

**Extraktion und Separation  
von Konturen und Formen  
in realen Szenen  
mit dynamischen neuronalen Netzen**

Dissertation  
zur  
Erlangung des Doktorgrades  
der Naturwissenschaften  
(Dr. rer. nat.)

dem  
Fachbereich Physik  
der Philipps-Universität Marburg  
vorgelegt von

**Lothar Weitzel**

aus Marburg

Marburg/Lahn 1998

Vom Fachbereich Physik der Philipps-Universität Marburg als Dissertation am 30. Juli 1998  
angenommen.

Erstgutachter: Prof. Dr. Dr. H.J. Reitböck

Zweitgutachter: Prof. Dr. R. Eckhorn

Tag der mündlichen Prüfung: 16. September 1998



# Inhalt

<b>1. Einleitung</b>	<b>1</b>
<b>2. Grundlagen</b>	<b>5</b>
2.1 Das biologische Neuron	5
2.2 Visuelle Neuronenbahnen im Gehirn	6
2.3 Das verwendete Modellneuron	13
2.4 Das Acceleratorneuron	16
2.5 Objekterkennungsmodelle	17
2.5.1 Direkter Vergleich mit Eingangsbild	18
2.5.2 Vergleich selektierter Merkmale	18
2.5.3 Vergleich mit Merkmalen und ihren Relationen	20
2.6 Hinweise auf die Bedeutung von Vertices aus Wahrnehmungsexperimenten	22
<b>3. Bildvorverarbeitung</b>	<b>25</b>
3.1 Bildsequenzen	25
3.2 Modellierung der Photorezeptoren	27
3.3 Modellierung der X-Zellen	30
3.3.1 Räumliche Eigenschaften der Rezeptiven Felder	30
3.3.2 Zeitliche Eigenschaften der Rezeptiven Felder	31
3.4 Untersuchung von Eigenschaften der X-Zellen	33
3.4.1 Stationäres Antwortverhalten der Modell-X-Zelle	34
3.4.2 Transientes Antwortverhalten der Modell-X-Zelle	38
3.5 Diskussion	42
<b>4. Kontur/Form-System</b>	<b>43</b>
4.1 Kontrastkantendetektion	43
4.1.1 ON-OFF-Kantendetektor	47
4.1.2 ON-ON-Kantendetektor	53
4.1.3 ON <sup>2</sup> -OFF <sup>2</sup> -Kantendetektor	59
4.1.4 Eine Netzwerkstruktur zur Unterstützung von Kontur-Kontinuität und zur Erzeugung eindeutiger Antworten von Kantendetektoren	64
4.1.5 Polaritätsinvariante Modellneuronen	74
4.1.6 Diskussion	75
4.2 Detektion von Linienenden und Krümmungen	76
4.3 Detektion von Vertices	81
4.3.1 Repräsentationen von Vertices	81
4.3.2 Konturüberschneidungen zweier Objekte (T-Stücke)	85
4.3.3 Simulationen	85
4.4 Zusammenbinden und Separieren von Konturen im Zeitbereich	87
4.4.1 Zusammenbinden von Kontursegmenten	87
4.4.2 Separation von Konturen unterschiedlicher Objektkandidaten im Zeitbereich	91
4.4.3 Simulationsergebnisse	94
4.4.4 Diskussion	108

---

<b>5. Ein Texturanalyseverfahren basierend auf Orientierungscontrast und dessen Ankopplung an das Kontur/Form-System</b>	<b>113</b>
5.1 Detektion lokalen Orientierungscontrasts in künstlichen Texturen	114
5.1.1 Liniendetektion	114
5.1.2 Räumliche Tiefpaßfilterung	120
5.1.3 Orientierungscontrastdetektion	123
5.1.4 Liniendetektion auf Grundlage von Orientierungscontrast	127
5.1.5 Simulationsergebnisse	131
5.2 Einpassung des Texturmoduls in das Kontur/Form-System	132
5.2.1 Veränderungen im Netz	132
5.2.2 Simulationsergebnisse	133
5.3 Diskussion	136
<b>6. Abschließende Diskussion</b>	<b>139</b>
6.1 Allgemeine Diskussion	139
6.2 Biologische Relevanz	140
6.3 Technische Aspekte	142
6.4 Sonstige Einsetzbarkeit	144
6.5 Ausblick	144
<b>7. Zusammenfassung</b>	<b>146</b>
<b>8. Literaturverzeichnis</b>	<b>147</b>

## 1 Einleitung

Es ist bisher nicht gelungen, mit technischen Mustererkennungs- und Szenenanalysesystemen die Leistungsfähigkeit des menschlichen Sehsystems - und selbst des Sehsystems der meisten Tiere - bei der Erkennung von Objekten in einer natürlichen Szene zu erreichen oder gar zu übertreffen. Gerade Aufgaben, die der Mensch unbewußt und scheinbar mühelos bewältigt, bereiten künstlichen Systemen große Probleme.

Um nämlich ein Objekt erkennen zu können, benötigt eine Maschine eine geeignete Repräsentation, in der das Objekt zuvor vom Hintergrund getrennt wurde. Auf der anderen Seite muß das Objekt aber der Maschine bereits bekannt sein, um vom Hintergrund getrennt werden zu können. Bisher wurde keine Möglichkeit gefunden, aus diesem Dilemma der Figur/Hintergrundtrennung auszubrechen. Da aber die Figur/Hintergrundtrennung unabdingbare Voraussetzung für die Objekterkennung ist, besteht ein großes Interesse darin, zu wissen, wie das menschliche Sehsystem dieses Problem löst. Große Schwierigkeiten sind auch mit der Erkennung von Objekten verbunden, wenn sie in unterschiedlichen Größen, relativen Positionen oder Ansichten, besonders aber in verschiedenen Formen auftreten, die sich möglicherweise sogar dauernd verändern. Hier ist es notwendig, durch eine Reduktion der Objekte auf die wesentlichen Merkmale Invarianzgruppen zu bilden. Dabei ist insbesondere zu berücksichtigen, welche Merkmale zu einem Objekt gehören.

Wie kodiert nun das menschliche Sehsystem solche Zusammengehörigkeiten und wie führt es die Figur/Hintergrundtrennung durch? Einen Ansatz zur Klärung dieser Fragen liefert die Hypothese der zeitlich korrelierten Aktivität (Reitboeck 1980, Reitboeck 1983, von der Malsburg 1981, von der Malsburg und Schneider 1986), wonach Zusammengehörigkeit durch synchrones Auftreten neuronaler Aktionspotentiale kodiert wird. Diese Hypothese wurde später durch neurophysiologische Entdeckungen unterstützt (z.B. Eckhorn et al. 1988), die zeigten, daß Neurone, die ähnliche Merkmale kodieren, ihre Aktivitäten teilweise synchronisieren können. Als Folge dieser Korrelationshypothese läßt sich eine Objekt-Repräsentation aufbauen, indem die zum Objekt gehörenden charakteristischen Merkmale gleichzeitig repräsentiert werden. Auch die Repräsentation mehrerer Objekte oder Objektkandidaten ist durch Dekorrelierung der Aktivitäten von Neuronen, die an der Repräsentation verschiedener Objekte beteiligt sind, möglich. Eine Reihe theoretischer Studien, die sich mit der Verarbeitung von Eingangsinformation mit wenigen verschiedenen Amplituden (z.B. Eckhorn et al. 1990; von der Malsburg und Buhmann 1992) oder binärer Eingangsinformation (Ostkamp 1996; Ostkamp und Reitboeck 1994) beschäftigten, konnten zeigen, daß das Konzept der Korrelation erfolgreich anwendbar ist.

In der vorliegenden Arbeit wird das Problem der Figur/Grundtrennung auf der Grundlage von quasianaloger, aus natürlichen Szenen gewonnener Konturinformation behandelt. Ein aus dynamischen Neuronen aufgebautes Netz wird vorgestellt, das ausschließlich anhand dieser

Konturinformationen (also ohne Rückkopplungen von höheren Verarbeitungsstufen, in denen Objekte erkannt werden) Repräsentationen möglicher Objekte liefert (Weitzel et al. 1996, Weitzel et al. 1997a, Weitzel et al. 1997b). Dabei werden bereits Annahmen über wahrscheinliche Objektverdeckungen gemacht. In Anlehnung an die Korrelationshypothese wird Zusammengehörigkeit von Merkmalen eines Objektkandidaten durch (nahezu) gleichzeitige neuronale Aktivität definiert, und die Separation verschiedener Objektkandidaten geschieht durch deutlichen zeitlichen Versatz ihrer Repräsentationen. Eine Repräsentation beinhaltet nur noch die Information der "wesentlichen" Objektmerkmale, wie Konturen oder Ecken, die auch für eine spätere invariante Objektdefinition von Bedeutung sein könnte. Das aus mehreren Verarbeitungsstufen bestehende Netz trennt somit Objektkandidaten voneinander bzw. vom Hintergrund.

Kapitel 2 beinhaltet Beschreibungen des verwendeten Neuronenmodells und der sonstigen zugrundeliegenden Voraussetzungen. Insbesondere wird auch auf die besondere Bedeutung von Vertizes für das Sehen eingegangen. Als Grundlage für die Konturgewinnung aus Szenen werden in der Vorverarbeitungsstufe (Kapitel 3) Eingangsbilder hexagonal unterabgetastet, der lokale Intensitätskontrast durch eine "Mexikanerhut"-ähnliche Filterung extrahiert und als Impulssequenzen an die nachfolgenden Schichten zur Kantenextraktion weitergegeben (Kapitel 4.1). Die Neuronen der Kantendetektorschichten sind mit in Vorzugsrichtung koaxial angeordneten, benachbarten Neuronen wechselseitig gekoppelt, so daß ein Gruppierungsmechanismus zur verstärkenden Repräsentation ausgedehnter Kanten implementiert wird. Aus der Aktivität der Kantendetektoren werden in weiteren Verarbeitungsstufen Repräsentationen von Vertizes gewonnen (Kapitel 4.3), die für den Aufbau von Objektrepräsentationen erforderlich sind. Da hier die Vertizes durch das gleichzeitige Auftreten von mindestens zwei Linienenden in etwa demselben Ort definiert werden, müssen zuvor Repräsentationen von Linienenden erzeugt werden (Kapitel 4.2). Die Informationen der Vertizes lassen sich auch zur symbolischen Beschreibung von Objektkandidaten durch relationale Zeiger verwenden (Spengler 1996). Durch Wechselwirkungen zwischen den Neuronen der Vertexkodierung mit den Neuronen einer getrennten Kontur-Synchronisationsstufe wird korrelierte Aktivität zur Definition von Objektkandidaten (hypothetische Objekte) erzeugt (Kapitel 4.4). Eine inhibitorische Rückkopplung von der Synchronisationsstufe auf die Vertex-Repräsentationen ermöglicht schließlich die Separation dieser Objektkandidaten im Zeitbereich.

Das Netzwerk wurde im Rahmen des BMBF-Projektes "Elektronisches Auge" entwickelt, dessen Ziel die Implementierung eines aktiven künstlichen Sehsystems war. In Anlehnung an das Vorbild des menschlichen Sehsystems ist neben einem künstlichen *Transientensystem* (oder: *schnelles System*, Schott et al. 1996, Schott et al. 1997) und einem System zur *Blicksteuerung* (Pauly et al. 1997) mit dem vorliegenden Netz auch ein *Kontur/Form-System* (oder: *langsames System*) entwickelt worden. Die Realisierbarkeit eines solchen künstlichen Sehsystems wurde bereits anhand eines speziellen Hardwarebeschleunigers (Frank und

Hartmann 1995, Frank et al. 1998, Schaefer 1995, Wolff 1996) gezeigt, auf den das Kontur/Form-System erfolgreich portiert werden konnte.

In vielen komplexen Szenen reicht die Gewinnung von Konturinformation allein nicht aus, um Objekte "richtig" extrahieren zu können, und die Extraktion weiterer Merkmale, wie z.B. Textur, ist notwendig. In Kapitel 5 wird deshalb ein für bestimmte Texturarten geeignetes Analyseverfahren vorgestellt und das Kontur/Form-System um ein entsprechendes zusätzliches neuronales Modul erweitert. Hier soll vor allem ein Weg gezeigt werden, wie prinzipiell neben dem Intensitätskontrast weitere Merkmalskontraste im Kontur/Form-System verarbeitet werden können.





## 2 Grundlagen

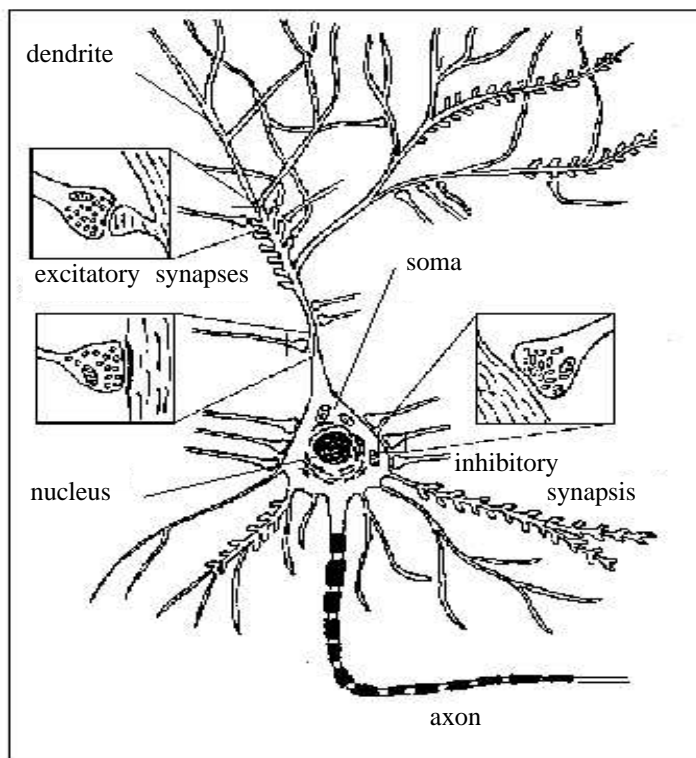
Wie bereits in Kapitel 1 erwähnt wurde, vermag das menschliche Sehsystem komplexe Szenen erfolgreich zu verarbeiten, bei denen künstliche Systeme bisher versagen. Daher liegt es nahe, die Funktionsprinzipien des Sehsystems zu untersuchen, um aus ihnen zu lernen und die wesentlichen Eigenschaften auf künstliche Systeme zu übertragen. Um die Prinzipien des Zentralnervensystems wenigstens teilweise im Großen verstehen zu können, muß zunächst das Gehirn im Kleinen untersucht werden. Nach der Darlegung der für diese Arbeit wichtigsten biologischen Grundlagen (Kapitel 2.1 und 2.2) wird das daraus abgeleitete und für den Aufbau des Netzes bestimmte Modellneuron beschrieben (Kapitel 2.3 und 2.4). Schließlich wird auch auf bekannte Modelle (Kapitel 2.5) und psychophysische Phänomene eingegangen, die für das Modell relevant sind (Kapitel 2.6).

### 2.1 Das biologische Neuron

Den elementaren Baustein des Gehirns stellt die Nervenzelle oder das Neuron dar. Über  $10^{11}$  Nervenzellen sind im Zentralnervensystem des Menschen enthalten und über etwa  $10^{14}$ - $10^{15}$  Synapsen miteinander verbunden (Mel 1997). Die Anzahl der Neuronen in einem künstlichen neuronalen Netz, wie es im folgenden vorgestellt wird, ist hingegen aus Gründen der Rechenkapazität um mehrere Größenordnungen kleiner.

Nervenzellen lassen sich in drei wesentliche Bereiche (Kompartimente) unterteilen: Dendrit, Soma (Zellkörper) und Axon (vgl. Abb. 2.1.1). Der Dendrit entspricht dem Eingangsbereich der Nervenzelle. Über seine vielen Verzweigungen empfängt er von zahlreichen Neuronen Signale, die zeitlich und räumlich integriert werden und das Potential der Zellkörpermembran beeinflussen. Diese Membran umgibt den Zellkörper, so daß Innen- und Außenraum voneinander getrennt werden. Sowohl intra- als auch extrazellulär liegen ionenhaltige Flüssigkeiten vor. Während das Außenmedium vorwiegend Natrium- ( $\text{Na}^+$ ) und Chlorionen ( $\text{Cl}^-$ ) enthält, überwiegen innerhalb der Zelle die Kaliumionen ( $\text{K}^+$ ). Durch einen aktiven Transportvorgang, in dem dauernd Natriumionen und Kaliumionen aus der bzw. in die Nervenzelle transportiert werden, entsteht ein (Membran-) Ruhepotential von etwa -60 bis -80 mV zwischen Innen- und Außenraum. Die über den Dendritenbaum erhaltenen Signale können die Membranspannung beeinflussen. Überschreitet die Membranspannung rasch einen bestimmten Wert (Schwellenpotential), so erhöht sich die Permeabilität der Membran für Natriumionen. Natriumionen gelangen durch die Membran in den Innenraum, wodurch die Membranspannung stark ansteigt (Depolarisation) und einen Wert von etwa 20-30 mV (innen) erreicht. Dieser etwa 1 ms dauernde elektrische Impuls wird Aktionspotential oder Spike genannt. Zeitlich etwas verzögert wird der Kaliumionenstrom in den Außenraum verstärkt,

während die Membran für Natriumionen undurchlässiger wird. Das Membranpotential kehrt dadurch rasch wieder zum Ruhewert zurück, kurzzeitig wird er sogar unterschritten (Hyperpolarisation). Solange die Permeabilität der Membran für Kaliumionen erhöht ist, ist die Nervenzelle kaum erregbar, d.h. refraktär. Der hier beschriebene Vorgang der Entstehung eines Aktionspotentials wird auch als Hodgkin-Huxley-Zyklus bezeichnet. Das so generierte Aktionspotential wandert vom Ort seiner Entstehung über das Axon bis hin zu den Synapsen oder Endknöpfchen, wo es die Ausschüttung von Neurotransmittern in den jeweiligen synaptischen Spalt bewirkt.



**Abb. 2.1.1:** Schematischer Aufbau einer Nervenzelle. Dargestellt sind die Dendriten (dendrites), Soma, Zellkern (nucleus), das Axon und hemmende und erregende Synapsen (inhibitory and excitatory synapses) einer Nervenzelle. Nach: Zell 1994.

Die Neurotransmitter bewirken eine Verringerung oder eine Erhöhung des Membranpotentials der postsynaptischen Zellen, je nachdem, ob eine inhibitorisch wirkende (hemmende) oder eine exzitatorisch wirkende (erregende) Synapse vorliegt, wodurch das entsprechende Signal weitergeleitet wird.

## 2.2 Visuelle Neuronenbahnen im Gehirn

In diesem Abschnitt soll kurz auf den Aufbau der ersten Verarbeitungsstufen des visuellen Systems eingegangen werden. Dabei werden diejenigen Aspekte hervorgehoben, die für das nachfolgende Modell von besonderer Bedeutung sind. Ein Schema des menschlichen Sehsystems zeigt die Abbildung 2.2.1.

Die Signale der lichtempfindlichen Netzhaut (Retina) des Auges werden über den Sehnerv (Nervus opticus) durch die Sehnervenkreuzung (Chiasma opticum) an verschiedene Schaltstellen weitergeleitet. Als wichtigste Schaltstellen sind die beiden Kniehöcker (Corpus geniculatum laterale, CGL) zu nennen. Die Lichtreize des rechten Gesichtsfeldes verursachen eine Signalausbreitung von beiden Augen an den Kniehöcker der linken Hirnhälfte und die Lichtreize des linken Gesichtsfeldes eine Signalausbreitung von beiden Augen an den Kniehöcker der rechten Hirnhälfte. Von den Kniehöckern werden die Signale schließlich zum visuellen Kortex weitergeleitet. Auf der Netzhaut befinden sich mehr als 100 Millionen Photorezeptorzellen. Sie antworten auf einfallende Lichtquanten mit Hyperpolarisation ihrer Membranspannung. Dabei liegt näherungsweise ein logarithmischer Zusammenhang zwischen Lichtintensität  $I$  und Rezeptorpotential  $P$  vor (Schmitt 1985):

$$P = k \cdot \log \frac{I}{I_s} \quad (\text{Gl.2.1.1})$$

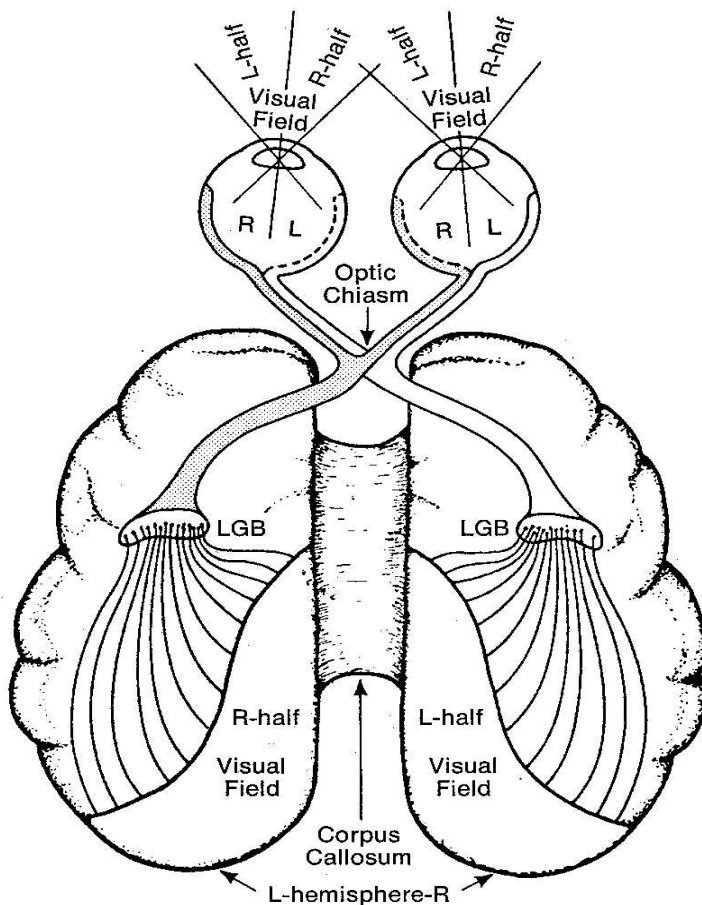
wobei  $I_s$ : Schwellenintensität.

Dieser Zusammenhang entspricht einem von mehreren Adaptationsmechanismen, die im Verlauf der Evolution entwickelt wurden, um das visuelle System an sehr verschiedene Helligkeiten anpassen zu können, denn die mittlere Beleuchtungsstärke der natürlichen Umwelt des Menschen schwankt bereits um einen Faktor von mehr als  $10^8$ . Es gibt zwei grundsätzlich verschiedene Rezeptortypen, die Stäbchen und Zapfen genannt werden. Während die etwa 6 Millionen Zapfen über die gesamte Netzhaut, jedoch mit hoher Konzentration in der Fovea, verteilt sind, befinden sich die etwa 120 Millionen Stäbchen nur außerhalb der Fovea, d.h. in der Peripherie. Grundsätzlich sind Zapfen für das Farbsehen und das Sehen bei Tageslicht, Stäbchen hingegen für das Dämmerungssehen und periphere Bewegungen zuständig. Ein weiterer wichtiger Unterschied ist der Grad der Konvergenz in den jeweils nachfolgenden Verschaltungen. Die Verschaltungen der Stäbchen sind deutlich konvergenter als die der Zapfen. Zapfen ermöglichen somit die höhere Sehschärfe.

Die Rezeptorzellen sind lateral mit Horizontalzellen verbunden und zusammen mit ihnen auf nachfolgende Bipolarzellen verschaltet. Dabei haben die Vorwärtsverbindungen der Horizontalzellen jedoch inhibitorische Wirkung, so daß die resultierende Verschaltung einen Antagonismus zwischen den Teilbereichen Zentrum (Center) und Peripherie (Surround) des rezeptiven Feldes einer Bipolarzelle bewirkt. Unter einem rezeptiven Feld wird der Bereich derjenigen Rezeptoren verstanden, die die Antwort des Neurons bei Stimulierung beeinflussen. Es ergeben sich zwei unterschiedliche Typen von Bipolarzellen, zum einen Neuronen mit exzitatorisch wirkendem Zentrum und inhibitorisch wirkender Peripherie des rezeptiven Feldes (ON-Center-Neuronen) und zum anderen, der inverse Fall (OFF-Center-Neuronen). Kuffler fand bereits 1953 bei Untersuchungen an Katzen, daß diese zwei

verschiedenen Arten rezeptiver Felder existieren (Kuffler 1953). Die Bipolarzellen sind lateral mit den Amakrinzellen verbunden. Die Amakrinzellen-Verschaltungen weisen dabei große Ähnlichkeiten zu den Horizontalzellen-Verschaltungen auf, so daß die Amakrinzellen möglicherweise eine ähnliche Rolle wie Horizontalzellen spielen (z.B. Kolb und Nelson 1993). Bipolarzellen geben ihre Signale außerdem an Ganglienzellen weiter.

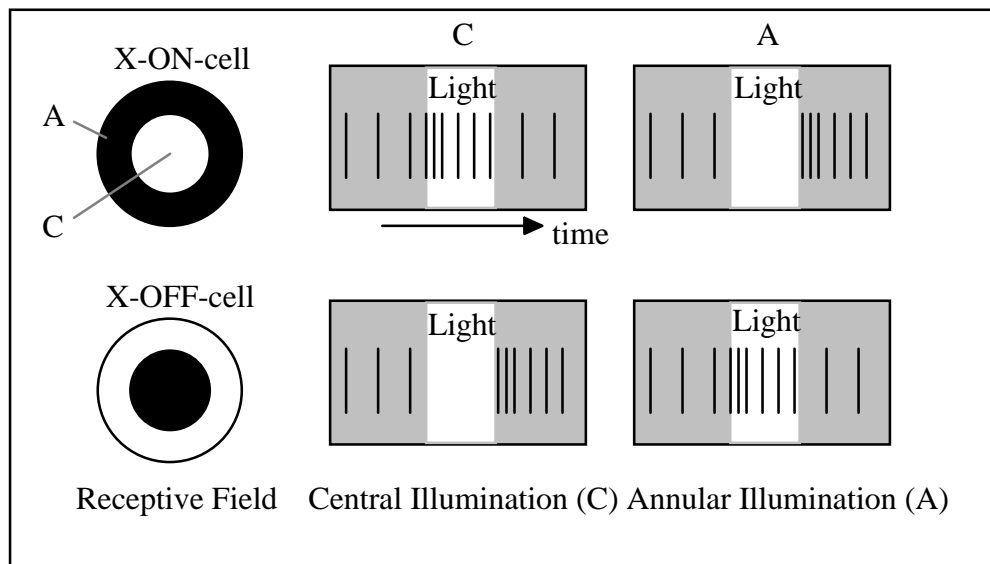
Es gibt viele verschiedene Arten von Ganglienzellen in der Netzhaut. Den größten Anteil haben die X-Zellen (Die Bezeichnung stammt aus der beim Sehsystem der Katze üblichen Terminologie). Sie besitzen, wie die Bipolarzellen, rezeptive Felder mit Center-Surround Struktur. Wird das exzitatorische Zentrum einer X-ON-Zelle beleuchtet (Einschaltvorgang), so erhöht sich die Rate ihrer Aktionspotentiale. Bei Beleuchtung der inhibitorischen Peripherie wird die Rate kleiner bzw. bleibt Null.



**Abb. 2.2.1:** Schema des menschlichen visuellen Systems. Die Lichtreize des linken Gesichtsfeldes verursachen eine Signalausbreitung von beiden Augen über die Sehnervenkreuzung (optic chiasm) an den Kniehöcker (LGB) der rechten Hirnhälfte und die Lichtreize des rechten Gesichtsfeldes eine Signalausbreitung von beiden Augen über die Sehnervenkreuzung an den Kniehöcker der linken Hirnhälfte. (Aus: Popper und Eccles 1977).

Wird das inhibitorische Zentrum einer X-OFF-Zelle verdunkelt (Ausschaltvorgang), so erhöht sich die Rate ihrer Aktionspotentiale. Bei Verdunklung der exzitatorischen Peripherie wird die Rate kleiner bzw. bleibt Null (z.B. Nicholls et al. 1992). Werden Zentrum und Peripherie gleichzeitig stimuliert, so antwortet die Zelle nicht oder nur schwach (z.B. Goldstein 1989). X-Zellen reagieren auf einen geeigneten Stimulus mit kurzzeitig stark erhöhter

Rate (transient response) und anschließender weniger starker, aber gleichbleibender Aktivität (sustained response). Abbildung 2.2.2 veranschaulicht einige dieser Zusammenhänge.



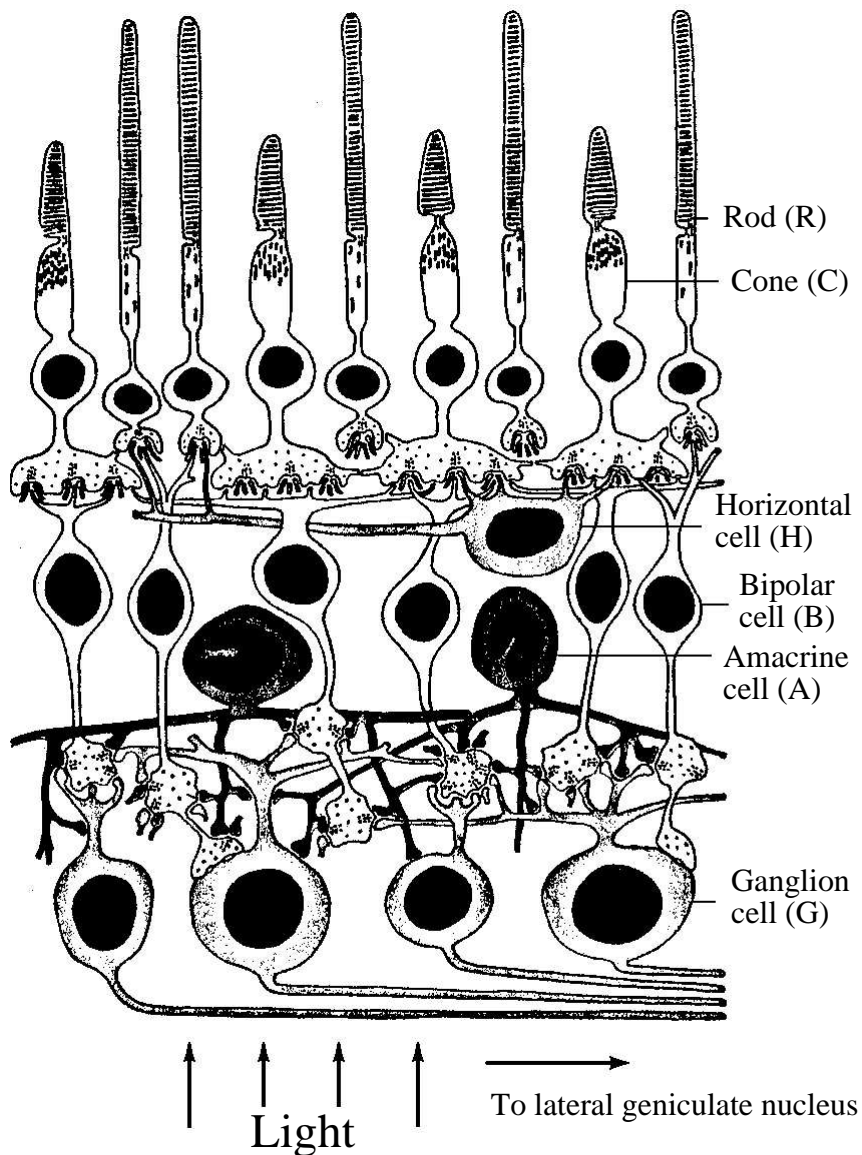
**Abb. 2.2.2:** Antwortverhalten von retinalen X-Zellen bei unterschiedlichen Reizbedingungen. C: Reizung des Zentrums, A: Reizung der Peripherie. (nach Nicholls et al. 1992)

Die Y-Zelle ist ein weiterer retinaler Ganglienzellentyp, dessen rezeptives Feld Center/Surround-Struktur zeigt. Sie antwortet auf gleichzeitige Stimulierung von Zentrum und Peripherie stärker als die X-Zelle und reagiert auf einen geeigneten Stimulus lediglich mit transienter Antwort.

Bei gleichem Stimulus ist die transiente Antwort einer Y-Zelle stets sehr viel stärker als die transiente Antwort einer X-Zelle. Axon, Zellkörper und rezeptives Feld der X-Zelle sind kleiner als die der Y-Zelle, außerdem hat das Axon einer Y-Zelle eine höhere Leitungsgeschwindigkeit als das einer X-Zelle. X-Zellen finden sich vor allem in der Fovea, Y-Zellen hingegen verstärkt in der Peripherie der Netzhaut. Dabei nimmt die Größe der rezeptiven Felder von innen nach außen zu. Die W-Zellen bilden eine weitere Ganglienzellenart. Einige von ihnen weisen keine Center/Surround-Struktur auf und könnten daher zur Ermittlung der mittleren Lichtintensität dienen. Die Abbildung 2.2.3 veranschaulicht den Aufbau der Retina.

Die Ausgänge der Ganglienzellen projizieren über den optischen Nerv auf die beiden CGL, die bei den Primaten jeweils aus sechs neuronalen Schichten bestehen. Zwei dieser Schichten (Schicht 1 und 2) bestehen aus großen Zellen und erhalten hauptsächlich Signale von den Y-Zellen (magnozellulärer Pfad, die Bezeichnungen "X-" und "Y-Zellen" werden der Einfachheit halber hier beibehalten), während die übrigen Schichten aus kleinen Zellen bestehen und vorwiegend Signale von X-Zellen erhalten (parvozellulärer Pfad). In beiden Fällen sind die rezeptiven Felder der CGL-Zellen denen der Ganglienzellen sehr ähnlich.

Das CGL projiziert auf die primären Areale, vor allem auf Schicht IV in V1, des visuellen Kortex. Die rezeptiven Felder der Zellen im visuellen Kortex unterliegen nicht mehr einer einfachen Center/Surround-Struktur, wie sie bei den Ganglienzellen in der Retina oder bei den Zellen im CGL gefunden wurde. Die kortikalen Zellen antworten bevorzugt auf komplexere Reize, wie zum Beispiel orientierte Balkenreize. Hubel und Wiesel entdeckten bei Untersuchungen an der Großhirnrinde der Katze orientierungselektive kortikale Zellen und unterschieden diese nach ihrem Antwortverhalten in Simple- und Komplex-Zellen. Die Simple-Zelle stellt den einfachsten Zelltyp dar.



**Abb. 2.2.3:** Schnitt durch die menschliche Retina. **G:** Ganglienzelle, **A:** Amakrinzelle, **B:** Bipolarzelle, **H:** Horizontalzelle, **C:** Zapfen, **R:** Stäbchen. (Nach: Dowling und Boycott 1966, Brause 1991)

Sie antwortet auf einen Balkenreiz oder eine Hell/Dunkel-Kante mit bestimmter Orientierung und bestimmter relativer Lage bzgl. ihres rezeptiven Feldes. Die dazu notwendige Richtungs-

selektivität erreicht eine Simple-Zelle durch ein länglich ausgeprägtes rezeptives Feld, das aus einer exzitatorischen "ON"- und einer inhibitorischen "OFF"-Region besteht. Diese beiden Regionen entstehen durch konvergente Verschaltung von kollinear angeordneten CGL-Zellen mit sich teilweise überdeckenden rezeptiven Feldern (Hubel und Wiesel 1962, 1965). Heggelund fand bei Zelleitungen an der Katze, daß eine Simple-Zelle sowohl exzitatorische als auch inhibitorische Signale ausschließlich von einem CGL-Zellentyp (entweder ON-Center-Zelle oder OFF-Center-Zelle) erhält (Heggelund 1981a). Neben der Beschaffenheit der rezeptiven Felder haben auch exzitatorische und inhibitorische Verbindungen der Simple-Zellen untereinander großen Einfluß auf die Richtungsselektivität (z.B. Maex und Orban 1996).

Eine Komplex-Zelle antwortet, wie eine Simple-Zelle, optimal auf einen Balkenreiz oder eine Kante mit bestimmter Orientierung. Die Antwort der Komplex-Zelle ist aber unabhängiger von der räumlichen Phasenlage der Kante im rezeptiven Feld. Die Komplex-Zelle erhält wahrscheinlich exzitatorische und inhibitorische Signale von beiden Zelltypen (von ON-Center-Zellen als auch von OFF-Center-Zellen; Heggelund 1981b). Sie besitzt zudem ein größeres rezeptives Feld und ein robusteres Antwortverhalten auf Rauschen als eine Simple-Zelle (Heggelund 1981b). Die rezeptiven Felder der Komplex-Zellen lassen sich meistens aus denen der Simple-Zellen aufbauen, d.h. es liegt eine hierarchische Struktur zugrunde.

Diese setzt sich fort in den Hyperkomplex- oder End-Stopped-Zellen. Solche Zellen mit endinhibierten rezeptiven Feldern antworten optimal auf Stimuli bestimmter Länge und können somit als Detektoren für Linien- oder Kantenenden und auch als Krümmungsdetektoren fungieren (Peterhans et al. 1986, Peterhans und von der Heydt 1989). Hubel und Wiesel fanden darüberhinaus auch Neuronen, die bevorzugt auf Ecken oder Winkel ansprechen, und sie entdeckten die sogenannten Kolumnen, innerhalb derer alle Zellen etwa dieselbe Vorzugsorientierung aufweisen und die Komplexität der rezeptiven Felder dieser Zellen senkrecht zur Kortexoberfläche sukzessive zunimmt (Hubel und Wiesel 1965). Andere Untersuchungen zeigten aber, daß sich längs einer Kolumne beim Übergang von niedrigeren zu höheren Schichten die Vorzugsorientierung auch ändern kann und daß in den unteren Schichten 5 und 6 eine andere Häufigkeitsverteilung der Vorzugsorientierung vorliegt (z.B. Bauer und Dow 1989).

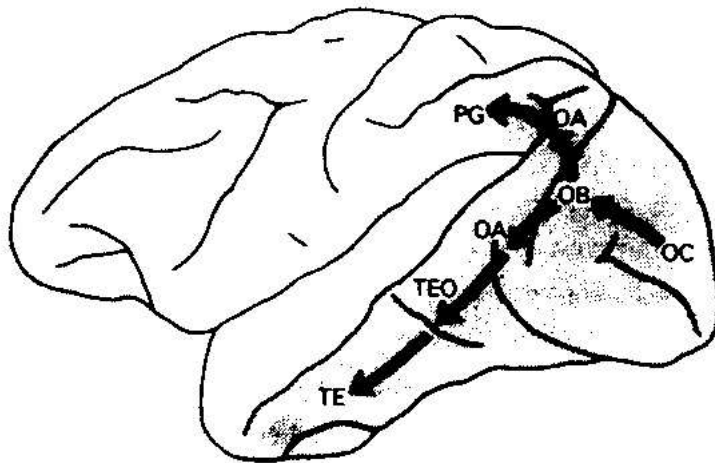
Ungerleider und Mishkin schlugen 1982 eine Einteilung des visuellen Kortex in zwei relativ unabhängige Pfade vor, zum einen in den ventralen Pfad, der ausgehend vom primären visuellen Kortex (V1) über weitere Areale (V2 und V4) läuft und im inferioren temporalen Kortex (IT) endet, und zum anderen in den dorsalen Pfad, der von V1 ausgehend auf den inferioren parietalen Lappen projiziert (Ungerleider und Mishkin 1982, Abbildung 2.2.4 zeigt eine schematische Darstellung der beiden Pfade). Dabei erhält jedes der vorwärtsprojizierenden Areale auch Rückkopplungsverbindungen (z.B. Distler et al. 1993). Die beiden Pfade lassen sich auch funktionell unterteilen. Ungerleider und Mishkin gingen zunächst davon aus, daß der ventrale Pfad für das Objektsehen zuständig ist. Viele Untersuchungen unterstützen



diese Annahme. Beispielsweise wurden Musterseparations- und Erkennungsleistungen von Affen nach Entfernung des IT stark beeinträchtigt (Gross 1973). Der dorsale Pfad ist nach Ungerleider und Mishkin für das Raumsehen zuständig. Eine solche Trennung in Objektsehen und Raumsehen erscheint jedoch problematisch (Goodale 1993, Milner und Goodale 1993). Milner und Goodale folgerten hingegen aus weiteren Untersuchungen, daß dorsaler und ventraler Pfad nicht unabhängig voneinander funktionieren. Jeder Pfad nutzt demnach die visuelle Information (wie Orientierung, Kontur und möglicherweise räumliche Relationen) auf verschiedene Weise. Während der ventrale Pfad Botschaften andauernder Charakteristiken visueller Objekte liefern kann, ist der dorsale Pfad mit der instantanen Charakteristik konfrontiert und mit der visuomotorischen Kontrolle gekoppelt. Diese Hypothese wird auch durch die folgenden Untersuchungsergebnisse unterstützt.

1. Der für das "stationäre Sehen" zuständige parvozelluläre Pfad (Die Bezeichnung stammt aus der beim Sehsystem des Affen üblichen Terminologie) bleibt auch noch in V1 von dem für "bewegtes Sehen" zuständigen magnozellulären Pfad getrennt und verläuft anschließend vorwiegend ventral, während der magnozelluläre Pfad vorwiegend mit dem dorsalen Pfad gekoppelt ist.

2. Affen, deren hinterer parietaler Lappen entfernt wurde, wiesen keine Leistungsmängel in Mustererkennungsaufgaben auf. Jedoch zeigten sie Störungen beim Greifen nach Objekten.



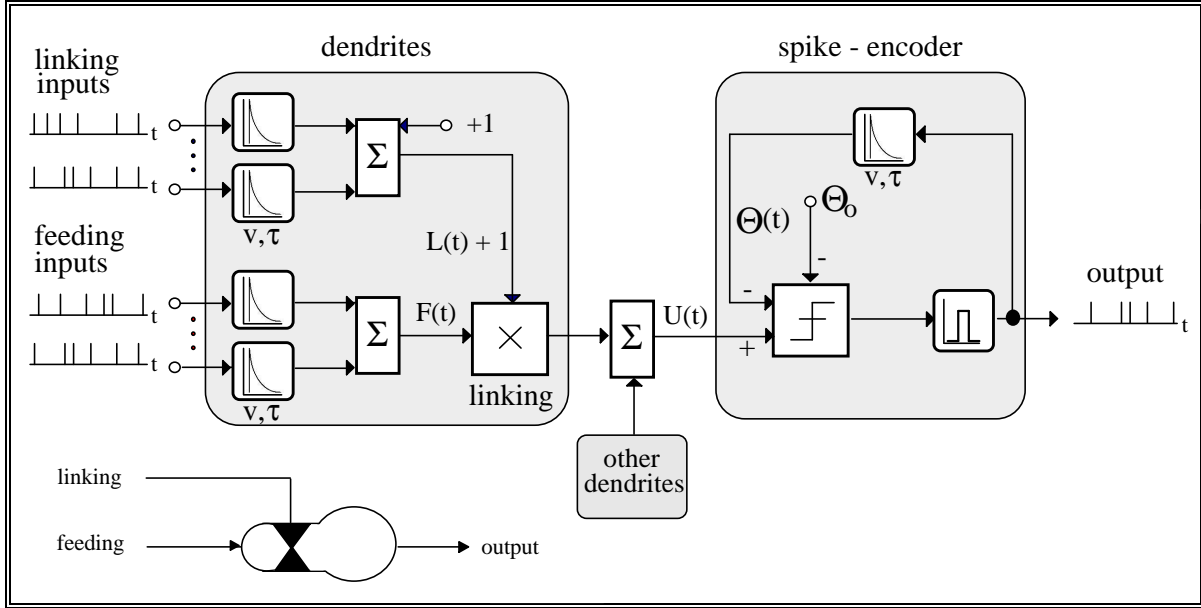
**Abb. 2.2.4:** Schematische Darstellung der beiden kortikalen visuellen Systeme nach Ungerleider und Mishkin (1982). Der dorsale Pfad projiziert vom primären visuellen Kortex (OC) auf den posterior parietalen Lappen (PG), während der ventrale Pfad vom primären visuellen Kortex auf den inferior temporalen Kortex (TE) projiziert (aus: Milner und Goodale 1993).

IT wird in das hintere Areal TEO und das vordere Areal TE unterteilt. Die Projektionsbahnen von V4 auf TEO sind dichter als die auf TE. IT projiziert auf viele Strukturen außerhalb des visuellen Kortex, wobei die Projektionen von TE zahlreicher sind als die von Arealen früherer Stufen. Deshalb läßt sich hier von einem sequentiellen Pfad von V1 zu TE sprechen, deren Ausgänge sich vorwiegend in TE befinden (Tanaka 1993). Auch in IT gibt es starke Hinweise auf funktionelle Unterschiede. Läsionen in Area TEO führten beim Affen zu einfachen Musterdefiziten, während Läsionen in Area TE zu Defiziten im assoziativen und visuellen Gedächtnis führten (Iwai 1978, 1981, 1985, Logothetis und Sheinberg 1996). Viele IT-Neuronen sind selektiv für Farbe, Orientierung, Textur, Bewegungsrichtung oder Kontur.

Konturselektivität ist zwar auch von früheren Arealen wie V4 bekannt (Desimone und Schein 1987), in IT finden sich jedoch erstmals Neuronen, die verstärkt selektiv für Kontur sind (Logothetis und Sheinberg 1996). Area TE ist, ähnlich wie V1, aus kolumnenartigen Modulen zusammengesetzt, in denen Zellen mit überlappender, aber wenig unterschiedlicher Selektivität in Clustern zusammengefaßt sind (Fujita et al. 1992). Diese Kolumnenorganisation weist darauf hin, daß ein Objektmerkmal durch Aktivität vieler Zellen eines einzelnen kolumnären Moduls repräsentiert wird. Eine solche Kodierung ist robust gegen kleine Veränderungen im Eingangsbild, und sie erreicht eine präzise Repräsentation. Die Kolumnenorganisation könnte für überlappende und möglicherweise kontinuierliche Repräsentationen von Objektmerkmalen sorgen (Tanaka 1993). In TE wird weiterhin ein gewisser Grad an Translations- und Größeninvarianz ermöglicht (Tanaka 1993, Logothetis und Sheinberg 1996). Die rezeptiven Felder der TE-Zellen sind viel größer als die der TEO-Zellen oder die von V4-Zellen, und die TEO-Zellen sind im Gegensatz zu den TE-Zellen retinotop angeordnet (Boussaoudd et al. 1991). Das führt zu der Annahme, daß in TEO zunächst die Selektivität für Stimuli konstruiert wird und anschließend in TE die Invarianzen erzeugt werden. Auf der Grundlage dieser Aktivitäten könnten schließlich weitere Operationen durchgeführt werden (Tanaka 1993), wie z.B. Vergleich der Objektrepräsentation mit gespeicherten Objektansichten (Ullman und Basri 1991). In einer ähnlichen Hypothese wird angenommen, daß das Erkennen visueller Objekte auf mindestens zwei verschiedenen Typen gespeicherter Repräsentationen beruht. Zum einen werden demnach grundsätzliche Objektformen durch neuronale Aktivität in verschiedenen Modulen von IT kodiert (strukturelle Primitive und ihren räumlichen Relationen) und zum anderen werden Objekte auf einer untergeordneten Stufe durch Kombinationen verschiedener Ansichten repräsentiert (Logothetis und Sheinberg 1996). Letzteres wird auch durch die Entdeckung einzelner IT-Zellen (in wachen Affen) unterstützt, die ansichtsabhängig auf spezifische Objekte antworten (z.B. Hasselmo et al. 1989, Logothetis und Pauls 1995).

### 2.3 Das verwendete Modellneuron

Entsprechend der zugrundeliegenden Arbeitshypothese der korrelierten Aktivitäten (vgl. Kap. 1) sind Synchronisationsvorgänge von besonderer Bedeutung für die neuronale Verarbeitung von Reizen. An ein künstliches Neuron, das zur Modellierung solcher Synchronisationsvorgänge verwendet werden soll, sind jedoch gewisse Anforderungen zu stellen. Zum einen sollte es nur zu festen Zeitpunkten Aktionspotentiale generieren können, um eine genaue Definition von Gleichzeitigkeit zu ermöglichen. Zum anderen muß das Modellneuron Eingänge besitzen, die den Aufbau rezeptiver Felder ermöglichen. Schließlich muß auch eine Möglichkeit zur (meist lateralen) Kopplung von Neuronen untereinander gegeben sein, so daß sich die neuronalen Ausgangsaktivitäten synchronisieren können. Diesen Anforderungen wird das biologienahe Marburger Modellneuron (Abb. 2.3.1, Eckhorn et al. 1990) gerecht.



**Abb. 2.3.1:** Das Marburger Modellneuron (Eckhorn et al. 1990). Dargestellt sind der dendritische Bereich und die zur Generierung von Aktionspotentialen verwendete dynamische Schwelle (Spike-Encoder). **Links unten:** Symbol des Marburger Modellneurons.

Hierbei handelt es sich um ein nichtlineares System, das als ein um den modulatorischen *Linking*-Eingang erweitertes "neural analog" (French und Stein 1970) angesehen werden kann. Für den hierarchischen Aufbau rezeptiver Felder besitzt das Modellneuron den *Feeding*-Eingang. Beide Typen synaptischer Eingänge werden durch Leckintegratoren (zeitliche Tiefpässe erster Ordnung) modelliert. Die postsynaptischen Potentiale  $F(t)$  (Feeding-Potential) und  $L(t)$  (Linking-Potential) ergeben sich aus den Faltungen der Eingangsimpulsfolgen  $x^F(t)$  und  $x^L(t)$  mit den Stoßantworten  $h^F(t, t^F, V^F)$  und  $h^L(t, t^L, V^L)$  der entsprechenden Leckintegratoren zu:

$$F(t) = x^F(t) * h^F(t, \tau^F, V^F), \quad (\text{Gl. 2.3.1})$$

$$L(t) = x^L(t) * h^L(t, \tau^L, V^L), \quad (\text{Gl. 2.3.2})$$

wobei

$$h^F(t, \tau^F, V^F) = \begin{cases} V^F \cdot \exp(-t/\tau^F), & \text{falls } t \geq 0 \\ 0, & \text{sonst} \end{cases}, \quad (\text{Gl. 2.3.3})$$

$$h^L(t, \tau^L, V^L) = \begin{cases} V^L \cdot \exp(-t/\tau^L), & \text{falls } t \geq 0 \\ 0, & \text{sonst} \end{cases}. \quad (\text{Gl. 2.3.4})$$

Dabei werden die Eigenschaften der Leckintegratoren durch die Verstärkungsfaktoren  $V^F$  und  $V^L$  und die Zeitkonstanten  $t^F$  und  $t^L$  bestimmt. Solche räumlichen und zeitlichen Integrationen

wurden z.B. von Foerster 1966 bereits modelliert. Feeding- und Linking-Eingang besitzen unterschiedliche Funktionalität. Ein Linking-Signal hat ohne ein gleichzeitiges Feeding-Signal keinen Einfluß auf das Membranpotential, wohingegen ein Feeding-Signal allein sehr wohl zu einem von Null verschiedenen Membranpotential führt. Das Linking-Potential hat nur modulatorischen Einfluß auf das Feeding-Potential. Diese Eigenschaften werden erreicht, indem sich das *Membranpotential*  $U(t)$  aus der Multiplikation von Feeding- und Linking-Potential unter Berücksichtigung eines zusätzlichen Offsets von +1 ergibt zu:

$$U(t) = \sum_{d=1}^D F_d(t) \cdot (1 + L_d(t)), \quad (\text{Gl. 2.3.5})$$

mit  $D$  : Anzahl der Dendriten.

Für die Erzeugung eines Ausgangssignals  $y(t)$  wird das Membranpotential in jedem Simulationszeitschritt (1 BIN = 1 ms) mit einer dynamischen Schwelle  $\Theta(t)$  zuzüglich eines Schwellenoffsets  $\Theta_0$  verglichen. Übersteigt das Membranpotential die Schwelle, so generiert das Modellneuron ein Aktionspotential und erhöht die Schwelle um den Wert des Verstärkungsfaktors  $V^\Theta$ . Die Ausgangsaktivität läßt sich somit beschreiben durch:

$$y(t) = \begin{cases} 1, & \text{falls } U(t) \geq \Theta(t) + \Theta_0 \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 2.3.6})$$

Die dynamische Schwelle  $\Theta(t)$  wird durch einen Leckintegrator ( $V^\Theta, \tau^\Theta$ ) modelliert und ergibt sich zu:

$$\Theta(t) = y(t) * h^\Theta(t, \tau^\Theta, V^\Theta), \quad (\text{Gl. 2.3.7})$$

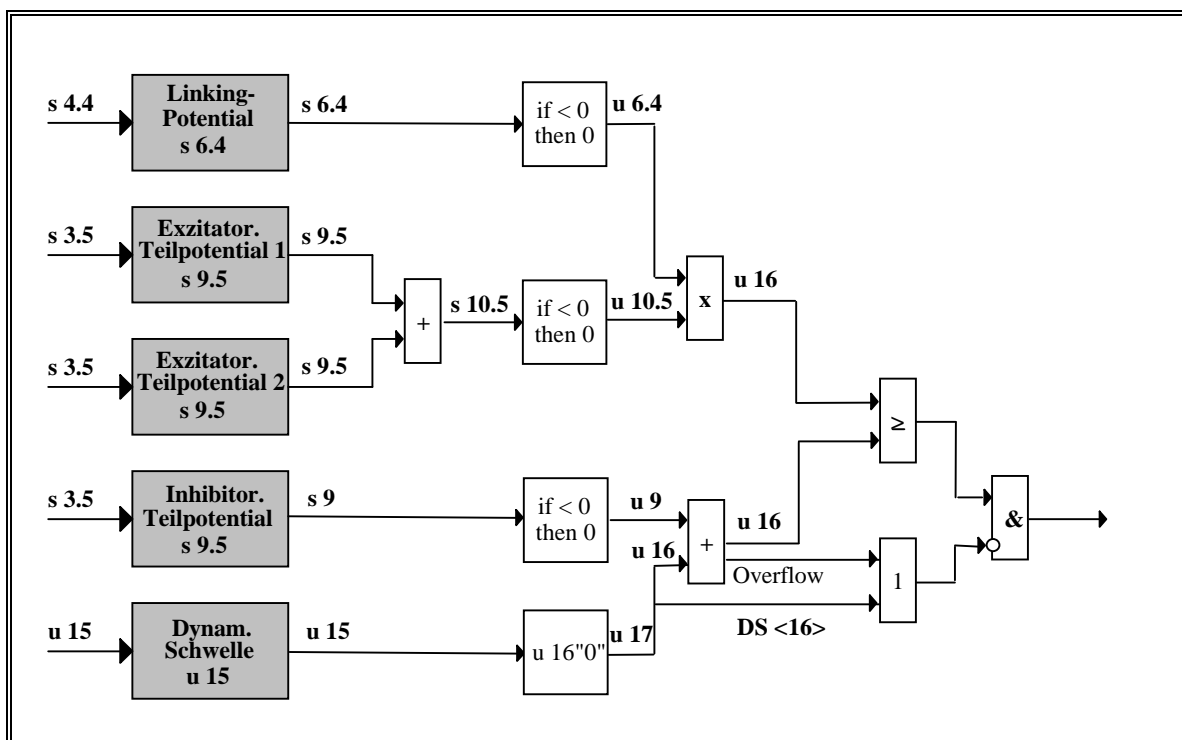
mit:

$$h^\Theta(t, \tau^\Theta, V^\Theta) = \begin{cases} V^\Theta \cdot \exp(-t/\tau^\Theta), & \text{falls } t \geq 0 \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 2.3.8})$$

Der hier gezeigte Schwellenmechanismus erlaubt die Simulation der relativen Refraktärzeit biologischer Neurone, da durch das Heraufsetzen der dynamischen Schwelle nach der Generierung eines Aktionspotentials die erneute Generierung für eine gewisse Zeit erschwert wird.

## 2.4 Das Acceleratorneuron

In der Einleitung wurde bereits darauf hingewiesen, daß für die Netze des angestrebten künstlichen Sehsystems auch eine spezielle Hardware (Accelerator) zur beschleunigten Verarbeitung entwickelt wurde, wobei die Entwicklungsarbeiten noch nicht abgeschlossen sind (Wolff 1998). Hardware und Software wurden zeitlich parallel entwickelt, was die Verwendung eines Simulators während der Entwicklung der Netze erforderlich machte (Spengler 1995, Gabriel 1995, Schott 1995, Möller 1995). Um eine besonders effiziente Verarbeitung des Accelerators gewährleisten zu können, werden einige Modifikationen am Marburger Modellneuron vorgenommen, die jedoch die grundsätzlichen Eigenschaften des Neurons nicht beeinträchtigen (Abb. 2.4.1). Der dendritische Bereich des Acceleratorneurons wird auf vier mögliche Teilpotentiale beschränkt, so daß zwei Feeding-, ein Linking- und ein inhibitorischer Eingang zur Verfügung stehen, deren Potentialverläufe durch Abklingtabellen realisiert werden. Eine weitere Einschränkung besteht in der Berechnung des Membranpotentials auf der Grundlage von 16 Bit-Operationen. Demzufolge ergeben sich für das Membranpotential sowie für die einzelnen Teilpotentiale maximale Wortbreiten, die kleiner oder gleich 16 Bit sind.



**Abb. 2.4.1:** Berechnung des Membranpotentials und Generierung von Aktionspotentialen im Acceleratorneuron. Neben der dynamischen Schwelle sind auch die übrigen Teilpotentiale mit ihren Auflösungen und erlaubten Gewichtungen dargestellt. Weiter werden die Verschaltung zur Berechnung des Membranpotentials und der Vergleich mit der dynamischen Schwelle gezeigt, die durch eine vorzeichenlose Fixpunktdarstellung mit 15 Vorkommastellen ( $u_{15}$ ) repräsentiert wird. Nach: Frank und Hartmann 1995.

Beispielsweise wird das Linking-Potential durch eine vorzeichenbehaftete Fix-Punkt-Darstellung mit sechs Vorkomma- und vier Nachkommastellen (s6.4) repräsentiert. Je mehr Nachkommastellen auftreten, desto kleinere Schritte können für die Dekrementierung in der Abklingphase verwendet werden, was wiederum zu einer größeren Empfindlichkeit gegenüber Gewichts- oder Signaländerungen führt. Für die Berechnung des Feeding- und des inhibitorischen Potentials stehen mehr Vor- und Nachkommastellen zur Verfügung. In Abbildung 2.4.1 sind die einzelnen Teilpotentiale und ihre Verschaltung dargestellt. Die beiden exzitatorischen Teilpotentiale werden addiert. Nur positive Teilpotentiale werden weiterverrechnet, um ausschließen zu können, daß bei gleichzeitig negativem Linking-Potential ein sehr großes, positives Membranpotential auftritt. Das Membranpotential ergibt sich auch hier gemäß Gleichung 2.3.5. Durch Abklingen des Linking-Potentials auf den Wert 1 wird die modulatorische Wirkung des Linking-Potentials auf die Feeding-Potentiale erhalten. Eine Option des Accelerators erlaubt es, daß das Linking-Potential, wie die anderen Teilpotentiale auch, auf den Wert 0 abgeklungen werden kann, wodurch eine "echte" Multiplikation von Feeding- und Linking-Potentialen erreicht wird. Das Acceleratorneuron fungiert dann als eine Art zeitliches AND-Gatter. Für bestimmte Anwendungszwecke erweist sich diese Option als nützlich (siehe Kapitel 4). Auch die dynamische Schwelle wird durch eine Abklingtabelle realisiert, die auf den Wert des Schwellenoffsets abzuklingen ist. Das Schwelleninkrement kann nur den Wert eines ganzzahligen Vielfachen von 16 annehmen. Dynamische Schwelle und inhibitorisches Teilpotential werden addiert und mit dem Membranpotential verglichen. Ist das Membranpotential größer als die Summe aus Schwellen- und inhibitorischem Potential, dann wird, falls kein Überlauf aufgetreten ist, ein Aktionspotential generiert. Eine detaillierte Beschreibung des Accelerators ist in Frank 1997 zu finden.

## 2.5 Objekterkennungsmodelle

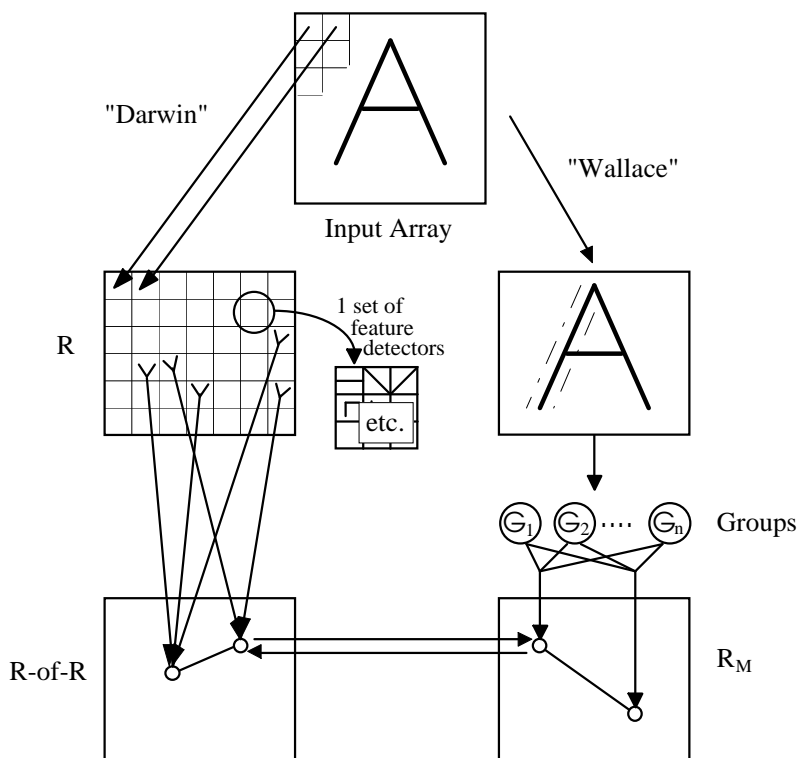
Um ein im Eingangsbild vorliegendes Objekt als solches wiederzuerkennen oder einer bestimmten Objektklasse zuzuordnen, vergleicht ein künstliches Erkennungssystem eine geeignete Repräsentation des Eingangsbildes mit bereits vorhandenen Repräsentationen bekannter Objekte. Eine Vielzahl von Modellen der Objekt- und Mustererkennung wurden entwickelt, die sich vor allem in der Art der aufzubauenden Repräsentationen unterscheiden. Den einfachsten Typ stellen Modelle dar, die das (nahezu) unveränderte Eingangsbild mit entsprechenden internen Repräsentationen vergleichen. In anderen Modellen wird die Erkennung von Objekten anhand bestimmter extrahierter Merkmale ermöglicht und in einer dritten Gruppe werden schließlich neben spezifischen Merkmalen auch deren räumliche Relationen zueinander berücksichtigt.

### 2.5.1 Direkter Vergleich mit Eingangsbild (Template Matching)

Ein relativ unverändertes Eingangsbild wird mit Repräsentationen einer bestimmten Ansicht eines Objekts verglichen. Das Objekt wurde erfolgreich erkannt, wenn eine Repräsentation gefunden werden konnte, die dem Objekt im Eingangsbild innerhalb gewisser Fehlergrenzen entspricht. Dieses Verfahren erfordert entweder für jede spezifische Ansicht eine eigene gespeicherte Repräsentation oder aber Möglichkeiten, spezifische interne Ansichten mit weniger Speicheraufwand gewinnen zu können (z.B. Ullman 1989, Ullman und Basri 1991, Bühlhoff et al 1995, siehe auch Kapitel 2.5.3). In jedem Fall müssen aber große Datenmengen bewältigt werden, da das vollständige Eingangsbild mit aufwendigen Repräsentationen verschiedener Objekte verglichen werden muß, insbesondere, wenn natürliche Szenen verarbeitet werden sollen. Daher bietet sich Datenreduktion an, wie sie in der folgenden Gruppe von Modellen durch Merkmalsextraktion erreicht wird.

### 2.5.2 Vergleich selektierter Merkmale

Zunächst werden bestimmte Merkmale aus dem Eingangsbild gewonnen, wie z.B. Konturen, Vertizes, Oberflächeneigenschaften wie z. B. Textur oder Farbe. Jedes Objekt wird durch einen oft translationsinvarianten Satz solcher visueller Merkmale charakterisiert (z.B. Selfridge und Neisser 1960). Der Erkennungsautomat *Darwin II* (Reeke et al. 1990) stellt ein neueres Beispiel für solche Mustererkennungsverfahren dar. Er besteht aus zwei parallel arbeitenden zweistufigen Pfaden, deren zweite Stufen miteinander wechselseitig verbunden sind (*reentry*, Sporns et al. 1991). Abbildung 2.5.1 zeigt schematisch den Aufbau des Systems. Der erste Pfad ("Darwin") antwortet auf Kombinationen lokaler Merkmale, die den individuellen Stimulus definieren. Dazu werden in einer ersten Schicht orientierte Liniensegmente, Linienenden und Ecken (Winkel) extrahiert. Die zweite Schicht erhält zahlreiche Verbindungen von zufällig ausgewählten Einheiten der ersten Schicht und antwortet daher nichtlinear auf nichtlokale Kombinationen der lokalen Merkmale. Der zweite Pfad ("Wallace") besitzt dieselbe Struktur wie der erste Pfad, jedoch werden hier in der ersten Schicht Merkmalskorrelationen einer gröberen Auflösung ermittelt, wobei die Merkmale relativ invariant gegen Translation und Rotation des Stimulus sind (z.B. Konturecken und T-Stücke, siehe Kap. 2.6). Somit antwortet der zweite Pfad auf Merkmalskorrelationen, die Klasseneigenschaften des Stimulus definieren. Die Antworten der wechselwirkenden Teilsysteme werden schließlich zur Klassifikation von Objekten herangezogen. Das Modell wurde jedoch vornehmlich für theoretische Zwecke konzipiert und verarbeitet daher nur relativ einfache künstliche Szenen.



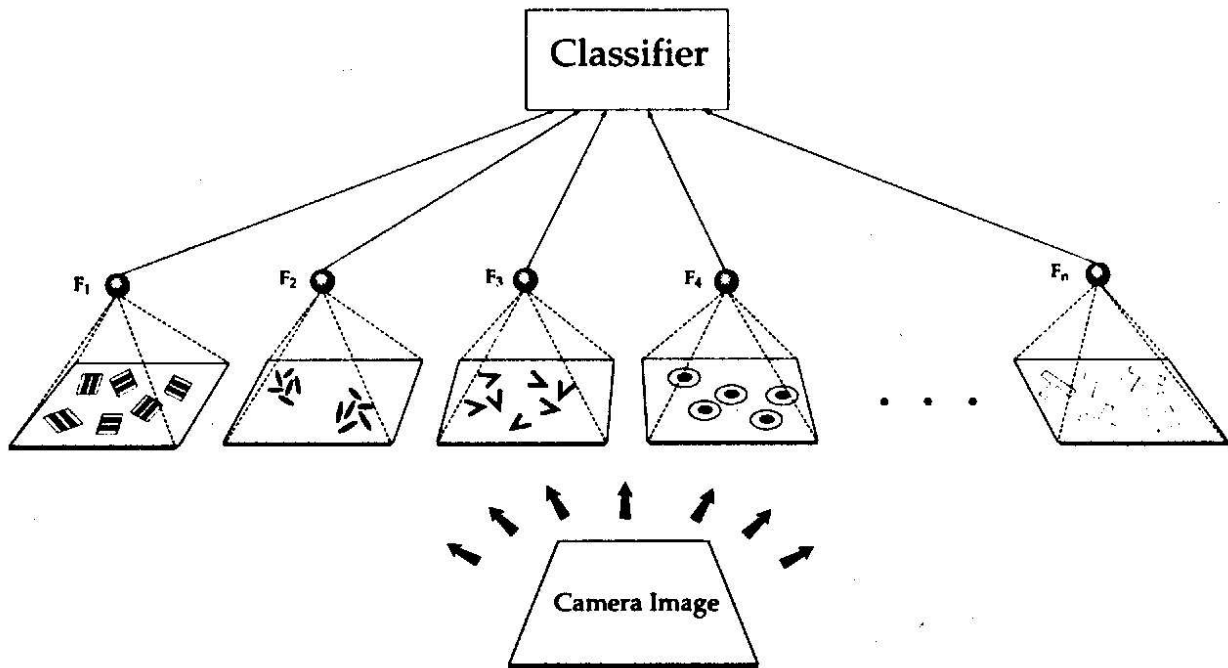
**Abb. 2.5.1:** Vereinfachte Darstellung der Konstruktion von Darwin II. Darwin II besteht aus zwei interagierenden Pfaden, "Darwin" und "Wallace". Während Darwin Kombinationen lokaler Merkmale bestimmt, antwortet Wallace auf Merkmalskorrelationen. Die auf Grundlage extrahierter Merkmale gewonnene Gesamtantwort des Systems wird schließlich zur Klassifikation von Objekten herangezogen (nach: Reeke et al. 1990).

Das Modell *Chorus* von Sh. Edelman und Duvdevani-Bar (1997) berechnet für ein aus einer beliebigen Ansicht gezeigtes Objekt einen Punkt oder Vektor in einem Merkmalsraum anhand detektierter Merkmale. Nach einer Transformation dieses Vektors in einen Ähnlichkeitsraum wird der diesem Vektor ähnlichste Vektor eines Prototyps einer Objektklasse ermittelt. Auf diese Weise können Objekte (wieder-)erkannt oder kategorisiert werden, d.h. sowohl bekannte als auch neue Objekte werden sinnvoll behandelt. Das Modell konnte anhand einer begrenzten Anzahl einzelner als 3D-Computergraphiken implementierter Objekte erfolgreich getestet werden.

Ein Beispiel für Modelle, die sich mit der Verarbeitung realer Objekte beschäftigen, ist *SEEMORE* (Mel 1997). Die visuelle Repräsentation wird von über 100 Merkmalskanälen generiert, die durch lokale Filterung die Merkmale Kontur, Farbe, Ecken, Blobs und Textur (Gaborfilter) aus einem von einer Kamera gelieferten Eingangsbild extrahieren. Die in Histogrammen zusammengefaßten Merkmale ermöglichen die Zuordnung des dargebotenen Objekts zu einem der gespeicherten Objekte anhand eines neuronalen Klassifizierers. Abbildung 2.5.2 zeigt eine schematische Übersicht des Modells. Das System wurde mit einem auf 100 echte Objekte unterschiedlichen Typs beschränkten visuellen Bereich trainiert und anschließend auch mit in gewissen Graden abweichenden Ansichten und Größen dieser Objekte getestet. Die Klassifikationsleistung lag dabei bei 97%. Obwohl das ausschließlich mit Vorwärtsverbindungen konstruierte System sehr erfolgreich einzelne Objekte auch unter erschwerten Bedingungen wie Größen- und Ansichtsveränderungen wiedererkennen kann, so wird doch das Problem der Verarbeitung mehrerer gleichzeitig im Bild enthaltener Objekte, d.h. der Figur/Hintergrundtrennung, nicht behandelt. Räumliche Invarianzen, wie sie durch



Verwendung solcher Merkmalsätze gewonnen werden, sind zudem zwangsläufig mit dem Verlust räumlicher Information verbunden, die aber weiterhin benötigt wird. Einen möglichen Ausweg bietet die nun folgende Gruppe von Modellen, die mehrere Objekte oder Objektteile gleichzeitig verarbeiten und auch die räumlichen Relationen der Merkmale untereinander berücksichtigen.



**Abb.: 2.5.2:** Architektur des Objekterkennungsmodells SEEMORE. Das Modell besteht aus einer Reihe von Merkmalskanälen, deren Antworten zur Klassifikation von Objekten verwendet werden. Die Antworten werden durch Vektoren  $\mathbf{F}$  repräsentiert, die relativ invariant gegen Rotationen und Translationen im Bildbereich sind, aber empfindlich gegen Änderungen der Objektqualität, d.h. Objektidentität (aus: Mel 1997).

### 2.5.3 Vergleich mit Merkmalen und ihren Relationen

Die bisher beschriebenen Mustererkennungsverfahren arbeiten auf einer Stufe, in der das zu erkennende Objekt bereits von allen anderen Strukturen derselben Szene getrennt vorliegt, so daß nur dieses einzelne Objekt behandelt werden muß. Wurde diese Vorarbeit der Figur/Hintergrundtrennung jedoch noch nicht geleistet, so kann ein Erkennungssystem auch mit mehreren Objekten oder Objektteilen gleichzeitig konfrontiert werden, die in räumlichen Beziehungen zueinander stehen. In den entsprechenden Modellen werden Objekte als Konfigurationen von Attributen mit spezifischen Relationen zueinander repräsentiert. Strukturelle Beschreibungen geben diese Relationen explizit an (z.B. Winston 1975, Hummel und Biederman 1992).

Im Modell von Hummel und Biederman wird die 3D-Struktur eines visuellen Eingangsbildes rekonstruiert und dann an eine gespeicherte Repräsentation angepaßt. Ein

Objekt wird durch eine strukturelle Beschreibung repräsentiert, die die Objektteile und die Beziehungen dieser Teile untereinander spezifiziert. Bei dieser relationalen Beschreibung wird das Objekt in seine *Geonen* zerlegt ("geon structural description") und die Kodierung der Relationen mehrerer Geone zueinander definiert. Synchrone neuronale Aktivität definiert dabei die Zusammengehörigkeit. Linienenden und Vertizes, die sich aus Linienzeichnungen komplexer Objekte extrahieren lassen, spielen eine entscheidende Rolle. Ein Vertex wird als Diskontinuität einer Objektkontur verstanden, die sich aus dem Zusammentreffen von zwei oder drei Linienenden unterschiedlicher Orientierung ergibt. Je nach Zusammensetzung aus Linienenden bestimmter ansichtsspezifischer Orientierung lassen sich Vertizes in verschiedene Klassen einteilen. Abbildung 2.5.1 zeigt eine Auswahl möglicher Vertizes.

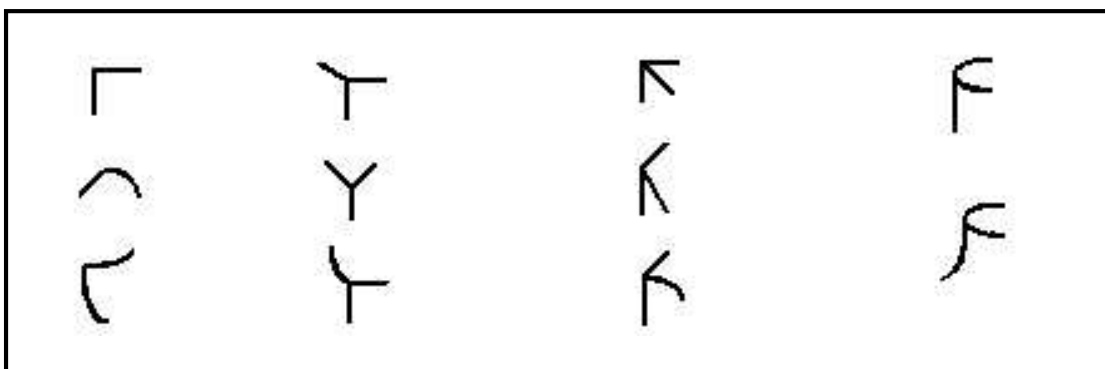


Abb. 2.5.1: Beispiele klassifizierter Vertizes. Aus: Hummel und Biederman 1992.

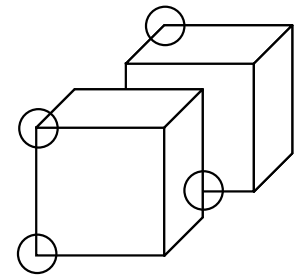
Die von diesem Modell geleistete strukturelle Beschreibung führt zu dreidimensionalen und ansichtsinvarianten Repräsentationen der 3D-Objekte. In anderen Theorien wird die Repräsentation dreidimensionaler Objekte als ansichtsspezifische 3D-Repräsentation erklärt, d.h. 3D-Objekte lassen sich durch ihre 2D-Bilder repräsentieren (Poggio und Edelman 1990). Das Modell von Ullman ("aligned approach", Ullman 1989, Ullman und Basri 1991) vergleicht Merkmale des 2D-Eingangsbildes mit denen der Projektion eines gespeicherten 3D-Modells. Tatsächlich deuten psychophysische Experimente stark auf eine Standpunktabhängigkeit der Repräsentation und Erkennung von Objekten hin (Bülthoff et. al 1995), was der Theorie der Ansichtsinvarianz widerspricht. Bülthoff, Edelman und Tarr erklären die Repräsentation von 3D-Objekten hingegen als ansichtsabhängige 2D-Repräsentation. 3D-Objekte werden durch einen Satz von 2D-Repräsentationen kodiert. Eine korrelationsähnliche Operation (Ansichtsinterpolation) vergleicht den jeweiligen Stimulus mit gespeicherten Repräsentationen verschiedener Ansichten, und die Repräsentation der ähnlichsten Ansicht wird aktiviert. Die Theorie der ansichtsbasierten Repräsentationen wird auch durch Ergebnisse aus neurophysiologischen Untersuchungen unterstützt (siehe Kapitel 2.2).

Ein weiteres Modell, das Repräsentationen zur Objekterkennung aufbaut, wurde 1996 von Ostkamp vorgestellt. Hier wird zunächst das grundlegende Problem der Extraktion mehrerer Objekte oder Objektkandidaten aus einer natürlichen Szene behandelt. Auf

Grundlage des binarisierten Eingangsbildes werden Merkmale wie Kontur, Liniensegmente und Ecken gewonnen. Ein Objekt wird durch synchrone neuronale Aktivität definiert. T-Stücke werden als Hinweis auf Verdeckungen interpretiert (siehe Kapitel 2.6) und zur aktiven Trennung der neuronalen Aktivitäten verschiedener Objektteile verwendet. Diese Art der Repräsentation beinhaltet nach wie vor die räumliche Information und damit implizit die Relationen der Objektkandidaten zueinander.

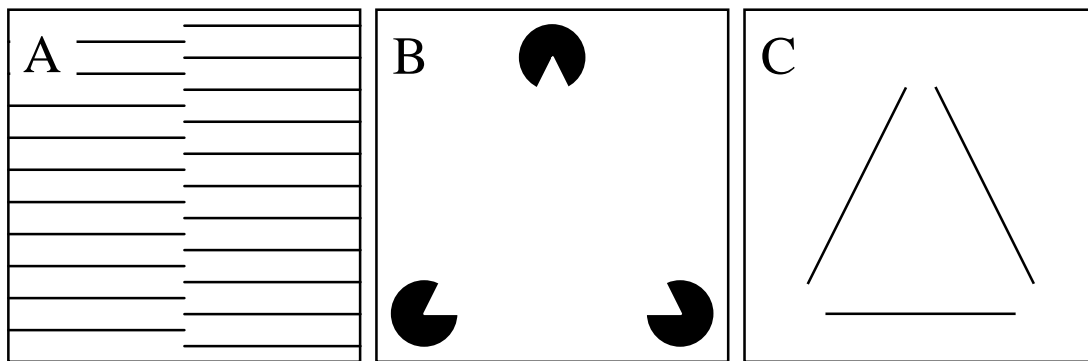
## 2.6 Hinweise auf die Bedeutung von Vertizes aus Wahrnehmungsexperimenten

Ein Vertex ergibt sich nach Biederman durch das Zusammentreffen von zwei oder drei Kantenenden an einem Ort (Biederman 1987, Hummel und Biederman 1992). Beispielsweise werden Ecken von zwei- oder dreidimensionalen Objekten als Vertizes betrachtet. Diese Definition werde nun modifiziert. Ein Vertex sei definiert durch das Zusammentreffen von mindestens zwei Kantenenden an etwa demselben Ort, wobei T-Stücke nicht als Vertizes angesehen werden (Weitzel et al. 1996, 1997 a, b, Spengler 1996). T-Stücke weisen auf Objektverdeckungen hin und sind daher gesondert zu behandeln (z.B. Heitger et al. 1992). Abbildung 2.6.1 zeigt ein Beispiel, das Vertizes und T-Stücke enthält.



**Abb.: 2.6.1:** Vertizes und T-Stücke in einer künstlichen Szene.

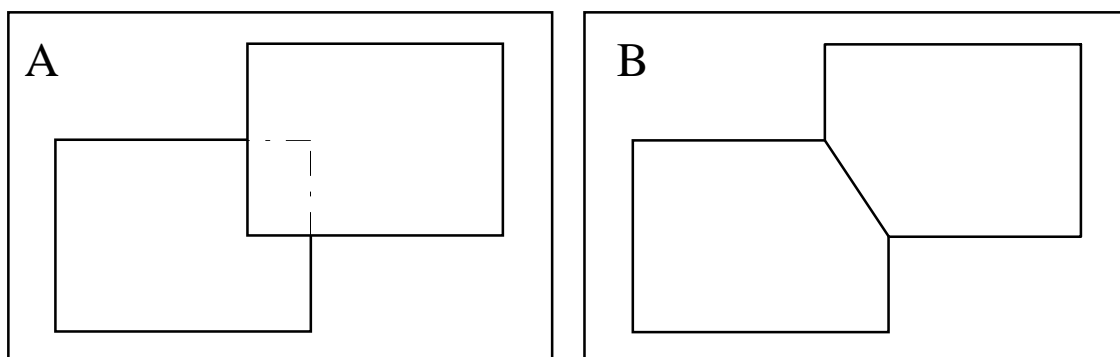
Vieles spricht für eine besondere Bedeutung von Vertizes in der Objekterkennung. Psychophysische Untersuchungen zeigen, daß das Auge des menschlichen Betrachters von Orten, an denen starke Konturänderungen auftreten "angezogen" wird (z.B. Attneave 1957, Bozkov et al. 1977, Zusne und Michels 1964). In weiteren psychophysischen Experimenten (Bülthoff et al. 1995) wurden Gitterdrahtmodelle auf verschiedene Arten deformiert (Rotation, Streckung, Scherung etc.) und Versuchspersonen präsentiert. Es stellte sich heraus, daß die Versuchspersonen die schlechtesten Wiedererkennungslieferungen lieferten, wenn das Objekt rotiert wurde. Bülthoff et al. folgerten, daß die (an den Vertizes auftretenden) Winkel die erbrachten Erkennungsleistungen am besten erklären. Weitere eindrucksvolle Beispiele liefert die Wahrnehmung von Scheinkonturen, die durch die Präsenz von Linienenden oder Vertizes ermöglicht wird. Ein Beispiel zeigt die Abbildung 2.6.2, in der der Betrachter eine senkrechte Linie in der Bildmitte wahrnimmt. Die die Vertizes verbindenden Konturen müssen für die Wahrnehmung nicht notwendigerweise explizit vorhanden sein. Dieses Phänomen der Musterergänzung läßt sich auch durch das Beispiel des in der Mitte der Abbildung 2.6.2 gezeigten Kanizsa-Dreiecks (Kanizsa 1976) illustrieren. Der Betrachter nimmt die (Schein-)Konturen eines weißen Dreiecks wahr.



**Abb. 2.6.2:** Psychophysische Beispiele, die die Bedeutung von Linienenden und Vertizes für die Wahrnehmung von Scheinkonturen zeigen. **A:** Die Anordnung der Linienenden führt beim Betrachter zur Wahrnehmung einer senkrechten Linie in der Bildmitte (nach: Leshner und Mingolla 1993). **B:** Kanizsa-Dreieck (Kanizsa 1976). Der Betrachter nimmt ein weißes Dreieck wahr. **C:** Der menschliche Betrachter nimmt kein vollständiges Dreieck wahr, wenn die Vertizes des Kanizsa-Dreiecks gegen Linien, die entlang der Konturen verlaufen, ausgetauscht wurden.

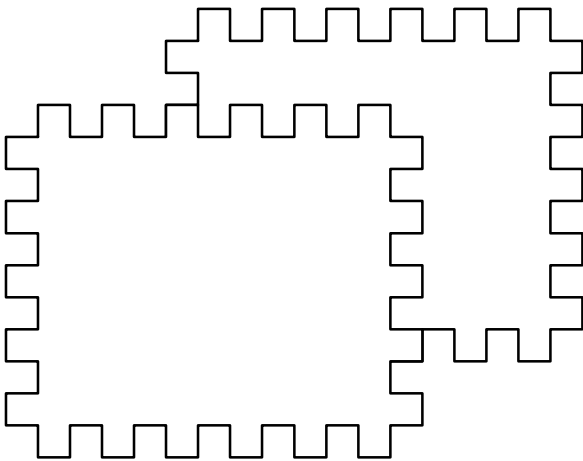
Die Wahrnehmung eines vollständigen Dreiecks wird hingegen erschwert, wenn die den Vertizes entsprechenden schwarzen Scheiben durch die Konturen der Dreiecksseiten ersetzt werden. Das wird auch durch andere psychophysische Experimente bestätigt: Fehlende Vertizes tragen stärker zur Fehlerrate in Erkennungsaufgaben bei als fehlende Kontursegmente (Biederman 1987). Die zur Erkennung wesentliche Information läßt sich daher in den Vertizes bzw. Ecken oder Orten starker Krümmung vermuten.

Einen Spezialfall unter den Vertizes stellt das T-Stück dar, das sich dadurch auszeichnet, daß zwei der drei zur Definition notwendigen Linienenden mit entgegengesetzter Orientierung im Ort des T-Stücks "terminieren", so daß im eigentlichen Sinne nur ein Linienende vorliegt. Alleinige Linienenden (T-Stücke) geben starke Hinweise auf Verdeckung und können zur Wahrnehmung von Scheinkonturen in Vorzugsorientierung des jeweiligen Linienendes führen, wie am Beispiel der Abbildung 2.6.3 A deutlich wird.



**Abb. 2.6.3:** Hinweise auf Verdeckung durch T-Stücke. **A:** T-Stücke können zur Wahrnehmung des vervollständigten Musters führen. Die entsprechenden Scheinkonturen sind durch gestrichelte Linien angedeutet. **B:** Die Schnittpunkte zwischen den Objekten sind zugleich Orte hoher Krümmung, so daß sich keine T-Stücke ergeben. Hier geht die Tiefeninformation verloren.

Auf eine gewisse Problematik weist die "Helmholtzregel" hin, wonach Tiefeninformation bei Objektverdeckung nur dann gewonnen werden kann, wenn der Ort, wo sich verdeckende und verdeckte Kontur schneiden, nicht gleichzeitig ein Ort mit hoher Krümmung ist. Ein Beispiel für eine solche Überdeckung zeigt Abbildung 2.6.3.B. In diesem Falle liegt jedoch meist kein T-Stück vor. Anhand eines weiteren Beispiels (Abb. 2.6.4) wird deutlich, daß auch die Art der Betrachtung eine gewisse Rolle spielen kann. Auf den ersten Blick sieht der Betrachter ein überdeckendes (linkes) Objekt und ein überdecktes (rechtes) Objekt. Dieser Eindruck wird durch "große" T-Stücke vermittelt. Bei genauerer Betrachtung sind jedoch auch "kleine" T-Stücke zu sehen, die auf eine Überdeckung des linken Objektes durch das rechte Objekt hinweisen. Welches Objekt nun im Vordergrund und welches im Hintergrund gesehen wird, ist also auch von der Betrachtungsweise abhängig.



**Abb. 2.6.4:** *Unterschiedliche T-Stücke am gleichen Ort.* Je nachdem welche der T-Stücke (die "großen" oder die "kleinen") der menschliche Beobachter betrachtet, sieht er das linke oder das rechte Objekt im Vordergrund. (Nach: Metzger 1975).

Anhand der zahlreichen psychophysischen Untersuchungen und Beispiele wird die besondere Bedeutung der Vertizes und T-Stücke deutlich. Vertizes erlauben sogar eine weitgehende (zweidimensionale) Beschreibung von Objekten, wenn auch die Information über die Orientierungen der die Vertizes induzierenden Linienenden berücksichtigt wird (Spengler 1996, Weitzel et al. 1996, Weitzel et al. 1997a, b). Entsprechend lassen sich Objektkonturen anhand der relationalen Zeiger ihrer Vertizes rekonstruieren, wenn das Objekt nur gerade Kanten besitzt.

### 3 Bildvorverarbeitung

In diesem Kapitel werden die Einzelheiten der Bildvorverarbeitung beschrieben, die die wesentliche Grundlage zur Gewinnung von Konturinformation aus einer Szene darstellt. Zunächst wird auf zu verarbeitende Grauwertbildsequenzen eingegangen. Anschließend werden die Modell-Photorezeptoren und Modell-X-Zellen beschrieben, die in Anlehnung an das Vorbild der natürlichen Retina den lokalen Intensitätskontrast gewinnen, der dann in Pulsfolgen den nachfolgenden Verarbeitungsstufen zur Verfügung gestellt wird. Schließlich werden die wesentlichen Eigenschaften der X-Zellen im Hinblick auf mögliche Auswirkungen für die weitere Verarbeitung untersucht.

#### 3.1 Bildsequenzen

Die Bildaufnahme erfolgt durch eine handelsübliche Videokamera. Sie liefert Grauwertbilder  $G(x,y,t)$  mit einer Taktfrequenz von  $f=25$  Hz, d.h. in einem zeitlichen Abstand von  $T=40$  ms. Dabei bleibt das Zeilensprungverfahren unberücksichtigt, indem nur jedes zweite Bild ausgewertet wird. Die nachfolgenden neuronalen Verarbeitungsstufen verwenden jedoch einen Simulationszeitschritt der Dauer von 1 ms bzw. 1 BIN, so daß die Kamera nur in jedem 40. Zeitschritt ein Grauwertbild zur Verfügung stellt. In den dazwischenliegenden Zeitschritten sollen aber ebenso Grauwertbilder dargeboten werden. Diese können auf verschiedene Arten ermittelt werden.

1. Darbietung desselben Grauwertbildes während der gesamten Dauer  $T$ . Dieses Verfahren hat jedoch den Nachteil, daß neben großen Fehlern relativ zur besten möglichen Schätzung (Interpolation) starke zeitliche Grauwertsprünge auftreten und das Netzwerk durch Taktvorgabe beeinflussen können.

2. Interpolation unter Berücksichtigung des Abtasttheorems. Das Abtasttheorem läßt sich wie folgt formulieren: Ist die Fouriertransformierte  $H(f)$  einer Funktion  $h(t)$  bandbegrenzt, d.h. es gibt eine Frequenz  $F$  mit  $H(f)=0$ , falls  $f > F$ , und wird die Funktion  $h(t)$  in Intervallen  $T$  mit:

$$T < \frac{1}{2F} \quad (\text{Gl.3.1.1})$$

(Nyquistbedingung) abgetastet, so läßt sich  $h(t)$  eindeutig aus den Abtastwerten  $h(n \cdot T)$  mit  $n \in N$  rekonstruieren, und es ist:

$$h(t) = T \cdot \sum_{n=-\infty}^{+\infty} h(nT) \cdot \frac{\sin(2\pi F \cdot [t - nT])}{\pi \cdot [t - nT]}. \quad (\text{Gl.3.1.2})$$

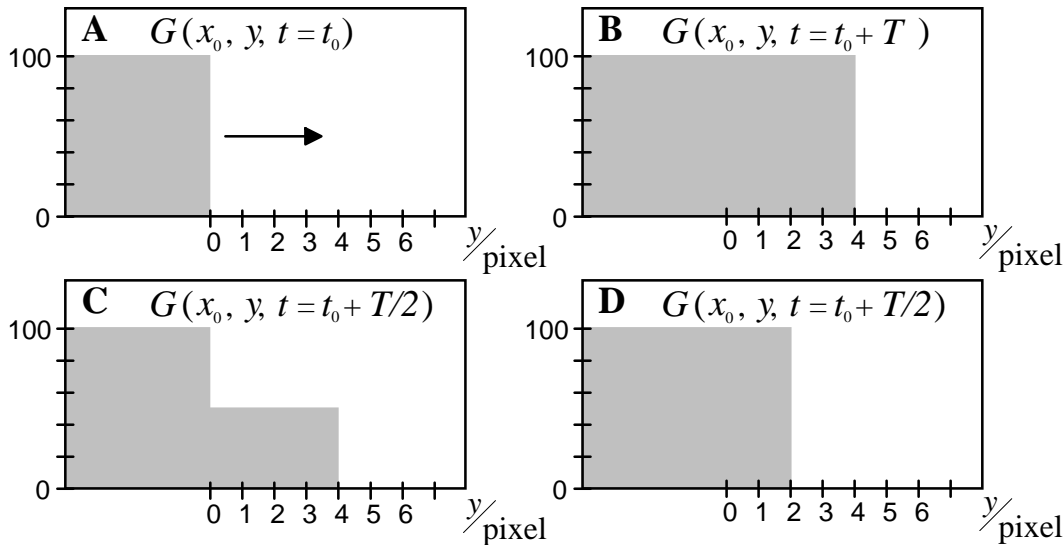
Daraus folgt, daß die zeitliche Interpolation mit  $\sin t/t$  prinzipiell die perfekte Rekonstruktion aller Zwischen-Grauwertbilder ermöglicht, jedoch sind die folgenden Aspekte zu berücksichtigen. Zum einen darf die sich hier aus der Nyquistbedingung ergebende Grenzfrequenz von  $F=12.5$  Hz nicht überschritten werden. Üblicherweise treten aber in dynamischen natürlichen Szenen hohe zeitliche Frequenzkomponenten auf, die eine Tiefpaßfilterung zur Vermeidung von Aliasing notwendig machen. Zum anderen liegt ein schwerwiegendes Problem in der Notwendigkeit, möglichst viele Abtastwerte in die Berechnung miteinbeziehen zu müssen. Die Rekonstruktion wird umso besser, je mehr Abtastwerte von "vergangenen" und "zukünftigen" Bildern miteinberechnet werden können. In der Praxis bedeutet das, daß das System nicht nur mehr Rechenaufwand bewältigen, sondern darüberhinaus auch auf Zukunftsdaten warten muß und zwar umso länger, je besser die Rekonstruktion werden soll. Auf der anderen Seite wird eine Verarbeitung in Echtzeit angestrebt, die eine rasche Vorverarbeitung verlangt. Daher wird auf die Interpolation mit  $\sin t/t$  verzichtet.

3. Lineare Interpolation (Schott 1995). Der Grauwert  $G(x_0, y_0, t)$  eines Bildpunktes  $(x_0, y_0)$  zur Zeit  $t$  ist gegeben durch:

$$G(x_0, y_0, t) = G(x_0, y_0, t \operatorname{div} T) + \frac{G((x_0, y_0, t \operatorname{div} T) \cdot T + T) - G((x_0, y_0, t \operatorname{div} T) \cdot T)}{T} \cdot (t \operatorname{mod} T) \quad (\text{Gl.3.1.3})$$

- mit     $\operatorname{div}$ :        Operator, der nur anwendbar auf ganzzahlige Operanden ist und den ganzzahligen Anteil des Ergebnisses einer Division liefert.
- $\operatorname{mod}$ :        Operator, der nur anwendbar auf ganzzahlige Operanden ist und den Rest einer Division liefert.
- $T$ :             zeitliches Abtastintervall.

Bei diesem Verfahren treten zwar immer in periodischen Abständen nicht differenzierbare Stellen auf, aber es führt nicht zu Grauwertsprüngen, wie sie beim ersten Verfahren entstehen können. Zudem weisen die X-Zellen, wie später noch gezeigt wird, zeitliche Tiefpaßeigenschaften auf, die solche nicht differenzierbaren Stellen glätten können. Ein weiteres Problem bei der Ermittlung der Zwischen-Grauwertbilder kann im Zusammenhang mit bewegten Objekten auftreten. Dies soll anhand eines einfachen Beispiels erläutert werden. Ein Objekt (Grauwert sei 100) bewege sich über eine Fläche mit Grauwert 0. Die Kamera liefere zur Zeit  $t = t_0$  ein Grauwertbild. Nach  $T = 40$  ms sei die Objektkante um 4 Pixel in  $y$ -Richtung verschoben worden, und die Kamera liefere ein zweites Grauwertbild. Die oben erwähnten Interpolationsverfahren werden nun bei der Bestimmung der Zwischen-Grauwertbilder die Grauwerte aller von dieser Objektverschiebung beeinflussten Bildpunkte gleichermaßen verändern.



**Abb. 3.1:** Ermittlung der Zwischen-Grauwertbilder im Zusammenhang mit bewegten Objekten. **A**, **B**: Kante, die in der Zeit  $T$  um 4 Pixel verschoben wird. **C**: Das (lineare) Interpolationsverfahren berechnet für alle in der Zeit  $T/2$  überstrichenen Bildpunkte denselben Grauwert  $G=50$ . **D**: Der (geschätzte) Aufenthaltsort der Kante zur Zeit  $T/2$  führt zur Berechnung unterschiedlicher Grauwerte ( $G=100$  bzw.  $0$ ) für die in der Zeit  $T$  überstrichenen Bildpunkte.

Ein wahrscheinlich wirklichkeitsnäheres Verfahren würde aber die Grauwerte ausgehend vom kantennächsten Bildpunkt sukzessive verändern (Abb. 3.1). Dieses Verfahren würde jedoch Kenntnisse über das Objekt und seine Bewegung benötigen, die erst auf höheren Verarbeitungsstufen des Netzwerkes zur Verfügung stehen. Daher können entsprechende Rekonstruktionsfehler in der Bildvorverarbeitung nicht ausgeschaltet werden. Sie treten jedoch nicht auf, wenn für die Verschiebungsgeschwindigkeit gilt:  $v = 0 \text{ Pixel} / T$  oder  $v = 1 \text{ Pixel} / T$ .

### 3.2 Modellierung der Photorezeptoren

Die Anordnung der Photorezeptoren in der Netzhaut läßt sich in sehr guter Näherung durch die Schnittpunkte eines hexagonalen Gitters beschreiben (Siminoff 1980). Ein entsprechendes Modell wurde beispielsweise 1982 von Hartmann vorgestellt. In der rechnergestützten Modellierung bietet sich die Einpassung des hexagonalen Gitters in ein kartesisches Koordinatensystem an. Das daraus resultierende Raster läßt sich für verschiedene Abtastweiten  $d$  beschreiben durch:

$$R(x, y) = \begin{cases} 1, & \text{falls } (x \bmod d = 0 \wedge y \bmod 2d = 0) \vee (x \bmod d = \frac{d}{2} \wedge y \bmod 2d = d) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 3.2.1})$$



Neben der Anordnung ist auch die Lichtempfindlichkeitscharakteristik der Rezeptoren von Bedeutung, die sich nach (Rodieck 1965) durch eine Gaußverteilung annähern läßt. Die Gaußverteilung ist gegeben durch:

$$Z(x, y, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \cdot \exp(-(x^2 + y^2)/2\sigma^2). \quad (\text{Gl. 3.2.2})$$

Die Potentiale  $P$  der Rezeptorzellen zur Zeit  $t$  werden im Modell durch diskretisierte Faltung der Gaußverteilung mit dem ermittelten Grauwertbild  $G(x, y, t)$  unter Berücksichtigung des Rasters  $R(x, y)$  ermittelt, und es ist:

$$P(x_0, y_0, t) = \sum_{x=x_0-\frac{m}{2}}^{x_0+\frac{m}{2}} \sum_{y=y_0-\frac{m}{2}}^{y_0+\frac{m}{2}} Z(x-x_0, y-y_0, \sigma) \cdot G(x, y, t) \quad \forall x_0, y_0 | R(x_0, y_0)=1, \quad (\text{Gl. 3.2.3})$$

wobei sich  $m$  aus der Größe  $(m+1) \cdot (m+1)$  der Gaußmaske  $Z$  ergibt und stets geradzahlig ist. Wie aus Gleichung 3.2.1 ersichtlich ist, können verschiedene Auflösungen durch unterschiedliche Abtastweiten  $d$  realisiert werden. Durch die gleichzeitige Verwendung mehrerer Auflösungen ( $d = 2, 4, 8$ ) bei jeweils gleicher Anzahl von Abtastpunkten und konzentrischer Überlagerung der verschiedenen Raster nimmt das Auflösungsvermögen der modellierten Retina vom Zentrum zur Peripherie hin ab (vgl. Abbildung 3.2.1). Auch in der natürlichen Retina nimmt das Auflösungsvermögen zur Peripherie hin ab, jedoch gibt es dort in jeder Exzentrizität jeweils nur eine Auflösung für X- und Y-Zellen.

Für die Realisierung der verschiedenen Auflösungen ist für jede Abtastweite eine geeignete Gaußverteilung, die auf eine Filtermaske abgebildet wird, zu ermitteln. Die Standardabweichung  $\sigma$  ist dabei so zu wählen, daß bei der Unterabtastung im hexagonalen Raster die Grauwerte aller Pixel mit gleichem Gewicht eingehen. Spengler errechnete daher für verschiedene Maskengrößen jeweils diejenige Standardabweichung  $\sigma$ , bei der die Streuung  $s$  des Ergebnisses aus linearer Superposition der Gaußmasken  $Z$  über alle Abtastpunkte ein Minimum annimmt (Spengler 1996). Tabelle 3.2.1 zeigt die entsprechenden Daten.

Auch eine zweite, physiologisch motivierte Bedingung für die Bestimmung der Standardabweichung  $\sigma$  wird dabei näherungsweise erfüllt:  $\sigma$  ist so zu wählen, daß sich die Empfindlichkeitsverteilungen zweier benachbarter Rezeptoren beim halben Wert des Maximums schneiden (Becker 1995). Damit ergibt sich die Standardabweichung zu:

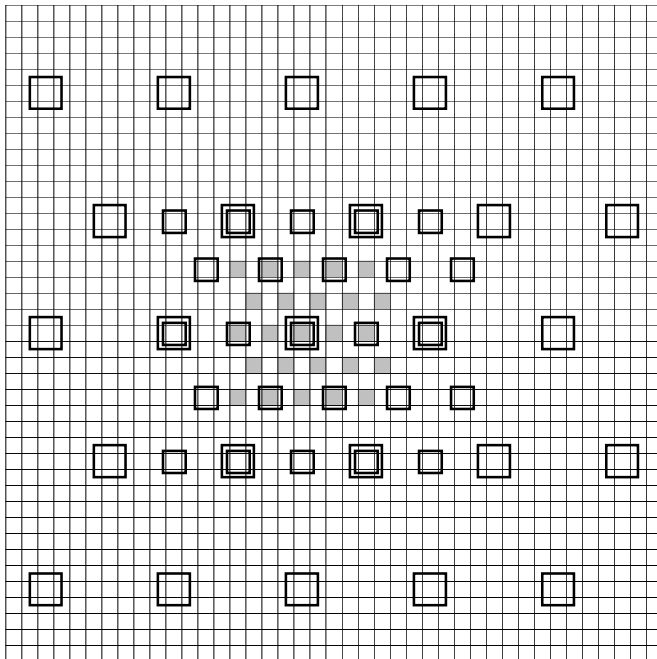
$$\sigma = \frac{1}{2} \cdot \frac{r}{\sqrt{2 \ln 2}}, \quad (\text{Gl. 3.2.4})$$

wobei  $r$ : Abstand der jeweils nächsten Schnittpunkte im hexagonalen Raster.

Für das in das kartesische Koordinatensystem eingebaute hexagonale Raster kann nur ein mittlerer Abstand  $\bar{r}$  ermittelt werden:

$$\bar{r} = \frac{2d + 4\sqrt{d^2 + (d/2)^2}}{6} \approx 1.08 d . \quad (\text{Gl. 3.2.5})$$

Die daraus resultierenden Werte für  $\sigma$  sind ebenso der Tabelle 3.2.1 zu entnehmen.

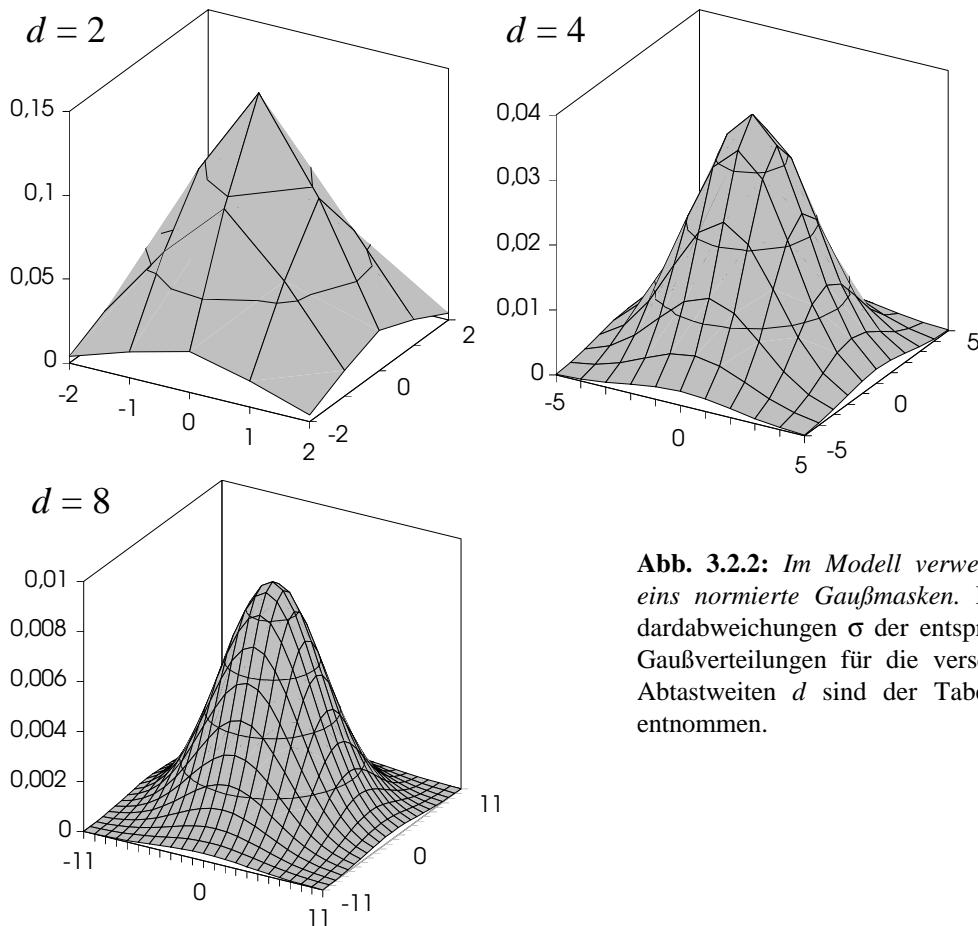


**Abb. 3.2.1:** Hexagonales Abtastraster der modellierten Retina. Dargestellt sind die Raster für die Abtastweiten  $d = 2$  (gefüllte Quadrate), 4 (kleine offene Quadrate), 8 (große offene Quadrate) jeweils mit  $5 \cdot 5$  Abtastpunkten auf dem kartesischen Koordinatensystem eines Grauwertbildes (Schott 1995, Pauly 1995).

Abtastweite $d$	Dimension von $Z$	$\sigma$ bei minimaler Streuung $s$	$\sigma$ nach Gl. 3.2.4
2	$5 \cdot 5$	1.05	0.92
4	$11 \cdot 11$	2.1	1.83
8	$23 \cdot 23$	4.2	3.67

**Tab. 3.2.1:** Standardabweichung  $\sigma$  der Gaußmaske  $Z$  bei minimaler Streuung  $s$  des Ergebnisses aus linearer Superposition der Gaußmasken  $Z$  unter Berücksichtigung der verschiedenen Abtastraster (Daten aus Spengler 1996).

Die sich aus den Standardabweichungen bei minimaler Streuung  $s$  ergebenden Gaußmasken werden im Modell verwendet und sind in Abbildung 3.2.2 dargestellt.

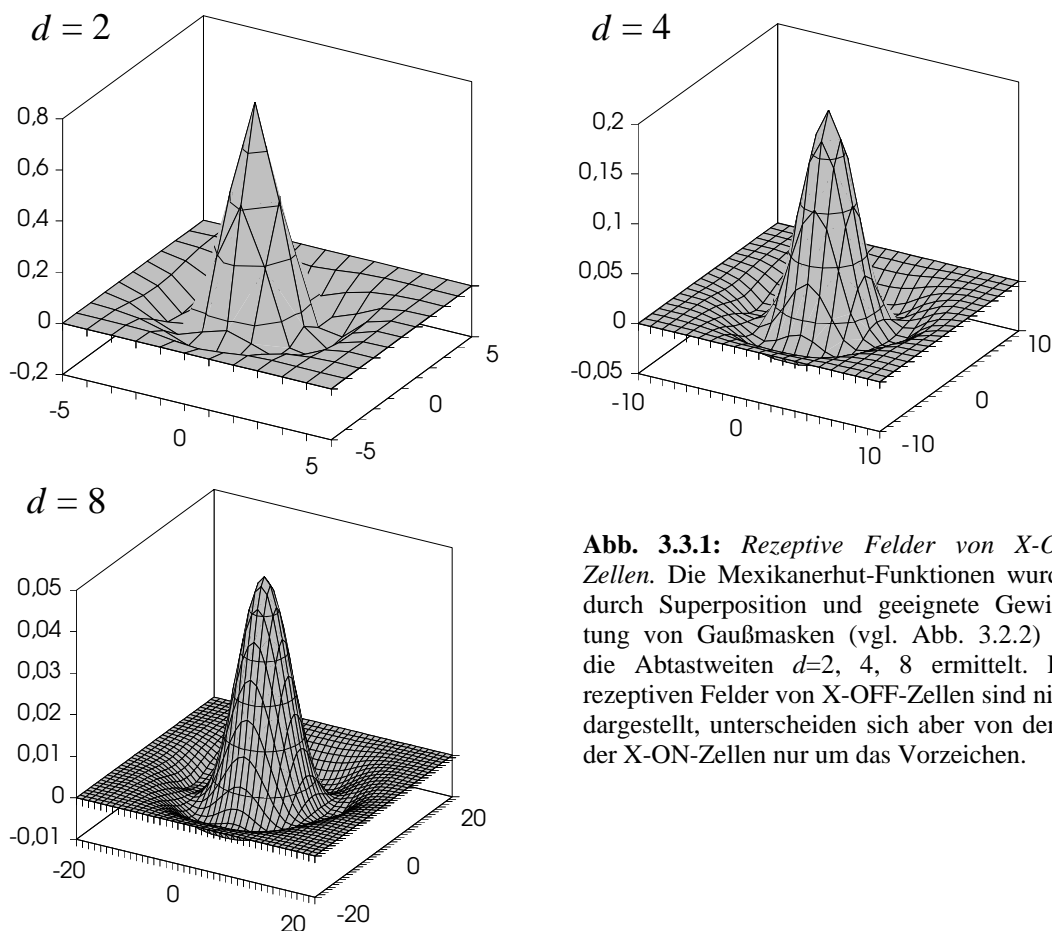


**Abb. 3.2.2:** Im Modell verwendete auf eins normierte Gaußmasken. Die Standardabweichungen  $\sigma$  der entsprechenden Gaußverteilungen für die verschiedenen Abtastweiten  $d$  sind der Tabelle 3.2.1 entnommen.

### 3.3 Modellierung der X-Zellen

#### 3.3.1 Räumliche Eigenschaften der Rezeptiven Felder

Konzentrische rezeptive Felder mit Center/Surround-Struktur, wie sie zum Beispiel bei retinalen Bipolarzellen oder Ganglienzellen auftreten, lassen sich durch Mexikanerhut-Funktionen, die der zweiten Ableitung einer Gaußverteilung entsprechen, annähern. Das hexagonale Gitter erlaubt auf einfache Weise die Konstruktion solcher Mexikanerhut-Funktionen durch Gewichtung und Überlagerung von Gaußfiltern. Dieses Vorgehen ist auch als DOOG-Modell (Difference Of Offset Gaussians) bekannt und wurde bereits von z.B. Hartmann 1982, Young 1986, Young 1988 verwendet. Im vorliegenden Modell wird jeweils ein sechsfach positiv gewichtetes Gaußfilter über den zentralen Abtastpunkt und sechs einfach negativ gewichtete Gaußfilter über die sechs peripheren Abtastpunkte superpositioniert, um das rezeptive Feld einer X-ON-Zelle zu modellieren. Die entsprechend inverse Gewichtung führt zur Modellierung der räumlichen Filtereigenschaften einer X-OFF-Zelle. Abbildung 3.3.1 zeigt die aus Superpositionen gewonnenen rezeptiven Felder von X-ON-Zellen.



**Abb. 3.3.1:** *Rezeptive Felder von X-ON-Zellen.* Die Mexikanerhut-Funktionen wurden durch Superposition und geeignete Gewichtung von Gaußmasken (vgl. Abb. 3.2.2) für die Abtastweiten  $d=2, 4, 8$  ermittelt. Die rezeptiven Felder von X-OFF-Zellen sind nicht dargestellt, unterscheiden sich aber von denen der X-ON-Zellen nur um das Vorzeichen.

### 3.3.2 Zeitliche Eigenschaften der Rezeptiven Felder

Das Produkt aus der einem Abtastpunkt zugeordneten Gaußfunktion und dem Grauwertbildausschnitt (Rezeptorantwort  $R$ ) wird auf zwei Leckintegratoren mit unterschiedlichen zeitlichen Eigenschaften gegeben. Diese Unterscheidung führt zu bestimmten zeitlichen Eigenschaften der X-Zelle, die im Kapitel 3.4 näher untersucht werden. Zunächst resultieren daraus jeweils ein Center-Potential  $\Phi_c$  und ein Surround-Potential  $\Phi_s$ , für die gilt:

$$\tau^c \cdot \frac{\partial}{\partial t} \Phi_c(t) = -\Phi_c(t) + V^c \cdot R \quad (\text{Gl.3.3.1})$$

$$\tau^s \cdot \frac{\partial}{\partial t} \Phi_s(t) = -\Phi_s(t) + V^s \cdot R \quad (\text{Gl.3.3.2})$$

mit  $\tau^c, \tau^s$ : Zeitkonstanten der Leckintegratoren,  
 $V^c, V^s$ : Verstärkungsfaktoren der Leckintegratoren.

Schließlich ergeben sich das Membranpotential einer X-ON-Zelle  $\Phi^{ON}(x, y, t)$  und das Membranpotential einer X-OFF-Zelle  $\Phi^{OFF}(x, y, t)$  am Ort  $(x, y)$  zur Zeit  $t$  im pseudohexagonalen Gitter bei Abtastweite  $d$  zu:

$$\Phi^{ON}(x, y, t) = 6 \cdot \Phi_c(x, y, t) - \Phi_s(x+d, y, t) - \Phi_s(x-d, y, t) - \Phi_s(x+d/2, y+d, t) - \Phi_s(x-d/2, y+d, t) - \Phi_s(x+d/2, y-d, t) - \Phi_s(x-d/2, y-d, t), \quad (\text{Gl.3.3.3})$$

$$\Phi^{OFF}(x, y, t) = -6 \cdot \Phi_c(x, y, t) + \Phi_s(x+d, y, t) + \Phi_s(x-d, y, t) + \Phi_s(x+d/2, y+d, t) + \Phi_s(x-d/2, y+d, t) + \Phi_s(x+d/2, y-d, t) + \Phi_s(x-d/2, y-d, t). \quad (\text{Gl.3.3.4})$$

Analog dem Marburger Modellneuron wurde auch die X-Zelle unter Verwendung eines Spike-Encoders modelliert und die Funktionsweise dabei weitestgehend übernommen. Das sich aus den Gleichungen 3.3.3 und 3.3.4 ergebende Membranpotential  $\Phi(t)$  der X-Zelle unterliegt dem ständigen Vergleich mit dem Wert der dynamischen Schwelle  $\Theta(t)$  zuzüglich einem Schwellenoffset  $\Theta_0$ . Bei Überschwelligkeit generiert die X-Zelle ein Aktionspotential am Ausgang und erhöht den Wert der Schwelle um das Schwelleninkrement  $V^\ominus$ . Das Ausgangssignal läßt sich somit beschreiben durch

$$y(t) = \begin{cases} 1, & \text{falls } \Phi(t) \geq \Theta(t) + \Theta_0 \\ 0, & \text{sonst} \end{cases}, \quad (\text{Gl. 3.3.5})$$

wobei sich die zeitliche Entwicklung der dynamischen Schwelle aus der Differentialgleichung

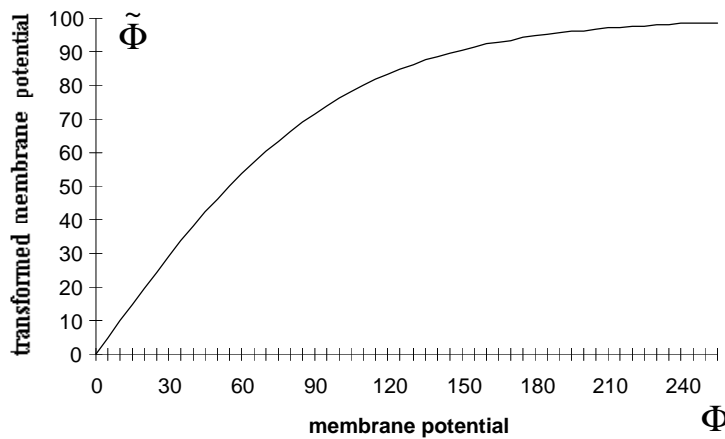
$$\tau^\ominus \cdot \frac{\partial}{\partial t} \Theta(t) = -\Theta(t) + V^\ominus \cdot y(t) \quad (\text{Gl. 3.3.6})$$

ergibt. Darin sind  $V^\ominus$  der Verstärkungsfaktor und  $\tau^\ominus$  die Zeitkonstante der dynamischen Schwelle.

Die modellierte X-Zelle stellt, ähnlich der biologischen X-Zelle, einen Detektor für Grauwertdifferenzen dar. Genauer gesagt spricht die biologische X-Zelle auf Intensitätskontraste an. Der Intensitätskontrast folgt aus der Differenz zwischen den im Zentrum und den in der Peripherie der Mexikanerhut-Funktion auftretenden Intensitäten, die ins Verhältnis zur Intensitätssumme gesetzt wird (*Michelson-Kontrast*).

Eine mögliche Erweiterung der modellierten X-Zelle ergibt sich aus den folgenden Überlegungen. Wie im nächsten Kapitel noch ausgeführt werden wird, werden die X-Zellen für die Realisierung von Kantendetektoren additiv verschaltet. Hier besteht die Gefahr, daß im zu verarbeitenden Grauwertbild ein einzelner Punkt oder eine einzelne Ecke mit starkem Intensitätskontrast auftritt und eine einzelne X-Zelle auf diesen Stimulus mit einer Ausgangsrate an Aktionspotentialen antwortet, die ausreicht, um den Kantendetektor überschwellig werden zu lassen. Auf der anderen Seite ist es möglich, daß bei einer in

Vorzugsrichtung des Detektors orientierten Kante mit vergleichsweise weniger starkem Intensitätskontrast die X-Zellen sehr geringe Ausgangsraten erzeugen, so daß der Kantendetektor keine Aktionspotentiale generieren kann. Um diesem Problem entgegenzuwirken, bietet sich eine Begrenzung der X-Zellen-Kennlinie an. Dies kann beispielsweise durch geeignet groß gewählte Parameter  $\tau^\ominus, V^\ominus$  des Spike-Encoders geschehen, was jedoch auch die Sensitivität des Detektors bei kleine Raten verringern würde. Eine weitere Möglichkeit ist die Begrenzung des Membranpotentials, die zum Beispiel durch eine logarithmische Skalierung des Potentials erreicht werden kann (Schott 1997). Eine solche Potentialbegrenzung ist zwar vor allem technisch motiviert, jedoch gibt es auch eine gewisse physiologische Motivation.



**Abb. 3.3.2:** Logarithmische Skalierungsfunktion zur Begrenzung des Membranpotentials.

Wie in Kapitel 2 beschrieben, ist das menschliche Sehsystem fähig, in einem sehr weiten Dynamikbereich unterschiedliche Intensitäten zu verarbeiten. Einen bestimmten Anteil haben daran auch die Zellen der Retina, wie zum Beispiel die Photorezeptoren, die bereits eine logarithmisch verlaufende Skalierung der Intensitäten auf ihre Potentiale aufweisen. Im Modell wird eine solche logarithmische Skalierung durch den positiven Teil einer Sigmoidfunktion angenähert (vgl. Abb. 3.3.2), und das *sigmoidal* abgebildete Membranpotential ergibt sich aus dem bisherigen Membranpotential der X-Zelle zu

$$\tilde{\Phi}(t) = \left( \frac{200}{1 + \exp(-\Phi(t)/50)} \right) - 100. \quad (\text{Gl. 3.3.7})$$

Die physiologische Motivation wird aber insofern etwas relativiert, als daß die Potentiale der natürlichen Photorezeptoren Beleuchtungsstärken, die über mehrere Dekaden schwanken, noch verarbeiten können (vgl. Kap. 2.2), während im Modell nur etwa zwei Dekaden zur Kodierung von Intensitäten zur Verfügung stehen.

### 3.4 Untersuchung von Eigenschaften der X-Zellen

Im folgenden werden die Eigenschaften der vier X-Zelltypen (ON-Center/OFF-Center bzw. unverändertes/modifiziertes Membranpotential  $\Phi, \tilde{\Phi}$ ) durch Anwendung verschiedener Stimuli in Simulationen untersucht und miteinander verglichen. Wie in Kapitel 2 beschrieben, besitzt die Antwort einer X-Zelle auf einen zeitlich nicht konstanten Stimulus einen transienten und einen stationären Anteil. Zunächst wird das stationäre, anschließend das transiente Antwortverhalten untersucht.

#### 3.4.1 Stationäres Antwortverhalten der Modell-X-Zelle

Beide Typen des Antwortverhaltens werden im Modell durch unterschiedlich aufeinander abgestimmte Verstärkungsfaktoren und Zeitkonstanten der Tiefpaßfilter für Center- und Surround-Potentiale ermöglicht. Beim Einschalten eines zeitlich und räumlich konstanten Stimulus, der das gesamte rezeptive Feld abdeckt, zeigt die X-Zelle nur eine transiente Antwort. Das Membranpotential der modellierten X-Zelle soll beim gleichen Stimulus entsprechend nur kurzzeitig deutlich größer Null sein. Um eine stationäre Antwort zu unterdrücken, ohne jedoch einen refraktären Zustand durch ein negatives Membranpotential zu erhalten, muß zusätzlich

$$\Phi_c(t) - \Phi_s(t) = 0 \quad \text{für } t \rightarrow \infty \quad (\text{Gl. 3.4.1})$$

gefordert werden. Mit dieser Bedingung ist einer der vier Parameter  $V^c, V^s, \tau^c, \tau^s$  bei Vorgabe der übrigen drei festgelegt. Der Zusammenhang ist über die Gleichung

$$V^c = V^s \cdot (1 - \exp(1/\tau^c)) / (1 - \exp(1/\tau^s)) \quad (\text{Gl. 3.4.2})$$

gegeben, die wiederum aus Gleichung 3.4.1 und dem Wert  $F(n \cdot \Delta t)$  eines Leckintegrators ( $V, \tau$ ) bei Dauererregung mit konstantem Stimulus nach  $n$  diskreten Zeitschritten der Dauer  $\Delta t$  folgt. Dieser ist gegeben durch:

$$F(n \cdot \Delta t) = V \cdot (1 - \exp(-n \cdot \Delta t / \tau)) / (1 - \exp(-\Delta t / \tau)) \quad (\text{Gl. 3.4.3})$$

(z.B. Blümler 1997). Die in Tabelle 3.4.1 aufgeführten und in den folgenden Simulationen verwendeten Parameter erfüllen den Zusammenhang der Gleichung 3.4.2.

Für die erste Simulation wird ein Pixel mit Grauwertdifferenz  $\Delta G$  gegenüber dem Hintergrund über 1100 Zeitschritte (BIN) an einer festen Position relativ zum rezeptiven Feld (Abtastweite  $d=2$ ) der X-ON-Zelle als Stimulus dargeboten. Um eventuell auftretende Initiali-

sierungseffekte und transiente Anteile durch Einschaltvorgänge bei dieser Simulation ausschließen zu können, finden die ersten 100 Simulationsschritte keine Berücksichtigung bei der Auswertung. Nach 100 BIN sind die beiden hier betrachteten Leckintegratoren in sehr guter Näherung in Sättigung, was aus Gleichung 3.4.3 mit den Parametern aus Tabelle 3.4.1 folgt.

Schwellen- verstärkung	Schwellen- zeitkonstante	Schwellen- offset	Center- Potential- verstärkung	Center- Potential- zeitkonstante	Surround- Potential- verstärkung	Surround- Potential- zeitkonstante
$V^\Theta$	$\tau^\Theta$	$\Theta_0$	$V^c$	$\tau^c$	$V^s$	$\tau^s$
58	15 BIN	10	1.0	3.0 BIN	0.5	6.5 BIN

**Tab. 3.4.1:** Parameter der Modell-X-Zelle.

Die daraus resultierenden größtmöglichen absoluten Fehler bei einem Stimulus mit  $\Delta G = 255$  von Center- und Surroundpotential ergeben sich zu

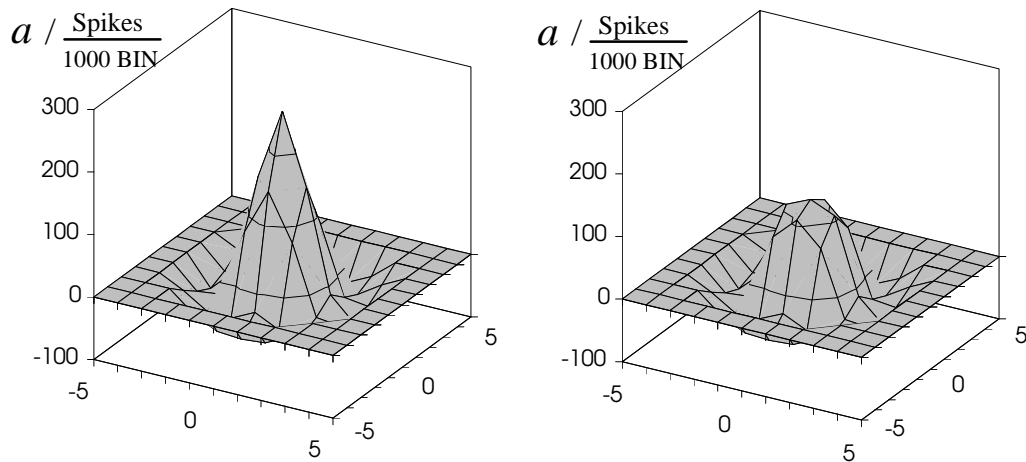
$$\Delta\Phi_c = \Phi_c(100 \cdot \Delta t) - \Phi_c(t \rightarrow \infty) \approx 10^{-12}, \quad (\text{Gl. 3.4.4})$$

$$\Delta\Phi_s = \Phi_s(100 \cdot \Delta t) - \Phi_s(t \rightarrow \infty) \approx 10^{-4} \quad (\text{Gl. 3.4.5})$$

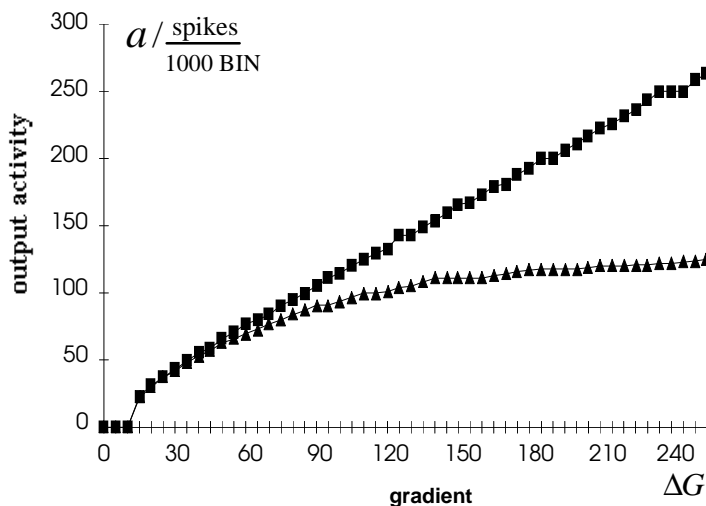
und sind im Vergleich zu  $\Delta G$  des Stimulus vernachlässigbar klein, so daß der transiente Anteil keine Rolle mehr spielt. Ermittelt wird die durchschnittliche Rate  $a$  der Aktionspotentiale am Ausgang der X-Zelle, die sich aus dem Verhältnis zwischen gemessener Summe der innerhalb des Beobachtungszeitintervalls generierten Aktionspotentiale und Länge des Beobachtungszeitintervalls von 1000 BIN ergibt. Die Simulation wird für jede Position relativ zum rezeptiven Feld wiederholt. Abbildung 3.4.1 zeigt die ermittelten Ausgangsraten bei unterschiedlichen Positionen des Punktstimulus. Die Ausgangsraten bei inversem Punktstimulus (negative Grauwertdifferenz) sind mit negativem Vorzeichen aufgetragen. Die gewonnene Kurve (Abbildung 3.4.1) approximiert erwartungsgemäß die errechnete Gewichtsfunktion (vgl. Abb.3.3.1), wobei jedoch sehr kleine Potentiale durch den Schwellenoffset unterdrückt werden. Es tritt eine Reduzierung großer Raten bei X-Zellen mit Membranpotential  $\tilde{\Phi}$  auf (Abbildung 3.4.1 rechts).

In einer weiteren Simulationsreihe wird der Punktstimulus nur im Zentrum des rezeptiven Feldes positioniert und die Grauwertdifferenz variiert. Das Ergebnis ist in Abbildung 3.4.2 dargestellt. Bei unverändertem Membranpotential zeigen die X-Zellen oberhalb eines durch den Schwellenoffset bedingten Schwellenwertes näherungsweise eine lineare Abhängigkeit der Ausgangsrate von der Grauwertdifferenz des Stimulus. Bei X-Zellen mit Membranpotential  $\tilde{\Phi}$  hingegen spiegelt sich der sigmoidale Verlauf der Potentialskalierungsfunktion in den Ausgangsraten wider.



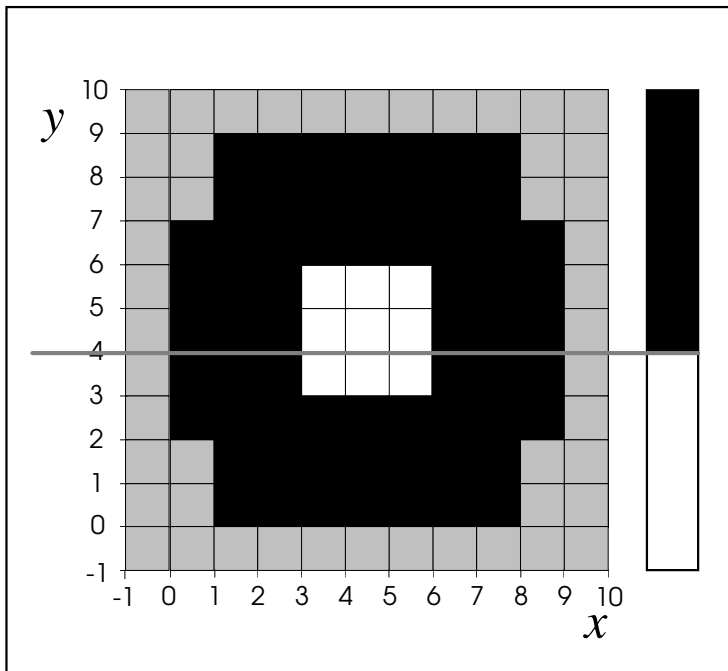


**Abb. 3.4.1:** Ausgangsraten  $a$  der stationären Antwort einer X-Zelle auf Punktstimulus mit  $\Delta G = 255$ . **Links:** bei X-Zelle mit Membranpotential  $\Phi$ . **Rechts:** bei X-Zelle mit Membranpotential  $\tilde{\Phi}$ .

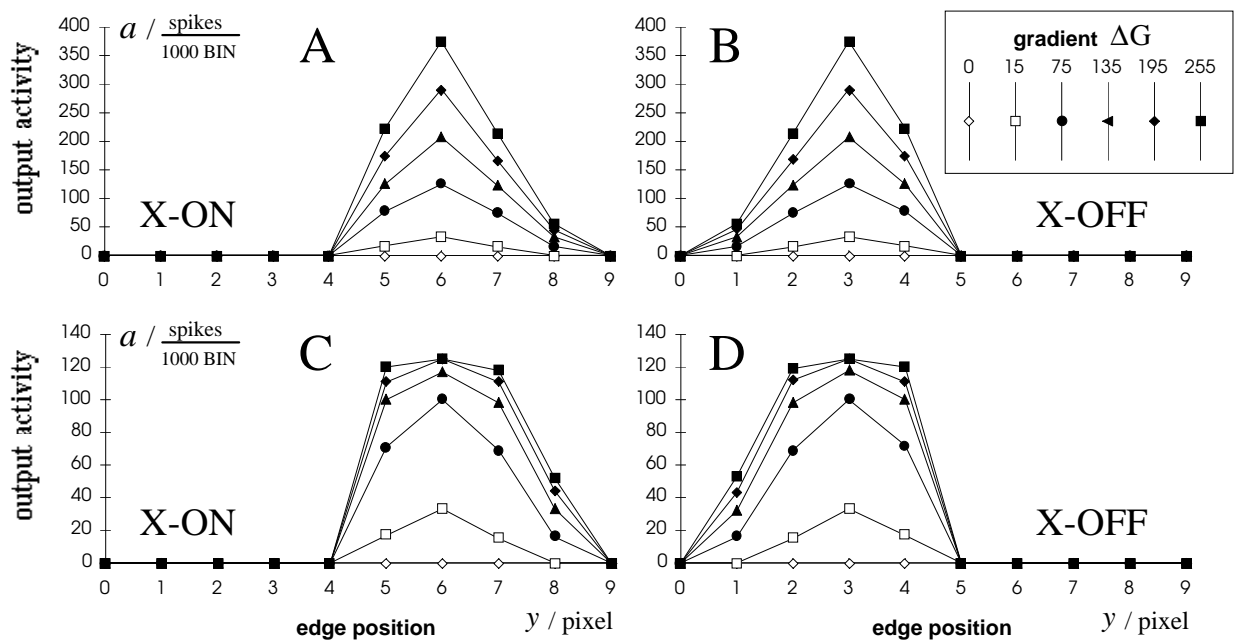


**Abb. 3.4.2:** Ausgangsraten  $a$  der stationären Antwort einer X-Zelle auf einen im Zentrum des rezeptiven Feldes positionierten Punktstimulus als Funktion der Grauwertdifferenz. **Kästchen:** X-Zelle mit unverändertem Membranpotential  $\Phi$ . **Dreiecke:** X-Zelle mit transformiertem Membranpotential  $\tilde{\Phi}$ .

Ein einzelner Punkt kann sich innerhalb einer Bildsequenz, möglicherweise auch durch Rauschen bedingt, sehr stark von seiner Umgebung im Grauwert unterscheiden. Daher ist es wichtig, zu wissen, wie eine X-Zelle auf einen Punktstimulus antwortet. In der Regel sollen aber Objektkanten detektiert werden, weshalb als nächstes das Antwortverhalten einer X-Zelle auf Kanten zu untersuchen ist. Bei dieser Simulation wird im Vergleich zur ersten Simulation lediglich ein anderer Stimulus verwendet. Eine ideale Kante mit einer Orientierung von  $0^\circ$  und bestimmter Grauwertdifferenz  $\Delta G$  wird relativ zum rezeptiven Feld der X-Zelle fest positioniert (statischer Stimulus). Die Simulation wird für unterschiedliche Positionen  $y$  der Kante relativ zum rezeptiven Feld wiederholt. Abbildung 3.4.3 zeigt die Anordnung für eine Kante der Position  $y=4$  auf dem rezeptiven Feld einer X-ON-Zelle, und Abbildung 3.4.4 zeigt die Ausgangsraten  $a$  der X-Zellen in Abhängigkeit von der Verschiebung  $y$  der Kante bei unterschiedlichen Grauwertdifferenzen als Ergebnis der Simulation.



**Abb. 3.4.3:** Ideale Hell/Dunkel-Kante mit Position  $y=4$  auf rezeptivem Feld der Abtastweite  $d=2$  einer X-ON-Zelle. Die negativen Werte (Surround) des rezeptiven Feldes sind schwarz, die positiven Werte (Center) sind weiß und Nullwerte sind grau dargestellt. Die graue Linie gibt die Position der idealen Kante an, wobei der schwarz-weiße Balken rechts andeutet, daß die Kante für  $y>4$  eine geringere Intensität (dunkel) aufweist, als für  $y\leq 4$  (hell).



**Abb. 3.4.4:** Ausgangsraten  $a$  von modellierten X-Zellen in Abhängigkeit von der Verschiebung  $y$  einer idealen Kante bei unterschiedlichen Grauwertintensitätsdifferenzen  $\Delta G$ . **A:** X-ON-Zelle mit Membranpotential  $\Phi$ . **B:** X-OFF-Zelle mit Membranpotential  $\Phi$ . **C:** X-ON-Zelle mit Membranpotential  $\tilde{\Phi}$ . **D:** X-OFF-Zelle mit Membranpotential  $\tilde{\Phi}$ .

Unabhängig von verwendetem Potential und verwendeter Polarität kann jede X-Zelle auf eine ideale Kante nur dann antworten, wenn diese in einer von jeweils vier  $y$ -Positionen dem rezeptiven Feld angeboten wird. Hierbei sind die Antworten der X-OFF-Zellen räumlich versetzt zu denen der im selben Ort angeordneten X-ON-Zellen. Diese Eigenschaft ergibt sich sofort aus der Geometrie der rezeptiven Felder. Weiterhin gibt es für alle untersuchten X-Zellen eine Kantenposition, bei der die X-Zelle am stärksten antwortet. Die X-Zellen mit

Membranpotential  $\tilde{\Phi}$  antworten bei starken Grauwertdifferenzen  $\Delta G$  mit ähnlich großen Raten für drei Kantenpositionen, wodurch eine gewisse Positionsinvarianz erzeugt wird. Bei starkem Intensitätskontrast hingegen weisen die X-Zellen mit Membranpotential  $\tilde{\Phi}$  erwartungsgemäß deutlich geringere Raten auf, als die X-Zellen mit unverändertem Potential  $\Phi$ . Bei geringen Grauwertdifferenzen ist das Antwortverhalten für die Zellen mit unterschiedlich ermitteltem Membranpotential sehr ähnlich, was auf die sigmoidale Potentialskalierungsfunktion zurückzuführen ist, die ja kleine Potentiale nahezu unverändert abbildet.

In einem zu verarbeitenden Bild treten in der Regel nicht nur Kanten oder einzelne Punktmuster auf. So kommen beispielsweise auch Ecken vor, auf die eine X-Zelle stärker antwortet als auf ein Kante. In der nächsten Simulation sollen daher die größtmöglichen Antwortraten der X-Zellen ermittelt werden. Dazu werden diese mit dem Stimulus gereizt, der dem rezeptiven Feld entsprechend die stärkste Antwort hervorruft. Der optimale Stimulus für eine Modell-X-ON-Zelle besteht aus einem  $3 \times 3$  Pixel großen Bereich mit maximalem Grauwert  $G=255$  auf einem Hintergrund mit minimalem Grauwert  $G=0$  (vgl. Abbildung 3.4.3). In verschiedenen Simulationsdurchgängen wird dabei die Grauwertdifferenz  $\Delta G$  variiert. Die X-Zelle mit Membranpotential  $\Phi$  zeigt, wie bei der Stimulierung mit einem einzelnen Punkt, eine lineare Abhängigkeit der Ausgangsaktivität von der Intensitätsdifferenz, wobei jedoch Raten bis zu 100% auftreten. Diese großen Raten sind nicht erwünscht, da eine spärliche Kodierung angestrebt wird. Das biologische System kommt mit Raten unter 10% aus, und bei der Verarbeitung im künstlichen System kann mit kleinen Raten unter Umständen Rechenzeit gespart werden. Darüberhinaus ist die Tatsache, daß die X-Zelle eine deutlich größere Rate erzielen kann, als sie bei Stimulierung mit einer idealen Kante erreicht wird, problematisch für die anschließende Liniendetektion. Durch geeignete Wahl kleiner Gewichte des rezeptiven Feldes ist zwar sofort eine Ratenreduktion möglich, - die aus Verringerung der Gewichte resultierende Kurve unterscheidet sich von der aufgetragenen Kurve lediglich in ihrer Steigung - aber sie führt auch zu äußerst kleinen Raten bei kleinen Grauwertdifferenzen. Im Gegensatz dazu antwortet die X-Zelle mit Membranpotential  $\tilde{\Phi}$  bereits deutlich auch auf geringe Grauwertdifferenzen, und zugleich ist ihre größtmögliche Ausgangsrate geeignet, um eine spärliche Kodierung zu ermöglichen.

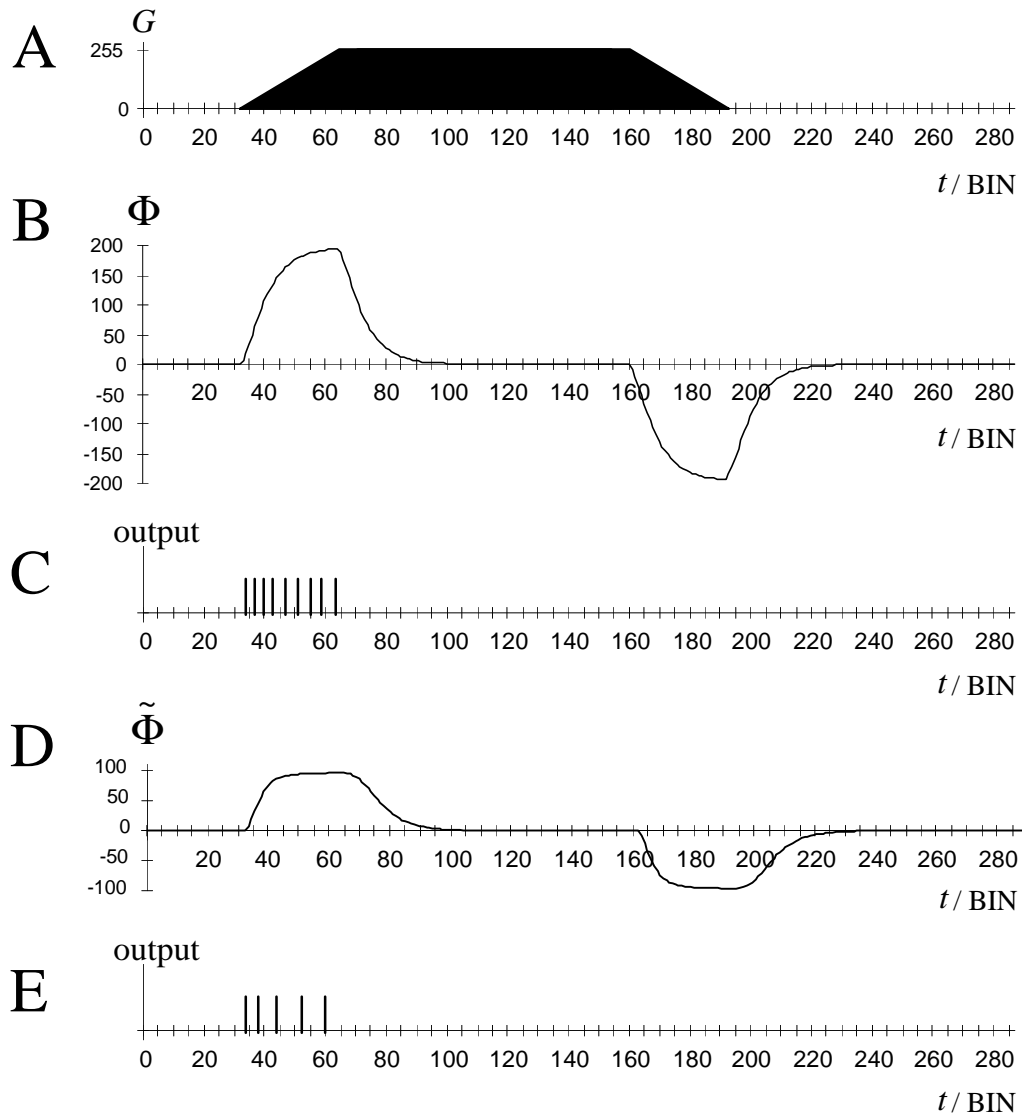
### 3.4.2 Transientes Antwortverhalten der Modell-X-Zelle

Nach der Untersuchung des stationären Anteils der Antwort einer X-Zelle, wird nun auf den transienten Teil der Antwort eingegangen. In der nächsten Simulation wird das Verhalten der X-Zelle bei einem eingangsseitigen Ein- bzw. Ausschaltvorgang untersucht. Der Stimulus ist ein zunächst homogen schwarzes Grauwertbild ( $G=0$ ), das nach einer bestimmten Zeit in ein vollständig weißes Grauwertbild ( $G=255$ ) übergeht und anschließend, nach einem weiteren Zeitintervall, wieder in ein schwarzes überführt wird. Die Übergänge verlaufen jeweils,

bedingt durch die Interpolation, linear über ein Zeitintervall von 32 BIN. Betrachtet werden das Membranpotential  $\Phi$  bzw.  $\tilde{\Phi}$  und die Ausgangsaktivität der X-Zelle, die in Abbildung 3.4.5 für die X-ON-Zellen dargestellt sind. Das Membranpotential der X-ON-Zelle wächst während des Einschaltvorgangs an und entlädt sich nach Ende des Einschaltvorgangs. Bedingt durch die geeignet gewählten Parameter fällt das Membranpotential auf den Wert Null ab. Beim Ausschaltvorgang zeigt die X-Zelle qualitativ das gleiche Verhalten wie beim Einschaltvorgang, jedoch mit umgekehrtem Vorzeichen. Die X-ON-Zellen generieren im Zusammenhang mit dem Einschaltvorgang eine kurze Folge von Aktionspotentialen, die beim Membranpotential  $\Phi$  stärker ausfällt als beim Membranpotential  $\tilde{\Phi}$ . Das Membranpotential einer X-OFF-Zelle ist zwar nicht explizit dargestellt, unterscheidet sich aber von dem einer X-ON-Zelle lediglich im Vorzeichen, so daß die X-OFF-Zellen im Gegensatz zu den X-ON-Zellen beim Ausschaltvorgang aktiv sind.

In der letzten Simulation wird die X-Zelle mit einer bewegten idealen Kante stimuliert. Diese Untersuchung ist aus zwei Gründen wichtig. Einerseits ist mit Antworten zu rechnen, die sowohl transiente als auch stationäre Anteile enthalten, andererseits soll auch Bewegung eine gewisse Rolle im Kontur/Form-System spielen. Dazu wird eine ideale Kante ( $\Delta G = 255$ ) der Orientierung  $\varphi = 0^\circ$  senkrecht zur Kantenorientierung über das rezeptive Feld der X-ON-Zelle (Abtastweite  $d=2$ ) mit der Geschwindigkeit  $v$  geschoben. Wie bereits oben beschrieben, liefert die lineare Interpolation keine exakte Rekonstruktion der fehlenden Zwischen-Grauwertbilder, falls die Verschiebungsgeschwindigkeit  $v$  von 1 Pixel/Bild abweicht. In diesen Fällen sind die Simulationsergebnisse auch von der Startposition der Kante abhängig. Für die Verschiebungsgeschwindigkeit  $v = 0, 1, 2, 3, \dots$  Pixel/Bild gibt es genau  $|v|$  diskrete Startpositionen, bei denen die Verschiebung ein prinzipiell unterschiedliches Ergebnis liefert. Daher ist es sinnvoll, nur gemittelte Werte mit ihren Abweichungen zu betrachten (Abb. 3.4.3). Die Ausgangsraten von X-ON- und X-OFF-Zelle werden ermittelt, wobei Membranpotentialtyp ( $\Phi$  bzw.  $\tilde{\Phi}$ ), die Verschiebungsrichtung der Kante (längs der y-Achse, vgl. Abb. 3.4.3), die Verschiebungsgeschwindigkeit ( $v = 1, 2, 3, 4$  Pixel/Bild) und entsprechend die Startposition der Kante variiert werden.

In Vorwärtsrichtung ist das rezeptive Feld zunächst vollständig hell ( $G = 255$ ) ausgeleuchtet und nach Überstreichen der Kante vollständig dunkel ( $G = 0$ ). Entsprechend umgekehrt verhält es sich bei Umkehrung der Bewegungsrichtung. Es zeigt sich, daß die Ergebnisse einer X-ON-Zelle identisch denen der X-OFF-Zelle bei umgekehrter Bewegungsrichtung sind. Daher werden im folgenden nur die Ergebnisse im Zusammenhang mit der X-ON-Zelle diskutiert. In Abbildung 3.4.6 sind die über die Startpositionen gemittelte Anzahl  $\bar{s}$  der in der Antwort enthaltenen Aktionspotentiale und die mittlere Dauer  $\bar{d}$  der Antwort in Abhängigkeit von der Geschwindigkeit  $v$  aufgetragen. Unabhängig vom X-Zellentyp beinhaltet die Antwort eine um so größere Anzahl an Aktionspotentialen und ist umso länger, je kleiner die Geschwindigkeit der Kante ist, was auf die immer länger werdende Dauer der Stimulierung zurückzuführen ist.

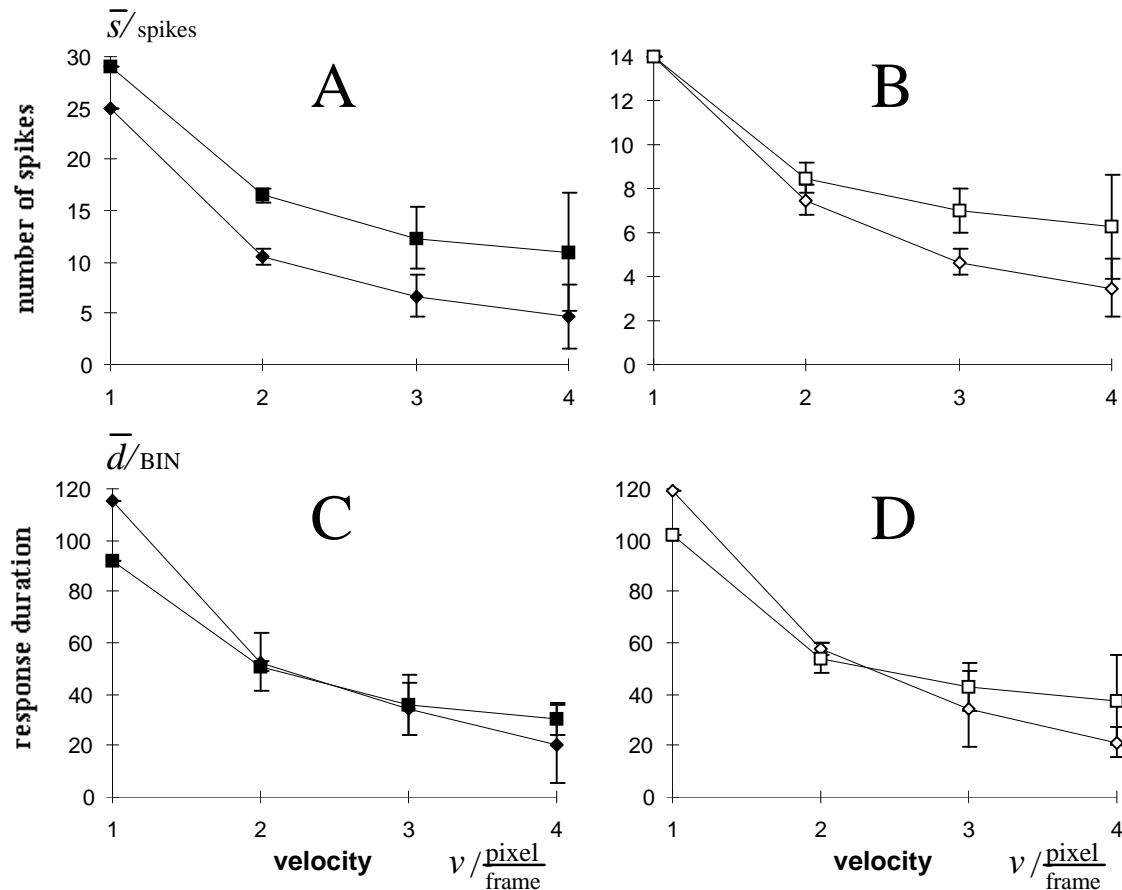


**Abb. 3.4.5:** Antwortverhalten der X-ON-Zelle bei Ein- und Ausschaltvorgang. **A:** Zeitliche Struktur des Stimulus, der Zentrum und Peripherie des rezeptiven Feldes homogen "beleuchtet". **B:** Zeitlicher Verlauf des Membranpotentials einer X-ON-Zelle. **C:** Aktionspotentialfolge der transienten Antwort. **D:** Zeitlicher Verlauf des logarithmisch skalierten Membranpotentials einer X-ON-Zelle. **E:** Aktionspotentialfolge der transienten Antwort bei logarithmisch skaliertem Membranpotential.

Im Grenzfall werden die Antwortdauer und die Anzahl der Aktionspotentiale unendlich groß, da, wie oben gezeigt, bei statischem Reiz eine stationäre Antwort erzeugt wird.

Der Vergleich der Antworten in Vorwärts- und Rückwärtsrichtung der Bewegung zeigt eine gewisse Richtungsselektivität. Im allgemeinen generieren die X-Zellen eine größere Anzahl an Aktionspotentialen bei Stimulierung in Rückwärtsrichtung. Die Dauer der Antwort ist bei kleinen Geschwindigkeiten in Vorwärtsrichtung länger, während sie bei größeren Geschwindigkeiten in Rückwärtsrichtung größer ist, was sich wie folgt erklären läßt. Überstreicht die Kante das rezeptive Feld in Vorwärtsrichtung, so wächst zunächst das Membranpotential durch Verringerung des inhibitorisch wirkenden Surround-Potentials  $\Phi_s$ , das durch einen Leckintegrator mit relativ großer Zeitkonstante  $\tau^s$  modelliert wird. Wenn die Kante anschließend das Zentrum der Zelle erreicht, wird auch das exzitatorische Center-Potential herabgesetzt, so daß sich das Membranpotential wieder verringert. Das Herabsetzen

des Center-Potentials  $\Phi_c$  geschieht aufgrund der kleinen Zeitkonstanten jedoch  $\tau^c$  vergleichsweise schnell. Eine schnell bewegte Kante wird das Membranpotential aufgrund des langsamen Anhebens und des anschließend raschen Absenkens nur relativ wenig vergrößern können.



**Abb. 3.4.6:** Anzahl der Aktionspotentiale und Dauer der X-Zellenantwort auf eine bewegte ideale Kante der Geschwindigkeit  $v$ . **A:** Über die Startpositionen der Kante gemittelte Anzahl der Aktionspotentiale in der Antwort einer X-Zelle mit Membranpotential  $\Phi$  bei Überstreichen des rezeptiven Feldes durch eine ideale Kante der Geschwindigkeit  $v$  in Vorwärts- (Rauten) und in Rückwärtsrichtung (Quadrate). **B:** wie A, jedoch für X-Zelle mit Membranpotential  $\tilde{\Phi}$ . **C:** Über die Startpositionen der Kante gemittelte Dauer der Antwort einer X-Zelle mit Membranpotential  $\Phi$  bei Überstreichen des rezeptiven Feldes durch eine ideale Kante der Geschwindigkeit  $v$  in Vorwärts- (Rauten) und in Rückwärtsrichtung (Quadrate). **D:** wie C, jedoch für X-Zelle mit Membranpotential  $\tilde{\Phi}$ . (Die Fehlerbalken geben die Standardabweichungen an).

Überstreicht hingegen die Kante das rezeptive Feld in Rückwärtsrichtung, so wird das Membranpotential zunächst durch langsames Anheben des inhibitorisch wirkenden Surround-Potentials verringert und anschließend durch einen Aufladeprozess des exzitatorisch wirkenden Centerpotentials, das mit relativ kleiner Zeitkonstante modelliert wird, vergrößert. Eine schnell bewegte Kante hat hier aufgrund des langsamen Absenkens und anschließend raschen Anhebens einen größeren Effekt auf das Membranpotential. Somit ist die Antwortstärke in Rückwärtsrichtung für große Geschwindigkeiten relativ zur Antwortstärke in Vorwärtsrichtung stärker. Der Vergleich von X-Zellen unterschiedlichen Membranpotential-

typs zeigt keine signifikanten Unterschiede bei der Antwortdauer. Die geringere Anzahl an Aktionspotentialen bei X-Zellen mit Membranpotential  $\tilde{\Phi}$  ist eine Folge der Potentialbegrenzung. Aufgrund der Interpolation der Grauwertbilder treten bei großen Geschwindigkeiten zum Teil deutliche Schwankungen (die Standardabweichungen sind eingezeichnet) der Werte auf, die sich durch Vergrößerung der Bildabtastrate verringern bzw. ausschließen ließen.

### 3.5 Diskussion

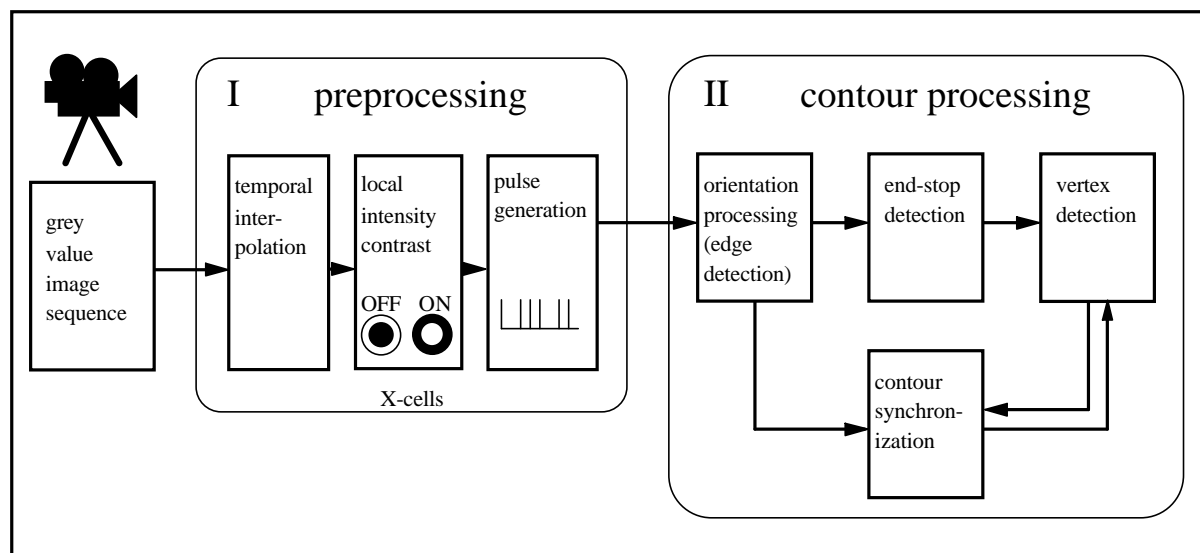
Im vorliegenden Modell werden viele wichtige Eigenschaften der rezeptiven Felder von natürlichen X-Zellen berücksichtigt. Die wesentlichen räumlichen Eigenschaften sind die Center/Surround-Struktur und die Unterscheidung zwischen ON- und OFF-Typ, was im Modell durch Überlagerung von Mexikanerhut-Funktionen (DOOG-Modell) mit unterschiedlichem Vorzeichen realisiert wird.

Auch wichtige zeitliche Eigenschaften werden modelliert. Bei Beleuchtung des Zentrums einer natürlichen X-ON-Zelle steigt deren Ausgangsrate stark an und durch Beleuchtung der Peripherie wird sie abgesenkt. Im Modell können durch Reizung des Zentrums bzw. der Peripherie vergleichbare Effekte erzielt werden (Abb. 3.4.2).

Neben der stationären Antwort spielt auch der transiente Teil der Antwort eine Rolle in der natürlichen X-Zelle, die sich beispielsweise beim Einschalt- oder Ausschalteffekt bemerkbar machen, jedoch fällt die transiente Antwort schwächer aus als bei einer natürlichen Y-Zelle. Im Modell wird ein vergleichbares Antwortverhalten durch Verwendung von Leckintegratoren mit unterschiedlichen Zeitkonstanten für Zentrum und Peripherie erzielt (Abb. 3.4.5), was auch Auswirkungen, wie z. B. eine geringe Richtungsselektivität, für das Antwortverhalten der X-Zellen auf bewegte Reizmuster hat (Abb. 3.4.6).

## 4 Kontur/Form-System

Nachdem im vorhergehenden Kapitel die Extraktion lokaler Intensitätskontraste durch die Vorverarbeitungsstufe beschrieben wurde, werden in diesem Kapitel nun die einzelnen nachfolgenden Verarbeitungsstufen des Kontur/Form-Systems untersucht. Abbildung 4.0.1 gibt eine erste Übersicht über das Gesamtsystem.



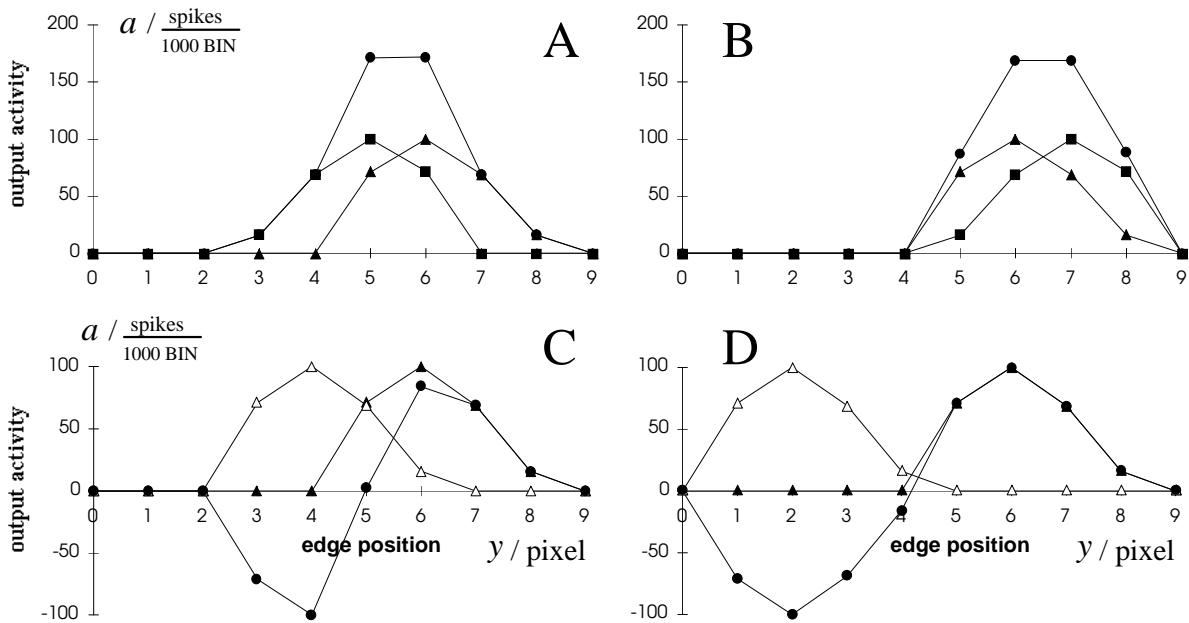
**Abb. 4.0.1:** Struktur des Kontur/Form-Systems. Die Vorverarbeitung (preprocessing) extrahiert aus Sequenzen von Grauwertbildern lokale Intensitätskontraste in Form von Pulsfolgen. Auf der Grundlage dieser Pulsfolgen werden orientierte Kanten extrahiert. Aus der Aktivität der Kantendetektoren werden in weiteren Verarbeitungsstufen Repräsentationen von Linienenden und daraus wiederum Vertizes gewonnen. Die Neuronen, die diese Vertizes kodieren, wechselwirken mit einer getrennten Kontur-Synchronisationsstufe, die die Information über die Kanten enthält, so daß synchrone Aktivität zur Definition von Objektkandidaten erzeugt wird. Eine inhibitorische Rückkopplung von der Synchronisationsstufe auf die Vertex-Repräsentationen ermöglicht schließlich die Separation dieser Objektkandidaten im Zeitbereich.

### 4.1 Kontrastkantendetektion

Wie bereits in Kapitel 2 erwähnt, fanden Hubel und Wiesel bei Untersuchungen an der Großhirnrinde der Katze orientierungsselektive kortikale Zellen, die auf Balken- oder Hell/Dunkel-Reize antworten und in Spalten, den Hyperkolumnen, angeordnet sind (Hubel und Wiesel 1962, Hubel und Wiesel 1965). Vor diesem Hintergrund werden nun auch im Modell Neuronen implementiert, deren Antwortverhalten spezifisch für orientierte Kontrastkanten ist. In jedem Abtastpunkt des hexagonalen Rasters werden Orientierungsneuronen mit unterschiedlichen, aber sich überdeckenden rezeptiven Feldern angeordnet. Die Vorzugsorientierung variiert dabei von  $0^\circ$  bis  $150^\circ$  in  $30^\circ$ -Schritten. Unter Berücksichtigung der zwei Kontrastpolaritäten ergeben sich somit zwölf nachbarschaftserhaltende (topologische) neuronale Schichten. Da CGL-Zellen ähnliche rezeptive Felder wie Ganglienzellen aufweisen,



werden sie im künstlichen Netzwerk aus Gründen der Rechenzeitersparnis nicht explizit modelliert. Zur Modellierung der rezeptiven Felder von Ganglienzellen werden weitgehend kollinear angeordnete X-Zellen konvergent auf die Kantendetektorneuronen verschaltet.

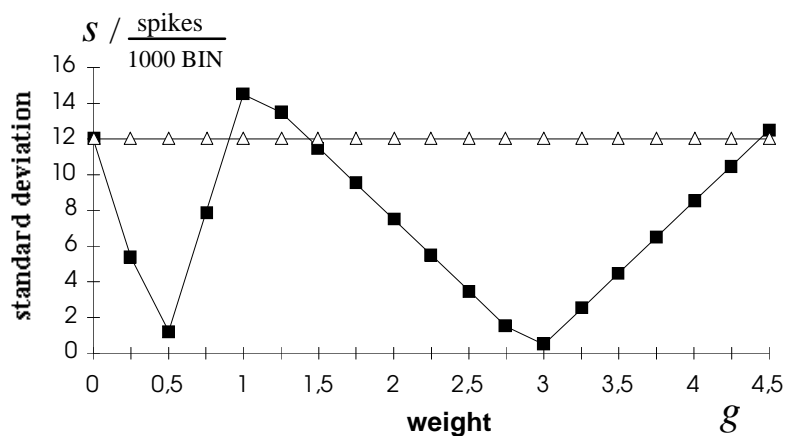


**Abb. 4.1.1:** Gesamtausgangsimpulsrate  $a$  von zwei zueinander um  $\Delta y$  versetzten X-Zellen auf eine ideale Kante der Position  $y$ . **A:** Summe (Kreise) der Ausgangsraten einer X-ON-Zelle (Dreiecke) und einer dazu um  $\Delta y = 2$  versetzten X-OFF-Zelle (Kästchen). **B:** wie A, jedoch  $\Delta y = 4$ . **C:** Summe (Kreise) der Ausgangsraten einer exzitatorisch wirkenden X-ON-Zelle (gefüllte Dreiecke) und einer dazu um  $\Delta y = 2$  versetzten inhibitorisch wirkenden X-ON-Zelle (offene Dreiecke). **D:** wie C, jedoch  $\Delta y = 4$ .

Um ein möglichst großes Auflösungsvermögen der vorverarbeitenden Schichten zu erreichen, müssen die kleinstmögliche Abtastweite ( $d=2$ ) und Kantendetektoren mit möglichst kurzen rezeptiven Feldern verwendet werden. Kantendetektoren mit längeren rezeptiven Feldern erhöhen zwar bei länger ausgedehnten Kanten die Genauigkeit der Detektion, sind aber kaum geeignet für die Detektion kurzer Kantenstücke. Kantendetektoren mit zu kurzen rezeptiven Feldern zeigen hingegen zu geringe Orientierungsselektivität. Ein geeigneter Kompromiß wird durch kollineare Verschaltung von jeweils etwa drei bis vier benachbarten X-Zellen zu einer entlang der räumlichen Vorzugsorientierung ausgerichteten Teilregion des rezeptiven Feldes erreicht. Für die Konstruktion solcher Teilregionen gibt es vier grundsätzliche Verschaltungstypen, nämlich exzitatorische oder inhibitorische Verbindungen von X-ON-Zellen oder X-OFF-Zellen zum Detektorneuron. Zunächst muß geklärt werden, welchen Abstand die Zentren der rezeptiven Felder von X-Zellen der beiden unterschiedlichen Teilregionen zueinander haben müssen, um ein geeignetes Antwortverhalten des Kantendetektors zu bewirken. Dazu können die Untersuchungsergebnisse der X-Zellen aus Kapitel 3 herangezogen werden. In Abbildung 3.4.4 sind die Ausgangsimpulsraten der X-Zellen in Abhängigkeit der Verschiebung einer stimulierenden idealen Kante dargestellt. Hier wird

deutlich, daß eine X-ON-Zelle eine um drei Pixel versetzte bevorzugte Kantenposition im Vergleich zur X-OFF-Zelle besitzt. Folglich ist die Summe der Ausgangsaktionspotentiale beider Zellen genau dann am größten, wenn die beiden X-Zellen um drei Pixel zueinander versetzt sind und die Kante sich in der nun gemeinsamen Vorzugsposition befindet. Das hexagonale Abtastraster der Abtastweite  $d=2$  erlaubt jedoch nur einen Versatz  $\Delta y$  um ganzzahlige Vielfache von  $d$ .

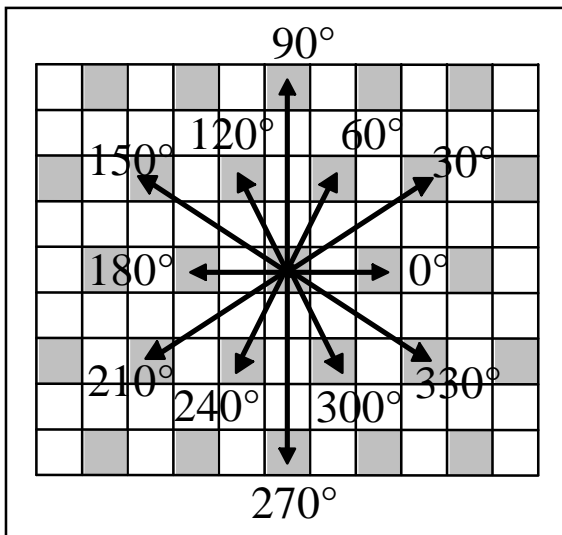
Die Abbildungen 4.1.1 A und B zeigen die Summen der generierten Aktionspotentiale von X-ON- und X-OFF-Zelle bei verschiedenen Kantenpositionen für  $\Delta y=2$  bzw. 4. Die Ergebnisse sind sich sehr ähnlich. In beiden Fällen gibt es zwei Kantenpositionen (y-Position 5 und 6 bzw. 6 und 7), die zur stärksten gemeinsamen Antwort führen. Darüberhinaus liegt ein bezüglich der Kantenposition symmetrisches Gesamtantwortverhalten vor, das eine Gleichbehandlung aller Kantenpositionen (in y-Richtung) gewährleistet, so daß hier die Streuung des jeweiligen Ergebnisses aus Superposition der Gesamtantworten unter Berücksichtigung des Abtastrasters ( $d=2$ ) verschwindet. Teilregionen des rezeptiven Feldes eines Kantendetektors, die sich nur in der Polarität der X-Zellen unterscheiden, können somit wahlweise im Abstand  $\Delta y=2$  oder  $\Delta y=4$  angeordnet werden. In den Abbildungen 4.1.1 C und D sind die Differenzen aus den Antwortraten einer exzitatorisch wirkenden X-ON-Zelle und den Antwortraten einer um  $\Delta y=2$  bzw. 4 versetzten inhibitorisch wirkenden X-ON-Zelle bei unterschiedlichen Kantenpositionen aufgetragen. In beiden Fällen wird näherungsweise ein ungerades Symmetrieverhalten erreicht, im Gegensatz zum Gesamtantwortverhalten, das sich aus der Überlagerung der Antworten aus X-ON- und X-OFF-Zelle ergibt. Ein kleinerer Versatz  $\Delta y$  führt zu einem stärkeren Gradienten und verringerten Amplituden in der Funktionskurve des Gesamtantwortverhaltens. Die Streuung des Ergebnisses aus Superposition der positiven Anteile in der Gesamtantwort unter Berücksichtigung des Abtastrasters bei  $\Delta y=4$  fällt zunächst ähnlich wie die entsprechende Streuung bei  $\Delta y=2$  aus (vgl. Abb. 4.1.2 für  $g=1$ ). Dies legt nahe, den Versatz zwischen exzitatorisch und inhibitorisch wirkenden Teilregionen des rezeptiven Feldes eines Kantendetektors beliebig mit  $\Delta y=4$  oder 2 wählen zu können.



**Abb. 4.1.2:** Streuung  $s$  des Ergebnisses aus Überlagerung der Gesamtantworten von jeweils einer exzitatorisch und einer inhibitorisch wirkenden und um  $\Delta y=2$  (Quadrate) bzw. um  $\Delta y=4$  (Dreiecke) versetzten X-ON-Zelle auf eine ideale statische Kante um die mittlere Gesamtrate  $\bar{a}$  unter Berücksichtigung des hexagonalen Abtastrasters in y-Richtung. Die Impulsraten der inhibitorisch wirkenden X-Zelle werden dabei mit dem Faktor  $g$  gewichtet.

Wird aber zusätzlich eine Gewichtung  $g$  des inhibitorischen Anteils eingeführt, so läßt sich die Streuung für  $\Delta y=2$  erheblich verringern. Der Gewichtungsfaktor hat jedoch keine Wirkung auf die Streuung bei  $\Delta y=4$ , da in diesem Fall der exzitatorische Bereich unabhängig vom inhibitorischen Bereich in der Gesamtausgangsimpulsrate bleibt. Mit einem Gewichtungsfaktor  $g$  von etwa 0,5 oder 3 wird in guter Näherung die geforderte Gleichbehandlung der Abtastpunkte bei einem Versatz von  $\Delta y=2$  erzielt, wie auch aus Abbildung 4.1.2 ersichtlich ist.

Nachdem Klarheit über geeignete Abstände der X-Zellen aus möglichen Teilbereichen des rezeptiven Feldes von Kantendetektoren geschaffen wurde, ist nun auf die Einpassung der zwölf Orientierungen in das hexagonale Abtastraster einzugehen. Während die Orientierungen  $0^\circ$ ,  $180^\circ$ ,  $90^\circ$  und  $270^\circ$  im Raster bereits zwischen direkt benachbarten Abtastpunkten gefunden werden können, lassen sich die übrigen acht Orientierungen nur mit Abweichungen realisieren. Diesen Sachverhalt veranschaulicht die Abbildung 4.1.3.



**Abb. 4.1.3:** Einpassung der zwölf Orientierungen in das Abtastraster. Die Orientierungen  $30^\circ$  und  $60^\circ$  lassen sich im Gegensatz zu den Orientierungen  $0^\circ$  und  $90^\circ$  bei Verwendung der nächsten Nachbarn im hexagonalen Abtastraster (graue Felder) nur in Näherung realisieren. Gleiches gilt entsprechend für alle größeren Winkel (nach: Spengler 1996). Diese Einpassung stellt ein rein technisches Problem dar, das aber aufgrund seiner möglicherweise einschränkenden Auswirkungen auf die Behandlung des eigentlichen Problems, d.h. der Kantendetektion, berücksichtigt werden muß.

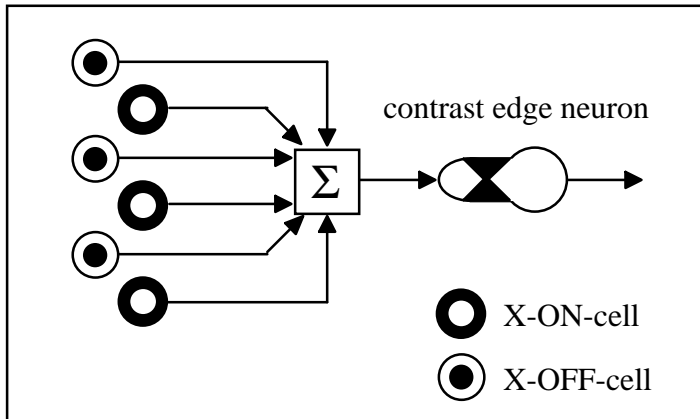
Die Orientierung von  $30^\circ$  läßt sich durch einen Winkel von etwa  $33,69^\circ$  und die Orientierung von  $60^\circ$  durch einen Winkel von etwa  $63,43^\circ$  auf einfache Weise im hexagonalen Abtastraster annähern, wobei dies auch analog für die Orientierungen  $120^\circ$ ,  $150^\circ$ ,  $210^\circ$ ,  $240^\circ$ ,  $300^\circ$  und  $330^\circ$  gilt. Zudem wird sich zeigen, daß diese Abweichungen keine gewichtigen Nachteile für die weitere Modellierung mit sich bringen.

Im folgenden werden einige Kantendetektortypen untersucht, deren rezeptive Felder sich in der Kombination ihrer Teilbereiche unterscheiden. Im einzelnen sind das:

1. *ON-OFF-Kantendetektor*, dessen rezeptives Feld aus einem exzitatorischen ON- und einem inhibitorischen OFF-Teilbereich besteht.
2. *ON-ON-Kantendetektor*, dessen rezeptives Feld sich aus einem exzitatorischen ON- und einem inhibitorischen ON-Teilbereich zusammensetzt.
3. *ON<sup>2</sup>-OFF<sup>2</sup>-Kantendetektor*, dessen rezeptives Feld aus je zwei ON- und zwei OFF-Teilbereichen besteht.

### 4.1.1 Simple-Zellen Modell: ON-OFF-Kantendetektor

Das rezeptive Feld des ON-OFF-Kantendetektors besteht aus zwei exzitatorisch wirkenden Teilbereichen, die aus der kollinearen Verschaltung von jeweils mindestens drei benachbarten X-ON- und X-OFF-Zellen auf das Detektorneuron resultieren (Abb. 4.1.4).



**Abb. 4.1.4:** Verschaltung von X-ON-Zellen (X-ON-cells) und X-OFF-Zellen (X-OFF-cells) auf ein ON-OFF-Kantendetektorneuron (contrast edge neuron). Die Ausgangssignale aller X-Zellen werden aufsummiert und als Eingangssignal (Feeding) an das Detektorneuron weitergegeben.

Die Vorzugsorientierung jedes Detektors wird maßgeblich durch die konkrete Anordnung der rezeptiven Felder der ihm vorgeschalteten X-Zellen bestimmt. In Abbildung 4.1.5 sind die Positionen dieser rezeptiven Felder relativ zur Position des entsprechenden Detektorneurons im hexagonalen Abtastraster für die zwölf verschiedenen Orientierungen schematisch dargestellt. Die Stärke  ${}^P w_{ijkl}^\varphi$  der Verbindungen zwischen einer X-Zelle der Polarität  $P$  im Ort  $(k, l)$  und einem Kantendetektorneuron der Orientierung  $\varphi$  im Ort  $(i, j)$  des hexagonalen Rasters ergibt sich zu:

$${}^{ON} w_{ijkl}^{0^\circ} = \begin{cases} 1, & \text{falls } (k = i-1 \vee k = i+1 \vee k = i+3) \wedge (l = j+2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.1})$$

$${}^{OFF} w_{ijkl}^{0^\circ} = \begin{cases} 1, & \text{falls } (k = i-1 \vee k = i+1 \vee k = i+3) \wedge (l = j-2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.2})$$

$${}^{ON} w_{ijkl}^{30^\circ} = \begin{cases} 1, & \text{falls } (k = i-2 \wedge l = j) \vee ((k = i-1 \vee k = i+1) \wedge l = j+2) \\ 0.5, & \text{falls } (k = i-4 \wedge l = j) \vee (k = i+2 \wedge l = j+4) \\ 0 & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.3})$$

$${}^{OFF} w_{ijkl}^{30^\circ} = \begin{cases} 1, & \text{falls } (k = i+2 \wedge l = j) \vee ((k = i+1 \vee k = i-1) \wedge l = j-2) \\ 0.5, & \text{falls } (k = i+4 \wedge l = j) \vee (k = i-2 \wedge l = j-4) \\ 0 & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.4})$$

$${}^{ON}W_{ijkl}^{60^\circ} = \begin{cases} 1, & \text{falls } (k = i - 2 \wedge l = j) \vee (k = i - 1 \wedge l = j + 2) \vee (k = i \wedge l = j + 4) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.5})$$

$${}^{OFF}W_{ijkl}^{60^\circ} = \begin{cases} 1, & \text{falls } (k = i + 1 \wedge l = j - 2) \vee (k = i + 2 \wedge l = j) \vee (k = i + 3 \wedge l = j + 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.6})$$

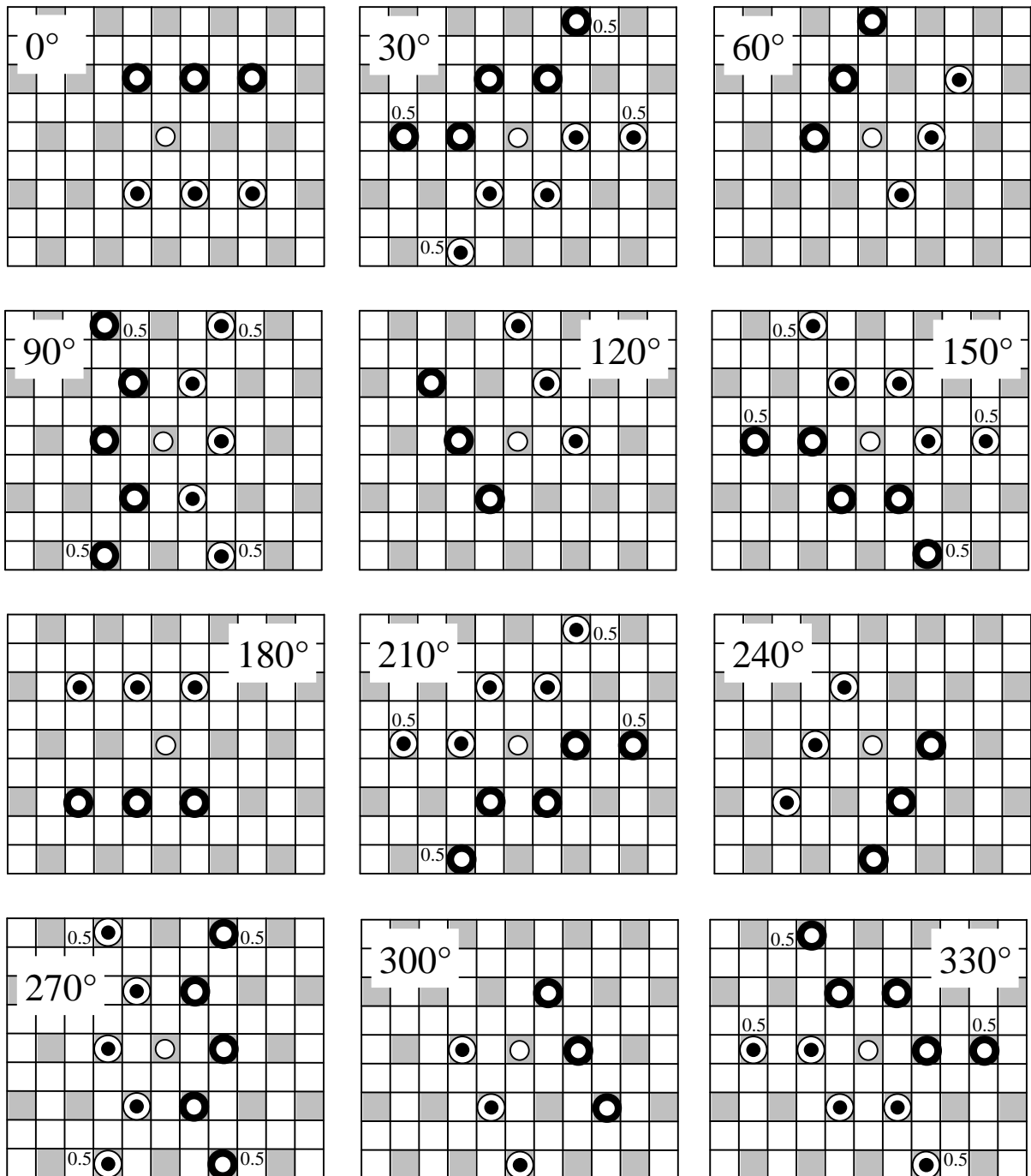
$${}^{ON}W_{ijkl}^{90^\circ} = \begin{cases} 1, & \text{falls } (k = i - 2 \wedge l = j) \vee ((k = i - 1) \wedge (l = j - 2 \vee l = j + 2)) \\ 0.5, & \text{falls } (k = i - 2 \wedge (l = j - 4 \vee l = j + 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.7})$$

$${}^{OFF}W_{ijkl}^{90^\circ} = \begin{cases} 1, & \text{falls } (k = i + 2 \wedge l = j) \vee ((k = i + 1) \wedge (l = j - 2 \vee l = j + 2)) \\ 0.5, & \text{falls } (k = i + 2 \wedge (l = j - 4 \vee l = j + 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.8})$$

Exemplarisch werden nur die Gewichte der Orientierungen  $0^\circ$ ,  $30^\circ$ ,  $60^\circ$  und  $90^\circ$  beschrieben.

Aus der Überlagerung der rezeptiven Felder der X-Zellen läßt sich das rezeptive Feld des jeweiligen Kantendetektors bestimmen. Bei dieser Vorgehensweise bleiben durch Potentialskalierung und Spike-Encoder verursachte Nichtlinearitäten unberücksichtigt. Die Ergebnisse der Superposition zeigt Abbildung 4.1.6 erneut nur für die Orientierungen  $0^\circ$ ,  $30^\circ$ ,  $60^\circ$  und  $90^\circ$ . Aus ihnen ergeben sich die entsprechenden Überlagerungsergebnisse für die übrigen Orientierungen durch Rotation in der xy-Ebene, die daher nicht explizit dargestellt sind. Bei der Berechnung der Superposition werden die rezeptiven Felder der X-ON-Zellen positiv, die der X-OFF-Zellen negativ gewichtet, so daß ein negativer Wert in der Darstellung der Superposition diejenigen Orte angibt, in denen bei Auftreten eines Reizes der Kantendetektor eine stärkere Antwort von X-OFF-Zellen als von X-ON-Zellen erhält. Die Anzahl der jeweils auf das Kantendetektorneuron verschalteten X-Zellen hängt von den Abständen der in Vorzugsorientierung kollinearen Abtastpunkte ab. Während bei einem  $0^\circ$ -Kantendetektor Abstände von genau 2 Pixel zwischen kollinear verschalteten X-Zellen auftreten, zeigen die in  $60^\circ$ -Richtung angeordneten Abtastpunkte Abstände von etwa 2,24 Pixel, in  $30^\circ$ -Richtung Abstände von etwa 3,6 Pixel und in  $90^\circ$ -Richtung sogar 4 Pixel-Abstände, die durch das Abtastraster vorgegeben sind. Würden bei der Konstruktion der rezeptiven Felder nur kollinear angeordnete X-Zellen verwendet werden, so führte das zu starken Schwankungen in den Amplituden der rezeptiven Felder insbesondere beim  $30^\circ$ - und beim  $90^\circ$ -Detektor. Um diese Schwankungen zu verringern, werden hier, wie in den Gleichungen 4.1.3 bis 4.1.8 beschrieben, fünf statt drei X-Zellen berücksichtigt. Weiterhin zeigen sich deutliche Unterschiede in den Längen der rezeptiven Felder. Für eine Annäherung der Längen, werden beim  $30^\circ$ - und beim  $90^\circ$ -Detektor die in Vorzugsrichtung äußeren X-

Zellen nur halb so stark gewichtet wie die übrigen X-Zellen. Entsprechendes gilt auch für die ON-OFF-Kantendetektoren der übrigen Orientierungen. Ein wichtiges Merkmal eines Kantendetektors ist seine Orientierungscharakteristik, die von Länge und Breite des rezeptiven Feldes maßgeblich beeinflusst wird.



**Abb. 4.1.5:** Schematische Darstellung der rezeptiven Felder von ON-OFF-Kantendetektoren für zwölf verschiedene Orientierungen im hexagonalen Abtastraster der Abtastweite  $d=2$ . Die Position des Detektorneurons im Abtastraster ist durch einen weißen Kreis gekennzeichnet. Die konzentrischen Kreise geben die Zentrumspositionen der rezeptiven Felder derjenigen X-Zellen an (außen schwarz, innen weiß: X-ON-Zellen; außen weiß, innen schwarz: X-OFF-Zellen), die ihre Ausgangssignale an das Detektorneuron senden. Bei fünfelementigen Teilbereichen werden die äußeren X-Zellen nur mit halbem Gewicht berücksichtigt. Die Zentren der entsprechenden X-Zellen sind mit "0.5" gekennzeichnet.

Der folgenden Bestimmung von Orientierungscharakteristiken der ON-OFF-Kantendetektoren liegt der in Tabelle 4.1.2 aufgeführte Parametersatz zugrunde. Eine ideale Kante der Grauwertdifferenz  $\Delta G=64$  wird als Stimulus in das rezeptive Feld des Kantendetektors gebracht und die Summe der vom Detektorneuron innerhalb eines Zeitintervalles von 500 BIN generierten Aktionspotentiale bestimmt. Die daraus resultierende Rate  $a$  wird für unterschiedliche Orientierungswinkel  $\varphi$  (in  $5^\circ$ -Schritten) und jeweils für alle relevanten Positionen der Kante ermittelt. In Abbildung 4.1.7 ist die bezüglich der Position maximale Rate des  $0^\circ$ -,  $30^\circ$ -,  $60^\circ$ - und des  $90^\circ$ -Kantendetektors als Funktion der Kantenorientierung dargestellt, wobei sich die Raten der übrigen Detektoren aus den gezeigten Raten durch Rotation der entsprechenden Detektoren ableiten lassen.

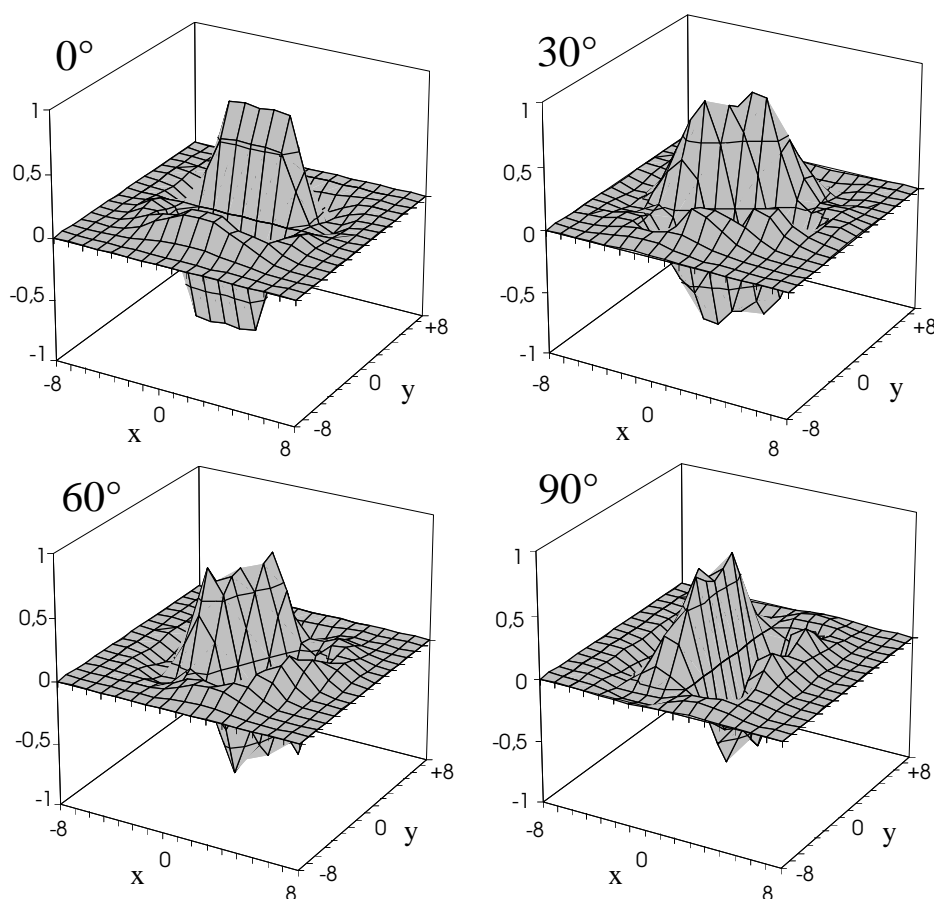
Schwellenverstärkung $V^\Theta$	Schwellenzeitkonstante $\tau^\Theta$	Schwellenoffset $\Theta_0$	Feeding-Potentialverstärkung $V^F$	Feeding-Potentialzeitkonstante $\tau^F$
32	10 BIN	0, 2, 4 und 6	0,7	20 BIN

**Tab. 4.1.2:** Parameter des ON-OFF-Kantendetektorneurons.

Auf diese Weise ergeben sich die Kurven der Orientierungscharakteristiken, die für unterschiedliche Werte des Schwellenoffsets  $\Theta_0$  aufgetragen sind. Bei kleinem Schwellenoffset fallen die Raten über den gesamten Bereich sehr ähnlich, in Vorzugsrichtung etwas stärker aus, was verdeutlicht, daß auch in Nichtvorzugsorientierung eine deutliche Aktivierung der Kantendetektoren gegeben ist. Bei Anhebung des Schwellenoffsets tritt jedoch eine zunehmende Verschärfung der Orientierungscharakteristiken ein. Die Stärke der Detektoraktivierung bei einem in Nichtvorzugsrichtung ausgerichteten Reiz genügt nun in den meisten Fällen nicht mehr, die Schwelle des Spike-Encoders zu überschreiten, so daß die Rate dort verschwindet. Auf eine um  $180^\circ$  zur Vorzugsorientierung gedrehte Kante antworten bevorzugt solche Kantendetektoren, die um die Breite des exzitatorisch wirkenden Teilbereichs von der eigentlichen Kantenposition entfernt angeordnet sind. Diese räumliche Abweichung der Antworten bei Umkehr der Kantenpolarität erschwert die spätere Weiterverarbeitung und ist daher nicht erwünscht, auch wenn sie sich zunächst durch geeignet großen Schwellenoffset weitgehend unterdrücken lassen.

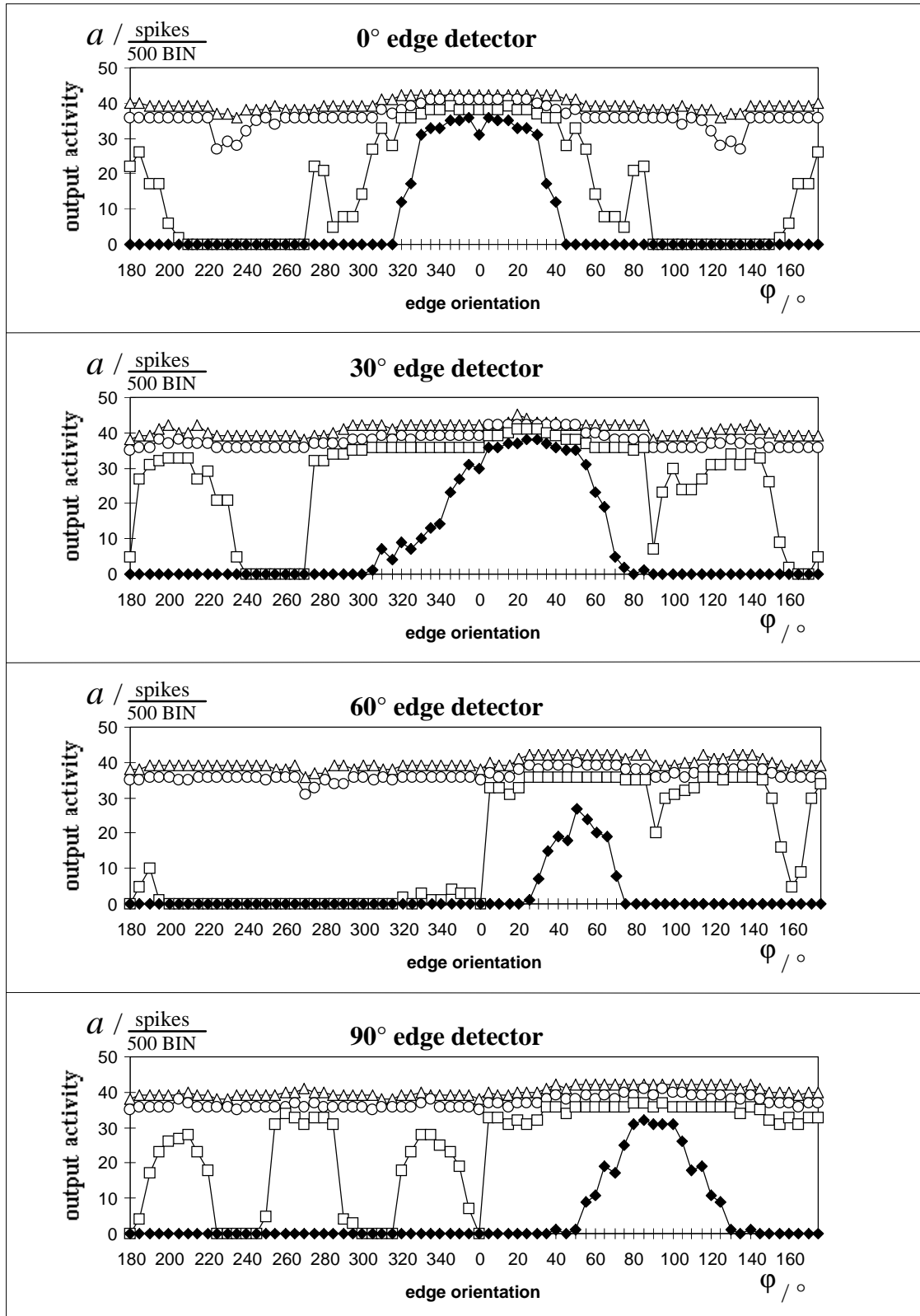
Der  $30^\circ$ - und der  $90^\circ$ -Detektor besitzen deutlich längere rezeptive Felder als der  $0^\circ$ - und der  $60^\circ$ -Detektor. Während die Zentren der äußersten im Teilbereich berücksichtigten X-Zellen etwa 7,2 Pixel beim  $30^\circ$ - und 8 Pixel beim  $90^\circ$ -Detektor voneinander entfernt sind, zeigt sich beim  $0^\circ$ -Detektor ein 4 Pixel-Abstand und beim  $60^\circ$ -Detektor ein Abstand von etwa 4,5 Pixel. Auf der anderen Seite sind bei allen Detektoren die Zentren der orthogonal zur Vorzugsorientierung äußersten X-Zellen etwa 3,6 ( $60^\circ$ -Detektor) bis 4,4 Pixel ( $30^\circ$ -Detektor) voneinander entfernt, so daß in allen Fällen eine ähnliche Breite vorliegt. Das Verhältnis zwischen der so ermittelten Länge und der Breite ist für den  $90^\circ$ -Detektor und den  $30^\circ$ -Detektor besonders groß. Entsprechend sollte eine vergleichsweise scharfe Richtungs-

charakteristik zu erwarten sein. Tatsächlich tritt das genaue Gegenteil auf. Ein Grund dafür ist darin zu sehen, daß diese Detektoren Eingangsaktivität von mehr als drei X-Zellen erhalten, so daß insgesamt eine stärkere Aktivierung zu erwarten ist, die hier zu einer Verbreiterung der Richtungscharakteristik führt. Ähnliche Auswirkungen hat die weniger starke Gewichtung der äußeren X-Zellen, die aber gerade für die große Länge des rezeptiven Feldes verantwortlich sind. Um dennoch vergleichbare Amplituden und Breiten in den Richtungscharakteristiken der ON-OFF Kantendetektoren zu erhalten, bietet sich beispielsweise eine abgestimmte Gewichtung der jeweiligen Gesamteingangsaktivität an.



**Abb. 4.1.6:** *Rezeptive Felder von ON-OFF-Kantendetektoren (Simple-Zellen).* Die rezeptiven Felder ergeben sich aus Überlagerung der rezeptiven Felder von X-Zellen und sind dargestellt für die Orientierungen  $0^\circ$ ,  $30^\circ$ ,  $60^\circ$  und  $90^\circ$ . Die Masken der X-ON-Zellen werden dabei positiv, die der X-OFF-Zellen negativ für die Superposition gewichtet.

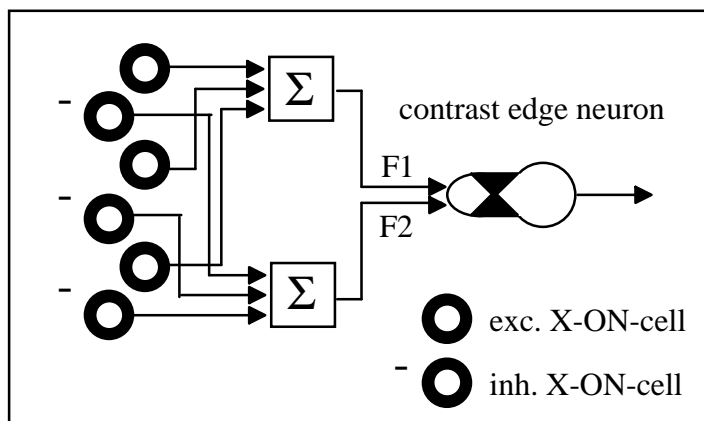




**Abb. 4.1.7:** Orientierungscharakteristiken der ON-OFF-Kantendetektoren (Simple-Zellen). Dargestellt ist jeweils die bezüglich des Ortes maximale Ausgangsimpulsrate  $a$  des 0°-, 30°-, 60°- und des 90°-Kantendetektors in Abhängigkeit vom Orientierungswinkel  $\varphi$  der stimulierenden idealen Kante mit Grauwertdifferenz  $\Delta G=64$  für die Schwellenoffsetwerte  $\Theta_0 = 0$  (Dreiecke), 2 (Kreise), 4 (Quadrate) und 6 (Rauten). (Nähere Erläuterungen siehe Text)

### 4.1.2 ON-ON-Kantendetektor

Das rezeptive Feld des ON-ON-Kantendetektors - ebenso hätte auch ein OFF-OFF-Kantendetektor betrachtet werden können - besteht aus einem exzitatorisch und einem inhibitorisch wirkenden Teilbereich, die aus der kollinearen Verschaltung von jeweils mindestens drei benachbarten X-ON-Zellen auf das Detektorneuron resultieren (Abb. 4.1.8).



**Abb. 4.1.8:** Exzitatorische und inhibitorische Verschaltung von X-ON-Zellen (X-ON-cells) auf ein ON-ON-Kantendetektorneuron (contrast edge neuron). Die Ausgangssignale der drei inhibitorisch und der drei exzitatorisch wirkenden X-Zellen werden getrennt aufsummiert und als Eingangssignale (über den Feeding1 bzw. den Feeding2-Eingang) an das Detektorneuron weitergegeben.

Der ON-ON-Kantendetektor unterscheidet sich vom ON-OFF-Detektor auf zwei Arten. Zum einen werden die exzitatorisch verschalteten X-OFF-Zellen durch inhibitorisch wirkende X-ON-Zellen ersetzt, zum anderen werden die Antworten der X-Zellen aus verschiedenen Bereichen getrennt aufsummiert und über unterschiedliche Feeding-Eingänge dem Detektorneuron zugeführt. Dies ist notwendig, um eine unterschiedliche zeitliche Behandlung von inhibitorischen und exzitatorischen Eingangssignalen zu ermöglichen. Der inhibitorische Bereich dient dazu, Antworten auf Kanten, die nicht in Vorzugsorientierung des rezeptiven Feldes angeordnet sind, zu unterdrücken. Die Verbindungsstärken zwischen exzitatorischen X-ON-Zellen und dem Kantendetektorneuron entsprechen denen der Verschaltung der ON-OFF-Detektoren, jedoch wird eine andere Position des Kantendetektors gewählt. Während bei ON-OFF-Kantendetektoren aus der Information von X-ON- und X-OFF-Zellen eindeutig auf die Position der Kante geschlossen werden kann - in dieser Position wird das ON-OFF-Detektorneuron angeordnet -, ergeben sich aus der Information von Zellen einer Polarität allein zwei Möglichkeiten der Kantenposition. Da somit a priori nicht klar ist, in welcher Position sich die Kante befindet, wird für das Kantendetektorneuron die mittlere Position gewählt, nämlich ein Ort im exzitatorischen Teilbereich des rezeptiven Feldes. Mit dieser Anordnung ist jedoch sofort ein weiteres Problem verbunden, denn bei Kontrastumkehr einer idealen Kante treten die Antworten nun räumlich versetzt auf.

Die Stärke  $w_{ijkl}^{exz}$  der Verbindungen zwischen einer exzitatorischen X-ON-Zelle am Ort  $(k, l)$  und einem Kantendetektorneuron der Orientierung  $\varphi$  am Ort  $(i, j)$  des hexagonalen Rasters läßt sich beschreiben durch:

$${}^{exz}W_{ijkl}^{0^\circ} = \begin{cases} 1, & \text{falls } (k = i - 2 \vee k = i \vee k = i + 2) \wedge (l = j) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.9})$$

$${}^{exz}W_{ijkl}^{30^\circ} = \begin{cases} 1, & \text{falls } (k = i - 1 \wedge l = j - 2) \vee ((k = i \vee k = i + 2) \wedge l = j) \\ 0.5, & \text{falls } (k = i - 3 \wedge l = j - 2) \vee (k = i + 3 \wedge l = j + 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.10})$$

$${}^{exz}W_{ijkl}^{60^\circ} = \begin{cases} 1, & \text{falls } (k = i - 1 \wedge l = j - 2) \vee (k = i \wedge l = j) \vee (k = i + 1 \wedge l = j + 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.11})$$

$${}^{exz}W_{ijkl}^{90^\circ} = \begin{cases} 1, & \text{falls } (k = i \wedge l = j) \vee (k = i + 1 \wedge (l = j - 2 \vee l = j + 2)) \\ 0.5, & \text{falls } (k = i \wedge (l = j - 4 \vee l = j + 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.12})$$

Weiterhin ergibt sich die Stärke  ${}^{inh}W_{ijkl}^\varphi$  der Verbindungen zwischen einer inhibitorischen X-ON-Zelle am Ort  $(k, l)$  und einem Kantendetektorneuron der Orientierung  $\varphi$  am Ort  $(i, j)$  des hexagonalen Rasters zu:

$${}^{inh}W_{ijkl}^{0^\circ} = \begin{cases} -1, & \text{falls } ((k = i - 3 \vee k = i - 1 \vee k = i + 1 \vee k = i + 3) \wedge (l = j - 2)) \\ & \vee ((k = i - 2 \vee k = i \vee k = i + 2) \wedge (l = j - 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.13})$$

$${}^{inh}W_{ijkl}^{30^\circ} = \begin{cases} -1, & \text{falls } ((k = i \vee k = i + 2) \wedge l = j - 4) \vee ((k = i + 1 \vee k = i + 3) \\ & \wedge l = j - 2) \\ -0.5, & \text{falls } (k = i - 2 \wedge l = j - 4) \vee (k = i - 1 \wedge l = j - 6) \\ & \vee (k = i + 4 \wedge l = j) \vee (k = i + 5 \wedge l = j - 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.14})$$

$${}^{inh}W_{ijkl}^{60^\circ} = \begin{cases} -1, & \text{falls } ((k = i + 2 \vee k = i) \wedge l = j - 4) \vee ((k = i + 3 \vee k = i + 1) \wedge l = \\ & j - 2) \vee ((k = i + 4 \vee k = i + 2) \wedge l = j) \vee ((k = i + 3 \wedge l = j + 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.15})$$

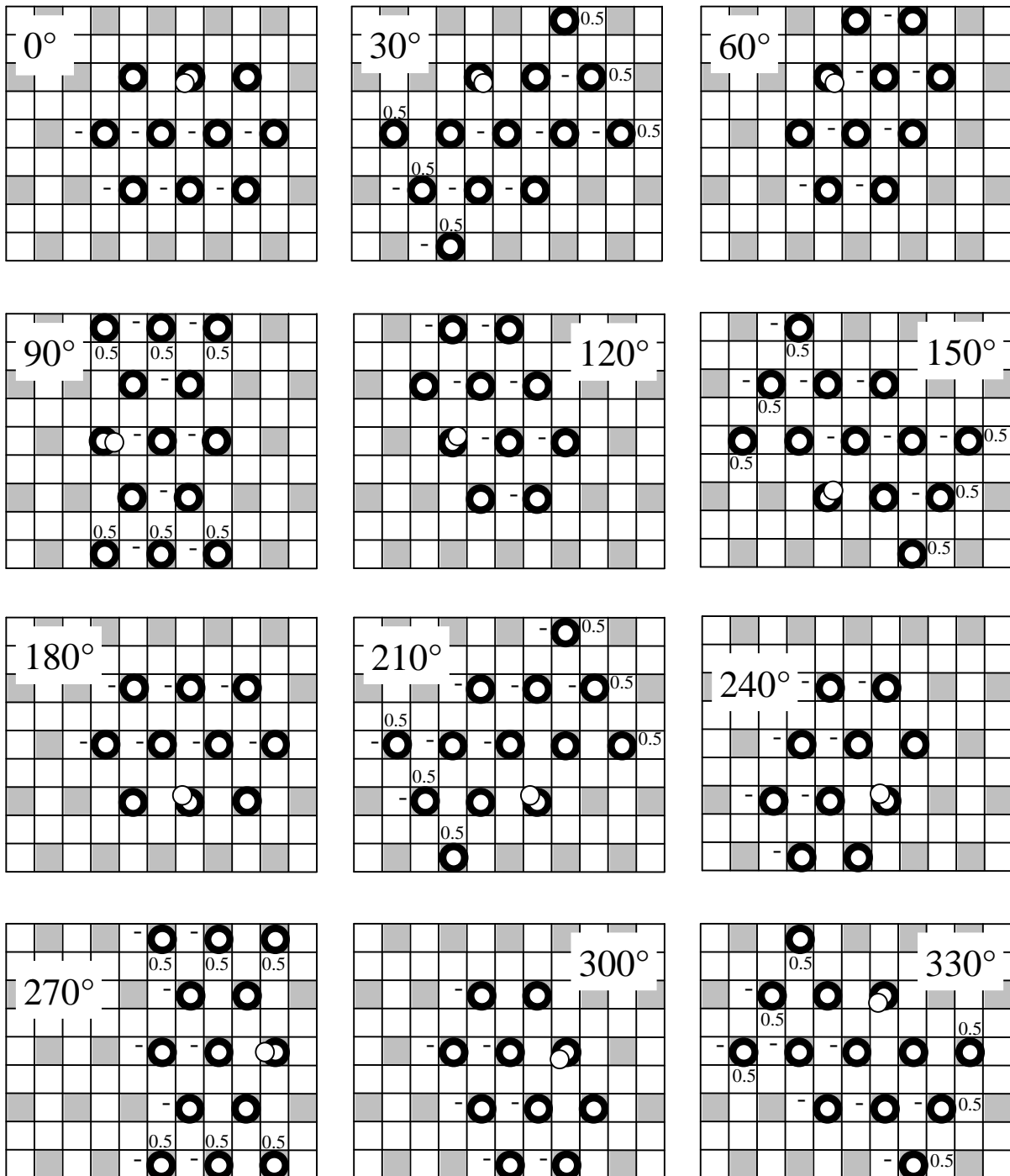
$${}^{inh}W_{ijkl}^{90^\circ} = \begin{cases} -1, & \text{falls } ((k = i + 4 \vee k = i + 2) \wedge l = j) \vee (k = i + 3 \wedge (l = j - 2 \\ & \vee l = j + 2)) \\ -0.5, & \text{falls } ((k = i + 4 \vee k = i + 2) \wedge (l = j - 4 \vee l = j + 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.16})$$

Neben der schematischen Darstellung (Abbildung 4.1.9) ist auch hier die Darstellung des rezeptiven Feldes als Überlagerung von rezeptiven Feldern der X-Zellen möglich. Abbildung 4.1.10 zeigt die entsprechenden Darstellungen der Superpositionen. Wie weiter oben in diesem Kapitel gezeigt wurde, kann eine Gleichbehandlung der Abtastpunkte bei gegebenen Bedingungen mit einem zusätzlichen Gewichtungsfaktor von etwa 0,5 für den inhibitorischen Teilbereich angenähert werden. Ein solcher Gewichtungsfaktor wird hier mit Hilfe der im Verhältnis von etwa 1:2 zueinander gewählten Feeding-Potentialverstärkungen  $V^{F2}$  und  $V^{F1}$  realisiert. Auch in der Darstellung der rezeptiven Felder wird dies berücksichtigt. Zur Bestimmung der Orientierungscharakteristiken wurde der Verstärkungsfaktor  $V^{F1}$  des ON-ON-Detektors zudem mit doppelter Größe des Verstärkungsfaktors  $V^F$  (bei ON-OFF-Detektor) gewählt, um die geringere Anzahl an Verbindungen von exzitatorisch wirkenden X-Zellen auszugleichen. Bei kleinem Schwellenoffset fallen die Raten in zur Vorzugsorientierung senkrechten Richtung deutlich geringer aus als bei ON-OFF-Detektoren (vergleiche die Abbildungen 4.1.11 und 4.1.7, insbesondere bei kleinerem Schwellenoffset). Jedoch tritt wegen der fehlenden Polarität neben der Antwort in Vorzugsrichtung auch eine ähnlich starke Antwort in der um  $180^\circ$  zur Vorzugsrichtung gedrehten Richtung (Rückwärtsrichtung) auf.

Schwellen- verstärkung $V^\Theta$	Schwellen- zeitkonstante $\tau^\Theta$	Schwellen- offset $\Theta_0$	Feeding- Potential- verstärkung $V^{F1}$	Feeding- Potential- zeitkonstante $\tau^{F1}$	Feeding- Potential- verstärkung $V^{F2}$	Feeding- Potential- zeitkonstante $\tau^{F2}$
32	10 BIN	0, 2, 4, 6	1.4	20 BIN	0.6 · 1.4	40 BIN

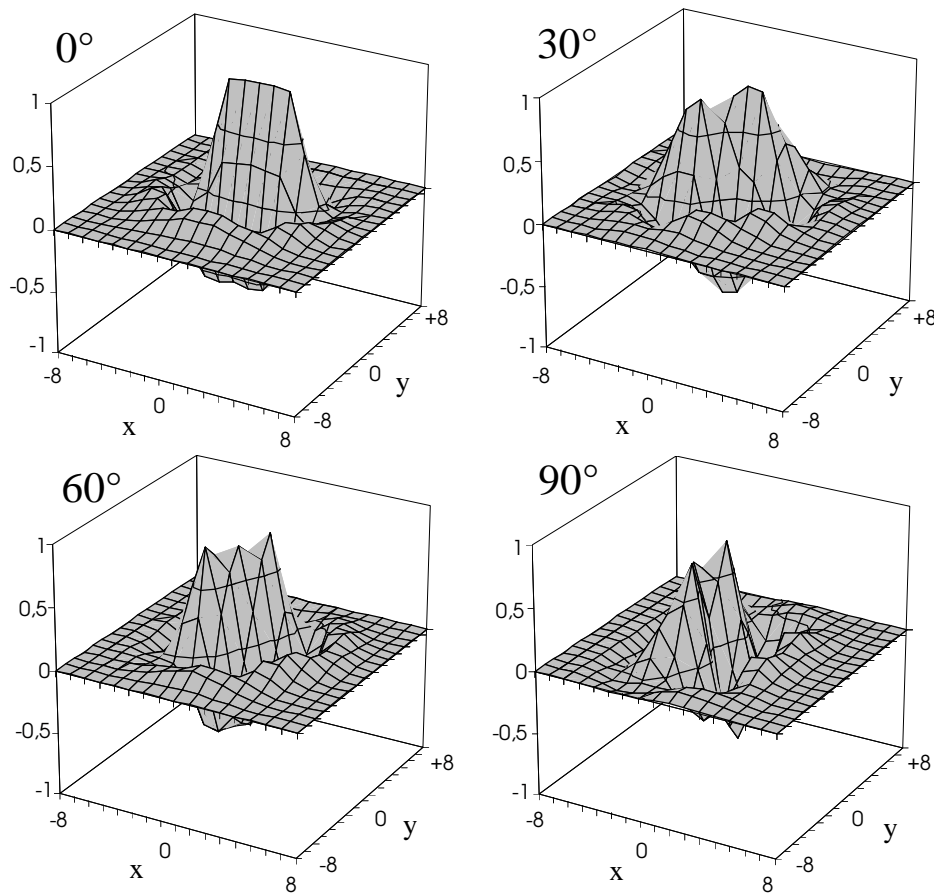
**Tab. 4.1.4:** Parameter des ON-ON-Kantendetektorneurons.

Der inhibitorische Teilbereich des rezeptiven Feldes sollte möglichst keine Lücken in der Anordnung aufweisen, um insbesondere Antworten auf (nahezu) senkrecht verlaufende Kanten sicher unterdrücken zu können. Dies wird im Bedarfsfall, analog der in Kapitel 4.1.1 beschriebenen Vorgehensweise, durch die Berücksichtigung weiterer inhibitorisch wirkender X-Zellen ermöglicht (vgl. Abb. 4.1.9). Neben der räumlichen muß auch eine zeitliche Kontinuität der inhibitorischen Wirkung gewährleistet sein, was durch eine geeignet große Zeitkonstante  $\tau^{F2}$  des hier inhibitorisch wirkenden Feeding2-Potentials realisiert wird. Dadurch soll insbesondere bei zeitlich und räumlich sich wenig ändernden Eingangsbildern verhindert werden, daß durch ein zeitliches Schwanken des inhibitorischen Potentials ein Kantendetektorneuron beispielsweise bei Reizung mit einer senkrecht zur Vorzugsorientierung ausgerichteten Kante kurzzeitig überschwellig werden kann. Die Zeitkonstante  $\tau^{F1}$  des Feeding1-Leckintegrators ergibt sich aus dem Kompromiß zweier gegensätzlicher Anforderungen. Einerseits muß durch eine große Zeitkonstante vermieden werden, daß ein Kantendetektorneuron nur bei nahezu zeitgleichem Eintreffen von Antworten verschiedener X-Zellen (Koinzidenzen) überschwellig werden kann.

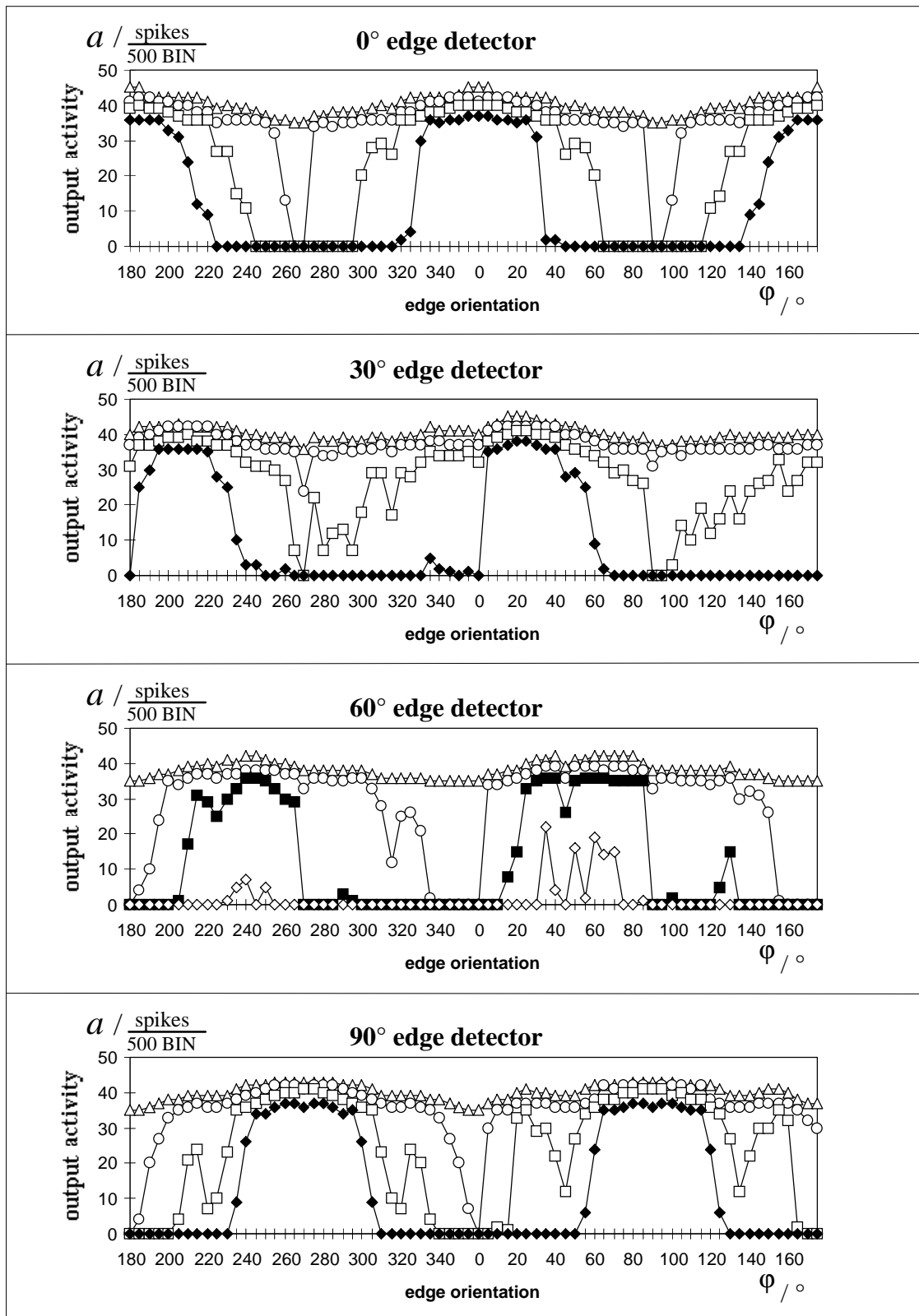


**Abb. 4.1.9:** Schematische Darstellung der rezeptiven Felder von ON-ON-Kantendetektoren für zwölf verschiedene Orientierungen im hexagonalen Abtastraster bei Abtastweite  $d=2$ . Die Position des Detektorneurons im Abtastraster ist durch einen weißen Kreis dargestellt. Die konzentrischen Kreise geben die Zentrumspositionen der rezeptiven Felder derjenigen X-ON-Zellen an, die ihre Ausgangssignale an das Detektorneuron senden. X-Zellen mit inhibitorischer Wirkung sind durch ein Minuszeichen und X-Zellen mit halber Gewichtung mit "0.5" gekennzeichnet.

Diese Gefahr ist insbesondere bei geringen Ausgangsraten der X-Zellen gegeben. Auf der anderen Seite ist eine kurze Zeitkonstante wünschenswert, um insbesondere bei Eingangsbildern, in denen starke zeitliche Veränderungen auftreten, die nötige Flexibilität erreichen zu können. In Tabelle 4.1.4 sind die entsprechenden Parameter des ON-ON-Detektorneurons zusammengestellt.



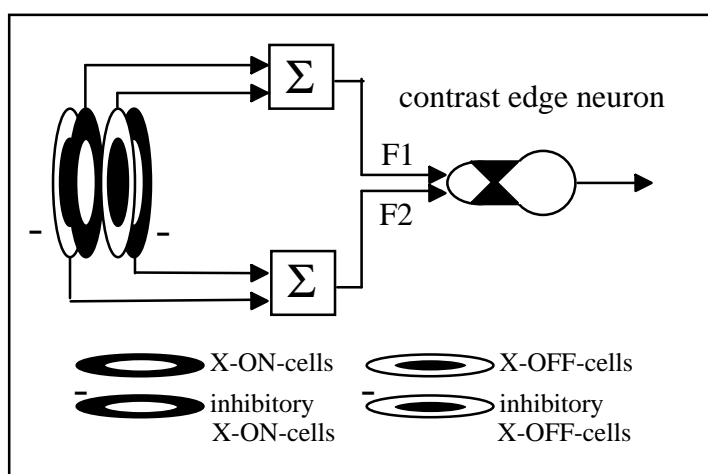
**Abb. 4.1.10: Rezeptive Felder von ON-ON-Kantendetektoren.** Die rezeptiven Felder ergeben sich aus Überlagerung der rezeptiven Felder von X-Zellen und sind dargestellt für die Orientierungen  $0^\circ$ ,  $30^\circ$ ,  $60^\circ$  und  $90^\circ$ . Die Masken der exzitatorisch wirkenden X-ON-Zellen werden dabei einfach positiv, die der inhibitorisch wirkenden X-ON-Zellen mit  $-0,6$  für die Superposition gewichtet. Im Unterschied zu den in Abb. 4.1.6 gezeigten rezeptiven Feldern der ON-OFF-Kantendetektoren ist der inhibitorische Bereich (entspricht in Abb. 4.1.6 dem OFF-Bereich) aufgrund der zusätzlichen Gewichtung weniger stark ausgeprägt, was auch geringe Auswirkungen auf den exzitatorischen Bereich hat.



**Abb. 4.1.11:** Orientierungscharakteristiken der ON-ON-Kantendetektoren. Aufgetragen sind die maximalen Ausgangsimpulsraten  $a$  von 0°-, 30°-, 60°- und des 90°- Kantendetektor in Abhängigkeit vom Orientierungswinkel  $\varphi$  der stimulierenden idealen Kante mit Grauwertdifferenz  $\Delta G=64$  für die Schwellenoffsetwerte  $\Theta_0 = 0$  (**Dreiecke**), 2 (**Kreise**), 4 (**Quadrate**) und 6 (**Rauten**). Die Charakteristiken sind gegenüber denen der ON-OFF-Detektoren verschärft. Jedoch treten hier auch Antworten bei Umkehrung der Kantenpolarität auf.

### 4.1.3 ON<sup>2</sup>-OFF<sup>2</sup>-Kantendetektor

Die beiden bisher untersuchten Detektortypen stellen noch keine befriedigenden Lösungen zur Kantendetektion dar. Denn während der ON-ON-Detektor bedingt durch die fehlende Polarität Kantenpositionen nicht eindeutig wiedergeben kann und immer auch bei Umkehrung der Kantenpolarität antwortet, sind beim ON-OFF-Kantendetektor räumlich versetzte Antworten auf umgekehrt ausgerichtete Kanten und Antworten auf senkrecht zur Vorzugsorientierung ausgerichtete Kanten insbesondere bei hohem Intensitätskontrast möglich. Einen Ausweg aus diesen Problemen liefert der ON<sup>2</sup>-OFF<sup>2</sup>-Kantendetektor, dessen rezeptives Feld sich aus den rezeptiven Feldern der obigen Kantendetektortypen zusammensetzen läßt.



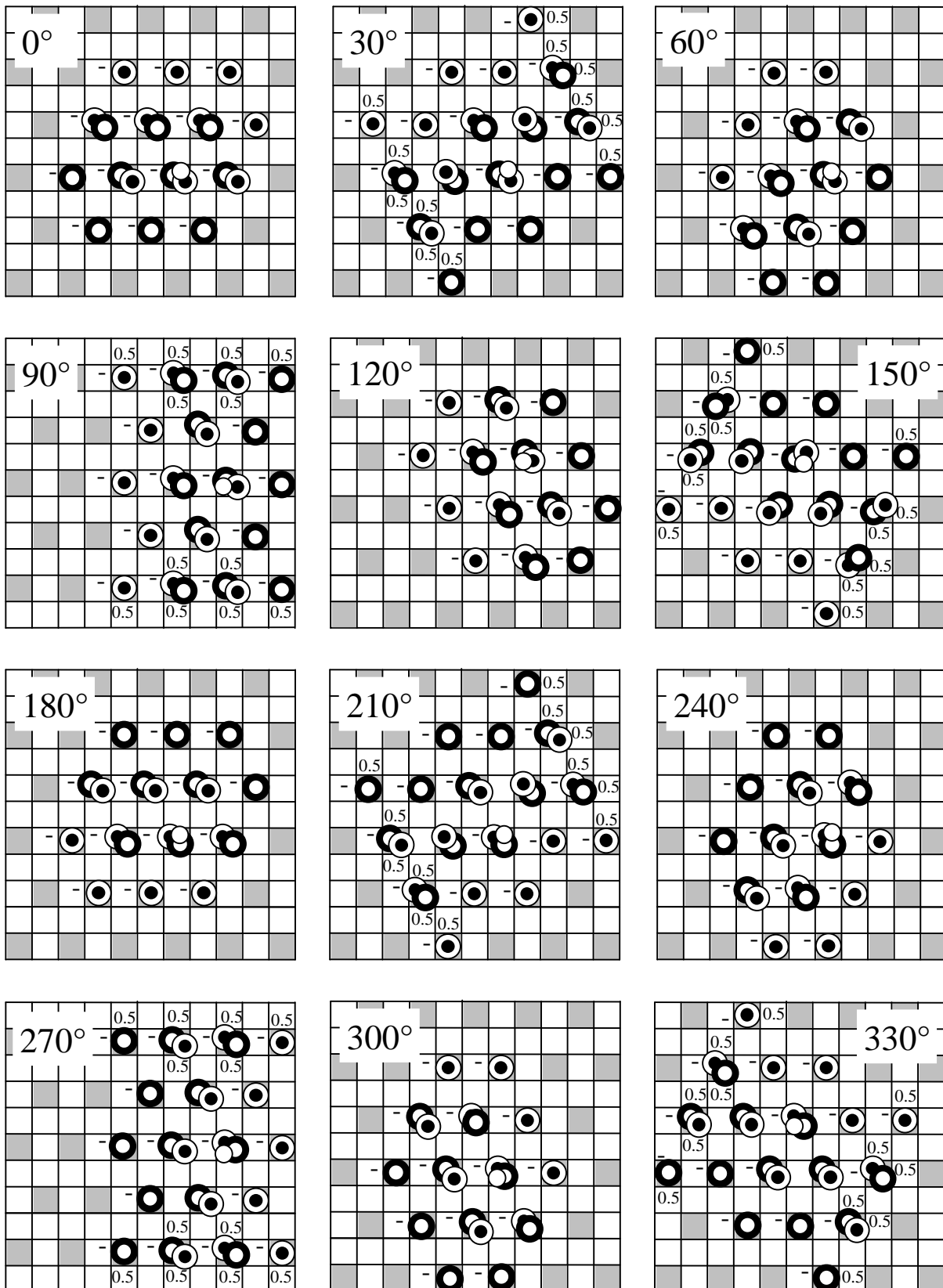
**Abb. 4.1.12:** Exzitatorische und inhibitorische Verschaltung von X-ON-Zellen und X-OFF-Zellen auf ein ON<sup>2</sup>-OFF<sup>2</sup>-Kantendetektorneuron (contrast edge neuron). Die Ausgangs-Signale der (mindestens) sechs inhibitorisch und der sechs exzitatorisch wirkenden X-Zellen werden getrennt aufsummiert und als Eingangssignale (über den Feeding1- bzw. den Feeding2-Eingang) an das Detektorneuron weitergegeben.

Es besteht aus zwei exzitatorisch und zwei inhibitorisch wirkenden Teilbereichen, die aus der kollinearen Verschaltung von jeweils mindestens drei benachbarten X-ON-Zellen und X-OFF-Zellen auf das Detektorneuron resultieren (Abbildung 4.1.12). Die Wirkungsweise des ON<sup>2</sup>-OFF<sup>2</sup>-Kantendetektors entspricht weitgehend der des ON-OFF-Detektors, wobei hier der Abstand zwischen den beiden exzitatorischen Teilbereichen mit  $\Delta y = 2$  gewählt wurde, um die Größe des gesamten rezeptiven Feldes zu begrenzen.

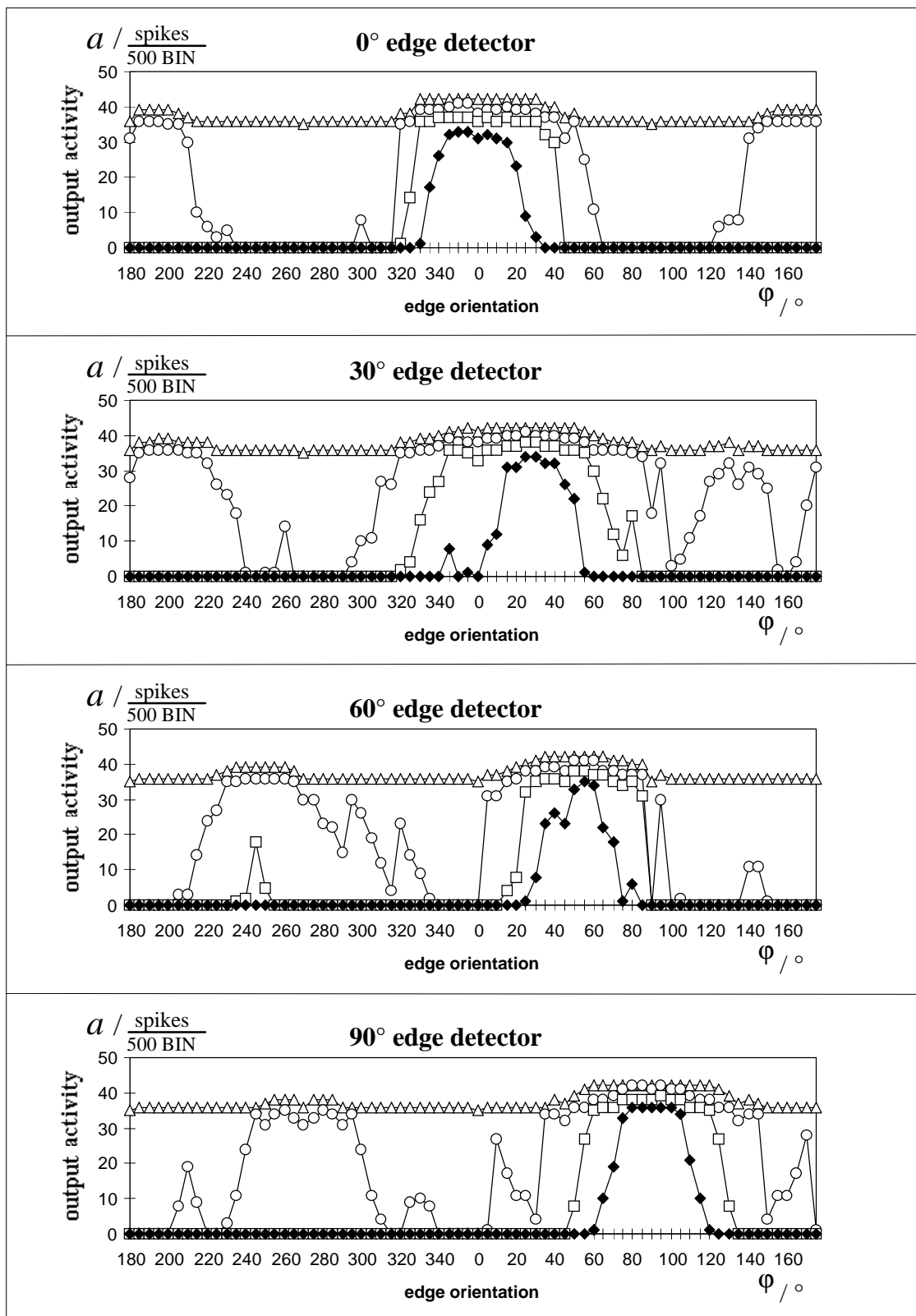
Vorzugs- richtung $\varphi$	Schwellen- verstärkung $V^\ominus$	Schwellen- zeitkonstante $\tau^\ominus$	Schwellen- -offset $\Theta_0$	Feeding- Potential- verstärkung $V^{F1}$	Feeding- Potential- zeitkonstante $\tau^{F1}$	Feeding- Potential- verstärkung $V^{F2}$	Feeding- Potential- zeitkonstante $\tau^{F2}$
0°	32	10 BIN	6	0,6	20 BIN	$0,6 \cdot V^{F1}$	40 BIN
30°	32	10 BIN	6	0,7	20 BIN	$0,6 \cdot V^{F1}$	40 BIN
60°	32	10 BIN	6	1,0	20 BIN	$0,6 \cdot V^{F1}$	40 BIN
90°	32	10 BIN	6	0,65	20 BIN	$0,6 \cdot V^{F1}$	40 BIN

**Tab. 4.1.5:** Parameter der ON<sup>2</sup>-OFF<sup>2</sup>-Kantendetektoren.





**Abb. 4.1.13:** Schematische Darstellung der rezeptiven Felder von ON<sup>2</sup>-OFF<sup>2</sup>-Kantendetektoren für zwölf verschiedene Orientierungen im hexagonalen Abtastraster bei Abtastweite  $d=2$ . Die Position des Detektorneurons im Abtastraster ist durch einen weißen Kreis dargestellt. Die konzentrischen Kreise geben die Zentrumspositionen der rezeptiven Felder derjenigen X-Zellen an (außen schwarz, innen weiß: X-ON-Zellen; außen weiß, innen schwarz: X-OFF-Zellen), die ihre Ausgangssignale an das Detektorneuron senden.



**Abb. 4.1.14:** Orientierungscharakteristiken der  $ON^2-OFF^2$ -Kantendetektoren. Aufgetragen sind erneut die maximalen Ausgangsimpulsraten  $a$  von 0°-, 30°-, 60°- und 90°- Kantendetektor in Abhängigkeit vom Orientierungswinkel  $\varphi$  der stimulierenden idealen Kante mit Grauwertdifferenz  $\Delta G=64$  für die Schwellenoffsetwerte  $\Theta_0 = 0$  (Dreiecke), 2 (Kreise), 4 (Quadrate) und 6 (Rauten). Die Detektoren antworten, ähnlich den ON-OFF-Kantendetektoren, nur bei einer Polarität der Kante, jedoch besitzen sie schärfere Charakteristiken als ON-OFF- oder ON-ON-Detektoren.

Zusätzlich verhindern die inhibitorischen Teilbereiche Antworten auf um  $180^\circ$  zur Vorzugsorientierung gedrehte Kanten (Rückwärtsantworten) von räumlich versetzten Detektoren und schwächen deutlich die Antworten der Detektoren in zur Vorzugsorientierung senkrechte Richtungen. Die Stärke  ${}^W P w_{ijkl}^\varphi$  der Verbindungen zwischen einer X-Zelle der Polarität  $P$  und der Wirkungsweise  $W$  (exzitatorisch oder inhibitorisch) am Ort  $(k, l)$  und einem Kantendetektor der Orientierung  $\varphi$  am Ort  $(i, j)$  des hexagonalen Rasters wird beschrieben durch:

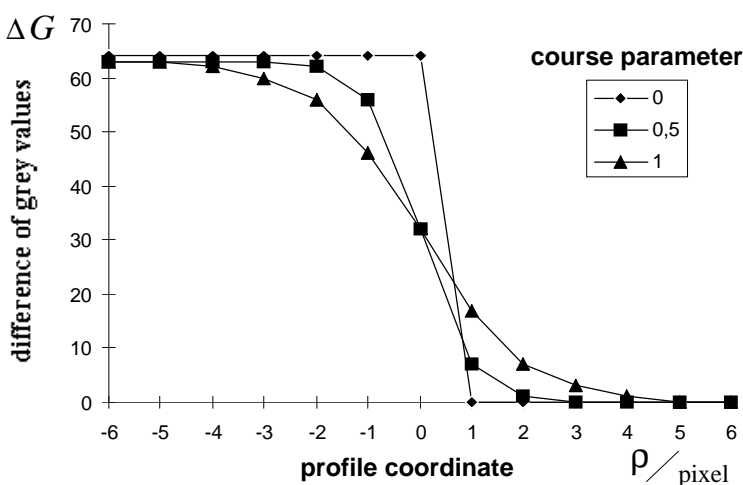
$${}_{ON}^W w_{ijkl}^\varphi = \begin{cases} {}^W w_{i-1, j+2, kl}^\varphi, & \text{falls } \varphi \in \{0^\circ, 30^\circ\} \\ {}^W w_{i-2, j, kl}^\varphi, & \text{falls } \varphi \in \{60^\circ, 90^\circ\} \end{cases} \quad (\text{Gl. 4.1.17})$$

und

$${}_{OFF}^W w_{ijkl}^\varphi = {}^W w_{-i-j-k-l}^\varphi \quad \forall \varphi \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ\}, \quad (\text{Gl. 4.1.18})$$

wobei  ${}^W w_{ijkl}^\varphi$  aus den Gleichungen 4.1.9 bis 4.1.16 hervorgeht.

Schematische Darstellungen der rezeptiven Felder zeigt die Abbildung 4.1.13 und die resultierenden Orientierungscharakteristiken sind in Abbildung 4.1.14 aufgetragen. Weiterhin sind die zur Berechnung der Orientierungscharakteristiken verwendeten Parameter in Tabelle 4.1.5 zusammengefaßt. Wie oben beschrieben, sind die rezeptiven Felder der verschiedenen Kantendetektoren unterschiedlich breit. Der  $60^\circ$ -Kantendetektor besitzt von allen Detektoren das schmalste rezeptive Feld, weil inhibitorische und exzitatorische Teilbereiche sehr dicht zusammenliegen, wodurch auch die Amplituden des rezeptiven Feldes verringert werden. Zur Kompensation dieser Verringerung werden die Verstärkungsfaktoren der Feedingpotentiale entsprechend erhöht.



**Abb. 4.1.15:** Sigmoidfunktionen als Definition des Kantenquerschnittes. Der Grauwert ist als Funktion der Querschnittskoordinate  $\rho$  für unterschiedliche Werte des Verlaufspareters  $P$  aufgetragen. Für sehr kleine Werte geht die Sigmoidfunktion in eine Sprungfunktion (ideale Kante) über, die gesondert berechnet wird.

Schließlich werden die so konzipierten Detektoren auf ihre Eigenschaften bezüglich der Detektion unterschiedlich breiter Kanten hin untersucht. Dabei interessiert insbesondere die

Fragestellung, wie breit eine Kante sein darf, um noch detektierbar zu sein. Als Stimulus wird eine zeitlich und räumlich schwach verrauschte Kante verwendet, deren Profil sich durch eine Sigmoidfunktion beschreiben lässt. Die Sigmoidfunktion  $S$  ist gegeben durch:

$$S(\rho, P) = \frac{1}{1 + \exp(-\rho/P)}, \quad (\text{Gl. 4.1.19})$$

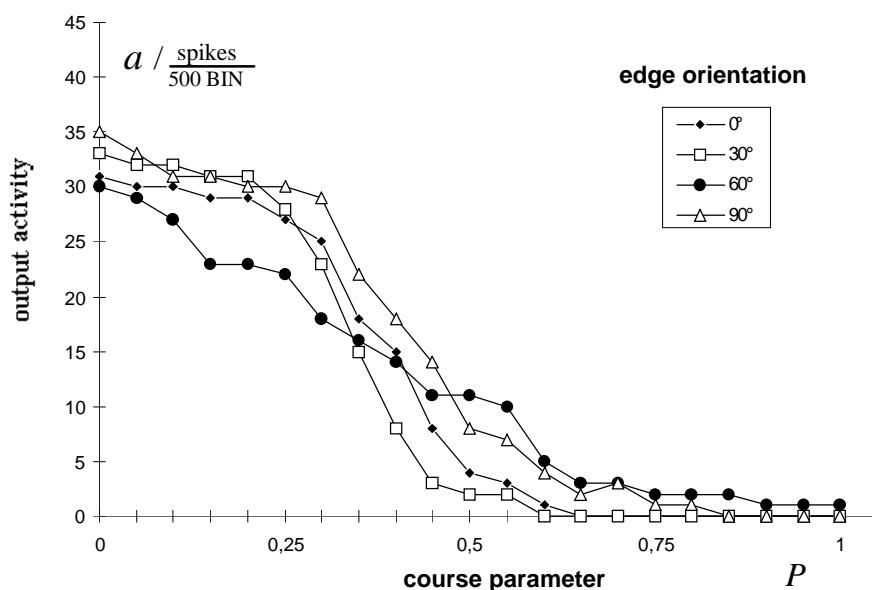
wobei  $P$ : Verlaufsparemeter, bestimmt die Ausdehnung des Kantenanstiegs (Breite).

Wie in Kapitel 3 beschrieben, liefert die Kamera in jedem 40. Zeitschritt ein neues Bild. Alle innerhalb dieses Zeitintervalls erzeugten Bilder ergeben sich aus linearer Interpolation. Entsprechend wird auch hier die Kante  $K$  in jedem 40. Zeitschritt neu berechnet zu:

$$K(x, y, \varphi, P) = S(x \cdot \cos \varphi + y \cdot \sin \varphi, P) \cdot \Delta G + G_0 + B \cdot R(x, y), \quad (\text{Gl. 4.1.20})$$

mit:  $\Delta G$  : Grauwertdifferenz (hier:  $\Delta G = 64$ ),  
 $G_0$  : Grauwertoffset (hier:  $G_0 = 20$ ),  
 $B$  : maximale Rauschamplitude (hier:  $B = 16$ ),  
 $R(x, y)$  : Ergebnis eines im Intervall  $[-1, 1]$  gleichverteilten Zufallsprozesses, das für jeden Punkt  $(x, y)$  gesondert berechnet wird,  
 $\varphi$  : Winkel der Vorzugsorientierung.

Geht der Verlaufsparemeter gegen 0, so liegt eine Sprungfunktion (ideale Kante) vor. Der Anstiegsverlauf der Kante ist für die Werte des Verlaufsparemters  $P = 0, 0.5$  und  $1$  in Abbildung 4.1.15 aufgetragen.



**Abb. 4.1.16:** Aktivität der Kantendetektoren bei Stimulation mit unterschiedlich breiten Kanten. Aufgetragen ist die Summe der innerhalb eines Zeitintervalls von 500 BIN generierten Aktionspotentiale des 0°-, 30°-, 60° und des 90°-Kantendetektors in Abhängigkeit vom Verlaufsparemeter  $P$  (Breite) von stimulierenden Kanten der jeweiligen Vorzugsorientierung

Abbildung 4.1.16 zeigt die Summe der innerhalb eines Zeitintervalls von 500 BIN generierten Aktionspotentiale des 0°-, 30°-, 60° und des 90°- Kantendetektors als Funktion des Verlaufsparmeters (Breite) von stimulierenden Kanten der jeweiligen Vorzugsorientierung. Die Detektoren zeigen nur bei Kanten mit Werten des Verlaufsparmeters unterhalb von etwa  $P=0.4$  signifikante Aktivität. Hier wird deutlich, daß diese Detektoren nur einen relativ kleinen Spielraum von etwa zwei Pixel in der Variation der Kantenbreiten erlauben. Für Kanten mit einem weniger steilen Anstieg müssen Detektoren einer größeren Auflösung konzipiert werden. Unter Vernachlässigung dieses breiteren Kantenanstiegs genügen die gezeigten Detektoren zur weiteren Verarbeitung.

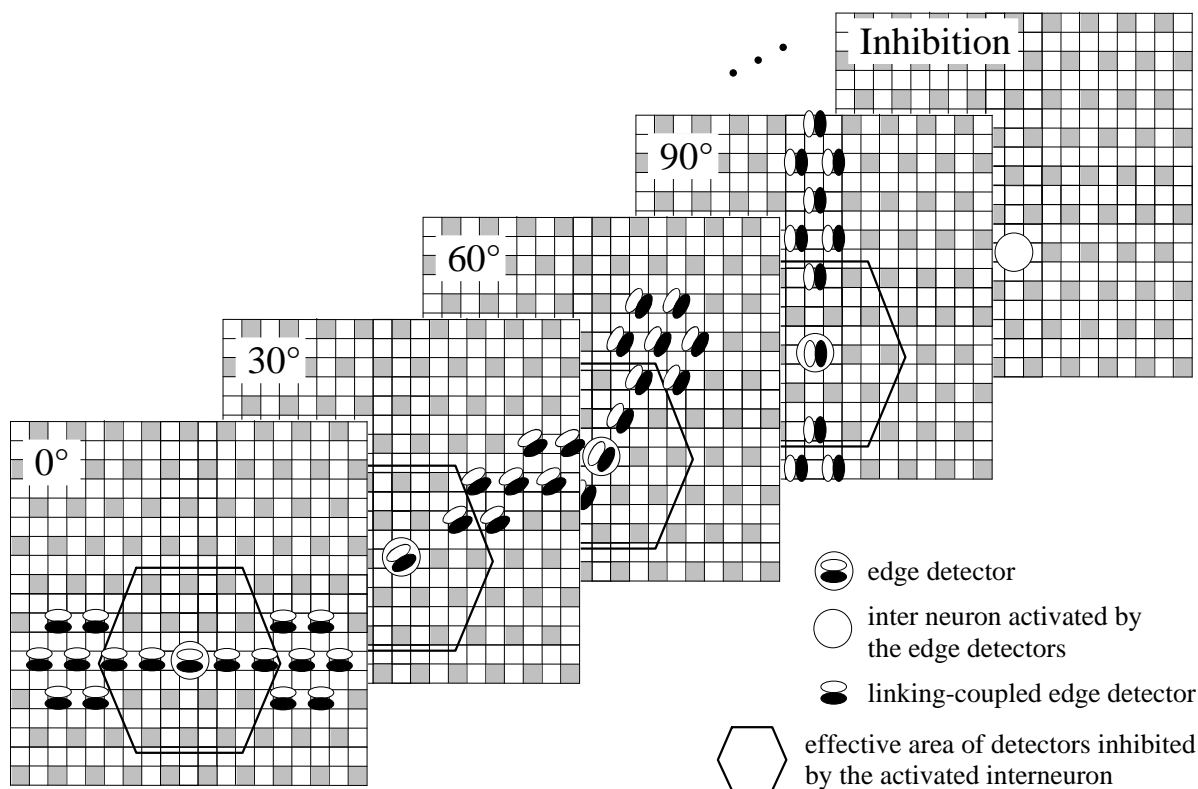
#### 4.1.4 Eine Netzwerkstruktur zur Unterstützung von Kontur-Kontinuität und zur Erzeugung eindeutiger Antworten von Kantendetektoren

Wie aus den im vorherigen Unterabschnitt gezeigten Orientierungscharakteristiken direkt hervorgeht, antworten in der Regel mehrere Kantendetektoren auf dieselbe Kante einer beliebigen Orientierung. Für die nachfolgende Detektion von Linienenden wird aber, wie weiter unten gezeigt wird, ein eindeutiges Antwortverhalten benötigt. Dies wird durch die Einführung eines gemeinsamen Interneurons für solche Kantendetektoren erreicht, die am gleichen Ort (und einer bestimmten Umgebung) lokalisiert sind, aber verschiedene Vorzugsorientierung besitzen. Dieses Interneuron wirkt inhibitorisch und unterdrückt die Antworten aller Kantendetektorneuronen desselben Ortes mit Ausnahme des Detektors mit der stärksten Antwort. Das entsprechende Gewicht  $^{inhib}w_{ghij}^\varphi$  der Verbindung von einem Kantendetektorneuron der Vorzugsorientierung  $\varphi$  am Ort  $(i, j)$  zum Feeding-Eingang eines Interneurons am Ort  $(g, h)$  ist gegeben durch:

$$^{inhib}w_{ghij}^\varphi = \begin{cases} 1, & \text{falls } ((i = g - 1 \vee i = g + 1) \wedge (j = h + 2 \vee j = h - 2)) \\ & \vee ((i = g - 2 \vee i = g \vee i = g + 2) \wedge j = h) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.1.21})$$

Auch die Verbindungen  $^{inhib}w_{ijgh}^\varphi$  der umgekehrten Richtung von einem Interneuron am Ort  $(g, h)$  zum inhibitorischen Eingang eines Kantendetektorneurons der Vorzugsorientierung  $\varphi$  am Ort  $(i, j)$  werden durch Gleichung 4.1.21 beschrieben, wenn zuvor die Indizes  $(g$  gegen  $i$  und  $h$  gegen  $j)$  vertauscht werden. Die inhibitorisch wirkende Antwort des Interneurons führt zu einer Unterdrückung der Antworten aller Kantendetektoren dieser Kolumne, so daß nur die Antwort des Detektors mit der stärksten Aktivität übrig bleibt (Winner-takes-all-Mechanismus).

Durch die wechselseitige Verschaltung der Kantendetektorneuronen mit den Interneuronen führt die Aktivität eines Kantendetektorneurons auch zu einer starken Unterdrückung der sechs in direkter räumlicher Nachbarschaft befindlichen Kantendetektoren.



**Abb. 4.1.17:** Linking- und Inhibitions-Architektur zur Verbesserung der Kantendetektion. Dargestellt sind Ausschnitte von Orientierungsschichten der Kantendetektoren. Die Kantendetektoren (edge detectors) und das Interneuron desselben Bezugsortes sind jeweils durch einen Kreis gekennzeichnet. Jeder Kantendetektor ist mit seinen in einer keulenförmigen Umgebung gelegenen benachbarten Kantendetektoren derselben Vorzugsorientierung über Linking-Verbindungen gekoppelt. Zudem aktiviert er das allen Detektoren desselben Ortes (und einer bestimmten Umgebung) gemeinsame Interneuron, das wiederum die Antworten sämtlicher in einer kreisförmigen Umgebung gelegenen Kantendetektoren inhibiert.

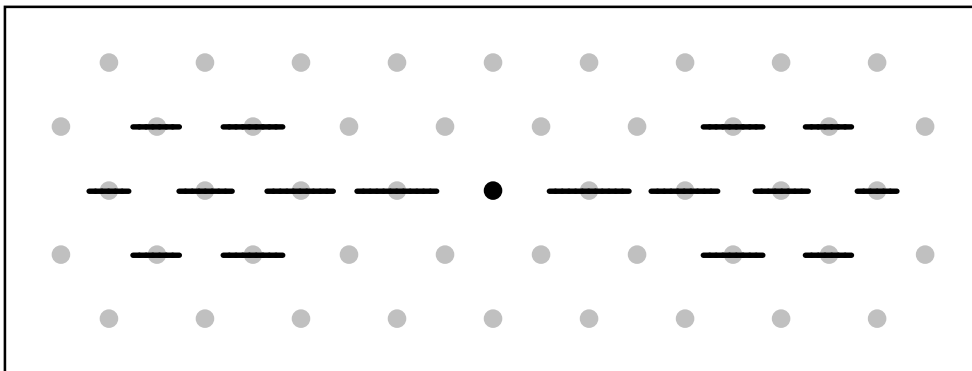
Darüberhinaus werden die Antworten der zwölf übernächsten Kantendetektoren der räumlichen Nachbarschaft jeder Orientierungsschicht unterdrückt, jedoch weniger stark als die direkten Nachbarneuronen. Entsprechend wird die Effektivität der Inhibition mit größerem Abstand des inhibierten Kantendetektors vom inhibierenden Interneuron geringer. Da aber a priori nicht klar ist, daß das Neuron mit der stärksten Antwort auch derjenigen Kantenorientierung entspricht, die sich am besten in die gegebene Kontur einfügt, wird hier zusätzlich ein zweiter Mechanismus eingeführt, der genau das gewährleisten soll. Innerhalb einer jeden Orientierungsschicht werden die Kantendetektorneuronen, die koaxial in Vorzugsrichtung - auch gewisse Abweichungen sind erlaubt - angeordnet sind, miteinander durch Linking-Verbindungen (Eckhorn et al. 1990) gekoppelt. Dadurch ergibt sich eine keulenförmige und mit größerem Abstand schwächer werdende Gewichtungsverteilung. Die Stärke  $^{link}w_{ijkl}^{\varphi}$  der Linking-Verbindung zwischen einem Kantendetektorneuron der Vorzugsorientierung  $\varphi$  am Ort  $(k, l)$  und einem Kantendetektorneuron der Vorzugsorientierung  $\varphi$  am Ort  $(i, j)$  ist gegeben durch:

$$\begin{aligned}
 \text{link } W_{ijkl}^{0^\circ} &= \begin{cases} 3.0, & \text{falls } k = i \pm 2 \wedge l = j \\ 2.5, & \text{falls } k = i \pm 4 \wedge l = j \\ 2.0, & \text{falls } k = i \pm 6 \wedge l = j \\ 1.5, & \text{falls } k = i \pm 8 \wedge l = j \\ 2.25, & \text{falls } k = i \pm 5 \wedge l = j \pm 2 \\ 1.75, & \text{falls } k = i \pm 7 \wedge l = j \pm 2 \\ 0, & \text{sonst} \end{cases} \\
 \text{link } W_{ijkl}^{30^\circ} &= \begin{cases} 3.5, & \text{falls } k = i + 3 \wedge l = j + 2 \vee k = i - 3 \wedge l = j - 2 \\ 2.5, & \text{falls } k = i + 6 \wedge l = j + 4 \vee k = i - 6 \wedge l = j - 4 \\ 1.0, & \text{falls } k = i + 9 \wedge l = j + 6 \vee k = i - 9 \wedge l = j - 6 \\ 3.25, & \text{falls } k = i + 4 \wedge l = j + 4 \vee k = i - 4 \wedge l = j - 4 \\ & \vee k = i + 5 \wedge l = j + 2 \vee k = i - 5 \wedge l = j - 2 \\ 2.0, & \text{falls } k = i + 7 \wedge l = j + 6 \vee k = i - 7 \wedge l = j - 6 \\ & \vee k = i + 8 \wedge l = j + 4 \vee k = i - 8 \wedge l = j - 4 \\ 0, & \text{sonst} \end{cases}
 \end{aligned}$$

(Gl. 4.1.22 a,b,c,d)

$$\begin{aligned}
 \text{link } W_{ijkl}^{60^\circ} &= \begin{cases} 3.0, & \text{falls } k = i + 1 \wedge l = j + 2 \vee k = i - 1 \wedge l = j - 2 \\ 2.5, & \text{falls } k = i + 2 \wedge l = j + 4 \vee k = i - 2 \wedge l = j - 4 \\ 2.0, & \text{falls } k = i + 3 \wedge l = j + 6 \vee k = i - 3 \wedge l = j - 6 \\ 1.5, & \text{falls } k = i + 4 \wedge l = j + 8 \vee k = i - 4 \wedge l = j - 8 \\ 2.25, & \text{falls } k = i + 1 \wedge l = j + 6 \vee k = i - 1 \wedge l = j - 6 \\ & \vee k = i + 4 \wedge l = j + 4 \vee k = i - 4 \wedge l = j - 4 \\ 1.75, & \text{falls } k = i + 2 \wedge l = j + 8 \vee k = i - 2 \wedge l = j - 8 \\ & \vee k = i + 5 \wedge l = j + 6 \vee k = i - 5 \wedge l = j - 6 \\ 0, & \text{sonst} \end{cases} \\
 \text{link } W_{ijkl}^{90^\circ} &= \begin{cases} 4.0, & \text{falls } k = i \wedge l = j \pm 4 \\ 2.5, & \text{falls } k = i \wedge l = j \pm 8 \\ 1.0, & \text{falls } k = i \wedge l = j \pm 12 \\ 3.25, & \text{falls } k = i \pm 1 \wedge l = j \pm 6 \\ 2.0, & \text{falls } k = i \pm 1 \wedge l = j \pm 10 \\ 0, & \text{sonst} \end{cases}
 \end{aligned}$$

Die Verbindungsstärken der übrigen Orientierungen ergeben sich sinngemäß. In Abbildung 4.1.17 sind die inhibitorischen Verschaltungen und die lateralen Linking-Verbindungen innerhalb der Orientierungsschichten der Kantendetektoren schematisch dargestellt, die hinsichtlich der Inhibitions- und Linkingreichweiten eine Erweiterung der Linking- und Inhibitionsverschaltungen nach Spengler darstellen (Spengler et al. 1996). Die Inhibitionsreichweite wurde effektiv vergrößert, und die bei Spengler ausschließlich längs der Vorzugsorientierung angeordneten Linking-Kopplungen wurden keulenformig angeordnet. Abbildung 4.1.18 zeigt schematisch die Verteilung der Linking-Gewichte für einen  $0^\circ$ -Kantendetektor.



**Abb. 4.1.18:** Verteilung und Stärke der Linking-Verbindungen eines  $0^\circ$ -Kantendetektors. Dargestellt sind die Gewichte der Linking-Verbindungen eines  $0^\circ$ -Kantendetektor-Neurons zu seinen im hexagonalen Abtastgitter benachbarten  $0^\circ$ -Detektorelementen. Die Länge eines dargestellten Linienelements ist dabei proportional zur Gewichtungstärke der entsprechenden Verbindung.

Diese Art der Nachbarschaftskopplung führt zu einer Verstärkung der Impulsraten von koaxial in Vorzugsrichtung angeordneten benachbarten Neuronen, insbesondere bei Anordnung dieser Neuronen entlang einer längeren Kante. Zum anderen lassen sich ihre Antworten auch synchronisieren. Im folgenden sollen diese Effekte anhand des Beispiels einer räumlich und zeitlich verrauschten Kante demonstriert werden. Wie in Kapitel 3 beschrieben wurde, liefert die Kamera in jedem 40. Zeitschritt ein neues Bild. Alle innerhalb dieses Zeitintervalls erzeugten Bilder ergeben sich aus linearer Interpolation. Entsprechend wird auch hier das Bild  $K$  einer Kante in jedem 40. Zeitschritt neu berechnet zu:

$$K(x, y, A, B) = A \cdot \Theta(y) + B \cdot R(x, y), \quad (\text{Gl. 4.1.23})$$

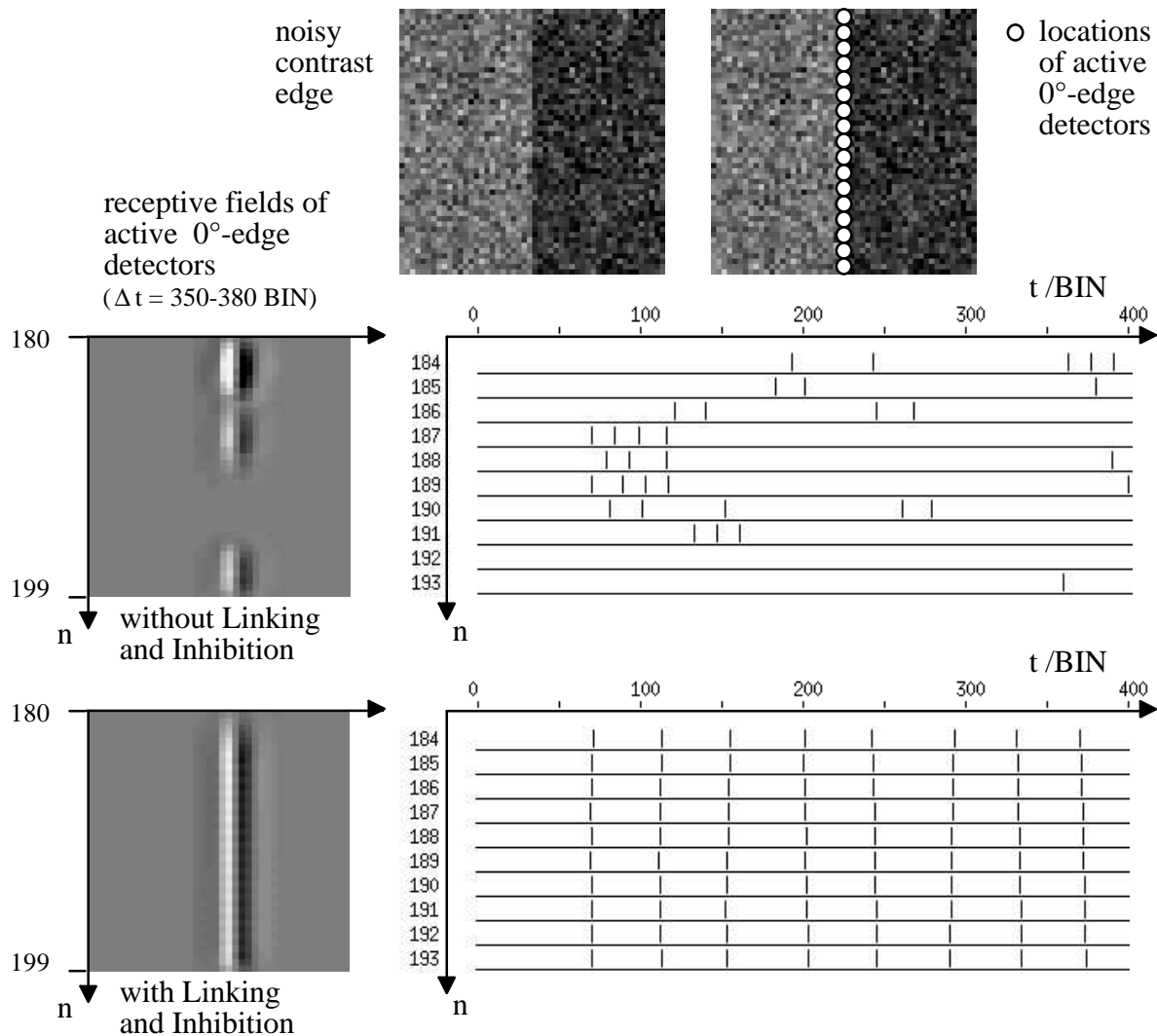
- mit:  $R$ : Ergebnis eines standardnormalverteilten (Gaußschen) Zufallsprozesses, das für jeden Punkt  $(x, y)$  gesondert berechnet wird,  
 $\Theta(y)$ : Sprungfunktion,  
 $A, B$ : Vorfaktoren, hier:  $A=40, B=20$ .

Abbildung 4.1.19 zeigt die verrauschte Kante und die rezeptiven Felder der aktiven  $0^\circ$ -Kantendetektorneuronen bei Anwendung von zwei verschiedenen Netzwerken. Die beiden auf das Bild angewandten Netze unterscheiden sich in der Verschaltung ihrer Kantendetektoren. Im ersten Netz sind die Kantendetektoren nicht miteinander gekoppelt. Hingegen sind im zweiten Netz die Kantendetektoren gemäß den oben beschriebenen Prinzipien sowohl über laterale Linking-Verbindungen als auch über inhibitorische Interneuronen miteinander verschaltet. Während ohne Linking-Verbindungen kaum synchrone Aktivität der Kantendetektoren und Anhebung bzw. Angleichung der Ausgangsraten möglich ist, gelingt dies mit Linking-Verbindungen sehr wohl. Für die Simulation werden die Parameter aus Tabelle 4.1.5 und 4.1.6 verwendet.

Kantendetektorneuron					
$\varphi$	$\tau^L$	$V^L$	$\tau^{\text{Inh}}$	$V^{\text{Inh}}$	$V^{\text{F1}}$
$0^\circ$	2.0 BIN	0.4	4.0 BIN	14.0	0.7
$30^\circ$	2.0 BIN	0.4	4.0 BIN	14.0	0.7
$60^\circ$	2.0 BIN	0.4	4.0 BIN	14.0	0.9
$90^\circ$	2.0 BIN	0.4	4.0 BIN	14.0	0.7
Interneuron					
$V^\Theta$	$\tau^\Theta$	$\Theta_0$	$V^{\text{F1}}$	$\tau^{\text{F1}}$	
32	1.0 BIN	2	1.0	20.0 BIN	

**Tab. 4.1.6:** Parameter der Kantendetektoren und Interneuronen innerhalb der Linking/Inhibitions-Architektur. Die Feeding1-Verstärkung ist für den  $60^\circ$ -Detektor, wie bereits in Tab. 4.1.5, stärker als für die übrigen drei Detektoren, um die kleineren Amplituden seines rezeptiven Feldes auszugleichen (vgl. Kap. 4.1.3).





**Abb. 4.1.19:** Synchronisation von Kantendetektoren durch Linking-Verbindungen bei Verwendung einer verrauschten Kontrastkante als Stimulus. **Oben:** Gezeigt wird das als Stimulus verwendete  $40 \cdot 40$  Pixel große Bild einer zeitlich und räumlich verrauschten Kontrastkante. **Links** sind jeweils die rezeptiven Felder der in einem bestimmten Zeitintervall aktiven  $0^\circ$ -Kantendetektorneuronen dargestellt. **Rechts** ist für zehn der entlang der Kante lokalisierten  $0^\circ$ -Kantendetektorneuronen das zeitliche Antwortverhalten aufgetragen. Eine Synchronisation der Detektorantworten bei gleichzeitiger Anhebung bzw. Angleichung der Raten wird hier erst durch die Verwendung von Linking-Verbindungen (**unten**) möglich.

Zur Darstellung der (exzitatorischen) rezeptiven Felder werden diese überlagert, entsprechend ihrer Impulsrate zeitlich tiefpaßgefiltert und in den Bildbereich zurückprojiziert (Kopecz et al. 1996, Mohraz 1996). Durch diese Methode wird eine Schätzung der Bildkonfiguration erhalten, die die Detektorantworten hervorgerufen haben könnte.

Im folgenden soll dieser durch Linking-Kopplungen hervorgerufene Effekt quantitativ untersucht werden. Dazu wird anhand des eben verwendeten Netzes (mit Linking/Inhibitions-Architektur) das Signal/Rausch-Verhältnis der neuronalen Antwort auf die in Gleichung 4.1.23 beschriebene zeitlich und räumlich verrauschte Kante  $K$  in Abhängigkeit von der Linking-Kopplungsstärke betrachtet.

Das Signal/Rausch-Verhältnis  $SNR$  werde definiert als:

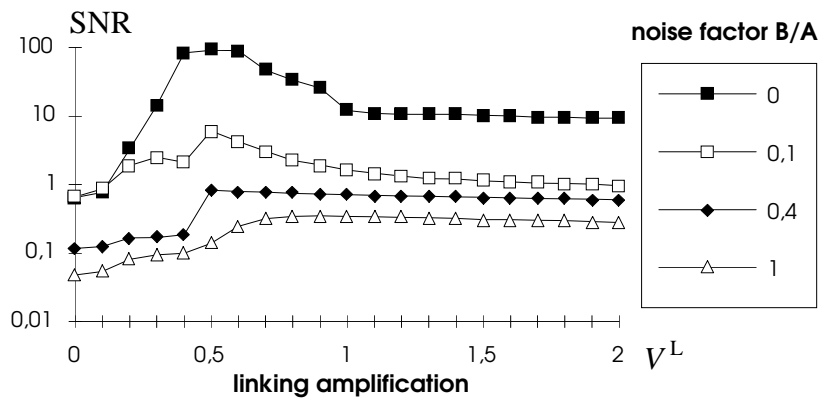
$$SNR = \frac{S(K, \Delta t)}{\Gamma(K_r, \Delta t) - S(K, \Delta t)}, \quad (\text{Gl. 4.1.24})$$

dabei sind:	$S(K, \Delta t)$	Summe aller von den entlang der unverrauschten $0^\circ$ -Kante $K(x, y, A=40, B=0)$ angeordneten $0^\circ$ -Kantendetektoren im Zeitintervall $\Delta t$ generierten Aktionspotentiale als Antwort auf diese Kante,
	$\Gamma(K_r, \Delta t)$	Summe aller von den Kantendetektoren beliebiger Vorzugs-Orientierung und beliebigen Ortes im Zeitintervall $\Delta t$ generierten Aktionspotentiale als Antwort auf die verrauschte $0^\circ$ -Kante $K_r = K(x, y, A=40, B > 0)$ ,
	$\Delta t$	Zeitintervall $t_2 - t_1 = 1000$ BIN, beginnend bei $t_1 = 200$ BIN, um mögliche Einschalt- und Einschwingeffekte ausschließen zu können.

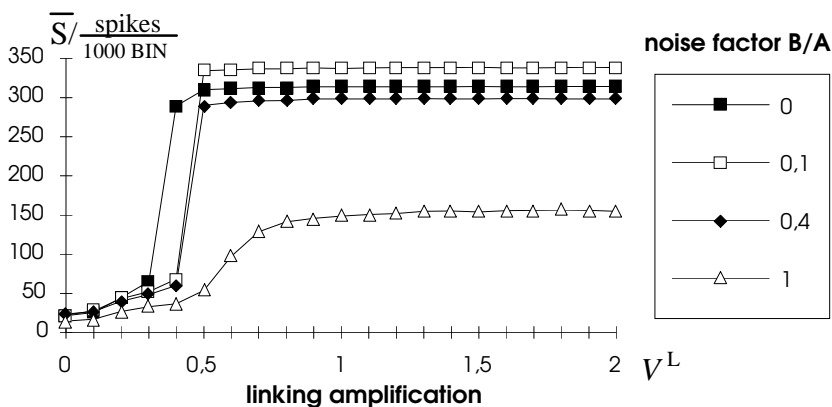
Die Verbindungsstärken der Linking-Kopplungen, die in den Gleichungen 4.1.22 a-d beschrieben sind, sind mit dem jeweiligen Linking-Verstärkungsfaktor  $V^L$  zu multiplizieren, um die tatsächliche Beeinflussung des Linking-Potentials zu erhalten. Da der Faktor  $V^L$  für alle Kantendetektoren identisch ist, eignet er sich zur Variation der Koppelstärken aller Linking-Verbindungen. In Abb. 4.1.20 ist das Signal/Rausch-Verhältnis  $SNR$  in Abhängigkeit vom Linking-Verstärkungsfaktor  $V^L$  aller Kantendetektoren für verschieden stark verrauschte Kanten dargestellt. Um das Rauschen nicht zum großen Teil durch einen hohen Schwellen-Offset zu unterdrücken, wird hier zu Demonstrationszwecken der Schwellen-Offset  $\Theta_0$  möglichst klein gewählt. Aufgrund des Schwellenvergleichs des Acceleratorneurons ist  $\Theta_0 > 0$  zu wählen, da sonst die Neuronen auch ohne ein Feeding-Signal Aktionspotentiale generierten. Daher wird in dieser Untersuchung der kleinste technisch mögliche Wert für den Schwellen-Offset, nämlich  $\Theta_0 = 2$ , gewählt.

Bereits durch kleine Linking-Verstärkungsfaktoren wird das  $SNR$  gegenüber dem  $SNR$  in einem Netz ohne Linking-Verbindungen deutlich angehoben. Die gezeigten Kurven weisen geringere  $SNR$  bei stärkerem eingangsseitigen Rauschen auf. Weiterhin zeigen sie jeweils ein globales Maximum, das für größere Rauschamplituden weniger stark ausgeprägt ist, aber in allen Fällen etwa bei  $V^L = 0.5$  liegt. Nach Überschreitung dieses Verstärkungsfaktors bleibt die Rate der Signalaktionpotentiale nahezu konstant (Abb. 4.1.21), während die Anzahl der durch Rauschen hervorgerufenen Aktionspotentiale langsam ansteigt, was in diesem Bereich zu schwach abfallenden  $SNR(V^L)$ -Kurvenverläufen führt. Aus Abb. 4.1.21 ist auch ersichtlich,

daß bei größeren Linking-Verstärkungsfaktoren die Raten sehr hoch ausfallen, daher wird im folgenden ein etwas geringerer Verstärkungsfaktor  $V^L=0.4$  verwendet. Zudem vergrößert sich das Signal/Rauschverhältnis zusätzlich durch Anhebung des Schwellen-Offsets.

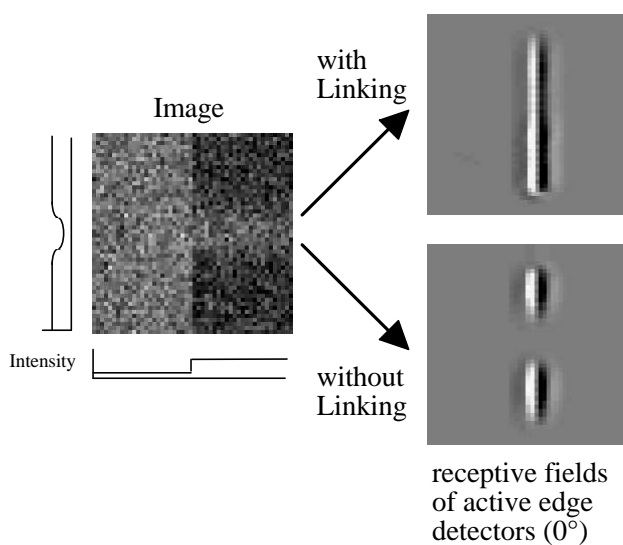


**Abb. 4.1.20:** Vergrößerung des Signal/Rausch-Verhältnisses durch Linking-Kopplungen. Aufgetragen ist logarithmisch das Signal/Rausch-Verhältnis der Kantendetektorantworten auf unterschiedlich stark verrauschte Kanten gegen den Linking-Verstärkungsfaktor  $V^L$ . Auf die unverrauschte  $0^\circ$ -Kante (gefüllte Kästchen) als auch auf die verrauschten  $0^\circ$ -Kanten antworten bei dem hier gewählten niedrigen Schwellen-Offset  $\Theta_0=2$  immer auch Kantendetektoren anderer Vorzugsorientierungen. Diese Antworten werden nicht als zum Signal gehörig, sondern als Fehlantworten interpretiert und verkleinern somit neben den durch Signalrauschen hervorgerufenen Antworten das SNR (Siehe Text).



**Abb. 4.1.21:** Signal-Verstärkung durch Linking-Kopplungen. Aufgetragen ist die durchschnittliche Rate an Aktionspotentialen eines entlang der  $0^\circ$ -Kante  $K$  angeordneten  $0^\circ$ -Kantendetektors bei unterschiedlich starkem überlagerten Rauschen gegen den Linking-Verstärkungsfaktor  $V^L$ . Alle Kurven zeigen ohne Kopplung nur sehr kleine Raten und etwa bei  $V^L=0.5$  einen starken Anstieg der Signalaktivität, wobei der Anstieg mit zunehmendem Rauschen bei etwas größeren Verstärkungsfaktoren erfolgt. Weiterhin zeigt sich, daß durch sehr schwaches Rauschen (offene Kästchen) die Signalaktivität gegenüber dem Ausgangssignal (gefüllte Kästchen) erhöht werden kann, da Rauschen dazu beiträgt, bei vorhandenem Signal die Schwelle (einschließlich des Schwellenoffsets) zu überwinden. Sind die Rauschamplituden stark genug, um auch ohne Signal die Schwelle zu überwinden, so führt das in der Verschaltung zu inhibitorischen Einflüssen, so daß die Raten vergleichsweise kleiner ausfallen (Rauten oder Dreiecke).

Die im Vergleich zu Feeding- und Schwellenzeitkonstanten kurze Linking-Zeitkonstante ( $\tau^L=2.0$  BIN) stellt eine wichtige Voraussetzung für das Erreichen fast synchroner Einzelspike-Aktivität dar. Sowohl Synchronisation als auch die gegenseitige Verstärkung benachbarter Neuronen verbessern die Kantendetektion und unterstützen die Kontinuität von Konturrepräsentationen durch Ausnutzen räumlicher Information, wie auch das folgende Beispiel zeigt. Eine verrauschte Kontrastkante mit kontrastschwächerem Verlauf im Zentrum wird als Stimulus für die beiden bereits oben erwähnten Netzwerke verwendet (Abbildung 4.1.22). Das Netz ohne Linking-Verbindungen kann in der Repräsentation diese sehr wahrscheinliche Störung der Kante nicht überbrücken.

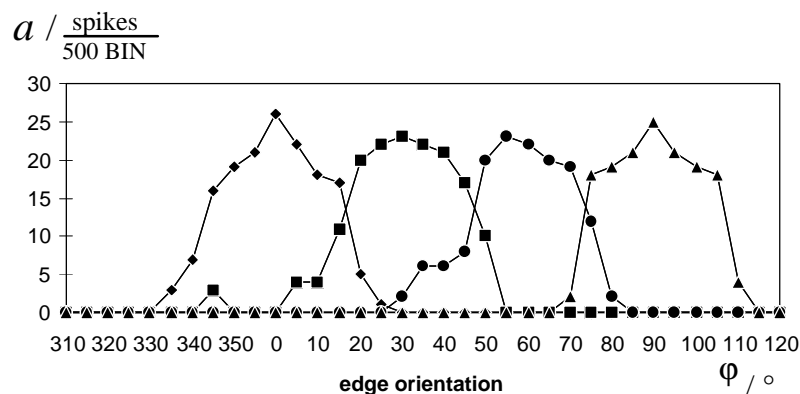


**Abb. 4.1.22:** Unterstützung der Kontinuität von Konturen durch Verwendung von Linking-Verbindungen. **Links:** Eine zeitlich und räumlich verrauschte Kontrastkante mit kontrastschwächeren Anteilen im Zentrum wird als Stimulus verwendet, wobei die Grauwertverläufe entlang der Kante und senkrecht zur Kante unterhalb bzw. links des Bildes skizziert sind. **Rechts** sind die rezeptiven Felder der aktiven  $0^\circ$ -Kantendetektoren dargestellt. Nur bei Verwendung von Linking-Verbindungen wird die Kante vollständig detektiert.

Hingegen sind die im Zentrum der Kante angeordneten Neuronen mit Linking-Verbindungen in der Lage, die kontrastschwächeren Teile der verrauschten Kante zu repräsentieren, was durch die Einbettung der Neuronen in den Kontext der beiden iso-orientierten Kanten mit starkem Kontrast ermöglicht wird. Die angrenzenden Neuronen erhöhen über ihre Linking-Verbindungen die Empfindlichkeit der in der Kantenmitte angeordneten Neuronen. Dieselbe Verbesserung läßt sich jedoch nicht durch eine Absenkung der Schwellen aller Detektorneuronen erreichen, da dies auch eine größere Empfindlichkeit des gesamten Netzes auf Rauschen mit sich bringt. Die Kontinuitätsunterstützung ist auch eine wesentliche Voraussetzung für die Vermeidung von Fehlern innerhalb der anschließenden Detektion von Linienenden (siehe Kap. 4.2).

Für die Kantendetektoren ergeben sich die in Abbildung 4.1.23 dargestellten Orientierungscharakteristiken. Daraus geht jedoch nicht die geforderte eindeutige Kantenrepräsentation hervor. Beispielsweise antworten sowohl der  $0^\circ$ - als auch der  $30^\circ$ -Kantendetektor auf eine Kontrastkante einer Orientierung von  $10^\circ$ . Der Grund für das Auftreten dieser Mehrdeutigkeiten trotz Anwendung des oben beschriebenen Inhibitionsmechanismus ist in der Synchronisation der Detektorantworten zu suchen. Antworten von Kantendetektoren

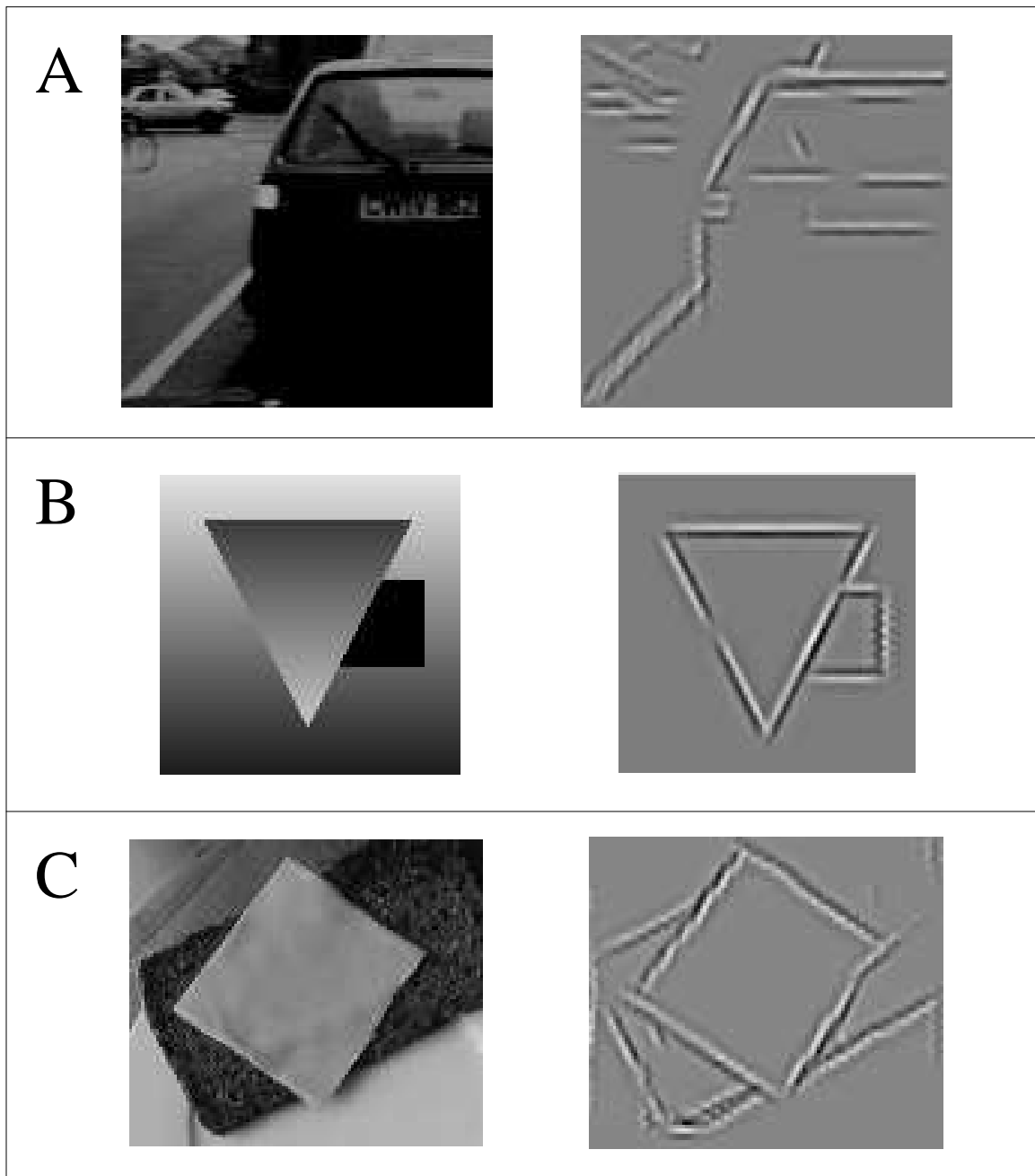
verschiedener Vorzugsorientierungen werden vom Netz meist in unterschiedliche Phasenlagen gebracht und entziehen sich dadurch teilweise der Inhibition, die durch Kantendetektoren anderer Orientierungen initiiert wird. Durch eine vergleichsweise lange Linking-Zeitkonstante läßt sich jedoch die Wechselwirkung zwischen Kantendetektoren unterschiedlicher Orientierungen verstärken. Auf diese Weise wird zwar Synchronisation weitgehend vermieden, auf der anderen Seite aber weiterhin die wechselseitige Verstärkung benachbarter Detektoren derselben Vorzugsrichtung gewährleistet. Synchronisation von Kantendetektor-Aktivität, die Kontur repräsentieren könnte, ist somit nicht als Aufgabe der Kantendetektorschichten anzusehen, vielmehr geht es hier zunächst nur um eine eindeutige Repräsentation.



**Abb. 4.1.23:** Orientierungscharakteristiken der Kantendetektoren bei Verwendung der Linking-/Inhibitionsarchitektur. Aufgetragen sind die maximalen Ausgangsimpulsraten  $a$  von 0°-, 30°-, 60° und 90°-Detektor in Abhängigkeit von der Kantenorientierung einer stimulierenden idealen Kante mit Grauwertdifferenz  $\Delta G = 64$  für eine kleine Linkingzeitkonstante  $\tau^L = 2.0$  BIN und den Schwellenoffsetwert  $\Theta_0 = 6$  (Parameter: Tab. 4.1.6).

Entsprechend lassen sich die Richtungscharakteristiken durch Verwendung einer relativ großen Linking-Zeitkonstanten (hier:  $\tau^L = 40.0$  BIN) verändern. Um die Gesamtaktivität auszugleichen, wird zusätzlich die Linking-Verstärkung verringert ( $V^L = 0.2$ ). Neben der Eindeutigkeit spielt auch die Gleichbehandlung von Kanten unterschiedlicher Orientierungen eine wichtige Rolle. Daher werden nicht nur die entlang der jeweiligen Vorzugsrichtung angeordneten Nachbarneuronen über Linking-Verbindungen gekoppelt, sondern auch solche, die davon im bestimmten Maße abweichen. Das führt schließlich zu der keulenförmigen Anordnung der über Linking-Verbindungen gekoppelten Kantendetektoren. In Abbildung 4.1.24 sind drei Simulationsbeispiele aufgeführt, die mit dem vorgestellten Netz zur Kantendetektion erzeugt wurden. Aufgrund des relativ hohen Schwellenoffsets werden sehr kontrastschwache Kanten nicht detektiert, wie aus den Rückprojektionen der Kantendetektorantworten in den beiden natürlichen Szenen (A und C) zu erkennen ist. Räumliches Rauschen oder Störungen, wie sie die Vollkornbrotoberfläche der Frühstücksszene liefert, werden weitgehend unterdrückt. Beispiel B zeigt eine weitere Anwendung für die Unterstützung von Kontinuität von Konturen, hier für den Fall der Kontrastumkehr einer Kante. An Orten, an denen Kantendetektoren kein Feeding-Eingangs-

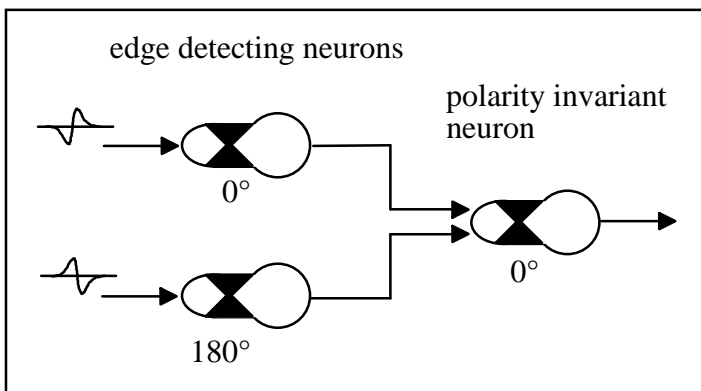
signal erhalten, weil kein Kontrast vorliegt, können diese Detektoren über Linking-Verbindungen nicht aktiv werden. Jedoch helfen weiterreichende Linking-Kopplungen angrenzenden Detektoren, die nur schwache Feeding-Signale erhalten, überschwellig zu werden und damit die Repräsentation der Kontrastumkehrlücke möglichst klein zu halten.



**Abb. 4.1.24:** *Simulationsbeispiele der Kantendetektion.* Als Stimulus wird (A) eine Verkehrsszene, (B) eine künstliche Szene und (C) eine Frühstücksszene verwendet. **Rechts** sind die rezeptiven Felder, der jeweilig aktiven Kantendetektorneuronen bei Verarbeitung dieser Szenen dargestellt (Rückprojektionen in den Bildbereich). Neben den Konturen des Autos im Vordergrund werden in der Verkehrsszene auch die Konturen des Mittelstreifens und die von Hintergrundstrukturen detektiert. Die künstliche Szene besteht aus einem Dreieck, das durch einen variablen Grauwertgradienten definiert wird und ein schwarzes Rechteck zu überdecken scheint. Die Konturen beider Objekte werden detektiert. In der Frühstücksszene überdeckt eine Scheibe Käse teilweise ein auf einem Teller liegendes Vollkornbrot. Die körnige Oberfläche des Brotes und zu kontrastschwache Kanten werden aufgrund des Schwellenoffsets  $\theta_0=6$  nicht detektiert.

#### 4.1.5 Polaritätsinvariante Modellneuronen

Den zwölf Orientierungsschichten sind weitere sechs neuronale Schichten derselben Größe nachgeschaltet. Durch Konvergenz von zwei Kantendetektorneuronen desselben Ortes, aber entgegengesetzter Polarität, auf ein nachgeschaltetes Neuron wird eine Invarianz gegenüber Kontrastpolarität erreicht. Diese Eigenschaft der Invarianz gegenüber der Richtung des Kontrastes ist ein wichtiger Unterschied zwischen einer Komplexzelle und einer Simplezelle (z.B. Hubel und Wiesel 1962, Gouras und Krüger 1979, Heggelund 1981b, Grossberg 1987). Ein zweiter wichtiger Unterschied ist die deutlich größere Positionsinvarianz der Komplexzelle, die hier aber nicht als solche modelliert wird. Daher wird ein nachgeschaltetes Neuron im folgenden als polaritätsinvariantes Neuron oder *PI-Neuron* bezeichnet.



**Abb. 4.1.25:** Verschaltung von Kantendetektoren entgegengesetzter Kontrastpolarität auf ein PI-Neuron. Die Ausgangssignale zweier Kantendetektoren desselben Ortes aber entgegengesetzter Vorzugsorientierung werden auf den Feedingeingang des PI-Neurons gegeben.

Zusätzlich läßt sich durch diese Konvergenz die Anzahl der Neuronen in den nachfolgenden Verarbeitungsstufen erheblich verringern, wodurch sich Erleichterungen für die Entwicklung und die praktische Anwendung ergeben.

Das PI-Neuron am Ort  $(m, n)$  und der Orientierung  $\varphi$  erhält Signale über seinen Feeding-Eingang vom Kantendetektorneuron  $(k, l)$  der Orientierung  $\varphi$  und vom Kantendetektorneuron  $(k, l)$  der Orientierung  $\varphi+180^\circ$  (Abbildung 4.1.25). Die PI-Neuronen sind, wie schon die Kantendetektoren, über Linking-Verbindungen gekoppelt, um auch die Repräsentation einer ausgedehnten Kante zu unterstützen, entlang der die Kontrastpolarität wechselt. Die konkreten Parameter der PI-Neuronen sind in Tabelle 4.1.7 zusammengefaßt.

Schwellen-Verstärkung $V^\Theta$	Schwellen-Zeitkonstante $\tau^\Theta$	Schwellen-Offset $\Theta_0$	Feeding-potential-Verstärkung $V^F$	Feeding-potential-Zeitkonstante $\tau^F$	Linking-potential-Verstärkung $V^L$	Linking-potential-Zeitkonstante $\tau^L$
128	10 BIN	6	1,0	40 BIN	3,0	2,0 BIN

**Tab. 4.1.7:** Parameter der PI-Neuronen.

#### 4.1.6 Diskussion

Die vorgestellten Kantendetektoren stellen bezüglich ihres räumlichen Sensitivitätsprofils eine Näherung an bekannte Gaborfunktionen dar. Aufgrund der durch das Modell gegebenen Randbedingungen kommt es aber zu gewissen Abweichungen. Als erste Restriktion sind die Asymmetrien des quasihexagonalen Gitters zu nennen. Unterschiedliche Verhältnisse von Breite zu Länge der Detektoren führen zu unterschiedlichen Breiten der Orientierungscharakteristiken, die sich aber weitgehend durch geeignete Gewichtungen ausgleichen lassen. Schließlich sind auch Rastereffekte zu berücksichtigen, die bei der Rotation der zur Bestimmung der Orientierungscharakteristiken verwendeten Kante gewisse Schwankungen in diesen Kurven hervorrufen.

Neben den räumlichen Eigenschaften spielen auch die zeitlichen Eigenschaften der Kantendetektoren eine wichtige Rolle. Bereits die X-Zellen kodieren im Modell die lokale Stärke des Intensitätskontrastes in Impulsraten des stationären Bereichs (anhaltende Antwort). Auch die Orientierung wird in Impulsraten kodiert. Dadurch läßt sich mit einem Detektor allein nicht eindeutig feststellen, ob seine Antwort durch eine optimal ausgerichtete, aber kontrastschwache Intensitätskante hervorgerufen oder ob sie von einer etwas ungünstiger ausgerichteten, aber kontraststarken Intensitätskante ausgelöst wurde. Auch ein einzelner Punkt könnte der Auslöser sein, da X-Zellen bevorzugt auf solche singulären Stimuli ansprechen. Bereits in Kapitel 3 wurde die Skalierung des X-Zellenpotentials bei Anwendung einer Sigmoidfunktion gezeigt, die die Gefahr einer Kantendetektorantwort bei Stimulierung einer der vorgeschalteten X-Zellen durch eine Singularität verringert. Weiterhin wird durch die Anwendung einer gemeinsamen inhibitorischen Rückkopplung Orientierung nur noch bedingt durch Impulsraten der Kantendetektoren kodiert und somit die hier angesprochenen Mehrdeutigkeiten weitgehend aufgehoben. Der auf diese Weise gewonnenen Eindeutigkeit steht das Prinzip der Kodierung von Information durch eine größere Anzahl von Neuronen (*Populationskode*) gegenüber, das aus vielen neurophysiologischen Untersuchungen vor allem am Kortex des Affen bekannt ist (z.B. Victor et al. 1994, Georgopoulos et al. 1986, Georgopoulos et al. 1992, Georgopoulos et al. 1993). Einzelne Neuronen besitzen dabei sehr breite Richtungscharakteristiken, dennoch kann die Population als Ganzes sehr präzise Antworten geben. Eine Reihe von Modellen wurde entwickelt, die versuchten, diesen Sachverhalt zu erklären (z.B. Modelle orientierungssensitiver kortikaler Zellen (Vogels 1990) oder Modelle kortikaler Zellen, die Merkmale höherer Ordnung kodieren (Mallot 1996)). Die Population liefert als Gesamtantwort einen *Populationsvektor*, der sich aus der Summe vieler neuronaler Einzelvektoren ergibt (Georgopoulos et al. 1986).

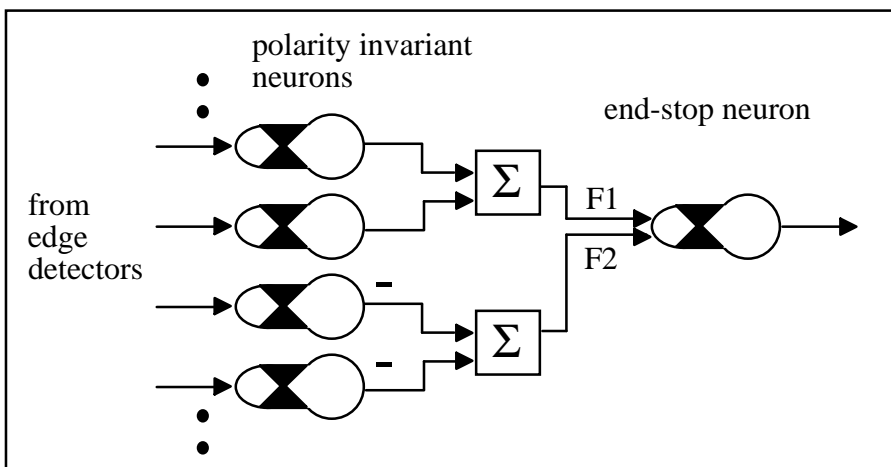
Ein wesentlicher Vorteil dieses Prinzips liegt in der Unabhängigkeit gegenüber Fehlen oder Falschantworten einzelner Neuronen. Im vorliegenden Modell können aus Gründen der Netzgröße keine großen lokalen Populationen gleichartiger Neuronen verwendet werden. Dennoch ist für jeden Ort eine Population zur Ermittlung der Kantenorientierung prinzipiell



denkbar. Durch Interpolation von Antworten benachbarter Neuronen mit überlappenden RF-Eigenschaften könnten eindeutige und genauere Resultate erhalten werden. Zu diesem Zwecke müßte lediglich die inhibitorische Wirkung des gemeinsamen Interneurons etwas reduziert werden, was beispielsweise auch, wie in Kapitel 4.1.4 angedeutet, durch Verwendung einer kurzen Linking-Zeitkonstanten realisierbar ist. Da aber die anschließende Detektion der Linienenden auf einer eindeutigen Repräsentation der Linien aufbaut, kann der Populationskode nicht verwendet werden. Für das Netz entstehen dadurch keine merklichen Nachteile, da zum einen der Ausfall einzelner Neuronen im Modell keine Rolle spielt und zum anderen das Netz Kanten beliebiger Orientierungen weitestgehend gleichberechtigt behandelt, trotz Diskretisierung der Orientierungen auf zwölf Vorzugsrichtungen. Darüberhinaus lassen sich die Detektoreigenschaften durch Verwendung eines Linking/Inhibitions-Mechanismus verbessern. Neben der Beeinflussung der Richtungscharakteristika (siehe auch Spengler et al. 1996) sind dadurch auch die Unterstützung der Repräsentationen von Konturkontinuität und die Verbesserung des Signal-/Rauschverhältnisses möglich.

## 4.2 Detektion von Linienenden und Krümmungen

Noch wichtiger als die Information über Kanten und Linien ist die Existenz starker Krümmungen, Ecken oder Linienenden innerhalb der zu verarbeitenden Szene für das visuelle System. Der Grund dafür ist, daß solche Vertizes und Linienenden einen besonders hohen Informationsgehalt besitzen (vgl. Kap. 2). Durch eine weitgehende Beschränkung der Kodierung von Bildinformation auf Bildpositionen mit hohem Informationsgehalt wird eine entscheidende Datenreduktion erreicht. Im Kortex könnten Zellen mit endinhibierten rezeptiven Feldern zur Kodierung von Linienenden und Krümmungen dienen (Peterhans et al. 1986, Peterhans und von der Heydt 1989, Heitger et al. 1992), die in älteren Modellen (Hubel und Wiesel 1965) als Hyperkomplex- oder Endstop-Zellen den Komplexzellen in der hierarchischen Struktur des Kortex folgen (vgl. Kap. 2.1).



**Abb. 4.2.1:** Exzitatorische und inhibitorische Verschaltung von PI-Neuronen auf End-Stop-Dektoren.

Auch im Modell werden Detektoren von Linienenden und Krümmungen implementiert, nämlich durch konvergente Feeding-Verbindungen von den Schichten der PI-Neuronen (Abb. 4.2.1). Ein hier modellierter End-Stop-Detektor erhält exzitatorische Eingangssignale von drei in Vorzugsorientierung kollinear angeordneten PI-Neuronen. Außerdem ist er über negative Gewichte mit PI-Neuronen verbunden, die in einem angrenzenden ausgedehnten Bereich angeordnet sind. Das entsprechende Gewicht  ${}^{ESD}W_{opmn}^\varphi$  der Verbindung von einem PI-Neuron der Orientierung  $\varphi$  am Ort  $(m, n)$  zu einem End-Stop-Detektorneuron der Orientierung  $\varphi$  am Ort  $(o, p)$  ist gegeben durch:

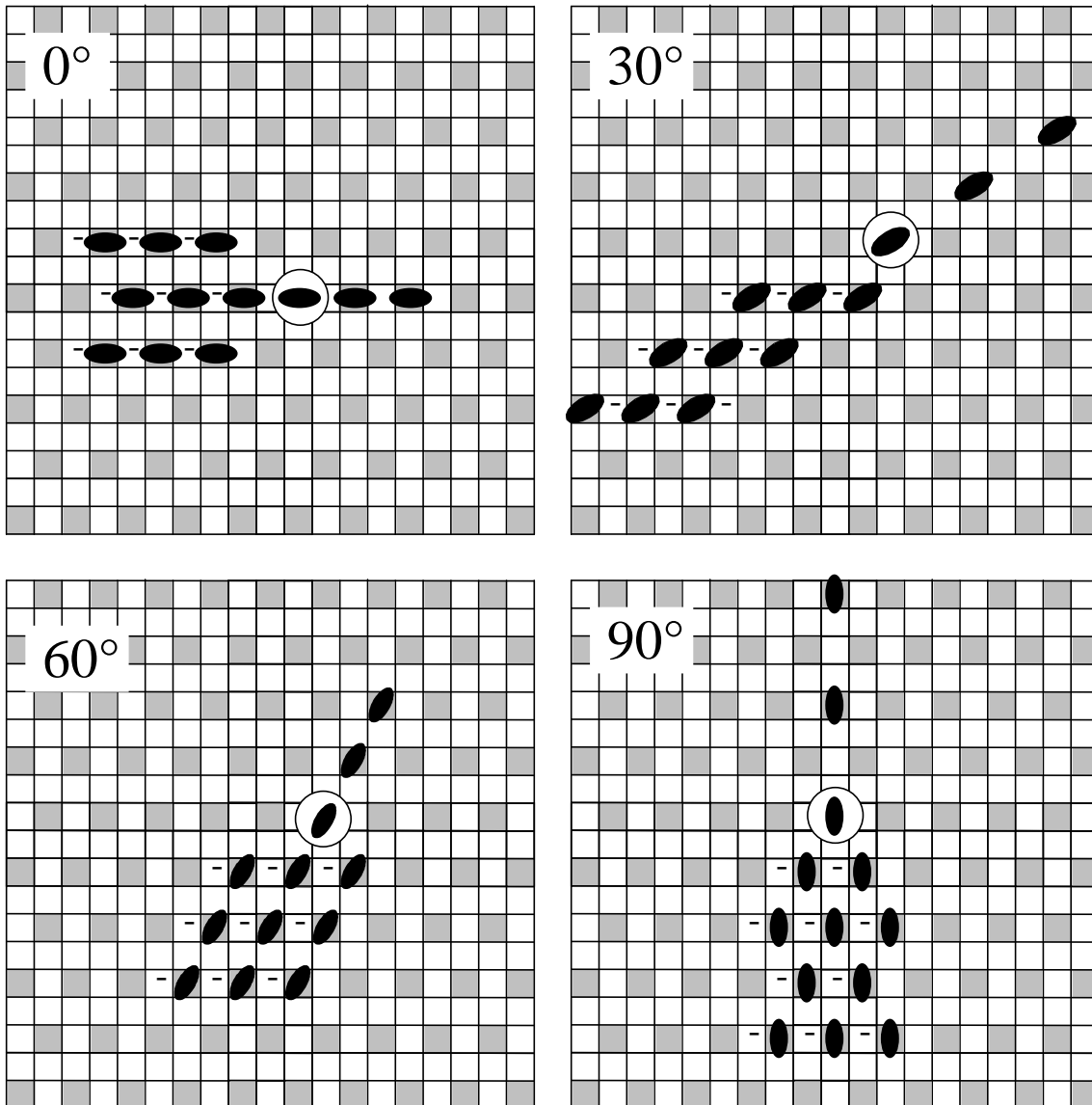
$${}^{ESD}W_{opmn}^{0^\circ} = \begin{cases} -7.0, & \text{falls } ((m = o - 3 \vee m = o - 5 \vee m = o - 7) \wedge n = p \pm 2) \\ & \vee ((m = o - 2 \vee m = o - 4 \vee m = o - 6) \wedge n = p) \\ 3.5, & \text{falls } (m = o \vee m = o + 2 \vee m = o + 4) \wedge n = p \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.2.1.a})$$

$${}^{ESD}W_{opmn}^{30^\circ} = \begin{cases} -7.0, & \text{falls } ((m = o - 1 \vee m = o - 3 \vee m = o - 5) \wedge n = p - 2) \\ & \vee ((m = o - 4 \vee m = o - 6 \vee m = o - 8) \wedge n = p - 4) \\ & \vee ((m = o - 7 \vee m = o - 9 \vee m = o - 11) \wedge n = p - 6) \\ 3.5, & \text{falls } (m = o \wedge n = p) \vee (m = o + 3 \wedge n = p + 2) \\ & \vee (m = o + 6 \wedge n = p + 4) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl.4.2.1.b})$$

$${}^{ESD}W_{opmn}^{60^\circ} = \begin{cases} -7.0, & \text{falls } ((m = o + 1 \vee m = o - 1 \vee m = o - 3) \wedge n = p - 2) \\ & \vee ((m = o \vee m = o - 2 \vee m = o - 4) \wedge n = p - 4) \\ & \vee ((m = o - 1 \vee m = o - 3 \vee m = o - 5) \wedge n = p - 6) \\ 3.5, & \text{falls } (m = o \wedge n = p) \vee (m = o + 1 \wedge n = p + 2) \\ & \vee (m = o + 2 \wedge n = p + 4) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl.4.2.1.c})$$

$${}^{ESD}W_{opmn}^{90^\circ} = \begin{cases} -7.0, & \text{falls } ((m = o \pm 2 \vee m = o) \wedge (n = p - 4 \vee n = p - 8)) \\ & \vee (m = o \pm 1 \wedge (n = p - 2 \vee n = p - 6)) \\ 3.5, & \text{falls } m = o \wedge (n = p \vee n = p + 4 \vee n = p + 8) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl.4.2.1.d})$$

Aus diesen folgen die Gewichte der übrigen Orientierungen sinngemäß durch Rotation. Für jede Kante einer bestimmten Orientierung können zwei Kantenenden bestimmt werden. Daher ergeben sich für die Detektion von Linienenden zwölf neuronale Schichten unterschiedlicher Vorzugsorientierung. Abbildung 4.2.2 zeigt entsprechend der Gewichte der Gleichungen 4.2.1 eine schematische Darstellung der rezeptiven Felder. In Abbildung 4.2.3 ist das Sensitivitätsprofil des  $0^\circ$ -End-Stop-Detektors exemplarisch dargestellt. Dieses ergibt sich aus der Superposition der rezeptiven Felder von  $0^\circ$ - und  $180^\circ$ - Kantendetektoren unter



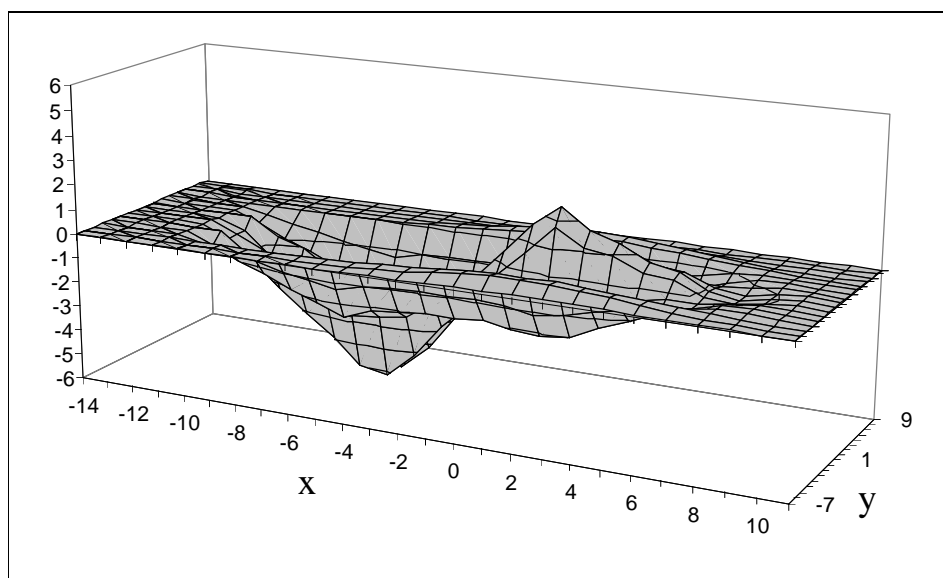
**Abb. 4.2.2:** Schematische Darstellung der rezeptiven Felder von End-Stop-Detektoren im hexagonalen Abtastraster für die Orientierungen  $0^\circ$ ,  $30^\circ$ ,  $60^\circ$  und  $90^\circ$ . Der Ort des Detektors ist durch einen weißen Kreis gekennzeichnet. Die schwarzen Ellipsen geben die Positionen der auf den End-Stop-Detektor inhibitorisch (Minus-Zeichen) oder exzitatorisch verschalteten PI-Neuronen an.

Berücksichtigung der in den Gleichungen 4.2.1 angegebenen Gewichtungen. Zudem werden die inhibitorischen Teilbereiche der negativ gewichteten rezeptiven Felder vernachlässigt, da diese im Profil als positive Bereiche sichtbar würden, obwohl eine Kante, die durch einen dieser Bereiche läuft, den Detektor nicht aktivieren kann. Die Aussagekraft des resultierenden Profils ist jedoch begrenzt, da weder die weiteren Nichtlinearitäten der drei zugrundeliegenden Verarbeitungsstufen, noch eine Unterscheidung zwischen ON- und OFF-Bereichen berücksichtigt werden können.

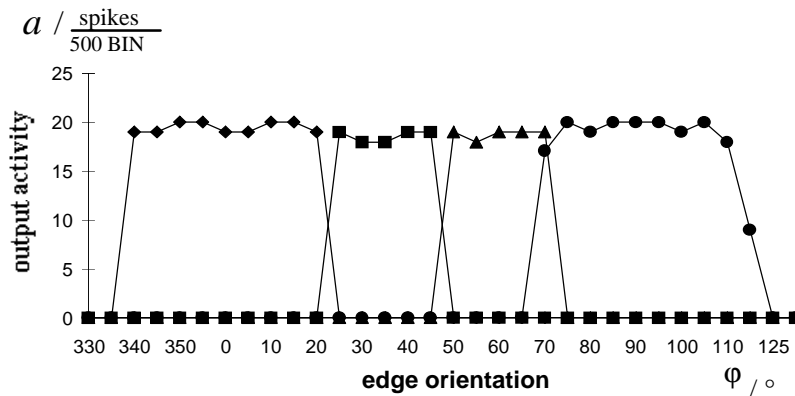
Eine kontinuierliche Kante, die durch das rezeptive Feld eines End-Stop-Detektors läuft, ruft keine Antwort hervor, da die Exzitation durch eine noch stärkere Inhibition aufgehoben wird. Der Detektor antwortet nur, wenn der inhibitorische Teilbereich nicht durch eine Kante

überdeckt wird. Die ausgedehnte Struktur des inhibitorischen Teilbereichs verringert deutlich die Empfindlichkeit auf Kantenstörungen.

Wenn zum Beispiel eine längere Kante nicht exakt in  $0^\circ$ - (oder  $90^\circ$ -) Orientierung ausgedehnt ist, was meistens der Fall ist, wird die Kante durch gestückelte Segmente in der  $0^\circ$ - ( $90^\circ$ -) Kantendetektorschicht repräsentiert. Daher dürfen gestückelte Repräsentationen nicht als Linienende interpretiert werden. Dies wird durch die räumliche Ausdehnung des inhibitorischen Teilbereichs gewährleistet. An dieser Stelle sei nochmals auf die Wichtigkeit von Linking-Kopplungen in der vorherigen Verarbeitungsstufe für die Detektion von Linienenden hingewiesen. Eine Kante, die nur stückweise detektiert wird, so daß sie größere Lücken aufweist, ruft Antworten von End-Stop-Detektoren an diesen Orten hervor. Sind die Lücken auf Rauschen oder Störungen zurückzuführen, so helfen die Linking-Verbindungen, die Kante vollständig zu detektieren und damit die Detektion falscher Linienenden zu vermeiden. Die vollständige Überbrückung der Lücken gelingt jedoch nur dort, wo Kantendetektoren Feeding-Eingangssignale erhalten. Bei Kontrastumkehr kann es einen Bereich geben, in dem (fast) kein Kontrast und damit auch kein Feeding-Signal vorliegt (vgl. Abb. 4.1.24 B). Hier läßt sich End-Stop-Detektion nur mit größeren inhibitorischen Teilbereichen der End-Stop-Detektoren vermeiden. Auf der anderen Seite verringert ein zu großer inhibitorischer Teilbereich das Auflösungsvermögen. Die hier gezeigten Detektoren stellen einen geeigneten Kompromiß dar.



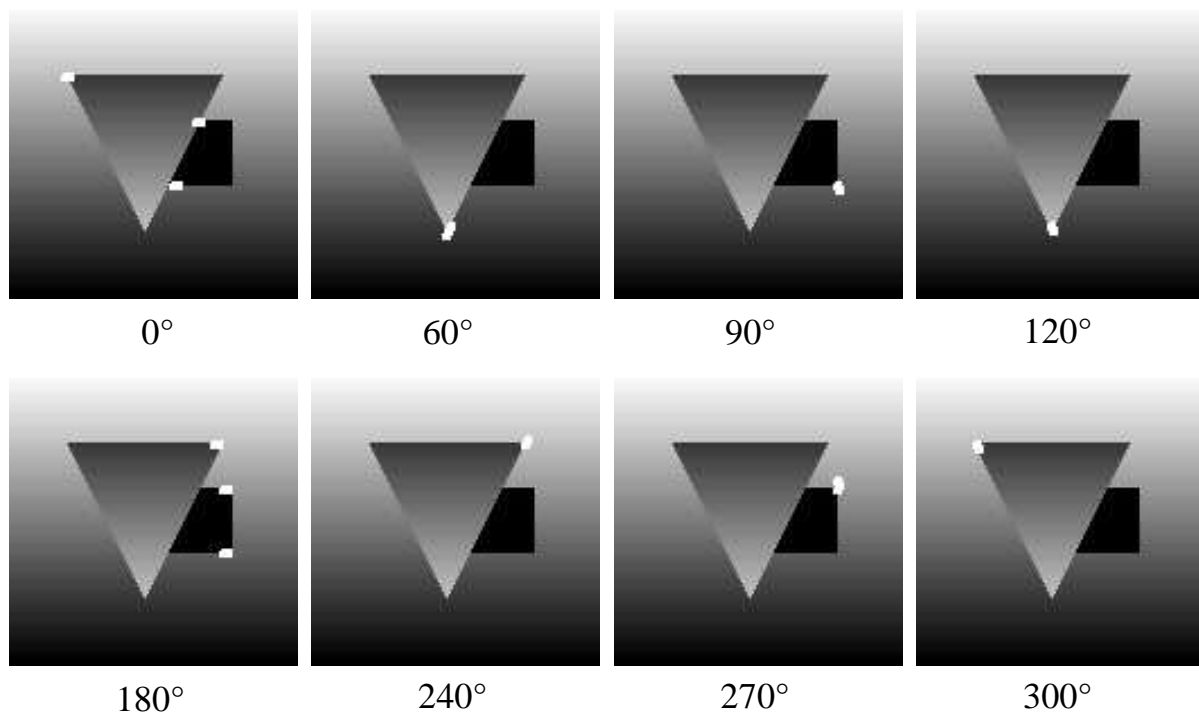
**Abb. 4.2.3:** *Sensitivitätsprofil des  $0^\circ$ -End-Stop-Detektors.* Es ergibt sich aus der Superposition der rezeptiven Felder von  $0^\circ$ - und  $180^\circ$ - Kantendetektoren. Dabei werden aber die inhibitorischen Teilbereiche der negativ gewichteten rezeptiven Felder vernachlässigt, da diese im Profil als positive Bereiche sichtbar würden, obwohl eine Kante, die durch einen dieser Bereiche läuft, den Detektor nicht aktivieren kann.



**Abb. 4.2.4:** Orientierungscharakteristiken von End-Stop-Detektoren. Aufgetragen ist die Ausgangsrate  $a$  des 0°- (Rauten), 30°- (Quadrate), 60°- (Dreiecke) und des 90°-End-Stop-Detektors (Kreise) als Funktion des Orientierungswinkels  $\phi$  einer idealen Kante, die zusammen mit einer zweiten Kante einen rechten Winkel (Ecke) bildet. Der jeweilige Detektor befindet sich am Ort dieser Ecke.

Zur Bestimmung der Orientierungscharakteristiken wird eine ideale Kante (Grauwertdifferenz  $\Delta G=64$ ) verwendet, die durch eine zweite zu ihr im 90°-Winkel angeordnete ideale Kante begrenzt wird, so daß eine Ecke (rechter Winkel) entsteht. Die Kante wird so in das rezeptive Feld des End-Stop-Detektors gebracht, daß dieser bezüglich der Position der Ecke optimal antwortet. Der Vorgang wird für verschiedene Orientierungen der Kante wiederholt. Die resultierenden Kurven der ersten vier Orientierungen sind in Abbildung 4.2.4 dargestellt.

visualization of active end-stop neurons:



**Abb. 4.2.5:** Detektion von Linienenden bei Verarbeitung einer künstlichen Szene. Die Orte der innerhalb eines Zeitintervalls von etwa 50 BIN aktiven Neuronen der End-Stop-Schichten sind weiß gekennzeichnet. Jede Kante innerhalb des Bildes wird durch zwei Linienenden in den End-Stop-Schichten repräsentiert.

Sie zeigen in Folge des Winner-takes-all-Mechanismus der vorhergehenden Verarbeitungsstufe weitgehend eindeutiges Antwortverhalten. Verbleibende geringe Abweichungen (Mehrdeutigkeiten) könnten durch weitere Verschärfung des Winner-takes-all-Mechanismus vermieden werden, sind aber unter bestimmten Voraussetzungen unproblematisch für die nachfolgenden Verarbeitungsstufen. Im Gegensatz zur geforderten Eindeutigkeit in der Repräsentation der Kantenenden gegenüber Vorzugsorientierungen ist eine räumliche Eindeutigkeit dieser Repräsentationen nicht gegeben. Die weitere Verarbeitung ist jedoch diesbezüglich tolerant (siehe Kap. 4.3).

Im folgenden soll anhand des Beispiels der in Abbildung 4.1.24 B gezeigten künstlichen Szene der Vorgang der Detektion von Kanten- beziehungsweise Linienenden verdeutlicht werden. Abbildung 4.2.5 zeigt die in den Sehraum zurückprojizierten Orte, in denen Neuronen der End-Stop-Schichten innerhalb eines zeitlichen Intervalls von etwa 50 BIN aktiv werden. Die Schichten der Vorzugsorientierungen von  $30^\circ$ ,  $120^\circ$ ,  $210^\circ$  und  $330^\circ$  zeigen keine Aktivität und sind daher nicht explizit aufgeführt. Da jedes Kantenende bzgl. der Orientierung eindeutig detektiert wird, wird jede Kante innerhalb des Bildes durch zwei Kantenenden in den End-Stop-Schichten repräsentiert. Die den End-Stop-Detektoren zugrundeliegenden Parameter sind in Tabelle 4.2.1 aufgeführt.

Schwellen- verstärkung $V^\Theta$	Schwellen- zeitkonstante $\tau^\Theta$	Schwellen- offset $\Theta_0$	Feeding- potential- verstärkung $V^{F1}$	Feeding- potential- zeitkonstante $\tau^{F1}$	Feeding- potential- verstärkung $V^{F2}$	Feeding- potential- zeitkonstante $\tau^{F2}$
32	10 BIN	6	1.0	10 BIN	1.0	40 BIN

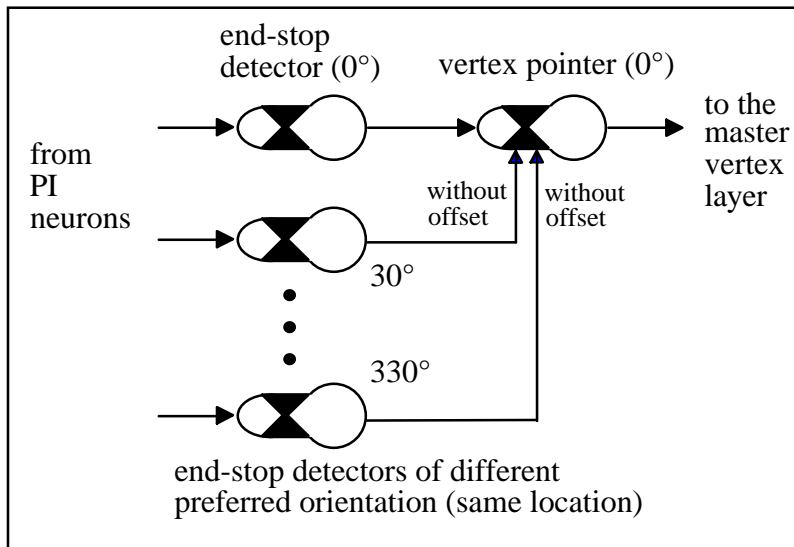
Tab. 4.2.1: Parameter des End-Stop-Detektorneurons.

### 4.3 Detektion von Vertizes

Wie bereits in Kapitel 2 erwähnt, sei ein Vertex hier definiert durch das Zusammentreffen von mindestens zwei Kantenenden an etwa demselben Ort, wobei T-Stücke nicht als Vertizes angesehen werden (Weitzel et al. 1996, 1997 a, b, Spengler 1996). T-Stücke weisen auf Objektverdeckungen hin und sind somit gesondert zu behandeln (z.B. Heitger et al. 1992).

#### 4.3.1 Repräsentationen von Vertizes

Zunächst werden zwei verschiedene Repräsentationen von Vertexinformation eingeführt: *Vertex-Zeiger-Schichten* und eine *Master-Vertex-Schicht*.



**Abb. 4.3.1:** Konvergente Verschaltung von End-Stop-Neuronen auf ein Vertex-Zeiger-Neuron. Die Antwort des 0°-End-Stop-Detektors (desselben Ortes wie das Vertex-Zeiger-Neuron) wird mit der Summe der Antworten der End-Stop-Neurone am selben Ort, aber verschiedener Orientierungspräferenz, multipliziert. Dazu werden Feeding- und Linking-Potential ohne Offset miteinander multipliziert. Deshalb wird das 0°-Vertex-Zeiger-Neuron bei Anwesenheit von mindestens zwei End-Stop-Neuronen (eines davon ist das 0°-End-Stop-Neuron) aktiv.

Ein Vertex-Zeiger-Neuron kodiert auf der Grundlage von Antworten der End-Stop-Detektoren Information über die Anwesenheit eines Vertex in seinem rezeptiven Feld. Darüber hinaus repräsentiert es einen gerichteten Teil (Zeiger) des Vertex, der einer der den Vertex bestimmenden Kante entspricht. Ein Vertex wird somit durch mehrere Vertex-Zeiger-Neuronen desselben Ortes vollständig kodiert. Diese Zeigerdarstellung ist der Konstruktion einer expliziten Repräsentation aller möglichen Vertizes in allen möglichen Zusammensetzungen und Orientierungen, die jede technische Anwendung durch übertriebene Netzgröße unmöglich machen würde, vorzuziehen.

Da die Detektion der Vertizes auf der Information über Kantenenden aufbaut, werden Zeiger-Neuronen durch konvergente Verbindungen von den End-Stop-Neuronen konstruiert. Das Vertex-Zeiger-Neuron übernimmt die Aufgabe einer neuronalen zeitlichen UND-Verschaltung. Dies wird durch Multiplikation von Feeding- mit Linkingpotential bei Vernachlässigung des Linkingpotential-Offsets von +1 erreicht. Auf den Feeding-Eingang des Zeiger-Neurons wird das Ausgangssignal des End-Stop-Detektors gegeben, dessen Vorzugsorientierung der Zeigerorientierung entspricht. Die Summe der Antworten der übrigen End-Stop-Neuronen desselben Ortes, aber unterschiedlicher Orientierungspräferenz, wird an den Linking-Eingang angelegt. Treffen an Feeding- und Linking-Eingang gleichzeitig Signale ein (Koinzidenzen), wird das Neuron überschwellig und generiert Aktionspotentiale. Um auch bei zeitlichem Versatz der Antworten von End-Stop-Detektoren noch einen Vertex detektieren zu können, muß mindestens eine der beiden Potentialzeitkonstanten relativ groß gewählt werden. Damit läßt sich die Erregung eines der Potentiale länger aufrechterhalten. Abbildung 4.3.1 zeigt die Konvergenz einiger End-Stop-Neuronen auf ein Vertex-Zeiger-Neuron, das die 0°-Orientierung kodiert.

Wie im vorherigen Unterkapitel (Kapitel 4.2) bereits deutlich wurde, werden Kantenenden meist durch mehrere End-Stop-Neuronen (derselben Vorzugsorientierung) repräsentiert. Insbesondere ist in echten Szenen mit gewissen räumlichen Abweichungen aufgrund weicher Kanten oder eventuell auftretender Bewegung zu rechnen. Ein strenger

räumlicher "Eins-zu-eins-Zusammenhang" zwischen End-Stop-Neuronen und Vertex-Zeiger-Neuronen ist somit nicht aufrechtzuerhalten. Daher konvergieren auch die räumlich angrenzenden End-Stop-Neuronen auf das Zeiger-Neuron. Die Stärke  ${}_{F}W_{qrop}^{\phi}$  der Feeding-Verbindung von einem End-Stop-Neuron der Orientierung  $\phi$  am Ort  $(o, p)$  auf ein Vertex-Zeiger-Neuron derselben Orientierung am Ort  $(q, r)$  wird beschrieben durch:

$${}_{F}W_{qrop}^{\phi} = \begin{cases} 1, & \text{falls } (o = q \pm 1 \wedge p = r \pm 2) \vee ((o = q \vee o = q \pm 2) \wedge p = r) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.3.1})$$

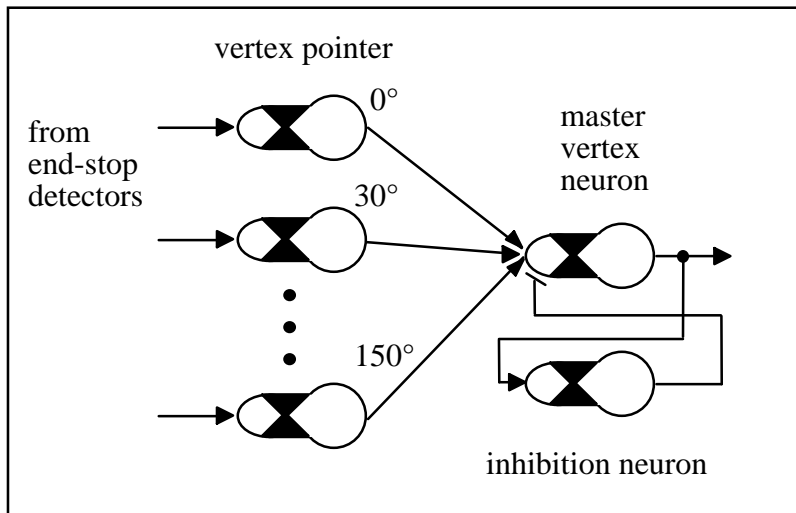
Sei weiterhin  $A = \{0^\circ, 30^\circ, 60^\circ, \dots, 300^\circ, 330^\circ\}$  die Menge aller verwendeten Vorzugsorientierungen. Dann ergibt sich die Stärke  ${}_{L}W_{qrop}^{\phi}$  der Linking-Verbindungen von einem End-Stop-Neuron der Orientierung  $\phi \in A$  am Ort  $(o, p)$  und einem Vertex-Zeiger-Neuron der Orientierung  $\alpha \in A$  am Ort  $(q, r)$  zu:

$${}_{L}W_{qrop}^{\alpha} = \begin{cases} {}_{F}W_{qrop}^{\phi}, & \text{falls } \phi \in A \setminus \alpha \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.3.2})$$

Bereits in Kapitel 4.2 wurde angedeutet, daß ein bezüglich der Vorzugsorientierung eindeutiges Antwortverhalten der End-Stop-Neuronen für die Detektion von Vertizes wichtig ist. Das hier vorgestellte Prinzip der Vertexdetektion würde bei der Repräsentation eines Kantenendes, beispielsweise mit einer Orientierung von  $15^\circ$ , durch End-Stop-Detektoren unterschiedlicher Orientierungspräferenz ( $0^\circ$  und  $30^\circ$ ) fälschlicherweise die Anwesenheit eines Vertex anzeigen, der sich auch aus dem Zusammenlaufen einer  $0^\circ$ - und einer  $30^\circ$ -Kante ergäbe. Üblicherweise kommen für solche Mehrdeutigkeiten nur bezüglich ihrer Orientierung benachbarte Schichten in Frage, da sich eine mögliche Überlappung in den Antwortkurven der End-Stop-Neuronen ebenso nur auf benachbarte Orientierungen erstreckt. Daher bietet sich die folgende Modifikation der Vertex-Zeiger-Verschaltung an. Auf den Linking-Eingang eines Zeiger-Neurons werden nur die End-Stop-Neuronen geschaltet, deren Orientierung um mehr als  $30^\circ$  von der Vorzugsorientierung des auf den Feeding-Eingang geschalteten End-Stop-Neurons abweicht. Diese Schaltung erlaubt mehrdeutiges Antwortverhalten der End-Stop-Detektoren, gestattet aber nicht mehr die Detektion sehr spitzer Winkel. Durch eine Erhöhung der Anzahl der diskreten Vorzugsorientierungen, d.h. mit entsprechender Vergrößerung des Netzes ließe sich diese "Detektionslücke" zumindest verkleinern. Für die weiteren Betrachtungen wird jedoch von einem eindeutigen Antwortverhalten der End-Stop-Neuronen ausgegangen. Entsprechend werden alle Winkel für die Vertex-Detektion zugelassen.

Aus Gründen der Netzwerkgröße werden darüberhinaus statt der beschriebenen zwölf nur sechs Vertex-Zeiger-Schichten verwendet, indem Zeiger mit entgegengesetzter Vorzugsorientierung innerhalb einer gemeinsamen Schicht bearbeitet werden. Die weitere Verarbeitung wird dadurch jedoch in keiner Weise beeinträchtigt.





**Abb. 4.3.2:** Verschaltung zwischen Vertex-Zeiger-Neuronen und einem Master-Vertex-Neuron desselben Ortes. Weiterhin wird auch die Eins-zu-eins-Verschaltung des Master-Vertex-Neurons zu seinem Inhibitionsneuron gezeigt, dessen Aufgabe es ist, das Master-Vertex-Neuron nach der Generierung einer Aktionspotentialfolge (Burst) für eine längere Zeit refraktär zu halten.

Den zweiten Typ zur Repräsentation von Vertizes liefern die Neuronen der Master-Vertex-Schicht (Abb. 4.3.2). Sie sind durch räumlich erweiterte Eins-zu-eins-Verbindungen mit den Vertex-Zeiger-Neuronen verschaltet und signalisieren die Anwesenheit von Vertizes in ihren rezeptiven Feldern. Informationen über die Zeiger der Vertizes enthält die Master-Vertex-Schicht hingegen nicht mehr. Die Gewichte  $^{Master}W_{stqr}^{\phi}$  zwischen einem Vertex-Zeiger-Neuron der Orientierung  $\phi \in A$  am Ort  $(q, r)$  und einem Master-Vertex-Neuron am Ort  $(s, t)$  lassen sich beschreiben durch:

$$^{Master}W_{stqr}^{\phi} = \begin{cases} 1, & \text{falls } (q = s \pm 1 \wedge r = t \pm 2) \vee ((q = s \vee q = s \pm 2) \wedge r = t) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.3.3})$$

Neuron	$V^{\theta}$	$\tau^{\theta}$	$\Theta_0$	$V^F$	$\tau^F$	$V^L$	$\tau^L$	$V^{Inh}$	$\tau^{Inh}$
Vertex-Zeiger	64	10 BIN	6	2	40 BIN	2	*)	-	-
Master-Vertex	2048	10 BIN	6	3	40 BIN	1	2 BIN	7.5	40 BIN
Vertex-Inhibition	32	1 BIN	2	7.5	9 BIN	-	-	-	-

**Tab. 4.3.1:** Parameter der in der Vertex-Detektion beteiligten Neuronen. \*) Ohne Offset +1

Für jedes Master-Vertex-Neuron gibt es darüberhinaus ein inhibitorisches Partnerneuron. Die Partnerneuronen sind in einer getrennten Inhibitionsschicht angeordnet und wechselseitig mit den Master-Vertex-Neuronen verschaltet. Aufgabe der Inhibitionsneuronen ist es, die Master-Vertex-Neuronen eine gewisse Zeit refraktär zu halten. Dies wird insbesondere durch eine relativ kurze Schwellen-Zeitkonstante realisiert, die gewährleistet, daß das Inhibitionsneuron sofort nach seiner Aktivierung mit einem kurzen Burst an Aktions-

potentialen antwortet und damit das Master-Vertex-Neuron nachhaltig inhibiert. Dieser Vorgang ist ein wichtiger Bestandteil der Dynamik, die die Separation von Konturzügen bewirkt und die in den nachfolgenden Abschnitten eingehend beschrieben und diskutiert wird. In Tabelle 4.3.1 sind die Parameter der hier beschriebenen Neuronen aufgeführt.

### 4.3.2 Konturüberschneidungen zweier Objekte (T-Stücke)

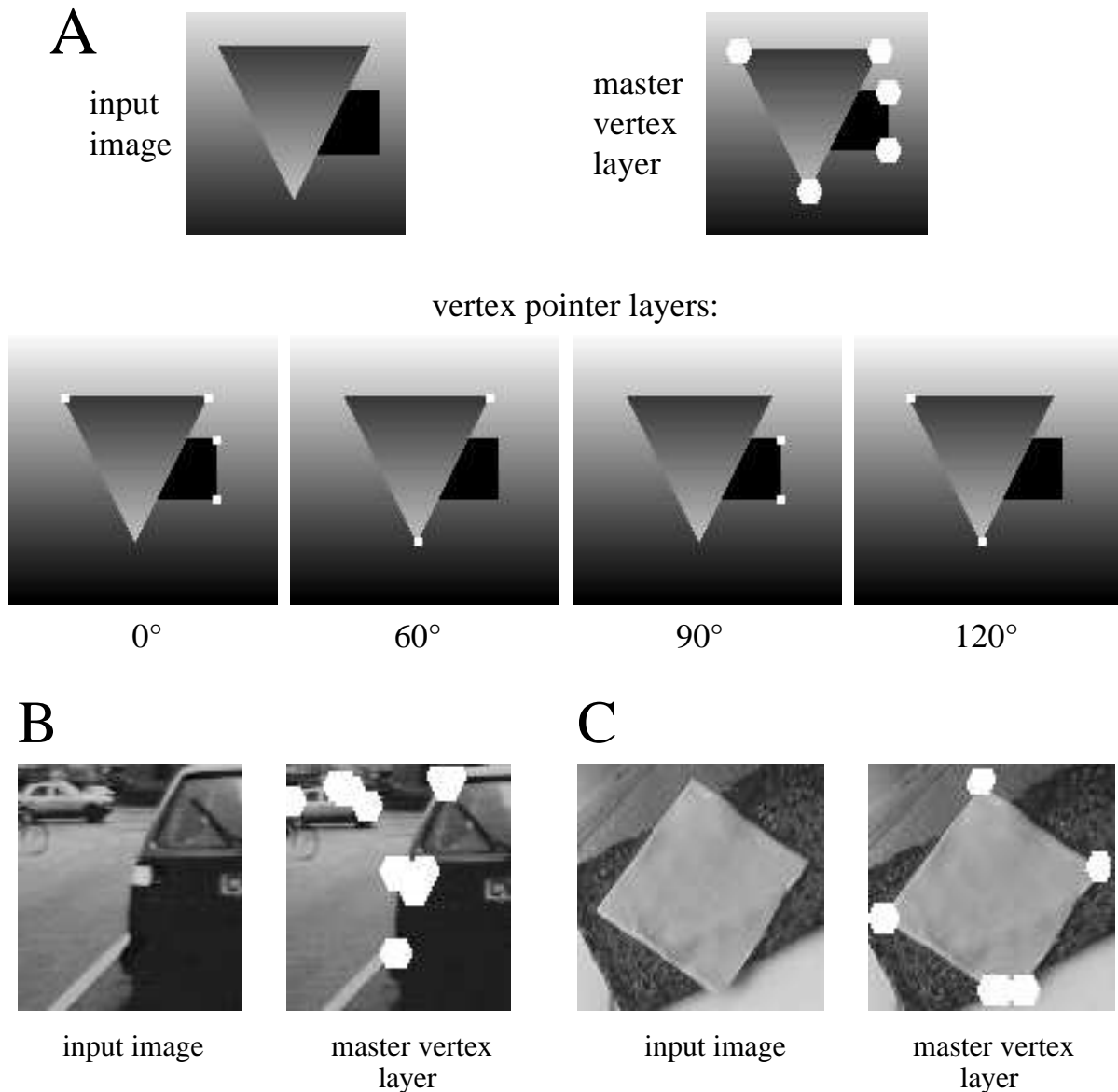
Wird an einem Ort nur ein Kantenende detektiert, so liegt dort offensichtlich keine Ecke und keine hohe Krümmung vor. Die Vertex-Zeiger-Neuronen benötigen Eingangssignale von mindestens zwei End-Stop-Detektoren, um aktiv werden zu können. Daher signalisieren sie hier nicht die Anwesenheit eines Vertex. In diesem Falle liegt ein T-Stück vor, das in der Regel auf eine Objektverdeckung hinweist.

Verschiedene Verfahren zur Extraktion von T-Stücken sind denkbar (z.B. Ostkamp 1996, Spengler 1996). Die weiteren Verarbeitungsstufen des vorliegenden Netzwerkes benötigen zwar die Information über Ecken und Orte hoher Krümmung zur Definition von Objekten, die Information über T-Stücke wird jedoch explizit nicht benötigt. Daher wird von einer weiteren Behandlung oder Extraktion der T-Stücke abgesehen. Auf diesen Zusammenhang wird auch in den nachfolgenden Abschnitten noch eingegangen werden.

### 4.3.3 Simulationen

Abbildung 4.3.3 illustriert anhand der aus Kapitel 4.1.4 bekannten Beispiele die Detektion von Vertices. Im ersten Beispiel (Abb. 4.3.3 A) wird die künstliche Szene zweier geometrischer Figuren als Eingangsbild verwendet. Weiterhin sind die Orte aktiver Vertex-Zeiger-Neuronen innerhalb der entsprechenden Schichten dargestellt. Beispielsweise sind in der  $0^\circ$ -Vertex-Zeiger-Schicht Neuronen genau dort aktiv, wo die Figuren des Bildes Ecken besitzen, die sich aus dem Zusammenlaufen einer  $0^\circ$ - oder  $180^\circ$ -Kante und einer weiteren Kante ergeben. Innerhalb der  $30^\circ$ - und der  $150^\circ$ -Schicht tritt in diesem Beispiel keine Aktivität auf. Daher sind diese Schichten hier nicht dargestellt. Jede Ecke führt zu neuronaler Aktivität in zwei Schichten gemäß der Orientierungen der zwei entsprechenden Kanten. Innerhalb der Master-Vertex-Schicht werden schließlich die Orte aller detektierten Ecken durch Bereiche aktiver Neuronen ("Blobs") repräsentiert. In Abbildung 4.3.3 A ist die entsprechende, in einem bestimmten Zeitintervall auftretende neuronale Aktivität der Master-Vertex-Schicht durch weiße Blobs im Eingangsbild dargestellt. Die T-Stücke, die hier zwischen Dreieck und dem scheinbar verdeckten Rechteck auftreten, haben keine Auswirkung auf die Vertex-Schichten.

Im nächsten Beispiel wird die Verkehrsszene als Stimulus verwendet. Abbildung 4.3.3 B zeigt neben der Szene auch die Aktivitätsverteilung innerhalb der Master-Vertex-Schicht.



**Abb. 4.3.3:** Beispiele für Vertexdetektion. **A:** künstliche Szene, **B:** Verkehrsszene, **C:** Frühstücksszene. Gezeigt werden die als Stimulus verwendeten Szenen und durch weiße "Blobs" die Orte neuronaler Aktivität der Master-Vertex-Schichten. Für die künstliche Szene sind die Vertex-Zeiger-Schichten mit den entsprechenden Orten der neuronalen Aktivität (weiß) im einzelnen dargestellt.

Sowohl Ecken der Hintergrundobjekte (Radfahrer, Auto) als auch des Autos im Vordergrund werden detektiert, wobei aber bereits deutlich wird, daß die Strukturen der Hintergrundobjekte für die verwendete Auflösung zu fein sind. Der Mittelstreifen liefert keine Ecken, da er vom PKW zum Teil verdeckt wird und dort somit nur T-Stücke auftreten.

In Abbildung 4.3.3 C ist der Stimulus des dritten Beispiels dargestellt. Als Eingangsbild wird ein Ausschnitt der Frühstücksszene verwendet. Die Neuronen der Master-Vertex-Schicht detektieren die vier Ecken der Käsescheiben, sprechen aber mit einer Ausnahme nicht auf die sich aus Überschneidung mit dem Vollkornbrot ergebenden T-Stücke an. Am unteren Bildrand wird an der Stelle eines T-Stückes fälschlicherweise ein Vertex detektiert, weil es sich in der unmittelbaren Nachbarschaft einer Ecke befindet und zudem ein Teil des T-Stückes nur sehr geringe Grauwertdifferenz aufweist, was die Detektion zusätzlich erschwert.

## 4.4 Zusammenbinden und Separieren von Konturen im Zeitbereich

### 4.4.1 Zusammenbinden von Kontursegmenten

Das Zusammenbinden und Trennen wird grundsätzlich durch dynamische Wechselwirkungen zwischen der Master-Vertex-Schicht und den *Synchronisationsschichten* erreicht. Die sechs Synchronisationsschichten sind im wesentlichen eine Kopie der sechs Schichten der PI-Neuronen, da sie über Eins-zu-eins-Gewichte miteinander verbunden sind. Die Stärke  ${}_{F}^{Syn}W_{uvmn}^{\varphi}$  der Feeding-Verbindung von einem PI-Neuron ( $m, n$ ) der Orientierung  $\varphi$  zu einem Synchronisationsneuron ( $u, v$ ) derselben Orientierung läßt sich beschreiben durch:

$${}_{F}^{Syn}W_{uvmn}^{\varphi} = \begin{cases} 1, & \text{falls } m = u \wedge n = v \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.4.1})$$

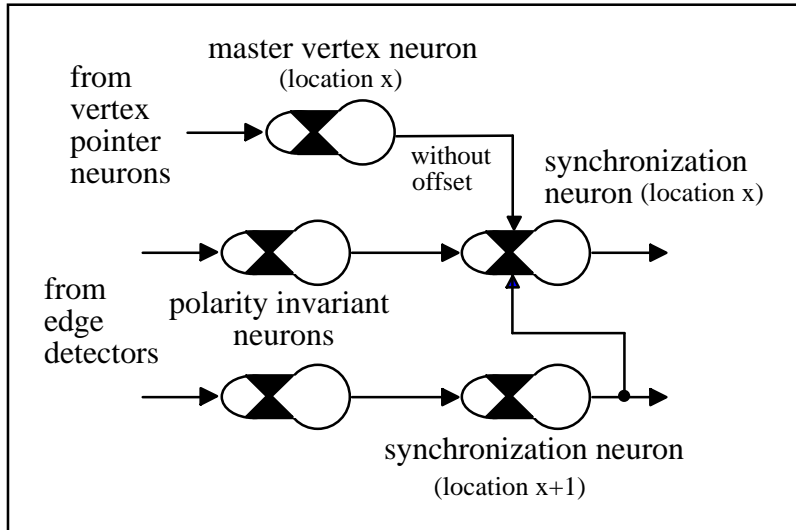
Die Neuronen der Synchronisationsschichten können nicht allein durch Eingangssignale von den PI-Neuronen aktiviert werden, aber sie werden durch die Kanteninformation voraktiviert und erst durch Signale aus der Master-Vertex-Schicht aktiviert, die an den entsprechenden Orten auftreten, wo Vertizes detektiert wurden. In den Synchronisationsschichten fungieren die Neuronen, ähnlich den Vertex-Zeiger-Neuronen, als zeitliche UND-Gatter, in dem der Linking-Eingang ohne einen Offset von +1 betrieben wird. In Abbildung 4.4.1 ist ein Teil der Verbindungen zwischen einem PI-Neuron, einem Synchronisationsneuron und einem Master-Vertex-Neuron desselben Ortes dargestellt. Das Gewicht  ${}_{L}^{M-Syn}W_{uvst}^{\varphi}$  der Verbindung von einem Mastervertexneuron ( $s, t$ ) auf den Linking-Eingang eines Synchronisationsneurons ( $u, v$ ) der Orientierung  $\varphi \in \{0^{\circ}, 30^{\circ}, \dots, 150^{\circ}\}$  ist gegeben durch:

$${}_{L}^{M-Syn}W_{uvst}^{\varphi} = \begin{cases} 15, & \text{falls } (s = u \pm 1, \pm 3 \wedge t = v \pm 2) \vee ((s = u \pm 2, \pm 4 \vee s = u) \wedge t = v) \\ & \vee ((s = u \pm 2 \vee s = u) \wedge t = v \pm 4) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.4.2})$$

Weiterhin sind die Synchronisationsneuronen mit in Vorzugsorientierung benachbarten Neuronen lateral gekoppelt. In Abbildung 4.4.1 ist nur eine der Verbindungen dargestellt. Die Stärke  ${}_{L}^{Syn-Syn}W_{abuv}^{\varphi}$  der Linking-Verbindungen von einem Neuron ( $u, v$ ) zu einem Neuron ( $a, b$ ) der Synchronisationsschicht mit Vorzugsorientierung  $\varphi$  ist gegeben durch:

$${}_{L}^{Syn-Syn}W_{abuv}^{\varphi} = \begin{cases} 4, & \text{falls } {}^{link}W_{ijkl}^{\varphi} > 0 \quad \text{mit } a = i, b = j, k = u, v = l \\ 0, & \text{sonst} \end{cases}, \quad (\text{Gl. 4.4.3})$$

wobei  ${}^{link}W_{ijkl}^{\varphi}$  aus den Gleichungen 4.1.22 hervorgeht und  $\varphi \in \{0^{\circ}, 30^{\circ}, \dots, 150^{\circ}\}$ .



**Abb. 4.4.1:** Verschaltungen zu einem Synchronisationsneuron. Ein PI-Neuron (polarity invariant neuron) ist mit dem Synchronisationsneuron desselben Ortes gekoppelt. Das Synchronisationsneuron kann aber nur dann aktiv werden, wenn neben dem Signal vom PI-Neuron auch ein Signal vom Master-Vertex-Neuron desselben Ortes und/oder von einem in Vorzugsorientierung räumlich benachbarten Synchronisationsneuron empfangen wird.

Ein durch das PI-Neuron voraktiviertes Synchronisationsneuron kann somit auch durch Aktivität aus der räumlichen Nachbarschaft aktiviert werden. Dadurch kann diese Aktivität innerhalb der Synchronisationsschichten in die beiden Richtungen der jeweiligen Vorzugsorientierung fortschreiten, wenn ein Synchronisationsneuron durch einen Vertex aktiviert wurde. Das ist aber nur dort möglich, wo die Synchronisationsneuronen durch Signale von den PI-Neuronen vorsensitiviert wurden, denn diese zeigen die Position der Kanten an. Daher breitet sich die Aktivität entlang der detektierten Kanten aus, was im folgenden als *Linking-Welle* bezeichnet wird.

Weiterhin erhalten auch die Synchronisationsschichten jeweils eine zusätzliche Inhibitionsschicht, die für Refraktärzeiten bestimmter Länge nach der Generierung von Aktionspotentialen innerhalb der entsprechenden Synchronisationsschicht zuständig ist. Jedoch werden hier im Gegensatz zur Verarbeitungsstufe der Kantendetektorschichten sechs getrennte Inhibitionsschichten verwendet, um einen Winner-takes-all-Mechanismus an dieser Stelle zu vermeiden. Grundsätzlich ist auch eine Verschaltung ohne Inhibitionsschichten vorstellbar, da durch geeignet große Schwellenverstärkungen und Schwellenzeitkonstanten jedes Synchronisationsneuron für sich längere Refraktärzeiten einhalten könnte.

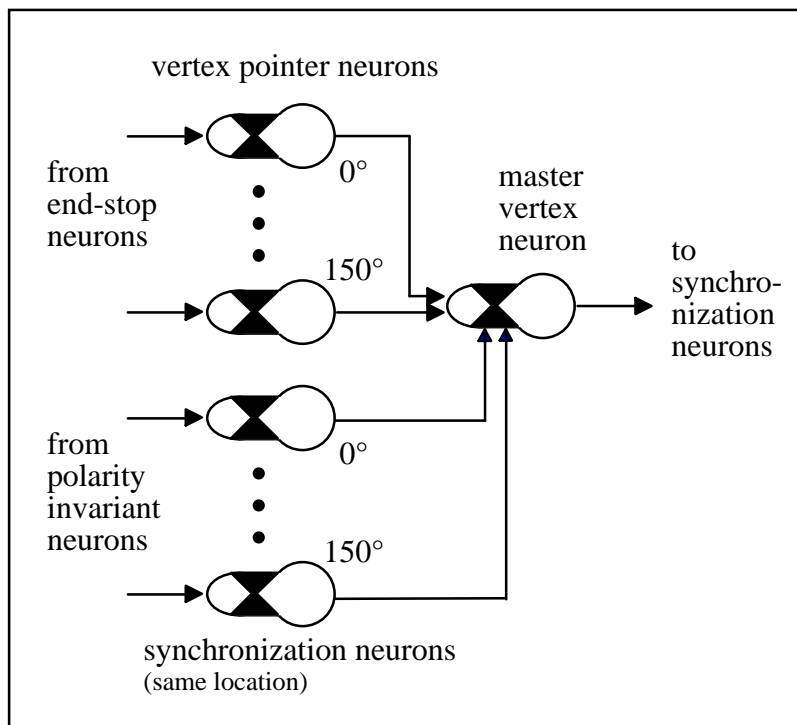
Für die angestrebte lokale Synchronisation ist es aber nützlich, die Inhibition benachbarter Neuronen anzunähern, was durch Ausdehnung der Eins-zu-eins-Verschaltungen zwischen Synchronisationsneuronen und Inhibitionsneuronen ermöglicht wird. Daher ergibt sich das Gewicht  ${}_{F}^{Syn-I}W_{xyuv}^{\phi}$  der Verbindung von einem Synchronisationsneuron der Orientierung  $\phi$  zum Feeding-Eingang eines Inhibitionsneurons  $(x, y)$  derselben Orientierung zu:

$${}_{F}^{Syn-I}W_{xyuv}^{\phi} = \begin{cases} 1, & \text{falls } (u = x \pm 1 \wedge v = y \pm 2) \vee ((u = x \vee u = x \pm 2) \wedge v = y) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.4.4})$$

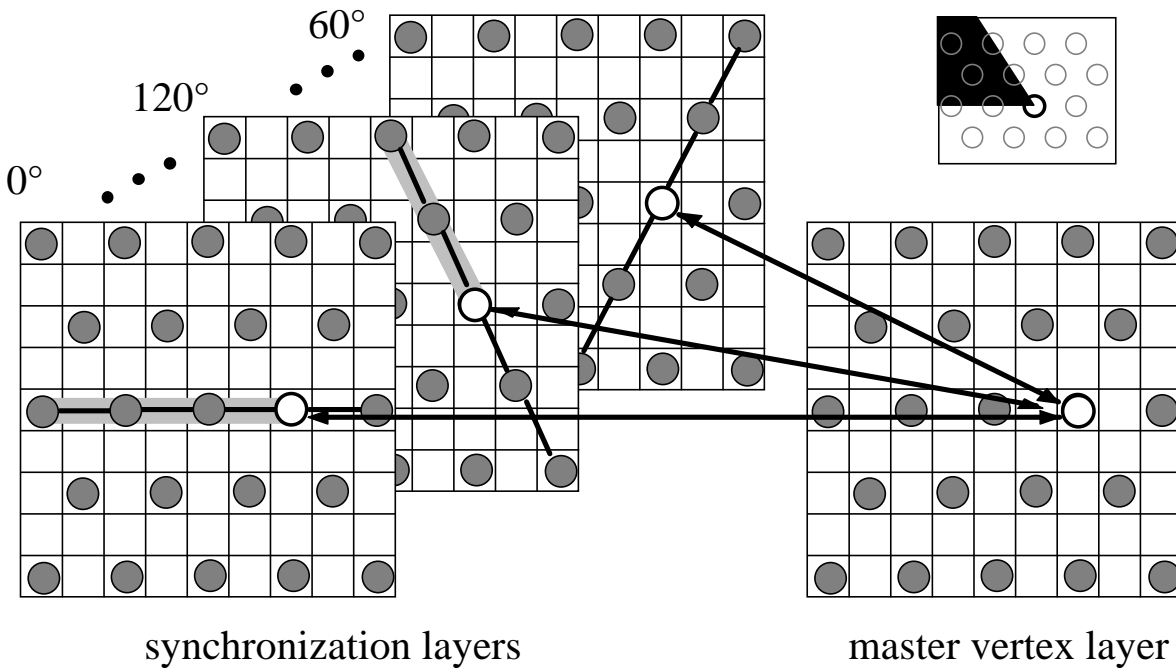
Analog ergibt sich die Stärke  ${}_{Inh}^{I-Syn} W_{uvxy}^{\phi}$  der Verbindung von einem Inhibitionsneuron  $(x, y)$  der Orientierung  $\phi$  zum Inhibitions-Eingang des Synchronisationsneurons  $(u, v)$  derselben Orientierung zu:

$${}_{Inh}^{I-Syn} W_{uvxy}^{\phi} = \begin{cases} 1, & \text{falls } (x = u \pm 1 \wedge y = v \pm 2) \vee ((x = u \vee x = u \pm 2) \wedge y = v) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.4.5})$$

Eine Linking-Welle endet zunächst, wenn an einem bestimmten Ort eine detektierte Kante endet oder eine detektierte Kante mit einer anderen Kante über einen Vertex (Ecke) verbunden ist. Im zweiten Fall muß die Linking-Welle um die Ecke herumgelenkt werden, um dann entlang der zweiten die Ecke definierenden Kante weiterzulaufen. Dieses *Linking um Ecken* wird durch Linking-Rückkopplungen von den Synchronisationsneuronen zu den Master-Vertex-Neuronen ermöglicht, die in Abbildung 4.4.2 für ein einzelnes Master-Vertex-Neuron skizziert sind.



**Abb. 4.4.2:** Verschaltungen auf ein Master-Vertex-Neuron. Dargestellt sind die Synchronisationsneuronen unterschiedlicher Vorzugsorientierung, aber gleichen Ortes, deren Ausgangssignale auf den Linking-Eingang des Master-Vertex-Neurons desselben Ortes gegeben werden. Diese Signale aktivieren das Master-Vertex-Neuron, wenn es zuvor durch Aktivität der Vertex-Zeiger-Neuronen über den Feeding-Eingang voraktiviert wurde, d.h wenn an diesem Ort ein Vertex vorliegt.



**Abb. 4.4.3:** Linking-Verbindungen zwischen den Synchronisationsschichten und der Master-Vertex-Schicht für zeitliches Zusammenbinden von Konturen durch Linking-Wellen. Dargestellt sind die im hexagonalen Abtastraster angeordneten Neuronen (**graue Kreise**) der Synchronisationsschichten (**links**) und der Master-Vertex-Schicht (**rechts**), wobei die Neuronen der Synchronisationsschichten jeweils zu den in Vorzugsorientierung benachbarten koaxial angeordneten (eigentlich: in Keulenform angeordneten) Neuronen über Linking-Verbindungen (**starke schwarze Linien**) gekoppelt sind. Weiterhin wird eine Ecke eines gegebenen schwarzen Objekts auf weißem Grund gezeigt (**rechts oben**), wobei auch die Lage der Ecke bezüglich des Abtastrasters durch gestrichelte offene Kreise (Lage der Neuronen) angedeutet wird. Die Position des an der Spitze der Ecke befindlichen Neurons ist mit einem schwarz umrandeten offenen Kreis markiert. Auch in den dargestellten Schichten wird diese Position jeweils durch einen **weißen Kreis mit schwarzem Rand** besonders gekennzeichnet. Die entlang der Objektkanten sitzenden Neuronen der Synchronisationsschichten werden bei geeigneter Vorzugsorientierung voraktiviert, was durch **breite graue Balken** verdeutlicht wird. In der Master-Vertex-Schicht wird das am Ort der Ecke gelegene Neuron aktiv. Folglich ruft dieses Neuron Aktivität von Synchronisationsneuronen via Linking-Verbindungen hervor. Durch die lateralen Linking-Verbindungen der Synchronisationsneuronen untereinander kann sich die hervorgerufene Aktivität entlang neuronaler Ketten (voraktivierte Neuronen), die gerade Kontursegmente repräsentieren, ausbreiten. Diese Ausbreitung von Aktivität wird *Linking-Welle* genannt. Es werde nun angenommen, daß sich eine Linking-Welle auf eine Ecke hin ausbreitet. Wenn das durch diese Linking-Welle repräsentierte Kontursegment (bestimmter Vorzugsorientierung) den Ort des detektierten Vertex erreicht, ruft das aktivierte Synchronisationsneuron am Ort der Ecke Aktivität in dem vorsensitvierten Vertex-Detektor über die Linking-Verbindung hervor. Das so aktivierte Master-Vertex-Neuron aktiviert nun seinerseits die voraktivierten Synchronisationsneuronen der übrigen Vorzugsorientierungen an diesem Ort. Auf diese Weise wird die Linking-Welle um die Ecke herumgelenkt und breitet sich weiter entlang nachfolgender Abschnitte der Kontur, d. h. in einer Synchronisationsschicht anderer Vorzugsorientierung, aus.

Die entsprechende Stärke  ${}_{L}^{Syn-M}W_{stuv}^{\varphi}$  der Verbindung von einem Synchronisationsneuron  $(u, v)$  der Orientierung  $\varphi \in \{0^{\circ}, 30^{\circ}, \dots, 150^{\circ}\}$  zu einem Master-Vertex-Neuron  $(s, t)$  wird beschrieben durch:

$${}_{L}^{Syn-M}W_{stuv}^{\varphi} = \begin{cases} 7.5, & \text{falls } (u = s \pm 1 \wedge v = t \pm 2) \vee ((u = s \vee u = s \pm 2) \wedge v = t) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 4.4.6})$$

Erreicht eine Linkingwelle einen Ort, an dem ein Vertex von den Vertex-Zeiger-Neuronen detektiert wurde, wird das entsprechende voraktivierte Master-Vertex-Neuron aktiviert. Das Master Vertex-Neuron erhält die dazu notwendige Voraktivierung von den Vertex-Zeiger-Neuronen und befindet sich nicht in einer durch frühere Aktivierung hervorgerufenen Refraktärperiode. Das so aktivierte Master-Vertex-Neuron induziert nun seinerseits, wie oben beschrieben, Aktivität in den voraktivierten Neuronen der Synchronisationsschichten, und die Linking-Welle wird um die dem Vertex entsprechende Ecke umgeleitet. Linking-Wellen umlaufen somit durch die eingeführten Kopplungen eine Kontur, die aus mehreren über Vertizes verbundenen Liniensegmenten besteht. Die Zeitspanne, innerhalb der dies geschieht, zeigt die Zusammengehörigkeit der detektierten Gruppe von Kontursegmenten (Konturzug) an. Linking-Wellen werden jedoch nicht an T-Stücken umgelenkt, weil diese nicht als Vertizes detektiert werden. Üblicherweise breitet sich die von einem Vertex initiierte Linking-Welle in zwei Richtungen gleichzeitig aus. Abbildung 4.4.3 zeigt ein Beispiel zur Verdeutlichung der beschriebenen Dynamik. Wie auch aus dem Beispiel hervorgeht, kann eine Linking-Welle nur von den Neuronen der Master-Vertex-Schicht, nicht aber von den Synchronisationsneuronen initiiert werden, da nur die Master-Vertex-Detektoren modulatorische Linking-Verbindungen besitzen, die auch eine Aktivierung über den Feeding-Eingang erlauben. Durch eine große Feeding-Zeitkonstante bei einer gleichzeitig kleinen Feeding-Verstärkung wird jedoch dafür gesorgt, daß eine solche Aktivierung über den Feeding-Eingang erschwert und in der Regel erst bei zusätzlichen Linking-Signalen ermöglicht wird. Dies spielt eine wichtige Rolle bei der zeitlichen Trennung von Linking-Wellen, die möglicherweise die Konturen unterschiedlicher Objekte repräsentieren. Die Synchronisationsneuronen besitzen hingegen nur rein multiplikative Linking-Verbindungen und stellen daher neuronale zeitliche UND-Gatter dar, wodurch sie nicht durch ihren Feeding-Eingang allein aktiviert werden können und folglich auch keine Linking-Wellen einleiten können.

#### 4.4.2 Separation von Konturen unterschiedlicher Objektkandidaten im Zeitbereich

Linking-Wellen, die wechselseitig nicht miteinander verbundene Konturen repräsentieren oder Konturen, die über T-Stücke miteinander verbunden sind, weil sich die entsprechenden Objekte (scheinbar) verdecken, werden nun im Zeitbereich separiert. Diese Separation wird durch die Einführung einer globalen inhibitorischen Rückkopplung von den Synchronisationsschichten zur Master-Vertex-Schicht erreicht. Entsprechend wird ein inhibitorisches Neuron eingeführt, das Signale von allen Synchronisationsneuronen empfängt und die daraus resultierende eigene Ausgangsaktivität an alle Master-Vertex-Neuronen austellt. Die Stärke  ${}_{F}^{Syn-globI}W_{uv}^{\varphi}$  der Verbindung von einem Synchronisationsneuron ( $u, v$ ) der Orientierung  $\varphi \in \{0^{\circ}, 30^{\circ}, \dots, 150^{\circ}\}$  zum Feeding-Eingang des globalen Inhibitionsneurons beträgt 1, und die Stärke  ${}_{Inh}^{globI-M}W_{st}$  einer Verbindung vom globalen Inhibitionsneuron zum Inhibitions-



Eingang eines Master-Vertex-Neurons am Ort  $(s, t)$  beträgt  $1/3$ . Die Inhibition verhindert, daß ein Master-Vertex-Neuron, das über seinen Feeding-Eingang voraktiviert wurde, bereits ohne ein Linking-Signal aktiv werden kann. Dadurch unterdrückt die Aktivität einer Linking-Welle die zeitgleiche Entstehung einer weiteren Linking-Welle, die ja nur durch ein Master-Vertex-Neuron initiiert werden kann. Auf der anderen Seite verhindert die Inhibition nicht die Aktivierung eines Master-Vertex-Neurons, wenn es zum einen über seinen Feeding-Eingang voraktiviert wurde und zum anderen über Linking-Verbindungen von einer Linking-Welle erreicht wird, so daß diese Linking-Welle auch um Ecken ungehindert weiterlaufen kann. Folglich wird innerhalb eines zeitlichen Fensters nur eine Linking-Welle zugelassen, so daß verschiedene Linking-Wellen, die wahrscheinlich Konturen verschiedener Objekte oder Objektkandidaten repräsentieren, zeitlich getrennt werden.

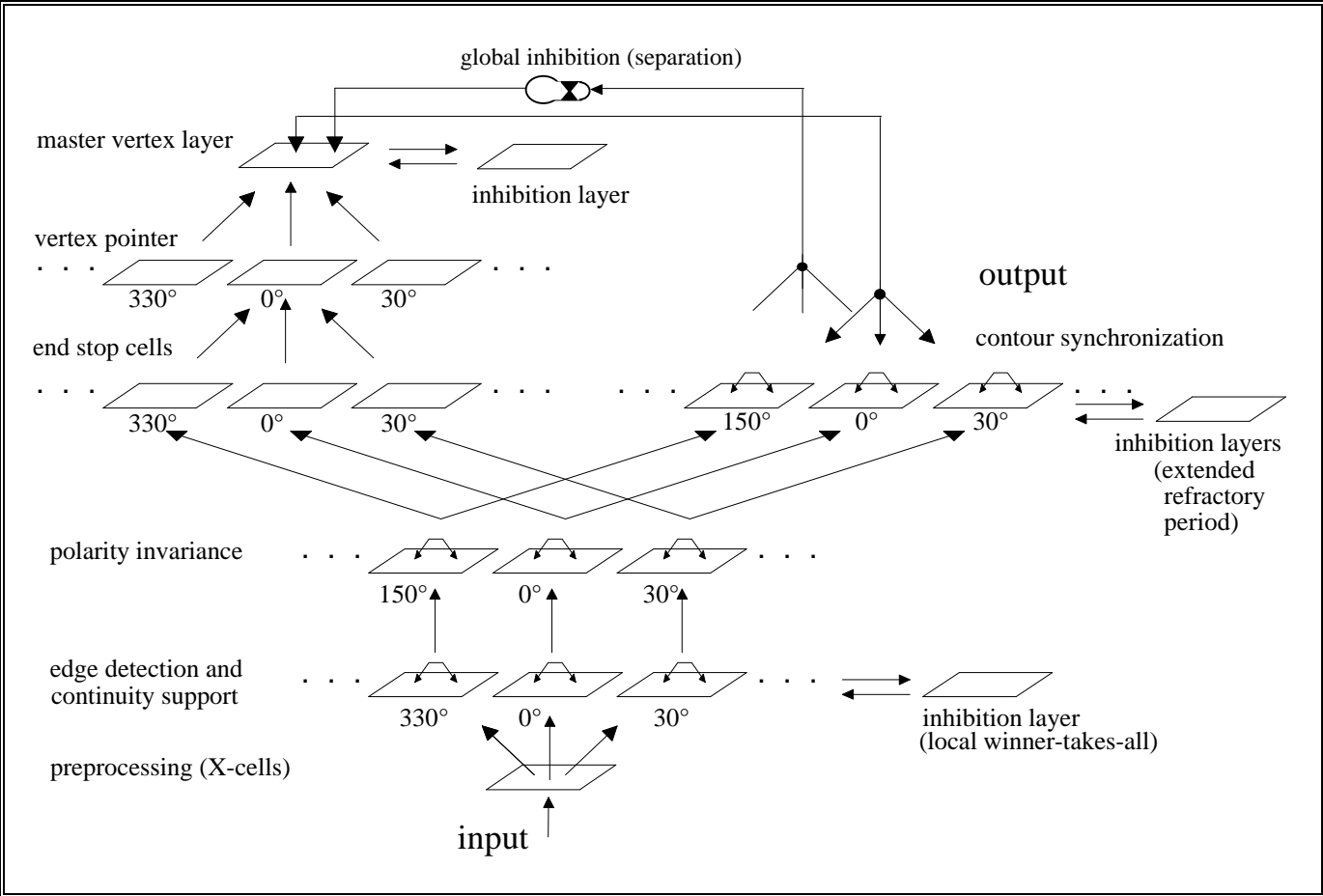
Die in den Simulationsbeispielen verwendeten Parameter der hier behandelten Neuronen sind in Tabelle 4.4.1 aufgeführt. Eine Übersicht über das gesamte Netzwerk ist in Abbildung 4.4.4 gegeben.

Neuron	$V^\Theta$	$\tau^\Theta$	$\Theta_0$	$V^F$	$\tau^F$	$V^L$	$\tau^L$	$V^{\text{Inh}}$	$\tau^{\text{Inh}}$
Synchr.-neuron	256	10 BIN	6	1	40 BIN	1	*) 1 BIN	2	40 BIN
Synchr.-Inhibition	32	1 BIN	2	7	20 BIN	-	-	-	-
Global-Inhibition	32	1 BIN	2	1	1 BIN	-	-	-	-

**Tab. 4.4.1:** Parameter der bei dem Zusammenbinden und Separieren von Kontur beteiligten Neuronen.

\*) ohne Offset

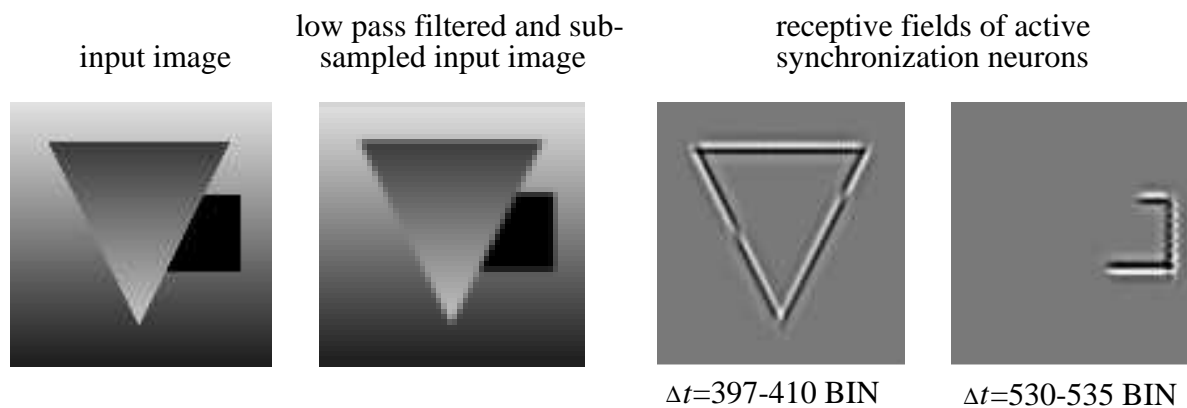
**Abb. 4.4.4** (nachfolgende Seite): *Architektur des Netzwerks.* Dargestellt sind die Schichten der einzelnen Verarbeitungsstufen. Weiterhin werden die Feeding-, Linking- und inhibitorischen Verbindungen durch Pfeile angedeutet. Den Ausgangssignalen des Netzes entsprechen die neuronalen Aktivitäten der Synchronisationsschichten.



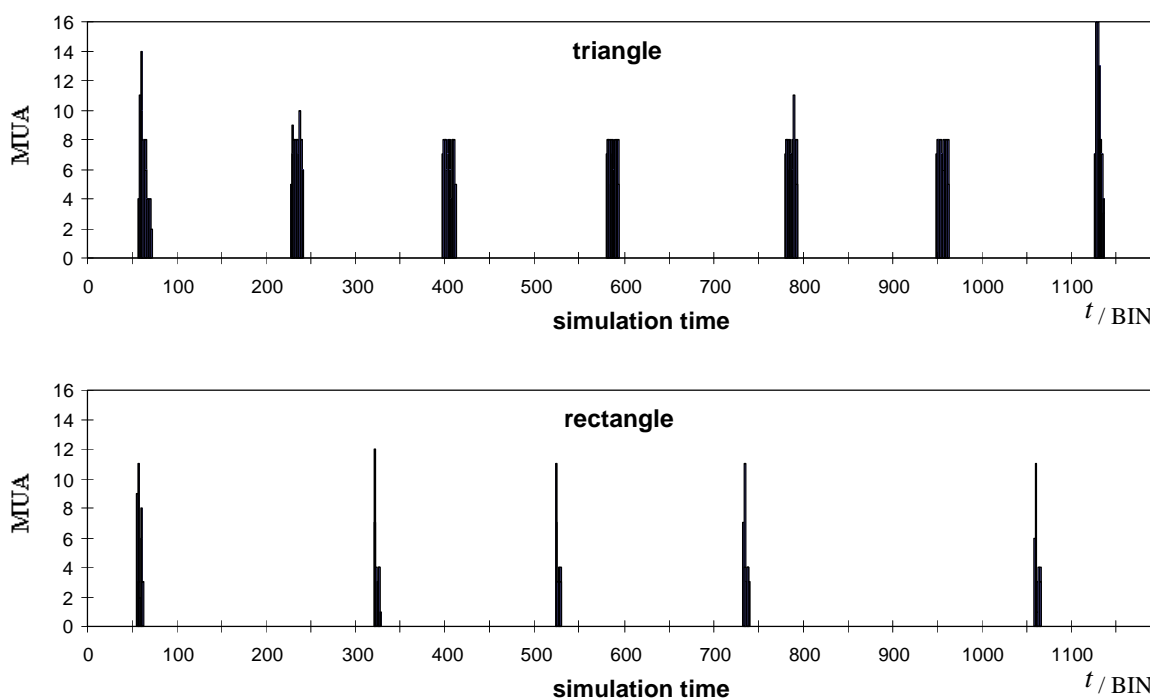
### 4.4.3 Simulationsergebnisse

#### Statische Szenen

In diesem Unterabschnitt sind einige Simulationsbeispiele aufgeführt. Das erste Beispiel ist das bekannte künstliche Bild (Abbildung 4.4.5 links), das bereits zur Illustration der Detektion von Linienenden und Vertizes verwendet wurde (vgl. Abb. 4.2.5 und 4.3.3 A). Das Bild enthält ein Dreieck, das durch einen variablen Grauwertgradienten definiert ist und das eine möglicherweise rechteckige Gestalt zu überdecken scheint. Vom Standpunkt der *Gestaltprinzipien* (geschlossene Form, Wertheimer 1923) gesehen, läßt sich das Dreieck durch Konturen, die über Ecken verbunden werden, definieren. Der Eindruck der Verdeckung wird durch zwei T-Stücke hervorgerufen (siehe dazu auch Kapitel 2.6). Menschliche Betrachter nehmen das Dreieck problemlos als solches wahr, obwohl der Kontrast längs der Kontur variiert und in der Mitte der linken Seite sogar verschwindet. Das Netz erreicht nach etwa 250 Zeitschritten einen stabilen Zustand, in dem alle Konturelemente in zwei Objekte gruppiert worden sind. Die über ein bestimmtes Zeitintervall aufintegrierten und in den Sehraum zurückprojizierten Aktionspotentiale aller Synchronisationsneuronen sind ebenso in Abbildung 4.4.5 dargestellt. Die Rückprojektionen enthalten auch die Information der Kontrastpolarität, die in den Synchronisationsschichten nicht mehr vorhanden ist. Um dies zu erreichen, greift das verwendete Verfahren auch auf die in den Kantendetektorschichten kodierte Information zurück. In Abbildung 4.4.6 ist die Summe aller zu einem Zeitschritt aktiven Synchronisationsneuronen (MUA, multiple unit activity) für die beiden Objekte aufgetragen. Wie daraus ersichtlich wird, werden die Repräsentationen der Objektkonturen deutlich in der Zeit separiert. Der zeitliche Abstand zwischen zwei Repräsentationszeitintervallen des Dreiecks ist sogar nahezu konstant 180 BIN im beobachteten Zeitraum, so daß hier von Oszillationen gesprochen werden kann. Bei den zeitlichen Repräsentationen des Rechtecks läßt sich jedoch keine feste Wiederholfrequenz finden. Der Grund dürfte in den unterschiedlich starken Eingangsaktivitäten der Neuronen, die verschiedene Objekte repräsentieren, zu suchen sein. Da erst Vertexdetektoren die Linkingwellen induzieren, ist insbesondere die Grauwertdifferenz an Ecken von Bedeutung. Entsprechend ergibt sich für jedes Objekt eine eigene Repräsentationsvorzugsfrequenz. Bei Nichtübereinstimmung der Vorzugsfrequenzen unterschiedlicher Objektrepräsentationen kommt es zwangsläufig zu Überschneidungen der Repräsentationszeitintervalle. Bedingt durch den Inhibitionsmechanismus wird dann eine der beiden Objektrepräsentationen kurzzeitig unterdrückt und damit zeitlich verschoben. Das führt insgesamt zu Aktivitäten, bei denen Angaben von festen Oszillationsfrequenzen wenig sinnvoll sind und die dennoch eine eindeutige zeitliche Separation der Repräsentationen von Objektkonturen bewirken.



**Abb. 4.4.5:** *Simulationsbeispiel: Objektverdeckung.* Links ist das als Stimulus verwendete Bild dargestellt: Ein durch einen variablen Grauwertgradienten definiertes Dreieck scheint ein schwarzes Rechteck zu überdecken. Um die vom Netz zu bewältigende Aufgabenstellung zu verdeutlichen, wird daneben das mit der in der Vorverarbeitung verwendeten Gaußmaske gefilterte und anschließend (aus Darstellungsgründen kartesisch) mit Abtastweite  $d=2$  unterabgetastete Eingangsbild gezeigt. In den beiden rechten Abbildungen sind die rezeptiven Felder der aktiven Synchronisationsneuronen für verschiedene Repräsentationszeitintervalle dargestellt, woraus die erfolgreiche zeitliche Gruppierung aller Konturelemente in zwei Objektrepräsentationen hervorgeht.

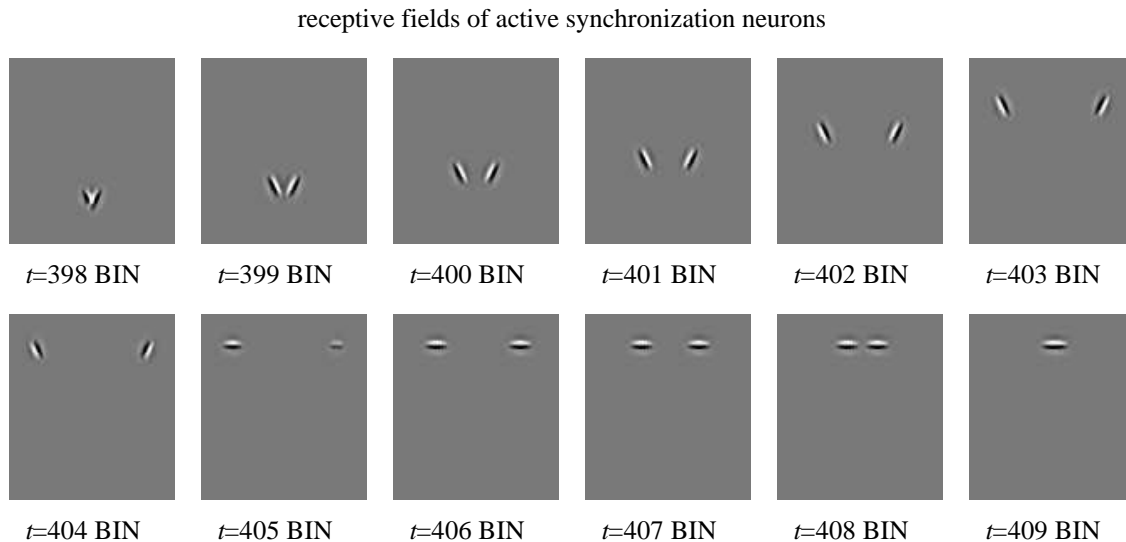


**Abb 4.4.6:** *Objektseparation im Zeitbereich.* Aufgetragen ist die Summe der Aktionspotentiale der das jeweilige Objekt repräsentierenden Synchronisationsneuronen in Abhängigkeit von der Zeit. Der Vergleich der Netzwerkaktivitäten bei Repräsentation des Dreiecks (**oben**) mit denen bei Repräsentation des Rechtecks (**unten**) zeigt, daß die Objekte deutlich im Zeitbereich getrennt werden.

Innerhalb eines Repräsentationszeitintervalls breiten sich Linkingwellen längs der Objektkonturen aus. In Abbildung 4.4.7 ist die innerhalb eines bestimmten Zeitintervalls das Dreieck repräsentierende Linkingwelle dargestellt. Für jeden Zeitschritt werden die rezeptiven Felder der aktiven Synchronisationsneuronen gezeigt.

Als zweites Beispiel wird das Bild eines Kekses (Abb. 4.4.8) verwendet. Obwohl auch auf der Keksoberfläche einzelne Kanten auftreten, liefert das Netz nur die (unregelmäßigen)

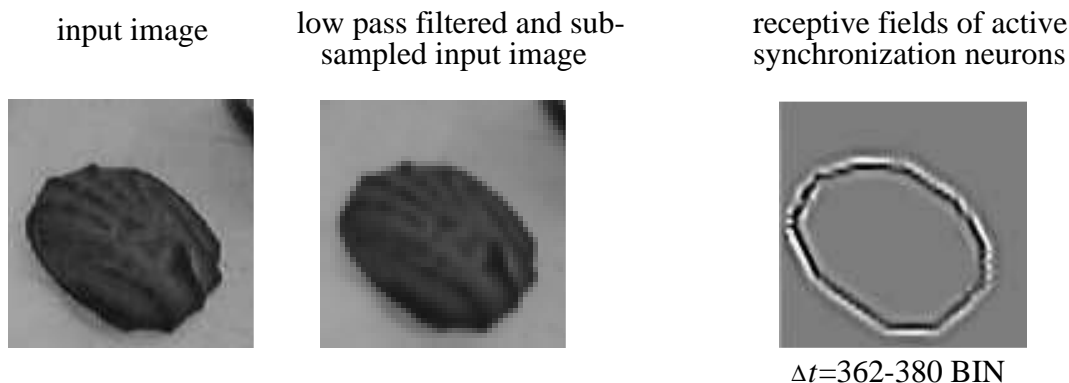
Konturen, d.h. die äußere Umrandung, des Objekts. Dies entspricht den Modellannahmen, da die Oberflächenlinien mit den äußeren Konturen nur T-Stücke und untereinander keine Ecken bilden.



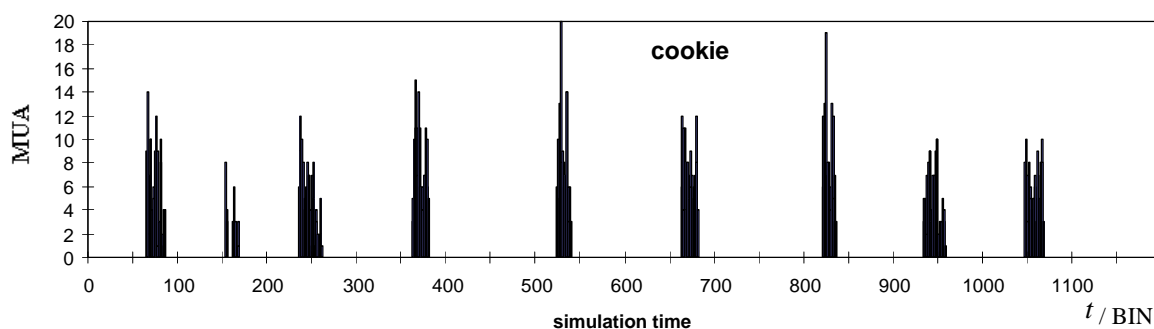
**Abb. 4.4.7:** *Ausbreitung einer Linkingwelle entlang der Konturen eines Dreiecks.* Dargestellt sind die rezeptiven Felder aller aktiven Neuronen der Synchronisationsschichten für jeden Zeitschritt eines bestimmten Zeitintervalls. Die Linkingwellen werden vom Vertexdetektor der unteren Ecke induziert und breiten sich längs der beiden Schenkel des Dreiecks mit einer Geschwindigkeit von etwa 3 Neuronen/BIN aus. Bei Erreichen der beiden oberen Ecken werden die entsprechenden Vertexdetektoren aktiviert. Diese aktivieren ihrerseits die 0°-Synchronisationsneuronen, so daß die Linkingwellen umgelenkt werden und entlang der oberen Seite des Dreiecks aufeinanderzulaufen und schließlich terminieren.

Linkingwellen lassen sich auch um runde Formen lenken, was durch die keulenförmige Anordnung der Linkingverbindungen in den verschiedenen Schichten und die Aktivitätsübertragung über Ecken ermöglicht wird. Ein rundes Objekt wird als Vieleck behandelt. Durch die entsprechenden Interaktionen zwischen der Master-Vertex-Schicht und den Synchronisationsschichten kann es zu einer geringen Verzögerung der Ausbreitung der entsprechenden Linkingwellen kommen. Auch die Größe eines Objekts relativ zur Bildgröße beeinflusst die Repräsentationsdauer. Beide Faktoren spielen im vorliegenden Beispiel eine Rolle und führen zu vergleichsweise längeren Repräsentations-Zeitintervallen (Abb. 4.4.9).

Abbildung 4.4.10 zeigt die bekannte Frühstücksszene und die rezeptiven Felder der in einem bestimmten Zeitintervall aktiven Synchronisationsneuronen. Eine Linkingwelle repräsentiert die Konturen einer Käsescheibe. Aufgrund der unmittelbaren räumlichen Nähe eines T-Stücks zu dem der unteren Ecke entsprechenden Vertex (vgl. Abb. 4.3.3 C) wird auch ein Konturstück der Vollkornbrotsscheibe von der Linkingwelle durchlaufen. Dieses Auflösungsproblem spielt in der nächsten Szene (Abb. 4.4.11 und 4.4.12) keine Rolle. Die "abgebissene" Käsescheibe wird eindeutig repräsentiert. Das Vollkornbrot findet keine Berücksichtigung, da es im bearbeiteten Bildausschnitt keine Vertizes besitzt und diese die Repräsentationen durch Linkingwellen erst induzieren.



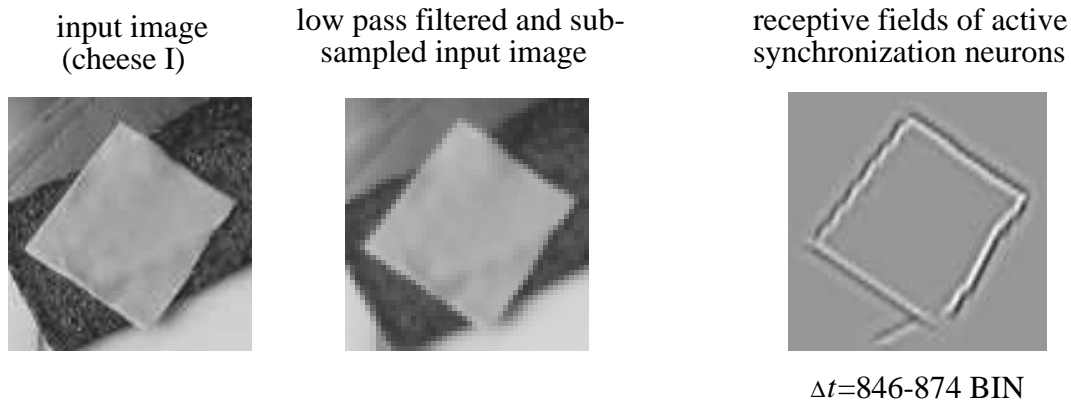
**Abb. 4.4.8:** *Simulationsbeispiel: Objekt mit unregelmäßigen Konturen.* **Links:** Darstellung des als Stimulus verwendeten Bildes eines Kekses. Um die vom Netz zu bewältigende Aufgabenstellung zu verdeutlichen, wird daneben das mit der in der Vorverarbeitung verwendeten Gaußmaske gefilterte und anschließend (aus Darstellungsgründen kartesisch) mit Abtastweite  $d=2$  unterabgetastete Eingangsbild gezeigt. **Rechts:** Innerhalb bestimmter Zeitintervalle umläuft eine Linkingwelle die unregelmäßigen Konturen des Kekses. Dargestellt sind die rezeptiven Felder der entsprechenden aktiven Synchronisationsneuronen.



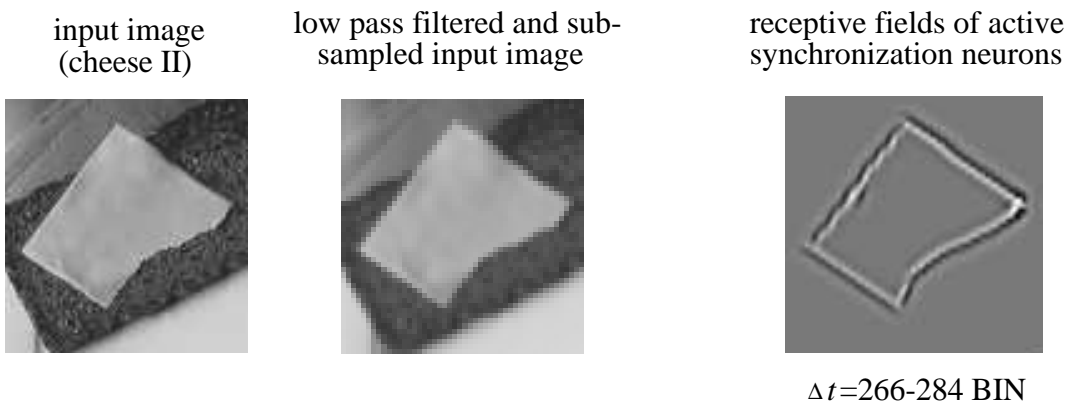
**Abb 4.4.9:** *Zusammenbinden von Objektkonturen im Zeitbereich.* Aufgetragen ist die Summe der Aktionspotentiale aller Synchronisationsneuronen in Abhängigkeit von der Zeit. Die relative Größe des Objekts und die erhöhte Anzahl an detektierten Ecken führen zu etwas längeren Repräsentations-Zeitintervallen.

Anhand eines weiteren Beispiels soll die Fähigkeit des Netzes, dreidimensionale Objekte als solche zusammenzubinden, gezeigt werden. Die künstliche Szene besteht aus einem Zylinder, der einen Würfel teilweise verdeckt (Abb. 4.4.13). Wie die Objektkonturen, die auch dreischenklige Vertizes enthalten, zusammengebunden werden, geht aus den in derselben Abbildung gezeigten Rückprojektionen hervor. Den zeitlichen Verlauf der Aktivitäten der Synchronisationsneuronen für beide Objekte zeigt die Abbildung 4.4.14, aus der auch die zeitliche Separation ersichtlich wird.

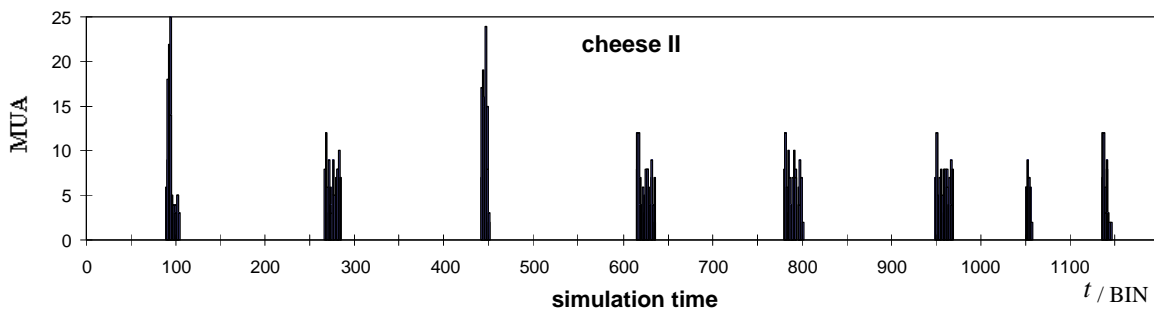
Abbildung 4.4.15 zeigt die Verarbeitung der bekannten Verkehrsszene. Auch hier werden die Konturen nach einiger Zeit in zwei Objekte gruppiert. Die Kontur des Autos im Vordergrund wird von den Hintergrundstrukturen (Abb. 4.4.16), die wegen der verwendeten räumlichen Auflösung nicht im Detail repräsentiert werden, getrennt.



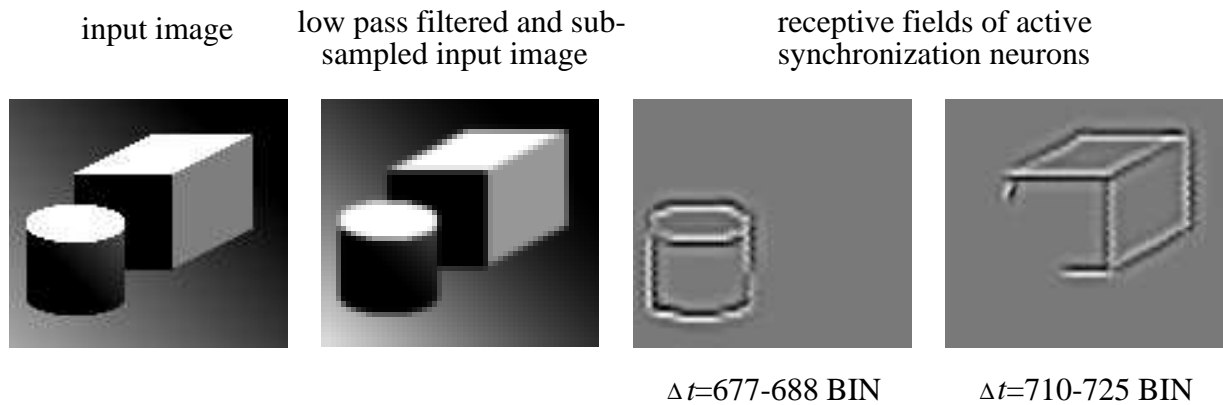
**Abb 4.4.10:** *Simulationsbeispiel: detektiertes Objekt vor nichtdetektiertem Objekt (Frühstücksszene I).* Das Eingangsbild zeigt eine Käsescheibe auf einer Vollkornbrotsscheibe. Die vom Netz zu bewältigende Aufgabenstellung soll auch hier verdeutlicht werden, daher ist neben dem Eingangsbild das mit der in der Vorverarbeitung verwendeten Gaußmaske gefilterte und anschließend (aus Darstellungsgründen kartesisch) mit Abtastweite  $d=2$  unterabgetastete Eingangsbild dargestellt. **Rechts** ist die Rückprojektion der in einem bestimmten Zeitintervall aktiven Synchronisationsneuronen dargestellt. Da im bearbeiteten Bildausschnitt das Hintergrundobjekt (Vollkornbrotsscheibe) keine Ecken besitzt, sollte es nicht als eigenständiges Objekt detektiert werden. Jedoch wird auch ein Stück der Vollkornbrotkontur aufgrund der unmittelbaren räumlichen Nähe eines T-Stücks zu dem die untere Ecke repräsentierenden Vertex von der entsprechenden Linkingwelle durchlaufen. Hier liegt ein Auflösungsproblem vor.



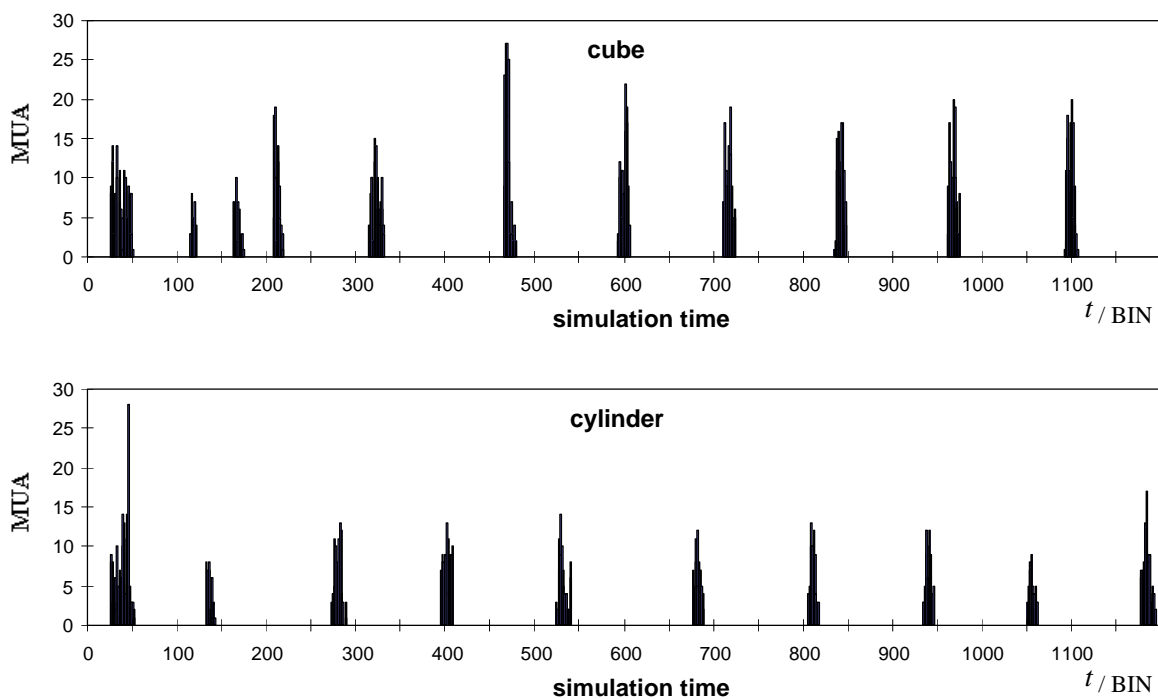
**Abb 4.4.11:** *Simulationsbeispiel: detektiertes Objekt vor nichtdetektiertem Objekt (Frühstücksszene II).* Das Eingangsbild zeigt eine "abgebissene" Käsescheibe auf einer Vollkornbrotsscheibe. Zur Verdeutlichung der vom Netz zu bewältigenden Aufgabenstellung ist auch das tiefpaßgefilterte und unterabgetastete Eingangsbild dargestellt. **Rechts** ist erneut die Rückprojektion der in einem bestimmten Zeitintervall aktiven Synchronisationsneuronen dargestellt. Im Gegensatz zum vorherigen Beispiel liegt kein T-Stück in direkter Vertexträhe vor, und die Konturen der Käsescheibe werden fehlerfrei detektiert. Da im bearbeiteten Bildausschnitt das Vollkornbrot keine Ecken besitzt, wird es nicht als eigenständiges Objekt detektiert.



**Abb 4.4.12:** *Zusammenbinden von Objektkonturen im Zeitbereich.* Aufgetragen ist die Summe der Aktionspotentiale aller Synchronisationsneuronen in Abhängigkeit von der Zeit. Die Konturen der Käsescheibe werden vollständig innerhalb der Repräsentationszeitintervalle von einer Linkingwelle umlaufen.



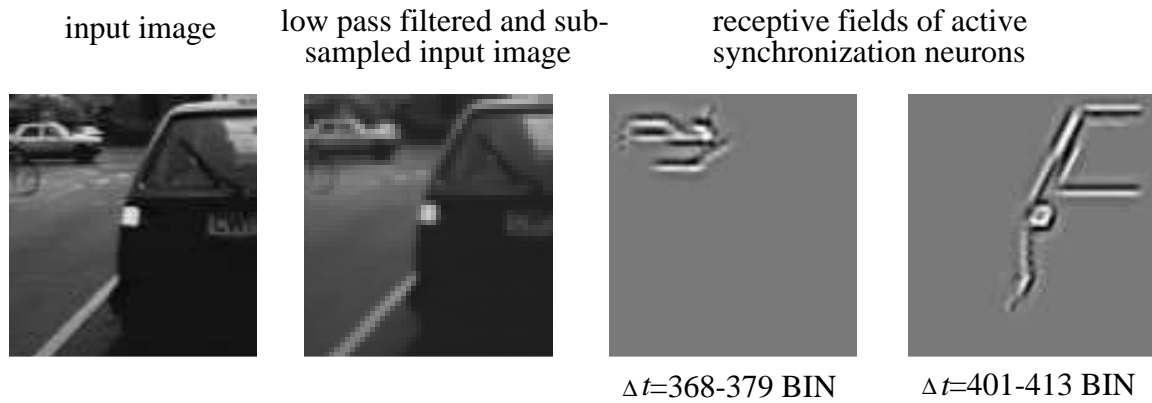
**Abb. 4.4.13:** *Simulationsbeispiel: 3D-Objekte.* Links ist die als Stimulus verwendete künstliche Szene eines Zylinders, der einen Würfel teilweise überdeckt, dargestellt. Daneben ist zur Verdeutlichung der Aufgabenstellung des Netzes das entsprechend der Vorverarbeitung tiefpaßgefilterte und unterabgetaste Eingangsbild dargestellt. Rechts sind die rezeptiven Felder der aktiven Synchronisationsneuronen für verschiedene Zeitintervalle dargestellt, woraus auch hier die erfolgreiche zeitliche Gruppierung aller Konturelemente in zwei Objektrepräsentationen deutlich wird.



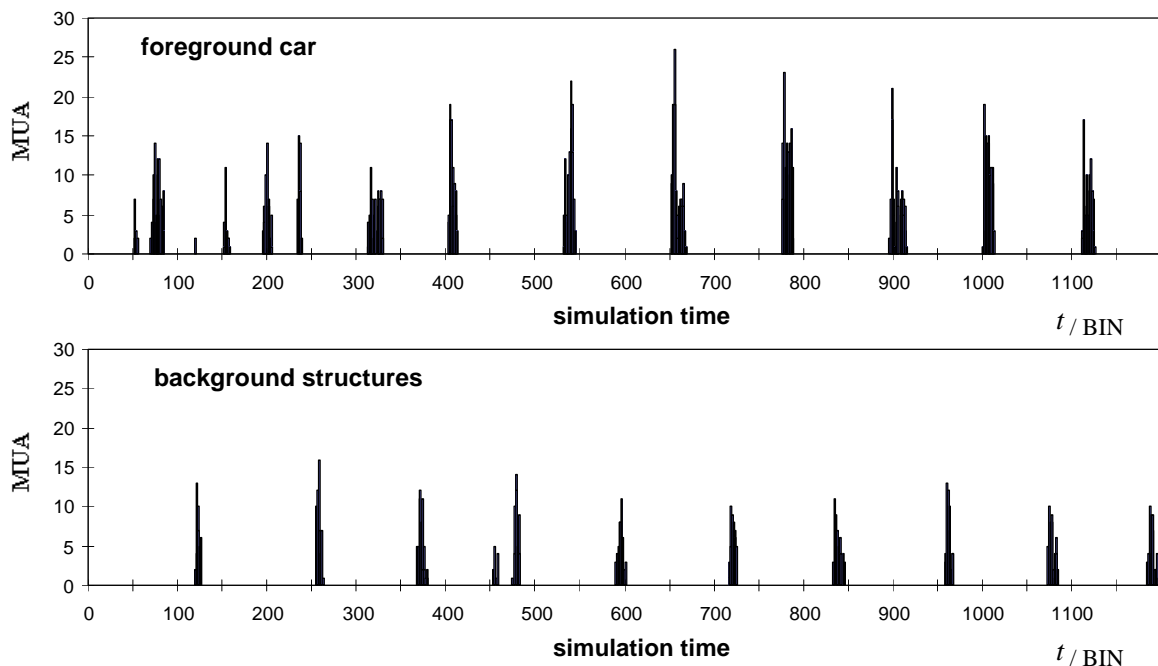
**Abb 4.4.14:** *Objektseparation im Zeitbereich.* Aufgetragen ist auch hier die Summe der Aktionspotentiale der das jeweilige Objekt repräsentierenden Synchronisationsneuronen in Abhängigkeit von der Zeit. Nach etwa 200 Zeitschritten sind die Repräsentation des Würfels (**oben**) und die Repräsentation des Zylinders (**unten**) zeitlich separiert.

Die Fahrbahnmarkierung bildet zunächst kein eigenes Objekt, da sie im bearbeiteten Bereich zwar T-Stücke, aber keine Vertizes bildet. Bei sehr genauer Betrachtung läßt sich jedoch auch eine Ecke erkennen, die die Fahrbahnmarkierung mit der Kontur des Vordergrundautos bildet und die auch durch geringe Aktivität von den Vertex-Zeiger-Schichten angezeigt wird. Daher kann es hier verschiedentlich zur Einbindung der Fahrbahnmarkierung in die Pkw-Kontur kommen.



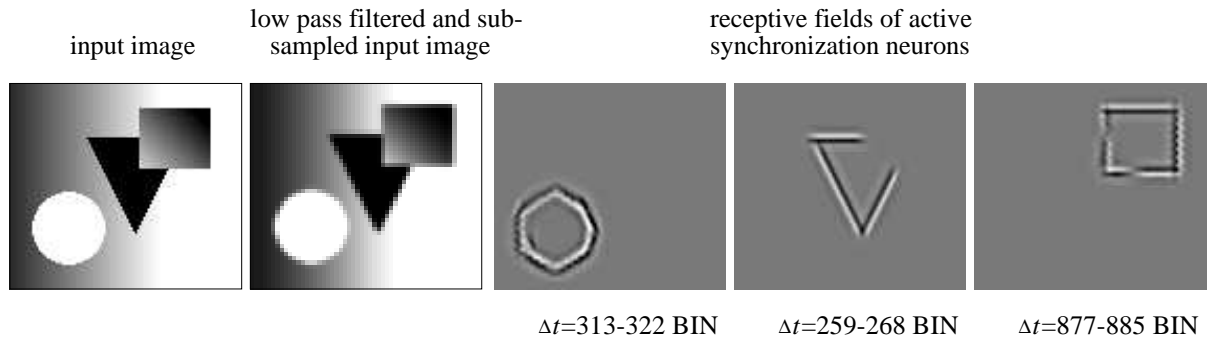


**Abb. 4.4.15:** Simulationsbeispiel: sich nicht verdeckende Objekte (Verkehrsszene). Das als Stimulus verwendete Bild zeigt ein Auto im Vordergrund, eine Fahrbahnmarkierung sowie verschiedene Strukturen im Hintergrund. Daneben ist auch hier zur Verdeutlichung der Aufgabenstellung des Netzes das entsprechend der Vorverarbeitung tiefpaßgefilterte und unterabgetastete Eingangsbild dargestellt. In der Mitte und rechts sind die rezeptiven Felder der aktiven Synchronisationsneuronen für verschiedene Zeitintervalle dargestellt. Auch in dieser Simulation werden die Konturelemente in zwei Objektrepräsentationen gruppiert. Die Fahrbahnmarkierung wird zunächst nicht als eigenständiges Objekt detektiert, da sie keine Vertizes, sondern nur T-Stücke im bearbeiteten Bildausschnitt aufweist (siehe auch Text).

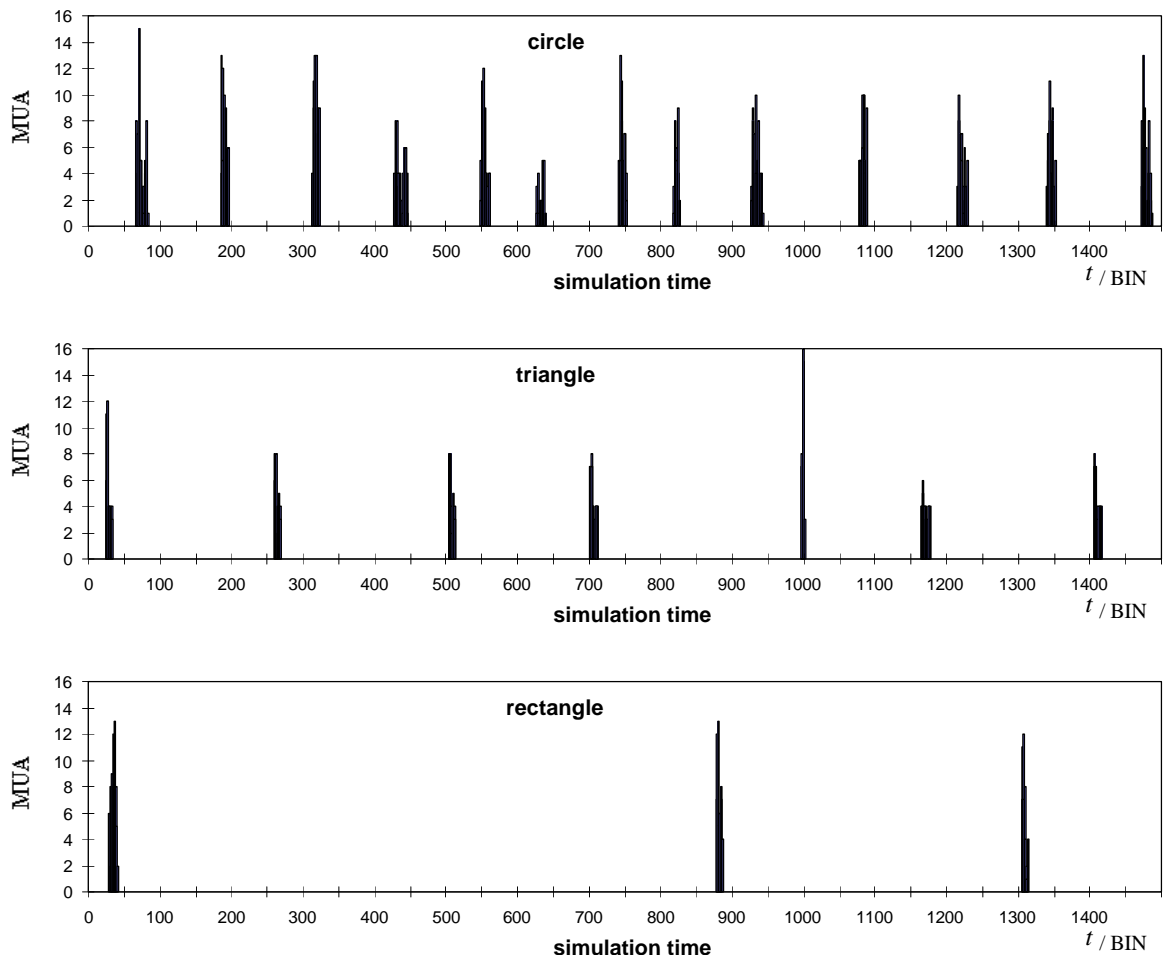


**Abb 4.4.16:** Objektseparation im Zeitbereich. Aufgetragen ist erneut die Summe der Aktionspotentiale der das jeweilige Objekt repräsentierenden Synchronisationsneuronen in Abhängigkeit von der Zeit. Nach etwa 500 Zeitschritten befindet sich das Netz in einem stabilen Zustand und liefert abwechselnd Repräsentationen der Konturen des Autos im Vordergrund (**oben**) und der Konturen der Hintergrundstrukturen (**unten**), was einer Separation der Objekte im Zeitbereich entspricht.

In einem weiteren Beispiel werden mehrere Objekte gezeigt, die vom Netz in sich zusammengebunden und wechselseitig zeitlich separiert werden (Abb. 4.4.17 und 4.4.18). Das entsprechende Bild zeigt einen Kreis, ein durch einen variablen Grauwertgradienten definiertes Rechteck und möglicherweise ein Dreieck, das vom Rechteck scheinbar teilweise verdeckt wird.

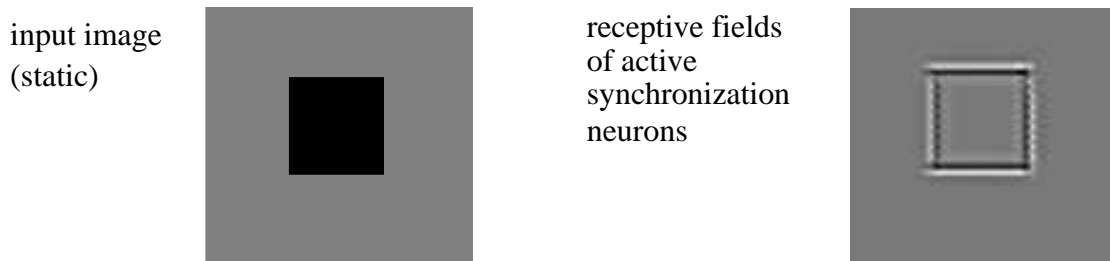


**Abb. 4.4.17:** *Simulationsbeispiel: drei Objekte.* Links ist die als Stimulus verwendete künstliche Szene dreier geometrischer Figuren dargestellt, wobei das mögliche Dreieck durch ein Rechteck teilweise verdeckt wird. Daneben wird zur Verdeutlichung der Aufgabenstellung des Netzes das entsprechend der Vorverarbeitung tiefpaßgefilterte und unterabgetastete Eingangsbild gezeigt. Rechts sind die Rückprojektionen der aktiven Synchronisationsneuronen für verschiedene Zeitintervalle dargestellt. Das Netz bindet die Konturen zu drei Objekten zusammen.

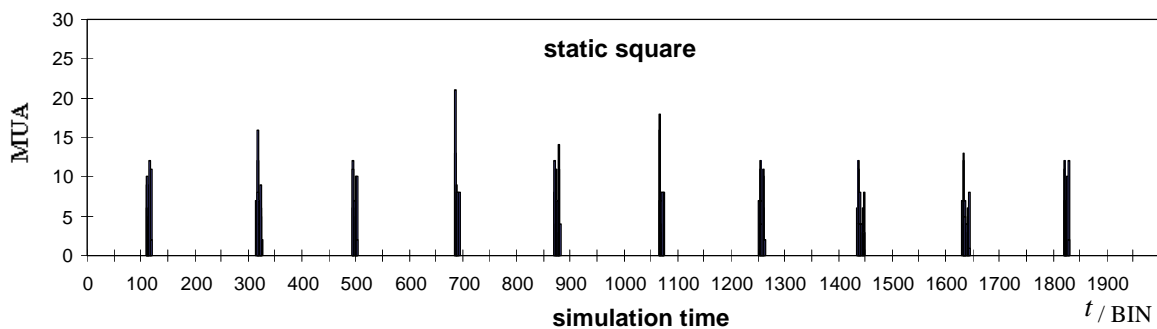


**Abb 4.4.18:** *Objektseparation im Zeitbereich.* Dargestellt ist die Summe der Aktionspotentiale der das jeweilige Objekt repräsentierenden Synchronisationsneuronen in Abhängigkeit von der Zeit. Die Konturen des Kreises (**oben**) werden deutlich häufiger repräsentiert als die Konturen des Dreiecks (**mitte**) oder die Konturen des Rechtecks (**unten**). Der Grund ist in der großen Anzahl an detektierten Vertizes des Kreises mit hoher Grauwertdifferenz zu suchen.

Die Konturen der drei Objekte werden vom Netz zu drei entsprechenden Repräsentationen zusammengebunden (Abb. 4.4.18). Dabei werden die Konturen des Kreises häufiger repräsentiert als die Konturen der übrigen Objekte, was auf die größere Anzahl an detektierten Vertizes des Kreises mit hoher Grauwertdifferenz zurückzuführen ist.



**Abb. 4.4.19:** Simulationsbeispiel: statisches Objekt (Referenzszene). Das statische Bild eines schwarzen Quadrats auf grauem Grund ( $\Delta G=100$ ) wird als Stimulus dargeboten. Die Synchronisationsneuronen repräsentieren die Konturen des Quadrats in bestimmten Zeitintervallen



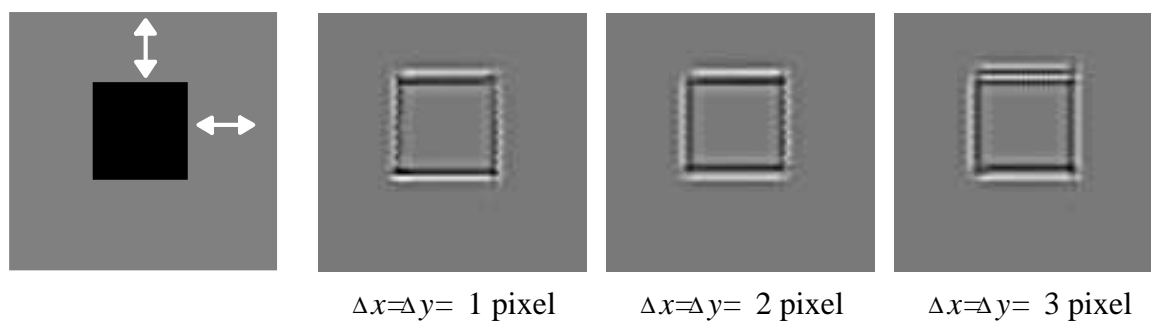
**Abb. 4.4.20:** Zusammenbinden von Objektkonturen im Zeitbereich. Das schwarze Quadrat wird in nahezu regelmäßigen zeitlichen Abständen durch Linkingwellen repräsentiert. Dies spiegelt sich in den hier in Abhängigkeit von der Zeit aufgetragenen Gesamtantworten der Synchronisationsneuronen wider.

### Dynamische Szenen

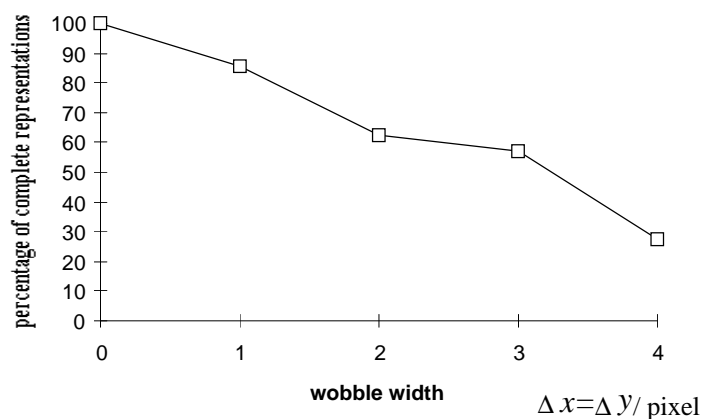
Bisher wurden ausschließlich statische Szenen dargeboten. Soll das Kontur/Form-System aber reale Kamerasequenzen verarbeiten, so wird es auch mit erschwerenden Faktoren, wie Wackeln der Kamera oder Bewegung von Objekten, konfrontiert. Anhand der folgenden Simulationsbeispiele wird das Netz auf Robustheit gegen solche Verwicklungen von Bildern und die Fähigkeit, bewegte Objekte zu verarbeiten, untersucht. Dazu wird zunächst als (statische) Bezugsszene ein schwarzes Quadrat auf grauem Hintergrund als Eingangssequenz verwendet. Abbildung 4.4.19 zeigt den Stimulus und die Antwort der Synchronisationsneuronen innerhalb eines Zeitintervalls. In Abbildung 4.4.20 ist das zeitliche Antwortverhalten der Synchronisationsneuronen aufgetragen. In jedem der Repräsentationszeitintervalle wird das Quadrat vollständig repräsentiert, d.h. eine Linkingwelle umläuft die vier Seiten des Quadrats.

wobbled input sequence

receptive fields of active synchronization neurons



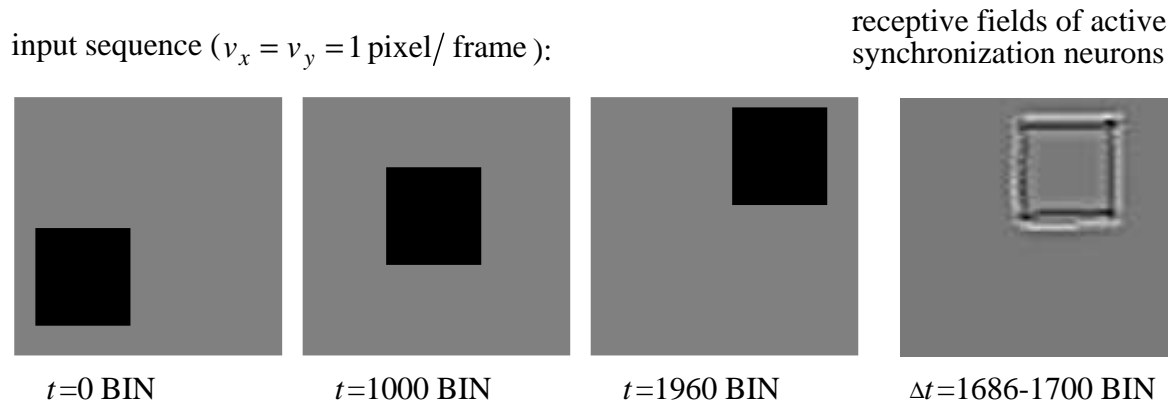
**Abb. 4.4.21:** Simulationsbeispiel: lateral bewegtes (verwackeltes) Objekt. Als Eingangssequenz wird die Bildfolge eines in x- und y-Richtung verwackelten Quadrats verwendet. Je nach verwendeten Sprungweiten  $\Delta x$  und  $\Delta y$  ergibt sich ein unterschiedliches Repräsentationsverhalten (siehe auch Text). Für die verschiedenen Sprungweiten sind die rezeptiven Felder aller in bestimmten Zeitintervallen aktiven Synchronisationsneuronen dargestellt.



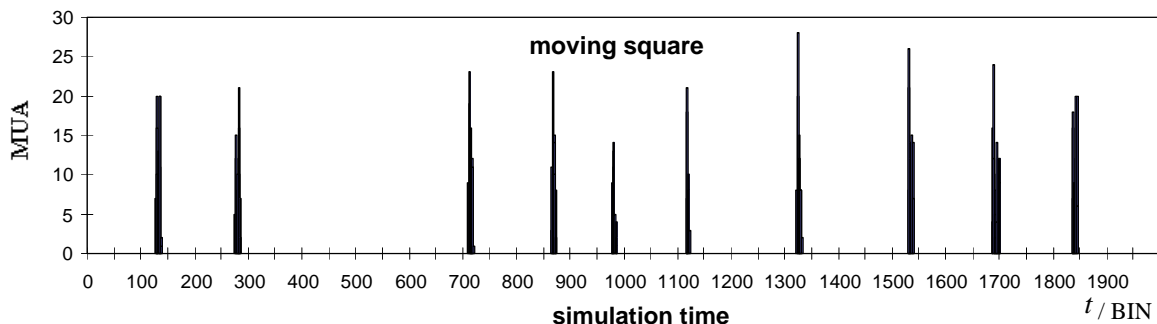
**Abb. 4.4.22:** Repräsentationsleistung bei verwackeltem Objekt. Aufgetragen ist der prozentuale Anteil der vollständigen Repräsentationen (Linkingwelle umschreibt die gesamte Kontur des Objekts) an der Gesamtheit aller Repräsentationen gegen die Sprungweiten  $\Delta x$  und  $\Delta y$ , die der Stärke der Verwacklung entsprechen. Mit zunehmender Verwacklung treten häufiger unvollständige Repräsentationen auf, und die Leistung nimmt ab.

Die statische Szene wird anschließend durch eine Folge "verwackelter" Bilder ersetzt. Dazu werden vier verschiedene Einzelbilder des vorherigen statischen Bildes (schwarzes Quadrat) erzeugt, die sich nur in der räumlichen Position des Quadrats unterscheiden. Das Quadrat bleibt in einem der Bilder unverändert am selben Ort, in den drei übrigen Bildern befindet es sich um einen Pixel in x-Richtung, um einen Pixel in y-Richtung oder um je einen Pixel in x- und y-Richtung versetzt. Von diesen vier Bildern wird in jedem 32. Zeitschritt eines zufällig ausgewählt, so daß eine Sequenz eines (um  $\Delta x = \Delta y = 1$  Pixel/Bild) verwackelten Quadrats dargeboten wird. Innerhalb der Simulationsdauer von 2000 BIN repräsentiert die Aktivität der Synchronisationsneuronen das Quadrat nur noch in fast 90% aller Repräsentationszeitintervalle vollständig. Die Simulation wird für weitere Sprungweiten  $\Delta x$  und  $\Delta y$  wiederholt. In Abbildung 4.4.21 sind bei verschiedenen Sprungweiten erhaltene vollständige Repräsentationen dargestellt, und in Abbildung 4.4.22 sind die Prozentsätze vollständiger Repräsentationen gegen die untersuchten Sprungweiten aufgetragen. Mit

zunehmender Verwacklung treten immer häufiger fehlerhafte, d.h. unvollständige Repräsentationen auf. Es wird nun davon ausgegangen, daß das Zusammenbinden von Objektkonturen dann erfolgreich ist, wenn mindestens 75% aller Repräsentationen vollständig sind. Daraus ergibt sich für das System, daß es Sequenzen mit Verwacklungen bis etwa 1 Pixel/Bild erfolgreich verarbeiten kann.



**Abb. 4.4.23:** *Simulationsbeispiel: bewegtes Objekt.* Dem Netz wird die Bildfolge eines sich mit  $v_x=v_y=1$  Pixel/Bild bewegenden schwarzen Quadrats dargeboten. Die neuronale Aktivität der Synchronisationsschichten vermag die Konturen des Quadrats trotz Bewegung vollständig zu repräsentieren. Exemplarisch sind die rezeptiven Felder der aktiven Synchronisationsneuronen für ein Repräsentationszeitintervall  $\Delta t$  dargestellt.



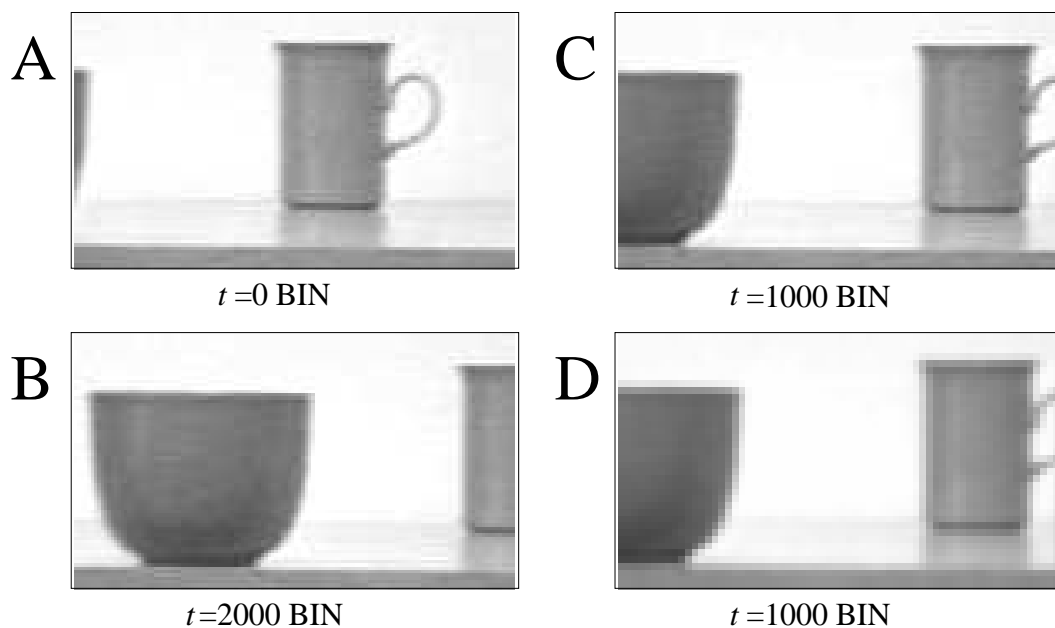
**Abb. 4.4.24:** *Zusammenbinden der Konturen eines bewegten Objekts im Zeitbereich.* Dargestellt ist die Summe der Aktionspotentiale aller Synchronisationsneuronen, die sich bei Präsentation der oben dargestellten Sequenz (Abb. 4.4.23) ergibt. Die Repräsentationszeitintervalle folgen nicht so regelmäßig aufeinander wie es bei Präsentation des statischen Quadrats der Fall ist (vgl. Abb. 4.4.20), dennoch liegt in jedem der Intervalle eine vollständige Repräsentation vor.

Neben der Verarbeitung verwackelter Szenen interessiert auch die Verarbeitung von Bildfolgen mit bewegten Objekten. Die in Abbildung 4.4.23 teilweise dargestellte Sequenz eines sich von einer Bildecke aus diagonal über das Bild bewegenden Quadrats (Verschiebungsgeschwindigkeit  $v_x=v_y=1$  Pixel/Bild) wird als Stimulus dargeboten. Das Objekt wird in den verschiedenen Zeitintervallen vollständig richtig repräsentiert (Abb. 4.4.23 und 4.4.24).

Bereits bei Verdopplung der Verschiebungsgeschwindigkeit antworten die Synchronisationsneuronen nicht mehr, so daß hier von einer maximal verarbeitbaren Geschwindigkeit von  $v_x=v_y=1$  Pixel/Bild ausgegangen werden muß.

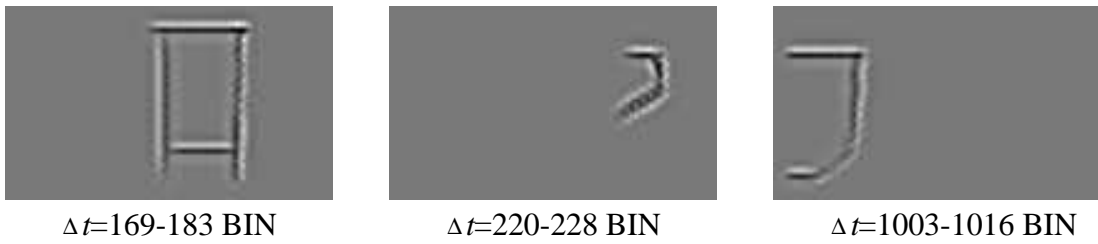
In den folgenden beiden dynamischen Simulationsbeispielen treten typischerweise lokale Geschwindigkeiten und Verwacklungen von etwa 1 Pixel/Bild auf. Wie sich hier bestätigt, kann das System solche Sequenzen verarbeiten. Abbildung 4.4.25 zeigt zeitliche Ausschnitte einer dynamischen Szene, in der eine Schüssel und eine Tasse auf einem Schreibtisch stehen. Die Kamera schwenkt dabei von rechts nach links. Zunächst wird der Rumpf der Tasse repräsentiert (Abb. 4.4.26). Der Tassenhenkel wird als eigenes Objekt behandelt, da seine Konturen ausschließlich T-Stücke mit den Konturen des Tassenrumpfes bilden. Nach einigen hundert Zeitschritten ist die Schüssel soweit in das Bild getreten, daß sie erstmals repräsentiert wird. Bei etwa 1800 Bin liegt eine unvollständige Repräsentation der Schüssel vor, was auf die zu diesem Zeitpunkt besonders ruckartige Kameraführung zurückzuführen sein dürfte. In Abbildung 4.4.27 sind die zeitlichen Verläufe der Gesamtaktivitäten aller Synchronisationsneuronen aufgetragen.

input sequence

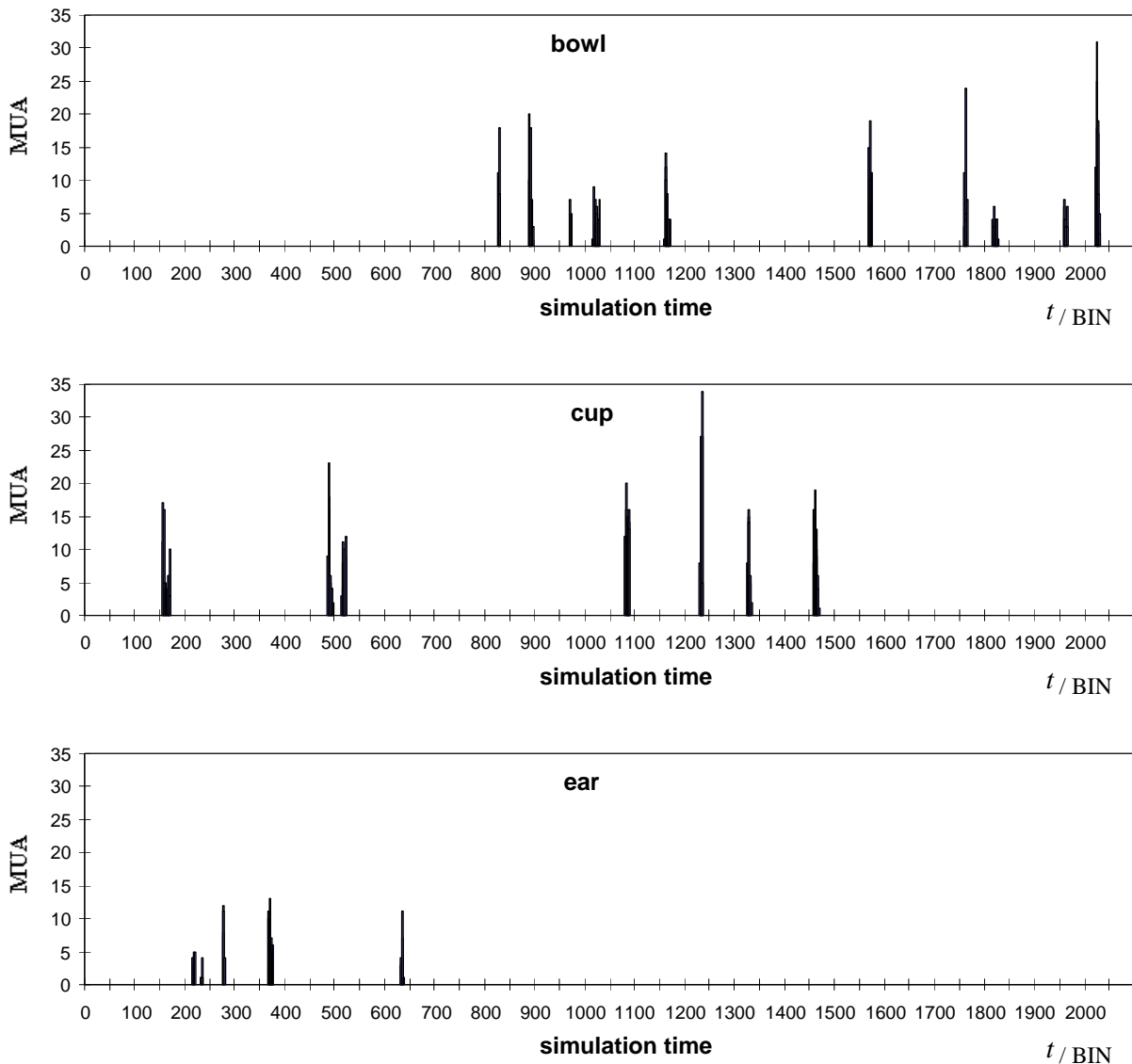


**Abb. 4.4.25:** *Simulationsbeispiel: dynamische Szene (Kameraschwenk).* **A, B, C:** Gezeigt werden verschiedene Bilder einer Sequenz in der sich der Beobachter an einer Tasse und einer Schüssel vorbeibewegt. **D:** Weiterhin ist für  $t=1000$  BIN das gemäß der Vorverarbeitung tiefpaßgefilterte und unterabgetastete Bild dargestellt, um auf die vom Kontur/Form-System zu bewältigende Aufgabe hinzuweisen.

receptive fields of active synchronization neurons

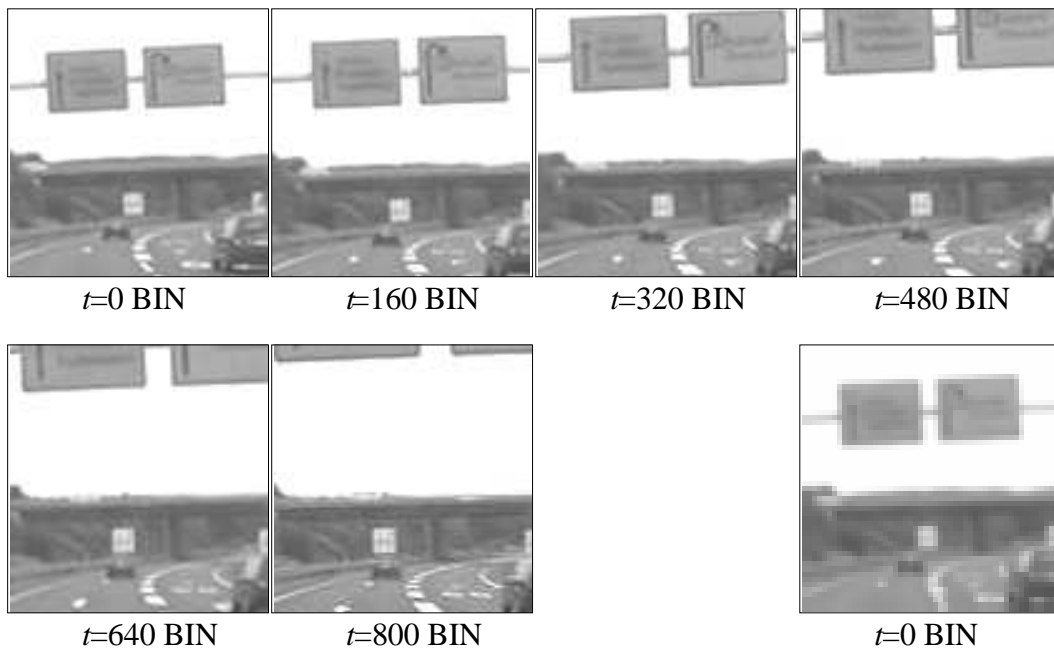


**Abb. 4.4.26:** *Simulationsergebnis: dynamische Szene (Kameraschwenk).* Dargestellt sind die rezeptiven Felder aller aktiven Synchronisationsneuronen für verschiedene Zeitintervalle. Die Szene wird in drei Objekte zerlegt: Tassenrumpf, Henkel und Schüssel. Der Henkel stellt ein eigenes Objekt dar, da seine Konturen mit denen des Tassenrumpfes ausschließlich T-Stücke bildet. Der Tassenrumpf wird einschließlich der Konturen seines Schattens repräsentiert, da die Konturen der Tasse mit denen des Schattens keine T-Stücke, sondern aufgrund vieler Linienenden Vertizes bildet.



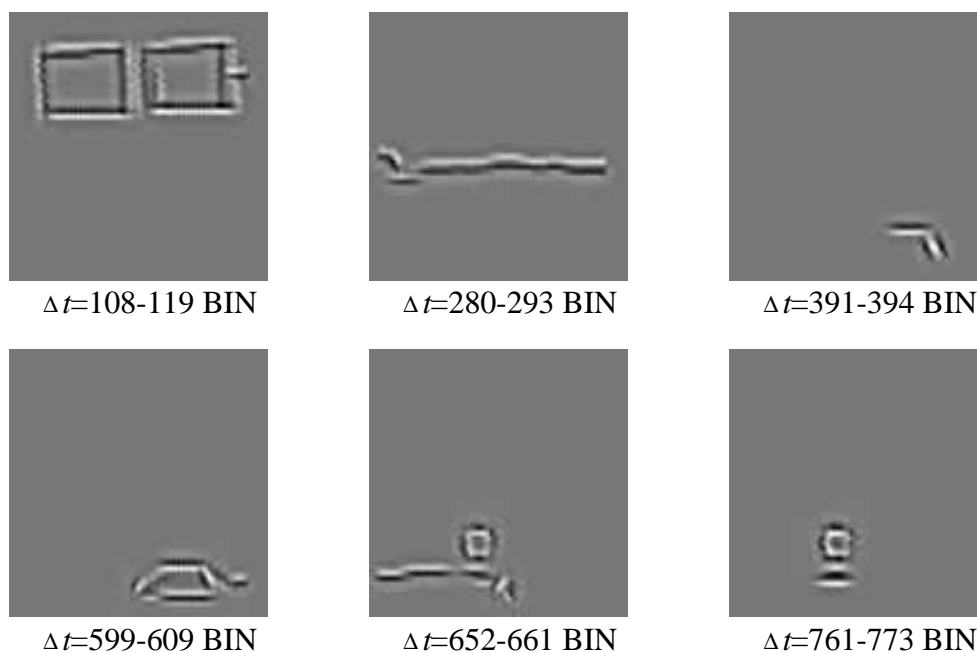
**Abb. 4.4.27:** *Objektseparation im Zeitbereich.* Aufgetragen ist die Summe der das jeweilige Objekt repräsentierenden aktiven Synchronisationsneuronen. Da sich zuerst die Tasse im Bild befindet, werden zunächst Tassenrumpf (cup) und Henkel (ear) repräsentiert. Später tritt auch die Schüssel hinzu und der Tassenhenkel aus dem Bild. In den meisten Fällen liegen vollständige Repräsentationen vor, lediglich bei etwa 1820 BIN (Schüssel) und 240 BIN (Henkel) treten unvollständige Repräsentationen auf.

input sequence:



**Abb. 4.4.28:** *Simulationsbeispiel: dynamische Verkehrsszene.* Dargestellt sind einige Bilder einer aus einem fahrenden Pkw aufgenommenen Sequenz. Außerdem wird für  $t=0$  das entsprechend der Vorverarbeitung mit einer Gaußmaske gefilterte und anschließend mit Abtastweite  $d=2$  (aus Darstellungsgründen kartesisch) unterabgetastete Eingangsbild gezeigt, um die Aufgabenstellung zu verdeutlichen.

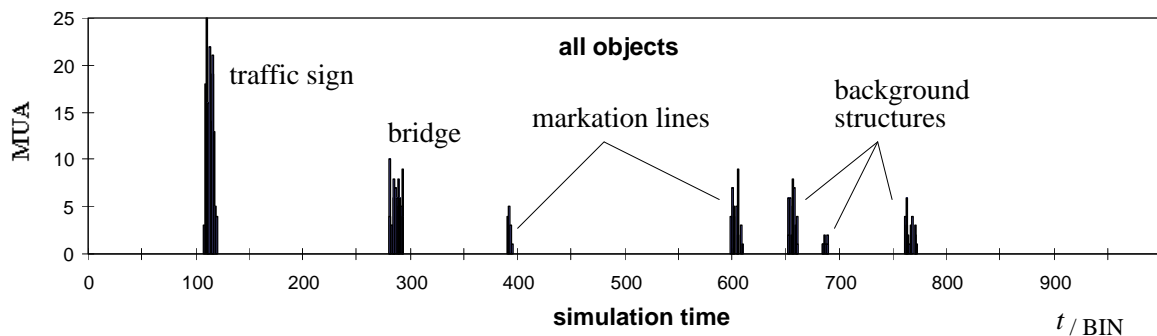
receptive fields of active synchronization neurons



**Abb. 4.4.29:** *Simulationsergebnis: dynamische Verkehrsszene.* In verschiedenen Zeitintervallen werden die Wegweisertafel, die Konturen der Brücke, Fahrbahnmarkierungen und das weiße Schild zusammen mit Konturen des Pkw und Leitplanken repräsentiert.



Abbildung 4.4.28 zeigt einige Bilder der letzten Beispielsequenz, die aus einem auf einer Kraftfahrstraße fahrenden Pkw aufgenommen wurde. Zunächst werden die relativ großen Wegweisertafeln (Abb. 4.4.29) und anschließend in verschiedenen Zeitintervallen verschiedene weitere Objekte repräsentiert: die Brücke (hier treten auch Vertizes auf), Fahrbahnmarkierungen und das weiße Schild im Hintergrund zusammen mit einem Pkw und Leitplanken. Die Objekte werden aufgrund kurzer Präsentation bzw. hoher Geschwindigkeit nur einmal (Wegweisertafel im Vordergrund), höchstens aber dreimal (weißes Schild im Hintergrund) repräsentiert.



**Abb. 4.4.30:** *Objektseparation im Zeitbereich.* Dargestellt ist die Summe aller Aktionspotentiale der Synchronisationsneuronen in Abhängigkeit von der Zeit. Da sich die Wegweisertafel relativ zum Benutzer mit hoher Geschwindigkeit bewegt und das Bild teilweise verläßt, wird sie nur einmal repräsentiert. Die Hintergrundstrukturen befinden sich nahezu im Fluchtpunkt und können aufgrund der geringen Geschwindigkeit mehrfach repräsentiert werden.

#### 4.4.3 Diskussion

Anhand der Simulationsbeispiele des vorangegangenen Unterabschnitts wurde die Leistungsfähigkeit des Kontur/Form-Systems verdeutlicht. Das System ermöglicht das Zusammenbinden von Konturen, die einzelne Objektkandidaten repräsentieren, durch Linkingwellen und weiterhin erlaubt es eine Trennung dieser Konturen je nach Zugehörigkeit zu Objektkandidaten untereinander durch einen globalen auf die Vertexrepräsentationen wirkenden Inhibitionsmechanismus.

Es läßt sich jedoch nicht von vornherein sagen, wieviele Objekte einer Szene höchstens separiert werden können. Die Anzahl der maximal separierbaren Objekte wird von zwei wesentlichen Faktoren bestimmt, nämlich erstens von den relativen Refraktärzeiten der Neuronen und zweitens von der räumlichen Auflösung. Um eine möglichst große Anzahl an Objekten separieren zu können, muß eine lange Refraktärzeit der entsprechenden Synchronisationsneuronen und Vertexdetektoren vorhanden sein, so daß viele verschiedene Repräsentationen in einem bestimmten zeitlichen Intervall Platz finden können. Sehr viel kritischer ist hingegen die Frage nach der Anzahl an Objekten, die bei gegebener Größe der neuronalen Schichten noch vernünftig aufgelöst dargestellt werden können. Nur wenn

genügend Neuronen zur Verfügung stehen, lassen sich entsprechend viele Objekte geeignet repräsentieren. Die gezeigten Simulationen wurden mit aus 2200 Neuronen bestehenden Schichten durchgeführt. Die Verarbeitung von Szenen mit sehr vielen Objektkandidaten durch einen Bildausschnitt mit solch einer relativ geringen Auflösung erlaubt kaum eine Separation vieler detaillierter Objekte. Insbesondere sei auf die Detektion von Vertices hingewiesen, die bereits einen gewissen räumlichen Toleranzbereich für sich beanspruchen. Es scheint jedoch nicht notwendig zu sein, sehr viele Objekte separieren zu können, da in einem Gesamtsystem ("elektronisches Auge") zunächst die Anzahl der in der Szene zu verarbeitenden Objekte durch ein Aufmerksamkeitsfenster deutlich reduziert wird.

In einigen Simulationsbeispielen wurden die Konturen runder Objekte, die als Vielecke behandelt werden, verarbeitet. Hier zeigte sich, daß die Repräsentationszeitintervalle dieser Objekte etwas länger als die nicht-runder Objekte ausfielen (vgl. Abb. 4.4.9). Zwar sind diese zeitlichen Ausdehnungen relativ klein, wenn sie mit typischen Refraktärzeiten verglichen werden, so daß hier kein dringliches Problem vorliegt. Dennoch ist eine Gleichbehandlung runder und nicht-runder Objektkonturen wünschenswert. Dies ließe sich beispielsweise durch den folgenden Ansatz, der eine Änderung in der Vertexdetektion und eine Erweiterung der Linkingverbindungen vorsieht, realisieren. Schließen zwei zusammenlaufende Kanten einen sehr stumpfen oder einen sehr spitzen Winkel ein, dann möge kein Vertex vorliegen. Linienenden dürfen nun in zwei End-Stop-Schichten, die ähnliche Orientierungen kodieren, repräsentiert werden (Populationskode). Weiterhin werden Linkingverbindungen zwischen Neuronen der Kantendetektorschichten und zwischen Neuronen der Synchronisationsschichten ähnlicher Vorzugsorientierungen eingeführt. Ein rundes Objekt läßt sich nun als Konturzug ohne Ecke repräsentieren, falls Linkingwellen auch von Synchronisationsneuronen initiiert werden dürfen. Jedoch ist dieses Vorgehen mit einem erheblichen technischen Mehraufwand verbunden. Die Anzahl der zu verarbeitenden Orientierungen müßte nämlich verdoppelt werden ( $0^\circ$  bis  $180^\circ$  in  $15^\circ$ -Schritten statt, wie bisher, in  $30^\circ$ -Schritten), da sonst Ecken, die Winkel von  $0^\circ$  bis zu etwa  $45^\circ$  bzw. von  $135^\circ$  bis  $180^\circ$  einschließen, nicht als solche detektiert würden. Durch die doppelte Anzahl würden statt dessen nur die Winkel  $0^\circ$  bis etwa  $22,5^\circ$  und  $157,5^\circ$  bis  $180^\circ$  keine Ecke mehr darstellen. Unter diesen Bedingungen verdoppelte sich die Anzahl der benötigten Neuronen zu Lasten der räumlichen Auflösung und die Anzahl der Verbindungen würde sich sogar nahezu vervierfachen. Da ein solcher Aufwand in keinem guten Verhältnis zum Gewinn (kleines Repräsentationszeitintervall) stünde, wurden die Simulationen nicht mit einem so modifizierten Netz durchgeführt.

Wie oben beschrieben, würden bei einer solchen Modifikation auch Synchronisationsneuronen eine Linkingwelle initiieren können. Das vorliegende Netz erlaubt jedoch nur Initiationen von Linkingwellen durch Master-Vertex-Neuronen, um zu vermeiden, daß auch einzelne Linien oder Kanten eigenständig repräsentiert werden. Die Konturen eines Objekts werden nur als solche repräsentiert, wenn sie im zu verarbeitenden Bildausschnitt mindestens einen Vertex aufweisen (vgl. Abb. 4.4.11). Diese Vorgabe ist insbesondere durch die

Einführung der Synchronisationsschichten neben den Kantendetektorschichten möglich. Der hauptsächliche Grund für die Unterscheidung zwischen Kantendetektorschichten und Kontur-Synchronisationsschichten liegt aber in der Verschiedenheit ihrer Aufgaben, die bei der Verarbeitung nichtbinärer Eingangssignale nur schwer in einer Verarbeitungsebene zugleich bewältigt werden können. Die für die End-Stop- und Vertex-Detektion notwendige Eindeutigkeit der Kantendetektorantworten erfordert eine große Linkingzeitkonstante, deren Einsatz aber gerade die Synchronisation der Aktivitäten erschwert. Daher wird unter Verwendung kleiner Linking-Zeitkonstanten zumindest lokale synchrone Aktivität gesondert in den Synchronisationsschichten generiert. Synchronizität im eigentlichen Sinne liegt jedoch nicht vor, da sich die Ausbreitung einer fortschreitenden Linkingwelle über ein Zeitintervall erstreckt. Ob nun fortlaufende Aktivität oder rhythmisch synchronisierte (oszillierende) Aktivität der Physiologie näher stehen, ist bisher nicht geklärt und daher Gegenstand aktueller experimenteller Untersuchungen. Im vorliegenden Modell ist Gleichzeitigkeit in der Repräsentation aller Konturen eines Objekts kaum zu realisieren, weil bei der Verarbeitung natürlicher Szenen üblicherweise die Amplituden der Eingangssignale stark schwanken und sich demzufolge die Feedingpotentiale der zu synchronisierenden Neuronen zu sehr unterscheiden. Für weiterverarbeitende Stufen dürfte die zeitliche Ausdehnung der Repräsentationen kein Problem darsellen, solange die Repräsentationszeitintervalle klein gegen die Zeitintervalle zwischen verschiedenen Repräsentationen sind, was in den gezeigten Simulationen immer der Fall ist.

In den letzten Simulationsbeispielen werden auch dynamische Szenen erfolgreich verarbeitet. Dazu ist eine gewisse Robustheit gegen Verwacklungen des ganzen Bildes oder Bewegungen einzelner Objekte zumindest für kleine Geschwindigkeiten unabdingbar. Im Modell sind Änderungen bis zu 1 Pixel/Bild= 1 Pixel/32 BIN erlaubt, was in der Regel für dynamische Szenen mit nicht sehr schnell bewegten Objekten ausreichend ist. Sogar eine Szene, bei der die Kamera geschwenkt wurde, ließ sich verarbeiten (Abb. 4.4.25 bis 4.4.27). Die dazu notwendige Ortstoleranz ist in der Natur typischerweise eine Eigenschaft von Komplex-Zellen, wird aber im vorliegenden Modell vor allem durch die Breite der rezeptiven Felder von Kantendetektoren ermöglicht. Im Modell wird auf die Modellierung solcher Komplex-Zellen verzichtet, da sie das Netz deutlich vergrößern würden und zudem eine gewisse Ortstoleranz bereits ohne Komplex-Zellen realisiert werden konnte. Darüberhinaus soll das Kontur/Form-System, wenn es im Gesamtsystem eingesetzt wird, einen zu verarbeitenden Bildausschnitt durch ein Aufmerksamkeitsfenster erhalten, das schnelle Objekte verfolgt und daher in der Regel keine großen lokalen Geschwindigkeiten mehr enthält (Pauly et al. 1997). Erhöhung der Ortstoleranz könnte zudem mit einer Verringerung des Auflösungsvermögens verbunden sein. Auflösungsprobleme zeigen sich auch in einigen Simulationen (Abb. 4.4.10, 4.4.15 und 4.4.29). In der Frühstücksszene I liegen ein T-Stück und ein Vertex zu dicht beieinander, und in den Verkehrsszenen können Hintergrundobjekte nur schlecht aufgelöst werden. Solche Ungenauigkeiten lassen sich bei der Verarbeitung realer

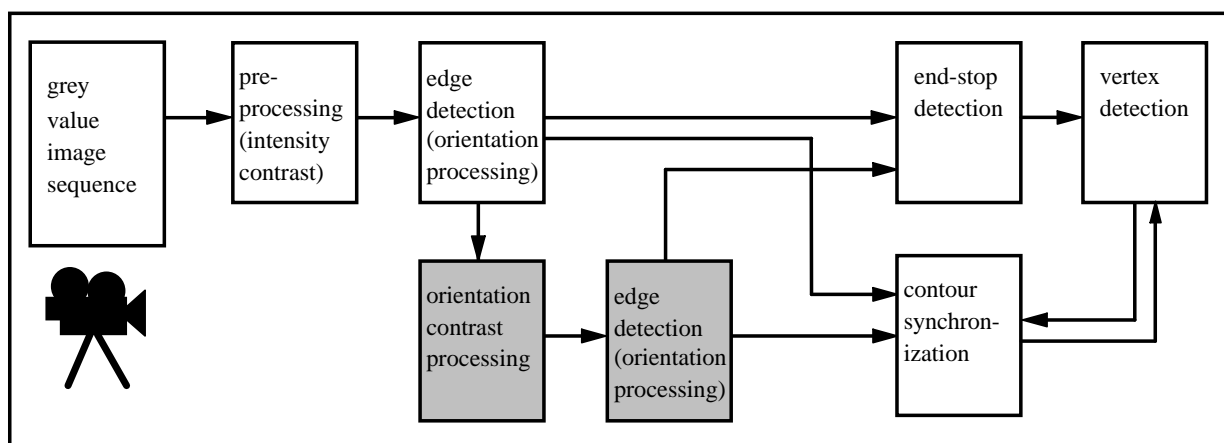
Szenen kaum ausschließen, jedoch kann durch eine größere Anzahl an Neuronen die räumliche Auflösung verbessert werden. Zudem wird im Modell nur eine Abtastweite, d.h. Auflösung aus Gründen der Rechnerkapazitäten realisiert, wodurch beispielsweise Kanten mit ausgedehntem Profil kaum detektiert werden können (vgl. Kap. 3). Durch die Verwendung weiterer Abtastweiten könnte die Kantendetektion durch Interaktionen zwischen Schichten verschiedener Auflösungen unterstützt werden.

Eine weitere Möglichkeit, die Leistungsfähigkeit des Kontur/Form-Systems zu verbessern, betrifft die Vorverarbeitung. Um eine Unterdrückung zeitlichen Rauschens zu ermöglichen, arbeiten X-Zellen bereits mit einem (kleinen) Schwellenoffset, der jedoch auch sehr schwache Signale unterdrückt. Durch Rückkopplungen von den Kantendetektoren oder Schichten höherer Verarbeitungsstufen ließe sich die bereits extrahierte Kontextinformation von den X-Zellen zugunsten der schwachen Signale nutzen. An der Verwirklichung dieser Ideen wird in der AG Neurophysik derzeit gearbeitet.



## 5 Ein Texturanalyseverfahren basierend auf Orientierungs- kontrast und dessen Ankopplung an das Kontur/Form-System

Oftmals kann der menschliche Betrachter Objekte nicht mehr aufgrund von Grauwertintensitätskontrast allein voneinander oder vom Hintergrund unterscheiden. Erst durch Verarbeitung zusätzlicher Merkmalskontraste wie Farbe oder Textur wird die Separation der Objekte möglich. Auch das Kontur/Form-System soll nun exemplarisch mit der Möglichkeit, einen weiteren Merkmalskontrast verarbeiten zu können, ausgestattet werden. Dazu wird zunächst ein eigenständiges aus dynamischen Neuronen aufgebautes Netz vorgestellt, das eine Erweiterung eines früheren aus McCulloch-Pitts-Neuronen aufgebauten Netzes darstellt (Weitzel 1995, Weitzel et al. 1995). Das Netz extrahiert orientierte Linienelemente von Texturen und bestimmt die lokale Orientierungsdifferenz zwischen benachbarten Linienelementen. Durch diesen Orientierungskontrast lassen sich oft Objektkonturen gewinnen. Die Erweiterung besteht nun vor allem darin, daß die Repräsentation der durch den stärksten im Bild enthaltenen Orientierungskontrast definierten Konturen eines Objekts die Repräsentation der aus schwächerem Orientierungskontrast erhaltenen Konturen eines anderen Objekts unterdrückt. Nach der Verdeutlichung der Leistungsfähigkeit des Netzes anhand künstlicher Szenen wird das Netz den Gegebenheiten des Kontur/Form-Systems angepaßt und als neuronales Modul eingegliedert. In diesem Kontext kann gezeigt werden, daß das Kontur/Form-System durch diese Erweiterung in der Lage ist, Texturen, die sich durch Orientierungskontraste voneinander unterscheiden, im Zeitbereich zu separieren. Abbildung 5.0.1 zeigt eine Übersicht über das um ein Texturmodul erweiterte Kontur/Form-System.

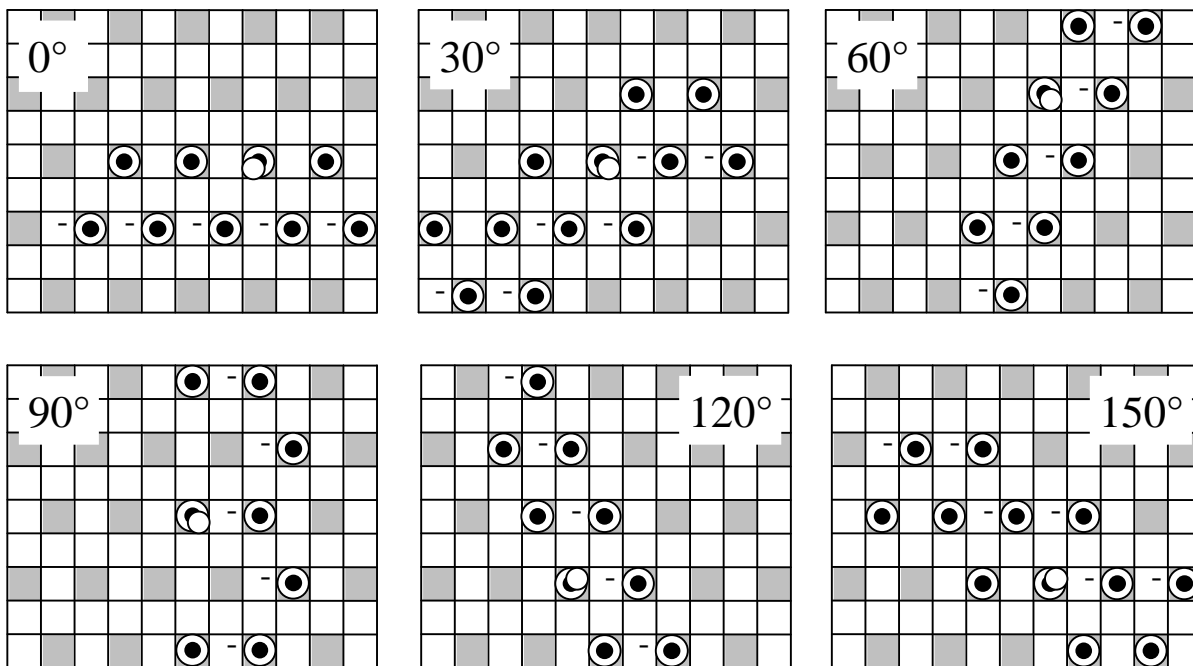


**Abb. 5.0.1:** *Erweitertes Kontur/Form-System.* Die Verarbeitungsstufen des bisher vorgestellten Kontur/Form-Systems sind durch weiße Kästchen dargestellt. Zusätzlich werden nun die Verarbeitungsstufen eines Texturmoduls eingegliedert (grau hinterlegte Kästchen). Sie gewinnen aus der Information der Kanten den Kanten-Orientierungskontrast, der zur Konturunterstützung an die Schichten der nachfolgenden Verarbeitungsstufen weitergegeben wird. Dieses Texturmodul ist ein Beispiel für die Integration der Verarbeitung weiterer Merkmalskontraste.

## 5.1 Detektion lokalen Orientierungskontrastes in künstlichen Texturen

### 5.1.1 Liniendetektion

Da das Netz auch Texturen verarbeiten können soll, die eine Orientierungsempfindlichkeit von  $15^\circ$  erfordern, kann der Satz von Kantendetektoren aus Kapitel 4, der nur  $30^\circ$ -Empfindlichkeit aufweist, nicht übernommen werden. Auf der anderen Seite handelt es sich bei den Testszenen um künstliche Texturen, die nur aus Linienelementen bestehen. Daher ist für diesen Unterabschnitt die Implementierung von zwölf relativ einfachen Liniendetektoren, die auf der Basis der in Kapitel 3 vorgestellten Vorverarbeitung arbeiten, ausreichend. Auf dem hexagonalen Abtastraster ( $d=2$ ) lassen sich zunächst nur Detektoren für sechs Vorzugsorientierungen ( $0^\circ, 30^\circ, \dots, 150^\circ$ ) mit stark überlappenden Orientierungscharakteristiken implementieren. Linien der übrigen sechs dazwischenliegenden Orientierungen werden im Nachhinein über die Antworten der Liniendetektoren der ersten sechs Orientierungen detektiert. Die Verschaltungen der Liniendetektoren der untersten Stufe gleichen denen der ON-ON-Kantendetektoren (Kap. 4.1.2), jedoch werden die Impulsraten der X-OFF statt der X-ON-Zellen als Eingangssignale und etwas anders geformte Teilbereiche der rezeptiven Felder verwendet. In Abbildung 5.1.1 sind die rezeptiven Felder schematisch dargestellt.



**Abb. 5.1.1:** Schematische Darstellung der rezeptiven Felder von Liniendetektoren der ersten Verarbeitungsstufe (LD1-Neuronen) für sechs verschiedene Orientierungen im hexagonalen Abtastraster der Abtastweite  $d=2$ . Die Position des Detektorneurons ist durch einen weißen Kreis dargestellt. Die konzentrischen Kreise geben die Zentrumspositionen der rezeptiven Felder derjenigen X-OFF-Zellen an, die auf das Detektorneuron verschaltet sind.

Die Stärke  ${}^{LD1}_{exz}W_{ijkl}^\varphi$  der Verbindungen zwischen einer exzitatorischen X-OFF-Zelle am Ort  $(k, l)$  und einem Liniendetektorneuron der Orientierung  $\varphi$  am Ort  $(i, j)$  des hexagonalen Rasters läßt sich beschreiben durch:

$${}^{LD1}_{exz}W_{ijkl}^{0^\circ} = \begin{cases} 1, & \text{falls } (k = i - 4 \vee k = i \pm 2 \vee k = i) \wedge l = j \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.1})$$

$${}^{LD1}_{exz}W_{ijkl}^{30^\circ} = \begin{cases} 1, & \text{falls } ((k = i \vee k = i - 2) \wedge l = j) \\ & \vee ((k = i - 5 \vee k = i - 3) \wedge l = j - 2) \\ & \vee ((k = i + 1 \vee k = i + 3) \wedge l = j + 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.2})$$

$${}^{LD1}_{exz}W_{ijkl}^{60^\circ} = \begin{cases} 1, & \text{falls } (k = i + v \wedge l = j + 2 \cdot v) \text{ mit } v \in \{-2, -1, 0, 1\} \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.3})$$

$${}^{LD1}_{exz}W_{ijkl}^{90^\circ} = \begin{cases} 1, & \text{falls } k = i \wedge (j = l \pm 4 \vee j = l) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.4})$$

Weiterhin ergibt sich die Stärke  ${}^{LD1}_{inh}W_{ijkl}^\varphi$  der Verbindungen zwischen einer inhibitorischen X-OFF-Zelle am Ort  $(k, l)$  und einem Liniendetektorneuron der Orientierung  $\varphi$  am Ort  $(i, j)$  des hexagonalen Rasters zu:

$${}^{LD1}_{inh}W_{ijkl}^{0^\circ} = \begin{cases} -1, & \text{falls } (k = i - 5 \vee k = i \pm 3 \vee k = i \pm 1) \wedge l = j - 2 \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.5})$$

$${}^{LD1}_{inh}W_{ijkl}^{30^\circ} = \begin{cases} -1, & \text{falls } ((k = i + 2 \vee k = i + 4) \wedge l = j) \\ & \vee ((k = i \pm 1) \wedge l = j - 2) \\ & \vee ((k = i - 4 \vee k = i - 2) \wedge (l = j - 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.6})$$

$${}^{LD1}_{inh}W_{ijkl}^{60^\circ} = \begin{cases} -1, & \text{falls } (k = i + v + 2 \wedge l = j + 2 \cdot v) \text{ mit } v \in \{-3, -2, -1, 0, 1\} \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.7})$$

$${}^{LD1}_{inh}W_{ijkl}^{90^\circ} = \begin{cases} -1, & \text{falls } (k = i + 2 \wedge (j = l \pm 4 \vee j = l)) \vee ((k = i + 3) \wedge (j = l \pm 2)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.8})$$

Die Verbindungsstärken der übrigen beiden Detektoren ergeben sich aus den Gewichten des 30°- bzw. 60°-Detektors durch Rotation. Die Parameter der Liniendetektoren der ersten Stufe sind in Tabelle 5.1.1 zusammengefaßt. Hier ist lediglich zu beachten, daß der exzitatorische



Teilbereich des rezeptiven Feldes eines 90°-Detektors nur drei X-Zellen berücksichtigt und daher der Verstärkungsfaktor  $V^{F1}$  entsprechend größer zu wählen ist.

$\varphi$	$V^\Theta$	$\tau^\Theta$	$\Theta_0$	$V^{F1}$	$\tau^{F1}$	$V^{F2}$	$\tau^{F2}$	$V^L$	$\tau^L$
0°	32	10 BIN	6	0.08	40 BIN	-0.08	40 BIN	0.7	40 BIN
30°	32	10 BIN	6	0.07	40 BIN	-0.07	40 BIN	0.7	40 BIN
60°	32	10 BIN	6	0.08	40 BIN	-0.06	40 BIN	0.7	40 BIN
90°	32	10 BIN	6	0.1	40 BIN	-0.05	40 BIN	0.7	40 BIN

**Tab. 5.1.1:** Parameter des Liniendetektorneurons der ersten Stufe.

Aus den bereits in Kapitel 4.1.4 genannten Gründen empfiehlt sich auch hier die laterale Kopplung von in Vorzugsorientierung coaxial angeordneten, benachbarten Detektorneuronen. Hier werden jedoch nur die beiden nächsten räumlichen Nachbarn berücksichtigt. Die Stärke  $^{link}w_{ijkl}^\varphi$  der Linking-Verbindung zwischen einem Liniendetektorneuron der Vorzugsorientierung  $\varphi$  am Ort  $(k, l)$  und einem Liniendetektorneuron derselben Vorzugsorientierung am Ort  $(i, j)$  ist gegeben durch:

$$^{link}w_{ijkl}^{0^\circ} = \begin{cases} 1, & \text{falls } (k = i \pm 2 \vee k = i \pm 4) \wedge l = j \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.9})$$

$$^{link}w_{ijkl}^{30^\circ} = \begin{cases} 1, & \text{falls } (k = i \pm 2 \wedge l = j) \\ & \vee ((k = i - 1 \vee k = i - 3) \wedge l = j - 2) \\ & \vee ((k = i + 1 \vee k = i + 3) \wedge l = j + 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.10})$$

$$^{link}w_{ijkl}^{60^\circ} = \begin{cases} 1, & \text{falls } (k = i + v \wedge l = j + 2 \cdot v) \text{ mit } v \in \{-2, -1, 1, 2\} \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.11})$$

$$^{link}w_{ijkl}^{90^\circ} = \begin{cases} 1, & \text{falls } (k = i \wedge l = j \pm 4) \vee (k = i \pm 1 \wedge l = j \pm 2) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.12})$$

Den bisher implementierten Detektorschichten werden aus Gründen der Netzgröße sechs Detektorschichten mit geringerer Abtastweite ( $d=4$ ) in einer zweiten Stufe nachgeschaltet. Diese Auflösungsverringerng wird erst hinter der ersten Stufe eingeführt, da die Liniendetektion der ersten Stufe durch eine grobe Auflösung sehr erschwert werden würde. Ein Liniendetektorneuron  $(m, n)$  der nachgeschalteten Schicht ist mit einem Liniendetektor-

neuron  $(i, j)$  derselben Vorzugsorientierung über Gewichte  ${}^{LD2}w_{mij}$  verbunden, deren Stärke sich berechnen läßt zu:

$${}^{LD2}w_{mij} = \begin{cases} 1, & \text{falls } ((i = m \vee i = m \pm 1 \vee i = m - 2) \wedge j = n) \\ & \vee ((i = m \pm 1 \vee i = m \pm 3) \wedge (j = n \pm 2)) \\ & \vee ((i = m \vee i = m + 2) \wedge (i = n \pm 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.13})$$

Diese Anordnung weicht geringfügig von einer Kreisform ab, erlaubt aber im Gegensatz zur Kreisform eine gleichberechtigte Behandlung aller Abtastpunkte auf dem hexagonalen Abtastraster. Ein Liniendetektorneuron der zweiten Stufe (*LD2-Neuron*) wird durch den in Tabelle 5.1.2 gezeigten Parametersatz bestimmt.

$V^\ominus$	$\tau^\ominus$	$\Theta_0$	$V^F$	$\tau^F$
128	1 BIN	2	1.0	10 BIN

Tab. 5.1.2: Parameter des Liniendetektorneurons der zweiten Stufe.

Jeder der bisher implementierten Detektoren antwortet auf eine Linie seiner Vorzugsorientierung und auf Linien, deren Orientierung bis etwa  $15^\circ$  von der Vorzugsorientierung abweicht (Abb. 5.1.2).

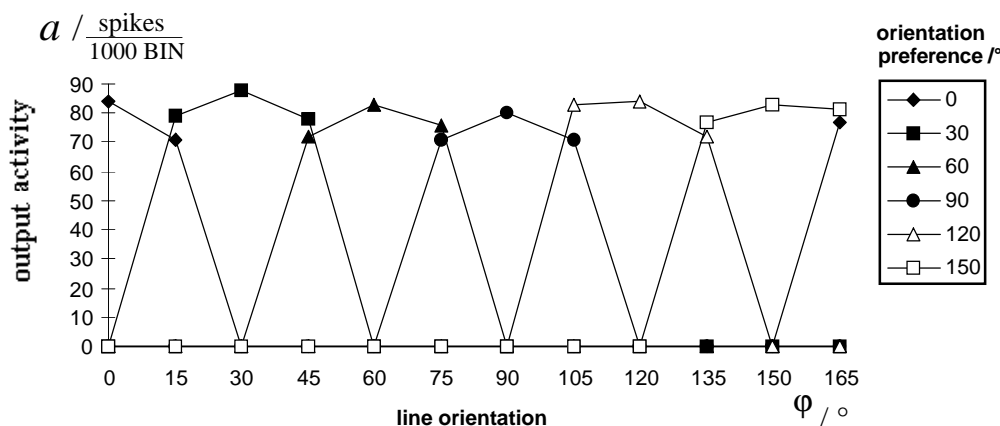
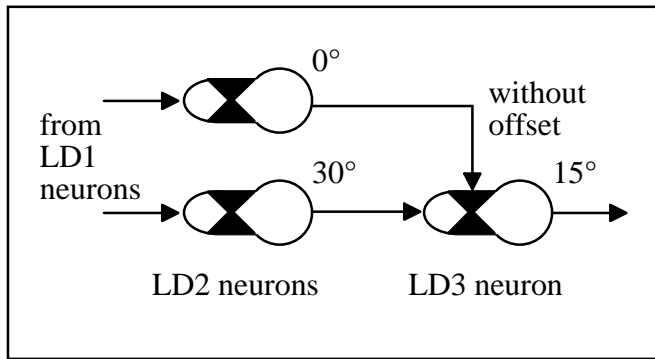


Abb. 5.1.2: Orientierungscharakteristiken der Liniendetektoren (*LD1*). Aufgetragen sind die Ausgangsraten der *LD1*-Neuronen bei Reizung mit einem optimal im jeweiligen rezeptiven Feld positionierten Linienelement der Orientierung  $\phi$ . Die Detektorneuronen antworten nur auf Linienelemente ihrer Vorzugsorientierung und auf Linienelemente deren Orientierung von der jeweiligen Vorzugsorientierung um  $15^\circ$  abweichen.

Antworten beispielsweise in einem Ort der  $0^\circ$ -Detektor und der  $30^\circ$ -Detektor, so muß von einer Linie mit einer Orientierung von etwa  $15^\circ$  ausgegangen werden. Dieses Prinzip nutzt die folgende Verschaltung zur Detektion von Linien der Orientierungen  $15^\circ, 45^\circ, \dots, 165^\circ$ . Den *LD2*-Schichten sind in der nächsten Stufe zwölf weitere Schichten (*LD3-Schichten*)

nachgeschaltet. Sei nun  $\Phi = \{15^\circ, 45^\circ, 75^\circ, 105^\circ, 135^\circ, 165^\circ\}$  die Menge aller "Zwischenorientierungen". Ein LD3-Neuron der Vorzugsorientierung  $\omega \in \Phi$  erhält Eingangssignale von zwei LD2-Neuronen desselben Ortes und der um  $15^\circ$  von  $\omega$  abweichenden Vorzugsorientierungen. Diese Eingangssignale werden im LD3-Neuron multiplikativ miteinander verrechnet, so daß das Neuron nur antwortet, wenn zuvor die beiden vorgeschalteten Neuronen (evtl. zeitlich versetzt) aktiv waren (Abb. 5.1.3).



**Abb. 5.1.3:** Multiplikative Verschaltung von LD2-Neuronen ähnlicher Vorzugsorientierung auf ein LD3-Neuron der mittleren Vorzugsorientierung. Diese Verschaltung ermöglicht die Kodierung der Linienorientierungen in zwölf statt in sechs Schichten unterschiedlicher Vorzugsorientierung

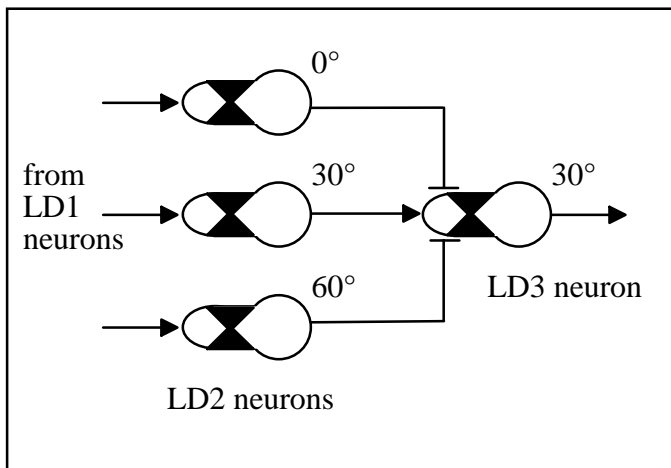
Die Stärke  ${}^{LD3a}_F W_{opmn}^{\omega\phi}$  der Verbindung von einem LD2-Neuron der Orientierung  $\phi$  am Ort  $(m, n)$  auf den Feeding-Eingang eines LD3-Neurons der Orientierung  $\omega$  am Ort  $(o, p)$  ergibt sich zu:

$${}^{LD3a}_F W_{opmn}^{\omega\phi} = \begin{cases} 1, & \text{falls } m = o \wedge n = p \wedge \phi = \omega + 15^\circ \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.14})$$

Ganz ähnlich läßt sich die Stärke  ${}^{LD3a}_L W_{opmn}^{\omega\phi}$  der Verbindung von einem LD2-Neuron der Orientierung  $\phi$  am Ort  $(m, n)$  auf den Linking-Eingang eines LD3-Neurons der Orientierung  $\omega$  am Ort  $(o, p)$  beschreiben durch:

$${}^{LD3a}_L W_{opmn}^{\omega\phi} = \begin{cases} 1, & \text{falls } m = o \wedge n = p \wedge \phi = \omega - 15^\circ \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.15})$$

Auf diese Weise werden sechs der zwölf LD3-Schichten implementiert. Die Schichten der übrigen sechs Orientierungen erhalten ihre Eingangssignale über Eins-zu-eins-Vorwärtsverschaltungen von den LD2-Schichten der jeweils gleichen Vorzugsorientierung. Um jedoch zu vermeiden, daß beispielsweise eine  $15^\circ$  orientierte Linie eine Antwort in den LD3-Schichten von  $0^\circ$  und  $30^\circ$  hervorruft, werden zusätzlich LD2-Neuronen desselben Ortes, aber ähnlicher Vorzugsorientierung inhibitorisch auf das LD3-Neuron verschaltet. Abbildung 5.1.4 zeigt das Beispiel eines  $30^\circ$  LD3-Neurons.



**Abb. 5.1.4:** Exzitatorische und inhibitorische Verschaltung von LD2-Neuronen auf LD3-Neuronen.

Die Stärke  ${}^{LD3b}_{exz} W_{opmn}$  der Verbindung von einem LD2-Neuron der Orientierung  $\phi$  am Ort  $(m, n)$  auf den exzitatorischen Feeding1-Eingang eines LD3-Neurons derselben Vorzugsorientierung am Ort  $(o, p)$  ergibt sich zu:

$${}^{LD3b}_{exz} W_{opmn} = \begin{cases} 1, & \text{falls } m = o \wedge n = p \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.16})$$

Außerdem folgt die Stärke  ${}^{LD3b}_{inh} W_{opmn}^{\omega\phi}$  der inhibitorischen Verbindung von einem LD2-Neuron der Orientierung  $\phi$  am Ort  $(m, n)$  auf den Feeding2-Eingang eines LD3-Neurons der Orientierung  $\omega$  am Ort  $(o, p)$  und einer gewissen Umgebung durch:

$${}^{LD3b}_{inh} W_{opmn}^{\omega\phi} = \begin{cases} -1, & \text{falls } (((m = o \pm 4 \vee m = o) \wedge n = p) \vee (m = o \pm 2 \wedge n = p \pm 4)) \\ & \wedge \phi = \omega \pm 30^\circ \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.17})$$

Auf diese Weise ergibt sich nun eine eindeutige Repräsentation der detektierten Linien in den zwölf Liniendetektionsschichten der dritten Stufe. Die Parameter der LD3-Neuronen sind in Tabelle 5.1.3 zusammengefaßt.

$\phi$	$V^\ominus$	$\tau^\ominus$	$\Theta_0$	$V^{F1}$	$\tau^{F1}$	$V^{F2}$	$\tau^{F2}$	$V^L$	$\tau^L$
$15^\circ, 45^\circ, \dots$	128	10 BIN	6	1.0	40 BIN	-	-	1.0	40 BIN*)
$0^\circ, 30^\circ, \dots$	128	10 BIN	6	2.0	40 BIN	7.0	10 BIN	-	-

**Tab. 5.1.3:** Parameter der Liniendetektorneuronen der dritten Stufe.

\*) ohne Offset +1

### 5.1.2 Räumliche Tiefpaßfilterung

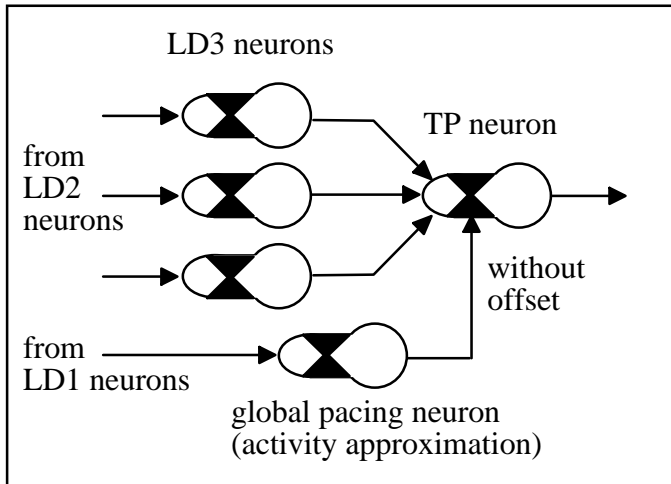
Die zwölf Liniendetektionsschichten geben ihre neuronalen Ausgangssignale an die zwölf Schichten der nachfolgenden Verarbeitungsstufe weiter. Hier wird eine ortstolerantere Repräsentation der detektierten Linien erzeugt, indem der Ausgang jedes Liniendetektors durch eine divergente Vorwärtsverschaltung mit den Neuronen einer bestimmten räumlichen kreisförmigen Umgebung der nachfolgenden Schicht (*TP-Schicht*) verbunden wird. Da dabei die Gewichte Verteilung näherungsweise einer Gaußmaske entspricht, wird somit eine räumliche Tiefpaßfilterung durchgeführt. Die Motivation für dieses Vorgehen liegt darin, daß Textur zwar lokale Eigenschaften besitzt, aber auch als Fläche auftritt. Vor allem letztere Eigenschaft wird von den TP-Neuronen durch diese Tiefpaßfilterung kodiert. Sei  $H_4$  die Menge aller Orte auf dem hexagonalen Raster der Abtastweite  $d=4$ , dann ergibt sich die Stärke  ${}^{TP}w_{qrop}$  der Verbindungen von einem LD3-Neuron ( $o, p$ ) auf den Feeding-Eingang eines TP-Neurons ( $q, r$ ) derselben Vorzugsorientierung zu:

$${}^{TP}w_{qrop} = \begin{cases} 2.0, & \text{falls } o = q + \mu \wedge p = r + \nu \text{ mit } (q, r) \in H_4 \text{ und } \sqrt{\mu^2 + \nu^2} \leq 5 \\ 1.0, & \text{falls } o = q + \mu \wedge p = r + \nu \text{ mit } (q, r) \in H_4 \text{ und } 5 < \sqrt{\mu^2 + \nu^2} \leq 10 \\ 0.2, & \text{falls } o = q + \mu \wedge p = r + \nu \text{ mit } (q, r) \in H_4 \text{ und } 10 < \sqrt{\mu^2 + \nu^2} \leq 14 \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.18})$$

Für die weitere Verarbeitung ist es vorteilhaft, wenn in jedem Ort höchstens ein TP-Neuron aktiv ist, da Mehrfachaktivität die nachfolgende Orientierungskontrast-Detektion verfälschen könnte, wie im nächsten Unterkapitel noch deutlich werden wird. Durch eine wechselseitige inhibitorische Verschaltung der 12 TP-Neuronen in jedem Ort wird erreicht, daß das am stärksten aktivierte Neuron die Aktivitäten der übrigen Neuronen im selben Ort vollständig unterdrückt. Dazu ist der Ausgang eines TP-Neurons ( $q, r$ ) mit dem Feeding-Eingang eines TP-Neurons desselben Orts, aber unterschiedlicher Vorzugsorientierung inhibitorisch (Gewichtung: -7.5) verbunden.

Ein TP-Neuron, das im Zentrum eines Linienelements lokalisiert ist, wird, bedingt durch die divergente Verschaltung, die Signale von mehr LD-Neuronen erhalten als ein TP-Neuron, das sich etwa am Ort eines Linieneendes befindet. Zur Vermeidung von Verfälschungen in der nachfolgenden Orientierungskontrastdetektion ist aber eine gleichmäßige Aktivierung aller Neuronen der kreisförmigen Umgebung wünschenswert. Daher wird ein einfacher Mechanismus zur räumlichen Angleichung der Raten eingeführt. Ein globales Neuron (*Takt-Neuron*), das über Linking-Verbindungen sein Ausgangssignal an alle TP-Neuronen weitergibt, wird in regelmäßigen Zeitabständen aktiv. Durch eine multiplikative Verschaltung (Linking ohne Offset) wird gewährleistet, daß TP-Neuronen nur dann aktiv werden können, wenn das globale Neuron zuvor aktiv war. Um einen solchen Takt vorgeben zu können, benötigt das globale Takt-Neuron seinerseits Feeding-Eingangssignale, die es von den

Liniendetektorschichten durch Einsammeln der Ausgangssignale aller LD1-Neuronen erhält. Abbildung 5.1.5 zeigt die Verschaltung für ein TP-Neuron. Die Parameter der in dieser Stufe verwendeten Neuronentypen sind in Tabelle 5.1.4 zusammengefaßt.



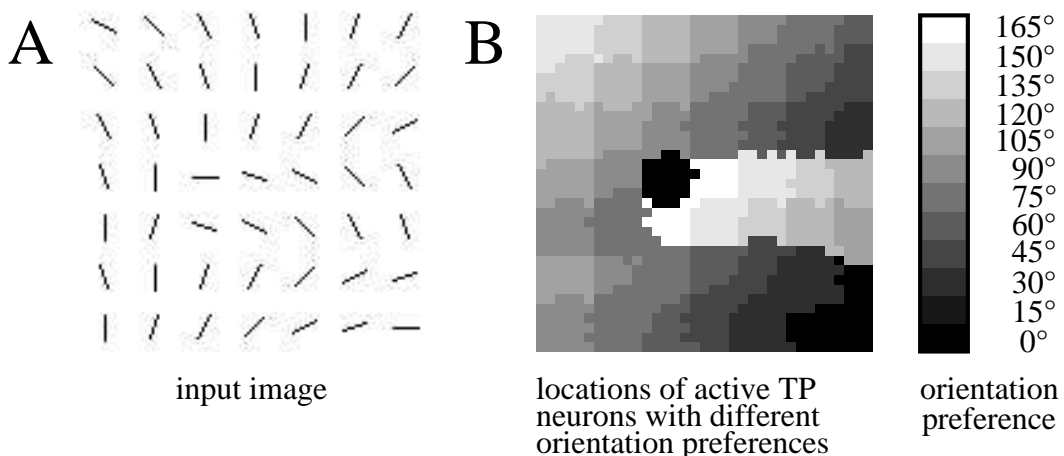
**Abb. 5.1.5:** Multiplikative Verschaltung von LD3-Neuronen und dem globalen Takt-Neuron auf ein TP-Neuron. Die Taktvorgabe ermöglicht eine Angleichung der Ausgangsraten von TP-Neuronen, die für die nachfolgende Orientierungs-kontrast-detektion hilfreich ist.

	$V^\theta$	$\tau^\theta$	$\Theta_0$	$V^{F1}$	$\tau^{F1}$	$V^L$	$\tau^L$
TP-Neuron	128	1.0 BIN	2	1.0	40 BIN	7.5	0.5 BIN *)
globales Takt-Neuron	512	10 BIN	6	0.2	10 BIN	-	-

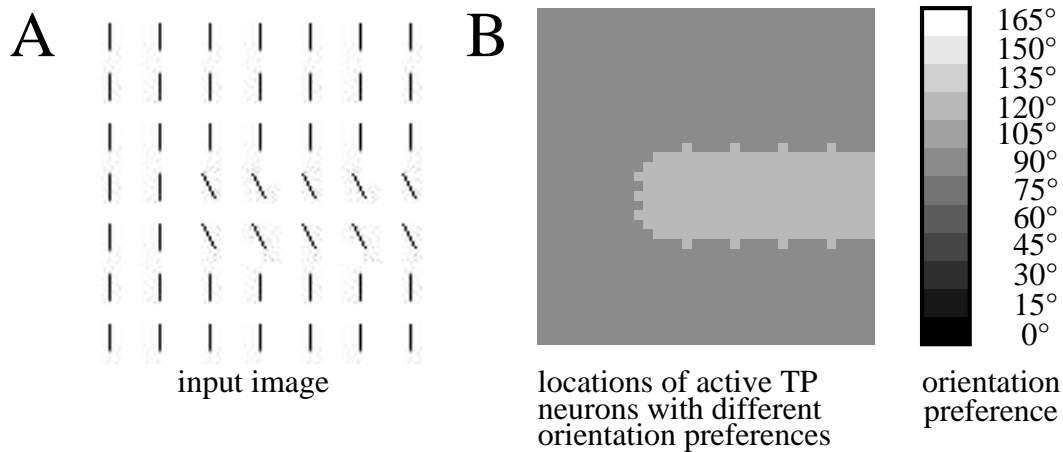
**Tab. 5.1.4:** Parameter der Neuronen der TP- Stufe.

\*) ohne Offset +1

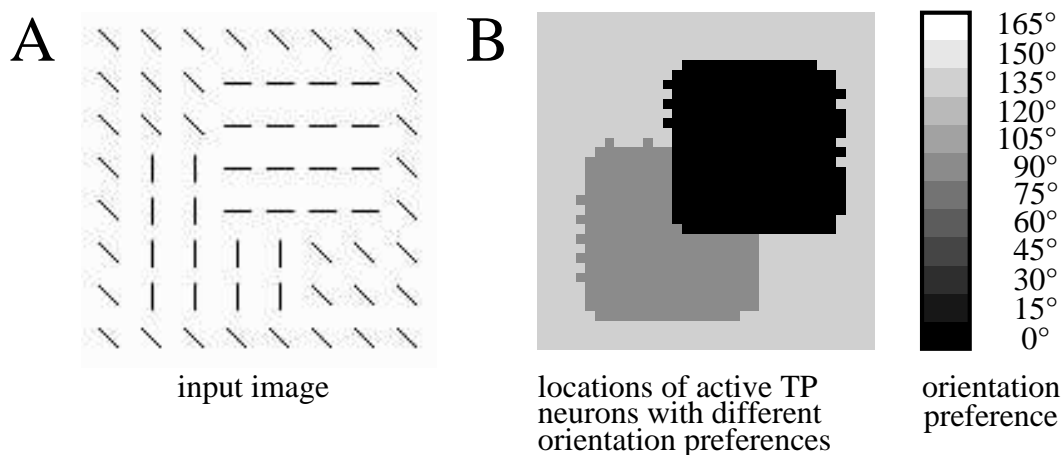
Im folgenden sind verschiedene Simulationsbeispiele aufgeführt. Abbildung 5.1.6 zeigt eine künstliche Textur (140 · 140 Pixel), die aus verschiedenen Linienelementen zusammengesetzt ist und in der menschliche Betrachter einen horizontalen "Balken" wahrnehmen. In den zwölf TP Schichten (jeweils 35 · 35 Neuronen) werden die Linienelemente durch Bereiche neuronaler Aktivitäten je nach Orientierung kodiert.



**Abb. 5.1.6:** Antwortverhalten aller TP-Neuronen bei Verarbeitung eines künstlichen Texturbeispiels. **A:** Die aus orientierten Linienelementen bestehende Textur wird als Eingangsbild verwendet (nach: Nothdurft 1993b). **B:** In den TP-Schichten werden die Linienelemente durch Bereiche neuronaler Aktivitäten je nach Orientierung kodiert.



**Abb. 5.1.7:** Antwortverhalten aller TP-Neuronen bei Verarbeitung eines künstlichen Texturbeispiels. **A:** Auch diese aus orientierten Linienelementen bestehende Textur wird als Eingangsbild verwendet (nach: Nothdurft 1993b). **B:** In den TP Schichten werden die Linienelemente durch Bereiche neuronaler Aktivitäten je nach Orientierung kodiert.



**Abb. 5.1.8:** Antwortverhalten aller TP-Neuronen bei Verarbeitung eines weiteren künstlichen Texturbeispiels. **A:** Als Eingangsbild wird diese aus orientierten Linienelementen bestehende Textur verwendet. Menschliche Betrachter nehmen hier zwei sich teilweise überdeckende Quadrate wahr. **B:** In den TP Schichten führt dies zu größeren Bereichen neuronaler Aktivitäten, die die Linienelemente je nach Orientierung kodieren.

Da die Orientierungen benachbarter Linienelemente sowohl im Hintergrund ( $15^\circ$ ) als auch entlang des wahrgenommenen Balkenrandes ( $90^\circ$ ) voneinander abweichen, treten keine größeren Bereiche neuronaler Aktivität in einer TP-Schicht auf. Größere zusammenhängende Bereiche zeigen die TP-Schichten bei Verarbeitung der in den Abbildungen 5.1.7 und 5.1.8 gezeigten künstlichen Texturen. In Abbildung 5.1.7 A nimmt der Betrachter ebenso einen horizontalen Balken und in Abbildung 5.1.8 A zwei Quadrate wahr, von denen eines teilweise überdeckt wird.

### 5.1.3 Orientierungskontrastdetektion

Die in den zwölf TP-Schichten vorliegenden Linienrepräsentationen werden in der folgenden Verarbeitungsstufe, die ebenso zwölf Schichten enthält, genutzt. In dieser Stufe wird prinzipiell eine modifizierte Filterung der Linienrepräsentationen mit einer Mexikanerhut-Funktion (Center/Surround-Antagonismus) durch Vorwärtsverbindungen durchgeführt. Innerhalb jeder TP-Schicht liegen in der Regel Aktivitätsbereiche vor, die (evtl. mehrere) Linien repräsentieren. Um die Grenzen dieser Repräsentationsbereiche zu detektieren, wird für jede der Vorzugsorientierungen getrennt eine Mexikanerhutfilterung durchgeführt. Auf diese Weise werden Bereiche definiert, in denen nur Linienelemente einer bestimmten Orientierung auftreten. Die Modifikation besteht nun darin, daß die Signale der im peripheren (inhibitorischen) Bereich des Mexikanerhuts befindlichen Neuronen mit ähnlicher Vorzugsorientierung bei der Berechnung des lokalen Faltungsprodukts (*lokaler Orientierungskontrast*) berücksichtigt werden. Dadurch können unmittelbar benachbarte Linienelemente, die eine ähnliche Vorzugsorientierung aufweisen, das lokale Faltungsergebnis und damit die lokale Repräsentationsstärke der Bereichsgrenzen verringern. Dort, wo sehr unterschiedlich orientierte Linienelemente aneinandergrenzen, ist eine solche Unterdrückung nicht möglich, und der lokale Orientierungskontrast bleibt unverändert. Die dazu notwendigen Verbindungen werden durch die folgenden Gewichte beschrieben, die einer Mexikanerhut-ähnlichen Funktion entsprechen. Somit ergibt sich die Stärke der exzitatorischen Verbindungen von einem Neuron  $(q, r)$  der TP-Schicht mit Vorzugsorientierung  $\varphi$  auf den Feeding1-Eingang eines Neurons  $(s, t)$  der Orientierungskontrast-Schicht (*OK-Schicht*) mit selber Vorzugsorientierung zu:

$${}_{center}^{OK} w_{stqr} = \begin{cases} 7.0, & \text{falls } q = s \wedge r = t \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.19})$$

Die inhibitorische Gewichtung über den Feeding2-Eingang wird beschrieben durch:

$${}_{surround}^{OK} w_{stqr} = \begin{cases} -0.2, & \text{falls } q = s + \mu \wedge r = t + \nu \\ & \text{mit } (s, t) \in H_4 \text{ und } 5 < \sqrt{\mu^2 + \nu^2} \leq 14 \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.20})$$

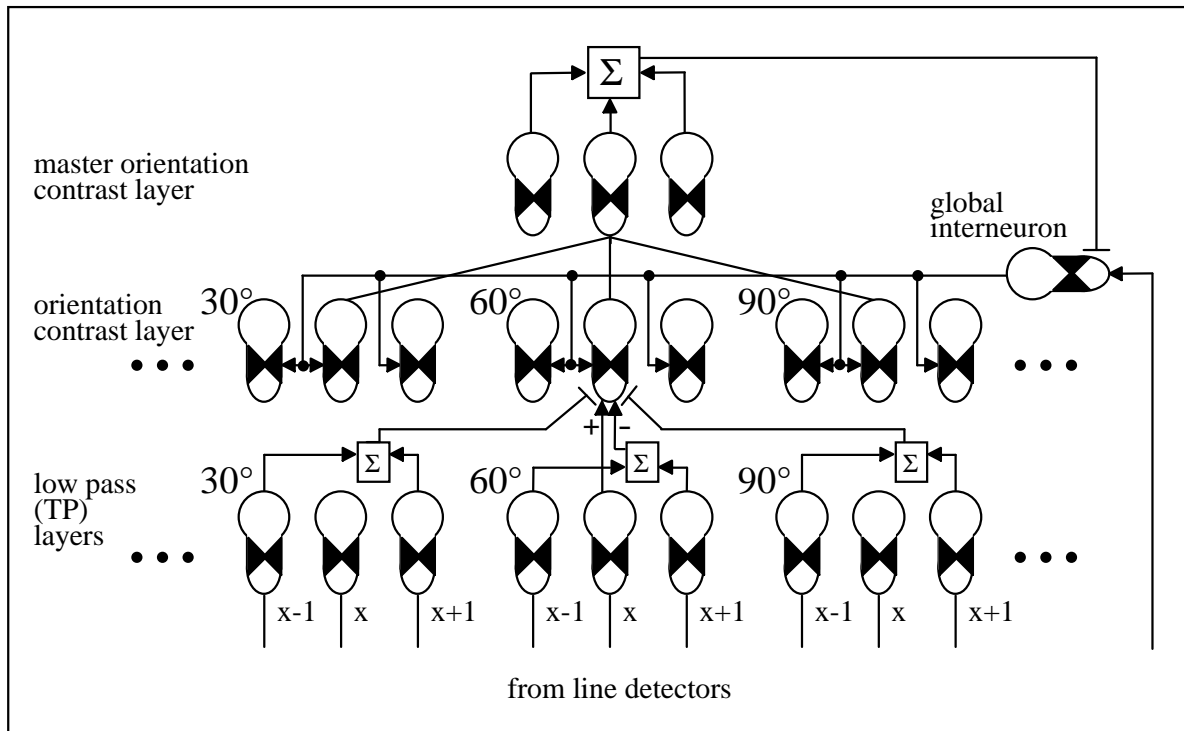
Durch diese Gewichtung ist gewährleistet, daß bei einer Multiplikation der resultierenden Mexikanerhut-ähnlichen Funktion (DOOG) mit einer einheitlichen Fläche das Produkt näherungsweise verschwindet. Die Stärke der Verbindungen von einem Neuron  $(q, r)$  der TP-Schicht mit Vorzugsorientierung  $\varphi$  auf den inhibitorischen Eingang eines Neurons  $(s, t)$  der OK-Schicht mit Vorzugsorientierung  $\omega$  ergibt sich zu:



$$\text{OK}_{\text{nachbar}} W_{\text{stgr}}^{\omega\phi} = \begin{cases} 0.15, & \text{falls } q = s + \mu \wedge r = t + \nu \wedge \phi = \omega \pm 15^\circ \\ & \text{mit } (s, t) \in H_4 \text{ und } 5 < \sqrt{\mu^2 + \nu^2} \leq 14 \\ 0.1, & \text{falls } q = s + \mu \wedge r = t + \nu \wedge \phi = \omega \pm 30^\circ \\ & \text{mit } (s, t) \in H_4 \text{ und } 5 < \sqrt{\mu^2 + \nu^2} \leq 14 \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.21})$$

Schließlich werden die Ausgangsaktivitäten aller zwölf OK-Schichten auf eine gemeinsame Orientierungskontrast-Master-Schicht gegeben, die den gesamten Orientierungskontrast repräsentiert. Das Gewicht der entsprechenden Eins-zu-eins-Verbindung von einem Neuron der OK-Schichten auf den Feeding1-Eingang eines Master-Neurons desselben Ortes beträgt eins.

Die Master-OK-Schicht repräsentiert durch starken lokalen Orientierungskontrast möglicherweise die Konturen von Objekten wie etwa in dem Texturbeispiel der Abbildung 5.1.6 A. Hier ist es sinnvoll, schwachen Orientierungskontrast durch starken im Bild zu unterdrücken, so daß nur die Konturen des durch den stärksten Kontrast definierten Objekts repräsentiert werden. Dazu wird eine Verschaltung implementiert, die prinzipiell einer inhibitorischen globalen Rückkopplung von der Master-OK-Schicht auf die OK-Schichten entspricht, jedoch die unterschiedliche Funktion von inhibitorischem Feeding- und Inhibitions-Eingang ausnutzt und damit gewährleistet, daß sich starker Kontrast durch Rückkopplung nicht selbst auslöschen kann. Ähnlich dem globalen Takt-Neuron wird auch hier ein globales Interneuron eingeführt, das jedoch nicht die Aufgabe hat, Raten anzugleichen, sondern die Aktivierung der OK-Neuronen über Linking-Verbindungen zu erhöhen. Seine Eingangssignale erhält es aus den LD-Schichten, und es wird von der Aktivität in der Master-OK-Schicht inhibiert. Die entsprechende Verschaltung ist in Abbildung 5.1.9 skizziert. Tritt besonders starke Aktivität in der Master-OK-Schicht auf, so antwortet das Zwischenneuron nicht mehr, und die Unterstützung der OK-Schichten entfällt, so daß schwache Kontraste nicht mehr repräsentiert werden. Besonders stark wirkt sich das Ausfallen der unterstützenden Linking-Signale auf die Inhibition von benachbarten Schichten aus, da im OK-Neuron das Potential des inhibitorischen Eingangs im Gegensatz zum exzitatorischen Feeding-Potential nicht durch den Ausfall des Linking-Potentials verringert wird. Die Hemmung durch ähnlich orientierte Linien ist somit besonders effektiv und schwache Kontrastkanten, die typischerweise zwischen Linienelementen ähnlicher Vorzugsorientierung auftreten, werden nicht mehr detektiert. Solange aber nur wenig Aktivität in der Master-OK-Schicht auftritt, bleibt das globale Interneuron aktiv und, da der Feeding-Eingang durch das modulatorische Linking-Potential entsprechend starke Wirkung erhält, kann der inhibitorische Eingang nur wenig Einfluß nehmen, so daß die Repräsentation einer schwachen Orientierungskontraste kaum gehemmt wird.

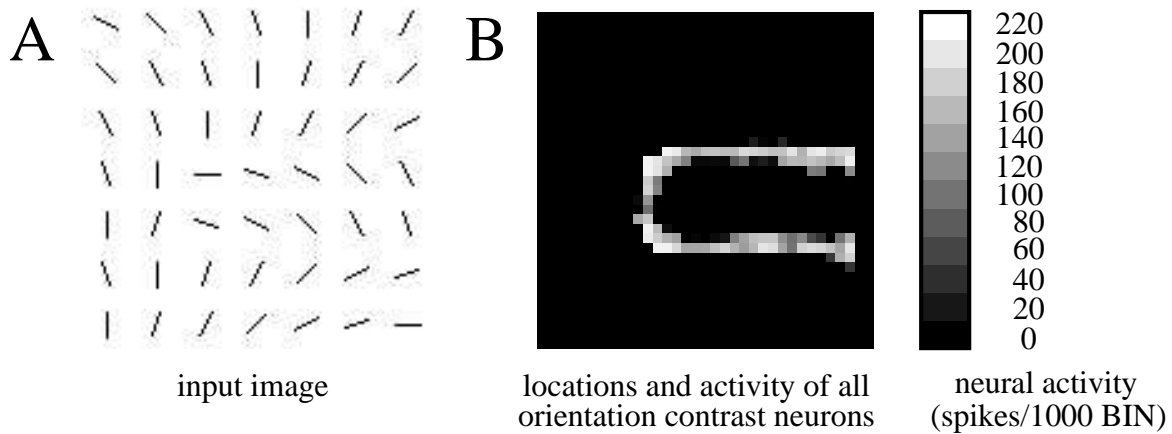


**Abb. 5.1.9:** Verschaltung der TP- und Orientierungskontrast-Neuronen. Die Orientierungskontrast-Detektion basiert auf einer Verschaltung, die einem räumlichen Center/Surround-Antagonismus entspricht und die exemplarisch für ein 60° OK-Neuron des Ortes  $x$  hier skizziert ist. Der über ein globales Interneuron implementierte Rückkopplungsmechanismus unterdrückt die Repräsentationen von schwachem Kontrast, wenn im zu verarbeitenden Bild bereits genügend starker Kontrast vorhanden ist.

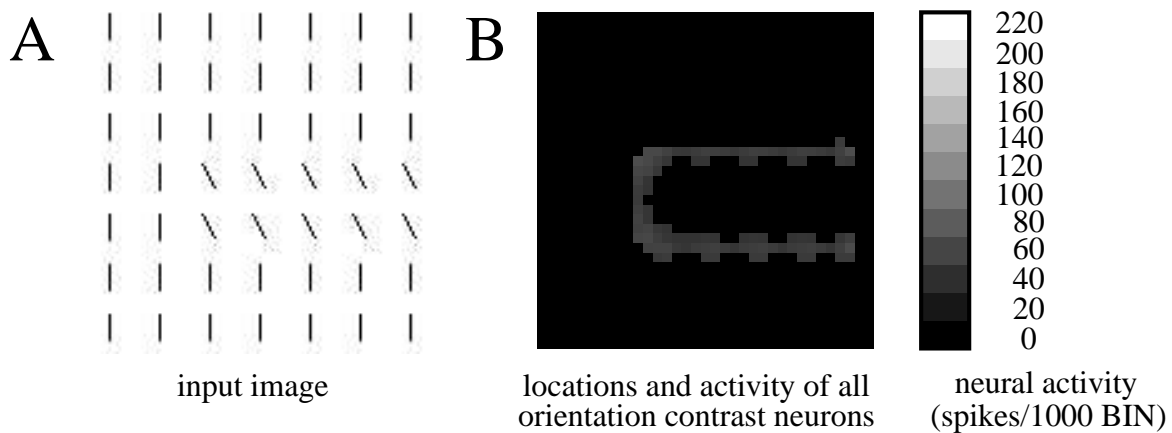
Da schwacher und starker Orientierungskontrast getrennt behandelt werden, wird je nach vorhandener Gesamtaktivität in der Master-OK-Schicht, (ausschließlich) Aktivität unterdrückt, die schwachen Orientierungskontrast repräsentiert. Das ist insbesondere dann hilfreich, wenn Objektkonturen durch starken Kontrast repräsentiert werden. Werden die Konturen der Objekte nur durch schwache Kontraste definiert, dann wird die Aktivität nicht gehemmt. Insgesamt liegt also ein Mechanismus vor, der nur die Repräsentation des stärksten im Bild vorhandenen Orientierungskontrasts erlaubt. Die Parameter der in dieser Stufe verwendeten Neuronentypen sind in Tabelle 5.1.5 zusammengestellt. In den Abbildungen 5.1.10 bis 5.1.12 wird das Antwortverhalten der OK-Schichten (jeweils  $35 \cdot 35$  Neuronen) bei Verarbeitung der bekannten Texturbeispiele anhand ihrer Ausgangsraten gezeigt.

	$V^\Theta$	$\tau^\Theta$	$\Theta_0$	$V^{F1}$	$\tau^{F1}$	$V^{F2}$	$\tau^{F2}$	$V^L$	$\tau^L$	$V^{Inh}$	$\tau^{Inh}$
OK-Neuron	32	1.0 BIN	2	1.0	40 BIN	1.0	40 BIN	7.5	5.0 BIN	1.0	40 BIN
Master-OK-Neuron	32	1.0 BIN	2	1.5	40 BIN	-	-	-	-	-	-
globales Exz.-Neuron	512	10 BIN	6	1.0	0.5 BIN	-	-	-	-	1.0	5.0 BIN

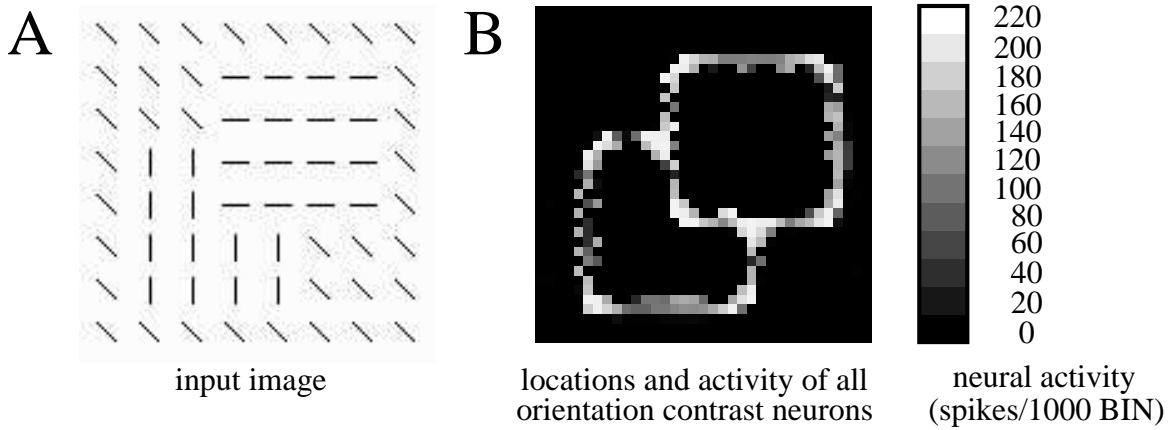
**Tab. 5.1.5:** Parameter der Neuronen der Orientierungskontrast-Stufe.



**Abb. 5.1.10:** Antwortverhalten aller Orientierungskontrast-Neuronen bei Verarbeitung eines künstlichen Texturbeispiels mit starkem und schwachem Orientierungskontrast. **A:** In dem bereits aus Abb. 5.1.6 bekannten Stimulus nehmen menschliche Betrachter einen horizontalen "Balken" wahr. Während die Orientierungen unmittelbar benachbarter Linienelemente im Hintergrund nur wenig voneinander abweichen ( $15^\circ$ ), treten entlang des wahrgenommenen Balkenrandes starke Orientierungssprünge ( $90^\circ$ ) auf (nach: Nothdurft 1993b). **B:** Die OK-Neuronen der zwölf Schichten zeigen den stärksten Kontrast an, da die Repräsentationen des schwachen Hintergrundkontrastes unterdrückt werden. Das Ergebnis entspricht der menschlichen Wahrnehmung.



**Abb. 5.1.11:** Antwortverhalten aller Orientierungskontrast-Neuronen bei Verarbeitung eines künstlichen Texturbeispiels mit schwachem Orientierungskontrast. **A:** In dem bereits aus Abb. 5.1.7 bekannten Stimulus nimmt der Betrachter einen horizontalen "Balken" wahr, der jedoch im Gegensatz zum vorangegangenen Beispiel durch schwachen Orientierungskontrast definiert wird (nach: Nothdurft 1993b). **B:** Die Repräsentation des schwachen Kontrastes wird nicht unterdrückt, und die OK-Neuronen zeigen die Konturen des Balkens an. Auch dieses Ergebnis entspricht der menschlichen Wahrnehmung.



**Abb. 5.1.12:** Antwortverhalten aller Orientierungskontrast-Neuronen bei Verarbeitung eines künstlichen Texturbeispiels mit starkem Orientierungskontrast. **A:** In dem bereits aus Abb. 5.1.8 bekannten Stimulus nimmt der Betrachter zwei Rechtecke wahr, von denen eines scheinbar teilweise überdeckt wird. **B:** Die Konturen der wahrgenommenen Objekte werden auch von den OK-Neuronen repräsentiert.

#### 5.1.4 Liniendetektion auf Grundlage von Orientierungskontrast

Wie im vorangegangenen Abschnitt deutlich wurde, liefert die Master-OK-Schicht bereits das Ergebnis des Texturanalyseverfahrens. Um dieses Ergebnis zur weiteren Verarbeitung an die sechs Synchronisations- und zwölf End-Stop-Schichten unterschiedlicher Vorzugsorientierungen des Kontur/Form-Systems weiterleiten zu können, muß zuvor eine Liniendetektion auf dem Analyseergebnis durchgeführt werden. Daher werden der Master-OK-Schicht sechs neuronale Schichten (*LDOK-Schichten*) der Vorzugsorientierungen von  $0^\circ$  bis  $150^\circ$  (in  $30^\circ$ -Schritten) nachgeschaltet. Abb. 5.1.13 zeigt schematisch die Verschaltungsgewichte der Detektoren.

Die Stärke  ${}_{exz}^{LDOK}W_{xyuv}^\varphi$  der exzitatorischen Verbindungen von einem Master-OK-Neuron am Ort  $(u, v)$  zum Feeding1-Eingang eines Liniendetektorneurons der Vorzugsorientierung  $\varphi$  am Ort  $(x, y)$  ist bei Abtastweite  $d=2$  gegeben durch:

$${}_{exz}^{LDOK}W_{xyuv}^{0^\circ} = \begin{cases} 1, & \text{falls } ((u = x \pm 2 \vee u = x) \wedge (v = y)) \\ & \vee ((u = x \pm 1 \vee u = x + 3) \wedge (v = y - 2)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.22})$$

$${}_{exz}^{LDOK}W_{xyuv}^{30^\circ} = \begin{cases} 1, & \text{falls } ((u = x \vee u = x \pm 2) \wedge v = y) \vee (u = x - 1 \wedge v = y - 2) \\ & \vee (u = x + 1 \wedge v = y + 2) \\ 0.5, & \text{falls } ((u = x - 3 \wedge v = y - 2) \vee (u = x + 3 \wedge v = y + 2)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.23})$$

$${}^{LDOK}_{exz} W_{xyuv}^{60^\circ} = \begin{cases} 1, & \text{falls } (u = x + \mu \wedge v = y + 2 \cdot \mu \text{ mit } \mu \in \{-1, 0, 1\}) \\ & \vee (u = x + \mu + 2 \wedge v = y + 2 \cdot \mu \text{ mit } \mu \in \{-2, -1, 0\}) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.24})$$

$${}^{LDOK}_{exz} W_{xyuv}^{90^\circ} = \begin{cases} 1, & \text{falls } (u = x \pm 1 \wedge v = y \pm 2) \vee (u = x \wedge v = y) \\ 0.5, & \text{falls } (u = x \wedge v = y \pm 4) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.25})$$

Die inhibitorischen Verbindungen zum Feeding2-Eingang des entsprechenden Liniendetektor-Neurons lassen sich beschreiben mit:

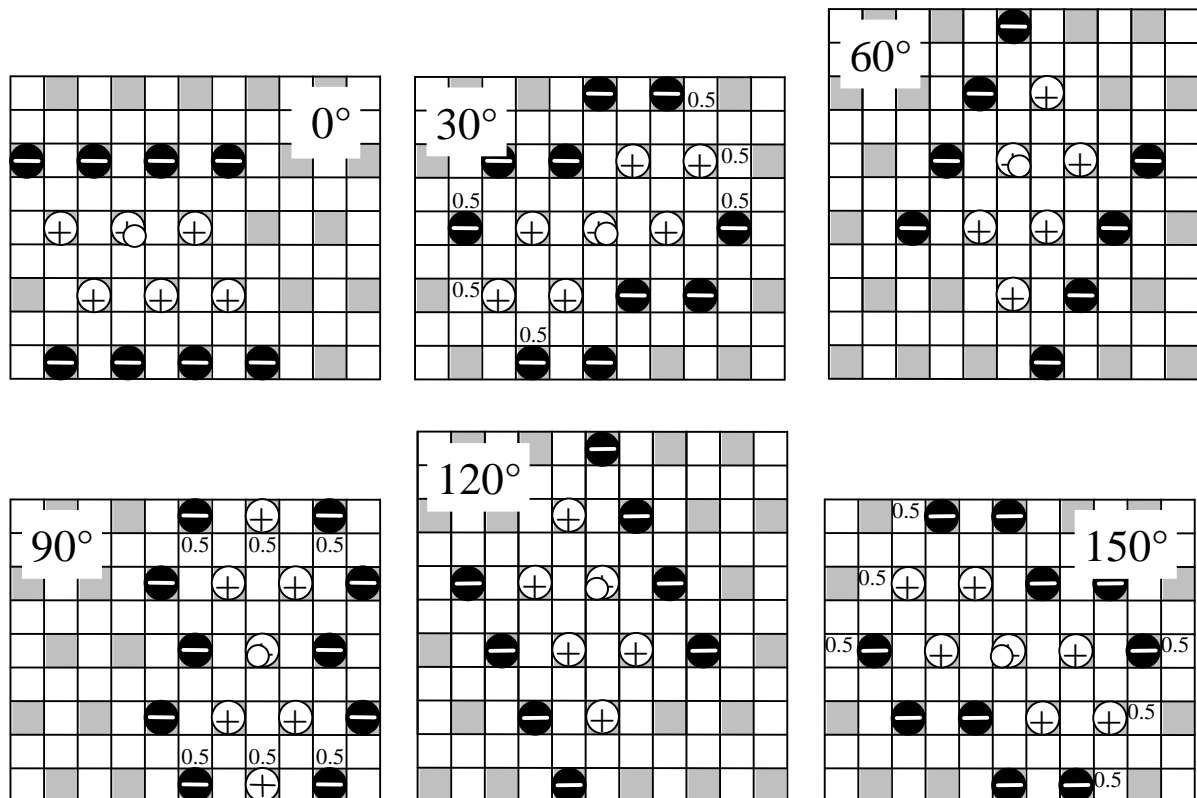
$${}^{LDOK}_{inh} W_{xyuv}^{0^\circ} = \begin{cases} -1, & \text{falls } ((u = x \pm 3 \vee u = x \pm 1) \wedge (v = y + 2)) \\ & \vee ((u = x \pm 2 \vee u = x \vee u = x + 4) \wedge (v = y - 4)) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.26})$$

$${}^{LDOK}_{inh} W_{xyuv}^{30^\circ} = \begin{cases} -1, & \text{falls } ((u = x - 1 \vee u = x - 3) \wedge v = y + 2) \vee (u = x \wedge v = y + 4) \\ & \vee ((u = x + 1 \vee u = x + 3) \wedge v = y - 2) \vee (u = x \wedge v = y - 4) \\ -0.5, & \text{falls } ((u = x \pm 4 \wedge v = y) \vee (u = x + 2 \wedge v = y + 4)) \\ & \vee (u = x - 2 \wedge v = y - 4) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.27})$$

$${}^{LDOK}_{inh} W_{xyuv}^{60^\circ} = \begin{cases} -1, & \text{falls } (u = x + \mu - 2 \wedge v = y + 2 \cdot \mu \text{ mit } \mu \in \{-1, 0, 1, 2\}) \\ & \vee (u = x + \mu + 4 \wedge v = y + 2 \cdot \mu \text{ mit } \mu \in \{-3, -2, -1, 0\}) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.28})$$

$${}^{LDOK}_{inh} W_{xyuv}^{90^\circ} = \begin{cases} -1, & \text{falls } (u = x \pm 2 \wedge v = y) \vee (u = x \pm 3 \wedge v = y \pm 2) \\ -0.5, & \text{falls } (u = x \pm 2 \wedge v = y \pm 4) \\ 0, & \text{sonst} \end{cases} \quad (\text{Gl. 5.1.29})$$

In den folgenden Simulationsbeispielen wird jedoch die Abtastweite  $d=4$  für die Master-OK- und LDOK-Schichten verwendet, so daß die jeweiligen Abstände in Abbildung 5.1.13 und in den Gleichungen 5.1.22 bis 5.1.29 zu verdoppeln sind. Weiterhin werden Linking-Verbindungen eingeführt, die sich durch die Gleichungen 4.1.22 beschreiben lassen. Auch hier sind lediglich die Abstände zu verdoppeln. Die Parameter der LDOK-Neuronen sind in Tabelle 5.1.6 zusammengefaßt. Einen Überblick über das gesamte Netz zur Texturanalyse aufgrund der Detektion von lokalem Orientierungskontrast gibt die Abbildung 5.1.14.

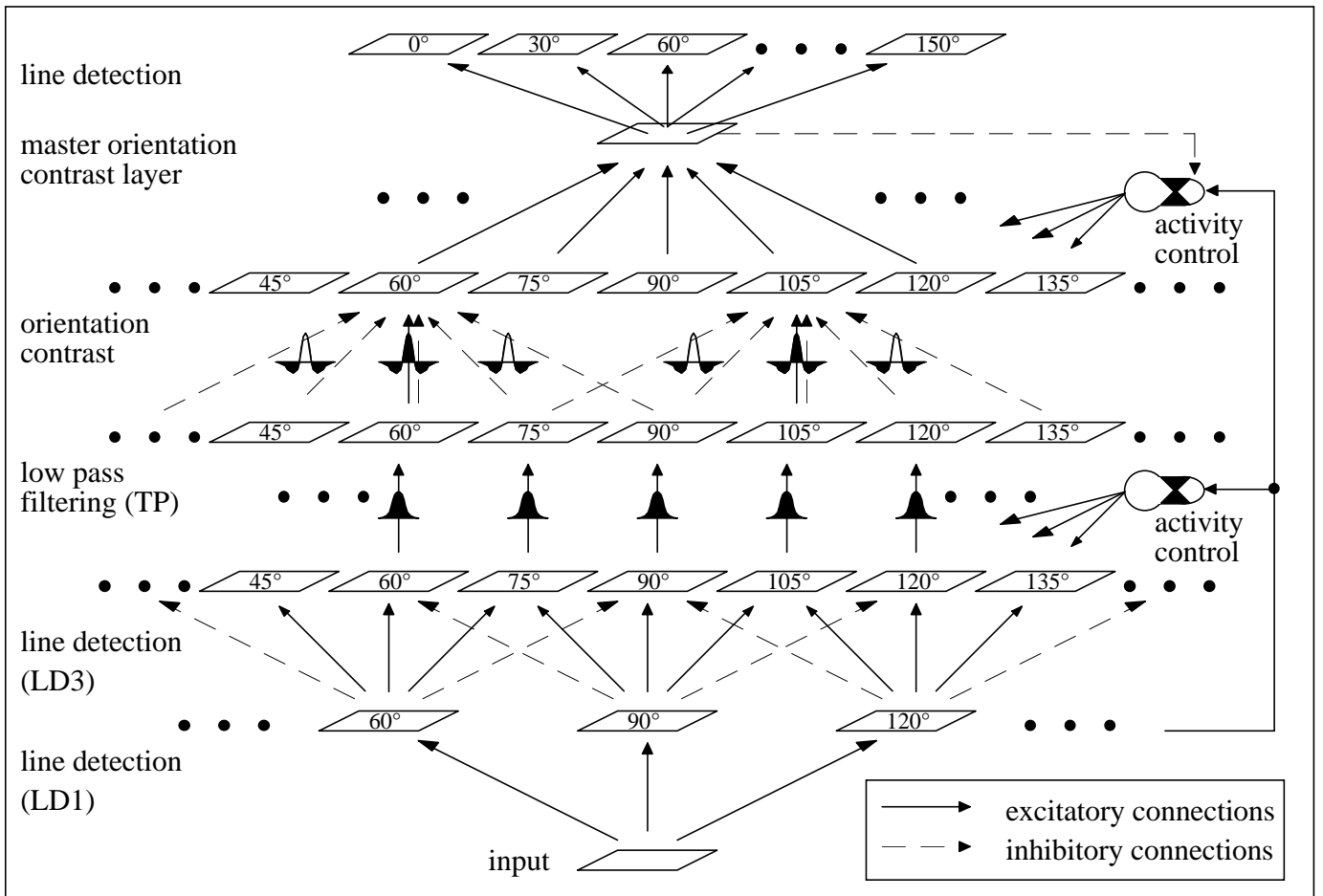


**Abb. 5.1.13:** Schematische Darstellung der Verbindungen von Master-Orientierungskontrast-Neuronen zu einem nachgeschalteten Liniendetektorneuron auf dem hexagonalen Gitter mit Abtastweite  $d=2$ . Die Position des Detektorneurons ist durch einen kleinen weißen Kreis markiert. Inhibitorische Gewichte werden durch schwarze gefüllte Kreise mit weißem "Minus"-Zeichen, exzitatorische durch weiße Kreise mit "Plus"-Zeichen gekennzeichnet. Gewichte der halben Stärke sind mit "0.5" versehen.

$V^\Theta$	$\tau^\Theta$	$\Theta_0$	$V^{F1}$	$\tau^{F1}$	$V^{F2}$	$\tau^{F2}$	$V^L$	$\tau^L$
32	10 BIN	6	0.5	40 BIN	0.6	40 BIN	0.5	5 BIN

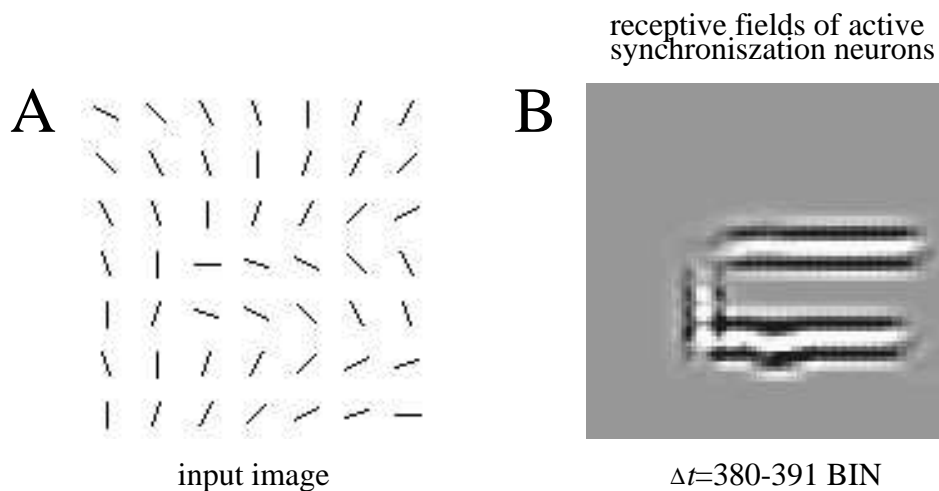
**Tab. 5.1.6:** Parameter der Liniendetektorneuronen (LDOK-Schicht).

**Abb. 5.1.14 (nächste Seite):** Architektur des Netzes zur Texturanalyse auf Basis der Detektion lokalen Orientierungskontrastes. Das Netz besteht aus mehreren Verarbeitungsstufen. In der untersten Stufe werden Linien extrahiert (line detection), die in der nächsten Stufe einer räumlichen Tiefpaßfilterung unterzogen werden (TP). Diese Signale verwendet die folgende Stufe zur Detektion des Orientierungskontrastes. In jedem Ort wird der lokale Orientierungskontrast durch eine modifizierte neuronale Mexikanerhutfilterung ermittelt (siehe Text). Die Master-OK-Schicht kodiert den gesamten Orientierungskontrast, der schließlich durch eine Liniendetektion an nachfolgende Schichten des Kontur/Form-Systems weitergeleitet werden kann.

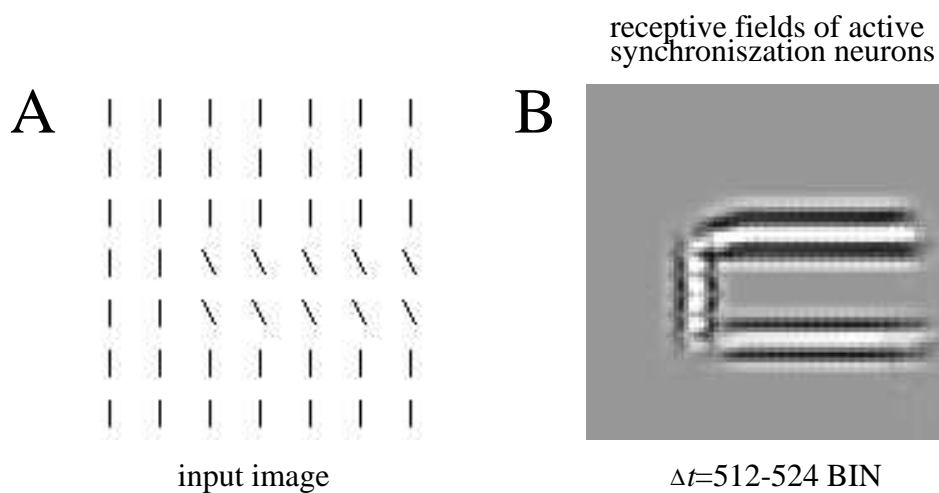


### 5.1.5 Simulationsergebnisse

Die in den letzten Schichten des Netzes zur Texturanalyse (LDOK-Schichten) folgenden Verarbeitungsstufen liefern zeitlich getrennte Repräsentationen der durch vorliegenden Orientierungskontrast definierten Objekte. Die Abbildungen 5.1.15 bis 5.1.17 zeigen neben den Eingangsbildern auch die neuronalen Aktivitäten der Synchronisationsschichten als Simulationsbeispiele.

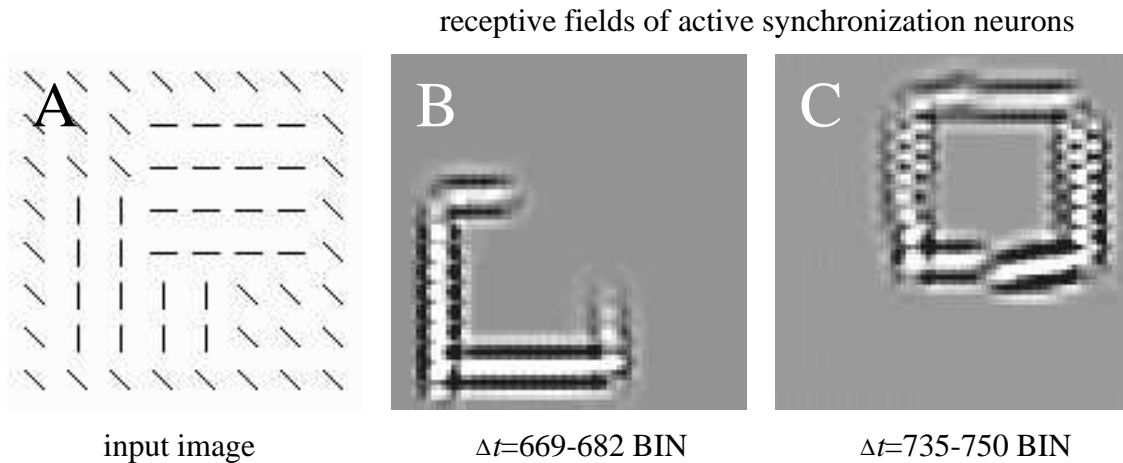


**Abb. 5.1.15:** *Simulationsbeispiel: Definition eines Objekts durch starken Orientierungskontrast.* In der aus orientierten Linienelementen bestehenden künstlichen Textur nimmt der menschliche Betrachter einen horizontalen "Balken" wahr (nach: Nothdurft 1993b, vgl. Abb. 5.1.10). In den Synchronisationsschichten wird dieser Balken im Zeitintervall  $\Delta t$  durch Impulsaktivitäten repräsentiert (B).



**Abb. 5.1.16:** *Simulationsbeispiel: Definition eines Objekts durch schwachen Orientierungskontrast.* Auch in dieser aus orientierten Linienelementen bestehenden künstlichen Textur nimmt der menschliche Betrachter einen horizontalen "Balken" wahr (nach: Nothdurft 1993b, vgl. Abb. 5.1.10). In den Synchronisationsschichten wird dieser Balken im Zeitintervall  $\Delta t$  durch Impulsaktivitäten repräsentiert (B).





**Abb. 5.1.17:** *Simulationsbeispiel: Separation von durch Orientierungscontrast definierten Objekten.* Aufgrund der Orientierungscontrastverteilung in einer durch orientierte Linienelementen bestehenden Textur nimmt der menschliche Betrachter zwei Quadrate wahr, von denen eines scheinbar teilweise überdeckt wird (A). Die Synchronisationsneurone repräsentieren die Konturen dieser Objekte während verschiedener Zeitintervalle  $\Delta t$  (B, C).

## 5.2 Einpassung des Texturmoduls in das Kontur/Form-System

Das im vorangegangenen Abschnitt vorgestellte Netz zur Texturanalyse kann, wie gezeigt wurde, künstliche Texturen, die aus Linienelementen bestehen, erfolgreich verarbeiten. Dazu wurde ein Satz spezieller Liniendetektoren implementiert, der eine Orientierungsempfindlichkeit von  $15^\circ$  besitzt. Diese Detektoren sind aber kaum für den Einsatz bei der Verarbeitung von Szenen, in denen natürliche Texturen auftreten, geeignet. Daher erhalten die höheren Schichten des Texturmoduls nun die Kanteninformationen von den in Kapitel 4 beschriebenen Kantendetektoren. Wird dann die Ausgangsaktivität der Master-OK-Schicht, wie schon in Kapitel 5.1 beschrieben, über Liniendetektoren an die Synchronisations- und End-Stop-Schichten weitergeleitet, so ist das Texturmodul bereits vollständig eingekoppelt.

### 5.2.1 Veränderungen im Netz

Das in Kapitel 5.1 vorgestellte Netz wird nun in den folgenden Punkten abgeändert:

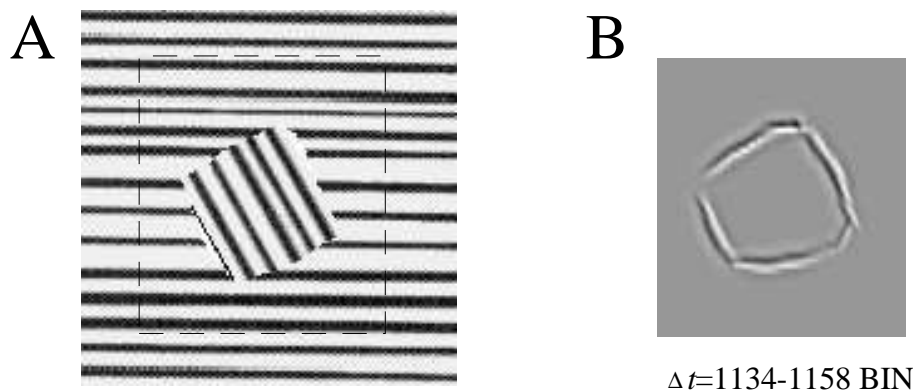
- 1.) Die Liniendetektion (Kap. 5.1.1) entfällt. Statt dessen erhalten die TP-Neurone gemäß Gleichung 5.1.18 ihre Eingangssignale von den Komplex-Neuronen (Kap. 4.1.5). In Gleichung 5.1.18 muß  $H_4$  gegen die Menge aller Orte auf dem hexagonalen Raster der Abtastweite  $d=2$  ersetzt werden. Die beiden globalen Neuronen erhalten Feeding-Signale von den Komplex-Schichten.
- 2.) Da nur sechs Kantendetektororientierungen vorliegen, werden auch nur sechs TP- und sechs OK-Schichten benötigt (Vorzugsorientierungen von  $0^\circ$  bis  $150^\circ$  in  $30^\circ$ -Schritten).

3.) Bei der OK-Detektion werden nur noch die Aktivitäten der ähnlichen Orientierungen ( $\varphi \pm 30^\circ$ ) berücksichtigt, da die "Zwischenorientierungen" fehlen, d.h. die Gleichungen 5.1.14 und 5.1.15 werden ungültig.

Durch diese wenigen Änderungen wird das Texturmodul vollständig in das Kontur/Form-System integriert.

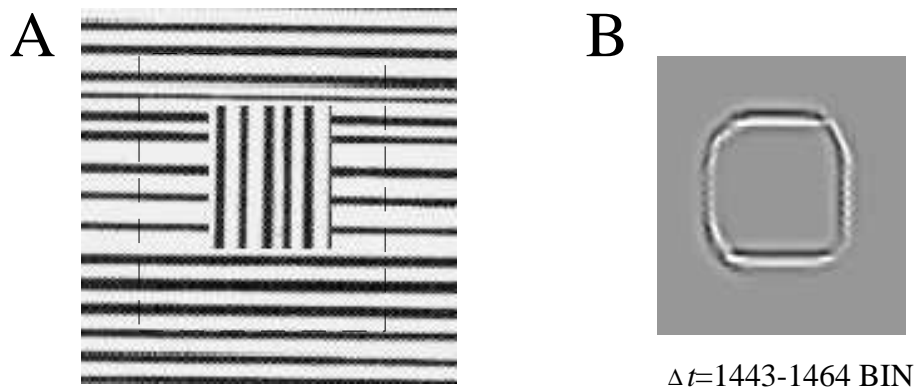
### 5.2.2 Simulationsergebnisse

Anhand einiger Texturbeispiele, die sich aus verschiedenen Brodatz-Texturen (Brodatz 1966) zusammensetzen, kann nun die Leistungsfähigkeit des erweiterten Systems hinsichtlich Texturanalyse verdeutlicht werden. In den Simulationen werden zunächst die direkten Verbindungen von den Komplex-Schichten zu den Synchronisationsschichten bzw. Endstop-Schichten nicht berücksichtigt. Der bearbeitete Bildausschnitt hat die Größe von  $88 \cdot 100$  Pixel, und das Simulationsergebnis liegt in den zwölf Synchronisationsschichten, die jeweils aus  $44 \cdot 50$  Neuronen bestehen, vor. Abbildung 5.2.1 A zeigt die Oberfläche eines Strohschirms. Ein um  $60^\circ$  gedrehtes nahezu quadratisches Oberflächenstück wurde nachträglich als Objekt eingefügt und ist auch deutlich als solches wahrzunehmen. Das Kontur/Form-System liefert ab etwa 500 Zeitschritten entsprechende Repräsentationen des Objekts in bestimmten Zeitintervallen. Auch bei einem um  $90^\circ$  gedrehten rechteckigen Oberflächenstück liefern die Synchronisationsneurone entsprechende Repräsentationen (Abb. 5.2.2).

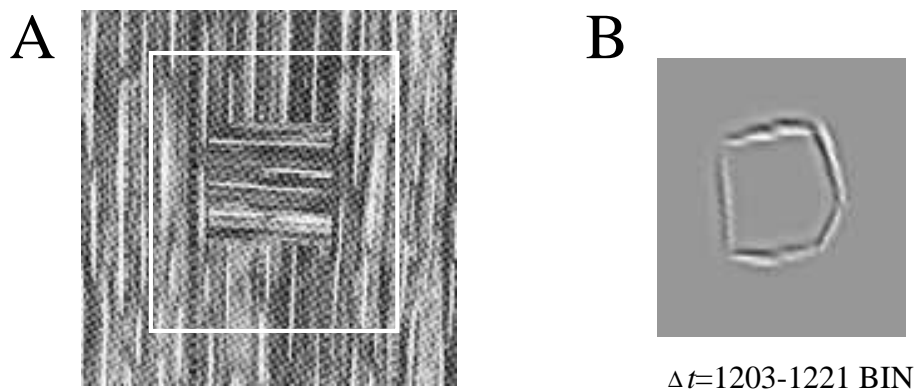


**Abb. 5.2.1:** *Simulationsbeispiel: Definition eines Objekts durch Orientierungskontrast in natürlichen Texturen.* Als Stimulus wird die Textur einer Strohschirmoberfläche verwendet, in die nachträglich ein rechteckiges um  $60^\circ$  gedrehtes Oberflächenstück eingesetzt wurde (A, nach: Brodatz 1966, Textur 49). In den Synchronisationsschichten wird das Rechteck durch neuronale Aktivität innerhalb des Zeitintervalls  $\Delta t$  repräsentiert (B).

Abbildung 5.2.3 zeigt ein weiteres Simulationsbeispiel. Ein durch unterschiedliche Orientierungen von Holzstrukturen wahrgenommenes Rechteck wird vom Kontur/Form-System als solches repräsentiert.



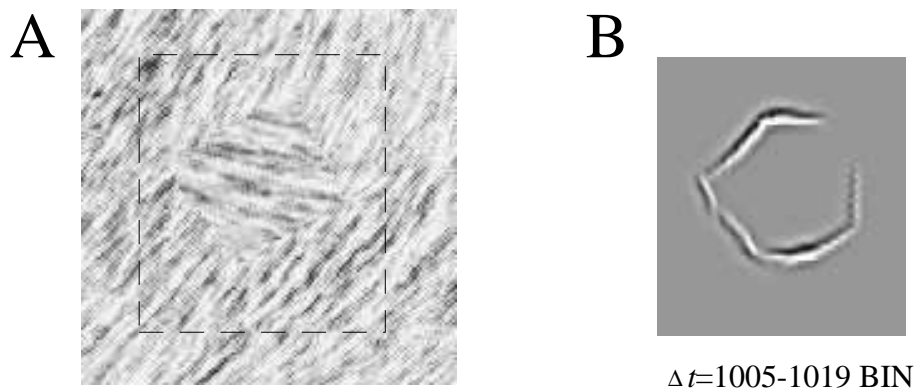
**Abb. 5.2.2:** *Simulationsbeispiel: Definition eines Objekts durch Orientierungskontrast in natürlichen Texturen.* Als Stimulus wird erneut die Textur einer Strohschirmoberfläche verwendet, in die jedoch nachträglich ein rechteckiges um  $90^\circ$  gedrehtes Oberflächenstück eingesetzt wurde (A, nach: Brodatz 1966, Textur 49). Auch hier wird in den Synchronisationsschichten das Rechteck durch neuronale Aktivität innerhalb bestimmter Zeitintervalle repräsentiert (B).



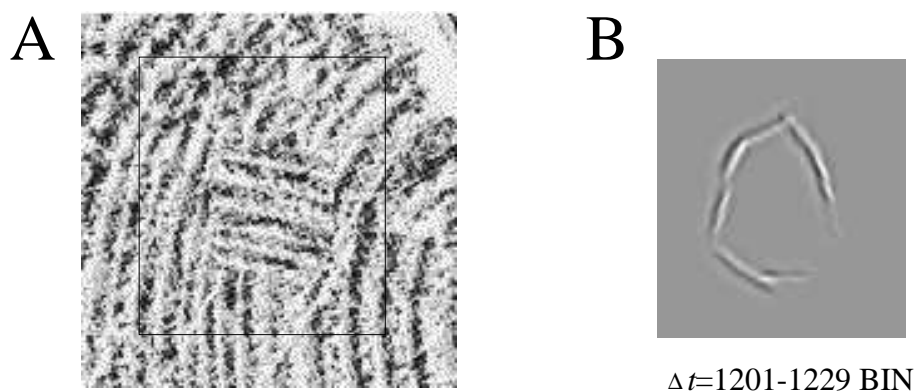
**Abb. 5.2.3:** *Simulationsbeispiel: Definition eines Objekts durch Orientierungskontrast in natürlichen Texturen.* Als Stimulus wird die Textur einer Holzoberfläche verwendet, in die auch hier nachträglich ein rechteckiges um  $90^\circ$  gedrehtes Oberflächenstück eingesetzt wurde (A, nach: Brodatz 1966, Textur 68). Das Rechteck wird in den Synchronisationsschichten innerhalb bestimmter Zeitintervalle repräsentiert (B).

Etwas schwieriger ist die Wahrnehmung des Rechtecks in Abbildung 5.2.4 A. Das Bild zeigt zunächst die Fellstrukturen eines ungeborenen Kalbes. Auch hier wurde nachträglich ein rechteckiges gedrehtes Oberflächenstück eingefügt. Der menschliche Betrachter nimmt das Rechteck aufgrund der unterschiedlichen Struktur-Orientierungen von Rechteck und Hintergrundtextur wahr. Die Synchronisationsschichten liefern fast vollständige Repräsentationen dieses Objekts in bestimmten Zeitintervallen. An einer Ecke des Rechtecks wird kein

Kontrast angezeigt, was sich auf die (fälschliche) Detektion ähnlicher Kantenorientierungen in diesem Bereich zurückführen läßt. Ebenso nahezu vollständige Repräsentationen der als Rechteck wahrgenommenen Konturen liefern die Synchronisationsschichten bei Verarbeitung des in Abbildung 5.2.5 gezeigten Beispiels. Das Rechteck wird durch starke Orientierungsunterschiede in den Strukturen eines Baumstumpf-Querschnittes definiert und als solches wahrgenommen.



**Abb. 5.2.4:** *Simulationsbeispiel: Definition eines Objekts durch Orientierungskontrast in natürlichen Texturen.* Als Eingangsbild wird die Textur der Felloberfläche eines ungeborenen Kalbes benutzt, in die nachträglich ein rechteckiges um  $60^\circ$  gedrehtes Oberflächenstück eingesetzt wurde (**A**, nach: Brodatz 1966, Textur 93). Das Rechteck wird nahezu vollständig in den Synchronisationsschichten innerhalb bestimmter Zeitintervalle repräsentiert (**B**).



**Abb. 5.2.5** *Simulationsbeispiel: Definition eines Objekts durch Orientierungskontrast in natürlichen Texturen.* Hier wird die Oberfläche eines Baumstumpf-Querschnitts als Eingangsbild verwendet, nachdem ein rechteckiges um  $90^\circ$  gedrehtes Oberflächenstück eingesetzt wurde (**A**, nach: Brodatz 1966, Textur 97). Das Rechteck wird nahezu vollständig in den Synchronisationsschichten innerhalb bestimmter Zeitintervalle repräsentiert (**B**).

### 5.3 Diskussion

Ein mit einer Textur ausgefüllter Bildbereich erscheint dem menschlichen Betrachter als eine perzeptuelle Einheit innerhalb einer Hintergrundtextur, wenn sich die beiden Texturen genügend unterscheiden. Die meisten Verfahren und Theorien zur Trennung solcher Texturen basieren auf dem Vergleich bestimmter Texturmerkmale (z.B. Julesz 1981a, 1981b, 1986, Reitboeck et al. 1987, 1988, Graham et al. 1992). Zur Extraktion von Merkmalen werden häufig Sätze von Gaborfiltern unterschiedlicher räumlicher Auflösung verwendet (z.B. Turner 1986, Fogel und Sagi 1989, Dunn et al. 1994, Graham et al. 1992, Graham et al. 1993), da sich die zugrundeliegenden Gaborfunktionen (Gabor 1946, Daugman 1985) besonders gut für eine Kantendetektion in künstlichen Systemen eignen. Außerdem gibt es eine Reihe von neurophysiologischen Untersuchungen, die zeigen, daß die räumlichen Eigenschaften von rezeptiven Feldern bestimmter Neuronen des visuellen Kortex durch Gaborfunktionen beschrieben werden können (z.B. Marcelja 1980, Daugman 1980). Es gibt jedoch auch einige psychophysische Studien, die zeigen, daß für das visuelle System nicht Merkmale als solche, sondern der Merkmals*kontrast* zur Trennung von Texturbereichen entscheidend ist (Nothdurft 1991a, 1991b, 1992, 1993). Ein solcher Merkmalskontrast ist z.B. der Orientierungsgradient oder Orientierungskontrast, der zwischen benachbarten (orientierten) Linienelementen einer künstlichen Textur auftritt (vgl. Abb. 5.1.10). Modelle, die zur Berechnung des lokalen Orientierungsgradienten entwickelt wurden, können bei der Verarbeitung solcher künstlichen Texturen Ergebnisse liefern, die der menschlichen Wahrnehmung entsprechen. Ein solches Modell wurde von Landy und Bergen 1991 vorgestellt. Es berechnet lokale Orientierungsgradienten durch Differenzenbildung zwischen lokalen Filtern, die bevorzugt auf orthogonal orientierte Linien(-elemente) ansprechen. Ein weiteres Modell ermittelt die Stärke des lokalen Orientierungskontrastes durch laterale inhibitorische Kopplungen zwischen Neuronen, die ähnliche Vorzugsorientierungen von Neuronen kodieren (Weitzel et al. 1995, Weitzel 1995). Die entsprechende Verschaltung wird durch eine Gewichte Verteilung realisiert, die sich durch Mexikanerhut-Funktionen beschreiben läßt. Das Modell erlaubt weiterhin eine erfolgreiche Verarbeitung von Texturen, deren Elemente aus mehreren Linienelementen zusammengesetzt sind. Auch andere Merkmalskontrasttypen können bei der Verarbeitung von Texturen von Bedeutung sein, wie z.B. der *Linienenden-Orientierungskontrast* (Weitzel 1995) oder ein Farbgradient.

Da das menschliche Sehsystem in der Lage ist, anhand von Texturinformation Objekte zu erkennen, ist auch eine Erweiterung des Kontur/Form-Systems um ein Verarbeitungsmodul, das zunächst wenigstens einen kleinen Teil dessen leistet, wünschenswert. Das in diesem Kapitel vorgestellte Texturanalyseverfahren ermittelt ausschließlich den lokalen Orientierungskontrast. Es kann auf künstliche und, bei geeigneter Vorverarbeitung, auch auf natürliche Texturen angewendet werden. Das Verfahren ist nicht grundsätzlich neu. Es stellt eine Erweiterung der oben erwähnten früheren Version dar (Weitzel et al. 1995, Weitzel

1995). Auch das hier gezeigte Netz ermittelt die Stärke des lokalen Orientierungskontrastes durch Gewichteverteilungen, die sich durch Mexikanerhut-Funktionen (DOOG) beschreiben lassen. Im Gegensatz zur früheren Version wurde das Netz aber mit dynamischen Neuronen implementiert. Es liefert nur den stärksten im Bild enthaltenen lokalen Orientierungskontrast, d.h. schwacher Hintergrundkontrast wird unterdrückt. Eine vergleichbare Unterdrückung verwendet auch das Texturanalyseverfahren von Landy und Bergen (Landy und Bergen 1991, Gonzales und Winz 1977). Außerdem ist das Texturmodul durch den Aufbau mit dynamischen Neuronen technisch einfacher in das Kontur/Form-System zu integrieren und kann damit auf reale Texturbeispiele angewendet werden.

Das visuelle System kann durch Texturinformation Objekte erkennen, d.h. Objektkonturen wahrnehmen, die nicht unbedingt durch Intensitätskontrast hervorgerufen wurden. Beispiele dazu zeigen die Abbildungen 5.1.10 A und 5.1.11 A, in denen Betrachter jeweils einen horizontalen Balken wahrnehmen, der nicht auf Intensitätskontrast zurückgeführt werden kann. Diese scheinbaren Konturen haben aber mit "echten" Konturen offensichtlich viele Eigenschaften gemeinsam. Beide Konturtypen können Ecken, T-Stücke, gerade Kanten oder Krümmungen beinhalten. Daher liegt es nahe, Repräsentationen solcher scheinbaren Konturen, die ein Texturanalyseverfahren als Ausgabe liefern soll, genauso weiterzuverarbeiten wie Repräsentationen "echter" Konturen, wie sie beispielsweise in den PI-Schichten des Kontur/Form-Systems signalisiert werden.

Das Texturanalysemodul wird daher zwischen die PI-Schichten und die Synchronisations- bzw. ES-Detektionsstufe geschaltet. Diese Art der Einkopplung ist weniger biologisch als technisch motiviert, denn es gibt keine Hinweise auf eine getrennte Verarbeitung von Orten hoher Krümmung im visuellen System, wie es im künstlichen System bei dieser Einkopplung der Fall ist (Detektion des lokalen Orientierungskontrasts in Texturen und Vertexdetektion). Auf der anderen Seite wird eine einfache Realisierung der Berücksichtigung von weiteren Merkmalskontrasttypen ermöglicht. Denn obwohl bestimmte Stimuli denkbar sind, bei denen es zu unerwünschten Wechselwirkungen zwischen durch Textur- und durch Konturverarbeitung gewonnenen Repräsentationen kommen kann, ist in der Regel eine Separation dieser unterschiedlichen Repräsentationen möglich.

Eine zweite Variante der Einkopplung würde zu einer Aufspaltung der Verarbeitung in einen Kontur- und einen Texturpfad nach der Vorverarbeitung führen. Die Kanteninformation würde einerseits direkt weiterverarbeitet und andererseits über ein Texturmodul an eine getrennte, aber gleichartige Weiterverarbeitung gegeben. Die Repräsentationen von Konturen, die durch Intensitätskontrast gewonnen werden, und die von Konturen, die durch Orientierungskontrast definiert werden, lägen schließlich in getrennten Synchronisationsschichten vor.

Ebenso würden zwei Synchronisationsschichten in einer dritten Kopplungsvariante existieren. Die End-Stop- und die Vertex-Schichten lieferten hier ihre Signale an das Texturmodul, dem eine eigenständige Weiterverarbeitung nachgeschaltet wäre. Diese dritte

Variante erlaubte auch die Detektion von Linienenden-Orientierungscontrast. Die zweite Variante wurde prinzipiell auch im vorliegenden Modell implementiert, jedoch nur einer der beiden Pfade tatsächlich simuliert. Sie könnte bei geeigneter großer Anzahl an zur Verfügung stehenden Neuronen auch technisch einfach und vollständig realisiert werden. Ein nachgeschaltetes System müßte jedoch die Ausgabesignale von zwei getrennten Ausgabestufen auswerten. Diese Vorstellung scheint aber nicht ganz abwegig zu sein. Das läßt sich beispielsweise anhand der Abbildungen 5.1.10 A und 5.1.11 A verdeutlichen. Hier nimmt der menschliche Betrachter zwar einerseits horizontale Balken wahr, andererseits sieht er auch Linienzüge. Die durch Intensitätscontrast definierten Linienzüge würden durch eine Synchronisationsstufe und der durch Orientierungscontrast wahrgenommene Balken durch die andere Synchronisationsstufe repräsentiert werden. Bei konsequenter Fortführung dieser Verschaltung müßte für jeden Merkmalscontrasttyp ein eigener Verarbeitungspfad implementiert werden, lediglich die Vorverarbeitung könnte weitgehend gemeinsam genutzt werden.

Ergebnisse biologischer Untersuchungen scheinen tatsächlich eine getrennte Verarbeitung zu unterstützen. Beispielsweise wird von Neuronen in höheren Arealen des visuellen Kortex (IT, wo auch der Aufbau entsprechender Repräsentationen vermutet wird) berichtet, die besonders selektiv für Kontur sind (vgl. Kap. 2.2). Bei Untersuchungen an V1 und V2 des visuellen Kortex des Affen fanden von der Heydt et al. (1991, 1992) einen Zelltyp (*grating-Zelle*), der verstärkt auf Gitter mit geeigneter Orientierung, Position und Periodizität antwortet, aber im Gegensatz zu anderen orientierungsselektiven Zellen kaum auf einzelne Kanten oder Linien. Außerdem fanden sie Zellen, die das genau umgekehrte Verhalten von *grating-Zellen* zeigten. Ähnliche Entdeckungen machten beispielsweise Schiller et al. bereits 1976, als sie Zellen in V1 fanden (*bar-Zellen*), die auf einzelne Balkenreize antworteten, nicht aber auf Sinus-Gitter. Die beiden Zelltypen scheinen somit komplementäre Rollen zu spielen, da *bar-Zellen* nur selektiv für Form- oder Konturinformation, d.h. für Orts-Linien, und *grating-Zellen* ausschließlich selektiv für orientierte Texturinformation sind, d.h. für Frequenz-Linien (Petkov und Krüzinga 1997), was als ein weiterer Hinweis auf eine getrennte Verarbeitung aufgefaßt werden kann.

Auch die dritte Variante der Einkopplung des Texturmoduls hätte zwei getrennte Ausgangsstufen. Prinzipiell scheint diese Methode die biologienäheste zu sein, da die Informationen über Orte hoher Konturkrümmung nicht mehrmals gewonnen werden müßten. Hier ist aber noch unklar, auf welche Weise die Vertexinformation genutzt werden könnte, um daraus einen lokalen Orientierungscontrast zu gewinnen, wie er auch aus Texturen detektiert wird.

## 6 Abschließende Diskussion

### 6.1 Allgemeine Diskussion

Damit ein künstliches System eine dem menschlichen Sehsystem qualitativ vergleichbare Leistungsfähigkeit in der Mustererkennung erreichen kann, müssen zuvor die Probleme der Figur/Hintergrundtrennung und der Invariantenbildung gelöst und so der Aufbau geeigneter Objektrepräsentationen ermöglicht werden. Es gibt eine Vielzahl von Objekterkennungsmodellen, die erfolgreich Objekte wiedererkennen oder klassifizieren können, wenn diese vom Hintergrund oder anderen Objekten getrennt präsentiert werden (z.B. Reeke et al 1990, Edelman 1997, Mel 1997). Die dabei erreichten Erkennungsleistungen sind bei einigen dieser Modelle auch in gewissem Rahmen robust gegen Größen- und Positionsveränderungen oder gegen Rotation. Andere Modelle beschäftigen sich vorwiegend mit der Bildung von Invarianzen (z.B. Reitboeck und Brody 1969, Fukushima 1980, Reitboeck 1989, Stöcker 1993).

In dieser Arbeit wurde ein Modell vorgestellt, das einen Ansatz zur Lösung des Figur/Hintergrundproblems zeigt, indem es Repräsentationen von Objekten bzw. Objektkandidaten, die innerhalb realer Szenen auftreten, liefert und dabei Annahmen über Zusammengehörigkeit von Konturelementen zu einzelnen Objektkandidaten macht. Einzelne Konturelemente werden bei Gleichmäßigkeit in ihren Orientierungen gemäß dem Gestaltsgesetz des *guten Verlaufs* über laterale Linking-Kopplungen zusammengebunden. T-Stücke werden vom Modell als Hinweis auf Verdeckung, d.h. Zugehörigkeit der angrenzenden Konturelemente zu verschiedenen Objekten, und andere Vertizes als Hinweis auf Zugehörigkeit zum selben Objekt interpretiert und die jeweiligen Konturelemente im Zeitbereich entsprechend zusammengebunden bzw. getrennt.

Für ein komplettes Mustererkennungssystem müßte eine Stufe der Invariantenbildung nachgeschaltet werden, und ein dieser Stufe wiederum nachfolgendes Erkennungssystem müßte anhand der resultierenden Repräsentationen entscheiden, um welches Objekt oder welche Objektklasse es sich dabei handelt oder ob überhaupt eine sinnvolle Repräsentation vorliegt. Zur vollständigen Lösung des Figur/Hintergrundproblems wären zusätzliche Rückkopplungen vom Erkennungssystem auf die vorgeschalteten Stufen notwendig, um die Trennung der Objekte vom Hintergrund oder anderen Objekten zu unterstützen. Vergleichbare Rückkopplungen wurden bereits in einem Modell von Stöcker 1993, das eine beschränkte Anzahl von Objekten in künstlichen Szenen erkennen kann, verwendet.

Im Kontur/Form-System signalisiert nahezu synchrone neuronale Aktivität Zusammengehörigkeit von Konturelementen. Ein Objekt wird durch eine Linking-Welle, die die Objektkonturen innerhalb bestimmter Zeitintervalle umläuft, repräsentiert und dadurch von anderen Objekten bzw. vom Hintergrund getrennt.



Neben der Repräsentation eines Objektes durch Kodierung seiner Konturen liefert das Modell immer auch eine Repräsentation durch Kodierung seiner Vertizes. Da die Master-Vertex-Neuronen, die die Vertizes kodieren, in denselben Repräsentationszeitintervallen aktiv werden, in denen die Linkingwelle die Objektkonturen umläuft, ist Eindeutigkeit bezüglich der Zusammengehörigkeit bzw. Zugehörigkeit der Vertizes gegeben. Unter Berücksichtigung der in den Vertex-Zeiger-Schichten enthaltenen Informationen lassen sich die Objektkonturen meist vollständig rekonstruieren. Zwei aufeinander gerichtete Zeiger signalisieren eine gerade Kante, soweit sich keine weiteren gleichgerichteten Zeiger zwischen ihnen befinden. Ein Zeiger ohne Gegenstück weist auf Verdeckung (T-Stück) oder Krümmungen hin. In einer Diplomarbeit wurde die Auswertung der Information solcher relationaler Zeiger, d.h. insbesondere die Rekonstruktion der Objektkonturen anhand der Zeigerinformation, näher untersucht (Spengler 1996). Dort stand jedoch die durch synchrone neuronale Aktivität kodierte Information der Zusammengehörigkeit von Vertizes nicht zur Verfügung. Durch diese zusätzliche Information scheinen die Vertex-Zeiger hervorragende Merkmale für die Erzeugung von größen-, verschiebungs- und rotationsinvarianten Repräsentationen von Objekten zu sein.

Synchronizität neuronaler Aktivität spielt bereits in den Modellen von Hummel und Biederman (1992) und Ostkamp (1996) eine wichtige Rolle. Beide Modelle verarbeiten nur Strichzeichnungen oder binarisierte Eingangsbilder, wodurch zwar die Generierung "echter" neuronaler Synchronizität erleichtert wird, die direkte Verarbeitung von Grauwertbildern aber nicht möglich ist. Das Modell von Hummel und Biederman spezifiziert Objektteile oder Geone und setzt sie explizit durch eine entsprechende Sprache zueinander in Relation. Objektattribute werden durch unabhängige Einheiten repräsentiert und durch synchrone oszillierende Aktivität zusammengebunden. Das Modell rekonstruiert 3D-Strukturen und paßt sie an gespeicherte Repräsentationen an.

Im vorliegenden Kontur/Form-System werden hingegen Objektkandidaten auf Grundlage von über Vertizes zusammenhängenden Konturzügen spezifiziert. Die räumlichen Relationen liegen implizit in den neuronalen Schichten vor. Außerdem liefert das Kontur/Form-System 2D-Repräsentationen und baut keine ansichtsinvarianten Repräsentationen auf. Das wiederum läßt sich sehr gut mit psychophysischen Experimenten vereinbaren, die auf eine Standpunktabhängigkeit der Repräsentationen und der Erkennung von Objekten hinweisen. 3D-Objekte scheinen demnach durch einen gespeicherten Satz von 2D-Repräsentationen kodiert zu werden (Bülthoff et al. 1995).

## 6.2 Biologische Relevanz

Bei der Umsetzung des Kontur/Form-Systems war die Berücksichtigung biologischer Aspekte ein zentrales Anliegen, da die Arbeit unter dem Leitgedanken "Lernen von der Biologie"

stand. Dies wirkte sich insbesondere in den Schichten der Vorverarbeitung, der Kantendetektoren und der End-Stop-Detektoren aus, denn auch kortikale Neuronen mit vergleichbaren Funktionen sind bekannt (Hubel und Wiesel 1965). Auch die Existenz von Neuronen, die ähnlich den Vertexdetektoren hohe Konturkrümmungen kodieren könnten, ist zumindest physiologisch plausibel. Die Extraktion von Merkmalen aus Szenen, wie sie in den neuronalen Schichten erfolgt, ist jedoch nicht als der entscheidende Teil dieser Arbeit zu verstehen, da bereits viele Algorithmen bekannt sind, die vergleichbare Extraktionen leisten. Ein wichtiger Vorteil der biologienahen Implementierung liegt hier vor allem in der Verbesserung der Detektion verrauschter oder unvollständiger Kanten durch die Verwendung von Linking-Verbindungen, durch die auch Gestaltprinzipien (z.B. Wertheimer 1923) als Annahmen im Modell berücksichtigt werden können. Überhaupt dürfte als wesentlicher Vorteil von Implementierungen mit biologienahen impulskodierenden Modellneuronen gegenüber technischen Realisationen, die in Spezialfällen stets effizienter sein können, die höhere Flexibilität des Linking-Ansatzes angesehen werden.

Die entscheidende Neuerung des Kontur/Form-Systems stellt jedoch die Dynamik zwischen den Vertex-Schichten und den Synchronisationsschichten zur Definition und Separation von Objekten im Zeitbereich dar. Es ist zwar unklar, ob im menschlichen Gehirn eine direkt vergleichbare Verschaltung vorliegt. Jedoch gibt es deutliche Hinweise auf Konturrepräsentationen auch in höheren Arealen des visuellen Kortex. Im inferior temporalen Kortex (IT) finden sich Neurone, die verstärkt selektiv für Kontur sind (Logothetis und Sheinberg 1996). Konturrepräsentationen, die denen im Kontur/Form-System ähnlich sind, könnten im posterioren IT (TEO) vorliegen, da hier die Zellen im Gegensatz zu denen des anterioren IT (TE) retinotop angeordnet sind (Boussaoudd et al. 1991). Die Annahme, daß in TEO zunächst die Selektivität für Objektmerkmale mittlerer Komplexität konstruiert wird und anschließend in TE Invarianzen aufgebaut werden (Tanaka 1993), unterstützt diese Hypothese.

In IT finden sich jedoch auch viele Neurone, die selektiv für Farbe, Bewegungsrichtung oder Textur sind. Außerdem konnten bereits in V1 und V2 von Affen Zellen gefunden werden, die ausschließlich selektiv für orientierte Texturinformationen, und Zellen, die ausschließlich selektiv für Konturinformationen sind (von der Heydt et al. 1991, 1992; Schiller et al. 1972). Daher ist es offensichtlich von Vorteil, auch solche Merkmale in einem biologienahen künstlichen Sehsystem verarbeiten zu können.

Um zu zeigen, auf welche Weise dies im Modell realisiert werden kann, wurde das Kontur/Form-System um ein Modul erweitert, das exemplarisch Orientierungskontrast, der zwischen Linienelementen in Texturen auftritt, extrahiert. Auch die Möglichkeiten, diesen Kontrast weiterzuverarbeiten und die Detektion von Konturen durch solche Zusatzinformation zu unterstützen, wurden gezeigt bzw. diskutiert (Abschnitt 5.3).

Das Zusammenbinden von verschiedenster zu einem Objekt gehörender Information könnte durch synchrone neuronale Aktivität ermöglicht werden (Reitboeck 1980, Reitboeck

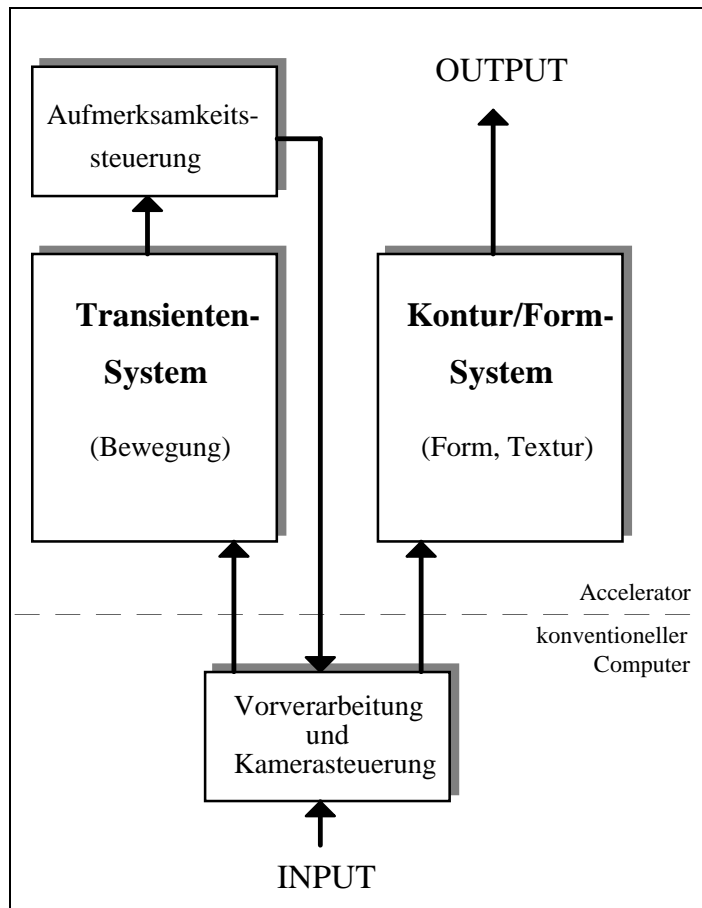
1983, von der Malsburg 1981, von der Malsburg und Schneider 1986). Entsprechende Oszillationen konnten zwar im visuellen Kortex der Katze gefunden werden (z. B. Eckhorn et al. 1988, Gray et al 1989), jedoch bisher nicht in IT. Synchrone neuronale Aktivität muß aber nicht periodisch sein, ebensogut ist das Auftreten nichtperiodischer Synchronisation vorstellbar. Eine solche Art der Synchronisation tritt in der Regel auch im Kontur/Form-System auf. Die einzelnen Objekte werden im Modell durch Ausbreitung von Linking-Wellen in meist nicht regelmäßigen zeitlichen Abständen repräsentiert. Eine solche Welle bedingt eine Repräsentation, die sich über ein bestimmtes Zeitintervall erstreckt. Insofern kann streng genommen nicht mehr von Synchronisation im eigentlichen Sinne, d.h. von gleichzeitiger Aktivität, gesprochen werden. Dennoch wird durch die Implementierung von sich entlang der Objektkonturen ausbreitender Aktivitäten, die im Modell überdies Nichtperiodizität zeigen, dem grundsätzlichen Gedanken des zeitlichen Zusammenbindens neuronaler Aktivität entsprochen, indem benachbarte Neuronen gleichzeitig aktiviert werden und Nachbarschaftsaktivität die Nachbarschaftskontinuität definiert. Es schließt sich sogar die Frage an, ob nicht auch im menschlichen Gehirn eher nichtperiodische Aktivitätsausbreitung als periodische Aktivität zu erwarten ist, wenn Synchronisation tatsächlich für das Zusammenbinden von Information verantwortlich sein sollte.

### 6.3 Technische Aspekte

Das Kontur/Form-System ist ein in sich abgeschlossenes System, es stellt zugleich aber auch den Kern eines Gesamtsystems dar, das im Rahmen des BMBF-Projekts "Elektronisches Auge" entwickelt wurde. Als weitere wesentliche Verarbeitungseinheiten beinhaltet das Gesamtsystem neben der Vorverarbeitung und dem Kontur/Form-System auch ein Transientensystem (Schott et al. 1996, Schott et al. 1997) und eine Aufmerksamkeitssteuerung (Pauly et al 1997). Eine schematische Darstellung des Gesamtsystems zeigt Abbildung 6.3.1.

Das Kontur/Form-System bearbeitet räumlich hochauflösend einen kleinen Teil des dem Gesamtsystem vorliegenden Eingangsbildes. Das jeweils interessierende Objekt muß daher im Bearbeitungsfenster des Kontur/Form-Systems zentriert werden. Bei sich bewegenden Objekten wird diese Zentrierung (Fixation) durch das Transientensystem in Verbindung mit der Aufmerksamkeitssteuerung und Folgebewegungen gewährleistet. Das Transientensystem verarbeitet das gesamte Eingangsbild in einer groben Auflösung und detektiert Kanten sich bewegender Objekte und deren Bewegungsrichtung und Geschwindigkeit. Die resultierende Repräsentation ist zwar in der Regel zu unvollständig, um eine Objekterkennung allein zu ermöglichen, die Information wird aber genutzt, um über die Aufmerksamkeitssteuerung die Kamera auf die Objektkandidaten auszurichten und gegebenenfalls zu verfolgen (dauerhafte Fixation und Verfolgung bewegter Objekte). Durch den Aufmerksamkeitsfokus und die entsprechende Kameraführung wird vor allem eine räumlich selektive Verarbeitung der

Bildinformation in hoher Auflösung durch das Kontur/Form-System ermöglicht, was auch zu einer starken Reduktion von Zeit- und Rechenaufwand im Vergleich zu einer Gesamtbildanalyse führt.



**Abb. 6.3.1:** Neuronales Netzwerk zur Objekt/Hintergrund-Separation. Dargestellt sind die vier Verarbeitungsbereiche des Gesamtsystems. Vorverarbeitung und Kameraansteuerung werden mit Hilfe eines konventionellen Computers durchgeführt. Die drei anderen Verarbeitungsmodule sind auf der speziellen Hardware (Accelerator) implementiert.

Die Gesamtheit aller Teilsysteme ohne das Kontur/Form-System bildet eine geschlossene Regelschleife, die aus Kamera, Kamerasteuerung, Merkmalsextraktion, Transientensystem, Zielauswahlschicht und Generierung von Motorkommandos für die Kamera besteht. Das System arbeitet außer einigen Annahmen über Gestaltprinzipien, die in die Verschaltung implementiert wurden, ohne weitere Vorinformation. Eine Erkennungsstufe wurde bisher nicht eingesetzt. Das Gesamtsystem leistet Figur/Hintergrundtrennung und Merkmalskopplung in dynamischen Szenen und kann in der technischen Bildverarbeitung eingesetzt werden. Jedoch ist in Zukunft eine größere Anzahl von Neuronen notwendig, um zumindest alle Teilsysteme gleichzeitig betreiben zu können.

An verschiedenen Stellen der Arbeit wurde bereits auf die technischen Randbedingungen hingewiesen. Ziel des Projekts war ein in Echtzeit arbeitendes Gesamtsystem. Aus diesem Grund wurde ein spezieller, an das Marburger Modellneuron angepaßter Hardware-Beschleuniger (Accelerator) konstruiert (Frank und Hartmann 1995, Frank et al. 1998, Schaefer 1995, Wolff 1996). Der Accelerator erlaubt die Verarbeitung von Netzen mit höchstens 128K Neuronen und 8M Gewichten. Im gesamten Netz können je Teilpotential

(Feeding, Linking, etc.) maximal vier verschiedene Zeitkonstanten verwendet werden. Das vorgestellte Netz des Kontur/Form-Systems besteht ohne Texturmodul aus 57 Schichten, die sich wiederum aus jeweils  $50 \cdot 44 = 2200$  Neuronen zusammensetzen, und einem globalen Inhibitionsneuron, so daß das Netz aus 125401 Modellneuronen mit 4,8 M Verbindungen besteht und damit der größtmögliche Bildausschnitt gewählt wurde. In einer der in Kapitel 4 gezeigten Simulationen wurde ein etwas länglicherer Bildausschnitt gewählt, um mehrere Objekte gleichzeitig im Verarbeitungsbildausschnitt plazieren zu können. Schon das Kontur/Form-System schöpft ohne Texturmodul allein die Acceleratorkapazität vollständig aus. Daher konnten bisher die Teilsysteme des Gesamtsystems nur einzeln betrieben werden. Eine Vergrößerung des Beschleunigersystems ist aber momentan in Arbeit. Für die Verarbeitung von 1000 Zeitschritten einer typischen Szene (Verkehrsszene, Abb. 4.1.15), bei der eine mittlere Feuerrate von 0,087 % auftritt, benötigt das System etwa 10 Sekunden (Wolff 1998). Die entsprechende Simulation auf einer handelsüblichen Workstation (DEC-Alpha, Typ 250-4/266) benötigt 212 Sekunden.

## 6.4 Sonstige Einsetzbarkeit

Das Prinzip der Definition von Objekten durch fortschreitende Aktivitäten (Linking-Wellen), wie sie vom Kontur/Form-System erzeugt werden, könnte auch für formale Betrachtungen von Interesse sein. Die Ausgangssignale der Synchronisations- und Vertex-Schichten lassen sich auch als gerichtete Graphen interpretieren, die wiederum als Terminale einer entsprechenden formalen Sprache aufgefaßt werden können. Die Positionen zweier aufeinanderzeigender Vertex-Zeiger zusammen mit der Richtung der zwischen ihnen laufenden Linking-Welle ist interpretierbar als ein Terminal einer kontextfreien Grammatik der *Picture Description Language* (PDL, Shaw 1968). Werden beispielsweise die Konturen eines Objektes vollständig und richtig von einer Wanderwelle umschrieben, so entspricht diese einem syntaktisch richtigen "Satz" der Sprache PDL. Insofern könnte das Modell auch als Anwendung in der syntaktischen Mustererkennung (Fu 1974) relevant sein.

## 6.5 Ausblick

Das Kontur/Form-System könnte seine hauptsächliche Anwendung im Zusammenhang mit den anderen Teilmodulen des oben erwähnten Gesamtsystems finden. Aufgrund des biologienahen und sehr allgemeinen Ansatzes ist damit zu rechnen, daß bei geeigneter Hardware-Kapazität und konsequenter Erweiterung des vorhandenen Systems um weitere Verarbeitungsmodule ein künstliches Sehsystem entsteht, das bisherigen Systemen in vieler Hinsicht überlegen ist. Die nächsten konkreten Schritte sind in der Erweiterung des

Kontur/Form-Systems um zusätzliche Auflösungen und Merkmalskontrastmodule zu sehen. Außerdem ist das System um die Fähigkeit der Invariantenbildung auf Grundlage der Vertex-Weiser zu ergänzen, sowie um eine Stufe, die die daraus resultierenden Repräsentationen mit gespeicherten vergleicht, und um Rückkopplungen, durch die die Konturgewinnung aufgrund bereits erkannter Strukturen unterstützt werden kann. Bei all diesen Ergänzungen sollte weiterhin der Blick auf aktuelle Forschungsarbeiten aus dem Bereich der Neurobiologie und Neurophysiologie geworfen werden, um die gewonnenen Erkenntnisse auch im Modell nutzen zu können. Umgekehrt gibt das Modell deutliche Anhaltspunkte, welches Ergebnis in einem Experiment zu erwarten ist und wie ein Experiment gestaltet werden sollte. Deshalb ist die weitere Entwicklung des Modells eng verbunden mit der weiteren Entwicklung der experimentellen Hirnforschung.

## 7. Zusammenfassung

In der technischen Mustererkennung stellt die Figur/Hintergrundtrennung ein zentrales und weitgehend ungelöstes Problem dar. Es ist daher von großem Interesse, zu wissen, wie im visuellen Kortex Merkmale, die zu einem individuellen Objekt oder zu verschiedenen Objekten im Sehraum gehören, als zusammengehörig bzw. als nicht-zusammengehörig definiert werden. Einen möglichen Lösungsansatz bietet die Synchronisationshypothese, nach der Zusammengehörigkeiten im Sehraum durch synchrone neuronale Aktivität kodiert werden (Reitboeck 1980, von der Malsburg 1981) und die durch die Beobachtung synchroner Aktivität im Sehsystem unterstützt wurde (z.B. Eckorn et al. 1988).

In der vorliegenden Arbeit wird ein mehrschichtiges neuronales Netz entwickelt (*Kontur/Form-System*) und anhand dessen untersucht, inwiefern durch neuronale Verschaltungen und Vorgänge synchrone Aktivität zur Figur/Hintergrundtrennung gemäß der Synchronisationshypothese generiert werden kann. Dieses Netz gewinnt quasianaloge Konturinformation aus Gradienten unterschiedlicher Merkmalstypen in natürlichen visuellen Szenen (Intensitätskontrast, Texturkontrast) und liefert auf dieser Grundlage Repräsentationen möglicher Objekte durch (fast) synchrone neuronale Aktivität. Dazu gewinnt das Netz durch Vorwärtsverbindungen relevante Objektmerkmale aus Intensitätskontrasten und repräsentiert sie durch neuronale Aktivität. Sukzessive werden dabei Merkmale mit immer höherem Informationsgehalt extrahiert, indem zunächst Konturen und daraus Konturstücke mit starken Krümmungen (z.B. Ecken) gewonnen werden. Durch das Einfügen spezieller lateraler und rückgekoppelter Modulationsverbindungen lassen sich die Konturelemente über zeitliche Synchronisation neuronaler Aktivität zu Repräsentationen verschiedener Objektkandidaten gruppieren. In diesem Zusammenhang werden auch Annahmen über wahrscheinliche Objektverdeckungen gemacht. Die Separation verschiedener Objektkandidaten, die sich möglicherweise auch teilweise überdecken, geschieht durch deutlichen zeitlichen Versatz der jeweiligen Repräsentationen und wird durch inhibitorische Verbindungen ermöglicht. Die Computersimulationen zeigen für Objekte realer Szenen nicht-periodische Synchronisation neuronaler Aktivitäten.

In vielen komplexen Szenen reicht die Gewinnung und Gruppierung von Konturinformation, die aus Intensitätskontrasten erhalten wurde, allein nicht aus, um Objekte trennen zu können, so daß die Extraktion von Konturinformation aus anderen Merkmalsgradienten, wie z.B. Textur-Kontrast, notwendig wird. Daher wird ein weiteres neuronales Netz entwickelt, das in der Lage ist, einen Merkmalsgradienten höherer Ordnung (exemplarisch: Linien-Orientierungskontrast) zu gewinnen. Dieses Textur-Modul wird untersucht und anschließend in das *Kontur/Form-System* integriert. Es zeigt sich, daß das *Kontur/Textur/Form-System* in der Lage ist, Objekte aufgrund unterschiedlichen Orientierungskontrastes voneinander zu trennen. Hier wird somit ein Weg gezeigt, wie prinzipiell verschiedene Merkmalskontrast-Typen nebeneinander verarbeitet und dann in unterstützende Wechselwirkung treten können.

## 8. Literaturverzeichnis

**F. Attneave (1957):** *Physical determinants of the judged complexity of shapes.* Journal of Experimental Psychology, 53, pp. 221-227, 1957.

**R. Bauer, B.M. Dow (1989):** *Complementary global maps of orientation coding in upper and lower layers of the monkey's foveal striate cortex.* Exp. Brain Res., Vol. 76, pp. 503-509, 1989.

**K. Becker (1995):** *Computersimulation der Vertebratenretina mit biologienahen Neuronen.* unveröffentlichte Diplomarbeit, Universität Marburg, FB Physik, 1995.

**I. Biederman (1987):** *Recognition by components: a theory of human image understanding.* Psychological Review, Vol. 94, pp. 115-145, 1987.

**M. Blümler (1997):** *Modellierung höherer Verarbeitungsmechanismen in der Hörbahn der Laubheuschrecke Tettigonia viridissima mit dynamischen Modellneuronen.* Dissertation Universität Marburg, Cuvillier Verlag Göttingen, 1997.

**D. Boussaoud, R. Desimone, L.G. Ungerleider (1991):** *Visual topography of area TEO in the macaque.* Journal of comparative neurology, Vol. 306, pp. 554-575, 1991.

**V. Bozkov, Z. Bohdanecky, T. Radil-Weiss (1977):** *Target point selection during scanning eye movements.* Biological Cybernetics, Vol. 27, pp. 215-220, 1977.

**H.H. Bülthoff, S.Y. Edelman, M.J. Tarr (1995):** *How are three-dimensional objects represented in the brain ?* Cerebral Cortex, Vol. 5, No. 3, May/June 1995.

**R. Brause (1991):** *Neuronale Netze.* B.G. Teubner, Stuttgart.

**J.G. Daugman (1980):** *Two-dimensional spectral analysis of cortical receptive field profiles.* Vision Res. Vol.20, pp. 847-856, 1980.

**J.G. Daugman (1985):** *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters.* J. Opt. Soc. Am. A 2, pp. 1160-1169, 1985.

**R. Desimone, S.J. Schein (1987):** *Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form.* Journal of Neurophysiology, Vol. 57, pp. 835-867, 1987.

**C. Distler, D. Boussaoud, R. Desimone, L.G. Ungerleider (1993):** *Cortical connections in inferior temporal area TEO in macaque monkeys.* Journal of Comp. Neurol., Vol. 334, pp. 125-150, 1993.

**J.E. Dowling, B.B. Boycott (1966):** *Organization of the primate retina: Electron microscopy.* Proc. of the Royal Society of London, Vol. B166, pp. 80-111, 1966.

**D. Dunn, W.E. Higgins, J. Wakeley (1994):** *Texture segmentation using 2-D Gabor elementary functions.* IEEE Trans. Pattern Anal. Machine Intell. Vol.16, pp. 130-149, 1987.



- S. Edelman, S. Duvdevani-Bar (1997):** *A model of visual recognition and categorization.* Phil. Trans. Royal Soc. (B), Vol. 352, pp. 1191-1202, 1997.
- R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, H.J. Reitboeck (1988):** *Coherent oscillation: A mechanism of feature linking in the visual cortex.* Biological Cybernetics, Vol. 60, pp. 121-130, 1988.
- R. Eckhorn, H.J. Reitboeck, M. Arndt, P. Dicke (1990):** *Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex.* Neural Computation, Vol. 2, pp. 293-307, 1990.
- L. Fogel, D. Sagi (1989):** *Gabor filters as texture discriminator.* Biol. Cybern. 61, pp.103-113, 1989.
- G. Frank (1997):** *Ein digitales Hardwaresystem zur echtzeitfähigen Simulation biologienaher neuronaler Netze.* Dissertation, FB Elektrotechnik, Universität Paderborn, 1997.
- G. Frank, G. Hartman (1995):** *An artificial neural network accelerator for pulse-coded model neurons.* In: Proc. of the Intl. Conf. on Neural Networks (ICNN), pp. (CD-ROM), IEEE, 1995.
- G. Frank, G. Hartmann, A. Jahnke, M. Schaefer (1998):** *An accelerator for neural networks with pulse-coded model neurons.* Journal of Artificial Neural Networks, Special Issue on Pulse Coupled Neural Networks, in press, 1998.
- A.S. French, R.B. Stein (1970):** *A flexible neural analog using integrated circuits.* IEEE, Vol. BME-17, No. 3, pp. 248-253, 1970.
- K.S. Fu (1974):** *Syntactic methods in pattern recognition.* Mathematics in Science and Engineering, Vol. 112, Richard Bellman (ed.), Academic Press, New York, 1974.
- I. Fujita, K. Tanaka, M. Ito, K. Cheng (1992):** *A Columns for visual features of objects in monkey inferotemporal cortex.* Nature, Vol. 360, pp. 343-346, 1992.
- K. Fukushima (1980):** *Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.* Biol. Cybern., Vol. 36, pp. 193-202, 1980.
- D. Gabor (1946):** *Theory of communication.* J IEE (London) 93, pp. 429-457, 1946.
- A. Gabriel (1995):** *Persönliche Kommunikation.*
- A.P. Georgopoulos, A.B. Schwartz, R.E. Kettner (1986):** *Neural population coding of movement direction.* *Sensation and perception.* Science, Vol. 233, pp. 1416-1419, 1986.
- A.P. Georgopoulos, J. Ashe, N. Smyrnis, M. Taira (1992):** *The motor cortex and the coding of force.* Science, Vol. 256, pp. 1692-1695, 1992.
- A.P. Georgopoulos, M. Taira, A. Lukashin (1993):** *Cognitive neurophysiology of the motor cortex.* Science, Vol. 260, pp. 47-52, 1993.

- M.A. Goodale (1993):** *Visual pathways supporting perception and action in the primate cerebral cortex.* Current Opinion in Neurobiology, Vol. 3, pp. 578-585, 1993.
- E.B. Goldstein (1989):** *Sensation and perception.* Brooks/Cole, Pacific Grove, California, 1989.
- R.C. Gonzales, P. Wintz (1977):** *Digital image processing.* Reading, MA: Addison-Wesley, 1977.
- P. Gouras, J. Krüger (1979):** *Responses of cells in foveal visual cortex of the monkey to pure color contrast.* Journal of Neurophysiology, Vol. 42, pp. 850-860, 1979.
- N. Graham, J. Beck, A. Sutter (1992):** *Nonlinear processes in spatial-frequency channel models of perceived texture segregation: effects of sign and of contrast.* Vis. Res. Vol. 32, No. 4, pp. 719-743, 1992.
- N. Graham, A. Sutter, C. Venkatesan (1993):** *Spatial-frequency- and orientation-selectivity of simple and complex channels in region segregation.* Vis. Res. Vol. 33, No. 14, pp. 1893-1911, 1993.
- C.M. Gray, P. König, A.K. Engel, W. Singer (1989):** *Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties.* Nature, Vol. 338 (6213), pp. 334-337.
- C.G. Gross (1973):** *Visual functions of inferotemporal cortex.* In R. Jung (ed.), Handbook of Sensory Physiology, Vol. 7, Part 3B, pp. 451-482, Springer-Verlag, Berlin, 1973.
- S. Grossberg (1987):** *The adaptive brain II: Vision, speech, language, and motor control.* Elsevier Science Publishers B.V., Amsterdam, North-Holland, 1987.
- G. Hartmann (1982):** *Recursive features of circular receptive fields.* Biol. Cybern. 43, pp. 199-208, 1982.
- M.E. Hasselmo, E.T. Rolls, G.C. Baylis, V. Nalwa (1989):** *Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey.* Experimental Brain Research, Vol. 75, pp. 417-429, 1989.
- P. Heggelund (1981a):** *Receptive field organization of simple cells in cat striate cortex.* Exp. Brain Res. 42, pp. 89-98, 1981.
- P. Heggelund (1981b):** *Receptive field organization of complex cells in cat striate cortex.* Exp. Brain Res. 42, pp. 99-107, 1981.
- F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans (1992):** *Simulation of neural mechanisms from simple to end-stopped cells.* Vision Res., Vol. 32, pp. 963-981, 1992.
- D.H. Hubel, T.N. Wiesel (1962):** *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.* Journal of Psychology. Vol. 160, pp. 106-154, 1962.

- D.H. Hubel, T.N. Wiesel (1965):** *Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat.* Journal of Neurophysiology. Vol. 26, pp. 229-289, 1965.
- J.E. Hummel, I. Biederman (1992):** *Dynamic binding in a neural network for shape recognition.* Psychological Review, Vol. 99, pp. 480-517, 1992.
- E. Iwai (1978):** *The visual learning area in the inferotemporal cortex of monkeys.* In: Integrative Control Functions of the Brain, M. Ito (ed.), pp. 419-427, Tokyo: Kodansha, 1978.
- E. Iwai (1981):** *Visual mechanisms in the temporal and prestriate association cortices of monkeys.* Adv. Physiol. Sci., Vol. 17, pp. 279-286, 1981.
- E. Iwai (1985):** *Neurophysiological basis of pattern vision in macaque monkeys.* Vis. Res., Vol. 25, pp. 425-439, 1985.
- B. Julesz (1981a):** *Textons, the elements of texture perception, and their interactions.* Nature 290, pp. 91-97, 1981.
- B. Julesz (1981b):** *A theory of pre-attentive texture discrimination based on first-order statistics of textons.* Biol. Cybern. 41, pp. 131-138, 1981.
- B. Julesz (1986):** *Texton gradients: the texton theory revisited,* Biol. Cybern. 54, pp. 245-251, 1986.
- G. Kanizsa (1976):** *Subjective contours.* Scientific American, Vol. 234, pp. 48-52, 1976.
- H. Kolb, R. Nelson (1993):** *Rod pathways in the retina of the cat.* Vision Res., Vol. 23, pp. 301-312, 1993.
- K. Kopecz, K. Mohraz, R. Eckhorn (1996):** *Visualizing information processing in neural vision networks by backprojection detector characteristics to visual space.* Technical Report June 1996, Universität Marburg, FB Physik, 1996.
- S.W. Kuffler (1953):** *Discharge patterns and functional organization of the mammalian retina.* J. Neurophysiol. 16, 37-68.
- M.S. Landy, J.R. Bergen (1991):** *Texture segregation and orientation gradient.* Vision Res., Vol. 31, No. 4, pp. 679-691, 1991.
- G.W. Leshner, E. Mingolla (1993):** *The role of edges and line-ends in illusory contour formation.* Vision Res., Vol. 33, pp. 2253-2270, 1993.
- N.K. Logothetis, J. Pauls (1995):** *Psychophysical and psychological evidence for viewer centered representations in the primate.* Cereb. Cortex, Vol. 5, pp. 270-288, 1995.
- N.K. Logothetis, D.L. Sheinberg (1996):** *Visual object recognition.* Annu. Rev. Neurosci., Vol. 19, pp. 577-621, 1996.
- R. Maex, G.A. Orban (1996):** *Model circuit of spiking neurons generating directional selectivity in simple cells.* Journal of Neurophysiology, Vol.75, No.4, pp.1515-1545, 1996.

- H.A. Mallot, F. Giannakopoulos (1996):** *Population networks: a large-scale framework for modelling cortical neural networks*. Biological Cybernetics, Vol. 75, pp.441-452, 1996.
- S. Marcelja (1980):** *Mathematical description of the responses of simple cortical cells*. J. Opt. Soc. Am. 70, pp. 1297-1300, 1980.
- B.W. Mel (1997):** *SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition*. Neural Computation, Vol. 9, pp. 777-804, 1997.
- W. Metzger (1975):** *Gesetze des Sehens*. Waldemar Kramer Verlag, Frankfurt, 1975.
- A. D. Milner, M.A. Goodale (1993):** *Visual pathways to perception and action*. Progress in Brain Research, Vol. 95, pp. 317-337, 1993.
- M. Möller (1995):** *Leistungsvergleich verschiedener Neuronenmodelle bei grundlegenden Aufgaben der visuellen Informationsverarbeitung*. Unveröffentlichte Diplomarbeit, Universität Marburg, FB Physik, 1995.
- K. Mohraz (1996):** *Persönliche Kommunikation*.
- J.G. Nicholls, A.R. Martin, B.G. Wallace (1992):** *From neuron to brain*. Sinauer Associates, Inc., Sunderland, Massachusetts, 1992.
- H.C. Nothdurft (1991a):** *The role of local contrast in pop-out of orientation, motion and color*. Investigative Ophthalmology and Visual Science, 32, 714, 1991.
- H.C. Nothdurft (1991b):** *Texture segmentation and pop-out from orientation contrast*. Vision Res. Vol.31, pp. 1073-1078, 1991.
- H.C. Nothdurft (1992):** *Feature analysis and the role of similarity in pre-attentive vision*. Perception and Psychophysics. Vol.52, pp. 355-375, 1992.
- H.C. Nothdurft (1993):** *The role of features in pre-attentive vision: comparison of orientation, motion and color cues*. Vision Res. Vol.33, No.14, pp. 1937-1958, 1993.
- R. Ostkamp (1996):** *Bildvorverarbeitung realer Szenen mit dynamischen Modellneuronen zur kontextabhängigen Merkmalsextraktion, Merkmalskopplung und Szenensegmentierung*. Dissertation, FB Physik, Universität Marburg, 1996.
- R. Ostkamp, H.J. Reitboeck (1994):** *A biology-oriented neural network that uses similarities in position and orientation for preprocessing tasks*. Proceedings of the 22nd Göttingen Neurobiology Conference, Vol. 2, p. 852, 1994.
- M. Pauly (1995):** *persönliche Kommunikation*.
- M. Pauly, K. Kopecz, R. Eckhorn (1997):** *A dynamic gaze control model network of pulse-coding neurons concurrent treatment of saccades and smooth pursuit*. In: Proceedings of the 25th Göttingen Neurobiology Conference 1997, Vol. 2, N. Elsner (eds.), Thieme Verlag, Stuttgart, New York, p.1022, 1997.

- E. Peterhans, R. von der Heydt, G. Baumgartner (1986):** *Neural responses to illusory contour stimuli reveal stages of visual cortical processing.* In: Visual Neuroscience, J.D. Pettigrew, K.J. Sanderson, W.R. Levick (eds.), Cambridge University Press, Cambridge, pp. 343-351, 1986.
- E. Peterhans, R. von der Heydt (1989):** *Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps.* The Journal of Neuroscience, May 1989, 9 (5), pp. 1749-1763, 1989.
- N. Petkov, P. Kruizinga (1997):** *Computational models of visual neurons specialised in detection of periodic and aperiodic oriented visual stimuli: bar and grating cells.* Biological Cybernetics, Vol. 76, pp. 83-96, 1997.
- T. Poggio, S. Edelman (1990):** *A network that learns to recognize three-dimensional objects.* Nature, Vol. 343, pp. 263-266, 1990.
- K.R. Popper, J.C. Eccles (1977):** *Das Ich und sein Gehirn.* Piper, München, 1977.
- G.N. Reeke, jr., O. Sporns, G.M. Edelman (1990):** *Synthetic neural modeling: the "Darwin" series of recognition automata.* Proc. of the IEEE, Vol. 78, No. 9, pp. 1498-1530, 1990.
- H.J. Reitboeck (1980):** In L. Altmann (1982): *Psychophysische Untersuchungen zur Rolle der Synchronisation bei der Mustererkennung mit Hilfe einer mikroprozessorgesteuerten LED-Matrix.* Unveröffentlichte Diplomarbeit, Universität Marburg, FB Physik.
- H.J. Reitboeck (1983):** *A multi-electrode matrix for studies of temporal signal correlations within neural assemblies.* In: Basar et al., editor, Synergetics of the Brain, pp. 174-181, Springer, Berlin, 1983.
- H.J. Reitboeck (1989):** *Invariances in pattern recognition.* In: Visuomotor Coordination. J.P. Ewert, M.A. Abib, editors, Plenum Publishing Cooperation, pp. 383-396, 1989.
- H.J. Reitboeck, T.P. Brody (1969):** *A transformation with invariance under cyclic permutation for applications in pattern recognition.* Inf. Control., Vol. 15, pp. 130-154, 1969.
- H.J. Reitboeck, R. Eckhorn, M. Pabst (1987):** *A model of figure/ground separation based on correlated neural activity in the visual system.* Computational Systems - Natural and Artificial, Haken, Springer, 1987.
- H.J. Reitboeck, M. Pabst, R. Eckhorn (1988):** *Texture description in the time domain.* Computer Simulation in Brain Science, Cotterill, Phys. Scripta, Univ. Press, Cambridge, 1988.
- R.W. Rodieck (1965):** *Quantitative analysis of cat retinal ganglion cell response to visual stimuli.* Vision Res., Vol. 5, pp. 583-601, 1965.
- M. Schäfer (1995):** *Systemintegrierte Hardware-Entwicklung und Inbetriebnahme eines multifunktionalen Arithmetik- und Speichermoduls.* Unveröffentlichte Diplomarbeit, Fachgebiet Grundlagen der Elektrotechnik, Universität-GH Paderborn, 1995.

**P.H. Schiller, B.L. Finlay, S.F. Volman (1976):** *Quantitative studies of single cell properties in monkey striate cortex.III. Spatial frequencies*, Journal of Neurophysiol., Vol. 39, pp. 1334-1351.

**U. Schott (1995):** *Persönliche Kommunikation.*

**U. Schott (1997):** *Persönliche Kommunikation.*

**U. Schott, R. Eckhorn, H.J. Reitboeck (1996):** *Extracting Motion Contrast by a Neural Network of Spiking Neurons.* In: Proceedings of the 24th Göttingen Neurobiology Conference 1996, Vol. 2, N. Elsner, H. Schnitzler (eds.), Thieme Verlag, Stuttgart, New York, p.795, 1996.

**U. Schott, R.Eckhorn, H.J.Reitboeck (1997):** *Motion Capture by Contour Terminators via Feature Linking. A Simulation of the Barberpole Illusion with a Neural Network of Spiking Neurons.* In: Proceedings of the 25th Göttingen Neurobiology Conference 1997, Vol. 2, N. Elsner (eds.), Thieme Verlag, Stuttgart, New York, p.1025, 1997.

**R.F. Schmidt (1985):** *Grundriß der Sinnesphysiologie.* Springer, Berlin, 1985.

**O.G. Selfridge, U. Neisser (1960):** *Pattern recognition by machine.* Scientific American, Vol. 203, pp. 60-68, 1960.

**A.C. Shaw (1968):** *The formal description and parsing of pictures.* Rep. SLAC-84, Stanford Linear Accelerator Center, Stanford Univ., California, 1968.

**R. Siminoff (1980):** *Modelling the vertebrate visual system. I. Wiring diagram of the cone retina.* J. theor. Biol., 86, pp. 673-708, 1980.

**O. Sporns, G. Tononi, G.M. Edelman (1991):** *Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections.* Proceedings of National Academy of Science of the United States of America, Vol. 88, pp. 129-133, 1991.

**C. Spengler (1995):** *Persönliche Kommunikation.*

**C. Spengler (1996):** *Relationale Kodierung von Objektmerkmalen für die visuelle Mustererkennung in biologienahen neuronalen Netzen.* Unveröffentlichte Diplomarbeit, Universität Marburg, FB Physik, 1996.

**C. Spengler, U. Schott, L. Weitzel, R. Eckhorn (1996):** *A model of visual cortical orientation tuning, including modulation by mutual synchronization among spiking neurons of iso-orientation preferences and inhibition among different orientations.* In: Proceedings of the 24th Göttingen Neurobiology Conference 1996, Vol. 2, N. Elsner, H. Schnitzler (eds.), Thieme Verlag, Stuttgart, New York, p. 792, 1996.

**M. Stoecker (1993):** *Höhere Mechanismen der visuellen Informationsverarbeitung in neuronalen Netzen.* Dissertation, FB Physik, Universität Marburg, 1993.

**K. Tanaka (1993):** *Neural mechanisms of object recognition.* Science, Vol. 262, pp. 685-688, 1993.

**M.R. Turner (1986):** *Texture discrimination by Gabor functions.* Biol. Cybern. 55, pp. 71-82, 1986.

**S. Ullman (1989):** *Aligning pictorial descriptions: an approach to object recognition.* Cognition Vol. 32, pp. 193-254, 1989.

**S. Ullman, R. Basri (1991):** *Recognition by linear combinations of models.* IEEE Trans. of Pattern Analysis and Machine Intelligence, Vol. 13, No. 10, 1991.

**L.G. Ungerleider, M. Mishkin (1982):** *Two cortical visual systems.* In: Analysis of visual behavior, D.J. Ingle (ed.), pp. 549-586, Cambridge, MA: MIT Press, 1982.

**J.D. Victor, K. Purpura, E. Katz, B. Mao (1994):** *Population encoding of spatial frequency, orientation, and color in macaque VI.* Journal of Neurophysiology, Vol. 72, No. 5, 1994.

**R. Vogels (1990):** *Population coding of stimulus orientation by striate cortical cells.* Biological Cybernetics, Vol. 64, pp. 25-31, 1990.

**R. von der Heydt, E. Peterhans, M.R. Dürsteler (1991):** *Grating-cells in monkey visual cortex: coding texture.* In: B. Blum (ed.), Channels in the visual nervous system: neurophysiology, psychophysics and models, Freund, London, pp. 53-73, 1991.

**R. von der Heydt, E. Peterhans, M.R. Dürsteler (1992):** *Periodic-pattern-selective cells in monkey visual cortex.* Journal of Neuroscience, Vol. 12, pp. 1416-1434, 1992.

**C. von der Malsburg (1981):** *The correlation theory of brain function.* Technical Report Internal Report 81-2, Abteilung für Neurobiologie, MPI für biophysikalische Chemie, Göttingen, 1981.

**C. von der Malsburg, W. Schneider (1986):** *A neural cocktail-party processor.* Biol. Cybern., Vol. 54, pp. 29-40, 1986.

**C. von der Malsburg, J. Buhmann (1992):** *Sensory segmentation with coupled neural oscillators.* Biological Cybernetics, Vol. 67, pp. 233-242, 1992.

**von Foerster (1966):** *Neural Networks.* Biological Computer Laboratory, Report No. 5.1, 1. April 1966, p. 46.

**L. Weitzel (1995):** *Bildsegmentierung durch Texturanalyse mit neuronalen Netzen.* Unveröffentlichte Diplomarbeit, Universität Marburg, FB Physik, 1995.

**L. Weitzel, M. Stoecker, U. Schott, H.J. Reitboeck (1995):** *Feature contrast: II. A neural network for texture segmentation via orientation contrast detection.* In: Proceedings of the 23th Göttingen Neurobiology Conference 1995, N. Elsner, Menzel (eds.), Thieme Verlag Stuttgart, New York, p. 890, 1995.

**L. Weitzel, R. Ostkamp, R. Eckhorn, H.J. Reitboeck (1996):** *Definition of Complex Visual Object Boundaries by Synchronized Activities in a Linking Architecture of Spiking Neurons.* In: Proceedings of the 24th Göttingen Neurobiology Conference 1996, Vol. 2, N. Elsner, H. Schnitzler (eds.), Thieme Verlag, Stuttgart, New York, p. 793, 1996.

- L. Weitzel, K. Kopecz, C. Spengler, R. Eckhorn (1997a):** *Contour Based Figure-Ground Segregation of Real World Images Modelled by a Linking Architecture of Pulse-Coding Neurons*. In: Proceedings of the 25th Göttingen Neurobiology Conference 1997, Vol. 2, N. Elsner (eds.), Thieme Verlag, Stuttgart, New York, p.1027, 1997.
- L. Weitzel, K. Kopecz, C. Spengler, R. Eckhorn, H.J. Reitboeck (1997b):** *Contour Segmentation with Recurrent Neural Networks of Pulse-Coding Neurons*. In: G. Sommer, K. Daniilidis, J. Pauli: Computer Analysis of Images and Patterns, Springer Verlag, pp. 337-344, 1997.
- M. Wertheimer (1923):** *Untersuchungen zur Lehre von der Gestalt II*. Psychologische Forschung, Vol. 4, pp. 301-350, 1923.
- P. Winston (1975):** *Learning structural descriptions from examples*. In: P. Winston (ed): The psychology of computer vision, New York: McGraw-Hill, pp. 157-209, 1975
- C. Wolff (1996):** *Entwicklung und Integration einer Prozessoreinheit zur Berechnung der Neuronenaktivitäten eines biologienahen neuronalen Netzwerksimulators*. Unveröffentlichte Diplomarbeit, Fachgebiet Grundlagen der Elektrotechnik, Universität-GH Paderborn, 1996.
- C. Wolff (1998):** *Persönliche Kommunikation*.
- R.A. Young (1986):** *The gaussian derivative model for machine vision: visual cortex simulation*. Journal of the optical society of america, GMR-5323, 1986.
- R.A. Young (1988):** *The gaussian derivative model for machine vision: I. retinal mechanisms*. Spatial Vision 2(4), pp. 273-293, 1988.
- A. Zell (1994):** *Simulation neuronaler Netze*. Addison-Wesley, Bonn, Paris, Reading Massachusetts, 1994.
- L. Zusne, K. Michels (1964):** *Non-representational shapes and eye movements*. Perception Motor Skills, Vol. 18, pp. 11-20, 1964.



## **Danksagung**

Ich möchte mich an erster Stelle bei meinem Doktorvater Herrn Prof. H.J. Reitböck und bei Herrn Prof. R. Eckhorn bedanken, die mir die Forschung in der Arbeitsgruppe Angewandte Physik: Neurophysik ermöglichten und mich bei der Erstellung meiner Doktorarbeit unterstützten.

Für die besonders gute Zusammenarbeit, die vielen fruchtbaren Diskussionen und die angenehme Arbeitsatmosphäre danke ich meinen Zimmerkollegen U. Schott, C. Spengler und J. Behring.

Mein besonderer Dank gilt T.C. Wölbern und H.J. Brinksmeyer, die mir nicht nur in "aufbauenden Definitionsfragen" zur Seite standen.

Weiterhin möchte ich den "Simulanten" M. Pauly, K. Mohraz, M. Saam, K. Kopecz und A. Gabriel für vielfältige Unterstützung - z.B. bei Rechnerproblemen - danken.

Für die ergiebigen Ferndiskussionen danke ich den ehemaligen Mitgliedern der Arbeitsgruppe M. Stöcker und R. Ostkamp.

Frau S. Thomas danke ich vor allem für die Mühen in organisatorischen und formellen Fragen.

Schließlich gilt mein Dank allen Mitgliedern der AG Neurophysik und der AG Grundlagen, die mir durch ein angenehmes Arbeitsklima das Erstellen der Arbeit erleichtert haben.