})) \cdot |G(\underline{k}')|^2 dk'_1 dk'_2 \quad (\text{D.24})$$

Sowohl Gl. D.23 als auch Gl. D.24 beschreiben einen Geradenfit (lineare Regression) im Fourierraum. Die Darstellung in Gl. D.24 allerdings weist eine direkte Analogie zur Mechanik des starren Körpers auf. Interpretiert man das Bildspektrum $|G(\underline{k}')|^2$ als Dichteverteilung eines (flächenhaften) Körpers, so ergibt das Integral das Trägheitsmoment J dieses Körpers bei Rotation um eine Achse entlang $\underline{k}_0$ (durch den Koordinatenursprung), wobei $\underline{k}_0$ derart gerichtet sein muß, so daß J minimal wird. Überträgt man diese Betrachtung auf Gl. D.8, so kann der symmetrische Tensor $\underline{J}(\underline{x})$ als *Trägheitstensor minimaler Trägheit* des aus der mechanischen Analogiebetrachtung heraus angenommenen Körpers angesehen werden.

Anhang E

Weitere tabellarische Ergebnisübersichten

Dieser Anhang enthält in Ergänzung zum Abschnitt 6.4 *Demonstration des realisierten Systems* eine Zusammenstellung weiterer Ergebnistabellen, die in Auswertung verschiedener Testreihen mit dem realisierten dynamischen Auffälligkeitssystem zur Personenlokalisierung entstanden sind.

E.1 Skalierungsinvarianz der Gesichts- und Konturdetektion

Es wurde für diese beiden Cues experimentell ermittelt, welche Invarianz sie bezüglich einer Entfernungs- und damit Skalierungsvariation aufweisen. Für alle fünf Auflösungsstufen wurde dazu der bei maximalem Objektöffnungswinkel¹ erfaßbare Entfernungsbereich bestimmt. Beide exemplarischen Meßreihen sind im Innenraum unter Leuchtstoffröhrenbeleuchtung anhand der Versuchsperson Christoph (*siehe z. B.* Abbn. 4.4 und B.1) ausgeführt worden. Die hierbei bestimmte Beleuchtungsstärke bildete den Referenzwert für das in dieser Arbeit verwendete Maß „relative Beleuchtungsstärke“ (*siehe auch* S. 86), die damit hier $e_S = 1$ beträgt, so daß damit $\text{ld}(e_S) = 0$ ist.

Der Exponent des Skalierungsmaßes in den Tabellen E.1 und E.2 entspricht dem relativen Oktavabstand zur Auflösung der Ebene 0. Absolut gesehen ist diese Auflösung bei beiden Cues verschieden. Bei der Gesichtsdetektion beträgt sie ein Drittel der Grundauflösung des Framegrabbers (768×572), bei der Konturdetektion (und ebenso bei der Hautfarbdetek-

¹ca. 45° vertikal und 55° horizontal

tion) ein Siebtel. Diese Unterscheidung ist deshalb erforderlich, da ansonsten entweder die Gesichtsauflösung zu gering und/oder die Konturauflösung unnötig fein ausfallen würde.

Tabelle E.1: Skalierungsinvarianz der Gesichtsdetektion aus Abschnitt 4.2.1.2 Normierte mittelwertfreie Skalarprodukte (NMS) für Versuchsperson Christoph: Erkennbar ist, daß die gewählte Abstufung der Pyramide für diesen Cue gerade Mindestanforderungen für einen minimalen Überlapp genügt. Unter diesen Versuchsbedingungen betragen die theoretisch zu erwartenden Mittelwerte der erfaßten Entfernungsbereiche: 50cm, 70,7cm, 100cm, 141,4cm und 200cm.

Ebene	Skalierung	erf. Bereich [cm]	Ø [cm]
4	2^{-2}	40... 60	50,0
3	$2^{-\frac{3}{2}}$	62... 84	73,0
2	2^{-1}	86...114	100,0
1	$2^{-\frac{1}{2}}$	125...158	141,5
0	2^0	177...214	195,5

Tabelle E.2: Skalierungsinvarianz der Konturdetektion aus Abschnitt 5.3 Konturmodellierung durch ein Arrangement orientierter Filter für Versuchsperson Christoph: Im Vergleich zur Gesichtsdetektion (Abb. E.1) erweist sich dieser Cue invarianter gegenüber Skalierung, so daß sich gute Anschlüsse mit Überlapp zwischen den Ebenen ergeben.

Ebene	Skalierung	erf. Bereich [cm]	Ø [cm]
4	2^{-2}	40... 65	52,5
3	$2^{-\frac{3}{2}}$	55... 90	72,5
2	2^{-1}	80...130	105,0
1	$2^{-\frac{1}{2}}$	120...170	145,0
0	2^0	160...220	190,0

E.2 Rotationsinvarianz der Personenlokalisierung

Für das Gesamtsystem wurde ermittelt, in welchem Maße die Voraussetzung, daß eine zu lokalisierende Person frontal zur Kamera ausgerichtet sein muß, verletzt sein darf. Die Versuchsperson Christoph nahm auf einem Drehstuhl mit bestimmtem Abstand zur Kamera Platz und behielt den Kopf gradeaus ausgerichtet. Nur der Stuhl wurde um bestimmte Winkel gedreht. Die relative Beleuchtungsstärke betrug unter denselben Innenraumbedingungen ebenfalls $e_S = 1$.

Da die Hautfarbdetektion nur vernachlässigbar rotationsvariant ist, wurde sie nicht mit in die Tabelle aufgenommen. In der jeweiligen Einstellung noch sichtbare Gesichtshautpartien wurden praktisch unabhängig vom Drehwinkel auf allen Auflösungsebenen ortsrichtig

klassifiziert. Da keine besonderen Unterschiede zwischen beiden Drehrichtungen festgestellt wurden, beschränkt sich Tab. E.3 auf Winkelbeträge. Die Tests zeigen, daß beträchtliche Rotationswinkel möglich sind, ohne daß das Verfahren falsche Ergebnisse liefert.

Tabelle E.3: Rotationsinvarianz der Personenlokalisierung (aufgegliedert in Gesicht, Kontur und Fusion/Selektion): Zwei Versuchspersonen wurden in einem Abstand von ca. 90cm zur Kamera platziert. Das exemplarische Ergebnis zeigt, daß bis zu Winkeln von mehr als 30° die Lokalisation richtig auf Ebene 2 erfolgt (größte Ausdehnung des Blobs auf dieser Ebene) und erst bei darüber hinausgehenden Winkeln sowohl die Gesichtsdetektion versagt und als auch falsche Entfernungsschätzungen auftreten (vgl. mit Tabn. E.1 und E.2).

Winkelbetrag [rad]	Gesicht	Kontur	Person
Christoph:			
0	2	2	2
$\frac{\pi}{16}$	2	2	2
$\frac{\pi}{8}$	2	2	2
$\frac{3\pi}{16}$	2	2	2
$\frac{\pi}{4}$	—	2	1
$\frac{\pi}{2}$	—	3	3
Autor:			
0	2	2	2
$\frac{\pi}{16}$	2	2	2
$\frac{\pi}{8}$	2	2-3	2
$\frac{3\pi}{16}$	2	3	2
$\frac{\pi}{4}$	—	3	—
$\frac{\pi}{2}$	—	3	2

Literaturverzeichnis*

- /Ama1977/ AMARI, SHUN-ICHI: *Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields*. Biological Cybernetics – Communication and Control in Organisms and Automata, 27:77–87, 1977. Zitiert auf den Seiten 9, 65 und 75.
- /BA1983/ BURT, PETER J. und ADELSON, EDWARD H.: *The Laplacian Pyramid as a Compact Image Code*. IEEE Transactions on Communications, 31(4):532–540, April 1983. Zitiert auf Seite 15.
- /Bar1993/ BARTSCH, HANS-JOCHEN: *Taschenbuch mathematischer Formeln*. Fachbuchverlag Leipzig, 1993.
- /Bau1972/ BAUM, L.: *An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes*. Inequalities, 3:1–8, 1972. Zitiert auf Seite xx.
- /BBBG1997/ BRAUMANN, ULF-DIETRICH; BRAKENSIEK, ANJA; BÖHME, HANS-JOACHIM und GROSS, HORST-MICHAEL: *Ein neuronales Verfahren zur Segmentierung hautfarbener Bildregionen*. In: PAULUS, DIETRICH und WAGNER, THOMAS (Herausgeber): *Tagungsband zum 3. Workshop „Farbbildverarbeitung“*, Seiten 67–72, 101. IRB-Verlag, Stuttgart, September 1997. Zitiert auf Seite 30.
- /BBBG1998/ BRAUMANN, ULF-DIETRICH; BRAKENSIEK, ANJA; BÖHME, HANS-JOACHIM und GROSS, HORST-MICHAEL: *Gaborfilterbasierte visuelle Personenlokalisierung mit dreidimensionalen Feldern dynamischer Neuronen*. In: NAUCK, DETLEF; KRELL, GERALD; KRUSE, RUDOLF und MICHAELIS, BERND (Herausgeber): *Proceedings of the Third International Workshop “Neural Networks in Applications” (NN’98, Magdeburg)*, Seiten 31–38, Februar 1998. Zitiert auf Seite 56.

*Die alphabetische Sortierreihenfolge im Literaturverzeichnis bezieht sich auf die *Schlüssel* der Literaturverweise (linke Spalte) und *nicht* auf die Erstautoren! – Abgesehen von einigen Nachschlagewerken sind für alle Literaturverweise die jeweiligen Stellen ihres Auftretens in der Arbeit mit Seitenzahl angegeben.

- /BBC⁺1998/ BRAKENSIEK, ANJA; BRAUMANN, ULF-DIETRICH; CORRADINI, ANDREA; BÖHME, HANS-JOACHIM und GROSS, HORST-MICHAEL: *Neuronale Verfahren zur Lokalisation und Bewertung von Handgesten in Real-World-Szenen*. In: NAUCK, DETLEF; KRELL, GERALD; KRUSE, RUDOLF und MICHAELIS, BERND (Herausgeber): *Proceedings of the Third International Workshop "Neural Networks in Applications" (NN'98, Magdeburg)*, Seiten 211–218, Februar 1998. Zitiert auf den Seiten 37 und 120.
- /BG1988/ BIGÜN, JOSEF und GRANLUND, GÖSTA H.: *Optimal Orientation Detection of Linear Symmetry*. In: *Proceedings of the 1st International Conference on Computer Vision (London)*, Seiten 433–438. IEEE Computer Society Press, 1988. Zitiert auf Seite 56.
- /BG1993/ BRUCE, VICKI und GREEN, PATRICK R.: *Visual Perception: Physiology, Psychology and Ecology*. Lawrence Erlbaum Associates, London, 2. Auflage, 1993. Zitiert auf den Seiten 38 und 49.
- /Bis1995/ BISHOP, CHRISTOPHER M.: *Neural Networks for Pattern Recognition*. Oxford University Press, 1995. Zitiert auf Seite 31.
- /Bra1995/ BRAUSE, RÜDIGER: *Neuronale Netze – Eine Einführung in die Neuroinformatik*, Kapitel 2.1.3 PCA und Dekorrelationsnetze, Seiten 99–106. Leitfäden der Informatik. B. G. Teubner Stuttgart, 2. Auflage, 1995. Zitiert auf Seite 119.
- /BSG⁺1996/ BRONSTEIN, ILJA NIKOLAJEWITSCH; SEMENDJAJEW, KONSTANTIN ADOLFOWITSCH; GROSCHE, GÜNTER; ZIEGLER, VIKTOR; ZIEGLER, DOROTHEA und ZEIDLER, EBERHARD (Herausgeber): *Teubner-Taschenbuch der Mathematik – Teil I*. B. G. Teubner Verlagsgesellschaft Leipzig, 1996.
- /Bur1984/ BURT, PETER J.: *The Pyramid as a Structure for Efficient Computation*. In: ROSENFELD, AZRIEL (Herausgeber): *Multiresolution Image Processing and Analysis*, Band 12 der Reihe *Springer Series in Information Science*, Seiten 6–35. Springer-Verlag, 1984. Zitiert auf Seite 15.
- /BVS⁺1996/ BARTLETT, MARIAN STEWART; VIOLA, PAUL A.; SEJNOWSKI, TERRENCE J.; GOLOMB, BEATRICE A.; LARSEN, JAN; HAGER, JOSEPH C. und EKMAN, PAUL: *Classifying Facial Action*. In: TOURETZKY, DAVID S.; MOZER, MICHAEL C. und HASSELMO, MICHAEL E. (Herausgeber): *Advances in Neural Information Processing Systems 8 (NIPS'95)*, Seiten 823–829. The MIT Press, 1996. Zitiert auf Seite 2.
- /CBB⁺1998/ CORRADINI, ANDREA; BRAUMANN, ULF-DIETRICH; BRAKENSIEK, ANJA; KRABBES, MARKUS; BOEHME, HANS-JOACHIM und GROSS, HORST-MICHAEL: *Visual Gesture Localization with Dynamic Neural Fields: To-*

- wards a Gesture Recognition System.* In: MARINARO, MARIA und TAGLIAFERRI, ROBERTO (Herausgeber): *Proceedings of the 10-th Italian Workshop on Neural Networks (WIRN'98)*, Vietri Sul Mare, Seiten 201–206. Springer-Verlag, Mai 1998. Zitiert auf Seite 56.
- /CBBG1998/ CORRADINI, ANDREA; BRAUMANN, ULF-DIETRICH; BOEHME, HANS-JOACHIM und GROSS, HORST-MICHAEL: *Contour-based Person Localization by 3D Neural Fields and Steerable Filters.* In: *Proceedings of the 6th IAPR Workshop on Machine Vision Applications (MVA'98)*, Seiten 93–96, November 1998. Zitiert auf Seite 56.
- /CBBG1999/ CORRADINI, ANDREA; BRAUMANN, ULF-DIETRICH; BOEHME, HANS-JOACHIM und GROSS, HORST-MICHAEL: *3D Neural Fields and Steerable Filters for Contour-based Person Localization.* In: LINDBLAD, THOMAS; PADGETT, MARY LOU und KINSER, JASON M. (Herausgeber): *Ninth Workshop on Virtual Intelligence – Dynamic Neural Networks (VIDYNN'98)*, SPIE Proceedings Series 3728, Seiten 476–485. The International Society for Optical Engineering, März 1999. Zitiert auf Seite 56.
- /Cor2000/ CORRADINI, ANDREA: *DYGEST: Neural Architectures for Visual-Based Dynamic Gesture Recognition.* Doktorarbeit, Technische Universität Ilmenau, 2000. Zitiert auf den Seiten 4 und 112.
- /Dau1980/ DAUGMAN, JOHN G.: *Two-Dimensional Spectral Analysis of Cortical Receptive Field Profiles.* *Vision Research - An International Journal for Functional Aspects of Vision*, 20:847–856, 1980. Zitiert auf den Seiten 39 und 56.
- /Den1997/ DENZLER, JOACHIM: *Aktives Sehen zur Echtzeitobjektverfolgung.* Doktorarbeit, Universität Erlangen-Nürnberg, 1997. Zitiert auf Seite 50.
- /DGWH1998/ DARRELL, TREVOR; GORDON, GAILE; WOODFILL, JOHN und HARVILLE, MICHAEL: *A Virtual Mirror-Interface using Real-time Robust Face Tracking.* In: *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, Seiten 616–621. IEEE Computer Society, April 1998. Zitiert auf Seite 8.
- /DI1997/ D'ZMURA, MICHAEL und IVERSON, GEOFFREY: *Color Constancy: Recovery of Spectral Properties of Lights and Surfaces.* Via Webpage <http://www.socsci.uci.edu/cogsci/projects/coco/fcoco/coco.html>, University of California at Irvine, Department of Cognitive Sciences, April 1997. Zitiert auf Seite 114.
- /DI1998/ D'ZMURA, MICHAEL und IVERSON, GEOFFREY: *A Formal Approach to Color Constancy: The Recovery of Surface and Light Source Spectral Properties using Bilinear Models.* In: DOWLING, CORNELIA E.; ROBERTS,

- FRED S. und THEUNS, PETER (Herausgeber): *Recent Progress in Mathematical Psychology, Part II: Psychophysics and Psychometrics*, Scientific Psychology. Lawrence Erlbaum Associates, Hillsdale, NJ, 1998. in press. Zitiert auf Seite 114.
- /Ede1995/ EDELMAN, SHIMON: *Receptive Field for Vision: From Hyperacuity to Object Recognition*. Technischer Bericht CS-TR 95-29, Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot, Israel, 1995. Zitiert auf Seite 54.
- /FA1991/ FREEMAN, WILLIAM T. und ADELSON, EDWARD H.: *The Design and Use of Steerable Filters*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 13(9):891–906, September 1991. Zitiert auf Seite 56.
- /Fra1997/ FRANKE, KARL-HEINZ: *Grundlagen der Farbbildverarbeitung*. Skript zur Vorlesung „Grundlagen der Farbbildverarbeitung für Informatiker“ an der TU Ilmenau, Oktober 1997. Zitiert auf Seite 16.
- /Fre1988/ FREY, HERBERT: *Digitale Bildverarbeitung in Farbräumen*. Doktorarbeit, Technische Universität München, 1988. Zitiert auf Seite 16.
- /Fre1990/ FREY, HERBERT: *Die Verarbeitung von Farbbildern nach Helligkeit, Sättigung und Buntton*. In: *Mustererkennung 1990 – 12. DAGM-Symposium*, Informatik aktuell, Seiten 324–331. Springer-Verlag, 1990. Zitiert auf Seite 16.
- /Fre1992/ FREEMAN, WILLIAM TAFEL: *Steerable Filters and Local Analysis of Image Structure*. Doktorarbeit, Massachusetts Institute of Technology, Cambridge, Juni 1992. Zitiert auf Seite 56.
- /Fri1995/ FRITZKE, BERND: *A Growing Neural Gas Network Learns Topologies*. In: TESAURO, GERALD; TOURETZKY, DAVID S. und LEEN, TODD K. (Herausgeber): *Advances in Neural Information Processing Systems 7 (NIPS'94)*, Seiten 625–632. The MIT Press, 1995. Zitiert auf den Seiten xix, 30, 37 und 120.
- /FZ1998/ FEYRER, STEFAN und ZELL, ANDREAS: *Ein integrierter Ansatz zur Lokalisierung von Personen in Bildfolgen*. In: POSCH, STEFAN und RITTER, HELGE (Herausgeber): *Dynamische Perzeption – Workshop der GI-Fachgruppe 1.0.4 Bildverstehen in Kooperation mit dem SFB 360: Situierete Künstliche Kommunikatoren*, Reihe Proceedings in Artificial Intelligence, Vol. 8 (Herausgeber: Freksa, Christian), Seiten 183–190, Sankt Augustin, Juni 1998. infix. Zitiert auf Seite 67.
- /FZ1999/ FEYRER, STEFAN und ZELL, ANDREAS: *Personentracking mit einer mobilen Roboterplattform unter Verwendung eines multimodalen Detektions-*

- ansatzes*. Künstliche Intelligenz, Seiten 7–12, Heft 1, März 1999. Zitiert auf den Seiten 5 und 50.
- /Gáb1946/ GÁBOR, DÉNES: *Theory of Communication*. Journal of the Institute of Electrical Engineers (IEE, London) – Part 3: Radio and Communication Engineering, 93:429–457, 1946. Zitiert auf den Seiten 39 und 56.
- /Gie1997/ GIESE, MARTIN A.: *A Dynamical Model for the Perceptual Organization of Motion*. Doktorarbeit, Ruhr-Universität Bochum, 1997. Zitiert auf Seite 74.
- /Gie1998/ GIESE, MARTIN A.: *Dynamic Neural Field Model Links Neural and Computational Theories of Motion Perception*. In: POSCH, STEFAN und RITTER, HELGE (Herausgeber): *Dynamische Perzeption – Workshop der GI-Fachgruppe 1.0.4 Bildverstehen in Kooperation mit dem SFB 360: Situierete Künstliche Kommunikatoren*, Reihe Proceedings in Artificial Intelligence, Vol. 8 (Herausgeber: Freksa, Christian), Seiten 159–166, Sankt Augustin, Juni 1998. infix. Zitiert auf Seite 74.
- /GKN1990/ GELLERT, WALTER; KÄSTNER, HERBERT und NEUBER, SIEGFRIED: *Lexikon der Mathematik*. Bibliographisches Institut Leipzig, 5. Auflage, 1990. Zitiert auf Seite 140.
- /Goo1997/ GOODRIDGE, STEVEN GEORGE: *Multimedia Sensor Fusion for Intelligent Camera Control and Human-Computer Interaction*. Doktorarbeit, North Carolina State University, Raleigh, 1997. On-line: <http://www.mindspring.com/~sggoodri/dissert/contents.htm>. Zitiert auf den Seiten 6 und 7.
- /Gra1978/ GRANLUND, GÖSTA H.: *In Search of a General Picture Processing Operator*. Computer Graphics and Image Processing (CGIP), 8:155–173, 1978. Zitiert auf Seite 56.
- /GSW1975/ GOREN, CAROLYN C.; SARTY, MERRILL und WU, PAUL Y. K.: *Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants*. Pediatrics, 56(4):544–549, 1975. Zitiert auf Seite 38.
- /GZZZ1995/ GROSCHE, GÜNTER; ZIEGLER, VIKTOR; ZIEGLER, DOROTHEA und ZEIDLER, EBERHARD (Herausgeber): *Teubner-Taschenbuch der Mathematik – Teil II*. B. G. Teubner Verlagsgesellschaft Leipzig, 1995.
- /Ham1995/ HAMMERSTROM, DAN: *A Digital VLSI Architecture for Real-World-Applications*. In: ZORNETZER, STEVEN F.; DAVIS, JOEL L.; LAU, CLIFFORD und MCKENNA, THOMAS (Herausgeber): *An Introduction to Neural and Electronic Networks*, Kapitel 17, Seiten 335–358. Academic Press, 2. Auflage, 1995. Zitiert auf Seite 133.

- /Ham1999/ HAMKER, FRED H.: *Visuelle Aufmerksamkeit und lebenslanges Lernen im Wahrnehmungs-Handlungs-Zyklus*. Doktorarbeit, Technische Universität Ilmenau, 1999. Zitiert auf Seite 74.
- /HBB1998/ HANCOCK, PETER J. B.; BRUCE, VICKI und BURTON, MIKE A.: *A Comparison of Two Computer-Based Face Identification Systems with Human Perceptions of Faces*. Vision Research - An International Journal for Functional Aspects of Vision, 38(15/16):2277–2288, August 1998. Zitiert auf Seite 38.
- /Heb1949/ HEBB, DONALD O.: *The Organization of Behavior*, Kapitel 4. The First Stage of Perception: Growth of an Assembly, Seiten 60–78. Wiley, New York, 1949. Zitiert auf Seite xx.
- /Hem1998/ HEMPEL, TORSTEN: *Implementation von Eigenfaces auf dem SIMD-Neurocomputer CNAPS*. Dokumentation zur Hauptseminarsarbeit am FG Neuroinformatik der TU Ilmenau, Juli 1998. Zitiert auf Seite 134.
- /Hem2000/ HEMPEL, TORSTEN: *Koordination und Implementation von Personenlokalisierung, -verifikation und -tracking für mobile Service-Systeme*. Diplomarbeit, Technische Universität Ilmenau, Dezember 2000. Zitiert auf Seite 111.
- /Her1905/ HERING, EWALD: *Grundzüge der Lehre vom Lichtsinn*. In: GRÄFE, A. und SÄMISCH, T. (Herausgeber): *Handbuch der Augenheilkunde*. I. Engelmann, Leipzig, 1905. Zitiert auf den Seiten 26 und 113.
- /HG1997/ HAMKER, FRED H. und GROSS, HORST-MICHAEL: *Object Selection with Dynamic Neural Maps*. In: GERSTNER, WULFRAM; GERMOND, ALAIN J.; HASLER, MARTIN und NICLOUD, JEAN-DANIEL (Herausgeber): *ICANN'97 – Proceedings of the 7th International Conference on Artificial Neural Networks*, Lecture Notes in Computer Science 1327, Seiten 919–924. Springer-Verlag, Oktober 1997. Zitiert auf Seite 74.
- /HKRS1995/ HERPERS, RAINER; KATTNER, HOLGER; RODAX, HOLM und SOMMER, GERALD: *GAZE: An Attentive Processing Strategy to Detect and Analyze the Prominent Facial Regions*. In: BICHSEL, MARTIN (Herausgeber): *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition (FG'95)*, Seiten 214–219. IEEE Computer Society, MultiMedia Laboratory, University of Zurich, Juni 1995. Zitiert auf Seite 2.
- /Hun1994/ HUNKE, H. MARTIN: *Locating and Tracking of Human Faces with Neural Networks*. Technischer Bericht CMU-CS-94-155, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, August 1994. Zitiert auf Seite 31.

- /IKN1998/ ITTI, LAURENT; KOCH, CHRISTOF und NIEBUR, ERNST A.: *A Model of Saliency-based Visual Attention for Rapid Scene Analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 20(11):1254–1259, 1998. Zitiert auf Seite 69.
- /Jäh1993/ JÄHNE, BERND: *Digitale Bildverarbeitung*. Springer-Verlag, 3. Auflage, 1993. Zitiert auf den Seiten 56 und 60.
- /Jäh1995/ JÄHNE, BERND: *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*. Springer-Verlag, 3. Auflage, 1995. Zitiert auf Seite 56.
- /Jäh1997/ JÄHNE, BERND: *Practical Handbook on Image Processing for Scientific Applications*. CRC Press LLC, Boca Raton, 1997. Zitiert auf Seite 56.
- /Jäh1998/ JÄHNE, BERND: *Moderne Konzepte der digitalen Bildverarbeitung*. Skript zum DAGM-Tutorium, September 1998. Zitiert auf den Seiten 3, 56 und 59.
- /JMNS1996/ JÄHNE, BERND; MASSEN, ROBERT; NICKOLAY, BERTRAM und SCHARFENBERG, HARALD: *Technische Bildverarbeitung – Maschinelles Sehen*. Springer-Verlag, 1996. Zitiert auf Seite 56.
- /JP1987a/ JONES, JUDSON P. und PALMER, LARRY A.: *An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex*. Journal of Neurophysiology, 58(6):1233–1258, 1987. Zitiert auf den Seiten 54 und 55.
- /JP1987b/ JONES, JUDSON P. und PALMER, LARRY A.: *The Two-Dimensional Spatial Structure of Simple Receptive Fields in Cat Striate Cortex*. Journal of Neurophysiology, 58(6):1187–1211, 1987. Zitiert auf Seite 54.
- /JR1994/ JOLION, JEAN-MICHEL und ROSENFELD, AZRIEL: *A Pyramid Framework for Early Vision: Multiresolutional Computer Vision*. Kluwer Academic Publishers, 1994. Zitiert auf den Seiten 7 und 13.
- /KA1979/ KISHIMOTO, K. und AMARI, SHUN-ICHI: *Existence and Stability of Local Excitations in Homogeneous Neural Fields*. Journal of Mathematical Biology, 7:303–318, 1979. Zitiert auf den Seiten 9 und 75.
- /Kar1984/ KARHUNEN, JUHA: *Recursive Estimation of Eigenvectors of Correlation Type Matrices for Signal Processing Applications*. Doktorarbeit, Helsinki University of Technology, 1984. Zitiert auf den Seiten xx, 119 und 127.
- /KD1990/ KOENDERINK, JAN J. und DOORN, ANDREA J. VAN: *Receptive Field Families*. Biological Cybernetics – Communication and Control in Organisms and Automata, 63:291–297, 1990. Zitiert auf Seite 54.

- /Kem1997/ KEMPE, HEIKO: *Effiziente Implementation der Tensormethode zur Orientierungsbestimmung im Ortsraum*. Zentrum für Bild- und Signalverarbeitung (ZBS) Ilmenau, 1997. persönliche Kommunikation. Zitiert auf Seite 59.
- /Key1998/ KEY, JÜRGEN: *Wettbewerb und Kooperation von Neuronen in Modellen zur Aufmerksamkeit*. Diplomarbeit, Technische Universität Ilmenau, März 1998. Zitiert auf Seite 76.
- /KHB1996/ KORTENKAMP, DAVID; HUBER, ERIC und BONASSO, R. PETER: *Recognizing and Interpreting Gestures on a Mobile Robot*. In: *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI'96) and the Eighth Innovative Applications of Artificial Intelligence Conference*, Seiten 915–921. AAAI/MIT Press, 1996. Zitiert auf Seite 2.
- /KK1994/ KASKI, SAMUEL und KOHONEN, TEUVO: *Winner-Take-All Networks for Physiological Models of Competitive Learning*. Neural Networks: Off. Journal of INNS, ENNS and JNNS, 7(6/7):973–984, 1994. Zitiert auf Seite 74.
- /Kof1935/ KOFFKA, KURT: *Principles of the Gestalt Psychology*. Harcourt, Brace & World, New York, 1935. Zitiert auf Seite 49.
- /Kop1992/ KOPECZ, JÖRG: *Modellierung lokaler Wechselwirkungen in diskreten corticalen Strukturen mittels nichtlinearer Dynamik*. Doktorarbeit, Ruhr-Universität Bochum, 1992. Zitiert auf den Seiten xiv, xix und 50.
- /Kop1995/ KOPECZ, KLAUS: *Unsupervised Learning of Sequences on Maps with Lateral Connectivity*. In: FOGELMAN-SOULIÉ, FRANCOISE und GALLINARI, P. (Herausgeber): *ICANN'95 – Proceedings of the 5th International Conference on Artificial Neural Networks*, Band 1, Seiten 431–436. EC2 & Cie., Paris, 1995. Zitiert auf Seite 74.
- /Kop1996/ KOPECZ, KLAUS: *Neural Field Dynamics Provide Robust Control of Attentional Resources*. In: MERTSCHING, BÄRBEL (Herausgeber): *Aktives Sehen in technischen und biologischen Systemen – Workshop der GI-Fachgruppe 1.0.4 Bildverstehen*, Reihe Proceedings in Artificial Intelligence, Vol. 4 (Herausgeber: Freksa, Christian), Seiten 137–144, Sankt Augustin, Dezember 1996. infix. Zitiert auf Seite 74.
- /KR1996/ KUBISCH, ROBERT und RITTER, HELGE: *Erkennen menschlicher Kopfhaltungen mittels künstlicher neuronaler Netze*. In: JÄHNE, B.; GEISLER, P.; HAUSSECKER, H. und HERING, F. (Herausgeber): *Mustererkennung 1996 – 18. DAGM-Symposium (Heidelberg)*, Informatik aktuell, Seiten 109–117. Springer-Verlag, 1996. Zitiert auf Seite 66.

- /Kra1997/ KRABBES, MARKUS: *Optimierungsmöglichkeiten lateraler Aktivierungsfunktionen innerhalb dreidimensionaler neuronaler Felder unter Berücksichtigung der zugrundeliegenden pyramidalen Inputtopologie*. FG Neuroinformatik, TU Ilmenau, 1997. persönliche Kommunikation. Zitiert auf Seite 107.
- /Kri1905/ KRIES, JOHANNES V.: *Influence of Adaptation on the Effects produced by Luminous Stimuli*. In: MACADAM, DAVID L. (Herausgeber): *Sources of Color Vision*. Massachusetts Institute of Technology, 1905. Zitiert auf Seite 113.
- /KS1995/ KOPECZ, KLAUS und SCHÖNER, GREGOR: *Saccadic Motor Planning by Integrating Visual Information and Pre-Information on Neural Dynamic Fields*. Biological Cybernetics – Communication and Control in Organisms and Automata, 73(1):49–60, 1995. Zitiert auf Seite 74.
- /KU1985/ KOCH, CHRISTOF und ULLMAN, SHIMON: *Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry*. Human Neurobiology, 4:219–227, 1985. Zitiert auf Seite 69.
- /Lan1995/ LANG, HEINWIG: *Farbwiedergabe in den Medien: Fernsehen, Film, Druck*. Muster-Schmidt Verlag, Göttingen - Zürich, 1995. Zitiert auf Seite 30.
- /LM1971/ LAND, EDWIN H. und MCCANN, JOHN J.: *Lightness and Retinex Theory*. Journal of the Optical Society of America, 61:1–11, 1971. Zitiert auf den Seiten 113 und 114.
- /LMW1992/ LÖFFLER, HELMUT; MEINHARD, JÜRGEN und WERNER, DIETER: *Taschenbuch der Informatik*. Fachbuchverlag Leipzig, 1992.
- /Loc1992/ LOCHER, FRANZ: *Numerische Mathematik für Informatiker*. Springer-Verlag, 1992. Zitiert auf Seite xix.
- /Loè1977/ LOÈVE, MICHEL M.: *Probability Theory I*, Band 45 der Reihe *Springer Graduate Texts in Mathematics*. Springer-Verlag, 4. Auflage, 1977. Zitiert auf den Seiten xx und 127.
- /LR1995/ LITTMANN, ENNO und RITTER, HELGE: *Neuronale und statistische Verfahren zur adaptiven Farbsegmentierung – ein Leistungsvergleich*. Technischer Bericht 95/3, Universität Bielefeld, SFB 360 (Situierete Künstliche Kommunikatoren), 1995. Zitiert auf den Seiten 27 und 31.
- /LR1997/ LITTMANN, ENNO und RITTER, HELGE: *Adaptive Color Segmentation – A Comparison of Neural and Statistical Methods*. IEEE Transactions on Neural Networks, 8(1):175–185, 1997. Zitiert auf Seite 31.
- /LS1998/ LEE, DANIEL D. und SEUNG, H. SEBASTIAN: *A Neural Network Based Head Tracking System*. In: JORDAN, MICHAEL I.; KEARNS, MICHAEL J.

- und SOLLA, SARA A. (Herausgeber): *Advances in Neural Information Processing Systems 10 (NIPS'97)*, Seiten 908–914. The MIT Press, Juni 1998. Zitiert auf den Seiten 9 und 66.
- /Mag1995/ MAGGIONI, CHRISTOPH: *GestureComputer – New Ways of Operating a Computer*. In: BICHSEL, MARTIN (Herausgeber): *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition (FG'95)*, Seiten 166–171. IEEE Computer Society, MultiMedia Laboratory, University of Zurich, Juni 1995. Zitiert auf Seite 2.
- /Mal1994/ MALSBURG, CHRISTOPH VON DER: *The Correlation Theory of Brain Function*. In: DOMANY, E.; HEMMEN, JAN L. VAN und SCHULTEN, KLAUS (Herausgeber): *Models of Neural Networks II*, Kapitel 2, Seiten 95–119. Springer-Verlag, 1994. Original: Report 81-2, Abt. Neurobiologie, Max-Planck-Institut für biophysikalische Chemie, Göttingen, 1981. Zitiert auf Seite 68.
- /Mal1998/ MALLOT, HANSPETER A.: *Sehen und die Verarbeitung visueller Informationen – Eine Einführung*, Kapitel 5 Farbkonstanz. Computational Intelligence (Herausgeber: Bibel, Wolfgang und Kruse, Rudolf). Verlag Vieweg, Wiesbaden, 1998. Zitiert auf Seite 114.
- /Mar1982/ MARR, DAVID: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, New York, 1982. Zitiert auf Seite 49.
- /Mel1997a/ MEL, BARTLETT W.: *SEEMORE: Combining Color, Shape, and Texture Histogramming in a Neurally-Inspired Approach to Visual Object Recognition*. Neural Computation, 9(4):777–804, 1997. Zitiert auf Seite 69.
- /Mel1997b/ MELLER, SEBASTIAN: *Segmentierungsfreie Personendetektion*. Diplomarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, November 1997. Zitiert auf Seite 51.
- /MMTM1998/ MÖLLER, RALF; MARIS, MARINUS; TOMES, JÖRG und MOJAEV, ALEXANDER: *A Strong Winner-Take-All Neural Network in Analogue Hardware*. In: SMITH, LESLIE S. und HAMILTON, ALISTER (Herausgeber): *Neuromorphic Systems – Engineering Silicon From Neurobiology*, Nummer 10 in *Progress in Neural Processing*. World Scientific, 1998. Zitiert auf Seite 75.
- /Mok1997/ MOKHTARIAN, FARZIN: *Silhouette-Based Occluded Object Recognition through Curvature Scale Space*. Machine Vision and Applications, 10(3):87–97, 1997. Zitiert auf Seite 48.

- /Möl1996/ MÖLLER, RALF: *Wahrnehmung durch Vorhersage – Eine Konzeption der handlungsorientierten Wahrnehmung*. Doktorarbeit, Technische Universität Ilmenau, 1996. Zitiert auf Seite 75.
- /MP1943/ MCCULLOCH, WARREN S. und PITTS, WALTER: *A Logical Calculus of the Ideas immanent in Nervous Activity*. Bulletin of Mathematical Biophysics, 5:115–133, 1943. Zitiert auf Seite xx.
- /MPW1993/ MILANESE, RUGGERO; PUN, THIERRY und WECHSLER, HARRY: *A Non-Linear Integration Process for the Selection of Visual Information*. In: ROBERTO, VITO (Herausgeber): *Intelligent Perceptual Systems – New Directions in Computational Perception*, Lecture Notes in Artificial Intelligence 745, Seiten 322–336. Springer-Verlag, 1993. Zitiert auf Seite 69.
- /MW1986/ MALONEY, LAURENCE T. und WANDELL, BRIAN A.: *Color Constancy: A Method for Recovering Surface Spectral Reflectance*. Journal of the Optical Society of America A, 3:29–33, 1986. Zitiert auf Seite 118.
- /NM1998/ NASTAR, CHAHAB und MITSCHKE, MATTHIAS: *Real-Time Face Recognition Using Feature Combination*. In: *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, Seiten 312–317. IEEE Computer Society, April 1998. Zitiert auf den Seiten 8 und 69.
- /OF1996/ OLSHAUSEN, BRUNO A. und FIELD, DAVID J.: *Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images*. Nature, 381(6583):607–609, Juni 1996. Zitiert auf Seite 54.
- /Oja1983/ OJA, ERKKI: *Subspace Methods of Pattern Recognition*. Research Studies Press, Letchworth, Hertfordshire, 1983. Zitiert auf Seite 119.
- /OPS⁺1997/ OREN, MICHAEL; PAPAGEORGIOU, CONSTANTINE; SINHA, PAWAN; OSUNA, EDGAR und POGGIO, TOMASO: *Pedestrian Detection Using Wavelet Templates*. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, Seiten 193–199. IEEE Computer Society, November 1997. Zitiert auf den Seiten xiv, 8 und 51.
- /OT1986/ OLZAK, LYNN A. und THOMAS, JAMES P.: *Seeing Spatial Patterns*. In: BOFF, KENNETH R.; KAUFMAN, LLOYD und THOMAS, JAMES P. (Herausgeber): *Handbook of Perception and Human Performance*, Band I – Sensory Processes and Perception: Basic Sensory Processes I, Kapitel 7, Seiten 7-1–7-55. Wiley, New York, 1986. Zitiert auf Seite 54.
- /Pen1996a/ PENTLAND, ALEX P.: *Intelligente Zimmer*. Spektrum der Wissenschaft, Seiten 44–51, Heft 6, Juni 1996. Zitiert auf den Seiten 2 und 68.

- /Pen1996b/ PENTLAND, ALEX P.: *Smart Rooms*. Scientific American, 274(4):54–62, April 1996. Zitiert auf den Seiten 2, 5 und 68.
- /Pen1996c/ PENTLAND, ALEX P.: *Zimmer denkt mit*. CHIP, Seiten 54, 56, Heft 6, Juni 1996. Zitiert auf Seite 68.
- /Pen1997/ PENTLAND, ALEX P.: *Intelligente Zimmer*. Spektrum der Wissenschaft, Seiten 84–91, Dossier „Kopf oder Computer“ 4/97 1997. Zitiert auf Seite 68.
- /PG1990/ POGGIO, TOMASO und GIROSI, FEDERICO: *Regularization Algorithms for Learning that are equivalent to Multilayer Networks*. Science, 247:978–982, 1990. Zitiert auf Seite xxi.
- /PG1996a/ POMIERSKI, TORSTEN und GROSS, HORST-MICHAEL: *Biological Neural Architecture for Chromatic Adaptation Resulting in Constant Color Sensations*. In: *Proceedings of the 1996 IEEE International Conference on Neural Networks (ICNN'96)*, Seiten 734–739. IEEE Service Center, 1996. Zitiert auf Seite 115.
- /PG1996b/ POMIERSKI, TORSTEN und GROSS, HORST-MICHAEL: *Neural Architecture for Chromatic Adaptation of Single Images and Picture Sequences*. In: *Tagungsband 2. Nationaler Workshop „Farbbildverarbeitung“*, Schriftenreihe des Zentrums für Bild- und Signalverarbeitung e.V. Ilmenau, Seiten 29–34, Ilmenau, 1996. Zitiert auf Seite 115.
- /PM1997/ PASCHKE, PETER und MÖLLER, RALF: *Simulation of Sparse Neural Networks on a CNAPS SIMD Neurocomputer*. In: SYDOW, ACHIM (Herausgeber): *Proceedings of the 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics*, Band 6, Seiten 529–534. GMD First Berlin, 1997. Zitiert auf Seite 133.
- /PM1998/ PASCHKE, PETER und MÖLLER, RALF: *Simulation of Sparse Random Networks on a CNAPS SIMD Neurocomputer*. In: SMITH, LESLIE S. und HAMILTON, ALISTER (Herausgeber): *Neuromorphic Systems – Engineering Silicon From Neurobiology*, Nummer 10 in *Progress in Neural Processing*, Seiten 251–260. World Scientific, 1998. Zitiert auf Seite 133.
- /Pom1996/ POMIERSKI, TORSTEN: *Neurophysiologisch motivierte Architektur zur Erzeugung stabiler Farb- und Texturrepräsentationen*. Doktorarbeit, Technische Universität Ilmenau, April 1996. Zitiert auf den Seiten xiii, xxi, 26, 115 und 116.
- /POP1998/ PAPAGEORGIOU, CONSTANTINE P.; OREN, MICHAEL und POGGIO, TOMASO: *A General Framework for Object Detection*. In: *Proceedings of the International Conference on Computer Vision (ICCV'98)*, Seiten 555–562, Januar 1998. Zitiert auf den Seiten 8 und 51.

- /Pöt1994/ PÖTZSCH, MICHAEL: *Die Behandlung der Wavelet-Transformation von Bildern in der Nähe von Objektkanten*. Diplomarbeit (IR-INI 94-04), Ruhr-Universität Bochum, Institut für Neuroinformatik, Mai 1994. Zitiert auf Seite 55.
- /PR1981/ POLLEN, DANIEL A. und RONNER, STEVEN F.: *Phase Relationships between adjacent Simple Cells in the Visual Cortex*. Science, 212:1409–1411, Juni 1981. Zitiert auf Seite 55.
- /RBK1996/ ROWLEY, HENRY A.; BALUJA, SHUMEET und KANADE, TAKEO: *Human Face Detection in Visual Scenes*. In: TOURETZKY, DAVID S.; MOZER, MICHAEL C. und HASSELMO, MICHAEL E. (Herausgeber): *Advances in Neural Information Processing Systems 8 (NIPS'95)*, Seiten 875–881. The MIT Press, 1996. Zitiert auf den Seiten 37 und 45.
- /RBLM1997/ RAUTERBERG, MATTHIAS; BICHSEL, MARTIN; LEONHARDT, UWE und MEIER, MARKUS: *BUILT-IT: A Computer Vision-Based Interaction Technique of a Planning Tool for Construction and Design*. In: HOWARD, STEVE; HAMMOND, JUDY und LINDGAARD, GITTE (Herausgeber): *INTERACT'97 – Proceedings of the IFIP TC 13 International Conference on Human-Computer Interaction*, Seiten 587–588. Chapman & Hall, Juli 1997. Zitiert auf Seite 2.
- /Rit1991/ RITTER, HELGE: *Learning with the Self-Organizing Map*. In: KOHONEN, T.; MÄKISARA, K.; SIMULA, O. und KANGAS, J. (Herausgeber): *Artificial Neural Networks – Proceedings of the 1991 International Conference on Artificial Neural Networks (ICANN'91)*, Band 1, Seiten 357–364. Elsevier Science Publishing Company B. V., 1991. Zitiert auf Seite xx.
- /Rit1992/ RITTER, WERNER: *Traffic Sign Recognition in Colour Image Sequences*. In: *Proceedings of the Intelligent Vehicle Symposium '92 (Detroit)*, Seiten 12–17, 1992. Zitiert auf Seite 31.
- /Rit1997/ RITTER, WERNER: *Automatische Verkehrszeichenerkennung*, Band 5 der Reihe *Koblenzer Schriften zur Informatik*. Verlag Dietmar Fölbach, Doktorarbeit an der Universität Koblenz-Landau, 1997. Zitiert auf Seite 31.
- /RMG1998/ RAJA, YOGESH; MCKENNA, STEPHEN J. und GONG, SHAO GANG: *Tracking and Segmenting People in Varying Lighting Conditions using Colour*. In: *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, Seiten 228–233. IEEE Computer Society, April 1998. Zitiert auf Seite 8.

- /Ros1962a/ ROSENBLATT, FRANK: *Strategic Approaches to the Study of Brain Models*. In: FOERSTER, H. v. (Herausgeber): *Principles of Self-Organization*, Seite 387 ff. Pergamon Press, Elmsford, New York, 1962. Zitiert auf Seite xx.
- /Ros1962b/ ROSENBLATT, FRANK: *Principles of Neurodynamics*. Spartan Books, New York, 1962. Zitiert auf Seite xx.
- /Ros1984/ ROSENFELD, AZRIEL: *Multiresolution Image Processing and Analysis*, Band 12 der Reihe *Springer Series in Information Science*. Springer-Verlag, 1984. Zitiert auf den Seiten 7 und 13.
- /San1989/ SANGER, TERRENCE DAVID: *Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network*. Neural Networks: Off. Journal of INNS, ENNS and JNNS, 2(6):459–473, 1989. Zitiert auf Seite 119.
- /Sch1920/ SCHRÖDINGER, ERWIN: *Grundlinien einer Theorie der Farbmatrik im Tagessehen*. Annalen der Physik (IV), 63:397–520, 1920. Zitiert auf Seite 23.
- /Sch1953/ SCHMIDT, HARRY: *Einführung in die Vektor- und Tensorrechnung*. Verlag Technik Berlin, 1953. Zitiert auf Seite 140.
- /Sch1977/ SCHÜRMANN, JÜRGEN: *Polynomklassifikatoren für die Zeichenerkennung*. Oldenbourg-Verlag, München, 1977. Zitiert auf Seite 31.
- /SEN1998/ STEFFENS, JOHANNES; ELAGIN, EGOR und NEVEN, HARTMUT: *PersonSpotter – Fast and Robust System for Human Detection, Tracking and Recognition*. In: *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, Seiten 516–521. IEEE Computer Society, April 1998. Zitiert auf den Seiten 8 und 39.
- /SG1997/ STEPHAN, VOLKER und GROSS, HORST-MICHAEL: *Forminhalte sequentielle visuelle Aufmerksamkeit in columnar organisierten neuronalen Feldern*. In: PAULUS, ERWIN und WAHL, FRIEDRICH M. (Herausgeber): *Mustererkennung 1997 – 19. DAGM-Symposium (Braunschweig)*, Informatik aktuell, Seiten 411–418. Springer-Verlag, 1997. Zitiert auf Seite 75.
- /SK1987/ SIROVICH, LAWRENCE und KIRBY, M.: *Low-Dimensional Procedure for the Characterization of Human Faces*. Journal of the Optical Society of America A, 4(3):519–524, März 1987. Zitiert auf Seite 121.
- /SKJ1997/ SCHARR, STEFAN; KÖRGEL, HANNO und JÄHNE, BERND: *Numerische Isotropieoptimierung von FIR-Filtern mittels Querglättung*. In: PAULUS, ERWIN und WAHL, FRIEDRICH M. (Herausgeber): *Mustererkennung 1997 – 19. DAGM-Symposium (Braunschweig)*, Informatik aktuell, Seiten 367–374. Springer-Verlag, 1997. Zitiert auf Seite 59.
- /SSMB1999/ SIM, TERENCE; SUKTHANKAR, RAHUL; MULLIN, MATTHEW D. und BALUJA, SHUMEET: *High-Performance Memory-based Face Recognition for*

- Visitor Identification*. Technischer Bericht JPRC-TR-1999-01, Just Research, Pittsburgh, PA, Juli 1999. Zitiert auf den Seiten 39 und 45.
- /TDA1998/ TERRILLION, JEAN-CHRISTOPHE; DAVID, MARTIN und AKAMATSU, SHIGERU: *Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments*. In: *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, Seiten 112–117. IEEE Computer Society, April 1998. Zitiert auf Seite 8.
- /TM1998/ TRIESCH, JOCHEN und MALSBURG, CHRISTOPH V. D.: *A Gesture Interface for Human-Robot-Interaction*. In: *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, Seiten 546–551. IEEE Computer Society, April 1998. Zitiert auf Seite 38.
- /TP1991/ TURK, MATTHEW und PENTLAND, ALEX: *Eigenfaces for Recognition*. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991. Zitiert auf den Seiten 37, 39, 119 und 122.
- /Tre1986/ TREISMAN, ANNE: *Properties, Parts, and Objects*. In: BOFF, KENNETH R.; KAUFMAN, LLOYD und THOMAS, JAMES P. (Herausgeber): *Handbook of Perception and Human Performance*, Band II - Cognitive Processes and Performance: Perceptual Organization and Cognition, Kapitel 35, Seiten 35-1–35-70. Wiley, New York, 1986. Zitiert auf Seite 22.
- /Tre1988/ TREISMAN, ANNE: *Features and Objects: The Fourteenth Bartlett Memorial Lecture*. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, 40A(2):201–237, 1988. Zitiert auf Seite 22.
- /Tri1999/ TRIESCH, JOCHEN: *Vision-Based Robotic Gesture Recognition*. Doktorarbeit, Ruhr-Universität Bochum, Mai 1999. Zitiert auf den Seiten 38 und 67.
- /Vei1997/ VEIT, DANIEL: *Existenz und Stabilität von lokalen Anregungen in homogenen neuronalen Feldern*. Seminararbeit, Justus-Liebig-Universität Gießen, Webpage <http://www.uni-giessen.de/~gcj7/seminare/seminare.html>, 1997. Zitiert auf Seite 76.
- /VF1992/ VIENNET, EMMANUEL und FOGELMAN SOULIÉ, FRANCOISE: *Scene Segmentation using Multiresolution Analysis and MLP*. In: ALEKSANDER, IGOR und TAYLOR, JOHN (Herausgeber): *Artificial Neural Networks – Proceedings of the 1992 International Conference on Artificial Neural Networks (ICANN'92)*, Band 2, Seiten 1599–1602. Elsevier Science Publishing Company B. V., 1992. Zitiert auf Seite 15.

- /WADP1997/ WREN, CHRISTOPHER RICHARD; AZARBAYEJANI, ALI; DARRELL, TREVOR und PENTLAND, ALEX PAUL: *Pfinder: Real-Time Tracking of the Human Body*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 19(7):780–785, Juli 1997. Zitiert auf den Seiten 8 und 68.
- /WD1985/ WEBSTER, MICHAEL A. und DE VALOIS, RUSSELL L.: *Relationship between Spatial-Frequency and Orientation Tuning of Striate-Cortex Cells*. Journal of the Optical Society of America A, 2:1124–1132, 1985. Zitiert auf Seite 54.
- /Wen1993/ WENDT, DAG: *Szenenanalyse mit selbstorganisierenden Merkmalsdetektoren unter Verwendung neuronaler Netzwerkstrukturen*. Diplomarbeit, Technische Universität Ilmenau, 1993. Inv.-Nr. 200-93D-91. Zitiert auf Seite 119.
- /Wer1974/ WERBOS, PAUL J.: *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Doktorarbeit, Harvard University, 1974. Zitiert auf Seite xx.
- /Wer1984/ WERNERS, BRIGITTE: *Interaktive Entscheidungsunterstützung durch ein flexibles mathematisches Programmiersystem*. Doktorarbeit, Universität Aachen, 1984. Zitiert auf Seite 72.
- /Wer1988/ WERNERS, BRIGITTE: *Aggregation Models in Mathematical Programming*. In: MITRA, G. (Herausgeber): *Mathematical Models for Decision Support*, Seiten 295–305, 1988. Zitiert auf Seite 72.
- /WFKM1997/ WISKOTT, LAURENZ; FELLOUS, JEAN-MARC; KRÜGER, NORBERT und MALSBURG, CHRISTOPH V. D.: *Face Recognition by Elastic Bunch Graph Matching*. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 19(7):775–779, Juli 1997. Zitiert auf Seite 38.
- /Wil1998/ WILHELM, TORSTEN: *Untersuchungen zu neuronalen Felddynamiken*. Studienjahresarbeit, Technische Universität Ilmenau, Fachgebiet Neuroinformatik, Juni 1998. Zitiert auf den Seiten 74, 81 und 82.
- /Wol1988/ WOLF, GERALD (Herausgeber): *BI-Lexikon Neurobiologie*. Bibliographisches Institut Leipzig, 1988. Zitiert auf Seite xxi.
- /YLW1997/ YANG, JIE; LU, WEIER und WAIBEL, ALEX: *Skin-Color Modeling and Adaptation*. Technischer Bericht CMU-CS-97-146, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, Mai 1997. Zitiert auf den Seiten 27 und 31.
- /You1986/ YOUNG, RICHARD A.: *Simulation of Human Retinal Function with the Gaussian Derivative Model*. In: *Proceedings of the IEEE Computer Socie-*

- ty Conference on Computer Vision and Pattern Recognition (CVPR'86)*, Seiten 564–569. IEEE Computer Society, 1986. Zitiert auf Seite 54.
- /You1987/ YOUNG, RICHARD A.: *The Gaussian Derivative Model for Spatial Vision: I. Retinal Mechanisms*. *Spatial Vision*, 2(4):273–293, 1987. Zitiert auf Seite 54.
- /YW1995/ YANG, JIE und WAIBEL, ALEX: *Tracking Human Faces in Real-Time*. Technischer Bericht CMU-CS-95-210, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, November 1995. Zitiert auf den Seiten 27 und 31.
- /YW1996/ YANG, JIE und WAIBEL, ALEX: *A Real-Time Face Tracker*. In: *Proceedings of the Third IEEE Workshop on Applications of Computer Vision (WACV'96)*, Seiten 142–147, Dezember 1996. Zitiert auf den Seiten 27 und 31.
- /ZH1996/ ZELINSKY, ALEXANDER und HEINZMANN, JOCHEN: *Real-Time Visual Recognition of Facial Gestures for Human-Computer Interaction*. In: *Proceedings of the Second IEEE International Conference on Automatic Face- and Gesture-Recognition (FG96)*, Seiten 351–356. IEEE Computer Society, Oktober 1996. Zitiert auf Seite 2.
- /Zim1991/ ZIMMERMANN, HANS-JÜRGEN: *Fuzzy Set Theory – and its Applications*. Kluwer Academic Publishers, 1991. Zitiert auf den Seiten 71 und 72.
- /ZZ1980/ ZIMMERMANN, HANS-JÜRGEN und ZYSNO, PETER: *Latent Connectives in Human Decision Making*. *Fuzzy Sets and Systems*, 4:37–51, 1980. Zitiert auf Seite 72.

Index

A

Aliasing 14 ff.
 Auflösungspyramide ... 13, 15, 19, 52,
 89 ff., 120

B

Backpropagation xx
 BP *siehe* Backpropagation

C

CCD xix, xxii
 CIE xix, xxi, xxii, 24, 26, 28, 61
 CNAPS *siehe* Hardware $\hookrightarrow$
 Parallelrechner CNAPS
 CPL xix, 134

D

DPR *siehe* Repräsentation $\hookrightarrow$ diskrete
 parametrische
 DSP *siehe* Signalprozessor $\hookrightarrow$ digitaler

E

EBU xix, xxi, 25, 27
 Eigenfaces 37, 119–129
 Eigenwertproblem 140
 Elementarfarbraum .. *siehe* Farbraum
 $\hookrightarrow RG/BY/WS$

Entropie 83, 84

F

Farbkonstanz 22, 28, 113 ff.
 referenzfreie 26, 113

Farbraum

$RG/BY/WS$ xxi, 26, 115, 117
 RGB xxi, 27
 RGB_{EBU} xxi, 25, 26, 116, 117
 RGB_{FCC} xxi, 116
 XYZ xxii, 22, 24

Farbwertanteil .. 23, 24, 27, 28, 31, 33,
 35

FCC xix, xxi, 27

Feld

dynamisches neuronales ... 65, 69,
 70, 74–84, 86, 89 ff., 107

FFT xix, 131

Filterarrangements 52–54

Fourierraum 57, 58, 137, 143

Fouriertransformation xix, 57, 58, 131,
 142, 143

Framegrabber *siehe* Hardware $\hookrightarrow$
 Framegrabber ITI IC-PCI & AM-
 CLR

Fuzzy-Operator

Min-Max 71–73

G

Glättung 14 ff.

GNG xix, 30, 37, 120

H

Hardware

Framegrabber ITI IC-PCI & AM-
CLR 117,
134

Mobiles Robotersystem MILVA
xx, 1, 4, 6

Parallelrechner CNAPS .. xix, 37,
132–135

Videokamera VC-C1 6, 25

Hauptachsentransformation 127

HMM xix, 4

Hyperkolumne 50

I

Intensität 27, 115

Interpolation 17 ff.

K

KLT xx, 119

KNN xx, 66, 132

Konturmodellierung 51 ff.

Konvexkombination 73

L

LLM xx, 31, 66

LUT xx, 31, 67

M

Mensch-Maschine-Interaktion xx, 1 ff.,
109

MILVA ... *siehe* Hardware $\leftrightarrow$ Mobiles
Robotersystem MILVA

MLP xx, 66, 132

MMI *siehe*
Mensch-Maschine-Interaktion

MMX xxi, 133

N

Nachbarschaft

lokale 137 ff.

NOA xxi, 132

Normalbeobachter xxii

Normfarbtafel 24, 26

Normvalenzsystem *siehe* Farbraum $\leftrightarrow$
XYZ

O

Orientierung 54

Orientierungsfilter 54–61

Tensormethode ... 56–61, 137–143

Trägheitstensoranalogie 142–143

Umsetzung 57–61

Orientierungsvektor 54

 mittlerer 140

ORL Database of Faces ... xxi, 40, 41

P

PCA xxi, 119, 127

PCI xxi, 132

PN xxi, 133

Primärvalenz xxi, xxii, 23–25

Primal Sketch 49

Q

Quadraturfilter 55, 56

R

RBF xxi, 4

Relaxation 69, 70, 75–79, 83

Repräsentation

 diskrete parametrische xix, 50

 pyramidale 13–19

Retinex-Theorie 113

Rezeptives Feld xxi, 54, 79, 80

RF *siehe* Rezeptives Feld

Richtung 54

S

Signalprozessor

 digitaler xix, 132

SIMD xxi, 132

Spektralwertkurve 23 ff.

T

Trägheitstensoranalogie *siehe* Orientierungsfilter $\leftrightarrow$ Tensormethode $\leftrightarrow$ Trägheitstensoranalogie

U

Unterabtastung 17 ff.

V

VC-C1 *siehe* Hardware $\leftrightarrow$ Videokamera VC-C1

Verlustfunktion

 lineare 75

 nichtlineare 74

W

Winner-Take-All .. xxii, 69, 75, 76, 78, 82, 84, 85, 91

WTA *siehe* Winner-Take-All