Elementare kognitive Mechanismen in neuronalen Architekturen

Förderkennzeichen 413-5839-01 IN 101D

Abschlußbericht

Auftragnehmer:

Prof.Dr.-Ing. Horst-Michael Groß

Institut für Theoretische und Technische Informatik
Fachgebiet Neuroinformatik
Technische Universität Ilmenau
98684 Ilmenau

Inhaltsverzeichnis

Projektzusammenfassung

Das im Rahmen dieses Berichtes vorgestellte Projekt *"Elementare kognitive Mechanismen in neuronalen Architekturen"* war Bestandteil des im Zeitraum von 1991 bis 1994 geförderten BMFT-Verbundprojektes NAMOS - *"Neuronale Architekturprinzipien für Autonome MObile Systeme"*. Gegenstand des Teilprojektes war die Entwicklung eines biologisch motivierten, verallgemeinerbaren Modellansatzes zur *Steuerung des autonomen Wissenserwerbs* für visuell explorierende, lernfähige visuo-motorische Systeme.

Basierend auf bekannten Konzepten aus Neurophysiologie, Neuroanatomie und Psychophysik wurden eine Arbeitshypothese und eine cortikal motivierte Modellarchitektur zur *on-line Selbstorganisation eines impliziten Objektverständnisses* im Kontext eines variablen Explorationsverhaltens bei der Analyse komplex strukturierter natürlicher Farbszenen entwickelt. In Ergänzung und Abgrenzung zu den auf "Active-vision"-Systemen basierenden Arbeiten im Rahmen des NAMOS-Verbundes wurde der hier vorgestellte Modellansatz allerdings weniger am Phänomen der sakkadischen Bildanalyse als an den Modellvorstellungen zum *internen Scanning fovealisierter Bilder* mit Mechanismen der selektiven visuellen Aufmerksamkeit, insbesondere der "Searchlight-Methapher", orientiert. In Erweiterung dieser Modellvorstellungen ging es im Rahmen dieses Projektes vor allem darum, grundlegende Aspekte der on-line Selbstorganisation von Wissen im Kontext eines attentiv steuerbaren räumlich-zeitlichen Szenenabtastverhaltens zu untersuchen und für autonom lernfähige visuo-motorische Systeme nutzbar zu machen.

Nach der dem Projekt zugrundeliegenden Arbeitshypothese kommt es im Verlaufe von Szenenanalysevorgängen auf der Grundlage von on-line Lernprozessen zu einer Veränderung des Abtastverhaltens im Sinne einer *Effektivierung des Abtastverlaufs*, was als Ausdruck für die fortschreitende Selbstorganisation eines *"impliziten Objektverständnisses"* interpretiert werden kann. Unter Effektivierung soll eine *aktive räumlich-zeitliche Umrangierung* des Abtastverlaufs verstanden werden, deren Ziel darin besteht, die in verschiedenen Szenen in stabilen räumlichen Relationen auftretenden und mit dem Aufmerksamkeitsfokus erfaßbaren, signifikanten Komponenten zusammengehöriger Bildstrukturen *zeitlich unmittelbar aufeinanderfolgend* abzutasten, um die intern ablaufenden Wahrnehmungs- und Entscheidungsprozesse deutlich beschleunigen zu können und damit den Analyseaufwand zu minimieren. Stabile objektspezifische räumliche Zusammenhänge werden dadurch in stabile zeitliche Zusammenhänge innerhalb des Abtastprozesses transformiert. Diese *zeitliche Nachbarschaft*, eine Vorstufe von zeitlicher Synchronität, wird im Rahmen der Arbeitshypothese als einziges, systemintern auswertbares Kriterium für Zugehörigkeit zu *ein und demselben* Szenenbestandteil angesehen und als *implizites, verhaltensorientiertes Verständnis* darüber interpretiert, welche lokalen Bildstrukturen unter zahlreichen anderen auffälligen Bildstrukturen zu einem Objekt gehören könnten. Durch *Gruppierung der etablierten zeitlichen Zusammenhänge* in einem weiteren Lernprozeß wird das System in die Lage versetzt, selbständig *Objekthypothesen* zu generieren und auf der Basis einer *"Senso-*

risch gesteuerten internen Simulation" Abtastverläufe zu antizipieren und gezielt zu verifizieren.

Da der Schwerpunkt des Projektes einerseits auf die Modellierung fundamentaler Architekturprinzipien zur Selbstorganisation eines "verhaltensorientierten Objektverständnisses" gelegt wurde, andererseits aber auch möglichst reale Farbszenen als Stimuli zum Einsatz kommen sollten, mußte bezüglich der Problemkomplexität ein Kompromiß derart gefunden werden, daß Rotations-, Skalierungs- und Perspektive-Aspekte bewußt ausgeklammert wurden. Stattdessen wurde die Aufmerksamkeit darauf gerichtet, die angestrebte Verhaltensleistung auch bei komplexen Hintergrundsituationen und willkürlichen Szenenarrangements sowie bei stark schwankenden Beleuchtungsverhältnissen zu erbringen.

Eine elementare Voraussetzung für einen derartigen, unter realen Bedingungen ablaufenden Selbstorganisationsprozeß ist eine möglichst gute *Reproduzierbarkeit des datengetriebenen Abtastverlaufs* bei der Analyse natürlicher Farbszenen – auch unter schwankenden Beleuchtungsverhältnissen und bei veränderlichen Szenarien. Daher beschäftigten sich im Rahmen des Projektes umfangreiche Arbeiten mit dem Problem der *Robustheit der implementierten Auffälligkeitsmechanismen*, wobei ein besonderer Schwerpunkt auf die Verarbeitung der Farbinformationen in einem speziellen, biologisch motivierten Farbraum gelegt wurde. Basierend auf neurophysiologischen Modellvorstellungen zur Farbverarbeitung in der Retina und im visuellen System wurde eine *neuronale Farbverarbeitungsarchitektur* entwickelt, die in einem neuartigen, physiologisch motivierten RG-YB-BW-Farbdifferenzraum eine sich selbständig *adaptierende Farbumstimmung* realisiert. Diese Farbumstimmung ist ein sehr leistungsfähiges, referenzfrei operierendes Verfahren, das auch unter stark schwankenden Beleuchtungsverhältnissen (bezüglich der spektralen Zusammensetzung) eine sehr flexible Korrektur der Farbinformationen ermöglicht.

Die Tragfähigkeit der Arbeitshypothese und die Leistungsfähigkeit der Modellarchitektur bezüglich der angestrebten Veränderung des Abtastverhaltens im Verlaufe eines on-line ablaufenden Selbstorganisationsprozesses konnten im Rahmen von Systemsimulationen über einer Galerie von realen Farbszenen erfolgreich demonstriert werden. Anhand von exemplarischen Szenenbeispielen kann die *Scan-Effektivierung* verdeutlicht werden, dies ist Ausdruck dafür, daß die *inhärenten Gesetzmäßigkeiten* der in den Szenen enthaltenen komplexen Bildstrukturen durch aktives Lernen gezielt extrahiert und im Rahmen der *Sensorisch gesteuerten internen Simulation* des nachfolgenden Abtastverlaufs selektiv genutzt werden können. Insofern erwies sich die im Rahmen des Projektes entwickelte Lösung als ein sehr leistungsfähiger Ansatz, der einem visuomotorischen System auch in stark strukturierter visueller Umgebung durch zeitliche Bindung stabiler objektspezifischer Zusammenhänge eine inhärente *Objekt/Objekt- und Objekt/Hintergrund-Segmentierung* ermöglicht und auch ohne eine spezielle Trainingsphase das Erlernen unbekannter visueller Strukturen erlaubt.

Kapitel 1

Elementare kognitive Mechanismen in neuronalen Architekturen – Selbstorganisation eines verhaltensorientierten Objektverständnisses

1.1 Biologische Informationsverarbeitung als Organisation von Verhalten – Prämissen des Ansatzes

Das Gehirn als zentrale informationsverarbeitende Struktur biologischer Systeme hat sich im Lauf der Evolution in stetiger Auseinandersetzung des biologischen Systems mit seiner Umwelt entwickelt. Dabei führte die Evolution zu einer gegenseitigen Anpassung von Sinnesorganen, Gehirnen und Verhaltensweisen und konservierte dabei die Entstehungsgeschichte in der Organisation des informationsverarbeitenden Apparates [Mallot et al. 1992]. Wesentlich ist, daß die aus der Umwelt mittels verschiedener Sensoriken aufgenommenen Informationen durch eine komplexe räumlich-zeitliche Charakteristik gekennzeichnet sind und nur in Ausnahmefällen ausschließlich singuläre sensorische Ereignisse bedeutungstragend für das Verhalten des Systems sind. Da in einem biologischen informationsverarbeitenden System jede Art der Repräsentation der Umwelt niemals arbiträr ist, sondern immer dem Ziel einer möglichst erfolgreichen Auseinandersetzung mit der Umwelt dient, kann zumindest bis zu einem bestimmten Grad aus der Charakteristik der sensorischen Informationen aus der Umwelt auf notwendige Verarbeitungsprinzipien und -mechanismen des biologischen Systems geschlossen werden. Mallot [Mallot et al. 1992] bezeichnet diese enge Wechselbeziehung zwischen externen und internen Datenformaten als *Isomorphie zwischen inneren und äußeren Prozessen.*

Biologische Systeme sind gekennzeichnet durch ein hohes Maß an Flexibilität, das heißt durch die Fähigkeit, unter variablen Randbedingungen gute und effektive Problemlösungsstrategien zu entwickeln. Sie zeigen *adaptives Verhalten* als Ausdruck eines flexiblen, effektiven und sinnvollen Interagierens mit der Umwelt, welches nach Beer [Beer 1991] bereits gleichbedeutend ist mit Intelligenz.

Eine entscheidende Voraussetzung für ein in unbekannter Umgebung operierendes System ist die Fähigkeit zum *autonomen Wissenserwerb.* Autonomes Lernen erfordert eine kontinuierliche Wechselwirkung zwischen Eingangsdatenstrom und systeminternem Entscheidungsprozeß, was mit dem Begriff der *sensorisch gesteuerten internen Simulation* (siehe auch [Groß et al. 1992a] und [Boehme et al. 1994]) charakterisiert werden soll. Bei diesem Prozeß der Interaktion mit der Umwelt erfolgt eine *verhaltensorientierte Bewertung der aktuellen sensorischen Situation.* Die Akquisition von Wissen ist also stets handlungsorientiert und eingebettet in einen Prozeß der Selbstorganisation, da neben der Eignung des bereits akkumulierten Wissens zur effektiven Auseinandersetzung mit der Umwelt keine weitere Bewertung des Lernerfolgs zur Verfügung steht. Kontextinformation und bereits akkumuliertes Wissen (Erfahrung) fließen daher mit in den Entscheidungsprozeß ein. Interne Hypothesen werden gebildet und über den aktiven Eingriff in den Prozeß der Informationsaufnahme verifiziert. Diese aktive Einflußnahme im Sinne einer selektiven Steuerung der Informationsauswahl gestattet dem System eine Effektivierung der Entscheidungsfindung, was wiederum eine Verbesserung der Fitness des Systems in Bezug auf seine Interaktion mit der Umwelt bewirkt.

1.2 Handlungs- und Verhaltensszenario

Für ein in unbekannter Umgebung explorierendes System erscheint die Detektion von stabil auftretenden, typischen sensorischen Situationen notwendig, da auf deren Basis bereits einfache Verhaltensweisen generiert werden können. Unter diesem Blickwinkel stellt sich der Problemkreis der Objekterkennung im Vergleich zum technischen Standpunkt für ein biologisches System völlig anders dar. Geht es bei der technischen Objekterkennung um das Auffinden meist bereits a priori bekannter und mehr oder weniger komplex kodierter Strukturen in Bildern ohne Verhaltensbezug, so besteht für das biologische System das Ziel vielmehr in der Extraktion verhaltensrelevanter Information. Beispiele für solche Verhaltensweisen in Bezug auf Objekte könnten die Vermeidung von Hindernissen, die Orientierung, die Bestimmung der eigenen Position oder ein zielgerichteter Zugriff sein.

Da einem biologischen System Objekte seiner Umwelt a priori unbekannt sind, muß es im Laufe der Auseinandersetzung mit seiner Umwelt ein *verhaltensorientiertes Objektverständnis* entwickeln. Die Interaktion mit der Umwelt soll sich im Rahmen dieses Projektes ausschließlich in Form eines *Abtastprozesses über einer visuellen Szene* vollziehen, der die Szene in einen räumlich-zeitlichen Datenstrom transformiert. Bei dieser Abtastung soll es sich nicht um ein externes Scanning mit dabei ausgeführten Augenbewegungen, sondern um ein *internes Scanning* als selektiven visuellen Aufmerksamkeitsprozeß handeln. Derartige Active-Vision-Prozesse können bereits als einfache Form der Interaktion mit der Umwelt angesehen werden. Obwohl die Prozesse des externen Scanning im Rahmen dieses Projektes nicht Gegenstand der Untersuchungen waren, sind die zahlreichen Bezüge zwischen externem und internem Scanning im Projektvorhaben berücksichtigt worden.

Selektive Aufmerksamkeitsprozesse (sowohl Active-Vision als auch das interne Scanning ohne Augenbewegungen) stellen Basismechanismen dar, die es gestatten, den hochparallelen visuellen Input in mehr oder weniger relevante Substrukturen zu zerlegen. Diese Zerlegung ist notwendig, um eine Szene nicht als „paralleles Ganzes“ repräsentieren zu müssen, sondern selektiv auf die Details zugreifen zu können, die für den aktuellen Interpretationsprozeß wesentlich sind. Damit wird einerseits die Datenflut enorm reduziert, andererseits sind nur auf diesem Wege flexible Gruppierungs- bzw. Bindingprozesse realisierbar, die auf der Grundlage des bereits akkumulierten Wissens eine Einordnung bekannter Szenenbestandteile in neue Arrangements ermöglichen.
Die Interaktion mit der Umwelt in Form dieses internen Scannings wurde bewußt gewählt, da der Schwerpunkt auf der Entwicklung neuronaler Architekturprinzipien und Verarbeitungsmechanismen lag, die eine effiziente Interaktion mit der Umwelt erst ermöglichen. Außerdem sollte die Interaktion mit der Umwelt eine externe Beobachtbarkeit erlauben, ansonsten aber so einfach wie möglich gehalten sein.

Als *Szenario* wurde die *Analyse visueller, stationärer 2D-Farbszenen* gewählt. Diese Szenen, in denen vor variablem, strukturiertem Hintergrund Objekte, die in ihrer Anordnung und ihrem Auftreten von Szene zu Szene variieren, enthalten sind, bilden die

Umwelt für das implementierte System.

Eine verbesserte Fitness des Systems aufgrund bereits akkumulierten Wissens muß sich in einer signifikanten, extern beobachtbaren Verhaltensänderung ausdrücken. Für den vorliegenden Fall bedeutet das, daß die in den Szenen enthaltenen komplexen visuellen Strukturen im Verlauf des Lernprozesses immer effizienter detektiert und identifiziert werden können. Das Abtastverhalten soll sich von einer anfänglich ungeordneten, „ziellosen" Abtastfolge hin zur bevorzugten Abtastung dieser stabilen visuellen Strukturen entwickeln. Dabei soll die Veränderung des Abtastverhaltens als Ausdruck für das sich entwickelnde Objektverständnis gewertet werden, da der anfänglich rein datengetriebene Abtastverlauf mehr und mehr durch den internen Entscheidungsprozeß und somit durch das bislang akkumulierte Wissen über die Umwelt (Szenen) beeinflußt wird.

Eine Grundvoraussetzung für die angestrebte adaptive Verhaltensänderung ist die robuste und vor allem unter variablen Randbedingungen möglichst reproduzierbare Extraktion von Informationen über lokale Bildgebiete („Elementarinformationen") aus den Farbszenen. Kann die Stabilität dieser Elementarinformation nicht oder nur unzureichend gewährleistet werden, so kann das System nicht in die Lage versetzt werden, stabil auftretende Merkmalskombinationen im Verlaufe des Lernprozesses zu extrahieren. Da die angestrebte sensorisch gesteuerte interne Simulation jedoch darauf angewiesen ist, einen neuen Eingang immer aufgrund des bereits akkumulierten Systemwissens zu interpretieren (Zerlegung des parallelen Inputs in mehr oder weniger bekannte Komponenten), wäre für den Fall vorwiegend instabiler Merkmale dieser Interpretationsprozeß nicht mehr sinnvoll möglich. Somit ließe sich eine kombinatorische Explosion möglicher Entscheidungsvarianten durch den Prozeß der sensorisch gesteuerten internen Simulation nicht verhindern.

Die mit dem Projektvorhaben verfolgten Zielrichtungen einer möglichst durchgängig biologisch-plausiblen Modellarchitektur einerseits und der Anwendung dieser Architektur auf eine möglichst reale Problemstellung andererseits mußten zwangsläufig zu Einschränkungen auf beiden Seiten führen. Daher soll die entwickelte Architektur nicht dem Anspruch auf ein umfassendes Modell des visuellen Systems erheben, da vorrangig die Teilaspekte der visuellen Verarbeitung berücksichtigt wurden, die vor dem Hintergrund der angestrebten Verhaltensleistung relevant waren.

1.3 Verhaltensorientiertes Objektverständnis – Zeitliche Objektrepräsentation

Für den angestrebten Selbstorganisationsprozeß wurden nachfolgende Überlegungen vorangestellt:
Die im gewählten Szenario enthaltenen Objekte sind hinreichend strukturiert und gekennzeichnet durch stabile räumliche Relationen zwischen lokalen Bildstrukturen. Sie variieren von Szene zu Szene in ihrer Anordnung und ihrem Auftreten. Bewußt ausgeklammert wurden Rotations- und Skalierungsaspekte sowie perspektivische Objektverzerrungen, da in diesen Fällen keine reproduzierbaren Merkmale für ein und dasselbe Objekt in verschiedenen Szenen extrahierbar wären, was die angestrebte Detektion stabiler visueller Strukturen durch den Selbstorganisationsprozeß unmöglich machen würde.
Da der Schwerpunkt des Projektvorhabens von Beginn an auf der Entwicklung und Untersuchung fundamentaler Lern- und Architekturprinzipien lag, mußten diese Einschränkungen, die zwangsläufig auf Kosten eines umfassenden und absolut realitätsnahen „Systementwurfs“ gingen, in Kauf genommen werden.

Zusammengehörende, stabil auftretende visuelle Strukturen im Sinne von Objekten sind dem System a priori nicht bekannt. Die Aufgabe besteht also darin, mit Hilfe des internen Scanprozesses Wissen über derartige Strukturen in selbstorganisierender Art und Weise zu akquirieren. Dazu muß die beim Scanprozeß über den Szenen aufgenommene Information (räumlich-zeitlicher Datenstrom) intern geeignet repräsentiert und für die Interaktion mit der Umwelt nutzbar gemacht werden.
An dieser Stelle soll die Isomorphie zwischen internen und externen Datenformaten wieder aufgegriffen werden. Überträgt man diesen Ansatz auf die vorliegende Problemstellung, so wird deutlich, daß im System eine *zeitliche Objektrepräsentation* entstehen muß, die den anfänglich rein datengetriebenen Analyseprozeß intern adäquat abbildet, um den Abtastprozeß geeignet beeinflussen zu können. Auf der Grundlage dieser internen zeitlichen Repräsentation muß das System nun versuchen, relevantes, d.h. eine Verbesserung der Fitness ermöglichendes Wissen, über seine Umwelt zu extrahieren. Verbesserte Fitness bedeutet im vorliegenden Fall eine Effektivierung des Scanprozesses und eine Beschleunigung des internen Entscheidungsprozesses, wie dies Abbildung 1.1 schematisch verdeutlicht.

Auf die Bedeutung der selektiven Aufmerksamkeit bei der Analyse komplex strukturierter visueller Szenen wurde im vorhergehenden Abschnitt bereits kurz eingegangen. Selektive Aufmerksamkeit gründet sich auf zwei wesentliche Prozesse. Diese sind die Ermittlung von auffälligen Szenengebieten (datengetrieben, bottom-up) und die Erzeugung von Erwartung (wissensbasiert, top-down). Da zu Beginn der Auseinandersetzung mit der Umwelt noch kein Wissen über diese vorliegt, zeigt das System zunächst rein datengetriebenes Abtastverhalten (Abbildung 1.1 links). Durch die Nutzung des intern repräsentierten Analyseprozesses zur Hypothesenbildung und die Verifizierung der Hypothesen in Form einer Modulation des Abtastverlaufs soll das System die lokalen

Bildstrukturen und deren räumliche Relationen erlernen, die in seiner Umwelt stabil auftreten. Das sich dabei entwickelnde emergente Objektverständnis soll sich dann in einer charakteristischen, zeitlichen Abtastung der Objekte als stabile visuelle Strukturen der Umwelt äußern. Dies bedeutet eine Umrangierung innerhalb des Abtastpfades, so daß zu einem Objekt gehörende Bildgebiete bevorzugt zeitlich unmittelbar aufeinanderfolgend abgetastet werden, wie dies in Abbildung 1.1 rechts in schematischer Form zu sehen ist.

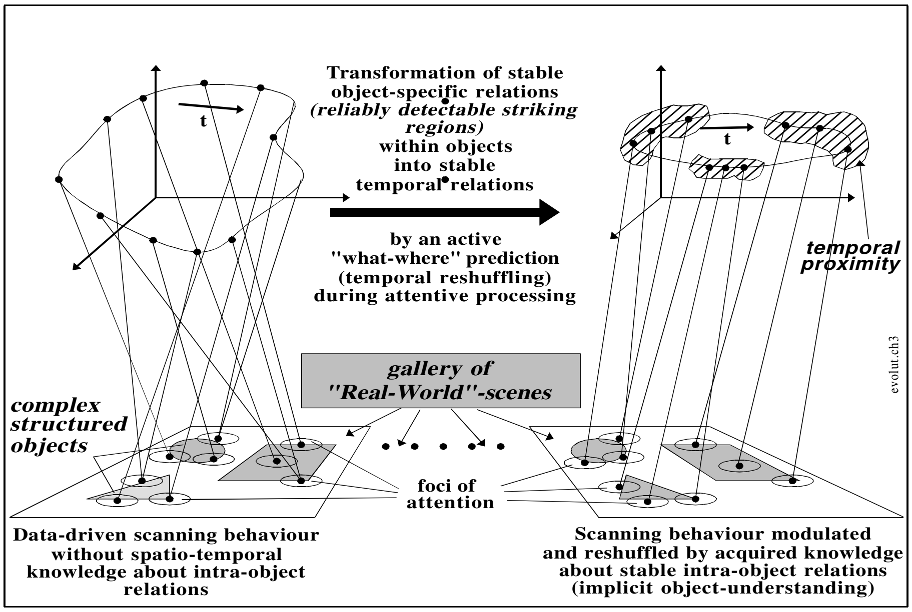


Abbildung 1.1: *Umrangierung bedeutungsmäßig zusammengehörender Szenenbestandteile in eine unmittelbare zeitliche Nachbarschaft beim Analyseprozeß*

Die Umrangierung der zu einem Objekt gehörenden Bestandteile in die unmittelbare zeitliche Nachbarschaft beim Analyseprozeß stellt nach unserer Auffassung eine elementare Verarbeitungsleistung dar, die einen wesentlichen Beitrag zur Lösung des Binding-Problems liefern kann. Weiterhin ermöglicht eine Umgruppierung elementarer Szenenbestandteile zu unterschiedlichen Bedeutungsinhalten eine kontinuierliche Modifikation des Systemwissens. Dies garantiert, daß eine Interpretation des aktuellen Inputs unter Nutzung des bereits vorhandenen Systemwissens überhaupt erst möglich wird.
Der durch die selektive Aufmerksamkeit realisierbare wahlweise Zugriff auf elementare Szenenbestandteile sichert dem System die Flexibilität, die notwendig ist, um beim Interpretationsprozeß eine kombinatorische Explosion möglicher Entscheidungsvarianten zu vermeiden.
Übertragen auf das gewählte Szenario, die Analyse komplex strukturierter Farbszenen, soll die Umgruppierung bedeutungsmäßig zusammengehörender Szenenbestandteile in die unmittelbare zeitliche Nachbarschaft beim Analyseprozeß die in Abbildung 1.2 dargestellte Verhaltensänderung bewirken, die dann auch extern beobachtbar ist.

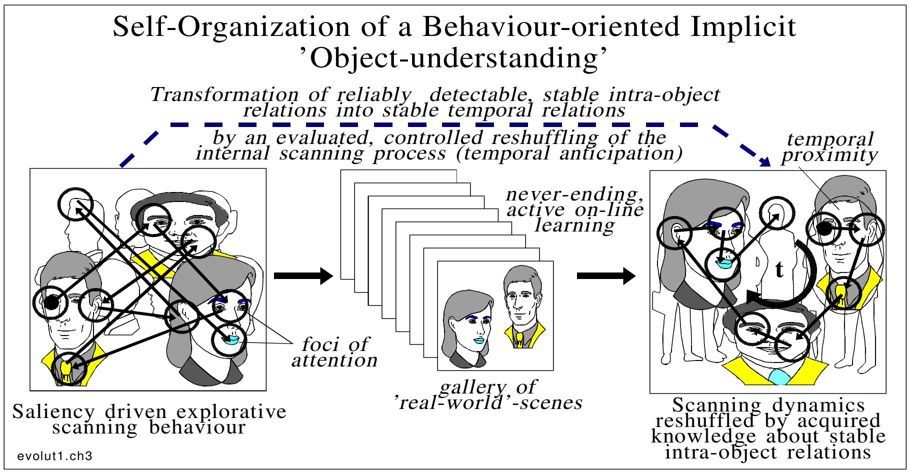


Abbildung 1.2: *Selbstorganisation eines verhaltensorientierten, impliziten Objektverständnisses*

Im nachfolgenden Kapitel 2 werden zunächst die neurobiologischen, psychophysischen und cognitiven Grundlagen der entwickelten Modellarchitektur dargestellt.
Daran schließt sich in Kapitel 3 die Ableitung der konkreten Modellarchitektur an. Weiterhin erfolgt hier eine formale Verhaltensdefinition, die die angestrebten Systemleistungen auf abstraktem Niveau festlegt. Das Kapitel 4 beschäftigt sich mit der robusten und reproduzierbaren Extraktion bildbeschreibender Merkmale als elementare Voraussetzung für die Selbstorganisation eines verhaltensorientierten Objektverständnisses. Im Kapitel 5 werden dann die Teile der Gesamtarchitektur behandelt, die in Form von zwei hierarchisch angeordneten attentiven Ebenen den eigentlichen Selbstorganisationsprozeß realisieren.

Kapitel 2

Neurobiologische, psychophysische und cognitive Grundlagen der Architektur

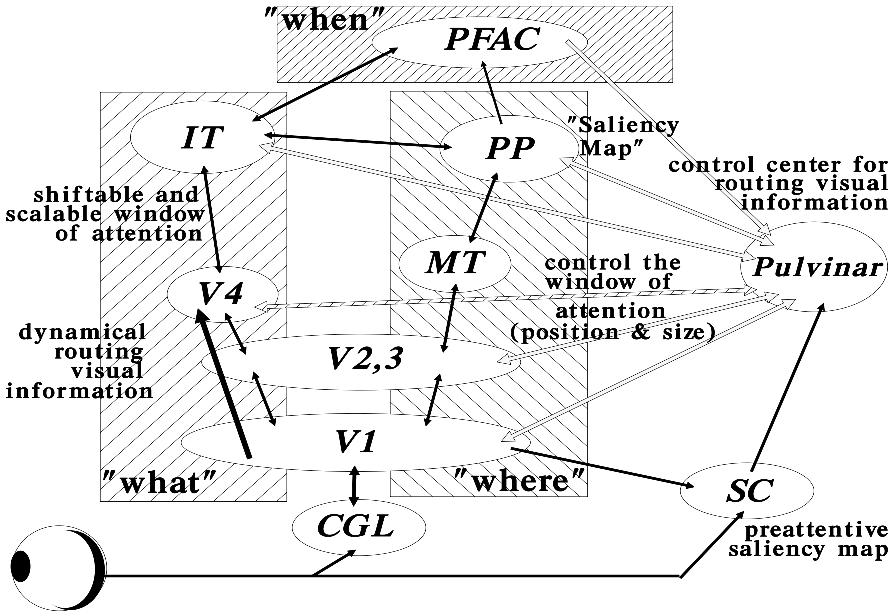


Abbildung 2.1: *Die für das Projektvorhaben relevanten biologischen Verarbeitungsstrukturen*

Die Abbildung 2.1 gibt einen Überblick über die biologischen Verarbeitungsstrukturen, die im Rahmen dieses Projektvorhabens berücksichtigt wurden. An dieser Stelle soll nochmals darauf hingewiesen werden, daß die Entwicklung einer umfassenden Modellarchitektur, die sich möglichst genau am biologischen Vorbild – dem visuellen System als Ganzes – orientiert, nicht das eigentliche Ziel dieses Projektes war und sein konnte. Vielmehr mußte zur Realisierung des Vorhabens die Problemkomplexität in ausreichendem Maße eingeschränkt werden, was bedeutete, daß nur die Aspekte der visuellen Verarbeitung Berücksichtigung finden konnten, die im Zusammenhang mit Aufmerksamkeitsprozessen gesehen werden können.

Relevant für das Projektvorhaben waren:

1. Hinweise aus der Neurobiologie und der Psychophysik zur selektiven visuellen Aufmerksamkeit.

2. Die Einbeziehung höherer Cortexareale zur Repräsentation elementarer visueller Basismerkmale und deren räumlicher Beziehung sowie zur Organisation eines emergenten Objektverständnisses.

3. Die biologischen Korrespondenzen zur frühen visuellen Verarbeitung von der Retina bis zum primären visuellen Cortex, insbesondere im Hinblick auf die robuste Extraktion von Basismerkmalen.

2.1 Neurobiologie der visuellen Aufmerksamkeit

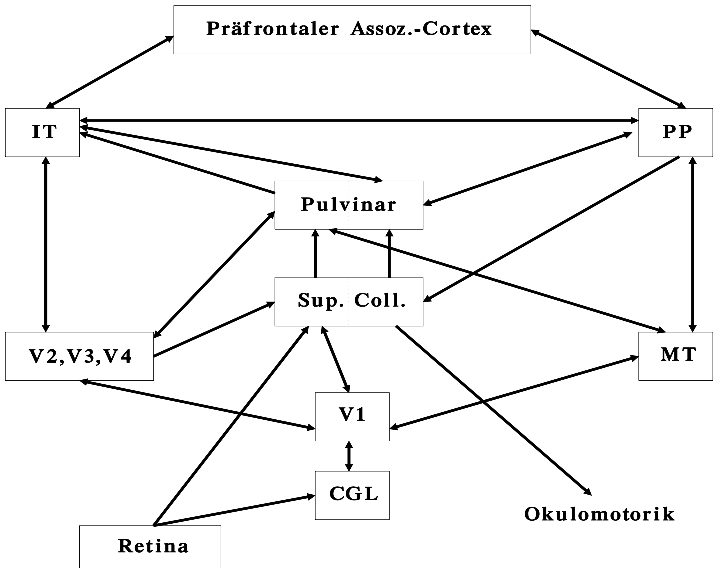


Abbildung 2.2: *Diese Abbildung gibt einen Überblick über die neuroanatomischen Strukturen, die wesentlich an dem Prozeß der internen Aufmerksamkeit beteiligt sind. Dabei wurde eine funktionale Aufteilung der Areale vorgenommen. So faßt z.B. der Block V2,V3,V4 die frühe Verarbeitung des visuellen Cortex zusammen.*

Die Abb. 2.2 gibt einen Überblick über die wesentlichen corticalen und subcorticalen Areale, von denen man annimmt, daß sie an dem Mechanismus der selektiven Aufmerksamkeit mitwirken. Diese Abbildung gibt gleichzeitig die Einschränkung bezüglich der betrachteten Areale, die für das Projektvorhaben getroffen wurde.
In den folgenden Abschnitten werden die Ergebnisse physiologischer Experimente und deren funktionelle Interpretation bezüglich der selektiven Aufmerksamkeit erläutert. Desweiteren werden die Bezüge zu den psychophysischen Modellen aufgezeigt und berechnungstheoretische Ansätze zur Beschreibung der physiologischen Vorgänge benannt. Diese Ansätze bilden u.a. die Grundlagen für die neuronale Modellarchitektur zur Steuerung der visuellen Aufmerksamkeit. Zusätzlich werden Ergebnisse vorgestellt, die im direkten Bezug zur zeitlichen Objektrepräsentation stehen. Einen guten Überblick über die Mechanismen findet man in [Colby 1991] und [Desimone et al. 1989].
Die Systeme der frühen visuellen Verarbeitung (Retina, CGL und visueller Cortex) bilden die Grundlage für die Identifikation von Objekten sowie zur Erzeugung einer Auffälligkeitskarte, die in der Modellarchitektur den rein datenbasierten Ausgangspunkt für die selektive Aufmerksamkeit darstellt. Diese Systeme wurden im vorherge-

henden Abschnitt bereits vorgestellt und sollen hier nicht noch einmal separat behandelt werden.

2.1.1 Frühe corticale Verarbeitung visueller Information

2.1.2 Zwei visuelle Bahnen

Wie in Abbildung 2.1 zu sehen ist, kann die V1 als Ausgangspunkt zweier funktionell unterschiedlicher Bahnen betrachtet werden [Mishkin et al. 1983], die vermutlich eng mit dem Parvo- und dem Magnozellularsystem korrespondieren [Essen et al. 1992]. Diese Einteilung beruht auf Untersuchungen der Wirkungen von Läsionen im temporalen und parietalen Lappen. Die erste Bahn läuft überwiegend ventral in den temporalen Lappen und spielt im wesentlichen eine Rolle bei der Objekterkennung. Aus diesem Grund werden die beteiligten corticalen Areale mit der Bezeichnung "What"-System oder ventrale Bahn zusammengefaßt.
Die zweite Bahn verläuft im wesentlichen dorsal in den parietalen Lappen und soll für die Verarbeitung von räumlicher und visuomotorischer Information verantwortlich sein. Somit wird dieses System als "Where"-System oder dorsale Bahn bezeichnet. Die einzelnen Areale der beiden Bahnen sind in der Regel rekurrent miteinander verbunden. Ebenso gibt es ausgeprägte Verbindungen zwischen den Arealen der einzelnen Bahnen. Trotzdem lassen sich den Arealen der einzelnen Bahnen überwiegende Funktionalitäten bezüglich "What" und "Where" zuordnen [Desimone et al. 1989].

What-System

Die ventrale Bahn (siehe Abbildungen 2.1 und 2.8) läuft von der V2 über V3 und V4 in den inferotemporalen Cortex (IT). Dabei nimmt die Größe der rezeptiven Felder von Areal zu Areal zu. Im IT überdecken die Felder das visuelle Feld vollständig. Diese Tatsache läßt sich mit "Stack"-Modellen (siehe auch Abschnitt 2.3.2) beschreiben (z.B. [Olshausen et al. 1993]), die eine Variation der Größe der rezeptiven Felder realisieren. Diese Modelle zeigen, daß die Variation der Größe der rezeptiven Felder möglicherweise eine Reduktion der visuellen Daten erreicht. Im Abschnitt 4.8.2 wird dieser Gesichtspunkt wieder aufgegriffen.
Das früheste Areal, in dem Effekte zur Aufmerksamkeit elektrophysiologisch nachgewiesen wurden, ist die V4: Im Rahmen einer Diskriminierungsaufgabe (z.B. unterscheide Rot von Orange) werden die rezeptiven Felder der korrespondierenden Zelle schmaler bezüglich der zu unterscheidenden Qualität und die Zelle antwortet stärker als im Normalfall. D.h. die Zellen zeigen die Auswirkungen, daß das Versuchstier seine Aufmerksamkeit erhöhen muß, um den Stimulus unterscheiden zu können [Desimone et al. 1989]. Sowohl im IT als auch in der V4 wurden Zellen gefunden, die abhängig vom Ort der Aufmerksamkeit reagieren [Moran et al. 1985]. Die Abbildung 2.3 illustriert das Experiment.
Zwei unterschiedliche Stimuli werden im rezeptiven Feld eines Neurons präsentiert. Auf den einen Stimulus reagiert die Zelle kaum (ineffektiver Stimulus), auf den anderen Stimulus antwortet die Zelle stark (effektiver Stimulus). Wenn die Aufmerksamkeit des Versuchstiers von dem effektiven Stimulus auf den ineffektiven Stimulus gelenkt wird,

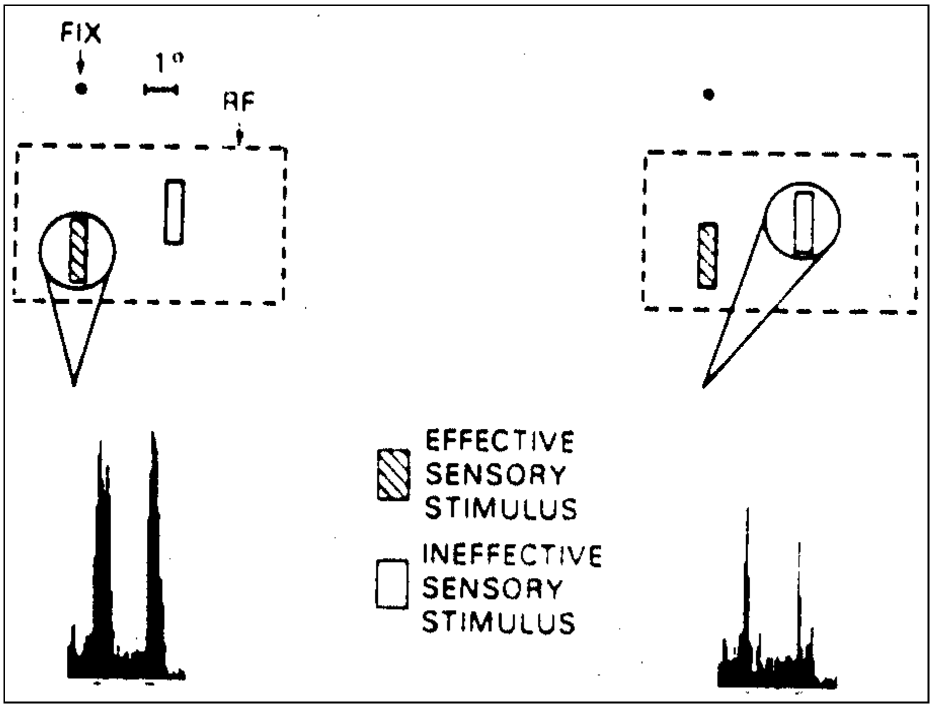


Abbildung 2.3: *Die Abbildung illustriert ein elektrophysiologisches Experiment zur selektiven Aufmerksamkeit im IT des Affen. Das Experiment zeigte, daß für die Modifikation der rezeptiven Felder im IT Aufmerksamkeitsprozesse eine wesentliche Rolle spielen (nach [Moran et al. 1985]).*

so reduziert sich die Antwort der Zelle. In der V4 wurde zusätzlich festgestellt, daß bei unterschiedlichem rezeptiven Feld je Stimulus die Antwort der Zelle unbeeinflußt vom Ort der Aufmerksamkeit ist. In der V4 bewirkt die Aufmerksamkeit eine Aufhebung einer zweideutigen Situation [Desimone et al. 1990]. Während der Experimente durfte das Versuchstier keine Augenbewegungen ausführen. Damit lassen sich die Ergebnisse mit der „searchlight“-Metapher [Treisman et al. 1980] (siehe Abschnitt 2.2) in Verbindung bringen.
Die Ergebnisse lassen weiterhin den Schluß zu, daß im IT eine Art „Suchscheinwerfer“ über das visuelle Feld verschoben wird. Damit lassen die Experimente den Schluß zu, daß der IT das erste Areal darstellt, in dem attentive Prozesse ablaufen.
In „delayed match-to-sample“-Aufgaben wird den Versuchstieren eine Sequenz von unterschiedlichen Stimuli präsentiert. Sie müssen dabei detektieren, ob einer der Stimuli mit dem ersten Stimulus übereinstimmt. Es gibt Zellen im IT, die aktiv bleiben, obwohl der erste Stimulus nicht mehr präsentiert wird. Wird dann ein zum ersten Stimulus passender präsentiert, so gibt es zwei unterschiedliche Reaktionen: Entweder erhöht sich die Aktivität der Zelle oder sie reduziert sich [Miller et al. 1994]. Es scheint, als ob die Zellen aktiv bleiben, um die Information über den zu suchenden Stimulus zu kodieren.

Where-System

Ausgehend vom striaten Cortex verläuft der dorsale Pfad (siehe Abbildungen 2.1 und 2.8) über den medialen temporalen Cortex (MT), den MST, PO, VIP zum postparietalen Cortex (PP) (siehe Abbildung 2.1). Die physiologischen Experimente zum postparietalen Cortex weisen dieses Areal als Schnittstelle zwischen der Verarbeitung sensorischer Information und der Ausführung von motorischen Bewegungen unter sensorischer Führung aus. So zeigen Personen mit bilateralen Läsionen am PP Defizite bei der willentlichen Ausführung von Augenbewegungen [Andersen et al. 1989]. Ein Teil der Zellen des PP besitzt visuelle rezeptive Felder, die vermutlich nicht retinotop organisiert sind. Vielmehr scheint die Position der Augen mit verrechnet zu werden, so daß sich eine kopfzentrierte Abbildung von visuellen Stimuli ergibt. Weitere physiologische Messungen zeigen „enhancement“-Effekte vor der Ausführung einer Sakkade. Dabei muß das Versuchstier eine Sakkade von einem zentralen Stimulus zu einem peripheren Stimulus ausführen. Etwa 55 ms vor der Sakkade erhöht sich die Aktivität der Zelle. Dieser Effekt scheint spezifisch bezüglich des Zielorts zu sein [Duhamel et al. 1992].
Neben den Effekten, die die externe Aufmerksamkeit betreffen, zeigen sich Aktivitäten im PP, die zur Steuerung der internen Aufmerksamkeit dienen können. Dabei spielt die „Posner Task“ eine wesentliche Rolle [Posner et al. 1980]. Die Versuchspersonen müssen eine Augenbewegung zu einem plötzlich erscheinenden, peripheren Stimulus ausführen. Die Reaktionszeit der Versuchspersonen sinkt, wenn die Position des peripheren Stimulus kurz vor der Präsentation durch einen kurz erscheinenden Lichtpunkt vorbereitet wird („precuing“). Die Reaktionszeit steigt, wenn die falsche Position vorbereitet wird. Dies bedeutet, daß die interne Aufmerksamkeit durch den vorbereitenden Stimulus an die entsprechende Position gelenkt wird, womit dann die externe Aufmerksamkeit die richtige Position erreicht. Läsionen im PP beeinträchtigen die Reaktionszeiten der Versuchstiere [Posner et al. 1984]. Ebenso zeigen PET-Messungen eine

erhöhte Aktivität im PP bei der Ausführung des "Posner-Task" [Corbetta et al. 1991]. Der oben erläuterte „enhancement“-Effekt im PP tritt auch ohne Augenbewegungen auf. Dabei muß das Versuchstier seine interne Aufmerksamkeit auf einen peripheren Stimulus richten.
Damit läßt sich vermuten, daß der PP das Ziel der internen Aufmerksamkeit angibt [Posner et al. 1988]. Dies geschieht in Analogie zu der Iniziierung der Sakkaden, nur fehlt das entsprechende „go“-Signal [Desimone et al. 1990]. Damit bildet der PP neben dem IT ein weiteres Areal, in dem eine attentive Verarbeitung erfolgt.
Zusätzlich wird diskutiert, ob der PP auch der Ort für eine Auffälligkeitskarte ist. Dies ist aber aufgrund der langen physiologischen Verzögerungszeiten fraglich [Olshausen 1994]. Für die Lokalisation einer Auffälligkeitskarte kommen ebenso subcortikale Strukturen in Frage, worauf im nachfolgenden Abschnitt eingegangen wird.

2.1.3 Subcorticale Strukturen

Es gibt im wesentlichen zwei subcorticale Strukturen, die an internen und externen attentiven Prozessen beteiligt sind: Pulvinar und Superior Colliculus (SC) [Robinson et al. 1989].

Superior Colliculus

Der Superior Colliculus (SC) läßt sich grob in eine obere und eine untere Gruppe von Schichten einteilen. Die unteren Schichten erhalten Projektionen vom frontalen Augenfeld (FEF) und dem postparietalen Cortex (PP). Die Retina und der visuelle Cortex projizieren sowohl in die unteren Schichten als auch in die oberen Schichten. Die unteren Schichten des SC iniziieren Sakkaden. Dabei zeigt sich eine topologische Kodierung von Richtung und Länge [Tweed et al. 1990]. Die visuellen rezeptiven Felder der Zellen der oberen Schichten sind retinotop organisiert. In diesen Schichten zeigen sich ähnliche „enhancement“-Effekte wie im PP, bevor das Versuchstier eine Sakkade zu einem peripheren Stimulus ausführt. Hält das Versuchstier beim Auftreten des peripheren Stimulus weiter die Fixation, so zeigt die korrespondierende Zelle ein kurzes Aufblitzen der Aktivität. Im Rahmen der „Posner Task“ zeigen die Zellen keine Reaktion auf den „precuing“-Stimulus. Diese Ergebnisse weisen daraufhin, daß der SC nicht in Prozesse der internen Aufmerksamkeit eingebunden ist. Vielmehr können die Ergebnisse als sensomotorische Steuerung einer Antwort auf einen visuellen Stimulus interpretiert werden.
Dagegen zeigen Verhaltensstudien zu pharmakologischen Läsionen am SC Defizite bei Augenbewegungen und der „Posner Task“. So wurde von [Posner et al. 1990] vorgeschlagen, daß der SC für die Verschiebung der internen Aufmerksamkeit zuständig ist. Allerdings stellen pharmakologische Läsionen eine sehr unspezifische experimentelle Methode dar, so daß davon ausgegangen wird, daß der SC primär für Augenbewegungen zuständig ist.

Pulvinar

Auf der Basis von cytoarchitektonischen Kriterien läßt sich die Pulvinar in fünf Kerne unterteilen: inferior (PI), lateral (PL), medial (PM), oral (PO) und medial dorsal (Pdm). PI und PL haben rekurrente Verbindungen mit den Arealen des ventralen Pfades und des MT, sowie Projektionen vom SC. Die Strukturen Pdm, PM und PO haben im wesentlichen rekurrente Verbindungen mit PP und IT, sowie Eingänge von den unteren Schichten des SC [Robinson et al. 1989]. Die PI/PL-Areale haben visuelle rezeptive Felder und zeigen „enhancement"-Effekte wie der SC, die stark mit Augenbewegungen verknüpft sind. Die Zellen des PM/PL/Pdm-Komplexes weisen ebenfalls „enhancement"-Effekte auf, die unabhängig davon sind, ob der Zielstimulus mit der internen Aufmerksamkeit oder mit Augenbewegungen erfaßt wird.
Studien zur Wirkung von Läsionen der Pulvinar ergeben eine höhere Reaktionszeit der Versuchstiere beim "Posner Task". Ebenso tritt eine erhöhte Fehlerrate auf, wenn bei der Diskriminierung von Farben ein zusätzlicher Distraktor im gleichen rezeptiven Feld präsentiert wird. Offensichtlich läßt sich die interne Aufmerksamkeit durch die Läsionen der Pulvinar nicht mehr so gut steuern. In PET-Studien zeigen sich erhöhte Aktivitäten in der Pulvinar, wenn die Versuchspersonen die Form eines Zielgegenstands innerhalb von Distraktoren unterscheiden müssen [LaBerge et al. 1990].
Zusammenfassend lassen sich die Ergebnisse der Experimente dahingehend interpretieren, daß die Pulvinar eine steuernde Wirkung ("gating") bei Prozessen der internen Aufmerksamkeit hat. Diese Hypothese läßt sich auch durch die Vielzahl von Verbindungen mit corticalen Arealen, insbesondere des visuellen Cortex und des postparietalen Cortex, unterstützen. Zusätzlich ist sie aufgrund ihrer Verbindungen ein möglicher Ort einer Auffälligkeitskarte und deren Auswertung [Olshausen 1994]. Gleichzeitig bildet die Pulvinar damit eine Schnittstelle zwischen attentiver und präattentiver Verarbeitungsebene.

2.1.4 Höhere Cortexareale

Für die angestrebte Verarbeitungsleistung des Modells – die Selbstorganisation eines verhaltensorientierten, emergenten Objektverständnisses – mußten auch neurobiologische und funktionelle Erkenntnisse über höhere Cortexareale für den Entwurf der Modellarchitektur berücksichtigt werden, da hierfür die frühe visuelle Informationsverarbeitung von der Retina bis zur V1 mit Sicherheit nicht ausreicht (siehe auch Abbildung 2.2).

Bereits im Abschnitt 2.1 wurden zwei relevante höhere Cortexareale benannt, die für die Identifikation von Objekten und die Steuerung der selektiven visuellen Aufmerksamkeit eine wesentliche Rolle spielen – der inferotemporale Cortex (IT) und der postparietale Cortex (PP). Interessant ist, daß diese beiden Areale Teile von sogenannten Assoziationscortizes darstellen. Der IT ordnet sich in den limbischen Assoziationscortex ein, der PP bildet einen Teil des parietalen Assoziationscortex [Kupfermann 1991].
Aktuelle physiologische Untersuchungen legen nahe, den IT als Struktur für ein elementares Alphabet komplexer visueller Muster anzusehen (siehe auch [Fujita et al. 1992] und [Rolls et al. 1995]), aus denen während der Perception Objekte aufgebaut werden

könnten. Besonders wichtig ist dabei die endliche Anzahl der im IT repräsentierten Icons, wobei exakte Zahlenangaben fehlen.

Der PP verarbeitet in erster Linie Ortsinformation. Die in Wechselwirkung mit dem IT realisierte Kopplung von Ort und Inhalt eines visuellen Stimulus existiert aus pragmatischen Gründen in der Modellarchitektur, ohne daß dafür eine biologische Korrespondenz auffindbar war.

Die Assoziationscortizes (präfrontal, limbisch und parietal-temporal-occipital) stellen nach aktuellen Erkenntnissen [Kupfermann 1991] die Gehirnstrukturen dar, die einerseits die Aktivitäten der primären sensorischen Areale integrativ verarbeiten und andererseits die Kopplung zwischen sensorischen und motorischen Cortizes realisieren. Sie bilden demnach das Substrat für die höchsten Gehirnfunktionen, für Wahrnehmung und Denken. Unterstützt wird diese These durch die Tatsache, daß die relative Ausdehnung der Assoziationscortizes im Laufe der Stammesgeschichte kontinuierlich zunahm und beim Menschen am größten ist.
Der präfrontale Assoziationscortex weist eine sehr intensive Verschaltung mit dem PP auf. Man vermutet, daß diese beiden Regionen die am meisten miteinander kommunizierenden Cortexareale darstellen [Kupfermann 1991]. Dem präfrontalen Assoziationscortex kommt eine entscheidende Rolle für kognitives Verhalten, insbesondere für die Handlungsplanung, zu. Kupfermann [Kupfermann 1991] schreibt:„*One important function of the frontal lobes is thought to be related to the capacity of the organism to weigh the consequences of future actions and to plan accordingly. The frontal lobes integrate the interoceptive and exteroceptive information that they receive so as to select the appropriate motor response from the many available.*“
Vor diesem Hintergrund kann vermutet werden, daß dem präfrontalen Assoziationscortex neben dem Where- und What-System die Rolle eines When-Systems zukommt. Neben den Frontal Eye Fields, die aufgrund ihrer ausschließlichen Zuständigkeit für die Sakkadengenerierung in der Modellarchitektur nicht berücksichtigt wurden, enthält auch der präfrontale Assoziationscortex vermutlich eine komplette Karte des jeweils contralateralen visuellen Feldes. Lang andauernde Aktivierungen von Neuronen im präfrontalen Assoziationscortex (auch nach dem Verschwinden des entsprechenden Stimulus) könnten die Basis für Kurzzeitspeicherung (Pufferung) von visueller Information bilden.
Hervorhebenswert sind die starken Projektionen vom präfrontalen Assoziationscortex in das Frontal Eye Field [Goldberg et al. 1989]. Da dem präfrontalen Assoziationscortex in erster Linie Verarbeitungsleistungen hinsichtlich Verhaltensorganisation und Handlungsplanung zugeordnet werden, könnte man vermuten, daß über diese Projektionen das Abtastverhalten (FEF's verantwortlich für die Generierung von Sakkaden) bei der Analyse des visuellen Inputs maßgeblich beeinflußt wird.

2.2 Psychophysik der selektiven Aufmerksamkeit

Zunächst wird die experimentelle Vorgehensweise bei der Analyse der selektiven Aufmerksamkeit vorgestellt. Anschließend wird ein grober Überblick über die "klassische" Modellvorstellung gegeben, um dann in den folgenden Abschnitten Details der Modellvorstellung zu erläutern. Abschließend werden alternative Ansätze zu diesem Modell diskutiert.

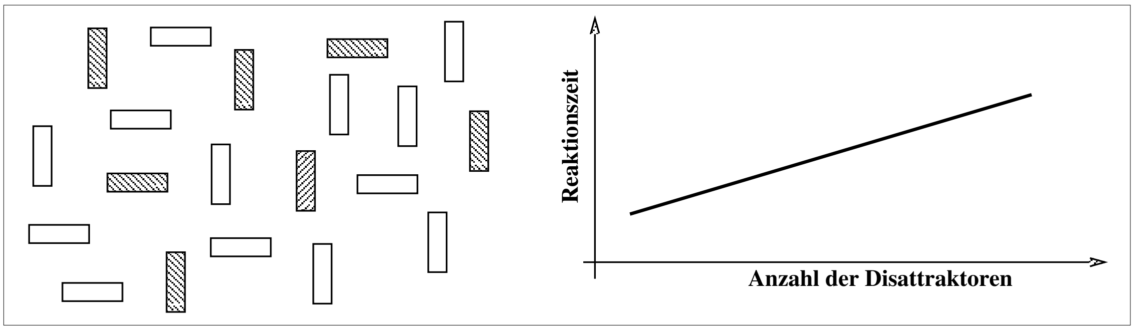


Abbildung 2.4: *Beispiel einer typischen visuellen Szene (links), wie sie in Experimenten zur Untersuchungen der internen Aufmerksamkeit verwendet wird. Rechts ist beispielhaft ein Ergebnis eines solchen Experiments dargestellt. Die Zeit, bis der Zielgegenstand (grauer horizontaler Balken) gefunden wird (Reaktionszeit), nimmt mit der Anzahl der Distraktoren zu.*

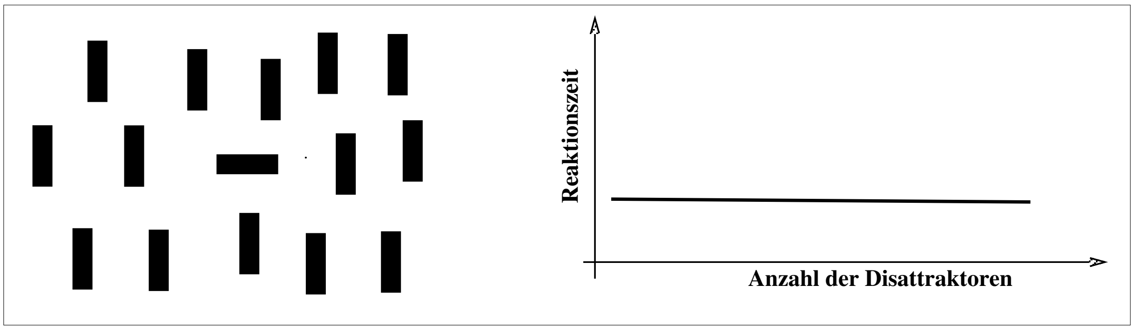


Abbildung 2.5: *Beispiel einer typischen visuellen Szene (links), bei der ein "pop-out"-Effekt auftritt. Die Reaktionszeit (rechts) ist unabhängig von der Anzahl der Distraktoren.*

2.2.1 Experimentelle Vorgehensweise

Zur Untersuchung der selektiven Aufmerksamkeit werden im wesentlichen visuelle Suchaufgaben gestellt. D.h. es werden Versuchspersonen Szenen gezeigt, in denen sie einen bestimmten Zielgegenstand (Attraktor) unter Distraktoren möglichst schnell finden müssen. In einem solchen Experiment kann u.a. die Zeit gemessen werden, bis der Attraktor gefunden wurde (Reaktionszeit) oder es wird die korrekte Lokalisation des Attraktors ermittelt. Diese Resultate werden in Abhängigkeit von Art und Anzahl der Gegenstände ermittelt. Die Abbildungen 2.4 - 2.6 zeigen typische Versuchsszenen. Zur Identifikation des Ziels lassen manche experimentelle Prozeduren Augenbewegungen zu. Andere unterdrücken sie durch unterschiedliche Verfahren. Allerdings zeigt es sich, daß die qualitativen Aussagen der Experimente kaum unterschiedlich sind (vgl. [Bergen et al. 1983] [Nothdurft 1993b]).

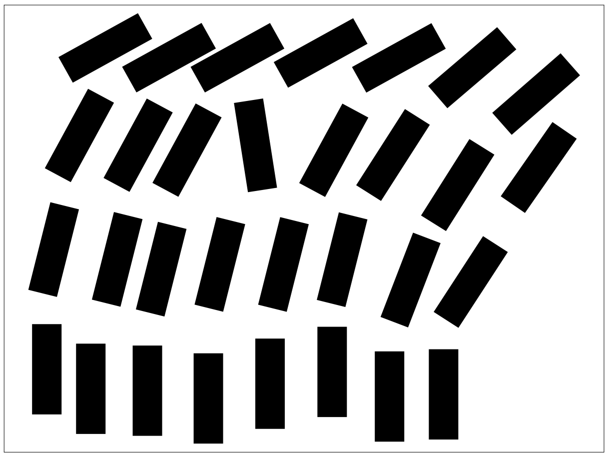


Abbildung 2.6: *Beispiel einer "texturierten" visuellen Szene, bei der aufgrund der Wechselwirkung zwischen den verschiedenen Merkmalskarten ein „pop-out"-Effekt auftritt.*

2.2.2 Zwei Ebenen der Verarbeitung

Die Ergebnisse der Suchaufgabe bezüglich der Reaktionszeit lassen sich folgendermaßen zusammenfassen: In manchen Szenen (Abb. 2.5) zeigt sich ein „pop-out"-Effekt, d.h. der Zielgegenstand „springt" der Versuchsperson sofort ins Auge. Dieser Effekt ist weitgehend unabhängig von der Anzahl der Distraktoren. Dabei liegen die Reaktionszeiten bei 30-50ms. In anderen Szenen steigt die Reaktionszeit in Abhängigkeit von der Anzahl der Distraktoren (Abb. 2.4).
Ausgehend von diesen experimentellen Tatsachen wird im allgemeinen davon ausgegangen, daß zwei Ebenen der Verarbeitung visueller Information existieren [Neiser 1967]. Die erste Verarbeitungsstufe wird als präattentive Verarbeitungsstufe bezeichnet. Sie läßt sich durch parallele, schnelle, lokale und unvollständige Verarbeitung charakterisieren. Sie lenkt die Aufmerksamkeit auf den entsprechenden Punkt in der visuellen Szene und bewirkt so den „pop-out"-Effekt. Die zweite Verarbeitungsebene wird als attentive Ebene bezeichnet, da die Identifikation des Zielgegenstands nur unter Aufbringung von Aufmerksamkeit erreicht wird. Sie wird aufgrund der längeren Reaktionszeiten als u.a. seriell charakterisiert. Dieser Modellansatz wird auch als parallel-serielle Dichotomie

bezeichnet.

Präattentive Verarbeitungsebene

Die Modelle zur präattentiven Verarbeitungsstufe (siehe [Wolfe et al. 1990] und [Treisman et al. 1980]) gehen davon aus, daß zunächst eine Vielzahl von parallel arbeitenden Prozessen Merkmale aus einer visuellen Szene extrahieren. Die Ergebnisse dieser Prozesse werden auch in Anlehnung an physiologische Ergebnisse als Merkmalskarten bezeichnet. Diese Merkmalskarten werden dann zu einer Auffälligkeitskarte verknüpft [Koch et al. 1985]. Diese Karte bewertet die Positionen einer Szene bezüglich ihrer Auffälligkeit im Rahmen der Gesamtszene. Die Auswertung der Karte bewirkt, daß die Aufmerksamkeit zunächst an die Stelle mit der größten Auffälligkeit gelenkt wird. Für diese Auswertung wurde von [Koch et al. 1985] ein WTA-Mechanismus vorgeschlagen. Dieser Vorgang wirkt reflexartig, d.h. selbst wenn den Versuchspersonen ein anderer Zielgegenstand genannt wird, identifizieren sie zuerst den auffälligsten Gegenstand.
Welche Prozesse bzw. Merkmale die Auffälligkeitskarte erzeugen, ist zur Zeit noch unklar. Experimente von [Wolfe et al. 1990] [Treisman 1990] zeigen, daß Abweichungen von einer szenenglobalen „Nulline“ eines Merkmals als auffällig bewertet werden. Dies gilt z.B. für einen grünen Gegenstand (Attraktor) unter roten Gegenständen (Distraktoren). Gleiches gilt auch für Orientierungen (Abb. 2.6). Hinzu kommen weitere Merkmale wie Geschlossenheit, Bewegung, Krümmung [Treisman 1990]. Für "texturierte" Szenen (Abb. 2.6) zeigt sich, daß lokale Merkmalsdifferenzen ebenfalls von Bedeutung sind [Bergen et al. 1983] [Nothdurft 1993b]. Dies gilt sowohl für Orientierungen als auch für lokale Farbdifferenzen und Helligkeitskontraste. Dabei scheint die Modellvorstellung von einer "additiven" Überlagerung der Merkmalskarten sinnvoll [Nothdurft 1993a]. Damit ist es jedoch kaum noch möglich, mit Hilfe von "pop-out"-Experimenten auf die Beiträge der einzelnen Merkmalskarten zu schließen, wie das in [Treisman et al. 1980] noch angenommen wurde, da offensichtlich eine Wechselwirkung der Merkmalskarten untereinander stattfindet.
Weiterhin zeigt sich, daß Gruppierungsprozesse präattentiv ablaufen [Julesz 1991]. So werden z.B. 3-dimensionale Illusionen präattentiv erzeugt [Enns et al. 1990] oder gleichfarbige Punkte, die in einem texturierten Hintergrund ein Dreieck bilden, als solche gruppiert [Nothdurft 1993b].
Neben der Reaktionszeit lassen sich die Experimente bezüglich der Lokalisation der Zielgegenstände auswerten. Diese Experimente ergeben, daß die Lokalisation präattentiv gefundener Gegenstände schlecht oder falsch ist [Bergen et al. 1983] [Treisman 1990]. D.h. Gegenstände, die durch die Auffälligkeit die Aufmerksamkeit anziehen, werden ohne Positionsinformation vom attentiven Prozeß verarbeitet. Ähnliches gilt für die Geometrie bzw. die Form von einheitlich texturierten und auffälligen Gebieten (siehe auch [Bergen et al. 1983]). Hier wird die Form nicht über präattentive Prozesse erfaßt.
Untersucht man die präattentive Identifikation von Gegenständen, z.B. eines orangefarbenen Dreiecks, so zeigen die Experimente, daß es zu illusionären Verknüpfungen kommt. Die Versuchsperson identifiziert ein rotes Dreieck, wenn ein roter Gegenstand in der Nähe ist. Dieser Effekt reduziert sich, wenn die attentive Verarbeitungsebene mit verwendet wird. Der Effekt relativiert sich ebenfalls, wenn typische Assoziationen

hinzugezogen werden können (z.B. wird das orangefarbene Dreieck der Versuchsperson als Mohrrübe benannt) [Tsal 1989] [Treisman 1990]. Weiterhin können Zahlen präattentiv identifiziert werden [Saarinen et al. 1991].
Diese Ergebnisse passen zu den Experimenten von [Schneider et al. 1977], die belegen, daß Objekte, die zunächst nur attentiv identifiziert werden können, durch Übung bzw. Lernvorgänge später auch präattentiv identifiziert werden.

Attentive Verarbeitungsebene

Die attentive Verarbeitungsebene wird von [Treisman et al. 1980] als serieller Prozeß aufgefaßt, der wie mit einem „Suchscheinwerfer" die aktuelle Szene nach dem Zielgegenstand absucht ("searchlight"-Metapher). Diese Form der Suche bewirkt, daß die Reaktionszeit von der Anzahl der Distraktoren (Szenenbestandteile, die vom eigentlichen Suchziel ablenken) abhängt. Allerdings variiert die Reaktionszeit von 30ms - 60ms pro Distraktor in Abhängigkeit vom Szeneninhalt [Julesz 1991]. Aufgrund dieser Variationen wird davon ausgegangen, daß keine rein serielle Suche stattfindet (siehe [Treisman et al. 1990] und [Wolfe et al. 1990]). Es wird vermutet, daß eine parallele Führung des Suchvorgangs erfolgt. Die parallele Führung hat zwei Quellen: die Auffälligkeitskarte ("bottom-up") und die Information über den Zielgegenstand ("top-down"). Die Information über den Zielgegenstand inhibiert in den korrespondierenden Merkmalskarten die entsprechenden Positionen der Distraktoren [Treisman et al. 1990] bzw. aktiviert die Merkmalskarten bezüglich des Zielgegenstandes [Wolfe et al. 1990]. Über diesen Einfluß der Merkmalskarten können Ziele schneller gefunden werden. Die Effektivität hängt dabei von den verwendeten Merkmalen bei Attraktor und Distraktor ab. Die Führung durch die Auffälligkeitskarte ist insbesondere dann sinnvoll, wenn nichts vorher über die Szene bekannt ist. Dann sollte die Aufmerksamkeit auf ungewöhnliche (auffällige) Positionen gelenkt werden, die sehr informativ sind, da sie sich von anderen Positionen in der Szene unterscheiden.

Obwohl diese bereits vor mehr als 20 Jahren postulierte Hypothese zur selektiven Aufmerksamkeit bis heute in der Literatur teilweise kontrovers diskutiert wird, bildete sie eine tragende Säule unseres Modellkonzepts.
Die Metapher vom „Suchscheinwerfer" wirft die Frage auf, wie sich dieser „Suchscheinwerfer" bewegt und welchen Durchmesser er hat? In [Remington et al. 1984] wird festgestellt, daß die Zeit, die der „Suchscheinwerfer" braucht, um sich von einer Position zur nächsten zu bewegen, unabhängig von der Entfernung zwischen den beiden Positionen ist. Dies deutet auf eine sprungartige Bewegung hin. Im Gegensatz dazu zeigen die Experimente von [Shulman et al. 1979], daß Gegenstände, die zwischen zwei Aufmerksamkeitspunkten liegen, ebenfalls wahrgenommen werden, was auf eine gleitende Bewegung hindeuten würde. Dies läßt vermuten, daß die Art der Bewegung abhängig von der Aufgabenstellung und dem Szeneninhalt ist. Die Experimente von [Saarinen 1993] ergeben, daß die Bewegungen über das ganze retinale Feld erfolgen.
Die Größe des „Suchscheinwerfers" variiert in Abhängigkeit von der Komplexität des Inhaltes [Eriksen et al. 1986]. So können z.B. Buchstaben unterschiedlicher Größe präattentiv identifiziert werden.

Weiterhin wurde experimentell nachgewiesen, daß es eine "inhibition of return" gibt [Klein 1988], d.h. Positionen, die bereits von dem „Suchscheinwerfer“ erfaßt worden sind, können nur unter Schwierigkeiten wieder erfaßt werden. Diese Inhibition kann bis zu 1,3 Sekunden anhalten.

2.2.3 Alternative Modelle

Die obigen Experimente haben bei einigen Arbeitsgruppen dazu geführt, andere Modellansätze vorzuschlagen, die nicht auf der parallel-seriellen Dichotomie aufbauen. Es werden im folgenden zwei Modellansätze kurz beschrieben, die jedoch nicht explizit in unserem Modellansatz berücksichtigt wurden.
Die Möglichkeit, Buchstaben unterschiedlicher Größe präattentiv zu identifizieren, bildet die Grundlage für das „zoom lens“-Modell [Eriksen et al. 1986]. Hier wird in Analogie zu einer Linse angenommen, daß die parallele Verarbeitung einem unscharfen Zustand der Linse entspricht und die serielle Verarbeitung einem scharfen Zustand der Linse.
In [Humphreys et al. 1993] wird ein Modell vorgestellt, daß im wesentlichen die präattentiven Gruppierungseffekte aufgreift. Dazu wird davon ausgegangen, daß die Szene zunächst aufgrund lokaler Ähnlichkeiten in einzelne Segmente zusammengefaßt wird, um dann die einzelnen Segmente in einem sequentiellen Prozeß zu durchsuchen. Ist die Anzahl der Segmente hoch, so dauert die Suche lange. Gibt es allerdings nur zwei Segmente, so tritt der „pop-out“-Effekt ein. Die Gruppierung erfolgt dabei ohne Wissen über den Zielgegenstand, d.h. die Segmente müssen während der Suche möglicherweise wieder aufgelöst werden, so daß die Suchzeit weiter steigt.

2.3 Frühe visuelle Verarbeitung

2.3.1 Signalverarbeitung in der Retina

Da ein wesentlicher Schwerpunkt des Projektvorhabens in der neurophysiologisch motivierbaren Analyse von Farbinformation bestand, soll auf die Verarbeitung von Farbinformation als Voraussetzung für die robuste Extraktion von Basismerkmalen innerhalb der Retina als elementare Basis der Farbwahrnehmung hier besonders eingegangen werden.

Zapfen

Derzeit ist allgemein anerkannt, daß die erste Stufe der visuellen Verarbeitung in der Lichtabsorption durch in den Stäbchen und Zapfen eingelagerte Photopigmentmoleküle besteht.
Da die Stäbchen bei tageslichtähnlichen Beleuchtungsverhältnissen, wie sie in diesem Projektvorhaben ausschließlich eine Rolle spielen, in die Sättigung gehen, wurde nur die Verarbeitung in den Zapfen berücksichtigt.

Bereits 1802 sagte Thomas Young die Existenz von drei Arten lichtdetektierender Elemente (Zapfen) voraus. Helmholtz (1924) entwickelte später diese Feststellung in eine umfassende Theorie, bekannt als Dreikomponenten- oder Young-Helmholtz-Theorie. Psychophysische Studien an normal und anormal farbsichtigen Personen führten zu einer Bestätigung dieser Theorie.
Da man annehmen kann, daß auch in den Zapfen, wie in den Stäbchen, das Licht durch Photopigmente absorbiert wird, folgt aus der Young-Helmholtz-Theorie, daß es in den Zapfen drei verschiedene Photopigmente mit unterschiedlichen Absorptionsspektren geben muß. Diese zusätzlich zum Rhodopsin existierenden Photopigmente werden Erythrolab (rotabsorbierend), Chlorolab (grünabsorbierend) und Cyanolab (blauabsorbierend) genannt.
Der Nachweis der einzelnen Photopigmente und deren spektraker Empfindlichkeitsverläufe erwies sich als relativ schwierig ([Brown et al. 1964], [Marks et al. 1964], [Lang 1978], [Smith et al. 1975]).
Letztlich führen jedoch alle Messungen auf die in der Abbildung 2.7 qualitativ dargestellten normierten oder nichtnormierten spektralen Kennlinien der einzelnen Photopigmente.
Der Gebrauch von Farbbezeichnungen (Rot, Grün, Blau) zur Kennzeichnung der unterschiedlichen Zapfentypen ist an dieser Stelle jedoch noch nicht gerechtfertigt, da erst Zellen in folgenden Verarbeitungsstufen Empfindlichkeitskurven besitzen, die empfindungsgemäße Farbverläufe repräsentieren. Naheliegender ist daher die Bezeichnung der Zapfen als L- (long-wavelength-sensitive), M- (middle-wavelength-sensitive) und S- (short-wavelength-sensitive) Zapfen.

In den folgenden Abschnitten soll kurz skizziert werden, wie die in den Photorezeptoren erzeugten Signale innerhalb der Retina durch verschiedene weitere Zelltypen verarbeitet und modifiziert werden, bevor sie zu höheren Gehirnarealen übertragen werden.

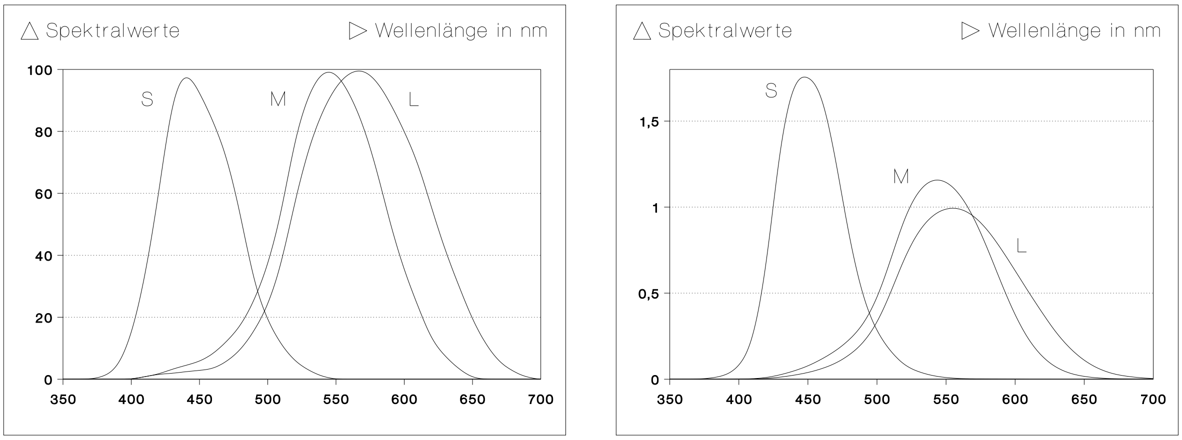


Abbildung 2.7: *Spektralwertkurven nach Smith und Pokorny (links) sowie nach Judd (rechts).*

Ganglienzellen

Die Ausgangsneuronen der Retina sind die Ganglienzellen, deren Axone den optischen Nerv bilden.
Schichtweise zwischen Photorezeptoren und Ganglienzellen sind die unterschiedlichen Typen von Interneuronen angeordnet: Bipolarzellen, Horizontalzellen und Amakrinzellen. Diese Zellen übertragen Signale von den Photorezeptoren zu den Ganglienzellen. Sie kombinieren Signale von verschiedenen Photorezeptoren, so daß die Aktivierung der Ganglienzellen direkt von der örtlichen Verteilung des visuellen Reizes auf der Retina und von seinem zeitlichen Verlauf beeinflußt wird.
Da Ganglienzellen die Ausgangsneuronen der Retina bilden, besteht der einfachste Weg der Bestimmung der Verarbeitungsleistungen der Retina in der Untersuchung, wie Ganglienzellen auf unterschiedliche Lichtreize reagieren.
Über Ganglienzellen ist verhältnismäßig viel bekannt. Es können zwei Typen von Ganglienzellen unterschieden werden, indem man ihre Antwort auf eine punktförmige Lichtreizung ihres rezeptiven Feldzentrums untersucht. On-Zentrum-Ganglienzellen feuern wenige Aktionspotentiale bei Dunkelheit. Auf ihr rezeptives Feldzentrum gerichtetes Licht erhöht die Feuerrate. (Sie sind praktisch bei on-geschaltetem Licht erregt.)
Dagegen werden Off-Zentrum-Ganglienzellen durch auf ihr rezeptives Feldzentrum gerichtetes Licht inhibiert. Während einer kurzen Periode nach dem Ausschalten des Lichtreizes ist ihre Feuerrate am größten. (Sie werden erregt, wenn das Licht offgeschaltet wird.) Bei beiden Zelltypen führt eine gleichzeitige Lichtreizung von Zentrum und Umfeld zu einer beinahe vollständigen Auslöschung der Zellaktivierung. Daher führt eine Reizung des gesamten rezeptiven Feldes mit diffusem Licht lediglich zu einer sehr schwachen Aktivierung.
On- und Off-Zentrum-Ganglienzellen sind in nahezu gleicher Anzahl vorhanden und bilden zwei parallele Pfade für die Verarbeitung visueller Information, da jeder Photorezeptor seine Aktivierung zu beiden Typen von Ganglienzellen weiterleitet.
Einige Ganglienzellen besitzen kein aus Zentrum und Umfeld bestehendes rezeptives Feld, sondern reagieren auf Änderungen der Gesamtbeleuchtung des Gesichtsfeldes und

sind somit für die Steuerung des Pupillenreflexes bedeutungsvoll.

Bipolar-, Horizontal- und Amakrinzellen

Horizontal- und Bipolarzellen liegen zusammen mit den Amakrinzellen in der inneren Schicht der Netzhaut. Jeder Typ dieser Interneuronen spielt eine bestimmte Rolle bei der Modifizierung des Photorezeptorsignals bei dessen Leitung durch die Retina, wobei das Wissen über Aktivierungsantworten dieser Interneuronen im Vergleich zu den Ganglienzellen bedeutend lückenhafter ist.
Zapfen im Zentrum eines rezeptiven Feldes einer Ganglienzelle besitzen eine direkte synaptische Verbindung zu Bipolarzellen, die wiederum direkt mit Ganglienzellen verbunden sind (direkte Bahn). Signale von Zapfen aus der rezeptiven Feldumgebung einer Ganglienzelle sind über Horizontal- und Amakrinzellen mit der Ganglienzelle verbunden (indirekte Bahn).
Die Bipolarzellen nehmen eine strategisch wichtige Position in der Retina ein, da alle Signale, die von den Rezeptoren kommen und zu den Ganglienzellen laufen, durch die Bipolarzellen fließen müssen. Diese Zellen sind an der direkten, wie auch an der indirekten Bahn beteiligt. Im Gegensatz dazu gehören die Horizontalzellen lediglich zur indirekten Bahn. Horizontalzellen sind viel seltener als Bipolarzellen, die in der inneren Schicht meist vorherrschen.
Wie bereits ausgeführt, spielte im Rahmen des Projektvorhabens ausschließlich das Tagessehen eine Rolle. Tagessehen beruht auf den Verbindungen von Zapfen zu Ganglienzellen. Die Analyse dieses Pfades soll nun im Mittelpunkt der folgenden Betrachtungen stehen.
Horizontalzellen übertragen die Information von entfernten Zapfen zu naheliegenden Bipolarzellen. Einige Typen von Amakrinzellen übertragen Informationen von entfernten Bipolarzellen zu Ganglienzellen. Der schichtweise Aufbau der Retina widerspiegelt diesen geordneten Informationsfluß.
Die Zellkörper der retinalen Neuronen sind in drei Kernschichten (so genannt, da sie die Zellkörper beinhalten) angeordnet:

- die äußere Kernschicht, die die Photorezeptoren beinhaltet,
- die innere Kernschicht, die Horizontal-, Bipolar- sowie Amakrinzellen beinhaltet und
- die Ganglienzellschicht.

Die Fortsätze dieser Zellen befinden sich in zwei Plexiform-Schichten, in denen die meisten synaptischen Kontakte stattfinden. Die äußere Plexiform-Schicht beinhaltet die Fortsätze von Rezeptoren, Bipolar- und Horizontalzellen, während die innere Plexiform-Schicht die Fortsätze von Bipolar-, Amakrin- und Ganglienzellen beinhaltet. Dieser Aufbau hebt die besondere Rolle der Bipolarzelle als Brücke zwischen den beiden Plexiform-Schichten hervor, da sie in beiden Fortsätze besitzt.
Horizontal- und Bipolarzellen reagieren auf Lichtreizung wie Photorezeptoren mit abgestuften Änderungen des Membranpotentials. Da diese kleinen Zellen kurze Fortsätze besitzen, werden Signale zu ihren synaptischen Kontakten ohne signifikante Verluste

übertragen. Passive Signalübertragung durch Zellen mit kurzen Fortsätzen erfolgt in vielen unterschiedlichen Teilen des Gehirns.
Im Gegensatz dazu projizieren die Axone von Ganglienzellen ihre Information zu verschiedenen Strukturen im Gehirn in Form von Aktionspotentialsequenzen. Viele Typen von Amakrinzellen feuern ebenfalls Aktionspotentiale.
Bipolarzellen besitzen ebenso wie Ganglienzellen komplexe rezeptive Feldeigenschaften ([Werblin et al. 1969], [Kaneko 1970]). Sie besitzen eine antagonistische rezeptive Feldstruktur und sind entweder On- oder Off-Zellen. Die Zapfen im rezeptiven Feldzentrum einer Bipolarzelle scheinen direkt mit der Bipolarzelle verbunden zu sein. On-Zentrum-Bipolarzellen depolarisieren, während Off-Zentrum-Bipolarzellen bei Stimulierung der Zapfen in ihrem rezeptiven Feldzentrum mit Licht hyperpolarisieren. Im Gegensatz dazu werden die Inputs von Zapfen in der rezeptiven Feldumgebung einer Bipolarzelle über Horizontalzellen umgeleitet. Diese können Inputs von entfernteren Quellen empfangen, da sie stark verzweigte, weiträumige Dendriten besitzen.
Wie bisher dargestellt, sind Haupteigenschaften der rezeptiven Felder der Ganglienzellen bereits bei den Bipolarzellen vorhanden. Dies wirft die Frage auf, welche Informationsverarbeitungsschritte zwischen Bipolar- und Ganglienzellen ablaufen. Betrachtet man die Komplexität der inneren Plexiform-Schicht, so erscheint es äußerst unwahrscheinlich, daß hier keine wesentliche Verarbeitungsleistung erbracht wird. An dieser Stelle sei auf die in Kapitel 4 behandelte Modellarchitektur zur Farbverarbeitung verwiesen, wo die Verarbeitung innerhalb der Retina noch einmal detailliert aufgegriffen wird.
Abgesehen von der Fovea, wo eine 1 : 1 Übertragung erfolgt, ist zwischen Bipolar- und Ganglienzellen in der direkten Bahn eine ausgeprägte Konvergenz vorhanden. Außerdem sind vielfach Amakrinzellen eingeschaltet, deren Funktion bisher nicht geklärt ist. Die Synapsen zwischen Bipolar- und Ganglienzellen sind wahrscheinlich alle exzitatorisch. Das bedeutet, daß On-Zentrum-Bipolarzellen zu On-Zentrum-Ganglienzellen führen und daß Off-Zentrum-Bipolarzellen zu Off-Zentrum-Ganglienzellen führen.

2.3.2 Postretinale Verarbeitung – Zwei parallele Systeme

Beginnend mit den Off- und On-Ganglienzellen der Retina lassen sich zwei Systeme (Bahnen) unterscheiden, die sowohl zytoarchitektonisch als auch funktionell voneinander abgrenzbar sind, das Magnozellularsystem und das Parvozellularsystem (siehe [Shapley et al. 1986] und Abbildung 2.8). Diese beiden Systeme setzen sich über den Corpus geniculatum laterale (CGL) und den primären visuellen Cortex bis in höhere Cortexareale fort. Die räumliche Trennung der beiden Systeme erfolgt ab dem CGL.
Ganglienzellen der fovealen Retina, die Inputs von unterschiedlichen Zapfentypen, im Zentrum jedoch von gerade einem einzigen Zapfen erhalten, projizieren in die parvozellulären Schichten des CGL und bilden damit den Ausgangspunkt des Parvozellularsystems.
Die Ganglienzellen der nichtfovealen Retina, die einen kombinierten Input von L- und M-Zapfen sowohl im Zentrum als auch in der antagonistischen Umgebung (siehe [Creutzfeldt et al. 1979] und [Perry et al. 1984]) erhalten, projizieren in die magnozellulären CGL-Schichten und bilden den Ausgangspunkt des Magnozellularsystems.

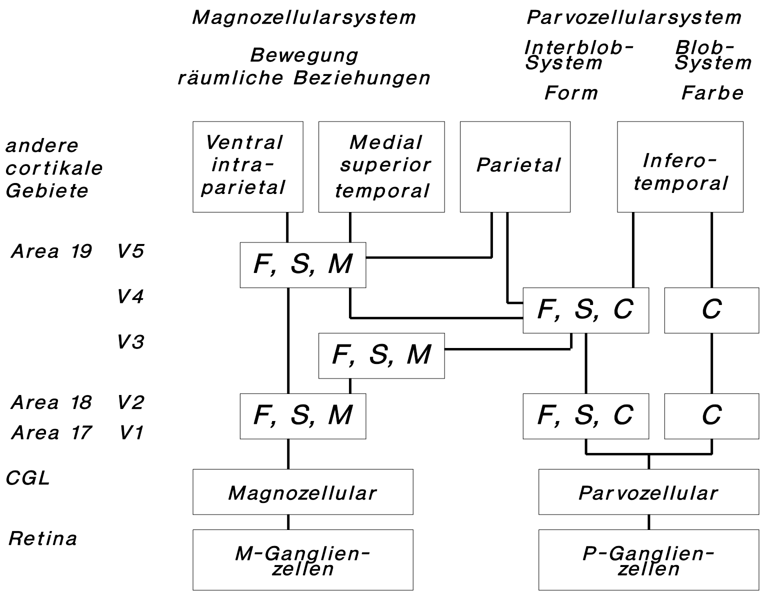


Abbildung 2.8: *Magno- und Parvozellularsystem beim Rhesusaffen (F:Form, M:Motion, S:Stereopsis, C:Colour) nach [Kandel & Schwartz 1991]*

Die retinale Ganglienzellpopulation besteht zu 80 % aus P- sowie zu 10 % aus M-Zellen. Beide Zelltypen leiten ihre Aktivierungen über den Sehnerv zum Corpus geniculatum laterale. Die restlichen Ganglienzellen projizieren in andere subcortikale Strukturen.

Corpus geniculatum laterale (CGL)

Beinahe alle der ungefähr 1,5 Millionen Zellen des Corpus geniculatum laterale (CGL) erhalten ihre Eingangsinformation von Sehnervfasern. Der CGL erhält zusätzlich reziproke Eingänge von der Großhirnrinde, auf die er selbst projiziert [Crick 1984], sowie von der Formatio reticularis im Hirnstamm [Burke et al. 1978], die an der Steuerung der Aufmerksamkeit mitwirkt.

Anatomisch besteht der CGL aus sechs übereinanderliegenden Zellschichten. Einzelne CGL-Zellen erhalten niemals Eingänge von Ganglienzellen beider Augen und sind damit entweder links- oder rechtsäugig. Diese zwei Zelltypen liegen jeweils getrennt voneinander in verschiedenen Schichten. Die gesamte Struktur ist topographisch einheitlich. Die beiden linken Retinahälften projizieren auf den linken CGL und die rechten Retinahälften entsprechend auf den rechten CGL.

Die beiden unteren, magnozellulären Schichten und die vier oberen, parvozellulären Schichten bilden jeweils eigenständige funktionelle Einheiten und stellen die Fortsetzung des Parvo-bzw. Magnozellularsystems dar [Zeki 1994].

Die Zellen des Corpus geniculatum laterale reagieren auf punktförmige Lichtreize ähnlich wie die Ganglienzellen der Retina. Sie besitzen ebenfalls rezeptive Felder mit On- und Off-Zentrum und sind damit nicht orientierungsselektiv [Levick et al. 1980].

Bei der Untersuchung von CGL-Zellantworten konnten 1966 erstmals von De Valois [De Valois et al. 1966] opponente + B − Y, + R − G, − B + Y, − R + G sowie achromatische Zellen festgestellt werden. Die Abkürzungen bedeuten Blau, Yellow, Rot und Grün. Es ist somit zu vermuten, daß es im Corpus geniculatum laterale zu einer Trennung von chromatischer und achromatischer Information kommt. Die visuelle Information scheint im CGL jedoch nicht tiefgreifend umgewandelt zu werden.
Die Untersuchungsergebnisse De Valois' [De Valois et al. 1958] [De Valois et al. 1966] zeigen interessante Korrespondenzen zu Modellvorstellungen des deutschen Physiologen Ewald Hering über das Farbensehen [Hering 1905], die dieser Ende des letzten Jahrhunderts entwickelte. Die Aussage Herings besteht darin, daß im Auge beim Sehen drei Paare von antagonistischen Prozessen beteiligt sind, die den drei Gegenfarbenpaaren entsprechen. De Valois' vorgenommene Ableitungen an CGL-Zellen schienen diese Theorie erstmals zu bestätigen.

Primäres Sehfeld V1

Vom Corpus geniculatum laterale gelangen die Nervensignale zuerst zum sogenannten visuellen Feld 1 (V1), dem primären Sehfeld des visuellen Cortex.
Die Signale des Magno-Systems werden in die obere Hälfte der mittleren Schicht des primären Sehfeldes (4Cα) geleitet, die des Parvo-Systems in die untere Hälfte (4Cβ).
Die Zellen der Schicht 4C projizieren in verschiedener Art und Weise auf die oberen Schichten: 4Cα sendet Signale zur Schicht 4B, 4Cβ zu den Schichten 2 und 3 (Abbildung 2.9).
Innerhalb der Schichten 2 und 3 treten leicht unregelmäßige, eiförmige Strukturen mit einem Durchmesser von ca. 0,2 mm auf. Diese sind halbregulär mosaikartig angeordnet. Livingston und Hubel bezeichneten sie wegen ihrer Form als Blobs und die helleren, sie umgebenden Flächen als Interblobs.
Magno- und Parvo-System, die bereits in den ersten Verarbeitungsstufen erkennbar sind, scheinen sich innerhalb der V1 in drei Subsysteme (Blob: Farbe, Interblob: Form und 4B: Bewegung) umzuorganisieren (Abbildung 2.9).
Die enorme Ausdehnung der V1 ist Ausdruck der ihr zukommenden Bedeutung als Region, die das Gesichtsfeld auf alle Aktivitäten, alle Submodalitäten des Sehens hin analysiert. Praktisch existiert eine Vielzahl von topographischen Merkmalskarten, die alle gleichzeitig unterschiedliche Aktivitäten innerhalb des Gesichtsfeldes feststellen. Diese funktionelle Spezialisierung des primären Sehfeldes setzt sich auch in folgenden visuellen Arealen, wie V3, V4 und V5 fort, auf die der primäre visuelle Cortex direkt projiziert.

Farbe Livingstone und Hubel äußerten die Vermutung, daß die Blobs einen Zweig der Sehbahn darstellen, der der Farbwahrnehmung gewidmet ist, wobei hier auch Schwarz-, Weiß- und Grautöne mit eingeschlossen sind [Livingstone et al. 1984a].
Dieses System scheint sich entweder bereits im CGL oder in den Schichten 2 und 3 des primären visuellen Cortex vom Rest der Sehbahn zu trennen.
Zwei Drittel der Blob-Zellen sind farbsensitiv. Sie reagieren hochselektiv auf Farbe oder Helligkeit, aber nicht auf Bewegung. Weiterhin fehlt ihnen, im Gegensatz zu Zel-

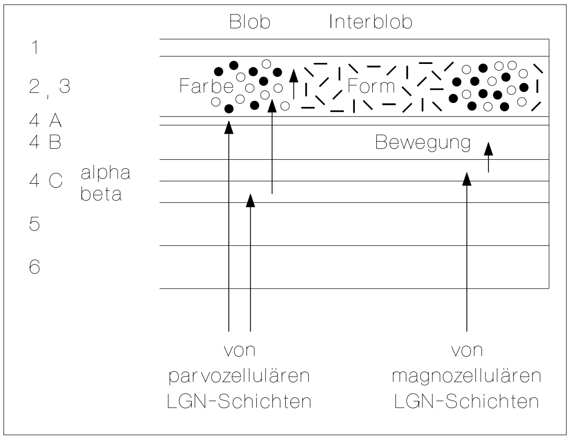


Abbildung 2.9: *Die Struktur des primären Sehfeldes V1 sowie die Einkopplung der parvo- und magnozellulären CGL-Aktivierungen.*

len außerhalb der Blobs, jegliche Orientierungsselektivität [Livingstone et al. 1984a] [Tso et al. 1988]. Vautin und Dow [Vautin et al. 1985] berichteten 1985 über die Existenz von vier grundlegenden Farbzelltypen im primären visuellen Cortex, deren spektrale Empfindlichkeitskurven in etwa den vier psychophysischen Farbempfindungen Rot, Yellow, Grün und Blau entsprechen. Auf weiße Lichtflecke beliebiger Größe und Form reagieren Blob-Zellen praktisch überhaupt nicht.

Die Größe der rezeptiven Felder der Zellen in der V1 variiert stark. Aus diesem Grund wurden auch hier von zahlreichen Autoren "Stack"-Modelle zur Beschreibung dieser Variation vorgeschlagen (z.B. [Burton et al. 1986] [Olshausen 1994], siehe auch Abschnitt 2.1.2).

Form Zellen der Interblob-Bereiche sind selektiv für die Orientierung eines Reizes, nicht aber für Farbe oder Bewegung.

Bewegung Zellen der Schicht 4B sind ebenfalls unspezifisch für Farbe, aber selektiv für Orientierung und Bewegung des Reizes.

Textur Psychophysische Untersuchungen der letzten Jahre lassen vermuten, daß das visuelle System eine quasilineare Kontrastempfindung, wahrscheinlich über Neuronen mit begrenzter Empfindlichkeit für bestimmte Abschnitte des räumlichen Frequenzspektrums, realisiert. Diese Studien [Sekuler 1974] [Robson 1975] [Braddick et al. 1978] [De Valois et al. 1980] deuten darauf hin, daß die Information im primären visuellen Cortex einer Filterung von Raumfrequenzen unterzogen wird. Die Existenz von frequenzselektiven Zellen deutet auf eine gleichzeitige Filterung lokaler Bildausschnitte auf mehreren Frequenzskalen hin.
Maffei und Fiorentini [Maffei et al. 1977] berichteten nach Einzelzellableitungen an Katzen über eine schichtbezogene Organisation frequenzselektiver Neuronen. Zellen der Schicht 4 des primären visuellen Cortex besitzen nach diesen Untersuchungen eine

Empfindlichkeit für höchste Raumfrequenzen, Zellen der Schichten 2 und 3 für mittlere und Zellen der Schichten 5 und 6 für niedrige Raumfrequenzen.
Im Gegensatz dazu fanden Tolhurst und Thompson [Tolhurst et al. 1981] innerhalb einer Cortexschicht verschiedene Zellen mit einer Selektivität für bestimmte Raumfrequenzen. Neuronen mit einer Empfindlichkeit für ein und dieselbe Raumfrequenz seien lediglich lose innerhalb von unregelmäßig verteilten Clustern organisiert. Diese Cluster seien weder eindeutig laminar noch columnar organisiert [Tolhurst et al. 1982].
Anatomische Studien mit einer 2-DG Autoradiographie deuten wiederum auf eine columnare Organisation von Zellen mit einer Empfindlichkeit für gleiche Raumfrequenzen hin. Die Präsentation eines Sinusgitters in allen Orientierungen mit lediglich einer Raumfrequenz führt im primären visuellen Cortex der Katze zu einer Aktivierung von columnaren Strukturen [Tootell et al. 1981] [Thompson et al. 1981]. Diese Columnen besitzen eine Periodizität von 1 mm und verlaufen durch alle Schichten. Columnen für hohe Raumfrequenzen waren lediglich im Zentralbereich des visuellen Feldes zu finden [Tootell et al. 1981].
Die Vielfalt der bekannten Untersuchungsergebnisse bezüglich einer möglichen Repräsentation von Raumfrequenzen innerhalb des primären visuellen Cortex lassen zur Zeit noch keine eindeutige Interpretation zu.
Allgemein scheint der primäre visuelle Cortex keine exakte Frequenzanalyse vorzunehmen. Für viele visuelle Anforderungen scheint die Feststellung der mittleren Raumfrequenz ohne Phase in einem lokalen Bildausschnitt zu genügen. Dieser Umstand würde eine erhebliche Vereinfachung der erforderlichen Analysestrukturen zulassen. Die Folge wären Doppeldeutigkeiten bei der Wahrnehmung von Raumfrequenzen, die objektiv auch auftreten. Für eine Objekterkennung oder eine Vorder- / Hintergrundseparierung ist die Kenntnis der mittleren Raumfrequenz oder Textur jedoch völlig ausreichend.
Frequenz-, aber nicht phasenspezifische komplexe Zellen erlauben es dem visuellen System, äußerst ökonomisch relevante, texturbezogene Informationen zu extrahieren. Die meiste Information über solche Texturen besteht in hohen Frequenzkomponenten, die durch komplexe Zellen des primären visuellen Cortex V1 detektiert werden können [De Valois et al. 1982] [Thorell et al. 1984].
Hingegen würde die Texturanalyse anhand von Kanten oder anderen Features in einem erheblich größeren Umfang verschiedenste Neuronen einbeziehen müssen.
Ein weiterer, wichtiger Grund für eine räumliche Frequenzfilterung besteht sicherlich in der Periodizität der visuellen Welt, die somit sehr effektiv kodiert werden kann. Da diese Periodizität lokal begrenzt ist, scheint auch nur eine lokale Filterung angebracht.

Lokal begrenzte Analyse Innerhalb des primären Sehfeldes V1 verlaufen die Hauptverbindungen generell auf- oder abwärts und verbinden die verschiedenen Schichten miteinander. Diagonale und seitwärts gerichtete Verbindungen erstrecken sich gewöhnlich 1 bis 2 mm. Der Informationsaustausch innerhalb der V1 bleibt systemintern: Blobs projizieren ausschließlich auf andere Blobs und Interblobs nur auf Interblobs [Livingstone et al. 1984b]. Die räumliche Einschränkung des seitlichen Informationsaustausches besitzt wichtige Konsequenzen.
Setzt man eine topographische Organisation des Inputs voraus, so muß diese auch für den Output des primären visuellen Cortex gelten. Die Analyse innerhalb der V1 kann

aufgrund des Fehlens langreichweitiger Verbindungen nur lokal sein.
Zwei im Cortex benachbarte Zellen besitzen stets eng beisammenliegende rezeptive Felder, die sich in der Regel sogar weitgehend, aber nie vollständig überlappen.

V4 – Farbe und Form mit Farbe

Die V4 stellt ein auf die Farbverarbeitung spezialisiertes Areal dar. Der Input der V4 kommt von der Retina über die parvozellulären CGL-Schichten zu den Blobs der V1 und erreicht von dort direkt oder über die dünnen Streifen der V2 die V4.
Zeki [Zeki 1980] vermutet, daß einzelne Neuronen innerhalb der V4 eher nach Farbkoordinaten als nach retinotopen Koordinaten angeordnet sind. Somit wären diese aufgrund des fehlenden örtlichen Bezuges zum Retinabild nicht in der Lage, das Formensehen zu analysieren, sondern würden eher für die Farbanalyse und -benennung eine Bedeutung besitzen. Zellen, deren Antwort mit der menschlichen Farbwahrnehmung korrelieren, lassen sich nur in der V4 [Zeki 1983] finden.
Diese Zellen reagieren stabil auf die spektrale Reflexion von einfarbigen Flächen vor buntem Hintergrund unabhängig von der spektralen Zusammensetzung des von einer solchen Fläche reflektierten Lichtes. Ihre Antwort wird jedoch von der farblichen Zusammensetzung des umgebenden bunten Hintergrundes beeinflußt, was auf eine Mitwirkung von Farbumstimmungsprozessen schließen läßt.
Interessant ist die Tatsache, daß die Betrachtung einer Szene, die als Farbbild eine starke Aktivierung der V4 hervorruft, lediglich ein Drittel dieser Aktivierung bei Betrachtung derselben Szene als Grauwertbild erzeugt. Das deutet sowohl auf eine Bestätigung Heringscher Vorstellungen als auch der Messungen De Valois hin, wonach jeder Bildpunkt der Fovea als RG-, BY- und WS-Aktivierung kodiert wird. Bei der Betrachtung eines Grauwertbildes sind nach einer Farbumstimmung die RG- und BY-Aktivierungen, als Träger der bunten Farbempfindungen, null und lediglich die WS-Aktivierungen ungleich von null.

Die neben der Farbverarbeitung gleichzeitig in der V4 realisierte Analyse von Form mit Farbe erhält ihren Input über die parvozellulären CGL-Schichten, die Interblobregionen der V1 sowie die blassen Streifen der V2.
Letztlich projiziert die V4 auf den inferotemporalen Cortex IT. Dieses System scheint für foveales Tagessehen verantwortlich zu sein, denn es besitzt seinen Ursprung im Midget-Zellsystem der Fovea des Primaten, das ein exzellentes Organ für detailliertes Form- und Farbensehen darstellt.

Farbumstimmung - Phänomene und mögliche Lokalisationen

Ein durchschnittlicher Beobachter sieht die Farbe eines hölzernen Schreibtisches immer in gleicher, hellbrauner Farbe, egal, ob er ihn im Morgenlicht, im stark rötlichen Licht des späten Nachmittags oder im gelben Licht der elektrischen Glühlampe betrachtet.
Tatsächlich reflektiert die Schreibtischplatte unter diesen Umständen jedesmal sehr verschiedene Wellenlängen. Auch wird die bezüglich eines Gegenstandes entstehende Farbempfindung nicht, wie man vermuten könnte, von dem durch dessen Oberfläche dominant reflektierten Licht bestimmter Wellenlänge festgelegt.

Die Farbwahrnehmung ermöglicht es, einen Gegenstand unabhängig vom eingestrahlten Wellenlängenspektrum im sichtbaren Bereich stets in gleicher Weise zu identifizieren. Das visuelle System ist bestrebt, sich durch Umstimmung an die sich ändernde spektrale Zusammensetzung der Allgemeinbeleuchtung anzupassen, sich praktisch auf ein mittleres Empfindungsniveau einzustellen. Damit sind die besten Voraussetzungen zum optimalen Erkennen von Farbunterschieden gegeben. Ohne funktionierende Farbumstimmung hätte das Farbensehen während der Evolution keine Bedeutung erlangen können.

Anatomisch werden mehrere Orte für die Realisierung der Farbumstimmung diskutiert. In der Retina, die traditionell als ein Ort für Mechanismen zur Farbumstimmung vorgeschlagen wird, sind es die Zapfen, die hier eine Rolle spielen könnten.
Entsprechend einer von Von Kries 1905 [Von Kries 1905] vogestellten Adaptation werden die Aktivierungen der Zapfen unter Berücksichtigung der allgemeinen Beleuchtungsverhältnisse gewichtet. Eine einseitige Dominanz bestimmter Bereiche des sichtbaren Spektrums innerhalb der Beleuchtung kann durch Verringerung der entsprechenden Zapfenempfindlichkeiten kompensiert werden.
Spektrale Reflexionsfunktionen von Oberflächen unter natürlichen Beleuchtungsverhältnissen können in der Tat durch Superposition der Aktivierungen dreier gewichteter Breitbandrezeptoren nachgebildet werden [Judd et al. 1964][Dixon 1978][Maloney 1986]. Die Berechnung der Farbumstimmung und damit die Erzeugung von Farbkonstanz erfordert die Kenntnis der spektralen Zusammensetzung des einfallenden Lichts. Präzise Farbkonstanz im Sinne der Von Kries-Adaptation der drei Rezeptoren ist nur beim Vorhandensein mehrerer reflektierender Oberflächen innerhalb einer Szene erzielbar [Hallet et al. 1988].
Bei einer Lokalisation der Farbumstimmung innerhalb der Retina wären große rezeptive Felder zur Feststellung der spektralen Zusammensetzung der Allgemeinbeleuchtung erforderlich, die die Neuronen der Retina aber nicht aufweisen. Andererseits bedarf die Farbumstimmung zu ihrer Realisierung keiner Augenbewegungen, die einen diesbezüglichen Vergleich unterschiedlicher Oberflächenreflexionen ermöglichen würden.
Somit kann eine retinale Lokalisation der Farbumstimmung im Sinne der Von Kries-Adaptation ausgeschlossen werden.
D'Zmura und Lennie (1986) [D'Zmura et al. 1986] vermuten die Mitwirkung farbopponenter Zellen des Corpus geniculatum laterale bei der Umsetzung der Farbumstimmung. Tatsächlich ist die Farbinformation auf ihrem Weg von den Zapfen der Retina über den CGL hin zum visuellen Cortex hier als RG-, BY- und WS-Aktivierung bereits so modifiziert bzw. aufbereitet, daß alle Voraussetzungen für eine Farbumstimmung gegeben sind. Obwohl die Information beider Gesichtsfeldhälften in den CGL-Hälften noch getrennt vorliegt, existieren CGL-Zellen mit kurzreichenden Axonen, die die erforderlichen synaptischen Verbindungen zu anderen CGL-Zellen bilden könnten. Unwahrscheinlich ist, daß die unabhängige Umstimmung der beiden Gesichtsfeldhälften unter allen denkbaren Beleuchtungsverhältnissen und Bildinhalten zu identischen Ergebnissen führen würde. Daher erscheint eine ausschließliche Lokalisation der Farbumstimmung innerhalb des CGL fraglich.
Nach unserer Auffassung naheliegender wäre vielmehr eine bezüglich der Retina nicht-

topographisch organisierte, cortikale Struktur, die in der Lage ist, den gesamten physiologischen Farbraum zu repräsentieren. Als direkten Input müßte eine solche Struktur parvozelluläre RG-, BY- und WS-Aktivierungen des vollständigen Gesichtsfeldes von beiden Hälften des CGL erhalten, diese durch Umstimmung modifizieren und rückwirkend in der V1 den Aufbau farbkorrigierter Merkmalskarten ermöglichen. Somit wäre eine Umstimmung in einem physiologischen Farbraum sinnvoll, in dem die Farbinformation des gesamten Gesichtsfeldes gleichberechtigt vorliegt. Auch dürfte die in einer topographischen Repräsentation enthaltene Größe von Regionen identischer Farbreize im bezüglich der Retina vermutlich nichttopographisch organisierten physiologischen Farbraum keine Rolle mehr spielen.
Physiologische Erkenntnisse [Land et al. 1983a] deuten darauf hin, daß die früheste, denkbare Struktur zur Repräsentation eines solchen physiologischen, nichttopographisch organisierten Farbraumes die V4 darstellen könnte. Nur sie besitzt gleichzeitig nichttopographisch angeordnete (nichtretinotope), farbsensitive Zellen [Zeki 1980] und die erforderlichen Verbindungen beider Hirnhemisphären in Form des Corpus callosum. Somit wäre hier erstmals Information über das gesamte Gesichtsfeld verfügbar. In Kapitel 4 wird diese Arbeitshypothese bei der Ableitung der Modellarchitektur zur Farbverarbeitung wieder aufgegriffen.

2.4 Zusammenfassung

Die Ausführungen zu den neurobiologischen, psychophysischen und cognitiven Grundlagen der entwickelten Architektur sollen die im folgenden Kapitel beschriebene Modellstruktur motivieren. Dabei ist die Anlehnung an das biologische Vorbild Gehirn auf den verschiedenen Verarbeitungsebenen unterschiedlich stark ausgeprägt. Dies liegt vor allem an dem oftmals unvollständigen und unsicheren Wissen bezüglich der Wechselwirkung der verschiedenen Substrukturen des Gehirns, was insbesondere auf die Interaktion der höheren Cortexareale zutrifft. Dieses unsichere, unvollständige Wissen eröffnet einen breiten Interpretationsspielraum und macht damit wiederum die Rückprojektion der Simulationsergebnisse auf die Biologie schwierig.
So sind z.B. die implementierten Strukturen und Mechanismen zur selektiven visuellen Aufmerksamkeit stärker von den psychophysischen Untersuchungen geprägt und orientieren sich daher nur relativ vage an detaillierten neuroanatomischen und neurophysiologischen Gegebenheiten. Es wurde jedoch darauf geachtet, daß das Modell durchgängig biologisch motivierbar bleibt.

Auf notwendige pragmatische Einschränkungen, die in erster Linie die simulationstechnische Beherrschbarkeit der Gesamtarchitektur dienten, wird an den entsprechenden Stellen in den einzelnen Kapiteln eingegangen.

Kapitel 3

Modellarchitektur und formale Verhaltensdefinition

Im ersten Teil dieses Kapitels werden noch einmal die Verhaltensaspekte ausführlich diskutiert, die den Ausgangspunkt für die zeitliche Objektrepräsentation und damit für die Organisation eines emergenten Objektverständnisses bilden. Zunächst wird auf der Basis der im Kapitel 2 dargelegten Grundlagen aus Psychophysik und Neurophysiologie ein funktionales Modell abgeleitet. Ausgehend von Untersuchungen zur sakkadischen Bildanalyse wird dann versucht, mögliche Parallelen zwischen externem und internem Scanning aufzuzeigen. Anhand dieser Überlegungen werden dann prinzipielle Verhaltensweisen abgeleitet, die sich nach unserer Hypothese durch die Selbstorganisation des Objektverständnisses ergeben müssen.

Um die abstrakte Ebene der Modellbeschreibung möglichst schnell zu verlassen und die Transparenz der Darstellung zu erhöhen, werden im zweiten Abschnitt dieses Kapitels bereits die zu erbringende Systemleistung anhand eines realen Szenarios erläutert und in diesem Kontext die im Rahmen des Selbstorganisationsprozesses verwendeten Szenen und ausgewählte Simulationsergebnisse einzelner Verarbeitungsstufen sowie des Gesamtmodells präsentiert. Die nachfolgenden Kapitel beschreiben dann detailliert die entsprechenden Teilstrukturen.

3.1 Funktionales Modell

3.1.1 Neurophysiologische und psychophysische Bezüge

Die Abbildungen 3.1 und 3.2 fassen die abgeleitete Modellstruktur unter physiologischen (Abbildung 3.1) und psychophysischen (Abbildung 3.2) Gesichtspunkten zusammen.

Die Objekte, die im Rahmen dieses Projektvorhabens verwendet wurden, besitzen hinreichend innere Struktur, um nur attentiv identifiziert werden zu können (siehe Abbildung 3.7). Damit bilden die attentiven Ebenen die wesentliche Grundlage für die zeitliche Objektrepräsentation. Von den attentiven Ebenen (Abbildung 3.2) wurde modellhaft angenommen, daß sie mit einem Aufmerksamkeitsfenster fester Größe arbeiten. Dies stellt zwar eine Vereinfachung gegenüber den aktuellen Modellen zur selektiven visuellen Aufmerksamkeit (siehe Kapitel 2) dar, die Einschränkung wurde aber bewußt in Kauf genommen, da es im Rahmen des Projektvorhabens nicht vordergründig um die Entwicklung eines Modells zur visuellen Aufmerksamkeit ging, sondern um die Selbstorganisation zeitlicher Objektrepräsentationen, wobei die selektive Datenaufnahme mittels des Aufmerksamkeitsfensters zwar wesentlich, aber nicht von zentraler Bedeutung ist.

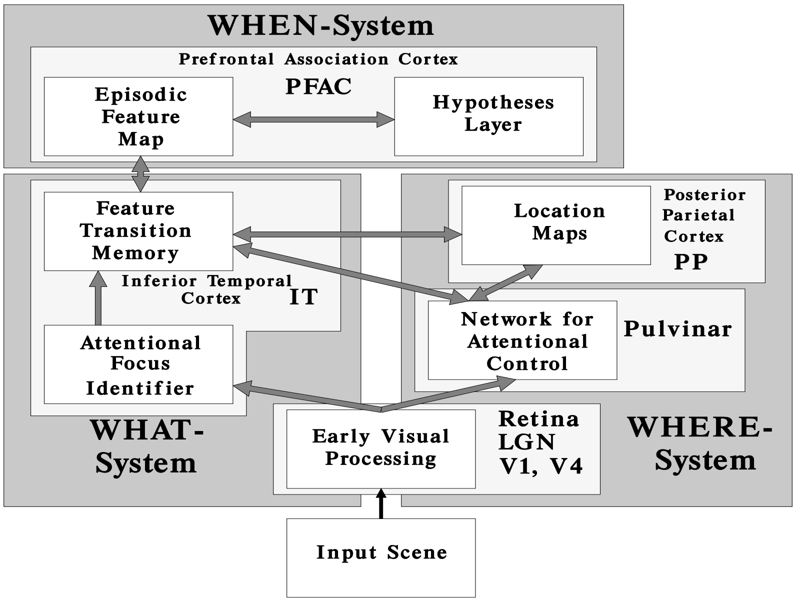


Abbildung 3.1: *Abbildung der neurobiologischen Grundlagen in die Modellarchitektur*

Die Größe des Aufmerksamkeitsfensters erfaßt in der Regel nur Teile eines Objekts. Das Aufmerksamkeitsfenster bewegt sich über die gesamte visuelle Szene, wobei die Art und Weise der Bewegung, ob gleitend oder sprunghaft, zunächst noch offen ist.

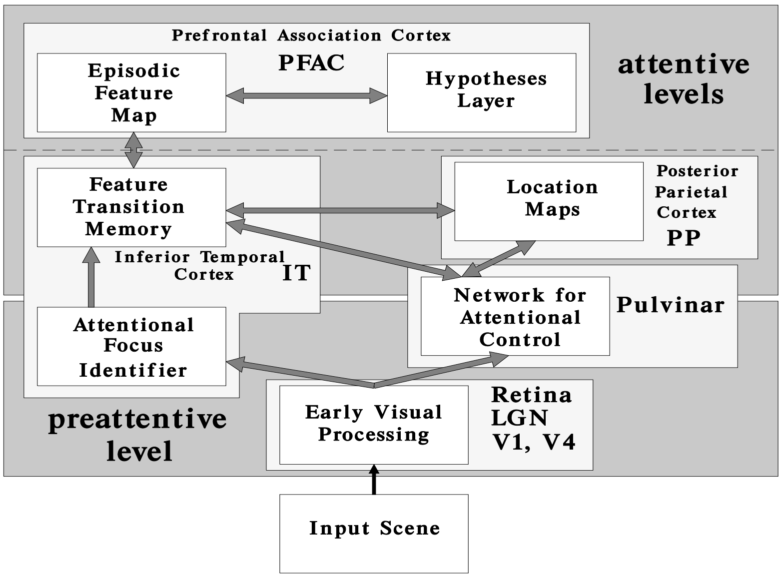


Abbildung 3.2: *Abbildung der psychophysischen Grundlagen in die Modellarchitektur*

Da die Wirkung der Auffälligkeitskarte in einer reflexartigen Positionierung des Aufmerksamkeitsfensters besteht, wird davon ausgegangen, daß die erste Position des Aufmerksamkeitsfensters durch die Position des Maximums der Auffälligkeitskarte bestimmt wird. Damit wird gleichzeitig festgelegt, daß die Auswertung der Auffälligkeitskarte durch einen WTA-Mechanismus (siehe auch [Koch et al. 1985]) erfolgen muß.
Es wird weiterhin davon ausgegangen, daß Positionen einer Szene, die bereits vom Aufmerksamkeitsfenster erfaßt wurden, kein weiteres Mal erfaßt werden sollten. Dies läßt sich durch eine (langanhaltende) Inhibition-of-Return motivieren, was in einem der folgenden Abschnitte noch einmal aufgegriffen wird.

Die attentive Verarbeitungsstufe der Modellarchitektur (erste und zweite attentive Ebene) wurde an den inferotemporalen Cortex (IT), den postparietalen Cortex (PP) und den präfrontalen Assoziationscortex (PFAC) angelehnt. Die präattentive Verarbeitungsstufe entspricht den Strukturen V1 und V4 im visuellen Cortex. Als Schnittstelle zwischen beiden Verarbeitungsstufen wird die Pulvinar angesehen.

Die robuste und reproduzierbare Extraktion von Merkmalen aus den präsentierten Szenen bildet eine entscheidende Voraussetzung für die Selbstorganisation eines verhaltensorientierten Objektverständnisses. Reproduzierbarkeit bedeutet dabei, daß z.B. die Merkmale bezüglich eines bestimmten Objekts von Szene zu Szene nur wenig variieren dürfen, da sonst keine stabilen Intraobjektbeziehungen zwischen Merkmalen als Basis des Objektverständnisses detektiert und repräsentiert werden können. Unter Robustheit soll die Stabilität der extrahierten Merkmale unter variablen Randbedingungen

verstanden werden, wobei im Rahmen des Projektes in erster Linie Variationen der Beleuchtungsverhältnisse in Bezug auf farbbasierte Merkmale behandelt wurden.
Die Extraktion von Merkmalen über das Bild vollzieht sich nach Abbildung 3.1 und 3.2 in der Teilstruktur *Early Visual Processing* (EVP). Implementiert wurden dabei neuronale Prinzipien und Mechanismen zur Farb- und Texturanalyse (siehe Kapitel 4) und zur Kantenextraktion.

Um einen schnellen Übergang von der frühen visuellen Verarbeitung zu einer an den IT angelehnten lokalen Repräsentation eines endlichen Satzes komplexer Merkmale zu vollziehen, wurde die Teilstruktur *Attentional Focus Identifier* (AFI) eingeführt. Diese transformiert die bezüglich einer Szenenlokalisation extrahierten Farb-, Helligkeits- und Texturmerkmale in ein begrenztes Alphabet von „Bildprimitiven“, die die verschiedenen Merkmalsausprägungen innerhalb eines lokalen Bildgebiets (des Aufmerksamkeitsfensters) in Form eines komplexen Merkmals beschreiben. Der Begriff des *komplexen Merkmals* soll hier für den Inhalt eines durch die Fokusregion erfaßten, lokalen Bildgebiets eingeführt werden. Die im Rahmen des Projektvorhabens modellierte Funktionalität des AFI läßt sich neurobiologisch dem IT zuordnen (siehe auch Abbildungen 3.1 und 3.2), wobei der AFI eher pragmatisch die Brücke vom *Early Visual Processing* zur zeitlichen Objektrepräsentation im *Feature Transition Memory* innerhalb des *What-Systems* bildet.

Das *Attention Control Network* (ACN) erstellt in seiner ersten Verarbeitungsstufe eine Auffälligkeitskarte der aktuellen Szene, deren Aktivitätswerte die rein datengetriebene Auffälligkeit der korrespondierenden Lokalisation im Eingangsbild repräsentieren. Diese Auffälligkeitskarte bildet zunächst die alleinige (präattentive) Basis für den Abtastprozeß über der momentan fovealisierten Szene, indem mit absteigender Auffälligkeit nacheinander verschiedene Lokalisationen analysiert werden. Dem ACN obliegt weiterhin die Aufgabe, die bei bereits akkumuliertem Systemwissen über die Umwelt mögliche wissensbasierte (attentive) Steuerung des Abtastpfades und die datengetriebene (präattentive) Steuerung des Abtastpfades zu koordinieren.

An den *Attentional Focus Identifier* (AFI) schließt sich unmittelbar innerhalb des *What-Systems* der biologisch an den IT angelehnte *Feature Transition Memory* an, welcher die eigentliche Basis einer zeitlichen Objektrepräsentation innerhalb der Gesamtarchitektur bildet. Die dafür notwendige Verarbeitung von Positionsinformation geschieht mit Hilfe der biologisch an den PP angelehnten *Location Maps* des *Where-Systems*. Der *Feature Transition Memory* bildet gemeinsam mit den *Location Maps* die erste attentive Ebene der Architektur, auf der Wissen über die zu analysierende, fovealisierte Szene akquiriert und für die wissensbasierte Steuerung des internen Abtastprozesses angewendet wird. Dabei werden *lokale* Hypothesen über den wahrscheinlichsten und damit erfolgversprechendsten nächsten räumlichen Übergang zu einem bestimmten Bildprimitiv gebildet und über die Modulation des Abtastprozesses verifiziert.

Die Teilstruktur *Episodic Object Memory* (EOM), die sich am präfrontalen Assoziationscortex orientiert, bildet die zweite attentive Ebene zur Repräsentation von Wissen innerhalb der Gesamtarchitektur. Ihre Notwendigkeit ergibt sich aus der Tatsache, daß im eigentlichen *What-System* ausschließlich Übergänge zwischen beim Abtastprozeß unmittelbar aufeinanderfolgenden komplexen Merkmalen repräsentiert werden können. Diese Repräsentation bildet eine unabdingbare Voraussetzung für die Organisation von Wissen über komplexe visuelle Strukturen der Umwelt (Objekte), stellt selbst aber erst eine Vorstufe des Objektwissens dar. Mit Hilfe des EOM kann das System jetzt mehrere, über den Abtastprozeß verifizierte Übergänge zwischen komplexen Merkmalen als mögliche Kandidaten für die Repräsentation von Objekten bzw. von Teilen von Objekten organisieren, die wiederum zur Bildung von *globaleren* Hypothesen über komplexe visuelle Strukturen herangezogen werden können. Die in Wechselwirkung mit der ersten attentiven Ebene stattfindende Einflußnahme auf den Abtastprozeß sichert auch hier die Verifikation der generierten Hypothesen in Form eines geschlossenen Hypothesentestzyklus.

Ausgehend von Neurobiologie und Psychophysik konnte modellhaft eine Zuweisung der Funktionalität einzelner Strukturen im Sinne von Funktionsblöcken (Abbildungen 3.1 und 3.2) sowie die Verschaltung der Funktionsblöcke untereinander geklärt werden. Allerdings bleiben noch eine ganze Reihe von Fragen offen: Welche Form hat die zeitliche Objektrepräsentation? Welche Funktion besitzen die Verbindungen zwischen den einzelnen Blöcken. Insbesondere, was bewirkt die Kopplung vom Attention Control Network zum Where-System? Welche Bewegungsart führt das Aufmerksamkeitsfenster beim internen Scanning aus? Wie funktioniert der Mechanismus zur Führung der attentiven Ebenen durch die Auffälligkeitskarte?
Im folgenden Abschnitt sollen insbesondere die verhaltensorientierten Grundlagen zur Beantwortung dieser Fragen diskutiert werden, um daran anschließend eine neuronale Struktur der Teilmodule der Modellarchitektur ableiten zu können. Dabei bildet die **Selbstorganisation** des emergenten, verhaltensorientierten Objektverständnisses als eigentliche Zielstellung des Projekts die Basis.

3.1.2 Externe und interne visuelle attentive Prozesse

Da angenommen werden kann, daß an der Erzeugung externer und interner attentiver Prozesse weitestgehend die gleichen biologischen Strukturen beteiligt sind (Where-System) [Desimone et al. 1989], sollen im folgenden experimentelle Untersuchungen zur sakkadischen Bildanalyse als Grundlage für die Motivation eines sinnvollen, beobachtbaren Gesamtverhaltens des Systems herangezogen werden.
Meßergebnisse zu Sakkaden, die [Yarbus 1967] entnommen wurden, sind in Abbildung 3.3 dargestellt. Die Verweildauer an einem Abtastpunkt lag bei diesem Experiment nach Angaben des Autors etwa zwischen 300ms bis 1.5s. Im allgemeinen wird davon ausgegangen, daß die Augenbewegungen ein Ausdruck cognitiver Prozesse im Gehirn sind. So variiert z.B. der Abtastpfad in Abhängigkeit von der gestellten Aufgabe („Wie alt sind die Personen?“ oder „Welche materiellen Umstände der Personen liegen vor?“).

Interessant sind die Ergebnisse zu sogenannten „fragmented“-Bildern [Stark et al. 1981]. Abbildung 3.4 zeigt ein solches Bild. Die Versuchspersonen wurden dabei aufgefor-

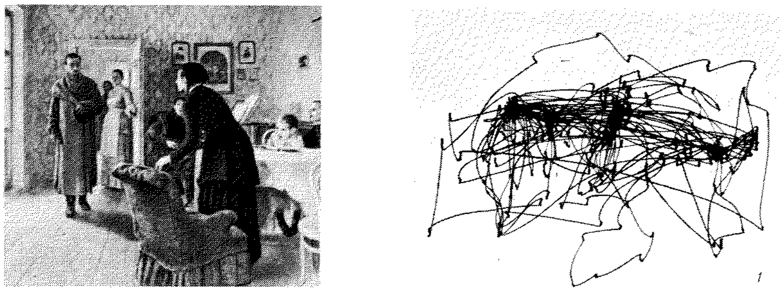


Abbildung 3.3: *Rechts sind die Augenbewegungen zu der Szene links dargestellt [Yarbus 1967].*

dert, ein Gesicht im Bild zu lokalisieren (oberer Bildbereich). Abbildung 3.5 zeigt die Veränderung des Abtastpfades nach der Identifikation des Gesichts. Der Abtastpfad ist nach der Identifikation deutlich auf das Gesicht konzentriert. Daraus kann geschlußfolgert werden, daß sich in diesem Falle eine zeitliche Objektrepräsentation des Gesichts im Gehirn ausgebildet hat (genauer im What-System).

Abbildung 3.4: *Diese Abbildung zeigt ein "fragmented"-Bild. Die Versuchspersonen sollen das Gesicht im oberen Bildbereich entdecken [Stark et al. 1981].*

Das Experiment zu dem „fragmented"-Bild zeigt, daß nach der Identifikation der Schwerpunkt des Abtastpfades auf dem Gesicht liegt. Geht man davon aus, daß die Sprünge, die vom Gesicht wegführen, keinen wesentlichen Beitrag zur Identifikation des Gesichts leisten, so reduziert sich das Abtastverhalten auf ein sukzessives Abtasten des Objekts in der Szene (Abbildung 3.5). Überträgt man dieses Verhalten auf die Bewegung eines internen Aufmerksamkeitsfensters, so kann dies die Grundlage für das Ausbilden einer zeitlichen Objektrepräsentation bilden. Die Abbildung 3.3 zeigt auch, daß diese Annahme eine starke Vereinfachung des tatsächlichen Verhaltens bei der Identifikation von Objekten in einer Szene ist.
Mit dem Verhalten des sukzessiven Abtastens von Objekten einer Szene könnte ein zeitliches Binden der einzelnen Objektbestandteile erreicht werden, was die Abbildung 3.6 noch einmal in abstrakter Form verdeutlicht. Notwendigerweise kann dieses Verhal-

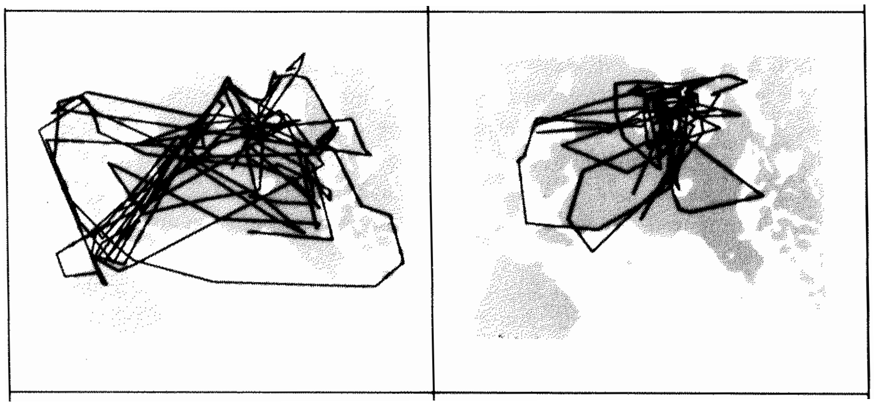


Abbildung 3.5: *Diese Abbildung zeigt die Veränderungen des Abtastpfades, die bei der Identifikation des Gesichts entstehen [Stark et al. 1981].*

ten nur durch einen Analyse**prozeß** über der Szene (und damit den Objekten) erfolgen. Hierbei müssen einzelne Bestandteile der Objekte und deren räumliche Relationen in der Szene gefunden und mit dem bisherigen Wissen über Objekte in Einklang gebracht werden. Aufgrund der Annahme einer Isomorphie zwischen Analyseprozeß und Repräsentation muß dieser räumlich-zeitliche Analyseprozeß auch intern eine zeitliche Objektrepräsentation erzeugen.

Abbildung 3.6: *Illustration einer Verhaltensänderung, der die Selbstorganisation der zeitlichen Objektrepräsentation zugrunde liegt.*

Im Experiment mit den „fragmented"-Bildern findet eine Veränderung des Abtastverhaltens statt: Das Verhalten vor der Identifikation kann vermutlich als Vergleich der vorhandenen Strukturen mit bereits bekannten Gesichtsstrukturen aufgefaßt werden. Bei der Identifikation „schlägt" das Verhalten in einen völlig anderen Abtastpfad um. Diese Verhaltensänderung soll auch auf die Selbstorganisation der zeitlichen Objektrepräsentation übertragen werden. Allerdings liegen hier die Dinge etwas anders: Die Selbstorganisation soll anhand einer Reihe von 2D-Szenen erfolgen, wobei die attentive Ebene erst Wissen über die in den Szenen enthaltenen typischen Objekte akquirieren muß. Am Anfang verfügt die attentive Ebene zunächst über kein Wissen über die Objekte. Dies bedeutet, daß sie die präattentive Ebene nicht beeinflussen kann. Damit kann eine Bewegung des Aufmerksamkeitsfensters zunächst nur auf der Basis des präattentiven Prozesses erfolgen, also auf der Grundlage von Auffälligkeitsprozessen. Dieser Zustand bildet den Ausgangspunkt der Selbstorganisation der zeitlichen Objektrepräsentation. Damit ergibt sich eine Verhaltensänderung von auffälligkeitsbasiertem Verhalten zu einem wissensbasierten Verhalten (Abbildung 3.6).

3.1.3 Formale Verhaltensdefinition

Ziel dieses Abschnitts ist es, die theoretischen Grundlagen zur Entwicklung einer neuronalen Modellarchitektur zur Selbstorganisation einer zeitlichen Objektrepräsentation zu legen. Dabei werden einige Annahmen aus dem funktionalen Modell verwendet. Nur implizit werden hier die strukturellen Randbedingungen, wie sie sich aus dem funktionalen Modell und den physiologischen Modellannahmen ergeben, berücksichtigt. D.h. die hier vorgestellten Überlegungen müssen zwar in eine neuronale Modellarchitektur umsetzbar sein, die konkrete Umsetzung wird jedoch erst in den folgenden Kapiteln diskutiert.
Um die theoretischen Grundlagen zu legen, wird die Vorstellung von einem Agenten, der sich in einer speziellen Umwelt aus fovealisierten visuellen Szenen bewegt, entwickelt. Dazu wird zunächst das aus den experimentellen Untersuchungen zu Sakkaden ableitbare zeitliche Verhalten mathematisch formalisiert und auf das Verhalten des Agenten übertragen, welches sich durch Selbstorganisation (Interaktion mit der Umwelt) ausbilden soll. Anschließend wird die grundlegende Strategie der Selbstorganisation in Form eines Reinforcement-Ansatzes motiviert. Abschließend werden einige Heuristiken vorgestellt, die sich auf Grund des funktionalen Modells und der 2D-Umwelt ergeben und die zu einer Erweiterung des Verhaltens des Agenten führen.

Definition des Zielverhaltens des Agenten

Um das Zielverhalten zu formalisieren, wird zunächst eine mathematische Beschreibung von Objekten in einer 2D-Umwelt vorgestellt. Anschließend wird diese Formalisierung genutzt, um das Zielverhalten des Agenten zu definieren.

Beschreibung von Objekten in einer 2D-Welt Im Rahmen des Projektvorhabens wurde die Repräsentation von Objekten auf eine translationsinvariante Repräsentation von 2D-Objekten beschränkt. Folgt man den Betrachtungen von [Haralick et al. 1993] und [Koenderink et al. 1979], so kann ein Objekt in einer 2D-Umwelt translationsinvariant durch einen „Aspekt Graph“ oder ein „View Class Modell“, d.h. durch eine Menge von räumlichen Relationen zwischen zwei Merkmalsvektoren, beschrieben werden:

$$O_k = \{(m, m^*, \Delta x, \Delta y) | m, m^* \in \mathbb{R}^n_D, \Delta x, \Delta y \in \mathbb{N}\}$$

Dabei sind m und m^* die beiden Merkmalsvektoren, Δx, Δy ist der räumliche Abstand zwischen den Merkmalsvektoren. Der räumliche Abstand bedeutet hier einen diskreten Abstand, der pixelweise bestimmt wird. $\mathbb{R}^n_D$ beschreibt eine Diskretisierung des $\mathbb{R}^n$. Die Form der Diskretisierung wird im Rahmen der lokalen Repräsentation komplexer Merkmale im Feature Transition Memory (inferotemporaler Cortex) diskutiert (Kapitel 5).

Zielverhalten Das Verhalten des Agenten korrespondiert mit einer Szenenanalyse bzw. Objekterkennung im "klassischen" Sinne [Haralick et al. 1993]. Eine typische Vorgehensweise in diesem Rahmen ist ein Graph-Matching zwischen den Eingangsdaten und den in Form von Graphenrepräsentationen gespeicherten Objekten. Dieses Verfahren gehört zu der Klasse der NP-vollständigen Probleme. In [Malsburg et al. 1986] wurde ein biologisch motivierter Ansatz vorgestellt, der mit Hilfe eines inexakten Graph-Matching die Problematik der NP-Vollständigkeit umgeht. Hierbei ist es notwendig, daß das System in einen Fixpunkt konvergiert, um eine Identifikation des Objekts zu erreichen. Das Modell vernachlässigt somit die zeitliche Struktur der biologischen Informationsverarbeitung.
Im Rahmen dieses Vorhabens wurde ein anderer Ansatz gewählt: Im funktionalen Modell wurde angenommen, daß sich ein Aufmerksamkeitsfenster fester Größe über eine Szene hinwegbewegt. Verbindet man diese Modellvorstellung mit der obigen Graphenbeschreibung eines Objekts, so liegt es nahe, daß die Merkmalsvektoren dieser Beschreibung den Inhalt eines Aufmerksamkeitsfensters darstellen. Verbindet man diese Feststellung mit der zeitlichen Bindung von Objektbestandteilen, so ergibt sich folgende formale Beschreibung des Zielverhaltens des Agenten (V_k) bei der Existenz eines Objekts (O_k) in einer Szene:

$$\begin{aligned} V_k &= \{(m(n), m^*(n+1), \Delta x(n), \Delta y(n))| \\ &\quad n = 0 \ldots l, \forall (m(n), m^*(n+1), \Delta x(n), \Delta y(n)) \in O_k\} \end{aligned}$$

Dabei bezeichnen $m(n)$, $m^*(n+1)$ den Merkmalsvektor am n-ten bzw. $(n+1)$-ten Abtastpunkt und $\Delta y(n)$, $\Delta x(n)$ die räumliche Distanz zwischen dem n-ten und dem $(n+1)$-ten Abtastpunkt. Wenn n Objekte in einer Szene vorkommen, wird das Verhalten zu:

$$R = \bigcup_{k=1}^{n} \quad \{(m(n), m^*(n+1), \Delta x(n), \Delta y(n))\}$$

mit

$$n = l_{k-1} \ldots l_k, (m(n), m^*(n+1), \Delta x(n), \Delta y(n)) \in O_k$$

Die Betrachtungsweise des Analyseprozesses als Sequenz von Bewegungen bringt ebenfalls kombinatorische Probleme (in Analogie zum "Graph-Matching"): Würde man z.B. jede mögliche Sequenz, in der ein Objekt abgetastet werden kann, speichern wollen, so ergäben sich pro Objekt l^l zu speichernde Sequenzen, wobei l die mittlere Anzahl der Merkmalsvektoren pro Objekt ist. Aus diesem Grund wird nur lokales Umweltwissen gespeichert:

$$U = \{(m_i, m_i^*, \Delta x_i, \Delta y_i, \alpha_i)| m_i, m_i^* \in \mathbb{R}^n_D, \Delta x_i, \Delta y_i \in N, i = 0 \ldots k\}$$

Wobei k zunächst die Anzahl aller möglichen, räumlichen Relationen ist, die in der 2D-Umwelt vorkommen können und α_i die i-te Relation dahingehend bewertet, inwieweit sie zu einem Objekt gehört. Deshalb wird α_i im folgenden als Objektzugehörigkeitswert bezeichnet.
Das Wissen über die Umwelt ist zunächst lokal (Wissen über einzelne, stabil auftretende Merkmalsübergänge). Diese Aussage bezieht sich bezüglich der Modellarchitektur (siehe Abbildungen 3.1 und 3.2) auf die erste attentive Ebene, die durch die beiden Strukturen Feature Transition Memory (FTM) und Location Maps (LM) gebildet wird. Obwohl mit Hilfe der zweiten attentiven Ebene (Episodic Object Memory (EOM), gebildet durch Episodic Feature Map (EFM) und Hypotheses Layer (HL)), das lokale Wissen zu einem mehr globalen Wissen über die Umwelt organisiert wird, bleibt die hier vorgestellte abstrakte Verhaltensdefinition in vollem Umfang gültig.

Bezüge zu offenen Fragen im funktionalen Modell Im Rahmen der oben getroffenen Definitionen lassen sich Antworten auf offene Fragen, die sich im Rahmen des funktionalen Modells stellen, angeben. So läßt sich entscheiden, ob eine sprunghafte oder gleitende Bewegung des Aufmerksamkeitsfensters sinnvoll ist. Eine gleitende Bewegung würde eine Vielzahl von räumlichen Merkmalsrelationen bedeuten. Dagegen kann ein sprunghaftes Verhalten eine deutliche Reduktion der Anzahl der Elemente der Menge U bewirken, d.h. man erreicht eine effektivere Repräsentation der Objekte.
Ebenfalls kann im Rahmen der zeitlichen Objektrepräsentation eine langanhaltende Inhibition of Return motiviert werden. Eine abklingende Inhibition würde ermöglichen, bereits erreichte Punkte in der Szene wieder anzuspringen. Dies hätte eine redundante Repräsentation zur Folge, da sich diese räumlichen Merkmalsrelationen aus bereits angesprungenen Positionen ermitteln lassen. Es ergibt sich somit eine effektivere Objektrepräsentation mit langanhaltender Inhibition of Return.

Selbstorganisation mit Reinforcement-Lernen

Definition eines Agenten und seiner Umwelt Die Betrachtungsweise der Szenenanlyse als Bewegung in der Umwelt "Szene" führt zur Beschreibung der neuronalen Modellarchitektur als Agenten, der sich in der Szene bewegt und die Aufgabe hat, eine zeitliche Objektrepräsentation zu erzeugen. Im Rahmen dieses Ansatzes läßt sich nun folgende mathematische Modellierung vornehmen [Kaelbling 1993]:
Die Umwelt besteht aus einem Tripel (S, A, W), wobei gilt:
S: Menge aller möglichen Zustände der Umwelt

A: Menge der möglichen Aktionen des Agenten

W: Zustandsübergangsfunktion $S \times A \to S$
Der Agent wird durch ein 4-Tupel beschrieben $(I, \mathcal{I}, R, B)$, wobei gilt:

$\mathcal{I}$: ist die Menge der möglichen Eingänge der Umwelt in den Agenten

I: ist die Abbildung von S nach $\mathcal{I}$, die bestimmt, welchen Eingang der Agent empfängt, wenn die Umwelt im Zustand S ist.

R: ist die Belohnungsfunktion des Agenten, die S auf eine reelle bzw. Bool'sche Zahl abbildet.

B: ist das Verhalten eines Agenten, das $\mathcal{I}^*$ nach A abbildet.

Damit ergibt sich folgender zeitlicher Ablauf: Zu einem Zeitpunkt t erhält der Agent den Eingang $i(t) \in \mathcal{I}$ und den Belohungswert $r(t)$, der sich aus der Funktion R ergibt. Die Funktion B erzeugt dann in Abhängigkeit von diesen Eingängen und möglicherweise internen Zuständen die Aktion $a(t) \in A$. Die Aktion bewirkt in der Umwelt gemäß der Funktion S eine Veränderung, die wiederum einen neuen Eingang $i(t+1)$ und $r(t+1)$ erzeugt.
Ziel des Agenten soll es sein, die Belohnungswerte $r(t)$ über der Zeit zu maximieren. Das Ereichen dieses Ziels kann durch folgende Randbedingungen erschwert sein:

1. Der Agent kann nicht alle Zustände der Welt unterscheiden. Dadurch erscheint es, als würde die Welt unterschiedlich auf für den Agenten gleiche Eingänge reagieren. Die Welt ist damit inkonsistent für den Agenten.

2. Die Sensoren bzw. Aktuatoren des Agenten sind fehlerbehaftet.

3. Die Übergangsfunktion S ist probabilistisch. Dieser Aspekt wird eine wesentliche Rolle spielen (siehe Kapitel 5).

Es gibt in der Literatur zahlreiche Lernverfahren, die das Verhalten B den Umweltrandbedingungen anpassen. Dabei sind insbesondere die Dyna-Architektur von Sutton [Sutton 1991], das Q-Learning von Watkins [Watkins 1992] und die Adaptive-Heuristic-Critic-Methode von Sutton [Sutton 1984] zu nennen. Diese Verfahren beschäftigen sich im wesentlichen mit der Problematik einer verzögerten Belohnung, d.h. der Agent erhält eine Belohnung für eine Aktion erst zu einem späteren Zeitpunkt. Diese Ansätze wurden nicht berücksichtigt, da eine direkte Verhaltensbewertung (Belohnung oder Bestrafung) notwendig war. Eine ausführliche Literaturübersicht zu Reinforcement-Ansätzen findet man in [Kaelbling 1993].

Agent in seiner speziellen Umwelt "Szene" Überträgt man die Modellierung eines autonomen Agenten in der Umwelt auf den Analyseprozeß über einer Szene, so ergeben sich die folgenden Beziehungen:

S: entspricht der Szene. Eine Szene ist eine zwei-dimensionale Anordnung von mehr oder weniger komplexen Merkmalen $m^s(x,y) \in \mathbb{R}^n$.

A: Menge der möglichen relativen Abtastbewegungen $(\Delta x, \Delta y)$ innerhalb der Szene

W: Transformation der relativen Abtastbewegungen $(\Delta x, \Delta y)$ in tatsächliche Szenenkoordinaten (x, y).

$\mathcal{I}$: $\mathcal{I} = \{m | m \in \mathbb{R}\}$ ist die Menge der möglichen Eingänge (mehr oder weniger komplexe Merkmale)

I: An einem Abtastpunkt (x, y) in der Szene wird der entsprechende Merkmalsvektor ausgelesen: $m^I = m^s(x, y)$

B, R sind so zu wählen, daß der Agent lernt, eine zeitliche Objektrepräsentation zu erzeugen.

Verhalten B Der Agent habe am Abtastpunkt n den Eingangsmerkmalsvektor

$$m^I(n) = m^s(x(n), y(n))$$

Dabei liegt $(x(n), y(n))$ auf einem Objekt. Das Verhalten B muß nun im Sinne der Zielstellung eine Aktion generieren, die erstens sehr wahrscheinlich zu einem Punkt in der Szene führt, der ebenfalls zu diesem Objekt gehört und zweitens zu einem Merkmalsvektor führt, der tatsächlich an dieser Szenenlokalisation vorhanden ist.
Die erste Bedingung kann B nur mit Hilfe seines impliziten Umweltwissens (U) entscheiden, insbesondere mit Hilfe der Objektzugehörigkeitswerte α_i.
Die zweite Bedingung stellt sicher, daß die zeitliche Objektrepräsentation tatsächlich mit dem in der Szene vorkommenden Objekt verknüpft ist.
Damit erzeugt B am Abtastpunkt n die Aktion $(\Delta x(n), \Delta y(n))$, für die gilt:

1. $(m_i(n), m_i^*(n+1), \Delta x(n), \Delta y(n), \alpha_i) \in U$, $m_i(n) \approx m^I(n)$, α_i maximal
2. $m_i^*(n+1) \approx m^s(x(n) + \Delta x(n), y(n) + \Delta y(n))$

$\approx$ soll andeuten, daß hier keine exakte Gleichheit im mathematischen Sinn gefordert werden kann, da der gespeicherte Vektor m_i^* und der Merkmalsvektor m_i^s in der Szene nie exakt, sondern nur im Rahmen der gewählten Diskretisierung übereinstimmen können.

Eine Abtastbewegung, für die die obigen Bedingungen zutreffen, wird im folgenden als *erfolgreicher Sprung* bezeichnet. Es wird im folgenden dabei auch von einer wissensbasierten Abtastbewegung gesprochen.
Die oben genannten Bedingungen lassen sich durch eine einfache Suche erfüllen. Dabei werden die α_i's der in Frage kommenden 4-Tupel in absteigender Reihenfolge durchsucht und jedesmal getestet, ob die 2. Bedingung erfüllt ist. Diese Suche erfolgt, bis die 2. Bedingung erfüllt ist. Der Fall, daß die 2.Bedingung nicht erfüllt werden kann, wird in Kapitel 5 diskutiert. Damit wird klar, daß die Position des aktuellen Abtastpunkts gespeichert werden muß, um einen erfolgreichen Sprung selektieren zu können.

Belohnungsfunktion R Der Suchvorgang läßt sich auch als Hypothesentestung verstehen, d.h. der Agent erwartet an der Position $(x(n) + \Delta x(n), y(n) + \Delta y(n))$ das Merkmal m_i^*. Ist dies nicht der Fall, so muß diese Erwartung bestraft werden. Aufgrund dieser Überlegungen muß für die Belohnungsfunktion R folgendes gelten:

$$r(t) = \begin{cases} 0 & \text{Sprung nicht erfolgreich} \\ 1 & \text{Sprung erfolgreich} \end{cases}$$

Dabei steht t nun für jeden Schritt im oben erläuterten Suchvorgang.

Lernregel Ausgangspunkt für den Selbstorganisationsprozeß ist die Annahme, daß stabile, räumliche Relationen von Merkmalen zu Objekten gehören (sogenannte Intraobjektbeziehungen). Daraus ergibt sich, daß häufig vorkommende, stabile räumliche Relationen von Merkmalen mit hohen α-Werten bewertet werden müssen, wohingegen räumliche Relationen zwischen Merkmalen auf Objekten und Merkmalen aus dem Hintergrund bzw. zwischen Merkmalen auf verschiedenen Objekten (sogenannte Interobjektbeziehungen) selten vorkommen (Instabilität dieser Relationen) und mit niedrigen α-Werten bewertet werden müssen.
Diese Überlegungen können auf das Auftreten der *erfolgreichen Sprünge* übertragen werden: Häufig auftretende *erfolgreiche Sprünge* gehören zu Objekten und selten auftretende *erfolgreiche Sprünge* gehören nicht zu Objekten. Damit wird der Mittelwert von $r(t)$ zum Maß für die Stabilität von Sprüngen (Stabilität des impliziten Umweltwissens).
Da ein wesentlicher Aspekt der Selbstorganisation das On-Line-Lernen ist, müssen die Objektzugehörigkeitswerte α_i permanent verändert werden. Dies läßt sich durch den gleitenden Mittelwert realisieren:

$$\alpha_i(t+1) = \alpha_i(t) + \tau \cdot (r(t) - \alpha_i(t)) \quad ; \quad \alpha_i(0) = 0$$

Dabei steht t nun für jeden Schritt im oben erläuterten Suchvorgang und α_i gehört zu dem aktuell getesteten Sprung.

Erweiterung des Verhaltens des Agenten

Das Verhalten B des Agenten berücksichtigt bisher nicht, wenn mehrere Objekte in einer Szene vorhanden sind und wann das Verhalten terminiert. Außerdem ist noch nicht erläutert worden, wie die Menge U mit Elementen gefüllt, d.h. wie Umweltwissen mittels on-line-Lernvorgängen erworben wird. Diese Aspekte werden in den folgenden Abschnitten diskutiert.

Abbruchkriterien Der Agent terminiert bisher nicht. Ein denkbares Abbruchkriterium wäre: "Objekt erkannt". Da aber das Umweltwissen eine solche Entscheidung nicht zuläßt, mußten im Rahmen des Projektvorhabens andere Kriterien gefunden werden. Dabei wurde darauf geachtet, daß möglichst kein Vorwissen über die Objekte verwendet wird. Es gab zwei Zielstellungen für die Abbruchkriterien: erstens "Verlasse das aktuelle Objekt" und zweitens "Beende das Abtasten dieser Szene". Die erste Zielstellung wird auch als die Iniziierung eines *erzwungenen Sprungs* bezeichnet. Die Anzahl der erzwungenen Sprünge ergibt sich aus den zeitlichen Relationen zwischen internem und externem Abtastpfad. Die Verweildauer an einem Abtastpunkt beim externen Abtasten liegt zwischen 300ms und 1.5s. Innerhalb dieser Zeit läuft das interne Abtasten ab. Hier beträgt die Verweildauer etwa 60 ms. D.h es kann davon ausgegangen werden, daß zwischen 5 und 26 Abtastpunkte pro Szene möglich sind. Damit erscheint es sinnvoll, die Anzahl der erzwungenen Sprünge auf eins zu beschränken, um die angesprungenen Objekte noch sinnvoll, d.h. genügend detailliert, repräsentieren zu können. Die Anzahl der abzutastenden Objekte einer Szene sollte daher zunächst auf zwei beschränkt werden. Ansatzpunkt für die Kriterien bildet die Auffälligkeitskarte. Die konkrete Berechnungsvorschrift für die Kriterien wird im Kapitel 5 erläutert.

Auffälligkeitsbasierte Sprünge Zwei Fragen sind bislang noch offen: Erstens, was passiert, wenn kein erfolgreicher Sprung gefunden werden kann? Zweitens, wie wird das implizite Umweltwissen aufgebaut? Für diese Fälle stellt die im ACN erstellte Auffälligkeitskarte (siehe Abschnitt 3.1, Abbildungen 3.1 und 3.2) die Basis für das Abtastverhalten dar. Die Abtastbewegungen sollen sich dabei nach folgendem Verfahren ergeben:

1. Finde die Position (x, y), an der die Auffälligkeit $a(x, y)$ maximal ist.
2. Markiere die Position (x, y) in $a(x, y)$ so, daß eine andere Position maximal wird.
3. Wenn kein Abbruchkriterium erfüllt ist, beginne wieder bei 1.

Der 2.Schritt entspricht der "inhibition of return". Dieser Mechanismus erzeugt also eine Folge von Abtastbewegungen:

$$\{(m(n), m^*(n+1), \Delta x(n), \Delta y(n)) | n = 0 \dots l\}$$

Diese Abtastbewegungen werden im folgenden im Gegensatz zu den wissensbasierten Sprüngen als *auffälligkeitsbasierte Sprünge* bezeichnet. Mit diesen *auffälligkeitsbasierten Sprüngen* kann die Menge U, das Umweltwissen, aufgebaut und erweitert werden. D.h. immer dann, wenn der Agent keinen erfolgreichen Sprung findet, wird eine auffälligkeitsbasierte Abtastbewegung ausgeführt. Dieser Sprung wird, wenn er bisher noch nicht vorgekommen ist, in das Umweltwissen eingebaut. Dies hat den Vorteil, daß die Anzahl der Elemente der Menge eingeschränkt wird, da immer versucht wird, zunächst aufgrund des bisher bereits akkumulierten Umweltwissens abzutasten.

Nachbarschaftsbeziehung Die Einschränkung der Anzahl der Elemente des Umweltwissens durch die auffälligkeitsbasierten Sprünge stellte sich im Verlauf des Projektes als zu stark heraus. Aus diesem Grund wurde die Anzahl der Einträge in das Umweltwissen sinnvoll erweitert. Dabei wurde die Eigenschaft der Umwelt ausgenutzt, daß sich die Merkmalsvektoren in der lokalen Nachbarschaft eines Abtastpunktes selten stark ändern. Damit kann jeder neue Eintrag in das Umweltwissen aus einer Menge von $(l \times l)$-Tupeln bestehen:

$$\{(m, m^*, \Delta x + i, \Delta y + j) | i = -l/2 \dots l/2, j = -l/2 \dots l/2, m, m^* \in \mathbb{R}^n_D, \Delta x, \Delta y \in N\}$$

Dabei ist das Element der Menge mit $i = 0$ und $j = 0$ das Ergebnis des auffälligkeitsbasierten Sprungs. Im Rahmen der neuronalen Modellarchitektur wurde l auf 3 gesetzt.

Einfluß der Auffälligkeitskarte bei der Suche Im Rahmen des funktionalen Modells wurde angenommen, daß die Suche nach einer Zielregion durch die Auffälligkeitskarte geführt wird. Diese Führung der Suche wird hier auf die Führung der Suche nach einem erfolgreichen Sprung übertragen. Da die Suche nach einem erfolgreichen Sprung wesentlich durch die Objektzugehörigkeitswerte bestimmt wird, läßt sich die Führung folgendermaßen formalisieren:

$$\alpha_i^* = f(\alpha_i, a(x(n) + \Delta x_i, y(n) + \Delta y_i))$$

Damit wird jeder Sprung des Umweltwissens mit dem Auffälligkeitswert verknüpft, der an der Position steht, zu der der Sprung führen würde. Da in die Berechnung der aktuelle Abtastpunkt $(x(n), y(n))$ eingeht, muß die Auffälligkeitskarte bezüglich dieses Punktes bestimmt werden. Dies wird im folgenden als relative Auffälligkeitskarte bezeichnet.

Erweiterung der Belohnungsfunktion durch die Auffälligkeitskarte Der Einfluß der Auffälligkeitskarte stellt einen Vorgang für eine Szene dar. Zusätzlich kann man vermuten, daß im Umweltwissen die typischen Auffälligkeitswerte zu den räumlichen Relationen von Merkmalen gespeichert werden. Dies kann durch die Modifikation der Belohnungsfunktion R erfolgen:

$$r(t) = \begin{cases} 0 & \text{Sprung nicht erfolgreich} \\ 1 + a(x(n) + \Delta x_i, y(n) + \Delta y_i) & \text{Sprung erfolgreich} \end{cases}$$

Damit erhält der Agent eine zusätzliche Belohnung, wenn er zu auffälligen Punkten in einer Szene springt.

Erweiterung der Lernregel Das bisherige Verhalten des Agenten führt dazu, daß der Agent sofort das Abtastverhalten bestimmt, es sei denn, das Umweltwissen U reicht nicht aus. Diese Vorgehensweise hat zwei Nachteile: Erstens ist das Umweltwissen (insbesondere die α_i's) zu Beginn der Auseinandersetzung mit der Umwelt noch unsicher und beeinflußt damit das Abtastverhalten möglicherweise in eine falsche Richtung. Zweitens schränkt das Verhalten des Agenten die Möglichkeiten zum Ermitteln von neuem Umweltwissen stark ein. Simulationen haben gezeigt, daß diese Nachteile oft dazu führen, daß keine adäquate zeitliche Objektrepräsentation aufgebaut werden kann.

Aus diesen Gründen wurde eine Unterscheidung in sichere und unsichere Sprünge eingeführt. Unsichere Sprünge werden nur getestet, ob sie erfolgreich wären (interne Simulation), aber nicht wirklich ausgeführt. Dagegen werden sichere Sprünge, wenn sie erfolgreich sind, auch ausgeführt. Diese Vorgehensweise bietet die Möglichkeit, daß der Agent einerseits sein Wissen über die Umwelt stabilisieren und andererseits neues Umweltwissen akquirieren kann (Sicherung von Plastizität und Stabilität des Umweltwissens). Die Menge des Umweltwissens U wird durch weitere Attribute erweitert:

$$\begin{aligned} U^* = \{(m_i, \gamma_i), (m_i, m_i^*, \Delta x_i, \Delta y_i, \alpha_i, \beta_i) | \\ m_i, m_i^* \in I\!R_D^n, \Delta x_i, \Delta y_i \in N, \alpha_i, \beta_i, \gamma_i \in I\!R, i = 0 \ldots k\} \end{aligned}$$

Die Lernregeln für β_i und γ_i lauten:

$$\beta_i(t+1) = \beta_i(t) + \tau_1 \cdot (1 - \beta_i(t))$$

$$\gamma_i(n+1) = \gamma_i(n) + \tau_2 \cdot (1 - \gamma_i(n))$$

γ_i gibt an, wie oft m_i Ausgangspunkt einer Suche nach einem erfolgreichen Sprung war. β_i gibt an, wie oft das 4-Tupel getestet worden ist. Damit stellen β_i und γ_i indirekt eine biologisch-plausible Schätzung der Größe des Konfidenzintervalls von α_i

dar [Bosch 1992]. Die Größe des Konfidenzintervalls liefert gleichzeitig eine Aussage darüber, wie sicher das Umweltwissen ist. Überschreiten nun β_i und γ_i die Schwellen s_1 bzw. s_2, so wird der dazugehörige Sprung ausgeführt (aktiv) und nicht nur getestet (passiv).
Die Unterscheidung in β_i und γ_i resultiert aus der Problematik, daß früh eingetragene Sprünge schneller "sicher" werden als spät eingetragene Sprünge. Um die negativen Auswirkungen dieses Effekts zu dämpfen, wurde γ_i eingeführt.

Vermeidung "lokaler Minima" Das bisher entwickelte Verhalten des Agenten mit noch unvollständigem Umweltwissen läuft nun folgendermaßen ab: Der Agent tastet ein Objekt mit einigen wenigen Sprüngen ab. Dabei beginnt er mit der auffälligsten Stelle des Objekts und springt dann bevorzugt weitere auffällige Bereiche des Objekts an. Gelangt er nach wenigen Sprüngen an eine Position, an der keine sinnvollen Abtasthypothesen mehr generiert werden können, so liefert die Auffälligkeitskarte die nächste anzuspringende Position. Diese liegt allerdings sehr wahrscheinlich nicht mehr auf dem gleichen Objekt, da andere Bereiche im Bild, z.B. ein zweites Objekt, ähnlich auffällige Punkte besitzen wie das soeben abgetastete Objekt. Damit liefert die Auffälligkeitskarte mit hoher Wahrscheinlichkeit eine Position, die von dem soeben abgetasteten Objekt wegführt. In dieser Situation kann der Agent keine weiteren stabilen räumlichen Zusammenhänge innerhalb des Objekts „kennenlernen“.
Um diesen Effekt zu reduzieren war es notwendig, die auffälligkeitsbasierten Abtastbewegungen und die vom Agenten wissensbasiert erzeugten Abtastbewegungen „parallel“ ablaufen zu lassen, so daß in der obigen Situation die Wahrscheinlichkeit dafür erhöht wird, daß der unmittelbar auf die wissensbasierte Abtastung folgende auffälligkeitsbasierte Sprung zum gleichen Objekt führt. Dieser Mechanismus führte zur Implementierung einer "internen" Inhibitionskarte, deren Funktion in Kapitel 5 näher erläutert ist.

Zweite Attentive Ebene Die bisherigen Betrachtungen beziehen sich im wesentlichen auf die erste attentive Ebene. Ziel der zweiten attentiven Ebene ist es, eine Vervollständigung des Umweltwissens der Modellarchitektur zu erreichen. Dazu muß insbesondere die Lokalität des Umweltwissens der ersten attentiven Ebene erweitert werden.
Um diese Erweiterung zu erreichen, wird zunächst die Mengenbeschreibung eines Objekts (Abschnitt 3.1.3) aufgegriffen:

$$O_k = \{(m, m^*, \Delta x, \Delta y) | m, m^* \in \mathbb{R}^n_D, \Delta x, \Delta y \in \mathbb{N}\}$$

Die erste attentive Ebene umfaßt nun nur das Umweltwissen, welches besagt, inwieweit die räumlichen Merkmalsrelationen zu einem beliebigen Objekt gehören. Für eine Erweiterung dieses noch objektunspezifischen Umweltwissens lernt die zweite attentive Ebene, inwieweit eine räumliche Merkmalsrelation zu einem speziellen Objekt bzw. Teilobjekt gehört:

$$U_2 = \{(m_i, m_i^*, \Delta x_i, \Delta y_i, \theta_{ik}) | m_i, m_i^* \in \mathbb{R}^n_D, \Delta x_i, \Delta y_i \in N, i = 0 \ldots n\},$$

wobei θ_{ij} die räumliche Merkmalsrelation i bewertet, inwieweit sie zu dem Objekt O_k gehört. Damit beschreibt das nunmehr spezifische Umweltwissen der zweiten attentive Ebene vollständig ein Objekt.

Dieses Umweltwissen muß zunächst aus dem sequentiellen Ausgang der ersten attentiven Ebene extrahiert werden (R in Abschnitt 3.1.3). Im Rahmen dieses Projektvorhabens wird dazu zunächst dieser sequentielle Prozeß in eine parallele Repräsentation abgebildet. Dabei wird die Abbildung so gewählt, daß das Problem der kombinatorischen Explosion umgangen wird (siehe Abschnitt 5.3.3). Eine Sequenz (V) als Ausgabe der ersten attentiven Ebene

$$\begin{aligned} V &= \{(m(n), m^*(n+1), \Delta x(n), \Delta y(n))| \\ &\quad n = 0 \ldots l, \forall (m(n), m^*(n+1), \Delta x(n), \Delta y(n)) \in U^*\} \end{aligned}$$

wird auf eine Menge von räumlichen Merkmalsrelationen (M):

$$\begin{aligned} M &= \{(m, m^*, \Delta x, \Delta y)| \\ &\quad \forall (m, m^*, \Delta x, \Delta y) \in V\} \end{aligned}$$

abgebildet. Die konkrete Form der Abbildung wird in den Abschnitten zur zweiten attentiven Ebene (5.3 und 5.4) erläutert.

Da allerdings neben der Ermittlung des "statischen" Umweltwissens ("ein Objekt X besteht aus den Bestandteilen Y") auch das Abtastverhalten gemäß dieses erweiterten Umweltwissens beeinflußt werden muß, ist es notwendig, daß die zweite attentive Ebene auf die erste attentive Ebene Einfluß nimmt und deren Verhalten moduliert. Da das Verhalten der ersten attentiven Ebene wesentlich von dem unspezifischen Objektzugehörigkeitswert α_i^* bestimmt wird, muß die zweite attentive Ebene über diesen Wert mit dem spezifischen Objektzugehörigkeitswert (θ_{ij}) Einfluß nehmen:

$$\alpha_i^{**} = g(\alpha_i^*, \theta_{ij})$$

Die konkrete Abbildung wird in Abschnitt 5.5 erläutert.

Zusammenfassung

Das Umweltwissen des Agenten besteht aus räumlichen Merkmalsrelationen (in der ersten und zweiten attentiven Ebene des Modells repräsentiert), deren unspezifischer (erste attentive Ebene) und spezifischer (zweite attentive Ebene) Objektzugehörigkeit und deren Sicherheit. Auf der Grundlage dieses Umweltwissens ergibt sich folgendes Verhalten: Zunächst wird auffälligkeitsbasiert ein Abtastpunkt selektiert. Anschließend sucht der Agent nach einem erfolgreichen Sprung, der sicher ist. Findet er keinen erfolgreichen Sprung, so erfolgt eine auffälligkeitsbasierte Abtastbewegung. Diese Abtastbewegung erweitert gebenenfalls das Umweltwissen. Das Abtasten einer Szene terminiert auf der Grundlage der Auffälligkeitskarte. Ebenfalls durch die Auffälligkeitskarte wird die Suche nach einem erfolgreichen Sprung beeinflußt. Während dieser Vorgänge lernt der Agent permanent, d.h. es findet ein On-Line-Lernen statt.

Im Zusammenhang mit dem Wechselspiel zwischen den beiden attentiven Ebenen ergibt sich im Rahmen dieses Projektvorhabens eine drei-stufige Vorgehensweise im Selbstorganisationsprozeß der neuronalen Modellarchitektur: Die erste Stufe (auffälligkeitsbasierte Abtastbewegungen) schränkt die Komplexität der Umwelt in einem ersten Schritt ein. In einer zweiten Stufe wird die Komplexität der Umwelt durch die Extraktion des räumlich lokalen, aber noch objekt-unspezifischen Umweltwissens weiter verringert (erste attentive Ebene). Die zweite attentive Ebene muß "nur" noch das Ergebnis der Komplexitätsreduktion der ersten beiden Stufen verarbeiten. Sie muß allerdings dieses Ergebnis zu einem objekt-spezifischen Umweltwissen zusammenfassen. Damit ergibt sich einerseits über die einzelnen Stufen der Verarbeitung eine Komplexitätsreduktion der Umwelt, aber andererseits eine Erhöhung der Komplexität der Zielstellung der einzelnen Verarbeitungsebenen.

3.2 Szenario und exemplarisches Systemverhalten

3.2.1 Szenario

In der Abbildung 3.7 sind exemplarisch 9 der insgesamt 100 in der Galerie enthaltenen Szenen gezeigt, die dem System als Umwelt zur Verfügung standen und die in zufälliger Reihenfolge präsentiert wurden.
Bei der Gestaltung der Szenen und der Auswahl der Objekte waren die folgenden Aspekte relevant: Die Objekte sollten möglichst anschaulich sein und auch von einem externen Beobachter als solche aufgefaßt werden können. Der Hintergrund, vor dem die Objekte erscheinen, sollte möglichst komplex strukturiert sein, um keine a priori Segmentierung der Objektgebiete zuzulassen. Die Größe der Objekte mußte deren variable Anordnung gestatten, auch sollten mehrere Objekte pro Szene enthalten sein können.

3.2.2 Exemplarisches Systemverhalten

Im ersten Simulationsbeispiel in Abbildung 3.8 wurde ein rein datengetriebener, auffälligkeitsbasierter Abtastverlauf einem wissensbasierten Abtastverlauf gegenübergestellt. Wesentlich ist bei diesem Beispiel, daß für die Modulation des Abtastprozesses ausschließlich das noch rein lokale Umweltwissen der ersten attentiven Ebene verwendet wurde.
Die Abbildung 3.9 verdeutlicht, wie sich die Steuerung des Abtastverhaltens unter Nutzung von lokalen und globalen Hypothesen (Wechselwirkung zwischen erster und zweiter attentiver Ebene) als Effektivierung der Suche nach erfolgreichen Sprüngen auswirkt. In der linken Darstellung sind im Vergleich zu Abbildung 3.8 rechts alle die Suchsprünge grün eingetragen worden, die ausschließlich aufgrund lokaler Hypothesen vorgeschlagen, aber nicht verifiziert werden konnten. Die rechte Darstellung zeigt das Abtastverhalten bei Nutzung lokaler und globaler Hypothesen. Deutlich zu sehen ist die Reduktion der nicht verifizierbaren Suchsprünge.
Anhand der beiden Abbildungen 3.8 und 3.9 wird ebenfalls deutlich, wie sich das Abtastverhalten im Verlauf des Selbstorganisationsprozesses von einer anfänglich rein datengetriebenen, ziellosen Abtastung hin zur bevorzugten Abtastung der in den Szenen enthaltenen Objekte als Ausdruck eines impliziten, durch Interaktion mit der Umwelt erworbenen, verhaltensorientierten Objektverständnisses entwickelt.

Abbildung 3.7: *Beispielszenen aus der Galerie, welche insgesamt 100 Szenen umfaßte*

Abbildung 3.8: *Beispiele für einen rein datengetriebenen Abtastverlauf (links) und einen wissensbasierten Abtastverlauf (rechts), nachdem alle Szenen der Galerie ein einziges mal präsentiert wurden, wobei hier nur Wissen über lokale Übergänge zwischen unmittelbar aufeinanderfolgenden komplexen Merkmalen (erste attentive Ebene) zur Steuerung verwendet wurde.*

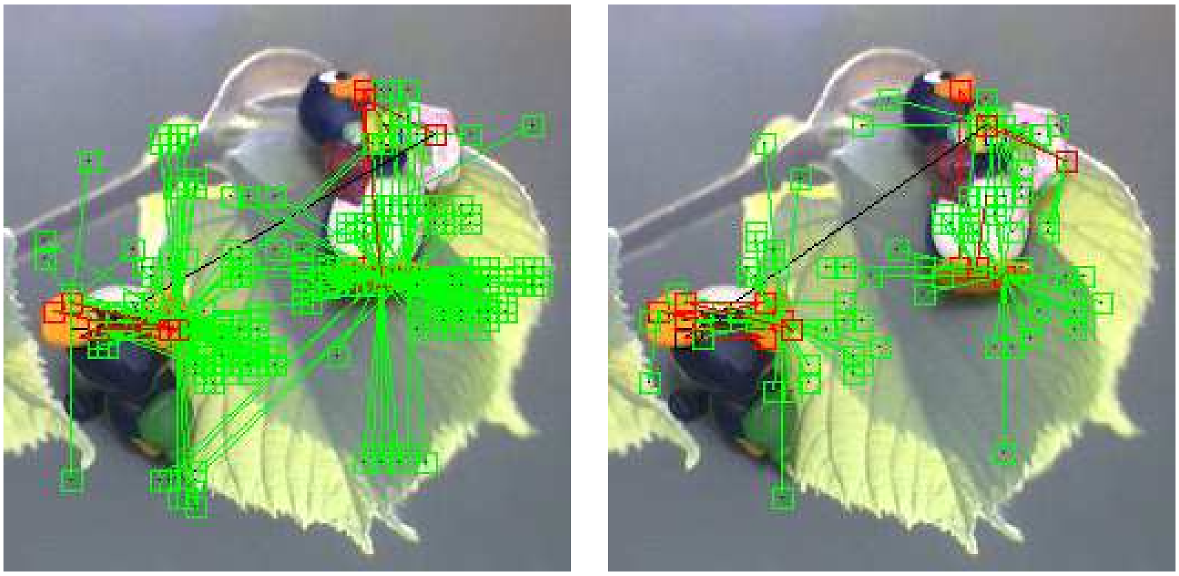


Abbildung 3.9: *Darstellung des Abtastverhaltens bei alleiniger Nutzung lokaler Hypothesen (links) und bei Nutzung lokaler und globaler Hypothesen (rechts)*

Kapitel 4

Robuste und reproduzierbare Extraktion bildbeschreibender Merkmale

Die robuste und reproduzierbare Extraktion bildbeschreibender Merkmale bildete eine tragende Säule des Projektvorhabens, da durch sie alle nachfolgenden Verarbeitungsstufen zur Selbstorganisation eines emergenten, verhaltensorientierten Objektverständnisses maßgeblich beeinflußt werden. Gelingt es beispielsweise nicht, auch unter variablen Randbedingungen (z.B. schwankende Beleuchtungsverhältnisse) für identische Szeneninhalte stets annähernd gleiche Merkmale zu extrahieren, so muß der Prozeß der Selbstorganisation von Wissen, der auf stabilen, wiederkehrenden „sensorischen Situationen“ aufbaut, konzeptionell zum Scheitern verurteilt sein.

Deshalb wurden in diesem Teilprojekt in Abstimmung mit den am Verbund beteiligten Partnern die biologisch motivierte Farb- und Texturverarbeitung stark in den Vordergrund gerückt.

Nach einem sehr straffen Überblick über die erarbeiteten Mechanismen zur Merkmalsextraktion befaßt sich der zweite Abschnitt dieses Kapitels mit den Untersuchungen zur biologisch motivierten Farb- und Texturanalyse, während im dritten Teil insbesondere die Auswirkungen der robusten und reproduzierbaren Bereitstellung von Farbmerkmalen auf die Identifikation lokaler Bildstrukturen im Zusammenhang mit dem „focus of attention“ behandelt werden.

4.1 Überblick

4.1.1 Farbverarbeitung

Durch abstrahierende Interpretation neurobiologischer und -anatomischer Erkenntnisse wurde ein 1993 vorgestelltes Modell De Valois' [De Valois et al. 1993] modifiziert und zu einem qualitativ neuen, umfassenderen Modell zur fovealen Wahrnehmung stabiler Farbempfindungen bei variierender spektraler Zusammensetzung der Allgemeinbeleuchtung weiterentwickelt.

Den Ausgangspunkt bilden die drei Zapfentypen der menschlichen Retina mit ihren spektralen Absorptionskurven (Abbildung 4.4 (rechts)). Der Kurvenverlauf wird als evolutionäre Anpassung an die Strahlungsverteilung des mittleren Tageslichts interpretiert. Obwohl L- und M-Zapfen bis zum heutigen Tage nicht anatomisch unterschieden werden können, lassen psychophysische Untersuchungen [Lennie et al. 1991] vermuten, daß L- und M-Zapfen in der Retina des Menschen in gleichen Verhältnissen vorhanden sind. Daher wird das Verhältnis der drei Zapfentypen L : M : S innerhalb der Fovea abweichend von Modellvorstellungen De Valois' als 6 : 6 : 1 zugrundegelegt.
Das entwickelte Modell orientiert sich in seinem stufenweisen Aufbau an der neuroanatomischen Struktur von fovealer Retina, CGL, V1 und V4 (siehe Kapitel 4) und folgt innerhalb dieser dem von Zeki [Zeki 1978a] [Zeki 1994] vermuteten, ausschließlich die Fovea analysierenden, direkten parvozellulären Farb-Path. Anatomiebezogene Erkenntnisse wurden in eine konkrete neuronale Verschaltung überführt und einer experimentellen Simulation unterzogen.

Aktuelle neurophysiologische Erkenntnisse bezüglich der Verschaltung von Zellen der Retina sowie des Corpus geniculatum laterale (CGL) des Menschen zeigen, daß das visuelle System eine Überführung der Zapfensignale in einen physiologischen Farbraum realisiert, die sich in einer Lineartransformation zusammenfassen läßt. Der entstehende Farbraum (siehe Abschnitte 4.4 und 4.5) wird durch die orthogonalen Achsen Rot/Grün (RG), Blau/Gelb (BY) sowie Weiß/Schwarz (WS) aufgespannt und umfaßt alle möglichen Farbempfindungen des Menschen. Er weist eine klare und plausible Geometrie auf, in der die sechs bunten und zwei unbunten, extremen Farbempfindungen des Menschen völlig gleichrangig und gleichberechtigt angeordnet sind. Die Kenntnis der Geometrie des physiologischen Farbraumes führt bei technischen Anwendungen zur Bestimmung einer direkten Lineartransformation zur Überführung von RGB-Werten in RG-, BY- und WS-Aktivierungen. Somit wird erst durch die Berücksichtigung der experimentell nachgewiesenen neurobiologischen Verschaltungsprinzipien und der daraus abgeleiteten mathematischen Abbildungsvorschrift in den physiologischen Farbraum ein direkter Zugang zur algorithmischen Umsetzung für technische Sehsysteme geschaffen.

Die in den physiologischen Farbraum transformierten Pixel eines einzelnen, übersteuerungsfreien RGB-Bildes ohne Gammakorrektur und Weißabgleich bilden dort eine Punktwolke. Die Position eines jeden Punktes der Wolke im Farbraum ist durch seine RG-, BY- und WS-Aktivierung eindeutig beschrieben.

Die Punktwolke ist entsprechend der bei der Bildaufnahme vorhandenen spektralen Zusammensetzung der Allgemeinbeleuchtung aus der WS-Achse (Unbuntachse) ausgelenkt. Diese Auslenkung der Punktwolke ist für jede spektrale Zusammensetzung der Allgemeinbeleuchtung charakteristisch und bei natürlichen Farbszenen mit üblicherweise einer Vielzahl unterschiedliche Strahlungsverteilungen reflektierenden, bunten Oberflächen unabhängig vom Bildinhalt.
Für jede ausgelenkte Punktwolke läßt sich eine ausgezeichnete Achse bestimmen, die in die WS- oder Unbuntachse überführt werden kann. Die Kenntnis dieser Hauptachse erlaubt die Optimierung der Lage eines jeden Punktes der Punktwolke bezüglich der RG-, BY- und WS-Achse des physiologischen Farbraumes. Verbunden mit einer psychophysisch motivierten Streckung der Punktwolke entlang der WS-Achse (koordinatenabhängige Dynamikanpassung) führt dies zu einer referenzfreien, automatischen Farbumstimmung des gesamten Bildes im physiologischen Farbraum und letztlich zu einem Weißabgleich. Eine inverse Transformation ermöglicht die Rücktransformation der in ihrer Lage optimierten Punkte der Wolke in den RGB-Farbraum und somit die Erzeugung einzelner oder einer Folge farbumgestimmter RGB-Farbbilder. Die Achsen des Farbraumes sind für eine Farbumstimmung optimal dekorreliert angeordnet. Die Geometrie des physiologischen Farbraumes ermöglicht unter allen Umständen die Bestimmung der relevanten Hauptachse. Der Farbfehler ist genau quantifizierbar. Der Einsatz der entwickelten Farbumstimmung zum automatischen, referenzfreien Weißabgleich stellt lediglich ein mögliches Anwendungsbeispiel unter anderen denkbaren dar.

Das entwickelte Verfahren zur Farbumstimmung stellt einen robusten Mechanismus zur Erzielung stabiler Farbempfindungen trotz wechselnder spektraler Zusammensetzung der Allgemeinbeleuchtung dar. Die Vorteile des Verfahrens bestehen zum einen in der Referenzfreiheit. Weiterhin gehen alle Bildpixel gleichberechtigt in die Bestimmung der Korrekturgröße ein. Die räumlichen Relationen der Punkte zueinander bleiben innerhalb der Wolke erhalten. Es kommt daher zu keiner Verfälschung der Farbinformation. Auf die Bedeutung der Farbumstimmung bzgl. des Projektziels wird im Abschnitt 4.8 ausführlich eingegangen.

Die Gesamtheit der für ein foveales Bild durch die Farbumstimmung optimierten RG-, BY- und WS-Aktivierungen können in drei bezüglich der Retina topographisch organisierten Aktivierungskarten zusammengefaßt werden. Als möglicher Ort einer derartigen Repräsentation der Aktivierungskarten wird die V1 nach einer Rückprojektion aus der V4 vermutet.

4.1.2 Texturanalyse

Das Modell zur fovealen Texturanalyse setzt auf der durch die Farbumstimmung optimierten WS-Aktivierungskarte auf und realisiert eine neurobiologisch motivierte Texturbeschreibung dieser Aktivierungsverteilung durch ein Array cortikaler Verarbeitungscolumnen, deren Aufbau sich strukturell und funktionell an den Minicolumnen des primären visuellen Cortex V1 orientiert. Dieses columnare Array analysiert parallel

die gesamte WS-Aktivierungskarte.
Jede Verarbeitungscolumne extrahiert dazu elementare Bildmerkmale aus ihrem lokalen Analysegebiet innerhalb der WS-Aktivierungskarte mit Hilfe komplexer rezeptiver Felder, wie sie erstmals in der V1 auftreten. Diese werden jedoch nicht algorithmisch vorgegeben, sondern in einem adaptiven Selbstorganisationsprozeß strukturiert.

Dazu wurde zunächst eine von [Oja 1982] vorgestellte und später von [Sanger 1989] modifizierte Hebbsche Lernregel untersucht. Dabei wurde bestätigt, daß mit dieser Lernregel die Eigenvektoren zu den größten Eigenwerten der Korrelationsmatrix mittelwertfreier Inputvektoren gelernt werden können. Ein speziell optimierter Lernratenverlauf ermöglicht eine erhebliche Beschleunigung des Lernprozesses. Die auf der Grundlage der WS-Aktivierungskarten gelernten Wichtungsvektoren werden als rezeptive Felder interpretiert.

Der Vorteil dieses Lernverfahrens gegenüber einer analytischen Lösung liegt in der online realisierbaren Belehrung, die keine Berechnung einer Korrelationsmatrix und der dazu erforderlichen Kenntnis der gesamten Lernstichprobe an Inputvektoren erfordert. Da der Input keinen Einschränkungen unterliegt und kein Vorwissen über dessen Struktur erforderlich ist, liegt die Vermutung nahe, daß mit diesem Verfahren nicht nur rezeptive Felder des visuellen Cortex, sondern beliebige rezeptive Felder anderer cortikaler Areale modelliert werden können.

Die Selbstorganisation der eigenwertgrößten Hauptkomponenten für zufällig gewählte lokale Ausschnitte der WS-Aktivierungskarte führt auch für sehr unterschiedliche natürliche Farbszenen zu qualitativ ähnlichen Sätzen rezeptiver Felder. Auf der Basis dieser rezeptiven Felder führt die erste Stufe der Modellarchitektur eine lokale Hauptkomponentenanalyse (innerhalb einer jeden Verarbeitungscolumne) aus, die eine lokale Beschreibung des zugehörigen lokalen Analysegebietes mit minimaler Anzahl beschreibender Parameter in Form von Fitwertaktivierungen liefert.
Alle Fitwerte einer Verarbeitungscolumne werden in einer zweiten Stufe in einer speziellen Weise zu einer einzigen columnaren Aktivierung zusammengefaßt, die ein Maß für die texturelle Rauhigkeit innerhalb des korrespondierenden Analysegebietes darstellt. Die Aktivierungen aller Columnen werden wiederum in einer bezüglich der Retina topographisch organisierten Merkmalskarte zur Beschreibung der texturellen Rauhigkeit natürlicher Farbszenen zusammengefaßt. Diese bildet neben den aus der Farbumstimmung hervorgegangenen RG-, BY- und WS-Aktivierungskarten eine weitere Merkmalskarte.
Als Ort einer neuroanatomischen Repräsentation dieser Merkmalskarte als auch der gesamten Modellarchitektur zur fovealen Texturanalyse wird im Modellansatz die V1 vermutet.

4.2 Primäre Farbverarbeitung im visuellen System und im Modellansatz

Stäbchen gehen bei Tageslicht in die Sättigung. Insofern ist Tagessehen gleichbedeutend mit ausschließlichem Zapfensehen. Die Grundlage der folgenden Betrachtungen bildet ein 1993 von [De Valois et al. 1993] vorgestelltes mehrstufiges Farbmodell. Unter den Gesichtspunkten,

- aktuelles Wissen über die Anatomie der Retina zu berücksichtigen,
- das Ausmaß zufälliger Verschaltungen der Nervenzellen, wie sie die Anatomie vermuten läßt, bei der Herausbildung eines sensiblen Farbsystems zu untersuchen sowie
- einen bezüglich älterer Modelle ausgewogeneren Input für das RG- bzw. BY-Farbsystem bereitzustellen,

ist gerade dieses Modell von hohem Interesse.

De Valois entwickelte einen bottom-up-Ansatz, indem er verschiedene mögliche Verbindungen vorgab, resultierende Ausgangsaktivierungen bestimmte und die Ergebnisse mit aktuellen physiologischen und psychophysischen Erkenntnissen über die visuelle Wahrnehmung des Menschen verglich. Auch die extrem geringe Anzahl der S-Zapfen innerhalb der Retina findet in seinem Modell ihre Berücksichtigung. Bisherige Standardmodelle besitzen ein Farbsystem, das RG-System, das auf den Aktivierungen der L- und M-Zapfen und somit auf 90 – 95 % der gesamten Zapfenpopulation beruht, während das BY-System auf den verbleibenden 5 – 10 % der Zapfen, den S-Zapfen, begründet ist. Solch ein Ungleichgewicht erscheint unrealistisch. Das Modell beinhaltet somit auch den Versuch der Bestimmung eines ausgewogeneren Inputs für das RG- und das BY-System.
Die anhand von Simulationen vorgenommene Untersuchung des Modells nach De Valois führte dennoch zu erheblichen qualitativen und quantitativen Modifikationen. Das daraus abgeleitete Modell wird im folgenden vorgestellt (Abbildung 4.1).
Den Ausgangspunkt und damit die erste Stufe der Modellarchitektur zur Farbumstimmung bilden die drei Zapfentypen der menschlichen Retina mit ihren spektralen Absorptionskurven. Der Kurvenverlauf wird als evolutionäre Anpassung an die Strahlungsverteilung des mittleren Tageslichts interpretiert. Das Verhältnis der drei Zapfentypen L : M : S wird innerhalb der Fovea als 6 : 6 : 1 zugrundegelegt.
Die zweite Stufe der Farbverarbeitung findet in der fovealen Retina statt und führt zur Entwicklung spektral opponenter und nicht opponenter Zellen. Es ist zu vermuten, daß die Ausgangsaktivierungen der sechs parvozellulären Midget-Bipolarzelltypen L_o, M_o, S_o und $-L_o$, $-M_o$, $-S_o$ der zweiten Stufe äquivalent mit der Aktivierung der parvozellulären Ganglienzellen und somit dem Output der fovealen Retina sind.
Diese Aktivierungen werden in einer dritten Stufe, die man sich im Parvo-CGL lokalisiert vorstellen kann, in verschiedenen Kombinationen aufsummiert.

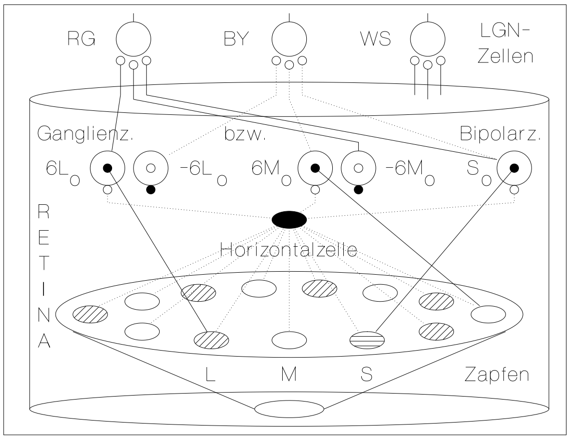


Abbildung 4.1: *Die dem Modell zur primären fovealen Farbverarbeitung zugrundeliegende Verschaltung von Parvo-Zellen der Retina und des CGL.*

Nach De Valois summiert $L_o - M_o$ Farbe und hebt Helligkeit auf, $L_o + M_o$ hingegen summiert Helligkeit und hebt Farbe auf. Das gleiche gilt offensichtlich für $M_o - L_o$ und $L_o + M_o$. Dieser Vorschlag ist der Vermutung von [Lennie et al. 1988] gleich, wie man Farb- und Helligkeitsinformationen von Parvo-LGN-Zellantworten trennen könnte.
Die Verschaltung von L_o- und M_o-Zellen in der Form $RG = 6\ L_o - 6\ M_o = 78\ L - 78\ M$ und $BY = 6\ M_o - 6\ L_o = 78\ M - 78\ L$ führt jedoch bei Darstellung aller möglichen Aktivierungen RG (Rot/Grün) und BY (Blau/Gelb) in einem karthesischen Koordinatensystem zur Abbildung aller denkbaren Farbreize auf ein und dieselbe Achse (Abbildung 4.2 (rechts)).

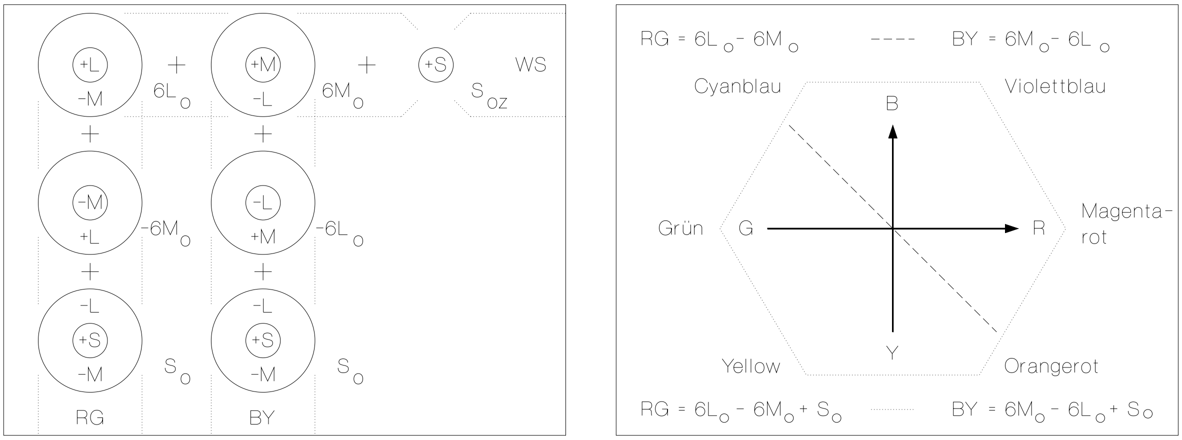


Abbildung 4.2: *Die schematische Verschaltung von Midget-Ganglienzellen der zweiten Stufe zur Erzeugung von RG-, BY- und WS-Aktivierungen der Parvo-CGL-Zellen der dritten Stufe (links) sowie die damit verbundene Herausbildung einer durch die Aktivierungsachsen RG und BY aufgespannten Farbfläche (rechts).*

Im Mittelpunkt der folgenden Betrachtungen steht die Aufspaltung dieser Farbachse. De Valois postuliert, daß die Aktivierung der S_o-Opponent-Zellen zu der der $(M_o - L_o)$-

sowie der ($L_o - M_o$)-Zellen addiert wird (Abbildung 4.2 (links)).
Demnach kommt es zu einer Überführung der Farbachse in eine Farbfläche und somit zur Abbildung aller Farbreize auf diese, durch die Aktivierungsachsen RG und BY aufgespannte Fläche (Abbildung 4.2 (rechts)).
Auch hier wird die bereits in den vorherigen Stufen eingeführte Wichtung der beteiligten Zellen von 6 L_o, 6 M_o und S_o beibehalten. Folglich gilt $RG = 6\ L_o - 6\ M_o + S_o$ und $BY = 6\ M_o - 6\ L_o + S_o$.
Prinzipiell existiert also lediglich eine Farbachse im Parvo-LGN-Path. Diese verläuft in etwa entlang der Farbempfindungen Cyanblau/Orangerot (Abbildung 4.2 (rechts)). Die Rolle der S_o-Opponent-Zellen besteht nun darin, diese eine dominierende Achse durch Modulation der LM-Opponent-Zellen in eine Farbfläche zu überführen.

Entgegen älteren Modellvorstellungen bilden die L_o- und M_o-Zellen in diesem Konzept den Hauptbeitrag sowohl zum RG- als auch zum BY-Farbsystem. Die schematische Verschaltung dieser Zellen ist in Abbildung 4.2 links dargestellt. Es ist offensichtlich, daß die unterschiedliche Verschaltung von Zellen der zweiten Stufe zu einer Trennung von Farbe und Helligkeit sowie zur Herausbildung orthogonaler RG- und BY-Farbachsen eines physiologischen Farbraumes führt.
Anhaltspunkte bezüglich einer über die qualitative Verschaltung hinausgehenden quantitativen Verkopplung der die WS-Aktivierung bildenden Zellen konnten bislang nicht gefunden werden. Eine denkbare, schematische Verschaltung, die zu einer bezüglich der RG- und BY-Achse orthogonalen Farbachse führt und daher von herausragender funktioneller Bedeutung für die Gesamttransformation wäre, zeigt ebenfalls Abbildung 4.2 (links).
Da auf jeder Stufe des Modells lediglich lineare Summationen vorgenommen wurden, lassen sich alle drei in einer einfachen Lineartransformation zusammenfassen (Tabelle 4.1).

Aktivierung der Midget-Bipolar- bzw. Midget-Ganglien-Zellen der Stufe 2										
			L_o	$=$	$13L$	$-(6L+6M+S)$				
			M_o	$=$	$13M$	$-(6L+6M+S)$				
			S_o	$=$	$13S$	$-(6L+6M+S)$				
			S_{oz}	$=$	$13S$					
Aktivierung der parvozellulären CGL-Opponent-Zellen der Stufe 3										
RG						BY				
$6L_o$	$=$	$42L$	$-36M$	$-6S$		$6M_o$	$=$	$-36L$	$+42M$	$-6S$
$-6M_o$	$=$	$36L$	$-42M$	$+6S$		$-6L_o$	$=$	$-42L$	$+36M$	$+6S$
S_o	$=$	$-6L$	$-6M$	$+12S$		S_o	$=$	$-6L$	$-6M$	$+12S$
RG	$=$	$72L$	$-84M$	$+12S$		BY	$=$	$-84L$	$+72M$	$+12S$

$$WS = 28{,}77L_o + 15{,}23M_o + 4{,}84S_{oz} = 110L - 66M + 19S$$

Tabelle 4.1: *Die zusammengefaßte Modelltransformation.*

4.3 Repräsentation retinotoper farbbasierter Merkmalskarten

Die mit dem vorgestellten Modellansatz extrahierbaren RG-, BY- und WS-Aktivierungen des fovealen Netzhautbildes lassen sich in drei separaten Merkmalskarten zusammenfassen. Die Abbildung 4.3 zeigt die entstehenden Aktivierungskarten bei fovealer Betrachtung eines auf der Spitze stehenden Farbwürfels und eines Graukeils (synthetische Referenzszene) unter Beleuchtungsverhältnissen, die denen des mittleren Tageslichts entsprechen.

Offensichtlich ist es dem visuellen System bereits auf dieser Verarbeitungsstufe gelungen, eine Dekorrelation der hochgradig korrelierten spektralen Empfindlichkeiten der Zapfen zu realisieren. Die unbunte WS-Aktivierungskarte als auch die bunte RG- und BY-Aktivierungskarte scheinen eine gleichberechtigte Rolle zu spielen. Psychophysische Erkenntnisse deuten darauf hin, daß sich im fovealen Bereich Luminanzkontrastdetektion und chromatische Kontrastdetektion ([Gouras et al. 1979],[Switkes et al. 1988]) gegenseitig unterdrücken können. Vautin und Dow [Vautin et al. 1985] berichteten 1985 über die Existenz von vier grundlegenden Farbzelltypen im primären visuellen Cortex, deren spektrale Empfindlichkeitskurven in etwa den vier psychophysischen Farbempfindungen Rot, Gelb, Grün und Blau entsprechen. Das würde wiederum für eine Aufspaltung der RG- als auch der BY-Aktivierungskarte in eine R-, eine G- sowie weitere B- und Y-Aktivierungskarten sprechen.

Somit existieren starke Anhaltspunkte für eine solche Lokalisation innerhalb des primären visuellen Cortex, wobei diese Aktivierungskarten lediglich drei dem Farbensehen gewidmete Karten unter vielen weiteren, auf andere Submodalitäten der Wahrnehmung spezialisierten, topographischen Karten, darstellen.
Die anatomische Struktur zur Repräsentation derartiger Karten können die Blobs der V1 bilden (siehe Kapitel 2).

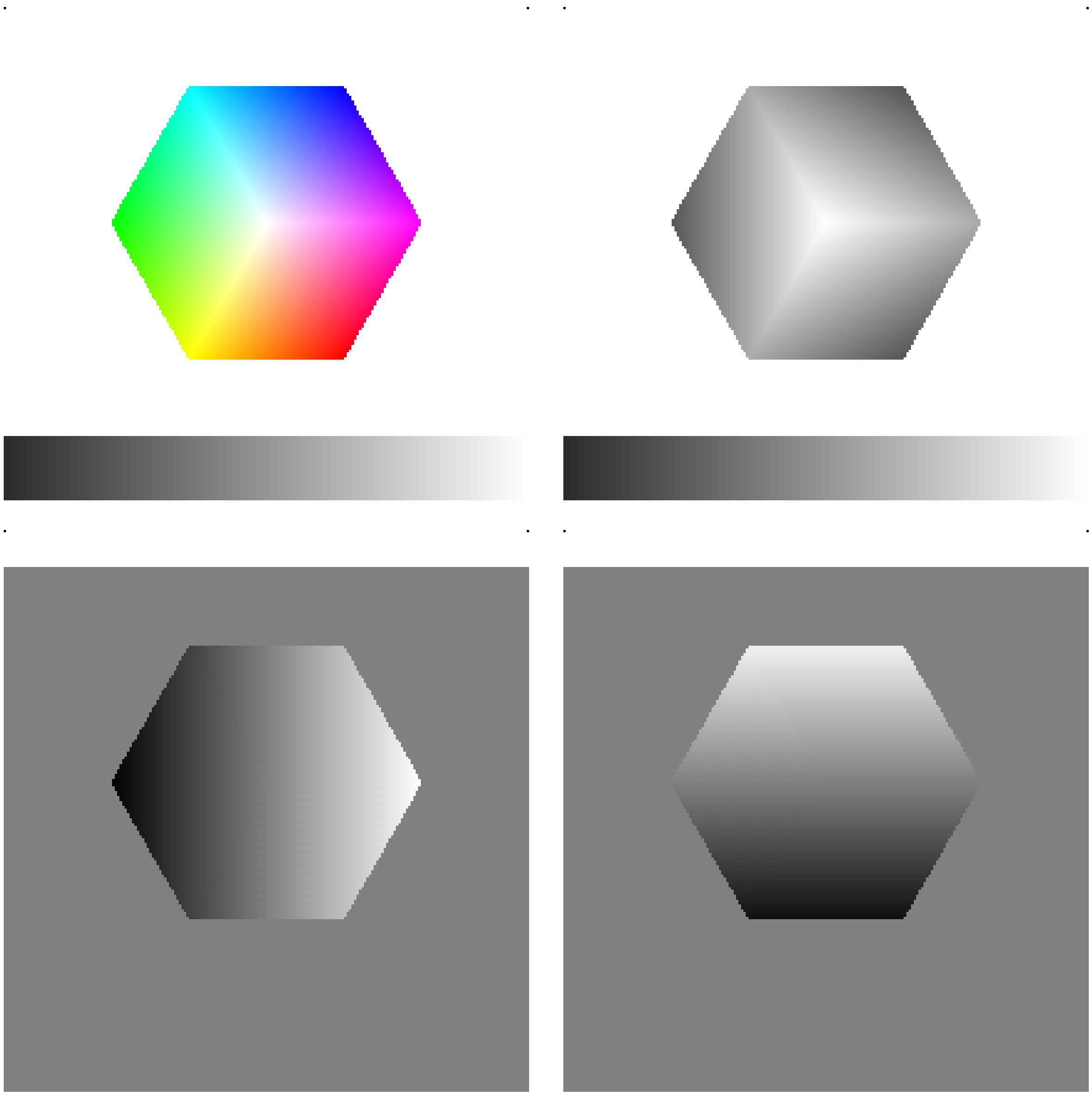


Abbildung 4.3: *Die bei fovealer Betrachtung der synthetischen Referenzfarbszene (links oben) durch Verschaltung von Parvo-Zellen der Retina und des CGL in der V1 entstehenden WS- (rechts oben), RG- (links unten) und BY-Aktivierungskarten (rechts unten). Da der Graukeil lediglich eine WS-Aktivierung hervorruft, findet er in den bunten Aktivierungskarten RG und BY keine Berücksichtigung.*

4.4 Der physiologisch motivierte Farbraum

4.4.1 Kodierung von Farbempfindungen durch neuronale Aktivierungen

Im Mittelpunkt der folgenden Betrachtungen steht die Untersuchung der in 4.2 vorgestellten Transformation. Der entstehende Farbraum wird, wie bereits 1905 von Hering [Hering 1905] vermutet, durch die orthogonalen Achsen Rot/Grün (RG), Blau/Gelb (BY) sowie Weiß/Schwarz (WS) aufgespannt und umfaßt alle möglichen Farbempfindungen des Menschen.
Von Interesse sind die Ausdehnung des entstehenden Farbraumes sowie die Lokalisation der möglichen Farbempfindungen. Dazu bedarf es einer Transformation aller denkbaren LMS-Reizkonfigurationen der topographisch angeordneten Zapfen in diesen bezüglich der Retina nichttopographisch organisierten Farbraum. Jeder transformierte Reiz läßt sich an der durch RG-, BY- und WS-Aktivierung bestimmten Position im Farbraum darstellen.
Für die Untersuchung wurde ein Farbmonitor mit einer Phosphorbeschichtung entsprechend der amerikanischen FCC-Norm genutzt. Die spektralen Empfindlichkeitskurven der FCC-Phosphore sind in der Abbildung 4.4 links dargestellt.

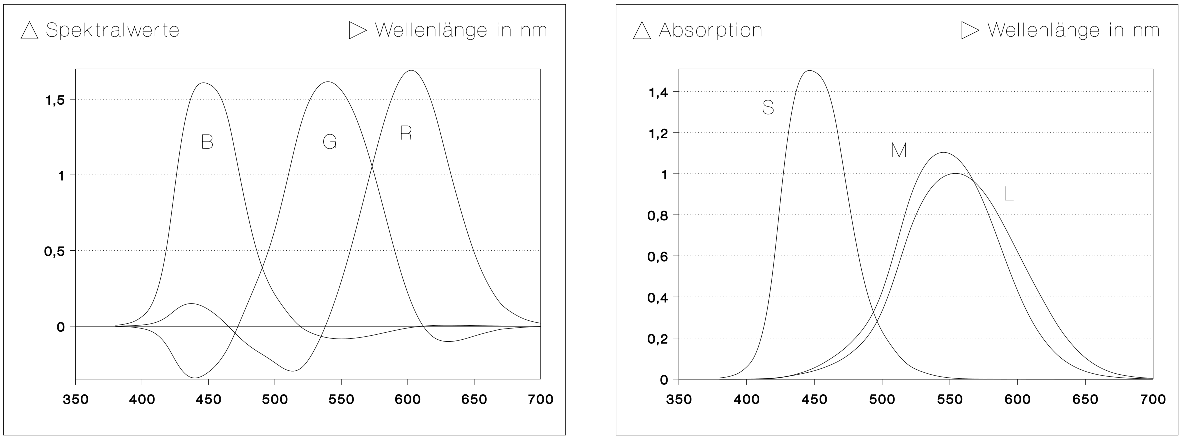


Abbildung 4.4: *RGB-Spektralwertkurvenzug nach FCC-Norm (links) sowie die LMS-Modell-Spektralwertkurven der menschlichen Zapfen (rechts) (leicht modifiziert nach Judd 1945).*

Auf der Basis der in Abbildung 4.4 dargestellten RGB- und LMS-Spektralwertkurven wurde folgende Transformationsmatrix für die lineare Abbildung von RGB-Werten in LMS-Werte mathematisch bestimmt.

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix}_{Modell} = \begin{pmatrix} 0,2915 & 0,5909 & 0,1176 \\ 0,1677 & 0,6803 & 0,1520 \\ 0,0 & 0,0549 & 0,9451 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}_{FCC}$$

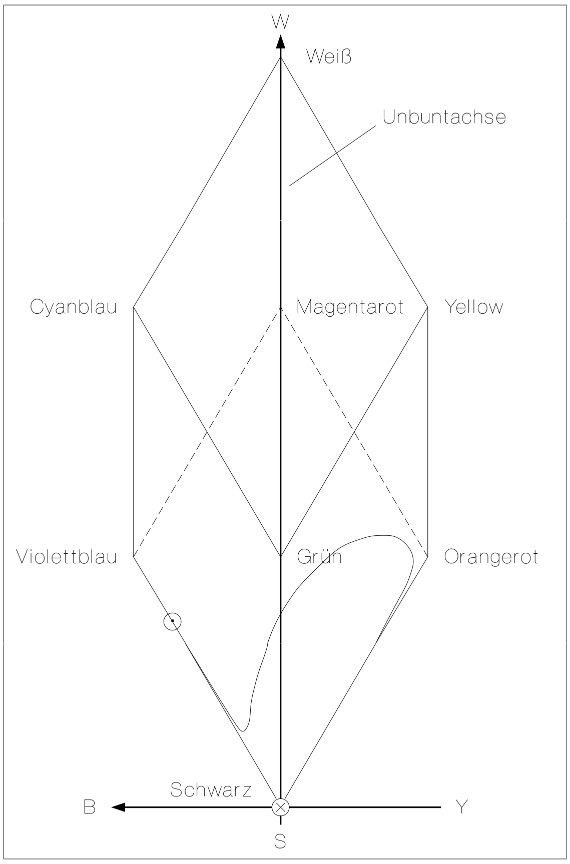
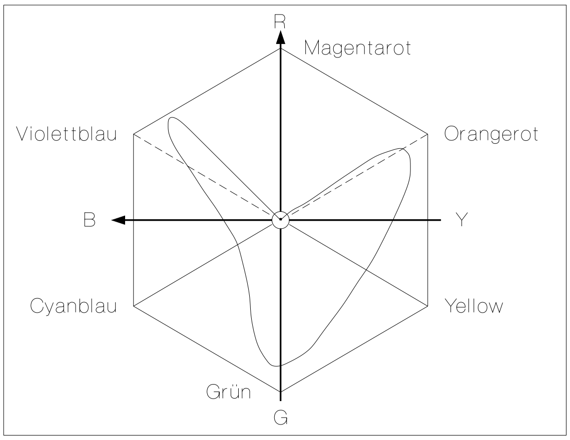
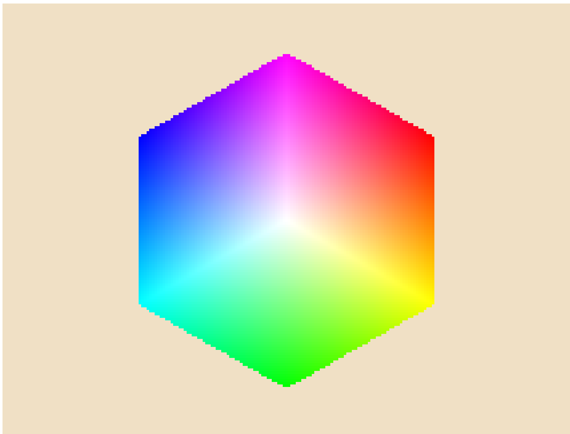


Abbildung 4.5: *Der entstehende physiologische Farbraum als Seitenansicht und Draufsicht. Er wird durch die drei orthogonalen Achsen RG, BY und WS aufgespannt. Jeder durch ein entsprechendes Aktivierungstripel RG, BY, WS beschriebene Punkt innerhalb dieses Farbraumes entspricht einer Farbempfindung. Der Kurvenzug innerhalb des Farbraumes verdeutlicht den Aktivierungsverlauf bei Anwendung der Transformation auf die spektralen Empfindlichkeitskurven des Menschen (Abbildung 4.4 (rechts)), bei Anwendung der direkten Transformation auf den Spektralwertkurvenzug nach FCC-Norm (Abbildung 4.4 (links)) bzw. durch Darstellung der Aktivierungstripel jeder Wellenlänge aus Abbildung 4.6. Dies entspricht einem Aktivierungsverlauf, wie er z.B. durch diskrete monochromatische Farbreize bei Betrachtung eines Linienspektrums hervorgerufen wird.*





Es erfolgt praktisch eine Transformation der spektralen Empfindlichkeiten der technischen CCD-Sensoren einer RGB-Farbkamera in die physiologischen spektralen Empfindlichkeiten der visuellen LMS-Sensoren des Menschen. Die Verschaltung der Parvo-Zellen innerhalb der Retina und des CGL führt zur Herausbildung einer RG-, BY- und WS-Aktivierung (Abbildung 4.6 (links)) entsprechend der in 4.2 hergeleiteten Transformation:

$$\begin{pmatrix} RG \\ BY \\ WS \end{pmatrix}_{Farbraum} = \begin{pmatrix} 72 & -84 & 12 \\ -84 & 72 & 12 \\ 110 & -66 & 19 \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix}_{Modell}$$

Alle möglichen RG-, BY- und WS-Aktivierungen bilden den in Abbildung 4.5 dargestellten Farbraum. Die Geometrie des physiologischen Farbraumes zeigt, daß das visuelle System zu acht extremen Farbempfindungen fähig ist. Diese sind Schwarz, Violettblau, Grün, Orangerot, Cyanblau, Gelb (Yellow), Magentarot sowie Weiß.
Aufgrund der entstehenden Geometrie dieses Raumes läßt sich eine lineare Transformation finden, die eine direkte Überführung der RGB-Spektralwerte für FCC-Phosphore

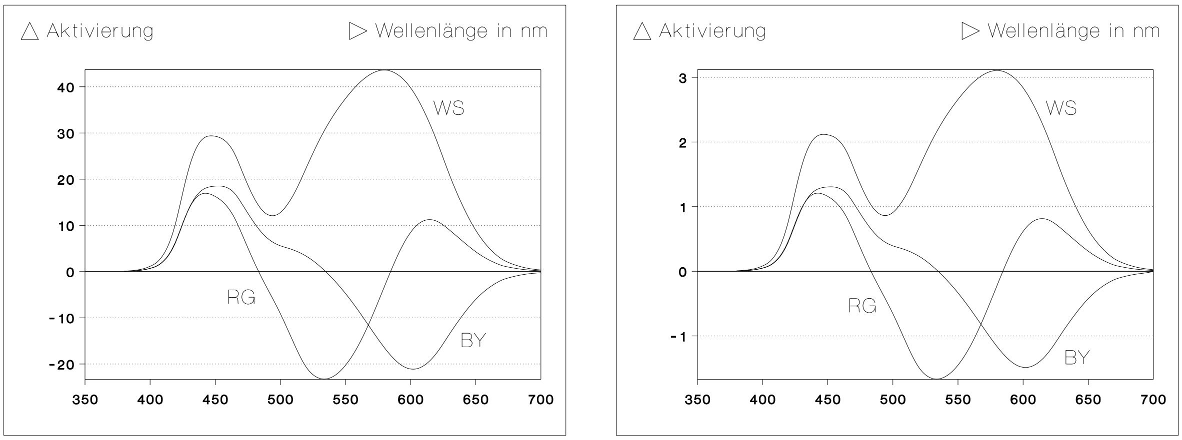


Abbildung 4.6: *Gegenüberstellung der auf die spektralen Empfindlichkeitskurven des Menschen (Abbildung 4.4 (rechts)) angewendeten Transformation und der auf den Spektralwertkurvenzug nach FCC-Norm (Abbildung 4.4 (links)) angewendeten direkten Transformation. Die resultierenden Aktivierungsverläufe der Parvo-CGL-Zellen sind äquivalent.*

in eine RG-, BY- sowie WS-Aktivierung (Abbildung 4.6 (rechts)) und damit eine Abbildung in den physiologischen Farbraum erlaubt:

$$\begin{pmatrix} RG \\ BY \\ WS \end{pmatrix}_{Farbraum} = \begin{pmatrix} 0,500 & -1,0 & 0,500 \\ -0,875 & 0,0 & 0,875 \\ 1,500 & 1,5 & 1,500 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}_{FCC}$$

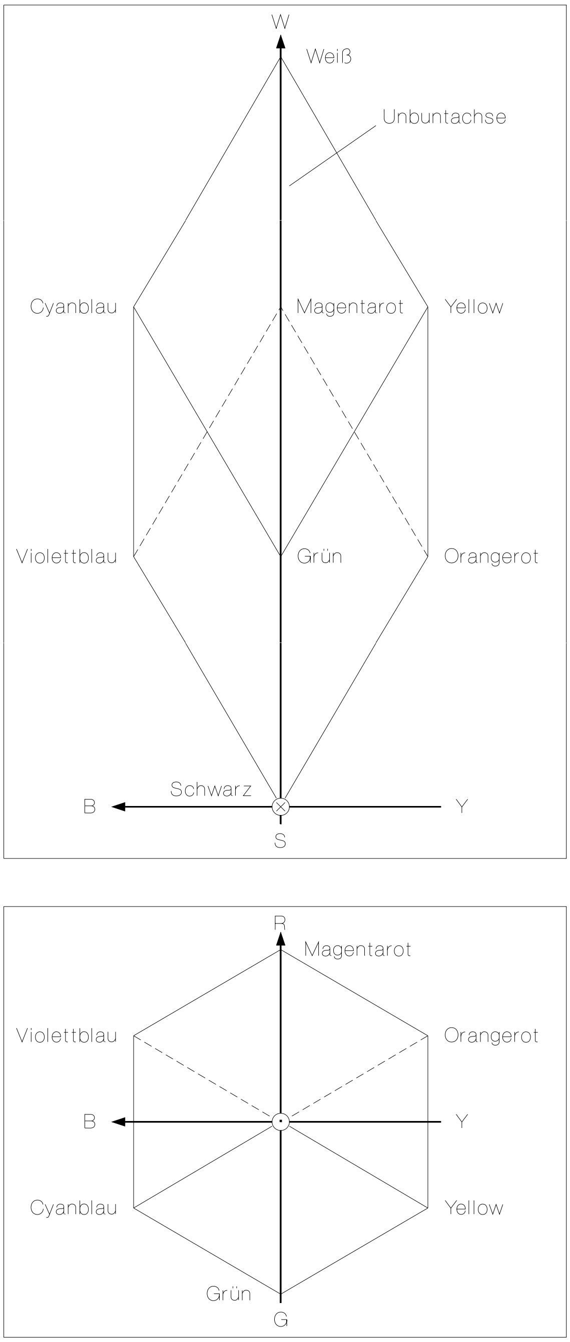
Die inverse Transformation lautet:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix}_{FCC} = \begin{pmatrix} 0,2222 & 0,5714 & 0,3333 \\ 0,2222 & 0,0000 & -0,6667 \\ 0,2222 & -0,5714 & 0,3333 \end{pmatrix} \begin{pmatrix} RG \\ BY \\ WS \end{pmatrix}_{Farbraum}$$

Erstaunlich ist die Tatsache, daß bei Betrachtung des sichtbaren Spektrums die dreidimensionale Darstellung des resultierenden Aktivierungsverlaufs der Parvo-CGL-Zellen (Abbildung 4.6) innerhalb des physiologischen Farbraumes (Abbildung 4.5) lediglich dessen unteres Drittel ausfüllt. Bis auf Magentarot, das im sichtbaren Spektrum auch nicht enthalten ist, kommt es bezüglich der RG- sowie BY-Achse (Abbildung 4.5 (unten)) zu einer Aktivierung aller im Spektrum auftretenden Farbempfindungen, wobei der Aktivierungsverlauf im Koordinatenursprung geschlossen wird. Die extremen Farbempfindungen am violettblauen und orangeroten Ende des sichtbaren Spektrums werden somit durch die unbunte Farbempfindung Schwarz und nicht durch die willkürliche Einfügung der dort nicht auftretenden Farbempfindung Magentarot verbunden.
Dennoch scheint die bloße Abbildung diskreter monochromatischer Farbreize, wie sie z.B. bei Betrachtung eines Linienspektrums auftreten, in den physiologischen Farbraum bezüglich der WS-Aktivierung augenscheinlich nicht zu gängigen Farbempfindungen eines farbnormalsichtigen Beobachters zu führen. Daß es dennoch zur Herausbildung dieser Farbempfindungen kommt, läßt sich nur mit dem Prozeß einer Dynamikanpassung aller durch Betrachtung einer Szene hervorgerufenen WS-Aktivierungen erklären, der ausführlich Gegenstand eines der folgenden Abschnitte ist.
Nur eine solche Dynamikanpassung würde dazu führen, daß der Aktivierungskurvenzug der Parvo-CGL-Zellen (Abbildung 4.6) bei Betrachtung des sichtbaren Spektrums den gesamten physiologischen Farbraum (Abbildung 4.5) ausnutzt und somit eine Aktivierung der gewohnten Farbempfindungen hervorruft.

4.4.2 Korrespondenzen zum Farbraum Küppers'

Der durch Analyse der Verschaltung der Parvo-Zellen der Retina sowie des CGL entstandene physiologisch motivierte Farbraum befindet sich in exzellenter Übereinstimmung mit 1973 veröffentlichten Überlegungen von Küppers [Kueppers 1973]. Dieser kam in einem kritischen Vergleich verschiedener Farbordnungssysteme sowie Farbraummodelle zu dem Schluß, daß all diese Lösungsvorschläge nur Teilwahrheiten sein konnten, hinter denen eine übergeordnete Gesetzmäßigkeit vorhanden sein müsse.
Nach seinen Vorstellungen sollte jede der sechs extremen, bunten Farbempfindungen einerseits einen besonderen Platz in einem Farbraum bekommen, andererseits aber zueinander völlig gleichrangig und gleichwertig angeordnet sein, was in einem geometrischen System nur bedeuten konnte, daß sie zueinander absolut symmetrische Abstände haben müßten. Diese Überlegungen führten zu der Erkenntnis, daß die drei Dimensionen im Farbraum durch die drei Buntachsen repräsentiert werden müßten, was so zu geschehen hätte, daß diese sich im Punkt der mittleren Farbempfindung Grau mit Winkelstellungen von 90° zueinander schneiden. Geometrisch entstand dabei der von Küppers entwickelte Oktaeder-Farbraum (Abbildung 4.7 (rechts)). Seine Außenflächen bilden acht gleichseitige Dreiecke.
Bei sorgfältiger Prüfung des Oktaeders stellte Küppers fest, daß es sich dabei lediglich um das mittlere Teilstück eines neuen Farbraumes handelte, da die Unbuntachse nach oben und nach unten weiterlief. Indem die fehlenden Teile, nämlich der Weiß- und der Schwarz-Tetraeder von ihm ergänzt wurden, war sein Rhomboeder-Farbraum gefunden (Abbildung 4.7 (rechts)).



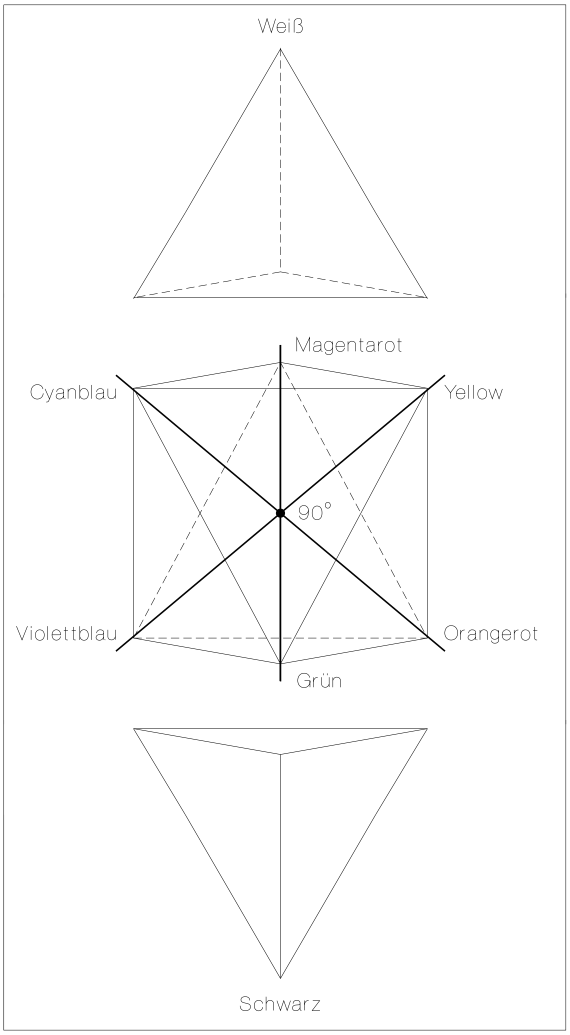


Abbildung 4.7: *Der durch sämtliche Aktivierungen der Parvo-CGL-Zellen aufgespannte physiologisch motivierte Farbraum (links). Der Rhomboeder-Farbraum nach Küppers (rechts).*

Küppers gewann im Laufe der folgenden Jahre durch farbmetrische Untersuchungen die Überzeugung, daß es sich bei diesem Farbraum um die endgültige Lösung des Ordnungsproblems der Farben handelt. Als Küppers 1973 die Lösung des Rhomboeder-Systems veröffentlichte, nahm die Wissenschaft davon kaum Notiz. Die These, daß es insgesamt acht extreme Farbempfindungen gäbe und daß die sechs extremen, bunten Farbempfindungen gleichwertig und gleichberechtigt seien, wurde von etablierten Farbmetrikern abgelehnt, weil sie nicht mit der Philosophie des CIE-Systems in Einklang zu bringen war.

Erst durch die Berücksichtigung der experimentell nachgewiesenen neurobiologischen Verschaltungsprinzipien und der daraus abgeleiteten mathematischen Abbildungsvorschrift in den physiologischen Farbraum (Abbildung 4.7 (links)) finden diese theoretischen Überlegungen ihre Rechtfertigung, und es wird ein direkter Zugang zur algorithmischen Umsetzung geschaffen. Gleichzeitig wird damit die Problematik der Farbfehlsichtigkeit besser interpretierbar, worauf im Rahmen dieses Projektberichts jedoch nicht näher eingegangen wird.

Physiologische Erkenntnisse [Land et al. 1983a] deuten darauf hin, daß die früheste, denkbare Struktur zur Repräsentation eines solchen Farbraumes die V4 darstellt. Nur sie besitzt gleichzeitig nichttopographisch angeordnete (nichtretinotope), farbsensitive Zellen [Zeki 1980] und die erforderlichen Verbindungen beider Hirnhemisphären über das Corpus callosum. Somit wäre hier erstmals Information über das gesamte Gesichtsfeld verfügbar.

4.5 Die Farbumstimmung im physiologischen Farbraum

4.5.1 Mittleres Empfindungsniveau

Sind alle Wellenlängen des sichtbaren Spektrums mit gleicher Intensität vorhanden, so spricht man von einem energiegleichen Spektrum.

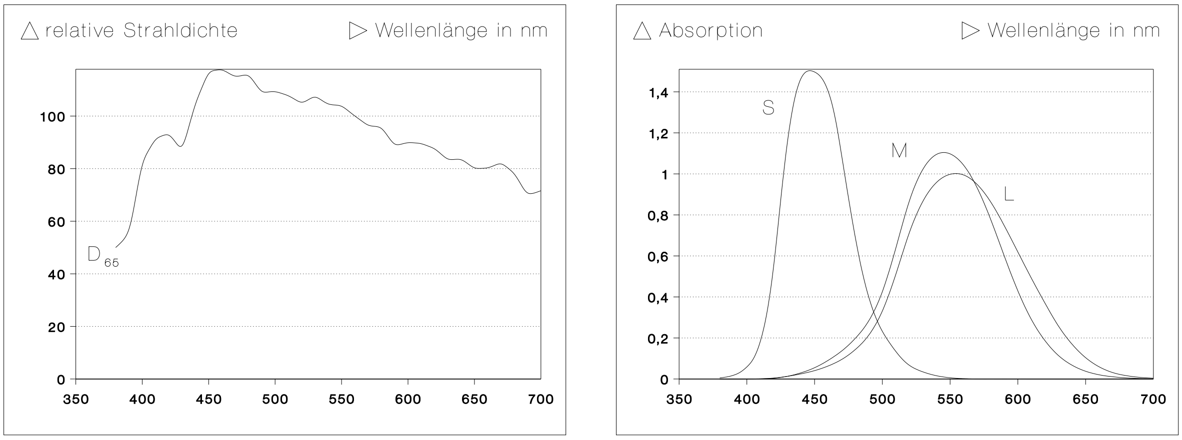


Abbildung 4.8: *Gegenüberstellung der Strahlungsverteilung des mittleren Tageslichts (links) sowie der bei der Modellbildung verwendeten Spektralwertkurven der Zapfen (rechts).*

Beim Tageslicht gibt es jedoch beträchtliche Schwankungen der spektralen Zusammensetzung. Sie verändert sich ununterbrochen im Verlauf eines Tages, mit der Jahreszeit und dem Breitengrad. Allgemein weist das Tageslicht unterschiedliche Strahlungsverteilungen mit zwischen 5.000 Kelvin und 10.000 Kelvin schwankenden Farbtemperaturen auf. Die Abbildung 4.8 zeigt links die relative Strahlungsverteilung des Tageslichts.
Nur die Strahlungsverteilung des mittleren Tageslichts führt gleichzeitig zu einer vollen, übersteuerungsfreien Ausnutzung der spektralen Empfindlichkeitskurven aller drei Zapfentypen (Abbildung 4.8 (rechts)) und somit nach Transformation der LMS-Zapfenaktivierungen entsprechend der Verschaltung der Zellen innerhalb der Retina und des CGL folgerichtig zu einer vollen Ausnutzung der Farbempfindungen des physiologischen Farbraumes in Form von RG-, BY- und WS-Aktivierungen. Nur unter diesen Beleuchtungsverhältnissen würde sich beim Menschen das mittlere Empfindungsniveau einstellen, das keiner Umstimmung bedarf. Alle anderen Beleuchtungsverhältnisse erfordern eine Umstimmung zur Erzielung äquivalenter Farbempfindungen bei Beobachtung ein und derselben Szene.

Für die technische Realisierung der folgenden Untersuchungen bilden jedoch Sensoren einer RGB-Farbbildkamera mit spektralen Empfindlichkeitskurven nach FCC-Norm (Abbildung 4.9 (links)) den Ausgangspunkt. Daher ist es erforderlich, die Strahlungsverteilung zu bestimmen, bei der diese ohne Übersteuerung voll ausgenutzt werden.

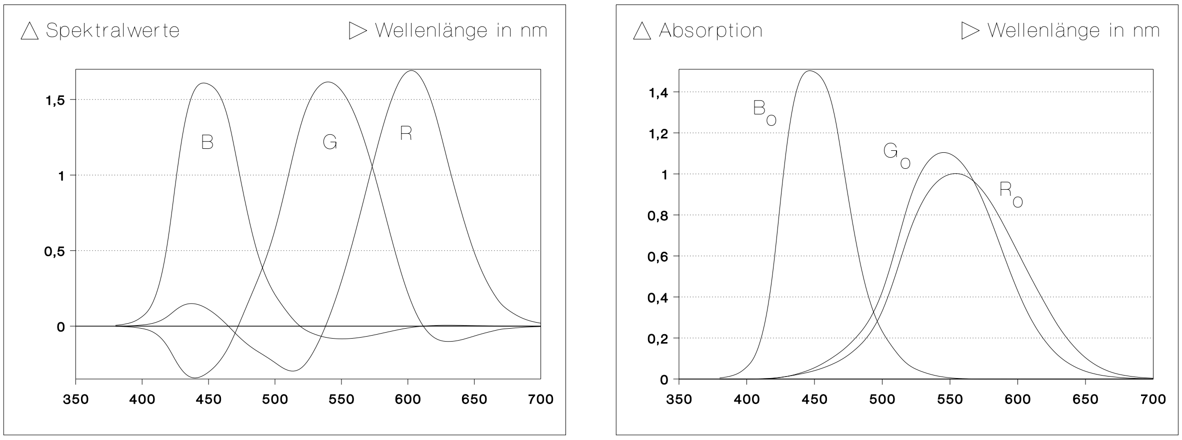


Abbildung 4.9: *Überführung der mit negativen Anteilen behafteten RGB-Spektralwertkurven nach FCC-Norm (links) in äquivalente $R_oG_oB_o$-Spektralwertkurven ohne negative Anteile (rechts).*

Das wird durch die negativen Anteile der Spektralwertkurven erschwert. Daher ist eine Umformung der spektralen Absorptionskurven nach FCC-Norm in eine gleichwertige Darstellung ohne negative Anteile sinnvoll.
Betrachtet man die Spektralwertkurven in Abbildung 4.9 links, so wird offensichtlich, daß der negative Anteil der Blaukurve im Bereich der positiven Anteile von Rot- und Grünkurve, die negativen Anteile der Grünkurve jeweils in den positiven Bereichen der Rot- bzw. Blaukurve und der negative Anteil der Rotkurve im positiven Bereich der Grün- und Blaukurve liegen. Somit lassen sich aus diesen Spektralwertkurven äquivalente Kurven ohne negative Anteile erzeugen, indem zu jeder Spektralwertkurve solange Anteile der beiden anderen addiert werden, bis die negativen Anteile verschwinden.
Es ist somit möglich, durch Linearkombination der Spektralwertkurven RGB zu neuen Spektralwertkurven $R_oG_oB_o$ zu gelangen, die keine negativen Werte mehr besitzen [Lang 1978]. Die resultierenden Kurvenzüge (Abbildung 4.9 (rechts)) verdeutlichen, daß auch diese Sensoren nur unter Beleuchtungsverhältnissen, die denen des mittleren Tageslichts entsprechen, voll ausgenutzt werden. Somit würde sich bei diesen Beleuchtungsverhältnissen auch unter Verwendung einer RGB-Farbbildkamera zur Bilderzeugung und der anschließenden direkten Transformation der RGB-Werte in den physiologischen Farbraum das mittlere Empfindungsniveau einstellen, das keiner Umstimmung bedarf.
Die spektralen Empfindlichkeiten der RGB-Sensoren nach FCC-Norm und die zur Modellbildung verwendeten Absorptionskurven der LMS-Zapfen sind somit äquivalent.
Weitere technische Anforderungen an die RGB-Farbbildkamera werden durch relevante Mechanismen des Auges festgelegt:

- in konstantem Verhältnis gehaltene, übersteuerungsfreie Verstärkung für R-, G- und B-Signal (begründet in der Adaptation des Auges)

- Ausschaltung der Gammakorrektur (lineares Sättigungsverhalten der LMS-Sensoren)

Eine erste Verarbeitungsstufe zur Untersuchung des mittleren Empfindungsniveaus besteht in der Transformation der ortsabhängigen, übersteuerungsfreien RGB-Werte natürlicher Farbszenen in den physiologischen Farbraum. Die in den Farbraum transformierten Pixel eines einzelnen RGB-Bildes bilden dort eine Punktwolke. Die Position eines jeden Punktes der Wolke im Farbraum ist durch seine RG-, BY- und WS-Aktivierung eindeutig beschrieben.
Eine ausgewogene, durchschnittliche Tagesbeleuchtung (Abbildung 4.8 (links)) führt zu einer Punktwolke, die senkrecht im Farbraum steht und diesen somit optimal ausnutzt (Abbildung 4.10 (rechts)). Eine Punktwolke innerhalb des physiologischen Farbraumes mit dieser Orientierung entspricht dem bisher nur vermuteten und nun plausibel geometrisch interpretierbaren mittleren Empfindungsniveau, das keiner Umstimmung bedarf. Alle unbunten Farbempfindungen befinden sich unter diesen Umständen auf der Unbuntachse.
Weitere Untersuchungen zeigten jedoch, daß es zu einer Auslenkung der Punktwolke entsprechend der bei der Bildaufnahme vorhandenen spektralen Zusammensetzung der Allgemeinbeleuchtung aus der WS-Achse (Unbuntachse) kommt (Abbildung 4.10 (links)). Die Auslenkung der Punktwolke ist für jede von der des mittleren Tageslichts abweichende spektrale Zusammensetzung der Allgemeinbeleuchtung charakteristisch und für natürliche Szenen mit üblicherweise einer Vielzahl unterschiedliche Strahlungsverteilungen reflektierenden, bunten Oberflächen unabhängig vom Bildinhalt. Somit erscheint eine Umstimmung im physiologischen Farbraum sinnvoll, da hier die Farbinformation des gesamten Gesichtsfeldes gleichberechtigt vorliegt. Auch spielt die Größe von Regionen identischer Farbreize im bezüglich der Retina nichttopographisch organisierten Farbraum keine Rolle mehr, da es hier zu einer Abbildung dieser Reize auf einen gemeinsamen Punkt kommt. Im Mittelpunkt der folgenden Betrachtungen sollen nun neurobiologisch motivierbare Mechanismen stehen, die das visuelle System in die Lage versetzen, fortlaufend räumliche und zeitliche Schwankungen in der Spektralverteilung der Beleuchtung so auszugleichen, daß zu betrachtende Gegenstände in Abstraktion von den momentanen Beleuchtungsverhältnissen nur auf der Grundlage ihrer optischen Charakteristika erfaßt werden können.

4.5.2 Neuronale Hauptkomponentenanalyse

Für jede Punktwolke läßt sich eine ausgezeichnete Achse bestimmen, die in die WS- oder Unbuntachse überführt werden kann (Abbildung 4.11), was, verbunden mit einer koordinatenabhängigen Reskalierung, zu einer referenzfreien, automatischen Farbumstimmung des gesamten Bildes im physiologischen Farbraum und letztlich zu einem Weißabgleich führt. Die Feststellung der ausgezeichneten Achse läßt sich durch ein bekanntes neuronales Netzwerk realisieren [Oja 1982] (siehe hierzu Abschnitt 4.6.2). Dieses bestimmt als Lageparameter der Punktwolke den eigenwertgrößten Eigenvektor $\mathbf{e}$ der Autokorrelationsmatrix $\mathbf{C} = E\{\mathbf{x}\mathbf{x}^T\}$ der Punkte der mittelwertfreien Wolke, d.h. die Hauptachse der Punktwolke in Form eines Wichtungsvektors $\mathbf{w}$ nach folgendem

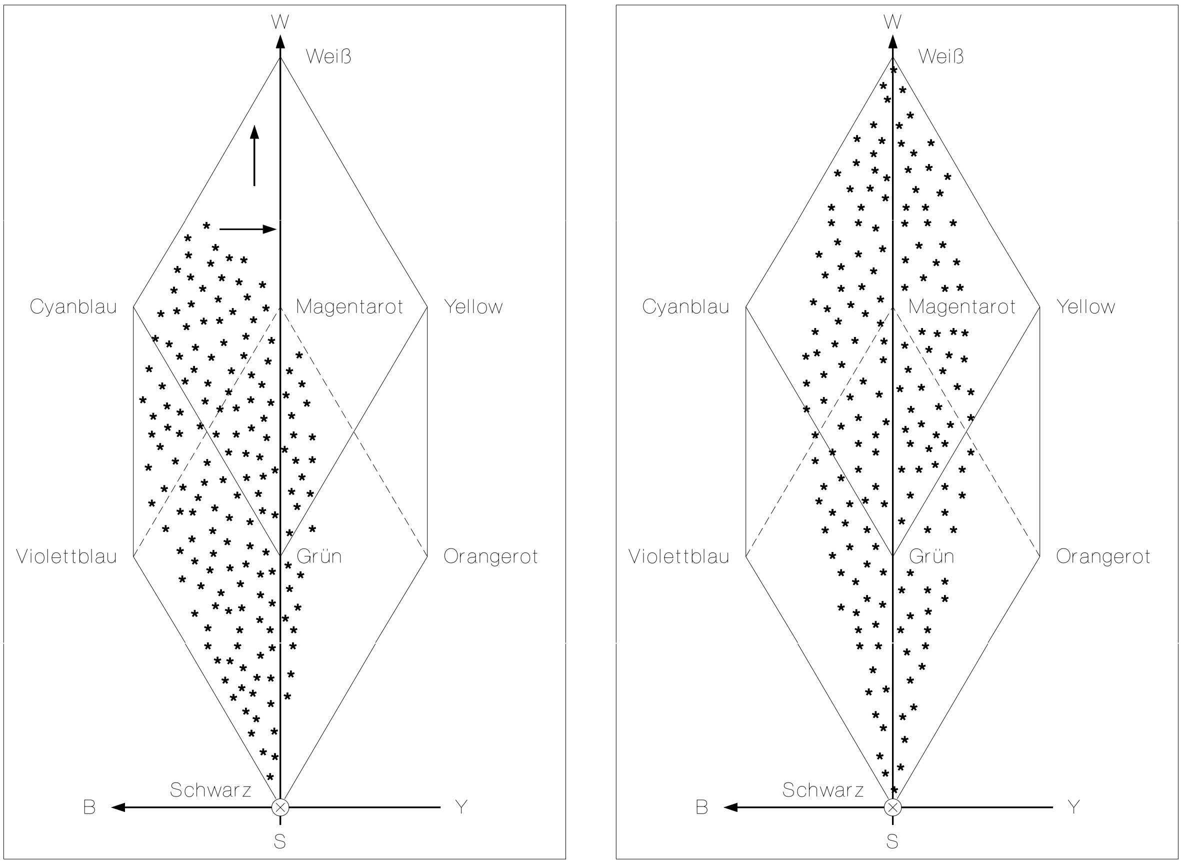


Abbildung 4.10: *Jedes in den Farbraum transformierte RGB-Bild bzw. die bei fovealer Wahrnehmung einer Farbszene in den Farbraum transformierten LMS-Zapfenaktivierungen führen dort zur Herausbildung einer Punktwolke (links). Diese ist entsprechend der spektralen Verteilung der vorherrschenden Allgemeinbeleuchtung aus der WS-Achse (Unbuntachse) ausgelenkt. Der Mechanismus der Farbumstimmung sollte es ermöglichen, das rechts dargestellte mittlere Empfindungsniveau, wie es ohne Farbumstimmung nur bei Beleuchtungsverhältnissen auftritt, die denen des mittleren Tageslichts entsprechen, unter allen anderen Beleuchtungsverhältnissen zu gewährleisten.*

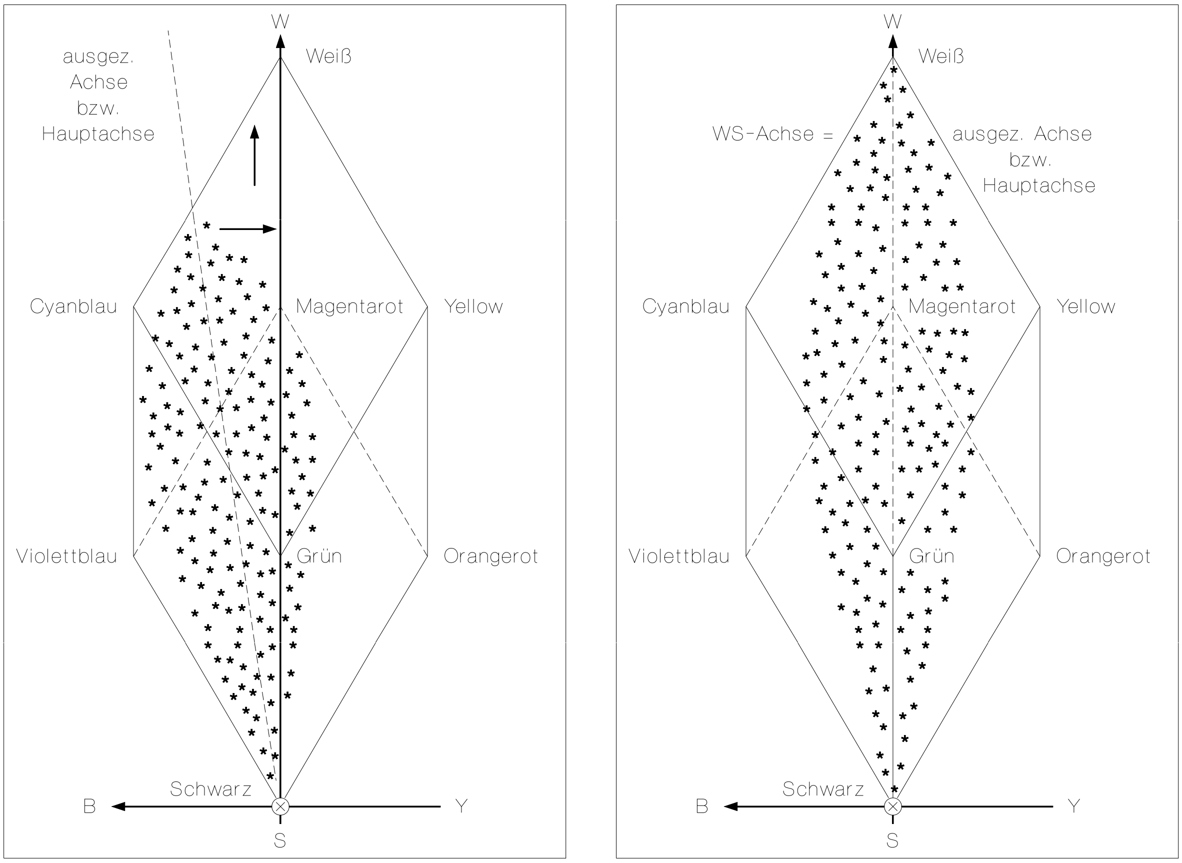


Abbildung 4.11: *Für jede Punktwolke läßt sich eine ausgezeichnete Achse bestimmen, die in die WS- oder Unbuntachse überführt werden kann, was, verbunden mit einer koordinatenabhängigen Reskalierung (Dynamikanpassung), zu einer referenzfreien, automatischen Farbumstimmung des gesamten Bildes im physiologischen Farbraum und letztlich zu einem Weißabgleich führt.*

Lernalgorithmus für den Gewichtsvektor.

$$\mathbf{w} := \mathbf{w} + \gamma\ y\ (\mathbf{x} - y\mathbf{w}) \qquad \text{mit} \qquad y = \mathbf{x}^T\mathbf{w}$$

Diese Lerngleichung läßt sich bezüglich des Wichtungsvektors $\mathbf{w}$ wie folgt in ihre Komponenten aufschlüsseln.

$$w_j(t+1) = w_j(t) + \gamma(t)\ y(t)\ \{x_j(t) - y(t)\ w_j(t)\} \qquad \text{mit} \qquad j = rg, by, ws$$

Hierbei bezeichnet $\mathbf{x} = (x_{rg}, x_{by}, x_{ws})^T$ den die Lage eines Punktes im physiologischen Farbraum beschreibenden und auf das Intervall $< -1, 1 >$ normierten Eingangsvektor, y das Ausgangssignal, w_j die Wichtung des Eingangsknotens x_j und γ die Lernrate.
Die Kenntnis der Hauptachse (Abbildung 4.11) in Form des extrahierten Wichtungsvektors $\mathbf{w} = (w_{rg}, w_{by}, w_{ws})^T$ erlaubt die Optimierung der Lage eines jeden Punktes bezüglich der RG-, BY- und WS-Achse des physiologischen Farbraumes.
Eine mit der WS-Aktivierung gewichtete Korrektur der RG- und BY-Aktivierungen aller Punkte der Wolke führt zu einer Rückführung der gesamten Punktwolke auf die WS-Achse.

$$y_{rg} = x_{rg} - w_{rg}\ \alpha_{Dyn}\ w(x_{ws}) \qquad \text{mit} \qquad w(x_{ws}) = \frac{x_{ws}}{w_{ws}\ \gamma_{Dyn}}$$

$$y_{by} = x_{by} - w_{by}\ \beta_{Dyn}\ w(x_{ws}) \qquad \text{mit} \qquad w(x_{ws}) = \frac{x_{ws}}{w_{ws}\ \gamma_{Dyn}}$$

Eine gleichzeitige Dynamikanpassung entlang der Unbunt- bzw. WS-Achse führt zu einer vollen Ausnutzung der Farbraumhöhe sowie zu einer Kontrastanhebung (Abbildung 4.11). Hierbei wird der eingeschränkte Dynamikumfang der ausgelenkten Punktwolke von $x_{ws.min}$ bis $x_{ws.max}$ auf den nun voll zur Verfügung stehenden Dynamikumfang entlang der Unbuntachse von $y_{ws.min} = x_{ws.min}$ bis $y_{ws.max}$ abgebildet.

$$y_{ws} = x_{ws.min} + (x_{ws} - x_{ws.min})\ \frac{y_{ws.max} - y_{ws.min}}{x_{ws.max} - x_{ws.min}}$$

Die Farbumstimmung realisiert somit für jeden Punkt eine Abbildung des Aktivierungsvektors $\mathbf{x} = (x_{rg}, x_{by}, x_{ws})^T$ auf einen neuen Aktivierungsvektor $\mathbf{y} = (y_{rg}, y_{by}, y_{ws})^T$. Da es sich um eine lineare Transformation handelt, bleiben die Relationen der einzelnen Punkte zueinander erhalten.
Die inverse Transformation ermöglicht die Rücktransformation der in ihrer Lage optimierten Punkte der Wolke in den RGB-Farbraum und somit die Erzeugung einzelner oder einer Folge aus Abbildung 4.12 hervorgehender, farbumgestimmter RGB-Farbbilder (Abbildung 4.13).
Die vorgestellte, biologisch motivierte Farbumstimmung gewährleistet somit bezüglich der vier Grundfarben stabile Farbempfindungen.

Abbildung 4.12: *Die dargestellten RGB-Bilder wurden mit einer 3-Chip-CCD-Farbbildkamera mit ausgeschalteter Gammakorrektur, ohne Weißabgleich und besondere Anforderungen an eine gleichmäßige Ausleuchtung übersteuerungsfrei unter verschiedenen Beleuchtungsbedingungen aufgenommen.*

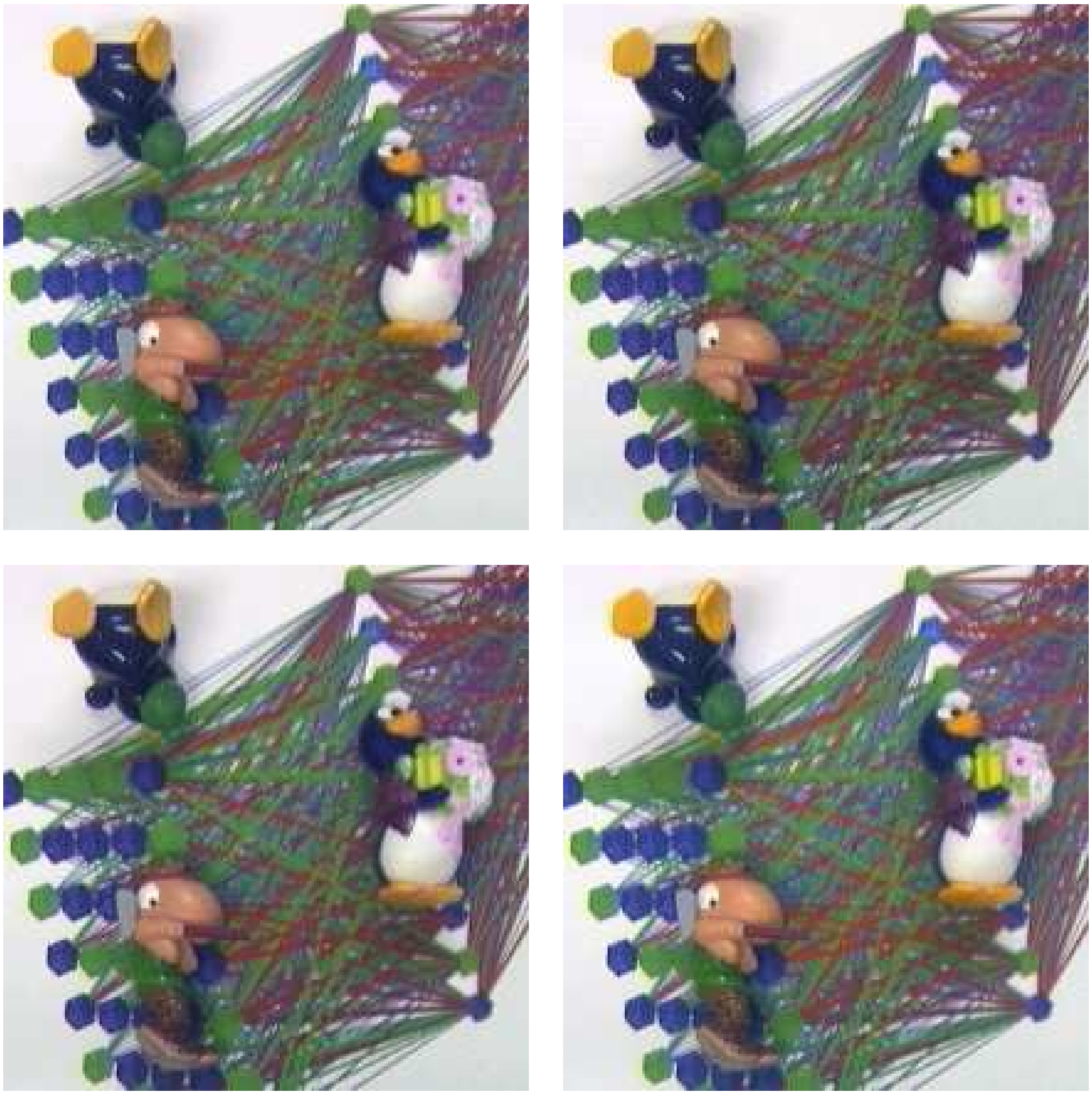


Abbildung 4.13: *Die RGB-Bilder zeigen das Ergebnis nach einer empfindungsgemäßen Farbumstimmung auf der Grundlage des vorgestellten Verfahrens. Es wird offensichtlich, daß das Verfahren trotz unterschiedlicher Beleuchtungsverhältnisse zu äquivalenten Farbempfindungen führt.*

4.5.3 Durch Farbumstimmung optimierte Aktivierungskarten der V1

In diesem Abschnitt soll in erster Linie die Auswirkung der Farbumstimmung auf die Funktion des Attentional Focus Identifier im Sinne einer Erhöhung der Robustheit der Analysevorgänge verdeutlicht werden.

Die Gesamtheit der für ein foveales Bild durch die Farbumstimmung optimierten RG-, BY- und WS-Aktivierungen können in drei bezüglich der Retina topographisch organisierten Merkmalskarten zusammengefaßt werden (Abbildungen 4.14 und 4.15). Als Ort einer neurophysiologischen Repräsentation der Merkmalskarten wird die V1 nach einer Rückprojektion aus der V4 vermutet.

Die Farbumstimmung sorgt für die Überführung der Punktwolke in das mittlere Empfindungsniveau und somit für eine Anordnung der Punkte innerhalb des physiologischen Farbraumes entlang der Unbuntachse. Erst durch diesen Prozeß kann der Simultankontrast innerhalb der bunten topographischen Merkmalskarten RG und BY voll zur Wirkung gelangen. Die Farbumstimmung ermöglicht für ein und dieselbe Szene unter verschiedensten spektralen Zusammensetzungen der Allgemeinbeleuchtung stabile Farbempfindungen in Form der Aktivierungen RG, BY und WS. Dies zeigen die in den Abbildungen 4.14 und 4.15 dargestellten Ergebnisse. Die Geometrie der Punktwolke zeigt aber auch, daß die vier Grundfarben innerhalb der zu analysierenden Szene nicht in ausgewogenen Verhältnissen vorhanden sein müssen. Dennoch werden die Ergebnisse der Farbumstimmung durch den Bildinhalt beeinflußt. Das macht sich besonders bemerkbar, solange im Gesichtsfeld nur wenige, bunte Oberflächen vorhanden sind. Ist im Gesichtsfeld lediglich eine bunte bzw. unbunte Fläche vorhanden, kommt es im physiologischen Farbraum aufgrund der unterschiedlichen Schattierungen zu einer Reduktion der ausgelenkten Punktwolke auf eine ausgelenkte Achse. Diese wird dann durch die Farbumstimmung in die Unbuntachse überführt und einer Dynamikanpassung unterzogen, so daß jede beliebige bunte oder unbunte Fläche als sehr helles Grau bis Weiß empfunden wird. Dieses Phänomen wurde eindrucksvoll durch neurophysiologische Untersuchungen von [Land 1983] vorgeführt.
Je größer die Vielfalt der bunten Bestandteile einer Szene und der folgerichtig ausgelösten Farbempfindungen innerhalb des physiologischen Farbraumes, desto unabhängiger wird das Umstimmungsergebnis vom Bildinhalt.
Auch spielt die in der topographischen Repräsentation der visuellen Szene enthaltene Größe von Regionen identischer Farbreize im nichttopographisch organisierten physiologischen Farbraum keine Rolle, da es hier zu einer Abbildung dieser Reize auf einen gemeinsamen Punkt kommt.
Somit spielt gerade bei natürlichen Farbszenen die Vielfalt der Schattierungen zwischen den bunten Objekten bei der Ausprägung einer stabilen Achse innerhalb der Punktwolke eine wesentliche Rolle. Diese Farbempfindungen befinden sich nach der Farbumstimmung auf der Unbuntachse bzw. in ihrer unmittelbaren Nähe.
Berücksichtigt man weiterhin, daß es sich bei der Farbumstimmung des Menschen um einen dynamischen Prozeß über einer Flut von visuellen Sinneseindrücken handelt, so

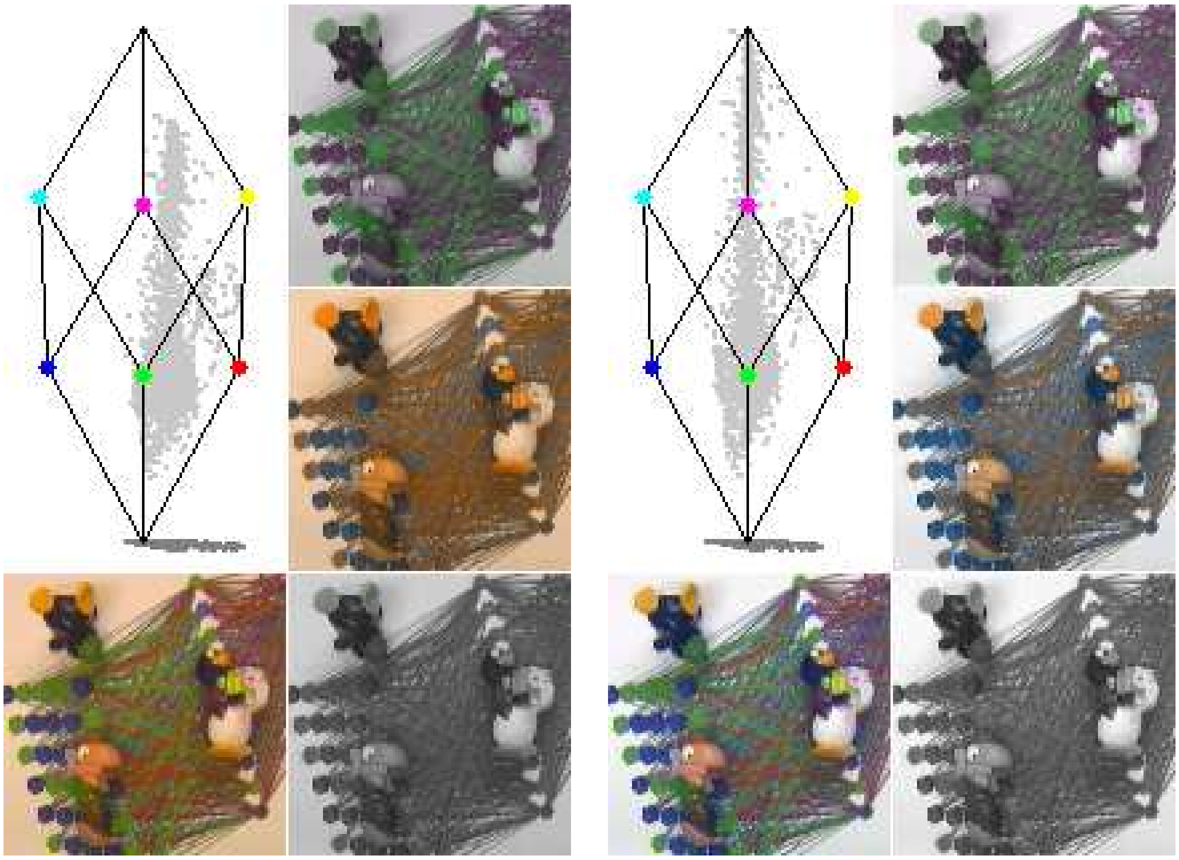


Abbildung 4.14: *(links) Das RGB-Bild unter Beleuchtungsverhältnissen, die von denen des mittleren Tageslichts abweichen sowie die folgerichtig im physiologischen Farbraum ausgelenkte Punktwolke. Weiterhin sind die RG-, BY- und WS-Aktivierungskarten dargestellt. Hierbei werden für die RG- sowie die BY-Merkmalskarte die Komplementärfarbpaare des physiologischen Farbraumes Grün/Purpur sowie Violettblau/Gelb genutzt. (rechts) Die durch Farbumstimmung optimierten Merkmalskarten sowie das nach der Rücktransformation entstehende RGB-Bild. Die Lage der Punktwolke repräsentiert nun das durch Umstimmung angestrebte mittlere Empfindungsniveau.*

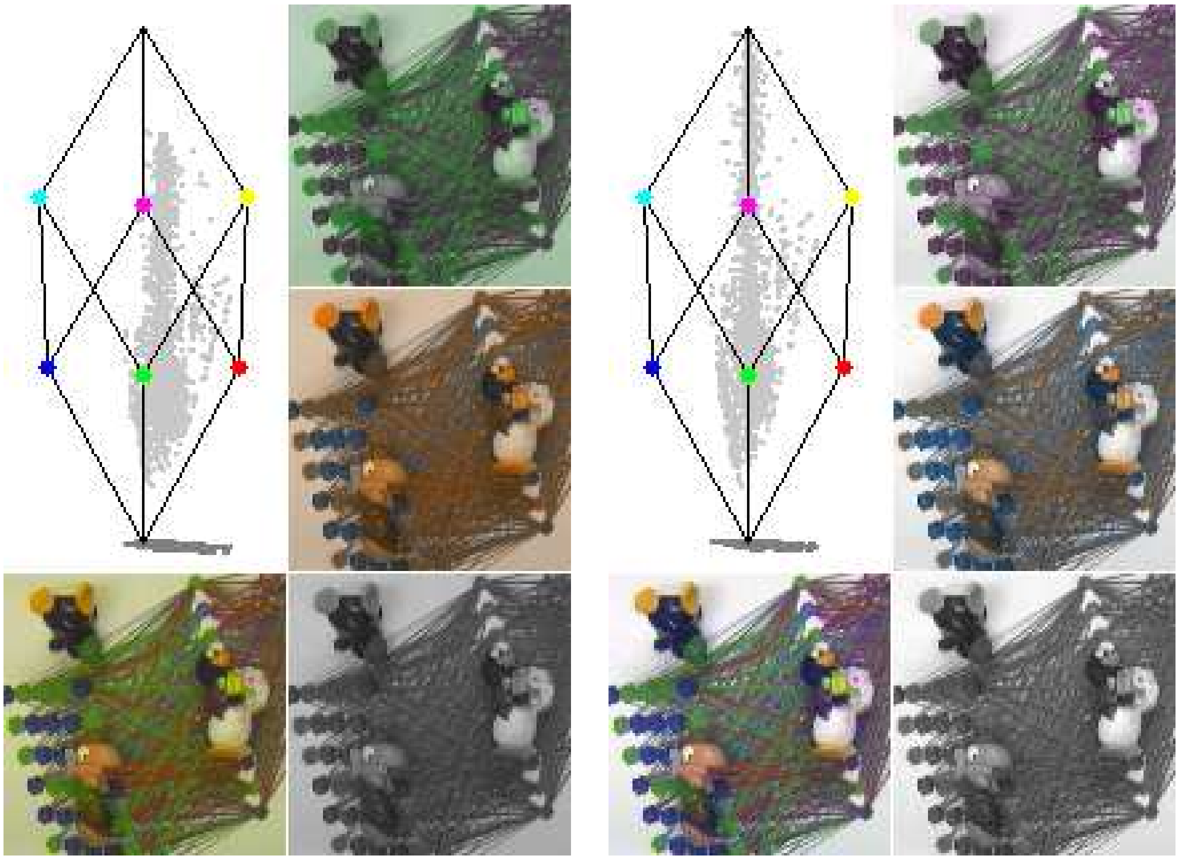


Abbildung 4.15: *(links) Das RGB-Bild unter Beleuchtungsverhältnissen, die von denen des mittleren Tageslichts abweichen sowie die folgerichtig im physiologischen Farbraum ausgelenkte Punktwolke. Weiterhin sind die RG-, BY- und WS-Aktivierungskarten dargestellt. Hierbei werden für die RG- sowie die BY-Merkmalskarte die Komplementärfarbpaare des physiologischen Farbraumes Grün/Purpur sowie Violettblau/Gelb genutzt. (rechts) Die durch Farbumstimmung optimierten Merkmalskarten sowie das nach der Rücktransformation entstehende RGB-Bild. Die Lage der Punktwolke repräsentiert nun das durch Umstimmung angestrebte mittlere Empfindungsniveau.*

lassen sich doch recht plausibel die durch Farbumstimmung erzielten stabilen Aktivierungen der RG-, BY- und WS-Aktivierungskarten und die damit verbundenen stabilen Farbempfindungen erklären.

Die vorgestellte, biologisch motivierte Farbumstimmung gewährleistet somit bezüglich der vier Grundfarben stabile Farbempfindungen.
Es wird offensichtlich, daß sich ein farbmetrisches Meßsystem grundsätzlich nicht auf ein physikalisches zurückführen läßt. Der Prozeß der Farbumstimmung schließt jegliche Korrelation zwischen Farbreiz und Farbempfindung aus.
Die Vorteile des entwickelten Verfahrens bestehen zum einen in der Referenzfreiheit. Weiterhin gehen alle Bildpixel gleichberechtigt in die Bestimmung der Korrekturgröße ein. Die räumlichen Relationen der Punkte zueinander bleiben innerhalb der Wolke erhalten. Es kommt daher zu keiner Verfälschung der Farbinformation.
Die Achsen des Farbraumes sind für eine Farbumstimmung optimal dekorreliert angeordnet. Die Geometrie des physiologischen Farbraumes gewährleistet unter allen Umständen die Möglichkeit einer Bestimmung der relevanten Hauptachse. Der Farbfehler ist genau quantifizierbar. Der Einsatz der entwickelten Farbumstimmung zum automatischen, referenzfreien Weißabgleich stellt lediglich ein mögliches Anwendungsbeispiel unter anderen denkbaren dar.

4.6 Neuronale Modellarchitektur zur Texturanalyse

4.6.1 Architekturkonzept im Überblick

In der im Rahmen des Projektes entwickelten Architektur erfolgt die Analyse der WS-Aktivierungskarte durch ein Array cortikaler Verarbeitungscolumnen, deren Aufbau sich strukturell und funktionell an den Minicolumnen des primären visuellen Cortex orientiert. Dieses columnare Array analysiert parallel die bezüglich der Retina topographisch organisierte WS-Aktivierungskarte (Abbildung 4.16 (links)). Jede Verarbei-

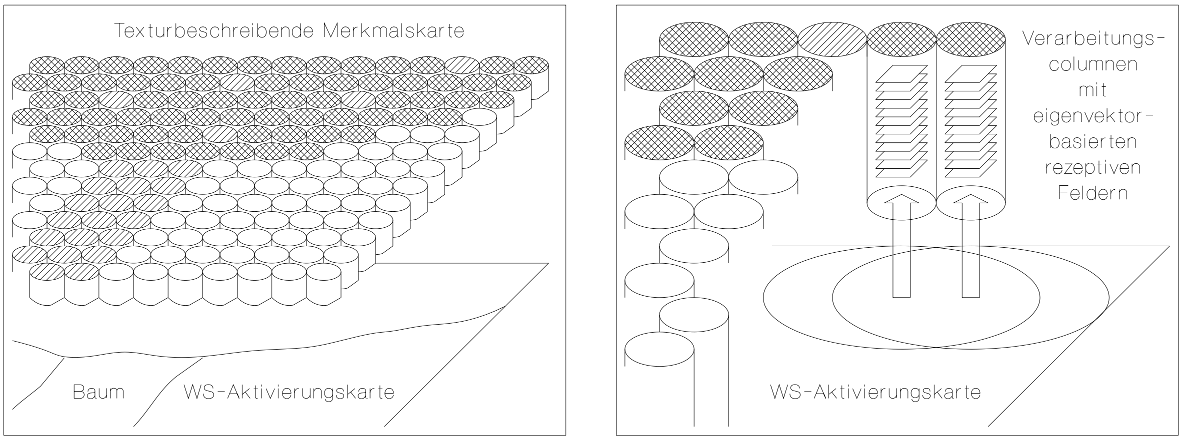


Abbildung 4.16: *Die neuronal motivierte Modellarchitektur zur Texturbeschreibung.*

tungscolumne extrahiert dabei elementare Bildmerkmale aus ihrem lokalen Analysegebiet innerhalb der Aktivierungskarte mit Hilfe komplexer rezeptiver Felder (Abbildung 4.16 (rechts)). Diese werden jedoch nicht als analytisch faßbare Filtermasken vorgegeben, sondern in einem adaptiven Selbstorganisationsprozeß auf der Basis einer neuronal motivierten Hauptkomponentenanalyse strukturiert. Die Selbstorganisation der eigenwertgrößten Hauptkomponenten für zufällig gewählte lokale Ausschnitte der WS-Aktivierungskarte führt auch für sehr unterschiedliche natürliche Farbszenen zu qualitativ ähnlichen Sätzen rezeptiver Felder. Dieses Ergebnis führte zu der Vermutung, daß ein universeller Satz komplexer rezeptiver Felder für jede cortikale Verarbeitungscolumne unabhängig von ihrer Lokalisation innerhalb der Modellarchitektur für eine Analyse der WS-Aktivierungskarte optimal geeignet wäre. Diese erste Stufe der Modellarchitektur realisiert somit eine lokale Hauptkomponentenanalyse innerhalb einer jeden Verarbeitungscolumne und ermöglicht damit eine lokale Beschreibung des zugehörigen lokalen Analysegebietes mit minimaler Anzahl beschreibender Parameter in Form von Fitwertaktivierungen. Alle Fitwerte einer Verarbeitungscolumne werden in einer zweiten Stufe zu einer einzigen, spezifischen columnaren Aktivierung zusammengefaßt, die ein Maß für die „texturelle Rauhigkeit" innerhalb des korrespondierenden Analysegebietes beschreibt.

Die spezifischen Aktivierungen aller Modellcolumnen werden in einer topographisch organisierten Merkmalskarte zur Beschreibung der texturellen Rauhigkeit natürlicher Farbszenen zusammengefaßt (Abbildung 4.16 (links)).
Ähnliche Analysevorgänge wurden auch für die RG- und BY-Aktivierungskarten untersucht. Diese besitzen jedoch, der Geometrie des physiologischen Farbraumes folgend, einen geringeren Dynamikumfang. Die meisten Farbempfindungen bei der Betrachtung natürlicher Farbszenen liegen generell nach erfolgter Farbumstimmung in unmittelbarer Umgebung der Unbuntachse (Abbildungen 4.14 und 4.15). Somit erwies sich die WS-Aktivierungskarte zur Ermittlung eines texturellen Rauhigkeitsmaßes als ausreichend.

4.6.2 Selbstorganisation der rezeptiven Felder

Auswahl und Vorverarbeitung des Inputs

Während der Trainingsphase wurde eine Vielzahl von RGB-Farbbildern mit 512×512 Pixeln sowie 8 bit Auflösung pro Band untersucht. Dabei wurde ganz bewußt kein Versuch unternommen, optische Unregelmäßigkeiten zu korrigieren. Die RGB-Werte eines jeden Bildes wurden mittels der vorgestellten direkten Transformation in den physiologischen Farbraum überführt und einer Farbumstimmung unterzogen. Von den drei resultierenden, bezüglich der Retina topographisch organisierten Aktivierungskarten (RG, BY, WS) wurde für die folgenden Untersuchungen die WS-Aktivierungskarte verwendet, da diese den höchsten Kontrast aufwies. Aus ihr wurden lokale Bildausschnitte mit einer Dimension von 16×16 Pixeln (somit $m = 256$) zufällig ausgewählt. Diese bildeten die Inputvektoren $\mathbf{x} = (x_1, ..., x_j, ..., x_m)^T$.
Bei Verwendung von Ausschnitten natürlicher Farbszenen als Inputvektoren zeigt der eigenwertgrößte Eigenvektor der Inputdaten $\mathbf{w}^{(1)}$ in die gleiche Richtung wie der Vektor zum Schwerpunkt der Verteilung (Abbildung 4.18). Somit bildet sich in einem solchen Fall auch ein optimaler Merkmalsraum heraus. Besitzt die Verteilung der Inputvektoren diese statistische Eigenschaft nicht, kann sich auch kein optimaler Merkmalsraum herausbilden, da sich jeder Eigenvektor orthogonal zum vorhergehenden strukturiert (Abbildung 4.17). Somit muß eine solche Verteilung der Eingangsdaten mittelwertfrei gemacht werden, bevor sie dem Lernalgorithmus angeboten wird.
Um den Ansatz universell auch für nichtmittelwertfreie Datenverteilungen einsetzen zu können, wurde eine prinzipiell zweistufige Vorverarbeitung realisiert. Während der Belehrung wurde dazu vor jedem Lernschritt ein gleitender Mittelwert der Inputvektoren nach

$$\bar{\mathbf{x}}(t) = \frac{1}{n+1}\,(n\;\bar{\mathbf{x}}(t-1) + \mathbf{x}(t)) \qquad \text{mit} \qquad \bar{\mathbf{x}}(0) = \mathbf{x}(0)$$

bestimmt. Hierbei beschreibt n die Nummer des zu betrachtenden Inputvektors der Lernstichprobe. Vom Inputvektor $\mathbf{x}$ wurde dieser Mittelwertvektor $\bar{\mathbf{x}}$ subtrahiert

$$\mathbf{x}' = \mathbf{x} - \bar{\mathbf{x}}\ .$$

Die Elemente des resultierenden Vektors $\mathbf{x}'$ wurden dann auf das Intervall $< -1, 1 >$ skaliert.

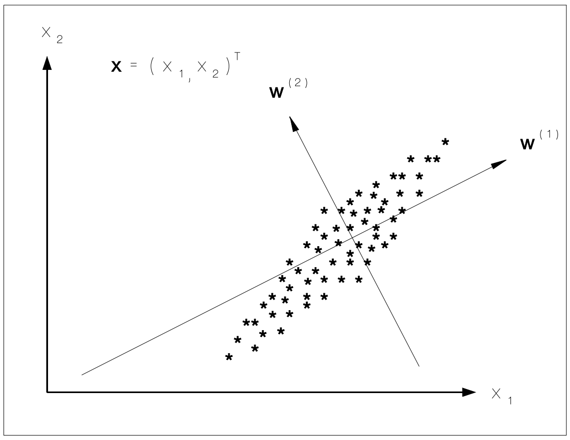


Abbildung 4.17: *Prinzipielle Darstellung der Korrelation zweier Variablen x_1 und x_2. Die Datenverteilung ist nicht mittelwertfrei. Somit ist hier zuerst der Mittelwert zu extrahieren, bevor die Bestimmung eines günstigeren orthogonalen Basissystems erfolgen kann. Wird dieser Umstand nicht berücksichtigt, ist die dekorrelierende Wirkung des neuen Basissystems nicht optimal.*

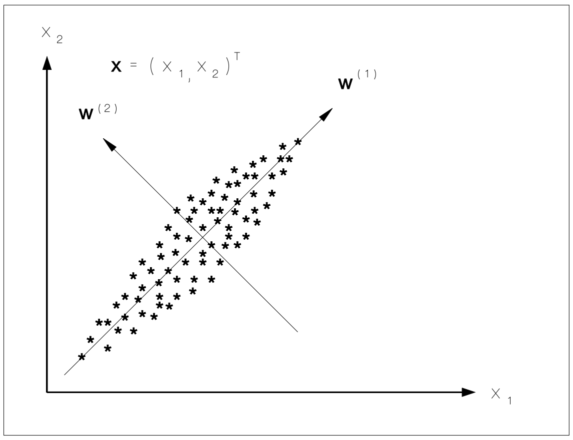


Abbildung 4.18: *Prinzipielle Darstellung der Korrelation zufällig gewählter Paare nebeneinanderliegender Grauwertpixel x_1 und x_2 einer natürlichen visuellen Szene. Hier sind Schwerpunktsvektor der Datenverteilung und eigenwertgrößter Eigenvektor $\mathbf{w}^{(1)}$ identisch.*

Wie aus dieser Vorgehensweise ersichtlich, ist keinerlei a-priori Information über den Input notwendig. Deshalb kann man schlußfolgern, daß dieses Verfahren zur Dekorrelation beliebiger Daten verwendet werden kann.

Lernprozeß

Für die Belehrung wurde die von [Oja 1982] für ein einzelnes Neuron vorgestellte Lernregel verwendet und im weiteren Verlauf der Arbeiten modifiziert.

$$w_j(t+1) = w_j(t) + \gamma(t)\; y(t)\; [x_j(t) - y(t)\; w_j(t)] \qquad \text{mit} \qquad j = 1, 2, ..., m$$

Hierbei bezeichnet $\mathbf{x} = (x_1, ..., x_j, ..., x_m)^T$ den aktuellen Eingangsvektor der auf das Intervall $< -1, 1 >$ skalierten, vom gleitenden Mittelwert befreiten Pixel eines zufällig gewählten lokalen Ausschnittes der WS-Aktivierungskarte, y das Ausgangssignal, w_j die Wichtung des Eingangsknotens x_j und γ die Lernrate. In [Oja 1982] wurde gezeigt, daß diese Lernregel einen Wichtungsvektor $\mathbf{w}$ erzeugt, der gerade dem Eigenvektor der Korrelationsmatrix aller dargebotenen Ausschnitte $\mathbf{x}$ mit dem größten Eigenwert entspricht. Sie extrahiert somit die dominanteste Hauptkomponente der Eingangsdaten. Gleichzeitig gewährleistet die Lernregel eine Normierung des Wichtungsvektors $\mathbf{w}$ auf die Einheitslänge.

Diese Lernregel wurde auf n Ausgangsneuronen erweitert [Sanger 1989], um so nacheinander die stärksten Hauptkomponenten bestimmen zu können. Dabei handelt es sich für jedes der n Neuronen um einen zweiphasigen Prozeß, in dem zunächst vom Eingangsvektor $\mathbf{x}^{(i-1)}$ die Anteile abgezogen werden, die durch den Wichtungsvektor $\mathbf{w}^{(i)}$ bereits dargestellt werden können.

$$\mathbf{x}^{(i)}(t) = \mathbf{x}^{(i-1)}(t) - y_i(t)\; \mathbf{w}^{(i)}(t) \qquad \text{mit} \qquad \mathbf{x}^{(0)} = \mathbf{x} \qquad \text{und} \qquad i = 1, 2, ..., n$$

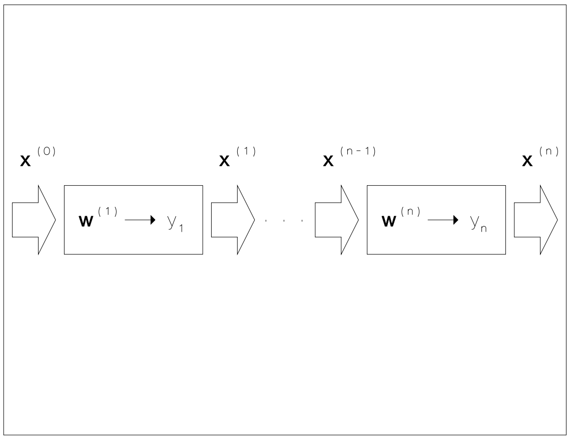
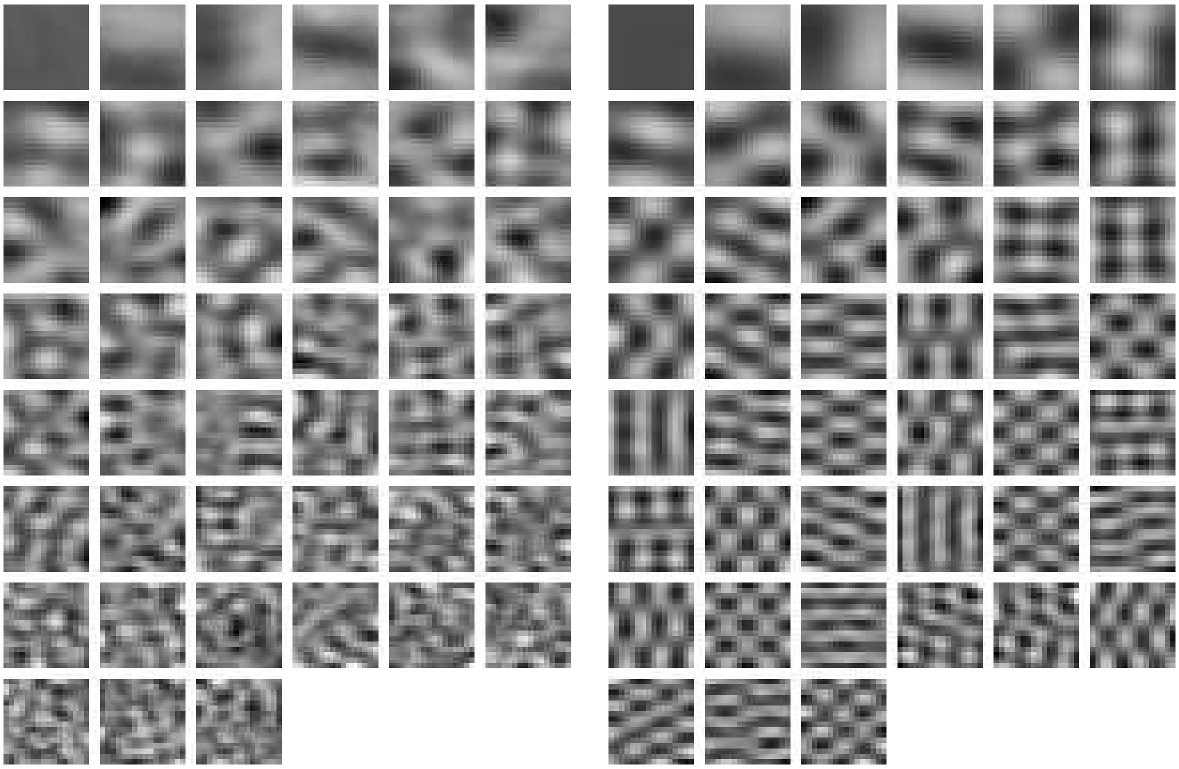


Abbildung 4.19: *Prinzipielle Vorgehensweise zur Feststellung des Umfanges der bereits vorhandenen Übereinstimmung des Inputvektors $\mathbf{x}^{(i-1)}$ mit dem Wichtungsvektor $\mathbf{w}^{(i)}$. Hieraus ergibt sich der korrigierte Inputvektor $\mathbf{x}^{(i)}$.*

Erst danach erfolgt die eigentliche Aktualisierung der Komponenten $w_j^{(i)}$ des Wichtungsvektors $\mathbf{w}^{(i)}$ mit dem korrigierten Vektor $\mathbf{x}^{(i)}$ als Input.

$$w_j^{(i)}(t+1) = w_j^{(i)}(t) + \gamma_i(t)\; y_i(t)\; x_j^{(i)}(t) \qquad \text{mit} \qquad i = 1, 2, ..., n; \; j = 1, 2, ..., m$$

Abbildung 4.20: *Für die Belehrung wurden zufällig gewählte lokale Ausschnitte der links dargestellten WS-Aktivierungskarte einer natürlichen Farbszene genutzt. Darunter befindet sich die Darstellung der Wichtungsvektoren* $\mathbf{w}^{(i)}$ *als rezeptive Felder. Die anfänglich zufällig gewichteten Vektoren strukturieren sich während des Lernprozesses aufgrund der Statistik der Eingangsdaten (links nach 1.000 Lernschritten, rechts nach 50.000 Lernschritten).*



Die Anzahl der Neuronen wurde experimentell bestimmt. Dabei zeigte sich, daß $n = 45$ Hauptkomponenten genügen, um nahezu die gesamte unbunte Information aus beliebigen 16×16 Pixel großen Bildausschnitten natürlicher RGB-Farbbilder zu extrahieren. Das Netz wurde durch Setzen aller Wichtungen $w_j^{(i)}$ auf einen zufällig kleinen Wert initialisiert, so daß die Quadratsumme der Komponenten eines jeden Wichtungsvektors $\mathbf{w}^{(i)}$ gleich eins war.

Das Training bestand in der Analyse der zufällig gewählten Ausschnitte der WS-Aktivierungskarte und einer sofortigen Aktualisierung aller Wichtungen der rezeptiven Felder. Zu Beginn der Belehrung schwingen die Wichtungsvektoren sehr stark (Abbildung 4.20 (links)).
Der Betrag der Wichtungsvektoren wird sehr groß, obwohl er sich im weiteren Verlauf des Trainings auf den Wert $1,0$ einpegelt. Das extreme Einschwingen ist dadurch zu erklären, daß die Eigenvektoren der erwarteten Korrelationsmatrix gelernt werden. Bei wenigen Trainingsmustern, wie es kurz nach dem Start eines Trainingsverlaufs der Fall ist, sind diese noch sehr ungenau und ändern sich ständig. Die Eigenvektoren stabilisieren sich erst nach einer bestimmten Anzahl von Trainingsmustern. Somit müssen sich die Wichtungsvektoren auf die sich anfänglich stark ändernden Eigenvektoren einschwingen.
Bei kleineren Lernraten ist der Einschwingvorgang nicht so stark ausgeprägt, so daß zumindest numerische Stabilität vorliegt. Eine kleine Lernrate hat jedoch den Nachteil, daß der Algorithmus nur langsam konvergiert und die Belehrung von vielen Neuronen nicht mehr praktikabel ist. Das Problem bestand also darin, das starke Einschwingen der Wichtungsvektoren abzufangen, ohne den Lernalgorithmus prinzipiell zu verändern. Aus der Überlegung heraus, daß sich synaptische Verbindungen in biologischen Systemen nicht sprunghaft ändern können, wurde eine maximale Wichtungsänderung $\Delta w_j^{(i)}$ für alle Wichtungen festgelegt. Überschritt der Betrag der berechneten Wichtungsänderung diese Schwelle, wurde die Wichtungsänderung auf diesen Maximalwert begrenzt. Nach dieser Modifikation blieb der Algorithmus auch bei Verwendung sehr großer Lernraten numerisch stabil. Desweiteren konvergiert er viel schneller, da die Einschwingphase erheblich verkürzt werden konnte.
Bei der Wahl der maximalen Wichtungsänderung ist darauf zu achten, daß einerseits der Einschwingvorgang genügend gedämpft wird und andererseits diese Schwelle im eingeschwungenen Zustand mit großer Wahrscheinlichkeit nicht erreicht wird. Mit einem Wert von $\Delta w_j^{(i)} = \pm 1$ ist dies gewährleistet, da die Wichtungsvektoren im eingeschwungenen Zustand auf den Betrag von eins normiert bleiben. Dadurch sind selbst im ungünstigsten Fall (nur eine Wichtung $w_j^{(i)} \neq 0$) nur zwei Lernschritte notwendig, um den gesamten Wertebereich der Wichtungsvektoren zu durchlaufen.

Verwendung einer modifizierten Zeitabhängigkeit der Lernrate

Bei den ersten Versuchen, mit Hilfe der Lernregel rezeptive Feldstrukturen zu erzeugen, zeigte sich ein recht träges Stabilisierungsverhalten der Neuronen höherer Ordnung. Somit war hier eine unverhältnismäßig hohe Anzahl von Lernschritten erforderlich.
Dieser Umstand läßt sich durch die Abhängigkeit der Neuronen höherer Ordnung von

denen niederer Ordnung erklären. Verändert das Neuron i seine Wichtungen, so müssen aufgrund der Lernregel alle Neuronen j mit $j > i$ ebenfalls ihre Wichtungen ändern.
Diesen Sachverhalt kann man sich auch anhand der Kopplung der Neuronen durch den korrigierten Input verdeutlichen. Es existiert praktisch eine Rückkopplung des Outputs eines Neurons auf den Input des folgenden

$$\mathbf{x}^{(i)}(t) = \mathbf{x}^{(i-1)}(t) - y_i(t)\,\mathbf{w}^{(i)}(t) \qquad \text{mit} \qquad \mathbf{x}^{(0)} = \mathbf{x} \qquad \text{und} \qquad i = 1, 2, ..., n\ .$$

Das Problem besteht nun darin, daß die Neuronen höherer Ordnung zunächst mit den Neuronen niederer Ordnung schwingen, ohne effektiv belehrt zu werden. Der Wichtungsvektor $\mathbf{w}^{(i)}$ des Neurons i muß sich erst stabilisieren, bevor der Wichtungsvektor $\mathbf{w}^{(i+1)}$ des Neurons $i+1$ mit der gleichen Effektivität belehrt werden kann.
Zu diesem Zeitpunkt hat sich jedoch die für alle Neuronen zur Verfügung stehende, ursprünglich gleichgroße Lernrate bereits verringert. Daraus folgt, daß mit steigender Ordnung der Neuronen die Lernrate effektiv immer kleiner wird.
Aus diesem Grund ist es zweckmäßig, nicht für alle Neuronen die gleiche Lernrate zu verwenden, sondern diese dem Lernprozeß anzupassen, sie praktisch in Richtung der Neuronen höherer Ordnung zeitverzögert abfallen zu lassen (Abbildung 4.21).

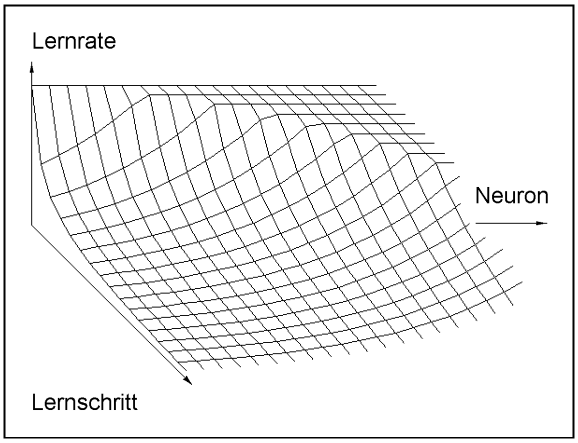


Abbildung 4.21: *Der kontinuierlich, jedoch mit zunehmender Ordnung der Neuronen zeitverzögert abfallende Lernratenverlauf.*

Somit wird die Konvergenz der Wichtungen durch eine sich kontinuierlich verringernde Lernrate bestimmt

$$\gamma_0(t) = \frac{a}{t+1}\ ,$$

$$\gamma_i(t) = \gamma_{i-1}(t)\ v \qquad \text{mit} \qquad i = 1, 2, ..., n\ .$$

Hierbei ist $\gamma_i(t)$ die Lernrate des Neurons i zum Zeitpunkt t, v der Rekursionsparameter, t der aktuelle Lernschritt und a der Ausgangswert. Nach Nutzung dieser Funktion zur Berechnung der Lernrate waren alle Neuronen zu jedem Zeitpunkt der Belehrung auf etwa dem gleichen Belehrungsniveau. Dadurch wurde es möglich, die Anzahl der Lernschritte erheblich zu verringern.

Nach zirka 50.000 Lernschritten kam es zu einer Stabilisierung der Wichtungen auch des 45. Neurons. Die Wichtungsänderungen unterschritten dann einen festgelegten Schwellwert, so daß hier der Lernvorgang abgebrochen werden konnte.
Mit einer für alle Neuronen gleichmäßig abfallenden Lernrate ist es nicht möglich, in vertretbarer Zeit mehr als acht bis zehn rezeptive Felder zu belehren. Die Ursache liegt in dem beschriebenen Einschwingvorgang der Neuronen niederer Ordnung begründet. Dieser macht eine sehr langsam abfallende Lernrate mit möglichst hohem Startwert erforderlich. Die Höhe des Startwertes wird jedoch durch Anforderungen an die numerische Stabilität des Algorithmus begrenzt.
Somit ist der Lernvorgang ohne modifizierte Zeitabhängigkeit der Lernrate ausgesprochen zeitintensiv.

In Abbildung 4.20 sind rechts die gelernten Wichtungsvektoren dargestellt. Der jeweilige Wert eines Elements (einer Synapse) eines Wichtungsvektors wird durch die Graustufe beschrieben. Eine helle Graustufe stellt somit einen hohen Wert dar. Bemerkenswert ist die gleichmäßige Struktur der Wichtungsvektoren, die sich mit immer höheren Frequenzen wiederholen. Fast jede Frequenz ist mit verschiedenen Orientierungen und Phasenverschiebungen vertreten.
In [Hancock 1992] werden Wichtungsvektoren vorgestellt, die anhand von gedruckten Texten belehrt wurden. Diese Wichtungsvektoren haben eine gewisse Ähnlichkeit mit den in [Turner 1986] vorgestellten Gaborfiltern. Die Vermutung liegt nahe, daß die herausgebildeten Wichtungsvektoren zumindest teilweise durch Gaborfunktionen approximierbar sind.

4.6.3 Fitwertbestimmung und Informationsextraktion

Der auf diesem Wege generierte Satz von Eigenvektoren (Abbildung 4.20 (rechts)) wird entsprechend der Modellvorstellung (Abbildung 4.16 (rechts)) in jede einzelne Verarbeitungscolumne des columnaren Arrays dupliziert. Jede Verarbeitungscolumne stellt somit ein elementares Modul zur Merkmalsextraktion dar, das innerhalb seines Analysegebietes die Untersuchung eines bestimmten lokalen Ausschnittes $\mathbf{x}^{(k)}$ der WS-Aktivierungskarte unabhängig von benachbarten Verarbeitungscolumnen realisiert. Um die Qualität des gewonnenen Merkmalsraumes einschätzen zu können, wurde die Orthogonalität der Wichtungsvektoren überprüft. Es wurde jeweils der Winkel zwischen den Wichtungsvektoren $\mathbf{w}^{(i)}$ und $\mathbf{w}^{(i+1)}$ berechnet. Hierbei konnte festgestellt werden, daß der durch die Wichtungsvektoren aufgespannte Merkmalsraum in sehr guter Näherung orthogonal ist.
Weiterhin ist es erforderlich festzustellen, inwieweit es mit Hilfe der einmal gelernten rezeptiven Felder auch für unterschiedlichste WS-Aktivierungskarten natürlicher Farbszenen gelingt, die beobachtete Variation von Pixeln lokaler Ausschnitte $\mathbf{x}^{(k)} = (x_1^{(k)}, ..., x_j^{(k)}, ..., x_m^{(k)})^T$ der WS-Aktivierungskarte auf eine Abhängigkeit von einer kleineren Anzahl unbekannter Parameter in Form eines Fitwertvektors $\mathbf{y}^{(k)} = (y_1^{(k)}, ..., y_i^{(k)}, ..., y_n^{(k)})^T$ mit $n < m$ zurückzuführen. Gegenüber den direkt zugänglichen Pixelaktivierungen $x_j^{(k)}$ würden die Variablen $y_i^{(k)}$ eine sparsamere Beschreibung der beobachteten lokalen Aktivierungsverteilungen ermöglichen.

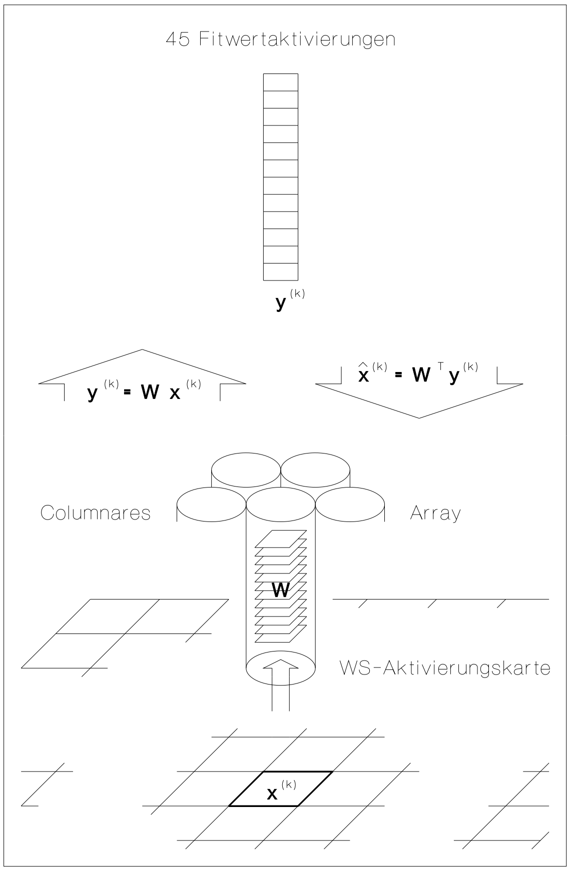


Abbildung 4.22: *Modellarchitektur zur Bestimmung der Abbildungsgüte.*

Zur Bestimmung der Abbildungsgüte des durch den Satz von Eigenvektoren $\mathbf{W} = (\mathbf{w}^{(1)}, ..., \mathbf{w}^{(i)}, ..., \mathbf{w}^{(n)})$ aufgespannten Merkmalsraumes wurden WS-Aktivierungskarten verschiedener natürlicher Farbszenen einer segmentweisen Hin- und Rücktransformation unter Verwendung des ersten Wichtungsvektors $\mathbf{w}^{(1)}$ unterworfen.

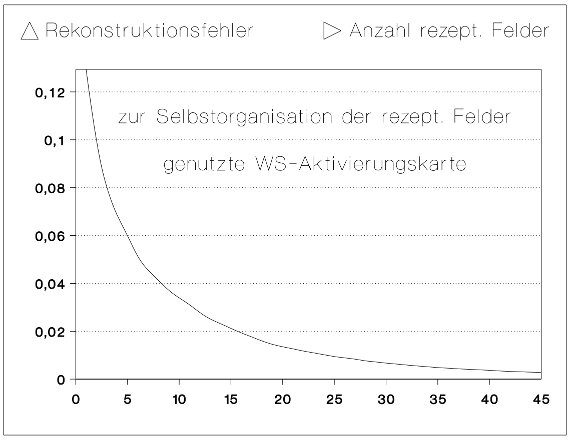
$$y_1^{(k)} = \mathbf{w}^{(1)\,T}\ \mathbf{x}^{(k)} \qquad \text{mit} \qquad k = 1, 2, ..., o$$

$$\hat{\mathbf{x}}^{(k)} = \mathbf{w}^{(1)}\ y_1^{(k)} \qquad \text{mit} \qquad k = 1, 2, ..., o$$

Die dabei entstehende Abweichung zwischen originaler und rücktransformierter Aktivierungskarte wurde als mittlerer Rekonstruktionsfehler ε_i nach

$$\varepsilon_i = \sum_k \sum_{x\ y} \frac{(x_{xy}^{(k)} - \hat{x}_{xy}^{(k)})^2}{(x_{xy}^{(k)})^2} \qquad \text{mit} \qquad i = 1, 2, ..., n$$

berechnet. Hierbei bezeichnet $\hat{x}^{(k)}$ die durch Hin- und Rücktransformation entlang der i rezeptiven Felder $\mathbf{w}^{(1)}$ bis $\mathbf{w}^{(i)}$ der k-ten Verarbeitungscolumne vorgenommene Schätzung des Pixels an der Position (x, y) innerhalb des k-ten lokalen Ausschnittes der WS-Aktivierungskarte (Abbildung 4.22).
Dadurch wird es möglich, den Beitrag eines jeden der n rezeptiven Felder $\mathbf{w}^{(i)}$ bei der Informationsextraktion einzuschätzen.
Somit kommt es zur Bestimmung eines durchschnittlichen Rekonstruktionsfehlers bezüglich eines jeden Wichtungsvektors $\mathbf{w}^{(i)}$ über die gesamte WS-Aktivierungskarte einer natürlichen Farbszene.



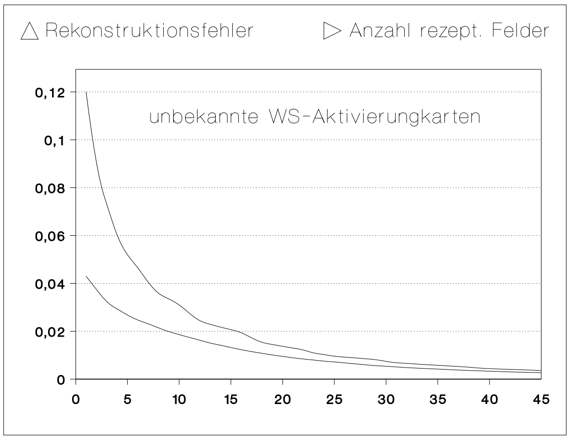


Abbildung 4.23: *(links) Darstellung des mittleren Rekonstruktionsfehlers bei Anwendung der Hin- und Rücktransformation auf die zur Belehrung genutzte WS-Aktivierungskarte (Abbildung 4.20 (links)). (rechts) Darstellung des mittleren Rekonstruktionsfehlers bei Anwendung auf unbekannte WS-Aktivierungskarten (Abbildungen 4.24 (links)) sowie Abbildung 4.25 (links)).*

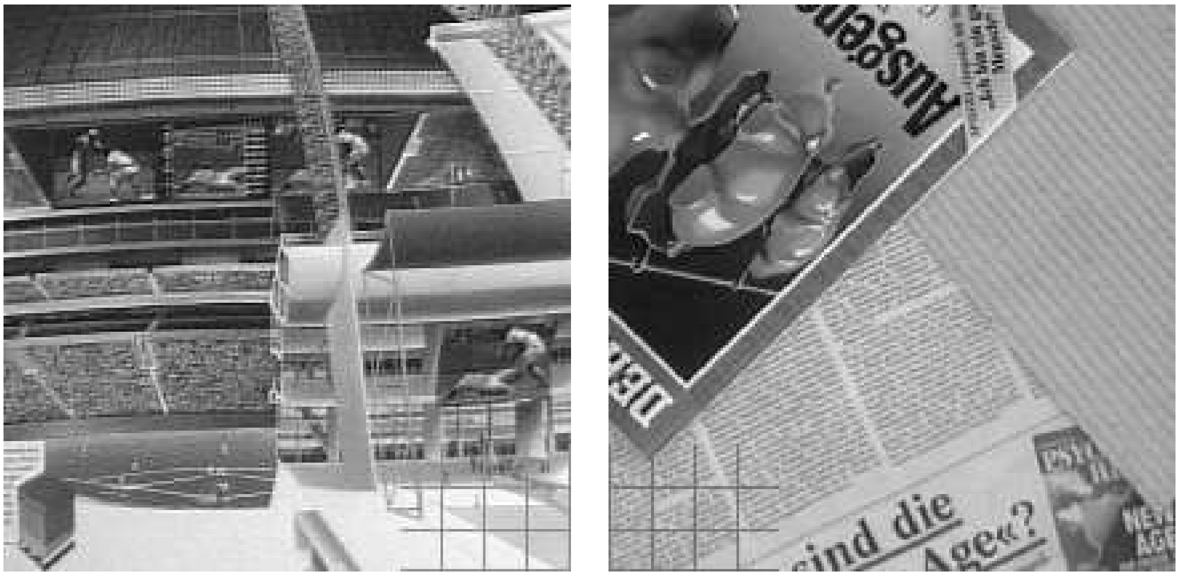


Abbildung 4.24: *Beide WS-Aktivierungskarten wurden, wie in Abbildung 4.22 angedeutet, segmentartig unter Nutzung der 45 für die in Abbildung 4.20 dargestellte WS-Aktivierungskarte gelernten rezeptiven Felder* $\mathbf{w}^{(i)}$ *analysiert. Die Projektion der* 16×16 *Aktivierungspixel eines Segments* $\mathbf{x}^{(k)}$ *entlang der 45 rezeptiven Felder ergab pro Segment einen 45-dimensionalen Fitwertvektor* $\mathbf{y}^{(k)}$*, der nahezu die gesamte Information des analysierten Aktivierungssegments beinhaltete. Die Ergebnisse der Rücktransformation der 45 Fitwertaktivierungen aller Verarbeitungscolumnen entlang ihrer rezeptiven Felder* $\mathbf{w}^{(i)}$ *sind hier dargestellt.*

Bemerkenswert ist der kontinuierliche Abfall des mittleren Rekonstruktionsfehlers bei Hinzunahme weiterer rezeptiver Felder $\mathbf{w}^{(i)}$ (Abbildung 4.23 (links)). Das deutet darauf hin, daß die Wichtungsvektoren $\mathbf{w}^{(i)}$ tatsächlich nach ihrer Relevanz bezüglich einer angestrebten optimalen Informationsextraktion geordnet vorliegen.
Dieser kontinuierliche Verlauf zeigt aber gleichzeitig, daß die experimentell realisierte Wahl der rezeptiven Feldgröße von 16×16 Synapsen für natürliche Szenen darstellende WS-Aktivierungskarten ein Optimum darstellt. Eine nichtoptimale Wahl würde sich in einem partiell linearen Verlauf des Rekonstruktionsfehlers bemerkbar machen, wie er z.B. bei einer rezeptiven Feldgröße von 32×32 Synapsen auftritt.
Auch die Universalität des gelernten Satzes rezeptiver Felder $\mathbf{W} = (\mathbf{w}^{(1)}, ..., \mathbf{w}^{(i)}, ..., \mathbf{w}^{(n)})$ zur Informationsextraktion aus beliebigen WS-Aktivierungskarten läßt sich anhand des mittleren Rekonstruktionsfehlers für unbekannte WS-Aktivierungskarten verdeutlichen (Abbildung 4.23 (rechts)).
Diese Ergebnisse werden durch eine rein visuelle Beurteilung der hin- und rücktransformierten WS-Aktivierungskarten beliebiger Farbzsenen bestätigt (Abbildung 4.24).

4.7 Die texturbeschreibende Merkmalskarte

Die Abbildung der Grauwerte eines lokalen Bildausschnittes der WS-Aktivierungskarte auf n Fitwertaktivierungen stellt eine erste erhebliche Datenreduktion bei maximaler Informationsübertragung dar. Das Ziel besteht jedoch darin, die n Fitwertaktivierungen einer Verarbeitungscolumne in einem zweiten Schritt zu einer einzigen columnaren Aktivierung zusammenzufassen, die ein Maß für die texturelle Rauhigkeit lokaler Ausschnitte der WS-Aktivierungskarte darstellt.
Um texturelle Merkmale unabhängig von der mittleren lokalen Aktivierung (helligkeitsunabhängig) bestimmen zu können, wurde die erste, den lokalen Gleichanteil beschreibende Komponente $y_1^{(k)}$ eines jeden Fitwertvektors $\mathbf{y}^{(k)}$ von der weiteren Verarbeitung ausgeschlossen.
Analysen zeigten weiterhin, daß sowohl die positive als auch die negative Übereinstimmung eines rezeptiven Feldes mit dem korrespondierenden Analysegebiet für die qualitative, translationsinvariante Beschreibung textureller Eigenschaften gleichbedeutend ist. Somit werden im folgenden nur noch die Beträge der Komponenten $y_i^{(k)}$ eines jeden Fitwertvektors $\mathbf{y}^{(k)}$ unter Ausschluß der den lokalen Gleichanteil beschreibenden ersten Fitwertaktivierung $y_1^{(k)}$ betrachtet.

$$y_i^{(k)} = \| \, \mathbf{w}^{(i)\,T} \; \mathbf{x}^{(k)} \, \| \qquad \text{mit} \qquad i = 2, ..., 45; \; k = 1, 2, ..., o$$

Diese Vorgehensweise bildet einen irreversiblen Verarbeitungsschritt. Ab diesem Punkt ist eine Rekonstruktion der WS-Aktivierungskarte aus den Fitwertaktivierungen nicht mehr möglich. Weitere Tests mit verschiedensten Texturen zeigten einen prinzipiell exponentiell fallenden Verlauf der Fitwertvektoren. Um die so modifizierten Fitwertaktivierungen einer Verarbeitungscolumne zu einer einzigen columnaren Aktivierung zusammenfassen zu können machte es sich erforderlich, diesen Trend zu eliminieren. Hierzu wurde für die unterschiedlichsten Texturen ein diese beschreibender mittlerer Fitwertvektor bestimmt.
Eine Normierung aller Komponenten $y_i^{(k)}$ auf die korrespondierenden Komponenten des mittleren Fitwertvektors führte dazu, daß nun jede normierte Komponente $(y_i^{(k)})_{norm}$ die Ausprägung des durch das rezeptive Feld $\mathbf{w}^{(i)}$ extrahierten spektralen Raumfrequenzanteils eines Texturverlaufs innerhalb des lokalen Analysegebietes beschreibt. Alle normierten Komponenten $(y_i^{(k)})_{norm}$ einer Verarbeitungscolumne können nach diesem Schritt zu einer einzigen columnaren Aktivierung $y_{column}^{(k)}$ zusammengefaßt werden.

$$y_{column}^{(k)} = \sum_{i=2}^{45} \, (y_i^{(k)})_{norm}$$

Alle columnaren Aktivierungen $y_{column}^{(k)}$ führen zu der angestrebten, bezüglich der Retina topographisch organisierten, texturbeschreibenden Merkmalskarte (Abbildung 4.16 (links)).
Die hier beschriebene Vorgehensweise zur Bestimmung der $y_{column}^{(k)}$ besitzt Korrespondenzen zum Verfahren der texturellen Energie [Laws 1980] [Laws 1985] [Unser 1986].
Die columnare Architektur zur empfindungsgemäßen Texturbeschreibung wurde anhand verschiedenster, natürliche Szenen darstellender WS-Aktivierungskarten getestet und erwies sich dabei als ausgesprochen robust.

Abbildung 4.25: *(unten) Darstellung zweier WS-Aktivierungskarten natürlicher Farbszenen. Beide wurden dem beschriebenen Verfahren zur columnaren, eigenvektorbasierten Texturanalyse unterzogen. Die resultierenden texturbeschreibenden Merkmalskarten sind oberhalb abgebildet. Regionen gleicher Texturempfindung besitzen gleiche Aktivierungen* $y_{column}^{(k)}$ *und somit gleiche Grauwerte. Dies trifft z.B. auf die Kieselsteine des linken Bildes als auch auf die Baumrinde des rechten Bildes zu. Zum besseren visuellen Vergleich wurde nachträglich eine texturelle Region des linken Bildes in das rechte integriert.*

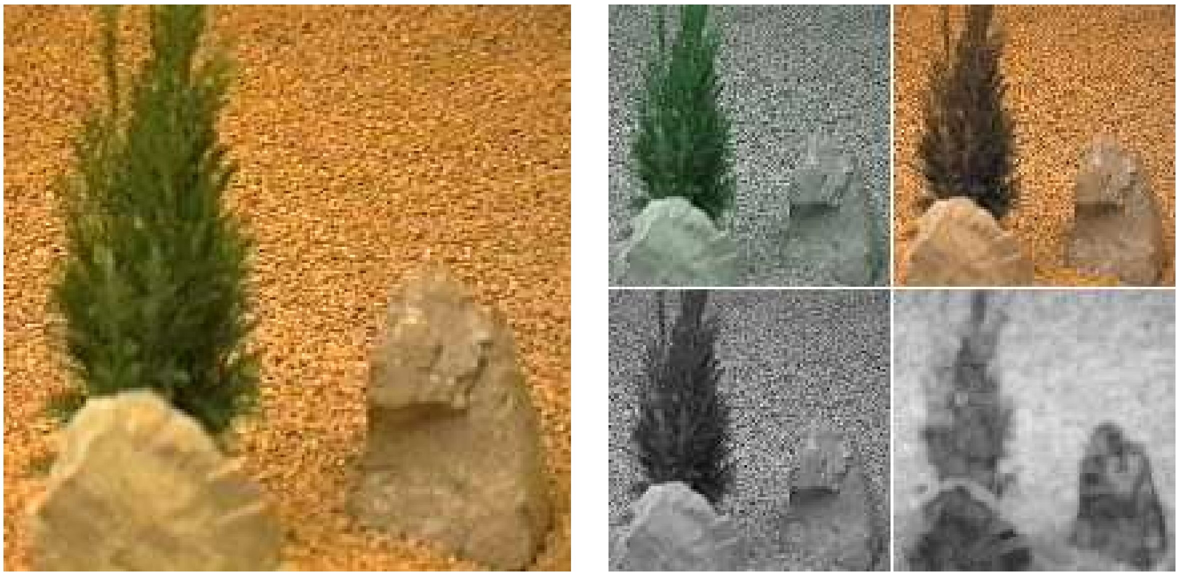


Abbildung 4.26: *Die für die links dargestellte Szene bestimmten Merkmalskarten RG, BY, WS sowie Textur. Hier wird offensichtlich, daß eine Segmentierung der Steine vor dem Kies lediglich auf der Grundlage der Texturkarte möglich ist.*

Typische texturbeschreibende Merkmalskarten sind in Abbildung 4.25 (oben) dargestellt.
Die columnaren Aktivierungsamplituden $y^{(k)}_{column}$ und damit die empfindungsgemäße texturelle Rauhigkeit wird durch den Grauwert kodiert. Ein zunehmender Grauwert beschreibt eine ausgeprägtere texturelle Rauhigkeit.
Hierbei ist zu beachten, daß die dargestellten texturbeschreibenden Merkmalskarten aus einer columnaren Analyse mit einem rezeptiven Feldüberlapp von 75% hervorgehen. Dies ist zum einen biologisch motiviert und führt zum anderen zu homogenen Aktivierungsverläufen innerhalb der Merkmalskarten.
Erst diese Merkmalskarten gestatten eine Segmentierung von Regionen, die sich weder farblich noch aufgrund des mittleren lokalen Grauwertes, sondern ausschließlich anhand textureller Eigenschaften segmentieren lassen (Abbildung 4.26). Somit kommt der texturbeschreibenden Merkmalskarte eine den RG-, BY- und WS-Karten gleichgestellte Bedeutung zu.
Hier soll nochmals darauf hingewiesen werden, daß diese Merkmalskarten nur auf der Grundlage lokaler Wechselwirkungen entstanden sind.

4.8 Identifikation lokaler Bildgebiete

4.8.1 Grundsätzliche Überlegungen

In Kapitel 2 wurde bereits vermutet, daß der IT die neurobiologische Struktur darstellen könnte, die zur Repräsentation eines elementaren Alphabets komplexer visueller Muster geeignet sei [Fujita et al. 1992] [Rolls et al. 1995]. Für das Projektvorhaben war es notwendig, einen möglichst schnellen Übergang von den szenenbeschreibenden Merkmalskarten zu einem endlichen Alphabet komplexer Merkmale zu realisieren, die dann die Basis zur Selbstorganisation des emergenten, verhaltensorientierten Objektverständnisses bilden sollten.

In der vorgestellten Gesamtarchitektur wird dieser Übergang durch den *Attentional Focus Identifier* realisiert (siehe auch Kapitel 3, Abbildungen 3.1 und 3.2), ein Modell, das funktionell im inferotemporalen Cortex (IT) angesiedelt sein könnte.

4.8.2 Attentional Focus Identifier (AFI)

Setzt man als Struktur des IT eine topologische Karte komplexer Merkmale an, was aufgrund der experimentellen Ergebnisse [Fujita et al. 1992] [Rolls et al. 1995] durchaus gerechtfertigt erscheint, und fordert zumindest biologische Plausibilität bei der Realisierung des Übergangs von den Merkmalskarten der V1 zum IT, so bietet sich das Prinzip der topologieerhaltenden Featuremaps, wie es für verschiedene sensorische Regionen des Cortex nachgewiesenermaßen existiert, geradezu an [Ritter et al. 1991]. Bei den Überlegungen zur Dimensionalität des Basisalphabets komplexer Merkmale spielten praktische Erwägungen die entscheidende Rolle. So mußte gesichert werden, daß sich die Netzwerkarchitektur zur weiteren Verarbeitung des Basisalphabets hinsichtlich ihrer Komplexität in Grenzen hielt, um die Untersuchungen zur Selbstorganisation des Objektverständnisses in den Vordergrund stellen zu können. Vor diesem Hintergrund erschien es sinnvoll, den mehrdimensionalen Merkmalsvektor, der ein lokales Bildgebiet beschreibt, auf einen eindimensionalen Vektor komplexer Merkmale abzubilden.

Reduktion der Dimension der Aktivierungs- und Merkmalskarten

Die gewonnenen Aktivierungskarten werden zu diesem Zweck zunächst in drei Gaußpyramiden [Jaehne 1989] weiterverarbeitet, wobei sich die Bildgröße auf 64x64 Pixel reduziert. Die Gaußpyramide stellt eine Form der Realisierung eines Stack-Modells (z.B. [Burton et al. 1986] [Olshausen 1994]) (siehe auch Abschnitt 2.3.2) dar. Damit wird einerseits eine Reduktion der Datenmenge erreicht, andererseits werden geringe Schwankungen der Merkmalsausprägungen ausgeglichen. Damit wird eine wesentliche Forderung erfüllt, die sich aus der Arbeitsweise des Agenten ergibt (siehe Kapitel 3). Das Shannon'sche Abtasttheorem bildet die theoretische Grundlage für die Gaußpyramide. Mit Hilfe einer Glättungsoperation wird die im Bild vorhandene Frequenz reduziert, so daß sich die höchste Frequenz halbiert. Danach kann ohne Informationsverlust die Größe des Bildes halbiert und dieses Verfahren iterativ fortgesetzt werden.

Hiermit ergibt sich eine Pyramidenstruktur der Bildrepräsentationen (Abbildung 4.27). Als Glättungsoperator wurde ein Binomialfilter (${}^nB$) der Ordnung $n = 2$ verwendet [Jaehne 1989]). Dieser Filter stellt eine diskrete Approximation der Gaußfunktion dar. Der Filter entspricht damit den typischen rezeptiven Filtern im Cortex (siehe auch [Marr 1982]). Dabei ergibt sich, daß durch die Glättungsoperation nicht nur der direkt unter einer Position der Pyramidenebene liegende 2x2-Bildauschnitt die Werte dieser Ebene beeinflußt, sondern auch die Umgebung in die Berechnung mit einbezogen wird. Dies wird noch im Zusammenhang mit dem Attention Control Network (ACN) eine Rolle spielen (siehe Kapitel 5).
Bezeichnet man die Halbierung des Bildes als Reduktionsoperator R, so ergibt sich für die $(i+1)$te Pyramidenebene folgender Formalismus:

$$G^{(i+1)} = (R^nB)^{(i)}G^{(i)}$$

Da das Originalbild die Dimension von 256x256 Pixel aufweist, wird nach drei Pyramidenebenen eine Bildgröße von 64x64 erreicht. In Korrespondenz zu den Farbkarten (Rot-Grün, Gelb(Yellow)-Blau, Helligkeit) ergibt sich somit pixelweise ein zunächst drei-dimensionaler Merkmalsvektor ($\underline{m}^*$). Zu diesem Merkmalsvektor wird der entsprechende Texturwert für die korrespondierende Bildregion hinzugefügt, sodaß sich letztendlich ein 4-dimensionaler Merkmalsvektor ($\underline{m}$) ergibt. Die Abbildung 4.27 stellt das Ergebnis der Verarbeitung der optimierten (nach der Farbumstimmung entstandenen) Helligkeitskarte durch die Gaußpyramide dar.

Attentional Focus Identifier (AFI)

Ziel des Attentional Focus Identifier ist es, den Inhalt einer lokalen Bildregion, des Aufmerksamkeitsfensters, in eine lokale Repräsentation abzubilden. Der Inhalt des Aufmerksamkeitsfensters wird durch den oben genannten 4-dimensionalen Merkmalsvektor beschrieben.
Für diese Abbildung wurde aus mehreren Gründen eine Kohonen Feature Map verwendet: Erstens korrespondieren die beiden Focus Identifier AFI und TFI (siehe Abschnitt 5.2.4) zum inferotemporalen Cortex und die Kohonen Feature Map hat sich bereits bei der Modellierung von kortikalen Strukturen als geeignet erwiesen [Ritter et al. 1991]. Zweitens ist sie in der Lage, eine lokale Repräsentation zu erzeugen, die möglichst gut das verwendete Bildmaterial beschreibt. Zu diesem Zweck wird die Fähigkeit der Kohonen Feature Map genutzt, die Eingangsdaten mit Hilfe der Gewichtsvektoren zu approximieren [Ritter et al. 1991]. Dies führt dazu, daß die Kohonen Feature Map den Eingangsvektorraum so diskretisiert, daß der Diskretisierungsfehler bezüglich der Statistik des Eingangsvektorraums relativ klein wird. Damit verringert sich ebenfalls der Fehler, der durch die lokale Repräsentation entsteht. Wird nun die Kohonen Feature Map mit dem Ergebnis der frühen visuellen Verarbeitung zu den verwendeten Bilddaten trainiert, kodiert die lokale Repräsentation das Bildmaterial relativ gut.

Da sich die lokale Repräsentation als eine eindimensionale Kodierung verstehen läßt, wurde eine eindimensionale Topologie gewählt. Es wird eine Abbildung von einem Eingangsvektor $\underline{x} \in \mathbb{R}^n$ auf einen Ausgangsvektor $\underline{y} \in \mathbb{R}^m$ mit Hilfe einer Gewichtsmatrix

Abbildung 4.27: *Ergebnis der Gaußpyramide für die Helligkeitskarte*

$W \in \mathbb{R}^n \times \mathbb{R}^m$ realisiert:

$$y_i = \begin{cases} 1 & ||\underline{w}_i - \underline{x}|| = \min_{j=1\ldots m} ||\underline{w}_j - \underline{x}|| \\ 0 & \text{sonst} \end{cases}$$

Die verwendete Norm $||\cdot||$ ist die Euklidische Metrik. Das i-te Neuron ist das „Gewinner-Neuron“ des AFI, dessen Gewichtsvektor $\underline{w}_i$ den geringsten Abstand zum Eingangsvektor $\underline{x}$ hat. Das Training der Kohonen Feature Map erfolgt nach [Ritter et al. 1991]:

$$\underline{w}_j^{neu} = \underline{w}_j^{alt} + \epsilon(t) h_{ji}(t)(\underline{x} - \underline{w}_j^{alt})$$

Wobei die Nachbarschaftsfunktion h_{ji} die bekannte Form hat:

$$h_{ji} = e^{-\frac{(i-j)^2}{2\sigma^2}(t)}$$

Die Lernrate ϵ und der Lernradius σ nehmen während des Lernens mit der Anzahl der Lernschritte q ab:

$$\epsilon = \epsilon_0 \left(\frac{\epsilon_1}{\epsilon_0}\right)^{\frac{t}{t_{\max}}} ; \sigma = \sigma_0 \left(\frac{\sigma_1}{\sigma_0}\right)^{\frac{t}{t_{\max}}}$$

Dabei sind ϵ_0 und σ_0 die Initialwerte und ϵ_1 und σ_1 die Endwerte. $t_{\max}$ ist die Anzahl der Lernschritte.
Da die frühe visuelle Verarbeitung 4 Merkmalskarten liefert, ergibt sich ein 4-dimensionaler Eingangsvektorraum. Die Ausgangsdimensionalität wurde empirisch ermittelt und auf 25 gesetzt. Die Anfangswerte der Gewichtsvektoren wurden zufällig gewählt. Die lokalen Merkmalsvektoren aller Szenen der Galerie wurden als Trainingsdaten verwendet. Die Abbildung 4.28 gibt zwei Beispiele für ein Diskretisierungsergebnis.

4.8.3 Auswirkung der Farbumstimmung auf die Identifikation lokaler Bildgebiete

Die durch die realisierte Farbumstimmung erzielbare Stabilität der Farbmerkmale läßt sich sehr anschaulich anhand des Identifikationsergebnisses für farbkorrigierte und nicht farbkorrigierte Szenen demonstrieren. Die Abbildung 4.29 stellt exemplarisch anhand von drei Szenen das durch den AFI erreichte Identifikationsergebnis ohne und mit erfolgter Farbumstimmung gegenüber. Die obere Reihe in Abbildung 4.29 zeigt die drei Ausgangsszenen, von denen die mittlere Szene ausgewogen beleuchtet wurde, die Szenen rechts und links dagegen unter deutlich unausgewogenen Beleuchtungsverhältnissen aufgenommen wurden. Die mittlere Reihe zeigt das Identifikationsergebnis des Focus Identifier für den Fall, daß keine Farbumstimmung der Eingangsszenen erfolgte. Deutlich zu erkennen ist die große Variabilität des Identifikationsergebnisses für alle drei unkorrigierten Szenen. Diese Variabilität würde für den angestrebten Selbstorganisationsprozeß eine enorme Schwierigkeit darstellen, da für gleiche Szenenbestandteile unter variablen Randbedingungen (Beleuchtungsverhältnisse) keine zumindest annähernd identischen komplexen Merkmale durch den Focus Identifier zur Verfügung gestellt

Abbildung 4.28: *Die Abbildung zeigt rechts in Form von Grauwerten das Ergebnis der lokalen Repräsentation des Attentional Focus Identifier für die links dargestellte Beispielszene aus der Galerie.*

werden könnten. Demgegenüber zeigen die Identifikationsergebnisse der unteren Reihe in Abbildung 4.29, bei denen vor der Transformation der Merkmalsvektoren zu einem komplexen Merkmal eine Farbkorrektur erfolgte, im Vergleich zu den Ergebnissen mit den unkorrigierten Szenen eine deutlich geringere Variabilität, sodaß hier von annähernd stabilen komplexen Merkmalen für den Selbstorganisationsprozeß ausgegangen werden kann. Diese Stabilität bei der Identifikation der lokalen Bildgebiete, die erst durch den Mechanismus der Farbumstimmung gewährleistet werden konnte, stellt nach unserer Auffassung ein wesentliches Ergebnis des Projektvorhabens dar (siehe auch Kapitel 6).

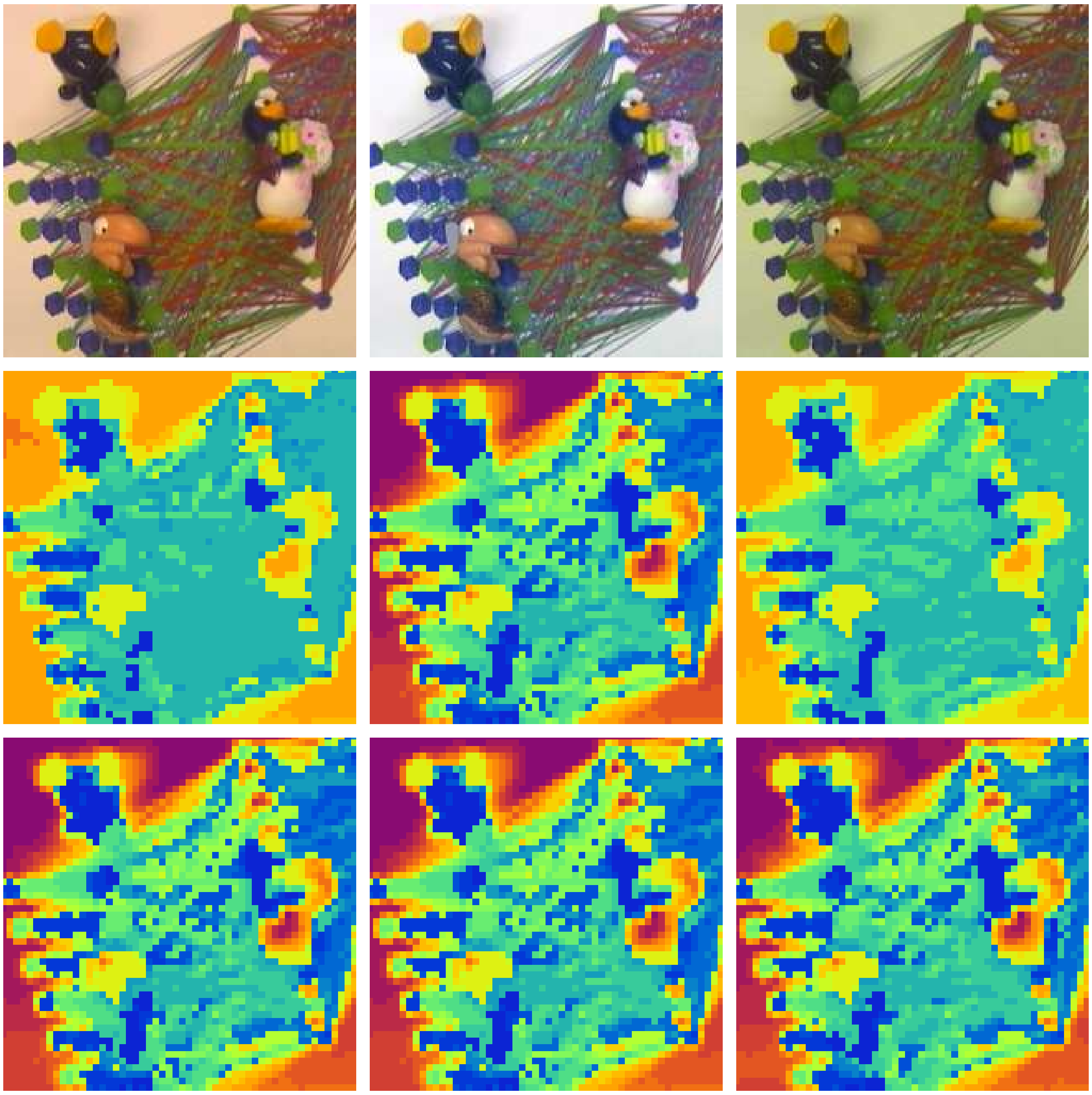


Abbildung 4.29: *Identifikationsergebnisse für Szenen, die unter unausgewogenen (oben links und rechts) und unter ausgewogenen (oben mitte) Beleuchtungsverhältnissen aufgenommen wurden. Die mittlere Reihe zeigt die Ergebnisse des Focus Identifier, ohne daß die Eingangsszenen einer Farbkorrektur unterworfen wurden. Die untere Reihe zeigt die Ergebnisse des Focus Identifier für die drei farbkorrigierten Eingangsszenen (weitere Eräuterungen im Text).*

Kapitel 5

Selbstorganisation eines verhaltensorientierten Objektverständnisses

Die Abschnitte dieses Kapitels beschäftigen sich mit den Teilen der Gesamtarchitektur, die die beiden attentiven Ebenen des Modells bilden. Außerdem wird dargelegt, wie aus den durch die frühe visuelle Verarbeitung bereitgestellten Merkmalskarten eine Auffälligkeitskarte berechnet werden kann, die die Basis für das zunächst rein datengetriebene Abtastverhalten bildet.

Die interne Repräsentation der Übergänge zwischen im Abtastprozeß unmittelbar aufeinanderfolgenden komplexen Merkmalen und deren räumlichen Relationen im **What**- und **Where**-System bildet die erste attentive Ebene und liefert die entscheidende strukturelle und funktionelle Grundlage für alle weiteren Lernprozesse innerhalb der Gesamtarchitektur. Die Generierung von lokalen Hypothesen über im Abtastprozeß unmittelbar aufeinanderfolgende lokale Bildstrukturen und deren Verifikation in Form einer wissensbasierten Modulation des anfänglich rein datengetriebenen Abtastprozesses stellt die erste Stufe des emergenten, verhaltensorientierten Objektverständnisses dar.
Einen weiteren Schwerpunkt bildet im Rahmen der ersten attentiven Ebene die Koordination von präattentiver und attentiver Verarbeitung.

Die sich daran anschließende Architektur, das **When**-System, das mehrere, im Abtastprozeß unmittelbar aufeinanderfolgende komplexe Merkmale repräsentieren und zur Bildung von globalen Hypothesen über Objekt- bzw. Teilobjektkandidaten heranziehen kann, stellt die zweite attentive Ebene des Gesamtsystems dar. Die antizipatorische Hypothesenbildung dieser höchsten Verarbeitungsebene und der mit der Hypothesenverifikation verbundene, in Wechselwirkung mit der ersten attentiven Ebene stattfindende, weitere wissensbasierte Eingriff in den Analyseprozeß führt dann zur eigentlichen Organisation von Wissen über komplexe visuelle Strukturen im Sinne von Objekten bzw. Teilobjekten.

Die beiden attentiven Ebenen der Modellarchitektur werden deshalb relativ separat behandelt, weil sie zum einen jeweils in sich abgeschlossene funktionelle Einheiten bilden und darum auch bis zu einem gewissen Grad unabhängig voneinander entwickelt wurden und weil zum anderen die Motivation für die zweite attentive Ebene nur plausibel wird, wenn die erste attentive Ebene zunächst ohne deren Einfluß dargestellt wird.
Zur besseren Überleitung wird an dieser Stelle die Abbildung 3.2 noch einmal dargestellt.

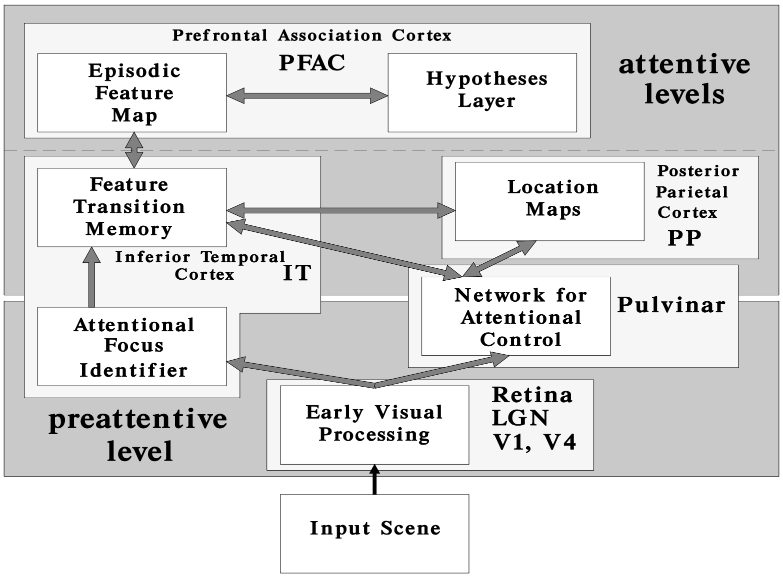


Abbildung 5.1: *Abbildung der psychophysischen Grundlagen in die Modellarchitektur*

5.1 Erste attentive Ebene

In diesem Abschnitt wird zunächst die Struktur der ersten attentiven Ebene des funktionalen Modells aus Kapitel 3 mit dem formal definierten Verhalten des Agenten verknüpft. Anschließend wird ein Überblick über das Verhalten der gesamten Modellarchitektur der ersten attentiven Ebene gegeben. Im Verlauf dieser Abschnitte werden weitere Fragen, die sich im Rahmen des Abschnitts zum funktionalen Modell stellten, beantwortet und abschließend noch einmal zusammengefaßt.

5.1.1 Allgemeine Vorbemerkungen

Die Abbildung 5.2 gibt einen Überblick über die Struktur der neuronalen Modellarchitektur der ersten attentiven Ebene.
Die Prozesse der frühen visuellen Verarbeitung wurden in Kapitel 4 umfassend dargestellt. Der Begriff des **What**-Systems umfaßt im folgenden nur noch die Verarbeitungs-

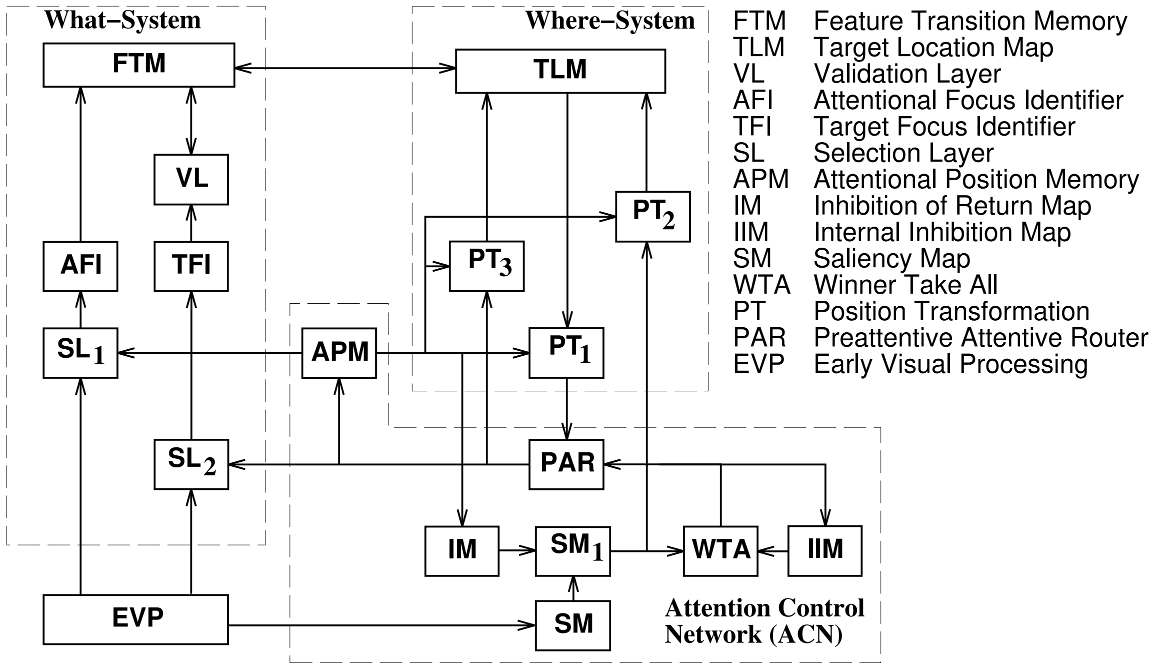


Abbildung 5.2: *Überblick über die Struktur der neuronalen Modellarchitektur der ersten attentiven Ebene.*

ebenen nach der frühen visuellen Verarbeitung. Gleiches gilt für die weitere Verwendung des Begriffs des **Where**-Systems (siehe auch Abbildung 5.2).

In der ersten attentiven Ebene des funktionalen Modells werden Merkmalsinformationen (What-System) und räumliche Relationen (Where-System) repräsentiert. Diese Informationen korrespondieren mit dem Umweltwissen des Agenten, das aus räumlichen Relationen zwischen bekannten lokalen komplexen Merkmalen besteht. Damit stellt die erste attentive Ebene eine optimale Basisstruktur dar, in der das Umweltwissen gespeichert wird. Dabei werden die komplexen Merkmale im What-System und deren räumliche Relationen im Where-System repräsentiert. Durch die Kooperation der beiden Systeme ergibt sich das vollständige Umweltwissen.

Das Attention Control Network (ACN) bildet u.a. den Ort der Auffälligkeitskarte. Damit erscheint es plausibel, dieser Struktur auch die Erzeugung der auffälligkeitsbasierten Sprünge funktionell zuzuordnen. Diese Erzeugung erfolgt aber in enger Kooperation mit den wissensbasierten Abtastbewegungen des Agenten bzw. der attentiven Ebene. Die Koordination von auffälligkeitsbasierter und wissensbasierter Abtastung erfolgt ebenfalls im ACN, weshalb dieses im funktionalen Modell (Kapitel 3) auch als Schnittstelle zwischen attentiver und präattentiver Ebene bezeichnet wurde.
Die Positionsinformation wird in Anlehnung an die Modellvorstellungen zum PP bzw. SC in topologischer Form repräsentiert, d.h. es existieren Positionskarten, deren Neuronen die (x, y)-Koordinaten der Szenen kodieren. Dies wird in den folgenden Abschnitten

noch präzisiert.

5.1.2 Frühe visuelle Verarbeitung

Die frühe visuelle Verarbeitung erzeugt eine Merkmalsrepräsentation der visuellen Eingangsszene, hier beispielhaft in Form von 4 Merkmalskarten (2 Farbkarten, Helligkeitskarte, Texturkarte). Das Ergebnis dieser Verarbeitungsstufe bildet den Eingang zum What-System (siehe Abschnitt 4.8.2) und zum ACN. Damit erzeugt die frühe visuelle Verarbeitung einerseits die Merkmalsvektoren für das Umweltwissen des Agenten und andererseits legt es die Grundlage für die Erzeugung der Auffälligkeitskarte (SM in Abbildung 5.2). Die in Kapitel 3.1.3 definierte Lernstrategie des Agenten setzt voraus, daß sowohl die komplexen Merkmale als auch deren räumliche Relationen in Bezug auf die Objekte möglichst stabil sind, d.h. nur geringen Schwankungen unterliegen. Diese Robustheitsanforderungen erfüllt zum einen die frühe visuelle Verarbeitung mit Hilfe der bereits behandelten Verarbeitungsmechanismen (physiologischer Farbraum, adaptive Farbumstimmung, Gaußpyramide), zum anderen sichern die vorab definierten Randbedingungen bezüglich des Szenarios (keine Skalierungen, Rotationen, perspektivische Verzerrungen) die räumliche Stabilität.
Bezüglich der Auffälligkeitskarte wurden die in der Psychophysik gefundenen qualitativen, experimentellen Befunde berücksichtigt. Da es nicht vordergründig um die Entwicklung eines möglichst umfassenden Aufmerksamkeitsmodells ging, wurden nicht alle diese Ergebnisse, sondern lediglich eine Auswahl von grundlegenden, insbesondere projektrelevanten Befunden, aufgegriffen.

5.1.3 What-System

Das What-System (Abbildung 5.2 und detailliert Abbildung 5.3) des funktionalen Modells stellt u.a. den Träger der zeitlichen Objektrepräsentation dar. Da diese Objektrepräsentation durch das Verhalten B des Agenten (Abschnitt 3.1.3) bestimmt wird, ist es sinnvoll, anzunehmen, daß das What-System das Verhalten B des Agenten (Suche nach einem erfolgreichen Sprung auf der Grundlage des Inhalts des aktuellen Aufmerksamkeitsfensters) realisiert. Der Inhalt des aktuellen Aufmerksamkeitsfensters wird ebenfalls im What-System mit Hilfe der Selection Layer (SL_1) extrahiert (4-dimensionaler Merkmalsvektor, bestehend aus 2 Farbmerkmalen, einem Textur- und einem Helligkeitsmerkmal).
In Anlehnung an die lokale Repräsentation komplexer Merkmale im IT (What-System) wird dieser Merkmalsvektor durch den Attentional Focus Identifier (AFI) in eine lokale Repräsentation abgebildet (siehe Abschnitt 4.8.2 in Kapitel 4). Damit wird die in Abschnitt 3 geforderte Diskretisierung zur Graphenbeschreibung eines Objekts realisiert. Somit bildet der AFI den ersten Schritt zu einer Speicherung des Umweltwissens.

Auf der Grundlage dieser Kodierung realisiert nun der Feature Transition Memory (FTM) das Verhalten B des Agenten, nämlich die Suche nach einem erfolgreichen Sprung. Dazu bildet jedes Neuron des AFI den Eingang zu einem Neuronen-Cluster

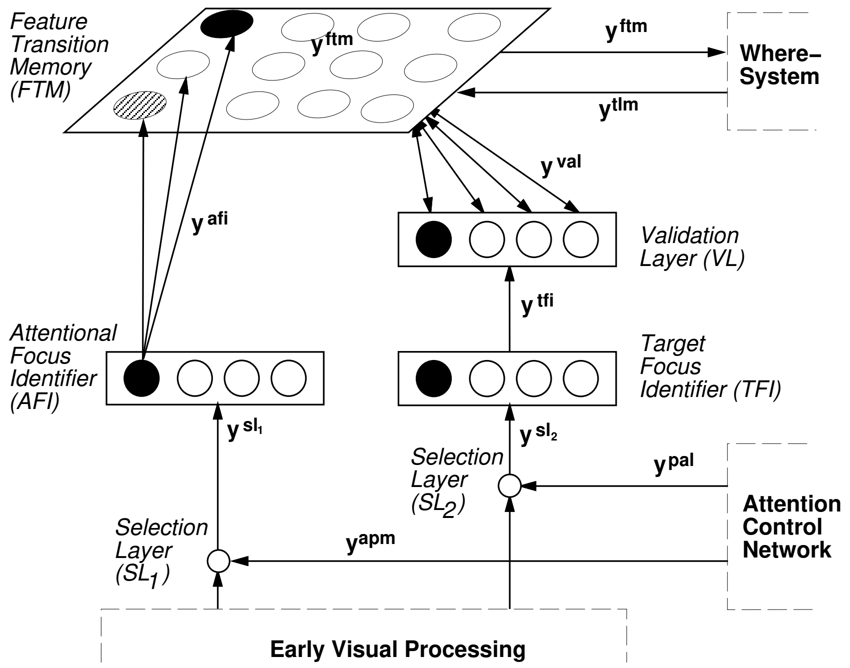


Abbildung 5.3: *What-System*

des FTM (entspricht einer Spalte im FTM in Abbildung 5.3). Jeder dieser Eingänge wird mit einem Gewicht bewertet. Diese Gewichte korrespondieren mit den Objektzugehörigkeitswerten (α_i). Die Cluster des FTM ergeben sich aus der Möglichkeit, daß ein Merkmalsvektor m_i mit mehreren Folgemerkmalsvektoren an unterschiedlichen relativen Positionen (lokale Sprunghypothesen) verbunden sein kann. Ein FTM-Cluster i kodiert damit alle Sprünge, die von einem Merkmalsvektor m_i im bisherigen Interaktionsprozeß mit der Umwelt ausgingen (bislang erworbenes Umweltwissen).

Ein WTA-Mechanismus selektiert ein FTM-Neuron. Dieses Neuron bewirkt über das Where-System, das Attention Control Network (ACN) und die Selection Layer SL_2 eine hypothetische Verschiebung des „Focus of Attention“ und damit die Extraktion eines weiteren Merkmalsvektors m_i^* aus dem Zielgebiet dieses Sprungs. Der Target Focus Identifier (TFI), dessen Aufbau identisch mit dem des AFI ist, erzeugt aus diesem Merkmalsvektor wiederum eine lokale Repräsentation in Form eines komplexen Merkmals. Parallel dazu erzeugt das selektierte FTM-Neuron eine Hypothese über den zu erwartenden Inhalt dieses Aufmerksamkeitsfensters in der Validation Layer (VL). Die VL überprüft, ob die beiden Repräsentationen übereinstimmen, d.h. ob ein erfolgreicher Sprung vorliegt oder nicht. Stimmen erwarteter und tatsächlicher Inhalt des Zielgebiets nicht überein, so liegt kein erfolgreicher Sprung vor. Dies führt dazu, daß die Aktivität des FTM-Neurons, welches für die Ausführung des gerade getesteten Sprunges verantwortlich war, zusammenbricht und ein neues FTM-Neuron wiederum über einen WTA-Mechanismus selektiert wird. Jedes FTM-Neuron steht damit für ein Tupel des Umweltwissens des Agenten. Dazu werden die Merkmalsvektoren (m_i und m_i^*) der Tupel im AFI und TFI gespeichert und die Gewichte zwischen AFI und FTM speichern die

Objektzugehörigkeitswerte. Das Gesamtverhalten des WTA-Mechanismusses kann als Max-to-Min-Verhalten (M2M-Verhalten) zusammengefaßt werden. Mit diesem M2M-Verhalten wird also das sukzessive Aktivieren von FTM-Neuronen bezeichnet, wobei sich die Reihenfolge der Aktivierungen aus den Aktivitätswerten ergibt.
Die rekurrente Verbindung zwischen VL und FTM dient der Erweiterung des Umweltwissens. Deren genaue Funktionalität wird beim Überblick über das gesamte Verhalten der ersten attentiven Ebene diskutiert.
Die Lernregeln, die im Zusammenhang mit dem Agenten formuliert worden sind (siehe Abschnitt 3.1.3) und auf die in einem gesonderten Abschnitt noch eingegangen wird, müssen aufgrund des lokalen Charakters dieser Lernregeln ebenfalls im FTM realisiert werden.

5.1.4 Where-System

Das Where-System verarbeitet Positionsinformationen, die in topologischer Art und Weise kodiert werden. Es werden relative Positionsinformationen (räumliche Relationen), also Positionen bezüglich der aktuellen Position des Aufmerksamkeitsfensters, und absolute Positionen (in Szenenkoordinaten) repräsentiert.
Den Kern des Where-Systems bildet die Target Location Map (TLM). Einen Eingang zur TLM bildet der FTM, mit dessen Neuronen die Merkmalsübergänge (m_i, m_i^*) des Umweltwissens verknüpft sind. Über die Verbindungen zwischen FTM und TLM erfolgt eine Speicherung der zu einem Merkmalsübergang gehörenden Positionsänderung $(\Delta x, \Delta y)$. FTM-Neuronen können also mit Hilfe dieser Verbindungen in der TLM eine Aktivität erzeugen, die der räumlichen Relation bezüglich eines bestimmten Merkmalsübergangs entspricht. Damit kodiert die Aktivität eines TLM-Neurons die relative Zielposition bei der Suche des FTM nach einem erfolgreichen Sprung. Dies entspricht der angenommenen Funktionalität der Neuronen im postparietalen Cortex (PP) (siehe auch Kapitel 2).

Da bei der Suche nach einem erfolgreichen Sprung die in der TLM kodierte relative Position in eine szenenadäquate absolute Position transformiert werden muß, wird die Aktivität der TLM durch die Positionstransformation (PT_1) in eine absolute Position (in der aktuell fovealisierten Szene) umgerechnet.
Die Verbindungen zwischen FTM und TLM fügen also der Merkmalsübergangsinformation die entsprechende Positionsinformation hinzu. Mit den entgegengesetzt gerichteten Verbindungen von der TLM zum FTM (Rückkopplung der FTM-TLM-Verbindungen) wird die Modulierung der FTM-Suche durch die Auffälligkeitskarte gewährleistet. Dazu wird die Auffälligkeitskarte (SM_1) des ACN (siehe folgenden Abschnitt) mit Hilfe der Positionstransformation PT_2 in eine relative Auffälligkeitskarte transformiert. Dieses Ergebnis bildet einen weiteren Eingang zur TLM. Diese Aktivität bewirkt dann durch die Verbindungen zum FTM eine Modulierung der FTM-Suche nach einem erfolgreichen Sprung.

Damit ergibt sich eine doppelte Funktionalität der Aktivitätskodierung innerhalb der TLM. Einerseits kodiert die Aktivität die relative Auffälligkeitskarte und andererseits

mögliche Zielpositionen der Suche des FTM. Dies macht es notwendig, bei der Bestimmung der Aktivierungen für die TLM-Knoten ein spezielles Aktivitätskodierungskonzept zu verwenden, um die verschiedenen Einflüsse noch unterscheiden zu können. Näheres hierzu findet sich im Rahmen der mathematischen Beschreibung der TLM.
Der von der Positionstransformation PT_3 kommende Eingang zur TLM dient der Erweiterung des Umweltwissens. Dessen genaue Funktionalität wird beim Überblick über das gesamte Verhalten der ersten attentiven Ebene erläutert.

5.1.5 Attention Control Network (ACN)

Bei der Entwicklung der Modellarchitektur wurde davon ausgegangen, daß das Attention Control Network die Erzeugung der Auffälligkeitskarte (Saliency Map (SM)) und die auffälligkeitsbasierte Selektion einer Position (WTA) realisiert und gleichzeitig die Schnittstelle zwischen attentiver und präattentiver Ebene bildet.
Aufgrund der Existenz der Auffälligkeitskarte im ACN liegt es nahe, die Funktionalität zur Erzeugung der auffälligkeitsbasierten Abtastbewegungen ebenfalls dem ACN zuzuordnen. Dazu ist es notwendig, bestimmte, bereits analysierte Bereiche in der Auffälligkeitskarte, zu markieren. Die Aktivität zur Markierung liefert die Inhibition of Return Map (IM). In ihr werden alle Positionen gespeichert, zu denen tatsächlich ausgeführte, d.h. erfolgreiche (wissensbasierte) oder auffälligkeitsbasierte Sprünge geführt haben. Damit realisiert die IM eine langanhaltende Inhibition, die verhindert, daß bereits analysierte Gebiete der aktuell fovealisierten Szene nochmals angesprungen werden. Der Inhalt der IM wird mit der Auffälligkeitskarte (SM) inhibitorisch zur SM_1 verknüpft. In dieser Karte sind nun alle Positionen einer Szene markiert, die von einer tatsächlich ausgeführten Abtastbewegung erreicht wurden.

Die Internal Inhibition Map (IIM) soll verhindern, daß die Selbstorganisation des Umweltwissens in ein „lokales Minimum“ konvergiert (siehe auch Abs. 3.1.3). Zu diesem Zweck speichert die Internal Inhibition Map (IIM) alle nicht ausgeführten auffälligkeitsbasierten Sprünge (siehe auch mathematische Beschreibung).
Die Arbeitsweise des Agenten macht es notwendig, die Position des aktuellen Abtastpunktes zu speichern, da von diesem aus nach einem erfolgreichen Sprung gesucht wird und die Transformation von der relativen zur absoluten Positionsinformation mit Hilfe dieser Position erfolgt. Dieses Speichern wird im Attentional Position Memory (APM) vorgenommen. Da diese Information sowohl im Where-System als auch im What-System verwendet wird und das ACN als Schnittstelle zwischen diesen beiden Systemen eine zentrale Position einnimmt, wurde der APM als Bestandteil des ACN aufgenommen.
Die Schnittstelle zwischen attentiver und präattentiver Ebene wird durch eine separate Struktur innerhalb des ACN, den Preattentive Attentive Router (PAR), realisiert, dessen Funktionalität sich aus der in Abschnitt 3.1.3 definierten Arbeitsweise des Agenten ergibt. Solange die attentive Ebene ihr Umweltwissen erfolgreich anwenden kann, muß der PAR dies ermöglichen. Ist dies nicht mehr der Fall, muß der PAR dafür sorgen, daß ein auffälligkeitsbasierter Sprung ausgeführt werden kann.
Wenn kein Umweltwissen vorhanden ist, verhält sich das ACN folgendermaßen: Der

WTA selektiert auffälligkeitsbasiert eine Position in der Szene. Diese Position wird im APM und anschließend in der IM gespeichert. Die IM bewirkt, daß der WTA eine neue Position selektiert u.s.w. Damit erzeugt das ACN eine Sequenz auffälligkeitsbasierter Sprünge.
Da die Abbruchkriterien für den Agenten auffälligkeitsbasiert bestimmt werden, werden diese ebenfalls im ACN ermittelt.

5.1.6 Überblick über das Gesamtverhalten der ersten attentiven Ebene

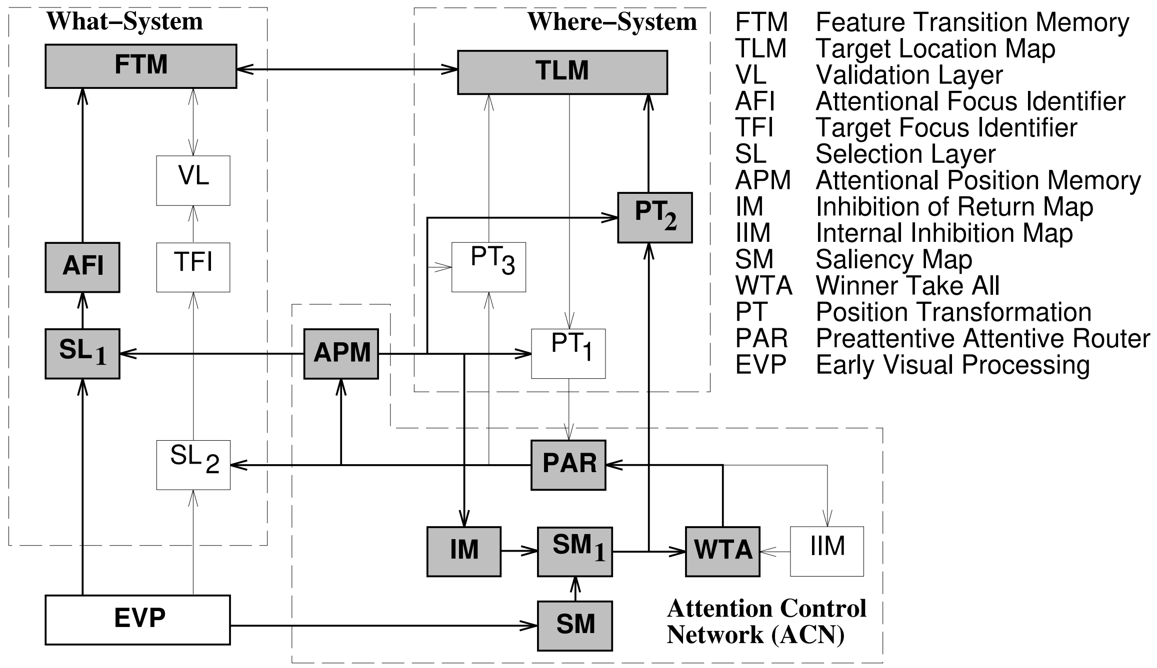


Abbildung 5.4: *Selektion der ersten Position des Aufmerksamkeitsfensters; die beteiligten Module sind schraffiert hinterlegt*

Die Abbildungen 5.4 – 5.7 geben einen Überblick über die Verhaltensweise der neuronalen Modellarchitektur der ersten attentiven Ebene. Die folgenden 4 typischen Fälle lassen sich dabei unterscheiden:

1. Selektion der ersten Position des Aufmerksamkeitsfensters (Abbildung 5.4)

2. Suche nach einem erfolgreichen Sprung (Abbildung 5.5)

3. Ausführung eines erfolgreichen Sprungs (Abbildung 5.6)

4. Ausführung eines auffälligkeitsbasierten Sprungs (Abbildung 5.7)

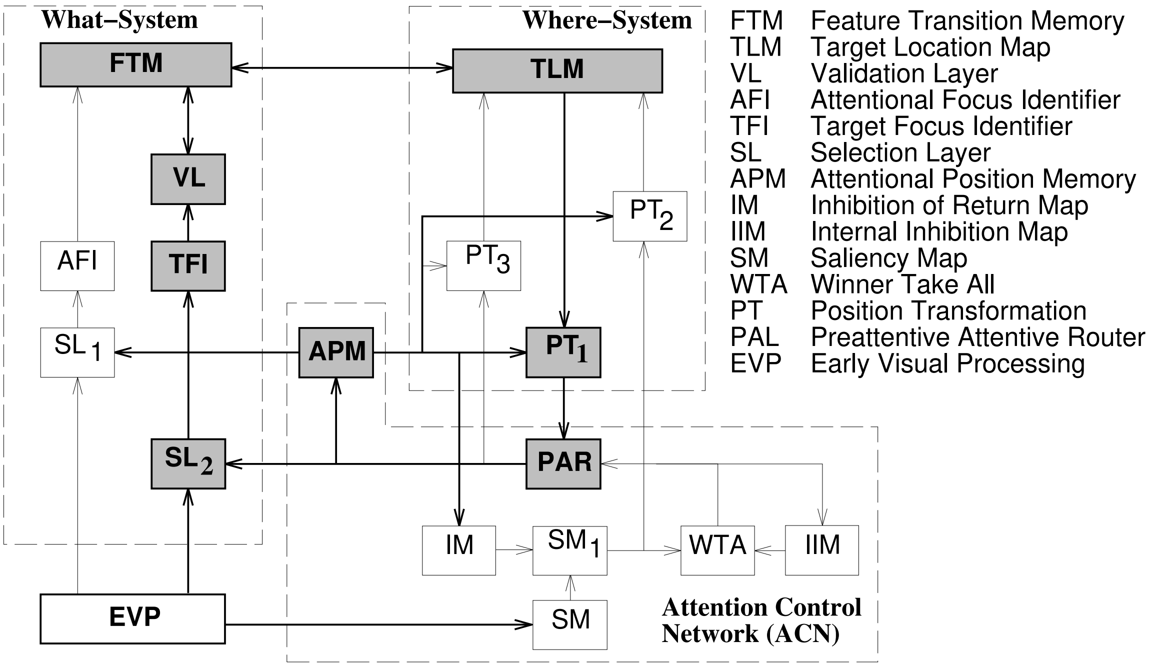


Abbildung 5.5: *Suche nach einem erfolgreichen Sprung*

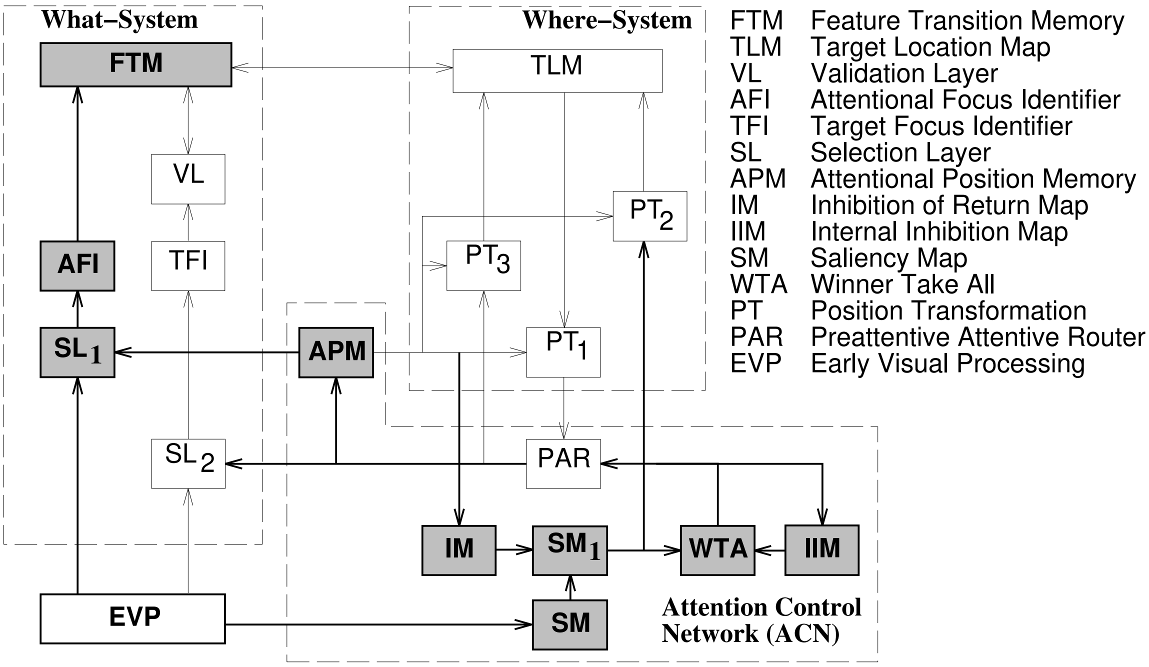


Abbildung 5.6: *Ausführung eines erfolgreichen Sprungs*

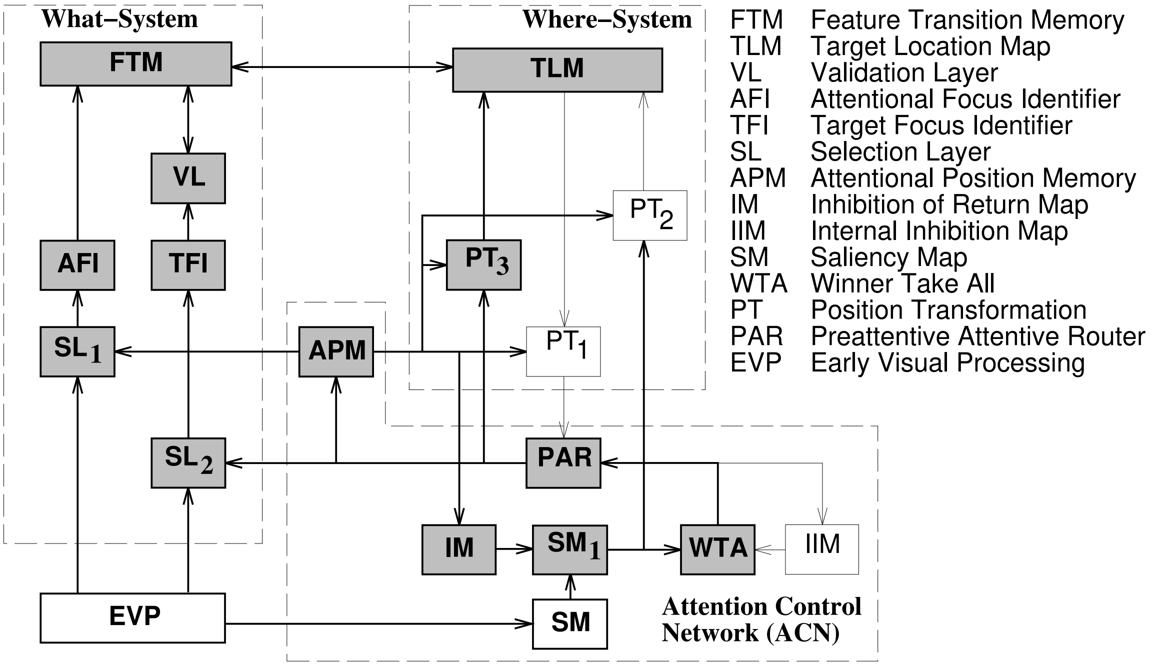


Abbildung 5.7: *Ausführung eines auffälligkeitsbasierten Sprungs, wenn ein hypothetischer attentiver Sprung nicht erfolgreich war*

In der ersten Phase (Fall 1, Abbildung 5.4) erfolgt zunächst über den WTA-Mechanismus des ACN eine Selektion der Position aufgrund der Auffälligkeitskarte (SM). Diese Position wird über den PAR in den APM und die IM eingetragen. Die Eintragung in der APM führt zu einer Selektion des Inhalts des Aufmerksamkeitsfensters (SL_1). Dieser Inhalt wird durch den AFI in eine lokale Repräsentation abgebildet. Parallel dazu wird die SM_1 über PT_2 in eine relative Auffälligkeitskarte transformiert. Diese Karte bildet den Eingang zur TLM. Die Verbindung zwischen TLM und FTM bildet die Auffälligkeitswerte der Karte auf die korrespondierenden FTM-Neuronen ab. Damit kann der WTA im FTM sowohl durch die Auffälligkeitskarte als auch durch die Objektzugehörigkeitswerte beeinflußt werden. Aufgrund der verwendeten statischen Modellneuronen ergibt sich für jede Schicht eine Verzögerung von einem Zeitschritt.

Mit Beginn der zweiten Phase (Fall 2, Abbildung 5.5) selektiert der WTA-Mechanismus im FTM ein Neuron und damit eine entsprechende Merkmalshypothese und eine Abtastbewegung. Diese Abtastbewegung wird über die TLM und die Positionstransformation PT_1 zum PAR geschickt. Da dieser Vorgang eine geraume Zeit dauert, muß der PAR die durch den WTA des ACN selektierte Position blockieren und verhindern, daß die falsche Position zum SL_2 gelangt. Damit wird es notwendig, daß der PAR drei Zustände hat: (1) Durchschalten einer wissensbasierten Position, (2) Durchschalten einer auffälligkeitsbasierten Position und (3) Blockieren einer auffälligkeitsbasierten Position, bis diese vom FTM zugelassen wird. Die SL_2 selektiert den Inhalt des Zielaufmerksamkeitsfensters. Der TFI bildet diesen entstehenden Merkmalsvektor in eine lokale Repräsentation ab. Die VL vergleicht diese mit der Hypothese des FTM.

Stimmen Hypothese und TFI-Ergebnis nicht überein, so selektiert der FTM eine neue Abtastbewegung.

Wenn ein erfolgreicher Sprung gefunden wurde, tritt Phase 3 ein (Fall 3, Abbildung 5.6). Der APM übernimmt die neue Position und trägt diese Position in die IM ein. Parallel dazu wird die vom WTA selektierte auffälligkeitsbasierte Position in die IIM eingetragen.

Wenn kein erfolgreicher Sprung gefunden wird, tritt Phase 4 ein (Fall 4, Abbildung 5.7), d.h. der FTM erzeugt in der TLM eine unspezifische, gleichmäßige Aktivität aller TLM-Neuronen und läßt so im PAR das Durchschalten einer auffälligkeitsbasierten Position zu. Diese Positionsinformation läuft über die PT_3 zur TLM und zum FTM. Parallel dazu gelangt der Inhalt des Aufmerksamkeitsfensters an dieser Position über SL_2 und TFI zum FTM. Diese Informationen aktivieren FTM-Neuronen und können so zu einer Erweiterung des Umweltwissens der attentiven Ebene führen. In diesem Zusammenhang zeigt sich auch die Notwendigkeit der beiden Focus Identifier AFI und TFI. Während der TFI benötigt wird, um die Hypothese über das komplexe Merkmal an einer bestimmten Szenenposition zu verifizieren und gegebenenfalls das Umweltwissen bei auffälligkeitsbasierten Abtastbewegungen zu erweitern, dient der AFI der Verarbeitung des aktuellen Fokusgebietes, von dem die Suche ausgeht. Bei der Erweiterung des Umweltwissens muß noch die Nachbarschaftbeziehung (siehe Abschnitt 3.1.3) berücksichtigt werden. Aus diesem Grund werden mehrere FTM-Neuronen eines FTM-Clusters (siehe insbesondere Abschnitte 5.2.4 und 5.2.6) und durch die Eingänge aus TLM und VL aktiviert. Gibt es die dazugehörigen Komponenten des Umweltwissens (Tupel aus Merkmals- und Positionsinformation) noch nicht, so werden diese durch noch nicht belegte FTM-Neuronen abgebildet, deren Auswahl zufällig erfolgt. Die Lernregeln, auf die im Abschnitt 5.2.6 eingegangen wird, müssen diese Belegung steuern.

5.1.7 Simulationsergebnisse zur ersten attentiven Ebene

Zeitliche Objektrepräsentation

Der Feature Transition Memory der ersten attentiven Ebene bildet den Träger der *zeitlichen* Objektrepräsentation, in dem er in Verbindung mit dem Where-System (insbesondere mit der Target Location Map) Übergänge zwischen beim Abtastprozeß unmittelbar aufeinanderfolgenden komplexen Merkmalen repräsentiert. Dabei werden die Merkmalsübergänge zunächst nur abgebildet. Die Spezifik der Repräsentation ergibt sich durch die aktive Einflußnahme auf das Abtastverhalten. Dieser Selbstorganisationsprozeß ermöglicht die Ausprägung von stabilen Übergängen als solche, die sich als zur Steuerung des Abtastprozesses geeignet erwiesen. Instabile Übergänge werden abgeschwächt, da sie sich nicht zur Steuerung des Abtastprozesses eignen. Das so organisierte Umweltwissen über stabil auftretende Merkmalsübergänge bildet die entscheidende Basis für die zweite attentive Ebene (siehe Abschnitt 5.3).

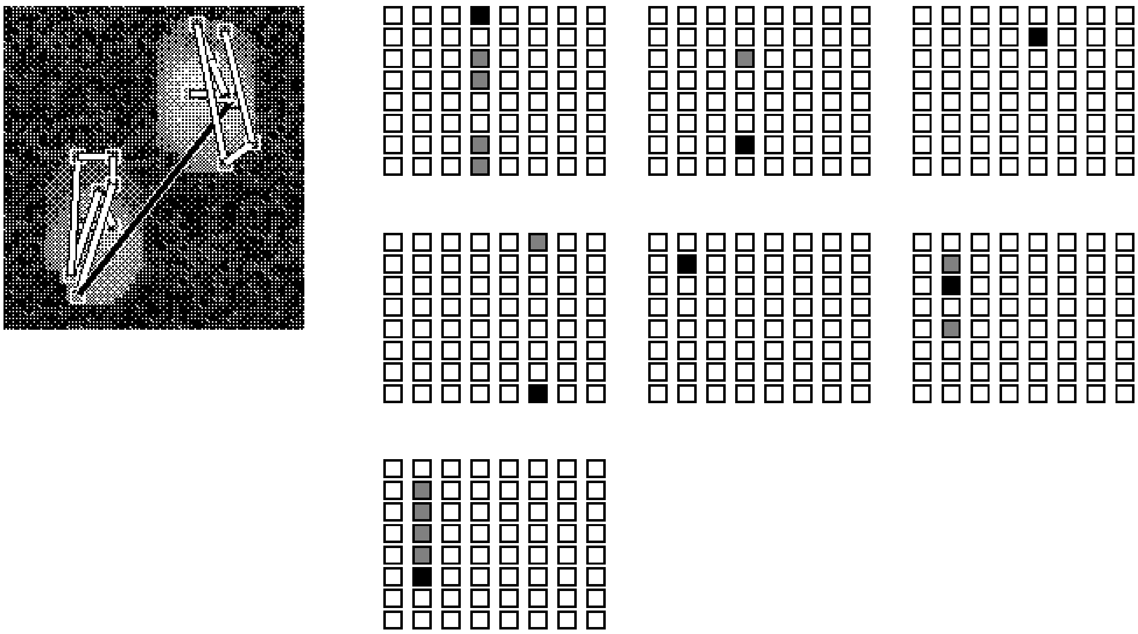


Abbildung 5.8: *Abtastverhalten auf einer synthetisch erzeugten Szene und Darstellung der zu den Abtastpunkten auf dem linken Objekt korrespondierenden FTM-Zustände (weiße Linie = wissensbasierter Sprung, schwarze Linie = erzwungener Sprung; ausführliche Erläuterung im Text)*

In den Abbildungen 5.8 und 5.9 wird das charakteristische Verhalten des FTM anhand einer synthetisch erzeugten Szene verdeutlicht. Dargestellt wurde die Aktivität der FTM-Knoten zu verschiedenen Abtastzeitpunkten nach einem abgeschlossenen Selbstorganisationsprozeß für die erste attentive Ebene über einer Galerie von 150 solcher synthetischen Szenen. Eine Spalte entspricht dabei einem FTM-Cluster. Die schwarz markierten FTM-Knoten symbolisieren die Repräsentanten der auf den Objekten der Szene ausgeführten Sprünge (Merkmalsübergänge). Die grau markierten Knoten stellen die nicht verifizierbaren Suchsprünge dar, die vor dem schwarz markierten erfolgreichen

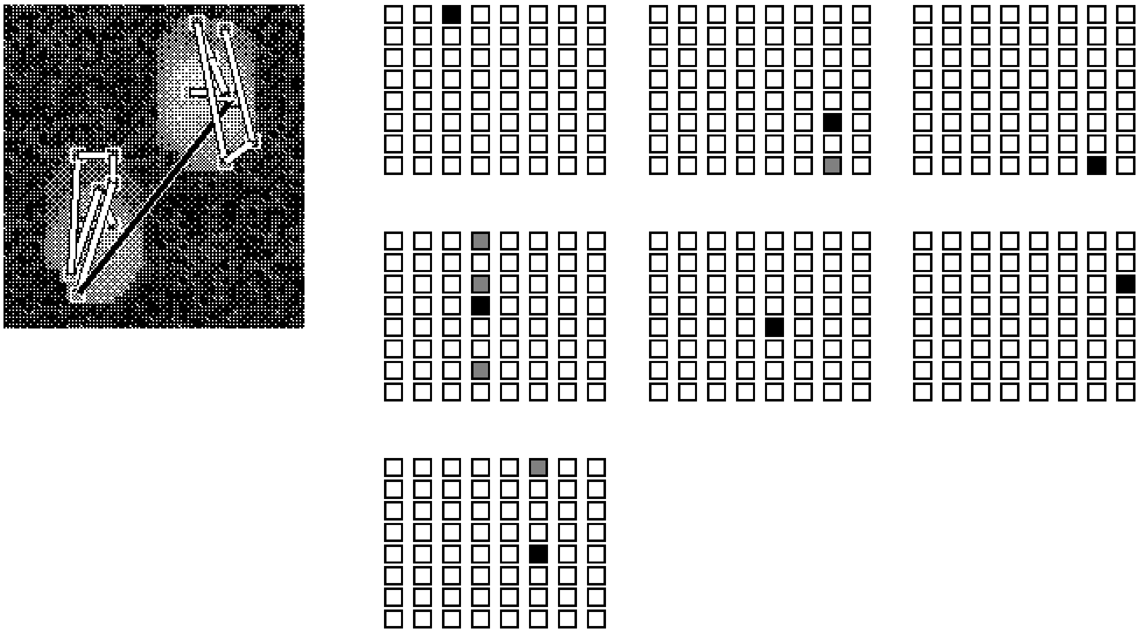


Abbildung 5.9: *Abtastverhalten auf einer synthetisch erzeugten Szene und Darstellung der zu den Abtastpunkten auf dem rechten Objekt korrespondierenden FTM-Zustände (weiße Linie = wissensbasierter Sprung, schwarze Linie = erzwungener Sprung; ausführliche Erläuterung im Text)*

Sprung getestet wurden. Abbildung 5.8 zeigt die Abfolge der FTM-Aktivität bei der Abtastung des ersten Objekts der Szene, Abbildung 5.9 die entsprechende Abfolge bei der Abtastung des zweiten Objekts, nachdem der mit einer schwarzen Linie gekennzeichnete erzwungene Sprung vom ersten zum zweiten Objekt führte. Auf beiden Objekten wurden jeweils 7 Sprünge ausgeführt.
Deutlich sichtbar wird der zeitliche Charakter der Repräsentation von Merkmalsübergängen im FTM. Entscheidend ist dabei die Abfolge der schwarz markierten Knoten zu den aufeinanderfolgenden Abtastzeitpunkten, der zeitliche Verlauf zwischen den Abtastzeitpunkten ist nur von untergeordneter Bedeutung. Ebenso zeigt sich die im Abschnitt 1.1 angenommene Isomorphie zwischen Verhalten (Abtastbewegungen) und Repräsentation (zeitliche Sequenz der FTM-Neuronen). Betrachtet man weiter die Sequenz der erfolgreiche Sprünge repräsentierenden FTM-Neuronen, so zeigt sich, daß sich eine Bindung der Bestandteile der einzelnen Objekte durch eine zeitliche Nachbarschaft der "erfolgreichen" Neuronen ergibt, was einen möglichen Lösungsansatz für das Binding-Problem bildet und für die Funktionsweise der zweiten attentiven Ebene entscheidend ist.
Die beiden Objekte werden durch die zeitliche Sequenz unabhängig von ihrer Lage aus der Szene „heraussegmentiert“. Die Repräsentation im FTM ist damit translationsinvariant.
Für die Darstellungen in den Abbildungen 5.8 und 5.9 wurde auf synthetisch erzeugte Szenen zurückgegriffen, da die Architektur der ersten attentiven Ebene teilweise durch Zuhilfenahme solcher Szenen getestet und weiterentwickelt wurde (siehe auch den letzten NAMOS-Zwischenbericht) und sich das FTM-Verhalten für die letztendlich verwendeten Realszenen aufgrund der erforderlichen Dimension des FTM nicht mehr anschaulich darstellen ließ.

Selbstorganisation der zeitlichen Objektrepräsentation

Abb. 5.10 zeigt die Veränderung des Abtastverhaltens durch die Selbstorganisation der zeitlichen Objektrepräsentation nach der einmaligen Präsentation der 100 Farbszenen der Galerie.
Den Verlauf der Selbstorganisation als Evolution des impliziten Umweltwissens der ersten attentiven Ebene illustriert die Abbildung 5.11. Dazu wurde jeweils nach 10 präsentierten Szenen der Zustand der ersten attentiven Ebene gespeichert und anschließend das Abtastverhalten über der ausgewählten Beispielszene zu den jeweiligen Zuständen protokolliert. Aus der Abbildung 5.11 lassen sich folgende Aussagen ableiten: Zunächst treten noch keine wissensbasierten Sprünge auf, da das Umweltwissen der ersten attentiven Ebene noch unsicher ist. Auf dem *Pinguin mit der Mütze* treten nach 20 Szenen die ersten wissensbasierten Sprünge auf, wohingegen beim *Ring* erst nach 30 Szenen erste wissensbasierte Sprünge auftreten. Die Ursache dafür ist, daß die Sprünge beim *Pinguin mit der Mütze* ebenfalls bei den anderen *Pinguin-Objekten* auftreten und diese Sprünge damit schneller sicher werden. Im Falle des *Rings* sind diese zunächst kurzreichweitig. Dieser Effekt liegt darin begründet, daß zunächst kurzreichweitige, auffälligkeitsbasierte Sprünge in das Umweltwissen eingetragen werden und damit früher sicher werden als die langreichweitigen Sprünge (Einfluß der invertierten

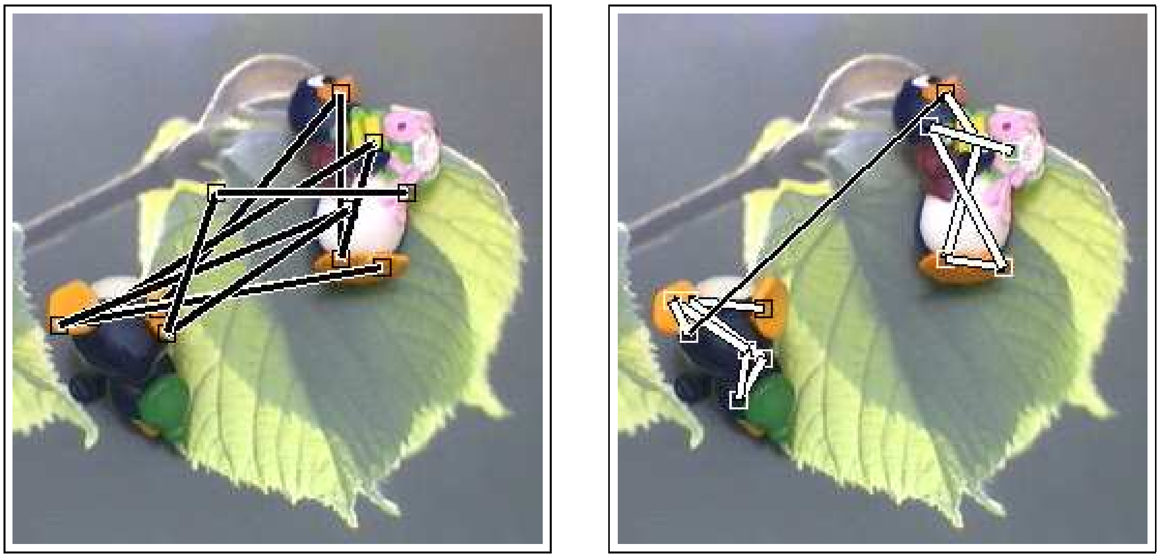


Abbildung 5.10: *Das linke Bild zeigt das rein auffälligkeitsbasierte Abtastverhalten. Das rechte Bild zeigt das Ergebnis der Selbstorganisation der zeitlichen Objektrepräsentation anhand des wissensbasierten Abtastverhaltens.*

Divergenz, siehe auch Abschnitt 5.2.5). Die langreichweitigen Sprünge kommen damit erst nach mehr als 50 Szenen zur wissensbasierten Ausführung. Nach etwa 70 Szenen finden nur noch leichte Umrangierungen innerhalb der wissensbasierten Abtastpfade auf den Objekten statt.

5.2 Detaillierte Beschreibung der ersten attentiven Ebene

5.2.1 Allgemeine Vorbemerkungen

Anfangszustand

Da während jeder Augenbewegung im gewissen Rahmen ein Reset angenommen werden kann, wird vor dem Beginn des Abtastens einer neuen Szene die Modellarchitektur in einen definierten Anfangszustand gesetzt (Zeit t auf null, Variablen wieder auf den entsprechenden Anfangswert).

Globale Signale

Zur Realisierung des Verhaltens der neuronalen Modellarchitektur erwies es sich als notwendig, zwei globale Signale einzuführen. Das erste Signal wird als A bezeichnet. A hat zwei Zustände mit folgender Bedeutung:

$$A = \begin{cases} 0 & \text{APM laden} \\ 1 & \text{sonst} \end{cases}$$

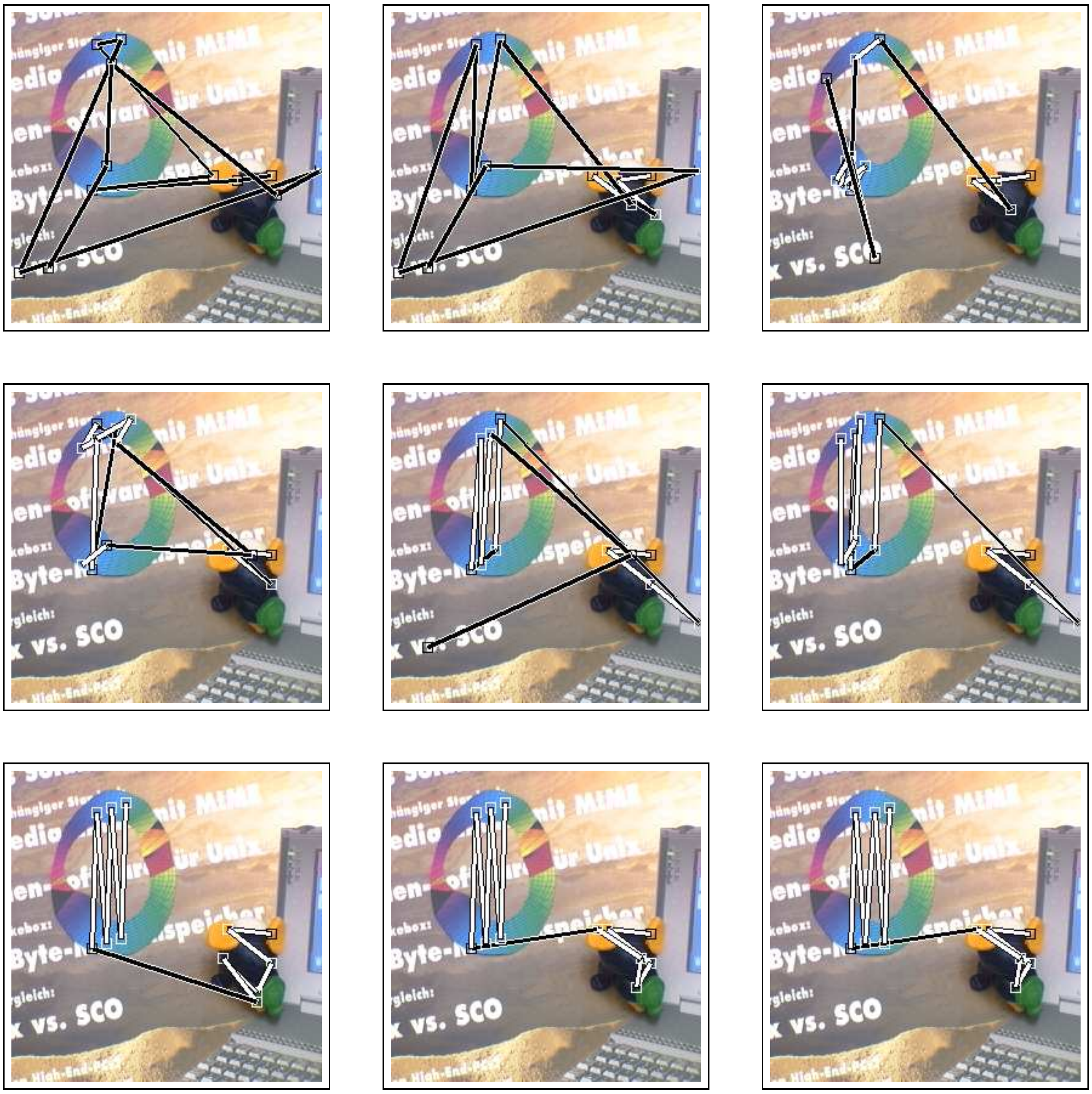


Abbildung 5.11: *Verlauf der Selbstorganisation als Entwicklung des impliziten Umweltwissens nach 0, 20, 30, 40,50, 60, 70, 80 und 90 Szenen (von links oben nach rechts unten).*

Die Berechnungsvorschrift für A ist im Abschnitt 5.2.4 angegeben.
Das zweite globale Signal wird mit D bezeichnet und im Abschnitt 5.2.3 ausführlich erläutert:

$$D = \begin{cases} 1 & \text{erzwungener Sprung} \\ 0 & \text{sonst} \end{cases}$$

Topologische Positionskodierung

In Anlehnung an die Positionskodierung im postparietalen Cortex und im Superior Colliculus wird in der Modellarchitektur eine zweidimensionale topologische Positionskodierung verwendet. Dabei werden eine relative und eine absolute Positionskodierung unterschieden. Im Rahmen der absoluten Positionskodierung kodiert jedes Neuron eine Position in den Merkmalskarten der frühen visuellen Verarbeitung (korrespondierend mit der Lokalisation in der aktuell fovealisierten Szene). Gilt z.B.

$$y_{ij} \begin{cases} 1 & \text{für} \quad i = l, j = m \\ 0 & \text{sonst} \end{cases}$$

so wird durch die Positionskodierung y_{ij} die Position (l, m) beschrieben.
Im Rahmen der relativen Positionskodierung steht die obige Kodierung für $(\Delta x, \Delta y) = (l - Dim_x - 1, m - Dim_y - 1)$, wobei die Merkmalskarten die Dimension $Dim_x \times Dim_y$ aufweisen (im konkreten Fall gilt $Dim_x = 64$ und $Dim_y = 64$, siehe Abschnitt 4.8.2).
In [Heinke et al. 1993] wurde gezeigt, daß diese Vorgehensweise eine unscharfe Kodierung der Position ermöglicht. Im Rahmen der Gesamtfunktionalität der neuronalen Modellarchitektur wurde dieser Ansatz bewußt nicht weiter verfolgt und theoretisch untersucht.

Verwendete Modellneuronen

Die verwendeten Modellneuronen beruhen auf dem Modell des formalen statischen Neurons. Diese Neuronen sind in homogenen Schichten angeordnet. Die Ausgangsaktivität des Neurons j wird folgendermaßen bestimmt:

$$y_j(t) = f(z_j(t))$$

z_j beschreibt dabei die Aktivierung oder den Zustand. Für die Ausgabefunktion $f(x)$ sowie die Berechnung der Aktivierungen wurden unterschiedliche Berechnungsvorschriften verwendet.
Bei der Berechnung der Zellaktivitäten wurden unterschiedliche synaptische Kopplungen berücksichtigt:

axodendritisch: In diesem Fall wird wieder von einer Gleichberechtigung der synaptischen Kopplung ausgegangen:

$$z_j(t) = \sum_{i=0}^{N} w_{ji} x_i(t-1)$$

präsynaptisch: Bei dieser synaptischen Kopplung kann von einer multiplikativen Verknüpfung der axonalen Eingangssignale ausgegangen werden. Diese Verknüpfungsart wird auch als "sigma-pi-Verknüpfung" bezeichnet [Rumelhart et al. 1988b].

$$z_j(t) = \sum_{i=0}^{N} x_i^{(1)}(t-1)x_i^{(2)}(t-1)$$

axosomatisch: Diese Kopplung verwendet im wesentlichen inhibitorische Synapsen ($x_i^{(1)}$). Diese dürften ebenfalls in multiplikativer Form die Aktivierung reduzieren:

$$z_j(t) = \sum_{i=0}^{N}(1 - x_i^{(1)}(t-1))x_i^{(2)}(t-1)$$

Dieser Zusammenhang wird auch als synaptische Modulation bezeichnet.

axoinitial: Diese Kopplung greift sehr stark inhibitorisch in die Erzeugung der Spikes ein und muß darum als Bestandteil der Ausgabefunktion verstanden werden (siehe unten).

Die Verzögerung von einem Zeitschritt ergibt sich aus einer modellhaften Berücksichtigung des langsamen Antwortverhaltens der Neuronen.
Folgende unterschiedliche Ausgabefunktionen zur Berechnung der Ausgangsaktivität werden für die Modellneuronen verwendet (siehe auch [Arbib 1989]):
Schwelle:

$$f_s(x) = \begin{cases} 1 & \text{falls} \quad x > s \\ 0 & \text{falls} \quad x \leq s \end{cases}$$

Rampe:

$$f_{be}(x) = \begin{cases} 0 & \text{falls} \quad x < 0 \\ x & \text{falls} \quad 0 < x < 1 \\ 1 & \text{falls} \quad x \geq 1 \end{cases}$$

Aktivitätsbereiche:

$$f_a(x) = \begin{cases} x & \text{falls} \quad x \leq 0 \\ x + a & \text{falls} \quad x > 0 \end{cases}$$

$$f_{a2}(x) = \begin{cases} 0 & \text{falls} \quad x \leq 1 \\ x - a & \text{falls} \quad x > 1 \end{cases}$$

Die axoinitiale Kopplung wird durch eine Blockierung beschrieben [Groß 1989]:

$$f_{bl}(x, A) = \begin{cases} 0 & \text{falls} \quad A \geq s \\ x & \text{falls} \quad A < s \end{cases}$$

Die Auswahl der verschiedenen Modellneuronen bezüglich synaptischer Kopplung und Ausgabefunktion wird funktionell bei der Beschreibung der ersten attentiven Ebene

motiviert.
Neben diesen "langsamen" Neuronen werden ebenfalls "schnelle" Interneuronen verwendet. Für diese Modellneuronen entfällt die Verzögerung von einem Zeitschritt. Sie sind jeweils Bestandteil einer Schicht von "langsamen" Neuronen und können nur eine einfache, axodendritische Summation mit unterschiedlichen Nichtlinearitäten ausführen.

Lernregeln

Eine wesentliche Modellannahme dieses Projektvorhabens ist, daß die Grundlage des Lernens die synaptische Plastizität darstellt. D.h. die Übertragungsstärke (Gewicht) der Synapse verändert sich. Ausgehend von der Modellvorstellung von [Hebb 1949] ist diese Veränderung eine Funktion von präsynaptischer Aktivität und Ausgangsaktivität. Dies bedeutet im Rahmen der oben verwendeten Notation folgende Abhängigkeit:

$$w_{ij}(t^*) = w_{ij}(t^* - 1) + \epsilon \cdot (b - w_{ij}(t^* - 1)) \cdot y_i(t^*) \cdot x_j(t^*)$$

Wobei ϵ die Lernrate und b der Endwert der Lernregel ist. t^* berücksichtigt die Zeitskala, auf der gelernt wird. Auf diesen Aspekt wird im folgenden Abschnitt genauer eingegangen. Die wesentliche Eigenschaft dieser Hebb'schen Lernregel ist die Lokalität des Lernvorgangs.
Neben diesem lokalen Lernvorgang können pysiologisch auch globale Lernvorgänge motiviert werden. Damit wird die oben formulierte Lernregel zu:

$$w_{ij}(t^*) = w_{ij}(t^* - 1) + \epsilon \cdot (R(t^*) - w_{ij}(t^* - 1)) \cdot y_i(t^*) \cdot x_j(t^*)$$

Wobei $R(t^*)$ den globalen Einfluß beschreibt.

Zeitskalen

Eine weitere Modellannahme im Rahmen dieses Projektvorhabens stellen die unterschiedlichen Zeitskalen dar [Groß 1989]. Das heißt, es werden die Ausgabewerte der verschiedenen Gleichungen der Modellarchitektur in unterschiedlichen Zyklen aktualisiert. Es werden in der ersten attentiven Ebene drei unterschiedliche Zeitskalen verwendet: Mit der langsamsten Zeitskala (t^*) werden die Lernregeln berechnet. Die mittlere Zeitskala umfaßt die Berechnung der Zustände der Modellneuronen (t). Die schnellste Zeitskala betrifft Schichten mit Winner-Take-All-Verhalten. Dieses Verhalten wird quasi ohne Verzögerung umgesetzt.

Divergenz

In der neuronalen Modellarchitektur wird an einigen Stellen das Prinzip der divergenten Verschaltung aufgegriffen. Typischerweise wird dieses durch eine Faltung im Ort mit einer Gaußfunktion beschrieben:

$$y_{ij} = \sum_{k=-N}^{N} \sum_{l=-N}^{N} A \cdot e^{-\frac{k^2+l^2}{2\sigma^2}} \cdot x_{i-k,j-l}$$

Aus praktischen Erwägungen werden hier die Parameter der Gaußfunktion durch die Vorgabe der Werte von N und $A \cdot e^{-\frac{N^2+N^2}{2\sigma^2}}$ bestimmt. Diese Parameter werden im folgenden als "Breite" und "Amplitude" bezeichnet. Die obige Ortsfaltung wird in den Gleichungen der neuronalen Modellarchitektur mit

$$y_{ij} = Div(x_{ij})$$

abgekürzt.

Konzept der separierbaren Aktivitätsbereiche

In dem Konzept der separierbaren Aktivitätsbereiche wird davon ausgegangen, daß unterschiedliche Bereiche von Aktivitätswerten unterschiedliche Funktionalitäten haben. Die Abbildung 5.31 illustriert ein Beispiel zu diesem Ansatz.

5.2.2 Schreibweise der Gleichungen

Notation von Gewichten und Ausgangsaktivitäten

Die zwei-dimensionale Schicht Quelle ($Y^{quelle} \in \mathbb{R}^N \times \mathbb{R}^N$) bildet den Eingang zu einer zwei-dimensionalen Schicht Senke ($Y^{senke} \in \mathbb{R}^M \times \mathbb{R}^M$). Die Schicht Senke ist mit der Schicht Quelle vollständig über die Gewichte ($W^{senke\_quelle} \in \mathbb{R}^M \times \mathbb{R}^M \times \mathbb{R}^N \times \mathbb{R}^N$) vernetzt. So kann dies folgendermaßen notiert werden:

$$y_{ij}^{senke} = \sum_{l=0}^{N} \sum_{k=0}^{N} w_{ijlk}^{senke\_quelle} \cdot y_{lk}^{quelle}$$

Wobei $i = 0 \ldots M$ und $j = 0 \ldots M$ gilt und hier zur Vereinfachung ein lineares Ausgangsverhalten von Schicht Quelle angenommen wird.
Um die Notation zu vereinfachen, wird im folgenden der Sachverhalt folgendermaßen beschrieben:

$$y_{ij}^{senke} = \sum_{l,k} w_{ijlk}^{senke\_quelle} \cdot y_{lk}^{quelle}$$

Die Indexgrenzen ergeben sich dann aus dem Zusamenhang, so gilt z.B. für die relative Positionsinformation $M = 127$, für die absolute Positionsinformation $M = 63$ und für die komplexe Merkmalsinformation $M = 24$. (Diese Parameter ergeben sich aus der Ausgangsdimensionalität der frühen visuellen Verarbeitung und den Parametern des Attentional Focus Identifier (siehe Kapitel 4).)
Außerhalb der Indexgrenzen ist der Wert der Variablen null.
Die Interneuronen der jeweiligen Schichten werden durch zusätzliches Anfügen eines Buchstaben gekennzeichnet: y_{ij}^{senkea}. Die Lernregel wird im Rahmen dieser Notation zu:

$$w_{ij}^{senke\_quelle}(t^*) = w_{ij}^{senke\_quelle}(t^*-1) + \epsilon \cdot (b - w_{ij}^{senke\_quelle}(t^*-1)) \cdot y_i^{senke}(t^*) \cdot y_j^{quelle}(t^*)$$

Externe und Interne Parameter

Die Gleichungen der Modellarchitektur haben eine Vielzahl von Parametern. Die meisten Parameter können durch die vorgegebene Funktionalität der Gleichung bestimmt werden. Dies erfolgt im Zusammenhang mit der Erläuterung der jeweiligen Gleichung. Die externen Parameter, also Parameter, die nicht aufgrund der Funktionalität der Gleichung erfaßt werden können, werden im Schreibmaschinenfont notiert (`Parameter`).

Anfangswerte

Die Anfangswerte der Gleichungen sind, sofern nicht explizit angegeben, null.

5.2.3 Attention Control Network (ACN)

Das Attention Control Network (Abbildung 5.12) realisiert folgende wesentliche Aspekte: Erstens werden die durch die frühe visuelle Verarbeitung erzeugten Merkmalskarten zu einer Auffälligkeitskarte, der Saliency Map (SM), verknüpft. Zweitens soll das Netzwerk auffälligkeitsbasierte Sprünge erzeugen. Drittens soll es zwischen wissensbasierten Sprüngen und auffälligkeitsbasierten Sprüngen umschalten (PAR). Viertens soll es die Kriterien für den erzwungenen Sprung und für den Abbruch des Abtastens einer Eingangsszene berechnen und fünftens die Speicherung des aktuellen Abtastpunkts (im Attentional Position Memory, APM) vornehmen.

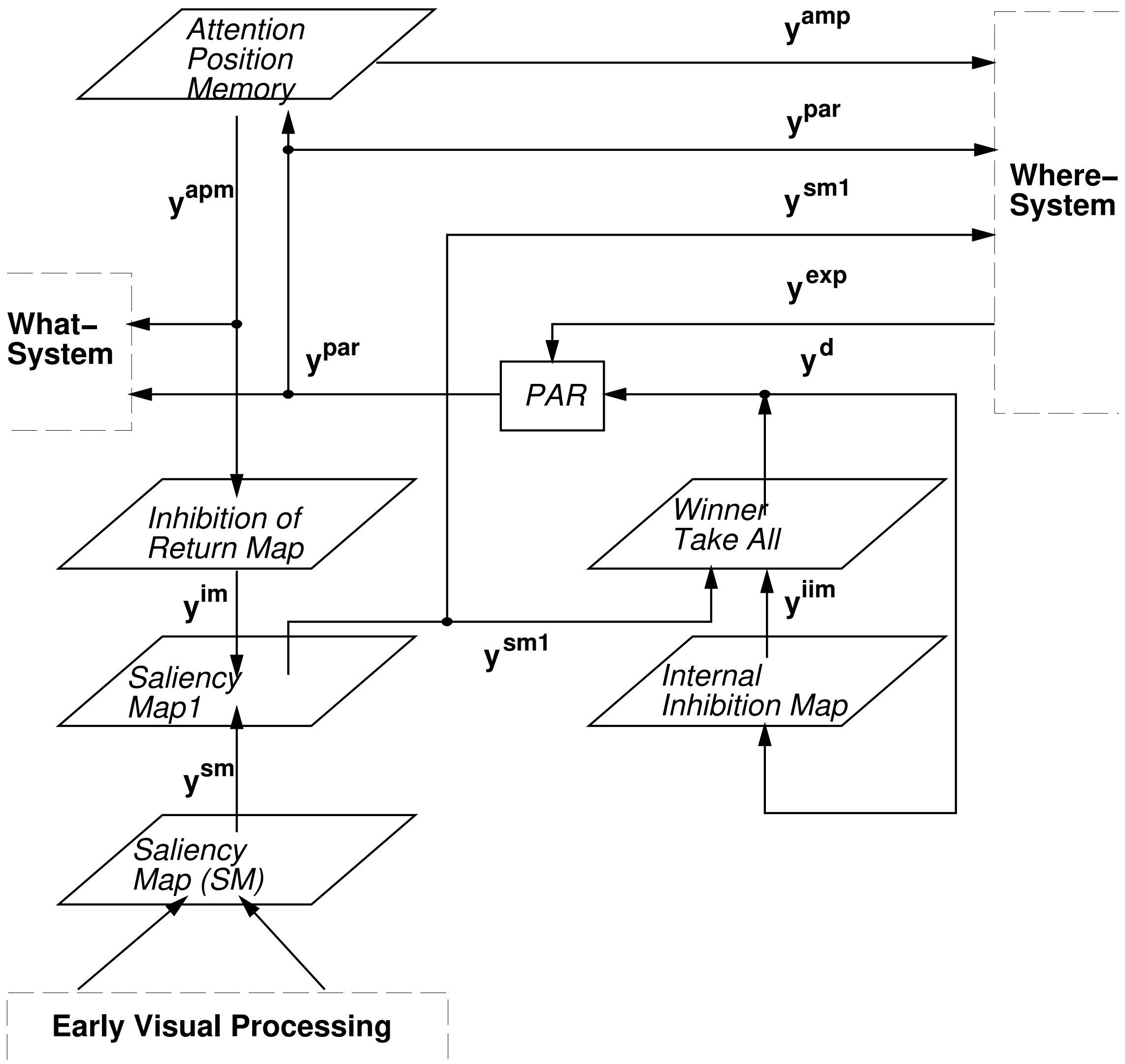


Abbildung 5.12: *Übersicht über das Attention Control Network*

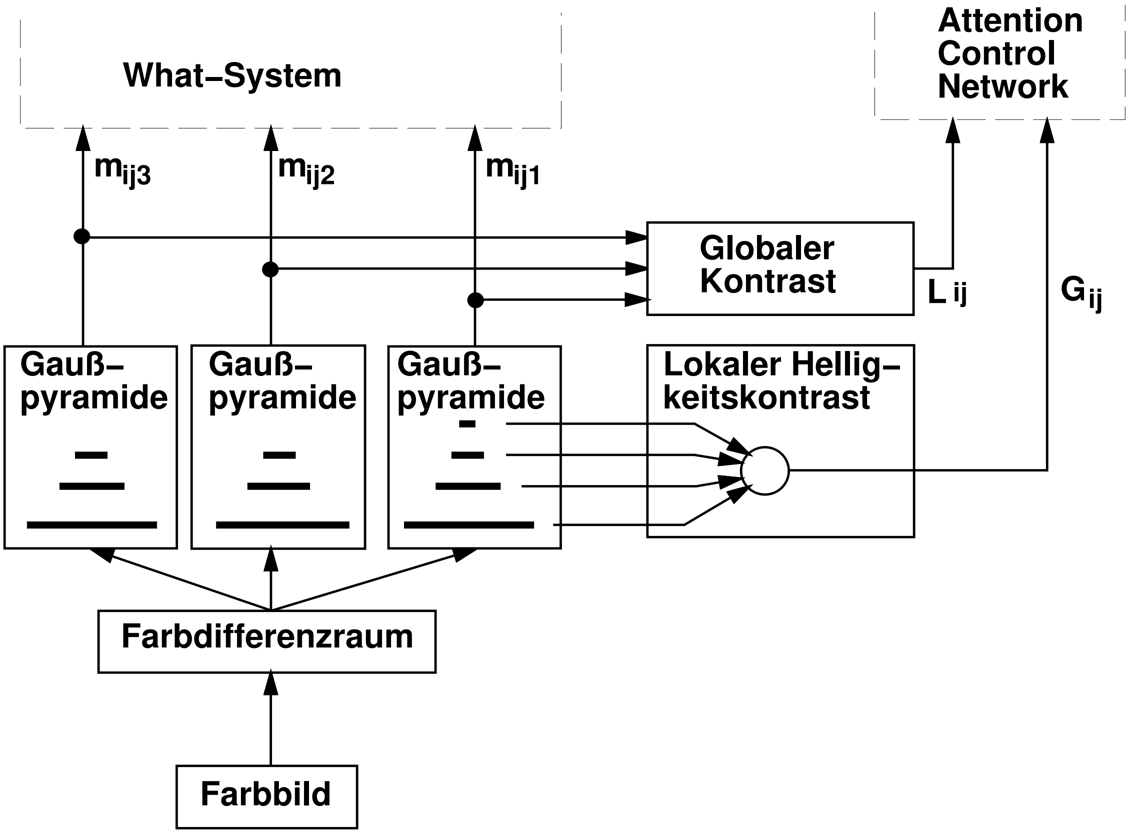


Abbildung 5.13: *Grundlagen aus der Early Visual Processing für die Erzeugung der Auffälligkeitskarte.*

Erzeugung der Auffälligkeitskarte

Bei der Implementation von Mechanismen zur Erzeugung der Auffälligkeitskarte (Abb. 5.13) wurden experimentelle Ergebnisse aus der Psychophysik (Kapitel 2) aufgegriffen, die besagen, daß sowohl lokale Helligkeitskonstraste als auch Abweichungen von einer szenenglobalen "Nulline" bezüglich farbbasierter Merkmale eine wesentliche Rolle spielen.

Lokale Helligkeitskontraste können relativ einfach mit Hilfe von Kantenfiltern ermittelt werden ([Jaehne 1989] [Marr 1982]). Im Rahmen des Projektvorhabens wurde ein biologisch-plausibler DOG-Filter („Difference of Gaussian"), approximiert durch ein Binomialfilter, zur Kantenfilterung verwendet [Marr 1982]. Da der DOG-Filter einem Ableitungsoperator zweiter Ordnung entspricht, entsprechen Kanten den Nulldurchgängen (zero crossings) der DOG-Filterantworten. Um jedoch den Kontrast als lokales Auffälligkeitsmaß zu erhalten, wurde anstelle einer Nulldurchgangsdetektion die lokale Varianz V der DOG-Filterantworten bestimmt [Jaehne 1989]:

$$V_{mn} = \frac{1}{P-1} \sum_{l,k \in B} (y^{WS}_{m-k,n-l} - \overline{y}^{WS}_{m,n})^2$$

Diese Summe läuft über alle P Punkte der WS-Aktivierungskarte der Dimension $B \times B$ (im konkreten Fall gilt $B = 5$).
Damit wird die Bestimmung des lokalen Helligkeitskontrasts HK_L zu:

$$HK_L =^{(B)} V((I -^n B) y^{WS})$$

Um den lokalen Kontrast für alle Auflösungsebenen zu berücksichtigen, werden die DOG- und die sich daran anschließende Varianzfilterung auf alle Ebenen der Pyramide angewandt:

$$HK_L^{(i)} =^{(B)} V((I -^n B) y^{WS^{(i)}})$$

Die Anzahl der Ebenen (Index i) ist im konkreten Fall 4. Für die Ordnung n wurde 5 verwendet.
Um eine einheitliche Bildgröße (64x64) zu erreichen, wurden die $L_K^{(i)}$ vergrößert bzw. verkleinert und additiv überlagert. Bei der Verkleinerung wurde der gleiche Mechanismus wie für die Gaußpyramide verwendet. Die Vergrößerung erfolgte mit dem Interpolationsfilter aus [Jaehne 1989]. Somit ergibt sich eine Karte $HK_L^{(i)}$ für den lokalen Helligkeitskontrast. Abbildung 5.14 zeigt ein Ergebnis dazu.

Die Abweichungen von einer szenenglobalen "Nullinie" bezüglich farbbasierter Merkmale wurden mit Hilfe eines szenenglobalen Mittelwertes bestimmt. Dabei wurde von folgenden Überlegungen ausgegangen: Ein grüner Punkt vor einem roten Hintergrund erscheint ebenso auffällig wie ein roter Punkt vor einem grünen Hintergrund. Daraus wurde geschlußfolgert, daß sich die Auffälligkeit einer Szenenlokalisation aus dem Abstand zwischen der dort vorhandenen Merkmalsausprägung und dem szenenglobalen Mittelwert für dieses Merkmal ergibt. Da weiterhin ein hellgrüner Punkt vor einem dunkelgrünen Hintergrund gleich auffällig erscheint wie ein dunkelgrüner Punkt vor einem hellgrünen Hintergrund, kann von einer additiven Überlagerung der Auffälligkeit in den einzelnen Dimensionen des Farbraums ausgegangen werden. Damit wird die Berechnung des globalen Kontrasts zu:

$$G_{ij} = \sqrt{(y_{ij}^{WS} - \overline{y}^{WS})^2 + (y_{ij}^{RG} - \overline{y}^{RG})^2 + (y_{ij}^{YB} - \overline{y}^{YB})^2}$$

$\overline{y}$ gibt den Mittelwert der Szenen in der jeweiligen Farbachse an. Die Abbildung 5.14 zeigt ein Ergebnis dieser Berechnungsvorschrift.

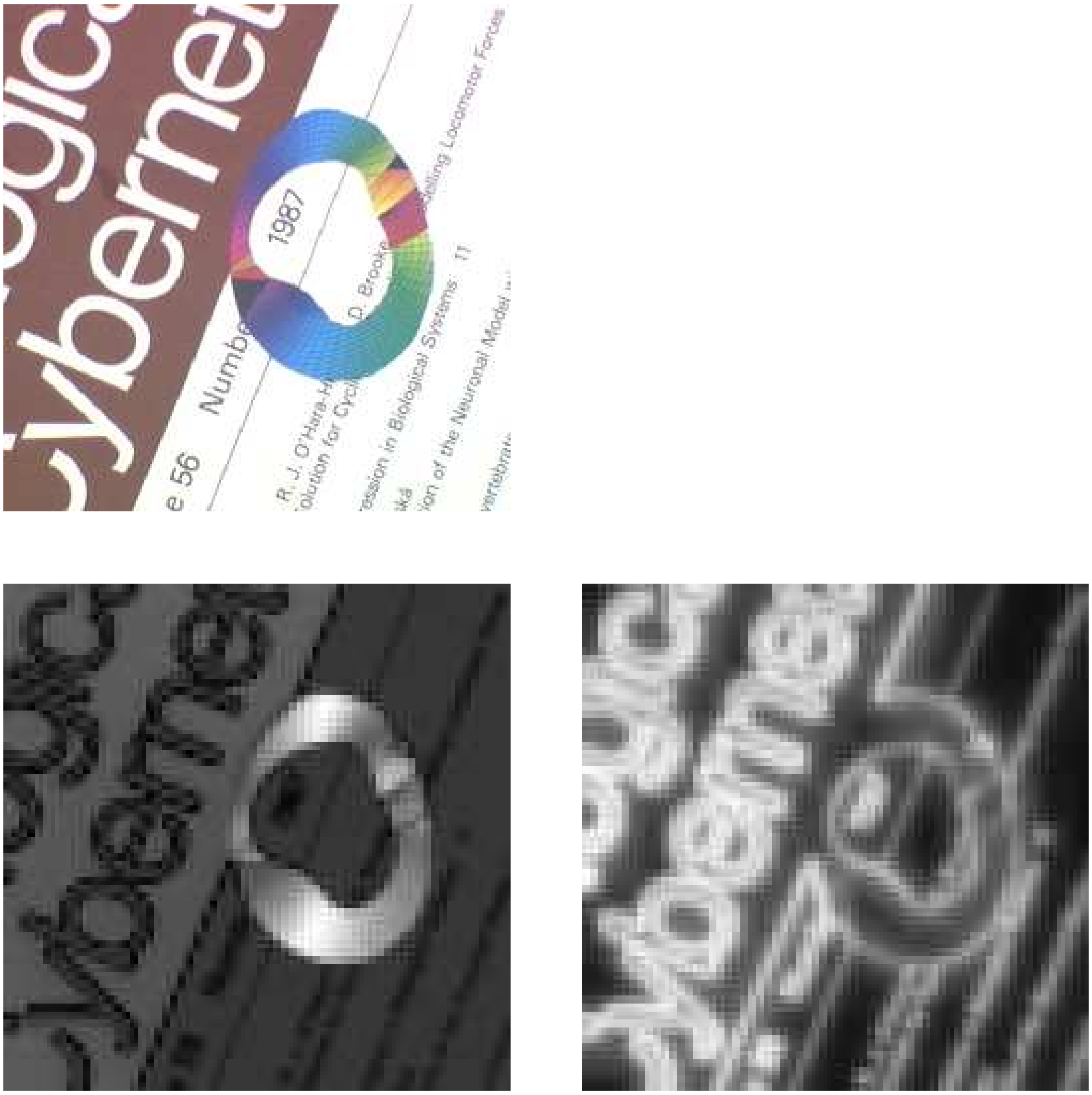


Abbildung 5.14: *Ergebnis des globalen Farbkontrasts (unten links) und des lokalen Helligkeitskontrasts (unten rechts) zur oben dargestellten Originalszene. Die Dimension der beiden Karten beträgt* 64×64

Die experimentellen Ergebnisse der Psychophysik (siehe Kapitel 2) legen nahe, daß die Auffälligkeitskarte durch eine additive Überlagerung der Ergebnisse der frühen visuellen

Verarbeitung erzeugt wird. Übertragen auf das Projektvorhaben bedeutet das, daß eine gewichtete Summe aus Kontrast und Helligkeit nach

$$y_{ij}^{sm} = w_1 \cdot G_{ij} + w_2 \cdot L_{ij}$$

berechnet wird. Die Parameter w_1 und w_2 wurden auf 1.5 und 1 gesetzt. Abbildung 5.15 zeigt am Beispiel einer Szene aus der Galerie das Ergebnis der Auffälligkeitsberechnung.

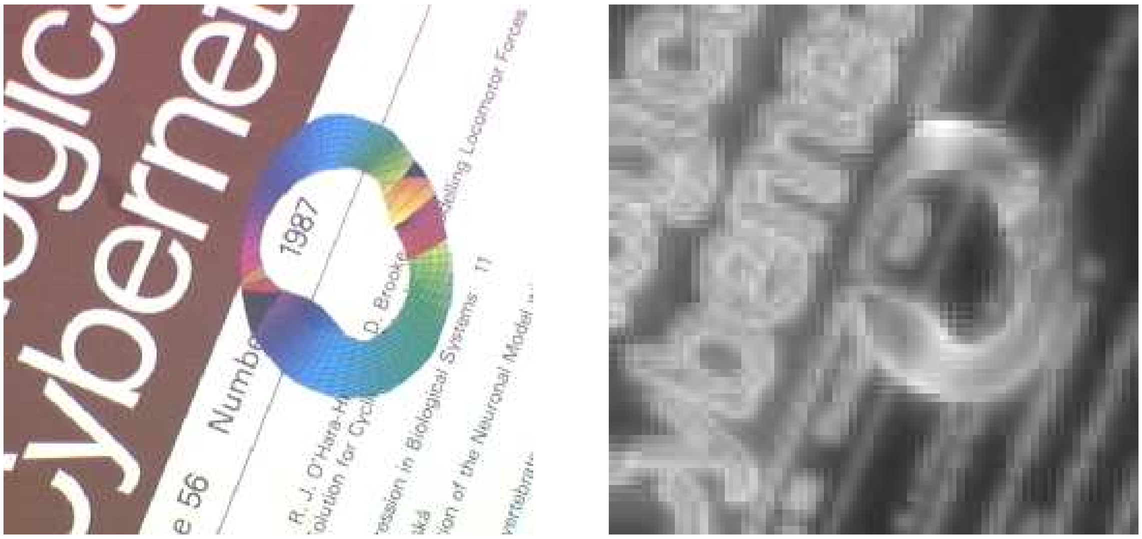


Abbildung 5.15: *Ergebnis der Auffälligkeitsberechnung an einer Beispielszene; links: Originalbild; rechts: korrespondierende Auffälligkeitskarte*

Auffälligkeitsbasierte Sprünge

Die WTA-Karte des ACN selektiert die auffälligste Position in der aktuell fovealisierten Szene. Dabei verhindern die beiden Inhibitionskarten IM und IIM, daß bereits selektierte Positionen erneut selektiert werden:

$$\begin{aligned} y_{ij}^{sm1}(t) &= y_{ij}^{sm}(t-1)(1 - y_{ij}^{im}(t-1)) \\ z_{ij}^{d}(t) &= y_{ij}^{sm1}(t-1)(1 - y_{ij}^{iim}(t-1)) \\ y_{ij}^{d}(t) &= WTA(z_{ij}^{d}(t)) \end{aligned}$$

Dabei geben die letzten beiden Gleichungen die Struktur des WTA-Mechanismus an. Der Index ^{d} korrespondiert hier nicht mit einer Teilstruktur im Modell, sondern symbolisiert lediglich den **d**atengetriebenen Aspekt.

Die Inhibition-of-Return-Map (IM) speichert die Positionen, die bereits den Ausgangspunkt einer Suche durch das What-System bildeten. Die Speicherung erfolgt durch eine rekurrente Verschaltung:

$$\begin{aligned} z_{ij}^{im}(t) &= y_{ij}^{im}(t-1) + y_{ij}^{ima}(t-1) \\ y_{ij}^{im}(t) &= f_{be}(z_{ij}^{im}(t)) \end{aligned}$$

Die interneuronale Schicht y_{ij}^{ima} trägt den Inhalt des APM ein:

$$y_{ij}^{ima}(t) = f_{be}(Div(y_{ij}^{apm}(t-1)) - y_{ij}^{im}(t-1))$$

Die Differenzbildung bewirkt, daß die divergente Charakteristik des Eintrags in die IM erhalten bleibt und nicht durch die rekurrente Verschaltung zum "quaderförmigen" Eintrag wird.
Die divergente Verschaltung hat drei Funktionalitäten: Erstens markiert sie die Umgebung des Abtastpunkts, die bereits über die Gaußpyramide mit verarbeitet wurde. Zweitens erreicht man so ein plausibles Abtastverhalten. Drittens läßt sich die Wirkung der Divergenz als modellhafte Berücksichtigung von Gruppierungseffekten verstehen.

Die Internal Inhibition Map (y^{iim}) soll verhindern, daß die Selbstorganisation in ein "lokales Minimum" konvergiert (siehe Abschnitt 3.1.3). Dazu speichert sie Abtastpunkte, die nicht von der attentiven Ebene angesprungen wurden, aber auffälligkeitsbasiert erreicht worden wären. Dazu wird das gleiche Verschaltungsprinzip angewandt wie das der Inhibition-of-Return-Map:

$$\begin{aligned} z_{ij}^{iim}(t) &= y_{ij}^{iim}(t-1) + y_{ij}^{iima}(t-1) \\ y_{ij}^{iim}(t) &= f_{bl}(f_{be}(z_{ij}^{iim}(t)), D(t)) \\ y_{ij}^{iima}(t) &= f_{bl}(f_{be}(Div(y^d(t)) - y_{ij}^{iim}(t))), A(t)) \end{aligned}$$

Die zwei zusätzlichen Blockierungen erreichen, daß nur Positionen eingetragen werden, wenn ein Eintrag im APM erfolgt ($A = 1$) und daß die Einträge gelöscht werden, wenn ein erzwungener Sprung ($D = 1$) iniziiert wird. Die letzte Funktionalität ist notwendig, um ein sinnvolles Abtasten des zweiten Objekts zu ermöglichen (siehe Abschnitt 3.1.3).

Die Parameter der Divergenz ("Breite" und "Amplitude") lassen sich nicht aufgrund der Funktionalität bestimmen. Damit hat die Modellarchitektur zwei weitere extern einzustellende Parameter:
`INHIBITION_DIV_BREITE` und `INHIBITION_DIV_AMP`.

Preattentive Attentive Router (PAR)

Die Verknüpfung zwischen y^{exp} und y^d (Abbildung 5.12) soll eine Auswahl zwischen auffälligkeitsbasierten und wissensbasierten Sprüngen treffen. Dabei sind folgende drei Situationen zu unterscheiden: Wenn die attentive Ebene keine Aktivität ($y^{exp} = 0$) liefert, darf auch keine Position ($y^{par} = 0$) selektiert werden. Diese Situation ergibt sich aus den Verzögerungen bei der Positionsinformation von der TLM zum PAR (Fall 2, Kap. 5.1.6).
Wenn die Aktivität der attentiven Ebene unspezifisch ist ($y^{exp} = 1$), muß $y^{par} = y^{sm1}$ gelten. Diese Situation tritt ein, wenn ein auffälligkeitsbasierter Sprung ausgeführt werden muß (Fall 4, Kap. 5.1.6).
In der dritten Situation übernimmt die attentive Ebene die Selektion der Position des Abtastpunktes. Dabei muß $y^{par} = y^{exp}$ gelten. Diese Situation tritt ein, wenn eine Suche nach einem erfolgreichen Sprung (Fall 2, Kap. 5.1.6) und das Ausführen

eines erfolgreichen Sprungs (Fall 3, Kap. 5.1.6) erfolgt. Die Funktionsweise für die drei Situationen wird durch folgende Gleichungen erreicht:

$$y^{para}(t) = f_s(\sum_{ij} y_{ij}^{exp}(t)))$$

$$y^{parb}(t) = f_s(\sum_{ij} y_{ij}^{exp}(t)))$$

$$\begin{aligned} y_{ij}^{par}(t) &= (1 - y^{para}(t-1)) \cdot y_{ij}^{exp}(t-1) \\ &+ (2 - D(t-1)) \cdot y^{parb}(t-1) \cdot y_{ij}^{d} \end{aligned}$$

Das Interneuron y^{para} läßt y_{ij}^{exp} im PAR durch. Die Schwelle $s1$ muß auf den Wert 7 gesetzt werden, um zu verhindern, daß beim Adressieren der attentiven Ebene durch einen auffälligkeitsbasierten Sprung am Ausgang des PAR eine Aktivität ungleich null erzeugt wird (Fall 4, Abschnitt 5.1.6 und Abschnitt 5.2.5, TLM). Das Interneuron y^{parb} läßt y_{ij}^{d} im PAR durch. Die Schwelle $s2$ muß den gleichen Wert haben, wie die Anzahl der Neuronen von y_{ij}^{exp}, um die unspezifische Aktivität zu detektieren. Die Multiplikation mit $(2 - D(t-1))$ hat zwei Wirkungen: Solange kein erzwungener Sprung erfolgt ($D = 0$), ist die Positionskodierung des PAR mit der Amplitude 2 als einen auffälligkeitsbasierten Sprung markiert. Damit erfolgt eine Erweiterung des Umweltwissens (Kap. 5.2.5, Positionstransformation). Tritt ein erzwungener Sprung ($D = 1$) auf, so darf dieser nicht zur Erweiterung des Umweltwissens führen.

Abbruchkriterien

Bei der formalen Verhaltensdefinition in Abschnitt 3.1.3 wurde gezeigt, daß es notwendig ist, auffälligkeitsbasierte Abbruchkriterien zu formulieren. Zwei Arten von Abbruchkriterien wurden vorgesehen: Mit Hilfe des ersten Abbruchkriteriums wird das Abtasten eines Objekts beendet (Generierung eines *erzwungenen Sprungs*). Mit Hilfe des zweiten Abbruchkriteriums wird das Abtasten einer Szene beendet.
Das erste Abbruchkriterium setzt sich aus zwei Teilen zusammen:

$$E(t) = E_a(t-1) + E_b(t-1)$$

Wobei $E = 1$ wird, wenn das Kriterium erfüllt ist.
Dem ersten Teil liegen folgende Überlegungen zugrunde: Geht man davon aus, daß ein Objekt bereits wissensbasiert abgetastet wird und die Objekte aus relativ auffälligen Bereichen bestehen, so wird die Summe der Aktivität in der Auffälligkeitskarte SM1 ($\sum y^{sm1}$) durch das Abtasten reduziert. Diese Reduktion kann als Maß dafür verwendet werden, in wie weit bereits auffällige Punkte in einem Objekt abgetastet wurden. Dies erhöht gleichzeitig die Wahrscheinlichkeit, daß das Ziel des *erzwungenen Sprungs*, ein weiteres Objekt in der Szene anzuspringen, erfüllt wird. Umgekehrt bedeutet dies, daß Punkte auf dem aktuellen Objekt nicht Ziele des *erzwungenen Sprungs* werden. Diese

Überlegungen lassen sich durch das Interneuron E_a mit folgender Berechnungsvorschrift umsetzen:

$$E_a(t) \quad = \quad f_{s1}(\sum_{i,j} y_{ij}^{sm}(t) - \sum_{i,j} y_{ij}^{sm1}(t))$$

Der zweite Teil des ersten Abbruchkriteriums beruht auf folgenden Überlegungen: Es besteht die Möglichkeit, daß das Objekt aus wenigen auffälligen Punkten besteht oder viele wissensbasierte Sprünge in wenig auffälligen Gebieten auftreten. In beiden Fällen greift die Reduktion des Auffälligkeitsmaßes eventuell so spät, daß bevorzugt wissensbasierte Sprünge in den Hintergrund erfolgen, da hier jetzt auffälligere Gebiete vorhanden sind. Um diesen Effekt einzuschränken, muß als zusätzliches Kriterium die Anzahl der Sprünge pro Objekt beschränkt werden. Die Summe über die Inhibition-of-Return Map (IM) gibt ein Maß für die Anzahl der Sprünge:

$$E_b(t) \quad = \quad f_{s2}(\sum_{ij} y_{ij}^{im}(t))$$

Da das Erfüllen des Kriteriums ($E = 1$) nur über ein begrenztes Zeitintervall das Verhalten der Modellarchitektur beeinflussen darf, muß das Signal $D(t)$ für den *erzwungenen Sprung* über folgende „Differenzierung“ berechnet werden:

$$D(t) = E(t - \Delta T) - E(t)$$

Für ein korrektes Ausführen eines erzwungenen Sprungs durch die Modellarchitektur muß die Dauer des Signals $D(t)$ 8 Zeitschritte betragen. Dies bewirkt $\Delta T = 8$.

Durch das zweite Abbruchkriterium ($F = 1$) soll das Beenden des Abtastens einer Szene gesteuert werden. Dafür gelten folgende Überlegungen: Die Summe über die Auffälligkeitskarte SM1 ($\sum y^{sm}$) gibt an, wie groß die Auffälligkeit einer Szene ist und damit, wie lange das Interesse an einer Szene anhält. Zusätzlich liefert diese Summe einen Hinweis darauf, wieviele Objekte in einer Szene vorhanden sind. Damit erscheint es sinnvoll, bei Auftreten des *erzwungenen Sprungs* die restliche Summe über y^{sm1} zu überprüfen und so festzustellen, ob noch ein weiteres Objekt in der Szene vorhanden ist:

$$F_a(t) \quad = \quad f_{s5}(\sum_{i,j} y_{ij}^{sm1}(t))$$

Wenn dies nicht mehr der Fall ist, sollte das Abtasten beendet werden. Da in Abschnitt 3.1.2 (Bezüge zwischen externem und internem Scanning) ausgeführt wurde, daß aufgrund der zeitlichen Randbedingungen nur zwei Objekte abgetastet werden können, muß für den Fall des Auftretens eines zweiten *erzwungenen Sprungs* das Abtasten der Szene ebenfalls beendet werden. Das Kriterium für den zweiten *erzwungenen Sprung* ergibt sich einfach dadurch, daß die Schwellen s_1 und s_2 vom Kriterium $E(t)$ mit dem Faktor zwei multipliziert werden:

$$F_b(t) \quad = \quad f_{s3}(\sum_{i,j} y_{ij}^{s}(t) - \sum_{i,j} y_{ij}^{sm1}(t))$$

$$F_c(t) \quad = \quad f_{s4}(\sum_{ij} y_{ij}^{im}(t-1))$$

Dabei gilt $s_3 = 2 \cdot s_1$ und $s_4 = 2 \cdot s_2$.
Damit lautet das Kritierium $F(t)$ für das Beenden des Abtastens einer Szene folgendermaßen:

$$F(t) \quad = \quad F_b(t-1) + F_c(t-1) + D(t) \cdot F_a(t-1)$$

Die Schwelle s_5 ist das Kriterium für das Beenden des Abtastens nach dem ersten *erzwungenen Sprung*. Die Multiplikation mit $D(t)$ ist notwendig, damit der Abbruch des Abtastens einer Szene erst dann erfolgt, wenn nach dem ersten Kriterium ein Objekt vollständig abgetastet wurde.
Da nicht alle Schwellen der Abbruchkriterien durch die Funktionalität der neuronalen Modellarchitektur bestimmt werden können, sind sie externe Parameter. Aus diesem Grund werden die Schwellen mit folgenden Namen bezeichnet: `ABBRUCH1` (s_1), `ABBRUCH2` (s_5) und `SPRUNG_MAX` (s_2).

Attention Position Memory (APM)

Der Attention Position Memory dient zur Speicherung der Position des aktuellen Abtastpunktes, um die Positionstransformation und die Suche nach einem erfolgreichen Sprung durch das What-System ausgehend von dieser Position zu ermöglichen. Die Speicherung wird durch eine rekurrente Verschaltung über eine interneuronale Schicht erzielt:

$$\begin{aligned} y_{ij}^{apm}(t) &= y_{ij}^{apmb}(t-1) + y_{ij}^{apma}(t-1) \quad ; \\ & \quad y_{00}^{apm}(0) = 1 \quad \text{sonst} \quad 0 \\ y_{ij}^{apma}(t) &= f_{bl}(y_{ij}^{apm}(t), 1 - A(t-1)) \end{aligned}$$

Der Anfangswert von y_{ij}^{apm} ermöglicht überhaupt erst einen Beginn der Suche durch das What-System, da sonst durch die Positionstransformation keine relative Auffälligkeitskarte erzeugt werden kann. Die Addition von y_{ij}^{apma} erreicht einen Eintrag in den APM. Die Blockierung löscht diesen Eintrag wieder, wenn ein neuer Sprung gespeichert wird ($A = 1$).
Damit der Eintrag in den APM unabhängig ist von auffälligkeitsbasierten und wissensbasierten Sprüngen, muß zunächst die unterschiedliche Aktivität der Positionskodierung für die beiden Fälle entfernt werden. Dies wird durch eine interneuronale Schicht mit einer Begrenzung erreicht. Da der Eintrag in den APM nur erfolgen darf, wenn ein erfolgreicher Sprung ($A = 1$) gefunden wurde, wird durch eine Blockierung die Weiterleitung der Aktivität an dieser interneuronalen Schicht verhindert:

$$y_{ij}^{apmb}(t)) \quad = \quad f_{bl}(f_{be}(y_{ij}^{par}(t - \Delta T)), A(t-1))$$

ΔT muß auf 4 gesetzt werden, da zum Zeitpunkt t $A = 1$ gilt, ist $y_{ij}^{par}(t)$ wieder null. Dies geschieht, weil die Blockierung im PAR erfolgt.

Auffälligkeitsbasierte Abtastpfade

Die Abbildung 5.16 zeigt zwei auffälligkeitsbasierte Abtastpfade, wie sie vom ACN erzeugt werden. Die relevanten Parameter wurden auf folgende Werte gesetzt:

```
INHIBITION_DIV_BREITE = 4
INHIBITION_DIV_AMP = 0.1
ABBRUCH1 = 67
ABBRUCH2 = 755
SPRUNG_MAX = 7
```

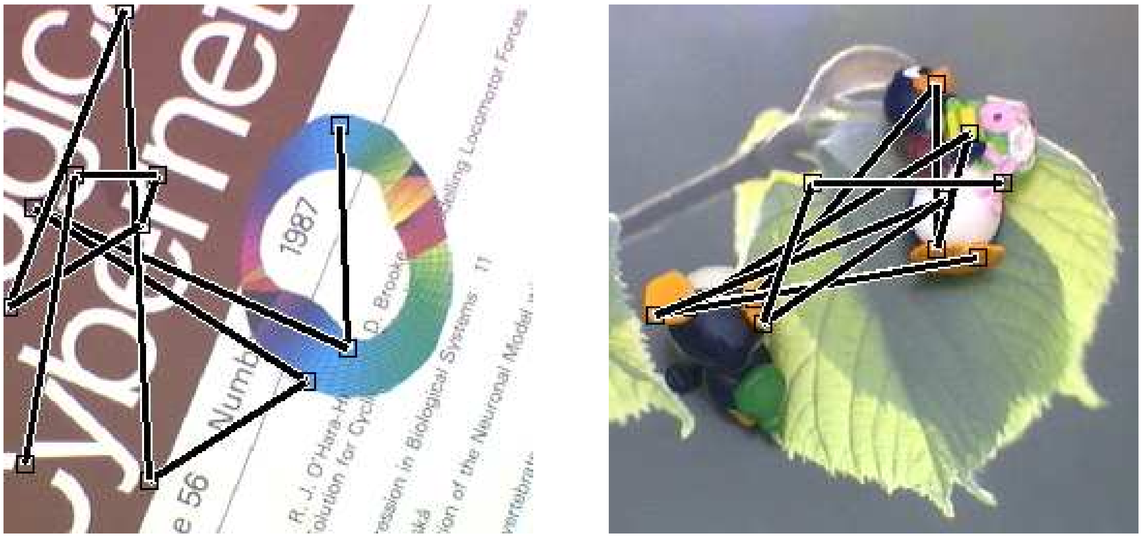


Abbildung 5.16: *Auffälligkeitsbasierte Abtastpfade für zwei Szenen aus der Galerie.*

Die Abbildung 5.17 illustriert die Verteilung der Abtastpunkte auf den verschiedenen Objekten der Szenen. Dabei zeigt sich, daß durch die Verwendung der Auffälligkeitskarte die Anzahl der verschiedenen Sprünge eingeschränkt ist. Damit hat die Auffälligkeitskarte die Wirkung, wie sie im Rahmen der formalen Verhaltensdefinition des Agenten erwartet wurde.

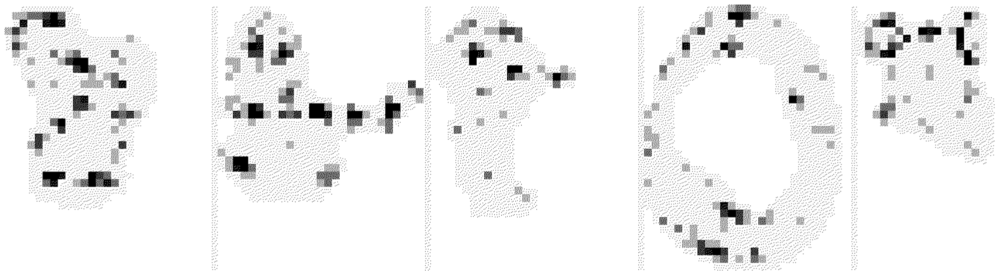


Abbildung 5.17: *Verteilung der Abtastpunkte auf den einzelnen Objekten*

5.2.4 What-System

Das What-System (Abbildung 5.18) hat drei Funktionen: Erstens extrahiert es mit Hilfe der Selection Layers (SL_1 und SL_2) aus den Merkmalskarten der frühen visuellen Verarbeitung den Inhalt des Aufmerksamkeitsfensters in Form eines Merkmalsvektors. Zweitens erzeugt es durch die Focus Identifier (AFI, TFI) eine lokale Repräsentation des Merkmalsvektors. Drittens realisiert es durch den Feature Transition Memory (FTM) die Suche nach einem wissensbasierten, erfolgreichen Sprung.

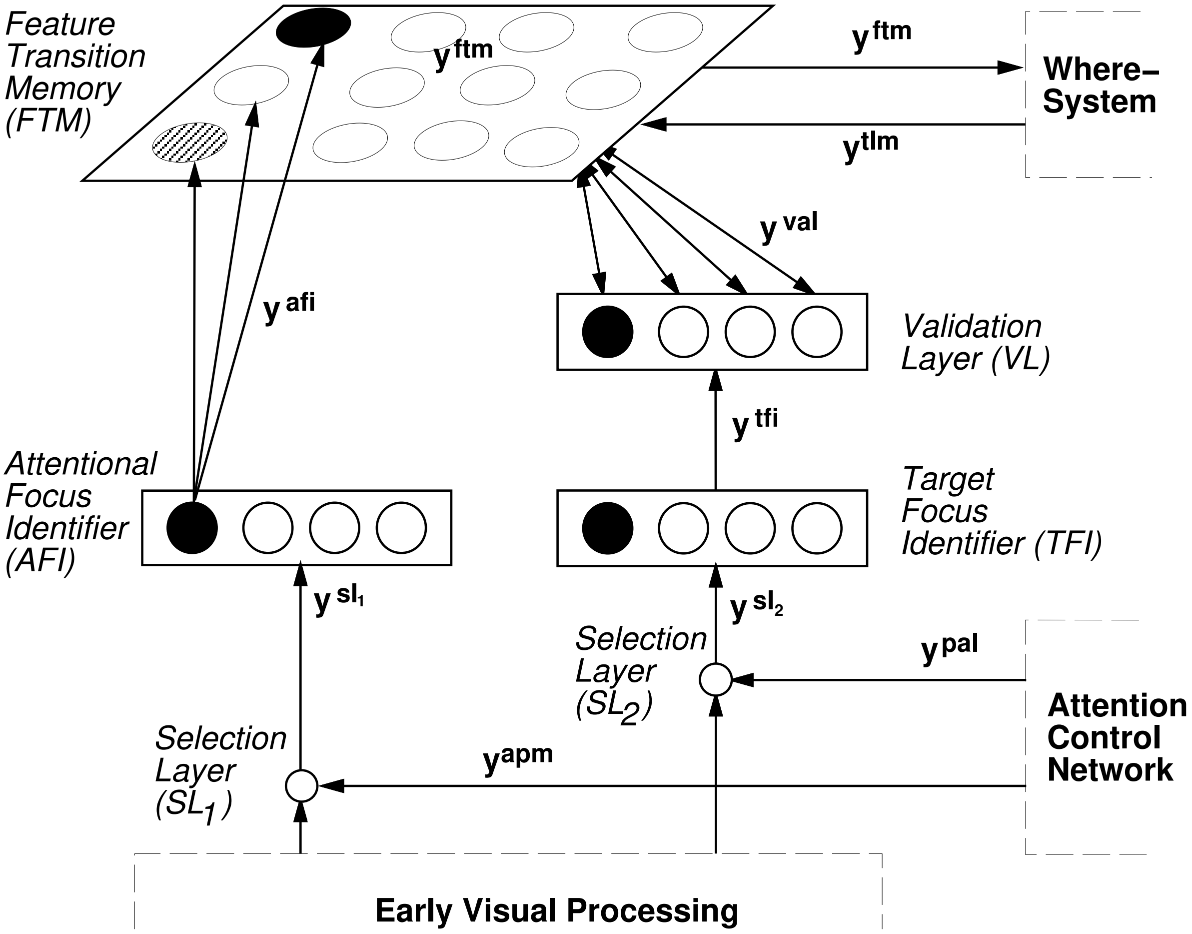


Abbildung 5.18: *What-System*

Selection Layers (SL1 und SL2)

Ziel der Selection Layer ist die Extrahierung des Inhalts des Aufmerksamkeitsfensters aus dem Ergebnis der frühen visuellen Verarbeitung. Dies wird durch eine sigma-pi-Verknüpfung (siehe Abschnitt 5.2.1) zwischen absoluter Positionsinformation (y_{ij}) und den Merkmalskarten der frühen visuellen Verarbeitung (m_{ijk}) erzielt:

$$y_k^{sl}(t) = \sum_{ij} m_{ijk} \cdot y_{ij}(t-1)$$

In [Olshausen 1994] wird ein ähnlicher Ansatz genutzt. Die Funktionsweise soll anhand eines Beispiels illustriert werden: Wenn an der Position (l, m) die Positionsinformation $y_{lm} = 1$ ist und sonst null, so ergibt sich die obige Gleichung zu: $y_k^{sl}(t) = m_{lmk}$. Der Zustand von y_{lm} illustriert gleichzeitig die topologische Kodierung der Position.
Wird die Selection Layer in der Modellarchitektur als SL_1 eingesetzt, so wird die obige Gleichung zu:

$$y_k^{sl_1}(t) = \sum_{ij} m_{ijk} \cdot y_{ij}^{apm}(t-1)$$

Für SL_2 ergibt sich der Zusammenhang zu:

$$y_k^{sl_2}(t) = \sum_{ij} m_{ijk} \cdot f_{be}(y_{ij}^{par}(t-1))$$

Die Begrenzung f_{be} ist notwendig, damit die Variation der Amplitude, die zur Unterscheidung zwischen wissensbasierten und auffälligkeitsbasierten Sprüngen notwendig ist, nicht das Extraktionsergebnis beeinflußt.

Focus Identifier (AFI, TFI)

Ziel der Focus Identifier ist es, den Inhalt des Aufmerksamkeitsfensters in eine lokale Repräsentation (auf ein komplexes Merkmal) abzubilden. Der Aufbau der beiden Focus Identifier ist identisch und wurde im Abschnitt 4.8.2 in Kapitel 4 bereits beschrieben.

Während der TFI benötigt wird, um die Hypothese über das komplexe Merkmal an einer bestimmten Szenenposition zu verifizieren und gegebenenfalls das Umweltwissen bei auffälligkeitsbasierten Abtastbewegungen zu erweitern, dient der AFI der Verarbeitung des Fokusgebietes, von dem die Suche gerade ausgeht.

Wird, wie in Kapitel 4 dargestellt, die Kohonen Feature Map als Attentional Focus Identifier (AFI) eingesetzt, so ergibt sich folgender Zusammenhang:

$$y_i^{afi}(t) = \mathrm{AFI}_i(y^{sl_1})$$

Für den Target Focus Identifier (TFI) wird die gleiche Kohonen Feature Map eingesetzt. Es ergibt sich aber ein etwas anderer Zusammenhang:

$$y_i^{tfi}(t) = \mathrm{TFI}_i(y^{sl_2}) \cdot f_{be}(\sum_{kj} y_{kj}^{par})$$

Die sigma-pi-Verknüpfung soll verhindern, daß keine lokale Repräsentation erzeugt wird, solange keine Positionsinformation vorliegt ($y_{ij}^{par} = 0$).

Zur besseren Überleitung stellt die Abbildung 5.19 noch einmal ein exemplarisches Ergebnis der durch die Focus Identifier vorgenommenen Merkmalsraumdiskretisierung dar.

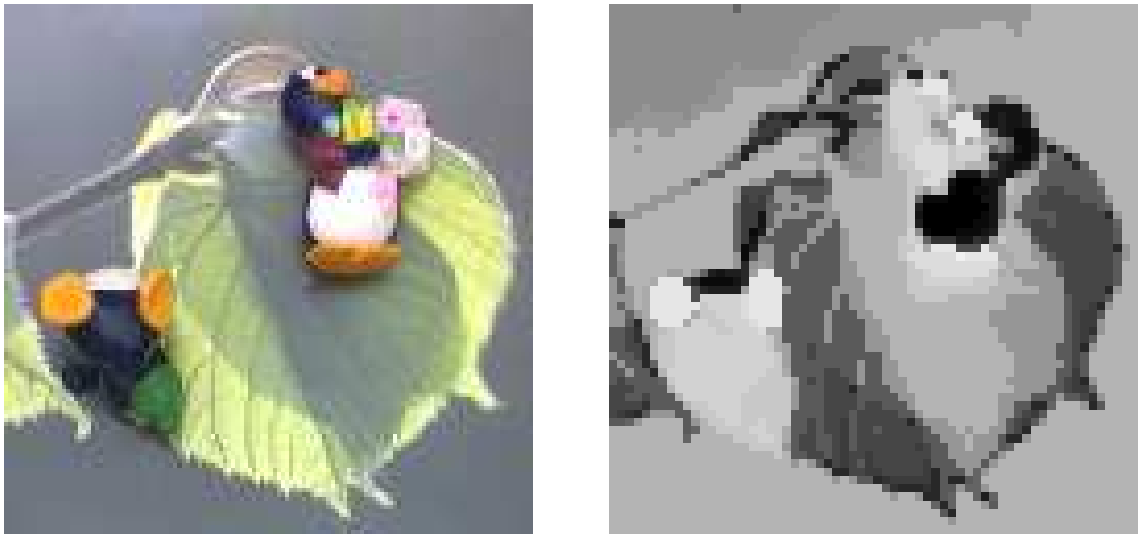


Abbildung 5.19: *Die Abbildung zeigt rechts in Form von Grauwerten das Ergebnis der lokalen Repräsentation des Attentional Focus Identifier für die links dargestellte Beispielszene aus der Galerie.*

Feature Transition Memory (FTM)

Der FTM hat zwei Aufgaben: Erstens muß er die Suche nach einem erfolgreichen Sprung auf der Grundlage seines Umweltwissens durchführen. Zweitens muß er in der Lage sein, dieses Umweltwissen zu erweitern. In Anlehnung an den inferotemporalen Cortex wurde eine columnare Struktur gewählt, d.h. eine FTM-Columne entspricht einem FTM-Cluster (Abb. 5.18). Die Zweiteilung der Aufgabe hat ebenfalls eine Zweiteilung der Aktivitätsberechnung der FTM-Neuronen zur Folge, da eine Teilaktivität die Suche nach einem erfolgreichen Sprung steuert, die andere Teilaktivität die Erweiterung des Umweltwissens ermöglicht. Da aber beide Funktionalitäten die gleichen Gewichte verwenden, müssen die Gewichte am gleichen Neuronen Einfluß nehmen, um die Lokalität der Lernregeln zu erhalten. Diese Strukturierung in Teilaktivitäten läßt sich physiologisch durch große dendritische Verzweigungen am Neuronen motivieren. Damit wird die Berechnung der Aktivität der FTM-Neuronen zu:

$$\begin{aligned} z_{kl}^{ftm}(t) &= z_{kl}^{ftma}(t) + z_{kl}^{ftmb}(t) \\ z_{k0}^{ftm}(t) &= 0 \end{aligned}$$

Die Funktion der Neuronen (k,0) wird im Zusammenhang mit der Berechnung der Ausgangsaktivität des FTM diskutiert.
Die Teilaktivität z_{kl}^{ftma} bestimmt das Verhalten des FTM während der Suche nach einem erfolgreichen Sprung:

$$z_{kl}^{ftma}(t) = y_l^{afi}(t-1) \cdot w_{kl}^{ftm\_afi}(t^*) \cdot \sum_{i,j} w_{klij}^{ftm\_tlm}(t^*) \cdot y_{ij}^{tlm}(t-1)$$

Da die Werte von $\underline{y}^{afi}(t)$ eine lokale Repräsentation darstellen (nur ein y_l^{afi} zum Zeitpunkt t aktiv), wird jeweils nur eine Columne (Cluster) des FTM durch die Multipli-

kation aktiviert. Dies ist notwendig, da nur ein FTM-Cluster alle Sprünge kodiert, die von einem Merkmal m_i ausgehen (Abschnitt 5.1.3).
Der Ausdruck $\sum_{i,j} w_{klij}^{ftm\_tlm}(t^*) \cdot y_{kl}^{tlm}(t-1)$ beruht auf einer sigma-pi-Verknüpfung. Er beschreibt die Modulierung der FTM-Suche durch die relative Auffälligkeitskarte. Damit führen wissensbasierte Sprünge bevorzugt zu auffälligen Positionen. Die Gewichte $w_{kl}^{ftm\_afi}$ entsprechen dabei den Objektzugehörigkeitswerten α_i des Agenten (siehe Abschnitt 3.1.3).
Die Teilaktivität z_{kl}^{ftma} ermöglicht eine Adressierung des FTM durch auffälligkeitsbasierte Sprünge (Abschnitt 5.1.6, Fall 4):

$$\begin{aligned} z_{kl}^{ftmb}(t) &= y_l^{afi}(t-1) \\ &\cdot \left[\sum_{i,j} w_{klij}^{ftm\_tlm2} \cdot y_{kl}^{ftmc}(t-1) \cdot (a + b \cdot \sum_j y_j^{val}(t-1) \cdot w_{klj}^{ftm\_val}(t^*))\right] \end{aligned}$$

y_l^{afi} bewirkt eine Selektion des FTM-Clusters. y_j^{val} erreicht eine Adressierung des FTM durch das Zielmerkmal (siehe Validation Layer).
Über den Ausdruck $\sum_{i,j} w_{klij}^{ftm\_tlm2} y_{kl}^{ftmc}(t)$ adressiert die TLM den FTM. Zu diesem Zweck läßt die interneuronale Schicht $y_{kl}^{ftmc}(t)$ nur den Aktivitätsbereich der TLM durch, der eine entsprechende Adressierung erreicht (Abschnitt 5.2.5):

$$y_{kl}^{ftmc}(t) = f_s(y_{ij}^{tlm}(t))$$

Damit muß $s = 2.1$ gelten. Außerdem sorgt der Ausdruck dafür, daß, solange keine Adressierung erfolgt, $z_{kl}^{ftmb}(t)$ null ist.
Bei der Adressierung des FTM durch auffälligkeitsbasierte Sprünge ergeben sich zwei Fälle: Entweder gibt es diese Sprünge bereits im Umweltwissen oder es sind neue, unbekannte Sprünge. Um für diese Fälle eine korrekte Adressierung des FTM zu erreichen, wurden die Parameter a bzw. b auf 2 gesetzt.

Die Ausgangsaktivität des FTM wird durch ein columnen(cluster)weise arbeitendes Max-to-Min-Netzwerk (M2M) erzeugt (Abschnitt 5.1.3):

$$y_{kl}^{ftm}(t) = \mathrm{M2M}_k(z_{kl}^{ftm}(t), D(t-1), A(t-1))$$

Dieses M2M-Netzwerk hat drei Aufgaben: Erstens muß es die Suche nach einem erfolgreichen Sprung realisieren. Zweitens muß es die Ausführung auffälligkeitsbasierter Abtastbewegungen unterstützen und drittens muß es die Erweiterung des Umweltwissens ermöglichen.
Das M2M-Netzwerk wurde "algorithmisch" realisiert, da die Verhaltensanforderungen an dieses Netzwerk sehr komplex sind, so daß sie mit den hier verwendeten Modellneuronen nicht realisierbar sind. Im Rahmen der Modellannahmen bedeutet dies, daß die Vorgänge im M2M-Netzwerk auf einer schnellen Zeitskala ablaufen müssen. Trotz der algorithmischen Realisierung des M2M-Verhaltens wurde der Zeitverlauf der Ausgangsaktivität dem Verhalten von dynamischen Neuronen nachempfunden (Abbildung 5.20).

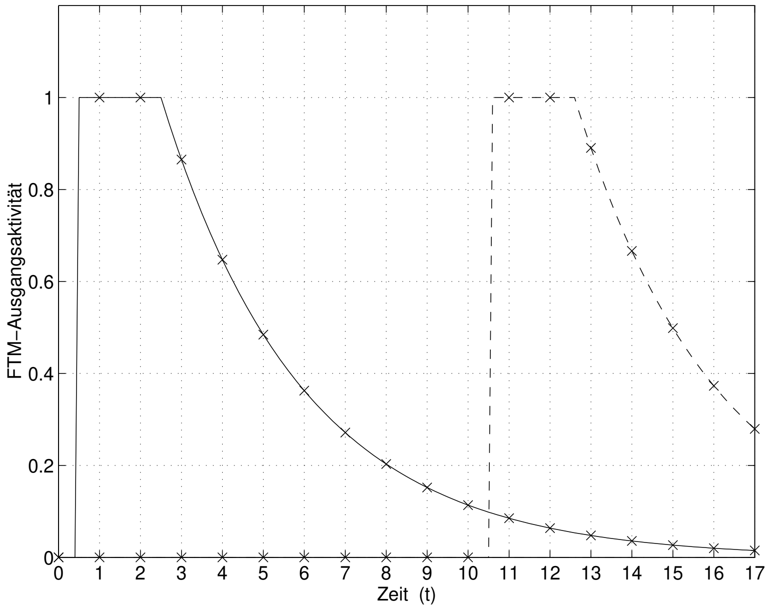


Abbildung 5.20: *Aktivitätsverlauf von $y_{kl}^{ftm}(t)$. Dieser Zeitverlauf wurde dem Antwortverhalten eines M2M-Netzwerkes, basierend auf dynamischen Neuronen, nachempfunden. Die Markierungen verdeutlichen die für die Funktionalität bedeutsamen Zeitpunkte und die dazugehörigen Aktivitätswerte.*

Neben dieser Anlehung an dynamische Modellansätze hat dieser Zeitverlauf auch eine Funktionalität im Rahmen der neuronalen Modellarchitektur. Der Ausgangswert von eins bleibt zwei Zeitschritte erhalten. In diesem Zeitraum kann die Validation Layer einen Vergleich zwischen Merkmalshypothese und tatsächlich vorhandenem komplexen Merkmal durchführen und der Attention Position Memory kann im Falle eines erfolgreichen Sprungs die neue Position als Ausgangspunkt für den nächsten Suchvorgang übernehmen. Außerdem können mit Hilfe der Lernregeln die entsprechenden Gewichte adaptiert werden.
Nach diesen zwei Zeitschritten fällt die Ausgangsaktivität des FTM ab. Dieser Vorgang dauert 8 Zeitschritte, danach wird der neue Gewinner selektiert. Damit blockiert der FTM im Attention Control Network die Selektion eines auffälligkeitsbasierten Sprungs, läßt aber weiterhin für die Lernregeln die Möglichkeit einer korrekten Zuordnung von Merkmals- und Positionsinformation offen.
War der Sprung nicht erfolgreich, dann wird das FTM-Neuron mit dem nächstkleineren z-Wert selektiert. Wird kein erfolgreicher Sprung gefunden, wird zuletzt das Neuron $(k, 0)$ selektiert. Diese Neuronen aktivieren in der TLM eine unspezifische Aktivität (Abschnitt 5.2.1) und erreichen so einen auffälligkeitsbasierten Sprung.
Neben dieser dynamischen Verhaltensweise des M2M-Netzwerkes wurden noch folgende "statische" Eigenschaften angenommen:
Wenn $A = 1$ wird, können in der Folgezeit wieder alle FTM-Neuronen am M2M-Verhalten teilnehmen. Dieser Vorgang kann auch als Löschen von inhibierender Aktivität bezeichnet werden.

Wenn alle y_l^{afi} null sind, bleibt y_{kl}^{ftm} auch null. Damit erfolgt die Suche nach einem erfolgreichen Sprung nur in dem zum aktuell aktivierten AFI-Neuronen korrespondierenden FTM-Cluster. So ergibt sich keine Überlagerung zwischen den FTM-Clustern.
Wenn D gleich eins ist, wird ein $y_{k0}^{ftm}(t)$ als Gewinner selektiert und über die Gewichte $w_{ijkl}^{tlm\_ftm}(t^*)$ eine unspezifische Aktivität in der TLM erzeugt. Damit kann der erzwungene Sprung ausgeführt werden.
Wenn $\sum_{ij} y_{ij}^{ftmc}(t) = l^2$ gilt (Adressierung durch einen auffälligkeitsbasierten Sprung), dann erfolgen l^2 WTA zwischen jedem l-ten FTM-Neuron. Dieser Mechanismus bewirkt, daß maximal l^2 Merkmalsrelationen durch die auffälligkeitsbasierte Adressierung im FTM eingetragen werden. Damit setzt er die lokale Nachbarschaftsbeziehung um 3.1.3) .l wurde im Rahmen der Simulationen auf 3 gesetzt.

Validation Layer (VL)

Aufgabe des Validation Layer (VL) ist es, einerseits einen Vergleich zwischen FTM-Hypothese und Inhalt des Zielaufmerksamkeitsfensters (TFI) durchzuführen (Fall 2, Abschnitt 5.1.6) und anderseits eine Adressierung des FTM durch auffälligkeitsbasierte Sprünge zu ermöglichen (Fall 4, Abschnitt 5.1.6). Die Gleichungen für die Validierungsebene lauten:

$$
\begin{aligned}
y_i^{val}(t) &= \sum_{k,l} w_{ikl}^{val\_ftm}(t^*) \cdot y_{kl}^{ftm}(t-1) \cdot y_{kl}^{vala}(t-1) \\
&+ a \cdot (1 - y^{valb}(t-1) + y^{valc}(t-1)) \cdot y_i^{tfi}(t-1)
\end{aligned}
$$

$$y_{kl}^{vala}(t) = f_{s1}(y_{kl}^{ftm}(t))$$

$$y^{valb}(t) = f_{s2}(\sum_j y_j^{val}(t))$$

$$y^{valc}(t) = f_{s3}(\sum_j y_j^{val}(t))$$

Die interneuronale Schicht $y_{kl}^{vala}(t)$ sorgt dafür, daß die Aktivität $y_i^{val}(t)$ gleich null wird, wenn die FTM-Ausgangsaktivität unter den Wert vom Zeitschritt 7 (Abbildung 5.20) gefallen ist. Damit überlagern sich die Merkmalshypothesen der verschiedenen FTM-Neuronen nicht und die Adressierung der FTM über einen auffälligkeitsbasierten Sprung wird möglich (siehe unten). Die Gewichte $w_{ikl}^{val\_ftm}(t^*)$ haben eine doppelte Funktionalität: Erstens kodieren sie die Merkmalshypothese des dazugehörigen FTM-Neurons. Zweitens stellen sie ein Maß für die Sicherheit des jeweiligen Sprungs (Abschnitt 3.1.3, β_i-Wert) dar.
Der zweite Summand von $y_i^{val}(t)$ ermöglicht die Adressierung des FTM durch den TFI (Fall 4, Abschnitt 5.1.6). Die synaptische Modulation durch die Interneuronen ($y^{valb}(t)$, $y^{valc}(t)$, siehe Abschnitt 5.2.1) soll bewirken, daß die Adressierung nur dann erfolgt, wenn keine Hypothese vom FTM erzeugt wird. Damit muß hier die Schwelle $s2$ vom Interneuron $y^{valb}(t)$ auf null gesetzt werden. Der Faktor a wird auf 2 gesetzt und

sichert so, daß anhand der Aktivität $y_i^{val}(t)$ unterschieden werden kann, ob die Erregung vom FTM oder vom TFI kommt (Konzept der separierbaren Aktivitätsbereiche). Der Einfluß des Interneuronen $y^{valc}(t)$ verhindert, daß, wenn eine auffälligkeitsbasierte Adressierung erfolgt, welche nur einen Zeitschritt lang anhält, die Lernregeln keine sinnvollen Lernvorgänge auslösen können. Deshalb wird der Wert der Schwelle $s3$ auf 1.5 gesetzt.

Der Vergleich zwischen Hypothese und tatsächlichem Inhalt des Aufmerksamkeitsfensters erzeugt ein Reinforcement-Signal $R(t)$ (siehe auch Abschnitt 3.1.3). Dieses Signal wird auch durch die Auffälligkeitswerte beeinflußt. Im Rahmen der neuronalen Modellarchitektur wird dazu zunächst über ein Interneuron der Vergleich zwischen Hypothese und tatsächlichem Inhalt des Aufmerksamkeitsfensters durchgeführt:

$$R^a(t) \quad = \quad f_s(\sum_i y_i^{val}(t) \cdot y_i^{tfi}(t))$$

Damit dieser Vergleich unabhängig von dem Sicherheitswert des aktuellen Sprungs ist, muß die Schwelle der Schwellwertfunktion f_s auf 0 gesetzt werden. Bei Übereinstimmung von Hypothese und Inhalt wird das Interneuronen eins, ansonsten null. Mit einer Modulation mit den Auffälligkeitswerten wird dann die Gleichung für $R(t)$ zu:

$$R(t) \quad = \quad R^a(t-1) \cdot (1 + \sum_{ij} y_{ij}^{sm} \cdot y_{ij}^{par}(t-1))$$

$R(t)$ wird ebenfalls eins, wenn eine Adressierung durch einen auffälligkeitsbasierten Sprung erfolgt.

Berechnung von $A(t)$

Voraussetzung für einen erfolgreichen Sprung ist neben der Übereinstimmung zwischen Hypothese und Inhalt des Aufmerksamkeitsfensters, daß der Sprung sicher ist. Das Kriterium $S(t)$, welches die Sicherheit eines Sprunges angibt, wird durch folgende Gleichungen ermittelt:

$$\begin{aligned} S^*(t) &= f_s(\sum w_i^{s\_afi} \cdot y_i^{afi}(t-1)) \\ S(t) &= f_s(\sum y_i^{val}(t-1)) \end{aligned}$$

Wobei $S(t)$ die Sicherheit eines jeden Sprungs des Umweltwissens und $S^*(t)$ die globale Sicherheit eines FTM-Clusters detektiert. Diese Werte entsprechen den β_i- und γ_i-Werten, die bei der formalen Verhaltensdefinition in Abschnitt 3.1.3 eingeführt wurden. Die Schwellen der beiden Funktionen f_s lassen sich nicht aus der Funktionalität der neuronalen Modellarchitektur herleiten. Damit sind es Parameter, die extern gesetzt werden müssen. Im folgenden werden sie als `s_lokal` und `s_global` bezeichnet.
Das Signal $A(t)$ muß den Wert null annehmen, wenn der Attention Position Memory (APM) mit einer neuen Position geladen werden soll. Dazu wird zunächst folgende interneuronale Schicht $A_{ij}^a(t)$ eingeführt:

$$A_{ij}^a(t) \quad = \quad f_s(y_{ij}^{ftm}(t - \Delta T))$$

Diese interneuronale Schicht detektiert, wann die Positionsinformation am Attention Position Memory (APM) vorliegt. Dazu muß $\Delta T = 5$ gelten, da es solange dauert, bis die Aktivität vom FTM bis zum Vergleich in der VL angelangt ist (Abschnitt 5.1.6) und die Schwelle muß 0.9 sein, da nur während der Zeit, in der die Ausgangsaktivität des FTM eins ist (Abb. 5.20), die Positionsinformationen vorliegen. Das Interneuron $A^b(t)$ wird eins, wenn ein auffälligkeitsbasierter Sprung ($s = 2$) erfolgt:

$$A^b(t) = f_s(\sum_i y_i^{val}(t))$$

$$\begin{aligned} A(t) &= f_s(-\sum_{ij} A_{ij}^a(t-1) - R(t) \cdot S(t) \cdot S^*(t) - D(t-1) - A^b(t-1)) \\ & A(0) = 1 \end{aligned}$$

Diese Gleichungen fassen alle Bedingungen zusammen, bei denen $A = 1$ gilt: Der Ausdruck $R(t) \cdot S(t) \cdot S^*(t)$ wird eins, wenn ein erfolgreicher Sprung vorliegt. Die Schwelle von $A(t)$ muß den Wert -1.5 haben, damit, wenn $\sum_{ij} A_{ij}^a(t-1) = 1$ gilt und eine der anderen Bedingungen eins wird, $A(t)$ gleich null wird.

5.2.5 Where-System

Die Aufgabe des Where-Systems (Abbildung 5.21) ist es, die Positionsinformation zu verarbeiten. Diese Aufgabe gliedert sich in zwei Teilaufgaben: Erstens muß die Positionsinformation aus dem Attention Control Network (ACN) in relative Positionsinformation transformiert werden und umgekehrt. Zweitens muß die merkmalsorientierte Information des What-Systems mit der Positionsinformation verknüpft werden.

Target Location Map (TLM)

In der TLM fließen alle Positionsinformationen zusammen. Um die unterschiedlichen Positionsinformationen trennen zu können, wurde auch hier das Konzept der separierbaren Aktivitätsbereiche angewandt. Hinsichtlich der Aktivitätsbereiche wurde folgende Aufteilung vorgenommen: $[0 \ldots 1]$ Auffälligkeitswerte, $(1 \ldots 2.1]$ räumliche Relation des Umweltwissens, $(2.1 \ldots 3]$ auffälligkeitsbasierte Sprünge.
Die Fusion aller Positionsinformation erfolgt durch folgende Gleichung:

$$\begin{aligned} y_{ij}^{tlm}(t) &= a \cdot Div(y_{ij}^{loc\_r}(t-1)) + y_{ij}^{sm1\_r}(t-1) \\ &+ \sum_{k,l} w_{ijkl}^{tlm\_ftm}(t^*) \cdot y_{ij}^{tlma}(t-1)) \cdot y_{ij}^{ftm}(t-1)) \end{aligned}$$

Mit der Schwelle $s = 0.9$ bewirkt die interneuronale Schicht, daß die Aktivität, die vom FTM erzeugt wird, zwei Zeitschritte vorhanden ist (Abb. 5.20):

$$y_{kl}^{tlm}(t) = f_s(y_{kl}^{ftm}(t))$$

Damit verhindert diese interneuronale Schicht, daß ungültige Daten zum SL_2 gelangen (Abschnitt 5.2.3, PAR).

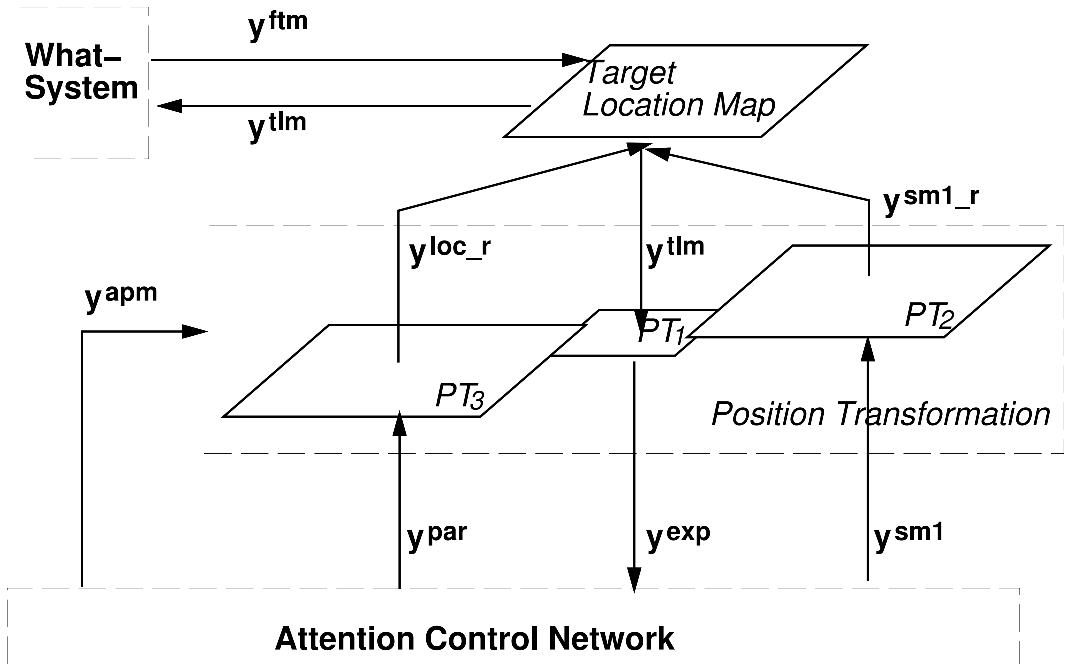


Abbildung 5.21: *Übersicht über das Where-System*

Der Aktivitätsbereich von y^{tlm} wird durch die entsprechenden Gewichte bestimmt.
Der erste Summand ermöglicht die Erweiterung des Umweltwissens durch einen auffälligkeitsbasierten Sprung. Die Divergenz dient zur Berücksichtigung der lokalen Nachbarschaft (Breite auf 3, Randwert auf 0.8 gesetzt). Der Faktor 3 sichert das Erreichen des entsprechenden Aktivitätsbereichs.
Der zweite Summand verkörpert den Einfluß der relativen Auffälligkeitskarte.
Der dritte Summand verknüpft die FTM-Neuronen mit den entsprechenden Positionen in der TLM. Die Gewichte kodieren die räumlichen Relationen $(\Delta x, \Delta y)$ der Elemente des Umweltwissens.

Positionstransformation (PT)

Die Positionstransformation umfaßt drei Vorgänge (Abbildung 5.21): Erstens wird die relative Positionsinformation der Target Location Map (TLM) in absolute Postionsinformation transformiert (PT_1). Zweitens wird die relative Aufmerksamkeitskarte aus der absoluten Aufmerksamkeitskarte erzeugt (PT_2). Drittens wird die Positionsinformation des ACN in eine relative Positionsinformation transformiert (PT_3).
PT_1 transformiert die relative Positionsinformation von der TLM durch eine sigma-pi-Verknüpfung in eine absolute Positionsinformation. Dabei muß die Aktivität der relativen Auffälligkeitskarte in der TLM wieder entfernt werden. Dies wird durch die Ausgabefunktion mit $s = 1$ erreicht:

$$
\begin{aligned}
z_{ij}^{exp}(t) &= \sum_{k,l} y_{k,l}^{tlm}(t-1) \cdot y_{i-k,j-l}^{apm}(t-1) \\
y_{ij}^{exp}(t) &= f_s(z_{ij}^{exp}(t))
\end{aligned}
$$

Diese Operation entspricht einer Faltung im Ort.

Bei der Erzeugung der relativen Auffälligkeitskarte (PT_2) wird ebenfalls eine Positionstransformation verwendet. Allerdings wird hier das Ergebnis noch zusätzlich mit einer "invertierten" Divergenz gewichtet:

$$y_{ij}^{sm1\_r}(t) \quad = \quad (1 - e^{-\frac{i^2+j^2}{2\sigma^2}}) \cdot \sum_{k,l} y_{k,l}^{sm1}(t-1) \cdot y_{i+k,j+l}^{apm}(t-1)$$

Die Parameter dieser Divergenz werden wie die Parameter der Divergenz $Div()$ ("Breite" und "Amplitude") bestimmt. Diese Parameter lassen sich nicht durch ihre Funktionalität bestimmen. Damit ergeben sich zwei weitere Parameter der neuronalen Modellarchitektur, die extern gesetzt werden müssen:
`INV_DIV_BREITE` und `INV_DIV_AMP`.

Die Einführung der "invertierten" Divergenz wurde aus folgenden Überlegungen notwendig: Aufgrund der Statistik von natürlichen Szenen [Ruderman et al. 1994] erhalten kurze Sprünge hohe Objektzugehörigkeitswerte. Damit werden diese Sprünge bevorzugt ausgeführt. Dies ist aus zwei Gründen nicht erwünscht: Erstens ist damit das Verhalten bezüglich einer Objektrepräsentation ineffektiv, da nur lokale Zusammenhänge repräsentiert werden. Zweitens wirkt sich das Verhalten negativ auf das Verhalten des erzwungenen Sprungs aus, d.h. der erzwungene Sprung führt seltener zu einem weiteren Objekt. Dies läßt sich folgendermaßen erklären: Da hohe Auffälligkeitswerte auf einem Objekt weitläufig verteilt sind, führen kurzreichweitige Sprünge nicht zur Inhibition aller höheren Werte. Damit ist die Wahrscheinlichkeit hoch, daß der erzwungene Sprung nicht das aktuelle Objekt verläßt. Mit der "invertierten" Divergenz werden langreichweitige Sprünge bevorzugt ausgeführt. So steigt die Wahrscheinlichkeit, daß alle hohen Auffälligkeitswerte eines Objekts inhibiert werden.

PT_3 wird durch eine sigma-pi-Verknüpfung von APM (y_{ij}^{apm}) und dem Ausgang des PAR ($y_{k,l}^{par}(t)$) realisiert:

$$\begin{aligned} z_{ij}^{loc\_r}(t) &= \sum_{k,l} y_{k,l}^{par}(t-1) \cdot y_{i+k,j+l}^{apm}(t-1) \\ y_{ij}^{loc\_r}(t) &= f_{a2}(z_{ij}^{loc\_r}(t)) \end{aligned}$$

Diese Art der Verknüpfung entspricht einer räumlichen Korrelation. Die Ausgabefunktion $f_{a2}(x)$ soll bewirken, daß nur Positionsinformationen von auffälligkeitsbasierten Sprüngen in die TLM gelangen. Aus diesem Grund wurde a auf eins gesetzt.

5.2.6 Zusammenstellung der Lernregeln der ersten attentiven Ebene

Die Lernregeln müssen einerseits die Bewertung des bereits bestehenden Umweltwissens (Objektzugehörigkeitswert, Sicherheit) realisieren und andererseits die Erweite-

rung des Umweltwissens auf der Grundlage der auffälligkeitsbasierten Abtastbewegungen ermöglichen. Die Gewichte $w_{jkl}^{s\_afi}$ und $w_{kl}^{ftm\_afi}$ bewerten das Umweltwissen. Die Gewichte $w_{ijkl}^{ftm\_tlm2}$, $w_{ijkl}^{tlm\_ftm}$, $w_{klji}^{ftm\_tlm}$ und $w_{kli}^{ftm\_val}$ sorgen für die Erweiterung des Umweltwissens. Die Gewichte $w_{jkl}^{val\_ftm}$ haben beide Aufgaben.
Die Zeitskala t^*, auf der gelernt wird, ist 11 mal langsamer als die Zeitskala der Aktivitäten. Wenn noch eine Startverzögerung von 7 Zeitschritten berücksichtigt wird, wird jedesmal gelernt, wenn das aktuelle Gewinner-Neuron des FTM sieben Zeitschritte aktiv war (Abb. 5.20). Zu diesem Zeitpunkt kann die neuronale Modellarchitektur in zwei verschiedenen Zuständen sein. Der erste Zustand umfaßt insbesondere das Ergebnis des Vergleichs zwischen Merkmalshypothese und Inhalt des Auffälligkeitsfensters. Der zweite Zustand stellt die Adressierung durch einen auffälligkeitsbasierten Sprung dar, wobei zu diesem Zeitschritt die Merkmalsinformation und die Positionsinformation den FTM erreicht haben und die korrespondierenden FTM-Neuronen aktiv sind.
Die Lernregeln haben die Form (siehe Abschnitt 5.2.1):

$$w_{ij}^{senke\_quelle}(t^*) = w_{ij}^{senke\_quelle}(t^*-1) + \epsilon \cdot (b - w_{ij}^{senke\_quelle}(t^*-1)) \cdot y_i^{senke}(t^*) \cdot y_j^{quelle}(t^*)$$

Tritt die Ausgangsaktivität des FTM als y_j^{quelle} bzw. y_i^{senke} auf, so erhält man einen Faktor, der sich aus der Abb. 5.20 zum Zeitpunkt $t' = 7$ ergibt und somit die Lernrate ϵ beeinflußt. Um für diesen Fall vergleichbare Lernraten mit anderen Lernregeln zu erhalten, wird in den Lernregeln mit Beeinflussung durch FTM-Neuronen ein Faktor a eingeführt. Dieser Faktor muß dann auf den reziproken Wert der Ausgangsaktivität des FTM zum Zeitpunkt t' gesetzt werden. Damit wird der Einfluß der FTM-Neuronen auf eins normiert.

FTM-Lernregeln

Die Lernregel für $w_{kl}^{ftm\_afi}$ (Objektzugehörigkeitswerte, Abschnitt 3.1.3) lautet:

$$w_{kl}^{ftm\_afi}(t^*) \quad = \quad w_{kl}^{ftm\_afi}(t^*-1) + \alpha(R(t^*) - w_{kl}^{ftm\_afi}(t^*-1)) \cdot a \cdot y_{kl}^{ftm}(t^*)$$

Damit wird der Objektzugehörigkeitswert von dem aktuellen Gewinner-Neuronen des FTM y_{kl}^{ftm} modifiziert. Dies erfolgt in Abhängigkeit von R, d.h. abhängig davon, ob die dazugehörige Hypothese erfolgreich war.
Die Gewichte $w_j^{s\_afi}$ bewerten die Sicherheit eines FTM-Clusters (Globaler Sicherheitswert):

$$w_j^{s\_afi}(t^*) \quad = \quad w_{jkl}^{s\_afi}(t^*-1) + \tau_2 \cdot (1 - w_j^{s\_afi}(t^*-1)) \cdot y_j^{afi}(t^*)$$

Der Faktor $y_j^{afi}(t^*)$ bewirkt, daß nur das Gewicht modifiziert wird, das mit dem FTM-Cluster verbunden ist, das den Ausgangspunkt der aktuellen Suche bildet.
Die Gewichte $w_{jkl}^{val\_ftm}$ verknüpfen einerseits die FTM-Modellneuronen mit einer Merkmalshypothese und anderseits bewerten sie die Sicherheit der räumlichen Merkmalsrelation, mit der dieses FTM-Modellneuron verknüpft ist:

$$w_{jkl}^{val\_ftm}(t^*) \quad = \quad w_{jkl}^{val\_ftm}(t^*-1) + \tau_1 \cdot (1 - w_{jkl}^{val\_ftm}(t^*-1)) f_s(y_j^{val}(t^*)) \cdot a \cdot y_{kl}^{ftm}(t^*)$$

Die Schwellwertfunktion $f_s(\cdot)$ bewirkt, daß die Lernregeln unabhängig von dem aktuellen Wert der Sicherheit sind, der durch die Aktivität des Validation Layer kodiert wird.
Die Gewichte $w_{kli}^{ftm\_val}$ ermöglichen eine Adressierung des FTM durch einen auffälligkeitsbasierten Sprung. Um eine korrekte Addressierung des FTM zu erreichen, müssen die Gewichte die Umkehrung der Abbildung der $w_{jkl}^{val\_ftm}$ darstellen. Damit werden FTM-Neuronen, die bereits mit einer Merkmalshypothese verknüpft sind, gegebenenfalls auch wieder gefunden. Dies wird durch eine einfache Vertauschung der Indizes erreicht. Damit wird die Lernregel zu:

$$w_{klj}^{ftm\_val}(t^*) \;=\; w_{klj}^{ftm\_val}(t^*-1) + (1 - w_{klj}^{ftm\_val}(t^*-1)) \cdot a \cdot y_{kl}^{ftm}(t^*) \cdot f_s(y_j^{val}(t^*))$$

Da die Gewichte keine weitere Funktionalität besitzen, werden sie in einem Lernschritt auf eins gesetzt. Dies erreicht die Lernrate von 1.

Lernregeln für die Verbindungen zwischen FTM und TLM

Die Verbindungen zwischen TLM und FTM weisen eine besondere Struktur auf. Sie zielen auf eine Realisierung der Nachbarschaftsbeziehung beim Eintrag der auffälligkeitsbasierten Sprünge in die erste attentive Ebene. Beim Eintrag werden im FTM und der TLM jeweils ein Cluster von neun Neuronen aktiv. Damit jedes FTM-Neuron nur mit einem TLM-Neuron verbunden wird, besteht nur zwischen jedem neunten FTM-Neuron und jedem neunten TLM-Neuron eine Verbindung. Und zwar ist jedes k-te FTM-Neuron mit den TLM-Neuronen $(3 \cdot i + (k \quad mod \quad 3), 3 \cdot j + (k \quad mod \quad 9)/3)$ verschaltet.
Diese Verschaltung von FTM und TLM wird bei den Gleichungen der Lernregeln implizit vorrausgesetzt und nicht weiter in den Gleichungen berücksichtigt.
Die Gewichte $w_{ijkl}^{tlm\_ftm}$ kodieren die räumlichen Relationen zwischen Merkmalen:

$$\begin{aligned} w_{ijkl}^{tlm\_ftm}(t^*) &= w_{ijkl}^{tlm\_ftm}(t^*-1) \\ &\quad + \; (b - w_{ijkl}^{tlm\_ftm}(t^*-1) \cdot a \cdot y_{kl}^{ftm}(t^*) \cdot f_s(y_{ij}^{tlm}(t^*))) \\ w_{ijk0}^{tlm\_ftm} &= b \end{aligned}$$

Wobei die Schwellenfunktion $f_s(y_{ij}^{tlm}(t^*)))$ die TLM-Neuronen detektiert, die bei der Adressierung durch einen auffälligkeitsbasierten Sprung aktiviert werden. Darum muß die Schwelle auf 2.2 gesetzt werden. Die Gewichte $w_{ijk0}^{tlm\_ftm}$ haben feste Werte, die eine unspezifische Aktivierung in der TLM erreichen und so die Iniziierung eines auffälligkeitsbasierten Sprungs ermöglichen. Für $b = 1.1$ erreichen die Gewichte, daß der FTM in der TLM den Aktivitätsbereich zur Kodierung der räumlichen Relation des Umweltwissens erreicht.
Die Gewichte $w_{klji}^{ftm\_tlm}$ kodieren die Umkehrung der Gewichte $w_{ijkl}^{tlm\_ftm}$, um die Modulierung der Suche nach einem erfolgreichen Sprung durch die Auffälligkeitskarte zu realisieren:

$$w_{klji}^{ftm\_tlm}(t^*) \;=\; w_{klji}^{ftm\_tlm}(t^*-1) + a \cdot y_{kl}^{ftm}(t^*)(1 - w_{klji}^{ftm\_tlm}(t^*-1)) \cdot f_s(y_{ij}^{tlm}(t^*))$$

Die Gewichte $w_{klji}^{ftm\_tlm2}$ sind notwendig, um die Erweiterung des Umweltwissens zu ermöglichen:

$$\begin{aligned} w_{klji}^{ftm\_tlm2}(t^*) &= w_{klji}^{ftm\_tlm2}(t^*-1) + a \cdot y_{kl}^{ftm}(t^*)(1 - w_{klji}^{ftm\_tlm2}(t^*-1)) \cdot y_{ij}^{tlmc}(t^*) \quad ; \\ & \quad w_{klji}^{ftm\_tlm2}(0) = \text{Zufall}(2\ldots3) \end{aligned}$$

$y_{ij}^{tlmc}(t^*)$ detektiert das Cluster der TLM-Neuronen, das die relative Positionsinformation des auffälligkeitsbasierten Sprungs kodiert. Aus diesem Grund muß die Schwelle auf 2.1 gesetzt werden.

5.3 Zweite attentive Ebene

5.3.1 Motivation

Um die Entwicklung der nachfolgend behandelten Architektur zu motivieren, wurde in Abbildung 5.22 ein mit der ersten attentiven Ebene erzieltes Ergebnis noch einmal unter einem etwas anderen Blickwinkel dargestellt.
Da die erste attentive Ebene aufgrund ihrer Architektur ausschließlich stabile Übergänge zwischen unmittelbar aufeinanderfolgenden komplexen Merkmalen (lokale Hypothesen) erlernt und zur Steuerung des Abtastprozesses verwendet, fehlt globales Wissen über größere Zusammenhänge. Dieses Wissen muß aber ebenfalls zur effektiven Steuerung des Abtastpfades herangezogen werden (siehe auch Abschnitt 3.1.3) und dazu führen, daß die Anzahl der pro Abtastpunkt zu verifizierenden Sprunghypothesen deutlich verringert werden kann (zeitliche Vorausschau aufgrund globaler Hypothesen, siehe Abbildung 5.22 unten rechts).
Die zweite attentive Ebene – der *Episodic Object Memory* (EOM) – vervollständigt die im Rahmen des Projektvorhabens entwickelte Gesamtarchitektur (siehe Abbildungen 3.1 und 3.2, Kapitel 3). Sie orientiert sich funktionell in erster Linie am präfrontalen Assoziationscortex (siehe Kapitel 2) und bildet als *When-System* die höchste Verarbeitungsstufe im Modell.
Das Ziel der Verarbeitung im Episodic Object Memory besteht nun darin, durch Wechselwirkung mit dem Feature Transition Memory der ersten attentiven Ebene **globale Hypothesen** über Folgen von erfolgreichen Sprüngen zu generieren und zu verifizieren (siehe auch Abschnitt 3.1.3). Damit soll eine Repräsentation komplexerer stabiler Bildstrukturen als Kandidaten für Objekte bzw. Teilobjekte erreicht werden, wie sie allein durch die erste attentive Ebene nicht möglich ist.

5.3.2 Modellarchitektur der zweiten attentiven Ebene

Die Abbildung 5.23 stellt die Modellarchitektur des Episodic Object Memory schematisiert dar.
Die Funktionsweise der zweiten attentiven Ebene kann im Gegensatz zur ersten attentiven Ebene nicht mehr separat betrachtet werden, sondern ergibt sich erst durch die enge Wechselwirkung mit der ersten attentiven Ebene. Der EOM setzt im wesentlichen auf dem Feature Transition Memory der ersten attentiven Ebene auf, da der FTM einerseits den eigentlichen Träger der zeitlichen Objektrepräsentation (Übergänge zwischen komplexen Merkmalen) darstellt und andererseits durch dessen Kopplung zum Where-System auch die Positionsinformation verfügbar ist.
Im FTM liegen Cluster vor, von denen jedes ein komplexes Merkmal als Ausgangspunkt für einen relativen Sprung kodiert. Jeder Knoten innerhalb eines solchen Clusters steht für genau einen relativen Sprung, der von dem durch das Cluster repräsentierten komplexen Merkmal ausgeht und über eine Positionsverschiebung $\triangle x$ und $\triangle y$ zu einem weiteren (oder dem gleichen) komplexen Merkmal führt. Das FTM gibt Information an den Episodic Object Memory (EOM) weiter und wird von diesem rückwirkend moduliert.

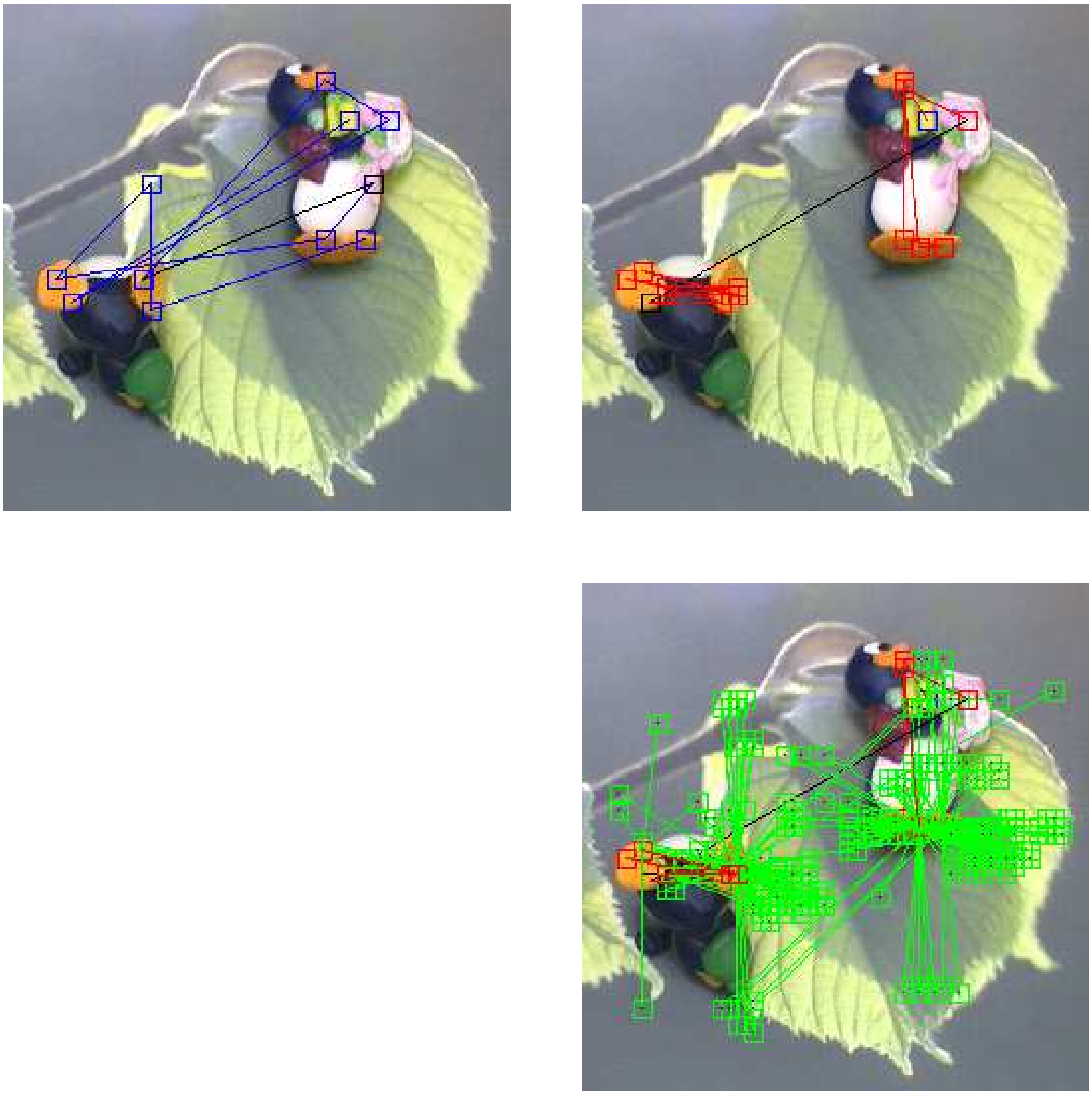


Abbildung 5.22: *Das linke obere Bild zeigt einen rein datengetriebenen Abtastpfad, das rechte obere Bild zeigt einen wissensbasierten Abtastpfad bei ausschließlicher Nutzung des Umweltwissens der ersten attentiven Ebene. Das rechte untere Bild zeigt das gleiche Simulationsergebnis, allerdings wurden dabei auch alle die Sprünge grün markiert eingetragen, die von der Wissensbasis der ersten attentiven Ebene zwar vorgeschlagen wurden, sich aber nicht als erfolgreich erwiesen (erfolgreiche Sprünge sind rot markiert). Mit dieser Darstellung soll verdeutlicht werden, daß das Wissen über beim Abtastprozeß unmittelbar aufeinanderfolgende komplexe Merkmale noch kein hinreichendes implizites Verständnis über komplexere visuelle Strukturen (im Sinne von Objekten) darstellt.*

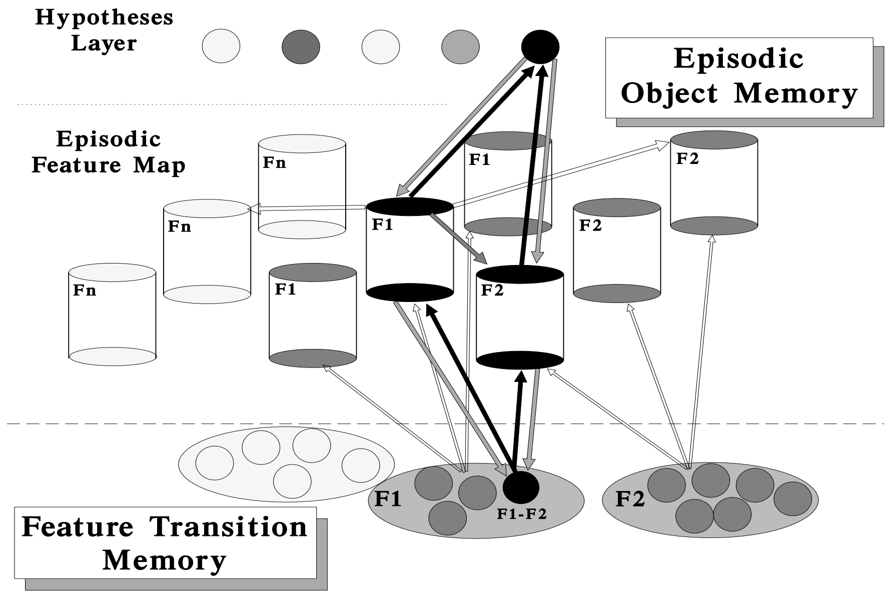


Abbildung 5.23: *Modellarchitektur der zweiten attentiven Ebene*

Der EOM besteht aus zwei Teilstrukturen, der *Episodic Feature Map* (EFM) und der *Hypotheses Layer* (HL).

Episodic Feature Map

Die EFM rekrutiert sich aus Columnenclustern, wobei auch hier jedes Cluster jeweils ein komplexes Merkmal kodiert. Die Columnen innerhalb eines EFM-Clusters realisieren sowohl intra- als auch intercolumnare Verarbeitungsprozesse und kommunizieren mit dem FTM und der Hypothesenebene HL. Die Struktur einer EFM-Columne zeigt die Abbildung 5.24.
Die columnare Organisation der EFM läßt sich biologisch durch die nachgewiesene columnare Struktur aller Assoziationscortizes motivieren. In der Modellarchitektur bilden die EFM-Columnen funktionell relativ abgeschlossene Verarbeitungseinheiten, die verschiedene Knoten mit unterschiedlichen Verarbeitungsaufgaben zusammenfassen. Die prinzipielle Überlegung bestand dabei darin, unterschiedliche Informationsqualitäten, die innerhalb der EFM auftreten können, durch verschiedene Knoten innerhalb einer EFM-Columne zu repräsentieren, um diese verschiedenen Informationsqualitäten separat modulieren zu können, ehe sie summarisch am sum-Knoten der EFM-Columne zusammengefaßt und weitergeleitet werden. Bei den unterschiedlichen Informationsqualitäten handelt es sich um

- die vom FTM kommende Information über ein das Focusgebiet beschreibendes komplexes Merkmal (unspezifischer Eingang vom FTM-Cluster (bezeichnet mit

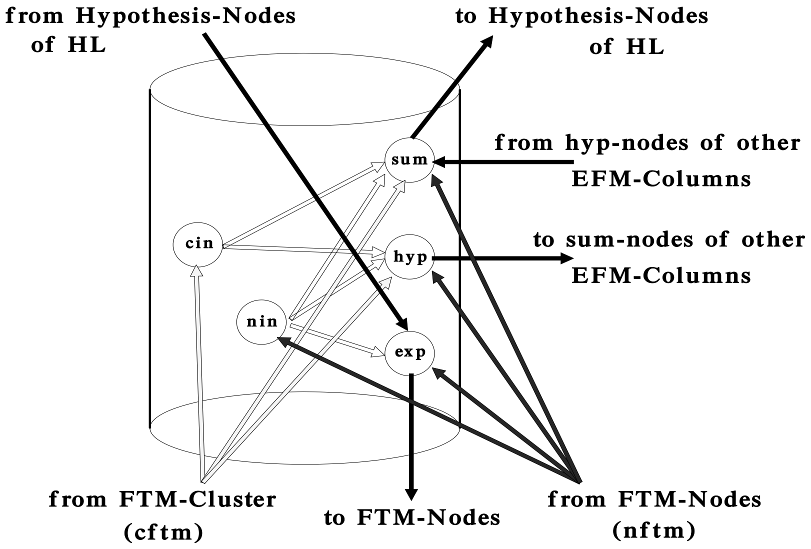


Abbildung 5.24: *Struktur einer EFM-Columne*

cftm),

- die vom FTM kommende Information über die Selektion eines konkreten Sprungs (spezifischer Eingang, kodiert durch einen aktiven FTM-Knoten und bezeichnet mit **nftm**),
- die Eingänge, die von anderen EFM-Columnen einlaufen (bezeichnet mit **hyp**) und
- die Projektionen aus der Hypotheses Layer (bezeichnet mit **exp**).

Eine EFM-Columne besteht somit aus 5 Knoten mit jeweils unterschiedlichen Verarbeitungsaufgaben. Der Knoten **cin** (**c** steht für **C**luster vom FTM, **in** steht für **in**tern) realisiert die columneninterne Repräsentation des vom FTM kommenden unspezifischen Input, der das Startmerkmal des entsprechenden FTM-Clusters darstellt. Da eine Korrespondenz zwischen den EFM- und FTM-Clustern, die das gleiche Merkmal bzw. das gleiche Startmerkmal kodieren, funktionell erforderlich ist, sich diese Korrespondenz aber nicht verändert, kann sie als Basisstruktur vorgegeben werden. Dementsprechend werden die Verbindungen von den FTM-Clustern zu den entsprechenden EFM-Clustern vorab angelegt und mit einem stationären Wert gleich 1 belegt. Der Knoten **nin** (**n** steht für **N**ode vom FTM, **in** steht für **in**tern) realisiert die columneninterne Repräsentation des vom FTM kommenden spezifischen Input. Die genaue Zuordnung zwischen FTM-Knoten und den entsprechenden EFM-Columnen vollzieht sich allerdings erst während der Organisation der EFM. Diese beiden Knoten sind

notwendig, da die Information über selektierte Merkmale und relative Sprünge über eine gesamte Abtastsequenz verfügbar sein muß, dies aber vom FTM aus funktionellen Gründen nicht gewährleistet werden kann.
Der Knoten **hyp** ist für das Aussenden von intercolumnaren Projektionen zuständig, wobei zu allen Columnen der EFM Projektionen möglich sind (also auch auf sich selbst; dies stellt zwar einen Spezialfall dar, der aber nicht ausgeschlossen werden darf). Die Stärke der ausgesendeten Projektionen ist dabei abhängig von den externen Eingängen *cftm* und *nftm* bzw. von den internen Eingängen *cin* und *nin*.
Der Knoten **exp** steht für die Erwartungshaltung (expectation) der EFM-Columne, die maßgeblich von den Aktivitäten der Hypothesenknoten der Hypotheses Layer bestimmt wird, die in diese EFM-Columne projizieren. Er sendet seine Aktivität an genau definierte FTM-Knoten, wobei sich die entsprechenden Korrespondenzen erst während des Selbstorganisationsprozesses herausbilden.
Der Knoten **sum** wird durch die Eingänge *cftm* bzw. *cin*, *nftm* bzw. *nin* und durch die *hyp*-Knoten anderer Columnen oder der gleichen Columne beeinflußt. Er sendet seine Aktivität an die Knoten der HL.

Hypotheses Layer

Die Hypotheses Layer wurde in erster Linie aus funktionellen Gründen eingeführt. Sie beobachtet die Assembly-Formierung innerhalb der EFM und faßt sich stabil entwickelnde EFM-Assemblies (Gruppen gemeinsam aktiver EFM-Columnen) als Kandidaten für Objekt- bzw. Teilobjektrepräsentationen zusammen. Damit bildet die Hypotheses Layer gemeinsam mit der EFM die eigentliche Ebene der Objektrepräsentation innerhalb der Gesamtarchitektur. Die Knoten der HL kodieren während des Abtastprozesses durch ihre Aktivität, inwieweit die sich aufbauenden EFM-Assemblies mit bereits in der HL repräsentierten Assemblies übereinstimmen (siehe insbesondere die entsprechende Aktivierungsfunktion). Die HL-Knoten projizieren ihre aktuelle Aktivität als Erwartung (globale Hypothesen) über die *exp*-Knoten der EFM-Columnen an die Cluster des FTM (siehe Abschnitt 5.5), wo sich dann die lokalen Hypothesen der ersten attentiven Ebene mit den globalen Hypothesen der zweiten attentiven Ebene überlagern. Im Ergebnis dieser Überlagerung entsteht eine *gemeinsame Hypothese beider attentiver Ebenen* (siehe Abschnitt 5.5), die dann das Abtastverhalten moduliert.

5.3.3 Gesamtverhalten der zweiten attentiven Ebene

Um das grundlegende Verständnis des Verhaltens der zweiten attentiven Ebene in Wechselwirkung mit der ersten attentiven Ebene zu erleichtern, wurde auch hier die detaillierte Beschreibung der Aktivierungsfunktionen und Lernregeln in einem separaten Unterpunkt zusammengefaßt.

Etablierung eines Hypothesenknotens

Die nachstehende Abbildung soll den Prozeß bei der Etablierung eines Hypothesenknotens im EOM verdeutlichen. Den Ausgangspunkt bildet die Annahme, daß die Wissensbasis (EFM) noch **unstrukturiert** ist. In den Teilabbildungen sind von unten nach oben jeweils der FTM, die eigentliche EFM und die Hypotheses Layer schematisch vereinfacht dargestellt. Die Spalten verdeutlichen die Cluster im FTM und in der EFM. Hellgraue Kästchen bedeuten inaktiv, mittelgraue Kästchen stehen für teilweise Aktivierung und schwarze Kästchen für vollständige Aktivierung der entsprechenden Knoten. Zum besseren Verständnis sind die jeweils korrespondierenden Cluster des FTM und der EFM unmittelbar übereinander dargestellt.

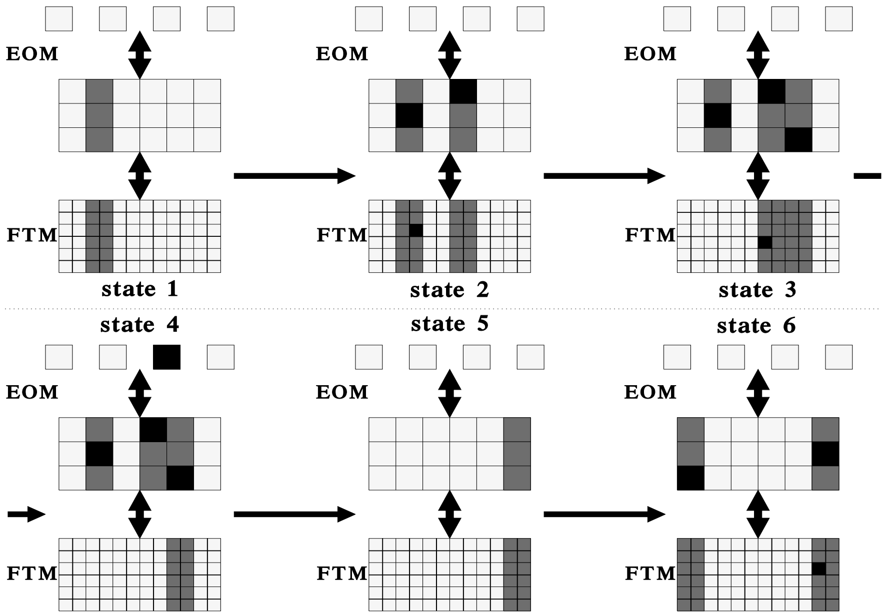


Abbildung 5.25: *Schematisierte Darstellung der Abfolge der Modellzustände bei der Etablierung eines Hypothesenknotens*

Zum **Zeitpunkt 1** wurde eine datengetrieben (auffälligkeitsbasiert) selektierte Fokusregion durch den AFI verarbeitet. Das Ergebnis in Form eines komplexen Merkmals liegt sowohl im FTM als auch in der EFM vor und führt in beiden zur Voraktivierung der entsprechenden Cluster.
Zum **Zeitpunkt 2** liegt eine maximale Aktivität eines FTM-Knotens vor. Das bedeutet, daß ein relativer Sprung zur nächsten Fokusregion im FTM ermittelt werden konnte, dessen Ausführung sich als erfolgreich erwiesen hat. Weiterhin liegt am FTM und an der EFM das komplexe Merkmal der Fokusregion an, zu der der Sprung erfolgte. Da für den maximal aktiven und damit diesen Sprung repräsentierenden FTM-Knoten noch keine Korrespondenzen zur EFM erlernt wurden, wird in den entsprechenden EFM-Clustern jeweils eine EFM-Columne zufällig ausgewählt. Diese beiden Columnen dienen jetzt zur Abbildung des gerade ausgeführten Sprungs. Die Verbindungen vom FTM-Knoten zu den *sum*-, *hyp*-, *exp*- und *nin*-Knoten der jeweiligen Columnen sowie die Verbindungen von den *exp*-Knoten der beiden Columnen zu dem FTM-Knoten werden jetzt modifiziert (siehe die entsprechenden Lernregeln). Außerdem setzt ein

intercolumnarer Lernprozeß zwischen den *hyp*- und den *sum*-Knoten der maximal aktiven EFM-Columnen ein (siehe entsprechende Lernregel), der bis zum Auftreten eines datengetriebenen Sprungs bzw. bis zum Wechsel der Eingangsszene andauert.
Zum **Zeitpunkt 3** wird im FTM wieder ein Knoten maximal aktiv. Das bedeutet, daß ein weiterer Sprung erfolgreich durch das FTM vorhergesagt werden konnte. Im entsprechenden EFM-Cluster wird wiederum eine Columne zufällig ausgewählt. Es erfolgt ein Lernen der Verbindungen analog zu Zeitpunkt 2.
Beim Wechsel von Zeitpunkt 3 zum **Zeitpunkt 4** tritt der Fall ein, daß vom FTM kein weiterer erfolgreicher Sprung ermittelt werden kann. Daraus folgt, daß der FTM die Kontrolle an den datengetriebenen (präattentiven) Prozeß abgibt. Dies dient gleichzeitig als Signal für den EOM, daß der gerade abgelaufene Abtastprozeß (die sequentielle Abfolge erfolgreicher Sprünge) beendet wurde. Daraufhin wird ein Hypothesenknoten in der HL etabliert, der jetzt diesen Abtastpfad kodiert. Die Verbindungen zwischen den entsprechenden *sum*-Knoten und dem HL-Knoten sowie die Verbindungen vom HL-Knoten zu den entsprechenden *exp*-Knoten werden jetzt ebenfalls modifiziert (siehe entsprechende Lernregeln im Abschnitt 5.4.2). Für diesen Lernprozeß ist entscheidend, daß sich ein HL-Knoten über eine extern vorgegebene Schwelle aktivieren läßt. Diese Schwelle hat eine ähnliche Funktion wie der Vigilance-Parameter der ART-Architektur nach [Carpenter et al. 88] und bestimmt, wie stark das in der EFM gerade vorhandene Columnen-Assembly von dem durch einen HL-Knoten repräsentierten Assembly abweichen darf, um noch als zu diesem HL-Knoten gehörend akzeptiert zu werden. Ist dies nicht der Fall (so wie im eben beschriebenen Fall), wird ein neuer HL-Knoten angelegt. Der Verlust der Kontrolle des Abtastprozesses durch den FTM triggert diesen Lernprozeß.
Zum **Zeitpunkt 5** liegt das Ergebnis des ausgeführten datengetriebenen Sprungs als komplexes Merkmal der entsprechenden Fokusregion wieder am FTM und an der EFM an und die Verarbeitung vollzieht sich wie eben beschrieben (Zeitpunkt 6 äquivalent zum Zeitpunkt 2).

Hypothesenbildung und Hypothesenverifikation mittels FTM

Die Hypothesenbildung und die Einflußnahme auf den FTM als Hypothesenverifikation bilden die hierarchisch höchsten Verarbeitungsleistungen der Gesamtarchitektur. Der eigentliche Prozeß der sensorisch gesteuerten internen Simulation als kontinuierliche Wechselwirkung von internem Entscheidungsprozeß und einkommender sensorischer Information findet erst mit der Wechselwirkung beider attentiver Ebenen statt.
Beginnend mit der Etablierung des ersten Hypothesenknotens im EOM setzt der Prozeß der Einflußnahme auf den Feature Transition Memory ein, der sich folgendermaßen charakterisieren läßt:
Durch die Projektionen aktiver FTM-Cluster und aktiver FTM-Knoten in den EOM erfolgt durch die intracolumnaren Verbindungen eine Ausprägung verschiedener paralleler Hypothesen in der EFM in Form von gemeinsam aktiven Columnen. Die Knoten der Hypothesenebene werden von diesen Columnen erregt. Ihre Aktivität hängt davon ab, inwieweit die sich in der EFM etablierenden Columnen-Assemblies den durch die Hypothesenknoten bereits repräsentierten Columnen-Assemblies entsprechen. Die Hy-

pothesenknoten projizieren nun abhängig von ihrer aktuellen Aktivität ihrerseits über die *exp*-Knoten der EFM-Columnen ihre Erwartungen in die entsprechenden FTM-Cluster zurück. Damit kommt es dort zur Überlagerung der lokalen Hypothesen vom FTM und den im globaleren Zusammenhang zu sehenden Hypothesen vom EOM, was letztendlich dazu führt, daß jetzt der aufgrund lokaler und globaler Hypothesen erfolgversprechendste Sprung ausgewählt und zur Steuerung des Abtastverhaltens (Hypothesentest) herangezogen wird (siehe Abschnitt 5.5).

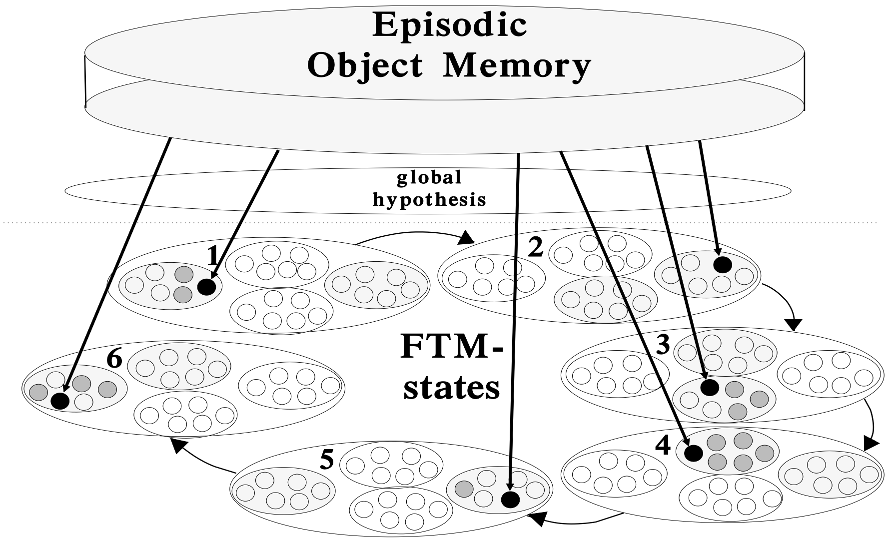


Abbildung 5.26: *Darstellung von sechs aufeinanderfolgenden Etappen der Einflußnahme des Episodic Object Memory auf die Selektion der Sprungkandidaten im FTM der ersten attentiven Ebene; grau hervorgehoben sind die gerade aktiven FTM-Cluster, die im Ergebnis der Analyse der Fokusregionen durch AFI und TFI aktiviert werden; mittelgraue Knoten bedeuten lokale Hypothesen, die ausschließlich aufgrund des innerhalb der ersten attentiven Ebene akkumulierten Umweltwissens selektiert worden wären; schwarz sind die Knoten markiert, an denen durch Zusammenwirken von lokalen und globalen Hypothesen der ensprechende Sprungkandidat ermittelt wurde*

Schematisiert ist der Prozeß der Einflußnahme des EOM auf den FTM in der Abbildung 5.26 dargestellt (siehe auch Abschnitt 5.5). Als Ergebnis der Rückwirkung soll sich die Knotenauswahl im FTM deutlich beschleunigen, da zur Erzeugung einer gemeinsamen Hypothese beider attentiven Ebenen jetzt lokales und globales Wissen zusammenfließt. Unter Beschleunigung der Knotenselektion im FTM soll dabei verstanden werden, daß sich die Anzahl der nicht verifizierbaren Sprunghypothesen im Vergleich zur ausschließlichen Nutzung von lokalen Hypothesen der ersten attentiven Ebene deutlich reduziert, was die nachfolgenden Simulationsbeispiele auch deutlich belegen.

5.3.4 Simulationsergebnisse zur zweiten attentiven Ebene

Die folgenden Abbildungen stellen anhand von zwei ausgewählten Beispielszenen aus der 100 Szenen umfassenden Galerie charakteristische Simulationsergebnisse dar, die mit der Gesamtarchitektur (erste und zweite attentive Ebene) gewonnen wurden. In Anlehnung an die Darstellungsweise der Simulationsergebnisse der ersten attentiven Ebene (Abbildungen 5.8 und 5.9) werden in den Abbildungen 5.27 und 5.29 die Aktivitätsverteilungen über alle bislang angelegten Hypothesenknoten (in beiden Beispielen jeweils 25) der Hypotheses Layer zu den aufeinanderfolgenden Abtastzeitpunkten gezeigt. In beiden Beispielszenen werden zwei Objekte abgetastet, wobei zu jedem dargestellten HL-Zustand diese Aktivitätsverteilung (globale Hypothesen) und die lokalen Hypothesen der ersten attentiven Ebene (aufgrund der Dimension des FTM nicht darstellbar) zur Ausführung des jeweiligen Sprungs führten (siehe auch Abbildung 5.26).

Bei der Erstellung der beiden gezeigten Simulationsbeispiele war der Selbstorganisationsprozeß der beiden attentiven Ebenen bereits abgeschlossen, was sich unter anderem auch in der konstanten Anzahl von Hypothesenknoten äußert.

Die Abbildungen 5.28 und 5.30 verdeutlichen den Einfluß der globalen Hypothesenbildung des Episodic Object Memory auf die Selektion der erfolgversprechendsten, nächstfolgenden Merkmalsübergänge (Sprünge) im FTM. Während in den Abbildungen 5.28 und 5.30 jeweils links das Abtastverhalten ohne den Einfluß der zweiten attentiven Ebene zu sehen ist, ist jeweils rechts das Abtastverhalten bei Wechselwirkung beider attentiver Ebenen dargestellt. Deutlich sichtbar wird die Reduktion der grün markierten Suchsprünge, die sich beim Abtastprozeß nicht verifizieren ließen. Diese grün markierten Suchsprünge entsprechen den in der Abbildung 5.8 grau markierten FTM-Knoten. Erfolgreiche, d.h. verifizierbare Sprünge, wurden rot markiert und entsprechen den in Abbildung 5.8 schwarz dargestellten FTM-Knoten. Damit ergibt sich eine weitere Effektivierung des Abtastprozesses, da jetzt nicht mehr ausschließlich das unspezifische Umweltwissen der ersten attentiven Ebene über in den Szenen stabil auftretende Merkmalsübergänge zur Steuerung des Abtastverhaltens verwendet wird, sondern auch das in der zweiten attentiven Ebene organisierte spezifische Wissen über Objekte bzw. Objektteile mit in die Steuerung des Abtastverhaltens einfließt. Der Abschnitt 5.5 beschreibt, wie das Zusammenspiel von lokalen und globalen Hypothesen im FTM konkret erfolgt.

Anhand von Abbildung 5.27 sollen einige wesentliche Aspekte der Hypothesenbildung und -verifikation im EOM etwas ausführlicher diskutiert werden. Die aufeinanderfolgenden HL-Zustände stellen die Aktivitätsverteilung über alle bislang angelegten Hypothesenknoten (25) dar, die verschiedene EFM-Columnen (Columnen-Assemblies) repräsentieren. Nachdem die Selektion des ersten Abtastpunktes auffälligkeitsbasiert (rein datengetrieben) zu dem rechten Objekt führte, beginnt aufgrund der über diese Aufmerksamkeitsregion ermittelten Merkmalsinformation unmittelbar der Hypothesenbildungsprozeß im EOM. Deutlich sichtbar wird die mit fortschreitendem Abtastprozeß zunehmende Differenzierung der sich parallel entwickelnden Hypothesen, die

alle in den FTM projizieren.
Die sich jeweils am stärksten etablierenden HL-Knoten wurden bei der Abtastung des ersten Objekts mit einem Stern, für das zweite Objekt (links) mit einem Kreuz gekennzeichnet. Während die mit einem bzw. zwei Sternen gekennzeichneten Hypothesen nicht ihre Maximalaktivität erreichen (Über- oder Unterrepräsentation des vollführten Abtastpfades, siehe insbesondere Abschnitt 5.4.1), setzt sich unmittelbar vor dem letzten Abtastschritt auf dem rechten Objekt die mit drei Sternen markierte Hypothese durch, die damit die volle Übereinstimmung des ausgeführten Abtastpfades mit dem durch sie repräsentierten Abtastverlauf signalisiert.
Mit dem achten HL-Zustand setzt nach erfolgtem erzwungenen (datengetriebenen) Sprung die Analyse des zweiten Objekts ein (sichtbar auch an dem geringen mittleren Aktivitätsniveau der HL-Knoten). Interessant ist bei diesem Abtastverlauf die mit dem vierten Abtastschritt auf dem zweiten Objekt (HL-Zustand 11) konstant auftretende, maximale Hypothese (mit Kreuz markiert). Dies kommt dadurch zustande, daß in der EFM lediglich Merkmalsinformation repräsentiert wird, da die Positionsinformation bereits durch die Wechselwirkung von FTM und TLM berücksichtigt ist. Da auf dem zweiten Objekt (abgesehen von der Positionsinformation) immer der gleiche Merkmalsübergang erfolgt (auf dem linken Objekt zwischen den beiden gelben Füßen), wurde dieser in der EFM durch zwei Columnen abgebildet. Die Aktivität dieser beiden Columnen, die bis zur Beendigung des Analyseprozesses des zweiten Objekts konstant bleibt, führt zur konstanten, maximalen Aktivierung des entsprechenden HL-Knotens.
Auf konzeptionelle Aspekte, die sich aus der Wechselwirkung der beiden attentiven Ebenen ergeben, wird im Kapitel 6 noch einmal eingegangen.

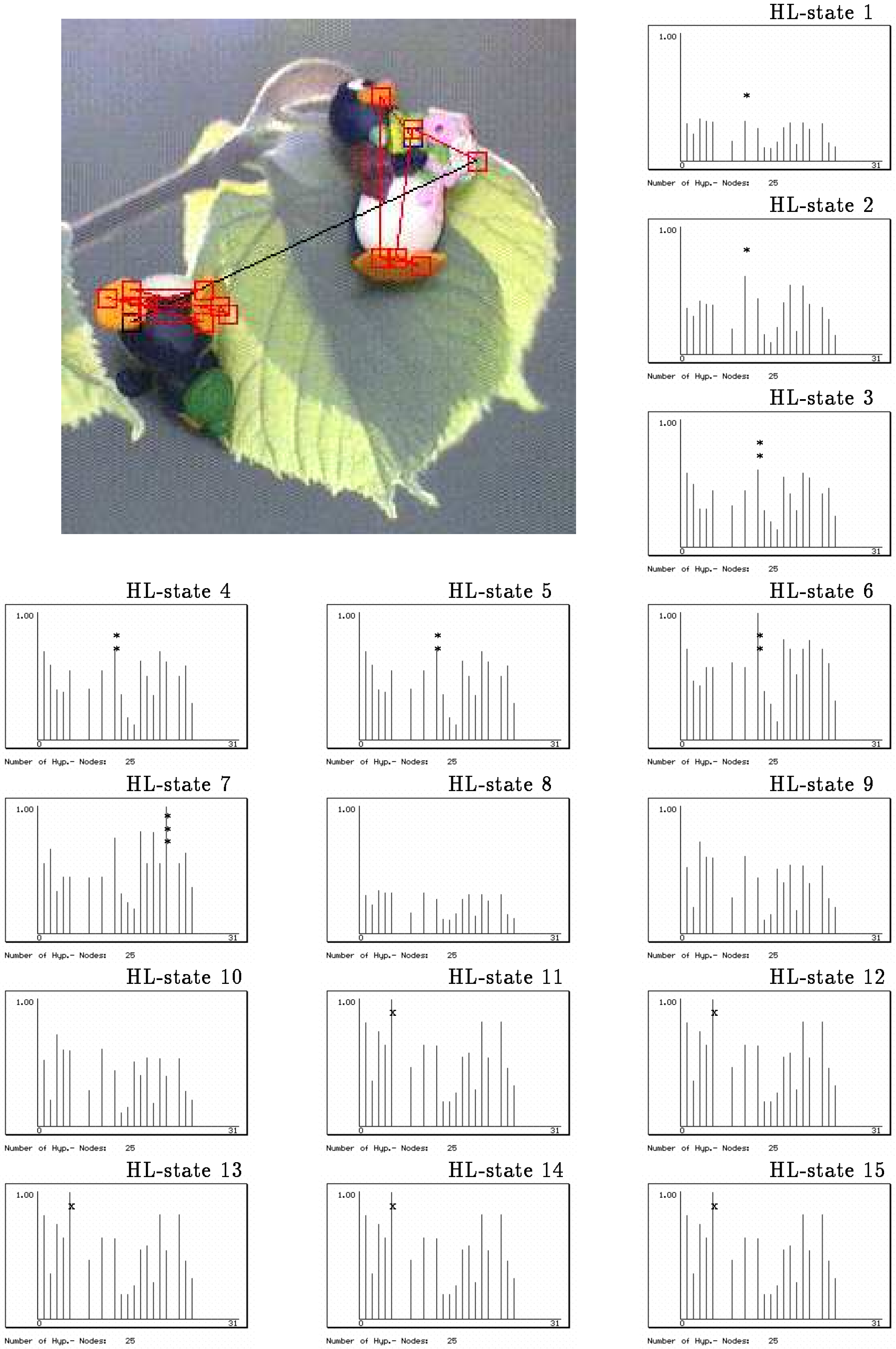


Abbildung 5.27: *Abfolge der Aktivitätsverteilung über die HL-Knoten entsprechend den nacheinander ausgeführten internen Abtastbewegungen; nach dem auffälligkeitsbasiert selektierten ersten Abtastpunkt auf dem rechten Objekt wird zunächst dieses Objekt, nach dem erzwungenen Sprung das linke Objekt wissensbasiert analysiert (weitere Erläuterungen im Text)*

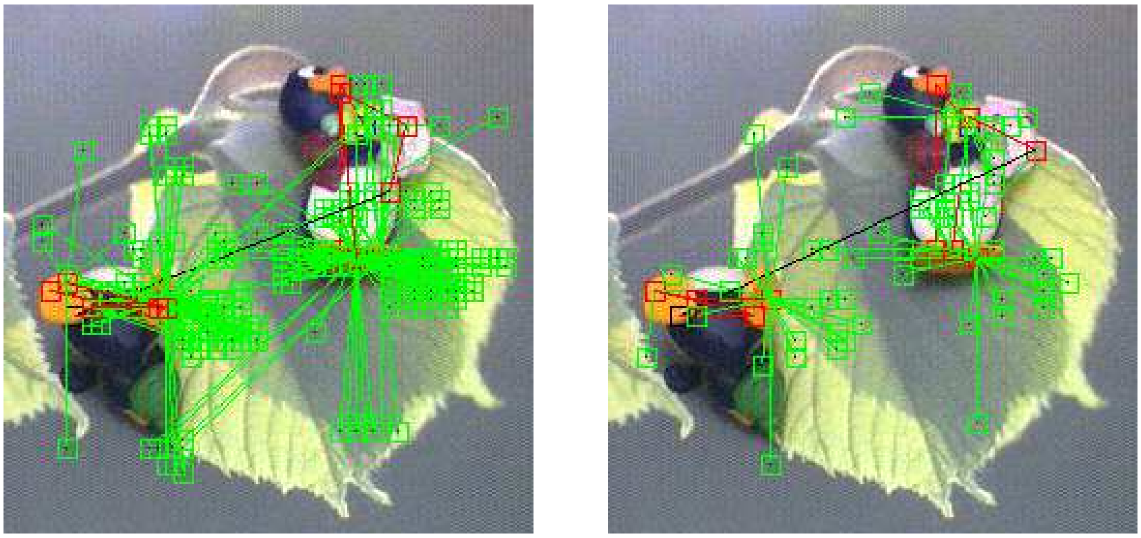


Abbildung 5.28: *Reduktion der nicht verifizierbaren Suchsprünge (korrespondierend zu Abbildung 5.27) durch Wechselwirkung beider attentiver Ebenen; links wurde das Abtastverhalten bei ausschließlicher Nutzung des lokalen Umweltwissens der ersten attentiven Ebene protokolliert, während rechts beide attentive Ebenen das Abtastverhalten modulieren*

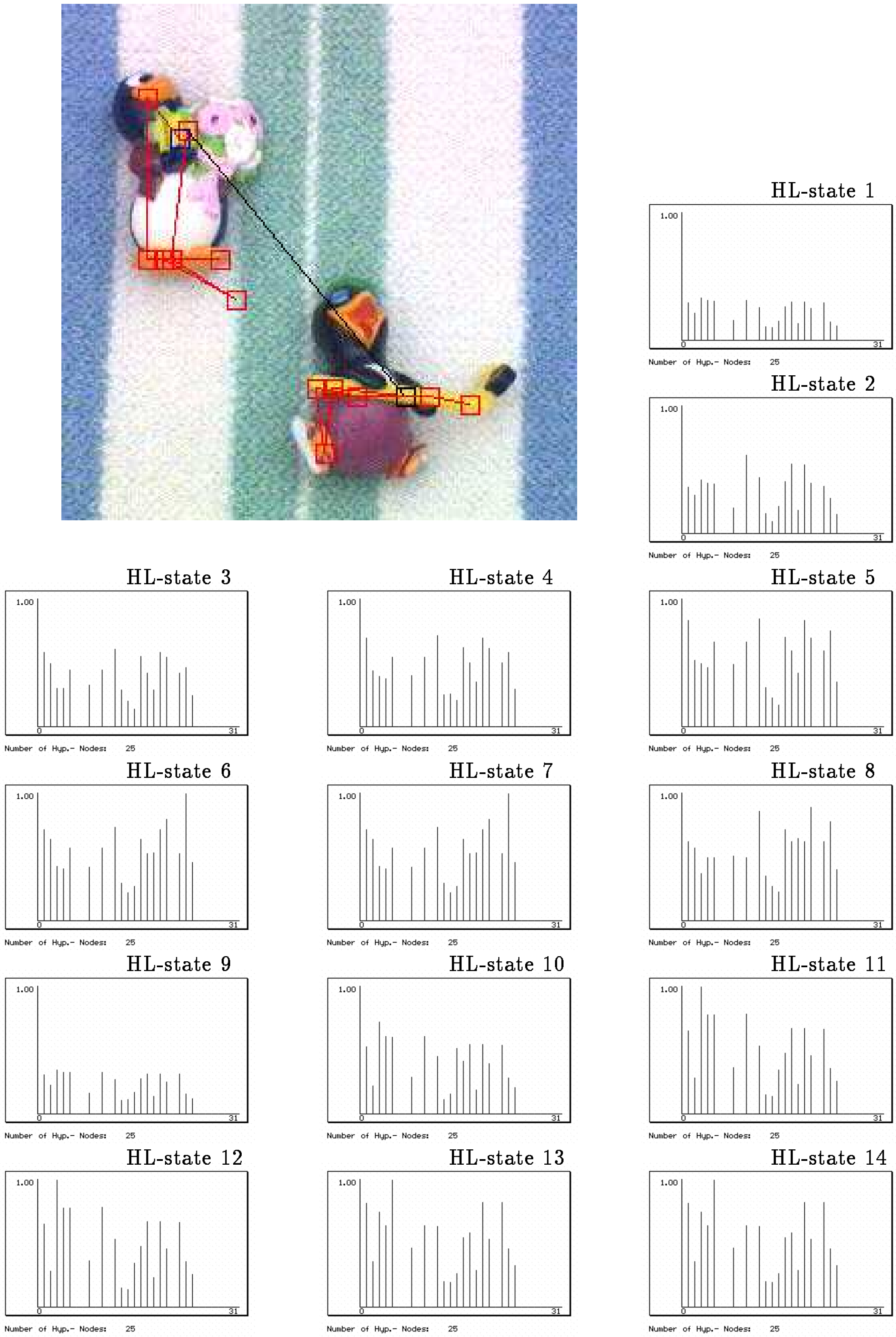


Abbildung 5.29: *Abfolge der Aktivitätsverteilung über die HL-Knoten entsprechend den nacheinander ausgeführten internen Abtastbewegungen; nach dem auffälligkeitsbasiert selektierten ersten Abtastpunkt auf dem linken Objekt wird zunächst dieses Objekt, nach dem erzwungenen Sprung das rechte Objekt wissensbasiert analysiert*

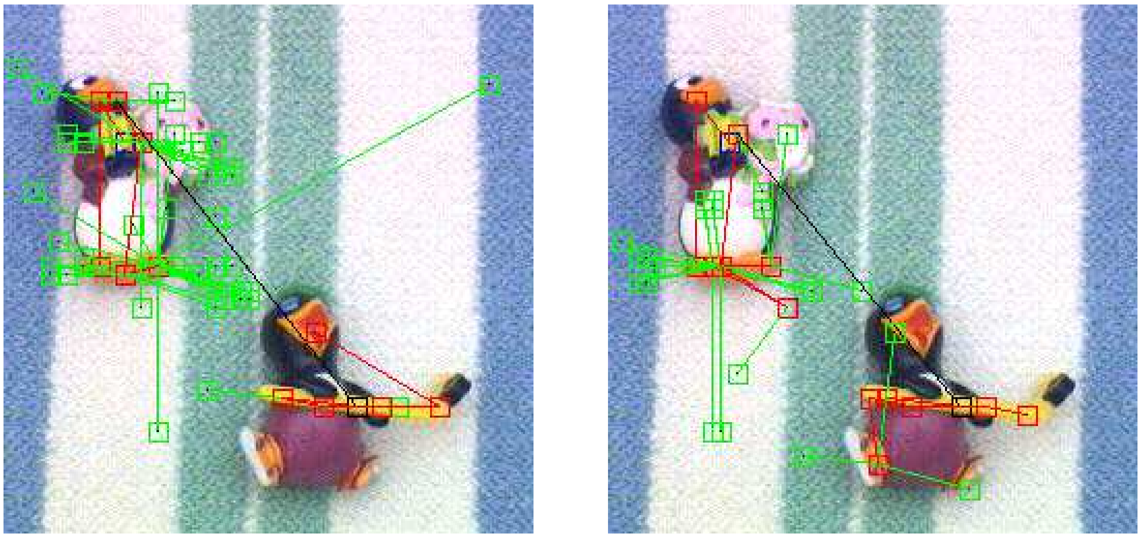


Abbildung 5.30: *Reduktion der nicht verifizierbaren Suchsprünge (korrespondierend zu Abbildung 5.29) durch Wechselwirkung beider attentiver Ebenen; links wurde das Abtastverhalten bei ausschließlicher Nutzung des lokalen Umweltwissens der ersten attentiven Ebene protokolliert, während rechts beide attentive Ebenen das Abtastverhalten modulieren*

5.4 Detaillierte Beschreibung der zweiten attentiven Ebene

5.4.1 Berechnungsvorschriften für die Knoten der einzelnen Ebenen

Berechnungsvorschrift für *sum*-Knoten

$$\begin{aligned} y_{ij}^{sum}(t) &= F\Big\{\Big(w_{ijkl}^{sum-nftm}(t)\cdot y_{kl}^{ftm}(t-1)+w^{sum-nin}\cdot y_{ij}^{nin}(t-1)\Big) \\ &+ f\Big[w^{sum-cin}\cdot y_{ij}^{cin}(t-1)+w_{ijk}^{sum-cftm}\cdot y_k^{ftm}(t-1) \\ &+ \sum_{mn} w_{ijmn}^{sum-hyp}(t)\cdot y_{mn}^{hyp}(t-1)\Big]\Big\} \end{aligned}$$

mit

$$f(x)=\begin{cases} x & : \; 0\le x\le \Theta_{sum} \\ \Theta_{sum} & : \; x>\Theta_{sum} \end{cases}$$

$$\Theta_{sum}=0.95$$

$$F(x)=\begin{cases} x & : \; 0\le x\le 1 \\ 1 & : \; x>1 \end{cases}$$

Die unspezifische Erregung *cftm* bzw. *cin* (siehe Abschnitt 5.4.2 und Abbildung 5.24) darf am *sum*-Knoten eine im Vergleich zum spezifischen Input *nftm* bzw. *nin* nur geringe Aktivierung bewirken. Ebenso wie der *hyp*-Knoten soll auch der *sum*-Knoten nur dann maximal aktiv werden können, wenn der entsprechende spezifische Input *nftm* vorlag. Die beiden Gewichtswerte $w^{sum-nftm}$ und $w^{sum-cftm}$ werden nur einmalig mit Langzeitspeichercharakteristik modifiziert. Die intracolumnaren Gewichte (intracolumnare Parameter) $w^{sum-cin}$ und $w^{sum-nin}$ werden mit den entsprechenden Werten fest eingestellt.

Die gleichzeitige Aktivierung des *sum*-Knotens durch *cftm* bzw. *cin* und interne Projektionen von *hyp*-Knoten soll bei fehlendem spezifischen Input auf keinen Fall dazu führen, daß der *sum*-Knoten maximal aktiv wird, damit eindeutig getrennt werden kann, aufgrund welcher Erregungsquelle der *sum*-Knoten aktiviert wurde (Konzept der separierbaren Aktivitätsbereiche). Diese Forderung wird realisiert, indem die Summe der Einflüsse von *cftm* bzw. *cin* und *hyp* auf einen *sum*-Knoten durch die Funktion f auf den Wert 0.95 begrenzt wird.

Die Funktion F ist notwendig, um den Gesamteinfluß aller gleichzeitig möglichen Eingänge auf den *sum*-Knoten auf 1 zu begrenzen. Zur Erläuterung dient die folgende Abbildung, in der die in dieser Verarbeitungsebene verwendeten, separierbaren Aktivitätsbereiche graphisch dargestellt sind.

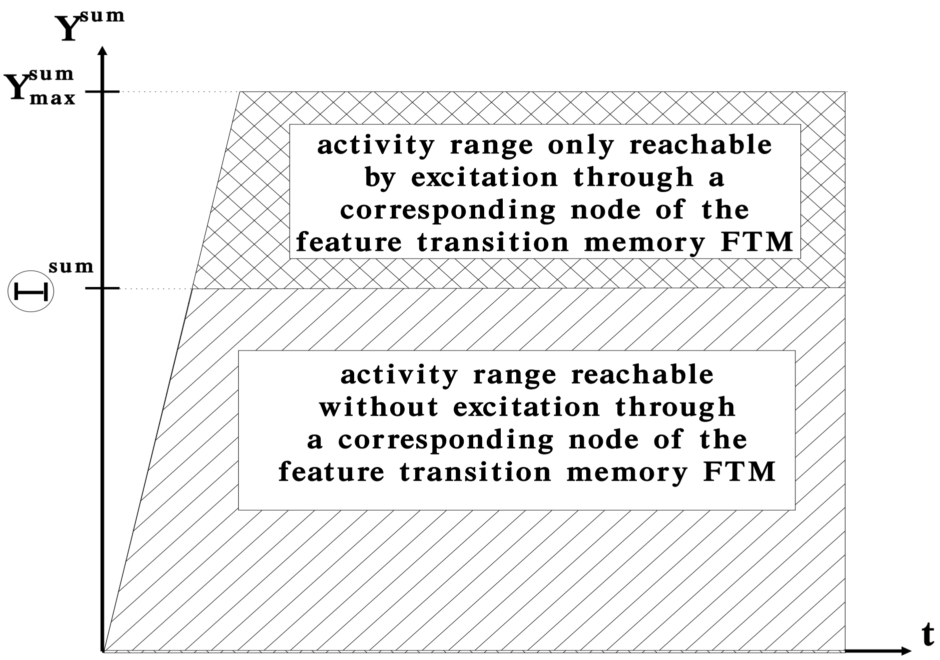


Abbildung 5.31: *Graphische Veranschaulichung der separierbaren Aktivitätsbereiche an einem sum-Knoten einer EFM-Columne*

Berechnungsvorschrift für *hyp*-Knoten

$$\begin{aligned} y_{ij}^{hyp}(t) &= w_{ijkl}^{hyp-nftm}(t) \cdot y_{kl}^{ftm}(t-1) + w_{ijk}^{hyp-cftm}(t) \cdot y_{k}^{ftm}(t-1) \\ &+ w^{hyp-nin} \cdot y_{ij}^{nin}(t-1) + w^{hyp-cin} \cdot y_{ij}^{cin}(t-1) \end{aligned}$$

Die beiden Gewichtswerte $w^{hyp-nftm}$ und $w^{hyp-cftm}$, die auch hier nur einmalig modifiziert werden und danach statisch sind (siehe Abschnitt 5.4.2), werden so festgelegt, daß der unspezifische Feature-Input (der mit angenommener Aktivität 1 vom AFI bzw. vom TFI kommt) den *hyp*-Knoten in die Lage versetzt, mit einem Viertel seiner maximal möglichen Aktivität Projektionen an andere *sum*-Knoten zu senden. Wird ein mit dem *hyp*-Knoten korrespondierender FTM-Knoten maximal aktiv, so liegt für den *hyp*-Knoten eine spezifische Erregung *nftm* vor, sodaß dieser entsprechend maximal aktiviert wird (siehe Abschnitt 5.4.2). Die Koppelparameter (intracolumnar) $w^{hyp-cin}$ und $w^{hyp-nin}$ werden ebenfalls mit den äquivalenten Werten fest eingestellt (vorgegebene Basisstruktur der EFM-Columne).

Berechnungsvorschrift für *exp*-Knoten

Die Einführung der Schwelle für die *exp*-Knoten erwies sich als relativ schwierig, da an ihnen die Rückprojektionen mehrerer HL-Knoten zusammenlaufen. Damit ergäbe sich der Effekt, daß mehrere schwach auf einen *exp*-Knoten einwirkende HL-Knoten die gleiche Wirkung bzgl. der Rückprojektion in den FTM erzielen wie ein einziger, stark rückprojizierender HL-Knoten. Um dies zu vermeiden, wurde eine Aktivierungsfunktion implementiert, die die Aktivität des *exp*-Knotens auf die Aktivität des jeweils am stärksten einwirkenden HL-Knotens festlegt.
Allerdings bestand ein Problem darin, daß die Ausbildung von Verbindungen zwischen ftm- und *exp*-Knoten (Rückprojektion der Erwartungen) erfolgt, noch bevor die *exp*-Knoten überhaupt aktiviert wurden. Damit ließe sich keine Hebb'sche Lernregel anwenden. Um zu sichern, daß bei Aktivierung eines FTM-Knotens auch ein bzw. zwei *exp*-Knoten aktiviert werden, wurde eine Aktivierung der *exp*-Knoten durch den spezifischen Input *nftm* bzw. *nin* implementiert. Um dabei eindeutig unterscheiden zu können, ob der *exp*-Knoten von HL-Knoten oder FTM-Knoten erregt wird, wurde die Funktion f eingeführt, die den Einfluß der HL-Projektionen auf den *exp*-Knoten auf den Wert 0.95 begrenzt. Man erhält somit wieder zwei separierbare Aktivitätsbereiche. Nur wenn ein *exp*-Knoten maximal aktiv ist, werden die Verbindungen $w^{exp-ftm}$ und $w^{ftm-exp}$ modifiziert. Der Einfluß von y^{nin} auf den *exp*-Knoten soll nur ein enable-Signal zur Modifikation der Gewichte w^{exp-hl} und w^{hl-exp} liefern und sich nicht unmittelbar auf die Aktivität des *exp*-Knotens auswirken, da diese nur von den HL-Aktivierungen abhängen soll. Ein Sonderfall mußte dabei jedoch noch abgefangen werden, bei dem das enable-Signal zur Modifikation der Gewichte w^{exp-hl} und w^{hl-exp} durch $y^{nin} = 1$ vorliegt, der *exp*-Knoten aber noch gar keine Projektionen von HL-Knoten erhält und demzufolge inaktiv ist. Deshalb wurde ein sehr kleines Gewicht $w^{exp-nin}$ implementiert, welches für diesen Fall eine Aktivität $y^{exp} > 0$ sicherstellt (siehe auch die entsprechende Lernregel).

Damit ergibt sich:

$$\begin{aligned} y_{ij}^{exp}(t) &= F\Big(f\Big(\max\Big(\Big(w_{ijm}^{exp-hl}(t)\cdot y_m^{hl}(t-1)\Big)\,\forall m\Big)\Big) \\ &+ w_{ijkl}^{exp-nftm}(t)\cdot y_{kl}^{ftm}(t-1) + w^{exp-nin}\cdot y_{ij}^{nin}(t-1)\Big) \end{aligned}$$

mit

$$F(x) = \begin{cases} 0 & : \quad x < Schwelle \\ x & : \quad Schwelle \leq x \leq 1 \\ 1 & : \quad x > 1 \end{cases}$$

und

$$f(x) = \begin{cases} x & : \quad 0 \leq x \leq 0.95 \\ x & : \quad x > 0.95 \end{cases}$$

Berechnungsvorschrift für HL-Knoten

$$y_{ij}^{hl}(t) = \frac{\sum_{mn} w_{ijmn}^{hl-sum}(t)\cdot y_{mn}^{sum}(t-1)}{\sum_{mn} w_{ijmn}^{hl-sum}(t)\cdot y_{mn}^{sum}(t-1) + d_i(\underline{\tilde{w}}_{ij}^{hl-sum}, \underline{\tilde{y}}^{sum})}$$

mit

$$d_i(\underline{\tilde{w}}^{hl-sum}, \underline{\tilde{y}}^{sum}) = \sum_{mn}\Big(\tilde{w}_{imn}^{hl-sum}(t) - \tilde{y}_{mn}^{sum}(t-1)\Big)^2$$

$$\tilde{w}_{ijmn}^{hl-sum}(t) = \begin{cases} 1 & : \quad w_{ijmn}^{hl-sum}(t) > 0 \\ 0 & : \quad w_{ijmn}^{hl-sum}(t) = 0 \end{cases}$$

$$\tilde{y}_{mn}^{sum}(t-1) = \begin{cases} 1 & : \quad y_{mn}^{sum}(t-1) = 1 \\ 0 & : \quad y_{mn}^{sum}(t-1) < 1 \end{cases}$$

Bei der Entwicklung der Aktivierungsfunktion für die HL-Knoten waren folgende Aspekte zu berücksichtigen:

- Es soll keine Normierung sowohl der Gewichte von den *sum*- zu den HL-Knoten als auch der *sum*-Aktivität vorgenommen werden, da eine solche Normierung erstens biologisch nicht motiviert werden kann und sie sich zweitens auch in funktioneller Hinsicht negativ auswirkt, da die sich unterschiedlich ausprägenden *sum*-Aktivierungen durch eine globale Normierung wieder geglättet werden würden.

- Die Aktivität eines HL-Knotens soll proportional dazu sein, wie gut der gerade ausgeführte Abtastpfad mit dem durch den HL-Knoten gespeicherten Abtastpfad übereinstimmt.

- Werden mehr *sum*-Knoten durch den gerade vollzogenen Abtastpfad aktiviert, als zu einem bestimmten HL-Knoten gehören, so muß sich das ebenso als Verringerung der betreffenden Aktivität bemerkbar machen wie der Fall, daß weniger *sum*-Knoten in den aktuellen Abtastpfad einbezogen waren, als zu dem betreffenden HL-Knoten gehören müßten. Dies ist als Teilmengenproblem bekannt und spielte bei der Entwicklung der ART-Architekturen von Carpenter und Grossberg ebenfalls eine zentrale Rolle. Daraus läßt sich unmittelbar ableiten, daß eine Skalarproduktaktivierung ohne vorherige Normierung der beiden Faktoren allein nicht ausreicht, da mit dieser Aktivierung der Fall von zuviel auftretenden *sum*-Knoten nicht detektiert werden kann.

- Um zu detektieren, ob zuviele oder zuwenige *sum*-Knoten maximal aktiviert wurden, als der aktuell betrachtete HL-Knoten zusammenfaßt, bietet sich die Verwendung eines Abstandsmaßes auf der Basis des euklidischen Abstandes an, auf dessen Grundlage dann die Aktivität eines HL-Knotens bestimmt wird. Die ausschließliche Verwendung des Abstandsmaßes würde aber den Nachteil mit sich bringen, daß sich für alle HL-Knoten mit dem gleichen Abstand zwischen Gewichtsvektor ($\underline{w}^{hl-sum}$) und Eingangsvektor (aktive *sum*-Knoten) identische Aktivitätswerte ergeben würden, unabhängig davon, wieviele EFM-Columnen von den einzelnen HL-Knoten zusammengefaßt werden. Dies ist jedoch insofern nicht sinnvoll, da sich bei einem HL-Knoten, der z.B. 10 *sum*-Knoten zusammenfaßt, ein Abstand von 1 weniger auswirken sollte als bei einem HL-Knoten, der nur 2 *sum*-Knoten zusammenfaßt.

Die implementierte Aktivierungsfunktion berücksichtigt im enthaltenen Skalarprodukt die Anzahl der durch den HL-Knoten kodierten *sum*-Knoten (bzw. EFM-Columnen). Sind zum Beispiel alle Gewichte stark ausgeprägt (nahe der Sättigung), so ist der Wert des Skalarprodukts identisch mit der Anzahl der durch den HL-Knoten repräsentierten *sum*-Knoten. Durch die Summe von Skalarprodukt und Abstand im Nenner wird folgendes erreicht: Ist der Abstand Null, so handelt es sich bei den maximal aktivierten *sum*-Knoten genau um die, die auch von diesem HL-Knoten repräsentiert werden. Der Aktivitätswert des HL-Knotens ist dann maximal. Die Aktivität verringert sich umso stärker, je größer der Abstand dem Eingangs- und Gewichtsvektor wird (Über- oder Unterrepräsentationen). Der Grad des Einflusses des Abstandes hängt dabei wie gewünscht davon ab, wieviele *sum*-Knoten der HL-Knoten repräsentiert.

5.4.2 Lernregeln der zweiten attentiven Ebene

Bei der Zuordnung der EFM-Columnen zu den aktiven FTM-Knoten (spezifischer Input *nftm*) wird folgendermaßen verfahren: Existiert im EFM bei Auftreten eines aktiven FTM-Knotens keine Columne, deren Knoten maximal erregt werden können, so bedeutet das, daß der durch das FTM gerade selektierte Übergang noch nicht im EFM repräsentiert wurde. Für diesen Fall wird aus den durch den Feature-Input *cftm* bereits vorselektierten EFM-Clustern jeweils eine Columne **zufällig** ausgewählt. Zu diesen beiden Columnen werden dann die in die Columne führenden Links w^{-nftm} bzw. w^{-cftm}

ausgebildet (in einem Schritt auf den Wert 1 hochgelernt und dann unverändert belassen). Die Links $w^{ftm-exp}$ werden ebenfalls nach dem beschriebenen Verfahren angelegt.

Lernregel für die Verbindungen $\mathbf{w^{cin-cftm}}$ und $\mathbf{w^{nin-nftm}}$

$$w_{ijkl}^{nin-nftm}(t) = \begin{cases} 1 & : \text{ zum cin-Knoten der zufällig ausgewählten EFM-Columne} \\ 0 & : \text{ sonst} \end{cases}$$

$$w_{ijk}^{cin-cftm} = \begin{cases} 1 & : \text{ für alle Verbindungen des ftm-Clusters } k \text{ zu allen } \textit{cin}\text{-Knoten } ij \\ 0 & : \text{ sonst} \end{cases}$$

Lernregel für die Verbindungen $\mathbf{w^{hyp-cftm}}$ und $\mathbf{w^{hyp-cin}}$

$$w_{ijk}^{hyp-cftm} = \begin{cases} 0.25 & : \text{ für alle Verbindungen des ftm-Clusters } k \\ & : \text{ zu allen } \textit{hyp}\text{-Knoten } ij \\ 0 & : \text{ sonst} \end{cases}$$

$$w^{hyp-cin} = 0.25$$

Lernregel für die Verbindungen $\mathbf{w^{sum-cftm}}$ und $\mathbf{w^{sum-cin}}$

$$w_{ijk}^{sum-cftm} = \begin{cases} 0.25 & : \text{ für alle Verbindungen des ftm-Clusters } k \\ & : \text{ zu allen } \textit{hyp}\text{-Knoten } ij \\ 0 & : \text{ sonst} \end{cases}$$

$$w^{sum-cin} = 0.25$$

Lernregel für die Verbindungen $\mathbf{w^{hyp-nftm}}$ und $\mathbf{w^{hyp-nin}}$

$$w_{ijkl}^{hyp-nftm}(t) = \begin{cases} 0,75 & : \text{ zum } \textit{hyp}\text{-Knoten der zufällig ausgewählten EFM-Columne} \\ 0 & : \text{ sonst} \end{cases}$$

$$w^{hyp-nin} = 0.75$$

Lernregel für die Verbindungen $\mathbf{w^{sum-nftm}}$ und $\mathbf{w^{sum-nin}}$

$$w_{ijkl}^{sum-nftm}(t) = \begin{cases} 0,75 & : \text{ zum } \textit{sum}\text{-Knoten der zufällig ausgewählten EFM-Columne} \\ 0 & : \text{ sonst} \end{cases}$$

$$w^{sum-nin} = 0.75$$

Lernregel für die Verbindungen $\mathbf{w}^{\mathbf{exp-nftm}}$ und $\mathbf{w}^{\mathbf{exp-nin}}$

$$w_{ijkl}^{exp-nftm}(t) = \begin{cases} 1 & : \text{ zum exp-Knoten der zufällig ausgewählten EFM-Columne} \\ 0 & : \text{ falls die Verbindung nicht existiert} \end{cases}$$

$$w^{exp-nin} = 0.01$$

Lernregel für die Verbindungen $\mathbf{w}^{\mathbf{sum-hyp}}$

$$w_{ijmn}^{sum-hyp}(t) = w_{ijmn}^{sum-hyp}(t-1) + \triangle w_{ijmn}^{sum-hyp}(t)$$

$$\triangle w_{ijmn}^{sum-hyp}(t) = \eta \cdot \tilde{y}_{mn}^{hyp}(t) \cdot \tilde{y}_{ij}^{sum}(t) \cdot aff(w_{ijmn}^{sum-hyp}, t)$$

$$\tilde{y}_{ij}^{sum}(t) = \begin{cases} 1 & : \quad y_{ij}^{sum}(t) > 0.95 \\ 0 & : \quad y_{ij}^{sum}(t) < 0.95 \end{cases}$$

$$\tilde{y}_{mn}^{hyp}(t) = \begin{cases} 1 & : \quad y_{mn}^{hyp}(t) = 1 \\ 0 & : \quad y_{mn}^{hyp}(t) < 1 \end{cases}$$

Dabei handelt es sich um die intracolumnaren Verbindungen, die zum Aufbau der Parallelhypothesen innerhalb der EFM (Columnen-Assemblies) verwendet werden. Die Lernregel soll sichern, daß nur zwischen den *sum*- bzw. *hyp*-Knoten Verbindungen etabliert werden, die auch wirklich einen spezifischen Input (*nftm*) durch den FTM erhalten haben. Ansonsten würden auch Verbindungen zwischen den Knoten aufgebaut, die nur durch interne Hypothesenbildung innerhalb des EFMs und/oder nur durch den Feature-Input (*cftm*) vom FTM aktiviert wurden.
Durch die Verwendung der beiden Hilfsgrößen $\tilde{y}^{sum}$ bzw. $\tilde{y}^{hyp}$ wird gesichert, daß die *hyp*- und *sum*-Knoten erst dann ihre Gewichte modifizieren können, wenn sie ihre jeweilige Maximalaktivität erreicht haben (Nutzung des Konzepts der separierbaren Aktivitätsbereiche).

Da die unterschiedliche Ausprägung der Gewichte für die EFM-interne parallele Hypothesenbildung eine wesentliche Rolle spielt, wurde bewußt auf eine Begrenzung des Wertebereichs der Gewichte verzichtet (Hebb'sches Lernen), die Inkremente zur Gewichtserhöhung sind deshalb entsprechend klein gewählt. Selbst sehr große Gewichtswerte wirken sich nicht störend aus.

Auf einen weiteren Aspekt muß an dieser Stelle noch eingegangen werden. Bei der angegebenen Lernregel würde folgender Effekt eintreten: Da generell das Lernen der Gewichte $w^{sum-hyp}$ zwischen zwei datengetriebenen Sprüngen im FTM zu jedem Zeitschritt erfolgt, wird beispielsweise das Gewicht zwischen den beiden zuerst aktivierten *hyp*- bzw. *sum*-Knoten stärker inkrementiert als die Gewichte zwischen den später aktivierten Knoten. Um diesen Effekt auszuschließen war zu sichern, daß das Gewicht zwischen einem *hyp*- und einem *sum*-Knoten innerhalb von zwei datengetriebenen Sprüngen nur ein einziges mal modifiziert wird. Dies wird dadurch erreicht, indem ein spezieller Mechanismus eingeführt wurde, der dafür sorgt, daß die entsprechende Synapse auf den ersten Lernvorgang sehr stark und auf alle weiteren Lernvorgänge innerhalb der aktuellen Abtastsequenz nicht mehr reagiert. Zu diesem Zweck wird der Term $aff(w_{ijmn}^{sum-hyp}, t)$ eingeführt, der die Affinität der Synapse beschreibt. Beim globalen Reset aller Aktivitäten im EOM (nach einem datengetriebenen Sprung oder nach Szenenwechsel) wird dieser Term wieder von Null auf Eins gesetzt. Er stellt also eine Art Kurzzeitspeicher an der Synapse dar.

Lernregel für die Verbindungen $\mathbf{w^{hl-sum}}$

$$w_{ijmn}^{hl-sum}(t) = w_{ijmn}^{hl-sum}(t-1) + \triangle w_{ijmn}^{hl-sum}(t)$$

$$\triangle w_{ijmn}^{hl-sum}(t) = \eta \cdot y_{ij}^{hl}(t) \cdot \left[\tilde{y}_{mn}^{sum}(t) - w_{ijmn}^{hl-sum}(t)\right]$$

mit

$$\tilde{y}_{mn}^{sum}(t) = \begin{cases} 1 & : \quad y_{mn}^{sum}(t) > 0.95 \\ 0 & : \quad y_{mn}^{sum}(t) < 0.95 \end{cases}$$

Die Lernregel sichert, daß nur die Gewichte der Verbindungen zwischen maximal aktiven (d.h. durch den spezifischen Input *nftm* erregten) *sum*-Knoten und dem entsprechenden HL-Knoten modifiziert werden. Ansonsten würden auch Verbindungen zwischen HL-Knoten und solchen *sum*-Knoten aufgebaut, die nur durch interne Hypothesenbildung innerhalb der EFM und/oder nur durch den Feature-Input (*cftm*) vom FTM aktiviert wurden.
Was die Bereitstellung der $\tilde{y}^{sum}$ betrifft, so wird wie im vorhergehenden Abschnitt geschehen argumentiert.

Lernregel für die Verbindungen $\mathbf{w^{exp-hl}}$

$$w_{ijmn}^{exp-hl}(t) = w_{ijmn}^{exp-hl}(t-1) + \triangle w_{ijmn}^{exp-hl}(t)$$

$$\triangle w_{ijmn}^{exp-hl}(t) = \eta \cdot \tilde{y}_{ij}^{exp}(t) \cdot (\tilde{y}_{mn}^{hl}(t) - w_{ijmn}^{exp-hl}(t))$$

mit

$$\tilde{y}_{ij}^{exp}(t) = \begin{cases} 1 & : \quad y_{ij}^{exp}(t) \geq 0.95 \\ 0 & : \quad y_{ij}^{exp}(t) = 0 \end{cases}$$

und

$$\tilde{y}_{mn}^{hl}(t) = \begin{cases} 1 & : \quad y_{mn}^{hl}(t) = max(y_{mn}^{hl}(t)) \forall m, n \\ 0 & : \quad y_{mn}^{hl}(t) < max(y_{mn}^{hl}(t)) \end{cases}$$

Diese Lernregel sichert, daß nur die Gewichte der Verbindungen zwischen HL-Knoten und maximal aktiven (d.h. durch den spezifischen Input *nftm* bzw. *nin* erregten) *exp*-Knoten verstärkt werden, andernfalls erfolgt eine Abschwächung.

Lernregel für die Verbindungen $\mathbf{w}^{\mathbf{ftm-exp}}$

$$w_{ijmn}^{ftm-exp}(t) = w_{ijmn}^{ftm-exp}(t-1) + \triangle w_{ijmn}^{ftm-exp}(t)$$

$$\triangle w_{ijmn}^{ftm-exp}(t) = (1 - w_{ijmn}^{ftm-exp}(t-1)) \cdot \tilde{y}_{mn}^{exp}(t) \cdot y_{ij}^{ftm}(t)$$

mit

$$\tilde{y}_{mn}^{exp}(t) = \begin{cases} 1 & : \quad y_{mn}^{exp}(t) = 1 \\ 0 & : \quad y_{mn}^{exp}(t) < 1 \end{cases}$$

Es handelt sich dabei um eine simple Hebb-Regel mit Einschrittlernen und zwangsweiser Konvergenz der Gewichtswerte auf den Wert 1.

5.5 Wechselwirkung beider attentiver Ebenen

Die Wechselwirkung zwischen beiden attentiven Ebenen bei der Auswahl des erfolgversprechendsten nächsten Sprungs vollzieht sich im FTM, der den Träger der zeitlichen Objektrepräsentation bildet. Dabei kommt es zur Überlagerung von lokalen Hypothesen, die aufgrund des bislang akkumulierten Umweltwissens der ersten attentiven Ebene generiert werden können, und globalen Hypothesen der zweiten attentiven Ebene, die ein mehr objektbezogenes Umweltwissen (Wissen über komplexe visuelle Strukturen als gemeinsam in einer Szene auftretende, unmittelbar aufeinanderfolgende erfolgreiche Sprünge) darstellen.

Die Erwartungshaltung der zweiten attentiven Ebene wird durch die aktuelle Aktivitätsverteilung über die HL-Knoten repräsentiert. Die HL-Knoten projizieren ihre Aktivität über die *exp*-Knoten der EFM-Columnen an die entsprechenden FTM-Knoten und stellen dort einen zusätzlichen erregenden Einfluß dar. Die Berechnungsvorschrift für die Aktivität eines FTM-Knotens, die bereits im Abschnitt 5.2.4 im Rahmen der Behandlung der ersten attentiven Ebene erläutert wurde, wird hier wieder aufgegriffen und wie folgt erweitert:

$$y_{ij}^{ftm}(t) = \alpha \cdot H_{l_{ij}}(t) + (1 - \alpha) \cdot H_{g_{ij}}(t)$$

mit

$$0 \leq \alpha \leq 1$$

In dieser vereinfachten Notation symbolisiert H_l die **l**okale **H**ypothese, die sich aus der Berechnungsvorschrift für die FTM-Knoten innerhalb der ersten attentiven Ebene ergibt. H_g symbolisiert die **g**lobale **H**ypothese vom EOM, die der Aktivität des auf den betrachteten FTM-Knoten projizierenden *exp*-Knotens einer EFM-Columne entspricht (siehe Abschnitt 5.4.1). Der WTA-Mechanismus innerhalb des entsprechenden FTM-Clusters findet erst nach der gewichteten Summation von lokaler und globaler Hypothese statt.

Entscheidend für die Funktionalität beider attentiver Ebenen ist die Wahl des Parameters α, der die Gewichtung von lokaler und globaler Hypothese bei der Sprungauswahl bestimmt. Bei der Untersuchung der Wechselwirkung von lokalen und globalen Hypothesen im FTM trat der Effekt auf, daß bei sofortigem parallelen Wirken beider attentiver Ebenen für verschiedene Werte von α weder in der ersten noch in der zweiten Ebene ein zufriedenstellendes Ergebnis des Selbstorganisationsprozesses erreichbar war. Dies kommt dadurch zustande, daß zu Beginn des Analyseprozesses das bislang akkumulierte Umweltwissen beider Ebenen unsicher ist und somit einerseits keine bereits als stabil detektierten Merkmalsübergänge von der ersten zur zweiten Ebene gelangen können und damit andererseits in der zweiten Ebene keine stabilen Objekt- oder Teilobjektkandidaten organisiert werden können.
Nach umfangreichen Tests ergab sich, daß die besten Ergebnisse beim Zusammenwirken

beider attentiver Ebenen dann erreicht werden, wenn die erste attentive Ebene bereits über gesichertes Umweltwissen verfügt, bevor der Selbstorganisationsprozeß der zweiten attentiven Ebene einsetzt. Bei der konkreten Simulation wurde dies durch eine einmalige Präsentation der Szenen ohne Mitwirkung des EOM realisiert, woran sich eine weitere einmalige Präsentation der gesamten Galerie mit stattfindender Wechselwirkung beider attentiver Ebenen anschloß. Weitere konzeptionelle und bewertende Überlegungen dazu finden sich im Kapitel 6.

Kapitel 6

Diskussion des Ansatzes und Ausblick

6.1 Wertung des entwickelten Modellansatzes

Zu Projektabschluß kann eingeschätzt werden, daß die im Rahmen des Vorhabens verfolgten *wesentlichen Projektziele* erreicht wurden. Diese Ziele waren insbesondere:

- die Entwicklung einer beobachtbaren Verhaltensänderung des Gesamtsystems als Ausdruck eines erworbenen impliziten Objektverständnisses
- die *Effektivierung des Abtastverhaltens* (Scan-Effektivierung) als Ausdruck dafür, daß die *inhärenten Gesetzmäßigkeiten* der Objekte in den Szenen
  - während des *datengetriebenen Abtastens* erfolgreich *beobachtet*,
  - durch *aktives Lernen* gezielt *extrahiert*,
  - geeignet *intern repräsentiert* (Qualitäten, Relationen)
  - in *Hypothesize-Verify-Zyklen* (Reinforcement) aktiviert und selektiv genutzt werden können.

Es entstand ein leistungsfähiger Modellansatz, der auch in *stark strukturierter visueller Umgebung* eine implizite *Objekt/Objekt-* und *Objekt/Hintergrundseparierung* durch *zeitliche Bindung* stabiler, objektspezifischer Zusammenhänge ermöglicht und das Erlernen *unbekannter Zusammenhänge* ohne eine spezielle *überwachte Trainingsphase* erlaubt.

Hervorgehoben werden soll die Entwicklung eines leistungsfähigen Selbstorganisationskonzeptes, welches durch die kontinuierliche Wechselwirkung des Systems mit seiner Umwelt (sensorisch gesteuerte interne Simulation) das System in die Lage versetzt, ein verhaltensorientiertes, implizites Umweltverständnis ohne explizit vorgegebene Annahmen über die Objekte zu entwickeln.
Die implementierte Modellarchitektur kann zu großen Teilen biologisch motiviert werden und beinhaltet eine funktionell hierarchisch organisierte Wissensrepräsentation, die die (hypothetisch vorausgesetzte) Isomorphie von Analyseprozeß und interner Repräsentation berücksichtigt und eine vertikale Kontrollstruktur aufweist, was der untergeordneten Ebene bei Ausfall der übergeordneten Ebene eine, wenn auch eingeschränkte, Funktionsfähigkeit ermöglicht.

6.2 Verwertbare Teilergebnisse des Projektes

Für nachfolgende, sowohl applikations- als auch grundlagenorientierte Projektvorhaben sind die folgenden erbrachten Teilleistungen von besonderer Relevanz:

- *Neuartige Farbraumtransformation* in einen neurobiologisch motivierten Farbdifferenzraum
- auf dieser Farbraumtransformation basierende, *autonom adaptierende Farbumstimmung* bei schwankenden Beleuchtungsverhältnissen incl. der Optimierung des Helligkeitssignals
- weitere spezielle *visuelle Merkmalskarten* für die Auffälligkeitsanalyse und die Identifikation lokaler Bildgebiete
  - Texturanalyse über der korrigierten Helligkeitskarte mittels selbstorganisierter, eigenvektorbasierter rezeptiver Felder
- *Biologisch motiviertes Auffälligkeitssystem* mit folgenden Verarbeitungsleistungen
  - *Reproduzierbarkeit* des datengetriebenen Abtastverlaufs bei der Analyse natürlicher Farbszenen – auch bei schwankenden Beleuchtungsverhältnissen und veränderlichen Szenarien
  - *Robustheit* der implementierten Auffälligkeitsmechanismen, insbesondere durch die Farbumstimmung im Farbraum
- *leistungsfähiges Selbstorganisationskonzept* zur Erzielung eines verhaltensorientierten, impliziten Umweltverständnisses durch
  - Annahme einer *Isomorphie* von Analyseprozeß und interner Repräsentation
  - Kopplung von datengetriebenen (auffälligkeitsbasierten) und wissensbasierten Aufmerksamkeitsprozessen
  - neurobiologisch und psychophysisch hinterlegbare *hierarchische Wissensrepräsentation*
  - *sensorisch gesteuerte interne Simulation* als kontinuierliche Interaktion zwischen einkommendem Datenstrom und systeminternem Entscheidungsprozeß

6.3 Kritische Diskussion und Ausblick

Neben der bereits mehrfach begründeten Einschränkung, nicht vordergründig ein in allen Belangen exakt biologieäquivalentes Modell visueller Aufmerksamkeitsprozesse zu entwickeln, ergaben sich im Verlauf der Bearbeitung einige Aspekte, die bis zum Projektabschluß nur unzureichend bzw. unbefriedigend gelöst werden konnten. Auf die aus unserer Sicht wesentlichsten offenen Probleme soll in diesem Abschnitt eingegangen werden.

6.3.1 Verhaltensorganisation im Gesamtsystem

Im Rahmen des Projektvorhabens ergab sich eine dreistufige Vorgehensweise im Selbstorganisationsprozeß: Die erste Stufe, die die Erzeugung auffälligkeitsbasierter Abtastbewegungen beinhaltet, schränkt die Komplexität der Umwelt in einem ersten Schritt ein. In der zweiten Stufe, verkörpert durch die erste attentive Ebene der Gesamtarchitektur, wird die Komplexität der Umwelt durch die Extraktion des räumlich lokalen, aber noch objekt-unspezifischen Umweltwissens weiter verringert. Die zweite attentive Ebene als dritte und gleichzeitig höchste Verarbeitungsstufe faßt das Ergebnis der ersten beiden Stufen zu einem objekt-spezifischen Umweltwissen zusammen, was zu einer weiteren Komplexitätsreduktion führt.

Als verbesserungswürdig wird vom Auftragnehmer in erster Linie die Wechselwirkung zwischen beiden attentiven Ebenen angesehen. Da die beiden Ebenen relativ unabhängig voneinander (ausgenommen die Definition der notwendigen Schnittstellen) entwickelt und erst in der letzten Projektphase zu einem attentiven Gesamtsystem verschmolzen wurden, konnte deren konzeptionelle und funktionelle Einheit noch nicht optimal hergestellt werden, obwohl die mit beiden attentiven Ebenen erzielten Simulationsergebnisse den gewählten Ansatz prinzipiell bestätigen.

Ein offenes Problem betrifft die Interaktion der Lernprozesse zwischen den beiden attentiven Ebenen. So konnte noch nicht abschließend untersucht werden, wie stark das Umweltwissen der ersten attentiven Ebene ausgeprägt sein muß, damit die zweite attentive Ebene ihrerseits Wissen organisieren und den Hypothesenbildungsprozeß beider Ebenen sinnvoll verbessern kann. Die derzeitige Vorgehensweise (siehe Abschnitte 5.3.4 und 5.5) stellt lediglich einen ersten Ansatz dar. Insbesondere eine günstigere Wahl der Gewichtung des Einflusses von lokalen und globalen Hypothesen bzw. eine mögliche adaptive Zeitabhängigkeit dieser Gewichtung mit fortschreitendem Selbstorganisationsprozeß versprechen eine Verbesserung der Performance beider attentiver Ebenen.
Mit den eben genannten Fragestellungen setzt sich eine Thematik auseinander, die im Rahmen der Landesgraduiertenförderung Thüringen am Fachgebiet aktuell bearbeitet wird.

6.3.2 Auffälligkeitsorientierte Verhaltensorganisation

Die stark auffälligkeitsorientierte Verhaltensorganisation, insbesondere innerhalb der ersten attentiven Ebene, erschwert teilweise den Selbstorganisationsprozeß. Dies stellt jedoch ein eher konzeptionelles Problem dar, da die impliziten Objektrepräsentationen und die damit verbundene verhaltensorientierte „Detektion“ oder „Erkennung“ von Objekten derzeit nicht in einem darauf basierenden Zielverhalten (z.B. zielgerichtetes Greifen eines erkannten Objekts) münden. Dieses fehlende Zielverhalten führte dazu, daß keine von einem solchen Zielverhalten ableitbaren Abbruchkriterien für den Analyseprozeß über einer Szene oder einem Objekt zur Verfügung standen und damit quasi künstlich generiert werden mußten. Da diese Abbruchkriterien derzeit auffälligkeitsbasiert ermittelt werden, kann der interne Interpretationsprozeß an dieser Stelle nur ungenügend Einfluß auf das Verhalten des Gesamtsystems nehmen.

Der erste Abtastpunkt einer Szene wird derzeit immer, also auch nach der Akkumulation von Wissen, auffälligkeitsbasiert bestimmt. Daraus folgt unmittelbar, daß die verwendeten Objekte zumindest Regionen aufweisen müssen, die auffälliger sind als der Hintergrund. Gilt diese Voraussetzung nicht, so kommt es dazu, daß die Objekte nicht oder nur unzureichend erlernt werden können, was auf eines der fünf im Vorhaben verwendeten Objekte auch teilweise zutraf. Geht man auch hier wieder von einem auf der Erkennung von Objekten basierenden Zielverhalten aus, so erscheint es sinnvoll, eine wissensbasierte Modulation der jetzt rein datengetriebenen Auffälligkeit einzuführen. Damit würden die Szenenbestandteile, nach denen ausgehend von einer bestimmten Zielstellung gerade gesucht wird, im Vergleich zu anderen Szenenbestandteilen auffälliger. Dieser Ansatz würde eine Erweiterung des bislang genutzten Konzepts der räumlich-zeitlichen Aufmerksamkeit um Konzepte der qualitativen Aufmerksamkeit unter Nutzung von Kontextwissen und Erfahrung bedeuten.

6.3.3 Grenzen der Architektur

Sowohl die Form als auch die Größe des Aufmerksamkeitsfensters können mit der hier vorgestellten Architektur nicht variiert werden. Obwohl diese Einschränkung bei der Definition der Randbedingungen für den hier gewählten Ansatz formuliert wurde, soll hier festgehalten werden, daß die Architektur eine Erweiterung um diese Aspekte derzeit nicht zuläßt bzw. in größerem Umfang konzeptionell überarbeitet werden müßte.

Die Organisation von parallelen Repräsentationen für Objekte bzw. Teilobjekte innerhalb des EOM würde die Möglichkeit bieten, wissensbasiert wieder parallel auf verschiedene Bereiche der Szene im Sinne einer *parallelen selektiven Aufmerksamkeit* (random access) zuzugreifen und sich damit zumindest teilweise vom Searchlight-Ansatz zu lösen. Auch diese mögliche Erweiterung läßt die gegenwärtige Architektur nicht ohne weiteres zu. Einschränkend muß an dieser Stelle angemerkt werden, daß seitens von Psychophysik und Neurobiologie zunächst fundamental neue Konzepte bereitgestellt werden müßten, zumindest was den experimentellen Nachweis derartiger Phänomene betrifft.

6.4 Im Rahmen des Projekts entstandene Publikationen

Gesamtarchitektur und konzeptionelle Einbindung

1995

1. **H.-J. Böhme, U.-D.Braumann, H.-M.**
A neural network model for scene interpretation. In: *Proc. of the Third Annual SNN Symposium on Neural Networks*, Nijmegen, The Netherlands, pp. 173-176, 1995

2. **H.-M. Gross, H.-J. Böhme, U.-D. Braumann, D. Heinke, T. Pomierski**
A behaviour-oriented approach to an implicit "object-understanding" in visual attention. to appear in: *Proc. of ICNN'95, IEEE-Intern. Conf. on Neural Networks*, Perth, 1995

1994

3. **H.-M. Groß, H.-J. Böhme, D. Heinke, T. Pomierski, R. Möller**
Self-Organizing a Behaviour-Oriented Interpretation of Objects in Active-Vision. In: *Marinaro & Morasso (eds.), Proc. of the ICANN'94*, Vol. 1, pp. 58-61, Springer, 1994

1992

4. **H.-M. Groß, E. Körner, H.-J. Boehme, T. Pomierski.**
A Neural Network Hierarchy for Data and Knowledge Controlled Selective Visual Attention. In: *I. Aleksander and J. Taylor (eds.), Proc. of the ICANN'92*, Vol. 1, pp. 825-828, North-Holland, 1992

5. **H.-J. Böhme, E.Körner, H.-M. Groß, R. Möller.**
A Neural Network Architecture which uses Cellular Automata to get Context-Specific Associative Memory. In: *I. Aleksander and J. Taylor (eds.), Proc. of the ICANN'92*, Vol. 2, pp. 1381-1384, North-Holland, 1992

6. **H.-M. Groß, H.-J. Böhme, D. Heinke, R. Möller, T. Pomierski.**
Steuerung parallel-sequentieller Verarbeitungsprozesse und Strukturierung dynamischer Repräsentationen. In: *Informatik-Fachberichte,* pp. 97-108, BMFT-Neuroinformatik-Statusseminar, 1992

1991

7. **H.-M. Groß, E. Körner, T. Pomierski.**
GNOM – A Modular Network Architecture for Adaptive Parallel–Sequential Pattern Recognition. In: *T. Kohonen, Mäkisara, Simula, Kangas, (eds.), Proceedings of the ICANN'91*, Vol. 1, pp. 747-752, North-Holland, 1991

8. **E. Körner, H.-M. Groß, H.-J. Böhme.**
Elementary Cognitive Mechanisms for Knowledge Based Image Interpretation. In: *Proceedings of 3. International Workshop on Adaptive Learning and Neural Networks*, Schloß Reisensburg, pp. 163-181, 1991

Frühe visuelle Verarbeitung und Auffälligkeitssteuerung

1994

9. **H.-M. Groß, H.-J. Böhme, K. Debes.**
A Neural Network Hierarchy for Data and Knowledge Controlled Selective Visual Attention. In: *Chinese Journal of Biomedical Engineering*, 4, pp. 363-368, 1994

10. **H.-M. Groß, D. Heinke, H.-J. Böhme, T. Pomierski**
A Behaviour-oriented Approach to an „Object-understanding“ in Visual Attention. In: *Proc. of 39th Int. Scient. Coll. Ilmenau*, pp. 95-103, 1994

11. **T. Pomierski, H.-M. Groß**
Feature Extraction using Self-Organizing Receptive Fields. In: *Proc. of 39th Int. Scient. Coll. Ilmenau*, pp. 117-123, 1994

1993

12. **D. Heinke, H.-M. Groß**
A Simple Selforganizing Neural Network Architecture for Selective Visiual Attention. In: *Gielen & Kappen (eds.), ICANN '93 – Proc. of the Int. Conf. on Artificial Neural Networks*, Amsterdam, pp. 63-66, Springer, 1993

13. **T. Pomierski, H.-M. Groß, D. Wendt**
A Distributed Multicolumnar System for Primary Cortical Analysis of Real-World Scenes. In: *Gielen & Kappen (eds.), ICANN '93 – Proc. of the Int. Conf. on Artificial Neural Networks*, Amsterdam, pp. 142-147, Springer, 1993

14. **T. Pomierski, H.-M. Groß, D. Wendt**
Ein verteiltes modulares System zur eigenvektorbasierten Analyse natürlicher Szenen mit selbstorganisierten rezeptiven Feldern. In: *Pöppel & Handels (Hrsg.), DAGM'93 – Mustererkennung im Dienste der Gesundheit*, S. 281-288, Springer, 1993

15. **T. Pomierski, H.-M. Groß**
Merkmalsextraktion mit selbstorganisierenden, eigenvektorbasierten rezeptiven Feldern. In: *Tagungsmappe Neuro-Mini-Messe – Tools und Anwendungen Neuronaler Netze*, Erlangen-Tennenlohe, Bayerisches Forschungszentrum für Wissensbasierte Systeme, 1993

1992

16. **H.-M. Groß, R. Franke, H.-J. Böhme, Claudia Beck.**
A Hierarchical Neural Network for Data Driven and Knowledge Controlled Selective Visual Attention. In: *Proc. of DAGM'92*, pp. 341-346, Springer, 1992

17. **H.-M. Groß, E. Körner.**
A Hierarchical Neural Network Model for Selective Visual Attention. In: *Proceedings of International Conference on Modelling Problems in Bionics – BIOMOD'92*, St. Petersburg, Russia, 1992

1992

18. **H.-J. Böhme, H.-M. Groß, E. Körner.**
Ein active-vision System zur wissensbasierten Bildinterpretation. In: *36th Int. Scient. Coll. Ilmenau,*, S. 656-661, 1991

Robuste Merkmalsgewinnung und Farbumstimmung

1995

19. **T. Pomierski, H.-M. Groß**
Verfahren zur empfindungsgemäßen Farbumstimmung. In: *40th Int. Scient. Coll. Ilmenau, 1995*

20. **T. Pomierski, H.-M. Groß**
Verfahren zur empfindungsgemäßen Farbumstimmung.
In: *Tagungsband der DAGM'95*, Bielefeld, 1995

Neuronale Architekturen zur on-line Wissensorganisation

1995

21. **U.-D. Braumann, H.J. Böhme, H.-M. Groß**
An Episodic Knowledge Base for Object Understanding. In: *Proc. of ESANN'95*, pp. 175-180, Brussels, 1995

1994

22. **H.-J. Böhme, U.-D. Braumann, H.-M. Groß**
A Neural Network Architecture for Episodic Feature Representation. In: *Proc. of 39th Int. Scient. Coll. Ilmenau*, pp. 11-18, 1994

23. **H.-J. Böhme, U.-D. Braumann, H.-M. Groß**
A Neural Network Architecture for Sensory Controlled Internal Simulation. In: *Marinaro & Morasso (eds.), ICANN '94 – Proc. of Int. Conf. on Artificial Neural Networks*, Sorrento, Italy, Vol. 2, pp. 1189-1192, Springer, 1994

Literaturverzeichnis

[Abeles 1991] Abeles,M. (1991). *Corticonics — Neural Circuits of the Cerebral Cortex.* Cambridge University Press.

[Andersen et al. 1989] Andersen, R.A. & J. Gnadt (1989). *Posterior parietal cortex.* in: WURTZ, R.H. & M. GOLDBERG, (eds.): *The Neurobiology of Saccadic Eye Movements*, Chapter 8, pp. 315–334. Elsevier Science Publisher.

[Arbib 1989] Arbib,M.A. (1989). *The Metaphorical Brain 2 - Neural Networks and Beyond.* John Wiley and Sons, New York.

[Beer 1991] Beer, R.D. (1991). *Intelligence as Adaptive Behavior.* Academic Press.

[Bergen et al. 1983] Bergen, J.R. & B. Julesz (1983). *Rapid Discrimination of Visual Patterns.* IEEE Transactions on System, Man and Cybernetics, 13(5), pp. 857–863.

[Boehme et al. 1994] H.-J. Boehme, U.-D. Braumann, H.-M. Gross. (1994). *A Neural Network Architecture for Sensory Controlled Internal Simulation.* In: Proc. of ICANN-94, Vol. 2, pp. 1189-1192

[Bosch 1992] Bosch, K. (1992). *Statistik-Taschenbuch.* Oldenbourg.

[Braddick et al. 1978] Braddick, O., Campbell, F. W. and Atkinson, J. (1978). *Channels in vision: basic aspects.* In: Handbook of Sensory Physiology VII. eds. Held, R., Leibowitz, H. W., Teuber, H. L., pp. 3-38.

[Brown et al. 1964] Brown, P. K. and Wald, G. (1964). *Visual pigments in single rods and cones of the human retina.* Science, 144, pp. 45-51.

[Burke et al. 1978] Burke, W., Cole, A. M. (1978). *Extra-retinal influences on the lateral geniculate nucleus.* Review of Physiology, Biochemistry and Pharmacology, 80, pp. 105-166.

[Burton et al. 1986] Burton, G.J., N. Haig, & I. Moorhead (1986). *A Self-Similar Stack Model for Human and Machine Vision.* Biological Cybernetics, 53:397–403.

[Colby 1991] Colby, C.L. (1991). *The Neuroanatomy and Neurophysiology of Attention.* J. of Child Neurology, 6.

[Carpenter et al. 88] Carpenter, G.A., Grossberg, S. (1988). *The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network.* IEEE Computer 3(88), pp. 77-88

[Corbetta et al. 1991] Corbetta, Maurizio, F. M. Miezin, S. Dobmeyer, G. L. Shulman, & S. E. Petersen (1991). *Selective and Divided Attention during Visual Discriminations of Snape, Color, and Speed: Functional Anatomy by Positron Emission Tomography.* The Journal of Neuroscience, 11(8):2383–2402.

[Creutzfeldt et al. 1979] Creutzfeldt, O. D., Lee, B. B. and Elepfandt A. (1979). *A quantitative study of chromatic organisation and receptive fields of cells in the lateral geniculate body of the rhesus monkey.* Exp. Brain Res., 35, pp. 527-545.

[Crick 1984] Crick,F. (1984). *Function of the thalamic reticular complex: The searchlight hypothesis*. Proc.Natl.Acad.Sci.UsS.A., 81:4586–4590.

[Desimone et al. 1989] Desimone, R. & L. Ungerleider (1989). *Neural mechanisms of visual processing in mokeys*.

[Desimone et al. 1990] Desimone, R., M. Wessinger, L. Thomas, & W. Schneider (1990). *Attentional Control of Visual Perception: Cortical and Subcortical Mechanisms*. in: *Cold Spring Habor Symposia on Quantitative Biology*, pp. 963–971. Cold Spring Habor Laboratory Press.

[De Valois et al. 1958] De Valois, R. L., Smith, C. J., Kitai, S. T., Katoly, S. J. (1958). *Responses of single cells in different layers of the primate lateral geniculate nucleus to monochromatic light.* Science, 127, pp. 238-239.

[De Valois et al. 1966] De Valois, R. L., Abramov, I, Jacobs, G. H. (1966). *Analysis of response patterns of LGN cells.* Journal of the Optical Society of America, 56, pp. 966-977.

[De Valois et al. 1980] De Valois, R. L., De Valois, K. K. (1980). *Spatial Vision.* Ann. Rev. Psychiol., 31, pp. 309-341.

[De Valois et al. 1982] De Valois, R. L., Albrecht, D. G., Thorell, L. G. (1982). *Spatial frequency selectivity of cells in macaque visual cortex.* Vision Res., 22, pp. 545-559.

[De Valois et al. 1993] De Valois, R. L., De Valois, K. K. (1993). *A multi-stage color model.* Vision Res., 3, pp. 1053-1065.

[Dixon 1978] Dixon, E. R. (1978). *Spectral distribution of Australian daylight.* J. Opt. Soc. Am., 68, pp. 437-450.

[Duhamel et al. 1992] Duhamel, J.-R., L. Carol, & M. Goldberg (1992). *The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements*. Science, 255:90–92.

[D'Zmura et al. 1986] D'Zmura, M., Lennie, P. (1986). *Mechanisms of color constancy.* J. Opt. Soc. Am., 3A, pp. 1662-1672.

[Enns et al. 1990] Enns, J.T. & R. A. Rensink (1990). *Influence of Scene-Based Properties on Visual Search*. Science, 247:721–723.

[Eriksen et al. 1986] Eriksen, C.W. & J.D.S. James (1986). *Visual attention within and around the field of focal attention: A zoom lens model*. Perception and Psychophysics, 40(4):225–240.

[Essen et al. 1992] Essen, van D.C., C.H. Anderson, & D.J. Felleman (1992). *Information Processing in the Primate Visual System: An Integrated Systems Perspective*. Science, 255:419–423.

[Fujita et al. 1992] Fujita, I., Tanaka, K., Ito, M., Cheng, K. (1992). *Columns for visual features of objects in monkey inferotemporal cortex.* In: Nature (360), pp. 343-346

[Goldberg et al. 1989] Goldberg, M.E. & Segraves, M.A. (1989). *The visual and frontal cortices*. in: WURTZ, R.H. & M. GOLDBERG, (eds.): *The Neurobiology of Saccadic Eye Movements*, Chapter 7, pp. 283–315. Elsevier Science Publisher.

[Gouras et al. 1979] Gouras, P., Krüger, J. (1979). *Responses of cells in foveal visual cortex of the monkey to pure color contrast.* J. Neurophysiol., 42, pp. 850-860.

[Groß 1989] Groß,H.-M. (1989). *Simulation eines Arrays corticaler Prozessoren zur inhaltsgesteuerten parallelen Informationsverarbeitung nach dem Vorbild des primären visuellen Cortex*. PhD Thesis, Technische Hochschule Ilmenau.

[Groß et al. 1992a] Groß,H.-M., Böhme,H.-J., Heinke,D., Möller,R., and Pomierski,T. (1992). *Steuerung parallel-sequentieller Verarbeitungsprozesse und Strukturierung dynamischer Repraesentationen.* In: Informatik-Fachberichte, BMFT-Neuroinformatik-Statusseminar, pp. 97–108

[Hallet et al. 1988] Hallet, P., Jepson, A., Gershon, G. (1988). *Color constancy is at the level of the cones.* Investigative Ophthalmology and Visual Science (Supplement), 29, pp. 163.

[Hancock 1992] Hancock, P. (1992). *The principal components of natural images.* Network, 3,

[Haralick et al. 1993] Haralick, R.M. & L. Shapiro (1993). *Computer and Robot Vision*, Vol. 2. Addison-Wesley.

[Hebb 1949] Hebb,D.O. (1949). *The Organization of Behaviour*. Wiley, New York.

[Heinke et al. 1993] D. Heinke, H.-M. Gross. (1993). *A Simple Selforganizing Neural Network Architecture for Selective Visiual Attention.* In: Proc. of ICANN-93, pp. 63-66

[Hering 1905] Hering, E. (1905). *Grundzüge der Lehre vom Lichtsinn.* In Handbuch der gesamten Augenheilkunde. ed. Saemisch, T. Leipzig: Verlag von Wilhelm Engelmann.

[Humphreys et al. 1993] Humphreys, Glyn W. & H. J. Müller (1993). *SEarch via Recursive Rejection (SERR): A Connectionist Model of Visual Search*. Cognitive Psychology, 25:43–110.

[Jaehne 1989] Jaehne, B. (1989). *Digitale Bilderverarbeitung*. Springer.

[Judd 1945] Judd, D. B. (1945). *Standard response functions for protanopic and deuteranopic vision.* J. Opt. Soc. Am., 35, pp. 199-220.

[Judd et al. 1964] Judd, D. B., MacAdam, D. L. and Wyszecki, G. (1964). *Spectral distribution of typical daylight as a function of correlated color temperature.* J. Opt. Soc. Am., 54, pp. 1031-1040.

[Julesz 1991] Julesz, B. (1991). *Early vision and focal attention*. Reviews of Modern Physics, 63(3):735–772.

[Kaelbling 1993] Kaelbling, L.P. (1993). *Learning in Embedded Systems*. The MIT Press.

[Kaneko 1970] Kaneko, A. (1970). *Physiological and morphological identification of horizontal, bipolar and amacrine cells in goldfish retina.* J. Physiol. (Lond.), 207, pp. 623-633.

[Klein 1988] Klein, R. (1988). *Inhibitory tagging system facilitates visual search*. Nature, 334:430–431.

[Koch et al. 1985] Koch,C. & Ullman,S. (1985). *Shifts in selective visual attention: towards the underlying neural circuitry*. Human Neurobiology, 4:219–227.

[Koenderink et al. 1979] Koenderink, J.J. & A. v. Doorn (1979). *The Internal Representation of Solid Shape with Respect to Vision.* Biological Cybernetics, 32:211–216.

[Kueppers 1973] Küppers, H. (1973). *Color - Origin, system, uses.* New York, London, Toronto, Melbourne: Van Nostrand Reinhold Ltd.

[Kupfermann 1991] Kupfermann, I. (1991). *Localization of Higher Cognitive and Affective Functions: The Assiciation Cortices.* In: Candel, E.S., Schwartz, J.H., Jessell, T.M. (eds.) Principles of Neural Science, Third Edition, Appleton & Lange, pp. 823–838

[LaBerge et al. 1990] LaBerge, D. & M. Buchsbaum (1990). *Positron Emission Tomographic Measurements of Pulvinar Activity During an Attention Task.* J. of Neuroscience, 10(2):613–619.

[Land 1983] Land, E. H. (1983). *Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image.* Proc. Natl. Acad. Sci. USA, 80, pp. 5163-5169.

[Land et al. 1983a] Land, E. H., Hubel, D. H., Livingstone, M. S., Perry, S. H. and Burns, M. S. (1983). *Colour-generating interactions across the corpus callosum.* Nature, 303, pp. 616-618.

[Lang 1978] Lang, H. (1978). *Farbensehen.* In: Farbmetrik und Farbensehen, 1. Auflage. Gottwald, A. (ed.), pp. 263-325. München, Wien: Oldenbourg.

[Laws 1980] Laws, K. I. (1983). *Rapid texture identification.* Proceedings of the SPIE Conference on Image Processing for Missile Guidance, San Diego, pp. 376-380.

[Laws 1985] Laws, K. I. (1985). *Goal-directed textured-image segmentation.* Applications of Artificial Intelligence II, SPIE, 548, 19-26.

[Lennie et al. 1988] Lennie, P., D'Zmura, M. (1988). *Mechanisms of color vision.* Critical Reviews of Neurobiology, 3, pp. 333-400.

[Lennie et al. 1991] Lennie, P., Haake, P. W., Williams, D. R. (1991). *The design of chromatically opponent receptive fields.* In: Computational models of visual processing. Landy, M. S. and Movshon, J. A. (eds.), Cambridge, Mass.: MIT Press.

[Levick et al. 1980] Levick, W. R., Thibos, L. N. (1980). *Orientation bias of cat retinal ganglion cells.* Nature, 286, pp. 389-390.

[Livingstone et al. 1984a] Livingstone, M. S. and Hubel, D. H. (1984). *Anatomy and physiology of a color system in the primate visual cortex.* J. Neurosci., 4, pp. 309-356.

[Livingstone et al. 1984b] Livingstone, M. S. and Hubel, D. H. (1984). *Specificity of intrinsic connections in primate primary visual cortex.* J. Neurosci., 4, pp. 2830-2835.

[Maffei et al. 1977] Maffei, L. and Fiorentini, A. (1977). *Spatial frequency rows in the striate visual cortex.* Vision Res., 17, pp. 257-264.

[Maloney 1986] Maloney, L. (1986). *Evaluation of linear models of surface spectral reflectance with small number of parameters.* J. Opt. Soc. Am., 3A, pp. 1673-1683.

[Mallot et al. 1992] Mallot,H.A., Kopecz,J. & von Seelen,W. (1992). *Neuroinformatik als empirische Wissenschaft*. Kognitionswissenschaft, 3(1):12–23.

[Malsburg et al. 1986] Malsburg,C.v.d. & Bienenstock,E. (1986). *Statistical Coding and Short-Term Synaptic Plasticity: A Scheme for Knowledge Representation in the Brain.* in: Bienenstock,E., (ed.): Disordered Systems and Biological Organization, pp. 248 – 272. Springer, Berlin, Heidelberg.

[Marks et al. 1964] Marks, W. B., Dobelle, W. H. and MacNichol, E. F., Jr. (1964). *Visual pigments of single primate cones.* Science, 143, pp. 1181-1183.

[Marr 1982] Marr,David (1982). *Vision.* Freeman, San Francisco.

[Miller et al. 1994] Miller, E.K. & R. Desimone (1994). *Parallel Neuronal Mechanisms for Short-Term Memory*. Science, 263:520–522.

[Mishkin et al. 1983] Mishkin, M., L. Ungerleider, & K. Macko (1983). *Object vision and sptial vision: two cortical pathways*. Trends Neuroscience, 6:414–417.

[Moran et al. 1985] Moran, J. & R. Desimone (1985). *Slective Attention Gates Visual Processing in the Extrastriate Cortex*. Science, 229:782–784.

[Neiser 1967] Neiser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts, New York.

[Nothdurft 1993a] Nothdurft, H.-Chr. (1993a). *Saliency Effects Across Dimensions in Visual Search*. Vision Res., 33(5/6):839–844.

[Nothdurft 1993b] Nothdurft, H.-Chr. (1993b). *The Role of Features in Preattentive Vision: Comparison of Orientation, Motion and Color Cues*. Vision Res., 33(14):1937–1958.

[Oja 1982] Oja, E. (1982). *A simlified neuron model as a principal component analyzer.* J. Math. Biology, 15, pp. 267-273.

[Olshausen et al. 1993] Olshausen, B., C. Anderson, & v. D. Essen (1993). *A Neural Model of Visual Attention and Invariant Pattern Recognition*. CNS Memo 18.

[Olshausen 1994] Olshausen, B.A. (1994). *Neural Routing Circuits for Forming Invariant Representations of Visual Objects*. Masters Thesis, California Institute of Technology.

[Perry et al. 1984] Perry, V. H., Oehler, R. and Cowey, A. (1984). *Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey.* Neuroscience, 12, pp. 1110-1123.

[Posner et al. 1984] Posner, M. I., J. A. Walker, F. J. Friedrich, & R. D. Rafal (1984). *Effects of parietal injury on convert orienting of attention.* J. of Neuroscience, 4(7):1863–1874.

[Posner et al. 1990] Posner, M.I. & S. Petersen (1990). *The attention system of the human brain*. Annu. Rev. Neurosci., 13:25–42.

[Posner et al. 1980] Posner, M.I., C. Snyder, & B. Davidson (1980). *Attention and the detection of signals*. J. Exp. Psychol., 109:160–174.

[Posner et al. 1988] Posner, Michael I., S. E. Petersen, P. T. Fox, & M. E. Raichle (1988). *Localization of Cognitive Operations in the Human Brain.* Science, 240:1627–1631.

[Remington et al. 1984] Remington, R. & L. Pierce (1984). *Moving attention: Evidence for time-invariant shifts of visual selective attention*. Perception and Psychophysics, 35(4):393–399.

[Ritter et al. 1991] Ritter, H., Schulten, K., Martinetz, T. (1991). *Neuronale Netze.* Bonn: Addison-Wesley.

[Robinson et al. 1989] Robinson, D.L. & J. McClurkin (1989). *The visual superior colliculus and pulvinar*. in: WURTZ, R.H. & M. GOLDBERG, (eds.): *The Neurobiology of Saccadic Eye Movements*, Chapter 9, pp. 337–358. Elsevier Science Publisher.

[Robson 1975] Robson, J. G. (1975). *Receptive fields: neural representation of the spatial and intensive attributes of the visual image.* In: Handbook of Perception V. eds. Carterette, E. C. and Freedman, M. P., pp. 81-112. New York: Academic Press.

[Rolls et al. 1995] Rolls, E.T. & Tovee, M.J. (1995). *Sparseness of the Neural Representation of Stimuli in the Primate Temporal Visual Cortex.* Journal of Neurophysiology, 73(2), pp. 713-725

[Ruderman et al. 1994] Ruderman, D.L., Bialek, W. (1994). *Statistivs of Natural Images: Scaling in the Woods.* Physical Review Letters, 73(6), pp. 814-817

[Rumelhart et al. 1988b] Rumelhart,D.E., McClelland,J.L., & PDP Research Group (1988b). *Parallel Distributed Processing; Explorations in the Microstructure of Cognition; Volume 1: Foundations*. A Bradford Book, The MIT Press.

[Saarinen 1993] Saarinen, J. (1993). *Shifts of Visual Attention at Fixation and Away from Fixation*. Vision Res., 33(8):1113–1117.

[Saarinen et al. 1991] Saarinen, J. & B. Julesz (1991). *The speed of attentional shifts in the visual field*. Proc. Natl. Acad. Sci. USA, 88:1812–1814.

[Sanger 1989] Sanger, T.D. (1989). *Optimal unsupervised learning in a single-layer linear feedforward neural network.* Neural Networks., 2, pp. 459-473.

[Schneider et al. 1977] Schneider, W. and Shiffrin, R.M. (1977). *Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention.* Psychological Review, 84(1977), pp. 1-66.

[Sekuler 1974] Sekuler, R. (1974). *Spatial Vision.* Ann. Rev. Psychol., 25, pp. 195-232.

[Shapley et al. 1986] Shapley, R. and Pery, V. H. (1986). *Cat and monkey retinal ganglion cells and their visual functional roles.* Trends Neurosci., 9, pp. 229-235.

[Shulman et al. 1979] Shulman, G.L., R. Remington, & J. McLean (1979). *Moving attention through visual space*. J. Exp. Psychol Human Percept, 5:522–526.

[Smith et al. 1975] Smith, V. C., Pokorny, J. (1975). *Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm.* Vision Res., 15, pp. 161-171.

[Stark et al. 1981] Stark, L. & S. Ellis (1981). *Scanpaths Revisited: Cognitive Models Direct Active Looking*. in: FISHER, MONTY, & SENDERS, (eds.): *Eye Movements, Cognition and Visual Perception*, pp. 193–226. Erlbaum Press.

[Sutton 1984] Sutton, R.S. (1984). *Temporal Credit Assignment in Reinforcement Learning*. Technical Report, University of Massachusetts at Amherst.

[Sutton 1991] Sutton, R.S. (1991). *Integrated Modeling and Control Based on Reinforcement Learning and Dynamic Programming*. in: LIPPMANN, RICHARD P., J. E. MOODY, & D. S. TOURETZKY, (eds.): *Advances in Neural Information Processing 3*, pp. 471–478, San Mateo, California. Morgan Kaufmann Publishers.

[Switkes et al. 1988] Switkes, E., Bradley, E., De Valois, K. K. (1988). *Contrast dependene and mechanisms of masking interactions amoung cromatic and luminance gratings.* J. Opt. Soc. Am., 5, pp. 1149-1162.

[Thompson et al. 1981] Thompson, I. D. and Tolhurst, D. J. (1981). *Columnar organization for optimal spatial frequency in cat striate cortex.* J. Physiol. (Lond.), 319, pp. 79.

[Thorell et al. 1984] Thorell, L. G., De Valois, R. L., Albrecht, D. G., (1984). *Spatial mapping of monkey V1 cells with pure color and luminance stimuli.* Vision Res., 24, pp. 751-769.

[Tolhurst et al. 1981] Tolhurst, D. J. and Thompson, I. D. (1981). *On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat.* Proc. Roy. Soc. London, B 213, pp. 183-199.

[Tolhurst et al. 1982] Tolhurst, D. J. and Thompson, I. D. (1982). *Oganization of neurons prefering similar spatial frequencies in cat striate cortex.* Exp. Brain Res., 48, pp. 217-227.

[Tootell et al. 1981] Tootell, R. B. H., Silverman, M. S. and De Valois, R. L. (1981). *Spatial frequency columns in primary visual cortex.* Science, 214, pp. 813-815.

[Treisman et al. 1980] Treisman, A.M. & G. Gelade (1980). *A Feature-Integration Theory of Attention.* Cognitive Psychology, 12:97–136.

[Treisman 1990] Treisman, A.M. (1990). *Merkmale und Gegenstände in der visuellen Verarbeitung.* Gehirn und Kognition.

[Treisman et al. 1990] Treisman, A.M. & S. Sato (1990). *Conjunction Search Revisited.* Journal of Experimental Psychology: Human Perception and Performance, 16(3):459–478.

[Tsal 1989] Tsal, Y. (1989). *Do Illusory Conjuctions Support the Feature Integration Theory? A Critical Review of Theory and Findings.* J. of Experimental Psychology: Human Perception and Performance, 15(2):394–400.

[Tso et al. 1988] Ts'o, D. Y. and Gilbert, C. D. (1988). *The organisation of chromatic and spatial interactions in the primate striate cortex.* J. Neurosci., 8, pp. 1712-1728.

[Turner 1986] Turner, M. R. (1986). *Texture discrimination by Gabor Functions.* Biological Cybernetics,

[Tweed et al. 1990] Tweed, D.B. & T. Vilis (1990). *The Superior Colliculus and Spatiotemporal Translation in the Saccadic System.* Neural Networks, 3:75–86.

[Unser 1986] Unser, M. (1986). *Local linear transforms for texture analysis.* Proceedings of the Seventh International Conference on Pattern Recognition, Montreal, pp. 1206-1208.

[Von Kries 1905] Von Kries, J. (1905). *Die Gesichtsempfindungen.* In Handbuch der Physiologie des Menschen. Nagel, W. (ed.), Brunswick: Vieweg.

[Vautin et al. 1985] Vautin, R. G. and Dow, B. M. (1985). *Color cell groups in foveal striate cortex of the behaving macaque.* J. Neurosci., 54, pp. 373-292.

[Watkins 1992] WATKINS, C.J.C.H. (1992). *Q-Learning.* Machine Learning, 8:279–292.

[Werblin et al. 1969] Werblin, F. S. and Dowling, J. E. (1969). *Organization of the retina of the mudpuppy.* Necturus maculosus. II. Intracellular recording. J. Neurophysiol., 32, pp. 339-355.

[Wolfe et al. 1990] Wolfe, J.M. & K. R. Cave (1990). *Deploying Visual Attention: The Guided Search Model.* in: BLAKE, A. & T. TROSCIANKO, (eds.): *AI and the Eye*, pp. 79–103. John Wiley and Sons.

[Yarbus 1967] YARBUS, A.L. (1967). *Eye Movements and Vision.* Plenum Press, New York.

[Zeki 1978a] Zeki, S. M. (1978). *The cortical projections of foveal striate cortex in the rhesus monkey.* J. Physiol. (Lond.), 277, pp. 227-244.

[Zeki 1980] Zeki, S. M. (1980). *The representation of colours in the cerebral cortex.* Nature, 284, pp. 412-418.

[Zeki 1983] Zeki, S. (1983). *Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelength and colours.* Neuroscience, 9, pp. 741-756.

[Zeki 1994] Zeki, S. (1994). *A vision of the brain.* Oxford: Blackwell Scientific Publications.