

# **Biologienahe Verarbeitung visueller Signale zur robusten Extraktion von Konturinformationen**

Zur Erlangung des akademischen Grades

**DOKTORINGENIEUR (Dr.-Ing.)**

der Fakultät für Elektrotechnik, Informatik und Mathematik

Universität Paderborn

genehmigte Dissertation

von

**Dipl.-Ing. Jörg Thiem**

aus Hamm

Referent: Prof. Dr. rer. nat. Georg Hartmann

Korreferent: Prof. Dr.-Ing. Bernd Henning

Tag der mündlichen Prüfung: 19. Februar 2004

Paderborn, den 13. April 2004

D 14/192









# Danksagung

Die vorliegende Arbeit entstand während meiner Zeit im Graduiertenkolleg „Parallele Rechnernetzwerke in der Produktionstechnik“ am Heinz Nixdorf Institut und im Fachgebiet „Grundlagen der Elektrotechnik“ der Universität Paderborn in den Jahren 1997 bis 2003. An dieser Stelle bedanke ich mich bei all jenen, die mich – in welcher Form auch immer – bei der Anfertigung der Arbeit unterstützt haben.

Herrn Professor Dr. Georg Hartmann gilt mein ganz besonderer Dank für diese äußerst interessante und interdisziplinäre Themenstellung und den hierfür notwendigen Freiraum. Seine stetige Diskussionsbereitschaft und seine vielen Anregungen und Betrachtungsweisen trugen wesentlich zum Gelingen der vorliegenden Arbeit bei. Es war für mich sowohl fachlich als auch persönlich äußerst wertvoll diesen Lebensabschnitt in seinem Fachgebiet verbracht zu haben.

Herrn Professor Dr.-Ing. Bernd Henning danke ich für die freundliche Übernahme des Korreferates, der damit verbundenen Mühe und das an der Arbeit gezeigte Interesse. Herrn Professor Dr.-Ing. Klaus Meerkötter möchte ich ebenfalls für die nützlichen fachlichen Diskussionen und das Interesse danken.

Für den uneingeschränkten Einsatz im Fachgebiet und das ausgezeichnete Klima möchte ich Herrn Dr.-Ing. Siegbert Drüe danken. Er begleitete mich in vielen fachlichen Gesprächen und hatte stets ein offenes Ohr.

Allen Kollegen und Studenten des Fachgebietes möchte ich ebenfalls für die freundliche Atmosphäre und den unvergleichlichen Zusammenhalt danken. Hervorheben möchte ich Nan-Shan Chen, Jian Cheng, Nicolai Götze, Martin Schäfer, Ralf Stemmer und Joachim Waßmuth, die mich durch wertvolle Diskussionen bzw. das Korrekturlesen beim Anfertigen der Dissertationsschrift unterstützt haben.

Ferner danke ich meinem Diplomanden Thomas Eisenbach, der mir durch sein engagiertes Mitwirken in der Schlußphase dieser Arbeit und sein schier unerschöpfliches Interesse die „Puste“ für den Endspurt gegeben hat.

Meiner Frau Tanja schließlich gilt mein herzlichster Dank. Nur durch Ihre Geduld und Ihre Liebe gab Sie uns die nötige Kraft in den letzten Jahren.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Biologische Sehsysteme . . . . .	1
1.2	Robuste Bildverarbeitung . . . . .	2
1.3	Ziel der Arbeit . . . . .	3
1.4	Gliederung und Abgrenzung . . . . .	4
<b>2</b>	<b>Mehrdimensionale Digitale Signalverarbeitung</b>	<b>7</b>
2.1	Mehrdimensionale Signale und Systeme . . . . .	8
2.1.1	Nichtrekursive Digitalfilter . . . . .	12
2.1.2	Rekursive Digitalfilter . . . . .	13
2.2	Abtastung zweidimensionaler Signale . . . . .	15
2.2.1	Umrechnung auf ein hexagonales Abtastraster . . . . .	18
2.2.2	Eine Strategie zur effizienten Interpolation . . . . .	21
2.2.3	Experimentelle Ergebnisse . . . . .	23
2.3	Biologische Verarbeitungssysteme . . . . .	28
2.3.1	Neurophysiologische Verarbeitung visueller Signale . . .	29
2.3.2	Künstliche neuronale Netzwerke . . . . .	35

---

<b>3</b>	<b>Lineares Modell der frühen visuellen Verarbeitung</b>	<b>39</b>
3.1	Bildgewinnung in einem biologischen Sehsystem . . . . .	40
3.1.1	Abbildungseigenschaften des Auges . . . . .	40
3.1.2	Abtastung mittels Photorezeptoren . . . . .	45
3.2	Modellierung retinaler Neuronen . . . . .	48
3.2.1	Photorezeptoren . . . . .	49
3.2.2	Horizontal- und Ganglienzellen . . . . .	51
3.3	Modellierung kortikaler simpler Neuronen . . . . .	59
3.4	Eigenschaften der Modellierung . . . . .	69
3.5	Reduktion der Netzstruktur . . . . .	73
<b>4</b>	<b>Realisierung pulscodierter Mechanismen</b>	<b>79</b>
4.1	Modellneuron von ECKHORN . . . . .	80
4.2	Erweiterung des linearen Modells . . . . .	83
4.3	Experimentelle Ergebnisse einer Neurocomputerimplementierung	85
4.4	Modifikation des ECKHORN-Modellneurons . . . . .	87
<b>5</b>	<b>Adaption bei variierenden Beleuchtungsbedingungen</b>	<b>93</b>
5.1	Verfahren zur Rauschunterdrückung . . . . .	93
5.2	Rauschprozesse bei der Bildakquisition . . . . .	97
5.3	Adaptive Filterung zur Rauschunterdrückung . . . . .	100
5.3.1	SNR einer gefilterten Szene . . . . .	101
5.3.2	Konstruktion eines Steuermechanismus . . . . .	106
5.3.3	Vergleich mit dem biologischen Vorbild . . . . .	110
5.4	Erweiterungen für eine adaptive Verarbeitung . . . . .	111
5.4.1	Einfluß auf nachgeschaltete Verarbeitungsschichten . . . .	112
5.4.2	Modellierung kortikaler Verknüpfungsstrukturen . . . . .	114

---

5.4.3	Rekursive Strukturen retinaler Neuronen . . . . .	124
<b>6</b>	<b>Experimentelle Ergebnisse</b>	<b>127</b>
6.1	Verarbeitung realer Kameradaten . . . . .	128
6.1.1	Meßbilder zur Schätzung des SNR . . . . .	128
6.1.2	Kompensation des AGC . . . . .	129
6.1.3	Kompensation des FPN . . . . .	131
6.2	Schätzung der Rauschparameter . . . . .	132
6.3	Adaptive Filterung . . . . .	134
6.3.1	Verbesserung des SNR . . . . .	134
6.3.2	Robuste Konturextraktion . . . . .	137
6.4	Fazit . . . . .	141
<b>7</b>	<b>Zusammenfassung und Ausblick</b>	<b>143</b>
	<b>Literaturverzeichnis</b>	<b>147</b>



# Einleitung

## 1.1 Biologische Sehsysteme

Das Sehen ist für den Menschen eine so alltägliche Fähigkeit, daß wir uns gar nicht oder nur selten über die ihm zugrunde liegende Komplexität der neuronalen Verschaltungen und Prozesse bewußt werden. Aber „*unsere Intuition sagt uns, daß das Gehirn kompliziert sein muß. Wir tun komplizierte Dinge, und das in enormer Vielfalt*“ [Hub90]. Mittlerweile weiß man auch, daß das menschliche Gehirn in der Tat etwa  $10^{12}$  (Billionen) Nervenzellen besitzt, die auf geschickte Weise durch  $10^{14} \dots 10^{15}$  Verbindungen miteinander gekoppelt sind. Dieses Wissen ist das Ergebnis jahrzehntelanger neurophysiologischer Forschung – getrieben von dem Wunsch die Funktionsweise des menschlichen Sehsystems zu verstehen oder zumindest ansatzweise erklären zu können. So muß auch der Nobelpreisträger DAVID HUBEL zugeben, daß seine stärkste Motivation „... *die reine Neugier auf die Arbeitsweise der kompliziertesten Struktur ist, die wir kennen*“ [Hub90]. Aber nicht nur das reine Erlangen von Erkenntnissen über die Funktionsweise unseres Gehirns, sondern auch die Möglichkeiten für Mediziner und Biologen, neurologische und psychiatrische Krankheiten zu heilen oder zu verhüten, ist überzeugender Motor neurologischer Studien. Damit ist die Erforschung des menschlichen Gehirns sicher eine der ehrgeizigsten Bestrebungen seit jeher und scheint nichts von seiner Faszination zu verlieren.

Trotz der erwähnten Komplexität ist zumindest das visuelle System höher entwickelter Säugetiere einer der am besten verstandenen Teile des Gehirns. Hier kommt der Wissenschaft entgegen, daß auf das Auge projizierte Lichtreize bestimmte Antwortaktivitäten in gewissen Hirnbereichen hervorrufen. Werden diese meßtechnisch erfaßt, so können daraus funktionale Zusammenhänge über den Aufbau der neuronalen Verschaltung abgeleitet werden. Modellierungen der visuellen Verarbeitung scheinen daher recht erfolgversprechend. Das menschliche

Sehsystem ist zudem offensichtlich in der Lage, relevante Informationen auf äußerst effiziente Weise aus natürlichen Szenen zu extrahieren, um hiermit eine robuste Repräsentation und Objekterkennung durchzuführen. Die Kenntnisse über neuronale Verarbeitungsstrukturen sind daher, wie sich vielfach gezeigt hat, auch für technische Sehsysteme von großem Interesse. Biologisch motivierte Methoden spielen daher zunehmend eine bedeutende Rolle in der industriellen Bildverarbeitung [EPdRH02]. Das ehrgeizige Ziel ist, zumindest Teile der biologischen Mechanismen zu beschreiben und für technische Bilderkennungsaufgaben nutzbar zu machen. Darüber hinaus können solche Modellierungen und deren Simulation auf leistungsfähiger Hardware auch dazu beitragen, biologische Abläufe besser zu verstehen.

## 1.2 Robuste Bildverarbeitung

Die klassische Bildverarbeitung ist Gegenstand der benachbarten Disziplinen Mathematik, Informatik und der Ingenieurwissenschaften, wobei die Schwerpunkte und die Terminologie unter Umständen verschieden sein können. Hierbei ist der Zweck vieler bisheriger und aktueller Forschungsaufgaben die effiziente und effektive Verarbeitung von Kameradaten einer betrachteten Szene für die unterschiedlichsten Anwendungsgebiete. So ist die robuste Extraktion von Objektmerkmalen in einer betrachteten Bildszene, insbesondere Linien- und Konturinformationen, offenbar stets ein wichtiges Thema vieler Forschungsgruppen. Denn nur durch solche robusten Merkmale einer komplexen Szene sind ehrgeizige Systeme realisierbar, welche z. B. durch eine automatische Objekterkennung einer Maschine eine begrenzten visuelle Wahrnehmung verleihen. Ein Beispiel sei hier das in der Arbeitsgruppe HARTMANN entwickelte, visuell gesteuerte Robotersystem zur automatisierten Demontage von Altfahrzeugen [STDH98, GDH98]. So steht auch in dieser Arbeit die robuste Extraktion von Konturmerkmalen im Vordergrund.

Ein Verfahren soll hier robust genannt werden, wenn es unempfindlich gegenüber äußeren Veränderungen oder Störeinflüssen ist, wie z. B. einer schlechten Beleuchtung und dem damit verbundenen Bildrauschen. Verwendet z. B. eine automatische Objekterkennung die extrahierten Objektkonturen aus einer Bildszene, so ist meist die Objekterkennung selbst nur robust, wenn die Konturen ebenfalls robust extrahiert wurden. Für die Gewinnung stabiler Objektkonturen benötigt man folglich die unverzichtbare Vorverarbeitung, welche die Bilddaten von evtl. Störungen im Vorfeld bereinigt. Erst durch diese Maßnahmen können robuste intelligente Systeme entwickelt werden. Exemplarisch sei an dieser Stelle die Unterdrückung bzw. Minderung von Störungen bei bildgebenden Verfahren für die



medizinische Diagnostik genannt. Hier erfolgt nicht unbedingt eine automatisierte Auswertung der Daten durch einen Computer. Der Mediziner wird hier die Diagnose stellen. Allerdings kann es dennoch nützlich sein, die Bilddaten zuvor vom Computer aufbereiten zu lassen und in einer verbesserten Qualität dem Mediziner für die weitere Auswertung zur Verfügung zu stellen. Die Erforschung und die Entwicklung neuer Verfahren zur Rauschunterdrückung scheint daher stets ein aktuelles Thema zu sein.

Die Entwicklung leistungsfähiger Algorithmen für die Bildverarbeitung erfordert aber i. d. R. ein solides theoretisches Fundament, welches durch die (Bild-) Signalverarbeitung geschaffen wird. Durch diese meist theoretische Beschreibung der zu manipulierenden Bilddaten und den gewünschten Funktionen können dann z. B. sogenannte Operatoren oder Digitalfilter entwickelt werden, die auf Standardrechnern eingesetzt werden können. Als Werkzeug in der Bildverarbeitung ist eine detaillierte Kenntnis über den theoretischen Hintergrund oft nicht mehr notwendig. Neben der klassischen Signalverarbeitung konnte sich auch die Theorie der neuronalen Netze, deren Ursprünge in der Informatik liegen, für den Einsatz in der Bildverarbeitung durchaus etablieren. Man erhofft sich hierdurch, wie oben erwähnt, wichtige vorteilhafte Eigenschaften und Strukturen des überlegenen biologischen Vorbildes nutzbar machen zu können. Somit ist sowohl mittels der Digitalfilter einerseits als auch mittels neuronaler Netze andererseits eine Verarbeitung von Bilddaten möglich. Die Verwandtschaft dieser beiden Herangehensweisen verdeutlicht das folgende Kapitel, in dem die Analogie zwischen bestimmten Digitalfiltern und künstlichen neuronalen Netzen aufgezeigt wird. Daraus können sich Vorteile bei der Entwicklung neuronaler Netze ergeben, wenn man diese z. B. als Digitalfilter interpretiert. So wird im Rahmen der Arbeit konsequent sowohl die systemtheoretische als auch die neuronale Terminologie parallel verwendet, um jeweils die Vorteile beider Betrachtungsweisen auszunutzen.

## 1.3 Ziel der Arbeit

Ziel der Arbeit ist eine systemtheoretische Beschreibung fundamentaler Verarbeitungspfade höher entwickelter biologischer Sehsysteme. Hier scheint es sehr attraktiv und lehrreich, Hypothesen zu biologischen Strukturen systemtheoretisch zu modellieren und vorteilhafte Strukturen in geeigneter Form technisch umzusetzen. Diese können dann einen Beitrag zum Verständnis der neurologischen Mechanismen liefern oder als Grundlage für neurophysiologische Simulationen und Experimente dienen, wobei die positiven Eigenschaften des biologischen Vorbildes hinsichtlich der Extraktion von orientierten Konturmerkmalen aufgezeigt werden. Da die Photorezeptoren des menschlichen Auges im Bereich des schärf-

sten Sehens eine hexagonale Anordnung aufweisen, wie in [Yel81] gezeigt, bildet solch ein hexagonales Abtastraster die grundlegende Basis für die dargestellten Betrachtungen. Diese Form der Signalabtastung liefert eine überlegene Symmetrie, eine einfache, wohl definierte Nachbarschaftsbeziehung und eine Datenreduktion im Vergleich zu einer gewöhnlichen Repräsentation unter Einhaltung des Abtasttheorems von SHANNON. Des weiteren werden aus der biologischen Modellierung Algorithmen für technische Bildverarbeitungssysteme und für eine Spezialhardware, die der Simulation großer neuronaler Netze dient, abgeleitet. Um robuste Merkmale aus den Bilddaten zu extrahieren, werden schließlich Mechanismen untersucht, die eine adaptive Rauschunterdrückung ermöglichen.

## 1.4 Gliederung und Abgrenzung

In Kapitel 2 folgen zunächst die systemtheoretischen Grundlagen für eine allgemein mehrdimensionale Signalverarbeitung, soweit sie für das Verständnis der Ausführungen notwendig sind. Ein Abschnitt widmet sich insbesondere der hexagonalen Abtastung und Verarbeitung, da dies die wesentliche Eigenschaft aller entwickelten Strukturen und Algorithmen dieser Arbeit ist. Darauf werden die neurophysiologischen Zusammenhänge und die formale Analogie zwischen künstlichen neuronalen Netzen und Digitalfiltern bzgl. der Ortskoordinaten der Bilddaten beschrieben.

Das Kapitel 3 stellt ein massiv paralleles Modell der frühen visuellen Verarbeitung im menschlichen Sehsystem vor, welches orientierte Konturmerkmale in einem Bild extrahiert. Das beschriebene Modell stellt eine Hypothese auf, welches zu erklären vermag, wie die Zellen in der Netzhaut (Retina) und in der Großhirnrinde (Cortex), möglicherweise in einer kaskadierten Struktur von Neuronenschichten organisiert sein könnten. Darüber hinaus werden die Vorteile der Architektur für biologische und technische Sehsysteme herausgearbeitet.

Existierende Ansätze für ähnliche Zielsetzungen betrachten meist nur sehr eingeschränkte Bereiche, wie z.B. die Retina, oder sind nicht systemtheoretisch fundiert. Unter diesen sei die recht detaillierte Abhandlung über die Netzhaut des Primaten [SL96a, SL96b] hervorzuheben, welche sowohl örtliche Filtercharakteristika als auch das zeitliche Verhalten berücksichtigt. Einige grundlegende Ideen wurden hieraus für die vorliegende Arbeit übernommen, um eine neue Hypothese der biologischen Signalverarbeitung von der Abtastung durch die Rezeptoren bis hin zu den Verknüpfungsschemata im visuellen Cortex aufzubauen. Im Unterschied zu anderen retinalen Modellen wird hierbei ein hexagonales Abtastraster der Photorezeptoren auf der Netzhaut biologisch und signaltheoretisch motiviert, welches

somit ein wesentliches Merkmal der vorliegenden Arbeit darstellt. Die Verbindungsgewichte der retinalen Neuronenstruktur werden dann anhand von neurophysiologischen Fakten durch ein Approximationsproblem gefunden. Für die Zusammenhänge in der Großhirnrinde werden im wesentlichen die anerkannten Forschungsarbeiten von POLLEN und RONNER [PR81, PR82, PR83] zugrunde gelegt, welche die Existenz sogenannter *simpler Zellen* im *primären visuellen Cortex* nachgewiesen haben. Die rezeptiven Felder dieser Neuronen können näherungsweise durch GABOR-Funktionen [Gab46] beschrieben werden und weisen ein signifikantes Antwortverhalten auf orientierte Linien oder Kanten auf (siehe auch [Dau80, Mar80, KMB82, JP87]). Aus diesen Erkenntnissen heraus hat sich eine neue Klasse der Digitalfilter, die GABOR-Filter, etabliert, deren Vorteile für viele Bildverarbeitungsaufgaben genutzt werden. Die Realisierungen bei technischen Systemen, wie z.B. [Tra98, LVB<sup>+</sup>93, WFKvdM97, Wis97, Bie87, SvdM97], finden allerdings weder auf einem hexagonalen Abtastschema, noch in Verbindung mit neuronalen Vorverarbeitungsstufen, wie sie in der Retina zu finden sind, statt.

Die theoretisch fundierte Arbeit von SABATINI [Sab96] sollte in diesem Zusammenhang nicht ungenannt bleiben, in der ein Modell simpler Neuronen auf der Basis interkortikaler inhibitorischer Kopplungen vorgestellt wird. Die rezeptiven Felder von Neuronen im seitlichen Kniehöcker (*Corpus geniculatum laterale*) sind hier allerdings der Einfachheit halber orientiert gaußförmig ausgeprägt. Es entsteht eine rekursiv gekoppelte Struktur simpler Neuronen, deren sich ergebenden rezeptiven Felder - je nach Umgebungseinfluß - den zweidimensionalen orientierungsselektiven GABOR-Funktionen sehr ähnlich sein können. Im Gegensatz zu einer realitätsnahen biologischen Modellierung zeigt die Studie vielmehr auf, welche rezeptiven Felder simpler Zellen mit der Hilfe interkortikaler Einflüsse entstehen können, wobei ein gaborähnliches Verhalten als Sonderfall betrachtet werden kann. Hier bleibt aber die berechtigte Frage offen, ob dieses Ergebnis immer noch gültig ist, wenn die rezeptiven Felder der Neuronen im seitlichen Kniehöcker Zentrum-Umfeld-Charakter haben, wie es bei den Gangliensignalen tatsächlich der Fall ist. Dennoch eröffnen die Betrachtungen rekursiver Strukturen neue Möglichkeiten, die es noch im Detail zu erforschen gilt. Die Interpretation als rekursive Digitalfilter können sicher auch hier hilfreich und ratsam sein.

In Kapitel 4 wird gezeigt, wie das zuvor entwickelte visuelle neuronale Netzwerk soweit erweitert werden kann, daß ab den Gangliensignalen in der Retina eine pulscodierte Verarbeitung stattfindet. Damit gewinnt die Modellierung weiter an Biologienähe. Die Informationsverarbeitung bis zu den Ganglienzellen ist meist durch kontinuierliche Signale modelliert, die die Erregungen der Zelle repräsentieren. Jedoch wird die Verarbeitung im Gehirn nach den Ganglienzellen als

pulscodiert angenommen. In einigen Untersuchungen konnte gezeigt werden, daß neben der Pulsrate dieser Signale auch insbesondere die Phaseninformation für visuelle Aufgaben genutzt wird [ERAD89, ERAD90]. Neuronen, welche ähnliche Merkmale repräsentieren wie z.B. eine kontinuierliche Linie in ihren rezeptiven Feldern, synchronisieren beispielsweise ihre Pulse. Dieser Mechanismus birgt potentielle Vorteile für viele Wahrnehmungsprozesse, wie der Objektsegmentation. Wegen dieser Tatsache wird in diesem Kapitel das vorhandene Netzwerk zu einem pulscodierten neuronalen Netzwerk ausgebaut. Die Ganglienzellen konvertieren dabei ihre Aktivität in eine Pulsrate und die simplen Zellen werden als pulscodierte ECKHORN-Neuronen entworfen [ERAD89, ERAD90]. Die auf diese Weise erzielte Kombination eines gaborähnlichen Verhaltens auf einem hexagonalen Abtastraster mit einer pulscodierten Informationsverarbeitung wurde ebenfalls auf einer Spezialhardwarearchitektur implementiert und simuliert. Die damit verbundenen Ergebnisse und Erkenntnisse werden hier ebenfalls dargestellt.

Wie eingangs erwähnt, rufen variierende Beleuchtungsbedingungen meist Schwierigkeiten hervor, wenn reale Szenen mit einem Digitalrechner verarbeitet werden sollen. Besonders Rauschprozesse, die mit der Bildentstehung selbst einhergehen, können bei dunklen Lichtverhältnissen zu fehlerhaften Resultaten führen. Eine automatische Auswertung oder Interpretation des Bildinhaltes durch ein technisches Sehsystem wird damit meist unmöglich. Das menschliche Sehsystem scheint jedoch in der Lage zu sein, sich auf diese Veränderungen in der wahrgenommenen Szene anzupassen und auf diese Weise die Qualität aller Verarbeitungsschritte zu gewährleisten – es findet eine robuste Verarbeitung statt. Obwohl dieser adaptive Vorverarbeitungsmechanismus in der Literatur vereinzelt erwähnt wird [SL96a, BC92], sind Untersuchungen oder technische Implementationen bisher ausgeblieben. In Kapitel 5 wird daher eine biologisch motivierte Vorverarbeitungsstrategie entwickelt, die sich an variierende Lichtverhältnisse adaptiert und auf diese Weise eine effektive Unterdrückung des Rauschens in Bilddaten erzielt. Wie weiter gezeigt wird, ist erst mit diesem Ansatz eine robuste Extraktion von Konturen möglich, die auch bei extrem dunklen und damit stark verrauschten Szenen zum Erfolg führt. Auf diesen Erkenntnissen aufbauend wird die Erweiterung des bisherigen Modells um diese adaptive Komponente untersucht und daraus eine alternative kortikale Verknüpfungsarchitektur abgeleitet.

Die Leistungsfähigkeit der in dieser Arbeit vorgestellten biologienahen Strukturen zur Bildvorverarbeitung wird in Kapitel 6 anhand realer Kameradaten demonstriert. Der Schwerpunkt wird hier insbesondere auf die adaptive Vorfilterung mit anschließender Konturextraktion gelegt.

# Mehrdimensionale Digitale Signalverarbeitung

In diesem Kapitel wird zunächst die Theorie mehrdimensionaler Signale und linearer Systeme behandelt, soweit es für das Verständnis der systemtheoretischen Betrachtungen dieser Arbeit wichtig ist. Im späteren Verlauf der Untersuchungen wird ggf. an verschiedenen Stellen auf diesen Grundlagen aufgebaut und die Thematik vertieft. Der Großteil der Arbeit bedient sich dieser fundierten Beschreibungsmethode, um verschiedene Bilddaten oder biologische visuelle Reize als ortsabhängige (ortsdiskrete) Signale auf beliebigen periodischen Abtastrastern aufzufassen und mit Digitalfiltern zu verarbeiten. Die wesentlichen Begriffe der nichtrekursiven und rekursiven mehrdimensionalen Digitalfilter müssen daher in diesem Kapitel ebenfalls kurz formuliert werden.

Des weiteren wird in einem nächsten Abschnitt insbesondere die hexagonale Abtastung eines zweidimensionalen (Bild-)Signals als das effizienteste periodische Raster näher betrachtet. Die mathematischen Zusammenhänge und die für technische und biologische Bildverarbeitungssysteme wichtigen Vorzüge werden dabei herausgestellt. Ein Schwerpunkt bildet hier die Umrechnung auf solch ein hexagonales Abtastraster, da üblicherweise die Bilddaten einer Kamera auf einem rechtwinkligen Raster (z.B. eines CCD oder CMOS-Chips) vorliegen. Durch experimentelle Untersuchungen werden verschiedene Verfahren zur Umrechnung vergleichend gegenübergestellt.

Für die wesentliche Zielsetzung dieser Arbeit, die systemtheoretische Beschreibung biologischer Sehsysteme, sind abschließend die Grundlagen der neurophysiologischen Informationsverarbeitung zu erörtern. Offensichtlich ist hier speziell die visuelle Wahrnehmung des Menschen oder anderer höher entwickelter Lebewesen von großem Interesse. Außerdem wird gezeigt, inwieweit künstliche neuro-

nale Netze als Digitalfilter interpretiert werden können, was für die Modellierung der biologischen Vorgänge und Architekturen im Laufe dieser Arbeit wertvolle Vorteile bietet.

## 2.1 Mehrdimensionale Signale und Systeme

Zu betrachten sind hier zunächst kontinuierliche Signale  $x_c(\mathbf{t})$  des Ortes  $\mathbf{t} \in \mathbb{R}^2$ , die mit Hilfe der *ortskontinuierlichen FOURIER-Transformation*

$$X_c(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} x_c(\mathbf{t}) e^{-j\boldsymbol{\omega}^T \mathbf{t}} d\mathbf{t} \quad (2.1)$$

$$x_c(\mathbf{t}) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} X_c(\boldsymbol{\omega}) e^{j\boldsymbol{\omega}^T \mathbf{t}} d\boldsymbol{\omega} \quad (2.2)$$

in der gewöhnlichen (*Orts-*)Frequenz  $\boldsymbol{\omega} \in \mathbb{R}^2$  dargestellt werden können<sup>1</sup>.

Erfolgt nun eine sogenannte *periodische Abtastung* des Signals in der Form<sup>2</sup>  $x(\mathbf{k}) = x_c(\mathbf{T}\mathbf{k})$  mit  $\mathbf{k} = [k_1 k_2]^T \in \mathbb{Z}^2$ , wobei hier  $\mathbf{T} = [\mathbf{t}_1 \mathbf{t}_2] \in \mathbb{R}^{2 \times 2}$  die *Abtastmatrix* sei, so verwenden wir die *ortsdiskrete FOURIER-Transformation*

$$X(\mathbf{v}) = \sum_{\mathbf{k}} x(\mathbf{k}) e^{-j\mathbf{v}^T \mathbf{k}} \quad (2.3)$$

$$x(\mathbf{k}) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} X(\mathbf{v}) e^{j\mathbf{v}^T \mathbf{k}} d\mathbf{v}. \quad (2.4)$$

Die Transformierte ist durch ihre Eigenschaft

$$X(\mathbf{v}) = X(\mathbf{v} + 2\pi\mathbf{n}) \quad , \forall \mathbf{n} \in \mathbb{Z}^2 \quad (2.5)$$

periodisch in der (normierten) Frequenz  $\mathbf{v} = [v_1 v_2]^T$ , so daß das Integrationsgebiet für die Rücktransformation z. B. zu  $-\pi \leq v_{1,2} \leq \pi$  gewählt werden kann. Wie in [DM84] nachgelesen werden kann, ist dann gemäß

$$X(\mathbf{v}) = \frac{1}{|\det \mathbf{T}|} \sum_{\boldsymbol{\mu}} X_c(\mathbf{T}^{-T}\mathbf{v} - \boldsymbol{\Omega}\boldsymbol{\mu}) \quad , \boldsymbol{\mu} \in \mathbb{Z}^2 \quad (2.6)$$

<sup>1</sup>Zur Abkürzung sollen hier Mehrfachintegrale vielfach nur durch die Dimension der Integrationsvariablen allein, z. B. durch  $\mathbf{t}$  bzw.  $\boldsymbol{\omega}$  gekennzeichnet und auf die Symbolik  $\int \int$  verzichtet werden, da die Betrachtungen meist nicht auf zweidimensionale Fälle beschränkt sind.

<sup>2</sup>Falls es aus dem Zusammenhang hervorgeht, kann der Index „c“ zur Abkürzung weggelassen werden, auch wenn der Ausdruck „ $x(\mathbf{k}) = x(\mathbf{T}\mathbf{k})$ “ natürlich mathematisch nicht korrekt ist.

die Transformierte des abgetasteten Signals eine periodische Fortsetzung der Transformatierten des kontinuierlichen Signals. Hierbei ist die Beziehung zwischen der gewöhnlichen und der normierten Frequenz

$$\mathbf{v} = \mathbf{T}^T \boldsymbol{\omega} \quad (2.7)$$

durch die Abtastmatrix  $\mathbf{T}$  gegeben. Die sogenannte *Kreisfrequenzmatrix*  $\boldsymbol{\Omega}$  ergibt sich dabei zu

$$\boldsymbol{\Omega} = 2\pi \mathbf{T}^{-T}, \quad \boldsymbol{\Omega} = [\boldsymbol{\omega}_1 \boldsymbol{\omega}_2] \in \mathbb{R}^{2 \times 2}. \quad (2.8)$$

Eine alternative Definition und ebenfalls gebräuchliche Form der diskreten Transformation lautet

$$\hat{X}(\boldsymbol{\omega}) = \sum_{\mathbf{k}} x(\mathbf{k}) e^{-j\boldsymbol{\omega}^T \mathbf{T} \mathbf{k}}, \quad (2.9)$$

so daß sich hier die Zusammenhänge zu

$$\hat{X}(\boldsymbol{\omega}) = X(\mathbf{T}^T \boldsymbol{\omega}) = \frac{1}{|\det \mathbf{T}|} \sum_{\boldsymbol{\mu}} X_c(\boldsymbol{\omega} - \boldsymbol{\Omega} \boldsymbol{\mu}) \quad (2.10)$$

ergeben.

In Abb. 2.1 ist die Verarbeitung eines kontinuierlichen Signals  $x_c(t)$  mit Hilfe eines diskreten Systems und anschließender Interpolation dargestellt. Wir beschränken uns hier auf lineare und ortsinvariante (engl. shift invariant) Systeme (LSI-Systeme), welche mit Hilfe einer *Impulsantwort*  $h(\mathbf{k})$  beschrieben werden können und nennen diese Klasse von Systemen auch *Digitalfilter*.

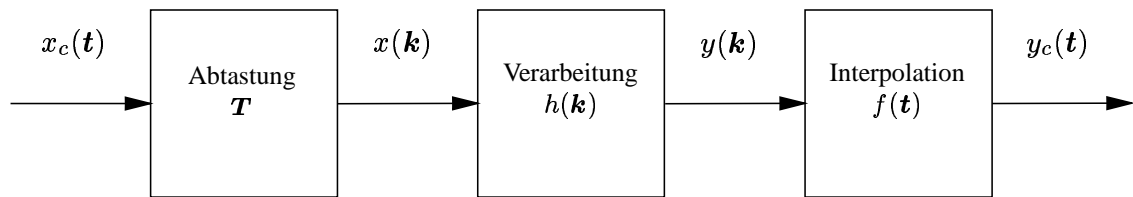


Abbildung 2.1: Digitale Verarbeitung eines kontinuierlichen Signals.

Die Verarbeitung entspricht dann einer diskreten Faltung des abgetasteten Signals  $x(\mathbf{k})$  mit der Impulsantwort, also

$$y(\mathbf{k}) = x(\mathbf{k}) * h(\mathbf{k}) = \sum_{\mathbf{n}} x(\mathbf{n}) h(\mathbf{k} - \mathbf{n}) = \sum_{\mathbf{n}} x(\mathbf{k} - \mathbf{n}) h(\mathbf{n}). \quad (2.11)$$

Die (diskrete) Transformation der Impulsantwort

$$H(\mathbf{v}) = \sum_{\mathbf{k}} h(\mathbf{k}) e^{-j\mathbf{v}^T \mathbf{k}}, \quad (2.12)$$

die bei der Bestimmung des Ausgangssignals eine bedeutende Rolle spielt, nennen wir *Übertragungsfunktion*<sup>3</sup> des Systems, da mit ihrer Hilfe das Systemverhalten im Frequenzbereich beschrieben wird. Bei einem Einsatz des Systems als Digitalfilter kann mittels  $H(\mathbf{v})$  die Filtereigenschaft analysiert oder auch speziell entworfen werden.

Hat das kontinuierliche Eingangssignal nun die spezielle Form  $x_c(\mathbf{t}) = e^{j\boldsymbol{\omega}^T \mathbf{t}}$ , so ergibt sich das diskrete Ausgangssignal zu

$$y(\mathbf{k}) = H(\mathbf{T}^T \boldsymbol{\omega}) x(\mathbf{k}) = H(\mathbf{T}^T \boldsymbol{\omega}) e^{j\boldsymbol{\omega}^T \mathbf{T} \mathbf{k}} \quad (2.13)$$

bzw. mit entsprechender Interpolation das kontinuierliche Ausgangssignal

$$y_c(\mathbf{t}) = H(\mathbf{T}^T \boldsymbol{\omega}) x_c(\mathbf{t}) = H(\mathbf{T}^T \boldsymbol{\omega}) e^{j\boldsymbol{\omega}^T \mathbf{t}}. \quad (2.14)$$

Aus diesem Grund werden Signale dieser Form auch *Eigenfunktionen* linearer ortsinvarianter Systeme genannt. Ein beliebiges Eingangssignal führt dann zu dem diskreten Ausgangssignal

$$y(\mathbf{k}) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} H(\mathbf{v}) X(\mathbf{v}) e^{j\mathbf{v}^T \mathbf{k}} d\mathbf{v}, \quad (2.15)$$

so daß sich die Transformation des Ausgangssignals als

$$Y(\mathbf{v}) = H(\mathbf{v}) X(\mathbf{v}) \quad (2.16)$$

bzw. mit (2.7) als

$$Y_c(\boldsymbol{\omega}) = H(\mathbf{T}^T \boldsymbol{\omega}) X_c(\boldsymbol{\omega}) \quad (2.17)$$

schreiben läßt.

Neben der eingeführten FOURIER-Transformation spielt die (hier zweidimensionale) *z-Transformation*

$$X^{[z]}(z_1, z_2) = \sum_{k_1} \sum_{k_2} x(k_1, k_2) z_1^{-k_1} z_2^{-k_2}, \quad z_{1,2} \in \mathbb{C}, k_{1,2} \in \mathbb{Z} \quad (2.18)$$

---

<sup>3</sup>Streng genommen handelt es sich um den *Frequenzgang* des Systems.



bei der Beschreibung diskreter Signale und insbesondere diskreter Systeme eine wichtige Rolle. Die z-Transformierte der Impulsantwort

$$H^{[z]}(z_1, z_2) = \sum_{k_1} \sum_{k_2} h(k_1, k_2) z_1^{-k_1} z_2^{-k_2}, \quad z_{1,2} \in \mathbb{C} \quad (2.19)$$

bezeichnet man als *z-Übertragungsfunktion* und wird für den Filterentwurf und für Stabilitätsbetrachtungen herangezogen, wie im nächsten Abschnitt deutlich wird. Zwischen der z-Übertragungsfunktion und der Übertragungsfunktion in der normierten Frequenzvariablen  $\mathbf{v} = [v_1 v_2]^T$  gilt der Zusammenhang

$$H(v_1, v_2) = H^{[z]}(e^{jv_1}, e^{jv_2}) \quad (2.20)$$

$$= \sum_{k_1} \sum_{k_2} h(k_1, k_2) e^{-jv_1 k_1} e^{-jv_2 k_2}, \quad (2.21)$$

wobei das komplexe Argument  $z_i$  auf dem Einheitskreis ausgewertet wird. Durch die Beziehung  $\mathbf{v} = \mathbf{v}(\omega) = \mathbf{T}^T \omega$  läßt sich die Übertragungsfunktion auch in der gewöhnlichen Frequenz ausdrücken

$$\hat{H}(\omega_1, \omega_2) = H^{[z]}(e^{jv_1(\omega)}, e^{jv_2(\omega)}) \quad (2.22)$$

$$= \sum_{k_1} \sum_{k_2} h(k_1, k_2) e^{-jv_1(\omega)k_1} e^{-jv_2(\omega)k_2}, \quad (2.23)$$

welche sich bei einer rechtwinkligen Abtastung (siehe Abschnitt 2.2)

$$\mathbf{T} = \begin{bmatrix} T & 0 \\ 0 & T \end{bmatrix} \quad (2.24)$$

zu

$$\hat{H}(\omega_1, \omega_2) = H^{[z]}(e^{j\omega_1 T}, e^{j\omega_2 T}) \quad (2.25)$$

$$= \sum_{k_1} \sum_{k_2} h(k_1, k_2) e^{-j\omega_1 T k_1} e^{-j\omega_2 T k_2}, \quad (2.26)$$

vereinfacht.

In diesem Zusammenhang sollte erwähnt werden, daß die z-Transformation selbst losgelöst von dem gewählten Abtastraster gesehen werden kann. So erfordern z.B. Stabilitätsbetrachtungen nicht die Kenntnis des Rasters. Erst bei der Interpretation als Übertragungssystem wirkt sich die Abtastmatrix auf das Verhalten im Frequenzbereich aus. In den folgenden Abschnitten wird der Zusammenhang der z-Übertragungsfunktion und dem Berechnungsalgorithmus eines diskreten Systems hergestellt.

### 2.1.1 Nichtrekursive Digitalfilter

Nichtrekursive diskrete Systeme oder nichtrekursive Digitalfilter sind in der Digitalen Signalverarbeitung weit verbreitet, da sie einfach zu implementieren sind und stets einen stabilen Algorithmus aufweisen. Da Kausalität in der Bildverarbeitung keine Bedeutung hat, wird hier der allgemeinere Fall der nichtkausalen Systeme betrachtet. Ein eindimensionales Übertragungssystem läßt sich als sogenannte *Differenzengleichung*

$$y(k) = \sum_{\mu=-N}^N a_{\mu} x(k - \mu) \quad (2.27)$$

angeben, bei der das Ausgangssignal  $y(k)$  lediglich von bestimmten mit reellen Filterkoeffizienten  $a_{\mu}$  gewichteten Eingangswerten  $x(k)$ , aber nicht von anderen Ausgangswerten selber abhängt. Der Übersichtlichkeit wegen seien hier die untere und obere Grenze des Summenindex symmetrisch zu  $-N$  bzw.  $N$  gewählt. Die direkte Realisierung dieser Differenzengleichung führt auf einen Algorithmus mit dem gewünschten Verhalten. Die Realisierungen alternativer Filterstrukturen mit dem selben Filterverhalten können aber durchaus für verschiedene Einsatzgebiete, wie z. B. für die Implementierung mit begrenzten Wortlängen der Signale und Koeffizienten Vorteile aufweisen [Fet86]. Die z-Übertragungsfunktion kann durch z-Transformation der Differenzengleichung mittels der Eigenschaften dieser Transformation direkt berechnet werden und lautet

$$H^{[z]}(z) = \sum_{\mu=-N}^N a_{\mu} z^{-\mu} \quad , z \in \mathbb{C}. \quad (2.28)$$

Die Übertragungsfunktion in der normierten Frequenzvariablen  $v$  kann mit

$$H(v) = H^{[z]}(e^{jv}) = \sum_{\mu=-N}^N a_{\mu} e^{-jv\mu} \quad (2.29)$$

direkt aus der z-Übertragungsfunktion gewonnen werden. Hieraus ist ablesbar, daß  $h(k) = a_k$  gilt, d.h. die Filterkoeffizienten stellen die endliche Impulsantwort dar. Daher wird für diese Klasse der Digitalfilter auch oft die Bezeichnung *FIR Filter* (finite impulse response) verwendet.

Im zweidimensionalen Fall kann das Ausgangssignal ebenfalls durch eine Differenzengleichung

$$y(k_1, k_2) = \sum_{\mu_1=-N_1}^{N_1} \sum_{\mu_2=-N_2}^{N_2} a_{\mu_1 \mu_2} x(k_1 - \mu_1, k_2 - \mu_2) \quad (2.30)$$

berechnet und das Verhalten im Frequenzbereich durch die zweidimensionale z-Übertragungstransformation

$$H^{[z]}(z_1, z_2) = \sum_{\mu_1=-N_1}^{N_1} \sum_{\mu_2=-N_2}^{N_2} a_{\mu_1\mu_2} z_1^{-\mu_1} z_2^{-\mu_2}, \quad z_{1,2} \in \mathbb{C} \quad (2.31)$$

charakterisiert werden. In der Bildsignalverarbeitung wird für die gespiegelte Version der zweidimensionalen Impulsantwort  $h(k_1, k_2) = a_{k_1 k_2}$  üblicherweise der Begriff *Faltungsmaske* oder *Faltungskern* verwendet und wie das Bildsignal als Matrix formuliert.

Für den besonderen Fall einer separierbaren Impulsantwort gilt  $a_{\mu_1\mu_2} = a_{\mu_1} a_{\mu_2}$ , und die Matrix läßt sich als dyadisches Produkt schreiben. Hieraus folgt weiter

$$y(k_1, k_2) = \sum_{\mu_1=-N_1}^{N_1} a_{\mu_1} \underbrace{\left( \sum_{\mu_2=-N_2}^{N_2} a_{\mu_2} x(k_1 - \mu_1, k_2 - \mu_2) \right)}_{\tilde{y}(k_1 - \mu_1, k_2)} \quad (2.32)$$

$$= \sum_{\mu_1=-N_1}^{N_1} a_{\mu_1} \tilde{y}(k_1 - \mu_1, k_2), \quad (2.33)$$

d.h. die beiden Faltungssummen können hintereinander (in beliebiger Reihenfolge) getrennt voneinander ausgeführt werden. In diesem Fall läßt sich die zweidimensionale Filterung auf zwei nacheinander ausgeführte eindimensionale Filterungen in den beiden Koordinatenrichtungen reduzieren. Hierdurch wird zwar die Menge aller realisierbaren Übertragungsfunktionen eingeschränkt, der Entwurfsprozeß des Filters vereinfacht sich aber enorm, da mit

$$H^{[z]}(z_1, z_2) = \left( \sum_{\mu_1=-N_1}^{N_1} a_{\mu_1} z_1^{-\mu_1} \right) \left( \sum_{\mu_2=-N_2}^{N_2} a_{\mu_2} z_2^{-\mu_2} \right) \quad (2.34)$$

$$= H_1^{[z]}(z_1) \cdot H_2^{[z]}(z_2) \quad (2.35)$$

der Entwurf ebenso auf zwei eindimensionale z-Übertragungsfunktionen reduziert wird.

### 2.1.2 Rekursive Digitalfilter

Bei einem rekursiven Digitalfilter werden neben den Eingangswerten  $x(k)$  auch Ausgangswerte benötigt, um das Ausgangssignal zu bestimmen. Die Differenzen-

gleichung eines kausalen rekursiven Digitalfilters lautet

$$y(k) = \sum_{\mu=0}^N a_{\mu} x(k - \mu) - \sum_{\nu=1}^M b_{\nu} y(k - \nu), \quad (2.36)$$

und die z-Übertragungsfunktion entsprechend

$$H^{[z]}(z) = \frac{\sum_{\mu=0}^N a_{\mu} z^{-\mu}}{1 + \sum_{\nu=1}^M b_{\nu} z^{-\nu}} = \frac{\sum_{\mu=0}^N a_{\mu} z^{-\mu}}{\sum_{\nu=0}^M b_{\nu} z^{-\nu}} \quad \text{mit } b_0 := 1. \quad (2.37)$$

Da die rekursiven Digitalfilter meist eine unendlich lange Impulsantwort besitzen, nennt man sie auch *IIR Filter* (infinite impulse response). Obwohl dies aber nicht unbedingt der Fall sein muß, wird diese Bezeichnung in der Literatur vielfach verwendet.

Damit der rekursive Filteralgorithmus bzw. das Digitalfilter stabil ist, müssen die Pole von  $H^{[z]}$  im Einheitskreis liegen. Dies kann insbesondere bei einer Direktrealisierung der Differenzengleichung auf einem Digitalrechner nicht immer gewährleistet werden. Zudem können die nichtlinearen Rundungsoperationen, die mit den begrenzten Signalwortlängen einhergehen, zu einem instabilen Verhalten führen. Hierfür bieten sich dann alternative Filterstrukturen an, deren Vorzüge vielfach in der Literatur diskutiert sind [Fet86, Gaz85] und die es ermöglichen, stabile rekursive Digitalfilter zu implementieren.

Für die Realisierung eines in der Bildverarbeitung oft gewünschten nichtkausalen Filters kann z. B. die Hintereinanderschaltung (*Kaskade*) eines kausalen und eines antikausalen Filters verwendet werden, wobei die Filterrichtung bei dem antikausalen Filter gedreht ist. In [vVYV98] ist ein Filterentwurf für sogenannte Gaußtiefpafilter gezeigt, welcher im Rahmen dieser Arbeit auch Verwendung findet (siehe Kapitel 5 und Kapitel 6).

Der Einsatz rekursiver Digitalfilter ist natürlich auch für die Bildverarbeitung von großem Interesse, da hiermit im Vergleich zu den FIR Filtern unter Umständen wesentlich weniger Aufwand für die Signalverarbeitung notwendig ist. Die Stabilität zweidimensionaler rekursiver Systeme zu gewährleisten, erfordert allerdings i. allg. viel mehr Mühe als bei eindimensionalen Systemen. Der Fundamentalsatz der Algebra z. B. besagt, daß ein Polynom vom Grad  $N$  (der Nenner der Übertragungsfunktion) in  $N$  Polynome vom Grad eins faktorisiert werden kann, so daß

jede Singularität von  $H^{[z]}(z)$  isoliert betrachtet werden kann. Für mehrdimensionale Polynome existiert aber kein solches Theorem, und daher ist hier selten eine Faktorisierung möglich, es sei denn die Übertragungsfunktion wird bereits bei dem Entwurf entsprechend konstruiert. Gleichwohl können wie im eindimensionalen Fall auch bei zweidimensionalen Systemen Kriterien für einen Stabilitätstest aufgestellt werden [DM84], was für den Filterentwurf entsprechend berücksichtigt werden muß. Weitere Besonderheiten eines mehrdimensionalen Filterentwurfes, wie die rekursive Berechenbarkeit und die Bedeutung der Anfangswerte des Randbereiches, sollen hier erwähnt werden, sind aber für den Rahmen dieser Arbeit nicht relevant (siehe dazu z. B. [DM84, Lim90, Wah80, Wah89]). Die Aufgaben konnten nämlich auf separierbare Impulsantworten und damit auf die Realisierung eindimensionaler rekursiver Filter reduziert werden. Sind diese nichtkausal realisiert, so werden auf diese Weise auch nichtkausale zweidimensionale rekursive Filter möglich.

## 2.2 Abtastung zweidimensionaler Signale

In der Bildverarbeitung wird meist eine *rechtwinklige* Abtastung vorgenommen, d.h. die Abtastmatrix hat z.B. die vereinfachte Form

$$\mathbf{T}_r = \begin{bmatrix} T_r & 0 \\ 0 & T_r \end{bmatrix}. \quad (2.38)$$

Hieraus resultiert nach (2.8) eine Kreisfrequenzmatrix

$$\mathbf{\Omega}_r = \begin{bmatrix} \frac{2\pi}{T_r} & 0 \\ 0 & \frac{2\pi}{T_r} \end{bmatrix}, \quad \|\boldsymbol{\omega}_{1,2}\| = \frac{2\pi}{T_r}, \quad (2.39)$$

die die Periodizität der Transformatierten festlegt. Im allgemeinen kann die Periode in den beiden Dimensionen sogar unterschiedlich sein. Wir wollen uns aber hier auf den obigen Spezialfall beschränken.

Als vorteilhafte Alternative zur rechtwinkligen Abtastung betrachten und verwenden wir hier allerdings die sogenannte *hexagonale* Abtastung mit

$$\mathbf{T}_h = \begin{bmatrix} T_h & T_h \\ T_h\sqrt{3} & -T_h\sqrt{3} \end{bmatrix} \quad (2.40)$$

bzw.

$$\mathbf{\Omega}_h = \begin{bmatrix} \frac{\pi}{T_h} & \frac{\pi}{T_h} \\ \frac{\pi}{T_h\sqrt{3}} & -\frac{\pi}{T_h\sqrt{3}} \end{bmatrix}, \quad \|\boldsymbol{\omega}_{1,2}\| = \frac{\pi}{T_h} \frac{2}{\sqrt{3}}. \quad (2.41)$$

Es kann gezeigt werden, daß die hexagonale Abtastung die aufwandsgünstigste periodische Abtastmethode ist, wenn die Transformierte des kontinuierlichen Signals, welches abgetastet werden soll, kreisförmig bandbegrenzt ist. Durch die Tatsache, daß i. allg. die Optik des bildgebenden Systems, sei es das Kameraobjektiv oder die Optik des Auges (siehe auch Kapitel 3) ein rotationssymmetrisches Tiefpaßverhalten aufweist, ist das resultierende Spektrum der Bildszene in guter Näherung kreisförmig bandbegrenzt. Ein hexagonales Abtastraster bietet sich daher i. d. R. immer an.

Um die Effizienz einer hexagonalen Abtastung unter diesen Bedingungen zu verdeutlichen, betrachten wir Abb. 2.2, in der die diskrete FOURIER-Transformation jeweils für die rechtwinklige und hexagonale Abtastung skizziert ist. Die durch den Abtastvorgang periodische Fortsetzung des kreisförmig begrenzten Frequenzspektrums ( $\omega_{1,2\max} = \pi/T_r$ ) eines kontinuierlichen Signals  $X_c(\omega)$  ist dabei angedeutet. Hieraus ist erkennbar, daß bei einer hexagonalen Abtastung die kreisförmigen Spektren eine dichtere Anordnung ergeben können ohne sich zu überschneiden, was für die Rekonstruktion des kontinuierlichen Signals von Bedeutung ist. Hieraus resultiert aber, daß die einzelnen Abtastwerte weiter auseinander liegen, d. h. man benötigt weniger Abtastwerte, um dasselbe kontinuierliche Signal zu repräsentieren. In [DM84] ist nachzulesen, daß bei einer hexagonalen Abtastung 13.4% weniger Abtastwerte anfallen, als bei einer rechtwinkligen Abtastung, man also einen geringeren Speicheraufwand betreiben muß. Da aber diese Daten auch verarbeitet werden müssen, ergeben sich auch deutlich weniger Rechenoperationen. Bei der Faltungssumme (2.11) sind z. B. etwa 25% weniger Additionen und Multiplikationen auszuführen, was zu einer beachtlichen Reduktion der Berechnungsdauer führen kann. Die wesentlichen Vorteile einer hexagonalen Abtastung können wie folgt zusammengefaßt werden:

- weniger Bilddaten nach SHANNON ( $\approx 13.4\%$ )
- weniger Berechnungsaufwand bei der Signalverarbeitung ( $\approx 25\%$ )
- Filterantworten weisen höhere Rotationssymmetrie auf
- jedes Pixel besitzt 6 eindeutige nächste Nachbarn

Hier sollte betont werden, daß die Signalbeschreibung und der Filterentwurf auf hexagonalem Abtastraster, und zudem auch die Verwaltung der Daten in Digitalrechnern (hier ist alles auf matrizenförmige Strukturen zugeschnitten) i. allg.

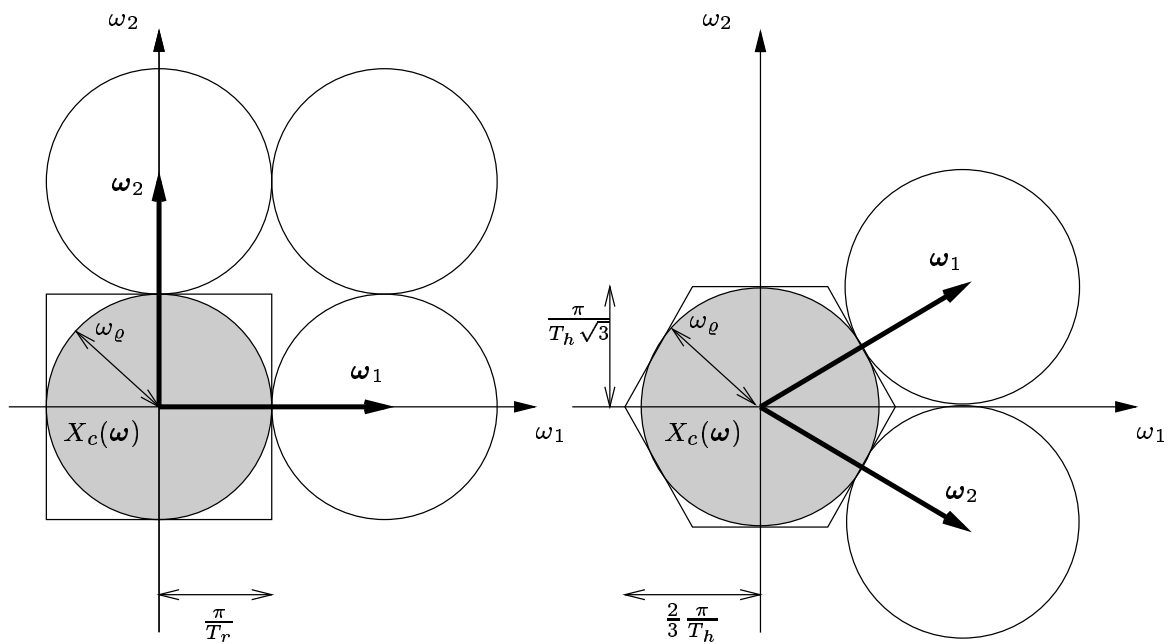


Abbildung 2.2: Periodizität der diskreten FOURIER-Transformation bei einer rechtwinkligen und hexagonalen Abtastung

weitaus komplizierter ist im Vergleich zu dem gewöhnlichen rechtwinkligen Raster. Zusätzlich stehen die Bilddaten der Kamera i. d. R. nicht auf einem hexagonalen Abtastraster zur Verfügung, so daß eine einmalige Umrechnung notwendig ist. Dennoch ist dieses vorteilhafte hexagonale Abtastraster als Strategie in der Bildverarbeitung für spezielle technische Lösungen zu finden, wie z. B. in [HDD<sup>+</sup>94, MDH94, FG96, Sta96, Sta89, GF]. Zudem hat sich das hexagonale Abtastraster, wie im späteren Verlauf dieser Arbeit noch beschrieben wird, im biologischen Sehsystem höher entwickelter Lebewesen, aber auch bei Insekten im Laufe der Evolution etabliert. So macht sich das biologische Vorbild die o. g. Vorteile dieser Form der Signalrepräsentation und Signalverarbeitung effizient zunutze. Um die mathematischen Details muß es sich dabei nicht kümmern.

Ist nun das kontinuierliche Signal mit

$$X_c(\omega) = 0 \quad \forall \|\omega\|^2 = \omega_1^2 + \omega_2^2 > \omega_\rho^2 \quad (2.42)$$

kreisförmig bandbegrenzt, was bei den meisten bildgebenden Verfahren näherungsweise erfüllt ist, so fordert das Abtasttheorem von SHANNON, daß

$$\|\omega_{1,2}\| > 2\omega_\rho \quad (2.43)$$

gilt und somit auch

$$T_r < \frac{\pi}{\omega_\varrho} \quad \text{bei einer rechtwinkligen Abtastung bzw.} \quad (2.44)$$

$$T_h < \frac{\pi}{\sqrt{3}\omega_\varrho} \quad \text{bei einer hexagonalen Abtastung.} \quad (2.45)$$

In der Literatur ist oft ein *pseudohexagonales* Abtastraster (auch oft *Quincunx-* oder *Zeilensprungraster* genannt) [Ohm99, HY94] anzutreffen, bei dem mit

$$\mathbf{T}_q = \begin{bmatrix} T_r & T_r \\ T_r & -T_r \end{bmatrix} \quad (2.46)$$

die Daten durch direkte Abtastung des rechtwinkligen Rasters gewonnen werden. Da die neuen Rasterorte auf dem ursprünglichen rechtwinkligen Raster liegen, entfällt eine aufwendige Interpolation. Diese Methode besticht damit durch seine sehr einfache Implementation und damit sehr niedrigem Berechnungsaufwand. Wie aus der Kreisfrequenzmatrix

$$\mathbf{\Omega}_q = \begin{bmatrix} \frac{\pi}{T_r} & \frac{\pi}{T_r} \\ \frac{\pi}{T_r} & -\frac{\pi}{T_r} \end{bmatrix}, \quad \|\boldsymbol{\omega}_{1,2}\| = \sqrt{2} \frac{\pi}{T_r} \quad (2.47)$$

ersichtlich ist, müssen allerdings mit hoher Wahrscheinlichkeit große Aliasingfehler in Kauf genommen werden, was sich auch durch experimentelle Ergebnisse bestätigt (siehe Abb. 2.5 und Tabelle 2.1). Eine vorherige Tiefpaßfilterung des rechtwinklig abgetasteten Signals ist dabei wenigstens vorzuschlagen. Die Symmetrieeigenschaften und die vorteilhafte Nachbarschaftsbeziehung des hexagonalen Abtastrasters können jedoch auch dann nicht erzielt werden.

### 2.2.1 Umrechnung auf ein hexagonales Abtastraster

Herkömmliche bildgebende Verfahren, wie z. B. CCD-Sensoren, liefern i. d. R. Bildsignale auf einem rechtwinkligen Abtastraster. Deshalb soll an dieser Stelle eine genauere Untersuchung und ein Vergleich einiger Methoden zur Umrechnung zwischen rechtwinklig und hexagonal abgetasteten Bildern gegeben werden, wobei diverse Gesichtspunkte, wie Effizienz, Berechnungsaufwand und Umrechnungsfehler berücksichtigt werden.

Die einfachste hier betrachtete Methode, ein hexagonalähnliches Bildsignal zu erzeugen, ist die Verwendung eines pseudohexagonalen Rasters. Wie in Abb. 2.3 (links) zu erkennen, liegen die gesuchten Abtastorte auf dem ursprünglichen rechtwinkligen Raster. So können durch ein lineares Mapping direkt die gesuchten hexagonalen Bilddaten angegeben werden. Wie sich in diesem Abschnitt aber zeigt,



muß für die Einfachheit und Schnelligkeit des Verfahrens aber auch ein beachtlicher Umrechnungsfehler in Kauf genommen werden. In [HY94] ist die Umrechnung für ein ähnliches Raster beschrieben, bei dem aber immer zwei benachbarte Pixel des rechtwinkligen Raster zu einem Wert des hexagonalen Rasters kombiniert werden. Dieses hexagonale Raster selber entsteht durch Verschieben jeder zweiten Spalte oder Zeile. Die Verrechnung zweier benachbarter Pixel kann als eine einfache Tiefpaßfilterung (Interpolation) interpretiert werden. Allerdings sind die Ergebnisse der betrachteten Interpolationsarten hier nur qualitativ gegenübergestellt.

Unter Zuhilfenahme des Abtasttheorems von SHANNON kann aber auch die Frage geklärt werden, welche Abtastperioden für eine korrekte Umrechnung höchstens notwendig wären, um das kontinuierliche Bildsignal fehlerfrei zu repräsentieren. Da das Spektrum des kontinuierlichen Signals a priori nicht bekannt ist und ja auch von der betrachteten Szene abhängt, können dazu nur Abschätzungen über das zugrunde liegende rechtwinklige Abtastraster gemacht werden. Im folgenden wird dieser Frage nachgegangen. Die resultierenden Abtastmatrizen bzw. Abtastorte fallen dann allerdings nicht mehr auf ursprüngliche Orte des rechtwinkligen Rasters wie bei dem pseudohexagonalen Abtastraster, so daß die Abtastwerte zudem durch Interpolationsverfahren bestimmt werden müssen.

Zunächst sei nach der notwendigen Abtastperiode  $T_h$  für ein hexagonales Raster gefragt, wenn ein rechtwinklig abgetastetes Signal, z.B. von einer CCD-Kamera, vorliegt. Wenn wir davon ausgehen, daß das mit  $T_r$  rechtwinklig abgetastete Signal kreisförmig begrenzt war, dann ist aus (2.44) ersichtlich, daß  $\omega_{\varrho, \max} = \frac{\pi}{T_r}$  die größte Frequenz ist, die in dem abgetasteten kontinuierlichen Signal vorkommen kann. Mit (2.45) ergibt sich dann die notwendige Abtastperiode zu

$$T_h < \frac{\pi}{\sqrt{3}\omega_{\varrho, \max}} = \frac{T_r}{\sqrt{3}}. \quad (2.48)$$

In Abb. 2.3 (mitte) ist das rechtwinklige Raster und das bei kreisförmig bandbegrenzten Signalen notwendige hexagonale Abtastraster gezeigt.

Es soll hier nicht unerwähnt bleiben, daß die Annahme, das kontinuierliche Signal sei kreisförmig bandbegrenzt, in technischen Systemen nicht unbedingt immer erfüllt ist, und somit unter Umständen ein gewisser Aliasingfehler in Kauf genommen werden muß. Untersuchungen mit Bilddaten einer CCD-Kamera zeigen aber, daß die Anteile in den Frequenzbereichen, die bei einer Interpolation auf hexagonalem Raster zu Fehlern führen, durchaus vernachlässigbar sind. Will man diese Voraussetzung dennoch nicht machen, und läßt so eine quadratische

Bandbegrenzung bis  $\omega_{1,2,\max} = \frac{\pi}{T_r}$  zu, erhalten wir die strengere Bedingung

$$\frac{\pi}{T_h} \frac{1}{2} > \omega_{1,\max} \quad \wedge \quad \frac{\pi}{T_h \sqrt{3}} > \omega_{2,\max} \quad (2.49)$$

und damit

$$T_h < \frac{\pi}{2\omega_{1,2,\max}} = \frac{T_r}{2}. \quad (2.50)$$

Dieses Abtastraster erfüllt zwar theoretisch optimal das Abtasttheorem, benötigt aber deutlich mehr Abtastwerte pro Flächeneinheit, was in Abb. 2.3 (rechts) zu sehen ist.

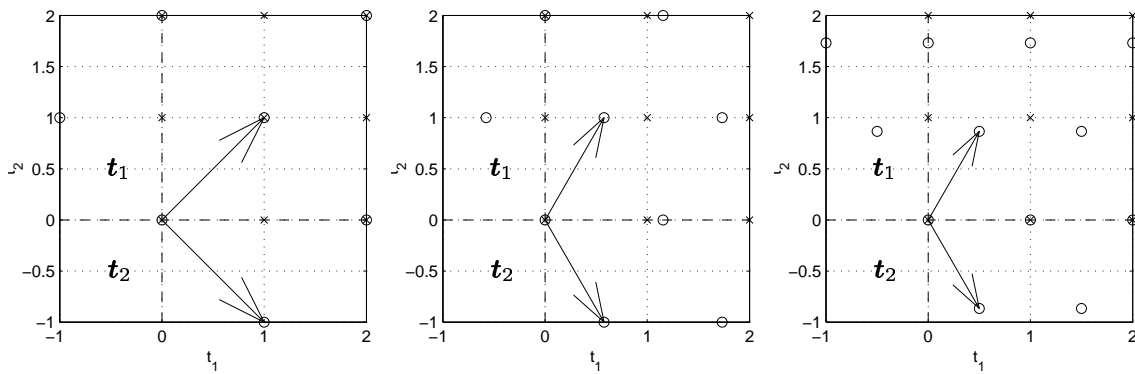


Abbildung 2.3: Rechtwinkliges Abtastraster ( $\times$ ) und verschiedene hexagonale Zielraster ( $\circ$ ). links: pseudo-hexagonales Raster. mitte/rechts: hexagonales Raster bei Annahme einer kreisförmigen (2.48) bzw. rechteckigen (2.50) Bandbegrenzung.

Der Vollständigkeit halber sei hier auch kurz die Rücktransformation behandelt, d. h. es liegt ein Bildsignal auf hexagonalem Raster vor, und die Abtastperiode  $T_r$  für eine Umrechnung auf ein rechtwinkliges Raster sei gesucht. Machen wir wieder die Annahme, es handle sich um eine kreisförmige Bandbegrenzung, so ist lt. (2.45) die größte zu erwartende Frequenz  $\omega_{\varrho,\max} = \frac{\pi}{\sqrt{3}T_h}$  und wir erhalten aus (2.44)

$$T_r < \frac{\pi}{\omega_{\varrho,\max}} = \sqrt{3}T_h. \quad (2.51)$$

Möchte man sich nicht auf o.g. Annahme festlegen, so erhalten wir für die beiden Koordinatenrichtungen die Schranken  $\omega_{1,\max} = \frac{2}{3} \frac{\pi}{T_h}$  und  $\omega_{2,\max} = \frac{\pi}{\sqrt{3}T_h}$  und folglich die kleinere Abtastperiode

$$T_r < \frac{\pi}{\max\{\omega_{1,\max}, \omega_{2,\max}\}} = \frac{3}{2}T_h. \quad (2.52)$$

### 2.2.2 Eine Strategie zur effizienten Interpolation

Wir wissen aus dem vorherigen Abschnitt, wie die Abtastperiode  $T_h$  bzw.  $T_r$  zu wählen ist, wenn Bilddaten auf ein hexagonales bzw. rechtwinkliges Abtastraster umzurechnen sind. Damit kennen wir also die neuen Orte der Abtastpunkte. In einem weiteren Schritt ist aber jetzt noch die eigentliche Interpolation zu leisten, also die Berechnung der jeweiligen Abtastwerte, deren Orte i. allg. nicht auf dem alten Raster liegen, wie aus Abb. 2.3 (mitte/rechts) ersichtlich ist. Die folgenden Methoden sollen hier vergleichend gegenübergestellt werden:

- **Bilineare Interpolation**  
Eine recht schnelle Methode ist die *bilineare Interpolation*, bei der ein neuer Abtastwert aus vier benachbarten Werten des alten Abtastrasters linear berechnet wird.
- **Bispline Interpolation**  
Die zweidimensionale kubische Spline Interpolation benötigt viel Rechenzeit, liefert dafür zwischen den Stützstellen im Gegensatz zur linearen Interpolation ein glatteres Ergebnis, was für die vorliegende Aufgabe von Vorteil sein könnte.
- **Bilineare oder Bispline Interpolation mit vorheriger Verdoppelung der Abtastrate**  
Durch Erhöhung der Abtastrate des rechtwinkligen Rasters ist eine Verringerung des Umrechnungsfehlers zu erwarten.
- **Matched Interpolation**  
Bei dieser neuartigen Methode soll zunächst das Abtastraster des rechtwinklig abgetasteten Bildsignals derart erhöht werden, so daß ein Raster entsteht, in dem möglichst genau die Abtastorte des gewünschten hexagonalen Rasters wiederzufinden sind. Hierfür ist eine effiziente rationale Abtastratenerhöhung notwendig, die im folgenden Abschnitt erläutert wird. In einem zweiten Schritt können durch ein einfaches lineares Mapping die Abtastwerte bestimmt werden.

Zur Verbesserung der Bilinearen und Spline Interpolation, aber vor allem für die Matched Interpolation, ist eine Abtastratenänderung auf rechtwinkligem Raster notwendig. Hierbei soll aus dem bereits abgetasteten Signal  $x(\mathbf{k}) = x_c(\mathbf{T}_r \mathbf{k})$  ein neues Signal  $\tilde{x}(\mathbf{n})$  der Abtastmatrix  $\tilde{\mathbf{T}}_r$  berechnet werden, welches aber identisch ist mit einer Abtastung des ursprünglichen kontinuierlichen Signals durch  $\tilde{\mathbf{T}}_r$ , also

$\tilde{x}(\mathbf{n}) = x_c(\tilde{\mathbf{T}}_r \mathbf{n})$ . Theoretisch muß also das kontinuierliche Signal rekonstruiert und erneut abgetastet werden, was auf einem Digitalrechner allerdings nicht direkt möglich ist. Wir beschränken uns aus diesem Grund auf Abtastratenänderungen mit rationalem Interpolationsfaktor, d. h. für das neue Abtastraster  $\tilde{\mathbf{T}}_r$  gilt die Beziehung

$$\mathbf{T}_r = \mathbf{D} \tilde{\mathbf{T}}_r \quad (2.53)$$

mit der Interpolationsmatrix

$$\mathbf{D} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}, \quad r_{1,2} \in \mathbb{Q}^{>0}, \quad (2.54)$$

wobei  $r_{1,2} > 1$  eine Erhöhung (Upsampling) und  $r_{1,2} < 1$  eine Verringerung (Downsampling) der Abtastrate, respektive der Auflösung in der entsprechenden Koordinate bedeutet.

Ein vor allem in der Onlineverarbeitung, wie z. B. der Sprachsignalverarbeitung, gängiges Vorgehen ist die Separation der Ratenänderung in einen Upsampling- und einen anschließenden Downsampling-Prozeß [PM92], gemäß der Eigenschaft

$$\mathbf{D} = \begin{bmatrix} p_1/q_1 & 0 \\ 0 & p_2/q_2 \end{bmatrix} = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} \begin{bmatrix} 1/q_1 & 0 \\ 0 & 1/q_2 \end{bmatrix}, \quad p_{1,2}, q_{1,2} \in \mathbb{N}^{>0}. \quad (2.55)$$

Da durch die Separation nur ganzzahlige Interpolationsfaktoren verwendet werden dürfen, ist diese Art von Ratenänderung umsetzbar. Bei der Ratenerhöhung werden zunächst Nullen an den entsprechenden zusätzlichen Abtastorten eingefügt. Dies staucht allerdings lediglich das Spektrum, so daß man das sich ergebende Signal anschließend mittels eines Tiefpaß-Filters (Interpolationsfilter) bearbeiten muß. Der Ratenverringerung geht dagegen eine Tiefpaß-Filterung voraus, um dann die benötigten neuen Abtastwerte durch ein lineares Mapping direkt abzulesen. Dieses Vorgehen der Ratenänderung ist dabei nicht auf ein bestimmtes Abtastraster beschränkt. Auch für die Änderung der Abtastrate auf einem hexagonalen Raster, z. B. für eine Rücktransformation auf rechtwinklige Koordinaten ist ein Interpolationsfilter, allerdings auf hexagonalem Raster notwendig (siehe dazu auch [LF97]). Diese Methode hat aber wesentliche Nachteile, wenn kleine nichtrationale Interpolationsfaktoren approximativ realisiert werden sollen, da hierdurch  $p$  und  $q$  recht groß werden können. Zudem ist die Qualität der Interpolation stark von der Güte des verwendeten Filters abhängig.

Im vorliegenden Fall handelt es sich allerdings nicht um eine Onlineverarbeitung im eigentlichen Sinne, da das Bildsignal als Matrix i. d. R. erst nach der Bildaufnahme vorliegt - dann allerdings komplett. Somit kann hier die Abtastratenänderung im Frequenzbereich nach Transformation mit der FFT vollzogen werden.

Dazu muß das Signal zunächst in den Frequenzbereich transformiert werden, um dort an der Nyquistfrequenz bzgl. beider Frequenzkoordinaten passend viele Nullen einzufügen (Ratenerhöhung) oder auch Werte des Spektrums wegzuschneiden (Ratenverringern). Bei letzterem wird immer eine ideale Tiefpaßfilterung vollzogen, was die Qualität der Interpolation mit wenig Aufwand erheblich verbessert. Es sei darauf hingewiesen, daß durch die Anwendung der FFT implizit das Bildsignal im Ortsbereich periodisch fortgesetzt wird, was an den Rändern eine entsprechende Extrapolation bewirkt und ggf. Artefakte verursachen kann. Die Rücktransformation ist dann allerdings i. d. R. keine schnelle FFT mehr, so daß mit zunehmender Bildgröße auch dieses Verfahren mehr Rechenzeit benötigt. Das beschriebene Verfahren und Weiterentwicklungen können in [Haw97, WSM92] nachgelesen werden.

Für die vorliegende Arbeit erwies sich die beschriebene Abtastratenerhöhung mittels der FFT als sehr effizient und wurde vor allem für die Matched Interpolation eingesetzt. Hierzu wurde das rechtwinklige Raster derart modifiziert (Resampling), daß das hexagonale Zielraster möglichst nah in beiden Koordinatenrichtungen getroffen wird. Aus Tabelle 2.1 ist dazu die erforderliche hexagonale Abtastmatrix für eine kreisförmige und quadratische Bandbegrenzung angegeben zu

$$\mathbf{T}_h^{(k)} = \begin{bmatrix} \frac{T_r}{\sqrt{3}} & \frac{T_r}{\sqrt{3}} \\ T_r & -T_r \end{bmatrix}, \quad \mathbf{T}_h^{(q)} = \begin{bmatrix} \frac{T_r}{2} & \frac{T_r}{2} \\ \frac{T_r\sqrt{3}}{2} & -\frac{T_r\sqrt{3}}{2} \end{bmatrix}. \quad (2.56)$$

Im ersten Fall ist abzulesen, daß lediglich in der ersten Koordinate eine Abtastratenerhöhung um  $\sqrt{3}$  erforderlich ist (vgl. auch Abb. 2.3). Bei der quadratischen Bandbegrenzung ist hingegen eine Ratenänderung in beiden Koordinaten nötig, so daß sich folgende Interpolationsmatrizen ergeben:

$$\mathbf{D}^{(k)} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{D}^{(q)} = \begin{bmatrix} 2 & 0 \\ 0 & \frac{2}{\sqrt{3}} \end{bmatrix}. \quad (2.57)$$

### 2.2.3 Experimentelle Ergebnisse

An dieser Stelle soll nun ein Vergleich der einzelnen Methoden zur Umrechnung auf ein hexagonales Raster gegeben werden, da deren Beurteilung für den Einsatz in einem technischen Bildverarbeitungssystem, aber auch für die Analyse biologischer Vorgänge wichtig ist. Allerdings scheint der Vergleich zweier Signale, die auf unterschiedlichen Abtastrastern liegen, nicht direkt möglich. Denkbar wäre aber, das rechtwinklig abgetastete Signal (einer Kamera) auf ein zu bewertendes

Raster zu transformieren, um es dann wieder auf das ursprüngliche Raster zurückzurechnen. Das rücktransformierte Signal kann dann mit dem Originalsignal verglichen werden. Da aber durch dieses Vorgehen sowohl Hin- als auch Rücktransformation jeweils einen Einfluß auf den berechenbaren Fehler ausüben und nicht mehr getrennt betrachtet werden können, erweist sich dieser Weg als unbrauchbar für die Untersuchungen. Aus diesem Grund soll hier eine Betrachtung im Frequenzbereich durchgeführt werden, um einen quantitativen Vergleich der verschiedenen Raster untereinander (hexagonales Raster mit kreisförmiger bzw. quadratischer Annahme über die Bandbegrenzung und das pseudohexagonale Raster) aber auch der notwendigen Interpolationsmethoden (Bilinear, B spline und Matched Interpolation) zu ermöglichen. Diese Betrachtung ist möglich, da das Definitionsgebiet  $\omega$  für alle Abtastraster einheitlich ist und das Spektrum den Informationsgehalt eines Signals vollständig widerspiegelt.

Die FOURIER-Transformation des auf rechtwinkligem Raster  $\mathbf{T}_r$  definierten Signals

$$\hat{X}_r(\omega) = \sum_{\mathbf{k}} x_r(\mathbf{k}) e^{-j\omega^T \mathbf{T}_r \mathbf{k}} = \frac{1}{|\det \mathbf{T}_r|} \sum_{\mu} X_c(\omega - \Omega_r \mu) \quad (2.58)$$

stellt in dem Basisband  $\omega_B = \{[w_1, w_2]^T | w_{1,2} = 0 \dots \pi\}$  die Frequenzverteilung  $X_c$  des zu Grunde liegenden kontinuierlichen Signals zur Verfügung. Somit kann mittels der DFT in beliebiger Frequenzauflösung ein Referenzspektrum berechnet werden, was aber auch nach einer evtl. Umrechnung auf ein anderes Abtastraster möglichst genau erreicht werden sollte. Dies gibt dann die Güte der Umrechnung wieder. Das Spektrum auf dem hexagonalen Abtastraster kann dann ebenso durch

$$\hat{X}_h(\omega) = \sum_{\mathbf{k}} x_h(\mathbf{k}) e^{-j\omega^T \mathbf{T}_h \mathbf{k}} = \frac{1}{|\det \mathbf{T}_h|} \sum_{\mu} X_c(\omega - \Omega_h \mu) \quad (2.59)$$

berechnet werden. Schließlich kann durch ein normiertes Fehlermaß

$$F_{\text{NMSE}} = \frac{\int_{\omega \in \omega_B} \left| \hat{X}_r(\omega) |\det \mathbf{T}_r| - \hat{X}_h(\omega) |\det \mathbf{T}_h| \right|^2 d\omega}{\int_{\omega \in \omega_B} \left| \hat{X}_r(\omega) \right|^2 \cdot |\det \mathbf{T}_r|^2 d\omega}, \quad (2.60)$$

welches die normierte mittlere Abweichung im Frequenzbereich angibt, der Informationsgehalt auf dem hexagonalen Raster bzw. die Güte der Umrechnung bewertet werden.

In Abb. 2.4 ist nun ein Bildsignal auf rechtwinkligem Raster (Bild einer CCD-Kamera) und das zugehörige Betragsspektrum (logarithmisch aufgetragen) dargestellt. Die Ergebnisse der Umrechnung dieses Bildsignals auf ein hexagonales

Abtastraster sind in Abb. 2.5 zu sehen. Der für das Fehlermaß berücksichtigte Bereich von 0 bis zur Nyquistfrequenz ist entsprechend gekennzeichnet.

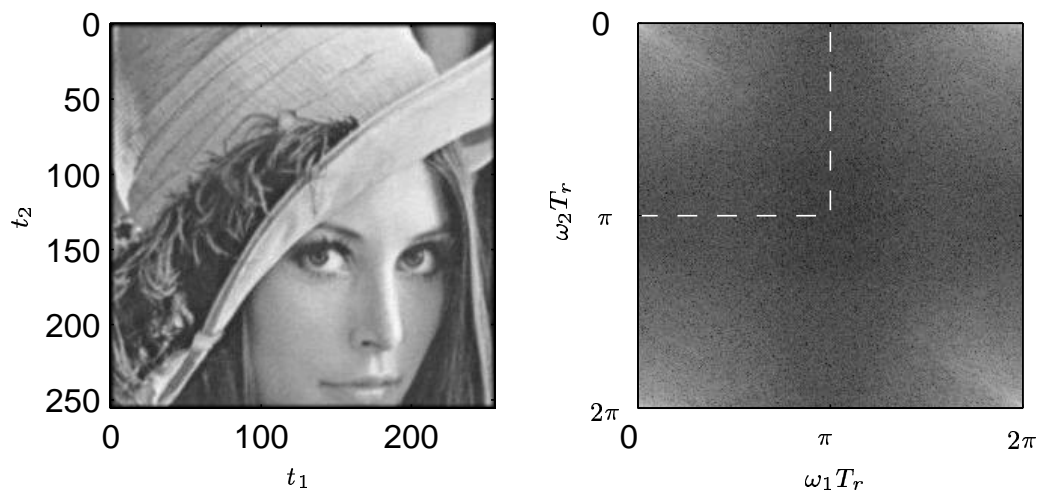


Abbildung 2.4: CCD-Bild und das zugehörige Spektrum.

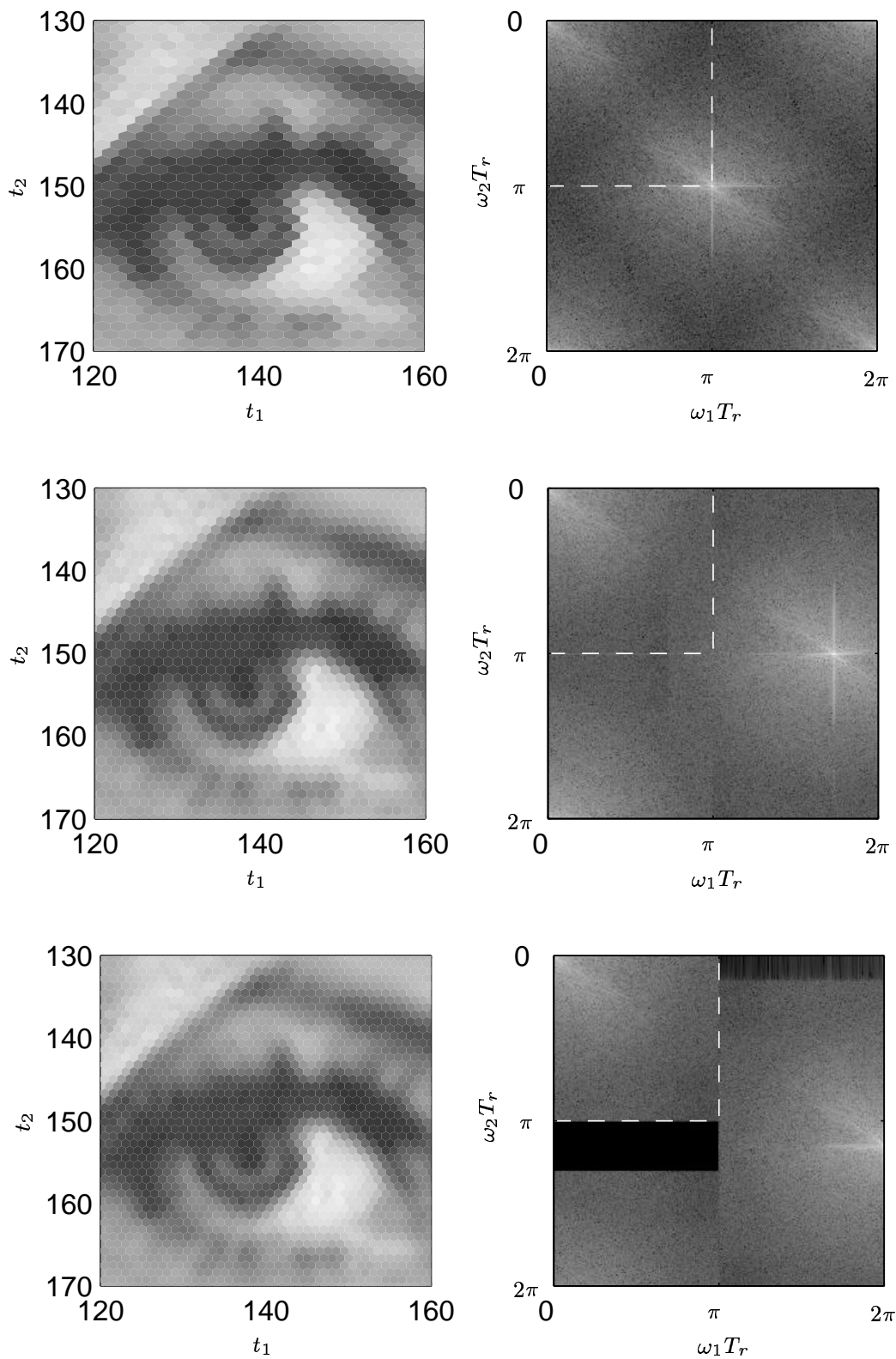


Abbildung 2.5: Umrechnung auf ein hexagonales Raster. oben: pseudohexagonales Raster (2.46). mitte/unten: hexagonales Raster bei Annahme einer kreisförmigen (2.48) bzw. rechteckigen (2.50) Bandbegrenzung.



Auffallend und für die Qualität der Umrechnung von großer Bedeutung sind hier die unterschiedlichen Periodizitäten der Spektren im Frequenzbereich. Bei dem pseudohehexagonalen Raster (oben) ist die Verletzung des Abtasttheorems deutlich zu sehen, da hier die höheren Spektren in das Basisspektrum hineinragen. Eine angenommene kreisförmige Bandbegrenzung (mitte) des kontinuierlichen Signals scheint zumindest qualitativ betrachtet näherungsweise erfüllt zu sein. Bei einer quadratischen Bandbegrenzung (unten) sind theoretisch keine Aliasingfehler zu erwarten, da hier die Periodizität im Frequenzbereich optimal ist.

In Tabelle 2.1 sind die Ergebnisse aller untersuchten Interpolationsmethoden für ein hexagonales Abtastraster gegenübergestellt. Die Effizienz gibt hier die Dichte der Abtastpunkte relativ zum ursprünglichen rechtwinkligen Raster an. Je kleiner dieser Wert ist, desto weniger Abtastpunkte müssen verwaltet und verarbeitet werden. Die Berechnungsdauer  $T_{\text{Ber}}$  für die Umrechnung der Daten wurde auf einer Sun Ultra 6 unter MATLAB ermittelt, kann also nicht als Leistungsbewertung für ein Bildverarbeitungssystem herangezogen werden, ermöglicht aber den Vergleich der Methoden untereinander. Das Fehlermaß  $F_{\text{NMSE}}$  gibt die Qualität der Umrechnung auf das hexagonale Abtastraster wieder und eignet sich für den Vergleich der Verfahren untereinander. Zusätzlich sind die Ergebnisse des pseudohehexagonalen Rasters gezeigt, das sich durch eine hohe Effizienz, schnelle Berechnung aber auch durch einen großen Umrechnungsfehler auszeichnet.

Aus der Tabelle Tabelle 2.1 wird ersichtlich, daß durch vorherige Abtastratenerhöhung bei der Bilinearen und Bispline Interpolation (Bilinear 2, Bispline 2) der Umrechnungsfehler wie vermutet verringert werden kann. Allerdings wird hierfür auch mehr als die doppelte Berechnungszeit benötigt. Die hier vorgeschlagene Matched Interpolation zeigt allerdings sowohl bzgl. der Berechnungsdauer als auch bei dem resultierenden Umrechnungsfehler ihre Vorteile. Bei der Annahme einer kreisförmigen Bandbegrenzung liefert diese Methode nur noch einen Fehler von nur  $\approx 4.95 \cdot 10^{-5}$ , was mehr als eine 10er Potenz besser ist als die einfache bilineare Interpolation bei vergleichbarem rechnerischem Aufwand. Wird eine quadratische Bandbegrenzung angenommen, so erhält man mit  $7.29 \cdot 10^{-29}$  praktisch keinen meßbaren Fehler mehr. Die anderen Verfahren hingegen liefern hier noch einen größeren Fehler als bei der Matched Interpolation bei kreisförmiger Bandbegrenzung. Aus diesem Grunde überzeugt die vorgeschlagene Methode in jeder Hinsicht und sollte für eine ressourceneffiziente und robuste Umrechnung herkömmlicher Bilddaten auf ein hexagonales Raster verwendet werden.

Annahme über Bandbegrenzung	Interpolationsmethode	Effizienz [%]	$T_{\text{Ber}}$ [s]	$F_{\text{NMSE}}$
kreisförmig: $\mathbf{T}_h^{(k)} = \begin{bmatrix} \frac{T_r}{\sqrt{3}} & \frac{T_r}{\sqrt{3}} \\ T_r & -T_r \end{bmatrix}$ $\mathbf{\Omega}_h^{(k)} = \begin{bmatrix} \frac{\pi\sqrt{3}}{T_r} & \frac{\pi\sqrt{3}}{T_r} \\ \frac{\pi}{T_r} & -\frac{\pi}{T_r} \end{bmatrix}$	Bilinear Bispline	86	0.17	$7.778531 \cdot 10^{-4}$
		86	2.46	$7.686476 \cdot 10^{-4}$
	Bilinear 2 Bispline 2	86	0.41	$1.298692 \cdot 10^{-4}$
		86	6.01	$1.399655 \cdot 10^{-4}$
	Matched	86	0.22	$4.945777 \cdot 10^{-5}$
quadratisch: $\mathbf{T}_h^{(q)} = \begin{bmatrix} \frac{T_r}{2} & \frac{T_r}{2} \\ \frac{T_r\sqrt{3}}{2} & -\frac{T_r\sqrt{3}}{2} \end{bmatrix}$ $\mathbf{\Omega}_h^{(q)} = \begin{bmatrix} \frac{2\pi}{T_r} & \frac{2\pi}{T_r} \\ \frac{2\pi}{T_r\sqrt{3}} & -\frac{2\pi}{T_r\sqrt{3}} \end{bmatrix}$	Bilinear Bispline	114	0.22	$1.920650 \cdot 10^{-3}$
		114	3.19	$1.900647 \cdot 10^{-3}$
	Bilinear 2 Bispline 2	116	0.47	$1.234659 \cdot 10^{-4}$
		116	6.75	$1.236909 \cdot 10^{-4}$
	Matched	115	0.29	$7.288345 \cdot 10^{-29}$
Quincunx (Pseudohexagonal): $\mathbf{T}_q = \begin{bmatrix} T_r & T_r \\ T_r & -T_r \end{bmatrix}$ $\mathbf{\Omega}_q = \begin{bmatrix} \frac{\pi}{T_r} & \frac{\pi}{T_r} \\ \frac{\pi}{T_r} & -\frac{\pi}{T_r} \end{bmatrix}$		50	0.07	$2.867551 \cdot 10^{-2}$

Tabelle 2.1: Vergleich der Verfahren zur Umrechnung auf ein hexagonales Raster.

## 2.3 Biologische Verarbeitungssysteme

Die *neuronalen Netze* (NN) sind teilweise vereinfachte Beschreibungen der vernetzten und massiv parallelen Informationsverarbeitung in biologischen (Nerven-) Systemen, wie im menschlichen Gehirn oder in Gehirnen recht artverwandter Lebewesen (Primaten). Der Begriff *neuronal* ist von der Bezeichnung der Nervenzelle (griech. *Neuron*) abgeleitet. Neuronale Netze werden allgemein als eine große Ansammlung informationstransportierender und informationsverarbeitender Neuronen verstanden, die über ein Netzwerk miteinander verbunden sind.

Die ursprünglich in der Informatik eingesetzten *künstlichen neuronalen Netze* (KNN) sind modellhafte Konstrukte, welche die für die Informationsverarbeitung wesentlichen Aspekte beinhalten, und repräsentieren damit meist einen numeri-

schen Algorithmus oder eine Berechnungsvorschrift. Dabei dient dieses Modell zum einen dem Verständnis biologischer Mechanismen, was vielfach durch die digitale Simulation der künstlichen neuronalen Netze unterstützt werden kann. Neurologen können z. B. auf diese Art die Verarbeitung tiefer studieren, zuverlässigere Diagnosen stellen, und durch Simulationen medikamentöse Behandlungen im Digitalrechner erproben. Zum anderen können diese künstlichen Netzwerke auch eingesetzt werden, um technische Aufgaben, wie z. B. im Bereich der (diskreten) Optimierung, der Funktionsapproximation, der Mustererkennung und Klassifizierung [EPdRH02] zu lösen. Hierdurch sind in der Vergangenheit diese biologisch inspirierten Verarbeitungssysteme auch für ingenieurwissenschaftliche Fragestellungen interessant geworden.

Im Rahmen dieser Arbeit steht die Verarbeitung visueller Signale und Informationen im Vordergrund, um Zielsetzungen der Bildsignalverarbeitung mit Hilfe des biologischen Vorbildes zu verfolgen. Daher werden im folgenden Abschnitt die neurophysiologischen Grundlagen speziell für die Verarbeitung visueller Reize gelegt.

### 2.3.1 Neurophysiologische Verarbeitung visueller Signale

Zunächst wird die Funktionsweise neuronaler Netze und der grundlegende Aufbau des visuellen Systems im Gehirn von Primaten in kompakter Form dargestellt. Im Verlauf dieser Arbeit werden dann einzelne für die Zielsetzung interessante Aspekte vertieft behandelt.

#### Neurophysiologische Grundlagen

In Abb. 2.6 ist der funktionelle Aufbau eines Neurons abgebildet [Hub90], welches aus *Dendriten*, einem *Zellkörper* und einem *Axon* besteht. Reize anderer Neuronen gelangen zu den Dendriten eines Neurons und können zu einer Reizweiterleitung über das Axon führen, wenn die zeitliche und örtliche Summation der Eingangsreize groß genug ist. Das Innere eines Neurons ist mit einer Lösung aus ionisierten Salzmolekülen gefüllt und weist eine elektrische Potentialdifferenz zu der dem Neuron umgebenden Salzlösung auf. Dieses *Membranpotential* wird durch dendritische Reizungen anderer Neuronen beeinflusst und führt ab einer gewissen Schwelle zu einer Öffnung von Ionenkanälen in der Zellmembran, wodurch positive Natriumionen in die Zelle hineinfließen. Dadurch erhöht sich das Membranpotential weiter und schließlich öffnen sich weitere Kanäle, die positive Kaliumionen nach außen leiten und das Ruhepotential wieder herstellen. Diese lokale Änderung des elektrischen Potentials pflanzt sich örtlich fort

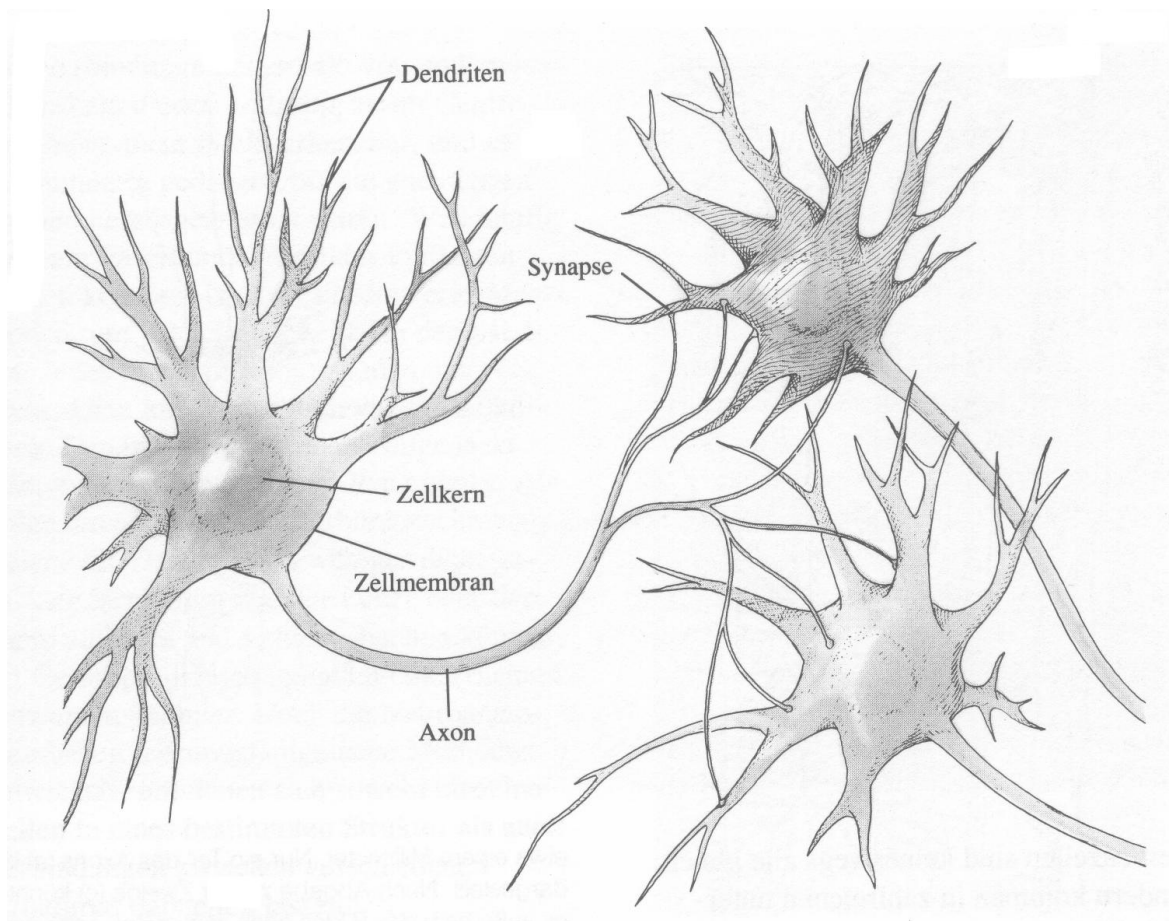


Abbildung 2.6: Aufbau und Verschaltung eines Neurons.

und wird ggf. über das Axon bis zu anderen Neuronen weitergeleitet, dient damit also als wesentlicher Informationsträger. Diese Potentialänderung wird auch als *Aktionspotential*, *Puls* oder *Spike* bezeichnet, und die Anzahl der erzeugten Pulse pro Zeiteinheit gibt die *Aktivierung* des Neurons wieder. Wird ein künstliches neuronales Netz (KNN) auf diese besonders biologienahe Weise modelliert, so spricht man auch von *pulscodierten* oder *pulsgekoppelten neuronalen Netzen* (PCNN). Dies ist eine Art der KNN, welche nicht nur die Aktivierung der Neuronen selbst, sondern auch das zeitliche Auftreten des Reizes berücksichtigt. Somit kommt die Zeit als weiterer Informationsträger in die Verarbeitungskette hinein. Untersuchungen haben beispielsweise gezeigt, daß Neuronen, die ähnliche Eingangsinformationen repräsentieren, wie eine kontinuierliche Linie, sich synchronisieren können, um so eine Art zeitliche Segmentierung durchzuführen [EKGS90, ERAD89, ERAD90, Eck94]. Dies kann natürlich bei der Entwicklung künstlicher Sehsysteme zur Interpretation einer Bildszene von großem Interesse sein [HD90, Har91].

## Das visuelle System

Das visuelle System des Primaten ist einer der am besten erforschten Teile des Gehirns. Bereits 1960 lieferte der Nobelpreisträger DAVID H. HUBEL durch seine Forschungsergebnisse einen großen Beitrag zum Verständnis der neuronalen Mechanismen [HW62, Hub90]. Die Signalverarbeitung der wahrgenommenen Szene beginnt bereits im Auge, wenn das reflektierte Licht der Szene durch das optische System von Iris und Linse auf die *Netzhaut (Retina)* projiziert wird (siehe Abb. 2.7 [Hub90]). Dort befinden sich die lichtempfindlichen Zellen (*Photorezeptoren*), die

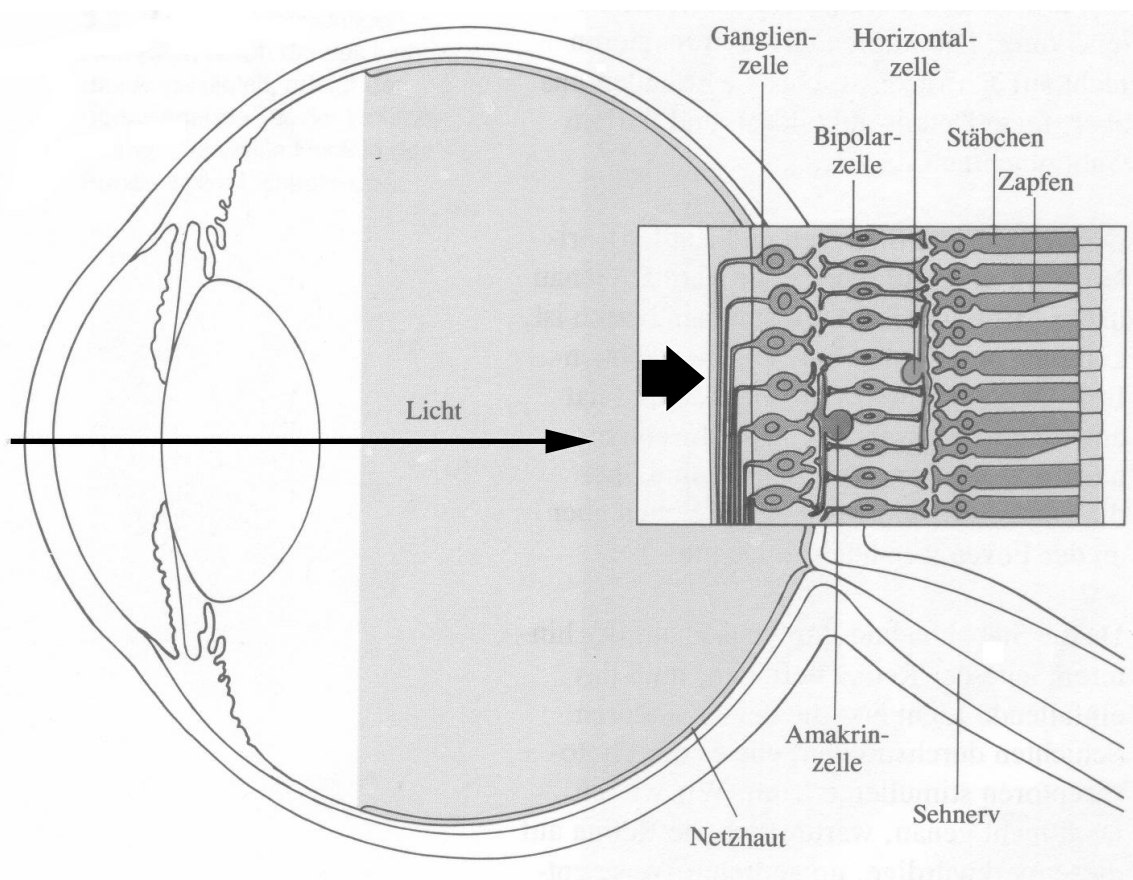


Abbildung 2.7: Aufbau des Auges und der Netzhaut.

die auftreffende Lichtintensität in ein elektrisches Signal umwandeln. Man unterscheidet zwischen *Zapfen*, die bevorzugt auf den roten, grünen und blauen Wellenlängenbereich reagieren und *Stäbchen*. Die Stäbchen sind in einer weitaus größeren Anzahl vorhanden ( $\approx 120 \cdot 10^6$ ) als die Zapfen ( $\approx 6 \cdot 10^6$ ), und sind auch anders auf der Retina verteilt, wie aus Abb. 2.8 deutlich wird [Köh90, ST97]. Das Auflösungsvermögen der Stäbchen ist wesentlich geringer als das der Zapfen, da-

für sind sie aber wesentlich lichtempfindlicher, wodurch sie im wesentlichen für das *Nachtsehen* zum Einsatz kommen. In der *Sehgrube (Fovea)* ist die Dichte der Zapfen am größten. Dies ist der Bereich des *schärfsten Sehens*.

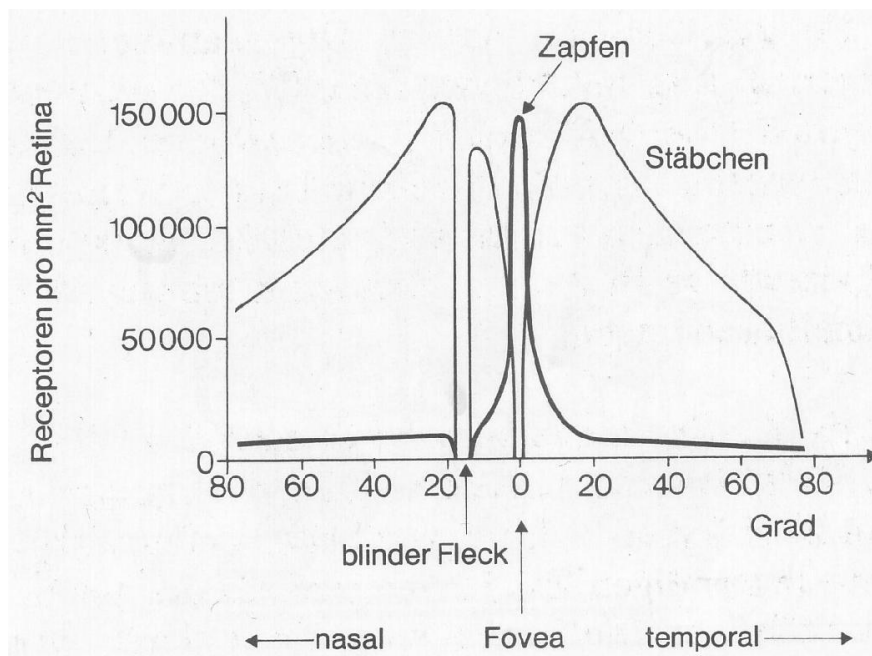


Abbildung 2.8: Verteilung der Photorezeptoren auf der Netzhaut.

Wie die Detailzeichnung der Retina in Abb. 2.7 erkennen läßt, leiten die Rezeptoren ihren Reiz weiter an die Schicht der *Bipolarzellen*, welche wiederum mit den *Ganglienzellen* verbunden sind. Darüber hinaus existieren *Horizontal-* und *Ama-krinzellen*, die eine weitere Verkopplung der Rezeptoren und der Bipolarzellen verursachen. Die verschiedenen Zellen können meist weiter in unterschiedliche Typen unterteilt werden.

An dieser Stelle ist erwähnenswert, daß durch die Verkopplung der einzelnen Schichten und durch die inhomogene Verteilung der Rezeptoren auf der Netzhaut bereits eine enorme Datenreduktion im Auge stattfindet. Die Signale von fast 130 Millionen Rezeptoren konvergieren auf ca. 1 Million Gangliensignale, die über den *Sehnerv (Nervus opticus)* zur *Großhirnrinde (Cortex)* übertragen werden (siehe Abb. 2.9 [ST97]). Hierbei kreuzen sich die Sehnerven der beiden Augen im *Chiasma Opticum* derart, daß in jeder Gehirnhälfte das gleiche Gesichtsfeld beider Augen zur Verfügung steht, um u. a. das Tiefensehen zu ermöglichen. Die Signale gelangen zunächst zu dem *seitlichen Kniehöcker (Corpus geniculatum laterale)*, was im wesentlichen eine Schaltstelle darstellt. Von dort werden die In-

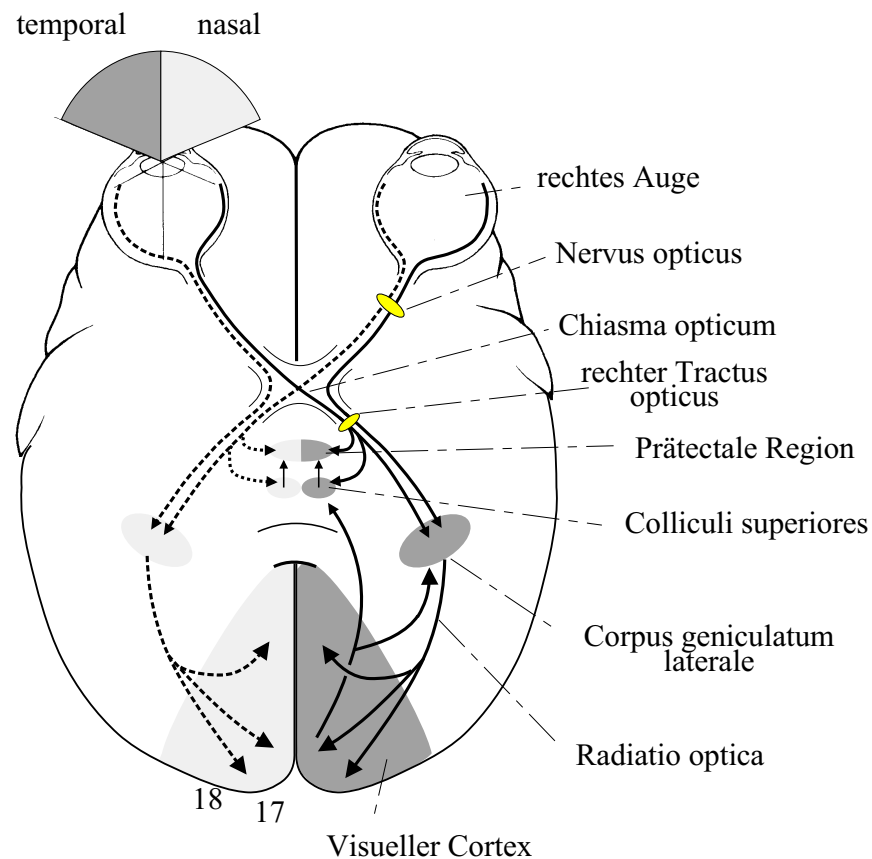


Abbildung 2.9: Schematische Darstellung der Sehbahn im menschlichen Gehirn.

formationen in verschiedene Bereiche des *visuellen Cortex* projiziert. Dort sind u.a. Zellen nachgewiesen, die auf Kontraständerungen in der auf der Netzhaut abgebildeten Bildszene ansprechen. So existieren Zellen, die auf Konturverläufe mit spezifischer Orientierung, auf bewegte Konturverläufe oder auch Konturverläufe begrenzter Länge reagieren. Die erste Gruppe wird als *einfache* oder *simple Zellen* charakterisiert, auf die im Laufe dieser Arbeit noch genauer eingegangen wird.

Durch aufeinander aufbauende Neuronenverkopplungen sind die Neuronen in der Lage, eine immer komplexer werdende Reizspezifität zu erhalten. Die Beschreibung der Reizspezifität eines Neurons erfolgt über das *rezeptive Feld*. Dieses gibt zum einen den Bereich (Ort und Größe) des Bildausschnittes respektive der Rezeptoren an, auf den das Neuron überhaupt reagiert, d. h. der dem Neuron Eingangssreize liefern kann. Zum anderen beschreibt das rezeptive Feld die optimale Form und Ausprägung des Lichtreizes auf der Netzhaut, der eine maximale Erregung des Neurons verursacht. So weisen Ganglienzellen ein *Zentrum-Umfeld*-

*Verhalten* auf. Bei einer *On-Zentrum-Zelle* beispielsweise besitzt das rezeptive Feld ein erregendes (*exzitatorisches*) Zentrum und ein hemmendes (*inhibitorisches*) Umfeld, so daß ein begrenzter heller Stimulus in einer dunklen Umgebung das Neuron am stärksten erregt. Ein dunkler Stimulus in einer hellen Umgebung hingegen hemmt das Neuron am stärksten, was in Abb. 2.10 (links) [ST97] exemplarisch dargestellt ist. Eine *Off-Zentrum-Zelle* verhält sich genau entgegengesetzt.

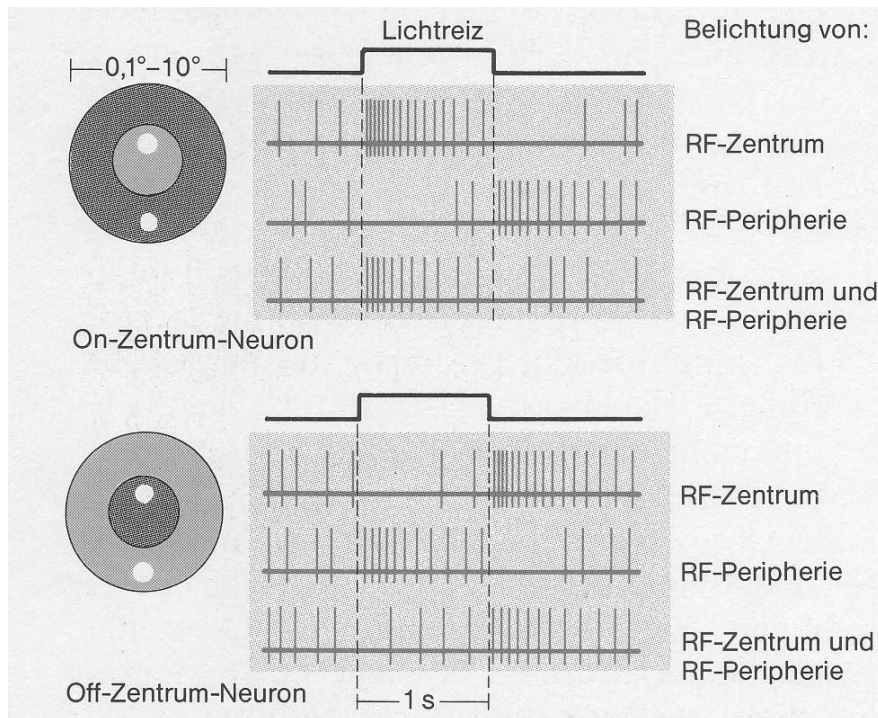


Abbildung 2.10: Rezeptives Feld und Antwortverhalten einer On- bzw. Off-Zentrum-Zelle.

## Terminologie beim menschlichen Auge

Zum Abschluß dieses Abschnittes über die neurophysiologische Verarbeitung sollen noch kurz ein paar weitere wichtige Fachausdrücke und deren Zusammenhang zu der ingenieurwissenschaftlichen Beschreibung gegeben werden.

In Kapitel 3 wird ein systemtheoretisches Modell des frühen visuellen Systems entwickelt werden. Dieses basiert darauf, daß die lichtempfindlichen Rezeptoren in der zentralen Fovea auf einem annähernd hexagonalen Raster liegen, was zu den genannten Vorteilen führt. Bei einem hexagonalen Abtastraster wird gewöhn-



lich der kleinste Abstand

$$d_{cc} = 2T_h \quad (2.61)$$

zwischen einem Abtastpunkt zu seinen 6 Nachbarpunkten meist gesondert als *spacing* benannt (cc=center-to-center spacing). Im Kontext der Photorezeptoren (Zapfen) des menschlichen Auges spricht man dann von *cone spacing*, was dann auch oft als Maßeinheit [csp] fungiert, wenn man die verwendeten Größen auf  $d_{cc}$  normiert.

Weiter ist es oft notwendig, aus der Dichte  $\rho$  der Abtastpunkte, welche z. B. aus einer Zählung der Zapfen gewonnen wird, auf die Abtastperiode  $T_h$  bzw. das cone spacing  $d_{cc}$  zu schließen. Die entsprechende Beziehung hierfür lautet [DM84, HM87]

$$d_{cc} = 2T_h = \sqrt{\frac{2}{\sqrt{3}\rho}}. \quad (2.62)$$

Die Größe eines Objektes, welches auf die Retina projiziert wird, ist vom Abstand des Objektes vom Auge abhängig. Um diese Abhängigkeiten zu vermeiden, verwendet man meist Winkelgrößen. Im Frequenzbereich gelangt man somit von der gewöhnlichen Einheit  $1/m$  entsprechend zu der Einheit  $1/^\circ$  bzw. *cpd* oder *c/deg* (*cycles per degree*). Die Umrechnung erfolgt durch die Beziehung

$$f_{cpd} = f \frac{2\pi b}{360^\circ}. \quad (2.63)$$

Hierbei ist  $b$  der Radius der gekrümmten Netzhaut, auf die das Bild projiziert wird. Ein typischer Wert eines Menschen liegt etwa bei  $b = 16.7$  mm, d. h. ein Sehwinkel von  $1^\circ$  entspricht dann einer Strecke von etwa 0.29 mm auf der Netzhaut [ST97].

### 2.3.2 Künstliche neuronale Netzwerke

Für die Untersuchung der biologischen Abläufe, aber auch für die Umsetzung und technische Nutzbarmachung bestimmter perzeptiver Fähigkeiten des biologischen Vorbildes bedient man sich abstrakter Modelle, den sogenannten *künstlichen Neuronen* bzw. den *künstlichen neuronalen Netzwerken*. Diese stellen die Arbeitsweise des Netzes als arithmetische Rechenschritte dar, welche dann theoretisch analysiert oder auf einem Digitalrechner nachgebildet und simuliert werden können. So gibt es eine Vielzahl von unterschiedlichen Modellen, die je nach Aufgabenstellung nur gewisse Funktionen und Mechanismen der biologischen Systeme nachbilden. Bei der überwiegenden Anzahl technischer Fragestellungen kann

z. B. der Einfachheit wegen auf die pulscodierte Verarbeitung verzichtet werden, um so bereits die meisten technischen Probleme erfolgreich zu lösen. Dieser Weg wird in dieser Arbeit ab Kapitel 3 zunächst ebenfalls gewählt. Daß dies aber nicht zwingend eine zu starke Einschränkung bzw. Sackgasse sein muß, wird dann im Laufe der Arbeit ersichtlich. In Kapitel 4 wird gezeigt, wie die Modellierung in einer späteren Entwurfsphase bei Bedarf für pulscodierte Mechanismen erweitert werden kann, um so nach einer soliden Modellierung ebenfalls die Vorteile der Pulscodierung nutzen oder untersuchen zu können.

Die grundlegende Arbeitsweise eines einfachen künstlichen neuronalen Netzes ist in Abb. 2.11 schematisiert. Bei der Verarbeitung werden die ankommenden Signale der Vorgängerneuronen gewichtet und verarbeitet, wie in Abb. 2.11 gezeigt. Die Gewichte entsprechen hier der synaptischen Erregung oder Hemmung beim biologischen Vorbild entsprechend dem rezeptiven Feld des Neurons. Sind die Eingangsreize ausreichend, so wird das Neuron selbst einen Ausgangsreiz zu nachfolgenden Zielneuronen weiterleiten. Die *Propagierfunktion*

$$\text{net}_i = \sum_j o_j \cdot w_i(j) \quad (2.64)$$

eines betrachteten Neurons  $i$  ist dabei meist die gewichtete Summierung der Eingangsreize, d. h. Ausgangssignale der Vorgängerneuronen. Die Verbindungsgewichte  $w_i(j)$  zwischen dem Vorgängerneuron  $j$  und dem betrachteten Zielneuron  $i$  gibt dabei an, wie sehr das Zielneuron  $i$  von einem Vorgängerneuron erregt ( $w_i(j) > 0$ , exzitatorische Verbindung) oder gehemmt ( $w_i(j) < 0$ , inhibitorische Verbindung) wird. Der Aktivierungszustand  $a_i$  entspricht dem Membranpotential der biologischen Zelle und ist von dem vergangenen Zustand, der Netzeingabe  $\text{net}_i$  und ggf. einem Schwellwert abhängig. Der Ausgangsreiz  $o_i$  ergibt sich aus dem Aktivierungszustand der Zelle über einen linearen oder nichtlinearen funktionalen Zusammenhang.

Die dieser Arbeit zugrundeliegenden biologischen örtlichen Filtercharakteristiken bis zu den simplen Zellen weisen in weiten Bereichen ein lineares Verhalten auf (siehe Kapitel 3), so daß hier im wesentlichen die linearen Neuronen mit

$$o_i = \sum_j o_j \cdot w_i(j) \quad (2.65)$$

zum Einsatz kommen. Speziell im biologischen Sehsystem (siehe Aufbau der Retina in Abb. 2.7) sind oft viele Neuronen gleicher Gattung in Schichten angeordnet, so daß jedes Neuron mit dem Ort eines Bildbereiches korrespondiert. Hier

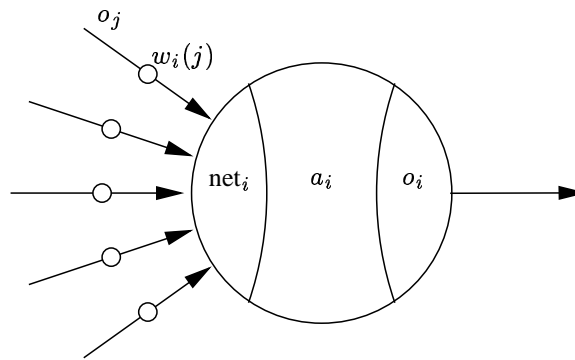


Abbildung 2.11: Schematisierung eines künstlichen Neurons und dessen Verschaltung mit Vorgängerneuronen.

empfängt jedes Neuron meist nur Eingangsreize von Neuronen einer anderen Gattung, d.h. einer anderen Schicht, was als vorwärtsgerichtetes Netz bezeichnet wird (*Feed-Forward Netz*). Ordnet man nun die Neuronen jeder Schicht jeweils in einer Matrix an und bezeichnet die Schichten mit 1 und 2, so kann die Summe in (2.65) als zwei Summen

$$o_{i_1, i_2}^{(2)} = \sum_{j_1} \sum_{j_2} o_{j_1, j_2}^{(1)} \cdot w_{i_1, i_2}(j_1, j_2) \quad (2.66)$$

für die Zeilen- und Spaltenkoordinate geschrieben werden. Darüber hinaus weisen die Verbindungsgewichte innerhalb einer Neuronenschicht ein sehr reguläres Schema auf, wodurch mit der Eigenschaft

$$w_{i_1, i_2}(i_1 - p, i_2 - q) = w_{k_1, k_2}(k_1 - p, k_2 - q) \quad \forall p, q, i_{1,2}, k_{1,2} \quad (2.67)$$

eine ortsinvariante Verarbeitung entsteht, d. h. das Verbindungsgewicht ist nicht mehr abhängig von dem Ort des Zielneurons selbst, sondern nur noch von der örtlichen Relation ( $p$  bzw.  $q$ ) zu den verbundenen Vorgängerneuronen. Wir kürzen dies für  $j_1 = i_1 - p$  und  $j_2 = i_2 - q$  mit

$$w_{i_1, i_2}(i_1 - p, i_2 - q) =: \tilde{w}(p, q) \quad (2.68)$$

ab und erhalten

$$o_{i_1, i_2}^{(2)} = \sum_p \sum_q o_{i_1 - p, i_2 - q}^{(1)} \cdot \tilde{w}(p, q). \quad (2.69)$$

Auf diese Weise geht die Berechnungsvorschrift des linearen Neurons formal in die Faltungssumme des nichtrekursiven Digitalfilters (2.30) über, und die Verbin-

dungsgewichte respektive das rezeptive Feld können als Faltungsmaske interpretiert werden. Auf ähnliche Weise läßt sich auch die Arithmetik einer rekursiv gekoppelten Neuronenstruktur mit sogenannten *lateralen Neuronenverknüpfungen* auf die Differenzengleichung eines rekursiven Digitalfilters zurückführen. Diese Zusammenhänge liefern damit wertvolle Ergebnisse für die Interpretation neuronaler Architekturen der biologischen (visuellen) Informationsverarbeitung. Hierdurch wird ersichtlich, daß lineare neuronale Netzwerke mit ihren lokal regulären Verbindungsgewichten als ortsinvariante Digitalfilter interpretiert werden können. Dies birgt zum einen für die Simulation der Netzwerke große Vorteile, da verschiedenste Methoden der effektiven Filterimplementierung auf Digitalrechnern und Signalprozessoren zur Verfügung stehen. Zudem kann hier von der großen Anzahl an Analysemethoden und Entwurfsverfahren im Orts- und Frequenzbereich profitiert werden, um die Modellierung neuronaler Netze zu unterstützen. Aus diesen Gründen bildet auch die systemtheoretische Betrachtung des biologischen Sehsystems die Grundlage der vorliegenden Arbeit.

# Lineares Modell der frühen visuellen Verarbeitung

In diesem Kapitel wird ein Modell auf der Basis eines hexagonalen Rasters vorgestellt, welches viele informationsverarbeitende Funktionen beinhaltet, die bei den Sehsystemen höherer Lebewesen erbracht werden [TH00, Thi99]. Hierbei betrachten wir die Verarbeitungskette von den Photorezeptoren in der Netzhaut (Retina) über die retinalen Horizontal- und Ganglienzellen bis hin zu den simplen Zellen in der Großhirnrinde (visueller Cortex). Dieses Modell hat u. a. die Eigenschaft, daß es massiv parallel arbeiten kann. Zudem werden aber auch wertvolle Zwischensignale produziert, die in weiteren Verarbeitungsaufgaben nützlich sind. Hier seien Blicksteuerungskonzepte, Adaptionsmechanismen bei Beleuchtungsänderungen oder die Komposition simpler Zellen anderer Auflösungsstufen genannt. Die Eigenschaften des Modells zeigen auf, wie effizient die Verarbeitungsprozesse im menschlichen Gehirn aufgebaut sein könnten.

Die Verarbeitung der visuellen Signale wird in dieser Arbeit systemtheoretisch modelliert, d.h. das Verbindungsnetzwerk zwischen Schichten meist gleichartiger Neuronen wird als lineares System mit Eingangs- und Ausgangsgrößen dargestellt. Da hierdurch, wie im vorangegangenen Kapitel aufgezeigt, die Signalverarbeitung solch eines Neuronennetzwerkes als Digitalfilter interpretiert werden kann, ergeben sich wesentliche Vorteile im Kontext dieser Arbeit. Die Beschreibung der Signalverarbeitung im Frequenzbereich und die Vielzahl an etablierten Verfahren zur Synthese bzw. zum Entwurf von Digitalfiltern sind hier besonders erwähnenswert. Darüber hinaus ist aber auch die Simulation auf Digitalrechnern eines systemtheoretisch vorliegenden Modells eines Neuronalen Netzwerkes mit Standardoperationen der Digitalen Signalverarbeitung leicht möglich.

Im folgenden Abschnitt wird zunächst die Umsetzung der visuellen Informationen einer betrachteten Szene durch die lichtempfindlichen Rezeptoren im Auge betrachtet, um einige wesentliche Konsequenzen der weiteren Signalverarbeitung herauszuarbeiten. In den beiden folgenden Abschnitten wird die Topologie des Modells mit den zugehörigen optimalen Modellparametern entwickelt, welche zu den gewünschten Verarbeitungseigenschaften bzw. rezeptiven Feldern des biologischen neuronalen Netzwerkes führt. Abschließend werden die wesentlichen Eigenschaften des vorgestellten Modells zusammengefaßt, eine Simulation mit einem Eingangsreiz durchgeführt und gezeigt, wie die Netzstruktur und damit die Anzahl der Neuronenverbindungen reduziert werden kann. Letzteres ist für die effiziente Simulation des Netzwerkes auf gewisser Spezialhardware für Neuronale Netze nützlich. So wird im nächsten Kapitel 4 die Implementation des Modells auf eine in der Arbeitsgruppe HARTMANN entwickelte Neurocomputerarchitektur beschrieben.

### **3.1 Bildgewinnung in einem biologischen Sehsystem**

In diesem Abschnitt wird zunächst die Bildentstehung auf der menschlichen Netzhaut betrachtet und die wesentlichen Konsequenzen für die weitere Modellierung aufgezeigt. Zum einen wird eine hexagonale örtliche Abtastung mittels der lichtempfindlichen Photorezeptoren, wie sie in der Biologie beobachtet wurde, motiviert. Dieses hexagonale Abtastraster soll dann Grundlage für das zu entwickelnde visuelle Verarbeitungssystem sein. Zum anderen wird der Einfluß der Rezeptorapertur untersucht und die Betrachtung dieser im ortsdiskreten Bereich begründet.

#### **3.1.1 Abbildungseigenschaften des Auges**

Die Photorezeptoren auf der Retina tasten das kontinuierliche Bild ab, welches durch das Auge als „optisches Übertragungssystem“ auf die Retina projiziert wird. In den meisten Fällen handelt es sich um ein lineares und ortsinvariantes Bildentstehungssystem, wie in Abb. 3.1 dargestellt. Die Abbildungseigenschaften des Auges werden dabei im wesentlichen durch die folgenden Effekte bestimmt:

- Beugungseffekte  
FRAUNHOFERSche Beugung an einer Lochblende
- sphärische Aberrationen  
Breckkraft der Pupille ist abhängig vom Eintrittsort des Lichtes

- chromatische Aberrationen

Brechkraft der Pupille ist abhängig von der Wellenlänge des Lichtes



Abbildung 3.1: Das Auge als optisches Übertragungssystem.

Ist der Pupillendurchmesser sehr klein ( $a < 2 \text{ mm}$ ), so sind die letzten beiden Effekte vernachlässigbar und es handelt sich näherungsweise um ein rein beugungsbegrenztes System. Das Übertragungsverhalten des Systems wird dann durch einen Beugungsprozeß bestimmt und kann mit Hilfe der *FOURIER-Optik* berechnet werden. Wir betrachten hierzu Abb. 3.2, in der das abstrahierte Auge und die schematisierte Beugungsfigur zu sehen sind. Trifft eine ebene Lichtwellenfront auf die Pupille (Lochblende), so handelt es sich um die Wellenfront eines unendlich weit entfernten Lichtpunktes, und die entstehende Beugungsfigur ist die Impulsantwort des Systems. In [KF88, Hec98] ist gezeigt, daß sich die kreis-

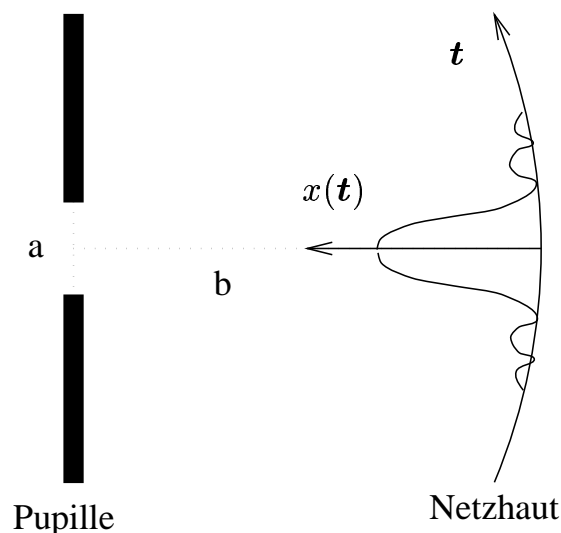


Abbildung 3.2: Beugungsfigur bei der Bildentstehung im Auge.

symmetrische Impulsantwort des Systems zu

$$h_O(\mathbf{t}) = h_O(\mathbf{0}) \left( \frac{2J_1\left(\frac{\pi}{\lambda} \frac{a}{b} \|\mathbf{t}\|\right)}{\frac{\pi}{\lambda} \frac{a}{b} \|\mathbf{t}\|} \right)^2 \quad (3.1)$$

ergibt, wobei  $J_1$  die Besselfunktion erster Gattung mit der Ordnung eins und  $\lambda$  die Wellenlänge des verwendeten Lichtes ist. Der konstante Faktor  $h_O(\mathbf{0})$  hängt u. a. von der Lichtstärke. Die ebenfalls kreissymmetrische Übertragungsfunktion des Systems kann zu

$$H_O(\boldsymbol{\omega}) = \frac{4h_O(\mathbf{0})}{\left(\frac{\pi}{\lambda} \frac{a}{b}\right)^2} (2\beta(\boldsymbol{\omega}) - \sin(2\beta(\boldsymbol{\omega}))) \quad , \quad \beta(\boldsymbol{\omega}) = \arccos\left(\frac{\|\boldsymbol{\omega}\|}{2\frac{\pi}{\lambda} \frac{a}{b}}\right) . \quad (3.2)$$

hergeleitet werden und wird in der Optik als *MTF (modulation transfer function)* bezeichnet.

In Abb. 3.3 ist u. a. das theoretische Übertragungsverhalten für verschiedene Pupillendurchmesser für ein rein beugungsbegrenztes System (—) nach (3.1) und (3.2) bei  $\lambda = 500 \text{ nm}$  dargestellt. Ebenfalls hier zu sehen ist das reale Verhalten der Augenoptik, welches experimentell bestimmt werden kann [Hau94] (vgl. auch [ILGA98, Bro79, CG66]). Es ist erkennbar, daß für größere Pupillenöffnungen ( $a > 2 \text{ mm}$ ) die experimentellen Ergebnisse immer mehr von dem rein beugungsbegrenzten Verhalten abweichen. Die Abberationseffekte können folglich nur bei kleinen Pupillenöffnungen vernachlässigt werden, da sie offensichtlich im Gegensatz zum Beugungsverhalten mit zunehmender Öffnung ein stärkeres Tiefpaßverhalten bzw. eine abnehmende Grenzfrequenz verursachen.

Die wahre (experimentelle) Übertragungsfunktion soll deshalb an dieser Stelle mit einer rotationssymmetrischen Exponentialfunktion der Form

$$H_O(\boldsymbol{\omega}) = e^{-\alpha \|\boldsymbol{\omega}\|} \quad , \quad \alpha \in \mathbb{R}^{>0} \quad (3.3)$$

beschrieben werden. Dies stellt eine gute Näherung der experimentellen Daten in [Hau94] dar. Die Form und die Ausdehnung der Impulsantwort des Systems, welche die Intensitätsverteilung bei einer punktförmigen Lichtquelle beschreibt, ist für weitere Schlußfolgerungen von Interesse. Um eine analytische Form zu erhalten, sind mit Hilfe von [GH73] einige Rechenschritte notwendig, die hier allerdings nicht im einzelnen gezeigt werden können. Als Resultat der Herleitung



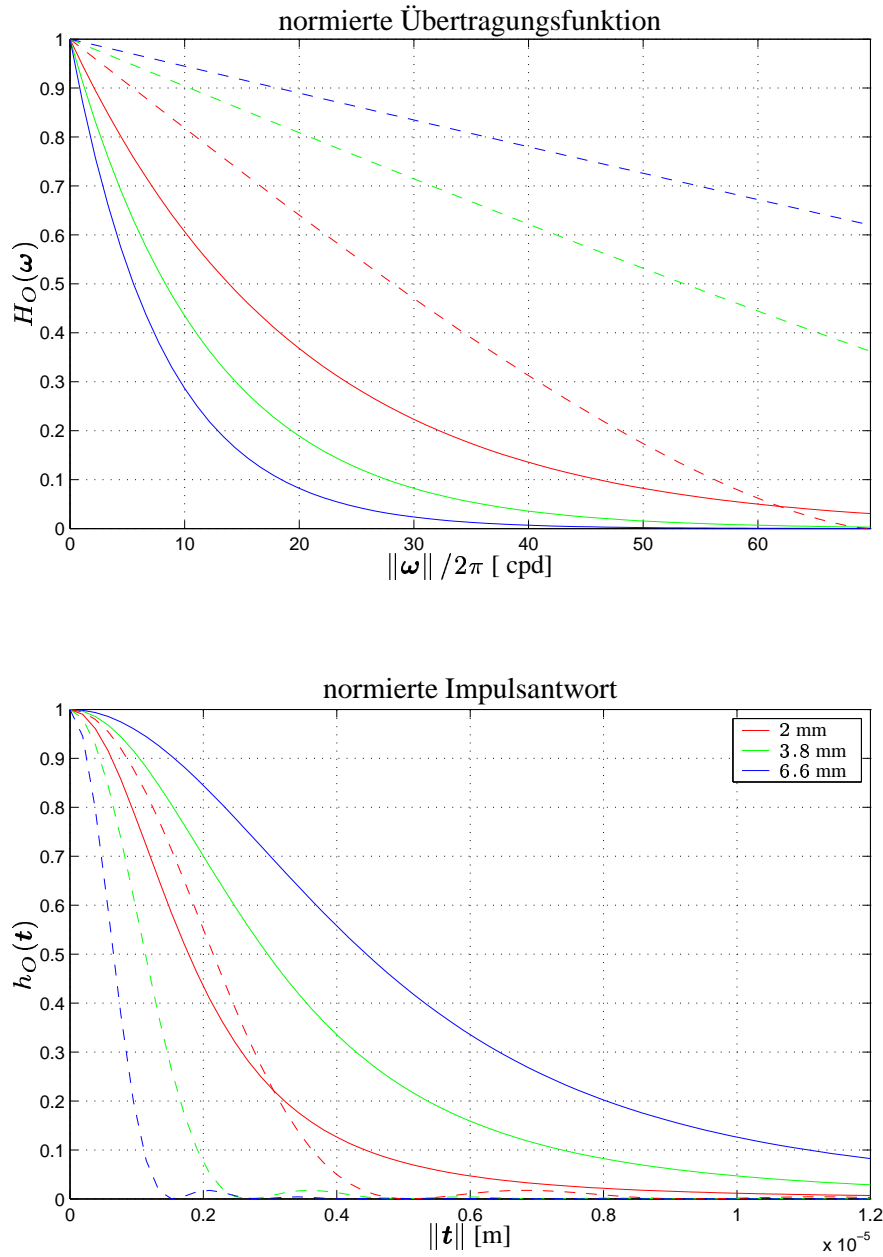


Abbildung 3.3: Übertragungsverhalten der Augenoptik verschiedener Pupillendurchmesser basierend auf experimentellen Untersuchungen (-) und die theoretischen Kurven für ein rein beugungsbegrenztes System (- -), jeweils bei einer Wellenlänge von  $\lambda = 500$  nm des verwendeten Lichtes.

ergibt sich die Impulsantwort zu

$$h_O(\mathbf{t}) = \frac{1}{2\pi} \frac{1}{\alpha^2 + \|\mathbf{t}\|^2} \sum_{\nu=0}^{\infty} \frac{(2\nu)!}{(\nu!)^2} \frac{1}{(1-2\nu)2^{2\nu}} \left( \frac{\|\mathbf{t}\|^2}{\alpha^2 + \|\mathbf{t}\|^2} \right)^{\nu} \quad (3.4)$$

$$= \frac{1}{2\pi} \frac{1}{\alpha^2 + \|\mathbf{t}\|^2} \frac{\left( \alpha^2 + \alpha \sqrt{\alpha^2 + \|\mathbf{t}\|^2} \right)}{\left( \alpha^2 + \alpha \sqrt{\alpha^2 + \|\mathbf{t}\|^2} \right) + \|\mathbf{t}\|^2}, \quad (3.5)$$

welche ebenfalls rotationssymmetrisch ist. Mit Hilfe der Impulsantwort könnte u. a. das Auflösungsvermögen des Auges untersucht werden [KF88, Hec98]. Im Rahmen dieser Arbeit stellen sich aber das Verhalten der Optik im Frequenzbereich und die daraus resultierenden Konsequenzen für eine örtliche Abtastung durch die Photorezeptoren in den Vordergrund. Im folgenden werden dazu einige Betrachtungen angestellt und mit verschiedenen, in der Literatur zu findenden experimentellen Ergebnissen verglichen.

Da es sich bei dem experimentell bestimmten Verhalten im Frequenzbereich und bei unserem Ansatz um ein kreissymmetrisches Tiefpaßverhalten handelt, ist folglich jedes auf die Netzhaut projizierte kontinuierliche Signal immer näherungsweise kreisförmig frequenzbegrenzt. Wie aus Kapitel 2 bekannt, wäre für diesen Fall eine hexagonale Abtastung des kontinuierlichen Signals die effektivste Form der Ortsdiskreten Repräsentation. Unter der Annahme, daß das biologische Vorbild hoch entwickelt sei und daher effizient arbeite, kann an dieser Stelle die Hypothese geäußert werden, daß sich im Laufe der Evolution eine hexagonale oder zumindest hexagonalähnliche Anordnung der Photorezeptoren ausgebildet haben könnte. In der Tat ist in zahlreichen Untersuchungen gezeigt worden, daß die Anordnung der Zapfen in der Fovea der Retina eine sehr hohe Regularität aufweist [Wil85, WC83, Wil86, WC87, Wil88, HM87, Yel81]. Es konnte in diesen Arbeiten darüber hinaus festgestellt werden, daß das Mosaik der Zapfen in sehr guter Näherung einem hexagonalem Abtastraster in der Fovea entspricht (siehe Abb. 3.4). Durch Zählung der Zapfen ist z.B. eine Dichte  $\rho = 14.7341/100 \mu\text{m}^2$

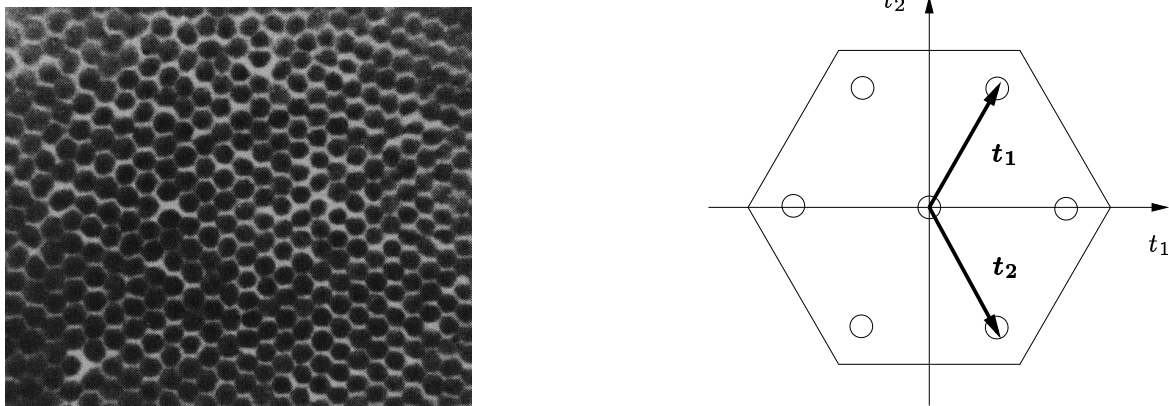


Abbildung 3.4: Anordnung der Photorezeptoren (Zapfen) in der *Fovea centralis* eines Affens [HM87]. Die hexagonale Gitterstruktur ist deutlich erkennbar.

ermittelt worden, was nach (2.62) einem mittleren Zapfenabstand (cone spacing)  $d_{cc} = 2.8 \mu\text{m}$  entspricht. Um das Abtasttheorem von SHANNON zu erfüllen, ergä-

be sich dann mit (2.45) eine Nyquistfrequenz von  $f_{\varrho, \max} = 56$  cpd. Da das kontinuierliche Bild auf der Retina, wie aus Abb. 3.3 (reales Verhalten) hervorgeht, keine nennenswerten Frequenzanteile über 50 cpd aufweisen kann, sind keine oder nur kaum Aliasingfehler in der *zentralen Fovea* (*Fovea centralis*) zu vermuten. Diese Aussage wird in der Tat durch die o. g. experimentellen Untersuchungen über Aliaseffekte, vor allem von WILLIAMS [Wil85], bestätigt.

Außerhalb der Fovea nimmt die Unregelmäßigkeit des Mosaiks allerdings zu. Berechnet man dennoch die theoretische Nyquistfrequenz aus der Zapfendichte [Yel81] (Rhesusaffe) unter Annahme eines hexagonalen Rasters, so ergeben sich die Werte in der folgenden Tabelle:

Exzentrizität	0°	6°	35°
Zapfendichte [ $1/100 \mu\text{m}^2$ ]	19.4	1.4	0.6
Nyquistfrequenz $f_{\varrho, \max}$ [cpd]	53	14	9.5
mittl. Zapfenabstand [csp]	2.5		

Ab der *Parafovea* (6°) müßten demnach Aliasingeffekte auftreten, was aber nicht in dem Maße der Fall ist. Wie in [Yel81] gefolgert wird, beugt das Sehsystem offensichtlich Aliaseffekten mit Hilfe der Irregularität des Mosaiks vor, da eine Frequenz  $> f_{\varrho, \max}$  keine konkrete Spiegelfrequenz ergibt, sondern durch das unregelmäßige Abtastraster zu einem gestreuten breitbandigen Rauschen führt. Hieraus läßt sich aber auch folgern, daß das biologische Vorbild außerhalb der Fovea auf ein periodisches Abtastraster verzichtet und auf eine unregelmäßige, zufällige Verteilung übergeht, um weniger Rezeptoren spendieren und damit weniger Daten verarbeiten zu müssen. Eine Interpretation, die die Leistungsfähigkeit eines biologischen Sehsystems unterstreicht. Darüber hinaus könnte diese Erkenntnis durchaus auch für technische Systeme von Interesse sein. Dies wäre eine Alternative zu anderen inhomogenen Abtaststrategien (z.B. logarithmisch-polare Abtastraster), so daß vergleichende Untersuchungen lohnen könnten.

### 3.1.2 Abtastung mittels Photorezeptoren

Die Abtastung durch die lichtempfindlichen Rezeptoren in der Netzhaut soll nun systematisch betrachtet werden, um die erste Stufe eines Modells zur biologienahen visuellen Verarbeitung schrittweise zu entwickeln. Die Photorezeptoren der Retina absorbieren die einfallenden Lichtquanten und setzen diese in ein elektrisches Potential für die Weiterverarbeitung um. Da die Rezeptoren auf einem diskreten Raster angeordnet sind, sorgen sie also für eine örtliche Diskretisierung des kontinuierlichen Signals auf der Retina und dienen somit im wesentlichen der

Abtastung. Das Abtasttheorem von SHANNON geht gewöhnlich davon aus, daß ein kontinuierliches Signal mit einem idealen Impulskamm, d. h. mit unendlich kleiner *Apertur* abgetastet wird. Dies ist aber in der Natur und bei technischen Systemen niemals der Fall. So besitzt zum einen die Fläche eines einzelnen Rezeptors – sei es auf der Netzhaut oder auch auf einem CCD-Chip – immer eine endliche Ausdehnung und zudem fällt stets Streulicht seitlich in den Rezeptor ein. Die resultierende *Abtastapertur* bzw. das *Empfindlichkeitsprofil* der Photorezeptoren ist ebenfalls endlich und bewirkt somit vor der eigentlichen Abtastung einen Tiefpaßeffekt durch eine gewisse örtliche Mittelung der Lichtintensität. Der Vorgang der Abtastung ist in Abb. 3.5 als Signalflußgraph schematisiert.

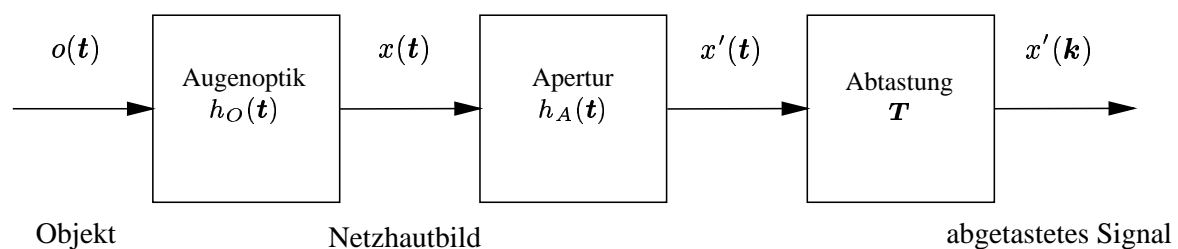


Abbildung 3.5: Abtastung des Retinabildes mit vorheriger Aperturfilterung.

Der Tiefpaßeffekt der Apertur kann in einem Abtastsystem unter Umständen als *Anti-Aliasing-Filter* dienen, was auch beim menschlichen Sehsystem auf den ersten Blick plausibel erscheint. Deshalb soll zunächst kurz geklärt werden, wie diese Aperturfunktion der Rezeptoren aussieht und inwieweit sie für die visuelle Signalverarbeitung von Bedeutung ist. Durch Untersuchungen von Aliaseffekten kann ansatzweise das Frequenzverhalten der einzelnen Rezeptoren im menschlichen Auge geschätzt werden. Die physiologischen Experimente in [Wil85], bei denen der Einfluß der Augenoptik durch die direkte Projektion von Interferenzmustern auf die Retina kompensiert wird, lassen z. B. den Schluß zu, daß die Apertur der Rezeptoren einen Tiefpaßcharakter mit hoher Grenzfrequenz und einer Nullstelle bei ca. 150 ... 160 cpd besitzt. Auch Modellierungen der Zapfen als dielektrische Lichtwellenleiter [SP94] weisen ein Tiefpaßverhalten mit einer recht langsam abfallenden Frequenzcharakteristik und einer Nullstelle bei hohen Frequenzen auf. Diese Untersuchungen widerlegen die oft anzutreffende Vereinfachung einer gaußförmigen Apertur mit hohem Überlappingsgrad benachbarter Rezeptoren. Im Gegenteil, die Ergebnisse deuten auf sehr schmale Aperturfunktionen in der Größenordnung der Rezeptorausdehnung hin, wodurch die Apertur eine eher untergeordnete Rolle bei dem Abtastvorgang zu spielen scheint. Falls dennoch der Einfachheit wegen eine gaußförmige Apertur bei der Modellbildung

zum Einsatz kommt, so kann demnach nur eine entsprechend geringe Überlappung, d. h. hohe Grenzfrequenz die biologischen Verhältnisse widerspiegeln. In [TWCH91] wird beispielsweise für die gaußförmige Aperturfunktion eine Standardabweichung von  $\sigma_{\text{Apertur}} = 17.6'' \approx 1.4 \mu\text{m}$  im Ortsbereich verwendet, was im Frequenzbereich einer Standardabweichung von 32.5 cpd entspricht (Exzentrizität  $3.8^\circ$ ). Hier könnte auch ein Vergleich mit [HM98] von Interesse sein.

Wie bereits erwähnt, treten bei der Abtastung in der Fovea keine oder nur kaum Aliasingfehler auf, d. h. die Tiefpaßwirkung der Augenoptik ist für das vorhandene Abtastraster bereits ausreichend, um das Abtasttheorem zu erfüllen. Es ist somit davon auszugehen, daß der Apertur der Rezeptoren zumindest bei der Abtastung keine große Bedeutung zukommt, und es ist daher möglich und zweckmäßig, den Einfluß der Apertur im diskreten Bereich zu betrachten, was in Abb. 3.6 verdeutlicht ist.

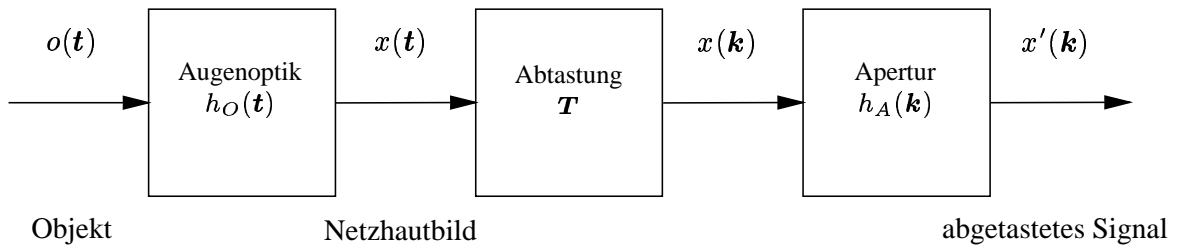


Abbildung 3.6: Ideale Abtastung des Retinabildes und Berücksichtigung der Apertur im diskreten Bereich.

Im weiteren ist darauf hinzuweisen, daß die Aperturfunktion der einzelnen Rezeptoren nicht zu verwechseln ist mit dem Übertragungsverhalten, d. h. mit der Impulsantwort oder dem rezeptiven Feld der Rezeptoren. Hier spielt nämlich eine laterale Kopplung benachbarter Rezeptoren eine wichtige Rolle, die in Verbindung mit der Apertur ein resultierendes rezeptives Feld erzeugt. Die Kopplung kann mit Hilfe eines weiteren Filters beschrieben werden, was in Abb. 3.7 zu sehen ist.

Das resultierende rezeptive Feld eines Rezeptors ergibt sich mit

$$h_R(k) = h_A(k) * h_K(k) \quad (3.6)$$

als Kaskadierung des Apertur- und des Kopplungsfilters und ist die Impulsantwort des Systemverhaltens zwischen dem ideal abgetasteten Signal  $x(k)$  und dem Signal  $x_R(k)$  am Ausgang des Rezeptors, was in Abb. 3.8 symbolisiert ist.

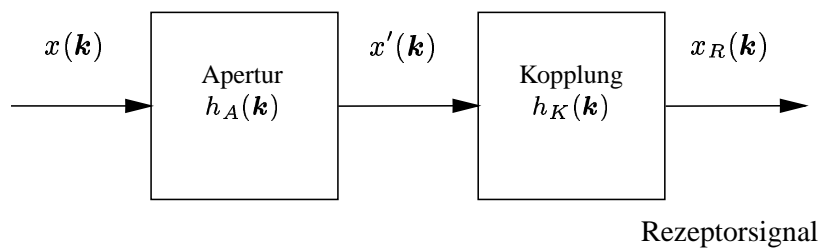


Abbildung 3.7: Einfluß der lateralen Kopplung benachbarter Rezeptoren.

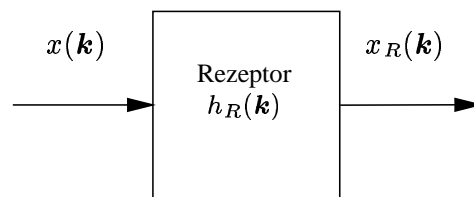


Abbildung 3.8: Zusammenfassung der Apertur und der lateralen Kopplung zur Impulsantwort eines Rezeptors.

In [SS90] wird das rezeptive Feld der Zapfen einer Katzenretina durch die Differenz zweier Gaußfunktionen (*DoG, difference of gaussian*) modelliert. Meist ist die Beschreibung des Übertragungsverhaltens mit Hilfe einer Gaußfunktion aber ausreichend [SL96a, SL96b]. Diese Näherung soll auch im Rahmen dieser Arbeit für die Modellierung der retinalen Signalverarbeitung als Grundlage dienen.

## 3.2 Modellierung retinaler Neuronen

Die Verschaltung der Zellen selbst zu Beginn der Verarbeitungskette in der Retina ist in biologischen Sehsystemen sehr komplex und bis heute noch nicht vollständig verstanden. Obwohl in dieser Arbeit eine biologienahe Betrachtung und Modellierung vorgenommen wird, sollen lediglich die für uns wichtigen Funktionen der retinalen Informationsverarbeitung berücksichtigt werden. So liegt der Schwerpunkt auf der Nachbildung rezeptiver Felder, also auf der örtlichen Filterwirkung. Diese Filtereffekte sind auch für technische Bildverarbeitungssysteme in Form von Digitalfilterbänken von großem Interesse, um diverse Vorverarbeitungsschritte, wie Rauschunterdrückung oder Merkmalsextraktion bei statischen Szenen zu realisieren. Zeitliche Vorgänge und ebenso nichtlineare Effekte sollen bewußt nicht in das Modell einfließen.

Der schematische Aufbau des Retinamodells ist in Abb. 3.9 dargestellt. Wesentliche Bestandteile der Verarbeitungskette sind die Rezeptoren, Horizontalzellen und Ganglienzellen. In der Fovea besteht darüber hinaus eine eins-zu-eins Verbindung von den Bipolarzellen zu den Ganglienzellen mit nur wenig Einfluß der Amakrinzellen [SL96a]. Deshalb ist das grundlegende Übertragungsverhalten bzgl. der örtlichen Filterung und damit das rezeptive Feld der Ganglienzellen und der Bipolarzellen näherungsweise gleich, weshalb die Bipolar- und Amakrinzellen hier nicht betrachtet werden sollen.

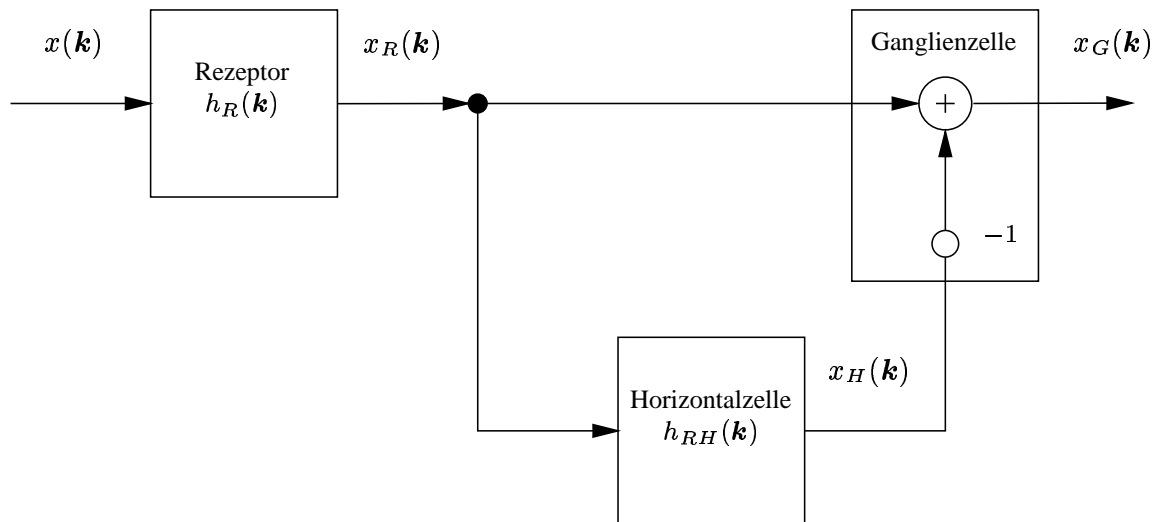


Abbildung 3.9: Modellierung der retinalen Verarbeitung durch Rezeptoren, Horizontal- und Ganglienzellen, die in Zellschichten organisiert sind.

### 3.2.1 Photorezeptoren

In Anlehnung an [SL96a, SL96b] beschränken wir uns hier auf das rot-grün-Zapfensystem in der zentralen Fovea, da dieses im Gegensatz zu dem Stäbchensystem die höchste örtliche Auflösung ermöglicht. In diesem Bereich der Retina sind keine oder nur kaum blaue Zapfen vorhanden. Betrachtet man nämlich die Absorptionskurven der drei Zapfenarten in Abb. 3.10 [Hau94] (siehe auch [ST97]), welche die spektrale Empfindlichkeit der Rezeptoren widerspiegeln, so ist erkennbar, daß die Kurven der roten und grünen Rezeptoren einen weiten Spektralbereich umfassen und sich auch relativ stark überlappen. Aus diesem Grund kann davon ausgegangen werden, daß das Signal dieses Zapfensystems auch genügend nutzbare Intensitätsinformationen beinhaltet, auch wenn der Farb- und Intensitätsgehalt gemischt in einem Pfad zu höheren Ebenen des Gehirns übertragen wird. Die Rezeptoren sollen zudem, wie im vorangegangenen Abschnitt motiviert, ein

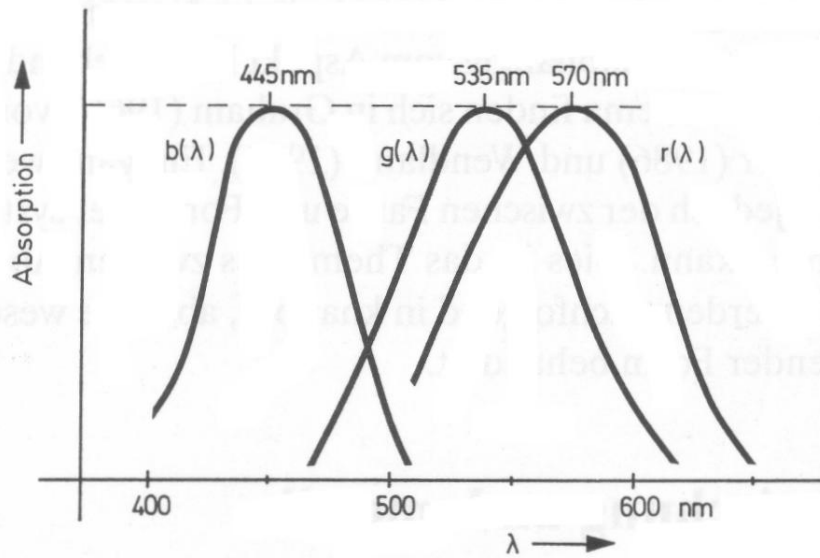


Abbildung 3.10: Absorptionskurven der Zapfen. Die spektrale Empfindlichkeit der roten und grünen Rezeptoren weisen eine große Überlappung auf.

hexagonales Abtastraster mit einer Abtastperiode von  $T_h = 1.25 \mu\text{m}$ , also einem cone spacing von  $d_{cc} = 2.5 \mu\text{m}$  bilden. Dies stimmt gut mit zahlreichen Untersuchungen [WC83, WC87, Wil88, HM87, Yel81] und den Ergebnissen aus Abschnitt 3.1 überein.

Wie im Abschnitt über den Abtastvorgang durch die Rezeptoren motiviert, kann das rezeptive Feld bzw. die Impulsantwort der Zapfen

$$h_R(\mathbf{k}) = \frac{1}{\left(\sqrt{2\pi\sigma_R^2}\right)^2} e^{-\frac{\|\mathbf{T}\mathbf{k}\|^2}{2\sigma_R^2}} \quad (3.7)$$

in guter Näherung als gaußförmig angenommen werden. Im Frequenzbereich ergibt sich durch die örtliche Diskretisierung eine periodisch fortgesetzte Gaußfunktion. Im interessierenden Nyquistbereich erhalten wir entsprechend approximativ die rein reelle Übertragungsfunktion des Systems

$$H_R(\mathbf{v}) \approx \frac{1}{|\det \mathbf{T}|} e^{-\frac{\|\mathbf{T}^{-T}\mathbf{v}\|^2 \sigma_R^2}{2}} \quad , \quad -\pi \leq v_{1,2} \leq \pi. \quad (3.8)$$

Es ist weiter bekannt, daß mit abnehmender mittlerer Beleuchtung die Größe der rezeptiven Felder adaptiv zunimmt, um für eine größere Lichtausbeute bzw. ein besseres Signal-zu-Rauschverhältnis zu sorgen, da auf diese Weise die Rauschkomponente stärker gemittelt wird. Natürlich verschlechtert sich hierdurch auch



das Auflösungsvermögen des Auges. Allerdings wird dieser Nachteil ein wenig mit der Tatsache kompensiert, daß bei schlechteren Lichtverhältnissen auch die Pupillenöffnung und damit auch die Impulsantwort der Augenoptik geweitet wird (siehe Abb. 3.3). So könnte sich nach [SL96a] die Standardabweichung  $\sigma_R$  z. B. in einem Bereich

$$\sigma_R = 1.5 \dots 12 \text{ csp} \quad (3.9)$$

bewegen. Auch im Frequenzbereich ist zu erkennen, daß die Übertragungsfunktion des Rezeptors an die Augenoptik angepaßt ist, da sich nämlich für die Standardabweichung von  $H_R(\mathbf{v})$  Werte zwischen 12 cpd (starke Beleuchtung) und 1.6 cpd (schwache Beleuchtung) ergeben würden. In der vorliegenden Modellierung lassen wir zunächst aber zur Vereinfachung keine Adaption zu und verwenden einen festen Wert von  $\sigma_R = 1.5 \text{ csp}$ , wie bei einer starken Beleuchtung. Erst in Kapitel 5 bzw. Abschnitt 5.4 gehen wir auf ein adaptives Verhalten der Rezeptoren ein. In Abb. 3.11 ist das sich ergebende rezeptive Feld und die Übertragungsfunktion als Funktion des Ortes

$$\mathbf{t}(\mathbf{k}) = \begin{bmatrix} t_1(\mathbf{k}) \\ t_2(\mathbf{k}) \end{bmatrix} = \mathbf{T}\mathbf{k} \quad (3.10)$$

bzw. der Frequenz

$$\mathbf{f}(\mathbf{v}) = \begin{bmatrix} f_1(\mathbf{v}) \\ f_2(\mathbf{v}) \end{bmatrix} = \frac{1}{2\pi} \mathbf{T}^{-\text{T}} \mathbf{v} \quad (3.11)$$

aufgetragen.

Abschließend ist erwähnenswert, daß das menschliche Sehsystem offensichtlich nicht versucht, die Tiefpaßverzerrungen der Augenoptik bei höheren Frequenzen bis zur Nyquistfrequenz  $50 \dots 60 \text{ cpd}$  durch die Verarbeitung in der Retina zu entzerren. Im Gegensatz zu technischen Systemen, in denen je nach Aufgabenstellung durchaus eine Kompensation der Verzerrungen des bildgebenden Prozesses nützlich sein kann, scheint das Übertragungsverhalten der Rezeptorschicht bereits eine erste Stufe in einer komplexeren Verarbeitungskette zu sein.

### 3.2.2 Horizontal- und Ganglienzellen

Die Ganglienzellen werden in drei Klassen unterteilt [ST97, DS94, Sla90]. X-Ganglienzellen ( $\beta$ -Zellen, parvozelluläre Zellen) weisen weitgehend ein lineares zeitliches und örtliches Verhalten auf. Sie sind dicht angeordnet, besitzen kleine rezeptive Felder und vermitteln hauptsächlich Farb- und Forminformationen.

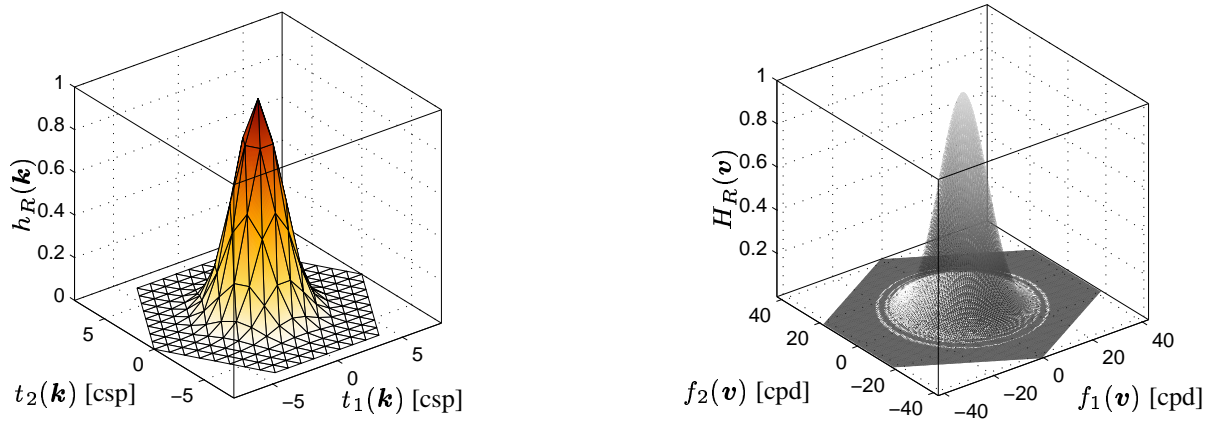


Abbildung 3.11: Normierte Impulsantwort und Übertragungsfunktion der Rezeptoren auf hexagonalem Abtastraster.

Aus diesem Grund ist dieser Zelltyp Bestandteil des Modells. Die nichtlinearen Y-Ganglienzellen ( $\alpha$ -Zellen, magnozelluläre Zellen) hingegen besitzen große rezeptive Felder und sind nicht so dicht gepackt. Sie gehören zu den bewegungsempfindlichen Zellen. Eine weitere Klasse sind die W-Ganglienzellen ( $\gamma/\delta$ -Zellen), die ebenfalls große rezeptive Felder aufweisen.

In Abb. 3.9 ist zu erkennen, daß die Ganglienzelle als Eingabe das Signal  $x_R(\mathbf{k})$  einiger Rezeptoren erhält und davon das Signal  $x_H(\mathbf{k})$  einer Horizontalzelle subtrahiert. Systemtheoretisch ist demzufolge die Ganglienzelle ein System mit zwei Eingangs- und einem Ausgangssignal  $x_G(\mathbf{k})$ . Dieses Gangliensignal berechnet sich dann zu

$$x_G(\mathbf{k}) = h_G(\mathbf{k}) * x(\mathbf{k}) \quad \text{mit} \quad h_G(\mathbf{k}) = \underbrace{[1 - h_{RH}(\mathbf{k})]}_{h_{RG}(\mathbf{k})} * h_R(\mathbf{k}), \quad (3.12)$$

wobei  $h_G(\mathbf{k})$  das rezeptive Feld der Ganglienzelle ist. Aus physiologischen Untersuchungen ist bekannt, daß das rezeptive Feld der Ganglienzellen eine Zentrum-Umfeld-Charakteristik aufweist, welche entweder als Differenz zweier Gaußfunktionen (*DoG, difference of gaussians*) oder als Laplace-Ableitung einer Gaußfunktion (*LoG, laplacian of gaussian*) approximiert werden kann. Das gewünschte rezeptive Feld (hier durch des=desired gekennzeichnet) soll im folgenden eine LoG-Funktion

$$h_{G,\text{des}}(\mathbf{k}) = -\frac{1}{\sigma_G^2} \left[ 2 - \frac{\|\mathbf{T}\mathbf{k}\|^2}{\sigma_G^2} \right] \frac{1}{\left( \sqrt{2\pi\sigma_G^2} \right)^2} e^{-\frac{\|\mathbf{T}\mathbf{k}\|^2}{2\sigma_G^2}}, \quad (3.13)$$

sein. In [Hub90] ist nachzulesen, daß das Zentrum des rezeptiven Feldes bei voller Beleuchtung etwa  $10 \mu\text{m}$  (4 csp) beträgt, was die Nullstelle des rezeptiven Feldes bei  $5 \mu\text{m}$  (2 csp) festlegt und damit zu einer Standardabweichung von

$$\sigma_G = \frac{5 \mu\text{m}}{\sqrt{2}} = \sqrt{2} \text{ csp} \quad (3.14)$$

führt. Die gewünschte Impulsantwort bzw. die Übertragungsfunktion der Ganglienzelle ist in Abb. 3.12 dargestellt.

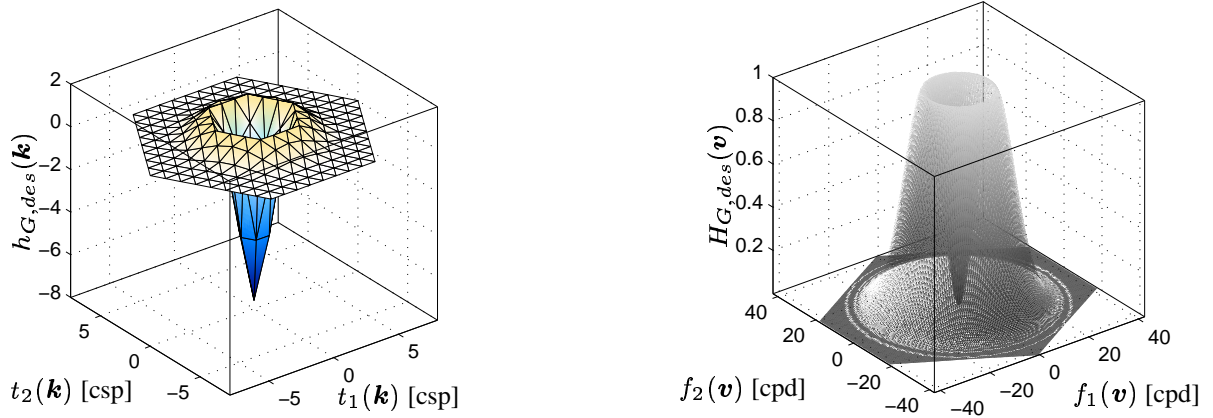


Abbildung 3.12: Normierte Impulsantwort und Übertragungsfunktion der Ganglienzellen auf hexagonalem Abtastraster.

Das retinale Modell ist nun bis auf die Verknüpfung zwischen den Rezeptoren und Horizontalzellen beschrieben, und es stellt sich die Frage, wie das gesuchte System  $h_{RG}(\mathbf{k})$  bzw.  $h_{RH}(\mathbf{k}) = 1 - h_{RG}(\mathbf{k})$  bestimmt werden muß, daß sich das gewünschte rezeptive Feld der Ganglienzelle  $h_{G,des}(\mathbf{k})$  „optimal“ ergibt. Wir machen also den Ansatz

$$\begin{aligned} h_{G,des}(\mathbf{k}) &= h_G(\mathbf{k}) \\ &= h_{RG}(\mathbf{k}) * h_R(\mathbf{k}) = \sum_{\mathbf{n}} h_{RG}(\mathbf{n}) \cdot h_R(\mathbf{k} - \mathbf{n}) \quad \forall \mathbf{k} \end{aligned} \quad (3.15)$$

und reduzieren die Doppelsumme über den vektoriellen Laufindex  $\mathbf{n}$  auf eine Einfachsumme über  $j = 1 \dots N_G$  durch Indizierung der Vektoren  $\mathbf{n}$ . Entsprechend beschreiben wir  $\mathbf{k}$  durch eine Indexmenge  $i = 1 \dots M$  und erhalten dadurch

$$h_{G,des}(\mathbf{k}_i) = \sum_{j=1}^{N_G} h_{RG}(\mathbf{n}_j) \cdot h_R(\mathbf{k}_i - \mathbf{n}_j) \quad , i = 1 \dots M. \quad (3.16)$$

Mit  $M > N_G$  liegt nun ein überbestimmtes lineares Gleichungssystem vor, welches z.B. mittels der *Pseudoinversen* im Sinne der kleinsten Fehlerquadrate gelöst werden kann. Das auf diese Weise gefundene System  $h_{RG}(\mathbf{k})$  minimiert die Fehlerquadratsumme bzw. den mittleren quadratischen Fehler (*MSE, mean squared error*)

$$F_{\text{MSE}} = \frac{1}{M} \sum_{\mathbf{k}} [h_{G,\text{des}}(\mathbf{k}) - h_G(\mathbf{k})]^2. \quad (3.17)$$

In (3.15) wird deutlich, daß das rezeptive Feld der Ganglienzelle  $h_G(\mathbf{k})$  eine lineare Überlagerung von Basisfunktionen ist. Die Basisfunktionen  $h_R(\mathbf{k} - \mathbf{n})$  sind um  $\mathbf{n}$  verschobene Versionen des rezeptiven Feldes der Rezeptoren und werden mit einem Faktor  $h_{RG}(\mathbf{n})$  gewichtet. Mathematisch betrachtet, handelt es sich demnach um die Approximation einer ortsdiskreten LoG-Funktion mittels einer Superposition von abgetasteten Gaußfunktionen. Hier ist allerdings anzumerken, daß die rezeptiven Felder auch als kontinuierliche Funktionen betrachtet und die Ermittlung der Verbindungsgewichte durch eine Approximationsaufgabe kontinuierlicher Funktionen gelöst werden können. Wenn die Funktionen aber nach SHANNON abgetastet werden dürfen (Signale sind näherungsweise bandbegrenzt), kann der Einfachheit wegen ebenso eine diskrete Approximation durchgeführt werden. Man kann hier zeigen, daß die Normalgleichungssysteme mit kontinuierlichen und abgetasteten Funktionen identisch sind und damit auf die selben Lösungen führen. Ein formaler Beweis soll der Übersichtlichkeit wegen an dieser Stelle aber ausbleiben.

Die Anzahl der Rezeptoren  $N_G$ , die auf eine Horizontalzelle konvergieren bzw. die Anzahl der zu berücksichtigenden Koeffizienten der Horizontalzelle, ist für die Approximation entsprechend zu beschränken. Es läßt sich aber schon im Vorfeld die Vermutung äußern, daß der Restfehler der Approximation umso geringer und damit das Approximationsergebnis umso besser ausfällt, je größer das  $N_G$ , also die Anzahl der Filterkoeffizienten von  $h_{RG}(\mathbf{k})$  gewählt wird. Dies scheint plausibel, da mit größerem  $N_G$  mehr Basisfunktionen für die Approximationsaufgabe zur Verfügung stehen. In Abb. 3.13, Abb. 3.14 und Abb. 3.15 werden die Ergebnisse der Approximation mit unterschiedlichem  $N_G$  vorgestellt. Für eine genauere Beurteilung sind hier die Impulsantworten als Aufsicht und zudem die Profilan-sicht in zwei Schnittrichtungen zu sehen. Die Horizontalzelle erhält jeweils einen Input von dem direkt „darüber liegenden“ Rezeptor (der den selben Ort auf der Netzhaut repräsentiert), sowie von sechs Nachbarrezeptoren erster und/oder zwölf Nachbarrezeptoren zweiter Ordnung. Der *Nachbarschaftsring* der Ordnung  $i$  sei mit  $R_i \subset \{\mathbb{Z}^2\}$  bezeichnet. In Tabelle 3.1 ist jeweils der Restfehler der Appro-

ximation angegeben, und es fällt auf, daß mit zunehmendem  $N_G$  tatsächlich der mittlere quadratische Fehler kleiner wird.

$N_G$	$\mathbf{n} \in$	Restfehler $F_{\text{MSE}}$	
7	$\{R_0; R_1\}$	$1.1059 \cdot 10^{-9}$	Abb. 3.13
13	$\{R_0; R_2\}$	$4.0212 \cdot 10^{-10}$	Abb. 3.14
19	$\{R_0; R_1; R_2\}$	$4.5345 \cdot 10^{-11}$	Abb. 3.15

Tabelle 3.1: Restfehler bei der Approximation des rezeptiven Feldes der Ganglienzelle.

Es liegt damit ein Modell zur Aufnahme und frühen Verarbeitung visueller Informationen in der Netzhaut höher entwickelter Lebewesen vor. Mit dessen Hilfe kann eine biologienahe Simulation und Signalverarbeitung über eine Kaskade von drei Teilsystemen (Rezeptor-Horizontalzelle-Ganglienzelle) stattfinden, wenn die Bildinformation auf einem hexagonalen Abtastraster vorliegt. Die Parameter des Modells konnten entweder direkt durch physiologische Untersuchungen oder mittels einer Approximationsaufgabe bestimmt werden. Im Hinblick auf eine Nutzung für ein technisches Bildverarbeitungssystem ergeben sich diverse positive Eigenschaften. Zunächst spiegelt die massive Parallelisierbarkeit der neuronalen Netze die hohe Verarbeitungsgeschwindigkeit biologischer Sehsysteme wider, was einer technischen Realisierung bei einer parallelen Verarbeitungsarchitektur grundlegende Vorteile bietet. Darüber hinaus stellt das Signal der Ganglienzelle ein wesentliches Basissignal für die „höheren“ Verarbeitungsstufen dar, da es über den Sehnerv zum Gehirn geleitet wird. Somit ist es möglich, sogenannte *Zwischensignale* zu berechnen, welche für viele komplexe Aufgaben im biologischen Sehsystem aber auch für technische Systeme nützlich sein können. Dies seien z.B. Blicksteuerungsmethoden, Adaption bei Beleuchtungsänderungen oder die effiziente Komposition simpler Zellen im visuellen Cortex des Gehirns, was im folgenden Abschnitt detaillierter beschrieben werden soll.

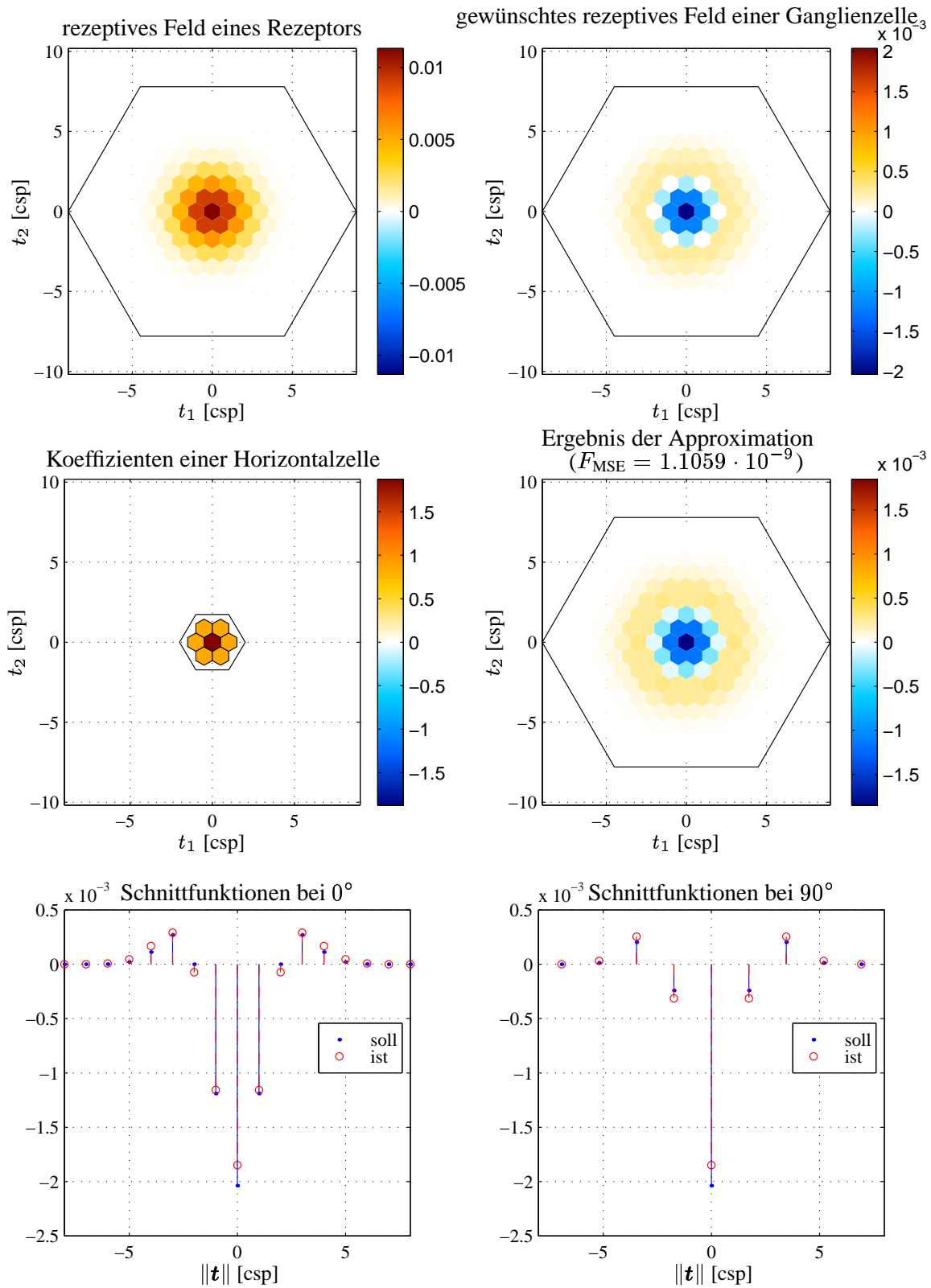


Abbildung 3.13: Approximation des rezeptiven Feldes einer Ganglienzelle, bei der  $N_G = 7$  Rezeptoren auf eine Horizontalzelle konvergieren.

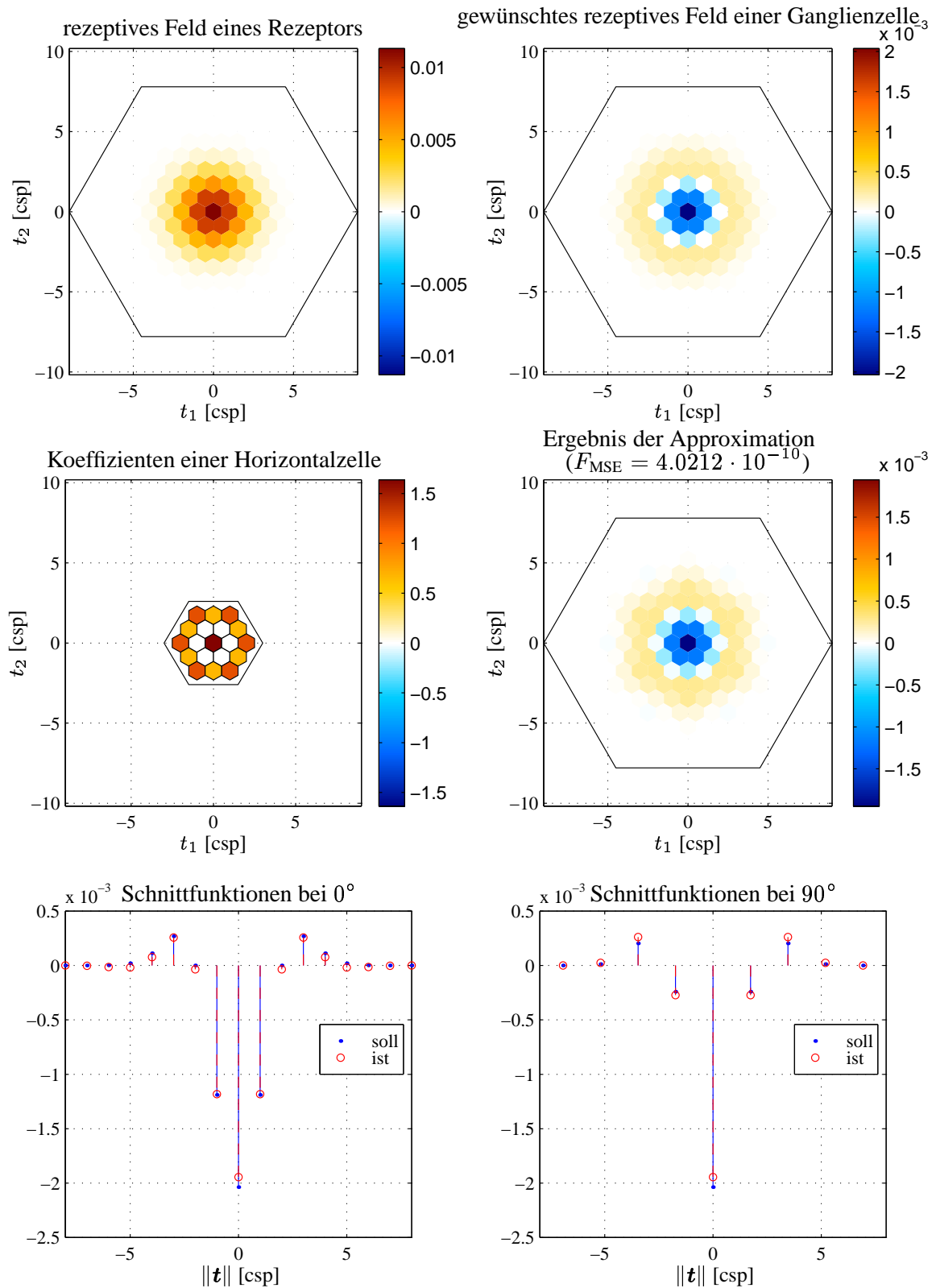


Abbildung 3.14: Approximation des rezeptiven Feldes einer Ganglienzelle, bei der  $N_G = 13$  Rezeptoren auf eine Horizontalzelle konvergieren.

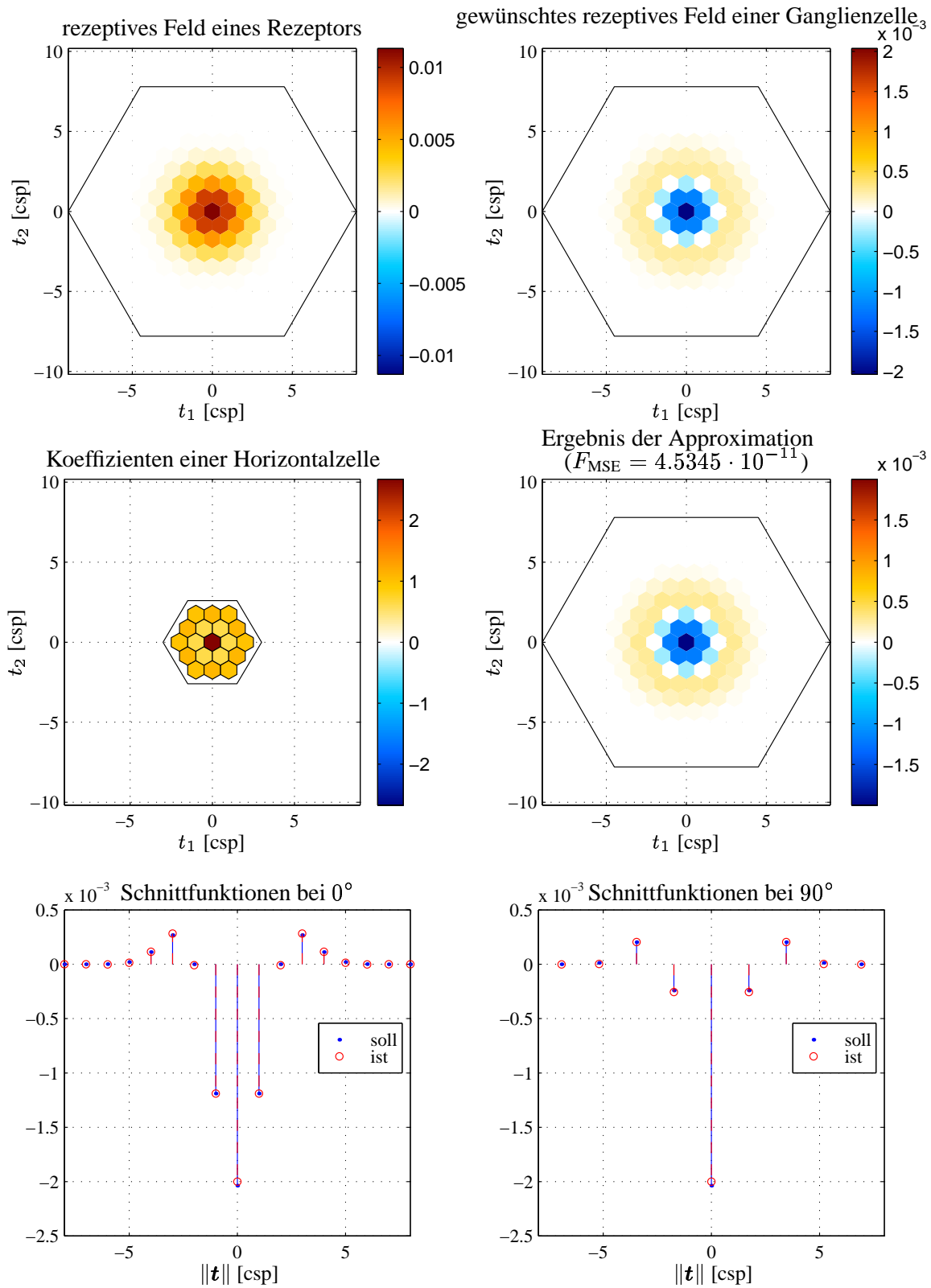


Abbildung 3.15: Approximation des rezeptiven Feldes einer Ganglienzelle, bei der  $N_G = 19$  Rezeptoren auf eine Horizontalzelle konvergieren.



### 3.3 Modellierung kortikaler simpler Neuronen

Die Ausgangssignale der Ganglienzellen höherer Lebewesen werden zunächst durch den Sehnerv zum *seitlichen Kniehöcker* (*Corpus geniculatum laterale*) übertragen. Die Verschaltung geschieht hier größtenteils eins-zu-eins, so daß diesem Ort eher die Funktion einer Signalverstärkung oder Signalverteilung als einer Signalverarbeitung zugeschrieben wird [DS94]. So kann die synaptische Übertragung an die jeweiligen Umweltbedingungen angepaßt werden, also z. B. abgeschwächt oder sogar abgeschaltet werden. Vom Kniehöcker aus bestehen zwar Verbindungen zur *Colliculi superiores*, wo Mechanismen zur Blicksteuerung ablaufen. Der größte Teil der Signale im Kniehöcker wird aber auf die Großhirnrinde, den visuellen Cortex abgebildet.

Im *primären visuellen Cortex* (genauer Area 17 oder V1) konnten Neuronen nachgewiesen werden, dessen Eingangssignale von den parvozellulären Ganglienzellen stammen und deren rezeptive Felder den sogenannten *GABOR-Funktionen* ähneln [PR81, PR82, PR83, KMB82, JP87]. Die Felder dieser *einfachen* bzw. *simplen Zellen* besitzen entweder gerade oder ungerade Profile, so daß sie entweder ausgeprägt auf Linien oder Kanten im Eingangsbild antworten. Zudem weisen jeweils genau zwei benachbarte Zellen dieses entgegengesetzte Verhalten für den selben Ort auf, so daß diese formal als (komplexwertiges) GABOR-Filter zusammengefaßt und damit annähernd als sog. *Quadraturfilterpaar* [AKM95] aufgefaßt werden können. Die Antwort der simplen Zellen ist sowohl richtungs- als auch frequenzselektiv, so daß die Gesamtheit der simplen Zellen schließlich als eine Art Bandpaß-Filterbank von Linien- und Kantendetektoren unterschiedlicher lokaler Orientierungen und Frequenzkanäle beschrieben werden kann.

Hier ist die theoretische Arbeit von SABATINI [Sab96] nennenswert, in der ein Modell simpler Neuronen auf der Basis interkortikaler inhibitorischer Kopplungen vorgestellt wird. Die rezeptiven Felder von Neuronen im *Corpus geniculatum laterale* sind hier allerdings der Einfachheit halber orientiert gaußförmig ausgeprägt. Ein betrachtetes simples Neuron erhält darüber hinaus nun weiteren inhibitorischen Einfluß aus der (nicht unmittelbaren) Umgebung des visuellen Cortex. Diese somit rückwirkenden simplen Neuronen der Umgebung liegen hier in mehreren, kreisförmigen Einzugsgebieten, wobei die synaptischen Verbindungen jeweils ebenfalls gaußförmig, aber isotrop gewählt sind. Es entsteht eine rekursiv gekoppelte Struktur simpler Neuronen, deren sich ergebenden rezeptiven Felder - je nach Umgebungseinfluß - den zweidimensionalen orientierungsselektiven GABOR-Funktionen sehr ähnlich sein können. Ziel ist es aber offenbar nicht gewesen, ein realistisches biologisches Modell des primären visuellen Cortex zu

erstellen. Die Studie zeigt vielmehr auf, welche rezeptiven Felder simpler Zellen mit der Hilfe interkortikaler Einflüsse entstehen können, wobei ein gaborähnliches Verhalten als Sonderfall betrachtet werden kann. Es bleibt hier die berechnete Frage offen, ob dieses Ergebnis immer noch gültig ist, wenn die rezeptiven Felder der Neuronen im seitlichen Kniehöcker Zentrum-Umfeld-Charakter haben, wie es bei den Gangliensignalen tatsächlich der Fall ist. Dennoch eröffnen die Betrachtungen rekursiver Strukturen neue Möglichkeiten, die es noch im Detail zu erforschen gilt. Die Interpretation als rekursive Digitalfilter können sicher auch hier hilfreich und ratsam sein.

Die hier folgende kortikale Modellierung soll aber ausschließlich nichtrekursive Strukturen verwenden, da der Filterentwurf dann analog zu dem vorangegangenen Abschnitt durchgeführt werden kann. Darüber hinaus könnten später dann immer noch die entstehenden endlichen Impulsantworten durch rekursive Digitalfilter realisiert werden, um deren Vorteile auszuschöpfen. In der vorliegenden Modellierung konvergieren eine Vielzahl von Ganglienzellen auf eine simple Zelle, deren Ausgangssignal sich somit zu

$$x_S(\mathbf{k}) = h_{GS}(\mathbf{k}) * x_G(\mathbf{k}) = \underbrace{h_{GS}(\mathbf{k}) * h_G(\mathbf{k})}_{h_S(\mathbf{k})} * x(\mathbf{k}) \quad (3.18)$$

ergibt, was in Abb. 3.16 veranschaulicht ist.

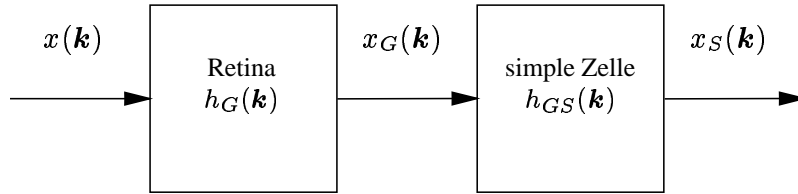


Abbildung 3.16: Modellierung der kortikalen Verarbeitung.

Es sollen hier gemäß der physiologischen Erkenntnisse beide Klassen der simplen Zellen betrachtet werden. Das gewünschte gerade und ungerade rezeptive Feld (RF) der simplen Zelle sei also

$$h_{S,\text{des}}(\mathbf{k}) = \begin{cases} \text{Re } h_{S,\text{des}}^Q(\mathbf{k}) & \text{rez. Feld gerade,} \\ \text{Im } h_{S,\text{des}}^Q(\mathbf{k}) & \text{rez. Feld ungerade} \end{cases} \quad (3.19)$$

mit

$$h_{S,\text{des}}^Q(\mathbf{k}) = \frac{1}{(\sqrt{2\pi\sigma_1\sigma_2})^2} \cdot e^{-\frac{1}{2}(\mathbf{T}\mathbf{k})^T \mathbf{R}^T \mathbf{P} \mathbf{R}(\mathbf{T}\mathbf{k})} \cdot e^{j\omega_0^T(\mathbf{T}\mathbf{k})}, \quad (3.20)$$

wobei  $h_{S,\text{des}}^Q(\mathbf{k})$  die komplexwertige Impulsantwort des Quadraturfilters sein soll. Durch eine Drehmatrix  $\mathbf{R}$  und eine Parametermatrix  $\mathbf{P}$  [Tra98] der Form

$$\mathbf{R} = \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix} \quad \text{bzw.} \quad \mathbf{P} = \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \quad (3.21)$$

handelt es sich bei den beiden rezeptiven Felder i. allg. um gedrehte ( $\phi \neq 0^\circ$ ), elliptische ( $\sigma_1 \neq \sigma_2$ ) zweidimensionale Gaußfunktionen. Im Nyquistbereich ergibt sich dann die Übertragungsfunktion

$$H_{S,\text{des}}(\mathbf{v}) = \begin{cases} \text{Ra } H_{S,\text{des}}^Q(\mathbf{v}) = \frac{1}{2} H_{S,\text{des}}^Q(\mathbf{v}) + \frac{1}{2} H_{S,\text{des}}^{Q*}(-\mathbf{v}) & \text{RF gerade,} \\ \text{Ia } H_{S,\text{des}}^Q(\mathbf{v}) = \frac{1}{2j} H_{S,\text{des}}^Q(\mathbf{v}) - \frac{1}{2j} H_{S,\text{des}}^{Q*}(-\mathbf{v}) & \text{RF ungerade} \end{cases}$$

mit

$$H_{S,\text{des}}^Q(\mathbf{v}) \approx \frac{1}{|\det \mathbf{T}|} \cdot e^{-\frac{1}{2}(\mathbf{T}^{-\text{T}}\mathbf{v}-\boldsymbol{\omega}_0)^{\text{T}}\mathbf{R}^{\text{T}}\mathbf{P}^{-1}\mathbf{R}(\mathbf{T}^{-\text{T}}\mathbf{v}-\boldsymbol{\omega}_0)} \quad , \quad -\pi \leq v_{1,2} \leq \pi,$$

deren Bandpaßcharakter in Abb. 3.17 deutlich wird.

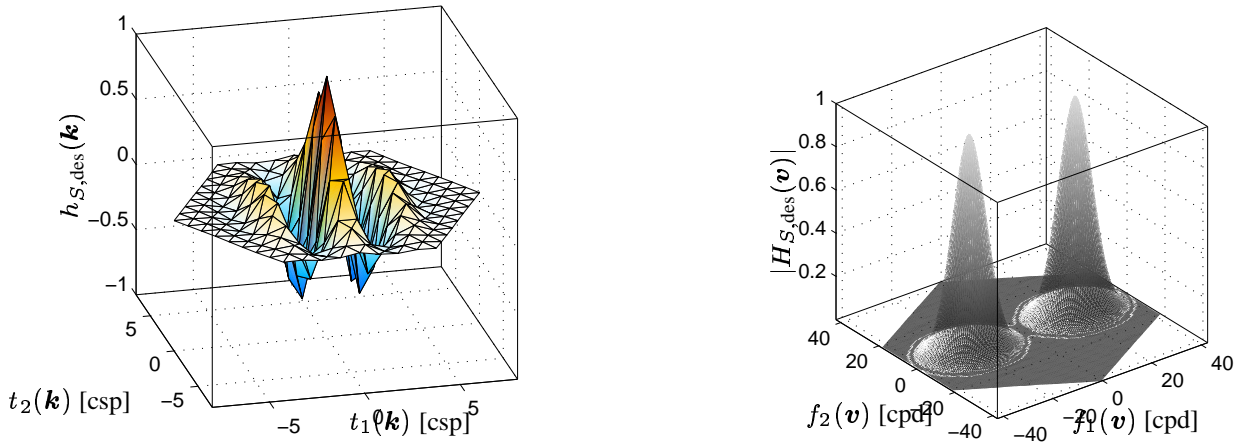


Abbildung 3.17: Normierte Impulsantwort und Übertragungsfunktion der simplen Zellen mit geradem rezeptiven Feld für das erste Richtungsband ( $\phi = 0^\circ$ ).

Die rezeptiven Felder der einfachen Zellen im Cortex von Katzen sind z. T. recht detailliert vermessen. In [JP87] beispielsweise ist die größte Modulationsfrequenz  $\|\boldsymbol{\omega}_0\|/2\pi$ , die durch Approximationsmethoden im Orts- und im Frequenzbereich bestimmt wurde, mit etwa 0.8 cpd angegeben. Im Vergleich zu den Grenzfrequenzen der bereits vorgestellten Komponenten unseres Retinamodells, wie z. B. den Ganglienzellen, scheint dieser Wert jedoch recht klein zu sein. Die Gründe hierfür

sind bislang nicht bekannt und können nur durch detailliertere Studien und physiologische Messungen geklärt werden. Um eine bzgl. des gewählten Abtastrasters möglichst gute örtliche Auflösung zu erreichen, soll im weiteren allerdings der in [Tra98] vorgestellte Filterentwurf verwendet werden, der einen weiten Frequenzbereich nutzt. In der vorliegenden Modellierung soll eine Orientierungsauflösung von  $\Delta\phi = 30^\circ$  gewählt werden, woraus sich 6 Orientierungsbänder (auf  $180^\circ$ ) für beide Typen der simplen Zellen (gerade und ungerade), also insgesamt 12 Zellschichten ergeben.

Wie bei der Bestimmung der Impulsantwort zwischen Rezeptoren und Ganglienzellen, erfolgt nun in gleicher Weise eine Approximation des rezeptiven Feldes der simplen Zellen durch Faltung des rezeptiven Feldes der Ganglienzellen mit einem noch unbekannten System  $h_{GS}(\mathbf{k})$ , also

$$h_{S,\text{des}}(\mathbf{k}) = h_S(\mathbf{k}) \quad (3.22)$$

$$= h_{GS}(\mathbf{k}) * h_G(\mathbf{k}) = \sum_{\mathbf{n}} h_{GS}(\mathbf{n}) \cdot h_G(\mathbf{k} - \mathbf{n}). \quad (3.23)$$

Nach Umschreiben der Doppelsumme in eine Einfachsumme erhalten wir für  $M > N_S$  wieder ein überbestimmtes Gleichungssystem

$$h_{S,\text{des}}(\mathbf{k}_i) = \sum_{j=1}^{N_S} h_{GS}(\mathbf{n}_j) \cdot h_G(\mathbf{k}_i - \mathbf{n}_j) \quad , i = 1 \dots M, \quad (3.24)$$

und bestimmen das System  $h_{GS}(\mathbf{k})$  so, daß der mittlere quadratische Fehler  $F_{\text{MSE}}$  minimal wird. Hierbei sei  $h_G(\mathbf{k})$  nicht das ideale rezeptive Feld der Ganglienzelle, sondern das Ergebnis der Approximation für unsere Modellierung im vorherigen Abschnitt. In Tabelle 3.2 ist der Restfehler der Approximation abzulesen, jeweils in Abhängigkeit von  $N_G$ , der Anzahl der Rezeptoren bei der Approximation der Ganglienzelle im vorherigen Kapitel. Für den Fall  $N_G = 7$  Rezeptoren, die auf eine Ganglienzelle konvergieren, und  $N_S = 37$  bzw.  $N_S = 61$  Ganglienzellen, die auf eine einfache Zelle konvergieren, sind die Approximationsergebnisse in Abb. 3.18 bzw. Abb. 3.19 dargestellt.

Analog zu der Approximation des rezeptiven Feldes der Ganglienzelle führt eine Erhöhung der Anzahl der Ganglienzellen  $N_S$ , die auf eine simple Zelle konvergieren, zu einem kleineren Restfehler. Hier konnte allerdings weiter beobachtet werden, daß eine Verbesserung des Approximationsergebnisses bei der Ganglienzelle bei festem  $N_S$  nicht unbedingt auch eine bessere Approximation für die einfache Zelle bedeutet. Erhöht man z. B. die Zahl der Rezeptoren, die auf eine Ganglienzelle konvergieren von  $N_G = 7$  auf  $N_G = 13$ , was im vorherigen Abschnitt eine

$N_G$	$N_S$	$\mathbf{n} \in$	Restfehler $F_{\text{MSE}}$	
7	37	$\{R_0; R_1; R_2; R_3\}$	$4.4015 \cdot 10^{-7}$	Abb. 3.18
	61	$\{R_0; R_1; R_2; R_3; R_4\}$	$2.8677 \cdot 10^{-8}$	Abb. 3.19
13	37	$\{R_0; R_1; R_2; R_3\}$	$1.4742 \cdot 10^{-6}$	
	61	$\{R_0; R_1; R_2; R_3; R_4\}$	$3.3335 \cdot 10^{-7}$	
19	37	$\{R_0; R_1; R_2; R_3\}$	$2.7315 \cdot 10^{-7}$	
	61	$\{R_0; R_1; R_2; R_3; R_4\}$	$1.3635 \cdot 10^{-8}$	

Tabelle 3.2: Restfehler bei der Approximation des rezeptiven Feldes der geraden simplen Zelle (basierend auf der Approximation der Ganglienzelle mit  $N_G$  Rezeptoren).

bessere Approximation bedeutete (siehe Tabelle 3.1), so verschlechtert sich für ein festes  $N_S = 37$  das Ergebnis bei der einfachen Zelle von  $F_{\text{MSE}} = 4.4015 \cdot 10^{-7}$  auf  $F_{\text{MSE}} = 1.4742 \cdot 10^{-6}$ . Dies kann dadurch begründet sein, daß sich bei  $N_G = 13$  zwar ein kleinerer mittlerer quadratischer Fehler ergibt, sich die Rotationssymmetrie des rezeptiven Feldes der Ganglienzelle jedoch qualitativ etwas verschlechtert (vgl. Abb. 3.13 und Abb. 3.14). Die Verwendung eines alternativen Fehlermaßes wäre zumindest für die Approximation der Ganglienzelle ggf. von Vorteil.

Für die weitere Modellierung sei  $N_G = 7$  oder  $N_G = 19$  gewählt, je nach geforderter Genauigkeit. Hier sollte allerdings immer berücksichtigt werden, daß die in der Biologie gemessenen rezeptiven Felder im Laufe der Evolution entstanden sind, aber auch nur näherungsweise mit GABOR-Funktionen in der Literatur beschrieben werden konnten, so daß das tatsächliche Verhalten natürlich davon abweichen kann. Unter diesem Aspekt ist die geringere Anzahl der Verbindungen  $N_S = 37$  in den meisten Fällen vorzuziehen, da mit  $N_S = 61$  der mittlere quadratische Fehler lediglich etwas mehr als halbiert wird, die Anzahl der notwendigen Berechnungen aber auch fast verdoppelt wird.

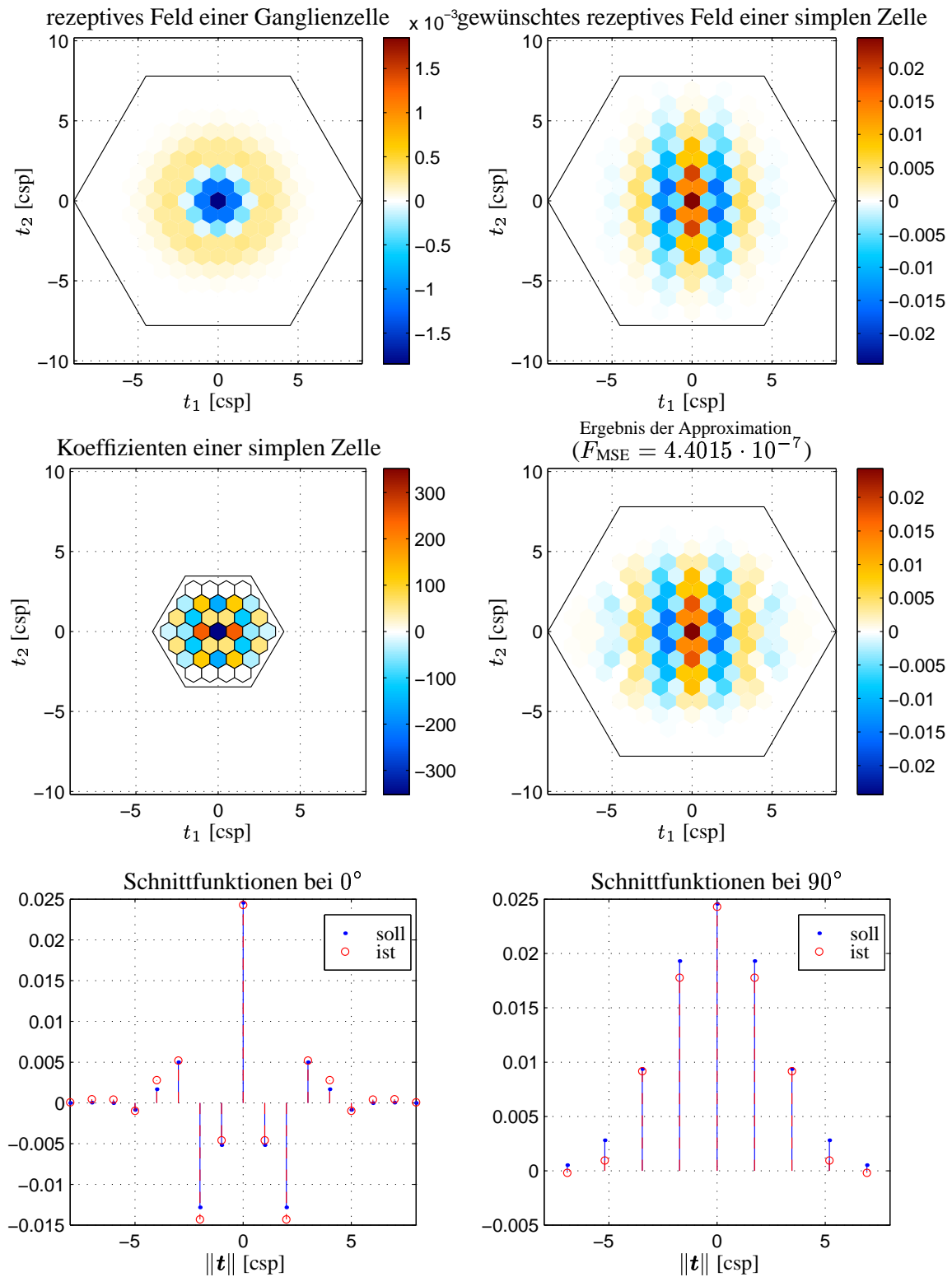


Abbildung 3.18: Approximation des rezeptiven Feldes einer simplen Zelle, bei der  $N_S = 37$  Ganglienzellen auf eine simple Zelle konvergieren ( $N_G = 7$ ).

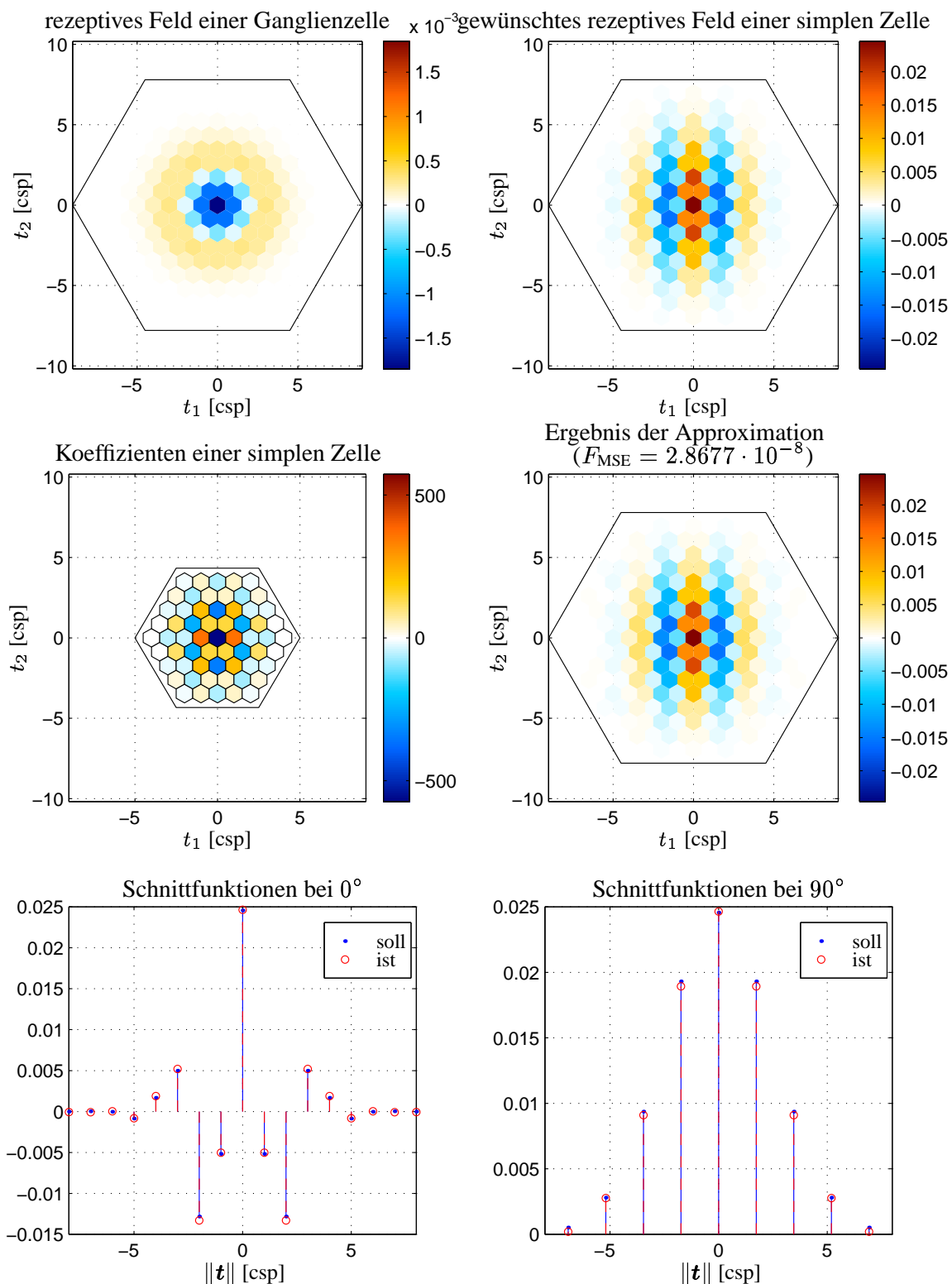


Abbildung 3.19: Approximation des rezeptiven Feldes einer simplen Zelle, bei der  $N_S = 61$  Ganglienzellen auf eine simple Zelle konvergieren ( $N_G = 7$ ).

In Abb. 3.20 ist schließlich der Signalflußgraph und in Abb. 3.21 sind die rezeptiven Felder der gesamten bisherigen Modellierung des biologisch motivierten Sehsystems für  $N_G = 7$  und  $N_S = 37$  zu sehen.

Dargestellt sind die rezeptiven Felder eines Rezeptors, einer Ganglienzelle und zweier Paare einfacher Zellen  $h_{S,0}^Q$  und  $h_{S,1}^Q$  in einer Orientierung von  $0^\circ$  und  $30^\circ$ . Die Gesamtheit aller 6 Orientierungen auf  $180^\circ$  ergibt sich jeweils aus einer Koordinatendrehung um  $60^\circ$  auf dem hexagonalen Raster der beiden Grundorientierungen gemäß

$$h_{S,i}^Q(\mathbf{k}) = \begin{cases} h_{S,0}^Q(\tilde{\mathbf{R}}_{60}^{i/2} \mathbf{k}) & , i = 0, 2, \dots, 10, \\ h_{S,1}^Q(\tilde{\mathbf{R}}_{60}^{(i-1)/2} \mathbf{k}) & , i = 1, 3, \dots, 11. \end{cases} \quad (3.25)$$

Hierbei ist  $\tilde{\mathbf{R}}_{60}$  eine diskrete Drehmatrix, welche eine Koordinatendrehung auf dem diskreten Abtastraster  $\mathbf{k}$  um  $60^\circ$  bewirkt. Sie ergibt sich aus der Drehmatrix

$$\mathbf{R}_{60} = \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix} \quad (3.26)$$

für die kontinuierlichen Koordinaten  $\mathbf{t} = \mathbf{T}\mathbf{k}$  durch die Identität

$$\mathbf{R}_{60}\mathbf{T}\mathbf{k} \stackrel{!}{=} \mathbf{T}\tilde{\mathbf{R}}_{60}\mathbf{k} \quad \forall \mathbf{k} \quad (3.27)$$

$$\iff \tilde{\mathbf{R}}_{60} = \mathbf{T}^{-1}\mathbf{R}_{60}\mathbf{T} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}. \quad (3.28)$$



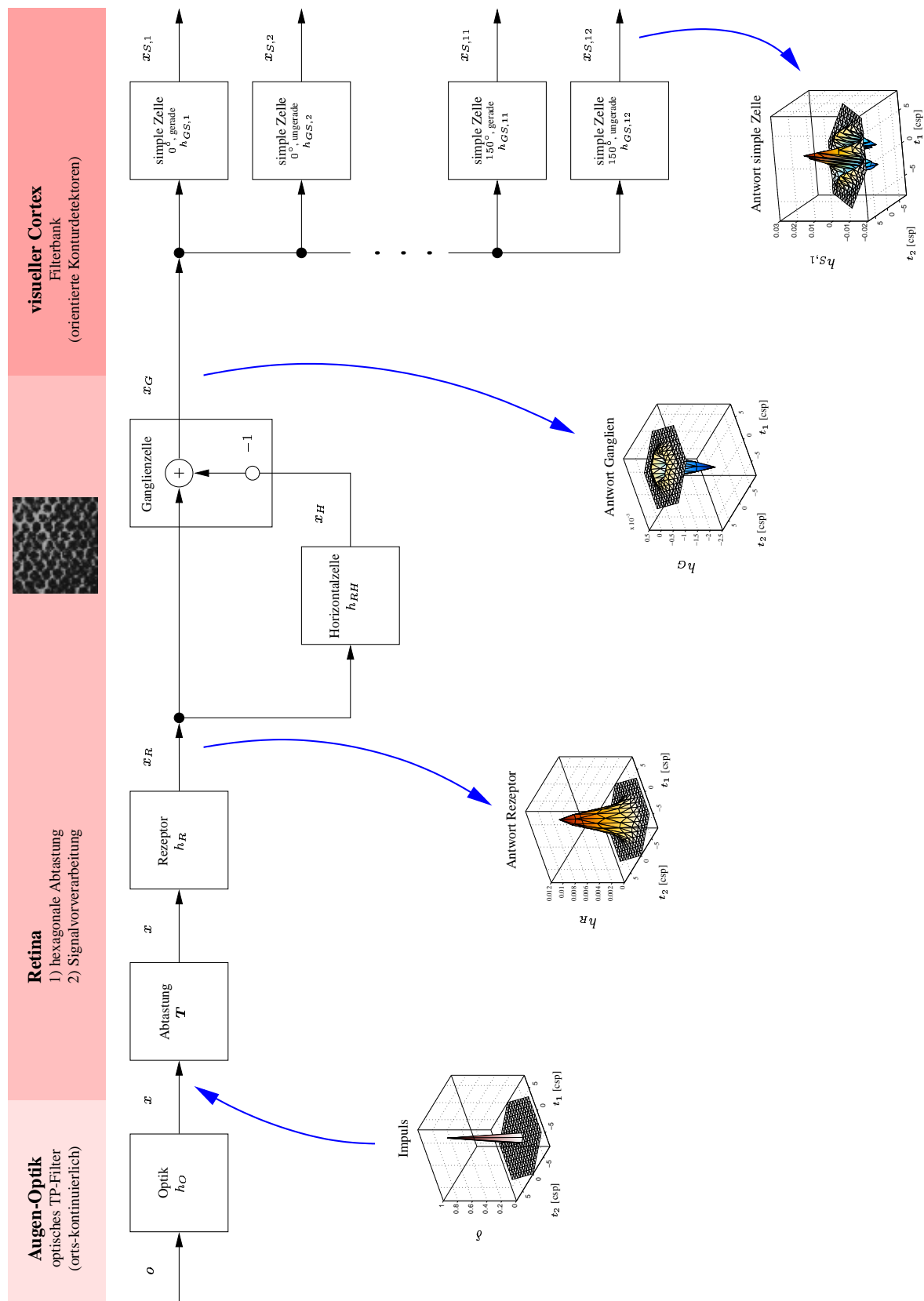


Abbildung 3.20: Signalflußgraph der Modellierung. Über einen impulsartigen Eingangsreiz läßt sich das Antwortverhalten der Verarbeitungskette schichtweise messen.

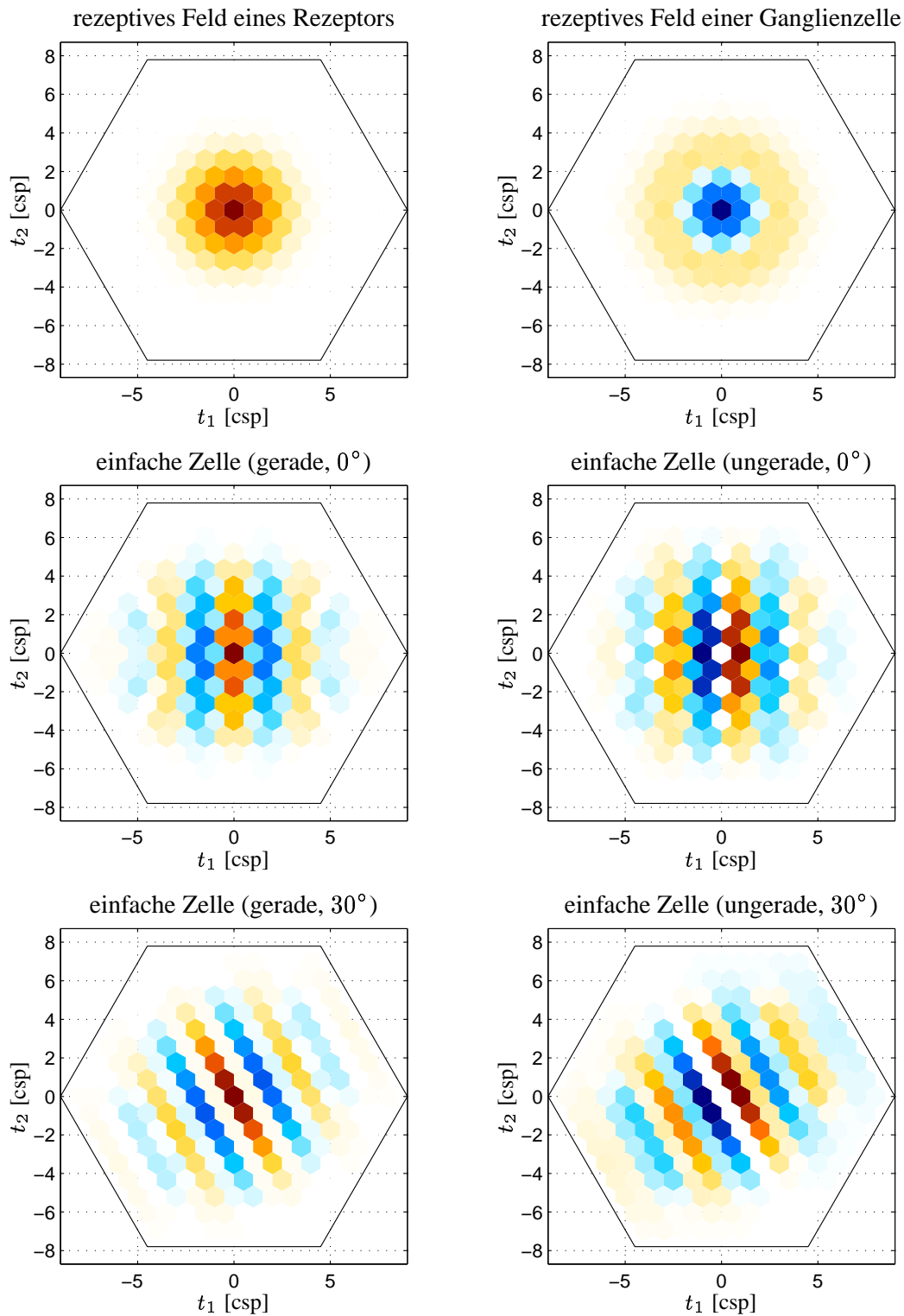


Abbildung 3.21: Approximierte rezeptive Felder des biologischen Sehsystems. oben: Rezeptor und Ganglienzelle, mitte/unten: simple Zellen der Orientierung  $0^\circ$  und  $30^\circ$ .

### 3.4 Eigenschaften der Modellierung

Die Architektur des entwickelten Verarbeitungssystems, welches einige Aspekte der biologischen Verhältnisse bei höheren Lebewesen widerspiegelt, weist einige Vorzüge auf, die nicht nur für das biologische Vorbild, sondern auch für technische Systeme von Interesse sind. In diesem Zusammenhang ist herauszustellen, daß nicht nur die eigentlichen Ausgangssignale der einfachen Zellen berechnet werden. Die Verarbeitungskette liefert, wie bereits geschildert, darüber hinaus weitere wertvolle Signale. Diese sogenannten Zwischensignale werden beim biologischen Vorbild für diverse visuelle Aufgaben eingesetzt. Dem Ausgangssignal der Ganglienzellen kommt hierbei eine besondere Bedeutung zu. Es wird beispielsweise für Blicksteuerungsmechanismen und für die Adaption bei sich ändernden Lichtverhältnissen verwendet. Wie in [Har82] gezeigt, erscheint es aber auch plausibel, daß mit Hilfe des Gangliensignals rezeptive Felder gleicher Form aber größerer örtlicher Ausdehnung erzeugt werden können, die wiederum auf einfache Zellen konvergieren. Auf diese Weise ergeben sich ähnliche parallele Verarbeitungspfade, aber mit unterschiedlichen Ortsauflösungen. Diese automatisch mitberechneten Zwischensignale, die in der Biologie einer enormen Einsparung von Neuronenverbindungen zugute kommen, sollten demnach auch bei der Realisierung eines technischen Gesamtsystems berücksichtigt und ausgenutzt werden.

Eine weitere wichtige Eigenschaft der kaskadenförmigen Struktur ist die hierin implizierte Aufwandsreduktion gegenüber einer direkten Implementierung der einfachen Zellen als Feed-Forward Netz mit einer Eingangsneuronenschicht (abgetastete Bilddaten) und der Ausgangsneuronenschichten (einfache Zellen verschiedenen Typs und Orientierung) bzw. einem nichtrekursiven Digitalfilter. Der Einfachheit halber wollen wir folgende Abschätzung betrachten, in der die Impulsantworten der Systeme jeweils quadratisch begrenzt sein sollen. Das rezeptive Feld einer einfachen Zelle hat eine radiale Ausdehnung bis etwa  $R_7$ , die Impulsantwort  $h_{S,des}(\mathbf{k})$  ist also vereinfacht mit dem quadratischen Gebiet  $k_{1,2} = -N, \dots, N$  bis  $N = 7$  zu berücksichtigen. Daraus ergeben sich  $(2N + 1)^2 = 225$  synaptische Verbindungen bzw. Multiplikationen und Additionen. Dieser Aufwand muß natürlich für jede Orientierung spendiert werden, woraus sich dann insgesamt  $12 \cdot 225 = 2700$  Verbindungen ergeben. Der erforderliche Aufwand bei der vorliegenden Modellierung ist in Tabelle 3.3 für zwei Genauigkeitsstufen nachzuvollziehen. Wird ein großer Modellierungsaufwand betrieben, um eine hohe Approximationsgüte zu erreichen, so sind bei der Simulation im Vergleich zur direkten Implementierung lediglich nur noch etwa 40% an synaptischen Verbindungen zu spendieren, was einer beachtlichen Reduktion an arithmetischen Operationen gleichkommt. Bei der etwas schlechteren Modellgüte beträgt die Anzahl

Anzahl der Orientierungen	Anzahl der synaptischen Verbindungen $(2N + 1)^2 [N]$			
	$h_R$	$h_{RG}$	$h_{GS}$	$\sum$
1	81 [4]	9 [1]	81 [4]	171
12			12·81	1062 (39.3%)
1	81 [4]	9 [1]	49 [3]	139
12			12·49	678 (25.1%)

Tabelle 3.3: Berechnungsaufwand für die Simulation des Modells. Die Anzahl der synaptischen Verbindungen gibt die Anzahl der Multiplikationen und Additionen wieder.

der Berechnungen noch weniger. Die wesentlichen Vorteile einer Kaskadierung in Subsysteme können wie folgt zusammengefaßt werden:

- wertvolle Zwischensignale werden mitberechnet (Adaption, Blicksteuerung)
- andere Auflösungsebenen können mit wenig Aufwand berechnet werden
- zusätzliche Parallelisierbarkeit (Bildsequenzen)
- weniger Neuronenverbindungen gegenüber einer Direktrealisierung ( $\approx -60 \dots 75\%$ )

Abschließend sollen die Ergebnisse einer Simulation des Systems mit einem Beispieleringangsbild (Szene einer Raddemontage) betrachtet werden. In Abb. 3.22 (oben) ist zunächst das zur Verfügung stehende Bildsignal einer CCD-Kamera auf rechtwinkligem Raster und die auf das hexagonale Raster umgerechneten Bilddaten dargestellt. Darunter sind die Aktivitäten der Rezeptoren und Ganglienzellen der Modellierung zu sehen.

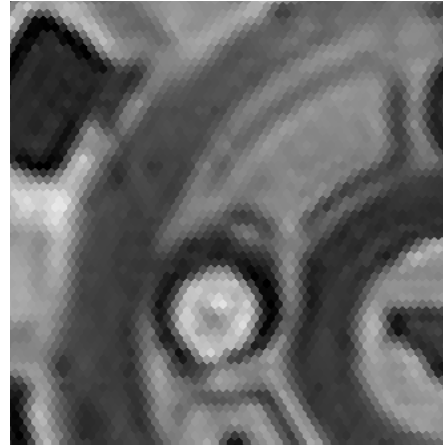
Die Ausgangssignale der kortikalen Verarbeitungssysteme sind in Abb. 3.23 zu sehen. Der Übersichtlichkeit wegen sind zunächst nur die Antworten der geraden und ungeraden einfachen Zellen für eine Orientierung ( $0^\circ$ ) dargestellt, welche auf vertikale Konturinhalt der Szene ansprechen. In der unteren Zeile sind darüber hinaus gewisse Verknüpfungen dieser Signale zu sehen, welche bereits als eine Art komplexe Zellen aufgefaßt werden können. Dies ist zunächst die Summe der quadrierten Signale der geraden und ungeraden simplen Zellen einer bestimmten Orientierung (in der Abbildung:  $0^\circ$ )

$$x_{C,i}(\mathbf{k}) = (x_{S,2i-1}(\mathbf{k}))^2 + (x_{S,2i}(\mathbf{k}))^2, i = 1 \dots 6 \quad (3.29)$$

CCD-Bild (rechtwinkliges Raster)



hexagonales Raster (Ausschnitt)



Rezeptoren



Ganglienzellen

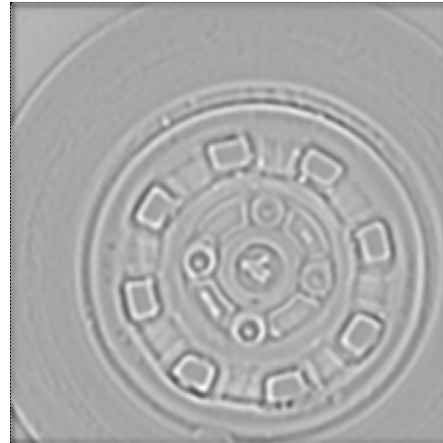


Abbildung 3.22: Simulation des Modells (Retina). oben: Bild einer CCD-Kamera und die auf ein hexagonales Raster umgerechneten Bilddaten. unten: Ausgangssignal der Rezeptor- und Ganglienzellenschicht auf dem hexagonalen Raster.

was bei der technischen Realisierung einer GABOR-Filterbank dem Betrag der komplexen Filterantwort entspricht. Schließlich ist in Abb. 3.23 noch zur Veranschaulichung der Arbeitsweise die lineare Überlagerung aller Orientierungen

$$\tilde{x}_C(\mathbf{k}) = \sum_{i=1}^6 x_{C,i}(\mathbf{k}) \quad (3.30)$$

dargestellt. Der Betrag der komplexen GABOR-Filterantwort wird in technischen

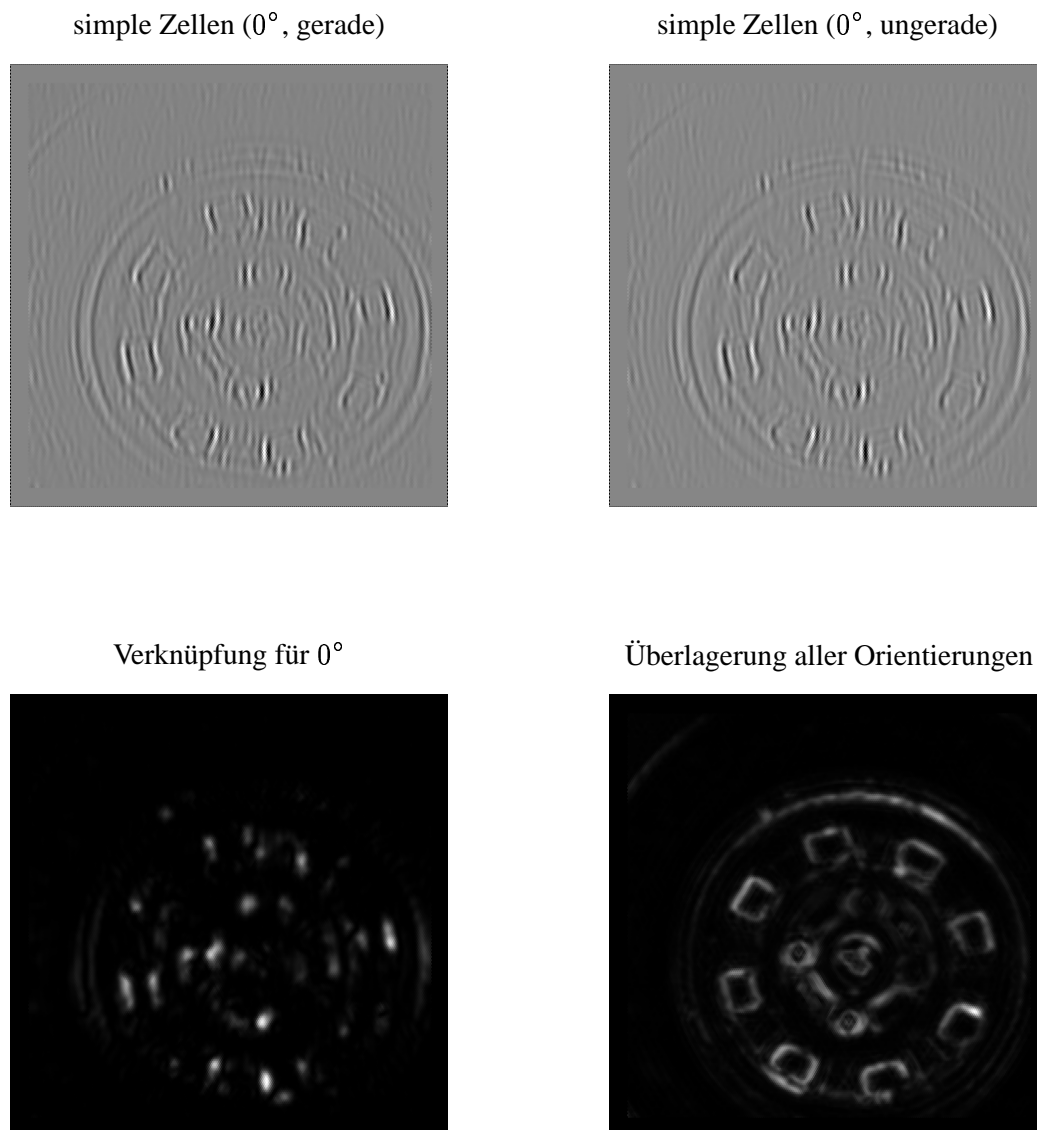


Abbildung 3.23: Simulation des Modells auf hexagonalem Raster (Cortex). oben: Antwortsignal der geraden und ungeraden simplen Zellen ( $0^\circ$ ). unten: Verknüpfung der Signale gerader und ungerader simplen Zellen ( $0^\circ$ ) bzw. Überlagerung aller Orientierungen.

Systemen vielfach als einfaches Modell für die *komplexen Neuronen* verwendet [SvdM97]. Neurophysiologische Untersuchungen des Antwortverhaltens komplexer Neuronen von POLLEN und RONNER [PR83, PR82] führen hingegen auf ein *periodisches komplexes Modellneuron*, welches die exzitatorischen Anteile der Antworten von insgesamt vier simplen Neuronen erhält. Es bleibt allerdings unklar, ob das von POLLEN und RONNER nachgewiesene periodische komplexe Mo-

dellneuron Vorteile gegenüber dem Betrag der GABOR-Antwort für technische Systeme aufweisen könnte. Abschließend sei hier erwähnt, daß der Begriff der „komplexen Zelle“ von der Idee des *Neocognitron* [FOH94, Fuk95, MV96] abzugrenzen ist. Bei diesem hat das komplexe Neuronen im wesentlichen die Aufgabe, eine Lageinvarianzbildung zu realisieren. Die Weiterverarbeitung zu komplexen Neuronen soll im Rahmen dieser Arbeit allerdings nicht näher behandelt werden.

### 3.5 Reduktion der Netzstruktur

Bei der beschriebenen Vorgehensweise zur Approximation der gewünschten rezeptiven Felder und damit der Bestimmung der synaptischen Gewichte bzw. Filterkoeffizienten wurden die synaptischen Verbindungen bislang immer so gewählt, daß sie innerhalb bestimmter Hexagone mit einem Radius  $R_v$  liegen. Die Koeffizienten wurden dann so errechnet, daß der mittlere quadratische Fehler minimal ist. Allerdings sollte hier nicht unerwähnt bleiben, daß bei fest vorgegebener Anzahl der synaptischen Verbindungen die Orte dieser Verbindungen einen Einfluß auf das Approximationsergebnis haben können. Bei der Approximation der Ganglienzelle war dies weniger deutlich, da die zu approximierende Impulsantwort rotationssymmetrisch ist. Bei der Modellierung der einfachen Zellen allerdings haben die Orte einen großen Einfluß. Um die optimalen Orte zu finden, kann z.B. wie folgt vorgegangen werden. Möchte man eine feste Anzahl an Verbindungen  $N_S = 37$  im Modell verwenden, so könnte zunächst die optimale Lösung mit deutlich mehr Verbindungen, z.B.  $N_S = 61$ , berechnet werden. Anschließend werden alle 61-37 Verbindungen, die das kleinste Gewicht besitzen, vernachlässigt. Wie in Abb. 3.19 zu erkennen, sind einige Koeffizienten der einfachen Zelle in der Tat betragsmäßig sehr klein. Auf diese Weise erhält man allerdings nur ein suboptimales Ergebnis.

Die gegebene Problemstellung ist in der linearen Algebra unter dem Begriff *subset selection* bekannt. Das überbestimmte Gleichungssystem

$$h_{S,\text{des}}(\mathbf{k}_i) = \sum_{j=1}^{N_S} h_{GS}(\mathbf{n}_j) \cdot h_G(\mathbf{k}_i - \mathbf{n}_j) \quad , i = 1 \dots M \quad (3.31)$$

ist eine lineare Regressionsaufgabe und kann als Matrixgleichung

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad , \mathbf{A} \in \mathbb{R}^{M \times N_S}, \mathbf{x} \in \mathbb{R}^{N_S}, \mathbf{b} \in \mathbb{R}^M \quad (3.32)$$

geschrieben werden. Bestimmt man den Lösungsvektor im *Least-Square Sinne* optimal, so ergibt sich ein Restfehler

$$\|\mathbf{A}\mathbf{x}_{LS} - \mathbf{b}\|^2 = \mathbf{b}^T \mathbf{b} - \mathbf{b}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \geq 0. \quad (3.33)$$

Wird nun eine Untermenge der Spalten von  $\mathbf{A}$  mit einer festen Anzahl  $N'_S < N_S$  an Spaltenvektoren ausgewählt, so ergibt sich i. allg. ein größerer Restfehler. Die Aufgabe besteht somit darin, diejenigen  $N'_S$  Spalten auszuwählen, die den kleinsten Restfehlerzuwachs verursachen.

GOLUB stellt in [GVL96] eine Methode zur subset selection mit Hilfe der Singulärwertzerlegung vor, die aber den Zielvektor  $\mathbf{b}$  nicht weiter betrachtet. Hier werden diejenigen Spalten von  $\mathbf{A}$  gesucht, die am stärksten linear unabhängig sind und damit  $\mathbb{R}^M$  am besten aufspannen. Dies führt aber nicht unbedingt zu einer Spaltenauswahl mit dem kleinsten Restfehler bzgl.  $\mathbf{b}$ , was ebenfalls von GOLUB diskutiert wird. Es ist weiter bekannt, daß die Lösung des least-square Problems mit Hilfe einer orthogonalen Zerlegung der Matrix  $\mathbf{A}$  bestimmt werden kann, was oft als *orthogonal-least-square* bezeichnet wird. Hierauf aufbauend wird in [CBL89, CW95] ein Verfahren zur subset selection gezeigt, welches sukzessive die besten Spalten von  $\mathbf{A}$  sucht (*forward selection*) und den Lösungsvektor bestimmt. In [SS97, VHV87, KDB93, CCG91] sind verwandte Strategien zu diesem Thema veröffentlicht. Für weitergehende Untersuchungen sei hier auch das Buch [Mil90] genannt, welches sich ausschließlich mit diesem Thema befaßt und eine gute Übersicht über bestehende und neuere Verfahren geben kann.

Die genannten Methoden müssen allerdings nicht notwendigerweise das globale Optimum finden. In der Tat zeigen heuristische Untersuchungen, daß in unserem Fall lediglich suboptimale Lösungen erreicht werden. Nur der Vergleich aller  $\binom{N_S}{N'_S}$  möglichen  $N'_S$ -dimensionalen Untermengen der Spalten von  $\mathbf{A}$  liefert die optimale Auswahl. Will man z. B. bis zu einem Radius  $R_4$  (also  $N_S = 61$ ) die optimale Untermenge mit  $N'_S = 37$  Elementen suchen, so ergeben sich bereits  $\binom{61}{37} = 5.94 \cdot 10^{16}$  mögliche Kombinationen. Diese Suche (*exhaustive search*) führt zwar gewiß auf die optimale Auswahl, allerdings nur mit erheblichem Rechenaufwand. Im vorliegenden Fall kann allerdings die Symmetrie der zu approximierenden Funktionen eine erhebliche Reduktion des Suchaufwandes bedeuten. Hierzu ist die Darstellung in Abb. 3.24 hilfreich, in der die optimalen Verbindungen nach solch einer Suche gezeigt sind. Dabei wird der Suchraum jeweils auf den ersten Quadranten reduziert ( $\bullet$ ), da sich die Orte in den übrigen Quadranten (graue Kreise) durch entsprechende Spiegelungen an den Achsen ergeben. Die Spiegellachsen sind hier immer die Symmetrieachsen der zu approximierenden Funktion.

Der erste Quadrant inklusive der Achsen umfaßt in unserem Beispiel 19 Koordinaten, aus denen alle möglichen Untermengen mit  $N''_S$  Elementen ( $\bullet$ ) für die Suche in Frage kommen. Hier ist zu beachten, daß die Positionen auf den Achsen nicht zweimal, sondern lediglich einmal gespiegelt werden, der Ursprung



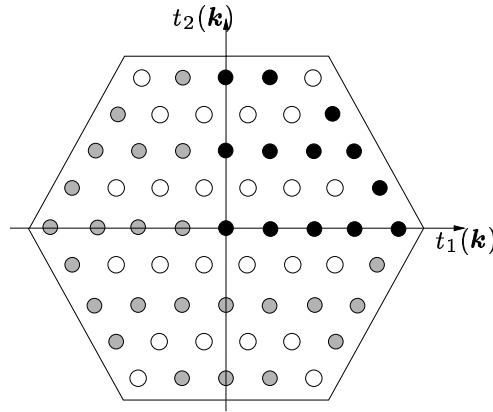


Abbildung 3.24: Reduktion der Suche nach den optimalen synaptischen Verbindungen auf den ersten Quadranten.

hingegen wird nicht gespiegelt. Damit nach der Spiegelung dennoch insgesamt  $N'_S = 37$  Koordinaten entstehen, wäre  $N''_S = \lfloor \frac{N'_S}{4} \rfloor = 9$  zu klein dimensioniert. Im schlimmsten Fall sind nämlich die Achsen vollständig besetzt, so daß sich aus der Bedingung

$$N'_S = \underbrace{2 \cdot 4}_{t_1\text{-Achse}} + \underbrace{2 \cdot 2}_{t_2\text{-Achse}} + \underbrace{1}_{\text{Ursprung}} + \underbrace{(N''_S - 7) \cdot 4}_{\text{Quadranten}} \quad (3.34)$$

$$= 13 + (N''_S - 7) \cdot 4 \quad (3.35)$$

$$\Leftrightarrow N''_S = \frac{N'_S - 13}{4} + 7 \quad (3.36)$$

und mit  $N'_S = 37$  die Dimension der reduzierten Untermenge zu  $N''_S = 13$  ergibt. Es sind dann natürlich nur diejenigen Kombinationen zu betrachten, die nach der Spiegelung insgesamt  $N'_S = 37$  Koordinaten aufweisen. Wird also diese Symmetrie ausgenutzt, so reduziert sich der Suchaufwand auf  $\binom{19}{13} = 27132$ , was recht schnell mit der heutigen Rechenleistung bewältigt werden kann. In Abb. 3.25 ist zu erkennen, inwieweit sich der Restfehler durch die Suche der optimalen synaptischen Verbindungen weiter um etwa die Hälfte reduzieren läßt (vgl. hierzu Tabelle 3.2).

Bei der Verwendung spezieller *Neurohardware* zur Simulation neuronaler Netze, wie z. B. in [Fra97, Wol01] und im folgenden Kapitel beschrieben, welche beliebige, auch spärliche synaptische Verbindungen realisieren kann, schlägt sich solch eine Reduktion der Netztopologie direkt in der benötigten Rechenzeit nieder und führt demzufolge zu drastischen Geschwindigkeitsvorteilen, die auf jeden Fall genutzt werden sollten. Wird die Simulation des neuronalen Netzes bzw. Digital-

filters mittels diskreter Faltung im Ortsbereich ausgeführt, so bleiben auch hier die Vorteile durch die beschriebene Reduktion der Netzstruktur erhalten, wenn die auf null gesetzten Filterkoeffizienten entsprechend berücksichtigt werden und nicht zu einer Multiplikation bei der Filteroperation führen. Bei einer Filterung durch Multiplikation im Frequenzbereich, hat die Anzahl der Filterkoeffizienten ungleich null allerdings i.allg. keine Auswirkungen auf die Rechenzeit.

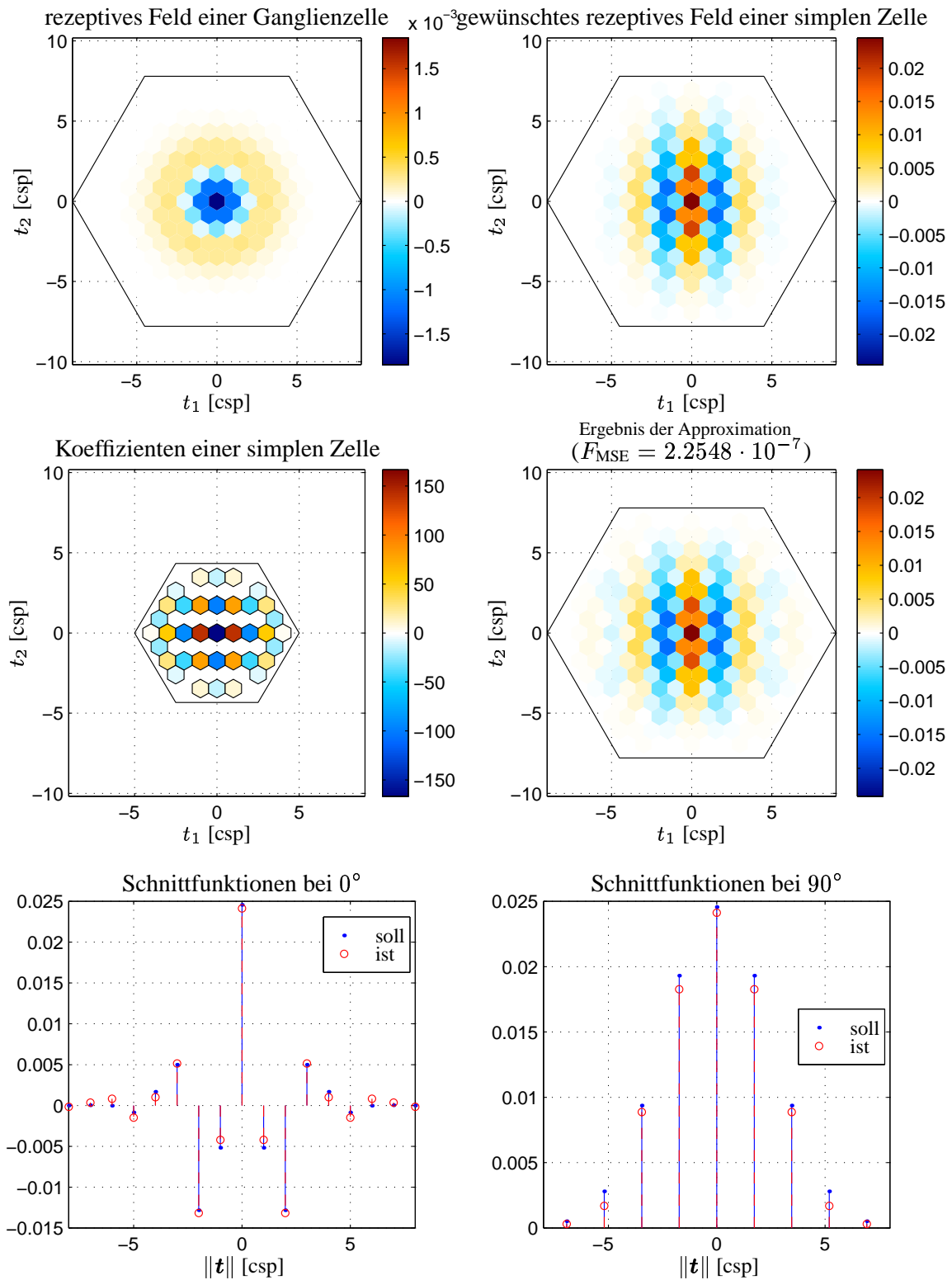


Abbildung 3.25: Approximation des rezeptiven Feldes einer simplen Zelle mit optimaler Wahl der synaptischen Verbindungen ( $N_G = 7$ ,  $N'_S = 37$ ).



# Realisierung pulscodierter Mechanismen

Durch neurophysiologische Untersuchungen weiß man, daß die Informationsverarbeitung im biologischen Sehsystem ab den Ganglienzellen in einer pulscodierten Form geschieht. Dabei codiert das Neuron seinen momentanen Aktivitätszustand in ein zeitliches Signal, eine Pulsfolge oder Spikes, deren Pulsfrequenz die Erregung der Zelle widerspiegelt. Dieses Signal wird über das Axon zu den Nachfolgeneuronen übertragen. An den Dendriten der Zielneuronen wird dann durch Aufsummierung oder Integration der Spikes die Aktivität des Vorgängerneurons rekonstruiert. Hierdurch kommt die Zeit als weitere Dimension und auch als weiterer Informationsträger hinzu. Wie nämlich verschiedene Studien zeigen, wird auch die Phaseninformation der Pulsfolgen für Wahrnehmungsmechanismen genutzt [ERAD89, ERAD90]. Diejenigen Neuronen, die ein ähnliches Merkmal repräsentieren, wie eine kontinuierliche Linie in ihrem rezeptiven Feld, synchronisieren so ihre ausgesendeten Pulsfolge. Dieser Mechanismus ist natürlich für viele Aufgaben nützlich. Die Segmentation einzelner Objekte oder die Trennung von Objekt und Hintergrund seien hier als Beispiel genannt.

Die genannten Synchronisationseffekte selbst sind zwar nicht Gegenstand der vorliegenden Arbeit (der interessierte Leser sei auf andere Arbeiten verwiesen, wie [Sch00]). Dennoch ist es von großem Interesse, das im vorangegangenen Kapitel vorgestellte Netzwerk, welches das lineare Filterverhalten der frühen visuellen Verarbeitung beschreibt, um diese pulscodierenden Strukturen zu erweitern. Durch die Kombination eines gaborähnlichen Verhaltens zur Konturextraktion auf einem hexagonalen Raster und eine pulscodierte Verarbeitung ergibt sich hierdurch ein wertvolles Vorverarbeitungssystem für pulscodierte neuronale Netzwerke (PCNN) [TWH00, Thi00, Wol01]. Dabei setzen die Ganglienzellen ihre Aktivitäten in Pulsraten um und die simplen Zellen können als ECKHORN-Neuronen

entworfen werden. Auf dieses pulscodierte Vorverarbeitungssystem könnten dann weitere Studien aufbauen, um z. B. die Synchronisationsmechanismen näher zu untersuchen.

Im folgenden Abschnitt wird das für pulscodierte Netzwerke vielfach eingesetzte ECKHORN- Neuron vorgestellt. Dieses Neuronenmodell ist zunächst für die Aufgabenstellung so zu parametrisieren, daß die Codierung und Decodierung der Membranpotentiale auf eine lineare Weise geschieht. Dies ist wichtig, da das Verhalten bzgl. der örtlichen Filterwirkung bis hin zu den simplen Zellen allgemein als näherungsweise linear angenommen wird.

Der konsequente Schritt ist dann die Erweiterung des linearen Modells auf eine pulscodierte Verarbeitung, was theoretisch hergeleitet werden soll und auf eine spezielle Netzstruktur führt. Die Betrachtungen zeigen, daß prinzipiell jedes lineare neuronale (Feed-Forward) Netzwerk (oder nichtrekursive Digitalfilter), wie die simplen Zellen, komplett mittels ECKHORN-Neuronen realisiert werden kann, ohne daß dabei die Linearität verloren geht.

Erst durch den Einsatz von Spezialhardware ist aber die Simulation solch großer pulscodierter neuronaler Netze in akzeptabler Rechenzeit möglich. Abschließend ist deshalb die Implementierung auf einer in der Arbeitsgruppe HARTMANN entwickelten Neurocomputerarchitektur *SPIKE* bzw. *ParSPIKE* gezeigt, welche das ECKHORN-Neuron als grundlegenden Baustein verwendet [FHJS99, HFSW97, WHR99]. Wie experimentelle Untersuchungen zeigen, ist allerdings das ECKHORN-Neuron für die geforderte Umsetzung nicht ausreichend. Erst eine Modifikation durch Verbesserung des Spike-Decoders kann die numerischen Probleme der Digitalrealisierung mit Erfolg beheben.

## 4.1 Modellneuron von ECKHORN

Die ECKHORN-Neuronen besitzen Dendriten für verschiedene dendritische Potentiale. Für die vorliegende Modellierung in Abb. 4.1 werden nur zwei Dendritenbäume (*EP1* und *IP*) berücksichtigt. Einer dieser beiden Dendritenbäume sorgt für einen exzitatorischen (*EP1*) und der andere für einen inhibitorischen Einfluß (*IP*) auf das Membranpotential (*MP*), wobei die Dendritenpotentiale über sogenannte *Leckintegratoren* akkumuliert werden. Für die Aussendung einer Pulsfolge (Spikes) wird das Membranpotential des Neurons in eine rückgekoppelte Verarbeitungsstufe mit einer dynamischen Schwelle (*DT*) geführt, die als *Spike-Encoder* bezeichnet wird und sehr ähnlich zu der Struktur von FRENCH und STEIN ist [FS70]. Dieses weit verbreitete Modell Neuron wird für die Simulation pulscodierter Mechanismen eingesetzt und wurde bereits auf spezieller

Neurohardware implementiert [FHJS99, HFSW97, WHR99]. Aus diesem Grund soll es auch hier Anwendung finden und etwas detaillierter diskutiert werden.

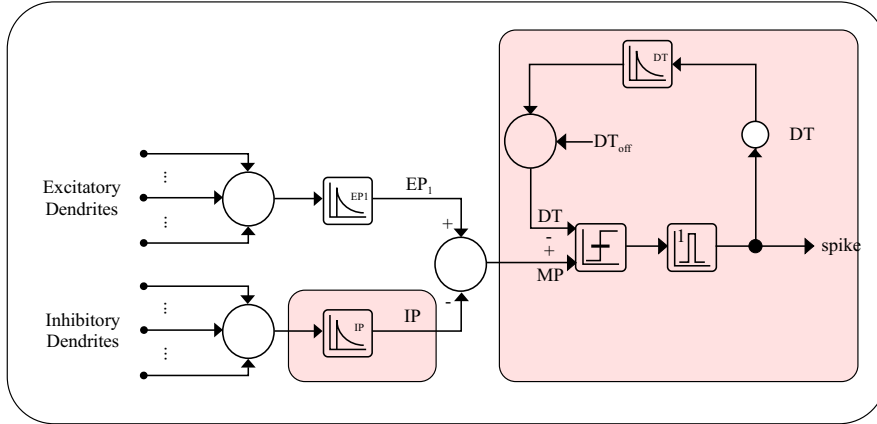


Abbildung 4.1: Das ECKHORN-Modellneuron. Der Spike-Decoder (links) und Spike-Encoder (rechts) sind farbig hinterlegt.

In der Abb. 4.1 sind die beiden wesentlichen Komponenten des Modell Neurons gekennzeichnet: der *Spike-Encoder* (rechts) und der *Spike-Decoder* (links). Der Spike-Encoder eines Neurons  $i$  (an der Position  $k_i$ ) erzeugt ein Zeitsignal, die sogenannte Pulsfolge oder Spikes, so daß die Frequenz dieser Spikes das (mehr oder weniger stationäre) Membranpotential  $MP$  bzw. die Erregung des Neurons codiert. So wird das ortsdiskrete Signal  $x(\mathbf{k})$  in ein (Puls-) Signal  $x_p(\mathbf{k}, t)$  des Ortes und der Zeit umgesetzt, was in Abb. 4.2 (links) schematisch dargestellt ist. Diese Operation kann auch als Raten-Puls-Umsetzung bezeichnet werden. Der Spike-Decoder in Abb. 4.2 (rechts) hat die entgegengesetzte Aufgabe, die Rückgewinnung der Erregung eines Vorgängerneurons zu rekonstruieren. Dazu werden die eintreffenden Spikes derart integriert, daß das Ausgangssignal des Spike-Decoders mit

$$\tilde{x}(\mathbf{k}) := \lim_{t \rightarrow \infty} \overline{x_p(\mathbf{k}, t)} \approx x(\mathbf{k}) \quad (4.1)$$

gegen den Mittelwert der Pulsfolge konvergiert, da dieser die gewünschte codierte Information bereithält. Im ECKHORN-Neuron ist dieser Decoder durch einen sogenannten *Leckintegrator* realisiert, was ein rekursives Tiefpaßfilter erste Ordnung darstellt. Durch die zugehörige Differenzengleichung

$$y(n) = e^{-1/\tau} x(n-1) + e^{-1/\tau} y(n-1) \quad (4.2)$$

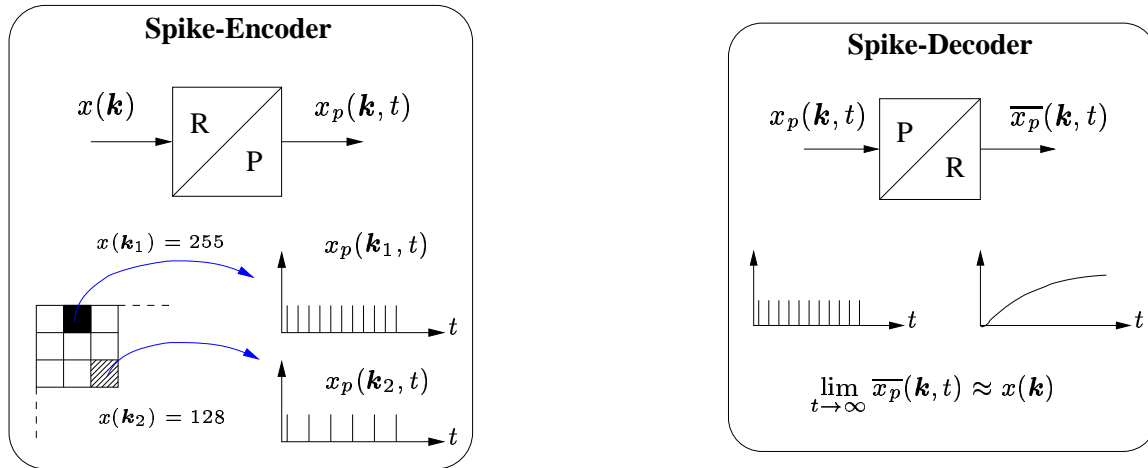


Abbildung 4.2: Prinzipielle Funktionsweise eines Spike-Encoders (links) und eines Spike-Decoders (rechts).

kann auf die z-Übertragungsfunktion

$$H^{[z]}(z) = \frac{e^{-1/\tau} z^{-1}}{1 - e^{-1/\tau} z^{-1}} \quad (4.3)$$

geschlossen werden, wobei die Zeitkonstanten der exzitatorischen und inhibitorischen Eingänge gleich gewählt werden sollen ( $\tau_{EP1} = \tau_{IP} = \tau$ ).

Um eine sichere Codierung und Decodierung zu gewährleisten, sind einige Parameter zu dimensionieren. Dazu muß zunächst eine passende Einstellung für den Spike-Encoder gefunden werden, welcher die Aktivität der Ganglienzellen auf eine möglichst lineare Weise in eine Pulsfolge passender Pulsfrequenz umsetzt. Die Parameter der dynamischen Schwelle ( $DT$ ) sind die Zeitkonstante  $\tau_{DT}$  des Leckintegrators, das Inkrement  $\Delta DT$  und der Offset  $DT_{\text{off}}$  der Schwelle. Wie Untersuchungen zeigen, führt ein  $\tau_{DT} = 32$  in Kombination mit  $\Delta DT = 16$  auf eine nahezu lineare Umsetzung der Ganglienaktivität in eine Pulsfrequenz. Sehr geringe Aktivitäten werden mit  $DT_{\text{off}} = 4$  unterdrückt. In einem weiteren Schritt sind die Leckintegratoren für das dendritische Potential der simplen Zellen passend zu dimensionieren. Diese dendritischen Potentiale müssen gegen die selben Werte wie die codierten Ganglienaktivitäten konvergieren, die durch die Frequenzen der eintreffenden Pulsfolgen repräsentiert werden. Wie sich gezeigt hat, ist mit einer Zeitkonstanten  $\tau = 1000$  der Integrationseffekt ausreichend. Die gesamte Codierung und Decodierung kann auf diese Weise recht genau linear realisiert werden, was in Abb. 4.3 illustriert ist.



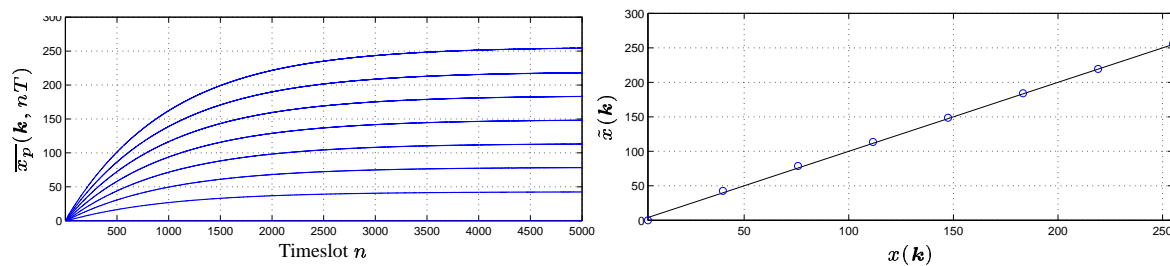


Abbildung 4.3: links: Rekonstruktion der Ganglienaktivität ( $0 \dots 255$ ) aus den Pulssignalen mit dem Leckintegrator nach ECKHORN. rechts: Lineare Codierung der Ganglienaktivität über die Spikfrequenz (– gewünscht,  $\circ$  erzielt).

## 4.2 Erweiterung des linearen Modells

Für die Implementierung mit ECKHORN-Neuronen muß das bestehende in Kapitel 3 entwickelte Modell für eine pulscodierte Informationsverarbeitung erweitert werden. Dazu muß zunächst berücksichtigt werden, daß Pulse oder Spikes in einem pulscodierten neuronalen Netzwerk zeitdiskrete Ereignisse ohne Vorzeichen und Amplitudenwerte sind. Das Ausgangssignal der Ganglienzellen und simplen Zellen (und auch Zellen anderer kortikaler Areale) in einer ratencodierten Beschreibung nehmen jedoch i. allg. durchaus sowohl positive als auch negative Werte an, z.B. im Bereich von  $-255$  bis  $255$ . Dies führt auf die Notwendigkeit, das bisherige Modell schrittweise zu erweitern. Der Einfachheit wegen beschränken wir uns im folgenden auf die für die vorliegende Arbeit interessierende Neuro-nenschicht der simplen Zellen, was in Abb. 4.4 schematisch für einen Typ simpler Zellen dargestellt ist. Das Ziel ist die Umsetzung der Antworten der Ganglien-

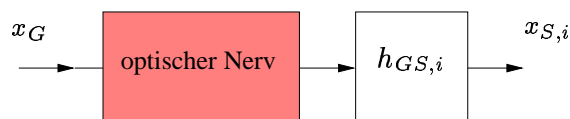


Abbildung 4.4: Die pulscodierten Gangliensignale werden über den optischen Nerv übertragen und konvergieren auf die simplen Zellen (hier ist nur ein Typ simpler Zellen dargestellt).

zellen und simplen Zellen in Pulsfolgen. Für diesen Zweck müssen sowohl die positiven als auch die negativen Anteile der Antworten separat in zwei Schichten aufgebaut werden. Generell kann aber jedes Signal  $x$  gemäß

$$x = x^+ - x^- \quad , \quad x^{+,-} \geq 0 \quad (4.4)$$

in zwei positivwertige Anteile

$$x^+ = x \cdot u(x) \quad \text{und} \quad x^- = -x \cdot u(-x) \quad (4.5)$$

aufgeteilt werden. Dabei sei

$$u(x) = \begin{cases} 1 & , x \geq 0 \\ 0 & , x < 0 \end{cases} \quad (4.6)$$

die Sprungfunktion. Diese Transformation wird in Abb. 4.5 (oben) für die Ganglienzellen und in Abb. 4.5 (unten) für die simplen Zellen eingesetzt. Da die ver-

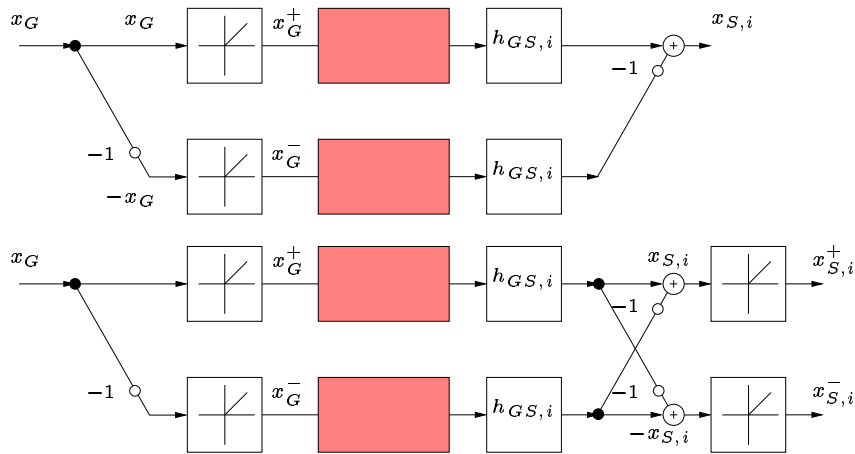


Abbildung 4.5: Repräsentation des positiven und negativen Anteils der Zellantworten in zwei Schritten für die Ganglienzellen (oben) und den simplen Zellen (unten).

wendeten Signale nun positiv sind, können sie mit Hilfe der bereits beschriebenen Spike-Encoder in Pulsfolgen codiert werden. Die Rekonstruktion der Membranpotentiale wird durch Spike-Decoder an einer geeigneten Stelle im Signalflußgraph nach der Verschaltung  $h_{GS,i}$  zur örtlichen Filterung durchgeführt. Diese Anpassung ist in Abb. 4.6 (oben) gezeigt. Abschließend ist mit der Hilfe von Abb. 4.1 ersichtlich, daß der resultierende Graph zwei bekannte Strukturen beinhaltet (die obere ist zur Verdeutlichung gekennzeichnet), die durch ECKHORN-Neuronen substituiert werden können, was in Abb. 4.6 (unten) gezeigt ist. Für die simplen Zellen, welche den positiven Anteil repräsentieren, werden die positiven Antworten der Ganglienzellen mit dem exzitatorischen Eingang des ECKHORN-Modells über die Verbindungsgewichte  $h_{GS,i}$  verbunden. Die negativen Antworten der Ganglienzellen gelangen zu dem inhibitorischen Eingang. Die negativen simplen Zellen werden entsprechend auf komplementäre Weise verschaltet.

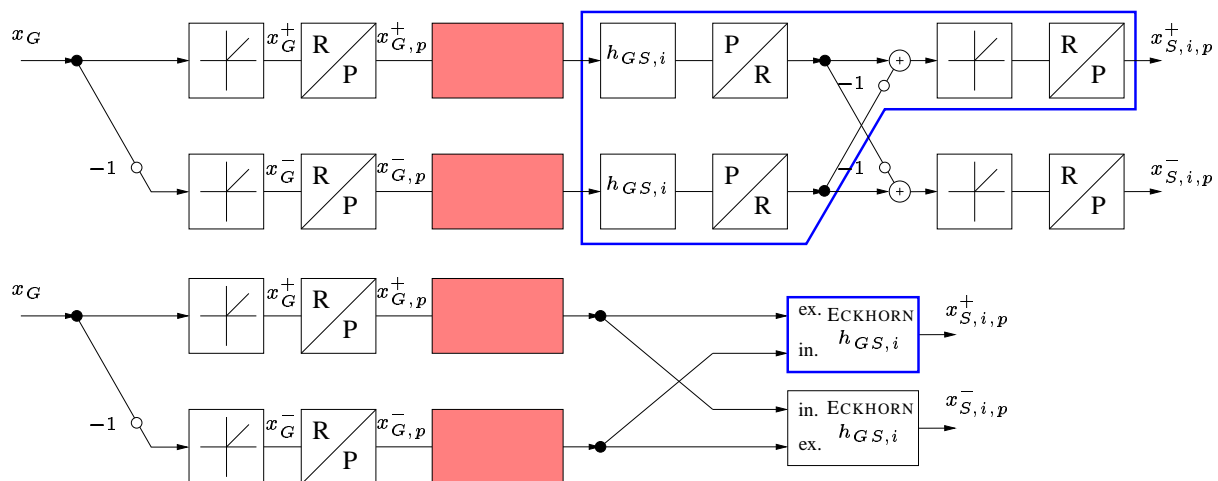


Abbildung 4.6: oben: Realisierung eines linearen Übertragungssystems für eine pulscodierte Verarbeitung. unten: Implementierung mit ECKHORN-Neuronen.

Zusammengefaßt ist festzuhalten, daß theoretisch jedes lineare System auf die beschriebene Vorgehensweise mit pulscodierten Mechanismen realisiert werden kann ohne die lineare örtliche Filtercharakteristik zu verlieren. Im folgenden Abschnitt wird allerdings gezeigt, daß die Qualität der resultierenden Filtercharakteristik stark von der Qualität der implementierten Spike-Decoder abhängt.

## 4.3 Experimentelle Ergebnisse einer Neurocomputerimplementierung

Die in der Arbeitsgruppe HARTMANN entwickelte Neurocomputerarchitektur *SPIKE* dient der schnellen Simulation großer pulscodierter neuronaler Netze. Eine Implementation ist das *SPIKE128K* System [FHJS99, HFSW97], welches eine schnelle Simulation von Netzstrukturen erlaubt, die bis zu 131072 Neuronen und 16 Millionen Verbindungen beinhalten dürfen. In diesem Kontext bedeutet Echtzeit eine Ausführdauer von 1 ms für einen Simulationszeitpunkt (Timeslot). Der *ParSPIKE* Entwurf erweitert die Architektur durch einen parallelen Ansatz auf Netzwerke bis zu einer Million Neuronen. Dieser wurde allerdings noch nicht als Hardware realisiert, liegt aber als Softwaresimulation vor, die eine Abschätzung der Leistungsfähigkeit erlaubt [WHR99]. In beiden Systemen werden die Neuronen durch die Werte ihrer dendritischen Potentiale und der aktuellen dynamischen Schwelle repräsentiert. Diese Zustände können in einem Speicherbereich organisiert werden, wobei der Zugriff auf diesen über eine spezielle Neuronenadresse für jedes Neuron erfolgt. Der Simulationsalgorithmus arbeitet dann auf

diesem Neuronenspeicher und muß diesen, falls erforderlich, für jeden Zeitpunkt aktualisieren. In Netzwerken für die Verarbeitung visueller Daten sind meist nur wenige Neuronen, die bestimmte Merkmale einer Szene beschreiben, in der Informationsverarbeitung involviert. Der Zustand der übrigen Neuronen wird dementsprechend in der Simulation nicht durch erregende Spikes verändert. Dies führt direkt auf einen ereignisgesteuerten Simulationszyklus, der nur die aktiven Neuronen berechnet. Dieser Algorithmus kann auf Hardware, wie dem SPIKE128K implementiert oder aber auch als Softwaresimulation einer parallelen DSP Architektur, dem ParSPIKE simuliert ausgeführt werden. Für die parallele Implementation verteilen sich die Neuronen auf viele DSPs, indem deren Zustände auf den Speicherbereichen dieser Prozessoren untergebracht werden. ParSPIKE basiert auf *SHARC* Prozessoren von *Analog Devices* mit großem On-Chip Speicher. Das System stellt zwei DSP Boards zur Verfügung. Ein Typ ist auf Vorverarbeitungsaufgaben spezialisiert und berechnet die Neuronenzustände durch Verbindungsmasken auf jedem Chip. Für die Kommunikation zwischen den parallelen DSPs ist nur der Austausch von sogenannten *Spikelisten* notwendig. Diese sog. (RC) Boards für reguläre Verknüpfungsmasken (rc=regular connections) eignen sich sehr für das vorgestellte Netzwerk der frühen visuellen Verarbeitung (siehe Abb. 4.7).

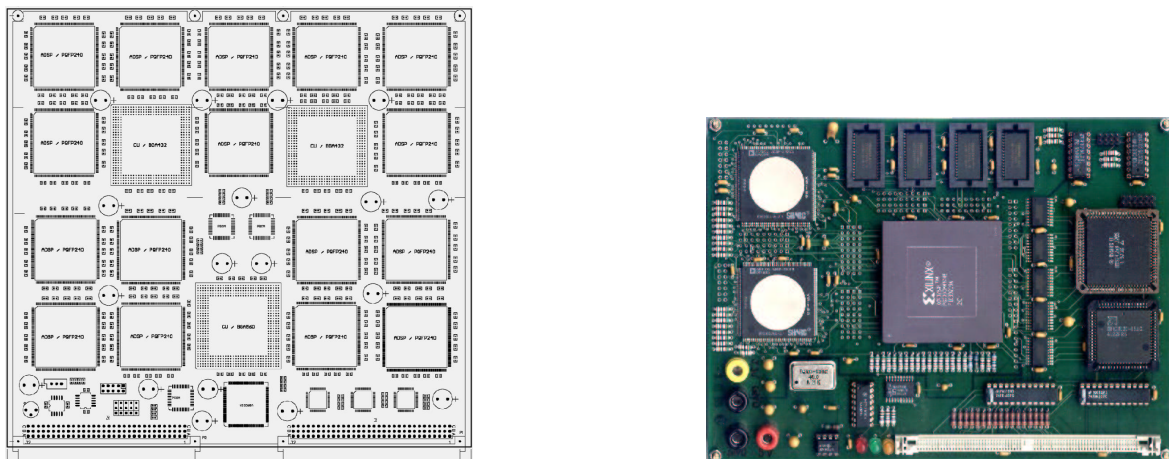


Abbildung 4.7: ParSPIKE Neurocomputer Architektur für PCNN. links: RC VME Board Implementation mit 32 DSPs (zweiseitig). rechts: Testboard für die Analyse der Leistungsfähigkeit.

Das Vorverarbeitungsnetzwerkes wurde im Rahmen dieser Arbeit als Softwaresimulation des ParSPIKE implementiert. Das entwickelte Konzept besteht zunächst aus 2 Schichten Ganglienzellen, um die positiven und negativen Anteile als Pulsfolgen zu codieren. Darüber hinaus werden 2 Typen simpler Zellen

(mit geradem und ungeradem rezeptiven Feld) für 6 verschiedene Orientierungen ( $0^\circ, 30^\circ, \dots, 150^\circ$ ) berücksichtigt, ebenfalls als Pulsfolge der positiven und negativen Anteile. Das führt insgesamt auf 24 Schichten simpler Zellen und 2 Schichten Ganglienzellen. Das ParSPIKE RC Board stellt 32 DSPs für die Simulation zur Verfügung. Es bietet sich daher an, für jede der 24 kortikalen Schichten einen DSP zu verwenden. Darüber hinaus kann über den VME Bus auf den On-Chip Speicher zugegriffen werden, um die Eingangsreize einzuspeisen. Diese Eingangsbilddaten können partitioniert und mit einer entsprechenden Überlappung auf die restlichen 8 DSPs für die Vorverarbeitung verteilt werden. So könnte die Umrechnung auf das hexagonale Abtastraster und die retinale Vorfilterung gemäß der rezeptiven Felder der Ganglienzellen auf einem Achtel des Bildes parallel ausgeführt werden. Letztlich werden die positiven und negativen Pulsströme der Ganglienzellen generiert. Durch das hexagonale Abtastraster werden die Eingangsbilddaten durch 14161 ( $119 \times 119$ ) simple Neuronen in jeder Schicht verarbeitet. Das gesamte Netzwerk umfaßt 368186 Neuronen inklusive der Ganglienzellen und könnte demnach auf einem ParSPIKE RC VME Board simuliert werden. Hierbei verwenden die simplen Zellen die selbe Spike-Encoder Struktur wie die Ganglienzellen.

## 4.4 Modifikation des ECKHORN-Modellneurons

Die SPIKE Architektur weißt, wie jede Digitalrechnerimplementation, gewisse Einschränkungen auf. So werden intern alle Berechnungen durch eine Fixed Point Arithmetik ausgeführt. Die Neuronengewichte  $h_{GS,i}$  sind als s3.5 Variablen (signed, 3 Bit + 5 Bit) und die dendritischen Potentiale ( $EP/IP$ ) als s9.5 Variablen quantisiert. Die Potentiale stellen dabei genau genommen die Zustandsvariablen der Leckintegratoren dar, welche Digitalfilter vom Grad eins sind. Wie die Untersuchungen anhand der Softwaresimulation zeigen, führen diese Restriktionen seitens der Hardware nicht auf die erwarteten Resultate. Stark quantisierte Verbindungsgewichte auf der einen Hand führen zu örtlichen Filtercharakteristiken, die vom Entwurf in Kapitel 3 abweichen. Entweder Gewichte mit höherer Auflösung oder ein spezieller Filterentwurf mittels diskreter Optimierung der Koeffizienten könnten bessere Resultate ergeben. Trotz dieser Verbesserungen würden die Einschränkungen des Leckintegrators dennoch nicht die gewünschten Signale oder gar Probleme mit der Konvergenz verursachen. Dieser Sachverhalt ist in Abb. 4.8 verdeutlicht, in der das Ausgangssignal des Leckintegrators als Spike-Decoder beim ECKHORN-Neuron dargestellt ist. Die Verknüpfung gewichteter Signale dieser Qualität führen offensichtlich nicht auf die gewünschte Filtercharakteristik, da die Ungenauigkeiten dadurch noch verstärkt werden.

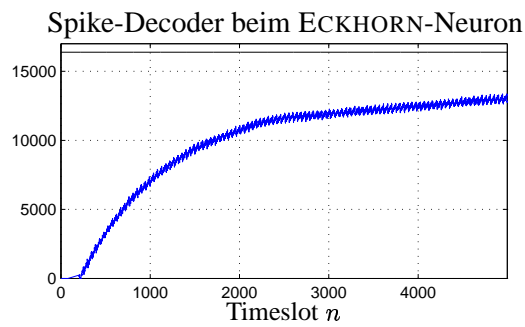


Abbildung 4.8: Integrationsleistung des Spike-Decoders beim ECKHORN-Neuron.

Da die Pulsströme, als Eingangssignale der Leckintegratoren, recht viel Energie auch bei hohen Frequenzen aufweisen, sind sie sicher als „schwierige“ Signale insbesondere für Tiefpaßfilter mit den genannten Restriktionen aufzufassen. Eine Kombination der quantisierten Zustandsvariablen mit den Nachteilen des Leckintegrators vom Grad eins führen auf entsprechend große numerische Ungenauigkeiten. Diese Fehler machen sich entweder durch stark oszillierende Anteile oder sogar durch Konvergenzprobleme des Filteralgorithmus bemerkbar. Diese Ergebnisse zeigen, daß eine Verbesserung des verwendeten Spike-Decoders notwendig ist, was in Abb. 4.9 angedeutet ist.

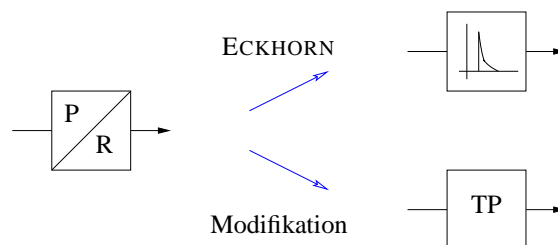


Abbildung 4.9: Modifikation des ECKHORN-Neurons. Der Leckintegrator wird durch ein „wohl entworfenes“ Tiefpaßfilter ersetzt.

Als verbesserter Spike-Decoder wird hier ein passend entworfenes Tiefpaßfilter vom Grad zwei vorgeschlagen. Es sollten zudem weitere positive Eigenschaften erfüllt werden. Dazu zählt zum einen eine große Dämpfung bei hohen Frequenzen, um die Oszillationen ausreichend zu unterdrücken. Eine kurze Einschwingzeit des Filters sollte zudem ebenfalls realisiert werden, um die Simulationsdauer zu verkürzen. Um auf der Neurohardware SPIKE überhaupt simuliert werden zu können, muß der Filteralgorithmus schließlich in eine Stimulations- und Abklingphase separierbar sein [WHR99, Wol01], was am Beispiel des Leckintegrators kurz erläutert werden soll. Bei dieser ereignisgesteuerten Signalverarbeitung wer-

den zunächst in einer ersten Phase alle Spikes gesammelt und die betroffenen Zielneuronen durch die eintreffenden Spikes erregt, d.h. ihre Zustandsspeicher  $z_w$  werden entsprechend modifiziert. Im Zustandsmodell aller Leckintegratoren wird dieser Schritt als

$$z_w(n+1) = x(n) + y(n), \quad (4.7)$$

berechnet, wobei  $x$  die gewichtete Summe aller eintreffenden Spikes, also das Eingangssignal, und  $y$  das Membranpotential bzw. das Ausgangssignal des Systems im vorherigen Zeitschritt sind. Erst dann werden in einer zweiten Phase, der Abklingphase, die Ausgangswerte der Spike-Decoder, also die Potentiale aller Neuronen mit

$$y(n+1) = e^{-1/\tau} \cdot z_w(n+1) \quad (4.8)$$

berechnet. Vergleiche dazu auch (4.2). Auf diese Weise ist eine parallele Abarbeitung der Daten möglich.

In Abb. 4.10 ist das Verhalten des Leckintegrators und zweier Tiefpaßentwürfe auf hochfrequente Pulsfolgen gezeigt. Das Spektrum der Pulsfolge (Impulskamm)

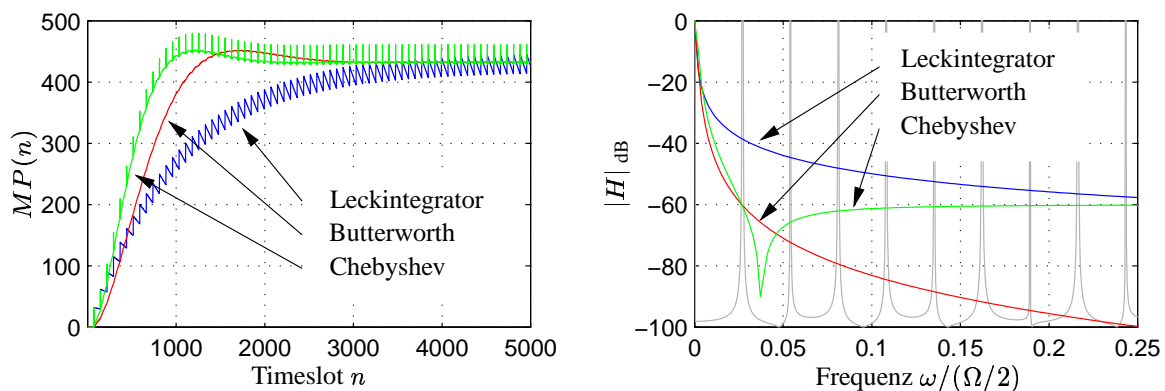


Abbildung 4.10: links: Das Einschwingverhalten verschiedener Spike-Decoder auf hochfrequente Pulsfolgen bei der Decodierung. rechts: Das Spektrum der Pulsfolge und die zugehörigen Übertragungsfunktionen im Frequenzbereich.

und die Übertragungsfunktionen in Abb. 4.10 (rechts) erklären die verbleibenden Oszillationen bei dem Leckintegrator und dem *Chebyshev-Entwurf*: die Dämpfung bei hohen Frequenzen reicht hier offensichtlich nicht aus. Eine Kaskadierung zweier Leckintegratoren könnte zwar ein Filter vom Grad zwei mit ausreichend hoher Dämpfung ergeben. Allerdings würde sich damit die Einschwingzeit noch weiter vergrößern. Bei dem Chebyshev Filter ergibt sich die selbe Problematik, wenn die geforderte Sperrdämpfung erhöht wird, da dieser Filtertyp



eher für andere Aufgabenstellungen geeignet ist. Einen Ausweg bietet hier jedoch ein *Butterworth-Filter* (Potenztiefpas) vom Grad zwei, welches eine kurze Einschwingzeit bei offensichtlich ausreichender Dämpfung liefert. Darüber hinaus läßt sich mit einem modifizierten Butterworth-Entwurf<sup>1</sup> eine Wellendigitalstruktur nach GRAY & MARKEL (siehe [Zal91, GM73]) entwickeln, mit der auch eine Separation in Erregungs- und Abklingphase für den Neurocomputeralgorithmus durchgeführt werden kann. Dies kann erreicht werden, wenn nur der Zählerkoeffizient  $a_0$  der z-Übertragungsfunktion (2.37) von Null verschieden ist und für die übrigen Koeffizienten  $a_{1,2} = 0$  gilt. Das Membranpotential eines Neurons ist in diesem Fall (wie bei der Struktur des Leckintegrators) bis auf einen Faktor eine Zustandsvariable des Filters, auf welchen in der Erregungsphase direkt die Eingangsspikes wirken. Dies ist von Vorteil, da diese spezielle Filterstruktur eine numerische Stabilität trotz starker Quantisierungseinschränkungen gewährleistet [Fet86, Gaz85]. Ob diese Filterstruktur, welche aus einer Kette von Leitungselementen abgeleitet werden kann [Zal91], biologisch plausibler ist als der Leckintegrator muß an dieser Stelle allerdings offen bleiben.

Die resultierenden Verbesserungen in Abb. 4.11 (rechts) sprechen für sich: hohe Frequenzen werden effektiv unterdrückt und der stationäre Zustand wird in etwa der halben Simulationszeit erreicht, verglichen mit dem Leckintegrator in Abb. 4.11 (links).

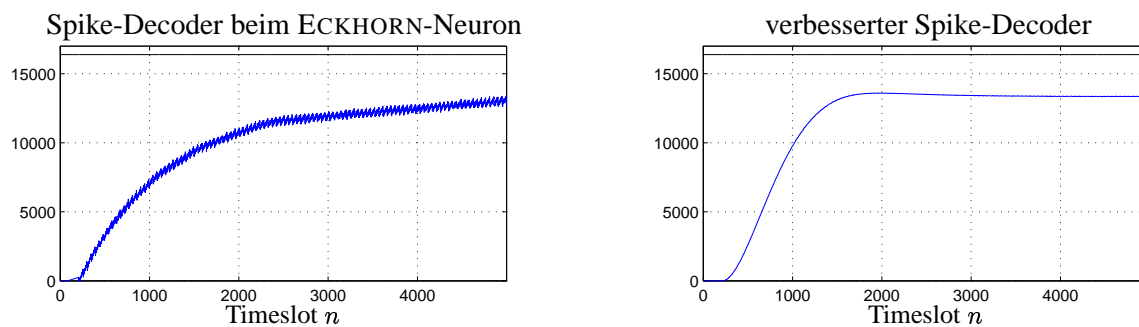


Abbildung 4.11: Integration einer Pulsfolge mittels des Spike-Decoders beim ECKHORN-Neuron (links) und die Leistungsfähigkeit des verbesserten Spike-Decoders (rechts).

In Abb. 4.12 sind abschließend einige Simulationsergebnisse gezeigt. Dazu wurde zunächst ein Bild als Eingangsreiz dem Netzwerk präsentiert und in den verschiedenen Zellschichten verarbeitet. Der Übersichtlichkeit wegen sind hier nur die Antworten der (geraden) simplen Zellen mit einer Orientierung von  $120^\circ$  abgebil-

<sup>1</sup>unter Verwendung einer reellen *charakteristischen Funktion*  $C(\psi) \sim \psi^2/(1 + \psi)^2$  mit  $\psi = (z - 1)/(z + 1)$  als äquivalente komplexe Frequenz



det. Um das Übertragungsverhalten des Netzwerkes beurteilen zu können, wurde zudem ein Impuls als Stimulus verwendet (rechte Spalte). Die daraus resultierenden Impulsantworten der simplen Zellen im Vergleich zu den gewünschten Antworten des linearen Modells geben Aufschluß über die erreichte Qualität der gesamten Simulation.

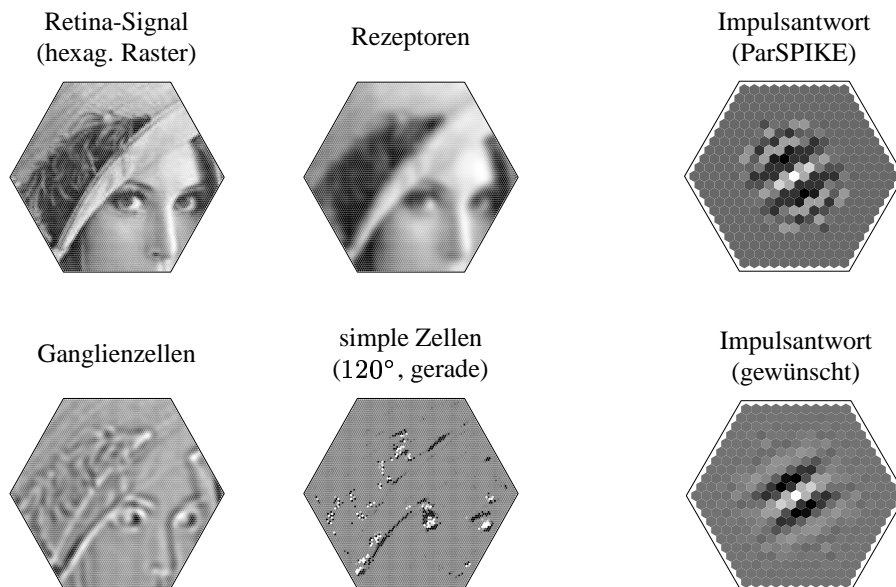


Abbildung 4.12: Simulation des Netzwerkes auf der ParSPIKE Architektur. links/mitte: Ergebnis bei einem Bild als Eingangsstimulus. rechts: Antwort der simplen Zellen bei einem pulsformigen Reiz (Impulsantwort) und Vergleich mit dem gewünschten Verhalten.

Die Geschwindigkeit der Simulation hängt stark von der parallelen Lastverteilung ab. In dem Simulationsbeispiel ist die Auslastung des am stärksten beanspruchten Prozessors nur etwa 4% größer als die durchschnittliche Belastung, was eine sehr gute Lastverteilung bedeutet. Die Messungen zur Lastverteilung entstanden durch Simulationen auf einer Sun Ultra60 Workstation. In Kombination mit Ergebnissen einer DSP Implementation mittels eines Testboards von Analog Devices konnten recht genaue Schätzungen für die Simulationsgeschwindigkeit gegeben werden [TWH00, Thi00, Wol01]. So benötigt die Simulation auf einem Prozessor der Sun Ultra60 etwa 1600 ms für einen Zeittakt. Das ParSPIKE RC VME Board hingegen, könnte diesen voraussichtlich in nur 36 ms berechnen, was eine beeindruckende Beschleunigung der Simulationsgeschwindigkeit um den Faktor  $\approx 50$  bedeutet.

In diesem Kapitel wurde eine Strategie zur Erweiterung linearer ortsdiskreter Übertragungssysteme für eine pulscodierte Verarbeitung auf der Basis des be-

kannten ECKHORN-Neurons entwickelt. Die Zielsetzung dieses Modellneurons, die Realisierung der Synchronisationseffekte, ist allerdings nicht Gegenstand dieser Arbeit und wurde daher nicht weiter betrachtet. Trotz des komplizierten und auf den ersten Blick nichtlinearen Verhaltens der pulscodierenden Strukturen kann dennoch eine lineare örtliche Filtercharakteristik realisiert werden. Damit genügt es, die Filterkoeffizienten des linearen Filtermodells ohne irgendeine Pulscodierung zu ermitteln, was den Entwurf (linearer) pulscodierter neuronaler Netze wesentlich vereinfacht. In diesem Rahmen wurde ein großes neuronales Netzwerk, welches einige wesentliche, für die Bildverarbeitung interessante Funktionen biologischer Sehsysteme modelliert, mit Erfolg auf einer Spezialhardware Architektur ParSPIKE implementiert. Auf diese Weise steht nun eine gaborähnliche Konturextraktion auf hexagonalem Abtastraster für ein pulscodiertes Bildverarbeitungsnetzwerk zur Verfügung. Aufgrund der Quantisierungseffekte der digitalen Umsetzung und einiger Einschränkungen der verwendeten Hardwareimplementation wichen die Antworten stark von den idealen GABOR Antworten ab. Diese numerischen Probleme konnten jedoch durch einen neuen Entwurf des Spike-Decoders effektiv minimiert werden, welcher auf diese Weise das bekannte ECKHORN- Modellneuron für die hier dargelegte Aufgabenstellung wesentlich verbessert. Die Simulationsergebnisse zeigen das gewünschte GABOR Verhalten und liefern eine Schätzung für die Simulationsgeschwindigkeit der verwendeten Spezialhardware.

# Adaption bei variierenden Beleuchtungsbedingungen

Variierende Beleuchtungsbedingungen sind stets eine Problematik, wenn reale Bildszenen auf Digitalrechnern verarbeitet werden. Insbesondere Rauschprozesse, die mit der Bildgebung selbst einhergehen, z. B. Ausleserauschen der CCD-Elektronik oder das Photonenrauschen, können zu falschen Verarbeitungsergebnissen bis hin zu unbrauchbaren Leistungen der Objekterkennung führen. Das Sehsystem des Menschen zeigt hingegen die vorteilhafte Fähigkeit, sich an Veränderungen in der wahrzunehmenden Szene, wie z. B. sich verändernde Lichtverhältnisse, anzupassen.

In diesem Kapitel wird eine biologisch motivierte Vorverarbeitung für eine effektive Rauschminderung in dunklen und daher stark verrauschten Bilddaten vorgestellt [TEDH02, Thi01]. Auf diese Weise sollen insbesondere Konturmerkmale robust extrahiert werden, um Mustererkennungsaufgaben auch in extremen Situationen durchführen zu können. Nach einer kurzen Einführung in die Problematik der adaptiven Rauschunterdrückung wird auf die Entstehung des Rauschens bei bildgebenden Verfahren selbst eingegangen. Die Analyse und Entwicklung eines neuartigen Adaptionskonzeptes zur Rauschunterdrückung bildet hier einen Schwerpunkt der Untersuchungen. Daraufhin wird ausführlich die Erweiterung des in Kapitel 3 entwickelten Modells behandelt, um das adaptive Verhalten bis hin zur Extraktion orientierter Konturmerkmale zu leisten.

## 5.1 Verfahren zur Rauschunterdrückung

Die Detektion von Kanten oder Linien in Kamerabildern ist ein wichtiges Mittel zur Interpretation einer Szene sowohl im menschlichen Sehsystem als auch in technischen Systemen. Üblicherweise werden in vielen Anwendungen dazu Gra-

dientenfilter eingesetzt, welche jedoch bekannterweise recht störempfindlich auf Rauschen in Kamerabildern reagieren. In Abb. 5.1 ist dazu ein einfaches Beispiel einer realen verrauschten Szene dargestellt, in der eine Kantendetektion durchgeführt wird. Ist die Szene sehr schwach beleuchtet, so führt eine Kantendetektion in der Regel auf unbrauchbare Ergebnisse. Eine Objekterkennung ist mit diesen Daten nicht möglich. Da es sich bei Rauschprozessen meist um hochfrequente

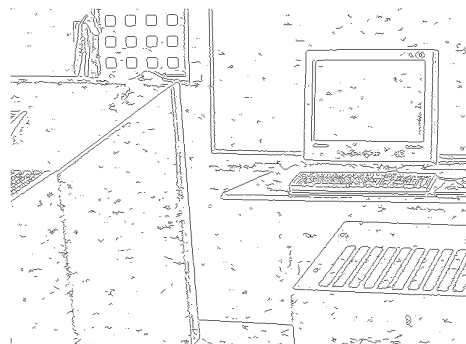


Abbildung 5.1: Kantendetektion in dunklen verrauschten Szenen. oben: Bei gedämpftem Licht mit niedrigem SNR kann noch ein brauchbares Kantenbild erzeugt werden. unten: Die dunkle Szene mit sehr niedrigem SNR führt zu einem stark verrauschten und damit unbrauchbaren Kantenbild.

Störungen handelt, die durch eine Gradientenbildung verstärkt werden, bringt eine vorherige Tiefpaßfilterung zur Unterdrückung des Rauschens in den meisten Fällen bessere Ergebnisse. Die Schwierigkeit liegt allerdings in der Wahl der Filterparameter, da eine zu starke Filterung auch mehr Details der Szene selbst unterdrückt. In der Bildverarbeitung hat sich die Klasse der sogenannten *Gaußableitungsfilter* (*gaussian derivatives*) in diversen Formen und Implementationen etabliert, bei der im wesentlichen ein Gaußtiefpaßfilter mit einem Gradientenoperator kombiniert wird. Viele weitere Details zu diesem Thema sind in [MH80, SM87, Der93, vVYV98] nachzulesen. Abgesehen davon, ob diese Imple-

mentierungen einen adaptiven Einsatz ermöglichen oder nicht, beschränken sich die Arbeiten bislang meist auf konkrete Skalierungen bzw. Auflösungspyramiden. Eine Adaption der Vorfilterung auf die Stärke des Rauschens in den Bildern bleibt meist aus.

Das klassische *WIENER-Filter* [Lim90] hingegen ist ursprünglich ein Ansatz zur Schätzung eines Signals in einer gestörten Beobachtung und diente als Basis für eine Vielzahl von Anwendungen in der Sprach- und Bildverarbeitung. Da es ein linearer Schätzer ist, welcher einen mittleren quadratischen Fehler zwischen dem zu rekonstruierenden Signal und der Schätzung verwendet, fällt es in die Klasse der *linear minimum mean square error estimator*. Bei gaußverteilten Prozessen handelt es sich sogar um den optimalen Schätzer. Die Übertragungsfunktion des WIENER-Filters zur Reduzierung von additivem Rauschen in Bildsignalen verwendet dabei Annahmen über die Leistungsdichtespektren des Signals bzw. der Störung. Da es sich um ein ortsinvariantes Filter handelt, werden Unterschiede in den einzelnen Bildbereichen wie helle und dunkle bzw. texturlose und strukturreiche Bereiche nicht berücksichtigt. Es wird überall die selbe Filterstärke verwendet. Hier etablierte sich das *adaptive WIENER-Filter* nach LEE [Lim90, KZ96, Lee80, Mat]. Dieses Verfahren verwendet einfache statistische Merkmale (Mittelwert und Varianz), um in einer lokalen Umgebung jedes Pixels eine Filterung durchzuführen. Dabei wird eine gewichtete Summe des Eingangssignals und des geschätzten Mittelwertes gebildet mit dem Ziel, in Regionen mit hoher Varianz eine geringe Filterwirkung zu verwenden, da dort Kanten vermutet werden. Nachteil hierbei ist, daß die Rauschleistung bekannt oder eine Annahme darüber vorhanden sein muß. Ist diese Annahme nicht korrekt, so hat das Verfahren allerdings den Effekt, daß bei Regionen hoher durch Rauschen verursachter Varianz - obwohl dort keine Kanten vorliegen - ebenfalls eine zu geringe Filterstärke eingesetzt wird. Darüber hinaus setzt diese Methode voraus, daß es sich um unkorreliertes Rauschen handelt, was bei der Bildaufnahme aber nicht gegeben sein muß. Als ein in der Bildverarbeitung etabliertes Verfahren wird es später als Referenz herangezogen, um einen Vergleich mit dem hier entwickelten Verfahren zu ermöglichen.

Das Verfahren zur *Glättung in einer selektierten Nachbarschaft* ist ebenfalls ein lokal arbeitendes Verfahren [KZ96, NM79, WW88] zur Rauschunterdrückung. Hier werden Nachbarn des aktuellen Grauwertes dahingehend untersucht, ob sie zu der selben Region gehören wie der aktuelle Grauwert, um die Gewichtungen der Nachbarn für den Glättungsprozeß abzuleiten. Die Schwierigkeit ist dabei, eine Schwelle für die Zugehörigkeit zu einer Region festzulegen. Wie bei allen Segmentierungsverfahren kann diese Vorgehensweise beliebig aufwendig sein. Der

Vollständigkeit halber sei hier auch der Ansatz mittels *MARKOV Random Fields* in [PK00] zur Rauschunterdrückung genannt.

Ein Vorzug des biologischen Sehsystems ist die automatische Anpassung der Modell- bzw. Filterparameter an Veränderungen der Beleuchtungssituation und daher auch an die Rauschstärke, um eine robuste Merkmalsextraktion oder Objekterkennung zu gewährleisten. Üblicherweise ist die wahrgenommene Szene mit Rauschen überlagert, welches in den lichtempfindlichen Sensorelementen und der Ausleseelektronik entsteht, und zu einer Verschlechterung des Signal-zu-Rauschverhältnisses (SNR) führt. Besonders in dunklen Szenen kann die Rauschleistung bereits in der Größenordnung der Signalleistung liegen. Am biologischen Vorbild allerdings wurde gezeigt, daß die rezeptiven Felder der retinalen Photorezeptoren mit der Verdunkelung einer Szene anwachsen [Hub88, SL96a]. Dieses Verhalten führt offensichtlich zwar zu einer Abnahme der Auflösung, durch die Mittelung über die eigentliche Signalinformation der Szene und zugleich des Rauschprozesses über ein größeres rezeptives Feld kann jedoch vor allem das Rauschen weiter unterdrückt werden. Zusätzlich wird bei dunklerer Szene aber auch die Öffnung der Iris geweitet, um mehr Licht zu sammeln, was ebenfalls zu einer Verringerung der Ortsauflösung des retinalen Intensitätssignals führt. Diese Beziehung ist in Abb. 3.3 zu erkennen, so daß hier die Komponenten des biologischen Systems aufeinander abgestimmt zu sein scheinen. Eine automatische Blende in einer Kamera agiert auf ähnliche Weise.

Die beschriebene Vorfilterung durch die Photorezeptoren kann wie in Kapitel 3 erwähnt als Gaußtiefpaßfilter modelliert werden [SL96a, TH00, TWH00, Thi00]. Um der Verschlechterung des SNR in dunklen Szenen entgegenzuwirken, wird die Größe der rezeptiven Felder bzw. die Koeffizienten dieses Gaußfilters adaptiv zur (lokalen) Helligkeit in der Szene verändert. Auf diese Weise versucht das System einen Kompromiß zwischen Auflösung und robuster Kantendetektion zu erreichen. Obwohl dieser adaptive biologische Mechanismus in einigen Arbeiten erwähnt ist [SL96a, BC92], sind weitere Untersuchungen und technische Implementationen bislang nicht bekannt. Im weiteren wird zunächst der theoretische Zusammenhang zwischen adaptiver Vorfilterung und dem Signal-zu-Rauschverhältnis näher betrachtet, um einen Algorithmus zu entwickeln, der das SNR verbessern oder gar auf einem bestimmten Niveau halten kann [TEDH02, Thi01]. Dabei soll bewußt zunächst aus der Sicht eines technischen Bildverarbeitungssystems das Signal-zu-Rauschverhältnis als objektives Beurteilungskriterium herangezogen werden. Anschließend wird die wichtige Erweiterung des Modells aus Kapitel 3 behandelt, um die Frage zu klären, wie das biologische Vorbild die adaptiven Mechanismen in einem Gesamtsystem zu realisieren vermag.

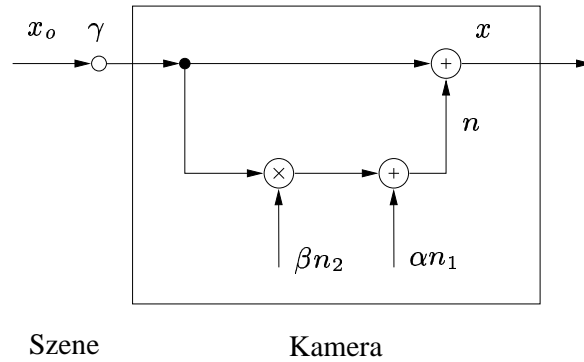
## 5.2 Rauschprozesse bei der Bildakquisition

Für die Entwicklung eines geeigneten Algorithmus zur Rauschunterdrückung ist es wichtig, die diversen Quellen des Rauschens in einem bildgebenden System zu betrachten. Bei der Bildaufnahme mit einer CCD-Kamera bilden sich durch einfallende Photonen Ladungspakete in den photoempfindlichen Zellen, die durch seriellles Auslesen in einen Datenstrom umgesetzt werden. Hierbei werden dem eigentlichen Informationsgehalt der betrachteten Szene eine Reihe von Störgrößen während des bildgebenden Prozesses überlagert, welche i. allg. stochastischer Natur sind, aber auch durchaus systematische Komponenten beinhalten können [BC92, Kam]. Bei guten Beleuchtungssituationen fallen diese Rauschanteile i. allg. nicht auf. Nimmt die Helligkeit der Szene aber ab, so ist bereits bei Innenaufnahmen ohne künstliche Beleuchtung (bei TV-Reportagen zu beobachten) ein gewisses Rauschen wahrnehmbar. Mit der Dunkelheit der Szene wird dieser Effekt immer stärker. Hier können im wesentlichen drei Rauscharten unterschieden werden, welche hier nur kurz genannt werden sollen:

- **Photonenrauschen (Photon shot noise)**  
Diese Rauschart wird durch die Quantennatur der Photonen hervorgerufen, da die Einfallrichtung und die Anzahl der Photonen auf ein lichtempfindliches Element während der Belichtung nicht konstant ist, sondern zufällig variiert.
- **Ausleserauschen (Read noise)**  
In diese Kategorie fallen eine Vielzahl unterschiedlicher Rauschquellen, die zum einen technischer Natur sein können (wie z. B. das Dunkelstromrauschen oder auch thermisches Rauschen), zum anderen aber auch von dem Design des Sensors oder der Signalverarbeitung abhängen (Rauschen des Ausgangsverstärkers). Die Art des Rauschens ist vor allem bei niedrigen Signalpegeln von Bedeutung. So ist bei extrem dunklen Aufnahmen, wie der Überwachung bei Nacht, oder aber auch bei der Astrofotografie die Rauschleistung im Vergleich zur Signalleistung derart hoch, daß die Geräte meist künstlich gekühlt werden müssen.
- **Musterrauschen (Pattern noise)**  
Durch die individuelle Empfindlichkeit der einzelnen Sensorelemente und irreguläre Kristallstrukturen ergeben sich signalabhängige (PRNU, Photo-response Nonuniformity) und signalunabhängige (FPN, Fixed Pattern Noise des Dunkelstromes) örtliche Ungleichmäßigkeiten. Diese systematischen

Muster können durch Mittelung von Serienaufnahmen gleichmäßiger Helligkeit oder von Dunkelbildern extrahiert und sichtbar gemacht werden.

In [Kam] ist die Komplexität der Rauschentstehung bei CCD- und Videobildern im Detail beschrieben. Für die folgenden Analysen dieser Arbeit ist es aber möglich und zweckmäßig, die Überlagerung des eigentlichen gewünschten Bildsignals mit den verschiedenen Rauscharten in dem vereinfachten Rauschmodell in Abb. 5.2 zu abstrahieren.



Kamerasignal:	$x(\mathbf{k}) = \gamma x_o(\mathbf{k}) + n(\mathbf{k})$
Gesamtrauschen:	$n(\mathbf{k}) = \alpha n_1(\mathbf{k}) + \gamma x_o(\mathbf{k}) \beta n_2(\mathbf{k})$
Rauschquellen:	$n_1 \perp n_2, \sigma_{n_{1,2}}^2 = 1, \mu_{1,2} = 0$
Rauschparameter:	$\alpha, \beta \in \mathbb{R}$
Originalszene:	$x_o (\perp n_{1,2})$
Beleuchtungsstärke:	$0 \leq \gamma \leq 1$

Abbildung 5.2: Das verwendete Rauschmodell.

Das Kamerasignal

$$x(\mathbf{k}) = \gamma x_o(\mathbf{k}) + n(\mathbf{k}) \quad (5.1)$$

$$= \gamma x_o(\mathbf{k}) + \alpha n_1(\mathbf{k}) + \gamma x_o(\mathbf{k}) \beta n_2(\mathbf{k}) \quad (5.2)$$

ist hierbei eine Summe aus Originalszene  $x_o$ , welche i. allg. um einen reellen Faktor  $0 \leq \gamma \leq 1$  verdunkelt ist, einem unkorrelierten Rauschanteil  $\alpha n_1$  und einem korrelierten Rauschterm  $\gamma x_o \beta n_2$ .  $\alpha, \beta$  seien reelle positive Faktoren. Die weißen Rauschprozesse  $n_{1,2}$  sind unkorreliert ( $\perp$ ), gaußverteilt, mittelwertfrei und haben die Varianz eins. Eine detaillierte Analyse des Kamerarauschens ist in [BC92, Kam] gegeben.



Das Signal-zu-Rauschverhältnis (SNR, in dB: Signal-zu-Rauschabstand,  $SNR_{\text{dB}}$ ) des zur Verfügung stehenden Kamerasignals  $x$ , hier als  $SNR_x$  bezeichnet, kann dann durch

$$SNR_x(\gamma) = \frac{E\{(\gamma x_o)^2\}}{E\{(\alpha n_1 + \gamma \beta x_o n_2)^2\}} \quad (5.3)$$

$$= \frac{\gamma^2 E\{x_o^2\}}{\alpha^2 E\{n_1^2\} + \gamma^2 \beta^2 E\{x_o^2 n_2^2\} + 2\gamma \alpha \beta E\{x_o n_1 n_2\}}, \quad (5.4)$$

berechnet werden und ist neben den Rauschparametern stark abhängig von der Leistung der Szene bzw. der Beleuchtung  $\gamma$ . Mit  $E\{\}$  ist der Mittelwertoperator bzgl. des Ortes abgekürzt. Die reellen Rauschparameter  $\alpha, \beta$  sind hierbei nicht eindeutig für ein bestimmtes SNR, da sie den Beitrag des korrelierten und unkorrelierten Rauschens zum Gesamttrauschen beschreiben. So ist beispielsweise

$$\beta = \sqrt{\frac{\gamma^2 E\{x_o^2\} - \alpha^2 SNR}{\gamma^2 E\{x_o^2 n_2^2\} SNR}} \quad \text{für } E\{x_o n_1 n_2\} \approx 0. \quad (5.5)$$

Die Abhängigkeit des SNR von der Beleuchtung ist hier von großem Interesse und muß genauer diskutiert werden. Im allgemeinen verringert (verschlechtert) sich das SNR, wenn  $\gamma$  kleiner wird. Genaugenommen ist das Verhalten von der Diskriminanten [Bro99]

$$\Delta = (2\alpha\beta)^2 (E\{n_1^2\} E\{x_o^2 n_2^2\} - E\{x_o n_1 n_2\}^2), \quad (5.6)$$

abhängig, welche aber auf Grund der Gleichung von CAUCHY-SCHWARZ nicht negativ sein kann. Der recht unwahrscheinliche Fall  $\Delta = 0$  besitzt 2 Pole in  $\gamma$  und wird nicht weiter betrachtet. Der wahrscheinlichere Fall  $\Delta > 0$  weist keinen reellen Pol auf und läßt das SNR gegen den Grenzwert

$$\lim_{\gamma \rightarrow \infty} SNR_x(\gamma) = \frac{E\{x_o^2\}}{\beta^2 E\{x_o^2 n_2^2\}}, \quad (5.7)$$

konvergieren. Dies geschieht entweder monoton steigend, wenn  $E\{x_o n_1 n_2\} \geq 0$  oder mit einem lokalen Maximum, wenn  $E\{x_o n_1 n_2\} < 0$ . Die Simulationen und Messungen an realen Kameradaten haben hier stets einen monotonen Verlauf gezeigt. Dieses Modell wird im weiteren Verlauf auch für die Generierung der Testbilder mit künstlichem Rauschen verwendet. Ein typischer Verlauf der Abhängigkeit und die beiden Extremfälle völlig korrelierten ( $\alpha = 0$ ) und unkorrelierten Rauschens ( $\beta = 0$ ) sind in Abb. 5.3 verdeutlicht.

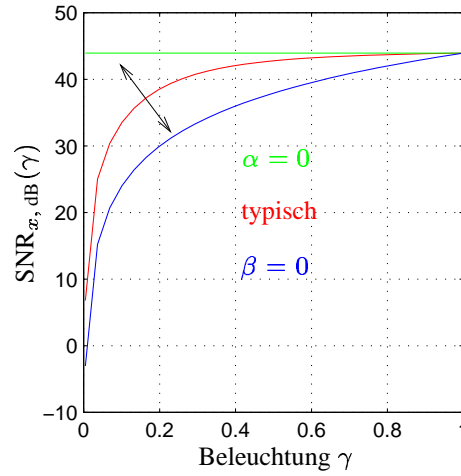


Abbildung 5.3: Verschlechterung des SNR bei abnehmender Beleuchtung  $\gamma$  für verschiedene Rauschparameter.

### 5.3 Adaptive Filterung zur Rauschunterdrückung

In Anlehnung an die Filtercharakteristik der Photorezeptoren in der menschlichen Retina wird im weiteren ein Gaußtiefpapfilter mit der Impulsantwort (bzw. dem rezeptiven Feld)

$$h(\mathbf{k}) = g(\mathbf{k}) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^2} e^{-\frac{\|\mathbf{k}\|^2}{2\sigma^2}} \quad (5.8)$$

und der Übertragungsfunktion

$$h(\mathbf{k}) \circ \bullet H(\mathbf{v}) \approx e^{-\frac{\|\mathbf{v}\|^2 \sigma^2}{2}} \quad , \mathbf{v} = [v_1 v_2]^T \quad (5.9)$$

$$, -\pi \leq v_{1,2} \leq \pi$$

als erste Stufe einer Prozeßkette angenommen, die, wie in Kapitel 3 beschrieben, Konturinformationen extrahiert [SL96a, TH00, TWH00, Thi00]. Zur Vereinfachung der Beschreibung aller Zusammenhänge in diesem Kapitel wird hier meist die Notation für ein technisches Bildverarbeitungssystem unter Verwendung eines rechtwinkligen Abtastrasters verwendet, so daß  $\mathbf{T}$  und  $\sigma$  in Pixelkoordinaten (statt in  $\mu\text{m}$  oder csp) angegeben sind und damit auch  $\mathbf{T} = \mathbf{T}^{-T}$  bzw.  $\mathbf{v} = \boldsymbol{\omega}$  gilt. Da dies allerdings die Ergebnisse nicht einschränkt, kann zu jeder Zeit eine Interpretation im Hinblick auf eine biologische Betrachtung gegeben werden.

Die Adaption im biologischen Sehsystem ist lokal, d. h. die Größe der einzelnen rezeptiven Felder der Photorezeptoren kann mit dem Ort variieren ( $\sigma = \sigma(\mathbf{k})$ ). Im

Rahmen dieser Arbeit liegt der Schwerpunkt allerdings auf der Entwicklung eines globalen Mechanismus. Das Ziel ist also die Adaption eines örtlich invarianten Filters mit der Hilfe eines Steuersignals, welches aus dem Eingangsbild gewonnen werden kann. Die Untersuchungen haben gezeigt, daß ein globaler Mechanismus ausreicht, um eine adaptive Rauschunterdrückung erfolgreich zu realisieren. Ein lokal arbeitender Algorithmus erhöht zudem im Hinblick auf eine technische Nutzbarmachung den Berechnungsaufwand auf sequentiell arbeitenden Digitalrechnern enorm und soll daher nicht weiter behandelt werden. Allerdings scheint dieser in der Biologie seine Berechtigung zu besitzen, da die Steuerung der Stärke der Rezeptorkopplung eigentlich nur aus dem lokalen Umfeld eines Rezeptors gewonnen werden kann. Ein globales Steuersignal wäre hingegen biologisch nicht plausibel, was in Abschnitt 5.4 näher erläutert wird. Weitere Untersuchungen und der Vergleich zwischen lokaler und globaler Rauschunterdrückung können in [Eis03] nachgelesen werden.

Im biologischen Vorbild variiert die Größe der rezeptiven Felder der Zapfen und damit der Filterparameter des Gaußfilters in einem Bereich von  $\sigma = 1.5$  bis 12 (Pixel oder cone spacing) für eine Adaption von hellen bis dunklen Szenen. Diese Randbedingung soll auch für die globale Filterung berücksichtigt werden.

Im folgenden wird der Einfluß einer Gaußfilterung auf das Signal-zu-Rauschverhältnis untersucht. Dabei kann experimentell und theoretisch gezeigt werden, daß durch solch eine Filteroperation das SNR bei dem verwendeten Rauschmodell verbessert werden kann. Aus den beschriebenen Zusammenhängen heraus wird dann ein plausibler Steuermechanismus abgeleitet, der zu einer gemessenen Beleuchtungssituation  $\gamma$  einer Szene ein notwendiges  $\sigma$  einer Gaußfilterung berechnet und somit eine adaptive Filterung ermöglicht.

### 5.3.1 SNR einer gefilterten Szene

In Abb. 5.4 ist das Beispiel der Filterung einer künstlich verrauschten Szene unter variablen Lichtverhältnissen gezeigt. Dazu wurde dem Originalbild (oben links) ein Rauschprozeß (hier mit  $\alpha = 1.8 \cdot 10^{-3}$ ,  $\beta = 9.5 \cdot 10^{-3}$ ) überlagert, so daß sich ein Signal-zu-Rauschabstand von  $\text{SNR}_{x, \text{dB}} = 40 \text{ dB}$  einstellt. Die Ermittlung der Rauschparameter einer realen Kamera wird in Kapitel 6 beschrieben. Dieser Szene wurde die Beleuchtung  $\gamma = 1$  zugeordnet (mitte links). Durch die Verdunkelung auf  $\gamma = 0.01$  (untere Reihe) verschlechtert sich das SNR um etwa 30 dB auf  $\text{SNR}_{x, \text{dB}} = 10 \text{ dB}$ . Eine anschließende Tiefpaßfilterung (hier mit  $\sigma = 1$ ) allerdings scheint objektiv betrachtet die Qualität des Bildes bzgl. des Rauschens zu erhöhen (unten rechts). Für beide Beleuchtungssituationen wurde schließlich

die Abhängigkeit des SNR von dem Filterparameter (im folgenden als  $\text{SNR}_y$  bezeichnet) experimentell bestimmt (oben rechts). Hierbei fällt auf, daß durch die Filterung Details zwar unter Umständen verloren gehen, das SNR aber jeweils verbessert werden konnte.

Numerische Untersuchungen des resultierenden SNR einer gefilterten Szene vieler gefilterter Kamerabilder führen zu der Annahme, daß das SNR in weiten Bereichen stets monoton steigend mit  $\sigma$  ist. Dies bedeutete, daß das SNR umso besser würde, je stärker das Tiefpaßverhalten ist, was auch in Abb. 5.4 (oben rechts) exemplarisch zu sehen ist. Wie man im weiteren allerdings erkennen wird, sind die Abhängigkeiten des SNR so komplex, daß die simulierten Ergebnisse nicht allgemein gezeigt werden können. Unter bestimmten Vereinfachungen können aber relativ gute Interpretationen geliefert werden, die zu sinnvollen und effektiven adaptiven Steueralgorithmen führen werden. Diese neuartigen Herleitungen sollen nun im einzelnen gegeben werden.

Die Verarbeitung des verrauschten und verdunkelten Bildsignals  $x(\mathbf{k})$  mit einem Filter beschrieben durch (5.9) führt auf das gefilterte Bild

$$y(\mathbf{k}) = h(\mathbf{k}) * x(\mathbf{k}) \quad (5.10)$$

$$= \gamma h(\mathbf{k}) * x_o(\mathbf{k}) + h(\mathbf{k}) * n(\mathbf{k}), \quad (5.11)$$

welches ebenfalls aus einem Signal- und einem Rauschanteil besteht. Das resultierende SNR läßt sich dann berechnen zu

$$\text{SNR}_y(\gamma, \sigma) = \gamma^2 \frac{\text{E} \{ (h * x_o)^2 \}}{\text{E} \{ (h * n)^2 \}}. \quad (5.12)$$

Mit Hilfe der PARSEVALSchen Gleichung kann das SNR im Frequenzbereich formuliert werden. Aus der Tatsache heraus, daß das Rauschen, auch wenn es aus einem korrelierten Anteil besteht, in allen Testreihen näherungsweise ein weißes Leistungsdichtespektrum  $\sigma_n^2 = \text{E} \{ n^2 \}$  aufweist, erhalten wir für reelle Signale weiter

$$\text{SNR}_y(\gamma, \sigma) = \gamma^2 \frac{\int |H(\mathbf{v})|^2 |X_o(\mathbf{v})|^2 d\mathbf{v}}{\int |H(\mathbf{v})|^2 \sigma_n^2 d\mathbf{v}} \quad (5.13)$$

$$= \frac{\gamma^2}{\sigma_n^2} \frac{\int |H(\mathbf{v})|^2 |X_o(\mathbf{v})|^2 d\mathbf{v}}{\int |H(\mathbf{v})|^2 d\mathbf{v}}. \quad (5.14)$$

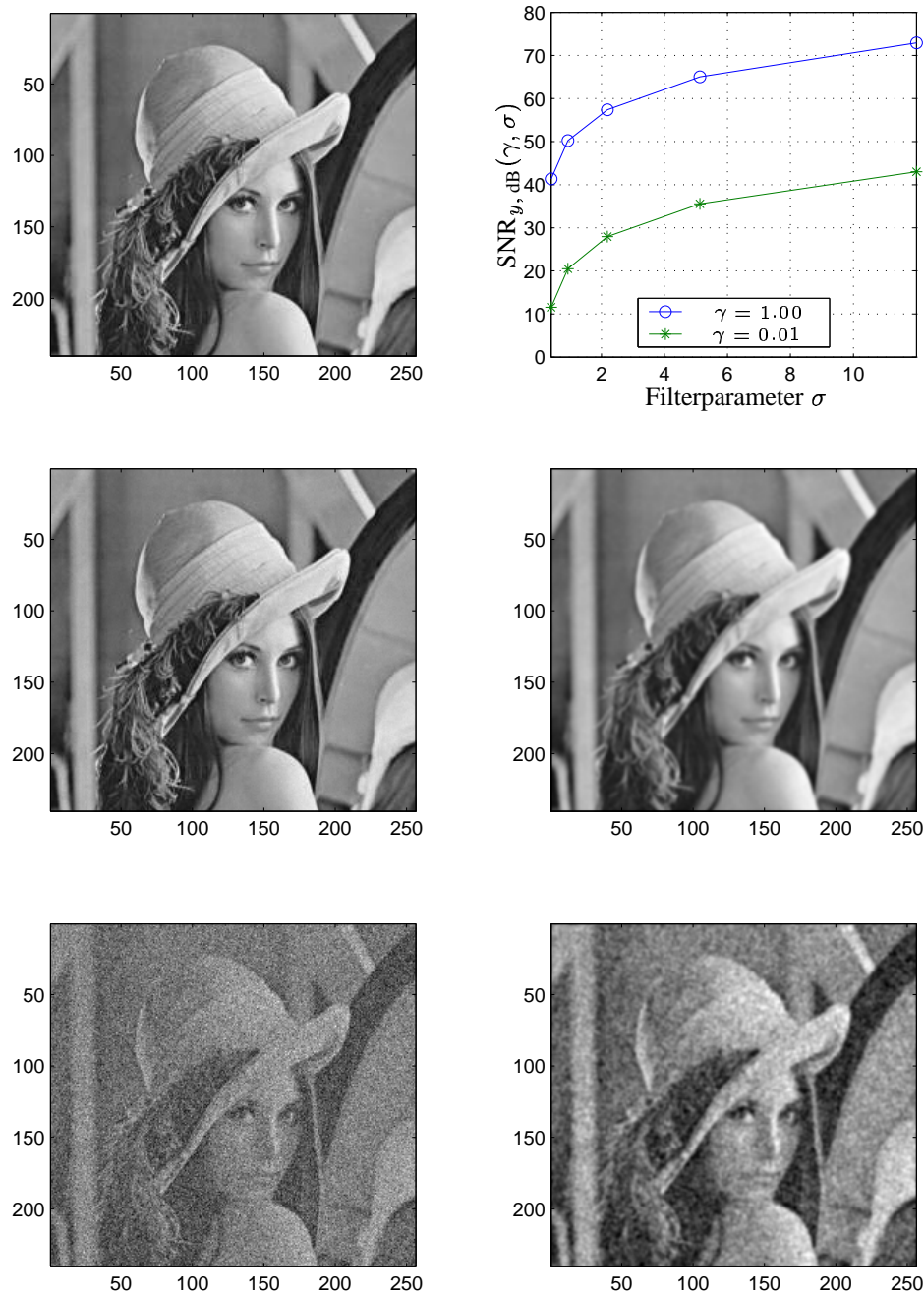


Abbildung 5.4: Verbesserung des SNR einer verdunkelten verrauschten Beispielszene durch Tiefpaßfilterung. oben: Unverraushtes Originalbild und SNR-Verbesserung für zwei unterschiedliche Beleuchtungen  $\gamma$ . mitte: Verrauschtes Originalbild  $SNR_{dB} = 40$  dB ( $\gamma := 1$ ) und das Filterergebnis mit  $\sigma = 1$ . unten: Verdunkelte Szene ( $\gamma = 0.01$ ) mit  $SNR_{dB} = 10$  dB und das Filterergebnis mit  $\sigma = 1$ .

### Annahme 1: Natürliche Bildszenen

Zur Vereinfachung lohnt es, sich auf eindimensionale Signale zu konzentrieren. Für den biologisch plausiblen Bereich (siehe Kapitel 3) des Filterparameters  $\sigma \geq 1.5$  können wir die Näherung

$$\int_{-\pi}^{\pi} |H(v)|^2 dv \approx \int_{-\infty}^{\infty} e^{-\sigma^2 v^2} dv = \frac{\sqrt{\pi}}{\sigma} \quad (5.15)$$

verwenden und erhalten für den eindimensionalen Fall

$$\text{SNR}_y^1(\gamma, \sigma) = \frac{2\gamma^2}{\sqrt{\pi}\sigma_n^2} \sigma \int_0^{\pi} e^{-\sigma^2 v^2} |X_o(v)|^2 dv. \quad (5.16)$$

Für die Untersuchung des Monotonieverhaltens des SNR bzgl.  $\sigma$  ist die partielle Ableitung

$$\frac{\partial \text{SNR}_y^1}{\partial \sigma} = \frac{2\gamma^2}{\sqrt{\pi}\sigma_n^2} \int_0^{\pi} e^{-\sigma^2 v^2} |X_o(v)|^2 (1 - 2\sigma^2 v^2) dv \quad (5.17)$$

zu diskutieren, welche die Grenzwerte

$$\lim_{\sigma \rightarrow 0} \frac{\partial \text{SNR}_y^1(\gamma, \sigma)}{\partial \sigma} = 2\sqrt{\pi} \text{SNR}_x^1(\gamma) > 0, \quad (5.18)$$

$$\lim_{\sigma \rightarrow \infty} \frac{\partial \text{SNR}_y^1(\gamma, \sigma)}{\partial \sigma} = 0. \quad (5.19)$$

zeigt. Die Annahme natürlicher Bilder erlaubt weitere Vereinfachungen. Dann nämlich klingt das Spektrum des Bildes näherungsweise mit  $|X_o(v)| \sim 1/(v + v_0)^p$ ,  $p \in \mathbb{R}^{>1}$  [TF97, WM97] ab und wir erhalten

$$\frac{\partial \text{SNR}_y^1(\gamma, \sigma)}{\partial \sigma} = \frac{2\gamma^2}{\sqrt{\pi}\sigma_n^2} \int_0^{\pi} e^{-\sigma^2 v^2} \frac{1 - 2\sigma^2 v^2}{(v + v_0)^{2p}} dv, \quad (5.20)$$

was für eine monoton steigende Funktion positiv sein sollte. Dieses Verhalten scheint gültig zu sein für alle untersuchten Bildszenen, da der gegebene Integrand schnell genug gegen null konvergiert. Für diesen Fall führt die Lage des kritischen Nulldurchganges

$$v_c = \min_{\sigma} \frac{1}{\sigma \sqrt{2}} = \frac{1}{12\sqrt{2}} \quad (5.21)$$

des Integranden zu einem positiven Beitrag für  $0 \dots v_c$  zum gesamten Integral, welches größer ist als der negative Beitrag im Bereich  $v_c \dots \pi$  und daher insgesamt zu einem positiven Ergebnis des Integrals und daher zu einer positiven Steigung des SNR.

Es ist erwähnenswert, daß die theoretische Interpretation, was das positive Vorzeichen der Steigung angeht, nicht von der Art des Rauschens abhängt, weder bei völlig unkorreliertem ( $\beta = 0$ ), noch bei korreliertem Rauschen ( $\alpha = 0$ ), so lange näherungsweise ein weißes Frequenzspektrum vorliegt. Untersuchungen haben bestätigt, daß die Leistung des originalen Bildsignals bei natürlichen Bildern auf niedrige Frequenzen konzentriert ist und damit auch der korrelierte Rauschanteil ebenfalls näherungsweise weiß bleibt. Die Größe der positiven Steigung  $\text{SNR}_y$  kann hingegen von der Art des Rauschens abhängig bleiben.

### Annahme 2: Konstantes Bildsignal

Neben den natürlichen Szenen sind Bildsignale mit konstanten Bildbereichen ein Sonderfall, für den eine theoretische Herleitung gegeben werden kann.

Das SNR des ungefilterten Bildes kann für  $x_o \equiv \text{const}$  nach (5.4) vereinfacht werden zu

$$\text{SNR}_x(\gamma) = \frac{\gamma^2 x_o^2}{\alpha^2 + \gamma^2 \beta^2 x_o^2}, \quad (5.22)$$

da die Rauschprozesse  $n_1$  und  $n_2$  als zueinander unkorreliert und mit einer Varianz eins angenommen werden können.

Für die gefilterte Szene ergibt sich dann für den zweidimensionalen Fall der Ausdruck

$$\text{SNR}_y(\gamma, \sigma) = \gamma^2 \frac{x_o^2}{\text{E}\{(h * n)^2\}} = \gamma^2 \frac{x_o^2}{\sigma_n^2 C \int |H(\mathbf{v})|^2 d\mathbf{v}} \quad (5.23)$$

$$= \gamma^2 \frac{x_o^2}{\sigma_n^2 C \pi / \sigma^2} = \frac{\gamma^2 x_o^2}{\sigma_n^2} \cdot \frac{\sigma^2}{C \pi}, \quad (5.24)$$

wobei  $C$  hier der Vorfaktor der gewählten FOURIER-Transformation ist. Da weiter die Leistung des ungefilterten Rauschens durch  $\sigma_n^2 = \alpha^2 + \gamma^2 \beta^2 x_o^2$  bestimmt ist, ergibt sich die Beziehung

$$\text{SNR}_y(\gamma, \sigma) = \frac{\gamma^2 x_o^2}{\alpha^2 + \gamma^2 \beta^2 x_o^2} \cdot \frac{\sigma^2}{C \pi} = \text{SNR}_x(\gamma) \cdot \frac{\sigma^2}{C \pi} \quad (5.25)$$

oder mit den logarithmischen Größen

$$\text{SNR}_{y, \text{dB}}(\gamma, \sigma) = \text{SNR}_{x, \text{dB}}(\gamma) + 10 \log_{10} \frac{\sigma^2}{C \pi}. \quad (5.26)$$

Dieses interessante Ergebnis ist einer Reihe von Annahmen und Näherungen unterworfen und gilt somit z. B. nur für  $\sigma > 1$ . Auch für beliebig große  $\sigma$  ist die Näherung nicht mehr sinnvoll, da nach (5.25) das SNR nicht gegen einen Grenzwert für wachsende  $\sigma$  konvergieren würde. Messungen bestätigen aber, daß die Kurvenform bzw. die Steigung des  $\text{SNR}_y(\sigma)$  nicht stark von  $\gamma$  abhängt, sondern die Beleuchtungsstärke hier lediglich einen konstanten Faktor ausmacht (bzw. eine additive Konstante in dB), was der obige Ausdruck auch bestätigt. Wie sich noch zeigen wird, ist dieses Ergebnis für praktische Zwecke durchaus ausreichend und von großem Interesse.

### 5.3.2 Konstruktion eines Steuermechanismus

Wie der vorherige Abschnitt zeigt, ist durch Filterung mit einem Gaußtiefpfaß eine Verbesserung des Signal-zu-Rauschverhältnisses in dunklen und verrauschten Bildszenen bis sehr große  $\sigma$  möglich. Offensichtlich kann es aber nicht zweckmäßig sein, stets das maximale  $\sigma$  zu verwenden. Dies liefert zwar das beste SNR, hierdurch würde jedoch immer die geringste örtliche Auflösung erzielt werden. Aus diesem Grund ist es sehr wichtig, einen Kompromiß zwischen „gutem“ SNR und ausreichender Ortsauflösung zu finden. In diesem Abschnitt wird daher eine mögliche Strategie vorgeschlagen, um ein  $\sigma$  für eine globale, d. h. ortsinvariante Filterung für eine bestimmte Beleuchtungsstärke  $\gamma$  des verrauschten Bildes zu finden. Ziel hierbei ist, durch die Filterung mit einem passend zu wählendem adaptiven  $\sigma_a(\gamma)$  das SNR der Szene soweit zu verbessern, daß ein gewünschtes SNR, z. B. das einer ungefilterten, gut beleuchteten Szene ( $\gamma = 1$ ) erreicht wird. Das Prinzip ist in Abb. 5.5 verdeutlicht. Die Bestimmung der notwendigen Filterstärke  $\sigma_a(\gamma)$  in Abhängigkeit von der Beleuchtung, was im folgenden *Steuersignal* genannt werden soll, kann z. B. durch einen Optimierungsprozeß mittels realen oder simulierten Kameradaten approximativ geschehen. Des weiteren kann aber auch die theoretische Herleitung des SNR eines gefilterten Bildes im vorherigen Abschnitt genutzt werden, um ein explizites Steuersignal aufzustellen. Beide Verfahren werden nun kurz vorgestellt.

#### Optimierungsprozeß

In einem Optimierungsprozeß wird die Abhängigkeit der adaptiven Filterparameter  $\sigma = \sigma_a$  von der Beleuchtung  $\gamma$  gesucht, um ein bestimmtes SNR zu erzielen, d. h. gesucht ist die Lösung

$$\sigma_a(\gamma) = \underset{\sigma}{\operatorname{argmin}} |\text{SNR}_y(\gamma, \sigma) - \text{SNR}_x(1)|^2. \quad (5.27)$$



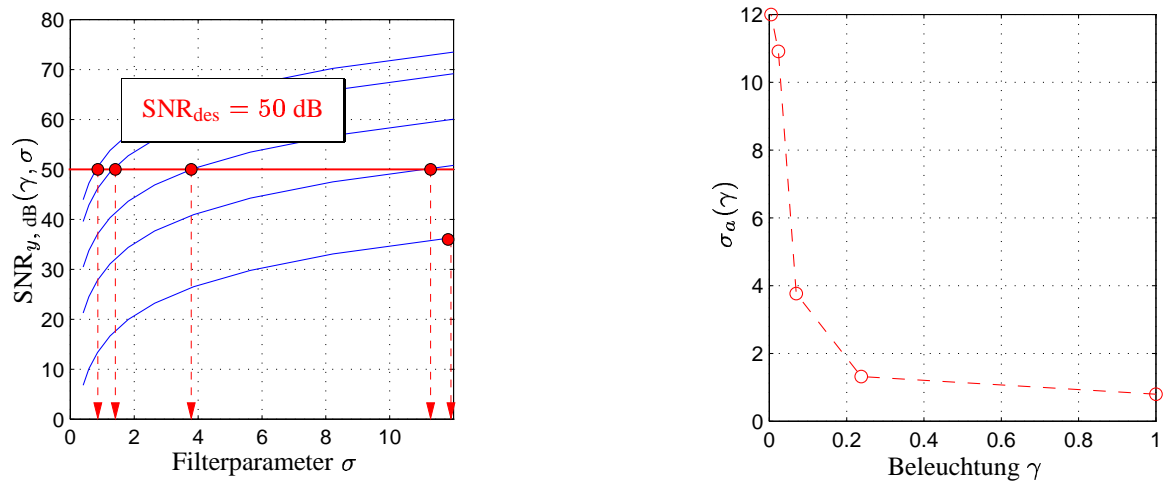


Abbildung 5.5: links: Bestimmung der notwendigen Filterstärke  $\sigma$  bei gegebener Beleuchtung und für ein gewünschtes SNR. rechts: Resultierendes Steuersignal.

Für diese Aufgabe wurde das Verhalten  $\sigma_a(\gamma)$  für  $\gamma = 0 \dots 1$  mit verschiedenen Testbildern berechnet und das adaptive  $\sigma_a$  durch einen Optimierungsprozeß bestimmt (siehe auch [Eis01]). Da die Ergebnisse für verschiedene Testbilder nur geringfügig voneinander abweichen, ist eine Approximation der Meßdaten durch eine Funktionsvorschrift

$$\sigma_a(\gamma) = (\sigma_{a,0} - \Sigma) \left( \frac{\gamma_0}{\gamma} \right)^\lambda + C \quad (5.28)$$

möglich und zweckmäßig, um jedem beliebigen  $\gamma$  einen Filterparameter zuzuordnen zu können. Mit der Konstanten

$$\Sigma = \frac{\sigma_{a,1} - \sigma_{a,0} \left( \frac{\gamma_0}{\gamma_1} \right)^\lambda}{1 - \left( \frac{\gamma_0}{\gamma_1} \right)^\lambda} \quad (5.29)$$

und der Vorgabe zweier Eckwerte  $\sigma_{a,i} = \sigma_a(\gamma_i)$ ,  $i = 1, 2$  ist es möglich die Abbildungsvorschrift auf zwei vordefinierte Werte zu „klemmen“. Zusätzlich kann ein gewichteter mittlerer Approximationsfehler während der Funktionsapproximation, der den Bereich kleiner Werte  $\sigma_a$  betont, die weiteren Ergebnisse verbessern. Das Ergebnis der Approximation ist in Abb. 5.6 (links) zu sehen. Auch wenn  $\text{SNR}(\sigma)$  immer eine positive Steigung hat, wie bereits diskutiert und in Abb. 5.4 gezeigt, ändert sich die Steigung mit unterschiedlichem Anteil von korreliertem und unkorreliertem Rauschen, also mit den Rauschparametern  $\alpha$  und  $\beta$ . Hier wird

also deutlich, daß immer eine Art Kamerakalibrierung zur Schätzung der Rauschparameter oder zur Gewinnung der Kurvenschar  $\text{SNR}(\sigma)$  im Vorfeld notwendig ist.

In Abb. 5.6 (rechts) ist eine Simulation einer adaptiven Filterung dargestellt. Der Mittelwert des Bildes wurde dabei benutzt, um die Beleuchtungsstärke  $\gamma$  zu schätzen. Der gut beleuchteten Szene wurde dabei  $\gamma = 1$  zugeordnet. Bei dieser Adaption wurde als kleinste Filterstärke  $\sigma = 1/\sqrt{2\pi}$  verwendet, wodurch auch das hellste Bild einer sehr geringen aber dennoch vorhandenen Filterung unterworfen ist. Aus diesem Grund ist das  $\text{SNR}_y$  etwas größer als das  $\text{SNR}_x$  bei  $\gamma = 1$ . Der Adaptionsmechanismus klemmt, wie zu sehen, das  $\text{SNR}_y$  gegen den optimalen Wert der hellen Szene. Die Abweichungen sind durch den Restfehler der Approximation in Abb. 5.6 (links) zu erklären.

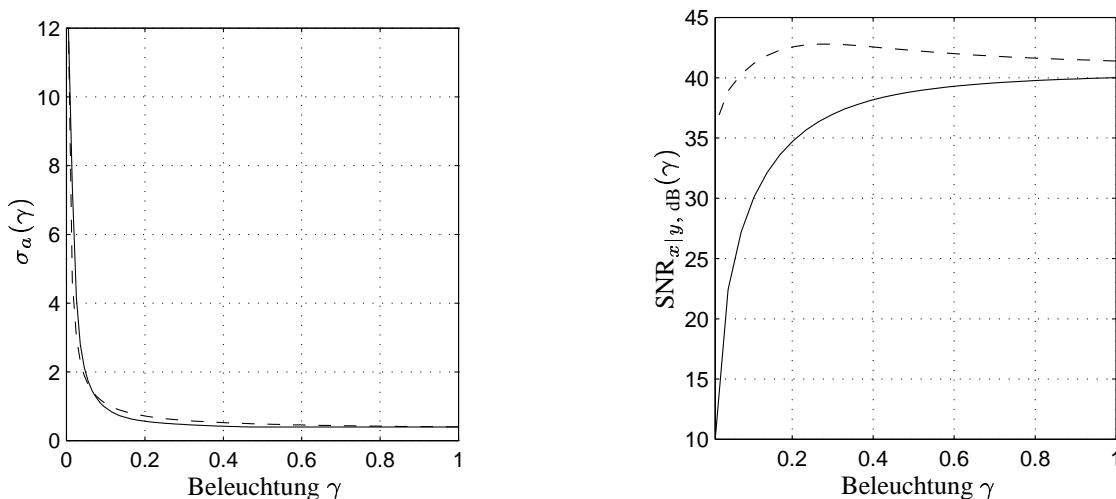


Abbildung 5.6: links: Sinnvolle Zuordnung der Beleuchtungsstärke (– gemessen, - - approximiert) durch Optimierung mittels simulierter Kameradaten. rechts: Verschlechterung (–) und adaptive Verbesserung (- -) bei abfallender Beleuchtung ( $\gamma = 1 \dots 0.01$ ) in einer simulierten Szene.

## Explizites Steuersignal

Unter der Voraussetzung, daß das Bildsignal der Szene konstant ist und die beiden Rauschprozesse weiß sind, kann man durch (5.25) ein explizites Steuersignal konstruieren. Unter Angabe eines zu erzielenden Signal-zu-Rauschverhältnisses

$\text{SNR}_{\text{des}} := \text{SNR}_y$  und der adaptiven Filterstärke  $\sigma_a$  erhält man die Beziehung

$$\text{SNR}_{\text{des}} = \text{SNR}_y(\gamma, \sigma_a) = \frac{\gamma^2 x_o^2}{\alpha^2 + \gamma^2 \beta^2 x_o^2} \cdot \frac{\sigma_a^2}{C\pi} = \text{SNR}_x(\gamma) \cdot \frac{\sigma_a^2}{C\pi} \quad (5.30)$$

und damit das explizite Steuersignal

$$\sigma_a^2(\gamma) := \frac{\text{SNR}_{\text{des}}}{\text{SNR}_x(\gamma)} \cdot C\pi. \quad (5.31)$$

Der Faktor  $C$ , der ursprünglich aus der FOURIER-Transformation resultierte, kann hierbei unter Umständen auch modifiziert werden, um die Näherungen (vor allem bei kleinen Filterstärken  $\sigma_a$ ), die zu dem Ergebnis in (5.25) führen, auszugleichen.

Mit diesem Ergebnis kann theoretisch ein beliebiges SNR eingestellt werden, wofür allerdings die Rauschparameter  $\alpha$  und  $\beta$  bekannt sein müssen, da sie bei der expliziten Schätzung der Größe  $\text{SNR}_x(\gamma)$  benötigt werden. Eine Simulation ist in Abb. 5.7 gezeigt. Sind die exakten Rauschparameter bekannt, so ergibt sich

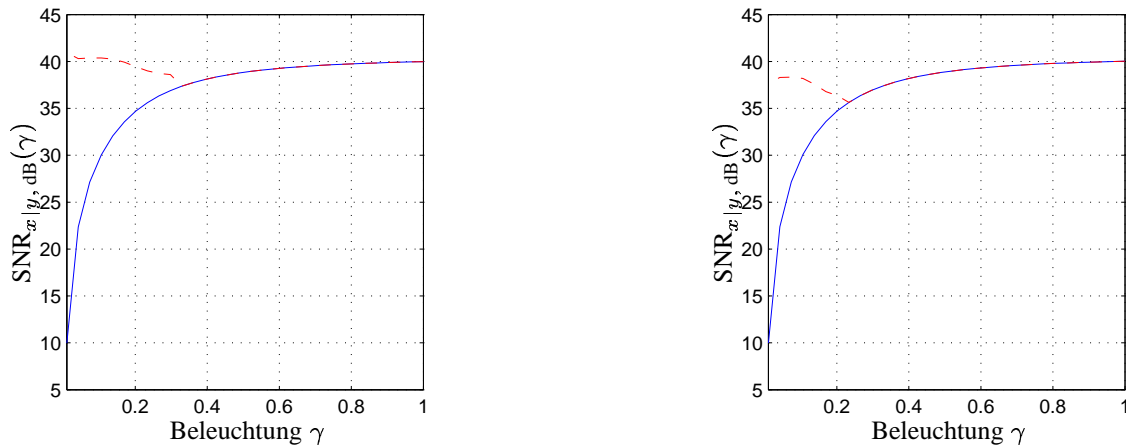


Abbildung 5.7: Verschlechterung (–) und adaptive Verbesserung (– –) bei abfallender Beleuchtung in einer simulierten Szene mit dem expliziten Steuersignal ( $\text{SNR}_{\text{des}} = 40$  dB). links: Ergebnis mit den exakten Rauschparametern. rechts: Einfluß unkorrekter Rauschparameter bei der Adaption.

das Ergebnis der Adaption in Abb. 5.7 (links). Der geschätzte adaptive Filterkoeffizient  $\sigma_a$  ist bis  $\gamma \approx 0.3$  so klein, daß keine Filterung stattfindet. Dies ist hauptsächlich darin begründet, daß der theoretische Ansatz für das explizite Steuersignal für kleine  $\sigma$  nur näherungsweise erfüllt ist. Für wachsende  $\sigma$  wird das gewünschte SNR dann schnell erzielt. In Abb. 5.7 (rechts) sind die Auswirkungen zu sehen, die sich durch unbekannte oder schlecht geschätzte Rauschparameter ergeben. Hier weichen die Parameter  $\alpha$  und  $\beta$  der Adaption um etwa 25% bzw. 50%

von den originalen Parametern ab, die zur Erzeugung des verrauschten Bildsignals verwendet wurden. Aus diesem Grund wird das gewünschte SNR nicht erreicht. Zur Schätzung der Rauschparameter ist also eine Art Kamerakalibrierung erforderlich, die im nächsten Kapitel genauer beschrieben wird. Dort ist auch gezeigt, daß trotz der genannten Näherungen und Annahmen durchaus auf sehr einfache Weise eine effektive Adaption erzielt werden kann.

### 5.3.3 Vergleich mit dem biologischen Vorbild

An dieser Stelle lohnt es, die Aufmerksamkeit auf das in der Biologie gemessene Steuersignal zur Generierung der adaptiven (lokalen) Rezeptorkopplung zu richten. Wie Untersuchungen in [SL96a] zeigen, kann die Abhängigkeit der Rezeptorkopplung von der *mittleren lokalen Helligkeit*  $I_a$  (gemessen in td=photopic trolands) hier durch die Gleichung

$$\sigma_R(I_a) = 1.5 \frac{I_a^n + A_0^n}{I_a^n + \delta^n}, \quad n \approx 0.5, \delta \approx 0.01, A_0 \approx 65 \text{ td} \quad (5.32)$$

näherungsweise approximiert werden, was in Abb. 5.8 (links) dargestellt ist.

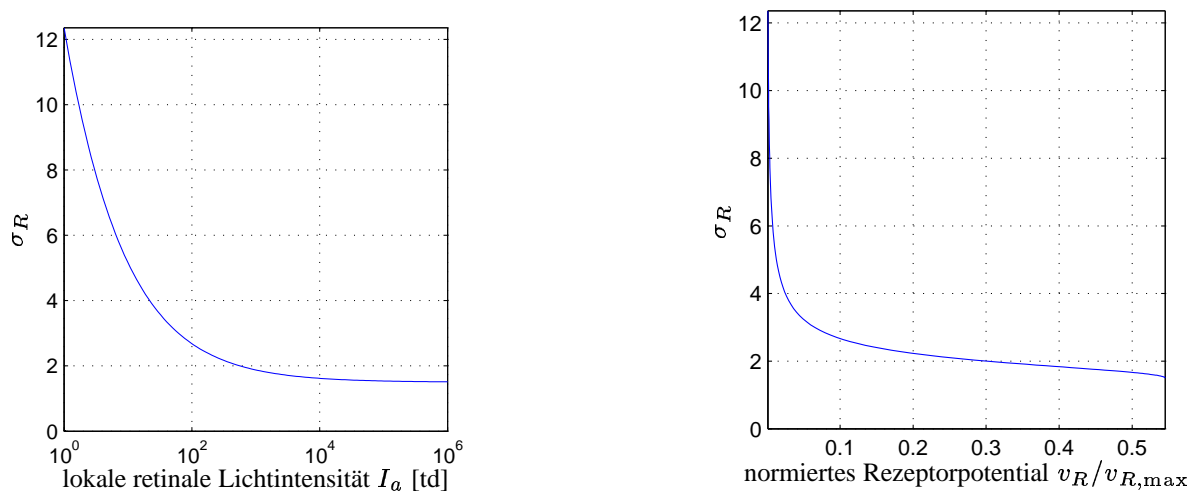


Abbildung 5.8: Adaptive Rezeptorkopplung des biologischen Vorbildes. links: Abhängigkeit von der lokalen mittleren Helligkeit. rechts: Abhängigkeit von dem (normierten) Rezeptorpotential (für ein gewähltes  $I_a$ ).

In einem weiteren Schritt muß allerdings die nichtlineare Umsetzung der Licht- bzw. Strahlungsintensität  $I$ , die auf einen Rezeptor fällt, in das resultierende Re-

zeptorpotential

$$v_R(I, I_a) \approx v_{R,\max} \frac{I}{I + (k_r/k_b I_a + k_r)} \quad , k_r \approx 833 \text{ td}, k_b \approx 10^3 \text{ td} \quad (5.33)$$

berücksichtigt werden, da erst dieses Potential  $v_R$  dem in dieser Arbeit verwendeten Rezeptorsignal  $x_R$  entspricht. Wird also die Rezeptorkopplung nicht über  $I_a$  sondern über  $v_R$  in Abb. 5.8 (rechts) aufgetragen, so verdeutlicht dies, daß der Mechanismus im biologischen Vorbild dem technischen Steuersignal (siehe Abb. 5.6 (links)) qualitativ sehr ähnlich ist. Damit scheint der in den vorherigen Abschnitten entwickelte Ansatz, in einem technischen System das SNR bei variierenden Beleuchtungsverhältnissen durch eine adaptive Filterung auf einen gewünschten Wert zu halten, weiter biologisch motiviert und sinnvoll zu sein.

## 5.4 Erweiterungen für eine adaptive Verarbeitung

In diesem Abschnitt wird nun die Frage behandelt, inwieweit das bereits vorgestellte lineare Modell aus Kapitel 3 um eine adaptive Komponente für variierende Beleuchtungsbedingungen erweitert werden kann. Dabei soll die Aufgabe, eine biologienahe Konturextraktion durchzuführen, stets im Blickpunkt bleiben. In Abb. 5.9 ist eine denkbare adaptive Verarbeitungskette dargestellt. Dabei wird das Rezeptorsignal  $x_R$  zurückgekoppelt, um ein Steuersignal zu erzeugen, was wiederum die Größe der Rezeptorkopplung  $\sigma_R$  adaptiv steuert. Im Gegensatz dazu wird in der biologischen Modellierung von SHAH [SL96a, SL96b] das Ausgangssignal der Horizontalzellen für die Adaption verwendet. Da dieses dort als Gaußfilter ausgelegt ist, wird so in Kombination mit der Rezeptorkopplung insgesamt wieder ein Gaußfilter größerer Filterstärke ausgebildet, um so eine mittlere lokale Helligkeit  $I_a$  zu schätzen. Dieses wird über eine nichtlineare Abbildung zur Berechnung der Rezeptorkopplung eingesetzt. Da im Rahmen dieser Arbeit allerdings die Übertragungseigenschaft der Horizontalzellen nicht direkt modelliert wurde, um dort keinen Freiheitsgrad zu verlieren, wird dieses Signal nicht verwendet. Vielmehr wird das Rezeptorsignal selbst mittels eines allgemeinen Verarbeitungsschrittes  $\sigma_R(x_R)$  in das Steuersignal überführt. Ist diese Verarbeitung als Gaußfilter ausgelegt, so entspricht dies aber wieder im wesentlichen der Funktionsweise des biologischen Vorbildes.

Der Mechanismus könnte wie in [SL96a, SL96b] ortsvariant sein und damit lokal adaptiv arbeiten. Dies kann Vorteile bei Szenen mit großem Dynamikbereich bieten, damit helle, aber auch dunkle Bereiche lokal unterschiedlich behandelt werden können. Wie aber bereits gezeigt, arbeitet bereits eine globale Adaption effektiv, wenn eine gewisse globale Helligkeit zugrunde gelegt wird. Diese kann

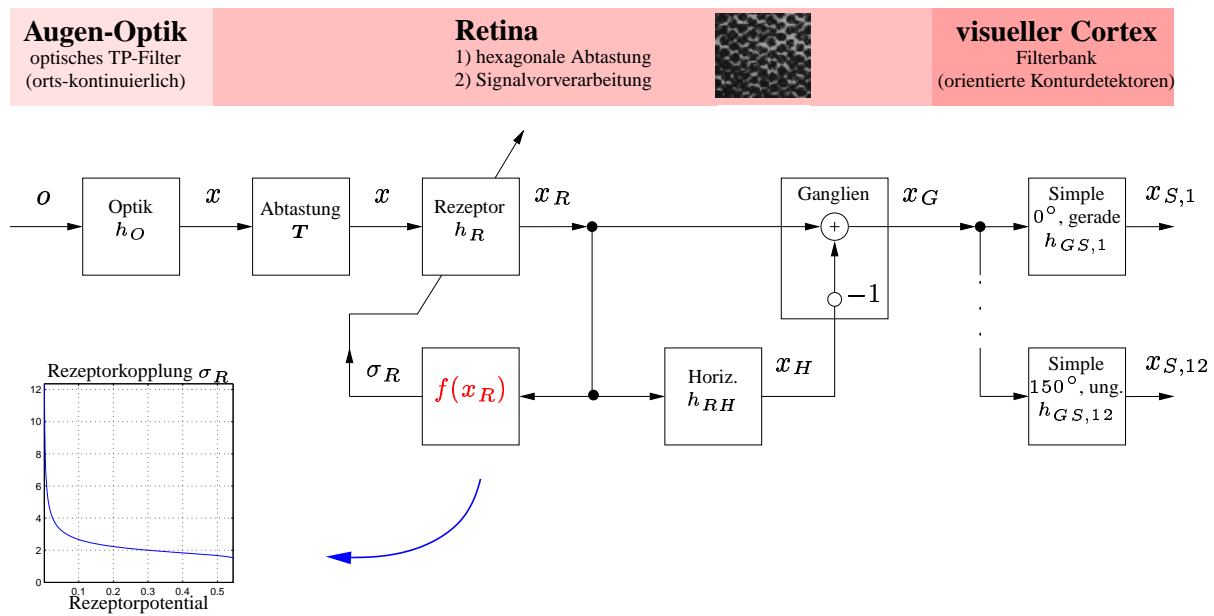


Abbildung 5.9: Erweiterung des bisherigen Modells um eine adaptive Rezeptorkopplung.

meßtechnisch durch den Mittelwert der kompletten Bildszene geschätzt werden, was einer Mittelung über die Gesamtheit aller Rezeptorpotentiale entspräche. Dies wäre aber biologisch nur sehr aufwendig implementierbar, da hierfür alle Rezeptorsignale auf ein resultierendes Neuron konvergieren müßten, was entweder sehr ausgedehnte synaptische Verbindungen bei einer einzelnen Verknüpfungsschicht oder aber eine sehr aufwendige Kaskade von vielen Neuronenschichten (die jeweils eine örtliche Tiefpaßwirkung besitzen) bedeuten würde. Dies wird ebenfalls ein Grund sein, warum in der Biologie ein lokaler Mechanismus zum Einsatz kommt. Da in technischen Bildverarbeitungssystemen aber im Gegensatz zu der retinalen Architektur eine globale mittlere Helligkeit  $\gamma$  der Szene leicht berechnet werden kann, beschränken wir uns hier auf einen globalen Steuermechanismus. Wie wir im weiteren sehen werden, stellt aber selbst diese ortsinvariante Adaption für die gesamte bisherige Verarbeitungskette eine zusätzliche Schwierigkeit im Entwurf dar.

### 5.4.1 Einfluß auf nachgeschaltete Verarbeitungsschichten

Ziel ist es, eine biologienahe Signalverarbeitungskette aufzustellen, die adaptiv zur Helligkeit der Szene eine robuste Konturextraktion gewährleistet. Dazu betrachten wir die Verarbeitung des bisherigen linearen Modells in Kapitel 3 und simulieren die Übertragungseigenschaften für eine adaptive Rezeptorkopplung  $\sigma_R$ . Da die nachgeschalteten Systeme  $h_{RH}$  und  $h_{GS,i}$  für ein spezielles  $\sigma_R =$

1.5 csp, also für die beste Ortsauflösung entworfen worden sind, ist es nicht unbedingt selbstverständlich, daß die gesamte Verarbeitungskette noch die gewünschten Übertragungseigenschaften für ein adaptives  $\sigma_R$  besitzt.

Untersuchungen zeigen, daß das Modell bis zu den Gangliensignalen die gewünschten Eigenschaften auch für eine adaptive Rezeptorkopplung beibehält. Dieses liegt darin begründet, daß die Filtereigenschaft  $h_G$  als Laplace-Ableitung einer Gaußfunktion (LoG) entworfen ist, das Subsystem  $h_{RG}$  sich damit näherungsweise (je nach Modellaufwand) als Laplace-Operator  $\Delta_t = \partial^2/\partial t_1^2 + \partial^2/\partial t_2^2$  ausgebildet hat

$$h_{RG}(\mathbf{k}) \circ \bullet H_{RG}(\mathbf{v}) \approx \frac{\|\mathbf{T}^{-T}\mathbf{v}\|^2}{|\det \mathbf{T}|} = \frac{\|\boldsymbol{\omega}\|^2}{|\det \mathbf{T}|}. \quad (5.34)$$

In Abb. 5.10 ist die (normierte) Übertragungsfunktion  $H_{RG}$  auf der Basis eines hexagonalen Abtastrasters dargestellt. Wird der Parameter des Gaußfilters nun adaptiv verändert, so bleibt insgesamt immer ein LoG-Operator bestehen.

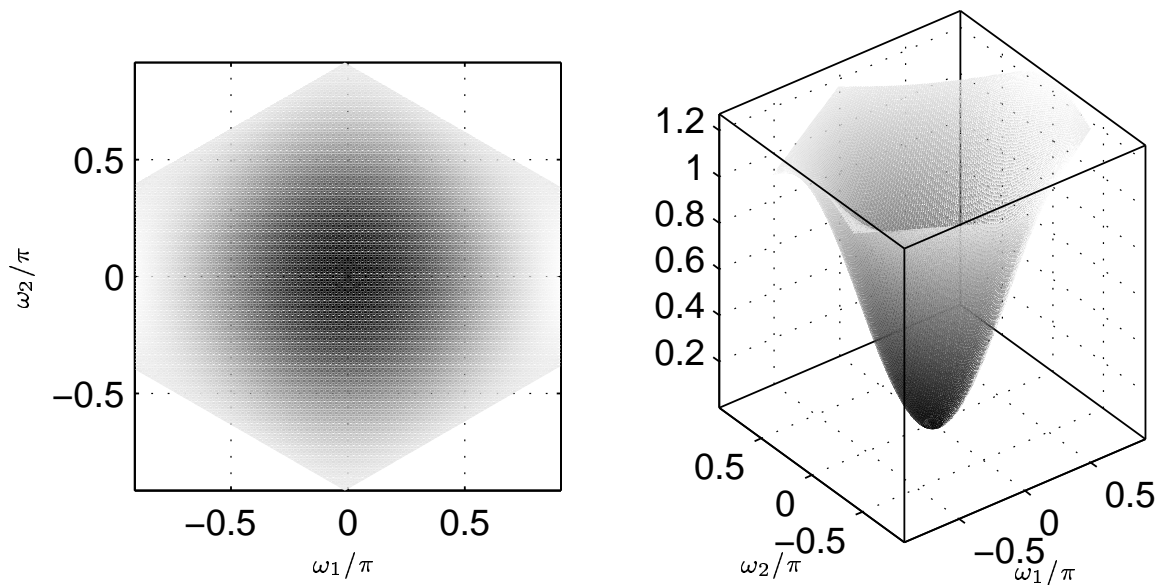


Abbildung 5.10: Die Übertragungsfunktion  $H_{RG}$  entspricht näherungsweise dem Verhalten eines Laplace-Operators.

Im Gegensatz dazu weichen die Resultate der Simplen Zellen aber für ein  $\sigma_R \neq 1.5$  csp von dem gewünschten „Gaborverhalten“ ab. In Abb. 5.11 ist dazu der Betrag der Übertragungsfunktion  $H_{S,1}$  für verschiedene Rezeptorkopplungen zu sehen. Da die kortikalen synaptischen Verknüpfungen bzw. die Filterkoeffizienten

der kortikalen Subsysteme für ein  $\sigma_R = 1.5$  csp ausgelegt und optimiert sind, führen diese bei größeren Rezeptorkopplungen nicht mehr zu dem gewünschten Verhalten.

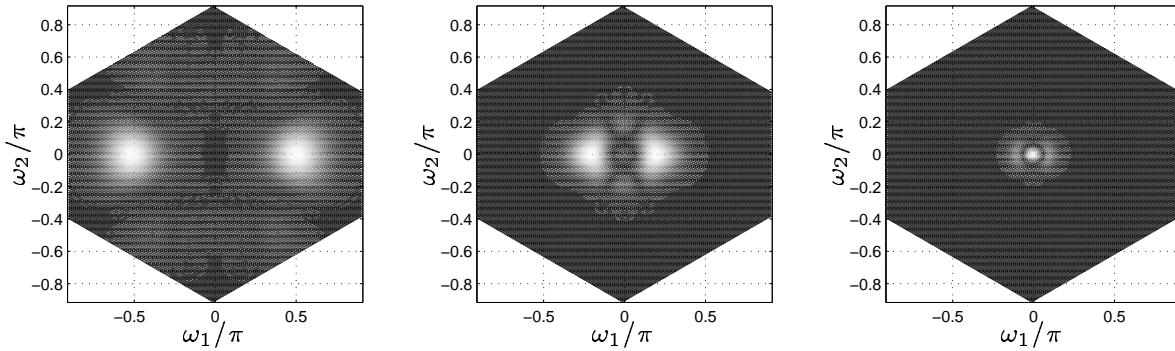


Abbildung 5.11: Übertragungsverhalten  $H_{S,1}$  der ursprünglichen Verarbeitungskette bei variablen Rezeptorkopplungen  $\sigma_R = 1.5$  csp (links),  $\sigma_R = 3$  csp (mitte) und  $\sigma_R = 6$  csp (rechts).

Natürlich könnte versucht werden, auf eine gewisse Art die Subsysteme  $h_{GS,i}$  ebenfalls adaptiv auszulegen, um stets das gewünschte Verhalten zu erreichen. Dies ist aber biologisch nicht erklärbar, da dies eine zusätzliche enorme Anzahl von synaptischen Verbindungen innerhalb des Cortex bedeuten würde, um die jeweiligen Verknüpfungen beider Typen Simpler Zellen (gerade/ungerade) für alle Orientierungen adaptiv auszulegen. Für das biologische Vorbild wäre deshalb nur ein Verarbeitungssystem sinnvoll und effizient, bei dem nur an einer Stelle, nämlich zu Beginn der Verarbeitungskette, sprich den Rezeptoren ein adaptiver Mechanismus zu implementieren ist. Auch für ein technisches System wäre nur solch eine Strategie interessant und überhaupt sinnvoll, da vom Aufwand her vertretbar.

## 5.4.2 Modellierung kortikaler Verknüpfungsstrukturen

Um die Problematik mit dem bisherigen Modell, die durch eine adaptive Verarbeitung entsteht, dennoch lösen zu können, wird im folgenden eine alternative kortikale Modellierung vorgestellt. Diese besteht aus einer Zusammenschaltung von Subsystemen, welche die örtlichen Gradienten des Intensitätssignals bilden. Damit wird die retinale Verarbeitung, die im wesentlichen eine Ableitung zweiter Ordnung bildet, im Cortex systematisch weitergeführt. Würde sich dann durch eine Adaption die Standardabweichung der Gaußfunktion ändern, sollte das prinzipielle Verhalten einer Gaborfunktion aber bestehen bleiben.

Um die Komplexität der Modellierung zu reduzieren, betrachten wir zunächst den



eindimensionalen Fall. Hier kann leicht veranschaulicht werden, daß die höheren Ableitungen einer Gaußfunktion recht gut der Kurvenform der Gaborfunktionen ähneln, so daß es zu untersuchen lohnt, die kortikale Modellierung als eine Kette von örtlichen Ableitungsfiltern aufzubauen. In Abb. 5.12 (links) sind exemplarisch die geraden und ungeraden Ableitungen einer Gaußfunktion bis zur Ordnung 4 bzw. 5 dargestellt. Die höheren Ableitungen scheinen dabei qualitativ einer Gaborfunktion ähnlich zu sein. Vergleicht man in Abb. 5.12 (rechts/oben) die 6. und 5. Ableitung mit dem Real- und Imaginärteil einer bestimmten Gaborfunktion, so wird dies nochmals deutlich. Werden nun aber jeweils die gerade und ungerade Ableitung kombiniert, wie in [KvD87] vorgeschlagen, so ergeben sich die Darstellungen in Abb. 5.12 (rechts/unten). Im Gegensatz zu dem Betrag der komplexen Gaborfunktion, welcher durch die Quadratureigenschaft von Real- und Imaginärteil eine unmodulierte Gaußfunktion ist, ähnelt die gezeigte Kombination der beiden Ableitungen nur in entfernter Weise einer Gaußfunktion. Da gerade diese Eigenschaft aber für eine leistungsfähige Konturextraktion von großer Bedeutung ist, soll an dieser Stelle ein neuer Ansatz zur Approximation der Gaborfunktionen mittels linear kombinierter Gaußableitungen vorgestellt werden.

### Approximationsansatz für den eindimensionalen Fall

Die Untersuchungen im eindimensionalen Fall, für die o. E.  $T = 1$  sei, zeigen, daß durch eine Linearkombination

$$\begin{aligned} h_{g,e}(k) &\approx \tilde{h}_{g,e}(k) = \left( \sum_{\nu} f_{2\nu} \partial_{2\nu} \right) \cdot g(k), \\ h_{g,o}(k) &\approx \tilde{h}_{g,o}(k) = \left( \sum_{\nu} f_{2\nu+1} \partial_{2\nu+1} \right) \cdot g(k), \end{aligned} \quad (5.35)$$

verschiedener Ableitungen einer Gaußfunktion die Approximationsgenauigkeit deutlich verbessert werden kann. Dabei seien  $\partial_i = \partial^i / \partial t^i$  die örtlichen Ableitungen der Ordnung  $i$  bzw. deren diskrete Realisierungen,

$$g(k) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{k^2}{2\sigma^2}} \quad \circ \longrightarrow \bullet \quad G(v) \approx e^{-\frac{v^2\sigma^2}{2}} \quad (5.36)$$

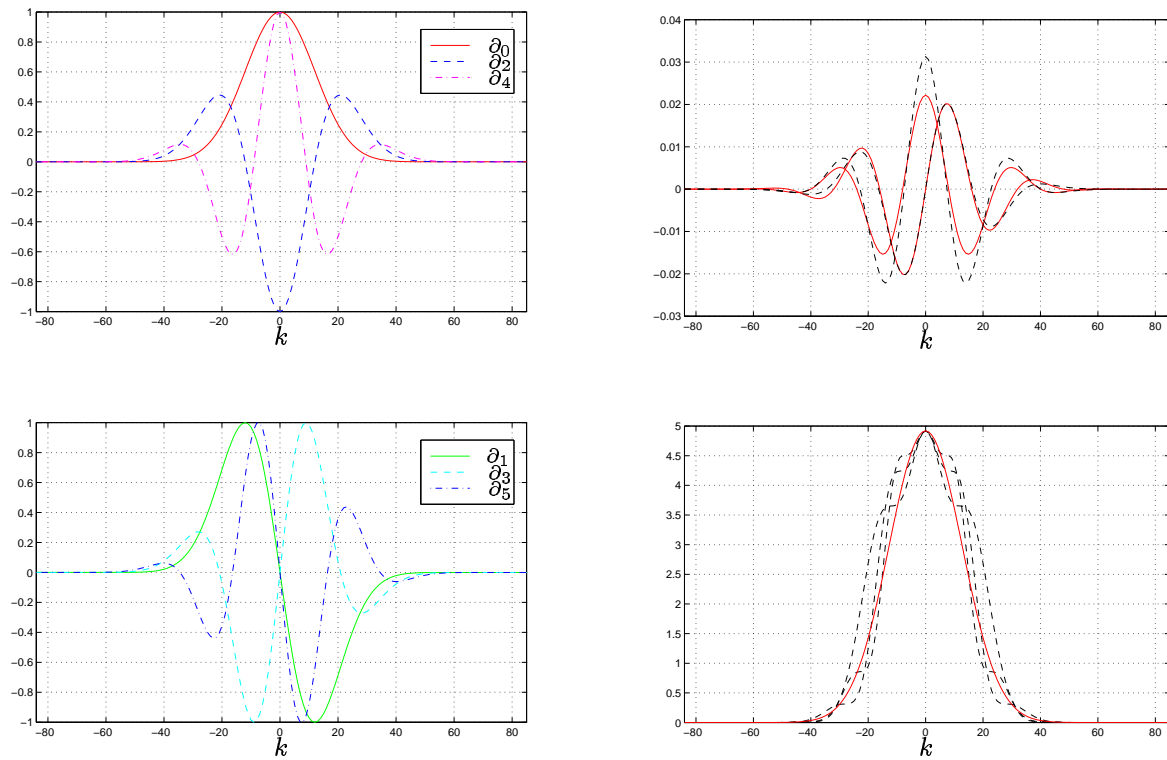


Abbildung 5.12: links: Gerade (oben) und ungerade (unten) örtliche Ableitungen einer Gaußfunktion. rechts: Vergleich der 6. und 5. Ableitung einer Gaußfunktion (- -) mit einer geraden und ungeraden Gaborfunktion (-) (oben) und Kombination der Gaborfunktionen (-) bzw. der beiden Ableitungen nach [KvD87] (- -) (unten).

die Gaußfunktion und

$$h_{g,e}(k) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \cdot e^{-\frac{k^2}{2\sigma_g^2}} \cdot \cos(v_0 k), \quad (5.37)$$

$$h_{g,o}(k) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \cdot e^{-\frac{k^2}{2\sigma_g^2}} \cdot \sin(v_0 k) \quad (5.38)$$

die zu approximierende gerade (even) bzw. ungerade (odd) Gaborfunktion. Die Koeffizienten  $f_i$  lassen sich durch ein Regressionsverfahren numerisch bestimmen. Es hat sich gezeigt, daß bereits drei Ableitungsfunktionen zu einer ausrei-

chend genauen Approximation führen. Wir machen deshalb den Ansatz

$$\begin{aligned}\tilde{h}_{g,e}(k) &= \left( \sum_{i=2,4,6} f_i \partial_i \right) \cdot g(k) = (f_2 \partial_2 + f_4 \partial_4 + f_6 \partial_6) \cdot g(k), \\ \tilde{h}_{g,o}(k) &= \left( \sum_{i=3,5,7} f_i \partial_i \right) \cdot g(k) = (f_3 \partial_3 + f_5 \partial_5 + f_7 \partial_7) \cdot g(k).\end{aligned}\quad (5.39)$$

Hierbei soll eine Ableitung  $< 2$  nicht verwendet werden, da im Cortex lediglich das Signal der Ganglienzellen zur Verfügung steht (laplacian of gaussian), auf das die Kaskade aufgebaut wird.

Die kaskadenförmige Verarbeitungskette ist in Abb. 5.13 durch einen denkbaren Signalflußgraph beschrieben. Neben dieser Implementierung gibt es weitere Varianten, die aber weniger effizient oder für eine Erweiterung auf den zweidimensionalen Fall weniger geeignet sind. Hier wird aber auch deutlich, daß durch die Linearkombination kein Mehraufwand in Kauf genommen werden muß, da durch eine Kaskade die Ableitungen niedriger Ordnung mitberechnet werden.

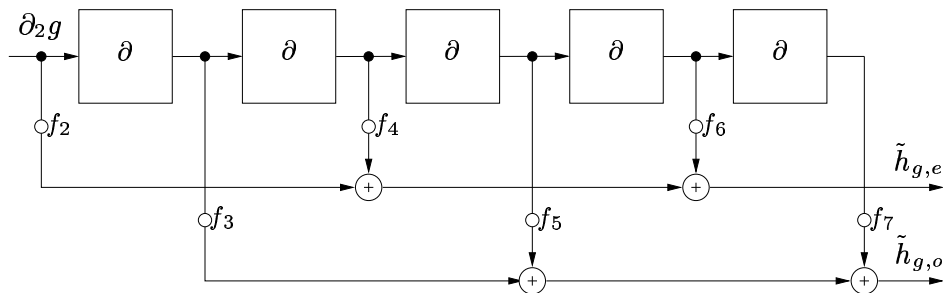


Abbildung 5.13: Signalflußgraph der eindimensionalen Verarbeitungskette.

Das Ergebnis der Approximation spiegelt Abb. 5.14 wider. Sowohl die beiden reellen Gaborfunktionen, als auch deren Kombination weisen eine hohe Approximationsgüte auf. Hierbei wurde als Entwurfsziel eine Gaborfilterimpulsantwort mit einer Mittenfrequenz von  $\pi/8$  (3. Band einer logarithmischen Frequenzaufteilung, siehe Abb. 5.15) und eine Standardabweichung der Gaußfunktion von  $\sigma = 6$  gewählt.

Verändert sich nun die Standardabweichung im Rahmen einer Adaption, so ergibt sich in einem weiten Rahmen immer eine Gaborfunktion als Resultat der Überlagerung der Gaußableitungen. Abb. 5.15 gibt diesen Aspekt im Frequenzbereich wieder, wobei die Standardabweichung hier in 4 Schritten von 0.4 bis 12 variiert.

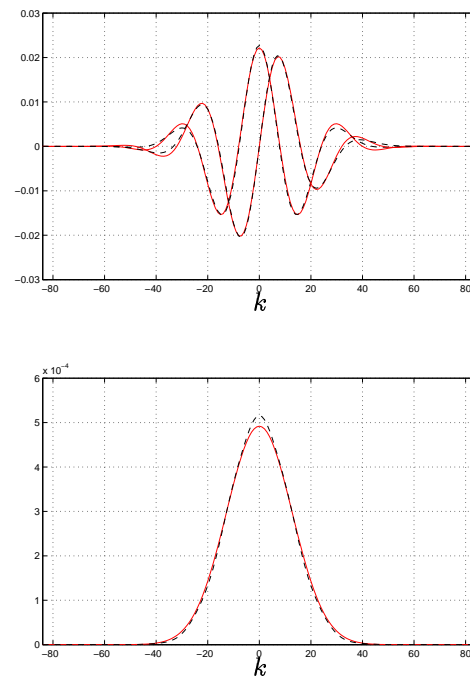


Abbildung 5.14: oben: Gerade und ungerade Gaborfunktion (—) im Vergleich zu dem vorgestellten Approximationsansatz (---) gemäß (5.39). unten: Kombination der Gaborfunktionen (—) und der Approximationsergebnisse (---).

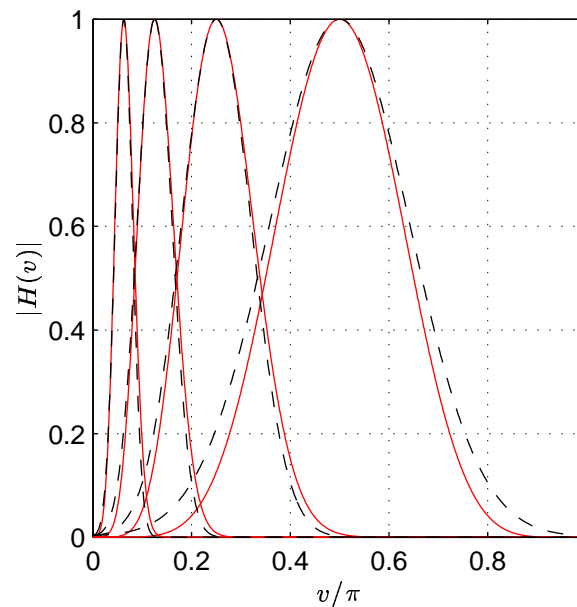


Abbildung 5.15: Vergleich der Gaborfunktionen (—) mit dem Ergebnis der Approximation (---) im Frequenzbereich für verschiedene Mittenfrequenzen.

Bei der biologischen Modellierung würde sich  $\sigma$  (bzw. die Rezeptorkopplung  $\sigma_R$ ) adaptiv einstellen. Für eine technische Implementierung kann es aber auch nützlich sein zu wissen, wie  $\sigma$  einzustellen ist, um eine gewisse gewünschte Mittenfrequenz  $v_0$  des Gaborfilters zu realisieren. Dazu ist das Verhalten der Ableitungskaskade im Frequenzbereich zu betrachten, da dort das Maximum bei der gesuchten Mittenfrequenz liegt. In der vorliegenden Modellierung kommt als AbleitungsfILTER ein zentrischer Gradient mit der Differenzengleichung

$$y(k) = \frac{x(k+1) - x(k-1)}{2T} \quad (5.40)$$

zum Einsatz, gleichwohl auch Alternativen dazu, wie ein rekursives nichtkausales Filter, untersucht wurden. Da neben der einfachen Implementierung der zentrische Gradient gegenüber Rauschprozessen durch die Dämpfung bei hohen Frequenzen zusätzlich robust ist, soll im weiteren diese Form verwendet werden. Die z-Übertragungsfunktion bzw. die gewöhnliche Übertragungsfunktion des zentrischen Gradienten läßt sich angeben zu

$$H_D^{[z]}(z) = \frac{z - z^{-1}}{2T} \quad \text{bzw.} \quad H_D(v) = j/T \sin(v). \quad (5.41)$$

Weiter erlaubt dies eine Beschreibung der verschiedenen (geraden und ungeraden) Ableitungsoperationen höherer Ordnung im Frequenzbereich mit

$$\partial_i \circ \bullet (\mathcal{F}\{\partial_1\})^i \approx (H_D(v))^i = (j/T)^i \sin^i(v). \quad (5.42)$$

Mit diesen Realisierungen ergeben sich die Approximationen der Gaborfunktionen im Frequenzbereich zu

$$\tilde{H}_{g,e}(v) = \left( \sum_{i=2,4,6} f_i (j/T)^i \sin^i(v) \right) \cdot G(v), \quad (5.43)$$

$$\tilde{H}_{g,o}(v) = \left( \sum_{i=3,5,7} f_i (j/T)^i \sin^i(v) \right) \cdot G(v). \quad (5.44)$$

Da die Übertragungsfunktionen am Rand des interessierenden Frequenzbereiches verschwinden

$$\tilde{H}_{g,e}(0) = \tilde{H}_{g,e}(\pi) = 0, \quad (5.45)$$

$$\tilde{H}_{g,o}(0) = \tilde{H}_{g,o}(\pi) = 0, \quad (5.46)$$

führt die notwendige Bedingung für eine Extremstelle

$$\begin{aligned} \frac{\partial}{\partial v} \tilde{H}_{g,e|o}(v) &\stackrel{!}{=} 0 \\ &= \left( \sum_i f_i(j/T)^i i \sin^{i-1}(v) \right) \cos(v) \cdot G(v) + \\ &\quad \left( \sum_i f_i(j/T)^i \sin^i(v) \right) (-v\sigma^2) \cdot G(v) \end{aligned} \quad (5.47)$$

auf eine allgemeine Berechnungsvorschrift für den gesuchten Filterparameter  $\sigma$  bei gegebener oder gewünschter Mittenfrequenz  $\tilde{v}_0$

$$\sigma = \sqrt{\frac{\sum f_i(j/T)^i i \sin^{i-1}(\tilde{v}_0) \cos(\tilde{v}_0)}{\sum f_i(j/T)^i \sin^i(\tilde{v}_0)} \frac{1}{\tilde{v}_0}}. \quad (5.48)$$

Der reduzierte Approximationsansatz für die gerade Gaborfunktion

$$\tilde{H}_{g,e}(v) = (-4f_2 \sin^2(v) + 16f_4 \sin^4(v) - 64f_6 \sin^6(v)) \cdot G(v) \quad (5.49)$$

führt damit nach entsprechenden Vereinfachungen auf

$$\sigma = \sqrt{\frac{-2f_2 + 16f_4 \sin^2(\tilde{v}_0) - 96f_6 \sin^4(\tilde{v}_0)}{-f_2 \sin(\tilde{v}_0) + 4f_4 \sin^3(\tilde{v}_0) - 16f_6 \sin^5(\tilde{v}_0)} \frac{\cos(\tilde{v}_0)}{\tilde{v}_0}}, \quad (5.50)$$

wobei hier o. E.  $T = 1$  (genormt auf Pixelkoordinaten) gesetzt sei. Die umgekehrte Fragestellung nach der resultierenden Mittenfrequenz  $\tilde{v}_0$  bei gegebenem Filterparameter  $\sigma$  kann durch die implizite Gleichung

$$\tilde{v}_0 = \frac{-2f_2 + 16f_4 \sin^2(\tilde{v}_0) - 96f_6 \sin^4(\tilde{v}_0)}{-f_2 \sin(\tilde{v}_0) + 4f_4 \sin^3(\tilde{v}_0) - 16f_6 \sin^5(\tilde{v}_0)} \frac{\cos(\tilde{v}_0)}{\sigma^2} \quad (5.51)$$

numerisch recht schnell gelöst werden. Die Zuordnung für eine logarithmische Frequenzbandaufteilung, wie es bei Filterbankrealisierungen i. d. R. üblich ist, ergibt bei der Wahl von zentrischen Gradienten die folgende Übersicht:

$\tilde{v}_0$	$\pi/2$	$\pi/4$	$\pi/8$	$\pi/16$
$\sigma$	0.3989	2.7584	6.0321	12.0777

(5.52)

### Approximationsansatz für den zweidimensionalen Fall

Was im eindimensionalen Fall die Ableitung bzgl. der Ortskoordinate  $t$  ist, muß im zweidimensionalen Fall durch die Ableitung in eine bestimmte (gewünschte) Richtung realisiert werden, um die Orientierungsselektivität der zweidimensionalen Gaborfunktionen zu approximieren. Die Ableitung eines kontinuierlichen Signals  $x_c(\mathbf{t})$  in eine bestimmte Richtung kann mittels eines Richtungsvektors

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix} \quad (5.53)$$

und des Gradienten  $\nabla$  des Signals durch

$$\partial^{(r)} x_c(\mathbf{t}) = \nabla^T x_c(\mathbf{t}) \cdot \mathbf{r} = [\partial^{(1)} x_c(\mathbf{t}), \partial^{(2)} x_c(\mathbf{t})] \cdot \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \quad (5.54)$$

bestimmt werden, was kurz Richtungsableitung genannt wird. Bei einem ortsdiskreten Signal

$$x(\mathbf{k}) = x_c(\mathbf{T}\mathbf{k}) \quad (5.55)$$

auf einem beliebigen Abtastraster  $\mathbf{T}$  müssen nun die Ableitungen bzgl. der kartesischen Koordinatenrichtungen  $\partial^{(i)}$  ( $i = 1, 2$ ) durch die entsprechenden diskreten Realisierungen auf den Koordinatenachsen des Abtastrasters ausgedrückt werden. Es ergibt sich durch Substitution die Beziehung

$$\partial^{(i)} x_c(\mathbf{t}) = \frac{\partial}{\partial t_i} x_c(\mathbf{T}\mathbf{k}) = \frac{\partial}{\partial \mathbf{k}} x(\mathbf{k}) \cdot \frac{\partial}{\partial t_i} \mathbf{k} = \nabla_{\mathbf{k}}^T x(\mathbf{k}) \cdot \underbrace{\frac{\partial}{\partial t_i} (\mathbf{T}^{-1}\mathbf{t})}_{=:\begin{bmatrix} \tau_{i1} \\ \tau_{i2} \end{bmatrix}}, \quad (5.56)$$

die die Berechnung der gesuchten Ableitung durch den Gradienten  $\nabla_{\mathbf{k}}$  auf dem Abtastraster und der ersten bzw. zweiten Spalte von  $\mathbf{T}^{-1}$  angibt. Zusammengefaßt gilt somit

$$\partial^{(r)} x(\mathbf{k}) = \nabla_{\mathbf{k}}^T x(\mathbf{k}) \cdot \mathbf{T}^{-1} \cdot \mathbf{r} \quad (5.57)$$

$$= [\partial^{(1)} x(\mathbf{k}), \partial^{(2)} x(\mathbf{k})] \cdot \begin{bmatrix} \tau_{11} & \tau_{21} \\ \tau_{12} & \tau_{22} \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}. \quad (5.58)$$

Für  $T = 1$  und einem rechtwinkligen Abtastraster ergibt sich eine starke Vereinfachung der Richtungsableitung, was in Abb. 5.16 zu sehen ist. Auf einem hexagonalen Abtastraster sind allerdings die allgemeine Form und die entsprechenden

Koeffizienten, die aus

$$\mathbf{T}_h^{-1} = \begin{bmatrix} \frac{1}{2T_h} & \frac{1}{2\sqrt{3}T_h} \\ \frac{1}{2T_h} & \frac{-1}{2\sqrt{3}T_h} \end{bmatrix} \quad (5.59)$$

resultieren, zu verwenden, was in Abb. 5.17 veranschaulicht ist. Letztlich ist in Abb. 5.18 der Signalflußgraph der resultierenden zweidimensionalen orientierungsselektiven Verarbeitungskette dargestellt.

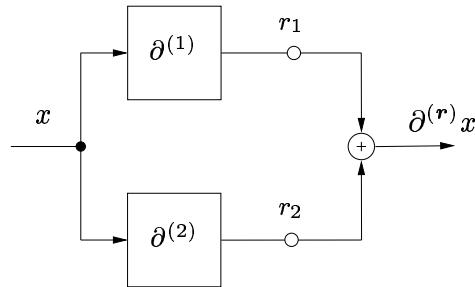


Abbildung 5.16: Vereinfachter Signalflußgraph der zweidimensionalen Richtungsableitung auf rechteckigem Raster mit  $T = 1$ .

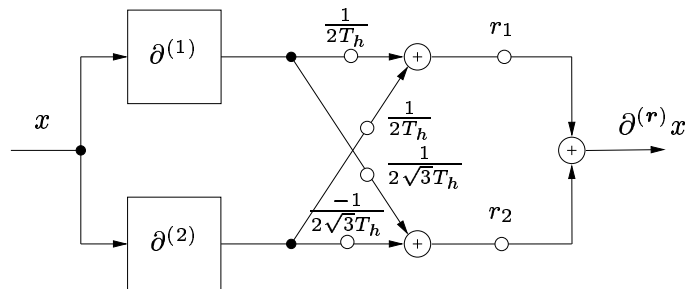


Abbildung 5.17: Signalflußgraph der zweidimensionalen Richtungsableitung auf hexagonalem Raster.

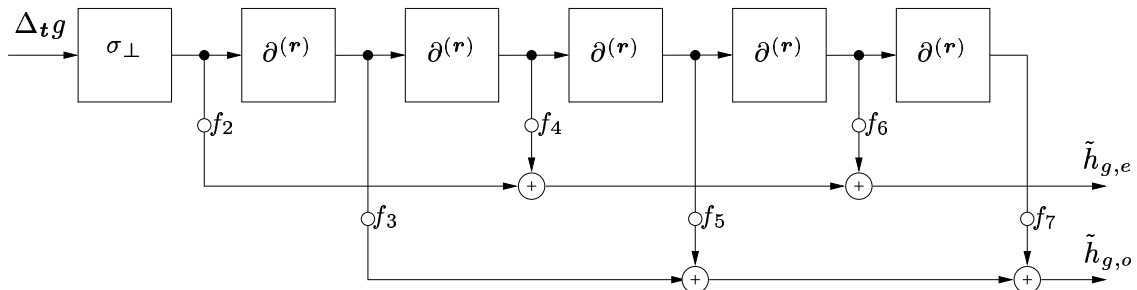


Abbildung 5.18: Signalflußgraph der zweidimensionalen Verarbeitungskette.

Bei der zweidimensionalen Filterung ist zusätzlich zu der Ableitungskette bzgl. der gewünschten Orientierung noch eine zu der Orientierung orthogonale Tief-



paßfilterung  $\sigma_{\perp}$  notwendig, um eine Orientierungsselektivität zu erreichen. Anschaulich wird dadurch eine kontrollierte Überschneidung zweier Frequenzbänder mit einer benachbarten Orientierung erreicht. So ist beispielsweise die Übertragungsfunktion bis zu der zweiten Richtungsableitung in der Kette für die Orientierung  $0^\circ$  (Ableitung in die erste Koordinatenrichtung) im wesentlichen eine Multiplikation einer isotropen Gaußfunktion mit  $\sin^2(v_1)$ . Dadurch ist aber die resultierende Übertragungsfunktion bzgl. der orthogonalen Frequenz (hier  $v_2$ ) für eine Orientierungsselektivität i. allg. zu breit. Die biologische Realisierung stellt eine zusätzliche Herausforderung dar, da durch das Gangliensignal als Eingangssignal der Ableitungskette immer eine Laplace-Ableitung eines gaußgefilterten Signals zur Verfügung steht. Dies bedeutet somit aber immer eine Fremdkomponente durch die 2. Ableitung bzgl. der orthogonalen Richtung. Durch die erwähnte orthogonale Tiefpaßfilterung, welche näherungsweise als Integrator, also als Umkehroperation zur Differentiation aufgefaßt werden kann, wird die störende Komponente aber weitgehend unterdrückt. In Abb. 5.19 sind exemplarisch die resultierenden Frequenzbänder für Orientierungen bis  $90^\circ$  und  $\tilde{v}_0 \in [\pi/2, \pi/4, \pi/8]$  der geraden und ungeraden Ableitungsketten dargestellt. Bei dem ersten Auflö-

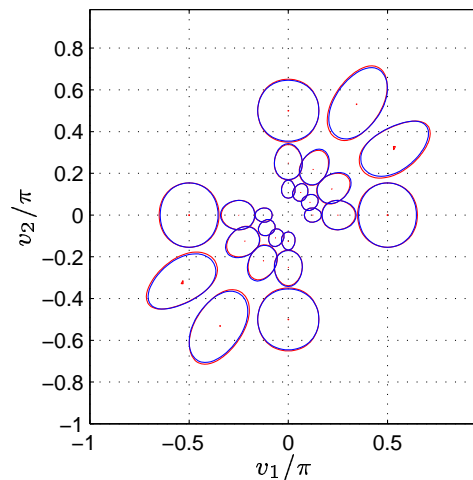


Abbildung 5.19: Verhalten einer frequenz- und orientierungsselektiven Filterbank im Frequenzbereich (Konturplot bei einer Dämpfung von 6 dB) auf der Basis der geraden und ungeraden Ableitungsketten.

sungskanal  $\tilde{v}_0 = \pi/2$  macht sich die Anisotropie des LoG der Ganglienzellen bemerkbar. Die Mittenfrequenzen liegen nicht auf einem Kreis und bei  $0^\circ$  sowie  $90^\circ$  müßte die orthogonale Filterung verstärkt werden. An dieser Stelle sollte aber daran erinnert werden, daß im biologischen Vorbild keineswegs geordnete Strukturen der Filterbank gemessen wurden. Deshalb ist eine Realisierung optimaler Frequenzbandaufteilung auch nicht unbedingt erstrebenswert. Hingegen ist mit

der vorliegenden Modellierung eine biologienahe Verarbeitungskette entwickelt worden, die den Bedürfnissen einer adaptiven Konturextraktion gerecht wird. In Abb. 5.20 ist noch einmal abschließend das Frequenzverhalten der ursprünglichen Modellierung und der unter Einsatz von Ableitungsoperatoren bei einer Adaption gegenübergestellt.

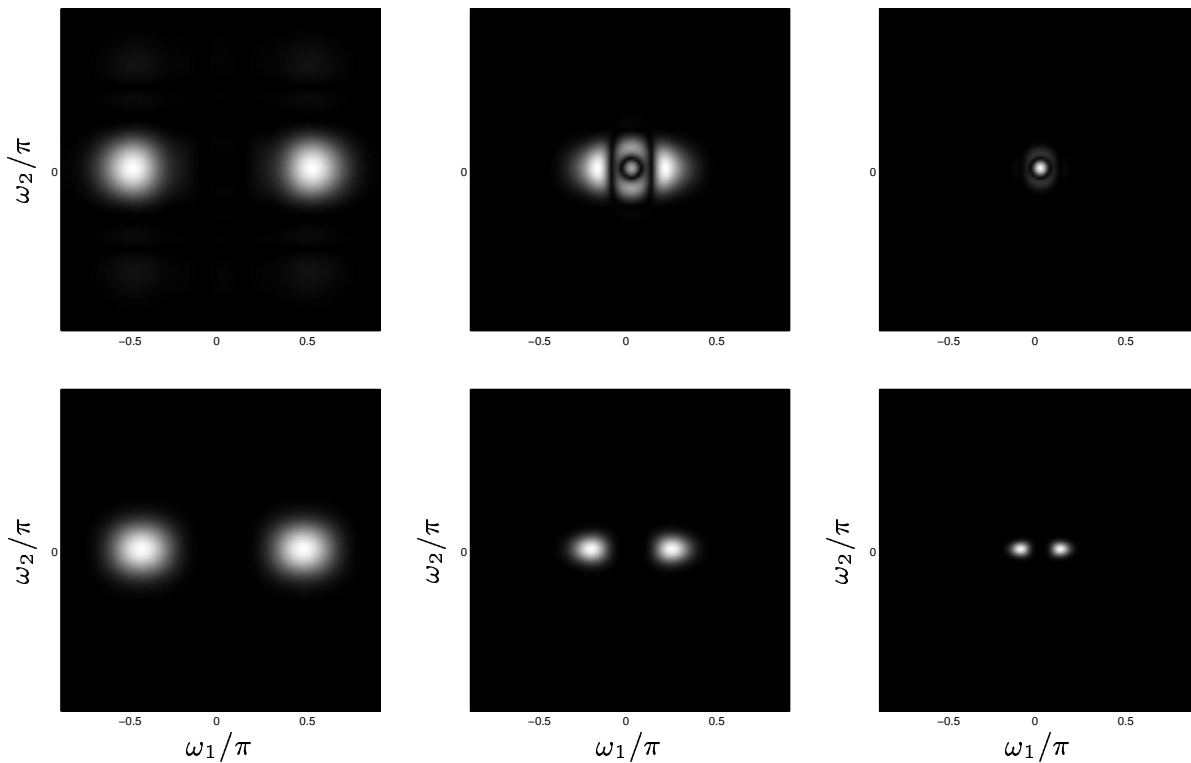


Abbildung 5.20: Übertragungsverhalten  $H_{S,1}$  der ursprünglichen Verarbeitungskette (oben) und der kortikalen Ableitungskette (unten) bei variablen Rezeptorkopplungen  $\sigma_R = 1.5$  csp (links),  $\sigma_R = 3$  csp (mitte) und  $\sigma_R = 6$  csp (rechts).

### 5.4.3 Rekursive Strukturen retinaler Neuronen

In der entwickelten Verarbeitungskette kann durch eine Adaption der ersten Verarbeitungsstufe (die Kopplung der Rezeptoren) insgesamt eine variable Extraktion von orientierten Konturmerkmalen realisiert werden. Für die Rezeptorkopplung, die in der Modellierung als Gaußtiefpaß eingeht, wurde bislang eine nichtrekursive Filterstruktur (bzw. ein neuronales Feed-Forward Netz) verwendet. Da die Filterwirkung adaptiv ist und die Größe der Impulsantwort (des rezeptiven Feldes) variabel ist, würde sich eine rekursive Struktur anbieten, biologisch gar als sehr plausibel erscheinen. Mit einem rekursiven Filter kann z.B. ein Tiefpaß realisiert werden, bei dem allein durch die Werte und nicht die Anzahl der Filterko-

effizienten die Grenzfrequenz des Filters eingestellt werden kann. Als rekursives neuronales Netz bliebe hierdurch dann die Anzahl der synaptischen Verbindungen, die auf ein Zielneuron konvergieren, konstant, nur die synaptischen Gewichte selbst verändern sich adaptiv und bewirken durch die rekursive Verschaltung die gewünschte Filterwirkung. Zudem kann ein rekursives (2D) Digitalfilter auch immer direkt durch seine Differenzengleichung als ein neuronales Verbindungsnetzwerk interpretiert werden (siehe Kapitel 2).

In [Der93] und [vVYV98] werden verschiedene Realisierungen zweidimensionaler Gaußtiefpaßfilter vorgeschlagen. Diese beruhen darauf, daß auf einem rechtwinkligen Abtastraster die Impulsantwort durch

$$e^{-\|\mathbf{t}\|^2} = e^{-\|\mathbf{T}_r \mathbf{k}\|^2} = e^{-T_r^2 \cdot (k_1^2 + k_2^2)} = e^{-T_r^2 \cdot k_1^2} \cdot e^{-T_r^2 \cdot k_2^2} \quad (5.60)$$

separierbar ist, und somit die zweidimensionale Filterung auf zwei eindimensionale Filteroperationen zurückgeführt werden kann. Auf einem hexagonalen Abtastraster ist allerdings die Impulsantwort nicht mehr in den Koordinatenrichtungen separierbar, da hier gilt:

$$e^{-\|\mathbf{t}\|^2} = e^{-\|\mathbf{T}_h \mathbf{k}\|^2} = e^{-T_h^2 \cdot ((k_1 + k_2)^2 + 3(k_1 - k_2)^2)} = e^{-4T_h^2 \cdot (k_1^2 + k_2^2 - k_1 k_2)}. \quad (5.61)$$

Im folgenden wird kurz vorgestellt, wie dennoch eine separierbare Verarbeitung erreicht werden kann. Betrachtet man nämlich die Diagonalkoordinaten  $\mathbf{d}$  des Rasters  $\mathbf{k}$ , die aus einer Drehung um  $45^\circ$

$$\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \quad (5.62)$$

hervorgehen, so ergibt sich die Beziehung

$$k_1^2 + k_2^2 - k_1 k_2 = \frac{1}{2} d_1^2 + \frac{3}{2} d_2^2 \quad (5.63)$$

und damit

$$e^{-\|\mathbf{t}\|^2} = e^{-4T_h^2 \cdot (\frac{1}{2} d_1^2 + \frac{3}{2} d_2^2)} = e^{-2T_h^2 \cdot d_1^2} \cdot e^{-2T_h^2 \cdot 3d_2^2}, \quad (5.64)$$

also ebenfalls eine separierbare Impulsantwort in den Diagonalkoordinaten. Somit kann durch eindimensionale Gaußfilterung in den beiden Diagonalkoordinaten des hexagonalen Rasters ebenfalls eine zweidimensionale Gaußfilterung erzielt werden. Dies kann dann auch mittels eindimensionaler rekursiver Filter geschehen. Hier ist zu erkennen, daß in den beiden Filterrichtungen allerdings unterschiedliche Filterstärken eingesetzt werden müssen, um insgesamt eine isotrope zweidimensionale Filterantwort zu erhalten. Damit ist eine rekursive Realisierung der Rezeptorkopplung auf der Basis eindimensionaler Filter auch auf einem hexagonalen Abtastraster möglich.



# Experimentelle Ergebnisse

In diesem Kapitel wird die Leistungsfähigkeit der in dieser Arbeit vorgestellten biologienahen Verfahren zur Bildvorverarbeitung mit realen Kameradaten untersucht. Der Schwerpunkt wird hier insbesondere auf die adaptive Vorfilterung mit anschließender Konturextraktion gelegt, welche, wie man zeigen kann, die wichtigste Komponente einer robusten Merkmalsextraktion ist. Die Untersuchungsergebnisse der übrigen in dieser Arbeit entwickelten Aspekte eines biologischen oder technischen Systems, wie die Interpolation für ein hexagonales Raster, die Simulation des nichtadaptiven Modells und der pulscodierten Mechanismen wurden bereits direkt in den jeweiligen Kapiteln gegeben. Für den Vergleich mit bestehenden Verfahren soll die Verarbeitung hier exemplarisch auf einem rechtwinkligen Abtastraster erfolgen. Wie im vorangegangenen Kapitel erläutert, sind die Strukturen jedoch jederzeit auf einem hexagonalen Raster implementierbar.

Bei der Gewinnung und Verarbeitung realer Kameradaten (Kameramodell *teli Micro-Camera CS3330*) für die adaptive Rauschfilterung sind zusätzliche Besonderheiten zu beachten, auf die im ersten Abschnitt eingegangen werden soll. Hier wird das Prinzip zur Messung des SNR mit gewissen Testmustern beschrieben, was die Grundlage der quantitativen Beurteilung der Algorithmen ist und deshalb viel Sorgfalt erfordert. Da die vorgestellte adaptive Strategie über die Helligkeit der Bildinhalte gesteuert wird, ist hier auch die Kompensation des AGC (automatic gain control) der Kamera von großer Bedeutung. Zudem wird anschließend die Kompensation des FPN (Fixed Pattern Noise) behandelt, weil bei sehr dunklen Szenen der systematische Anteil des Dunkelstromrauschens im Vergleich zum Signal nicht mehr zu vernachlässigen ist.

Für die vorgestellte Strategie zur adaptiven Rauschfilterung ist die Kenntnis der Rauschparameter der Kamera vorausgesetzt. Wie in einem weiteren Abschnitt gezeigt, können diese aber recht gut mit Hilfe von Testmustern und einem Regressi-

onsansatz ermittelt werden.

Abschließend wird eine adaptive Filterung mit der vorgestellten biologienahen Strategie durchgeführt und mit etablierten Standardverfahren verglichen. Dazu wird zunächst die Leistungsfähigkeit bei der Verbesserung des SNR untersucht. Anschließend werden die Ergebnisse der Konturdetektion vorgestellt.

## 6.1 Verarbeitung realer Kameradaten

Die vorgestellte Strategie zur adaptiven Rauschfilterung verwendet das Signal-zu-Rauschverhältnis und die Beleuchtungsstärke, um die Filterung zu steuern. Aus diesem Grund wird zunächst auf die schwierige Bestimmung des SNR und auf die Bestimmung des Verstärkungsfaktors der Kamera eingegangen. Anschließend wird die Kompensation des Dunkelstromrauschens besprochen, um die Gültigkeit des Rauschmodells auch in extrem dunklen Situationen zu gewährleisten.

### 6.1.1 Meßbilder zur Schätzung des SNR

Um die vorgestellten Algorithmen im Einsatz untersuchen zu können, ist die Verwendung von Testmustern notwendig, auf denen definierte weiße und schwarze Rechtecke oder Quadrate aufgedruckt sind (Abb. 6.1). Die aufgenommenen Bilddaten enthalten immer bereits die additive Überlagerung von Original- und Rauschsignal. Anhand der Testmuster ist es aber möglich, Schätzungen beider Signale getrennt voneinander zu berechnen, um dann z. B. eine Schätzung für das Verhältnis der Signalleistungen (SNR) angeben zu können.

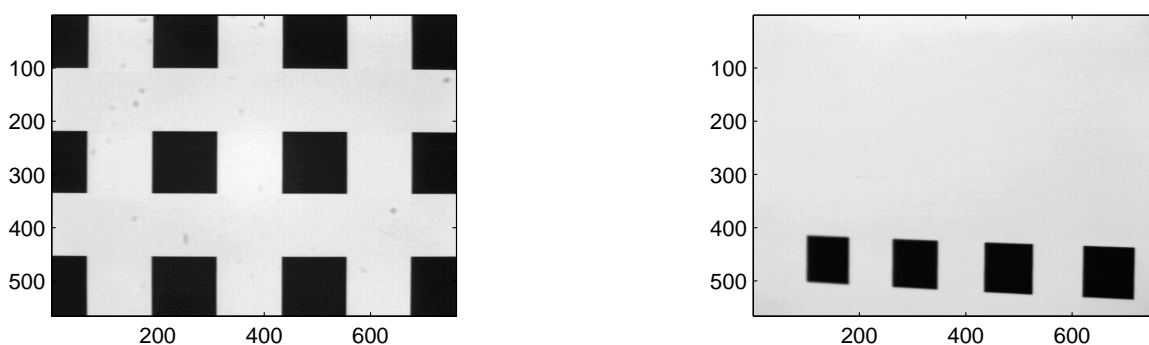


Abbildung 6.1: Kamerabilder verwendeter Testmuster. In Regionen mit näherungsweise konstantem Signal  $x_o \equiv \text{const}$  kann das SNR geschätzt werden.

Von dem Testmuster ist bekannt, daß das Originalsignal  $x_o$  der Szene (der Grauwert des Testmusters selbst) in definierten, uns bekannten Regionen konstant über

den Ort ist. Daher sollte eine Schätzung dieses Signals und des Rauschens durch einfache örtliche Mittelwertbildung

$$\begin{aligned}\gamma x_o &\approx E\{x\} \\ n &= \alpha n_1 + \gamma x_o \beta n_2 \approx x - E\{x\}\end{aligned}\tag{6.1}$$

möglich sein. Für diesen Zweck ist aber ein gleichmäßig ausgeleuchtetes Testbild erforderlich, was in der realen Messung natürlich nie erreicht werden kann. In der Tat hat sich gezeigt, daß diese Bedingung sehr schwer erfüllbar ist. Hierauf sollte also besonders geachtet werden. Trotz des Einsatzes spezieller Fotolampen, die ein diffuses Licht erzeugten, wiesen die Messungen leicht inhomogene Ausleuchtungen auf. Um dennoch eine gute Schätzung des Originalsignals und des Rauschens zu erhalten, musste das Meßfenster entsprechend klein gewählt werden, damit darin das Signal näherungsweise konstant ist.

Es wurde eine Vielzahl von Testbildern mit verschiedenen, abnehmenden Beleuchtungsstufen  $\gamma$  mittels einer dimmbaren Fotolampe in einer abgedunkelten Laborumgebung aufgenommen. Für Kompensationszwecke waren zu jeder Beleuchtung jeweils ein Bild mit und ohne AGC (automatic gain control) notwendig. Darüber hinaus diente eine Reihe von Dunkelbildern der Kompensation des Dunkelstrommusters FPN (fixed pattern noise).

Eine Mittelung einer statischen Szene über die Zeit scheint ebenfalls eine alternative Methode für die Schätzung der Signal- und Rauschleistung zu sein [Eis03]. Erste Versuche haben gezeigt, daß in dunklen verrauschten Situationen mindestens 100 Bilder benötigt werden, um mit einer pixelweisen Mittelung das Rauschen zu unterdrücken. Das Resultat ist dann eine Schätzung des eigentlichen Signals  $\gamma x_o$ , welches hier nicht zwingend konstant bzgl. des Ortes sein muß. Eine Subtraktion dieser Schätzung von einem verrauschten Bild würde dann auf das Rauschsignal führen. Der Szeneninhalt darf sich dabei allerdings nicht ändern. Hierfür muß die Kamera fest montiert sein, um Einflüsse durch Vibrationen zu vermeiden. Diese einschränkenden Bedingungen sind aber nur für Analysezwecke notwendig, nicht etwa in einer späteren Anwendung. Der Ansatz soll hier nicht weiter verfolgt werden, da er in [Eis03] detailliert untersucht wird.

### 6.1.2 Kompensation des AGC

In dunklen Szenen verstärkt die Elektronik der Kamera üblicherweise automatisch das gesamte Signal, d.h. das Originalsignal  $x_o$  und das Rauschsignal  $n$ , bevor es in ein Videosignal wie PAL umgesetzt wird. Natürlich ist dieses sogenannte AGC

(automatic gain control) notwendig, um den zur Verfügung stehenden Signalpegel auszunutzen und somit das Videosignal robust zu machen gegen Störungen auf dem Übertragungskanal (Kabel). Weiterhin ist es dadurch möglich, den maximalen Dynamik- oder Quantisierungsbereich der Signalamplitude bei der Digitalisierung durch den Framegrabber zu nutzen. In Abb. 6.2 ist ein Beispiel einer dunklen Szene ohne und mit AGC dargestellt. Es ist zu erkennen, daß ohne AGC

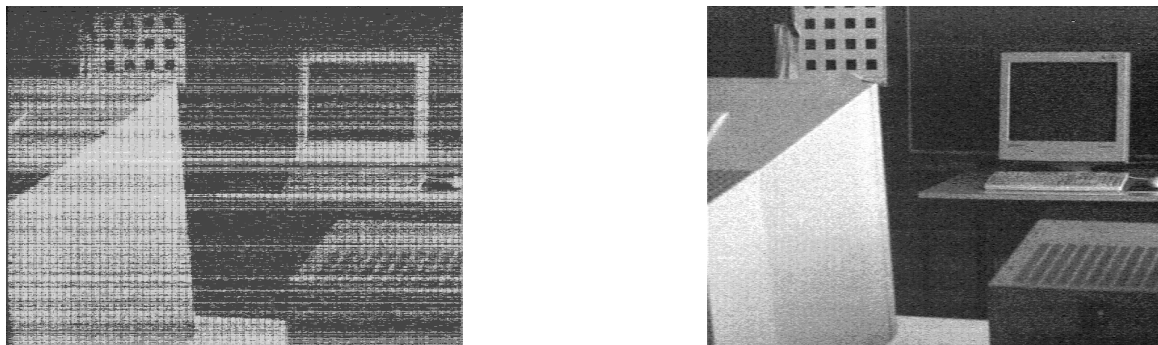


Abbildung 6.2: Aufnahme einer dunklen Szene ohne und mit automatischer Pegelverstärkung des Kamerasystems.

der Dynamikbereich viel zu gering für eine sinnvolle Signalverarbeitung ist. Das AGC hingegen verstärkt den Signalpegel, aber gleichzeitig auch das Rauschen. Zur besseren Darstellung wurde hier die Szene etwas aufgehellt.

Leider ist es aber nicht möglich, Szenen unterschiedlicher Beleuchtung  $\gamma$  zu analysieren und zu vergleichen, wenn die interne Verstärkung der Kamera nicht bekannt ist. Für diesen Zweck ist es wichtig, eine gewisse Kontrolle über das AGC der Kamera zu besitzen, um z. B. den Verstärkungsfaktor der Kamera auszulesen. Moderne Kameras verfügen meist über Schnittstellen zur Kontrolle des AGC [Ima]. Da die vorliegenden Untersuchungen aber mit statischen Bildszenen gemacht wurden, war eine echtzeitfähige Kontrolle des AGC nicht notwendig. Hier war ein Ausschalten des AGC ausreichend. Durch eine Aufnahme mit AGC und eine weitere Aufnahme mit ausgeschaltetem AGC war es hier möglich, implizit die Verstärkung herauszurechnen. Wie bereits erwähnt, ist das AGC für dunkle Szenen sehr wichtig, da sonst eine sehr schlechte Amplitudenauflösung nach der Digitalisierung zu unbrauchbaren Ergebnissen führt. Diese Tatsache spiegelt sich auch in dem Histogramm des Kamerabildes mit AGC in Abb. 6.3 (links) und in dem Bild ohne AGC (mitte) wider. Werden jedoch bei den Testmessungen jeweils ein Bild der Szene mit AGC  $x_{AGC}$  und eines ohne AGC  $x_{NoAGC}$  aufgenommen, so ist es möglich, das AGC-Bild (bei dunkler Szene) auf ein korrektes  $\gamma$  zu skalieren,



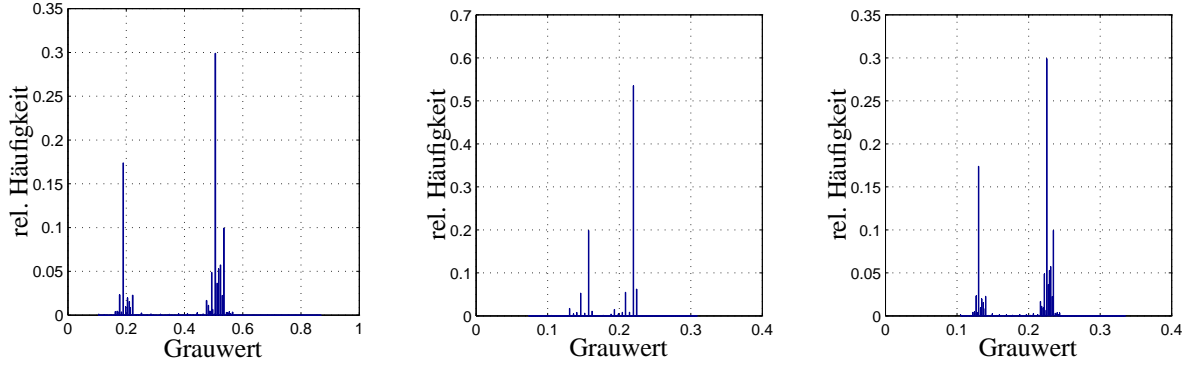


Abbildung 6.3: AGC Kompensation. links: Histogramm eines realen Kamerabildes bei eingeschaltetem AGC. mitte: Histogramm eines Bildes ohne AGC. rechts: Histogramm nach der AGC Kompensation.

um dessen hohe Amplitudenauflösung nutzen zu können. Mit der Transformation

$$x_{\text{comp}}(\mathbf{k}) = x_{\text{NoAGC},\min} + \dots + (x_{\text{AGC}}(\mathbf{k}) - x_{\text{AGC},\min}) \frac{\sum (x_{\text{NoAGC}}(\mathbf{k}) - x_{\text{NoAGC},\min})}{\sum (x_{\text{AGC}}(\mathbf{k}) - x_{\text{AGC},\min})} \quad (6.2)$$

und

$$x_{\text{NoAGC},\min} = \min_{\mathbf{k}} x_{\text{NoAGC}}(\mathbf{k}), \quad x_{\text{AGC},\min} = \min_{\mathbf{k}} x_{\text{AGC}}(\mathbf{k}), \quad (6.3)$$

ist diese AGC-Kompensation mit recht guter Näherung möglich. Bei der verwendeten Kamera war das Minimum der Grauwerte mit und ohne AGC etwa gleich groß, also  $x_{\text{NoAGC},\min} \approx x_{\text{AGC},\min}$ . Das Histogramm nach der Kompensation in Abb. 6.3 (rechts) hat offensichtlich den selben Signalpegel wie das Bild ohne AGC, weist aber die gute Amplitudenauflösung des AGC-Bildes auf (links).

### 6.1.3 Kompensation des FPN

Für extrem dunkle Szenen ist in einem weiteren Schritt eine Dunkelbildkompensation notwendig. Wir betrachten dazu im folgenden den *Dunkelstrom* (*dark current*), welcher als Überlagerung eines systematischen Anteils  $E_{\text{DC}}$  und eines mittelwertfreien Rauschprozesses aufgefaßt wird [Kam]. Letzterer kann in dem betrachteten Rauschmodell den übrigen Rauschprozessen zugeordnet werden. Um aber den systematischen Anteil des Dunkelstromes, der bislang vernachlässigt werden konnte, zu berücksichtigen, muß das Rauschmodell entsprechend erweitert werden zu

$$x = [\gamma x_o + E_{\text{DC}}] + \alpha n_1 + \gamma x_o \beta n_2. \quad (6.4)$$

Der Erwartungswert des Dunkelstromes  $E_{DC}$ , welcher auch als FPN (Fixed Pattern Noise) bezeichnet wird, variiert mit dem Ort  $k$  und hat seine Ursache in der irregulären Kristallstruktur des CCD. Deshalb muß zunächst mit einem Testdatensatz von (hier 10) Dunkelbildern, welche durch Verwendung eines lichtundurchlässigen Objektivdeckels gewonnen werden können, die systematische Komponente  $E_{DC}$  geschätzt werden. Dies geschieht durch pixelweise Mittelung über diese Bilder, um die Zufallskomponente zu eliminieren. Nach einer einfachen Subtraktion der gemittelten Dunkelbilder von dem aktuellen Kamerabild können wir wieder die Gültigkeit des Kameramodells (5.1) sicherstellen. In Abb. 6.4 ist die Schätzung des Dunkelbildes dargestellt.

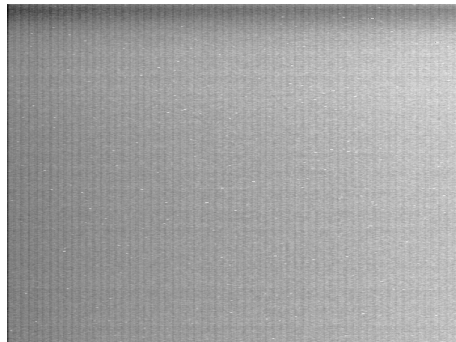


Abbildung 6.4: Schätzung des Dunkelbildes (FPN=Fixed Pattern Noise) der verwendeten Kamera.

Ohne die beschriebene Kompensation ist das verwendete Rauschmodell nicht gültig und die Ergebnisse weichen bei realen Kameradaten recht stark von den simulierten Szenen ab. Wie in [TEDH02] zu sehen, führt die Kompensation jedoch zu genaueren Schätzungen des Signals und des Rauschens, so daß die Simulationsergebnisse erreicht werden können. Dies ist ein Indiz dafür, daß das verwendete Rauschmodell eine gute Näherung der realen Verhältnisse wiedergibt. Im folgenden Abschnitt sollen anhand von realen Testmustern die Parameter  $\alpha$  und  $\beta$  des Modells geschätzt werden.

## 6.2 Schätzung der Rauschparameter

Für die adaptive Filterung mittels des expliziten Steuersignals (5.31), aber auch für die Verifikation des Modells ist es wichtig, die Rauschparameter  $\alpha$  und  $\beta$  zu schätzen. Dazu werden Meßbilder unterschiedlicher Beleuchtungen  $\gamma_i$  (hier  $i = 1 \dots 22$ ) aufgenommen, in denen Regionen mit einem konstanten Originalsignal  $x_o$  vorkommen. Anschließend wird eine Kompensation des AGC und FPN

durchgeführt. Eine Schätzung des Signals bzw. des Rauschens wird gemäß (6.1) berechnet, wobei sich die zugehörigen Leistungen entsprechend ergeben. Unter dieser Annahme ergäbe sich mit dem verwendeten Rauschmodell die Leistung des Gesamtrauschens für jede Beleuchtungssituation  $\gamma_i$  zu

$$\sigma_{n,i}^2 = \alpha^2 + \gamma_i^2 \beta^2 x_o^2 \quad (6.5)$$

und damit das überbestimmte Gleichungssystem

$$\begin{bmatrix} \sigma_{n,1}^2 \\ \vdots \\ \sigma_{n,N}^2 \end{bmatrix} = \begin{bmatrix} 1 & \gamma_1^2 x_o^2 \\ \vdots & \vdots \\ 1 & \gamma_N^2 x_o^2 \end{bmatrix} \begin{bmatrix} \alpha^2 \\ \beta^2 \end{bmatrix}. \quad (6.6)$$

Mittels linearer Regression kann dann eine Schätzung im Sinne der kleinsten Fehlerquadrate für die Unbekannten  $\alpha^2$  und  $\beta^2$  gefunden werden. Das Vorzeichen spielt hierbei keine Rolle, da es keinen Einfluß auf die Rauschleistung hat. Somit können die positiven Wurzeln der Regressionslösungen verwendet werden. Für die eingesetzte Kamera ergeben sich die folgenden Schätzungen für die Rauschparameter

$$\alpha \approx 1.47 \cdot 10^{-3} \quad (6.7)$$

$$\beta \approx 6.01 \cdot 10^{-3}, \quad (6.8)$$

wobei diese sich auf Bilddaten beziehen, die auf eins normiert sind. Das SNR kann für die Meßdaten ebenfalls angegeben werden, da die Leistung des Signals

$$\gamma_i^2 x_o^2 = E \{x_i\}^2 = \mu_{x_i}^2 \quad (6.9)$$

direkt durch das Quadrat des Mittelwertes der Grauwerte in den betrachteten Regionen geschätzt werden kann. Betrachtet man nun die Schätzungen des SNR anhand der Meßdaten und anhand des Modells in Abb. 6.5, so scheint das Modell des Rauschens hier sehr erfolgreich zu sein. Es wäre sicher sinnvoll, weitere Untersuchungen anzustellen, die Aufschluß darüber geben, ob die geschätzten Rauschparameter konstant über die Zeit sind oder sich unter Umständen durch Erwärmung der Kamera verändern. Sollte dies der Fall sein (Thermisches Rauschen), so ist allerdings nach einer gewissen Aufwärmphase ein stabiler Zustand zu erwarten. Für die vorgestellten experimentellen Ergebnisse kann daher stets davon ausgegangen werden, daß hier keine oder nicht meßbare Einflüsse vorhanden sind.

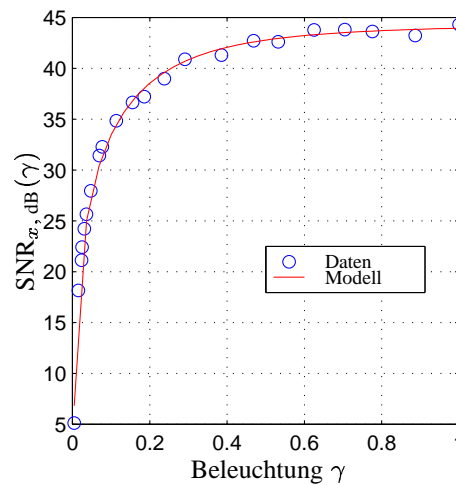


Abbildung 6.5: Messungen des SNR einer Verdunklungssequenz bei realen Kamerabildern (o) und mit geschätzten Rauschparametern (–).

## 6.3 Adaptive Filterung

Mit Hilfe der geschätzten Rauschparameter der verwendeten Kamera ist eine adaptive Rauschunterdrückung mittels des vorgestellten expliziten Steuersignals möglich. Anhand von Testmustern wird zunächst die erfolgreiche adaptive Verbesserung des Signal-zu-Rauschabstandes demonstriert und mit einem etablierten Standardverfahren, dem adaptiven *WIENER-Filter* nach LEE [Lim90, Lee80, Mat], verglichen. Anschließend werden komplexere Szenen auf diese Weise vorgefiltert, um eine Konturextraktion in extrem dunklen Situationen durchzuführen. Hier kommt zunächst der *CANNY-Operator* [Can86, Mat] zum Einsatz, um die Leistungsfähigkeit der adaptiven Rauschfiltermethoden bewerten zu können. Darüber hinaus werden die Ergebnisse des in Kapitel 5 vorgestellten biologienahen Systems zur Extraktion von orientierten Konturmerkmalen gezeigt, das für ein adaptives Verhalten entwickelt wurde.

### 6.3.1 Verbesserung des SNR

Wie in Abschnitt 5.3 an simulierten Szenen veranschaulicht und theoretisch analysiert, erwarten wir durch eine Tiefpaßfilterung eine Reduktion der Rauschleistung. Fällt die Leistung des Signals dann in kleinerem Maße ab, so wird das SNR verbessert. In Abb. 6.6 wird dieses Verhalten mit der Verdunklungssequenz experimentell bestätigt. Der Parameter der Kurvenschar ist die Beleuchtung  $\gamma$ . Um die Leistungen in den realen Kameradaten schätzen zu können, werden hier definierte Testmuster verwendet. Aus diesen Meßdaten könnte nun ein Steuersignal durch

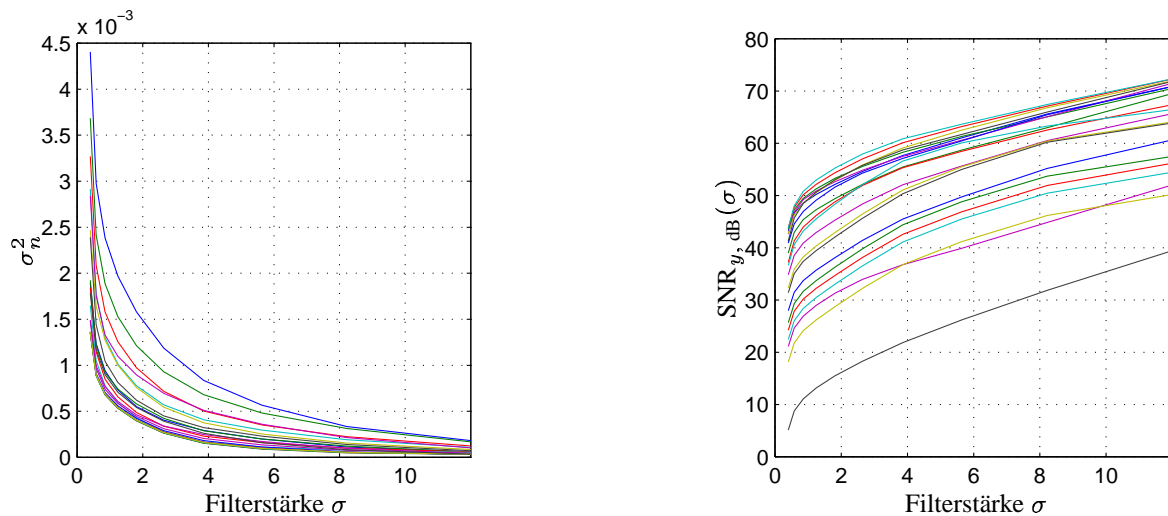


Abbildung 6.6: Auswirkung der Tiefpaßfilterung einer realen Szene unter verschiedenen Beleuchtungsverhältnissen. links: Reduktion der Rauschleistung. rechts: Verbesserung des SNR.

einen Optimierungsprozeß gemäß (5.28) abgeleitet und eine adaptive Filterung mit realen Kameradaten durchgeführt werden. Dies soll an dieser Stelle nicht im Detail betrachtet werden. Der interessierte Leser sei hier auch auf [Eis01] verwiesen. Das Ergebnis einer adaptiven Verbesserung des SNR einer simulierten Szene basierend auf solch einem Steuersignal wurde bereits in Abb. 5.6 verdeutlicht.

Wie eine Vielzahl weiterer Simulationen zeigt, sind die Ergebnisse einer adaptiven Verbesserung des SNR bei natürlichen Szenen einerseits und Testmustern mit konstanten Signalbereichen andererseits qualitativ recht ähnlich [Eis01]. Dies ist ein Grund, warum im weiteren Verlauf dieser Arbeit die Verwendung eines expliziten Steuersignals (5.31) der Konstruktion eines Steuersignals mittels eines Optimierungsprozesses (5.28) vorgezogen wird. Durch den Einsatz des expliziten Steuersignals kann zudem ein gewünschtes SNR vorgegeben werden, was für den praktischen Nutzen ein wichtiger Vorteil ist. Die Rauschparameter der Kamera, die hierfür notwendig sind, wurden bereits im vorangegangenen Abschnitt experimentell bestimmt.

In Abb. 6.7 (links) ist die adaptive Verbesserung unter Verwendung eines expliziten Steuersignals dem Ergebnis eines adaptiven WIENER-Filters und dem SNR ohne Filterung gegenübergestellt. Mit dem WIENER-Filter und einer Nachbarschaft von  $3 \times 3$  fällt das SNR unterhalb  $\gamma = 0.5$  zunächst langsam, dann stärker ab und erreicht bei der extrem dunklen Situation ( $\gamma_{\min} = 0.0045$ ) nur noch 10 dB. Durch Erhöhung des Nachbarschaftsparameters auf  $11 \times 11$  verschiebt sich die

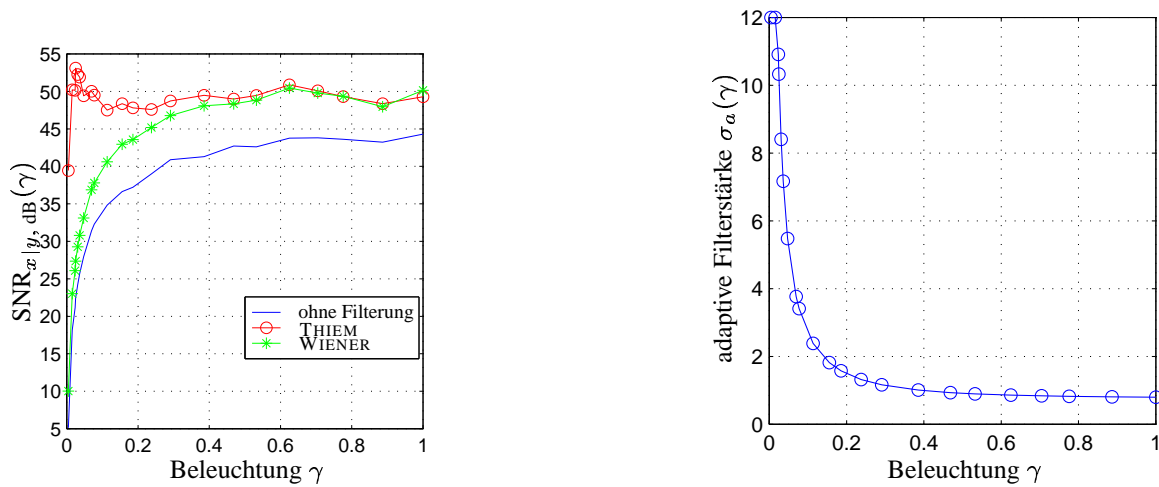


Abbildung 6.7: Adaptive Verbesserung des SNR. links: Vorgestellte Strategie (o) im Vergleich zu einem WIENER-Filter (\*) und dem Verhalten ohne Filterung. rechts: Adaptiv eingestellter Filterparameter (Steuersignal).

Kurve insgesamt einige dB nach oben und bei  $\gamma_{\min}$  werden noch 19.9 dB gemessen (siehe Tabelle 6.1). Bei dieser großen Nachbarschaftskopplung und der damit verbundenen Verschiebung nach oben wird allerdings bei helleren Szenen eine sehr starke Filterung durchgeführt, obwohl dies nicht unbedingt notwendig ist. Im Gegensatz dazu ist die vorgestellte adaptive Filterung auf der Basis des expliziten Steuersignals sehr erfolgreich. Das gewünschte SNR (zum Vergleich hier auf 50 dB gesetzt) kann hier über den gesamten Bereich nahezu gehalten werden. Erst bei extrem dunklen Szenen fällt das SNR schließlich auf etwa 40 dB ab. Dies ist aber im Vergleich zu der ungefilterten Szene ( $\text{SNR} = 5.1 \text{ dB}$ ) immerhin noch eine Verbesserung um etwa 35 dB. Der adaptiv eingestellte Filterparameter  $\sigma_a$  ist in Abb. 6.7 (rechts) gezeigt und spiegelt das verwendete Steuersignal wider.

Ein Vergleich der Berechnungsdauer der gegenübergestellten Verfahren zur adaptiven Rauschfilterung zeigt Tabelle 6.1. Die Berechnungen wurden mit Matlab durchgeführt. Die Zeitangaben lassen daher nur Aussagen über relative Verhältnisse zu, absolute Zeitmessungen insbesondere im Hinblick auf Echtzeitfähigkeit können nicht gemacht werden und waren hier auch nicht Gegenstand der Betrachtungen. Bei dem adaptiven WIENER-Filter kann durch manuelle Vergrößerung der Nachbarschaft auf  $11 \times 11$  eine Verbesserung der Qualität erreicht werden, wodurch der Berechnungsaufwand allerdings mehr als verdoppelt wird. Die Bildgröße geht etwa linear in die Berechnungsdauer ein. Die vorgestellte Strategie wurde mittels des expliziten Steuersignals durchgeführt und als nichtrekursives (FIR) und rekursives (IIR) Filter implementiert. Der Berechnungsaufwand ergibt

Filtermethode			Berechnungsdauer $T[s]$		SNR <sub>y</sub> , dB bei $\gamma_{\min}$
			$761 \times 566$	$512 \times 512$	
WIENER	$3 \times 3$		2.1	1.3	10
	$11 \times 11$		4.6	2.7	19.9
vorgestellte Strategie	45 dB	FIR	3.9	1.9	39.4
		IIR	1.7	1.0	39.4
	50 dB	FIR	7.2	3.9	39.4
		IIR	1.7	1.0	39.4

Tabelle 6.1: Vergleich der Resultate der adaptiven Rauschfilter unter Matlab.

sich hier hauptsächlich durch die Filterroutine und ist als FIR-Filter höher als bei dem WIENER-Filter. Insbesondere bei einem gewünschten SNR von 50 dB ergibt sich durch die langen Impulsantworten ein höherer Berechnungsaufwand. Allerdings lassen sich dadurch auch wesentlich bessere Ergebnisse bei der adaptiven Rauschfilterung erzielen. Daß man nicht unbedingt einen höheren Preis für die überzeugenden Filterergebnisse bezahlen muß, zeigt schließlich die Variante mittels einer IIR-Implementation. Hier sind die Berechnungszeiten nur abhängig von dem Filtergrad, also unabhängig von dem gewünschten SNR bzw. von der Filterstärke und erheblich kürzer als die des WIENER-Filters.

### 6.3.2 Robuste Konturextraktion

Abschließend soll eine Konturextraktion durchgeführt werden. Eine vorgeschaltete adaptive Rauschfilterung soll hierfür das Ergebnis verbessern. Es wird hier der Schwerpunkt auf eine qualitative Aussage gelegt. Quantitative Betrachtungen, wie z. B. die Verbesserung einer konturbasierten Objekterkennung bleiben hingegen weitergehenden Untersuchungen überlassen.

Exemplarisch wird eine sehr dunkle Szene mit einer Beleuchtung von  $\gamma \approx 0.03$  verwendet werden, die in Abb. 6.8 (links) gezeigt ist. Da es sich um eine komplexe Szene handelt, kann das Signal-zu-Rauschverhältnis nicht aus den Bilddaten geschätzt werden. Mittels  $\gamma$  kann aber aus den Messungen nach Abb. 6.5 ein Schätzwert von SNR  $\approx 23$  dB ermittelt werden. Da hier die Beleuchtungsstärke aus dem globalen Mittelwert der Szene ermittelt wird, kann das tatsächliche lokale SNR verständlicherweise in helleren oder dunkleren Bildregionen in geringem Maße nach oben oder unten davon abweichen. Die Extraktion von Kantenmerkmalen, z. B. mittels des CANNY-Operators, ergibt ein unbrauchbares Resultat, wenn sie direkt auf dem verrauschten Szenenbild durchgeführt wird (Abb. 6.8 (rechts)).

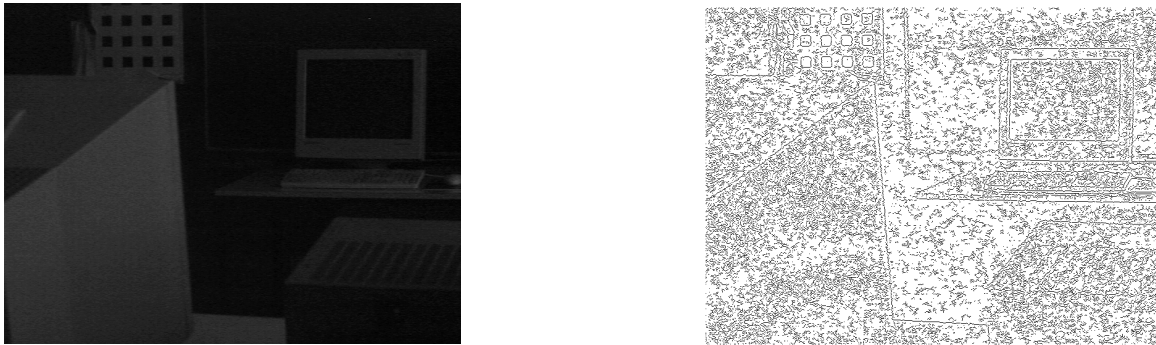


Abbildung 6.8: Beispiel einer Konturextraktion mittels des CANNY-Operators in einer sehr dunklen und daher sehr verrauschten Szene.

Wie zu erwarten, ist eine vorherige Rauschunterdrückung unerlässlich. Als Referenz verwenden wir hier zunächst das adaptive WIENER-Filter, welches sich als Standardwerkzeug etabliert hat und vielfach eingesetzt wird. Bei extrem schwach beleuchteten Szenen mit niedrigem SNR sind die Grenzen des Verfahrens aber schnell erreicht, wie in Abb. 6.9 zu sehen ist. Zwar kann mittels der Nachbarschaft, in der die lokale Bildstatistik berechnet wird, die Qualität etwas verbessert werden, in den Konturbereichen ist allerdings noch eine recht hohe Rauschleistung vorhanden. Zudem ist dieser Freiheitsgrad von Hand zu optimieren, so daß hier der Vorteil einer automatischen Adaption verloren geht.

Über die gemessene Beleuchtungsstärke  $\gamma \approx 0.03$  kann nun adaptiv über das explizite Steuersignal für ein gewünschtes SNR die erforderliche Filterstärke ermittelt werden. Da in gut ausgeleuchteten Daten der verwendeten Kamera ein typisches SNR von 45 . . . 50 dB gemessen werden konnte, sollen diese Werte hier als gewünschtes SNR der Adaption verwendet werden. Um dieses zu erreichen, wird eine Filterstärke von  $\sigma_a \approx 3 \dots 5.3$  adaptiv eingestellt, und es ergibt sich das Ergebnis der Rauschreduktion in der linken Spalte von Abb. 6.10 für 45 (oben) bzw. 50 dB (unten). In der rechten Spalte sind die Ergebnisse einer anschließenden Konturextraktion dargestellt. Hier ist gegenüber Abb. 6.8 und Abb. 6.9 eine enorme Verbesserung der Qualität des Kantenbildes feststellbar. Eine Verbesserung des SNR erhöht hier offensichtlich die Robustheit der Konturextraktion gegenüber Rauschen. Hochfrequente Details in der Szene gehen dabei aber natürlich weiter verloren. Eigene Untersuchungen haben gezeigt, daß ein  $\text{SNR}_{\text{des}} = 45$  dB für das vorliegende Kameramodell für die meisten Situationen durchaus ausreichend ist.



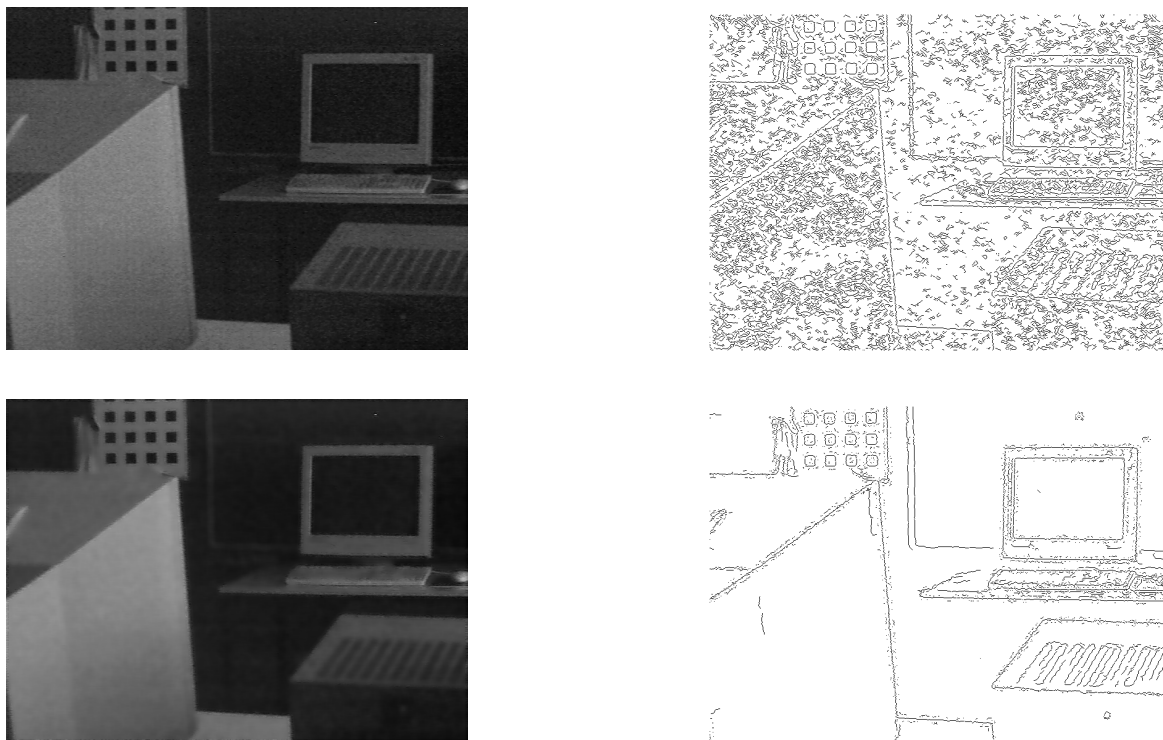


Abbildung 6.9: Rauschunterdrückung mittels eines adaptiven WIENER-Filters zur Verbesserung einer anschließenden Konturextraktion. Die Qualität kann durch die Nachbarschaftsgröße (oben: 3x3, unten: 11x11) manuell optimiert werden.

Statt des CANNY-Operators wird abschließend das in dieser Arbeit entwickelte biologische kortikale Modell auf der Basis der Ableitungsketten als Konturdetektion verwendet. Für ein gewünschtes SNR von 45 dB wurde eine Rauschunterdrückung durchgeführt und anschließend die Ableitungen in horizontaler Richtung  $0^\circ$  gebildet, was in Abb. 6.11 zu sehen ist. Hierdurch werden alle vertikalen Konturmerkmale aus der stark verrauschten Szene mit Erfolg extrahiert. Die Orientierungsauflösung wurde hier exemplarisch zu  $15^\circ$  ( $\sigma_\perp = 12$ ) und  $30^\circ$  ( $\sigma_\perp = 5.8$ ) gewählt. An dieser Stelle sollte auch erwähnt werden, daß eine Skelettierung wie in [Göt00] oder eine Non-Maximum-Suppression wie bei dem CANNY-Operator natürlich ebenfalls ein pixelbreites Konturergebnis liefern könnte, falls die aufbauende Objekterkennung des Bildverarbeitungssystems dies fordert. Darauf soll im Rahmen dieser Betrachtungen aber verzichtet werden.

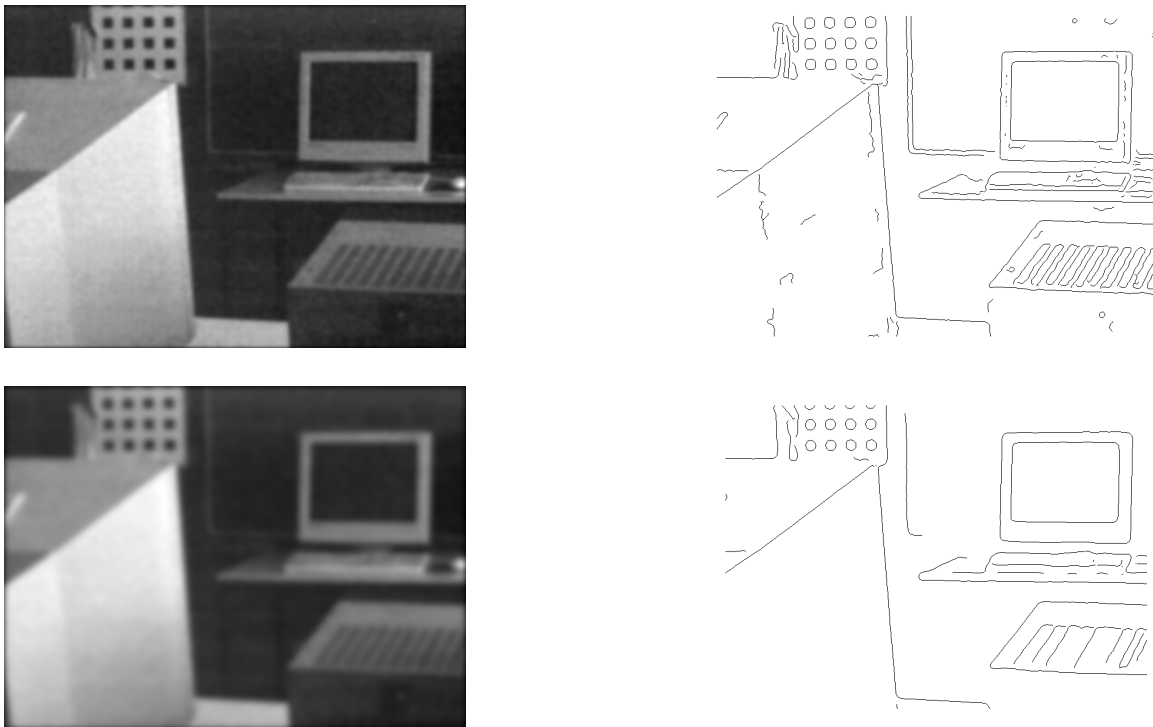


Abbildung 6.10: Rauschunterdrückung mittels der vorgestellten expliziten Adaption zur Verbesserung einer anschließenden Konturextraktion (oben: 45 dB, unten: 50 dB).



Abbildung 6.11: Adaptive Rauschunterdrückung ( $\text{SNR}_{\text{des}} = 45 \text{ dB}$ ) und anschließende orientierte Konturextraktion ( $0^\circ$ ) auf der Basis der Ableitungsketten mit einer Orientierungsauflösung von  $15^\circ$  bzw.  $30^\circ$ .

## 6.4 Fazit

Mit den vorliegenden experimentellen Ergebnissen konnte demonstriert werden, daß eine adaptive Verbesserung des SNR mit dem expliziten Steuersignal einfach und effizient durchgeführt werden kann. Trotz der gemachten Näherungen in der theoretischen Herleitung und dem Einsatz einer globalen Filterung auf der Basis des Bildmittelwertes überzeugen die Resultate auch bei realen Kameraszenen. Gerade bei extrem dunklen und damit extrem verrauschten Bilddaten versagen etablierte Standardverfahren, wie das adaptive WIENER-Filter. Zudem konnte die Strategie durch einen Vergleich mit den biologischen Mechanismen motiviert und begründet werden. Ein Nachteil des WIENER-Filters scheint die Einschränkung zu sein, daß die Rauschleistung bekannt sein muß, um diese Methode effizienter anzuwenden. Hier wäre eine Kombination aus dem vorgestellten expliziten Steuersignal und dem WIENER-Filter denkbar, da über die Helligkeit  $\gamma$  theoretisch auch die Rauschleistung geschätzt werden könnte. Dieses Vorgehen verspricht eine Verbesserung des lokal arbeitenden WIENER-Filters. Dieser Sachverhalt sollte in weiteren Arbeiten näher untersucht werden.

Es konnte experimentell gezeigt werden, daß durch eine Verbesserung des SNR, was die Grundlage der vorgestellten Strategie ist, ebenfalls eine deutliche Verbesserung in der Qualität von extrahierten Konturmerkmalen erzielt werden kann. Mit diesen Daten ergibt sich dann offensichtlich auch eine robustere Objekterkennung: Eine Kantendetektion ohne effektiv arbeitende, adaptive Rauschfilterung führt zu unbrauchbaren Ergebnissen und läßt keine robuste Repräsentation des Szeneninhaltes anhand der Konturmerkmale zu. Die vorgestellte Strategie verspricht daher nicht nur eine Verbesserung des SNR, sondern auch ein robustes Konturergebnis.

Wird die betrachtete Rauschfilterung mit nichtrekursiven Filtern implementiert, so ergeben sich sowohl im biologischen Vorbild als auch in technischen Systemen bei großen Filterstärken und dunklen Szenen ineffiziente Algorithmen. Zudem kann ein adaptives FIR-Filter nicht sehr gut als biologisches Neuronales Netz motiviert werden. Aus diesem Grund bieten sich gerade hier rekursive neuronale Strukturen bzw. rekursive Filter an. Bei diesen Strukturen ist der Aufwand an synaptischen Verbindungen bzw. an arithmetischen Operationen unabhängig von der Filterstärke. Deshalb ergeben sich hier wesentlich kürzere Berechnungszeiten auch im Vergleich zu dem WIENER-Filter (siehe Tabelle 6.1), was im Hinblick auf echtzeitfähige Anwendungen von großer Bedeutung ist.



# **Zusammenfassung und Ausblick**

Die Leistungsfähigkeit der visuellen Wahrnehmung des menschlichen Sehsystems ist technischen Ansätzen bislang noch weit überlegen. Ein Grund hierfür ist in der Komplexität biologischer Systeme zu suchen, deren Funktionsweise wir auch in naher Zukunft sicherlich noch nicht wirklich im Detail verstehen werden. Die Entwicklung von Modellen und Hypothesen biologischer Strukturen ist allerdings stets eines der ehrgeizigsten Bestrebungen aktueller Forschung, mit dem Zweck, unsere Kenntnisse über biologische Abläufe zu erweitern und zumindest Teile der überlegenen Strukturen für technische Aufgaben nutzbar zu machen. In der industriellen Bildverarbeitung spielen daher zunehmend biologisch motivierte Methoden eine bedeutende Rolle.

Ziel der vorliegenden Arbeit war eine systemtheoretische Beschreibung fundamentaler Verarbeitungspfade höher entwickelter biologischer Sehsysteme insbesondere unter dem Gesichtspunkt technischer Fragestellungen der Bildverarbeitung. In diesem Rahmen wurden die Eigenschaften des biologischen Vorbildes hinsichtlich der Extraktion von orientierten Konturmerkmalen herausgearbeitet. Als wesentliche Unterschiede zu bisherigen technischen Ansätzen sind die vorteilhafte hexagonale Bilderfassung und Verarbeitung, die vielversprechende neuronale Informationsverarbeitung durch pulscodierte Signale und wichtige adaptive Mechanismen zur robusten Rauschminderung bei extrem dunklen Szenen signaltheoretisch untersucht worden.

Auf einem biologisch und systemtheoretisch motivierten hexagonalen Abstrakter, welches viele Vorteile bietet, wurde eine plausible, massiv parallele Architektur der Signalverarbeitung vorgestellt, wie sie im menschlichen Sehsystem im Laufe der Evolution entstanden sein könnte. Das Modell ist durch eine kaskadier-

te Topologie gewisser Subfilter bzw. Neuronenschichten charakterisiert. Es konnte gezeigt werden, daß für die Realisierung von rezeptiven Feldern, die denen der Ganglienzellen in der Netzhaut und der simplen Zellen im visuellen Cortex entsprechen, lediglich nichtrekursive Neuronenverbindungen bzw. nichtrekursive Digitalfilter ausreichend sind. Die Parameter des Modells konnten entweder direkt mit der Hilfe von Ergebnissen physiologischer Experimente oder mittels eines Approximationsansatzes näherungsweise gefunden werden. Im Hinblick auf technische Sehsysteme bietet die präsentierte Architektur neben der überlegenen hexagonalen Datenrepräsentation eine Reihe nützlicher Eigenschaften, die im wesentlichen mit der kaskadierten Struktur zusammenhängen. Hierzu zählt zum einen die zusätzliche Möglichkeit der parallelen Abarbeitung der Daten bei Bildsequenzen, was direkt die hohe Effizienz des biologischen Systems widerspiegelt. Zum anderen werden nicht nur die eigentlichen Zielsignale (Konturmerkmale der simplen Zellen) durch das Netz berechnet, die Struktur bietet darüber hinaus den Zugang zu wertvollen Zwischensignalen, die, wie in der Biologie, für andere Bildverarbeitungsaufgaben weitergenutzt werden können. Als Beispiel sei hier das Signal der Ganglienzellen genannt, welches in der Natur u.a. für Blicksteuerungskonzepte eingesetzt wird. Wie in [Har82] gezeigt, sollten sich hieraus aber auch Auflösungspyramiden konstruieren lassen. Zwischensignale stellen daher einen Gewinn an Effizienz dar, da sie in der kaskadierten Struktur automatisch mitberechnet und genutzt werden können. Da die Signale der Ganglienzellen auf viele Arten simpler Zellen unterschiedlicher Orientierung und Selektivität konvergieren, birgt dies ebenfalls ein enormes Potential an Einsparungen von Neuronenverbindungen. Vergleicht man dies mit einer Implementation, bei der jeder Typ simpler Zellen für sich realisiert ist, so ist festzustellen, daß deutlich weniger arithmetische Operationen notwendig sind. Dies liegt darin begründet, daß in der Kaskade bereits berechnete Signale für aufbauende Verarbeitungsschritte vielfach benutzt werden können.

In einem weiteren Schritt konnte eine Strategie für die Erweiterung linearer Übertragungssysteme auf eine pulscodierte Verarbeitung entwickelt werden. Dadurch ist es möglich, bei Bedarf die Zeit als weitere Informationskomponente auch in das vorgestellte lineare Modell der frühen visuellen Verarbeitung einzubringen, was hierdurch natürlich an Biologienähe gewinnt. Da aber die Berechnung solch großer, massiv paralleler pulscodierter neuronaler Netze auf sequentiell arbeitenden Standardrechnern nur schwer möglich ist, wurde zudem das Netzwerk mit Erfolg auf einer in der Arbeitsgruppe entwickelten Neurocomputerarchitektur implementiert. Dieser ParSPIKE basiert auf parallel arbeitenden digitalen Signalprozessoren und verwendet das Modellneuron von ECKHORN. Die experimentellen Untersuchungen zeigten allerdings zunächst, daß durch die stets vorhande-

nen Quantisierungsmaßnahmen bei der digitalen Realisierung und den Einschränkungen der verwendeten Hardwarearchitektur die Antworten der simplen Zellen recht stark von den gewünschten GABOR-Antworten abwichen. Aus diesem Grund wurde das bekannte ECKHORN-Neuron für die vorliegende Implementation modifiziert und verbessert. Ein speziell entworfener Spike-Decoder, dessen Algorithmus als Tiefpaßfilter in Wellendigitalrealisierung interpretiert werden kann, konnte hierbei für diesen Zweck die numerischen Ungenauigkeiten minimieren. Darüber hinaus ließ sich auch die Einschwingzeit des Spike-Decoders, welche sich auf das dynamische Verhalten der Simulation und letztlich auch auf die gesamte Simulationsdauer niederschlägt, deutlich verkürzen. Hierdurch ist nun die effiziente Simulation eines komplexen Netzwerkes zur Verarbeitung visueller Daten möglich, welches ein gaborähnliches Verhalten auf einem hexagonalen Abtastraster mit einer pulscodierten Informationsverarbeitung kombiniert. Es steht also ein System zur Verfügung mit dem z.B. die für Bildverarbeitungsaufgaben interessanten Synchronisationseffekte simpler Zellen effizient studiert werden können.

Der dritte Schwerpunkt widmete sich der Rauschunterdrückung, welche stets als Vorstufe eines nachgeschalteten Systems zur Konturextraktion eingesetzt werden sollte, wie es auch in der bisherigen Modellierung stattgefunden hat. Um gegenüber Veränderungen der Beleuchtungsbedingungen einer betrachteten Szene robust zu sein, ist allerdings eine adaptive Rauschfilterung von großem Interesse. Die hierzu entwickelte Strategie basiert auf dem biologischen Mechanismus der Photorezeptoren, deren rezeptive Felder sich an die Lichtverhältnisse anpassen, um so bei dunklen Szenen, d. h. vermehrtem Rauschen eine stärkere Rauschfilterung durchzuführen. Theoretische Betrachtungen führten hier zu einem geeigneten und einfach zu implementierenden Kontrollmechanismus für ein gegebenes Rauschmodell. Trotz der notwendigen Näherungen in der theoretischen Herleitung und dem Einsatz einer globalen Filterung überzeugten dennoch die Resultate der experimentellen Untersuchungen mit realen Kameraszenen. Gerade bei extrem dunklen und damit extrem verrauschten Bilddaten versagen statt dessen etablierte Standardverfahren wie das adaptive WIENER-Filter. Allerdings scheint hier auch eine Kombination der entwickelten Strategie mit dem adaptiven WIENER-Filter denkbar, um diesen lokalen Ansatz zu verbessern, was in folgenden Arbeiten untersucht werden könnte. Für den Einsatz einer Konturextraktion in Kombination mit der entwickelten adaptiven Rauschfilterung wurden schließlich die Auswirkungen der Adaption auf das bisherige nichtadaptive Modell untersucht. Hier zeigte sich, daß die bisherigen kortikalen Neuronenverknüpfungen nicht ausreichen, um ein stabiles gaborähnliches Verhalten zu realisieren. Dieses Ergebnis rührt daher, daß die bisherige Modellierung für die kleinste Rezeptorkopplung

(und damit beste Ortsauflösung) entworfen und optimiert wurde. Durch einen alternativen Ansatz, der als eine Kaskadierung von Ableitungsoperatoren aufgebaut wurde, ist eine erfolgreiche adaptive Extraktion von Konturinformationen möglich, was durch die experimentellen Untersuchungen zusätzlich demonstriert werden konnte.

Die theoretischen und experimentellen Untersuchungen konnten die Leistungsfähigkeit der entwickelten Strategien zeigen. Damit stehen zum einen Hypothesen neuronaler Architekturen des biologischen visuellen Systems zur Verfügung, die evtl. einen Beitrag zum Verständnis dieser komplexen Materie liefern könnten. Darüber hinaus wurden einige wichtige Eigenschaften des biologischen Vorbildes aufgezeigt, die insbesondere für technische Umsetzungen von Interesse sein können. Im Rahmen dieser Arbeit ist eine biologienahe Strategie erarbeitet worden, die eine robuste Extraktion von Konturinformationen leistet. So beschäftigen sich derzeit laufende Arbeiten in der Arbeitsgruppe mit der Implementation der Verfahren auf einer mobilen Roboterplattform [Eis03]. Zum einen kann hiermit eine robuste Objekterkennung für autonome Einsätze in extrem dunklen Situationen, wie sie bei sicherheitsrelevanten Überwachungsaufgaben eintreten können, realisiert werden. Des weiteren soll ein telesensorischer Roboter (TSR) lediglich anhand der Konturinformationen in seinen Kamerabildern navigiert werden, welche für diesen Zweck über eine Funkstrecke (WLAN) übermittelt und der bedienenden Person visualisiert werden. Eine Reduzierung der Szene auf Konturinformationen erzielt dabei die notwendige Datenkompression für die Übertragungstrecke. Die vorgestellte adaptive Rauschunterdrückung sorgt dabei auch bei extrem dunklen bzw. verrauschten Situationen für eine robuste Extraktion der Konturen.



# Literaturverzeichnis

- [AKM95] Til Aach, Andre Kaup, and Rudolf Mester. On texture analysis: Local energy transforms versus quadrature filters. *Signal Processing*, 45:173–181, 1995.
- [BC92] R.A. Boie and I.J. Cox. An analysis of camera noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:671–674, June 1992.
- [Bie87] Irving Biedermann. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.
- [Bro79] J. L. Brown. Visual acuity and form discrimination. In R. E. Records, editor, *Physiology of the human eye and visual system*, chapter 15, pages 390–418. Harper & Row, 1979.
- [Bro99] I. N. Bronstein. *Taschenbuch der Mathematik*. Deutsch, 1999.
- [Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.
- [CBL89] S. Chen, S. A. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, 50(5):1873–1896, 1989.
- [CCG91] S. Chen, C. F. N. Cowan, and P. M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, 2(2):302–309, March 1991.

- [CG66] F. W. Campbell and R. W. Gubisch. Optical quality of the human eye. *The Journal of Physiology*, pages 558–578, 1966.
- [CW95] S. Chen and J. Wigger. Fast orthogonal least squares algorithm for efficient subset model selection. *IEEE Transactions on Signal Processing*, 43(7):1713–1715, July 1995.
- [Dau80] J.G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856, 1980.
- [Der93] R. Deriche. Recursively implementing the gaussian and its derivatives. Technical report, Institut National de Recherche en Informatique et en Automatique, April 1993.
- [DM84] D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice-Hall, 1984.
- [DS94] P. Deetjen and E. J. Speckmann. *Physiologie*. Urban & Schwarzenberg, 1994.
- [Eck94] R. Eckhorn. Oscillatory and non-oscillatory synchronizations in the visual cortex and their possible roles in associations of visual features. *Progress in Brain Research*, 102:405–426, 1994.
- [Eis01] Thomas Eisenbach. Untersuchungen zur adaptiven Signalverarbeitung verrauschter Bildszenen. Studienarbeit an der Universität Paderborn, 2001.
- [Eis03] Thomas Eisenbach. Entwicklung und Implementierung adaptiver Mechanismen zur Filterung verrauschter Bildszenen. Diplomarbeit an der Universität Paderborn, 2003.
- [EKGS90] A.K. Engel, P. König, C.M. Gray, and W. Singer. Stimulus-dependent neuronal oscillations in cat visual cortex: Inter-columnar interaction as determined by cross correlation analysis. *European Journal of Neuroscience*, 2:588–606, 1990.
- [EPdRH02] M. Egmont-Petersen, D. de Ridder, and H. Handels. Image processing with neural networks – a review. *Pattern Recognition*, 35:2279–2301, 2002.
- [ERAD89] R. Eckhorn, H.J. Reitböck, M. Arndt, and P. Dicke. Feature linking via stimulus-evoked oscillations: Experimental results for cat

- visual cortex and functional implications from a network model. In *Proc. of the International Joint Conference on Neural Networks (IJCNN 1989)*, volume I, pages 723–730, Washington D.C., USA, 1989.
- [ERAD90] R. Eckhorn, H.J. Reitböck, M. Arndt, and P. Dicke. Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Computations*, 2:293–307, 1990.
- [Fet86] Alfred Fettweis. Wave Digital Filters: Theory and Practice. *Proceedings of the IEEE*, 74(2):270–327, Februar 1986.
- [FG96] A.P. Fitz and R.J. Green. Fingerprint classification using a hexagonal fast fourier transform. *Pattern Recognition*, 29(10):1587–1597, 1996.
- [FHJS99] G. Frank, G. Hartmann, A. Jahnke, and M. Schäfer. An accelerator for neural networks with pulse-coded model neurons. *IEEE Trans. on Neural Networks, Special Issue on Pulse Coupled Neural Networks*, 10:527–539, 1999.
- [FOH94] Kunihiro Fukushima, Masata Okada, and Kazuhito Hiroshige. Neocognitron with dual c-cell layers. *Neural Networks*, 7(1):41–47, 1994.
- [Fra97] Gerald Frank. *Ein digitales Hardwaresystem zur echtzeitfähigen Simulation biologienaher neuronaler Netze*. PhD thesis, University of Paderborn, Germany, 1997.
- [FS70] A.S. French and R.B. Stein. A flexible analog using integrated circuits. *IEEE Transactions on Bio-Medical Engineering*, 17(3):248–253, 1970.
- [Fuk95] K. Fukushima. Neocognitron: A model for visual pattern recognition. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 613–617. MIT Press, Cambridge, Massachusetts, 1995.
- [Gab46] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946. London.
- [Gaz85] L. Gazsi. Explicit formulas for lattice wave digital filters. *IEEE Transactions on Circuits and Systems*, 32(1):68–88, January 1985.

- [GDH98] N. Götze, S. Drüe, and G. Hartmann. Erkennung und Lagebestimmung realer Objekte durch tolerante Repräsentation und modellbasiertes Gruppieren. In *Mustererkennung 1998*, volume 20, pages 580–587. Springer, 1998.
- [GF] R.J. Green and A. Fitz. A new approach to the analysis of fingerprints. Technical report, University of Bradford.
- [GH73] W. Gröbner and N. Hofreiter. *Integraltafel, Teil 2: Bestimmte Integrale*. Springer, 1973.
- [GM73] A.H. Gray and J.D. Markel. Digital lattice and ladder filter synthesis. *IEEE Transactions on Audio and Electroacoustics*, 21(6), 1973.
- [Göt00] Nicolai Götze. *Strategien zur ansichtenbasierten und ganzheitlichen invarianten Objekterkennung*. PhD thesis, University of Paderborn, Germany, 2000.
- [GVL96] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 1996.
- [Har82] G. Hartmann. Recursive features of circular receptive fields. *Biological Cybernetics*, 43:199–208, 1982.
- [Har91] G. Hartmann. Hierarchical neural representation by synchronized activity: a concept for visual pattern recognition. In J.G. Taylor, editor, *Neural Network Dynamics*, pages 356–370. Springer, 1991.
- [Hau94] G. Hauske. *Systemtheorie der visuellen Wahrnehmung*. Teubner, 1994.
- [Haw97] W.G. Hawkins. Fourier transform resampling: Theory and application. *IEEE Transactions on Nuclear Science*, 44(4):1543–1551, 1997.
- [HD90] G. Hartmann and S. Drüe. Self-organization of a network linking features by synchronization. In Eckmiller, Hartmann, and Hauske, editors, *Parallel Processing in Neural Systems and Computers*, pages 361–364. 1990.

- [HDD<sup>+</sup>94] G. Hartmann, S. Drüe, J. Dunker, K. O. Kräuter, B. Mertsching, and E. Seidenberg. The SENROB vision-system and its philosophy. In *Proc. of the 12th International Conference on Pattern Recognition (IAPR 1994)*, volume 2, pages 573–576, Jerusalem, Israel, 1994.
- [Hec98] E. Hecht. *Optics*. Addison-Wesley, 1998.
- [HFSW97] G. Hartmann, G. Frank, M. Schäfer, and C. Wolff. Spike128k - an accelerator for dynamic simulation of large pulse-coded networks. *Proc. of the 6th MicroNeuro*, pages 130–139, 1997.
- [HM87] J. Hirsch and W. H. Miller. Does cone positional disorder limit resolution. *Journal Optical Society America A*, 4(8):1481–1492, 1987.
- [HM98] S. He and D. I. A. Macleod. Local nonlinearity in S-cones and their estimated light-collecting apertures. *Vision Research*, 38(7):1001–1006, 1998.
- [Hub88] David Hubel. *Eye, Brain and Vision*. Freeman, 1988.
- [Hub90] David Hubel. *Auge und Gehirn*. Spektrum d. Wissenschaft, 1990.
- [HW62] D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.
- [HY94] I. Her and C.-T. Yuan. Resampling on a pseudo-hexagonal grid. *Graphical Models and Image Processing*, 56(4):336–347, July 1994.
- [ILGA98] I. Iglesias, N. Lopez-Gil, and P. Arttal. Reconstruction of the point-spread function of the human eye from two double-pass retinal images by phases-retrieval algorithms. *Journal Optical Society of America A*, 15(2):326–339, 1998.
- [Ima] The Imaging Source. *Users Manual DFM5303*.
- [JP87] Judson P. Jones and Larry A. Palmer. 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.

- [Kam] G. Kamberova. Understanding the systematic and random errors in video sensor data. Technical report, GRASP Laboratory, University of Pennsylvania.
- [KDB93] P. P. Kanjilal, P. K. Dey, and D. N. Banerjee. Reduced-size neural networks through singular value decomposition and subset selection. *Electronics Letters*, 29(17):1516–1518, August 1993.
- [KF88] M. V. Klein and T. E. Furtak. *Optik*. Springer, 1988.
- [KMB82] J. J. Kulikowski, S. Marcelja, and P. O. Bishop. Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics*, 41:187–198, 1982.
- [Köh90] M. Köhle. *Neuronale Netze*. Springer, 1990.
- [KvD87] J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.
- [KZ96] Klette and Zamperoni. *Handbook of Image Processing Operators*. Wiley, 1996.
- [Lee80] J. S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(3):165–168, 1980.
- [LF97] W. Li and A. Fettweis. Interpolation filters for 2-D hexagonally sampled signals. *International Journal of Circuit Theory and Applications*, 25:259–277, 1997.
- [Lim90] J. S. Lim. *Two-Dimensional Signal and Image Processing*. Prentice Hall, 1990.
- [LVB<sup>+</sup>93] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. v. d. Malsburg, R. P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.
- [Mar80] S. Marcelja. Mathematical description of the responses of simple cortical cells. *Journal Optical Society of America*, 70:1297–1300, 1980.

- [Mat] The Math Works Inc. *Image Processing Toolbox User's Guide (Matlab 5)*.
- [MDH94] B. Mertsching, S. Drüe, and G. Hartmann. Robot vision system learning and recognizing arbitrarily located objects from different camera positions. In *Computer Science and Technology*, volume 59, pages 121–130. Tanik, M. and Rossek, W. and Cooke, D.E., 1994.
- [MH80] D. Marr and E. Hildreth. Theory of edge detection. *Proc. of the Royal Society of London*, B 207:187–217, 1980.
- [Mil90] Alan Miller. *Subset selection in regression*. Chapman & Hall, 1990.
- [MV96] G. Murthy and Y. Venkatesh. Modified neocognitron for improved 2-D pattern. *IEE Proceedings on Vision, Image and Signal Processing*, 143(1):31–40, 1996.
- [NM79] M. Nagao and T. Matsuyama. Edge preserving smoothing. *Computer Graphics and Image Processing*, 9:394–407, 1979.
- [Ohm99] Jens-Rainer Ohm. *Skript zur Vorlesung: Bildverarbeitung II*. Technische Universität Berlin, Institut für Nachrichtentechnik und Theoretische Elektrotechnik, 1999.
- [PK00] S.C. Park and M.G. Kang. Noise-adaptive edge-preserving image restoration algorithm. *Optical Engineering*, 39:3124–3137, December 2000.
- [PM92] J.G. Proakis and D.G. Manolakis. *Digital Signal Processing*. Macmillan, 1992.
- [PR81] Daniel A. Pollen and Steven F. Ronner. Phase relationship between adjacent simple cells in the visual cortex. *Science*, 212:1409–1411, 1981.
- [PR82] D. A. Pollen and S. F. Ronner. Spatial computation performed by simple and complex cells in the visual cortex of the cat. *Vision Research*, 22:101–118, 1982.
- [PR83] Daniel A. Pollen and Steven F. Ronner. Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(5):907–916, 1983.

- [Sab96] Silvio P. Sabatini. Recurrent inhibition and clustered connectivity as a basis for gabor-like receptive fields in visual cortex. *Biological Cybernetics*, 74:189–202, 1996.
- [Sch00] Martin Schäfer. *Lernen in Neurocomputern für große pulscodierte neuronale Netze*. PhD thesis, University of Paderborn, Germany, 2000.
- [SL96a] Samir Shah and Martin D. Levine. Visual information processing in primate cone pathways - part i: A model. *IEEE Trans. on Systems, Man, and Cybernetics*, 26(2):259–274, April 1996.
- [SL96b] Samir Shah and Martin D. Levine. Visual information processing in primate cone pathways - part ii: Experiments. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(2):275–289, April 1996.
- [Sla90] Malcolm Slaughter. The vertebrate retina. In K. N. Leibovic, editor, *Science of Vision*, chapter 3, pages 53–83. Springer, 1990.
- [SM87] G. Sommer and G. Meinel. The design of optimal gaussian DOLP edge detectors. *Computer analysis of images and pattern*, pages 82–89, 1987.
- [SP94] A. Stacey and C. Pask. Spatial-frequency response of a photoreceptor and its wavelength dependence. i. coherent sources. *Journal Optical Society of America A*, 11(4):1193–1198, April 1994.
- [SS90] R. G. Smith and P. Sterling. Cone receptive field in cat retina computed from microcircuitry. *Visual Neuroscience*, 5:453–461, 1990.
- [SS97] A. Sarkar and K. M. S. Sharma. An approach to direct selection of best subset AR model. *Journal of Statistical Computation and Simulation*, 56:273–291, 1997.
- [ST97] R. F. Schmidt and G. Thews. *Physiologie des Menschen*. Springer, 1997.
- [Sta89] R.C. Staunton. Hexagonal image sampling: A practical proposition. In *Proceedings SPIE Conference: Expert Robots for Industrial Use*, volume 1008, pages 23–27, 1989.



- [Sta96] R.C. Staunton. An analysis of hexagonal thinning algorithms and skeletal shape representation. *Pattern Recognition*, 29(7):1131–1146, 1996.
- [STDH98] R. Stemmer, R. Trapp, S. Drüe, and G. Hartmann. 3D Lagebestimmung von zu demontierenden Objekten mit einem aktiven Stereokamerasystem. In *Mustererkennung 1998*, volume 20, pages 253–260. Springer, 1998.
- [SvdM97] L. Shams and C. von der Malsburg. Development of shape primitives from images of composite objects represented by complex cells. In *Proceedings of 7th International Conference on Artificial Neural Networks*, pages 895–900, Lausanne, Switzerland, October 1997.
- [TEDH02] J. Thiem, T. Eisenbach, S. Drüe, and G. Hartmann. Biology-inspired preprocessing of noisy visual signals. In *4th IASTED International Conference on Signal and Image Processing (SIP 2002)*, pages 294–299, Kauai, USA, August 2002.
- [TF97] M. Thomson and D. Foster. Role of second- and third- order statistics in the discriminability of natural images. *Journal of the Optical Society of America -A-*, 14(9):2081–2090, September 1997.
- [TH00] J. Thiem and G. Hartmann. Biology-inspired design of digital gabor filters upon a hexagonal sampling scheme. In *Proc. of the 15th International Conference on Pattern Recognition (ICPR 2000)*, volume 3, pages 449–452, Barcelona, Spain, September 2000.
- [Thi99] J. Thiem. Model of the human visual system for extraction of structural information. *7th Annual Report: Graduate College of the Heinz Nixdorf Institute, University of Paderborn*, November 1999.
- [Thi00] J. Thiem. Biology-inspired processing of pulscoded visual signals in cortical areas. *8th Annual Report: Graduate College of the Heinz Nixdorf Institute, University of Paderborn*, November 2000.
- [Thi01] J. Thiem. Biology-inspired adaptive preprocessing of noisy image data. *9th Annual Report: Graduate College of the Heinz Nixdorf Institute, University of Paderborn*, November 2001.

- [Tra98] Ralph Trapp. *Stereoskopische Korrespondenzbestimmung mit impliziter Detektion von Okklusionen*. PhD thesis, University of Paderborn, Germany, 1998.
- [TWCH91] C. L. M. Tiana, D. R. Williams, N. J. Coletta, and P. W. Haake. A model of aliasing in extrafoveal human vision. In M. S. Landy and J. A. Movshon, editors, *Computational models of visual processing*, chapter 3, pages 35–56. The MIT Press, 1991.
- [TWH00] J. Thiem, C. Wolff, and G. Hartmann. Biology-inspired early vision system for a spike processing neurocomputer. *Proc. of the 1st IEEE International Workshop on Biologically Motivated Computer Vision (BMCV 2000)*, 1:387–396, May 2000. Seoul, Korea.
- [VHV87] S. Van Huffel and J. Vandewalle. Subset selection using total least squares approach in collinearity problems with errors in the variables. *Linear Algebra and its Application*, 88/89:695–714, 1987.
- [vVYV98] L.J. van Vliet, I.T. Young, and P.W. Verbeek. Recursive gaussian derivative filters. In *Proc. of the 14th International Conference on Pattern Recognition (ICPR 1998)*, volume I, pages 509–514, Brisbane, Australia, August 1998.
- [Wah80] F. M. Wahl. *Der Entwurf zweidimensionaler rekursiver Filter und ihre Anwendung in der Digitalen Bildverarbeitung*. PhD thesis, University of Munich, Germany, 1980.
- [Wah89] F. M. Wahl. *Digitale Bildsignalverarbeitung*. Springer, 1989.
- [WC83] D. R. Williams and R. Collier. Consequences of spatial sampling by a human photoreceptor mosaic. *Science*, 221:385–387, 1983.
- [WC87] D. R. Williams and N. J. Coletta. Cone spacing and the visual resolution limit. *Journal Optical Society of America A*, 4(8):1514–1523, 1987.
- [WFKvdM97] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):774–779, 1997.
- [WHR99] C. Wolff, G. Hartmann, and U. Rückert. Parspike - A Parallel DSP-Accelerator for Dynamic Simulation of Large Spiking

- Neural Networks. In *Proc. of the 7th International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems*, pages 324–331, Granada, Spain, 1999.
- [Wil85] D. R. Williams. Aliasing in human foveal vision. *Vision Research*, 25(2):195–205, 1985.
- [Wil86] D. R. Williams. Seeing through the photoreceptor mosaic. *Trends in Neuroscience*, pages 193–198, May 1986.
- [Wil88] D. R. Williams. Topography of the foveal cone mosaic in the living human eye. *Vision Research*, 28(3):433–454, 1988.
- [Wis97] Laurenz Wiskott. Phantom faces for face analysis. *Pattern Recognition*, 30(6):837–846, 1997.
- [WM97] M. Webster and E. Miyahara. Contrast adaption and the spatial structure of natural images. *Journal of the Optical Society of America -A-*, 14(9):2355–2366, September 1997.
- [Wol01] Carsten Wolff. *Parallele Simulation großer pulscodierter neuronaler Netze*. PhD thesis, University of Paderborn, Germany, 2001.
- [WSM92] Z.W. Wang, J.J. Soltis, and W.C. Miller. Improved approach to interpolation using the fft. *Electronics Letters*, 28(25):2320–2322, 1992.
- [WW88] D. Wang and Q. Wang. A weighted averaging method for image smoothing. In *Proc. of the 8th International Conference on Pattern Recognition (ICPR 1988)*, pages 981–983, Paris, France, 1988.
- [Yel81] J. I. Yellot. Spectral consequences of photoreceptor sampling in the rhesus retina. *Science*, 212:382–385, 1981.
- [Zal91] Alejandro Zalnieriunas. *Adaptationsalgorithmen und Filterstrukturen zur Realisierung adaptiver Wellendigitalfilter*. PhD thesis, University of Bochum, Germany, 1991.