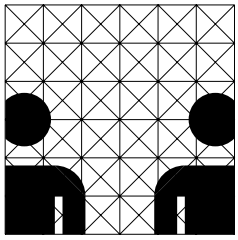
Entwicklung einer Aufmerksamkeitssteuerung für ein aktives Sehsystem

Dissertation
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt
dem Fachbereich Informatik
der Universität Hamburg



von
Maik Bollmann

Hamburg, März 1999

Gutachtende:	Prof. Dr.-Ing. Bärbel Mertsching
	Prof. Dr.-Ing. H. Siegfried Stiehl
Tag der Einreichung:	17.03.1999
Tag der Disputation:	25.11.1999

Danksagung

Ich möchte an dieser Stelle allen Personen danken, die auf die ein oder andere Weise zum Entstehen dieser Arbeit beigetragen haben. Zu nennen ist hier die Unterstützung durch verschiedene Fachbereichseinrichtungen, insbesondere die Anstrengungen der Fachbereichsbibliothek, mir die vielen Literaturwünsche zu erfüllen, sowie die Hilfe des Rechenzentrums beim Aufbau und der Pflege des Rechnernetzes unserer Arbeitsgruppe.

Mein Dank gebührt dem Gutachter Herrn Prof. Dr.-Ing. H. Siegfried Stiehl für die zahlreichen fruchtbaren Diskussionen und insbesondere Frau Prof. Dr.-Ing. Bärbel Mertsching für die Betreuung dieser Arbeit und anderer Aufgaben, die meiner wissenschaftlichen und/oder fachlichen Weiterbildung nützlich waren.

Besonderer Dank aber gilt Dipl.-Ing. Rainer Hoischen, Dipl.-Inform. Steffen Schmalz und Dipl.-Ing. Alexander Schwarz für die sehr gute Zusammenarbeit und gegenseitige Unterstützung bei den täglich anfallenden Aufgaben sowie die private Freundschaft. Desweiteren möchte ich den Studierenden danken, deren Arbeiten ich betreut habe und die so ebenfalls einen Anteil an der Entstehung dieser Arbeit haben. Insbesondere sind hier die Studienarbeiten von Thorsten Hempel und Christoph Justkowski zu nennen.

Nicht zuletzt möchte ich der Deutschen Forschungsgemeinschaft (DFG) danken, die diese Arbeit im Rahmen eines Projektes zum *E*ntwurf von *S*ystembausteinen der *A*ktiven *B*ildanalyse (ESAB-1) gefördert hat.

Kurzfassung

Das interdisziplinäre Forschungsthema Aktives Sehen reicht von theoretischen Fundierungen über die biologienahe Modellierung aktiver Sehsysteme von Lebewesen bis hin zu echtzeitfähigen systemtechnischen Realisierungen aus der ingenieurwissenschaftlichen Perspektive. All dies wird komplementiert durch die Entwicklung innovativer Produkte, wie etwa sensorgeführter, autonom agierender Service-Roboter.

Der vorliegenden Arbeit liegt eine anwendungsorientierte Sichtweise auf aktive Sehsysteme zugrunde. Ausgehend von bzw. aufbauend auf existierende Komponenten des gemeinsam mit der UGH Paderborn entwickelten Systems NAVIS wird eine Blicksteuerung entwickelt und in das bestehende System integriert. Diese Blicksteuerung kann in drei Teilsysteme untergliedert werden: eine datengetriebene ’Bottom-up‘-Komponente, eine modellgetriebene ’Top-down‘-Komponente und eine (bisher allerdings nur rudimentäre) Verhaltenssteuerung, die abhängig von der vorliegenden Situation die geeignetste Verhaltensklasse auswählt. Die ’Bottom- up‘-Komponente berechnet eine sogenannte Attraktivitätsrepräsentation, die verschiedene Merkmale und Auffälligkeiten enthält. Die Attraktivitätsrepräsentation bestimmt den nächsten Aufmerksamkeitspunkt, falls kein Wissen aus der ’Top-down‘-Komponente vorliegt. Liegt dagegen eine zu validierende Objekthypothese vor oder konnte Bewegung in der Szene detektiert werden, nimmt die ’Top-down‘-Komponente maßgeblichen Einfluß auf die Auswahl des nächsten Aufmerksamkeitspunktes. Unabhängig von der jeweils vorliegenden Verhaltensklasse, Objekterkennung oder Objektverfolgung, ermöglicht die Blicksteuerung geschlossene Perzeptions-Aktions-Zyklen und ein autonomes Verhalten.

Die entwickelte Blicksteuerung ist für general-purpose Sehen konzipiert. Das heißt insbesondere, daß sie im Gegensatz zu vielen aus der Literatur bekannten Blicksteuerungen aktiver Sehsysteme mehr als eine Verhaltensklasse unterstützt. Sie zeichnet sich weiterhin dadurch aus, daß sie leicht um zusätzliche Merkmale erweitert werden kann.

Inhaltsverzeichnis

Kapitel 1

Einleitung

Seit Ende der achtziger Jahre findet eine Besinnung darauf statt, daß biologisches Sehen nicht nur perzeptuelle, sondern auch aktorische Fähigkeiten umfaßt. Die Nachbildung solch aktiver Sehsysteme erlaubt es erstmals Maschinen, ihre Umgebung selbständig zu erkunden und als dreidimensionale Welt wahrzunehmen. Der Implementation von Blickbewegungen kommt hierbei eine wichtige Aufgabe zu ([Bajcsy 1980], [Clark 1988]). Zielgerichtetes Sehen erleichtert die räumliche Szenenrepräsentation, indem es einen Fixationspunkt liefert, der als Bezugspunkt in einem Koordinatenraum dienen kann. Die Anzahl der Freiheitsgrade reduziert sich auf diese Weise, wodurch die für eine Szenenanalyse notwendige Mathematik vereinfacht wird ([Aloimonos 1987], [Ballard 1988, 1991]). Desweiteren wird durch die Fovealisierung auf interessante Bildbereiche die zu verarbeitende Datenmenge erheblich reduziert. Dieser Effekt kann durch den Einsatz einer bezüglich des Auflösungsvermögens inhomogenen künstlichen Retina noch verstärkt werden.

Um interessante Regionen im Sehfeld eines Roboters zielorientiert und sequentiell selektieren zu können, ist die Nachbildung der visuellen Aufmerksamkeit erforderlich, die genau diese Fähigkeit besitzt. Es existieren zwei Arten visueller Aufmerksamkeit, die als offen bzw. verdeckt bezeichnet werden. Die verdeckte Aufmerksamkeit findet ohne motorische Aktionen statt. In dieser Phase wird abgeschätzt, ob sich ein mit erhöhtem Energieaufwand zu betreibender Blickwechsel "lohnt". Es finden im Gehirn Attentionsverschiebungen statt, die das Ziel der nächsten Augenbewegung aussuchen. Hierbei können nur Regionen untersucht werden, die sich im derzeit auf der Retina fixierten Bild befinden. Um weitere möglicherweise interessante Regionen zu finden, sind schnelle Augenbewegungen, die sogenannten Sakkaden, notwendig. Diese Attentionsverschiebungen werden als offene Aufmerksamkeit bezeichnet.

Biologisches Sehen kann sich nicht in Isolation entwickelt haben. Es findet statt, weil wir uns in einer dreidimensionale Welt bewegen. Sehen und insbesondere gerichtetes Sehen sollte daher immer eine Absicht verfolgen [Aloimonos 1987]. Ein Schwerpunkt dieser Arbeit ist deshalb neben dem Entwurf einer Aufmerksamkeitssteuerung auch deren Integration in ein bestehendes maschinelles Sehsystem, das zur Erkennung und Verfolgung von Objekten eingesetzt wird.

Kapitel 2

Übersicht

Dieses Kapitel gibt einen Überblick über die dieser Arbeit zugrundeliegende Gliederung, beschreibt die Aufgabenstellung und skizziert den verfolgten Lösungsansatz.

2.1 Gliederung der Arbeit

Der Mensch besitzt die Fähigkeit, seine Aufmerksamkeit visuell auf bestimmte Teile eines Raumes oder Objekts zu konzentrieren. Diese Aufmerksamkeitssteuerung ermöglicht dem Menschen eine effektive Suche nach Objekten bzw. eine effektive Rekonstruktion der beobachteten Umwelt. Durch solche Eigenschaften erscheint es interessant, den Menschen als biologisches Vorbild für den Entwurf eines maschinellen Aufmerksamkeitsmoduls zu betrachten. Einige der psychologischen, biologischen und maschinellen Modelle zur visuellen Aufmerksamkeit sind daher im Kapitel 3 unter der Überschrift "Stand der Forschung" zusammengefaßt.

Dem Kapitel über die psychophysischen und neurobiologischen Grundlagen folgt der eigentliche Kern dieser Arbeit, der Aufbau der Aufmerksamkeitssteuerung für das *N*euronale *A*ctive-*Vi*sion-*S*ystem NAVIS. NAVIS wird in Kooperation mit dem GET-Bildverarbeitungslabor der UGH Paderborn entwickelt. Besondere Aufmerksamkeit wird auf dessen biologische Adäquatheit, Modularität sowie die Integration der Module zu einem Gesamtsystem gelegt. Die Aufmerksamkeitssteuerung läßt sich in eine ‘Bottom up’-Komponente und eine von unterschiedlichen Modellen getriebene ‘Top down’-Komponente untergliedern. Im Kapitel 4 ist zunächst die bilddatengetriebene ‘Bottom up’-Komponente beschrieben. Die ‘Top down’-Komponente ist eng verzahnt mit anderen NAVIS-Modulen. Das Kapitel 5 beschreibt daher den derzeitigen Entwicklungsstand in NAVIS und die Integration der entwickelten Aufmerksamkeitssteuerung in dieses System. Die zur Realisierung der Aufmerksamkeitssteuerung umgesetzten Arbeitspakete sind im Abschnitt 2.2 detailliert aufgeführt.

Im Kapitel 6 wird zum einen die psychologische und biologische Plausibilität der entwickelten Aufmerksamkeitssteuerung diskutiert und zum anderen, inwieweit die im folgenden Abschnitt

2.2 aufgestellten Anforderungen an eine maschinelle Blicksteuerung erfüllt werden konnten. Die Limitierungen, denen das vorgestellte System unterliegt, werden genannt und aus diesen Einschränkungen werden Problemstellungen für zukünftige Aufgaben abgeleitet. Den Abschluß dieser Arbeit bildet die Zusammenfassung der wichtigsten Ergebnisse im Kapitel 7.

2.2 Aufgabenstellung und Lösungsansatz

Eines der Ziele des DFG-Projektes ESAB-1[1], dem diese Arbeit zuzuordnen ist, und der Schwerpunkt speziell dieser Arbeit ist die Implementation von attentiv gesteuerten Blickwechseln. Im Wachzustand vollziehen Menschen fortlaufend Blickwechsel. Sie erfolgen häufig kontextbasiert, wie in [Yarbus 1969] und [Noton 1970, 1971a+b] gezeigt wurde. Diese Eigenschaft sollten auch Blicksteuerungen für Robotersysteme beinhalten [Granlund 1994]. Abbott fordert in [Abbott 1988] weiterhin, daß ein Modell für Augenbewegungen folgende Eigenschaften aufweisen sollte:

- Berücksichtigung der Distanz und der Blickrichtung zwischen aktuellem Fixationspunkt und den Kandidaten für einen Blickwechsel
- Auswertung von Bildcharakteristiken wie z. B. Symmetrien
- Berücksichtigung von Objektgrenzen (Fixation auf Objekte)
- Berücksichtigung von zeitlichen Änderungen (Bewegung)
- Begrenzung der Komplexität (Anzahl der Fixationspunkte sollte möglichst minimal sein)

Die Aufgabe besteht nun in der Umsetzung der geforderten Eigenschaften bei der Entwicklung einer robusten Aufmerksamkeitssteuerung für das in der AG IMA aufgebaute aktive Sehsystem NAVIS. Hierzu ist zunächst nur eine Kamera eines zur Verfügung stehenden binokularen Kopfes zu regeln; die zweite Kamera wird in NAVIS zur Tiefenschätzung bei der Analyse komplexer 3D-Szenen eingesetzt.

Als erster wichtiger Bestandteil der Aufmerksamkeitssteuerung müssen Module zur Generierung topologischer Merkmals- und Auffälligkeitskarten erstellt werden. Die Merkmalskarten sollen möglichst disjunkte Merkmale enthalten, die auch für andere Module in NAVIS nutzbar sind. Insbesondere kann hier der Effekt der Kategorisierung ausgenutzt werden. Zu der Merkmalsdimension Farbe sollen so z. B. die Merkmalsausprägungen Rot, Blau, etc. auf verschiedene Karten verteilt werden. Hierdurch können andere Module in NAVIS gezielt auf bestimm-

[1] DFG-Projekt "Entwicklung von Algorithmen zur aktiven Bildanalyse und deren Beschleunigung durch Methoden des Hardware/Software-Codesign" (Me 1289/3-1); Projektpartner ist das FhG-Institut für Mikroelektronische Schaltungen und Systeme, Dresden (Zi 260/6-1); Laufzeit: 01.07.1996-31.03.1999 (Förderung: 2 Jahre)

te gewünschte Objekteigenschaften zugreifen. Die Auffälligkeitskarten sollen auf den Merkmalskarten aufsetzen. In ihnen werden Attraktivitätspunkte ermittelt, die eventuell später von dem Kamerasystem zu fixieren sind. Die Berechnung möglicher Kandidaten für einen Blickwechsel ist vergleichbar mit der verdeckten Aufmerksamkeitsbildung. Wichtig ist, daß die Auffälligkeitskarten nicht etwa lediglich durch Hemmungs- oder Verstärkungsmechanismen die lokalen Unterschiede in den Merkmalskarten erhöhen, sondern tatsächlich neue Eigenschaften, wie z. B. Symmetrien, beschreiben.

Da für die einzelnen Merkmalsklassen jeweils eigene Auffälligkeitskarten generiert werden, müssen geeignete Gewichtungsverfahren zur Bewertung von unter Umständen konkurrierenden Attraktivitätspunkten untersucht werden. Eine solche Untersuchung kann z. B. in einem besonderen Experimentierfeld, der aus psychologischen Untersuchungen bekannten visuellen Suche, erfolgen. Das Ergebnis dieser Bewertung soll der Aufbau einer globalen Attraktivitätskarte sein. Das Binding-Problem, also die Kombination der in den verschiedenen Merkmalskanälen generierten Attraktivitätspunkte, soll hier zunächst nur im 2D-Raum über die Bestimmung von Nachbarschaftsbeziehungen gelöst werden. Im Fortsetzungsprojekt ESAB-2[2] wird dann eine räumliche Integration der Attraktivitätspunkte durch die Auswertung von Tiefeninformation angestrebt.

Die Aufmerksamkeitssteuerung benötigt neben einer ‘bottom up’-generierten Attraktivitätskarte eine weitere gleich aufgebaute Karte zur Aufnahme bereits fixierter Punkte. Diese Inhibitionskarte sorgt dafür, daß bereits betrachtete Objekte oder Objektstrukturen innerhalb "kurzer Zeit", d. h. weniger Iterationsschritte, kein zweites Mal anvisiert werden. Da eventuell eine Objektregion für die Erkennung ein weiteres Mal angeschaut werden muß, ist ein Abklingen der Hemmung pro Iteration um einen prozentualen Wert erforderlich.

Bei komplexen Objekten kann nicht mehr davon ausgegangen werden, daß diese holistisch erkannt werden können, sondern es ist notwendig, Teilstrukturen auszuwerten. NAVIS erlernt die zu einem Objekt gehörenden Teilformen jeweils in einem Schritt. Zu jeder Teilform werden dabei auch die Attraktivitätspunkte gespeichert, die die Grundlage für das Bilden einer Hypothese "Objekt A befindet sich im Bild" liefern (der Erkennungsvorgang ist in Kapitel 5 genauer beschrieben). Konnte eine Objekthypothese aufgestellt werden, ist diese durch den sukzessiven Vergleich der gelernten mit den hypothesierten Teilformen zu validieren. Während des Erkennungsvorgangs ist sicherzustellen, daß das Kamerasystem nicht zwischen den Teilformen verschiedener Objekte "springt". Hierzu ist eine Steuerung notwendig, die wissengetriebene Information mit bilddatengetriebener Information verknüpft und die diversen Karten verwaltet. Diese Kontrollstruktur soll das zentrale Aufmerksamkeitsmodul sein.

Nachdem ein Aufmerksamkeitspunkt bestimmt worden ist, muß das Aufmerksamkeitsmodul die Blicksteuerung anstoßen, um die Kamera auf diesen Punkt zu richten. Die für die Ankopp-

[2] DFG-Projekt "Entwicklung von Algorithmen zur dynamischen Perzeption und deren Umsetzung in Systembausteine der aktiven Bildanalyse" (Me 1289/3-2); Förderzeitraum: 01.02.1999-31.01.2001

lung des Aufmerksamkeitsmoduls benötigte geeignete Schnittstelle ist bereits in NAVIS realisiert. Der rückgekoppelte Datenpfad "Bildaufnahme", "Aufmerksamkeitspunktbestimmung", "Kamerabewegung" und "erneute Bildaufnahme" ist zu realisieren. Dabei sind die Koordinaten der in die verschiedenen Karten eingetragenen Punkte entsprechend der Kamerabewegung zu verschieben. Die Blicksteuerung soll also geschlossene Umläufe ermöglichen. Weiterhin soll sie neben der Objekterkennung auch noch ein zweites Modell, die Objektverfolgung, zulassen. Hierzu muß sie die Objekterkennung unterbrechen und den initialen Punkt für einen Tracking-Algorithmus vorgeben.

Kapitel 3

Stand der Forschung

In diesem Kapitel werden Forschungsergebnisse zur visuellen Aufmerksamkeit, die auf den Gebieten der Psychophysik, der Neurophysiologie und des Computersehens erzielt worden sind, beschrieben. Sie mögen zum Verständnis der Thematik beitragen. Zur Vertiefung wird vielfach auf die Primärliteratur verwiesen. Zunächst einmal werden aber einige grundlegende Begriffe eingeführt, die die Abgrenzung der Aufmerksamkeit gegenüber anderen Phänomenen erleichtern. Anschließend werden einige psychophysische Modelle zur visuellen Aufmerksamkeit vorgestellt, die wichtige Hinweise geben, wie eine Aufmerksamkeitssteuerung für ein technisches System aussehen könnte. Der folgende Abschnitt beschäftigt sich mit den neurophysiologischen Regionen des menschlichen Gehirns, die an der Aufmerksamkeitsbildung oder den Blickwechseln beteiligt sind. Im letzten Abschnitt erfolgt eine Beschreibung und Bewertung der in der Literatur beschriebenen Computermodelle der visuellen Aufmerksamkeit.

3.1 Visuelle Aufmerksamkeit und Augenbewegungen

Posner [Posner 1980a] konnte bereits früh zeigen, daß sich der Fokus der Aufmerksamkeit im visuellen Feld bewegen kann, ohne daß hierfür Augenbewegungen nötig sind. Es existiert daher heutzutage eine weitgehende Übereinstimmung darüber, daß es sich bei den Augenbewegungen und der visuellen Aufmerksamkeit um zwei unterschiedliche Mechanismen handelt. Der eine Mechanismus bildet mittels Augenbewegungen sukzessiv periphere Bereiche in die zentrale Fovea ab und ermöglicht so, das inhomogene Auflösungsvermögen der Retina auszugleichen. Weitgehend unabhängig erlaubt der zweite Mechanismus, die interne Allokierung von Verarbeitungskapazitäten von einer Position zur nächsten zu verschieben, ohne die Blickrichtung zu wechseln [Jonides 1983]. Die Mehrzahl der Aufmerksamkeitsforscher und -forscherinnen priorisiert allerdings die sogenannte *Interdependence-Hypothese*, die besagt, daß Aufmerksamkeitsverschiebungen und Augenbewegungen weder identisch noch völlig unabhängig voneinander sind; vielmehr kann der eine Mechanismus den anderen erleichtern oder

behindern ([Remington 1980], [Sheperd 1986], [Umilta 1988]). Aufmerksamkeitsverschiebungen sind etwa viermal so schnell wie Augenbewegungen. Sie können deshalb sehr gut eingesetzt werden, um eine Szene grob zu untersuchen und die Augenbewegungen zu den Regionen vorzubereiten, für die sich eine genauere Untersuchung lohnt. Statt der Begriffe Aufmerksamkeit und Augenbewegungen findet man auch häufig die Synonyme *verdeckte Aufmerksamkeit* und *offene Aufmerksamkeit*.

3.1.1 Visuelle Aufmerksamkeit

Um das Phänomen visuelle Aufmerksamkeit zu beschreiben, bedienen sich zahlreiche Autoren und Autorinnen der Metapher eines *Scheinwerfers* (engl. Spotlight). Ein solches Spotlight bewegt sich analog von einer Position zur nächsten, überstreicht dabei also die Regionen, die zwischen Ausgangsposition und Ziel liegen, ist in sich homogen und besitzt eine kompakte Form bestimmter Größe. Eine zweite, häufig anzutreffende Bezeichnung für das Spotlight ist der '*Focus Of Attention*' (FOA). Die wesentlichste Annahme der *Spotlight-Hypothese* ist aber, daß sich Aufmerksamkeit nicht gleichzeitig für verschiedene, räumlich voneinander getrennte Regionen des visuellen Feldes allokieren läßt ([Posner 1980b], [Podgorny 1983]).

Untersuchungen von Shulman et al. [Shulman 1979] und Tsal [Tsal 1983] haben gezeigt, daß die Bewegungsart der Aufmerksamkeit analog ist. Das heißt, der FOA folgt einer Trajektorie, die alle Punkte zwischen Start- und Zielpunkt kontinuierlich und mit konstanter Geschwindigkeit durchläuft. Die Geschwindigkeit scheint sich dabei der Entfernung anzupassen, so daß die für den Aufmerksamkeitswechsel benötigte Zeit immer in etwa gleich groß bleibt. Ähnliche Ergebnisse zeigen sich auch für Augen- und Handbewegungen [Remington 1984]. Die Größe des FOA variiert und kann vermutlich ein Kontinuum einnehmen zwischen einer Verteilung über das gesamte visuelle Feld und einer Konzentration auf weniger als ein Grad des visuellen Feldes (siehe [Schneider 1984], [Eriksen 1985, 1986], [Mesulam 1985]). Die Ränder des FOA sind dabei wohl eher unscharf [Mangun 1988]. Allerdings ist die Spotlight-Hypothese nicht unumstritten. Eine Alternative ist z. B. die FINST-Hypothese von Pylyshyn [Pylyshyn 1994a+b], die besagt, daß eine kleine Zahl an Stimuli präattentiv und willkürlich über das visuelle Feld verteilt indiziert werden kann. Die Indizes sind hierbei so mit den Objekten verknüpft, daß sie insbesondere auch mit bewegten Objekten wandern. Gestützt wird die FINST-Hypothese von 'Multiple-Object Tracking'-Studien, die zeigen, daß Personen in der Lage sind, eine zuvor angezeigte Gruppe sich gleichzeitig bewegender Stimuli unter physikalisch identischen, sich aber nicht bewegenden Elementen zu verfolgen. Allerdings darf die Gesamtzahl der sich bewegenden Stimuli nicht zu groß sein.

Grundsätzliche Übereinstimmung findet sich in der Literatur darüber, daß bereits untersuchte Regionen des visuellen Feldes gehemmt werden und so Blickwechsel zu weiteren Regionen ermöglichen. Die Hemmung nimmt mit der Zeit wieder ab, so daß dieselben Regionen nach einiger Zeit wieder aufgesucht werden können ([Posner 1984a], [Maylor 1985, 1987], [Tassinari

1987]). Die Hemmung läßt sich zusätzlich durch willentliche Orientierung der Aufmerksamkeit überwinden. Hieraus ergibt sich eine weitere Differenzierung der visuellen Aufmerksamkeit: automatisch gegenüber willentlich (siehe [Schneider 1977], [Shiffrin 1977], [Jonides 1981], [Posner 1982, 1984b], [Norman 1985], [Bundesen 1990]). Die *automatische Aufmerksamkeit* wird ausschließlich durch die Stimuli im visuellen Feld gesteuert, deshalb wird diese häufig auch als *'Bottom up'-Aufmerksamkeit* bezeichnet und die Reizverarbeitung der Stimuli gar als präattentiv. *Willentliche Aufmerksamkeit* benötigt eine Erwartung, die durch Wissen, also durch *'Top down'-Information*, vorgegeben werden muß. Ein schönes Beispiel hierfür bilden psychophysische Experimente, die durch die Art des *Precuings* automatische und willentliche Aufmerksamkeit zu trennen versuchen. Unter Precuing versteht man die Vorgabe der Position, in der ein zu bestimmendes Target erscheinen wird. Periphere Cues (z. B. Lichtflecken) erregen die Aufmerksamkeit automatisch, indem sie direkt an der Stelle erscheinen, an der auch das Target erscheinen wird. Zentrale Cues (in der Regel Pfeile) geben dagegen nur die Richtung an, in die die Aufmerksamkeit anschließend willentlich zu lenken ist. Synonyme für 'Bottom up'- und 'Top down'-Aufmerksamkeit sind *exogene Aufmerksamkeit* bzw. *endogene Aufmerksamkeit* sowie *datengetriebene Aufmerksamkeit* bzw. *wissengetriebene Aufmerksamkeit.*

Die visuelle Aufmerksamkeit wird also nicht nur durch die physikalischen Eigenschaften der Stimuli, sondern auch durch nicht-visuelle Faktoren beeinflußt. Zwei bemerkenswerte nicht visuelle Faktoren bei der visuellen Suche nach einem Target sind der Kategorisierungseffekt sowie die Vertrautheit mit dem Target. Jonides und Gleitman [Jonides 1972] ließen Testpersonen nach einem O suchen unter Distraktoren, die entweder aus Buchstaben oder aus Zahlen bestanden. Einer Gruppe der Testpersonen wurde gesagt, daß es sich bei dem Target um den Buchstaben O handele, während die zweite Gruppe nach der Zahl Null suchen sollte. In den Experimenten zeigten sich anschließend deutliche Kategorisierungseffekte: So konnte z. B. eine höhere Reaktionszeit gemessen werden, wenn unter Buchstaben-Distraktoren nach dem Buchstaben O gesucht werden sollte, als wenn nach der physikalisch identischen Zahl Null gesucht werden sollte. Ähnliche Kategorisierungseffekte konnten auch von Brand [Brand 1971] und Egeth et al. [Egeth 1972] gezeigt werden.

Auch die Vertrautheit ist keine charakteristische Eigenschaft des visuellen Musters, sondern hängt von der individuellen Erfahrung einer Person mit eben diesem Muster ab. Wang et al. [Wang 1994] führten vier Experimente zum Phänomen der Vertrautheit durch und erhielten folgende Ergebnisse:

1. Wenn sowohl das Target als auch die Distraktoren der Testperson nicht vertraut sind, so ist die Suche nach dem Target seriell und vergleichsweise langsam.
2. Sind sowohl das Target als auch die Distraktoren der Person vertraut, ist die Suche zwar seriell, aber deutlich schneller als unter 1.
3. Ist das Target der Person vertraut, die Distraktoren aber nicht, so ist die Suche nach dem Target seriell und liegt bezüglich der durchschnittlich benötigten Suchzeit zwischen denen aus 1. und 2.

4. Ist das Target der Person im Gegensatz zu den Distraktoren nicht vertraut, so ist die Suchzeit unabhängig von der Anzahl der Distraktoren (*parallele Suche*).

Aus den Ergebnissen läßt sich ablesen, daß solche Stimuli bevorzugt verarbeitet werden, die der Testperson unbekannt sind: In Experiment 3 wird die Aufmerksamkeit zunächst auf die nicht vertrauten Distraktoren gelenkt, das Finden des Targets dauert entsprechend lange. In Experiment 4, in dem nur das Target nicht vertraut ist, ist die Suche dagegen parallel. Das heißt, das Target wird augenblicklich detektiert; man spricht hier auch vom sogenannten *'Pop out'-Phänomen*. Statt des Kriteriums Vertrautheit, kann man natürlich genauso gut von dem inversen Merkmal Neuheit sprechen. Tatsächlich gibt es neurophysiologische Untersuchungen, die gezeigt haben, daß die neuronale Aktivität einiger Zellen des inferior-temporalen Cortex (IT) proportional zur Neuheit oder auch Unerwartetheit des präsentierten Stimulus ist (siehe [Miller 1991] und Abschnitt 3.3).

Wang et al. zogen aus ihren Untersuchungen eine sehr bemerkenswerte Schlußfolgerung: Wenn nun die Vertrautheit ein so frühes Merkmal ist, daß sie zur parallelen Suche beiträgt, so muß auch ein sehr früher Zugriff auf das Gedächtnis stattfinden. Eine solche Vorstellung widerspricht der allgemeinen Annahme, daß der Zugriff auf das Gedächtnis, den Objektspeicher, der letzte Schritt im visuellen Erkennungsvorgang ist.

3.1.2 Augenbewegungen

Augenbewegungen ermöglichen es uns, den jeweils interessantesten Ausschnitt in unserem visuellen Feld mit der bestmöglichen Auflösung zu sehen. Dennoch kann man nicht sagen, daß der Sinn der Augenbewegungen in der Kompensation der inhomogenen Auflösung der Retina liegt. Eine solche Vorstellung betrachtet die inhomogene Auflösung der Retina als eine Art Defekt und erklärt nicht, warum niedere Lebewesen ohne eine inhomogene Retina ebenfalls Augenbewegungen durchführen. Augen- und Kopfbewegungen sind zunächstmal deshalb notwendig, weil wir uns in einer dreidimensionalen und dynamischen Welt bewegen. Augenbewegungen besitzen gegenüber Kopfbewegungen die gleichen Vorteile wie Aufmerksamkeitswechsel gegenüber Augenbewegungen: Sie sind schneller und mit geringerem Energieaufwand verbunden.

Die Augenbewegungen lassen sich in verschiedene Klassen unterteilen. Als *Sakkaden* bezeichnet man willentliche Augenbewegungen, die von einer Stelle im visuellen Feld zu einer anderen springen, um z. B. ein Objekt zu suchen oder zu analysieren. Sakkaden zeigen ballistisches Verhalten, d. h. der Zielpunkt wird vor der Ausführung der Sakkade programmiert und kann während der Bewegung nicht mehr verändert werden. Sakkaden treffen häufig nicht exakt das anvisierte Ziel, so daß eine nachfolgende kleinere Sakkade folgt, die den Fehler korrigiert. Während der Dauer einer Sakkade ist der Mensch blind. Sakkaden sind daher sehr schnell; sie erreichen bis zu 700 ˚/s. Sehr viel länger als die Sakkade selbst (10 - 70 ms, je nach Sprungwei-

te) dauert die Vorbereitung einer Sakkade (200 - 300 ms). Diese Zeit ist besonders für psychophysische Experimente zur visuellen Aufmerksamkeit wesentlich, in denen in der Regel Augenbewegungen ausgeschlossen werden. Der Mensch führt bis zu drei Sakkaden in der Sekunde durch. Die Augenbewegungen selbst können bis zu 10% der gesamten Beobachtungszeit ausmachen. Ausführliche Untersuchungen zu Sakkaden finden sich u. a. in [Carpenter 1988] und [Sekuler 1990]. Ähnlich den normalen Sakkaden verhalten sich die *Mikrosakkaden*. Sie weisen allerdings eine sehr viel geringere Sprungweite auf (1-25' gegenüber 1-20˚ bei etwa gleicher Geschwindigkeit). Eine spezielle Form der Sakkade ist die von Fischer und Weber entdeckte *Expreß-Sakkade* (siehe [Fischer 1993], [Weber 1994]). Diese Sakkade besitzt eine Präparationszeit von unter 100 ms und wird damit sehr viel schneller ausgelöst als eine normale Sakkade. Sie tritt allerdings nur unter der Bedingungen auf, daß der aktuell fixierte Stimulus kurz vor dem Erscheinen eines neuen peripheren Stimulus entfernt wird. Der periphere Stimulus muß zudem eine gewisse Größe aufweisen, um die Expreß-Sakkade auslösen zu können. Fischer und Weber vermuten, daß Expreß-Sakkaden deshalb eine kürzere Präparationszeit besitzen als übliche Sakkaden, weil die Inhibition des aktuellen FOA entfällt.

Eine zweite wichtige Klasse von Augenbewegungen ist die *langsame Folgebewegung* (engl. 'Smooth Pursuit'). Im Gegensatz zu den Sakkaden stellt sie eine kontinuierliche Bewegung dar. Sie setzt ein, wenn ein bewegtes Objekt mit den Augen verfolgt werden soll; sie ist reflektorisch, kann aber willentlich beeinflußt werden. Im allgemeinen muß zunächst eine Sakkade zu dem beweglichen Objekt ausgeführt werden, die anschließend von der langsamen Folgebewegung abgelöst wird. Dabei scheint die Geschwindigkeit des Objektes bereits beim Einsetzen der langsamen Folgebewegung bekannt zu sein, denn die Folgebewegung hat sich augenblicklich der Geschwindigkeit des Objektes angepaßt. Die Vorbereitungszeit einer langsamen Folgebewegung liegt bei etwa 100-150 ms. Ziele können bis zu einer Geschwindigkeit von 40 ^{o}/s getreu verfolgt werden. Erst bei höheren Geschwindigkeiten sind zusätzlich Sakkaden notwendig, um das Objekt wieder einzuholen.

Neben den Sakkaden und den langsamen Folgebewegungen existieren noch weitere Formen von Augenbewegungen. Hier sind insbesondere die *Vergenzbewegungen*, die *optokinetischen Bewegungen* und die *vestibulo-okulären Reflexe* zu nennen. Sie stellen zusammen mit den langsamen Folgebewegungen die Fixierungsmechanismen des menschlichen Sehsystems dar. Konvergente bzw. divergente Augenbewegungen ermöglichen die binokulare Fixation von nahen bzw. entfernten Objekten. Die Bewegungsrichtungen sind jeweils gegensinnig. Vergenzbewegungen weisen eine relativ kurze Latenz von 150-200 ms auf, sind langsam (max. 10 ^{o}/s) und reflektorisch. Sie können aber willentlich beeinflußt werden. Optokinetische Bewegungen stabilisieren die Bilder auf der Netzhaut, wenn sich ein Objekt relativ zum Betrachter bewegt. Sie sind reflektorisch und werden ausgelöst, wann immer es zur Verschiebung großer Abbildungsbereiche auf der Retina kommt. Sie besitzen eine hohe Latenz von bis zu 20 s und sind ebenfalls relativ langsam (max. 80 ^{o}/s). Vestibulo-okuläre Reflexe dienen zur Aufrechterhaltung der Fixation bei Kopf- und Körperbewegungen. Sie werden sehr schnell ausgelöst (max. 15 ms Latenz) und erreichen hohe Geschwindigkeiten (max. 500 ^{o}/s).

Die ersten Aufzeichnungen menschlicher Augenbewegungen stammen von Yarbus [Yarbus 1969], der Augenbewegungen analysierte, die Personen beim Betrachten komplexer Photographien durchführen (Abbildung 3.1). Das wesentlichste Ergebnis seiner Untersuchungen ist, daß Menschen zyklisch immer wieder dieselben Fixationspunkte aufsuchen. Das heißt, auch bei ausreichender Zeit werden keine neuen Fixationspunkte hinzugenommen, so daß einige Areale also niemals foveal, sondern immer nur extrafoveal analysiert werden. Die Dauer der Fixationen nimmt aber mit der Anzahl der durchgeführten Zyklen zu, während die Sprungweite der Sakkaden abnimmt.

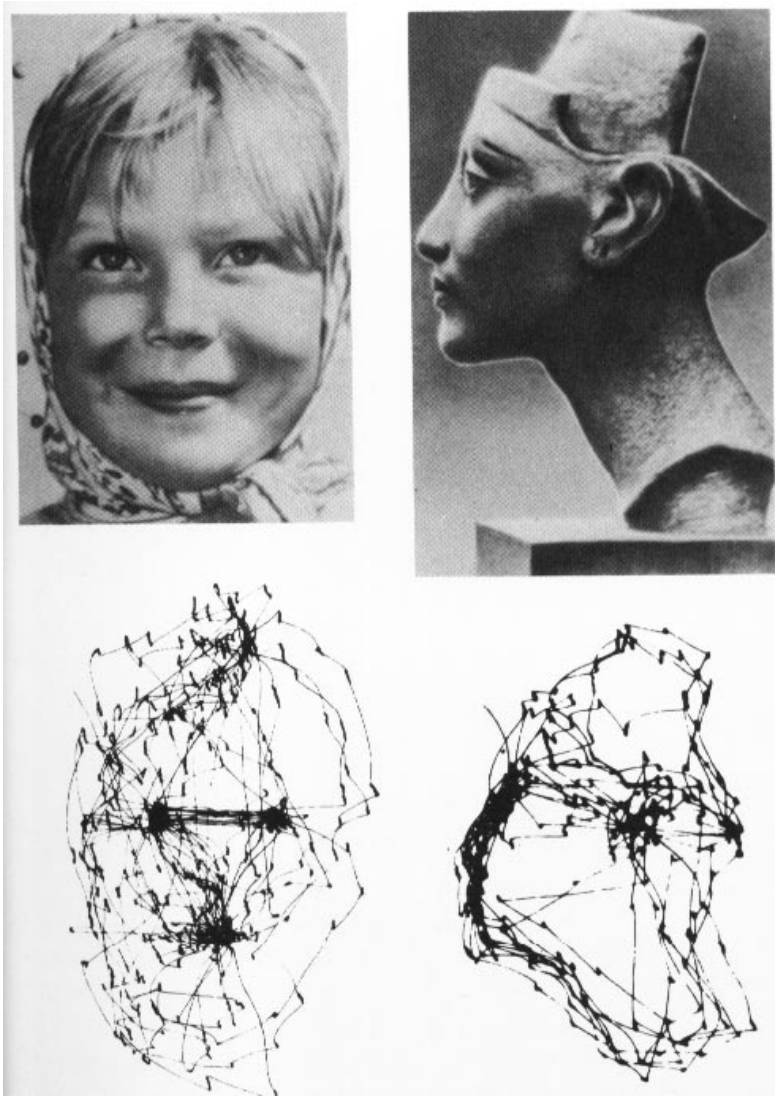


Abbildung 3.1: Aufzeichnungen von Augenbewegungen (aus [Kandel 1991], nach [Yarbus 1969])

Noton et al. [Noton 1970, 1971a+b] konnten auf der Basis der Ergebnisse von Yarbus zeigen, daß die Reihenfolge der Fixationen keinesfalls zufällig ist. Testpersonen führen deterministische Augenbewegungen durch, um Objekte zu lernen und wiederzuerkennen. Diese determini-

stischen Augenbewegungen werden zurückgehend auf Noton als Scanpath bezeichnet. Jeder Scanpath ist charakteristisch für eine bestimmte Szene und eine bestimmte Person. Die aus den genannten Beobachtungen entwickelte Scanpath-Theorie besagt, daß ein Objekt als eine Sequenz von Teilobjekten zusammen mit den zugehörigen Augenbewegungen in einer internen Repräsentation gespeichert wird. Eine solche interne Repräsentation könnte die Form eines Merkmalsringes, wie in Abbildung 3.2 gezeigt, haben.

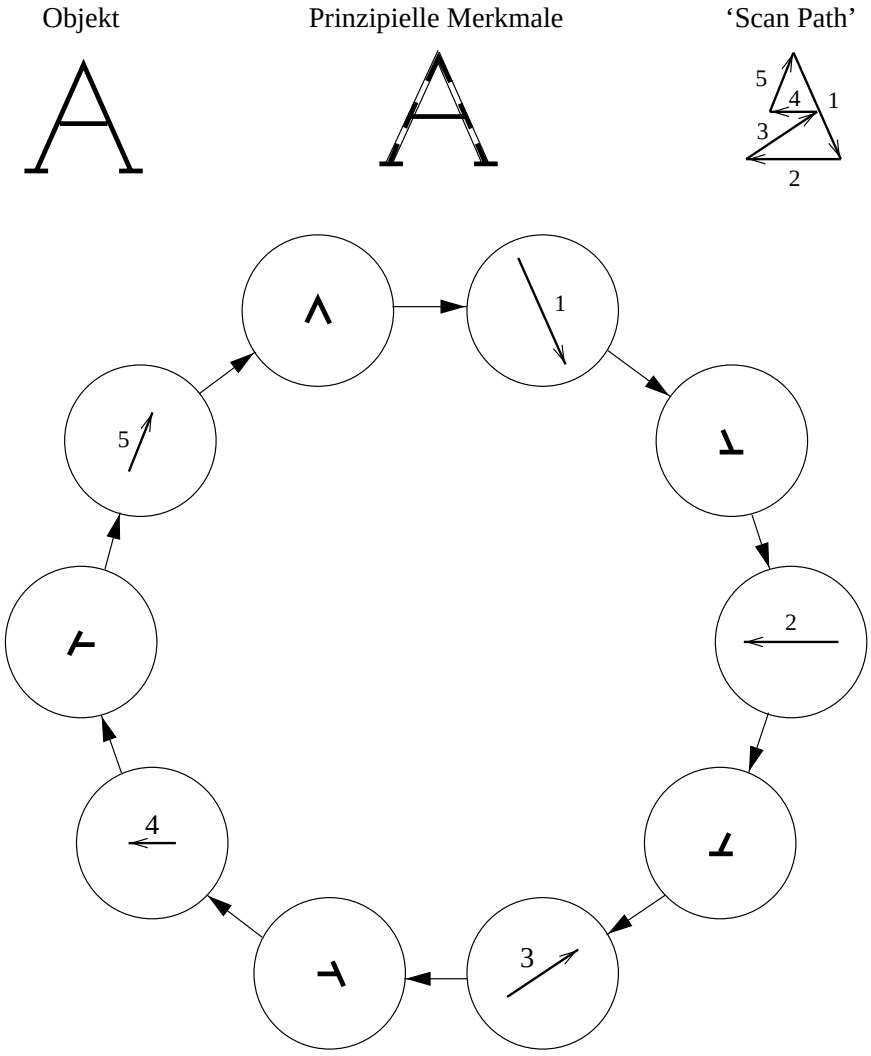


Abbildung 3.2: Merkmalsring der Scanpath-Theorie (aus [Noton 1971a])

Im wesentlichen gibt es zwei experimentelle Ergebnisse, die einen seriellen Erkennungsprozeß wie die Scanpath-Theorie stützen: Zum einen wird mehr Zeit benötigt, um ein komplexes Objekt zu erkennen, als um ein einfaches Objekt zu erkennen; zum anderen wird mehr Zeit benötigt, um ein Target zu erkennen, als um einen Distraktor zu verwerfen. Der Scanpath kann je nach Entfernung zwischen Betrachter/-in und Objekt eine Folge von Augenbewegungen oder von internen Aufmerksamkeitswechseln sein.

Ein erweitertes Modell der Scanpath-Theorie wurde von Groner et al. [Groner 1989] vorgestellt. Die Autoren schlagen eine Trennung zwischen lokalen und globalen Scanpathes vor. Die lokalen Scanpathes sind abhängig von den lokalen Beschaffenheiten der einzelnen Objekte und ihren nachbarschaftlichen Beziehungen, wogegen die globalen Scanpathes eine aufgabenspezifische Suchstrategie verfolgen. In der Literatur findet sich bereits eine Vielzahl visuomotorischer Experimente, in denen die Augenbewegungen bei der Durchführung unterschiedlichster Aufgaben analysiert wurden (siehe z. B. [O'Regan 1990], [Ballard 1995, 1997], [Tanenhaus 1995], [Furneaux 1997]). Zusammenfassend läßt sich feststellen, daß die Augenbewegungen genau wie die Aufmerksamkeitswechsel weitgehend aufgabenabhängig sind.

3.2 Ergebnisse der experimentellen Psychophysik

Befaßt man sich mit den Arbeiten der Psychophysik zur visuellen Aufmerksamkeit, so läßt sich ein wissenschaftlicher Disput über die Frage feststellen, ob dieser Prozeß früh oder spät im menschlichen Gehirn stattfindet. Die '*Early Selection Theory*' besagt im wesentlichen, daß die Selektion lokaler Information vor der Erkennung und Klassifizierung von Objekten stattfinden muß, wogegen die '*Late Selection Theory*' besagt, daß nur eine komplette Beschreibung der in der Szenen vorhandenen Objekte zur Selektion des Targets führen kann (Überblick siehe [Allport 1989], [Yantis 1990]). Beide Theorien werden von einer Vielzahl von Autoren unterstützt. Als Vertreter der 'Early Selection Theory' kann man z. B. [Broadbent 1982], [Hoffman 1986], [Johnston 1985, 1986], Treisman (siehe Abschnitt 3.2.3), [Ullman 1984], Wolfe (Abschnitt 3.2.4) und weitere ansehen. Die 'Late Selection Theory' wird u. a. gestützt durch die Arbeiten von [Coltheart 1984], Duncan (Abschnitt 3.2.5), [Mewhort 1984], [Neill 1987] und [Pashler 1987]. Eine endgültige Klärung dieser Frage scheint nicht in Sicht.

Einige psychophysische Arbeiten zur visuellen Aufmerksamkeit werden in den nächsten Abschnitten vorgestellt. Diese Arbeiten unterscheiden sich von der Vielzahl anderer Arbeiten dadurch, daß sie nicht nur Experimente, sondern auch daraus abgeleitete Theorien enthalten. Zunächst soll aber die experimentelle Form der visuellen Suche, aus der die Mehrzahl der psychophysischen Ergebnisse abgeleitet worden ist, näher beschrieben werden.

3.2.1 Die experimentelle Form der visuellen Suche

Die visuelle Suche ist eine der bedeutendsten visuellen Verhaltensweisen. Organismen suchen mit Hilfe ihres visuellen Systems nach Nahrung, Gefahrenquellen, Zufluchtsmöglichkeiten und Artgenossen. In der experimentellen Psychophysik ist unter dem Begriff *visuelle Suche* eine Klasse von Experimenten zusammengefaßt, in denen Testpersonen auf einem Display nach einem bestimmten Objekt, dem *Target*, suchen. Störobjekte, die sogenannten *Distrakto-*

ren, erschweren die Suche nach dem Target. Gemessen wird in der Regel entweder die Reaktionszeit in Abhängigkeit von der Displaygröße (Anzahl der Elemente im Display) oder die Fehlerrate. Die visuelle Suche ist beendet, wenn die Testperson das Target gefunden hat, alle (in Betracht kommenden) Stimuli im Display untersucht hat oder ein Target (falsch oder richtig) geraten hat.

Man unterscheidet im wesentlichen drei verschiedene Versuchsanordnungen: die ‘Feature Search’, die ‘Conjunction Search’ und die ‘Triple-Conjunction Search’. In der ‘Feature Search’ (deutsch *Merkmalssuche*) suchen die Testpersonen nach einem Target, das sich von den Distraktoren in genau einem Merkmal unterscheidet, also z. B. die Suche nach einem roten Kreis unter grünen Kreisen. Die Elemente in einer ‘Conjunction Search’ (deutsch *Kombinationssuche*) besitzen jeweils zwei Merkmale, wobei das Target kein exklusives Merkmal besitzt, sondern nur durch die Merkmalskombination definiert ist. Ein typisches Beispiel ist die Suche nach einer roten, vertikalen Linie unter grünen, vertikalen Linien und roten, horizontalen Linien. In der ‘Triple-Conjunction Search’ weisen alle Elemente des Displays drei Merkmale auf. Auch hier besitzt das Target kein einzelnes Merkmal, das es von den Distraktoren unterscheidet, sondern geht wie in der ‘Standard-Conjunction Search’ nur durch die Kombination der Merkmale aus den Distraktoren hervor. Zwei verschiedene Formen der ‘Triple Conjunction Search’ sind möglich: zum einen kann das Target mit jedem Distraktor ein Merkmal teilen, zum anderen kann das Target auch zwei Merkmale mit allen Distraktoren teilen. Ein Beispiel für die erste Form ist die Suche nach einem großen, roten Dreieck unter kleinen, grünen Dreiecken, kleinen, roten Kreisen und großen, grünen Kreisen. Ein Beispiel für die zweite Form ist die Suche nach einem großen, roten Dreieck unter großen, roten Kreisen, großen, grünen Dreiecken und kleinen, roten Dreiecken.

Die Bedingungen in ‘Visual Search’-Experimenten lassen sich noch durch eine Vielzahl weiterer Faktoren variieren. So kann man Experimente unterscheiden, in denen den Testpersonen das Target a priori nicht bekannt ist, und solche, in denen das Target vorher bekannt gegeben wurde. Für die zweite Form lassen sich neben den Suchen, in denen das Target im Display vorhanden ist, auch solche Suchen entwerfen, in denen das vorher bezeichnete Target nicht vorhanden ist. Eine weitere Variationsmöglichkeit bietet die Homogenität der Distraktoren. Eine Merkmalssuche mit inhomogenen Distraktoren wäre z. B. die Suche nach einem roten Kreis unter blauen, gelben und grünen Kreisen. Auch durch das Zahlenverhältnis der verschiedenen Distraktoren läßt sich Einfluß auf das Ergebnis der visuellen Suche nehmen. So ist die visuelle Suche in ‘Conjunction Search’-Experimenten dann am schwierigsten, wenn die verschiedenen Ausprägungen eines Merkmals im Verhältnis 1:1 vorkommen ([Egeth 1984], [Poisson 1992]).

Experimente zur visuellen Suche haben gezeigt, daß ein begrenzter Satz an relativ grundlegenden Merkmalen existiert und daß diese Merkmale parallel berechnet werden [Treisman 1980, 1985]. Es handelt sich hierbei allerdings nicht um primitive Merkmale, wie sie z. B. aus Theorien über die frühe visuelle Wahrnehmung bekannt sind. Es wird vielmehr vermutet, daß bereits ein Großteil der perzeptuellen Verarbeitung bis zur Entstehung dieser Merkmale stattgefunden hat ([Bravo 1990], [Cavanagh 1990], [Epstein 1990], [Enns 1991]). So besitzen Stimuli der vi-

suellen Suche Objekteigenschaften, die in frühen visuellen Repräsentationen fehlen, z. B. das Merkmal der Geschlossenheit auch für teilweise verdeckte Stimuli [Enns 1992]. Die visuelle Suche basiert demnach auf einer relativ späten Merkmalsrepräsentation.

Eine gewisse Einigkeit besteht auch darüber, daß die visuelle Suche von den lokalen Unterschieden der Stimuli, den Nachbarschaftsbeziehungen, bestimmt wird, wenn kein Target vorgegeben ist. Sie basiert dagegen zumindest teilweise auf dem kategorischen Match zwischen einem Target-Template und den Stimuli im Display, wenn das Target a priori bekannt ist [Wolfe 1993]. Vermutlich gibt es keinen qualitativen Unterschied zwischen paralleler und serieller Suche. Eine Kombinationssuche kann parallel sein, eine Merkmalssuche kann seriell sein, je nach Eindeutigkeit des Targets.

Eine gesonderte Bemerkung möchte ich an dieser Stelle noch zu der Verwendung des Begriffs Merkmal machen: Einige Autoren beschreiben etwa Rot und Vertikal als Merkmale und z. B. Farbe und Orientierung als Dimensionen, wogegen andere Autoren Rot und Vertikal als Ausprägungen und Farbe und Orientierung als Merkmale bezeichnen. Ich werde im folgenden, falls eine Unterscheidung zwischen diesen Begriffsklassen für das Verständnis wesentlich ist, Merkmalen wie Rot das Adjektiv *intradimensional* voranstellen und Merkmalen wie Farbe das Adjektiv *interdimensional* zuordnen.

3.2.2 Eriksens ‘Zoom Lens’-Modell

Eriksen et al. [Eriksen 1985, 1986] definieren Aufmerksamkeit als eine begrenzte Menge an Verarbeitungskapazitäten, die in variierenden Anteilen auf verschiedene Aufgaben verteilt werden kann. Insbesondere kann die visuelle Aufmerksamkeit über das visuelle Feld verteilt werden. Eine parallele Suche ist dann das Ergebnis einer gleichmäßigen Allokierung der Aufmerksamkeit über das gesamte visuelle Feld (*verteilte Aufmerksamkeit*). Eine serielle Suche ist nötig, wenn die Komplexität der Suchaufgabe steigt. Die visuelle Aufmerksamkeit wird dann auf ein begrenztes, bis zu einem Grad kleines Gebiet konzentriert und ihr Fokus sukzessive verschoben, bis das Target detektiert ist. Eriksen vergleicht die visuelle Aufmerksamkeit mit einer Zoom-Linse, die sowohl ein weites Blickfeld mit geringer Auflösung als auch ein nahezu beliebig eingeschränktes Blickfeld mit entsprechend besserer Auflösung ermöglicht.

Eriksen führte insbesondere eine Vielzahl von psychophysischen Experimenten unter Einsatz des sogenannten Precuings durch. Hierbei wird den Testpersonen die mögliche Position für das Auftreten eines Targets vorab, z. B. durch das kurze Aufblinken eines Cursors, angezeigt. Das Precuing beschleunigt die Suche, indem es den Testpersonen erlaubt, ihre Aufmerksamkeit vorab auf die geeignete Stelle zu konzentrieren. Die Ergebnisse aus den Eriksenschen Untersuchungen lassen sich wie folgt zusammenfassen: Findet kein Precuing statt, verwenden die Testpersonen die parallele Suche als Suchstrategie. Sie verteilen ihre Aufmerksamkeit also gleichermaßen über das gesamte Display. Werden dagegen mehrere Positionen für das mögli-

che Erscheinen des Targets angezeigt und gemäß der Auftrittswahrscheinlichkeit des Targets an der jeweiligen Position gewichtet, so untersuchen die Testpersonen die Positionen seriell in der erfolgversprechendsten Reihenfolge. Befindet sich das Target an keiner der vorgegebenen Positionen, wird die Suche seriell fortgesetzt. Umfaßt die a priori angezeigte Region das halbe Display oder noch mehr Objektpositionen, werden durch das Precuing keine verkürzten Antwortzeiten erzielt, denn in diesem Fall führen die Testpersonen wiederum eine parallele Suche durch.

Psychophysische Experimente haben gezeigt, daß die visuelle Aufmerksamkeit über den 'Focus of Attention' gleichverteilt ist, wenn die Größe und Lage des FOA durch das Precuing eingestellt wurde und das Auftreten des Targets an jeder dieser Positionen gleichwahrscheinlich ist. Die Größe des FOA entspricht hier der Anzahl der benachbarten Objektpositionen, die das Target gemäß des Precuings enthalten können. Alle vorab angezeigten Positionen werden also parallel untersucht. Der Einfluß der Distraktoren auf die Target-Identifikation ist dann nicht von ihrer relativen Position zum Target abhängig, sondern von ihrer Zugehörigkeit zum FOA. Zwei gleichartige Distraktoren mit gleicher relativer Position zum Target haben also unterschiedlichen Einfluß auf die Identifikation des Targets, wenn die Position des einen Distraktors beim Precuing markiert wurde und die andere nicht.

Eriksen nimmt an, daß die volle Kapazität der Aufmerksamkeit in der Regel nicht für eine einzige Aufgabe eingesetzt wird. Das heißt, daß mit steigender Komplexität der Aufgabe mehr Aufmerksamkeit allokiert werden kann. Eine sinkende Reaktionszeit oder Trefferquote bei steigender Größe des FOA ist demnach nicht immer meßbar. Mit dieser Annahme läßt sich das Eriksensche 'Zoom Lens'-Modell nur sehr schwer durch psychophysische Experimente validieren.

3.2.3 Treismans 'Feature Integration Theory'

Die 'Feature Integration Theory' ist die bekannteste und wohl auch die akzeptierteste Theorie auf dem Gebiet der visuellen Aufmerksamkeit (siehe [Treisman 1980], [Treisman 1984], [Treisman 1985], [Treisman 1988], [Treisman 1990], [Treisman 1992]). Es handelt sich hierbei um eine Theorie, die in den Jahren ständig modifiziert wurde, so daß eine Verständigung mit anderen Wissenschaftlern über diese Theorie mitunter schwierig ist. Ich beziehe mich bei der Darstellung der 'Feature Integration Theory' überwiegend auf die in [Treisman 1992] beschriebene Version.

Die 'Feature Integration Theory' besagt im wesentlichen, daß separate Merkmale früh, automatisch und parallel über das visuelle Feld verarbeitet werden, so daß ein Target in einer Merkmalssuche augenblicklich detektiert wird ('Pop out'). Die detektierten Merkmale werden in sogenannten 'Feature Maps' (deutsch Merkmalskarten) verwaltet. Diese Merkmalskarten sind intradimensional (eine Merkmalskarte für Blau, eine für Rot, eine für Vertikal, etc.) und retinotop

(Nachbarschaftsbeziehungen bleiben erhalten). Die Allokierung visueller Aufmerksamkeit in Form einer seriellen Suche ist notwendig, um ein Target in einer Kombinationssuche zu detektieren, oder bei geringer Unterscheidbarkeit der Stimuli, wenn also Stimuli das gleiche Merkmal zu einem unterschiedlichen Grad teilen. Gruppierungseffekte bewirken, daß in der Kombinationssuche eher homogene Gruppen denn einzelne Stimuli gleichzeitig erfaßt werden. Ist das Target a priori bekannt, entsteht eine Vorauswahl der zu untersuchenden Stimuligruppen durch die Reduktion der Aktivitäten solcher Distraktor-Positionen, die Merkmale enthalten, die inkonsistent mit der Target-Beschreibung sind (siehe Abbildung 3.3). So ist die Kombinationssuche nach einem bekannten Target im Durchschnitt schneller als die Suche nach einem unbekannten Target und kann im Extremfall sogar zum ‘Pop out’ des Targets führen. Ist das Target dagegen nicht a priori bekannt, ist die Suche seriell, linear (d. h. die Suchzeit steigt linear mit der Anzahl der Distraktorgruppen), und selbstbeendend (d. h. die Suchzeit ist durchschnittlich doppelt so groß, wenn kein Target vorhanden).

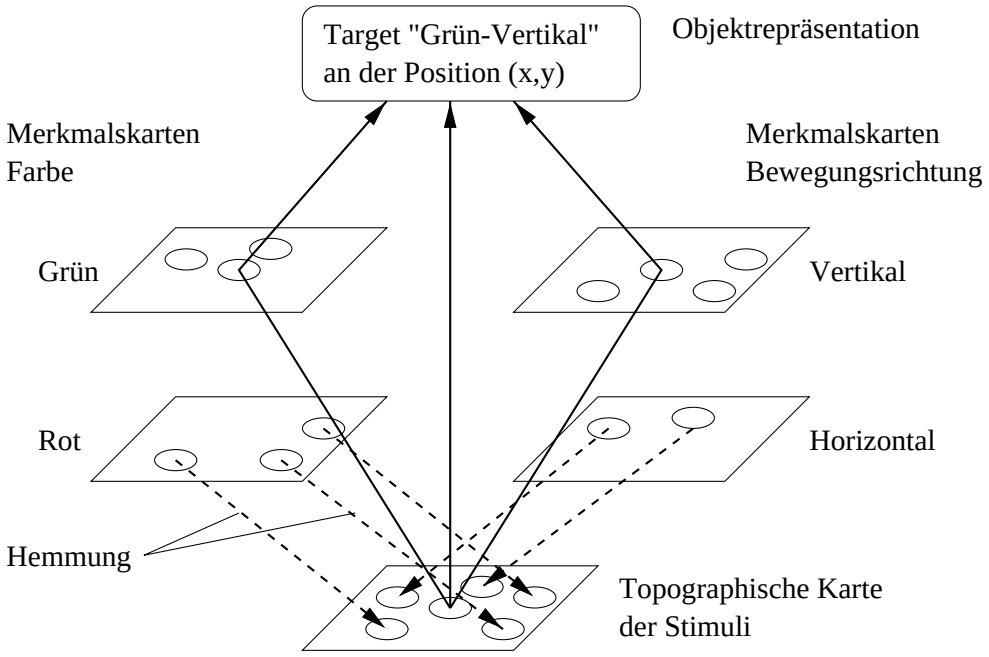


Abbildung 3.3: In der Kombinationssuche werden Gruppen von Distraktoren gehemmt, die als Target sicher ausgeschlossen werden können (aus [Treisman 1991])

Die ‘Feature Integration Theory’ postuliert in ihrer aktuellsten Version [Treisman 1992], daß Personen in ‘Visual Search'-Experimenten zunächst ihre Aufmerksamkeit über das gesamte Display verteilt haben, dieses global verarbeiten und in einer einzigen Beschreibung repräsentieren. Hierdurch können globale Eigenschaften, wie z. B. die Homogenität, bestimmt werden. Mit dem Begriff Homogenität ist hier gemeint, daß nur eine intradimensionale Merkmalskarte pro interdimensionalem Merkmal Aktivität aufweist (im Gegensatz zur Inhomogenität, wenn mehrere Merkmalskarten desselben interdimensionalen Merkmals Aktivität aufweisen). Zusätzlich kann die Position und Form der Grenzen innerhalb jeder Merkmalskarte bestimmt und

anschließend über mehrere Merkmalskarten vereinigt werden. Grenzen zwischen Merkmalskombinationen (engl. Conjunctions) können dabei nicht global detektiert werden, weil ihre Merkmale integriert werden müssen, was nicht parallel über das gesamte Display geschehen kann. Um die Anwesenheit eines ‘Feature Targets’ in einem ansonsten homogenen Feld detektieren zu können, wird die Aufmerksamkeit auf den Ort des ungleichartigen Merkmals verengt. Dieses kann schneller durchgeführt werden, wenn das relevante Merkmal a priori bekannt ist, weil die entsprechende Merkmalskarte dann mit erhöhtem Gewicht in die ‘Master Map of Location’ eingehen kann. Die ‘Master Map of Location’ ist eine zentrale, topographische Karte, in der die Informationen aus den Merkmalskarten zu Conjunctions integriert sind. Ist das Target unbekannt, müssen die verschiedenen interdimensionalen Merkmale überprüft werden, um festzustellen welches interdimensionale Merkmal zwei statt einer aktivierten Merkmalskarte besitzt. Auf die Merkmalskarten wird immer über die ‘Master Map of Location’ und die aktuelle Target-Beschreibung zugegriffen. Enthält der FOA allerdings mehr als ein Element, so ist die Position der Stimuli nicht verfügbar (es sei denn, ihre räumliche Anordnung läßt ein globales Merkmal entstehen). Die räumliche Unbestimmtheit kann zu inkorrekten Merkmalskombinationen, den sogenannten ‘*Illusory Conjunctions*’, führen. Die Antwort auf eine visuelle Suchaufgabe beruht also auf einer vereinigten Messung von Merkmalsaktivitäten innerhalb eines FOA. Visuelle Suche ist dann seriell, wenn fokussierende Aufmerksamkeit zur korrekten Integration der Merkmale benötigt wird.

Eine interessante Erweiterung zu Treismans ‘Feature Integration Theory’ findet sich in [Nakayama 1986]. Nakayama und Silverman bestätigen im wesentlichen die Treismanschen Untersuchungsergebnisse, indem sie zeigen, daß in ‘Feature Search’-Experimenten, in denen die Farbe oder die Bewegung das relevante Merkmal darstellen, das Target in einer parallelen Suche detektiert wird. In ‘Conjunction Search’-Experimenten, in denen die Kombinationen aus den Merkmalen Farbe und Bewegung bestehen, müssen die Personen dagegen eine serielle Suche bis zur Detektion des Targets durchführen. Bestehen die Merkmalskombinationen dagegen aus den Merkmalen Farbe und Tiefe oder Bewegung und Tiefe, so wird das Target wiederum in einer parallelen Suche detektiert. Eine erste Feststellung, die man zu diesen Beobachtungen machen kann, ist, daß visuelle Aufmerksamkeit räumlich, d. h. wohl auch bezüglich der Tiefe, begrenzt werden kann. Treisman hat gezeigt, daß Beobachter ihre Aufmerksamkeit seriell auf verschiedene zweidimensionale Gebiete einer komplexen Szene richten können. Die Suche innerhalb eines solchen Gebietes erfolgt dann aber parallel (siehe auch [Mordkoff 1990]). Ähnlich verhält es sich auch mit den verschiedenen Ebenen der stereoskopischen Tiefe. Innerhalb einer Tiefenebene erfolgt die Suche nach einfachen Merkmalen parallel. Nakayamas und Silvermans Untersuchungen stellen also eine Erweiterung der Treismanschen Experimente in den 3D-Raum dar. Tiefe ist ebenso wie die retinale Position ein primärer Index, auf den die anderen Merkmale bezogen werden, d. h. die Kombination aus Tiefe und retinaler Position dient der Verknüpfung/Indizierung aller übrigen Merkmale.

3.2.4 Wolfes ‘Guided Search’

Das von Wolfe und Cave entwickelte ‘Guided Search’-Modell läßt sich sowohl in die Reihe der psychophysischen als auch in die Reihe der maschinellen Aufmerksamkeitsmodelle einordnen. Das Modell basiert größtenteils auf den Ergebnissen psychophysischer Experimente, die von den Autoren selbst durchgeführt wurden [Wolfe 1989]. Gleichwohl ist es von den Autoren auch auf dem Computer simuliert worden [Cave 1990]. Primäres Ziel des ‘Guided Search’-Modells ist es, ‘Visual Search’-Experimente zu erklären. Allerdings soll das Modell auch auf reale Szenen übertragbar sein [Wolfe 1993, 1994].

Die Motivation für die Entwicklung eines Aufmerksamkeitsmodells bezogen Wolfe und Cave aus der Kritik an einer frühen Version der ‘Feature Integration Theory’ [Treisman 1985]. Diese Version bestand aus einem zweistufigen Modell, in dem die erste Stufe die Detektion der Merkmale repräsentierte und die zweite Stufe eine rein serielle Suche (nicht begrenzt parallel, sondern Stimulus für Stimulus) beinhaltete. Das heißt, die serielle Stufe profitierte nicht von den Ergebnissen der parallelen Stufe, sondern war von dieser entkoppelt. Mit einem solchen Modell läßt sich nicht erklären, warum unter bestimmten Voraussetzungen die Suchzeit in ‘Conjunction Search’-Experimenten nicht in dem zu erwartenden Maße, also linear, mit der Anzahl der Distraktoren steigt. Die Theorie konnte in dieser Form auch nicht voraussagen, warum die Suche nach Triple-Conjunctions, in denen jeder Distraktor eines seiner drei Merkmale mit dem Target teilt, einfacher ist als die Suche in Standard-Conjunction-Experimenten.

‘Guided Search’ ist ein zweistufiges Modell bestehend aus einer parallelen Stufe und einer nachfolgenden, seriellen Stufe, die allerdings von den Ergebnissen der ersten Stufe profitiert. Im Unterschied zur ‘Feature Integration Theory’ sind die Merkmalskarten der parallelen Stufe interdimensional (eine Merkmalskarte für Farbe, eine für Orientierung, etc.). Jede Position in einer Karte enthält genau einen Aktivierungswert, der sich aus einer ‘Bottom up’- und einer ‘Top down’-Komponente zusammensetzt. Die Höhe der Aktivierung an jeder Stelle der ‘Bottom up’-Komponente hängt zunächst von dem Unterschied zu den Elementen in der jeweils nächsten Nachbarschaft ab. Zusätzlich werden die Merkmale in Abhängigkeit von dem Grad, in dem sie einer Kategorie entsprechen, gewichtet (siehe Abbildung 3.4). In ‘Guided Search 2.0’ [Wolfe 1993, 1994] sind zwei Merkmalskarten, eine für die Orientierung und eine für die Farbe, implementiert. Für das Merkmal Orientierung existieren die Kategorien Steil, Flach, Links und Rechts; für das Merkmal Farbe existieren die Kategorien Rot, Gelb, Grün und Blau.

Die ‘Top down’-Aktivierung besteht aus der Auswahl einer Kategorie des Merkmals Orientierung und einer Kategorie des Merkmals Farbe. Hierbei gilt es, die Kategorien auszuwählen, die das Target am besten von den Distraktoren unterscheiden. Dieses müssen nicht notwendigerweise die Kategorien sein, die für das Target den größten Output liefern.

Die verschiedenen ‘Bottom up’- und ‘Top down’-Karten werden aufsummiert zur ‘Activation Map’, die die Aufmerksamkeit in der Reihenfolge der Maxima lenkt. Die verschiedenen Kom-

ponenten der Merkmalskarten können vor ihrer Aufsummierung noch je nach Bedeutung für die Target-Suche gewichtet werden. Desweiteren kann so psychophysischen Experimenten Rechnung getragen werden, die gezeigt haben, daß das Merkmal Farbe weniger rauschanfällig ist als das Merkmal Orientierung (siehe z. B. [Bundesen 1983], [Egeth 1984]).

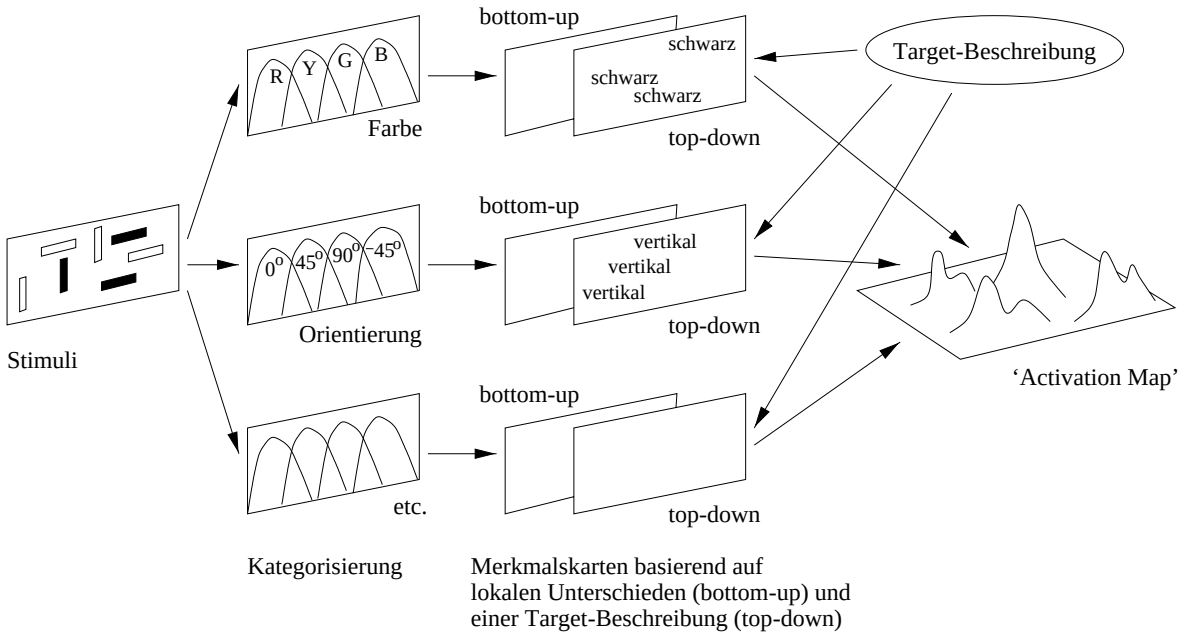


Abbildung 3.4: Das 'Guided Search'-Modell (aus [Wolfe 1994])

In dem 'Guided Search'-Modell profitiert die zweite, serielle Aufmerksamkeitsstufe von den Ergebnissen der ersten, parallelen Stufe. Der parallele Aufmerksamkeitsprozeß trifft bereits eine Vorauswahl möglicher Targets, so daß ein Großteil der Distraktoren in dem seriellen Prozeß nicht mehr untersucht werden muß. Allerdings ist die Übertragung der Ergebnisse des parallelen Prozesses (oder der parallele Prozeß selbst) durch eine Art Rauschen gestört. Ohne dieses Rauschen wäre die Reaktionszeit auch in der Kombinationssuche unabhängig von der Anzahl der Distraktoren. Wolfe und Cave vermuten, daß das Signal-zu-Rausch-Verhältnis in hohem Maße von der Auffälligkeit der Stimuli beeinflußt wird.

Jedesmal wenn der FOA während der seriellen Suche wechselt, richtet er sich auf die Stelle, die von dem parallelen Prozeß als die wahrscheinlichste für das Beinhalten des Targets angesehen wird. Hierbei werden bereits untersuchte Positionen nicht weiter betrachtet. Mit jedem Blickwechsel erhöht sich demnach die Wahrscheinlichkeit, daß der sich kontinuierlich erneuernde parallele Prozeß das Target liefert. In dieser Betrachtungsweise sind die Mechanismen, die während einer Merkmals- oder einer Kombinationssuche wirksam werden, also gleich. In jedem Fall liefert der parallele Prozeß das Signal, das die Aufmerksamkeit auf das Target lenkt. Ist der Signal-zu-Rausch-Abstand groß genug, so ist der Verlauf der Blickwechsel demnach auch in der Kombinationssuche nicht zufällig und die Anzahl der betrachteten Positionen liegt im Mittel unter der halben Anzahl der Distraktoren.

Die ‘Triple Conjunction Search’, in der das Target und die Distraktoren nur eines der drei Merkmale gemeinsam haben, liefert flachere "Reaktionszeit x Anzahl der Distraktoren"-Funktionen als die ‘Standard Conjunction Search’, in der Target und Distraktoren eines von zwei Merkmalen teilen. Der Grund hierfür liegt in der größeren Informationsmenge, die zur Trennung von Target und Distraktoren zur Verfügung steht und das Signal-zu-Rausch-Verhältnis positiv beeinflußt. ‘Guided Search’ sagt aber noch weitere Ergebnisse psychophysischer Untersuchungen richtig voraus: Teilen in einer ‘Triple Conjunction Search’ alle Distraktoren zwei Merkmale mit dem Target, so erhöht sich die Suchzeit gegenüber der ‘Standard Conjunction Search’. In den ‘Bottom up’-Komponenten der Merkmalskarten stimmen jeweils zwei Drittel der Distraktoren mit dem Merkmal des Targets überein. Demnach stellt sich erhöhte Aktivität an den Positionen der ‘Bottom up’-Komponenten ein, die eine vom Target unterschiedliche Eigenschaft besitzen. Die ‘Bottom up’-Komponenten verhindern also eine schnelle Target-Identifikation. In jeder ‘Top down’-Komponente erfahren die Elemente mit der Target-Eigenschaft erhöhte Aktivität. Während das Target also in allen drei ‘Top down’-Komponenten erhöhte Aktivierung erfährt, trifft dieses für die Distraktoren jeweils nur in einer Komponente zu. Insgesamt ergibt sich eine Suchzeit die gleich oder leicht höher liegt als in ‘Standard Conjunction Search’-Experimenten.

Als ‘Disjunctive Search’ bezeichnet man ‘Visual Search’-Experimente, in denen zwei oder mehr Targets definiert sind. Ein Target kann also z. B. rot oder vertikal sein. Das ‘Guided Search’-Modell sagt richtig voraus, daß ein solches Target schnell, ähnlich einer Merkmalssuche gefunden werden kann. Es stellt sich in einer der beiden ‘Bottom up’-Komponenten an der Stelle, an der sich das Target befindet, eine hohe Aktivität ein, da sich das Target in diesem einen Merkmal von allen Distraktoren unterscheidet. In der ‘Bottom up’-Komponente des anderen Merkmals ist die Aktivität dagegen gleichverteilt. ‘Top down’-Aktivität, durch eine der zwei Target-Eigenschaften hervorgerufen, unterstützt den Suchprozeß.

Die Plausibilität des ‘Guided Search’-Modells ist also mit der der ‘Feature Integration Theory’ vergleichbar. Beide Modelle besitzen allerdings auch ähnliche Charakteristika.

3.2.5 Duncans und Humphreys ‘Stimulus Similarity Theory’

Auch Duncan und Humphreys ([Duncan 1989, 1992], [Humphreys 1989]) führten verschiedene Experimente durch, deren Resultate sich nicht mit der ursprünglichen ‘Feature Integration Theory’ ([Treisman 1980], [Treisman 1985]) erklären ließen. Duncan und Humphreys stellten fest, daß die Suchzeit in ‘Conjunction Search’-Experimenten schnell und nicht linear ist, wenn hohe Ähnlichkeit zwischen den Distraktoren besteht (bei geringer Ähnlichkeit mit dem Target) oder das Display in Relation zur Stimulusgröße klein ist. Eine serielle, selbstbeendene Suche, wie von Treisman gemessen, stellt sich nur dann ein, wenn die Distraktoren hinreichend heterogen sind.

Aufgrund dieser Ergebnisse entwickelten Duncan und Humphreys ihre ‘Stimulus Similarity Theory’. Die Theorie basiert im wesentlichen auf der Ähnlichkeit der Stimuli. Sie widerspricht damit einer klaren Trennung zwischen einer parallelen Suche in einer ‘Feature Search’ und einer seriellen Suche in einer ‘Conjunction Search’; vielmehr ist die Effizienz einer Suche ein Kontinuum (siehe auch [Quinlan 1987]). Eine strikte Trennung von paralleler und serieller Suche wird auch in weiteren Arbeiten als Übervereinfachung kritisiert ([Cavanagh 1990], [Humphreys 1996]).

Die Effizienz einer Suche verringert sich, wenn die Ähnlichkeit zwischen den Targets und den Non-Targets (Distraktoren), die ‘T-N Similarity’, steigt. Sie verringert sich weiterhin, falls die Ähnlichkeit zwischen den Distraktoren, die ‘N-N Similarity’, sinkt; wobei sich beide Effekte gegenseitig beeinflussen. Steigende T-N-Ähnlichkeit hat nur einen geringen Einfluß auf die Suchzeit, wenn die N-N-Ähnlichkeit hoch ist. Sinkende N-N-Ähnlichkeit hat einen geringen Effekt, wenn die T-N-Ähnlichkeit niedrig ist (vgl. Abbildung 3.5).

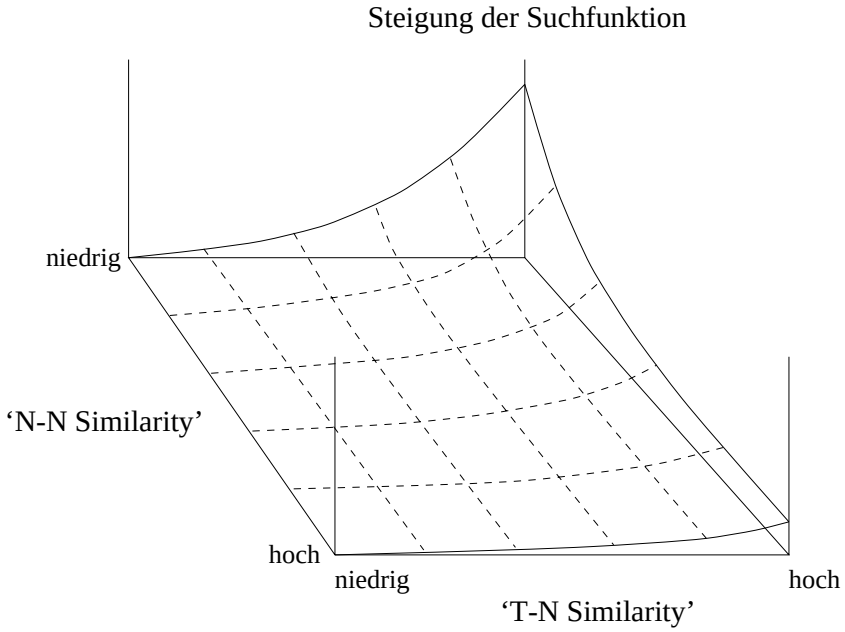


Abbildung 3.5: Qualitativer Einfluß der N-N- und T-N-Ähnlichkeit auf die visuelle Suche (aus [Duncan 1989])

Duncans und Humphreys Theorie zur visuellen Aufmerksamkeit besteht grundsätzlich aus drei Komponenten. Die erste Stufe erzeugt eine interne Repräsentation der vorhandenen Szene in verschiedenen Auflösungen. Die Zeit, die für das Formen der Repräsentation notwendig ist, ist unabhängig von der Komplexität der Szene. Die erste Stufe arbeitet also parallel. Eine Repräsentation entspricht einer kompletten Szenenbeschreibung, d. h. die Objekte in der Szene sind bereits erkannt, besitzen eine Bezeichnung. Allerdings ist der Person diese Szenenrepräsentation nicht bewußt.

In der zweiten Stufe findet ein Auswahlprozeß statt, der die strukturellen Einheiten der Szenenrepräsentation mit einem internen Template vergleicht. Die ausgewählte Information dient als Fokus weiterer Aktion. Das Template enthält die Information, die für das derzeitige Verhalten notwendig ist. Die strukturellen Einheiten, die eine gute Übereinstimmung erzielen, können ihre Gewichte erhöhen, wogegen die Gewichte der strukturellen Einheiten, die eine schlechte Übereinstimmung aufweisen, entsprechend verringert werden. Jede Änderung der Gewichte einer bestimmten strukturellen Einheit überträgt sich also auch auf die anderen Einheiten und zwar in dem Maße, in dem sie wahrnehmungsmäßig mit dieser gekoppelt sind (‘Spreading Suppression’).

Die ausgewählte Information wird in der letzten Stufe in einem visuellen Kurzzeitgedächtnis, dem sogenannten VSTM (‘*V*isual *S*hort *T*erm *M*emory’), gespeichert. Der VSTM macht die ausgewählte Information bewußt. Die Kapazität des VSTM ist allerdings begrenzt, so daß die verschiedenen strukturellen Einheiten der Szenenrepräsentation um den Zutritt zum VSTM konkurrieren müssen.

Mit Hilfe der beschriebenen Theorie lassen sich nun die experimentellen Ergebnisse erklären: Das Vergrößern der T-N-Ähnlichkeit erschwert die Suche, weil die Gewichte der Non-Targets durch den relativ guten Match mit dem Target-Template entsprechend hoch sind. Es konkurrieren also sehr viele strukturelle Einheiten um den Zugang zum VSTM. Das Verringern der N-N-Ähnlichkeit erhöht die Suchzeit, weil die Gewichte der Non-Targets relativ stark entkoppelt sind. Der schlechte Match einer strukturellen Einheit führt also nicht automatisch zur Unterdrückung der anderen Non-Targets, weil sich diese nicht mit dem untersuchten Non-Target wahrnehmungsmäßig gruppieren lassen.

An der Erzeugung des Templates sind nicht nur die im aktuellen Display vorhandenen Targets beteiligt, sondern alle vom Versuchsleiter genannten Targets. Ein Target-Template muß die Merkmale oder Merkmalskombinationen aufweisen, die die Targets von den Non-Targets unterscheiden. Je heterogener also die Targets und Non-Targets sind, um so komplexer ist das Template aufgebaut. Komplexe Templates aber erhöhen die Suchzeit. Zum einen benötigt der Match mit einem komplexen Template mehr Zeit und zum anderen erschwert er die Auswahl des Targets, weil das Template auch Merkmale mit den Non-Targets teilt.

In [Duncan 1992] ergänzen Duncan und Humphreys ihre Theorie durch die Unterscheidung zweier Ähnlichkeitsarten:

- ‘Interalternative Similarity’: Ähnlichkeit zwischen dem als Kategorie betrachteten Target-Template und den Non-Targets
- ‘Within-display Similarity’: Ähnlichkeit zwischen den Elementen im Display; diese kann Gruppierungseffekte auslösen

Die Reaktionszeiten in ‘Visual Search’-Experimenten hängen nach dieser modifizierten Theorie nicht mehr alleine von der Suche mit einem Template ab, sondern auch von der lokalen Detektion von Unstimmigkeiten. Hierdurch wird dem Einfluß der ‘bottom up’-gesteuerten Aufmerksamkeit auf die visuelle Suche eine höhere Bedeutung zugemessen als in der ursprünglichen ‘Stimulus Similarity Theory’. Hervorgerufen werden die Gruppierungseffekte nach den Untersuchungen von Humphreys et al. [Humphreys 1996] durch Flächeneigenschaften, wie z. B. gleiche Farbe, und Formeigenschaften, wie z. B. kollineare Kanten oder Einschlüsse.

Treisman untersuchte in [Treisman 1991] ebenfalls den Einfluß der ‘Target-Distractor Similarity’ und der ‘Distractor-Distractor Similarity’ auf die visuelle Suche. Sie konnte zwar ebenfalls einen meßbaren Effekt feststellen, weist ihm allerdings die gleiche Bedeutung wie etwa den Gruppierungseffekten zu. Treisman konnte zum einen zeigen, daß es auch dann noch einen meßbaren Unterschied zwischen Kombinations- und Merkmalssuche gibt, wenn man die Ähnlichkeitsbedingungen in beiden Experimenten angleicht, und zum anderen, daß der Einfluß der Ähnlichkeit abhängt von der An- bzw. Abwesenheit eines einzigartigen Merkmals, also von der Verwendung von Standardwerten (z. B. Blau oder Rot) gegenüber Nicht-Standardwerten (z. B. Violett).

Betrachtet man die beschriebenen psychophysischen Aufmerksamkeitsmodelle, so lassen sich einige Gemeinsamkeiten feststellen: Alle Modelle beinhalten die parallele und automatische Detektion grundlegender Merkmale. Diese werden kategorisiert und z. B. auf eine begrenzte Anzahl retinotoper Karten verteilt. Kategorisierungseffekte (Verwendung von Standardwerten gegenüber Nicht-Standardwerten) und Gruppierungseffekte (lokale Ähnlichkeiten unter den Merkmalen) können die Effizienz der Target-Suche beeinflussen. Zusätzlich nehmen ‘Top down’-Informationen in Form von Target-Beschreibungen (Templates) Einfluß auf die Antwortzeiten in ‘Visual Search’-Experimenten. Vermutlich haben also sowohl ‘Bottom up’- als auch ‘Top down’-Mechanismen Einfluß auf die Target-Suche. So scheinen sich auch die verschiedenen Versionen der ‘Feature Integration Theory’ und der ‘Stimulus Similarity Theory’ in dieser Frage immer weiter anzunähern (vgl. [Treisman 1980] vs. [Duncan 1984] und [Treisman 1992] vs. [Duncan 1992]). Strittig ist dagegen die Frage, ob es sich bei der Merkmals- und der Kombinationssuche um zwei prinzipiell identische Mechanismen handelt oder nicht. Hier sind noch weitere Untersuchungen notwendig.

Inwieweit die psychophysischen Modelle den Entwurf der Aufmerksamkeitssteuerung in NAVIS beeinflußt haben, wird in Kapitel 6 diskutiert.

3.3 Neurowissenschaftliche Ergebnisse

Neurowissenschaftliche Erkenntnisse haben großen Einfluß auf die Entwicklungen im Bereich des Computersehens. Interessante Eigenschaften des menschlichen Gehirns sind dessen Modu-

larität, Redundanz sowie parallele und gleichzeitig hierarchische Organisation. Besonders die Verteiltheit spezieller Funktionen zeigt sich auch in den neurophysiologischen Untersuchungen zur visuellen Aufmerksamkeit und den Augenbewegungen.

3.3.1 Anatomie des menschlichen visuellen Systems

Bevor im nächsten Abschnitt die Gehirnregionen beschrieben werden, denen man eine Beteiligung an der Bildung oder Verschiebung der Aufmerksamkeit oder an der Kontrolle der Augenbewegungen zuspricht, soll an dieser Stelle zunächst ein Überblick über die Anatomie des menschlichen visuellen Systems gegeben werden. Es werden an dieser Stelle nur die wesentlichsten Regionen und deren Verbindungen beschrieben, soweit dies zum Verständnis der neurophysiologischen Ergebnisse zur visuellen Aufmerksamkeit notwendig ist. Ausführlichere Beschreibungen finden sich in der Literatur, z. B. in [Kandel 1991] und [Kolb 1996].

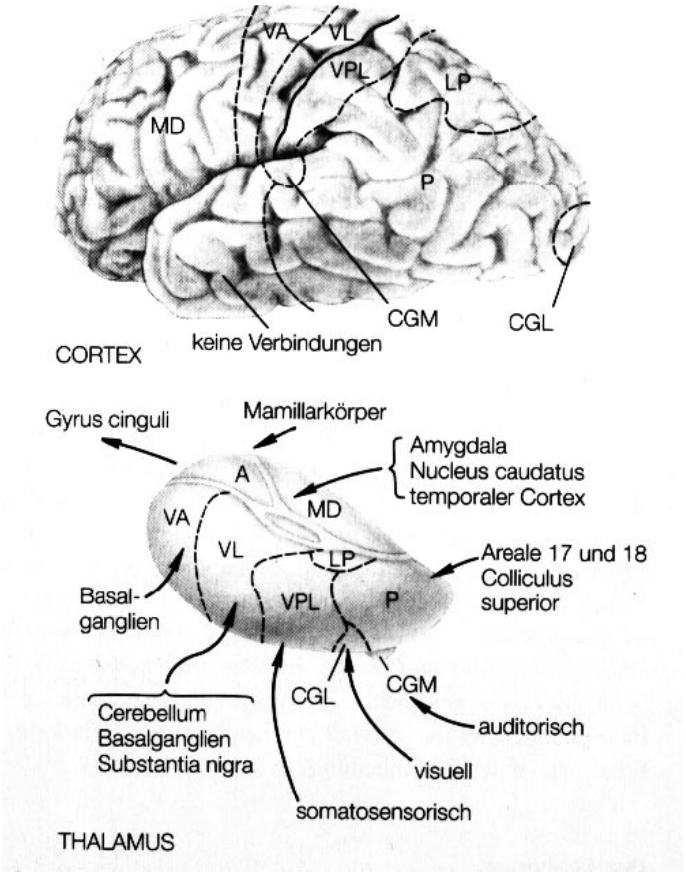
Ich verwende im folgenden die heute übliche Nomenklatur zur Beschreibung der verschiedenen Gehirnregionen: Dies ist zum einen die hierarchische Bezeichnung der visuellen Hirnrindenareale mit V1 bis V5 und die Definition der Areale über ihre Lage, z. B. medio-temporaler Cortex (MT) und inferior-temporaler Cortex (IT). Diese Systematik deckt sich nur selten mit der älteren Einteilung in Areae nach Brodman. So entspricht etwa die primäre Sehrinde V1 der Area 17, V2, V3 und V4 liegen in Area 18, V5 bzw. MT liegen in Area 19 und IT in Areae 20 und 21. Zur Festlegung räumlicher Beziehungen werden üblicherweise folgende Begriffe verwendet: *superior* (oben), *inferior* (unten), *lateral* oder *parietal* (seitlich), *median* (in der Mitte), *ventral* (zur Bauchseite hin), *dorsal* (zur Rückenseite hin), *anterior* (vorn) und *posterior* (hinten). Eine Struktur kann also in einer der genannten Relationen zu einer anderen liegen. Befinden sich zwei Strukturen auf derselben Körperseite, so nennt man diese Anordnung *ipsilateral*, andernfalls *kontralateral*. Sind die Strukturen auf beiden Seiten vorhanden, bezeichnet man diese als *bilateral*. Und schließlich kann sich ein Areal *proximal* (nahe) oder *distal* (entfernt) von einem anderen Areal befinden.

Visuelle Stimuli erreichen das menschliche Auge in Form von Licht. Dieses Licht wird durch die Zellschichten der Retina in Aktionspotentialfolgen gewandelt, die die visuelle Information repräsentieren. Die visuelle Information wird über den Sehnerv zum Corpus geniculatum laterale (CGL, seitlicher Kniehöcker, engl. Abkürzung LGN) und den Colliculi superiores[1] (SC, oberes Vierhügelpaar) transportiert. Der retino-geniculate Pfad setzt sich im Cortex (Neocortex, Großhirnrinde) fort, der 90% der visuellen Information verarbeitet (klassische Sehbahn, geniculo-striatales System). Das Verhältnis von ca. 130 Millionen Rezeptoren der Retina zu 1 Million Axone des Sehnervs zum CGL, der wiederum 200 Millionen Zellen des visuellen Cortex V1 versorgt, erzwingt eine effiziente Übertragung der Signale.

[1] Häufig wird statt des Plurals auch der Singular Colliculus superior verwendet.

Grundsätzlich unterteilt man sowohl den retino-geniculaten als auch den colliculären Pfad in einen magnozellulären (M-) und einen parvozellulären (P-) Strang. In der Retina unterscheiden sich die M- und P-Zellen bezüglich ihres Auflösungsvermögens, ihres zeitlichen Antwortverhaltens und ihrer räumlichen Verteilung. M-Zellen besitzen große rezeptive Felder und daher gegenüber den P-Zellen eine schlechtere örtliche Auflösung. M-Zellen reagieren besonders auf sich zeitlich ändernde Reize, sie zeigen daher schnelle, nur kurz anhaltende Reizmuster (*phasisches Verhalten*). Sie sind überwiegend in der peripheren Retina angesiedelt. P-Zellen dagegen reagieren sehr viel langsamer und anhaltend nur auf stationäre Stimuli (*tonisches Verhalten*). Sie sind mit ihrer guten räumlichen Auflösung vorwiegend in der Fovea, d. h. dem Punkt des schärfsten Sehens in der Retina, positioniert.

Sowohl der retino-geniculate als auch der colliculäre Pfad weisen Verbindungen zum Pulvinar (P) auf. Der Pulvinar ist ebenso wie der CGL ein *Nucleus* (Kern) des Thalamus, einer relativ großen anatomischen Struktur im Zentrum des Gehirns, der Verbindungen zum visuellen Cortex besitzt. Der Pulvinar unterscheidet sich vom CGL besonders dadurch, daß er keine direkten Verbindungen mit dem optischen Nerv aufweist. Er sei hier erwähnt, weil ihm eine wesentliche Beteiligung an der Aufmerksamkeit und den Augenbewegungen zugesprochen wird. Abbildung 3.6 zeigt die verschiedenen Nuclei des Thalamus und ihre Verbindungen zum Cortex.



A .. Nucleus anterior
MD .. Nucleus medialis dorsalis
VA .. Nucleus ventralis anterior
VL .. Nucleus ventralis lateralis
LP .. Nucleus lateralis posterior
VPL .. Nucleus ventralis postero-lateralis
P .. Pulvinar
CGL .. Corpus geniculatum laterale
CGM .. Corpus geniculatum mediale

Abbildung 3.6: Der Thalamus und seine Verbindungen zum Cortex (aus [Kolb 1996])

Der Cortex nimmt den größten Teil des Endhirns ein. Er umfaßt ca. 80% der gesamten Gehirnmasse, ist dabei aber nur 1,5 bis 3 mm dick. Seine stark gefaltete Oberfläche besteht aus *Fissuren* (Falten), *Sulci* (flachere Furchen) und *Gyri* (Windungen). Er besteht aus mehreren hierarchischen Schichten, die auch rückwärtige Verbindungen eingehen. Die einzelnen Schichten werden im folgenden kurz beschrieben. Eine Übersicht bietet Abbildung 3.7.

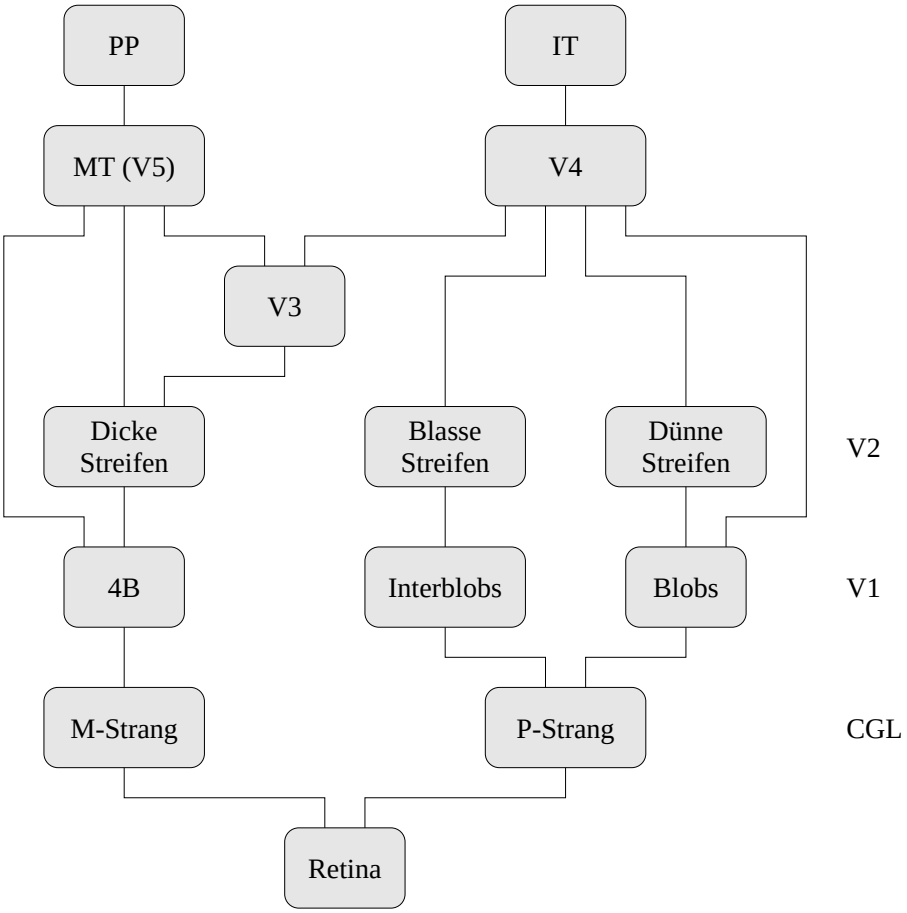


Abbildung 3.7: Neuroanatomische Verarbeitungsstränge (stark vereinfacht)

Die erste Schicht des visuellen Cortex wird als striärer Cortex, Streifenkortex, Area 17 oder V1 bezeichnet. Sie läßt sich wiederum in drei funktionale Einheiten unterteilen: Zellen, die mit der Bewegungsdetektion assoziiert werden, Zellen in den als Blobs bezeichneten Neuronennestern, die mit der Verarbeitung von Farbinformationen assoziiert werden, und Zellen in den Interblobs, die Orientierungs- und Tiefeninformationen verarbeiten ([Livingstone 1987], [DeYoe 1988]). Eine weitere wichtige Eigenschaft, die V1 zugesprochen wird, ist die Rekombination von Luminanz- und Farbkontrastinformationen, denn die Formerkennung als solche ist farbenblind [Davidoff 1995].

Neben der parallelen Unterteilung des V1 findet man in der Literatur eine hierarchische Unterteilung in einfache und komplexe Zellen [Hubel 1982]. Einfache Zellen reagieren optimal auf

Kanten oder Linien einer bestimmten Orientierung, ihr rezeptives Feld kann durch Gaborfunktionen beschrieben werden. Mehrere einfache Zellen projizieren auf eine komplexe Zelle. Auch komplexe Zellen sind orientierungsspezifisch. Der grundsätzliche Unterschied zwischen einfachen und komplexen Zellen besteht darin, daß bei einfachen Zellen eine optimal orientierte Linie nur in einer extrem kleinen Region des rezeptiven Feldes eine Reaktion auslöst; dagegen ruft bei einer komplexen Zelle eine richtig orientierte Linie überall im rezeptiven Feld Antworten hervor. Allerdings ist eine Reaktion häufig nur meßbar, wenn sich der Stimulus über das rezeptive Feld hinweg bewegt. Viele komplexe Zellen reagieren dabei auf eine Bewegungsrichtung besser als auf die entgegengesetzte. Einige einfache und komplexe Zellen reagieren nicht nur optimal auf Stimuli einer bestimmten Orientierung, sondern auch einer bestimmten Länge. Solche Zellen werden als *endinhibiert* bezeichnet.

Die extrastriären Schichten des visuellen Cortex sind V2, V3, V4, der inferior-temporale Cortex (IT), das medio-temporale Areal (MT oder V5) und der posterior-parietale Cortex (PP). V2 wird als die Verallgemeinerungsstufe des V1 angesehen. Das heißt, in V2 werden unterbrochene Konturen und andere Diskontinuitäten geschlossen (siehe z. B. [von der Heydt 1984]). Über die spezifischen Eigenschaften des V3 ist dagegen wenig bekannt. V4 wiederum ist recht gut untersucht. V4 scheint wesentlich an der Verarbeitung der Farbinformationen beteiligt zu sein und insbesondere einen Beitrag zur Farbkonstanz zu leisten [Zeki 1993]. Der IT stellt vermutlich die letzte Stufe der Objekterkennung dar. Die retinotopische Organisation der V-Areale ist hier aufgehoben. Viele Zellen im IT weisen rezeptive Felder auf, die das gesamte visuelle Feld abdecken, so daß die für die Objekterkennung notwendige Positionsinvarianz gewährleistet ist. Chelazzi [Chelazzi 1993] vermutet dagegen, daß IT an der Hemmung von Distraktoren und damit an der Selektion eines Targets beteiligt ist. MT ist besonders an der Analyse von Bewegungen beteiligt. Der PP nimmt eine wichtige Rolle bei der Lokalisierung einzelner Objekte und der Wahrnehmung räumlicher Konfigurationen ein [Goodale 1992].

Die Zellen aus V1 und V2, die Signale aus dem P-Strang weiterverarbeiten, projizieren überwiegend auf V4, dessen Zellen wiederum mehrheitlich auf den für die Objekterkennung relevanten IT projizieren. Die Zellen, die Signale aus dem M-Strang weiterverarbeiten, konzentrieren sich dagegen im MT, dessen Zellen wiederum auf den an der Lokalisierung beteiligten PP projizieren. In erster Näherung stellt diese Verzweigung eine anatomischen Trennung in einen sogenannten What-Pfad und einen Where-Pfad dar (siehe [Ungerleider 1982], [Mishkin 1983], [Milner 1993]).

Neben dem colliculären und dem geniculo-striatalen System existieren eine Reihe weiterer visueller Subsysteme, die z. B. für Kopf- und Augenbewegungen verantwortlich sind. In Abbildung 3.8 sind die verschiedenen visuellen Subsysteme des menschlichen Gehirns übersichtlich dargestellt. Inwieweit diese Subsysteme zur Bildung attentiver Blickwechsel beitragen, soll im folgenden Abschnitt erläutert werden. Zur Vertiefung in die Neuroanatomie sei hier noch einmal auf die Spezialliteratur, z. B. [Kandel 1991], verwiesen.

visuelles Subsystem	postulierte Funktionen
1. Nucleus supraopticus	Kontrolle tageszeitlicher Rhythmen (Schlaf, Essen, etc.) in Abhängigkeit vom Tag-Nacht-Zyklus
2. Prätectum	Veränderungen der Pupillengröße in Abhängigkeit von Helligkeitsänderungen
3. Colliculus superior	Ausrichtung des Kopfes, insbesondere auf Objekte im peripheren Gesichtsfeld
4. Epiphyse (Zirbeldrüse)	zirkadiane Dauerrhythmen
5. Nucleus opticus accessorius	Augenbewegung zum Ausgleich von Kopfbewegungen
6. visueller Cortex	Mustererkennung, Tiefenwahrnehmung, Farbensehen, Bewegungssehen
7. frontale Augenfelder	Willkürbewegungen der Augen

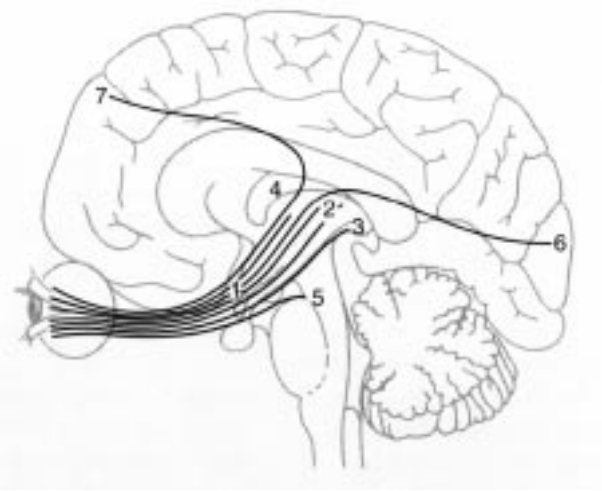


Abbildung 3.8: Schematische Darstellung visueller Subsysteme (aus [Kolb 1996])

3.3.2 Neurobiologische Aspekte der Aufmerksamkeit

Die Ergebnisse neurobiologischer Untersuchungen basieren auf verschiedenen Meßtechniken. Die elektrische Aktivität des Gehirns läßt sich mit Hilfe der *Elektroencephalographie* messen. Einzelne Messungen, die sogenannten *Elektroencephalogramme* (EEGs), eignen sich besonders, um verschiedene geistige Zustände, wie z. B. Erregung, Schlaf und Koma, aufzuzeichnen. Kurzzeitige Veränderungen im EEG lassen sich durch die Präsentation sensorischer Stimuli, z. B. Geräusche, verursachen. Die auf diese Weise erzeugten Reaktionen bezeichnet man als *ereigniskorrelierte Potentiale* (ERPs, engl. 'Event Related Potentials'). Durch die Einführung kleiner isolierter Drähte oder mit leitender Salzlösung gefüllter Pipetten in das Gehirn lassen sich Veränderungen des elektrischen Potentials einzelner Nervenzellen messen. *Einzelzellableitungen* liefern immer nur die Antwort des Gesamtsystems von den Photorezeptoren bis zum Meßpunkt. Sie werden an partiell narkotisierten Versuchstieren (vorrangig Rhesusaffen) durchgeführt. Mit Hilfe moderner bildgebender Verfahren, wie der *Computertomographie*

(CT) sowie der *Positronemissionstomographie* (PET) und der *Magnetresonanztomographie* (Kernspintomographie, MRI, engl. ‘Magnetic Resonance Imaging’) können das Gehirngewebe bzw. Gehirnaktivitäten visualisiert werden. Die PET mißt die Positronenemission eines zuvor injizierten schnell zerfallenden Radionukleids und zeichnet so die lokalen Veränderungen des cerebralen Blutflusses auf. Die MRI mißt die Dichte von Wasserstoffatomen, die unter Einfluß eines Magnetfelds und der Einstrahlung von HF-Impulsen eigene Radiowellen emittieren. Ein weiteres bildgebendes Meßverfahren ist die *funktionale MRI* (fMRI), die z. B. die funktionell induzierten Veränderungen der Sauerstoffsättigung des Blutes mißt.

Ein Beispiel für Einzelzellableitungen sind die Arbeiten von Moran und Desimone ([Moran 1985], [Desimone 1990], [Desimone 1992]). Sie stellten Aufmerksamkeitseffekte an Neuronen des V4 und IT, aber nicht an Neuronen des V1 und V2 fest. Im allgemeinen muß jedes Experiment, mit dem man zeigen will, daß Antworten eines Neurons vom Grad der Aufmerksamkeit abhängen, ein wichtiges Kriterium erfüllen: Derselbe Reiz muß das Neuron zu einem bestimmten Zeitpunkt aktivieren, zu einem anderen aber nicht. Genau diesen Effekt konnten Moran und Desimone an Zellen des V4 feststellen. Das Antwortverhalten eines V4-Neurons wird maßgeblich von der Lage des FOA innerhalb seines rezeptiven Feldes beeinflußt. Sind innerhalb dieses Feldes ein effektiver und ein ineffektiver Stimulus präsent, also ein Stimulus, für den die Zelle empfindlich ist, und einer, für den die Zelle nicht empfindlich ist, und die Aufmerksamkeit des Versuchsaffen wird auf den effektiven Stimulus gelenkt, so zeigt die Zelle eine deutliche Reaktion. Richtet man die Aufmerksamkeit des Tieres dagegen auf den ineffektiven Stimulus, so nimmt die Reaktion der Zelle deutlich ab, obwohl der effektive Stimulus nachwievor im rezeptiven Feld der Zelle liegt. Das beschriebene Ergebnis kann allerdings nur erzielt werden, wenn sowohl ein effektiver als auch ein ineffektiver Stimulus innerhalb des rezeptiven Feldes des Neurons liegt. Moran und Desimone folgerten hieraus, daß Aufmerksamkeit auch zur Auflösung von Mehrdeutigkeiten dient. Desweiteren zeigten V4-Neurone in den Experimenten erhöhte Reaktion, in denen eine hohe Ähnlichkeit zwischen dem Target und den Distraktoren bestand. Dieser Effekt wurde für die Merkmale Farbe und Orientierung (jeweils mit dem gleichen Ergebnis) untersucht. Das Maß an Aufmerksamkeit scheint also mit der Schwierigkeit der Aufgabe zu steigen.

Desimone schlägt in [Desimone 1992] zwei Modelle vor, mit denen die genannten Effekte erzielt werden könnten. Der erste Mechanismus ist das ‘Input Gating’-Modell, in dem eine zusätzliche Neuronengruppe den Input der V4-Neurone aus der Schicht V2 steuert. Der Aufmerksamkeitsmechanismus erregt ein Neuron aus dieser Gruppe, das wiederum alle Verbindungen der V2-Neurone zu den V4-Neuronen hemmt, die außerhalb seines eigenen rezeptiven Feldes liegen (vgl. Abbildung 3.9a). Das ‘Cell Gating’-Modell basiert auf einer Form von lateraler Hemmung zwischen Zellen, deren rezeptive Felder überlappen und die für verschiedene Reize empfindlich sind. In der Abwesenheit von Aufmerksamkeit kann sich kein Zelltyp gegen die anderen Zelltypen durchsetzen, so daß jede Zelle auf die für sie spezifischen Reizmuster antwortet. Erst der Aufmerksamkeitsmechanismus beeinflußt die Konkurrenz zugunsten eines bestimmten Zelltyps, indem er die Aktivität der Zellen erhöht, die innerhalb des Aufmerksam-

keitsbereichs liegen (vgl. Abbildung 3.9b, vier Zellen innerhalb des Aufmerksamkeitsbereichs werden durch den roten Stimulus aktiviert, nur eine Zelle durch den grünen Stimulus).

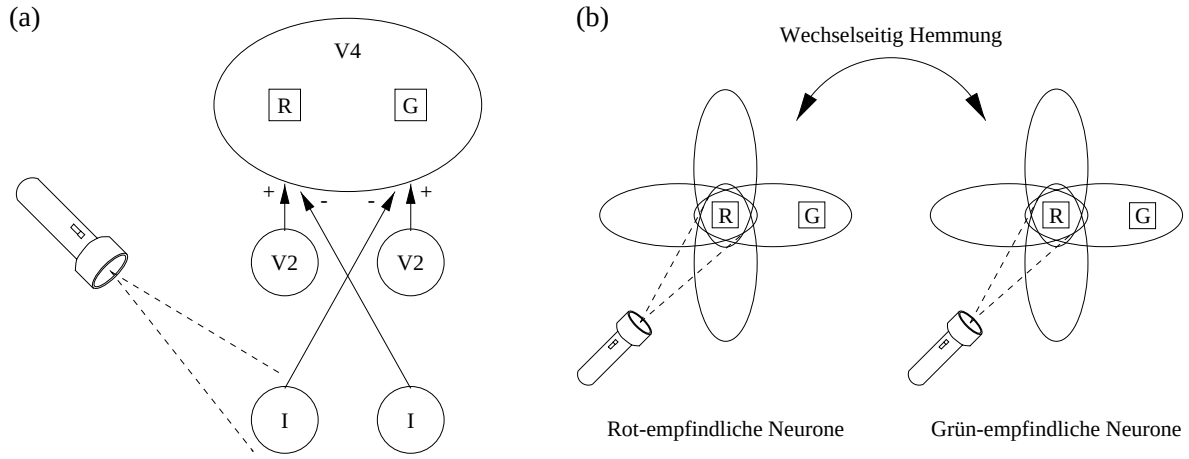


Abbildung 3.9: (a) 'Input Gating'-Modell, (b) 'Cell Gating'-Modell (aus [Desimone 1992])

Desimones Experimente bestätigen zudem die Bedeutung der Vertrautheit für die visuelle Aufmerksamkeit: IT-Neurone reagieren mit sehr viel höherer Aktivität auf neue, unerwartete oder längere Zeit nicht mehr präsentierte Stimuli als auf bereits bekannte oder erwartete Stimuli. Neuere Untersuchungen von Motter [Motter 1993] weisen aufmerksamkeitsspezifische Zellen nicht nur im V4, sondern auch im V1 und V2 nach. Die Sensitivität vieler dieser Zellen hängt nicht nur von der Position ab, auf die die Aufmerksamkeit gerichtet wird, sondern auch von der Anwesenheit konkurrierender Stimuli.

Wie wir bereits aus den Untersuchungen von Eriksen (siehe Abschnitt 3.2.2) wissen, dient die visuelle Aufmerksamkeit aber wohl nicht nur zur Selektion bestimmter Stimuli, sondern auch zur Zuweisung mentaler Ressourcen entsprechend des Schwierigkeitsgrades der gestellten Aufgabe. Eine derartige Beobachtung machten auch Spitzer et al. [Spitzer 1988]. Sie konnten zeigen, daß Zellen im V4 ihr Aktivitätsmuster an das Maß an Anstrengung, das zur Lösung eines visuellen Problems notwendig ist, anpassen können. So feuerten die untersuchten V4-Zellen in dem schwierigeren von zwei Experimenten mit 20% höherer Frequenz. Gleichzeitig erhöhten sie deutlich ihre Orientierungsselektivität, die zur Lösung der gestellten Aufgabe notwendig war.

Faßt man die beschriebenen Untersuchungsergebnisse zusammen, so läßt sich hier ein Widerspruch entdecken. Einerseits unterdrücken V4-Zellen die Antworten auf irrelevante Stimuli, wirken also hemmend, um einen bestimmten Stimulus besser selektieren zu können. Andererseits verstärken sie die Antworten auf relevante Reize und erhöhen ihre Selektivität, wenn der Schwierigkeitsgrad der Aufgabe steigt. Dieses Ergebnis legt nach [Kolb 1996] die Vermutung nahe, daß sich an V4-Zellen zwar Aufmerksamkeitseffekte messen lassen, aber andere neuro-

nale Strukturen diese Effekte hervorrufen, indem sie auf die V4-Zellen entsprechend einwirken.

Eine solche Struktur könnte der Thalamus sein (siehe z. B. [LaBerge 1990, 1992]), der häufig physiologisch in zwei Teile unterteilt wird, den dorsalen und den ventralen Thalamus. Einige Autoren bezeichnen auch den dorsalen Teil einfacher als Thalamus und den ventralen Teil als retikulären Komplex. Nahezu die gesamte Information, die den Cortex erreicht, passiert zunächst den Thalamus, der daher auch als eine Art Gateway betrachtet wird [Crick 1984]. Projektionen in die umgekehrte Richtung, also vom Cortex weg, müssen nicht notwendigerweise über den Thalamus gehen. Der retikuläre Komplex ist eine dünne Neuronenschicht, die den (dorsalen) Thalamus teilweise umgibt. Diese Schicht wird deshalb auch als Guardian für das Gateway Thalamus bezeichnet. Eine wichtige Aufgabe des retikulären Komplexes könnte die Begrenzung der Anzahl für den Thalamus relevanter Objekte sein. Die verschiedenen Nuclei des Thalamus scheinen dagegen eine wichtige Rolle bei der Detektion auffälliger Regionen und der Bildung und Verschiebung des FOA zu spielen. So soll z. B. der Pulvinar ebenfalls an der Ausblendung irrelevanter Stimuli beteiligt sein [Desimone 1990]. Der retikuläre Nucleus (RN) dagegen scheint die Bildung und Verschiebung des FOA durchzuführen. Die Aktivität der ohnehin schon hoch aktiven Neurone wird im RN verstärkt, wogegen gering gereizte Neurone gehemmt werden. Dieser Effekt läßt eine Art FOA entstehen, der die Projektionen zum visuellen Cortex bestimmt. Die am FOA beteiligten Neurone des RN gehen nach ihrer kurz anhaltenden Erregung (Bursts) in die Hemmung über und können erst nach einer Pause wieder feuern. Diese Pause kann von einer anderen Gruppe genutzt werden, um nun ihrerseits den FOA zu bestimmen. Auf diese Weise könnte die Verschiebung des FOA entstehen. Ein detaillierteres Modell des Thalamus in Verbindung mit visueller Aufmerksamkeit kann in den Arbeiten von LaBerge [LaBerge 1990, 1992] gefunden werden. Zusammenfassend läßt sich feststellen, daß der Thalamus wohl weniger an der Selektion einer auffälligen Region oder eines auffälligen Objekts beteiligt ist als an dem Filtermechanismus, der diese Region oder dieses Objekt gegenüber seiner Umgebung hervorhebt.

Klinische Studien der Gruppe um Posner [Posner 1987, 1988] an zerstörten posterior-parietalen Cortexarealen legen die Vermutung nahe, daß der PP an Aufmerksamkeitsverschiebungen beteiligt ist. Patienten mit der genannten Einschränkung konnten ihren FOA zwar ebenfalls zu auffälligen Regionen lenken, aber das Lösen vom jeweils aktuellen ‘Locus of Attention’ (LOA) fiel ihnen schwer. Posner war es auch, der die Aufmerksamkeitverschiebungen in drei Phasen aufteilte: Eine erste Phase, in der der FOA vom aktuellen LOA zu lösen ist, eine zweite Phase, in der der FOA zum nächsten LOA verschoben wird, und eine dritte Phase, in der der FOA mit dem neuen LOA verbunden werden muß. Posner leitete aus dieser Unterteilung ab, daß der PP an der ersten Phase der Aufmerksamkeitsverschiebung, dem Lösen vom aktuellen LOA, beteiligt ist. Das Mittelhirn und insbesondere der SC könnten dagegen die zweite Phase einer Aufmerksamkeitsverschiebung steuern [Rafal 1988]. Und die dritte Phase eines Aufmerksamkeitswechsels könnte nach Posner schließlich vom Pulvinar durchgeführt werden.

Corbetta et al. [Corbetta 1993] führten ebenfalls Untersuchungen zur Aufmerksamkeit am PP durch. In ihren PET-Studien präsentierten sie den Testpersonen in zwei getrennten Versuchsreihen physikalisch identische Reize, wiesen ihnen aber jeweils eine unterschiedliche Aufgabe zu. Sie konnten zeigen, daß die Aktivitäten im PP mit unterschiedlichen Anforderungen an die Aufmerksamkeit variieren. Insofern bestätigen diese Untersuchungen die von Moran, Desimone und Spitzer aus Einzelzellableitungen im V4 gewonnenen Ergebnisse. Auch im PP sind Aufmerksamkeitseffekte, die durch willentliche Steuerung entstehen, meßbar. Wie diese Effekte entstehen, bleibt aber unklar.

Corbetta et al. [Corbetta 1991] führten auch eine PET-Studie durch, die den Einfluß der Berücksichtigung verschiedener Stimuluseigenschaften zeigen sollte. Den Testpersonen wurden nacheinander zwei Displays mit jeweils homogenen Stimuli präsentiert. Die Stimuli im zweiten Display konnten sich in Form, Farbe und/oder Bewegungsgeschwindigkeit von denen des ersten Displays unterscheiden. In einer ersten Meßreihe mußten die Testpersonen dann entscheiden, ob sich beide Displays in einer bestimmten vorgegebenen Eigenschaft unterschieden. In einer zweiten Versuchsreihe mußten die Personen dagegen lediglich beantworten, ob sich die Displays überhaupt in irgendeiner Eigenschaft unterschieden. Nach Auffassung der Autoren, benötigt die erste Versuchsform eine eher gerichtete Aufmerksamkeit, wogegen die zweite Form mehr Gedächtnisleistung erfordert. PET-Untersuchungen zeigten, daß bei der Durchführung der ersten Aufgabe je nach relevantem Merkmal unterschiedliche Regionen aktiviert wurden. So wurden z. B. V4-Zellen aktiv, wenn die Farbe das relevante Merkmal war, und Zellen in V3 und IT, wenn die Form für die richtige Lösung der Aufgabe entscheidend war. Weiterhin wurden folgende Strukturen aktiviert: die Insula, der posteriore Thalamus (vermutlich Pulvinar), Colliculus superior und orbitaler Frontalcortex. Die zweite Aufgabe führte zu Aktivitäten des Gyrus cinguli anterior (cingulärer Cortex) und des dorsolateralen präfrontalen Cortex. Bemerkenswert ist wiederum, daß die Reize in beiden Versuchsreihen identisch waren. Dies zeigt den großen ‘Top down’-Einfluß auf die visuelle Aufmerksamkeit.

Neben den genannten visuellen Zentren, denen eine Bedeutung für die Aufmerksamkeit zugesprochen wird, existieren auch visuelle und vor allem motorische Zentren, die für Augenbewegungen sorgen. Dies sind z. B. die Colliculi superiores (SC), die Verbindungen zu Zentren besitzen, die an der Kontrolle der Augen- und Kopfbewegungen beteiligt sind [Sparks 1986]. Zudem sind im SC Zellen gefunden worden, deren Aktivität sich kurz vor der Ausführung einer Sakkade erhöht, und von denen daher vermutet wird, daß sie eine Art Warnsystem darstellen [Breitmeyer 1986]. Weiterhin müssen an dieser Stelle noch die frontalen Augenfelder[2] (FEFs, engl. ‘Frontal Eye Fields’) genannt werden, die direkt mit den okulomotorischen Zentren gekoppelt sind und deren Zellen erhöhte Aktivität vor der Ausführung von Augenbewegungen zeigen. Henik et al. [Henik 1994] konnten zeigen, daß die FEFs die Generierung voluntativer Sakkaden erleichtern und reflektorische Sakkaden zu exogenen Signalen hemmen. Etwa die Hälfte aller Neurone in den FEFs antworten auf visuelle Reize und die Hälfte dieser Neurone zeigt wiederum erhöhte Reaktion auf Stimuli, die das Ziel einer Sakkade sind. Anders als die

[2] Auch in der singularen Form gebräuchlich.

Neurone im PP zeigen sie aber keine erhöhte Reaktion auf Stimuli, auf die zwar Aufmerksamkeit gerichtet wird, aber keine Sakkade folgt. Eine zweite Klasse von Zellen in den FEFs feuert nicht nur vor der Ausführung von Sakkaden, die durch visuelle Reize initiiert werden, sondern auch vor solchen, die aus dem Gedächtnis abgerufen werden. Sie reagieren ebenfalls nicht auf Stimuli, die nicht das Ziel einer Sakkade sind. Im Gegensatz zu der ersteren Klasse projizieren sie auf den SC. Anders als die Zellen im SC, die während der Präparation jeder Sakkade feuern, feuern sie aber nur vor der Ausführung solcher Sakkaden, die für das Verhalten des Versuchstieres relevant sind. Zur Vertiefung in die Anatomie und die Funktionsweise der okulomotorischen Zentren sei wiederum die Spezialliteratur empfohlen, z. B. [Kandel 1991].

Aus den beschriebenen neurophysiologischen Untersuchungen ist unmittelbar ersichtlich, daß nicht etwa eine zentrale Gehirnregion existiert, die die visuelle Aufmerksamkeit steuert, sondern daß verschiedene Regionen interagierende Teilfunktionen übernehmen. Innerhalb dieser Regionen scheint zudem eine gewisse Redundanz zu bestehen, die zwar für neurobiologische Systeme typisch ist, aber die eindeutige Identifizierung der Funktionseinheiten erschwert. Cortikale Zellen, die in der menschlichen Sehbahn vorne plaziert sind, weisen kleine rezeptive Felder auf und antworten automatisch, wenn ihr bevorzugter Stimulus präsentiert wird. Entlang der Sehbahn steigt dann sowohl die Größe der rezeptiven Felder als auch die Empfindlichkeit der Zellen gegenüber attentiven Einflüssen. Zellen, die in der Sehbahn hinten plaziert sind, reagieren weniger effektiv auf die passive Präsentation eines Stimulus. Posner und Dehaene [Posner 1994] vermuten daher, daß attentive Effekte früh in der Sehbahn als Unterdrückung nicht relevanter Information erscheinen, da die Neurone in den ersten visuellen Schichten durch geeignete Stimuli auch in einer passiven Situation nahezu optimal stimuliert werden und daher nur wenig Spielraum für eine Erhöhung der Feuerungsrate auf das Target, aber ein großer Spielraum zur Unterdrückung von Non-Targets besteht. In späteren Schichten können dagegen relevante Informationen durch attentive Einflüsse erhöht werden, da diese Zellen nicht mehr automatisch auf bestimmte Stimuli reagieren.

Zusammenfassend läßt sich feststellen, daß ein sogenanntes posteriores attentives System, bestehend aus PP, Pulvinar und Colliculi superiores, weitestgehend verantwortlich ist für die Bildung und Verschiebung des FOA (gerichtete Aufmerksamkeit). Ein zweites System, das sogenannte anteriore attentive System, bestehend aus Gyrus cinguli anterior und den Basalganglien, erfüllt dagegen einen mehr exekutiven Zweck und steuert Gehirnregionen zur Ausführung komplexer kognitiver Aufgaben (Auswahl von Objekteigenschaften). So vermuten z. B. Posner und Dehaene [Posner 1994], daß hier die Definition eines Targets, z. B. durch Vorgabe einer Farbe und einer Orientierung, umgesetzt wird. Die Auswahl der Objekte selbst findet wahrscheinlich im IT statt, der mit der Objekterkennung assoziiert wird.

3.3.3 Organisation des attentiven Systems

In den vorangegangenen Abschnitten wurden die verschiedenen funktionellen Einheiten des menschlichen Gehirns beschrieben, denen eine Beteiligung an der visuellen Aufmerksamkeit zugesprochen wird. Gleichwohl die genaue Funktion der einzelnen Gehirnregionen und deren Interaktionen nicht bekannt sind, lassen sich auf der Basis der genannten Ergebnisse bereits Hypothesen über die neuronale Organisation des attentiven Systems aufstellen. Die hier vorgestellten Hypothesen gehen auf die Arbeiten von Milanese [Milanese 1993] und Goldberg et al. [Goldberg 1991] zurück.

Die cortikalen Regionen könnten in dem attentiven System in erster Linie zwei wichtige Funktionen übernehmen: die Erstellung von Merkmalskarten und die Selektion der auffälligsten Regionen. Die zweite Aufgabe ließe sich durch konkurrierende Neuronengruppen mit lateral hemmenden Verbindungen lösen. Die Wahl des nächsten FOA hängt allerdings nicht nur von der Aktivität in den Merkmalskarten, sondern auch von nachbarschaftlichen Beziehungen und der verfolgten Aufgabe ab. Für die Auswertung der räumlichen Anordnung verschiedener Objekte scheint der PP verantwortlich zu sein. Der SC nimmt wohl besonders dann Einfluß auf die Auswahl des nächsten FOA, wenn sich ein Objekt in der Szene zeitlich ändert, sich also z. B. in das Blickfeld des Betrachters bewegt.

Es ist denkbar, daß die genannten Gehirnregionen gleichzeitig für verschiedene Objekte im visuellen Feld Aufmerksamkeit allokieren möchten. Daher sind die "Anfragen" der verschiedenen funktionellen Einheiten in einer zentralen Struktur zu bündeln. Als die zentrale Struktur des attentiven System wird der Pulvinar angesehen, auf den viele cortikale Regionen projizieren. Der SC könnte neben seiner Funktion als eine Art Alarmsystem ebenfalls an dem Integrationsprozeß beteiligt sein. Die Organisation des attentiven Systems, die zur Lokalisierung eines Targets führt, ist in Abbildung 3.10 dargestellt.

Der Integrationsprozeß im Pulvinar und SC liefert die Position im visuellen Feld, auf die die Aufmerksamkeit als nächstes gerichtet werden soll. Hierbei kann die Aufmerksamkeit entweder verdeckt oder offen zu der neuen Position bewegt werden. Der SC mit seinen Verbindungen zum okulomotorischen System könnte für das Auslösen der Augenbewegungen verantwortlich sein, wogegen die Aufmerksamkeitsverschiebungen wohl eher durch die Projektionen des Pulvinars zu den verschiedenen Cortexarealen angestoßen werden (siehe Abbildung 3.11). Entsprechend Posners Unterteilung ist in der ersten Phase des Aufmerksamkeitswechsels die Aufmerksamkeit von dem derzeitigen LOA zu lösen. Diese Operation findet im PP statt. Der Pulvinar und der RN können dann, nachdem der neue LOA erreicht ist, die Aufmerksamkeit auf den neuen LOA konzentrieren.

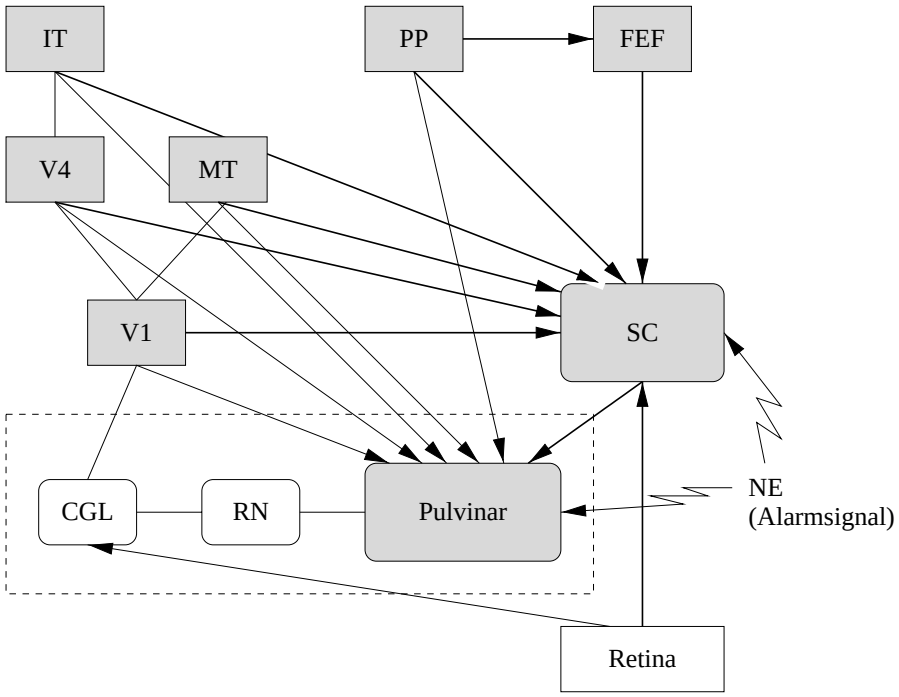


Abbildung 3.10: Darstellung der an der Lokalisierung eines Targets beteiligten funktionellen Strukturen (hier grau unterlegt, aus [Milanese 1993])

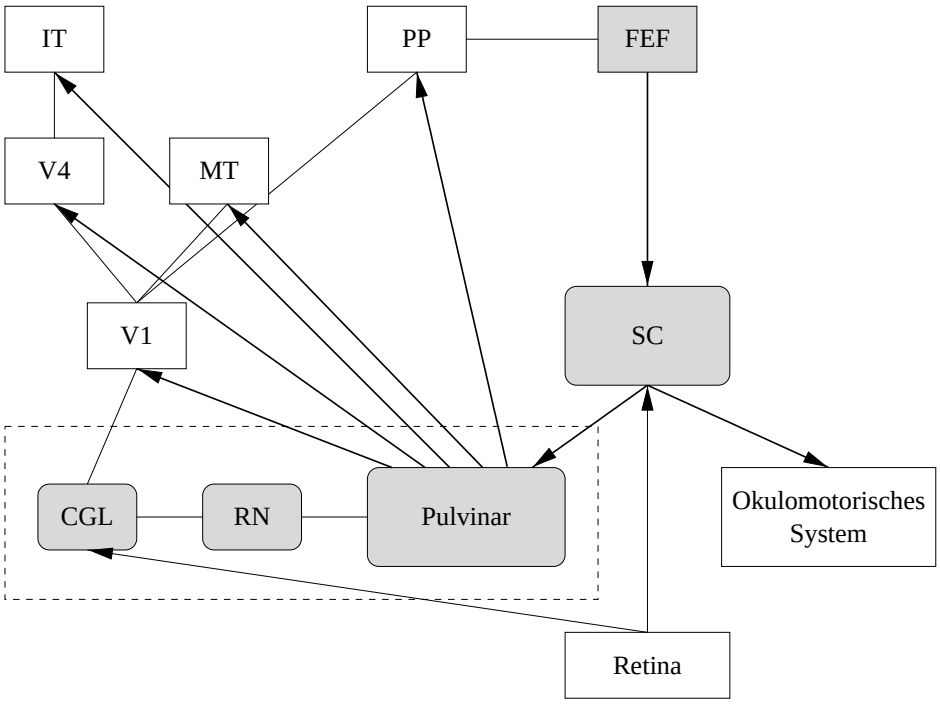


Abbildung 3.11: Darstellung der an Aufmerksamkeitsverschiebungen beteiligten funktionellen Strukturen (grau unterlegt, aus [Milanese 1993])

Goldberg et al. schlagen für offene Aufmerksamkeitsverschiebungen in Form von Sakkaden die in Abbildung 3.12 dargestellte Interaktion verschiedener okulomotorischer Strukturen vor. Der Gehirnstamm, in dem die Sakkaden generiert werden, erhält ein Motor-Kommando vom SC, auf den der PP und das FEF exzitatorisch und die Substantia nigra (schwarze Substanz, Nucleus des Mittelhirns) inhibitorisch projizieren. Der inhibitorische Einfluß der Substantia nigra auf den SC kann allerdings durch den Nucleus caudatus (Schweifkern, gehört zu den Basalganglien) unterdrückt werden, der wiederum Motor-Kommandos vom FEF erhält. Das FEF kontrolliert den SC also auf zwei Wegen: direkt sowie über den Nucleus caudatus und die Substantia nigra. Aktivitäten im FEF zur Präparation einer Sakkade erregen den SC und befreien ihn gleichzeitig über den Nucleus caudatus von der Hemmung durch die Substantia nigra.

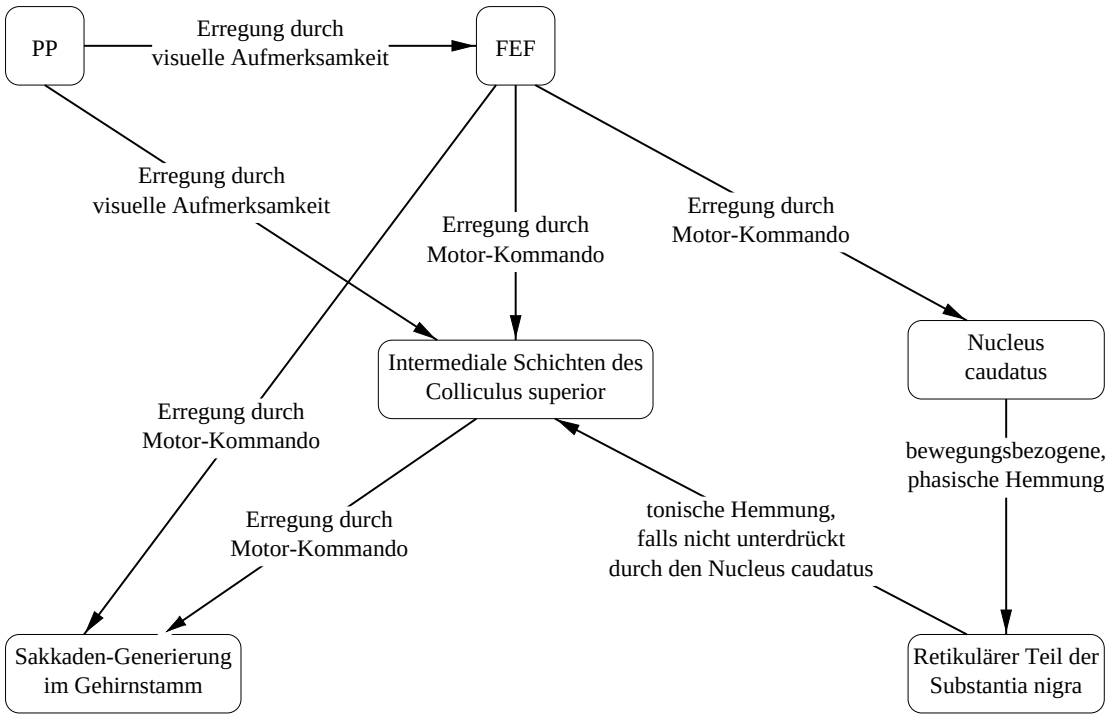


Abbildung 3.12: Okulomotorisches System zur Ausführung von Sakkaden (aus [Goldberg 1991])

3.4 Maschinelle Aufmerksamkeitsmodelle

Wie bereits erwähnt, ist die ‘Feature Integration Theory’ die wohl am weitesten akzeptierte Theorie zur visuellen Aufmerksamkeit. Das zeigt sich auch, wenn man die in diesem Abschnitt aufgeführten maschinellen Modelle betrachtet. Die Umsetzung einer ‘Early Selection Theory’ in ein technisches System bietet sich z. B. deshalb an, weil zur Merkmalsgewinnung auf bekannte Methoden der Bildverarbeitung (Filteroperationen, Regionensegmentierung, etc.) zu-

rück gegriffen werden kann. Weiterhin läßt sich ein solches System auch sehr schön modular aufbauen und gegebenenfalls durch die Hinzunahme weiterer Merkmale ergänzen. Die Umsetzung einer ‘Late Selection Theory’, wie z. B. der von Duncan und Humphreys vorgeschlagenen, in ein Computermodell erscheint dagegen sehr viel schwieriger, denn diese Theorie besagt ja, daß man zunächst eine komplette Beschreibung der in der Szene enthaltenen Objekte erstellen muß, bevor ein Auswahlprozeß dann das interessanteste Objekt selektieren kann. Ein auf dieser Theorie basierendes technisches System hätte also alle in der Szene vorhandenen Objekte holistisch zu erkennen. Dies ist besonders für komplexe Szenen häufig ein sehr schwieriges Problem. Die Vorteile aktiver Sehsysteme liegen ja aber gerade in der Möglichkeit, die relative Position des Beobachters zum Objekt verändern, mehrere Ansichten eines Objekts gewinnen und sich auf die relevanten Szenenausschnitte konzentrieren zu können. Aus dieser Sicht scheint auch für die Aufmerksamkeitssteuerung des aktiven Sehsystems NAVIS die Umsetzung eines auf der ‘Early Selection Theory’ basierenden Modells von Vorteil. In den folgenden Abschnitten werden aber zunächst bereits aus der Literatur bekannte maschinelle Aufmerksamkeitsmodelle vorgestellt.

3.4.1 Theoretische Überlegungen

Koch und Ullman

Eine grundlegende Arbeit auf dem Gebiet der maschinellen Aufmerksamkeitsmodelle ist von Koch und Ullman bereits Mitte der achtziger Jahre geschaffen worden [Koch 1985]. Nach Auffassung von Koch und Ullman sollte eine Aufmerksamkeitssteuerung für ein technisches System folgende Merkmale aufweisen:

1. Elementare Objekteigenschaften wie Farbe, Orientierung, Bewegungsrichtung, etc. werden parallel berechnet und in verschiedenen topographischen Merkmalskarten repräsentiert.
2. Es existiert ein Mapping von diesen Karten in eine zentrale nicht-topographische Darstellung, die zu jedem Zeitpunkt genau einen Punkt enthält; nämlich den Punkt, zu dem der nächste Blickwechsel erfolgen soll.
3. Die Lokalisierung des interessantesten Punktes wird mittels eines WTA-Netzwerks (’*W*inner *T*akes *A*ll‘, entspricht Maximumsuche) durchgeführt.
4. Nach der Fixation des interessantesten Punktes wird dieser gehemmt, wodurch der Blickwechsel zu dem nächst interessantesten Punkt ausgelöst wird.
5. Zusätzlich können Regeln über Nachbarschafts- und Ähnlichkeitsbeziehungen eingeführt werden.

Die Umsetzung der geforderten Eigenschaften in ein geschlossenes Modell ist in Abbildung 3.13 dargestellt.

Koch und Ullman postulieren weiter, daß auf der Ebene der ‘Feature Maps’ oder früher lokale inhibitorische Verbindungen existieren müssen, so daß Orte, die sich signifikant von ihrer Umgebung unterscheiden, herausgehoben werden. Um das globale Maximum der Aufmerksamkeitspunkte zu finden, muß eine weitere topographische Karte, die ‘Saliency Map’, existieren, die die Information der einzelnen Merkmalskarten kombiniert. Die Eigenschaften des über den WTA-Prozeß gefundenen interessantesten Punktes werden anschließend in eine zentrale Repräsentation kopiert und der Punkt wird in der ‘Saliency Map’ gehemmt.

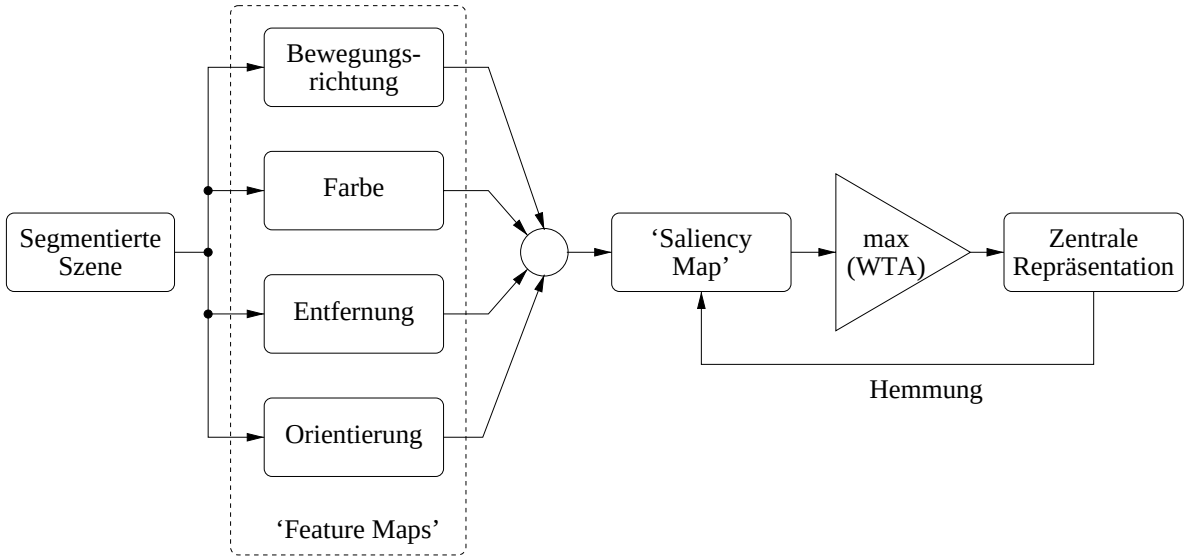


Abbildung 3.13: Prinzipieller Aufbau des Aufmerksamkeitsmodells von Koch und Ullman

Aus den genannten Eigenschaften wird deutlich, daß sich das Modell von Koch und Ullman stark an den frühen Versionen der ‘Feature Integration Theory’ ([Treisman 1980], [Treisman 1985]) orientiert. Implementationen eines solchen rein datengetriebenen Modells sind im Abschnitt 3.4.3 unter der Überschrift "Merkmalsbasierte Modelle" beschrieben.

Westelius

Das von Koch und Ullman vorgestellte Modell berücksichtigt allerdings nicht die starke Aufgabenabhängigkeit der Aufmerksamkeits- und Blickwechsel. Eine effektive Blicksteuerung sollte nach Westelius [Westelius 1991] mehrere gleichzeitig aktive Kontrollmechanismen besitzen, die sich grob wie folgt unterteilen lassen:

- (präattentive) bilddatengetriebene Kontrolle: Strukturelle Bildinformation und nicht vorhergesagte Ereignisse nehmen Einfluß auf die Blicksteuerung, damit eben diese Information analysiert wird.
- attentive, modellgetriebene Kontrolle: Die Blicksteuerung untersucht Regionen aufgrund vorhergesagter Ereignisse und/oder bereits gewonnener Information.
- Verhalten: Sind Bildstrukturen bereits analysiert und modelliert, sinkt ihr Einfluß auf die Blicksteuerung.
- Zufallskomponente: ermittelt Startwert, löst durch "Springen" eventuelle Deadlocks

Eine andere, aber doch ähnliche Unterteilung findet sich auch in [Takacs 1996]. Auch dieses Modell enthält drei parallel arbeitende Einheiten: eine bilddatengetriebene Komponente, die (präattentive) Merkmale detektiert, eine modellgetriebene Komponente, die die Merkmale oder Teilformen eines einzigen Objekts einsammelt, und eine aufgabengetriebene Stufe, die eine Art Verhalten einführt, indem sie dafür sorgt, daß nur die Objekte betrachtet werden, die für die derzeitige Aufgabe relevant sind. Alle in diesem Abschnitt vorgestellten maschinellen Aufmerksamkeitsmodelle haben aber gemeinsam, daß sie einen eher theoretischen Beitrag zur Entwicklung einer Blicksteuerung leisten. In den folgenden Abschnitten werden dagegen Implementationen konkreter Modelle vorgestellt.

3.4.2 Konnektionistische Modelle

Eine Klasse maschineller Aufmerksamkeitsmodelle bilden die in diesem Abschnitt beschriebenen konnektionistischen Modelle, das sind Modelle, die auf künstlichen neuronalen Netzwerken (KNN) basieren. Eine Vielzahl dieser Modelle befaßt sich mit der Transformation eines auffälligen Raum- oder Objektausschnitts in einen Referenzrahmen, dem sogenannten ‘Pattern Routing’. Eine derartige Transformation beinhaltet eine translations- und u. U. auch eine größeninvariante Abbildung. Für ein aktives Sehsystem scheint eine solche Modellierung allerdings nicht von großer Bedeutung zu sein, da die translationsinvariante Abbildung durch die Fokussierung auf das auffällige Objekt "augenblicklich" erzielt werden kann. Eine größeninvariante Abbildung soll zudem innerhalb des NAVIS-Projektes durch die Integration einer modifizierten logarithmisch-polaren Retina realisiert werden[3]. Die Mehrzahl der konnektionistischen Modelle vernachlässigt zudem die erste Stufe eines Aufmerksamkeitsmodells, die Extraktion interessanter Merkmale und die Lösung auffälliger Objekte oder Raumausschnitte aus dem Szenenhintergrund. Daher sind diese Modelle nur bedingt auf reale Szenen anwendbar. Sie entsprechen damit bezüglich Funktionalität und Entwurfsziel nicht den Anforderungen, die an eine Blicksteuerung für ein aktives Sehsystem gestellt werden müssen. Dennoch enthalten einige konnektionistische Modelle grundlegende Ideen, die auch für die Entwicklung einer Blicksteuerung für ein aktives Sehsystem nützlich sein können. Die Reihenfolge, in der die

[3] Dieser Teil wird von unserem Projektpartner, dem FhG-Institut für Mikroelektronische Schaltungen und Systeme in Dresden, bearbeitet.

konnektionistischen Aufmerksamkeitsmodelle im folgenden beschrieben werden, ist im wesentlichen willkürlich und stellt keine Wertung dar. Gleiches gilt auch für die im Abschnitt 3.4.3 beschriebenen merkmalsbasierten Modelle sowie für die Blicksteuerungen der in Abschnitt 3.4.4 beschriebenen aktiven Sehsysteme.

Tsotsos 'Inhibitory Attentional Beam'

Tsotsos stellt in [Tsotsos 1993] einen gegenüber dem Ansatz von Koch und Ullman modifizierten WTA-Prozeß zur Lösung des Aufmerksamkeitsproblems dar. Der modifizierte WTA-Prozeß ist dadurch gekennzeichnet, daß der gewinnende Netzknoten (engl. Unit) seinen ursprünglichen Wert beibehält (der Wert wird also nicht wie üblicherweise erhöht) und nur die Gewichte der verlierenden Netzknoten verringert werden. Desweiteren sind im Gegensatz zu dem von Koch und Ullman vorgeschlagenen WTA-Prozeß auch mehrere Gewinner zugelassen. Der vorgeschlagene Algorithmus läßt sich wie folgt beschreiben (siehe auch [Culhane 1992a+b]):

1. Empfange Stimulus auf der Ebene der Eingangsschicht (kontinuierlich über der Zeit).
2. Empfange Aufgabe auf der Ebene der höchsten Schicht (asynchron, wann immer erhältlich).
3. Führe Schritte 4. bis 8. kontinuierlich über der Zeit aus.
   4. Berechne die Repräsentation im Netz.
   5. Wende den 'Inhibitory Beam' an.
   6. Berechne die Repräsentation erneut.
   7. Extrahiere ausgewählte Region.
   8. Hemme die ausgewählte Region (oder möglicherweise auch mehrere Regionen) in der Eingangsschicht, falls sich die Werte der zugehörigen Netzknoten nicht geändert haben.

Der 'Inhibitory Beam' besteht aus einer Reihe von WTA-Prozessen, die sich von der obersten Schicht (engl. Layer) bis in die unterste Schicht fortpflanzen. In einem ersten Schritt wird in der obersten Schicht der globale Gewinner ermittelt. Dieser Gewinner wird als global bezeichnet, weil alle Knoten der Schicht an dem WTA-Prozeß beteiligt sind. In den nächst niedrigeren Schichten findet der WTA-Prozeß dann nur noch unter den Knoten statt, die mit dem Gewinner der jeweils höheren Schicht verbunden sind (lokale Gewinner). In der untersten Schicht, dem Rezeptor-Layer, ergibt sich letztendlich eine Ansammlung von Knoten, die mit der zum gegebenen Zeitpunkt interessantesten Region im Eingangsbild korrespondiert (siehe hierzu auch Abbildung 3.14, das Modell läßt sich entsprechend auf zweidimensionale Layer erweitern). Ein Blickwechsel wird durch die vollständige Hemmung der ausgewählten Region und erneute Berechnung der internen Repräsentation vollzogen.

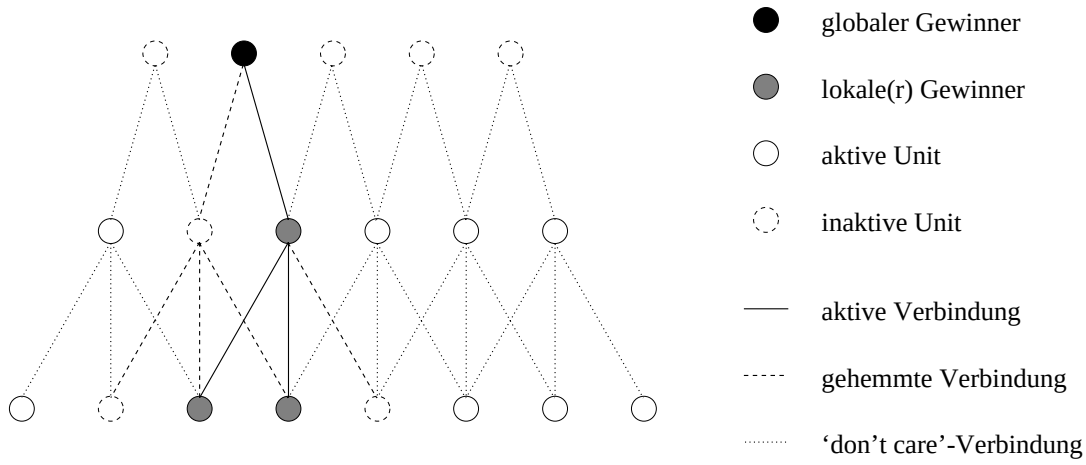


Abbildung 3.14: Eindimensionale Darstellung des ‘Inhibitory Beam’ (aus [Tsotsos 1993])

In [Tsotsos 1995] findet sich die Beschreibung einer weiteren Modifikation des konnektionistischen Netzwerks. Statt nur einer Unit an jeder Position im Netzwerk, besteht dieses Netzwerk aus einer Reihe solcher Units pro Netzwerkposition mit jeweils unterschiedlich großen rezeptiven Feldern. Unter den Units einer Netzwerkposition findet nun ebenfalls ein WTA-Prozeß statt. Es gewinnt diejenige Unit unter denen mit den größten Antworten, die das größte rezeptive Feld besitzt. Die Anzahl der Layer und Units pro Layer spielt für die Funktion des ‘Inhibitory Beams’ keine wesentliche Rolle und ist in den dargestellten Experimenten eher zufällig gewählt worden.

Olshausens ‘Dynamic Routing Circuits’

Anderson und Van Essen beschäftigen sich in [Anderson 1987] mit der Frage, wie ein gefundener FOA, also ein begrenzter Ausschnitt aus dem visuellen Feld, auf ein Zentrum in den höheren cortikalen Schichten abgebildet werden kann, ohne daß die Information über die räumliche Beziehung innerhalb der begrenzten Region verloren geht. Olshausen, Anderson und Van Essen [Olshausen 1993] erweiterten ihr Modell später um eine Version, die zusätzlich in der Lage ist, Objektrepräsentationen während der Überführung in die zentrale Darstellung zu skalieren. Eine Objektrepräsentation besteht hier nun nicht mehr nur aus Bildpunkten, sondern aus Merkmalsvektoren. Sie stellt ein orts- und größeninvariantes Referenzmuster für eine sich anschließende Erkennung, z. B. ein ‘Template Matching’, dar.

Die Autoren schlagen ein Modell vor, in dem der Informationsfluß durch die hierarchischen Schichten von Kontrollneuronen gesteuert wird. Jedes Kontrollneuron besitzt Verbindungen zu einer lokalen Gruppe sogenannter Synapsen (siehe Abbildung 3.15, das Modell läßt sich entsprechend auf zweidimensionale Layer erweitern).

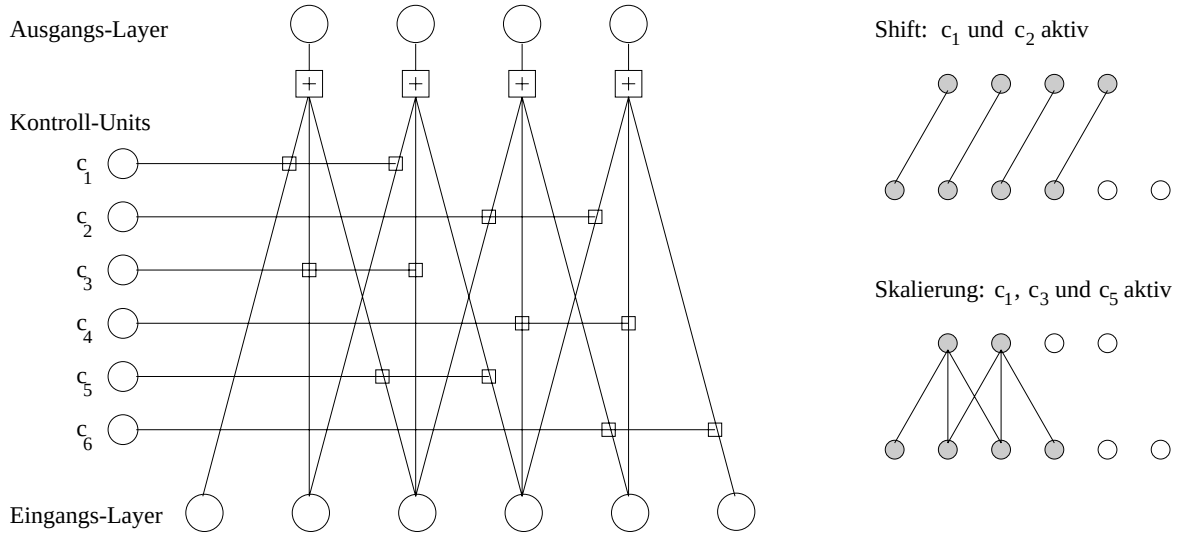


Abbildung 3.15: Eindimensionale Darstellung der ‘Dynamic Routing Circuits’ (aus [Olshausen 1993])

Als auffällige Regionen ergeben sich nach einer Tiefpaßfilterung kompakte Intensitätsblobs. Andere Merkmale als die Intensität sind nicht implementiert. Der hellste Intensitätsblob bestimmt die Position und Größe des FOA. Die zu dieser auffälligen Region gehörenden Kontrollneurone sind die Gewinner eines WTA-Prozesses, der die Position des FOA ermittelt.

Ein zweiter Satz Kontrollneurone ermöglicht die Überführung des FOA auf eine größeninvariante Objektrepräsentation. Kontrollneurone, die zu einem kleinen FOA führen, werden mit einem fein aufgelösten Intensitätsbild versorgt, während den Kontrollneuronen, die zu einem großen FOA gehören, ein grob aufgelöstes Intensitätsbild präsentiert wird. Ein WTA-Prozeß unter den Kontrollneuronen ergibt die Skalierung der Auffälligkeitsregion. Die für die Blickwechsel nötige Hemmung wird durch das Ausschalten der derzeit am FOA beteiligten Kontrollneurone realisiert.

In [Olshausen 1995] wird das Modell der ‘Dynamic Routing Circuits’ um eine Auflösungspyramide ergänzt. Hierdurch können sehr große Skalierungen, wie sie im oben beschriebenen Modell auftreten können, vermieden. Stattdessen wird die geeignetste, vorberechnete Auflösungsebene ausgewählt. Jedes Abtastgitter der Auflösungspyramide besitzt gleich viele Abtastpunkte, wobei die Auflösung von einem Gitter zum nächsten um den Faktor zwei sinkt. Abbildung 3.16 zeigt das erweiterte Schema.

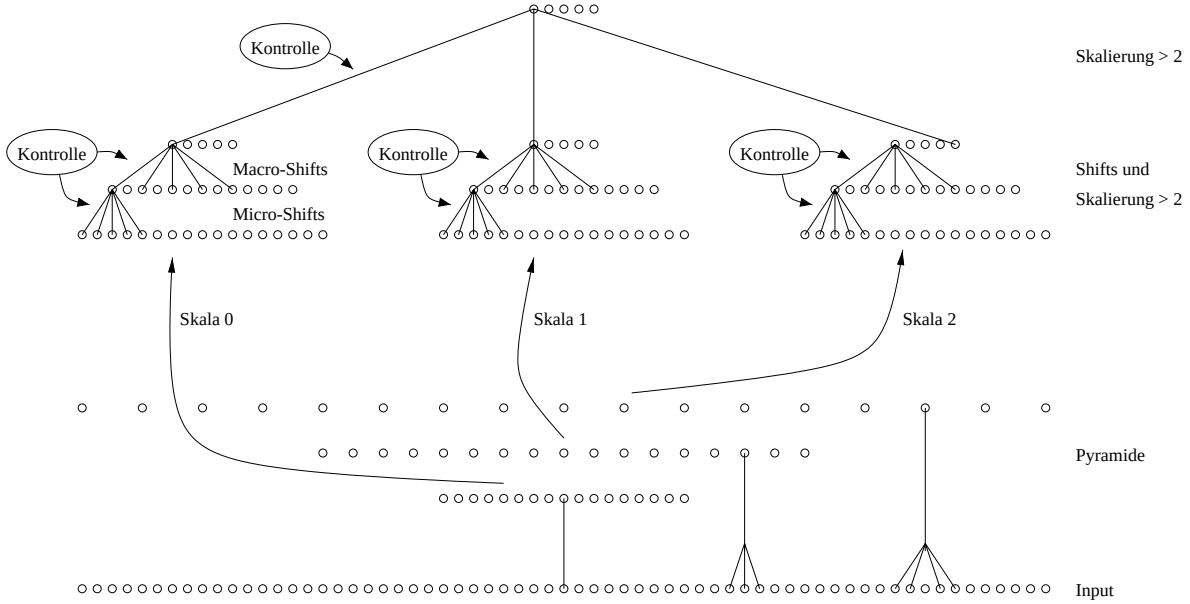


Abbildung 3.16: ‘Dynamic Routing Circuits’ unter Verwendung einer Auflösungspyramide (aus [Olshausen 1995])

Postmas SCAN

Auch Postmas SCAN [Postma 1994] ist insofern ein typisches konnektionistisches Modell, als es ein ‘Pattern Routing’ enthält, das zu einer translationsinvarianten Abbildung eines Ausschnitts des Eingangsmusters führt. Das Modell weist einige Parallelen zu den Modellen von Tsotsos und Olshausen auf. Das Gesamtsystem besteht aus zwei Teilen: einem Gating-Netzwerk, das das eigentliche ‘Pattern Routing’ übernimmt, und einem Klassifikationsnetzwerk. Das Gating-Netzwerk besteht aus mehreren hierarchischen Schichten, den sogenannten Gating-Gittern (engl. ‘Gating Lattices’), die wiederum aus mehreren Teilgittern (engl. Sublattices) bestehen. Die Anzahl der Gitter nimmt nach oben ab, so daß die oberste Schicht des Gating-Netzwerks aus nur einem Gitter besteht. Das heißt, das räumliche Einzugsgebiet eines Gitters nimmt mit jeder Schicht zu.

Jedes Gitter besteht aus einer konstanten Zahl an Teilgittern und ist regulär, d. h. der Abstand eines jeden Knoten (Gates) zu seinen Nachbarn ist konstant. Jedes Teilgitter besitzt nur Nachbarn, die nicht zum selben Teilgitter gehören. Abbildung 3.17 zeigt die verwendete Gitterstruktur.

Jedes Teilgitter routet genau ein Teilmuster in der Größe der Eingangsschicht des Klassifikationsnetzwerks. Alle Gates eines Teilgitters sind also entweder offen (Muster wird geroutet) oder zu (Muster wird gesperrt). Es ist immer nur das Teilgitter pro Schicht offen, das sich in einem WTA-Prozeß durchgesetzt hat. Den Datenteil des Gating-Netzwerks zeigt Abbildung 3.18.

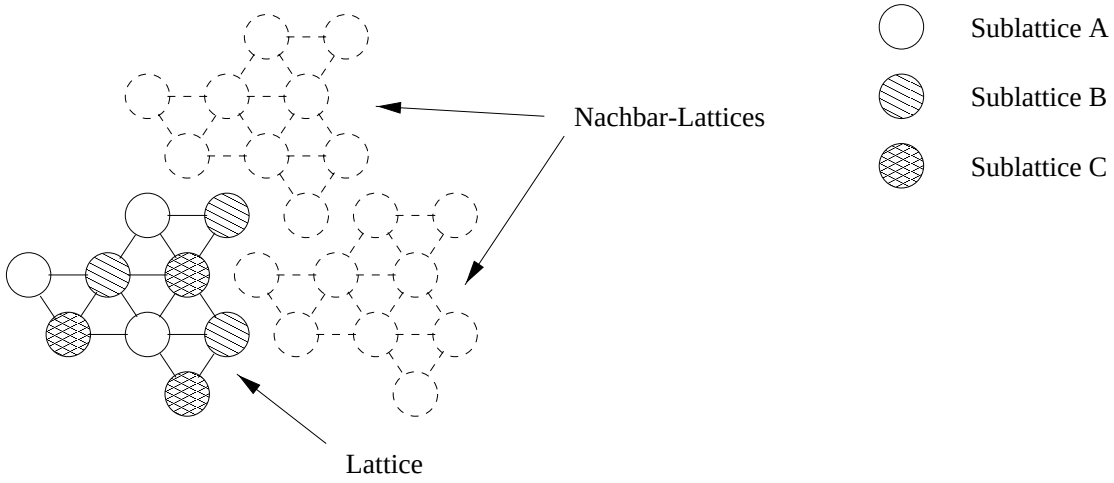


Abbildung 3.17: Gitterstruktur (aus [Postma 1994])

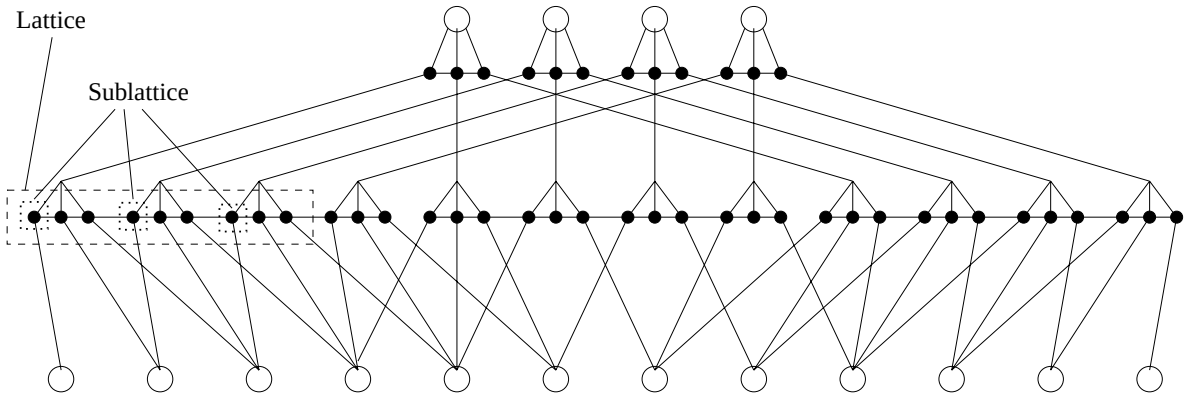


Abbildung 3.18: Datenteil des Gating-Netzwerks (aus [Postma 1994])

Ein Gitter wird durch ein ‘Gate Triplet’ (g_A,g_B,g_C) gesteuert. Geroutet wird immer das Teilgitter, dessen Kontrollsignal den höchsten Wert besitzt. Abbildung 3.19 zeigt den Kontrollteil des Gating-Netzwerks für ein Gitter:

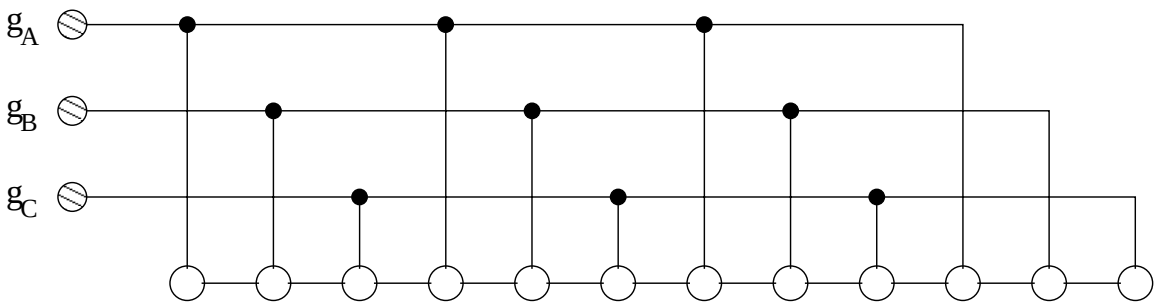


Abbildung 3.19: Kontrollteil des Gating-Netzwerks (aus [Postma 1994])

Das Klassifikationsnetzwerk ist ein vereinfachtes und abgewandeltes ART2-Netzwerk. Es besteht aus zwei Schichten: einer Eingangsschicht, die in einen Eingangs- und einen Ausgangsteil zerlegt werden kann, und einer Kategorisierungsschicht. Die Selektion eines Teilmusters geschieht ‘top down’-gesteuert. Ein aktiver Knoten der Kategorisierungsschicht liefert ein Erwartungsmuster, das mit allen Teilmustern (Sublattices) der Eingangsschicht des Gating-Netzwerks verglichen wird. Es handelt sich also um eine *systematische Suche*, d. h. um eine Suche, bei der alle Stimuli (hier Teilmuster) mit der Beschreibung des Targets (hier Erwartungsmuster) in einer bereits vor Beginn der Suche festgelegten Reihenfolge verglichen werden. Das Teilmuster, das die größte Übereinstimmung mit dem Erwartungsmuster erzielt, erzeugt das größte Kontrollsignal und ermöglicht so das Routing dieses Musters zum Eingang des Klassifikationsnetzwerks. Überschreitet das Kontrollsignal einen Schwellwert, wird das geroutete Muster mit dem erwarteten Muster verglichen. Andernfalls wird der aktuell aktive Knoten der Kategorisierungsschicht unterdrückt und ein anderer Knoten aktiv. Dieser generiert dann ein neues Erwartungsmuster. Der ganze Prozeß dauert an, bis ein Teilmuster klassifiziert werden konnte.

Egner

Das Modell von Egner [Egner 1997] stellt insofern eine Erweiterung der bisher vorgestellten konnektionistischen Aufmerksamkeitsmodelle dar, als es neben dem Routing-Mechanismus einen Zooming-Mechanismus besitzt, der durch die Ausnutzung gezielt gezoomter Detailinformation Ambiguitäten auflöst. Ambiguitäten treten dann auf, wenn ein Objekt nicht eindeutig klassifiziert werden konnte, sich also keine Kategoriehypothese gegen alle anderen Hypothesen durchsetzen konnte. Gezoomt wird auf den Bereich, in dem sich die aktuelle Formbeschreibung, d. h. die Formbeschreibung der derzeit beachteten Figur, und die Formbeschreibung der stärksten Kategoriehypothese maximal unterscheiden, so daß klassifikationsrelevante Forminformation gewonnen werden kann. Ist allerdings zu einer der an den Ambiguitäten beteiligten Kategorien bereits ein Zooming-Bereich mit einer hohen Klassifikationsrelevanz, d. h. ein Zooming-Bereich, der in einem zurückliegenden Erkennungsvorgang zu einer korrekten Klassifikation führte, gespeichert, so wird zunächst dieser Zooming-Bereich erneut aufgesucht. Es sind u. U. mehrere dieser Zooming-Bereiche innerhalb eines Klassifikationsvorgangs aufzusuchen. Daher werden alle im Verlauf eines Vorgangs gezoomten Bereiche in einer Inhibitionskarte gespeichert, um ein erneutes Zooming auf diese Bereiche zu unterbinden. Ist ein Zooming-Bereich ausgewählt worden, findet ein Match zwischen der aktuellen Detail-Formbeschreibung und der Detail-Formbeschreibung der stärksten Kategoriehypothese statt. Das Matchergebnis verändert die Stärke der aktuellen Kategoriehypothese so, daß entweder eine Klassifikation gelingt oder eine neue Kategoriehypothese zur stärksten wird, die dann eventuell zu einer korrekten Klassifikation führt. Das Matchergebnis beeinflußt in jedem Fall auch die zu dem aktuellen Zooming-Bereich gespeicherte (oder noch zu speichernde) Klassifikationsrelevanz. Der Erkennungsvorgang ist abgeschlossen, wenn eine Kategoriehypothese eine festgelegte Stärke erreichen konnte.

Ahmads VISIT

VISIT (‘*Vi*sual *S*earch *It*eratively’) ist ein konnektionistisches Modell, das von Ahmad [Ahmad 1991a+b] insbesondere für die visuelle Suche entworfen und auf seine biologische Plausibilität bezüglich der Lösung von visuellen Suchaufgaben überprüft wurde. Es untersucht seriell die als Target in Betracht kommenden Elemente des Display, bis es das gewünschte Element gefunden hat. Hierzu sind die Elemente geeignet zu gewichten. Im Gegensatz zum Modell von Postma ist in diesem Modell die Suche nach dem Target also keinesfalls systematisch, sondern hängt direkt von den Eigenschaften der Stimuli ab. Abbildung 3.20 zeigt die generelle Architektur.

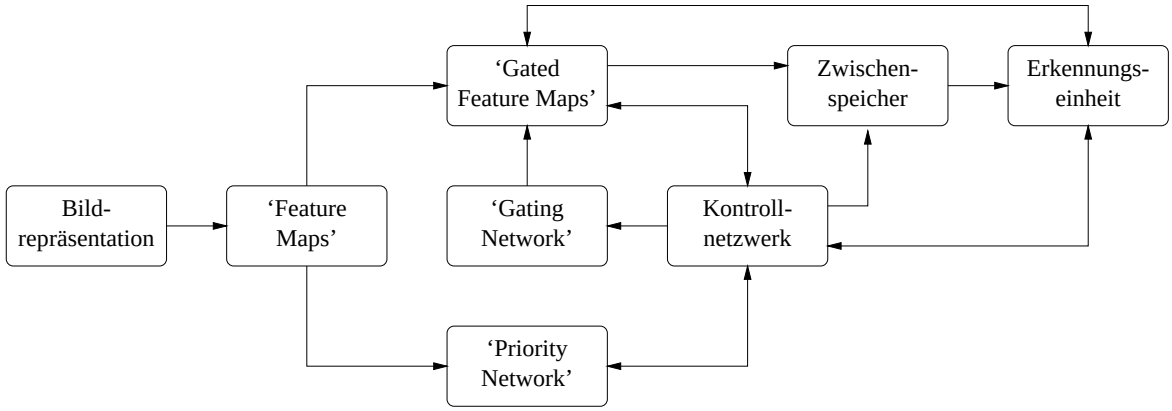


Abbildung 3.20: VISIT-Architektur (aus [Ahmad 1991b])

Das ‘Gating Network’ kontrolliert den Fokus, d. h. eine kreisförmige Aufmerksamkeitsfläche, deren Zentrum und Radius das Netz von drei externen Knoten erhält. Alle Knoten des Gating-Netzwerks, deren Abstand zum Zentrum des Fokus größer ist als der Fokusradius, werden aktiv und hemmen die Knoten der nachfolgenden Schichten. Das Gating-Netzwerk fungiert also als eine Art Lochmaske für die ‘Feature Maps’. Implementiert sind in VISIT vier Merkmalskarten, eine für die Farbe Rot, eine für Blau, eine für die Orientierung Horizontal und eine für Vertikal. Wie die Merkmalskarten entstehen, bleibt unklar. Vermutlich existiert in VISIT keine automatische Klassifizierung der Objekte, sondern diese werden von Hand in die Merkmalskarten eingetragen.

Das ‘Priority Network’ legt den Aufmerksamkeitspunkt fest, auf den der Fokus zu richten ist. Es klassifiziert also die Bildregionen hinsichtlich ihrer Bedeutung. Das Priority-Netzwerk besteht an jeder Position (Pixel im Eingangsbild) aus drei Units, die den Attraktivitätswert und den Abstand zum Zentrum des Fokus in x- und y-Richtung enthalten. Der Attraktivitätswert entspricht hier einfach der Pixeldichte in einer kleinen Region um die betrachtete Position. Ein Eingangsbild mit 256 Grauwertstufen müßte also vor der Verarbeitung mit VISIT erst binarisiert werden, z. B. mit Hilfe eines ‘Half Toning’-Verfahrens. Nur dann läßt sich eine Pixeldichte berechnen. Die Merkmalskarten können unterschiedlich gewichtet werden. Auf diese Weise

kann neben der ‘Bottom up’-Information auch ‘Top down’-Information in den Suchprozeß einbezogen werden.

Das Kontrollnetzwerk berechnet den Schwerpunkt des Fokus und bringt ihn mit dem Attraktivitätsschwerpunkt zur Deckung. Anschließend schrumpft dieses Modul den Fokusradius solange, bis die Summe der Attraktivitätswerte abnimmt (nicht mehr konstant bleibt). Jetzt befindet sich die interessierende Region genau im Fokus und eine Erkennung könnte stattfinden. Der nächste Fixationspunkt wird vom Lehrer vorgegeben. VISIT besitzt also keine automatische Blicksteuerung, die für die Blickwechsel sorgt. Auch ist dieses System vom Autor nicht auf reale Szenen angewendet worden. Stattdessen sind mit dem System Experimente zur visuellen Suche durchgeführt worden. Die für diese Experimente entwickelte Suchstrategie SWIFT (‘*S*earch *Wi*th *F*eatures *T*hrown out’) sorgt dafür, daß nur die Merkmalskarten zur Objektsuche herangezogen werden, die dem System vorgegeben werden. Soll das System z. B. einen roten, vertikalen Balken finden, werden nur die Merkmalskarten für die Farbe Rot und die Orientierung Vertikal untersucht. Die anderen beiden Karten bleiben unberücksichtigt. Desweiteren wird während des Suchprozesses nur die Merkmalskarte, die die wenigeren Objekte enthält, ausgewertet. Sind im Eingangsbildmaterial z. B. weniger rote als vertikale Objekte, werden alle roten Objekte seriell untersucht, ob sie die geforderten Eigenschaften (Rot und Vertikal) aufweisen. Ein solches Verhalten bestätigen auch psychophysische Experimente, siehe z. B. [Egeth 1984].

Eine ausführliche Analyse des Ahmadschen Modells und dessen Erweiterung für den Einsatz auf reale Echtfarbbilder findet sich in [Schmidt 1995].

Sandon

Sandon stellt in [Sandon 1990] die Simulation verschiedener, aus psychologischen Experimenten bekannter Ergebnisse zur visuellen Aufmerksamkeit mit einem zweistufigen neuronalen Netzwerk vor. Das hierfür entworfene Aufmerksamkeitsmodell wird im folgenden beschrieben. Es ist ebenfalls der Klasse der konnektionistischen Modelle zuzuordnen. Eine Übersicht über den Systemaufbau gibt Abbildung 3.21.

Das System arbeitet in vier verschiedenen Auflösungsebenen. Ein 64x64 Pixel großes Eingangsbild wird mit einer 3x3-Gaußmaske dreimal sukzessive gefiltert. Während der drei Filtervorgänge wird die Maske jeweils um zwei Pixel weitergeschoben, so daß sich Bilder der Größe 32x32, 16x16 und 8x8 ergeben. In den so generierten Eingangsbildern werden Kanten in acht Orientierungen detektiert. Das entspricht einer Winkelauflösung von 45^{o}. Über die verwendeten Kantenoperatoren wird keine Aussage gemacht.

Im folgenden wird nur die Verarbeitung in der 64x64-Auflösungsebene betrachtet. Die Kantenbilder oder deren Summe (abhängig vom Experiment) werden in 16x16 Elemente große Aufmerksamkeitsregionen aufgeteilt, wobei zwei benachbarte Regionen um acht Pixel überlappen. Damit ergeben sich für ein 64x64 Pixel großes Kantenbild 7x7 Regionen. Zu jeder lo-

kalen Aufmerksamkeitsregion gehört ein 9x9 Elemente umfassendes Aufmerksamkeitsfeld, wobei jedes Element des Arrays für einen 8x8 großen Ausschnitt der Aufmerksamkeitsregion verantwortlich ist (Abbildung 3.22).

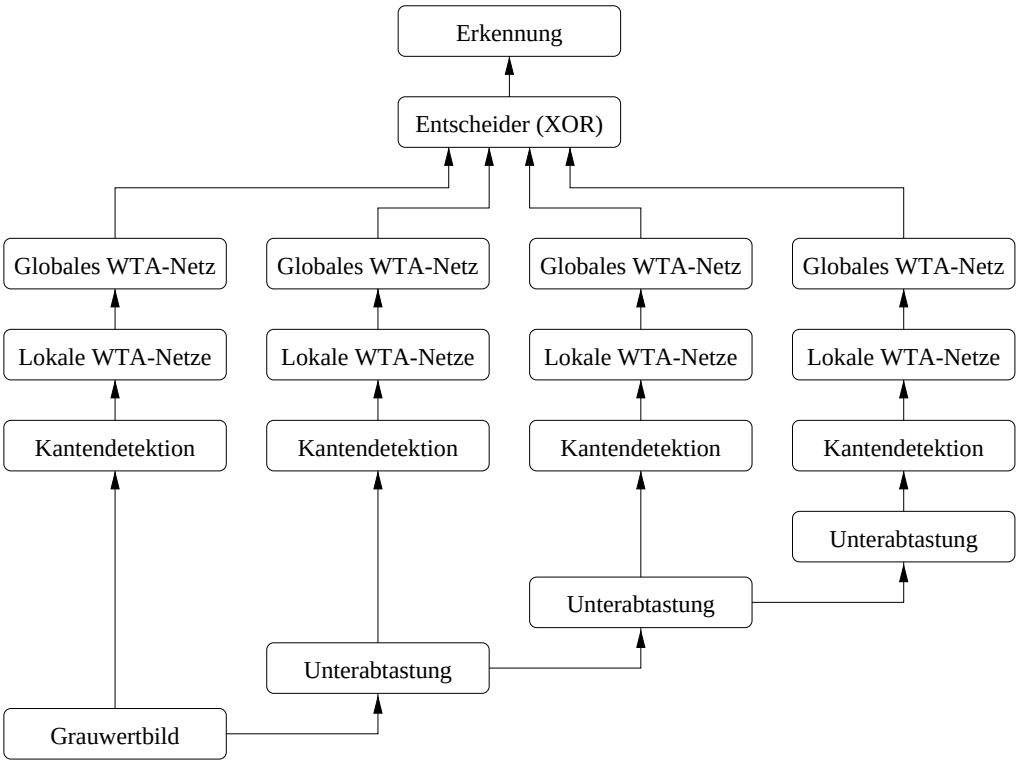


Abbildung 3.21: Architektur des Sandonschen Modells (aus [Sandon 1990])

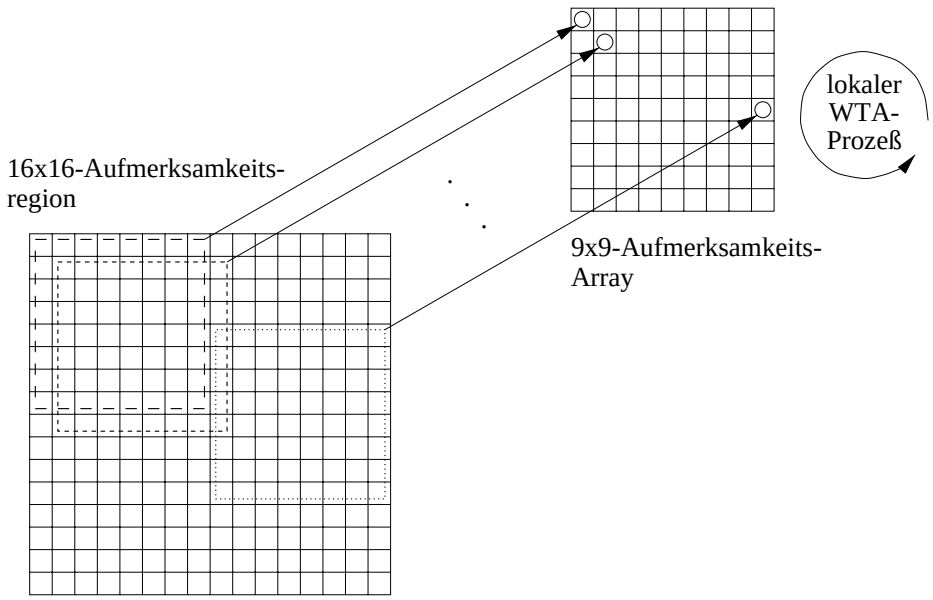


Abbildung 3.22: Ausschnitt aus der ersten Schicht des Sandonschen Netzwerks

In jedem Aufmerksamkeitsfeld findet ein WTA-Prozeß mit dem Ergebnis statt, daß jeweils alle bis auf den am stärksten aktivierten Knoten gehemmt werden. Der gewinnende Knoten in jedem Array gibt die 8x8 große Region der korrespondierenden lokalen Aufmerksamkeitsregion an, die die größte Kantenaktivität zeigt. Aus jeder der überlappenden 16x16 großen Aufmerksamkeitsregionen wird also ein 8x8 großes Aufmerksamkeitsfenster ausgewählt, das der weiteren Verarbeitung zugeführt wird.

In der zweiten Schicht des Sandonschen Netzwerks wird nun das auffälligste der 7x7 Aufmerksamkeitsfenster ausgewählt. Das für diesen Vorgang ausgewählte Merkmal ist die Parallelität zweier gegensätzlich orientierter, d. h. um 180^o gegeneinander verdrehter, Kanten. Die Berechnung der Parallelität findet ausschließlich in den Aufmerksamkeitsfenstern, die die erste Stufe liefert, statt. Für jedes Fenster existiert genau ein Knoten, der die Gesamtaktivität innerhalb des Fensters feststellt. Der Knoten der zweiten Schicht, der die höchste Aktivität aufweist, markiert das 8x8 große Aufmerksamkeitsfenster, das einer nachgeschalteten Erkennung zuzuführen wäre.

Es sind immer nur die Attraktivitätsergebnisse einer der vier Auflösungsebenen an die Erkennung weiterzugeben. Hierfür soll ein Entscheider sorgen, der die Aktivitäten in den verschiedenen Ebenen bewertet. Als vorrangig wird die gröbste Auflösungsebene angesehen, in der Aktivität gefunden werden konnte. Diese Ebene enthält die globalste Aussage über einen interessanten Bildausschnitt. Der Entscheider ist aber ebenso wie die Erkennung nicht implementiert. Weiterhin fehlt die Implementation von Blickwechseln.

3.4.3 Merkmalsbasierte Modelle

Eine weitere Klasse maschineller Aufmerksamkeitsmodelle bilden die von mir als merkmalsbasierte Modelle bezeichneten Verfahren. Diese Modelle beinhalten Filteroperationen, wie sie auch in vielen anderen Bereichen der Bildverarbeitung eingesetzt werden, und sind auf reale Szenen anwendbar. Der Schwerpunkt der merkmalsbasierten Modelle liegt auf der Beantwortung der Frage, wie ein auffälliger Szenenausschnitt überhaupt gefunden werden kann. Ein solches Vorgehen ist auch für die Blicksteuerung von NAVIS notwendig. Die merkmalsbasierten Modelle arbeiten zwar bilddatengetrieben, berücksichtigen aber nicht immer die Integration einer modellgetriebenen Komponente; ihnen fehlt zumeist die Kopplung mit einer Erkennungsstrategie.

Milanese

Milanese [Milanese 1991, 1993, 1994] entwickelte auf der Basis des Aufmerksamkeitsmodells von Koch und Ullman ein eigenes, erweitertes Modell. Das System von Milanese enthält reale Filteroperationen zur Gewinnung der Merkmals- und Auffälligkeitskarten und einen Relaxati-

onsprozeß zur Integration der verschiedenen Auffälligkeitskarten. Es läßt sich im wesentlichen in drei Verarbeitungsstufen unterteilen. In einer ersten Filterstufe werden ‘Feature Maps’ mit den Merkmalen Orientierung, Kanten, Krümmung, Rot-Grün- und Blau-Gelb-Farbkontrast erstellt. In diesen Merkmalskarten werden dann mittels weiterer Filteroperationen auffällige Regionen detektiert und in Auffälligkeitskarten, den ‘Conspicuity Maps’, gespeichert. Die dritte Stufe überführt die ‘Conspicuity Maps’ in eine einzelne zentrale Karte, die ‘Saliency Map’. Durch eine anschließende Binarisierung ergibt sich die finale ‘Attention Map’ (vgl. Abbildung 3.23).

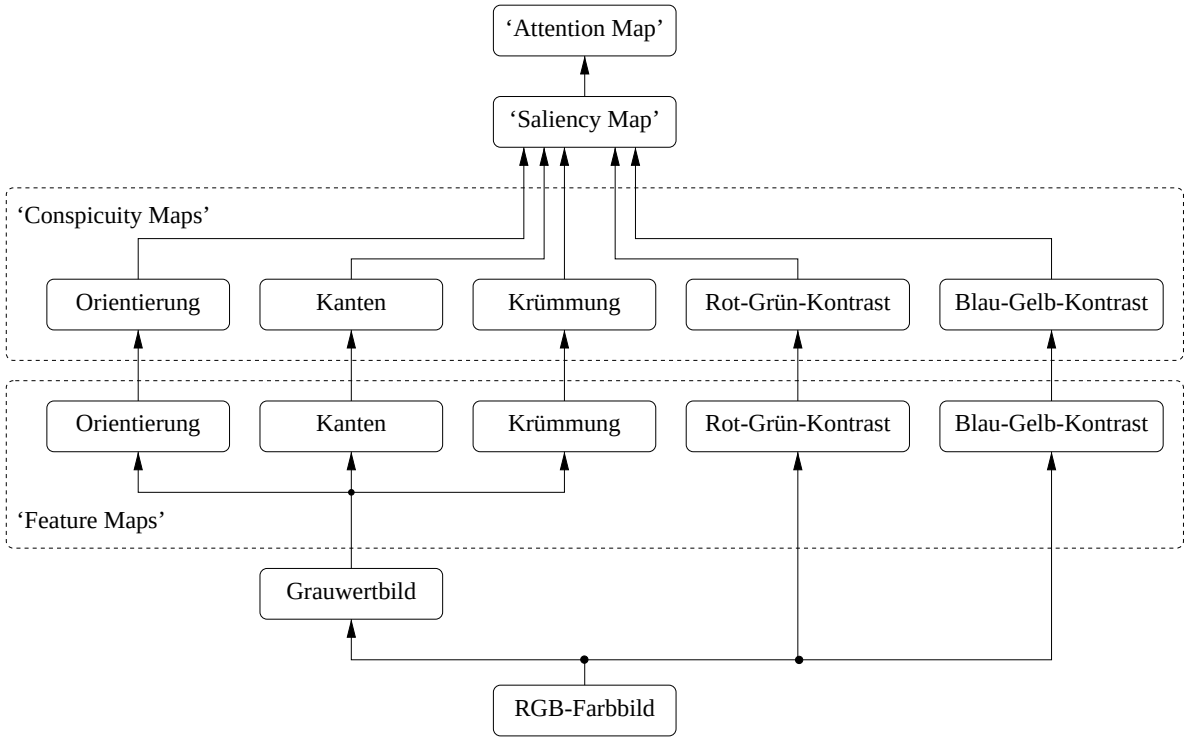


Abbildung 3.23: Architektur des Milaneseschen Aufmerksamkeitsmodells (aus [Milanese 1993])

Durch die Aufsummierung der drei Bänder des RGB-Eingangsbildes wird ein Grauwertbild erzeugt. Aus diesem Bild werden die Merkmale Orientierung, Kanten und Krümmung extrahiert. Zur Gewinnung der beiden erst genannten Merkmale entwirft Milanese eine Filterbank, die auf dem ‘Gaussian Derivative’-Modell von Young [Young 1986] basiert. Zur Detektion von lokalen Orientierungen und Kanten wurde ein Maskensatz auf einer Skala und bestehend aus 16 Orientierungen erzeugt. Die zweidimensionalen Faltungsmasken bestehen aus dem Produkt einer eindimensionalen Gaußfunktion und deren Ableitung, wobei beide Funktionen jeweils von der Koordinate einer der beiden orthogonalen Hauptachsen abhängen. Bildet man nach der Filterung des achromatischen Signals das Maximum über die 16 orientierten Kantenbilder, so erhält man die Merkmalskarte für das Merkmal Kante. Trägt man nun die Orientierungen der zu den gefundenen Maxima gehörenden Faltungsmasken in eine weitere Karte ein, so ergibt sich

die Merkmalskarte für das Merkmal Orientierung. Betrachtet man das Eingangsbild des achromatischen Kanals als Grauwertgebirge, so läßt sich die lokale Krümmung als Divergenz des normalisierten Gradientenbildes auffassen. Durch eine solche Divergenzberechnung wird schließlich die Merkmalskarte Krümmung generiert.

Die Merkmalskarten für die Merkmale Rot-Grün- bzw. Blau-Gelb-Kontrast werden wie folgt gewonnen: Das R-, G-, und B-Band des farbspezifischen Eingangsbildes werden jeweils auf die RGB-Summe normiert und anschließend gaußgefiltert. Aus den derartig vorverarbeiteten Bändern entsteht die Rot-Grün-Karte durch einfache Differenzbildung der entsprechenden manipulierten Bänder und nachfolgender Normierung auf das Intervall [0;1]. Einer solchen Normierung werden auch die Merkmalskarten des achromatischen Kanals unterzogen. Die Merkmalskarte für den Blau-Gelb-Farbkontrast wird auf die gleiche Weise wie die Merkmalskarte für den Rot-Grün-Kontrast erzeugt, wobei die Farbe Gelb als die arithmetische Mittelwertbildung über Rot und Grün definiert ist.

Die ‘Conspicuity Maps’ werden durch eine zweite Filterbank auf drei Skalen und bestehend aus jeweils 16 Orientierungen erzeugt. Die eingesetzten Filter sind ‘*D*ifference *O*f *Or*iented *G*aussians’ (DOOrG-Filter). Die 48 Filterergebnisse, die je ‘Feature Map’ entstehen, werden pixelweise quadriert und nach einer Maximumsoperation in die ‘Conspicuity Map’ eingetragen. In einem weiteren Schritt müssen die ‘Conspicuity Maps’ zu einer einzelnen ‘Saliency Map’ verrechnet werden. Hierzu werden die Auffälligkeitskarten zunächst tiefpaßgefiltert und dann gemeinsam einem Relaxationsprozeß unterzogen mit dem Ziel, kompakte und homogene Auffälligkeitsregionen zu erhalten (siehe [Milanese 1994]). Das Ergebnis des Relaxationsprozesses, die ‘Saliency Map’, wird anschließend binarisiert. Die binäre ‘Attention Map’ soll letztendlich die für eine Objekterkennung oder visuelle Suchaufgabe relevanten Regionen enthalten.

Ein Nachteil des Ansatzes von Milanese ist der hohe Rechenaufwand durch die Vielzahl der Filteroperationen (verschiedene Skalen und Orientierungen) sowie insbesondere durch den Relaxationsprozeß. Die ‘Conspicuity Maps’ stellen zudem lediglich kontrasterhöhte Merkmalskarten dar. Das heißt, die Ebene der ‘Conspicuity Maps’ profitiert nicht von der vorhergehenden Stufe, etwa durch die Gewinnung höherer Merkmale. Blickwechsel und die Kopplung mit einer Objekterkennung oder -verfolgung fehlen.

Itti, Koch und Niebur

Ein weiteres merkmalsbasiertes Aufmerksamkeitsmodell, das ausschließlich bilddatengetrieben arbeitet, ist in [Itti 1998] beschrieben. Das Modell besitzt einen zum Modell von Koch und Ullman sowie zum Modell von Milanese ähnlichen Aufbau (siehe Abbildung 3.24). Berücksichtigt sind in dem Modell die Merkmale Intensität, Farbe und Orientierung. Zentrum-Umfeld-Mechanismen, wie sie von Zellen der Retina und des CGL bekannt sind, werden durch einfache Differenzbildung von feinen und groben Auflösungsebenen modelliert. Hierzu wer-

den zunächst neun Auflösungsebenen mittels Gaußpyramide durch sukzessive Tiefpaßfilterung und Unterabtastung aus dem RGB-Eingangsbild erzeugt (acht Oktaven, Skalierungsfaktoren 1:1 bis 1:256).

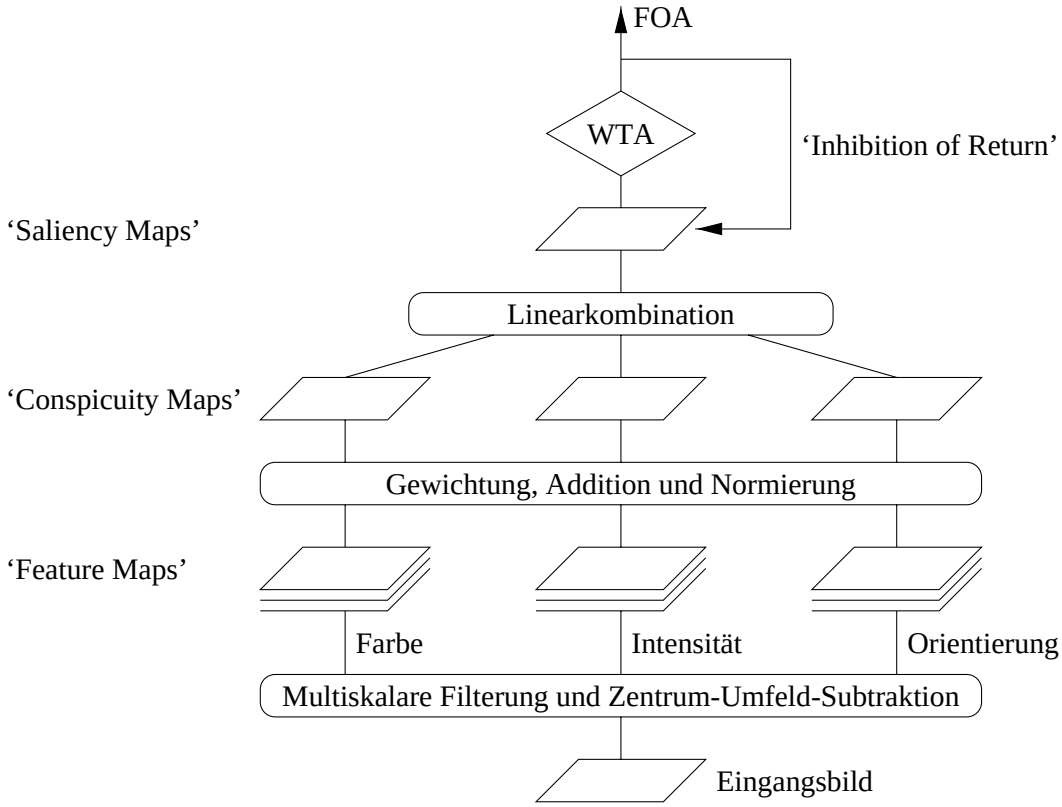


Abbildung 3.24: Aufmerksamkeitsmodell von Itti, Koch und Niebur (aus [Itti 1998])

Die Intensitäten ergeben sich durch die arithmetische Mittelwertbildung über die drei Farbbänder der jeweiligen Auflösungsebene. Für die Erzeugung der sechs Merkmalskarten Intensität werden drei benachbarte feine Auflösungsebenen ausgewählt, die mit jeweils zwei benachbarten grob aufgelösten und auf die notwendige größe interpolierten Auflösungsebenen pixelweise subtrahiert werden. Anschließend wird der Betrag gebildet. Nach demselben Verfahren wird auch ein Rot-Grün- und ein Blau-Gelb-Gegenfarbenkanal gebildet (12 Merkmalskarten Farbe). Zur Berechnung der Merkmalskarten Orientierung werden orientierte Gaborpyramiden verwendet (Winkelauflösung 45^o). Ansonsten wird das beschriebene Verfahren beibehalten. Für vier verschiedene Orientierung ergeben sich also insgesamt 24 Merkmalskarten Orientierung.

Die Berechnung der 'Conspicuity Maps' vollzieht sich in zwei Stufen: Zunächst werden die Merkmalskarten entsprechend der Streuung der in ihnen enthaltenen Werte gewichtet. Die Merkmalskarten, deren Werte stark streuen, also nur wenige sehr hohe Werte enthalten, werden in ihrer Wertigkeit erhöht, solche Karten mit gering streuenden Werten in ihrer Wertigkeit verringert. Anschließend werden die intradimensionalen Merkmalskarten zu jeweils einer 'Con-

spicuity Map' aufaddiert und normiert. Die 'Saliency Map' ergibt sich durch die arithmetische Mittelwertbildung über die drei 'Conspicuity Maps'. Ein WTA-Netzwerk ermittelt die Position des kreisförmigen FOA, dessen Größe auf ein Sechstel der Höhe oder Breite des Eingangsbildes, je nachdem welche kleiner ist, festgelegt ist. Eine transiente lokale Hemmung der Neurone im Bereich des FOA ermöglicht Blickwechsel. Durch die gleichzeitige transiente lokale Erregung der nahen Umgebung des FOA werden kurze Sakkaden langen Sakkaden vorgezogen. Dies entspricht der Berücksichtigung von Nachbarschaftsbeziehungen (eine der Forderungen aus [Koch 1985]).

Das Modell von Itti et al. unterscheidet sich von dem Modell von Milanese insbesondere dadurch, daß es eine flexiblere Skalierung bietet. So sind die DOOrG-Filter im Ansatz von Milanese zwar auf drei Skalen aufgetragen, aber das Verhältnis zwischen den Zentrums- und den Umfeldmasken ist konstant. Der Ansatz von Itti et al. ermöglicht dagegen die Kombination einer feinen Auflösung mit verschiedenen groben Auflösungen. Zudem scheint der Berechnungsaufwand durch die verwendeten Filtermechanismen und den sehr einfachen Bindungsmechanismus wesentlich geringer zu sein. Es fehlt diesem Ansatz allerdings ebenfalls die Berücksichtigung von 'Top down'-Information.

Giefing, Janßen und Mallot

Das Aufmerksamkeitsmodell von Giefing, Janßen und Mallot [Giefing 1991, 1992] unterscheidet sich von den bisher beschriebenen maschinellen Modellen insbesondere dadurch, daß es in ein aktives Bildanalysesystem integriert ist (Abbildung 3.25). Insofern ist das Gesamtsystem dem in Kapitel 5 beschriebenen NAVIS-System ähnlich.

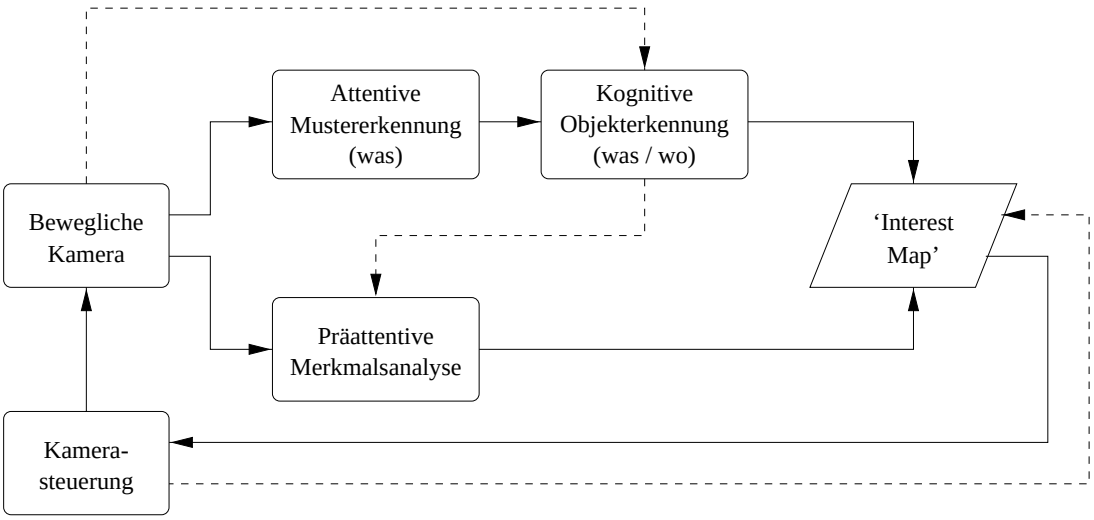


Abbildung 3.25: Übersicht über das aktive Sehsystem von Giefing, Janßen und Mallot (aus [Giefing 1991])

Das Erkennungssystem stellt aufgrund des ersten gefundenen Aufmerksamkeitspunktes eine Objekthypothese auf und versucht diese anhand des Vergleichs von gelernten mit aktuell vorhandenen Teilmustern zu validieren. Die gelernten Teilmuster sind in einer Art semantischen Netzwerk gespeichert. Die vermuteten relativen Positionen, die zur Validierung einer Hypothese von dem System angefahren werden müssen, werden in dem exzitatorischen Teil einer als 'Interest Map' bezeichneten Karte abgelegt. In den inhibitorischen Teil der 'Interest Map' trägt die Kamerasteuerung bereits angefahrende Positionen ein. Ein Relaxationsterm, der den Einfluß des inhibitorischen Teils der 'Interest Map' mit jeder Iteration senkt, ermöglicht das erneute Anfahren von Positionen.

Während die Objekterkennung in den exzitatorischen Teil der 'Interest Map' Punkte einträgt, die aus 'Top down'-Informationen gewonnen wurden (attentive Mustererkennung), verarbeitet das präattentive Aufmerksamkeitsmodul 'Bottom up'-Informationen. Die implementierte Merkmalsanalyse ist texelbasiert. Das heißt, die Merkmalskarten werden durch die Korrelation mit einer vorgegebenen Klasse von Linienelementen (Geraden, Kurven, Kreuzungen, etc.) gewonnen. Nach der Filterung mit grobauflösenden DoG-Filtern werden die Merkmalskarten pixelweise addiert und die Maxima in die 'Interest Map' eingetragen. Abbildung 3.26 zeigt schematisch das Aufmerksamkeitsmodul. Neben den örtlichen Merkmalen kann zusätzlich das Merkmal Zeit berücksichtigt werden. Sowohl im dynamischen Kanal als auch im statischen Kanal wird die extrafoveale Peripherie hoch gewichtet, wogegen sich die Erkennung auf den fovealen Bereich des Bildausschnittes konzentriert.

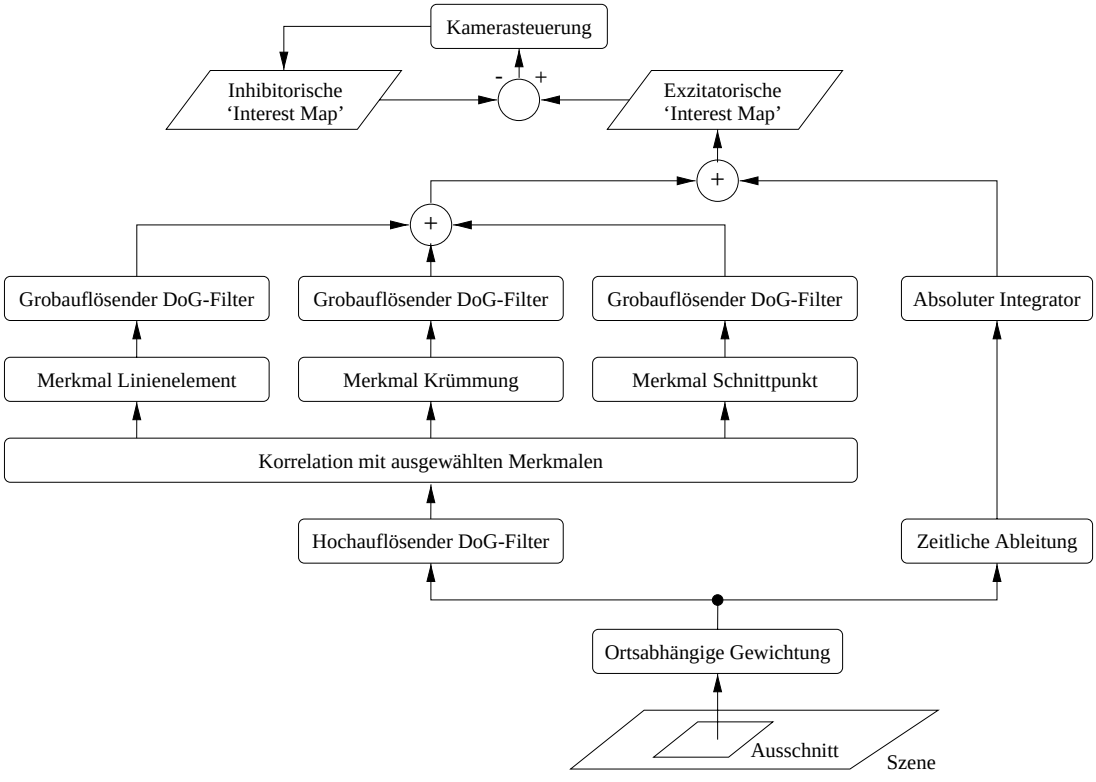


Abbildung 3.26: Prinzipielle Struktur des Aufmerksamkeitsmoduls (aus [Giefing 1991])

Das Modell von Giefing et al. kommt dem in dieser Arbeit vorgestellten System am nächsten, weil es ebenfalls mit einem aktiven Kamerasystem und einer Erkennungsstrategie gekoppelt ist. Ein Vergleich beider Systeme fällt schwer, da das Modell von Giefing et al. in der Literatur nur sehr spärlich beschrieben ist und das Aufmerksamkeitsmodul in aktuellere Arbeiten der Forschungsgruppe nicht einzugehen scheint.

3.4.4 Aktive Sehsysteme

Zu den Paradigmen des aktiven Sehens gehört die Vorstellung, daß Sehen nicht in Isolation stattfinden kann, sondern Teil eines komplexen Systems ist, welches mit seiner Umgebung interagiert ([Aloimonos 1987], [Bajcsy 1988]). Visuelle Aufmerksamkeit, Blicksteuerung, Datenselektion aus einer Vielzahl von Quellen, Berechnung von Tiefenhinweisen, Analyse von Bewegung und Hand-Auge-Koordination gehören zu den Eckpfeilern des aktiven Sehens (vgl. [Eklundh 1996]). Alle diese Techniken sind weitgehend aufgabenabhängig und erfordern Echtzeit-Lösungen. Um eine solche Funktionalität zu erreichen, werden seit Ende der achtziger Jahre besonders Kameraplattformen mit beweglichen Kameras entwickelt (siehe z. B. [Krotkov 1989], [Ballard 1991], [Vieville 1995], [Pahlavan 1996]).

Ein weiteres Paradigma des aktiven Sehens beinhaltet, daß jegliche Analyse visueller Daten immer ausgehend von einer Zielvorstellung, einer Absicht, erfolgt (‘Purposive Vision’, siehe [Aloimonos 1990]). Für die Entwicklung einer Blicksteuerung in einem aktiven Sehsystem bedeutet dieses die Beantwortung der Frage, wie ein für das derzeitige Ziel relevanter Szenenausschnitt gefunden werden kann. Die Aktorik des Systems ist also einzusetzen, um eine im System implizite Modellvorstellung zu validieren. Ballard [Ballard 1991] schlägt vor, ein aktives Sehsystem als eine Ansammlung vieler spezifischer Modelle zu betrachten, deren Ergebnis nur interpretiert werden kann, wenn das zugrundegelegte Verhalten bekannt ist.

Befaßt man sich mit den in der Literatur beschriebenen aktiven Sehsystemen, so fällt auf, daß viele dieser Systeme kein explizites Aufmerksamkeitsmodell besitzen und die Funktion der Blicksteuerung innerhalb des Systems somit schwer abschätzbar bleibt. Tsotsos et al. vertreten aber die Auffassung, daß ohne die Berücksichtigung der visuellen Aufmerksamkeit in einem Bildanalysesystem kein ‘General Purpose’-Sehen möglich ist [Tsotsos 1995]. Vor diesem Hintergrund wird klar, warum die Mehrzahl der aktiven Sehsysteme eben kein Aufmerksamkeitsmodell implizieren: Viele dieser Systeme sind im Gegensatz zu NAVIS nicht für ‘General Purpose’-Sehen konzipiert worden, sondern zur Lösung nur einer einzigen konkreten Aufgabe. Im folgenden werden dennoch einige Blicksteuerungen aktiver Sehsysteme kurz vorgestellt. Es handelt sich hierbei lediglich um eine Auswahl, die die vielfältigen Aufgaben zeigen soll, zu deren Lösung Blicksteuerungen eingesetzt werden. Entsprechend wird kein Anspruch auf Vollständigkeit erhoben.

CVAP, KTH Stockholm

Eine Vielzahl interessanter Arbeiten auf dem Gebiet aktiver Sehsysteme sind in der CVAP-Forschungsgruppe um Eklundh am KTH Stockholm entstanden. Im folgenden sollen zwei exemplarische Arbeiten beschrieben werden: Wie bereits erwähnt, macht es Sinn, die Blicksteuerung eines aktiven Sehsystems analog zum biologischen Vorbild aufgabenabhängig zu steuern. Die in [Brunnström 1994] zugrundegelegte Aufgabe besteht in der Erkennung künstlicher Objekte in einer statischen Szene. Die Blicksteuerung basiert auf der Analyse von Ecken- und Kreuzungspunkten sowie Kanten. Zunächst wird eine besonders auffällige Ecke oder Kreuzung fixiert und klassifiziert. Anschließend wird der Blick auf das Ende eines der (u. U. gekrümmten) Schenkel der Ecke oder Kreuzung gerichtet. Kann auf diese Weise eine neue Ecke oder Kreuzung fixiert werden, so wird diese ebenfalls klassifiziert und, falls sie in einem Schenkel mit der initialen Ecke oder Kreuzung übereinstimmt, mit dieser verknüpft. Ausgehend von der neuen Figur wird durch die Fixation auf ein noch loses Ende eine neue Ecke oder Kreuzung gesucht. Ecken und Kreuzungen, die verknüpft werden konnten, bilden eine immer komplexere Figur. Besitzt diese Figur keine losen Enden mehr, so stellt sie ein Objekt dar.

Eine Blicksteuerung, die speziell für die Verfolgung bewegter Objekte entworfen wurde, stellen Maki et al. in [Maki 1996] vor. Sie basiert auf der Auswertung der Merkmale Disparität, optischer Fluß und Bewegung. Wesentliche Aspekte dieses Modells sind also die zeitliche Veränderlichkeit der Aufmerksamkeit und die räumliche Integration der attentiven Hinweise. Das aktive Sehsystem, in das diese Blicksteuerung integriert ist, kennt zwei Betriebsarten: den Verfolgungs- und den Sakkadenmodus. Im Verfolgungsmodus bleibt die Aufmerksamkeit auf dem aktuellen Target konzentriert und der Rest der Szene wird ausgeblendet. Hierzu wird aus einer Disparitätskarte die Tiefenebene ausgewählt, in der das Target vermutet wird. Gleichzeitig wird aus der Merkmalskarte für den optischen Fluß eine Bewegungseinheit ausgewählt. Um den Berechnungsaufwand zu minimieren, sind nur horizontale Bewegungen zugelassen. Die beiden ausgewählten Masken werden anschließend zu einer sogenannten ‘Target Pursuit’-Maske UND-verknüpft, die dann eindeutig auf das Target hinweist, wenn nicht mehrere bewegliche Objekte in derselben Tiefenebene vorhanden sind. Im Sakkadenmodus wird die Aufmerksamkeit auf ein neues bewegtes Objekt gerichtet. Dieser Modus tritt ein, wenn ein neues Objekt ins Bild kommt und einen geringeren Abstand zum Betrachter aufweist als das bisher verfolgte oder wenn das aktuelle Target "verloren geht" und dadurch ein neues Objekt zum räumlich nächsten wird. Für den ersteren Fall wird die ‘Target Pursuit’-Maske negiert und mit dem aktuellen Bild UND-verknüpft, wodurch das vorherige Target ausgeblendet wird.

Computer Vision Laboratory, Linköping U.

Eine weitere Gruppe, die auf dem Forschungsgebiet aktive Sehsysteme arbeitet, ist das Computer Vision Laboratory unter Leitung von Granlund an der Linköping Universität. Ein Beispiel für den Einsatz einer Blicksteuerung ist die in [Westelius 1995] beschriebene Simulation einer Steuerung mit insgesamt vier Zuständen: "Verfolge Linie", "Suche Linie", "Lokalisiere

Objekt" und "Suche Objekt". Ein Objekt, im Sinne einer auffälligen Struktur, wird zunächst mit Hilfe des in [Reisfeld 1995] beschriebenen Symmetrieoperators gesucht. Konnte ein Objekt gefunden werden, geht die Blicksteuerung in den Zustand "Suche Linie" über. Konnte auch diese gefunden werden, so wird sie anschließend verfolgt (Zustand "Verfolge Linie"). Wird die Linie während der Verfolgung verloren oder das Ende einer Linie erreicht, so wird wieder in den Zustand "Suche Linie" übergegangen, um eine (eventuell neue) Linie in der lokalen Umgebung zu suchen. Wird ein Punkt zum zweiten Mal fixiert, z. B. weil die verfolgte Linie geschlossen ist, geht die Blicksteuerung in den Zustand "Vermeide Objekt" über. Können weitere Symmetrien detektiert werden, folgt anschließend wieder der initiale Zustand "Lokalisiere Objekt", andernfalls wird der Zustand "Suche Linie" angenommen. Beide Kameras des Systems arbeiten hierin zunächst unabhängig. Unterscheidet sich jedoch ihr Zustand, so wird der Zustand mit der höheren Priorität angenommen. Die Priorität der Zustände ist: "Lokalisiere Objekt", "Vermeide Objekt", "Verfolge Linie" und "Suche Linie".

Weitere aktive Sehsysteme

Die Forschungsgruppe um Groß (TU Ilmenau) beschäftigt sich im Rahmen des MILVA-Projektes mit aktiven Sehsystemen [Groß 1997]. Als Aufgabenszenarium wurde eine Transportsituation gewählt. Ein mobiler Roboter soll aufgrund der Analyse nonverbaler Interaktionsformen, wie z. B. Gestik, Mimik und Stimmelodie, auf Transportwünsche reagieren. Das Bildanalysesystem ermöglicht die Detektion von Handflächen und Gesichtern, die dann durch die Blicksteuerung fixiert und hinsichtlich der auszuführenden Transportaufgabe weiter analysiert werden. Ein anderes System, das ebenfalls menschliche Gesten interpretiert, ist in [Jenning 1997] beschrieben. Das System segmentiert zunächst Regionen entsprechend ihrer Farbwerte. Die Flächen mit einer typischen Hautfarbe werden anschließend weitergehend untersucht, um Hände und später auch die Fingerspitzen zu finden und die dadurch ausgedrückte Geste zu interpretieren. Die Fingerspitzen werden durch geometrische Constraints bestimmt. Wurde einmal eine Hand gefunden, wird ein Aufmerksamkeitsfenster um diese Hand generiert, das als räumlicher Index für weitere Fixationen fungiert. Ein weiteres ähnliches System entsteht im ARNOLD-Projekt des Instituts für Neuroinformatik, U Bochum. Hierin soll ein mobiler Roboter für Serviceleistungen eingesetzt werden [Bergener 1997]. Zunächst werden Handpositionen durch die Hautfarbe detektiert, anschließend durch die Blicksteuerung fixiert und interpretiert. Dasselbe Institut nutzt auch psychophysisch motivierte Merkmale zur Aufmerksamkeitssteuerung [von Seelen 1996]. Es werden lineare Filter zur Extraktion auffälliger statischer Merkmale und Bewegungen eingesetzt. Detektiert werden orientierte Kantenelemente, Farbe, Kantenkreuzungen und Ecken. Die Ergebnisse werden für die Steuerung eines autonomen Roboters innerhalb einer neuronalen Architektur eingesetzt.

In [Hamker 1997] wird ein System zur Sortierung von Wertstoffströmen (Altpapier) in Echtzeit vorgestellt. Zum Einsatz kommen hier Textur- und Farbmerkmale. Dabei wird ein Aufmerksamkeitsfenster aus dem Wettbewerb prozeßrelevanter Subziele abgeleitet. Auch Terzopoulos und Rabie [Terzopolous 1995] wählen interessante Regionen durch die Analyse des Merkmals

Farbe aus (Farbhistogrammauswertung nach Swain). Farbmerkmale werden ebenfalls in [Feyrer 1997] eingesetzt. Hier dienen sie zur Detektion von Objekten in dynamischen Umgebungen. Ein mobiler Roboter folgt den anhand der Farbe erkannten Objekten. Vieville et al. [Vieville 1995] stellen ebenfalls eine Blicksteuerung vor, die auf bewegte Objekte fokussiert. Zusätzlich wird die Entfernung des Objektes zum Kamerakopf berücksichtigt. Weiterhin wird die Szene auch nach Regionen mit einer hohen Dichte von Grauwertkanten durchsucht. Barile et al. [Barile 1997] kombinieren die Merkmale Farbe und Bewegung. Ihr Trackingsystem entscheidet in der initialen Phase anhand der Farbe, welches Objekt verfolgt werden soll. Es wird eine Farbsegmentierung der Szene vorgenommen, um den Startpunkt für das Tracking zu bestimmen. Das eigentliche Verfolgen wird dann durch ein binokulares, monochromes Kamerasystem durchgeführt. Die zwischen den S/W-Kameras angeordnete Farbkamera kommt nur während der Initialisierung zum Einsatz. In [Cahn von Seelen 1996] wird die enge Kopplung zwischen der Blicksteuerung und der Erkennungsaufgabe herausgestellt. Zur Suche nach dem Target werden einfache zweidimensionale ‘Template Matching’-Strategien eingesetzt. Die gefundenen Szenenpositionen werden durch den Kamerakopf fixiert.

Zusammenfassend läßt sich feststellen, daß die Mehrzahl der Blicksteuerungen sowohl bilddatengetrieben als auch modellgetrieben arbeitet. Die ‘Bottom up’-Komponente der meisten Steuerungen ist merkmalsbasiert unter Einsatz aus der Bildverarbeitung bekannter Filteroperationen. Konnektionistische Aufmerksamkeitsmodelle finden nur selten Eingang in die z. T. auf Echtzeit ausgelegten aktiven Sehsysteme. Die Anzahl der analysierten Merkmalen ist zudem häufig sehr gering und die Merkmale sind insbesondere für die Lösung nur einer einzigen Aufgabe ausgewählt und angepaßt worden. Ein ‘General Purpose’-Sehsystem wie NAVIS sollte nach unserer Auffassung aber ein separates Aufmerksamkeitsmodul besitzen, das einen möglichst großen Satz an Merkmalen abdeckt. Hierzu sollte das System auch leicht erweiterbar sein. Die zugehörige ‘Bottom up’-Komponente ist im folgenden Kapitel 4 beschrieben.

Es fällt auf, daß der überwiegende Teil der Blicksteuerungen nur durch ein einziges Modell getrieben wird. Damit unterliegt das zugehörige aktive Sehsystem einer sehr begrenzten Funktionalität. Zu den Paradigmen des aktiven Sehens gehört im Gegenteil die Vorstellung, ein aktives Sehsystem als eine Ansammlung vieler ‘Special Purpose’-Algorithmen anzusehen [Ballard 1991]. Es ist daher notwendig, eine intelligente Blicksteuerung, ein Verhalten, einzuführen, daß situationsabhängig eine Modellvorstellung aus einem größeren Satz an Modellen auswählt. So sind dann auch in NAVIS derzeit bereits drei Modelle vorhanden, die Objektsuche, die Identifikation und die Objektverfolgung. Die Blicksteuerung gibt situationsabhängig das jeweils geeignetste Modell vor (siehe Kapitel 5).

Kapitel 4

Die datengetriebene Aufmerksamkeitssteuerung in NAVIS

Die komplette für das aktive Sehsystem NAVIS entwickelte Aufmerksamkeitssteuerung läßt sich in drei Subsysteme unterteilen: Ein ‘Bottom up’-System, das bilddatengetrieben ist, ein ‘Top down’-System, welches modellgetrieben arbeitet, und eine (bisher nur rudimentäre) Verhaltenssteuerung, die situationsabhängig das jeweils geeignetste Modell der ‘Top down’-Komponente vorgibt. In diesem Kapitel wird die datengetriebene Aufmerksamkeitssteuerung beschrieben. Die modellgetriebene Komponente und die Verhaltenssteuerung werden erst in Kapitel 5 im Zusammenhang mit der Beschreibung des NAVIS-Gesamtsystems dargestellt, da sie eng mit anderen NAVIS-Modulen verzahnt sind. Hier wird zunächst vorausgesetzt, daß die zu explorierenden Szenen statisch sind und nach bereits gelernten Objekten untersucht werden.

Das ‘Bottom up’-Aufmerksamkeitssystem prüft den aktuellen Szenenausschnitt auf das Vorhandensein bestimmter Merkmale und bewertet deren Auffälligkeit. Unter dem Begriff *Merkmal* kann in unserem System jede Objekteigenschaft verstanden werden. Der Begriff *Auffälligkeit* bezeichnet den Interessantheitsgrad einer lokalen Struktur im Bild. Das zugrundeliegende Aufmerksamkeitsmodell zeigt Abbildung 4.1. Das Kamerasystem wählt einen Ausschnitt um den aktuellen Fixationspunkt zur weiteren Verarbeitung aus. Dieser Ausschnitt wird hinsichtlich verschiedener elementarer Merkmale, zunächst Kantenverläufe, homogene Grauwertflächen und Farben, untersucht. Diese Merkmale werden parallel in retinotope Karten eingetragen und bezüglich ihrer Auffälligkeit analysiert. Diese Analyse beinhaltet grob zusammengefaßt die Bewertung geometrischer Größen, wie z. B. Schwerpunkt, Symmetrie, Orientierung und Exzentrizität sowie ein Maß für die Farbdistanz. Die Auffälligkeitskarten beinhalten das Ergebnis der Auffälligkeitsanalyse, eine Liste gewichteter signifikanter Punkte. Diese Punkte werden von uns *Attraktivitätspunkte* genannt und in der sogenannten *Attraktivitätskarte* zusammengefaßt. Im folgenden wird die datengetriebene Komponente der Aufmerksamkeitssteuerung daher auch als *Attraktivitätsrepräsentation* bezeichnet. Die Attraktivitätspunkte sind potentielle Kandidaten für den nächsten *Aufmerksamkeitspunkt*. Dieser beinhaltet das Ziel für die folgende Sakkade und wird der motorischen Kamerasteuerung übergeben. Der Aufmerksamkeitspunkt wird aber nicht nur von den auffälligen Punkten in der Attraktivitätskarte bestimmt, sondern eben auch von dem derzeit vorliegenden Modell, das die rudimentäre Verhaltenssteue-

rung vorgibt. So trägt die Objekterkennung solche Punkte in eine sogenannte *Verstärkungskarte* ein, die zur Validierung einer aktuellen Objekthypothese relevant sind. Zusätzlich beeinflußt eine *Inhibitionskarte* die Blickwechsel, in dem sie bereits untersuchte Punkte hemmt und so sehr kleine Blickzyklen verhindert.

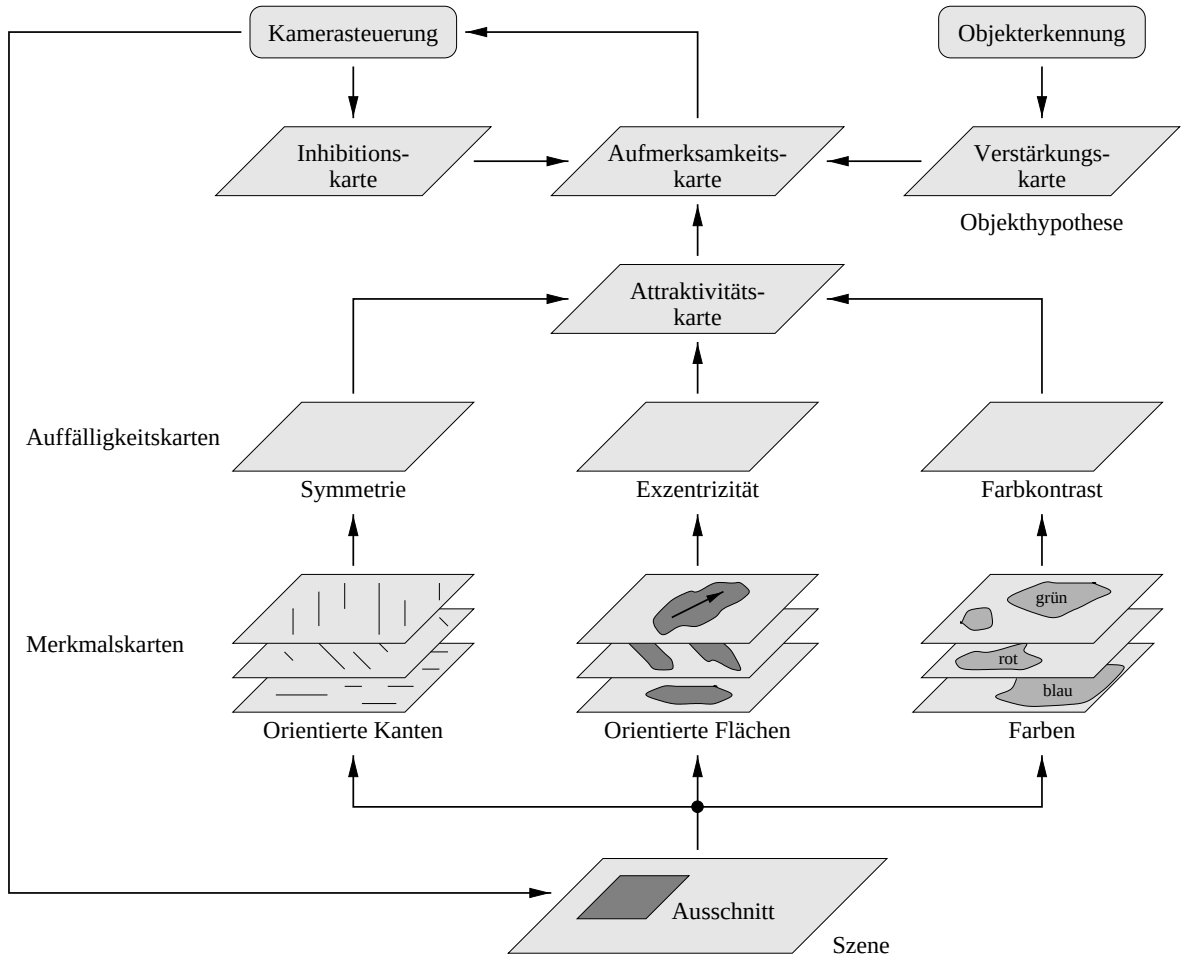


Abbildung 4.1: Schema der Aufmerksamkeitssteuerung in NAVIS

4.1 Die Merkmalskarten

Im folgenden wird die Generierung der verschiedenen Merkmalskarten beschrieben, die die erste Stufe des zugrundegelegten Aufmerksamkeitsmodells bilden. Die interdimensionalen Merkmalskarten bestehen jeweils aus einem ganzen Satz an intradimensionalen Karten, auf die die Merkmale entsprechend eines Klassifikationskriteriums verteilt werden.

4.1.1 Merkmalskarte "Orientierte Kanten"

Die Merkmalskarte "Orientierte Kanten" wird zur Berechnung der Auffälligkeitskarte Symmetrie benötigt. Sie liefert die für den lokalen Symmetrieoperator notwendigen Merkmale; dies sind die Antworten eines *orientierungsselektiven Gaborfiltersatzes*. Der in dieser Arbeit verwendete Symmetrieoperator wird in Abschnitt 4.2.1 vorgestellt. Die Gaborfilterantworten bilden die Datenbasis für eine Vielzahl der in NAVIS implementierten Mechanismen, so z. B. für die Objekterkennung ([Massad 1998a+b], [Mertsching 1998, 1999]), Objektverfolgung ([Springmann 1998], [Hoischen 1999]) oder die stereoskopische Signalauswertung ([Trapp 1995], [Lieder 1998]). Die verwendete Filterbank zur orientierungs- und frequenzselektiven Dekomposition monochromer Bilddaten wurde von Trapp [Trapp 1996] im Rahmen des NAVIS-Projekts an der UGH Paderborn entwickelt und ist dort ausführlich beschrieben.

Die Entwicklung des aktiven Sehsystems NAVIS beinhaltet, gestützt auf Erkenntnissen aus der Psychophysik und den Neurowissenschaften, den Entwurf neuroinformatischer Modelle der visuellen Informationsverarbeitung. Diese Modelle sollen für die technische Bildanalyse genutzt werden. Ein gutes Beispiel hierfür ist der Entwurf der Filterbank. Neurobiologische Untersuchungen [Campbell 1968] ergaben eine Anzahl unabhängiger orientierungs- und ortsfrequenzselektiver Kanäle im visuellen Cortex. Pollen und Ronner [Pollen 1981] führten elektrophysiologische Ableitungen von Zellpaaren mit jeweils gleicher Vorzugsortsfrequenz und -orientierung durch. Sie beobachteten bei Verschiebung eines Testmusters über einen großen Frequenzbereich eine konstante relative Phasenbeziehung von etwa 90˚ zwischen den Antworten der Zellpaare. Diese Tatsache läßt vermuten, daß einfache Zellen als konjugierte Paare mit Quadratur-Phasenbeziehung mit jeweils einem geraden und einem ungeraden rezeptiven Feld um eine gemeinsame Achse zentriert auftreten. Marcelja [Marcelja 1980] und Daugman [Daugman 1985] stellten eine große Ähnlichkeit zwischen den Eigenschaften dieser Zellen und den elementaren Signalen von Gabor [Gabor 1946] fest. Diese Hypothese konnte von Jones und Palmer [Jones 1987a-c] durch die Analyse des Antwortverhaltens einfacher Zellen im Orts- und Ortsfrequenzraum und den Vergleich mit dem Gabor-Formalismus bestätigt werden.

Zusammenfassend läßt sich feststellen, daß das Antwortverhalten einfacher Zellen einer lokalen Ortsfrequenzmessung entspricht, die durch die Gaborfilter umfassend beschrieben werden kann. Gaborfilter erfüllen die Unschärferelation des Orts-Bandbreiteprodukts an der unteren Grenze, d. h. sie sind bezüglich der Orts- und Ortsfrequenzauflösung optimale Filter.

Die Klasse von Signalen, die die untere Grenze des Orts-Bandbreiteprodukts erfüllen, werden auch als Elementarsignale bezeichnet:

$$g(x) = \frac{1}{\sqrt{2\pi}a} \cdot e^{-\frac{x^2}{2a^2}} \cdot e^{jk_0 x} . \qquad (4.1)$$

Der Parameter *a* beschreibt hierin die Breite der gaußförmigen Einhüllenden. Die Fouriertransformierte lautet:

$$G(k) = e^{-\frac{a^2 \cdot (k-k_0)^2}{2}} . \tag{4.2}$$

Hierin bezeichnet k_0 die Schwerpunktsortsfrequenz. Eine Gaborfunktion, die als Filter eingesetzt wird, weist bei einer gegebenen Frequenzauflösung die maximale Ortsauflösung auf. Die effektive Breite der Gewichtsfunktion ist somit minimal und damit ist auch der Berechnungsaufwand eines diskreten Faltungsprodukts minimal.

Eine weitere wichtige Eigenschaft der Gaborfilter ergibt sich aus der Phasenlage ihres Real- und Imaginärteils:

$$Re\{g(x)\} = \frac{1}{\sqrt{2\pi}a} \cdot e^{-\frac{x^2}{2a^2}} \cdot \cos(k_0 x) , \tag{4.3}$$

$$Im\{g(x)\} = \frac{1}{\sqrt{2\pi}a} \cdot e^{-\frac{x^2}{2a^2}} \cdot \sin(k_0 x) . \tag{4.4}$$

Das reellwertige Teilfilter ist gerade und erzeugt keine Phasenverschiebung, während das imaginäre Teilfilter ungerade ist und eine Phasenverschiebung von 90˚ erzeugt. Die Teilfilter besitzen somit ähnlich den von Pollen und Ronner untersuchten einfachen Zellen eine Quadratur-Phasenlage. Das reelle, gerade Teilfilter entspricht in der Bildverarbeitung einem frequenzselektiven Liniendetektor; wogegen das imaginäre, ungerade Teilfilter einen frequenzselektiven Kantenfinder darstellt. Um allerdings Gaborfilter auf Bildsignale anwenden zu können, müssen die Orts- und die Ortsfrequenzvariable als Vektoren im zweidimensionalen Raum betrachtet werden:

$$\boldsymbol{x} = [x_1, x_2]^T, \quad \boldsymbol{k} = [k_1, k_2]^T . \tag{4.5}$$

Für Beziehung (4.1) ergibt sich dann:

$$g(\boldsymbol{x}) = \frac{1}{2\pi a^2} \cdot e^{-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}} \cdot e^{j\boldsymbol{k}_0^T \boldsymbol{x}} . \tag{4.6}$$

Die Matrix $\boldsymbol{A}$ setzt sich aus einer Rotationsmatrix $\boldsymbol{R}$ und einer Parametermatrix $\boldsymbol{P}$ zusammen.

$$\boldsymbol{A} = \boldsymbol{R}\boldsymbol{P}\boldsymbol{R}^T = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} a^{-2} & 0 \\ 0 & b^{-2} \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} . \tag{4.7}$$

Der Parameter b bestimmt die Breite der Einhüllenden in der zusätzlichen Dimension, der Drehwinkel φ die Lage der Einhüllenden im zweidimensionalen Raum und $\boldsymbol{k}_0$ beschreibt die Schwerpunktortsfrequenz. Fordert man nun noch, daß die durch $\boldsymbol{k}_0$ bestimmte Modulationsrichtung orthogonal zur längeren Hauptachse der Einhüllenden ist ($a < b$), so ergibt sich für die Schwerpunktortsfrequenz:

$$\boldsymbol{k}_0 = \begin{bmatrix} |\boldsymbol{k}_0| \cdot \cos\varphi \\ |\boldsymbol{k}_0| \cdot \sin\varphi \end{bmatrix} . \tag{4.8}$$

Der Faktor $|\boldsymbol{k}_0|$ bezeichnet den Betrag der Schwerpunktsfrequenz und bestimmt die radiale Lage des Filters in der Ortsfrequenzebene. Die Übertragungsfunktion des Gaborfilters aus (4.6) ergibt sich dann zu:

$$G(\boldsymbol{k}) = e^{-\frac{1}{2}(\boldsymbol{k}-\boldsymbol{k}_0)^T(\boldsymbol{A}^{-1})^T(\boldsymbol{k}-\boldsymbol{k}_0)} . \tag{4.9}$$

Die grundlegende Idee des Filterentwurfs in NAVIS besteht nun darin, mit Hilfe eines Filtersatzes den Ortsfrequenzraum gleichmäßig zu bedecken. Prinzipiell gibt es hierzu zwei Möglichkeiten. Zum einen kann die absolute Bandbreite der Filter konstant gehalten werden. Dann besitzen zwar alle Filter die gleiche Orts- und Frequenzauflösung, die Anzahl der Schwingungszyklen der harmonischen Funktionen innerhalb der Einhüllenden nimmt dann allerdings mit zunehmendem Betrag der Schwerpunktortsfrequenz zu. Zweitens besteht die Möglichkeit, die relative Bandbreite für alle Filter konstant zu halten. Dies führt zu einer logarithmischen Bandaufteilung des Ortsfrequenzspektrums. Diese Möglichkeit wurde in NAVIS gewählt, weil sich so im Gegensatz zu dem ersteren Fall ein inhomogenes Auflösungsvermögen mit einem fovealen Bereich hoher Ortsauflösung und einer Peripherie niedrigerer Ortsauflösung konstruieren läßt. Der Entwurf der kompletten Filterbank ist in [Trapp 1996] beschrieben, worauf an dieser Stelle verwiesen werden soll. Der in NAVIS verwendete Filtersatz besteht aus vier Skalen und 24 Orientierungen, was einer Winkelauflösung von 15° entspricht. Abbildung 4.2 zeigt das prinzipielle Layout des Gaborfiltersatzes im Frequenzraum. Da die Aufmerksamkeitssteuerung aber sehr zeitkritisch ist, wird in dieser Arbeit nur eine Skala verwendet.

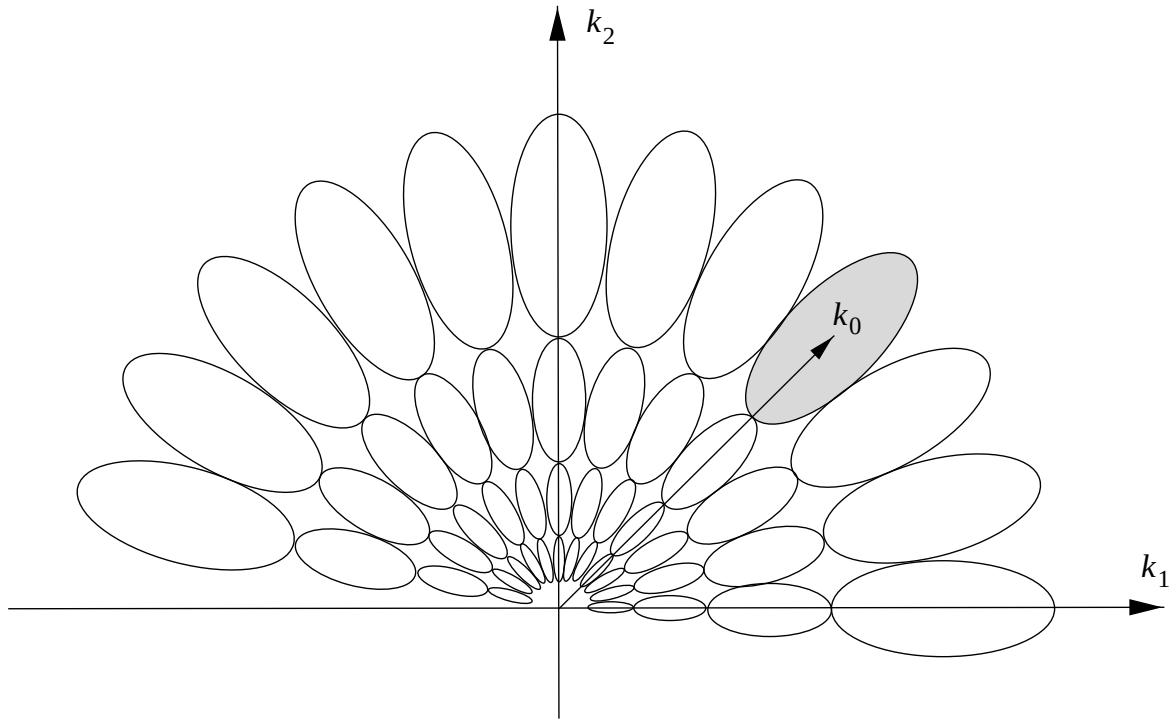


Abbildung 4.2: Filterdurchlaßbereich in der Ortsfrequenzebene

4.1.2 Merkmalskarte "Orientierte Flächen"

Dieser Abschnitt beschreibt die Entwicklung und Implementierung einer Merkmalskarte, die auf der Segmentierung von Zusammenhangsgebieten in Grauwertbildern basiert. Da diesem ‘Bottom up’-Prozeß keine Information über die im Bild vorkommenden oder gesuchten Objekte vorliegt, ist zur Erfassung der Objekte eine robuste Segmentierung notwendig, die auch Flächen mit teilweise starken Helligkeitsschwankungen oder stark gekrümmten Konturverläufen noch als zusammenhängend detektieren kann. Wesentliches Ziel der Forschung im Rahmen des NAVIS-Projekts ist die Entwicklung und Implementierung von Verfahren mit größtmöglicher Anwendungsbreite. Das Fehlen von Vorwissen führt u. a. auch dazu, daß die Zusammenhangsanalyse nur nach ganz allgemeinen Kriterien wie z. B. Varianz des Grauwerts, Kontrast zur Umgebung und ähnlichem durchgeführt werden kann.

Das in diesem Abschnitt und in Abschnitt 4.2.2 (Auffälligkeitskarte Exzentrizität) vorgestellte Verfahren liefert Attraktivitätspunkte, die das auf Symmetrien basierende Verfahren aus konzeptionellen Gründen nicht detektieren kann: Die Fovealisierung auf Grauwertflächen basiert auf einem ‘Region Growing’-Algorithmus mit anschließender Berechnung der Flächenschwerpunkte als interessant eingestufter Regionen. Die Flächenschwerpunkte sind potentielle Aufmerksamkeitspunkte, die, wenn sie an die Motorsteuerung übergeben werden, Objekte ins Zentrum des Interesses bringen können, die entweder keine Symmetrie aufweisen oder deren Sym-

metrie durch Verdeckungen nicht sichtbar ist. So kann beispielsweise auch ein Verkehrszeichen, das teilweise durch andere Objekte, z. B. Äste, verdeckt ist, analysiert werden. Ein mobiles aktives Sehsystems ermöglicht dann eventuell sogar die Umgehung der Verdeckung. Je nach Art der Verdeckungen kann das sich der Segmentierung anschließende Bereichswachstumsverfahren aber auch bereits zur Verschmelzung der Teilobjekte führen, so daß das geschlossene Objekt fixiert werden kann.

In der Literatur werden eine Vielzahl verschiedener Segmentierungsverfahren beschrieben. Im folgenden soll nur auf zwei prinzipielle Vorgehensweisen, die z. T. auch bei der Erstellung dieser Merkmalskarte zum Einsatz kommen, eingegangen werden. Das '*Region Split and Merge*' (z. B. [Gonzales 1987]) beruht auf einer rekursiven Unterteilung des Eingangsbildes in immer kleinere Bereiche mit anschließender Verschmelzung hinreichend ähnlicher, benachbarter Gebiete. Dabei wird im Verlauf der Segmentierung ein bestimmtes Prüfkriterium auf jedes entstehende Teilgebiet angewendet, um die Notwendigkeit einer weiteren Unterteilung zu beurteilen. Häufig entstehen benachbarte Teilregionen, die nahezu den gleichen mittleren Grauwert aufweisen. Wenn diese Regionen sowohl für sich allein betrachtet als auch als Gesamtgebiet gesehen das Prüfkriterium erfüllen, nimmt man an, daß die Teilgebiete zu einer Fläche gehören, und verschmilzt sie miteinander. Ein Nachteil dieses Verfahrens besteht u. a. darin, daß es durch die wiederholte Anwendung des Prüfkriteriums oft zu aufwendigen Summationen kommt, wobei ein großer Teil der Bildpunkte mehrmals abgearbeitet werden muß.

Das zweite hier vorgestellte Verfahren ist das sogenannte *Bereichswachstumsverfahren*. In diesem Verfahren werden bei ausgewählten Startpunkten beginnend Wachstumsprozesse initiiert, die zur Zusammenfassung von Punkten mit ähnlichen Eigenschaften zu Regionen führen. Wesentlich für dieses Verfahren ist die geschickte Reduktion der Startpunkte. In der Regel wird zunächst eine große Anzahl Startpunkte zufällig im Bild verteilt. Dabei werden häufig mehrere Startpunkte innerhalb eines Gebiets mit annähernd konstantem Grauwert gesetzt, so daß die Entfernung überschüssiger Startpunkte notwendig wird. Wenn eine beliebige Verbindung (bestehend aus einer Kette von Bildpunkten) zwischen zwei Initialpunkten gefunden wird, bei der die Grauwertdifferenz gewisse Grenzwerte nicht unter- oder überschreitet, wird einer der betreffenden Initialpunkte entfernt. Nachdem auf diese Weise die Anzahl der Initialpunkte minimiert wurde, beginnt ausgehend von den übrig gebliebenen Punkten ein quasi paralleles Flächenwachstum, wobei für die Ähnlichkeitsprüfung bei der Zuordnung eines Pixels an ein Segment zunächst ein relativ niedrig angesetzter Grenzwert verwendet wird. Lassen sich mit diesem Grenzwert keine Pixel mehr an Regionen anfügen, wird das Bereichswachstumsverfahren mit einem höheren Parameter weitergeführt. Durch schrittweises Erhöhen des Grenzwertes werden schließlich alle Bildpunkte den entsprechenden Regionen zugeordnet (siehe z. B. [Wahl 1984]). Die Vorgabe eines sinnvollen Parameters ist daher nicht Voraussetzung für diesen Segmentierungsprozeß. Ein solches Verfahren erfordert allerdings einen hohen Rechenaufwand und erscheint daher für ein aktives Sehsystem weniger geeignet. In dieser Arbeit ist deshalb ein einfacheres Segmentierungsverfahren mit einem statischen Grenzwert gewählt worden, wobei das Segmentierungsresultat jedoch noch einer Nachbearbeitung unterzogen wird,

indem in einem anschließenden Bereichsverschmelzungsverfahren benachbarte Flächen mit ähnlichem mittleren Grauwert miteinander verschmolzen werden.

Bei vielen in der Literatur beschriebenen Segmentierungverfahren wird der entsprechende Algorithmus direkt auf das Grauwertbild (Eingangsbild) angewendet. Ein Kennzeichen dieser Algorithmen ist, daß ein Maß für die Ähnlichkeit benachbarter Punkte bestimmt wird und daraus eine Entscheidung über die Zuordnung untersuchter Punkte zu bestehenden Regionen getroffen wird. Das hier eingesetzte Verfahren wertet dagegen die Gradientenfelder beider Raumrichtungen aus. Dies entspricht im wesentlichen einer Vorabberechnung des Grades der lokalen Unterscheidbarkeit benachbarter Punkte verbunden mit einer geringen Tiefpaßwirkung. Die Verarbeitungskette, die letztendlich zur Generierung der Merkmalskarte "Orientierte Flächen" führt, ist in Abbildung 4.3 dargestellt:

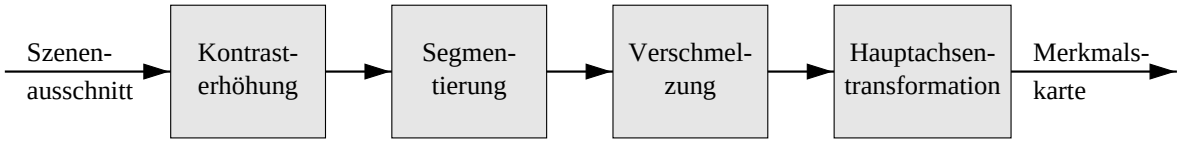


Abbildung 4.3: Verarbeitungskette zur Erstellung der Merkmalskarte "Orientierte Flächen"

Das erste Glied in der Verarbeitungskette ist die komponentenweise Berechnung der Gradientenfelder in horizontaler und vertikaler Richtung. Die komponentenweise Berechnung des Gradienten geschieht durch zweidimensionale Faltung mit den Sobel-Kernen für den horizontalen bzw. vertikalen Anteil. Eine besondere Randbehandlung erübrigt sich bei den verwendeten 3x3-großen Masken, wenn man die Faltung auf die Bildpunkte beschränkt, die innerhalb des Eingangsbildes liegen. Die Verkleinerung des Faltungsergebnisses um jeweils zwei Pixel in beide Dimensionen kann bei der vorliegenden Anwendung problemlos akzeptiert werden. An die nachfolgende Verarbeitungsstufe zur Flächensegmentierung werden die skalaren Felder beider Komponenten übergeben.

Ausgehend von den Gradientenbildern wird mittels eines einfachen Bereichswachstumsverfahrens eine erste Segmentierung vorgenommen. Als geeigneter Startwert für einen Wachstumsprozeß wird dabei jeder Punkt angesehen, der das ebenfalls für den Wachstumsprozeß selbst gültige folgende Kriterium erfüllt:

Bleibt der Betrag des Gradienten für beide Komponenten des Gradientenvektors unterhalb des festgelegten Schwellwertes, wird der betrachtete Punkt der Fläche zugeordnet bzw. ist als neuer Startwert geeignet.

Ist ein Startwert gefunden worden und sind alle benachbarten und hinreichend ähnlichen Punkte mit diesem Punkt zu einem Segment zusammengewachsen, muß ein neuer Startwert gesucht

werden. Die Suche nach geeigneten Startwerten läuft dabei zeilenweise ab und beginnt in der linken, oberen Bildecke. Gleichzeitig ermittelt das Verfahren verschiedene Flächenattribute wie den mittleren Grauwert, die Varianz, die Größe und den Flächenschwerpunkt für das wachsende Segment aus dem Originalbild.

In der Literatur finden sich auch Methoden zur Bestimmung eines ortsabhängigen Schwellwertes. So wird u. a. vorgeschlagen, das zu untersuchende Bild in kleine, sich nicht überlappende Bereiche zu unterteilen und für jeden Bildbereich einen lokalen Schwellwert z. B. durch die Auswertung des jeweiligen Sub-Histogrammes der Grauwerte zu bestimmen. Zur Gewinnung eines stetigen ortsabhängigen Schwellwertverlaufs können die lokalen Schwellwerte anschließend über das gesamte Bild interpoliert werden [Chow 1972]. Im Fordergrund des in dieser Arbeit angewendeten Verfahrens steht jedoch die für ein aktives Sehsystem notwendige geringe Berechnungszeit. Deshalb wird hier mit einem niedrigen statischen Grenzwert segmentiert. Die vielen kleinen Flächensegmente, die bei dieser einfachen Art der Segmentierung entstehen können, werden in der nachfolgend beschriebenen Verarbeitungsstufe soweit möglich noch miteinander verschmolzen oder bei Unterschreiten einer Mindestgröße entfernt.

Das Ergebnis der Segmentierung ist zunächst einmal eine relativ große Anzahl sehr unterschiedlich ausgedehnter Regionen mit den zugehörigen Flächendaten. Objekte enthalten mitunter ausgeprägte Strukturen wie z. B. Beschriftungen auf ansonsten homogenen Oberflächen. Für solche Objekte würden ohne weitere Vorverarbeitung mehrere Attraktivitätspunkte generiert. Um diesem Effekt zu begegnen, faßt das im folgenden beschriebene Bereichsverschmelzungsverfahren Segmente, die aufgrund ihrer Nachbarschaft und hinreichenden Ähnlichkeit zu einem durch den Segmentierungsvorgang fragmentierten Objekt gehören könnten, zu einem neuen, größeren Segment zusammen. Zu diesem Zweck müssen die Zusammengehörigkeitsverhältnisse benachbarter Segmente anhand der vorliegenden Segmentattribute ermittelt werden. Die zugrundeliegende Annahme ist, daß die Eigenschaften der im Grauwertbild sichtbaren Segmente einer Objektoberfläche meist konstant bleiben, auch wenn die betrachtete Oberfläche durch Gegenstände im Vordergrund teilweise verdeckt wird bzw. diese bereits eine gewisse Strukturierung aufweist. Dieses gilt allerdings besonders für künstlich beleuchtete Szenen nur im eingeschränkten Maße.

Es stellt sich nun die Frage, wie der Grad der räumlichen Nähe überhaupt festgestellt werden kann. Das implementierte Verfahren läßt hierzu alle Segmente quasi gleichzeitig in alle Richtungen expandieren. Beim Aufeinandertreffen zweier Wachstumsfronten wird dann getestet, ob die Segmente zu verschmelzen sind. Im Falle eines negativen Ausgangs des Tests wird der Wachstumsprozeß an dem betreffenden Frontverlauf eingestellt. Die Anzahl der Expansionsschritte kann über einen Parameter eingestellt werden. Um die Eigenschaften der zu verschmelzenden Segmente nicht zu verfälschen, werden die Flächenattribute während des Frontenwachstums nicht verändert. Jedoch werden bei einem Verschmelzungsvorgang sämtliche Attribute der neuen Region basierend auf den vorliegenden Werten der betreffenden Segmente neu berechnet. Diese Vorgehensweise führt dazu, daß nur segmentierte Bildpunkte und nicht durch den Expansionsprozeß zugeordnete Pixel in die Bestimmung der Flächenattribute einge-

hen. Dieses ist für die Prüfung auf hinreichende Ähnlichkeit zweier Segmente *A* und *B* wichtig. In der vorliegenden Implementierung besteht der Ähnlichkeitstest aus den folgenden zwei empirisch ermittelten Bedingungen, die beide zu erfüllen sind:

1. Die absolute Differenz der mittleren Grauwerte der Segmente *A* und *B* darf ca. 12% des maximal möglichen Wertes nicht überschreiten (entspricht ca. 30 Grauwertstufen bei einer maximalen Differenz von 255 Stufen).
2. Die Varianzen σ_A und σ_B der Pixelwerte der Segmente *A* und *B* müssen etwa in derselben Größenordnung liegen: $1/k < \sigma_A/\sigma_B < k$, wobei $k = 2$, 3 oder 4

Sind die Grauwertflächen segmentiert, Fragmente verschmolzen und zu kleine Segmente entfernt, werden die entstandenen Flächen mittels *Hauptachsentransformation* entsprechend ihrer Orientierung auf verschiedene Merkmalskarten (derselben Dimension) verteilt. Analog zu den 12 intradimensionalen Merkmalskarten "Orientierte Kanten" werden auch 12 topologische Karten der Dimension "Orientierte Flächen" mit einer Winkelauflösung von 15˚ erzeugt. Allerdings existiert zusätzlich eine 13. Karte, die die Flächen ohne auffällige Vorzugsrichtung enthält.

Die Orientierung eines Segments kann direkt aus den Zentralmomenten gebildet werden. Für den diskreten zweidimensionalen Fall lautet die Gleichung für die Zentralmomente:

$$m_{p,q} = \frac{1}{M_{00}} \cdot \sum_{(x,y) \in A} (x - M_{10})^p \cdot (y - M_{01})^q \ , \tag{4.10}$$

wobei die Fläche *A* genau M_{00} Pixel enthält. Das Koordinatenpaar bestehend aus den Momenten erster Ordnung (M_{10}, M_{01}) bezeichnet den Schwerpunkt der Fläche *A*:

$$M_{10} = \frac{1}{M_{00}} \cdot \sum_{x \in A} x \ , \tag{4.11}$$

$$M_{01} = \frac{1}{M_{00}} \cdot \sum_{y \in A} y \ . \tag{4.12}$$

Die Basis für die Hauptachsentransformation (engl. ‘Principal Component Analysis’) bildet die Kovarianzmatrix $\boldsymbol{C}$ mit den Zentralmomenten 2. Ordnung:

$$\boldsymbol{C} = \begin{bmatrix} m_{2,0} & m_{1,1} \\ m_{1,1} & m_{0,2} \end{bmatrix} \ . \tag{4.13}$$

Der Eigenvektor des größeren Eigenwertes der Kovarianzmatrix gibt dann die gesuchte Orientierung φ des Segments an. Zu lösen ist also die Eigenwertaufgabe:

$$(\boldsymbol{C} - \lambda_i \boldsymbol{E})\, \boldsymbol{b}_i = 0, \quad i = 1, 2 \,, \tag{4.14}$$

wobei $\boldsymbol{E}$ die Einheitsmatrix bezeichnet. Gleichung (4.14) hat nur dann eine Lösung, wenn gilt:

$$\det(\boldsymbol{C} - \lambda \boldsymbol{E}) = 0 \,. \tag{4.15}$$

Die beiden Lösungen der quadratischen Gleichung sind die Eigenwerte:

$$\lambda_{1,2} = \frac{m_{2,0} + m_{0,2}}{2} \pm \sqrt{\left(\frac{m_{2,0} + m_{0,2}}{2}\right)^2 - (m_{2,0} m_{0,2} - m_{1,1}^2)} \tag{4.16}$$

mit den zugehörigen Eigenvektoren:

$$\boldsymbol{b}_1 = \begin{bmatrix} m_{1,1} \\ \lambda_1 - m_{2,0} \end{bmatrix}, \quad \boldsymbol{b}_2 = \begin{bmatrix} \lambda_2 - m_{0,2} \\ m_{1,1} \end{bmatrix} . \tag{4.17}$$

Für die Orientierung der Eigenvektoren ergibt sich dann:

$$\varphi_1 = \operatorname{atan}\left(\frac{b_{1y}}{b_{1x}}\right) , \tag{4.18}$$

und da $\boldsymbol{b}_2$ orthogonal zu $\boldsymbol{b}_1$ ist:

$$\varphi_2 = \operatorname{atan}\left(\frac{b_{2y}}{b_{2x}}\right) = \operatorname{atan}\left(-\frac{b_{1x}}{b_{1y}}\right) . \tag{4.19}$$

Nach einigen Substitutionsschritten (siehe z. B. [Rosenfeld 1982]) ergibt sich schließlich:

$$\varphi = \begin{cases} \varphi_1 = \frac{1}{2}\operatorname{atan}\left(\frac{2m_{1,1}}{m_{2,0} - m_{0,2}}\right) & \text{für } \lambda_1 > \lambda_2 \\ \varphi_2 = \frac{1}{2}\operatorname{atan}\left(\frac{m_{0,2} - m_{2,0}}{2m_{1,1}}\right) & \text{für } \lambda_1 < \lambda_2 \\ \text{nicht definiert} & \text{für } \lambda_1 = \lambda_2 \end{cases} . \tag{4.20}$$

So lassen sich also die Orientierungen der Segmente bestimmen, die anschließend entsprechend ihrer Orientierung auf die verschiedenen Merkmalskarten verteilt werden.

4.1.3 Merkmalskarte Farbe

Eine Vielzahl der in der Natur vorkommenden Tiere und Pflanzen zeichnet sich durch eine auffällige Farbgebung aus, wodurch sich diese deutlich von ihrer Umgebung unterscheiden. Vielfach dient eine derartige Farbgebung anderen Arten als Warnung. Wir Menschen setzen Farbe ebenfalls als Gefahrenanzeige ein. So verwenden wir z. B. Signalfarben auf künstlichen Objekten wie Verkehrzeichen oder Warnschildern in industriellen Umgebungen. Alle diese genannten Aspekte lassen es sinnvoll erscheinen, Farbe und insbesondere Farbkontraste als Auffälligkeitsmaß auch in einer Aufmerksamkeitssteuerung zu nutzen.

Da unser System eine Klassifizierungsleistung aufweisen soll, die der menschlichen Klassifizierungsleistung möglichst nahe kommt, muß zunächst die Farbinformation in einen empfindungsgemäßen Farbraum transformiert werden. Um eine solche geeignete Repräsentation zu finden, wurden einige farbmetrisch und physiologisch begründete Ansätze auf ihre empfindungsgemäße Gleichabständigkeit untersucht. Als gleichabständige Referenz wurde das empirische Munsell-Farbordnungssystem verwendet (Munsell-Farbordnungssystem siehe z. B. [Hunt 1987]). Die Ergebnisse dieser Untersuchung zeigen, daß die farbmetrisch begründeten Farbräume, der *CIELAB* und der *CIELUV,* eine bessere empfindungsgemäße Gleichabständigkeit aufweisen als etwa die untersuchten physiologischen Farbräume von de Valois und de Valois [de Valois 1993], Bollmann [Bollmann 1995] und Pomierski [Pomierski 1996]. Eine relativ gute Charakteristik besitzt auch der *MTM*-Farbraum [Miyahara 1988], der eine approximierte Transformation des *RGB*-Farbraumes in das Munsell-System darstellt. Abbildung 4.4 zeigt die Untersuchungsergebnisse aus [Justkowski 1998] für den physiologischen Farbraum nach de Valois und de Valois, den *CIELAB*, den *CIELUV* und den *MTM*-Farbraum. Jeder Punkt in den Diagrammen steht für eine Farbe aus den ‘Munsell Value Planes’ 4, 6 bzw. 8 übertragen in den jeweiligen Farbraum. ‘Munsell Value Plane’ 0 steht für Schwarz, ‘Munsell Value Plane’ 10 für Weiß, entsprechend repräsentiert ‘Munsell Value Plane’ 4 dunkle, ‘Munsell Value Plane’ 6 mittlere und ‘Munsell Value Plane’ 8 helle Farben. Zur Bewertung der Gleichabständigkeit des jeweiligen Farbraumes ist die Äquidistanz der Punktabstände, die Geradlinigkeit der einzelnen Punktreihen sowie die Äquidistanz der Winkel zwischen den Punktreihen entscheidend. Weiterhin sollte der Schnittpunkt der Punktreihen, die Unbuntachse, in etwa im Zentrum der Diagramme liegen. Die Anzahl der Punkte je Reihe ist dagegen nicht entscheidend, denn die ‘Munsell Value Planes’ besitzen eine asymmetrische Farbverteilung. So weist z. B. die ‘Munsell Value Plane’ 4 mehr Farben im blau-roten Bereich als im gelb-grünen Bereich auf.

Aus Abbildung 4.4 läßt sich ablesen, daß die Farbdifferenzen in den physiologischen Farbräumen, hier stellvertretend durch den Farbraum von de Valois und de Valois, nicht äquidistant bewertet werden. Im *CIELUV* werden aufgrund seiner konischen Gestalt Farbdifferenzen

dunkler Töne schwächer bewertet als Farbdifferenzen heller Töne. Dagegen ist die Gleichabständigkeit der Farben im *CIELAB* und im *MTM*-Farbraum annäherungsweise gegeben. Da die Ergebnisse qualitativ ähnlich sind, kann für die Analyse des Merkmals Farbe innerhalb der Attraktivitätsrepräsentation zwischen diesen beiden Farbräumen gewählt werden. Abbildung 4.5 zeigt graphisch die Verarbeitungskette bis zur Generierung der farbbasierten Merkmalskarten (vgl. [Bollmann 1998]). Die entstehenden Merkmalskarten zeigen die für Farbsysteme wesentliche Eigenschaft der Kategorisierung.

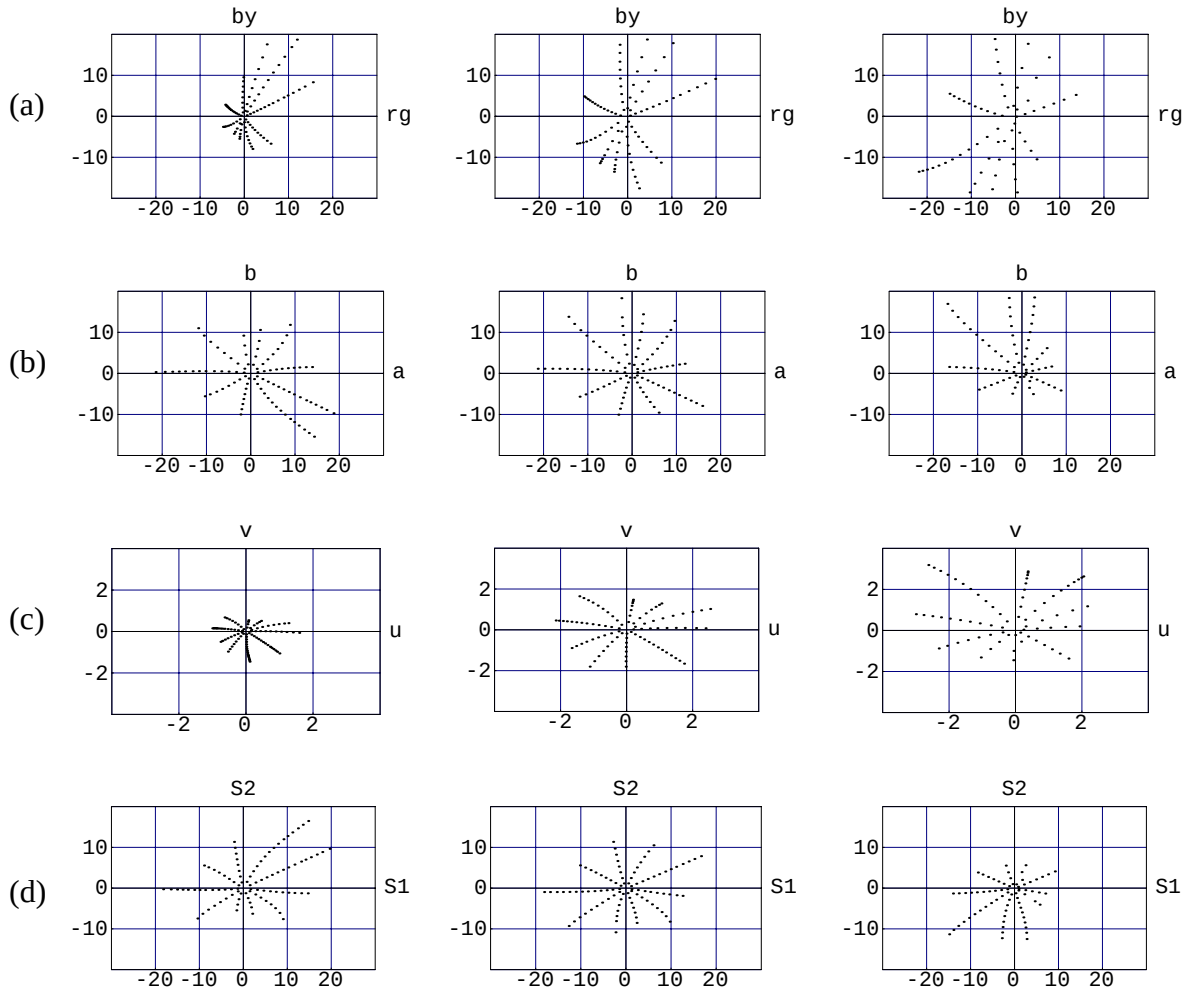


Abbildung 4.4: ‘Munsell Value Planes 4, 6 und 8 im (a) Farbraum von de Valois und de Valois, (b) *CIELAB*, (c) *CIELUV* und (d) *MTM*-Farbraum

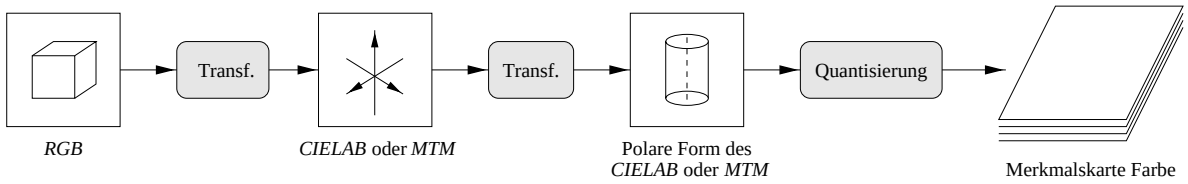


Abbildung 4.5: Verarbeitungskette zur Generierung der farbbasierten Merkmalskarten

Der *MTM*-Farbraum ist in der Bildverarbeitung trotz der erwähnten Eigenschaften bisher noch sehr wenig beachtet, weswegen er im folgenden ausführlicher beschrieben werden soll. Die *RGB*-*MTM*-Transformation basiert auf der Korrektur des Adams-Farbsystems (Adams-Farbsystem siehe [Adams 1942]). Ausgangspunkt der *RGB*-*MTM*-Transformation ist die Berechnung der *XYZ*-Tristimuluswerte aus den *RGB*-Eingangsdaten:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,608 & 0,174 & 0,200 \\ 0,299 & 0,587 & 0,144 \\ 0,000 & 0,066 & 1,112 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{NTSC} \quad . \tag{4.21}$$

Für das Adams-Farbsystem (M_1,M_2,M_3) besteht folgender Zusammenhang zu den *XYZ*-Tristimuluswerten:

$$\begin{aligned} M_1 &= V(X_c) - V(Y)\ , \quad X_c = 1,020\,X \\ M_2 &= 0,4 \cdot (V(Z_c) - V(Y))\ , \quad Z_c = 0,847\,Z \quad . \\ M_3 &= 0,23 \cdot V(Y) \end{aligned} \tag{4.22}$$

$V(x) = 11,6\,x^{1/3} - 1,6$ berücksichtigt hierin die Nichtlinearität des menschlichen visuellen Systems. Der *MTM*-Farbraum (S_1,S_2,L) entsteht schließlich durch die folgende empirisch ermittelte Korrektur des Adams-Farbsystems (Regressionsanalyse über 250 repräsentative Farbmuster, siehe [Miyahara 1988]):

$$\begin{aligned} S_1 &= (8,880 + 0,966 \cdot \cos\varphi) \cdot M_1 \\ S_2 &= (8,025 + 2,558 \cdot \cos\varphi) \cdot M_2 \quad , \\ L &= M_3 \end{aligned} \tag{4.23}$$

wobei $\varphi = \operatorname{atan}(M_1 / M_2)$. Anders als in dem beschriebenen Verfahren zur Generierung einer farbkontrastbasierten Auffälligkeitskarte, in dem die Farbmerkmale in einem orthogonalen Koordinatensystem repräsentiert werden, um die den Segmentierungsprozeß störende Singularität der Unbuntachse zu vermeiden, ist es für die Kategorisierung vorteilhaft, eine psychologisch

motivierte Darstellung in den Termen Farbwinkel, Sättigung und Helligkeit zu verwenden. Daher werden die orthogonalen Koordinaten in ein zylindrischen Koordinatesystem transformiert, in dem anschließend die Quantisierung erfolgt, die zur Generierung der binären Merkmalskarten führt. Die polare Form des *MTM*-Farbraumes, der *HVC*-Farbraum, ergibt sich zu:

$$\begin{aligned} H &= \operatorname{atan}\left(\frac{S_1}{S_2}\right) \\ V &= L \\ C &= \sqrt{S_1^2 + S_2^2} \end{aligned} \quad . \tag{4.24}$$

Auch die Transformation in die *CIE*-Farbräume basiert auf den *XYZ*-Tristimuluswerten. Da die Verwendung dieser Farbräume sehr viel verbreiteter ist, sollen hier nur kurz die Transformationsmatrizen angegeben werden (siehe z. B. [Hunt 1987]). Die Transformation in den *CIELUV* lautet:

$$\begin{aligned} L^* &= 116 \cdot \left(\frac{Y}{Y_w}\right)^{\frac{1}{3}} - 16\,, \quad \frac{Y}{Y_w} > 0{,}008856 \\ L^* &= 903{,}3 \cdot \left(\frac{Y}{Y_w}\right), \quad \frac{Y}{Y_w} \le 0{,}008856 \\ u^* &= 13L^* \cdot (u' - u'_w) \\ v^* &= 13L^* \cdot (v' - v'_w) \end{aligned} \quad . \tag{4.25}$$

Y_w ist hierin der *Y*-Tristimuluswert des Referenzweiß. u' und v' ergeben sich zu:

$$\begin{aligned} u' &= \frac{4X}{X + 15Y + 3Z} \\ v' &= \frac{9Y}{X + 15Y + 3Z} \end{aligned} \quad . \tag{4.26}$$

Die polare Form des *CIELUV*, der $L^*C^*_{uv}h_{uv}$-Farbraum ergibt sich zu:

$$\begin{aligned} C^*_{uv} &= \sqrt{u^{*2} + v^{*2}} \\ h_{uv} &= \operatorname{atan}\left(\frac{v^*}{u^*}\right) \end{aligned} \quad . \tag{4.27}$$

Die *L*-Komponente des *CIELAB* ist identisch mit der des *CIELUV*. Die anderen beiden Komponenten sind optimiert für kleine Farbdifferenzen und lauten:

$$\begin{aligned} a^* &= 500 \cdot \left(\left(\frac{X}{X_w}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_w}\right)^{\frac{1}{3}} \right) \\ b^* &= 500 \cdot \left(\left(\frac{Y}{Y_w}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_w}\right)^{\frac{1}{3}} \right) \end{aligned} \quad . \tag{4.28}$$

Die polare Form des *CIELAB*, der $L^*C^*_{ab}h_{ab}$-Farbraum, berechnet sich analog zu (4.27).

Nach der Koordinatentransformation werden die Farben entsprechend ihrer Lage auf dem Farbkreis den diversen Merkmalskarten zugeordnet. Hierzu wird die Farbtonkomponente *H* bzw. h_{ab} des polaren Farbraumes in äquidistante Winkel (Kreisausschnitte) unterteilt. In meinen Untersuchungen habe ich stets eine variable Unterteilung des Farbkreises in 6-12 Teile verwendet, um die Anzahl der Merkmalskarten überschaubar zu halten. Die Farben lassen sich so auch verbal bezeichnen. Allerdings würde man noch weitere Merkmalskarten für die Sättigung und die Helligkeit benötigen, um eine Zuordnung in ein Farbordnungssystem, wie z. B. das *CNS* (‘*C*olor *N*aming *S*ystem’), komplett durchführen zu können. Die Unterteilungsbereiche der Farbwinkelkoordinate sind in den Tabellen 4.1 und 4.2 zusammengefaßt.

Tabelle 4.1: Unterteilung des Farbkreises zur Generierung von sechs Merkmalskarten

Band	0	1	2	3	4	5
Winkelbereich / °	330-30	30-90	90-150	150-210	210-270	270-330

Tabelle 4.2: Unterteilung des Farbkreises zur Generierung von zwölf Merkmalskarten

Band	0	1	2	3	4	5
Winkelbereich / °	345-15	15-45	45-75	75-105	105-135	135-165
Band	6	7	8	9	10	11
Winkelbereich / °	165-195	195-225	225-255	255-285	285-315	315-345

Jedes Band entspricht einer Farbkarte. Die Reihenfolge der Farben hängt allerdings von dem ausgewählten Farbraum ab. Die Ursache hierfür ist die entgegengesetzte Ausrichtung der S_2-Achse des *MTM*-Farbraumes (positive Werte zeigen in Richtung Blau) gegenüber der Ausrich-

tung der analogen b^*-Achse im *CIELAB* (positive Werte zeigen in Richtung Gelb). Bei der Verteilung der Farbreize auf die sechs binären Merkmalskarten ergibt sich die in Tabelle 4.3 dargestellte Zuordnung.

Tabelle 4.3: Zuordnung zwischen Farben und Merkmalskarten in Abhängigkeit vom Farbraum

	0	1	2	3	4	5
MTM	Rot	Purpur	Blau	Grün	Gelb	Orange
CIELAB	Rot	Orange	Gelb	Grün	Blau	Purpur

Die Benennung der Farben ist allerdings relativ grob. Die Farbausschnitte stimmen zwischen den beiden Farbräumen aufgrund der unterschiedlichen Geometrie nicht exakt überein.

Einschränkend muß man allerdings feststellen, daß das Phänomen der Farbkonstanz in diesem Ansatz unberücksichtigt bleibt.

4.2 Die Auffälligkeitskarten

Die zweite Ebene der in der Einleitung zu diesem Kapitel kurz vorgestellten Attraktivitätsrepräsentation bilden die Auffälligkeitskarten. Sie setzen direkt auf den Ergebnissen der ersten Ebene, den Merkmalskarten, auf. So existiert zu jeder (interdimensionalen) Merkmalskarte genau eine Auffälligkeitskarte, die sich die spezifischen Eigenschaften des zugehörigen Merkmals zu nutze macht (siehe Abbildung 4.1). Im folgenden werden die verschiedenen Auffälligkeitskarten genauer beschrieben.

4.2.1 Auffälligkeitskarte Symmetrie

In der Psychophysik gibt es zahlreiche Untersuchungen, die bestätigen, daß der Mensch in der Lage ist, bilaterale Symmetrien zu detektieren (siehe z. B. [Locher 1987]). Diese Untersuchungen zeigen ebenfalls, daß die Detektion von Spiegelsymmetrien einen lokalen Charakter aufweist. Die Wahrnehmung einer Symmetrie kann sowohl von natürlichen als auch künstlichen Objekten ausgelöst werden. Dabei entsteht der Eindruck einer Symmetrie auch dann, wenn die Objekte bei genauer Vermessung keinesfalls symmetrisch sind. Die Wahrnehmung von Symmetrie unterliegt also einer Generalisierung, die vermuten läßt, daß Symmetrie ein wesentliches Prinzip menschlicher Wahrnehmung ist. Auch in der Bildverarbeitung wird die Symme-

trie bereits über einen längeren Zeitraum als bedeutende Eigenschaft eingestuft ([Brady 1984], [Atallah 1985], [Bigün 1988], [Marola 1989], [Hansen 1992], [Zabrodsky 1995]). Tatsächlich wird sie aber überwiegend genutzt, um Formen zu repräsentieren, zu vereinfachen oder durch symmetrische Templates zu ersetzen. Seit der Entstehung des Forschungsthemas aktives Sehen werden Symmetrieoperatoren allerdings auch zur Detektion interessanter Punkte eingesetzt ([Reisfeld 1995], [Yamamoto 1996], [Heidemann 1998]). Entsprechend werden Symmetrien auch in NAVIS genutzt ([Drüe 1994], [Bollmann 1997], [Mertsching 1997]).

Die Berechnung der Auffälligkeitskarte Symmetrie basiert, wie bereits erwähnt, auf den Ergebnissen der interdimensionalen Merkmalskarte "Orientierte Kanten". Die Filterergebnisse der Merkmalsebene sind im allgemeinen komplex; zur Berechnung der Symmetrie werden hier allerdings nur die Realteile der Filterantworten verwendet, um die Anzahl der detektierten Attraktivitätspunkte zu begrenzen. Für jeden Punkt in jeder der 12 intradimensionalen Merkmalskarten werden die Wertigkeiten in einem schmalen Bereich um einen festen Radius senkrecht zur Orientierung der Kanten aufaddiert. Die Ergebnisse aus allen 12 Merkmalskarten (Orientierungen) werden pixelweise aufaddiert und in einem sogenannten *Radiusbild* abgelegt. Diese Prozedur wird für verschiedene Radien durchgeführt, deren Additionsintervalle aneinandergrenzen, so daß keine Lücken entstehen. Eine Betragsbildung und anschließende Maximumsoperation über die verschiedenen Radiusbilder liefert die Attraktivitätspunkte der Auffälligkeitskarte Symmetrie (Abbildung 4.6).

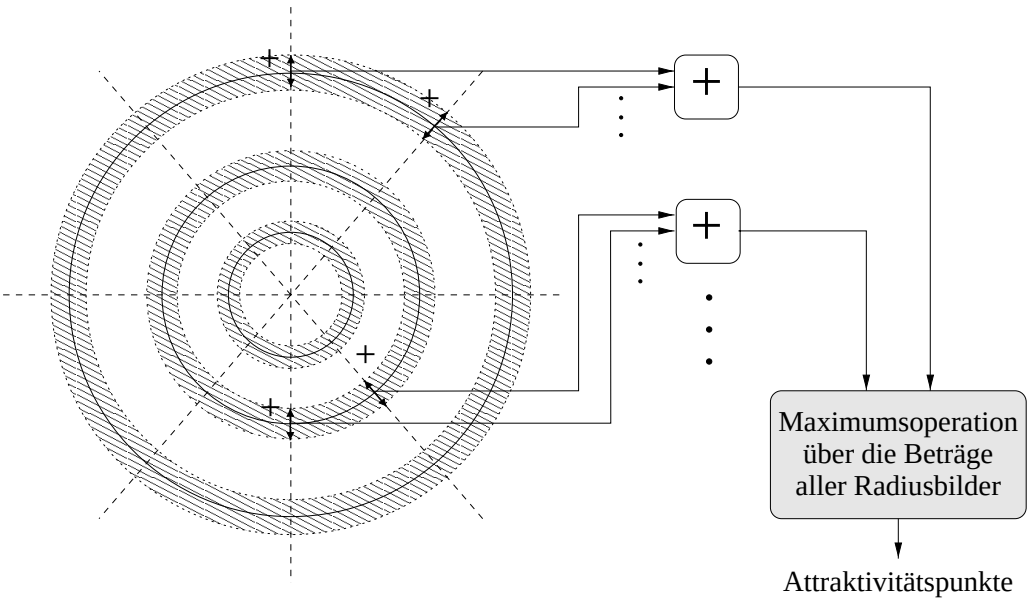


Abbildung 4.6: Detektion lokaler Symmetrien (Erläuterungen siehe Haupttext)

Neben der Position und der Wertigkeit (Auffälligkeit) der Attraktivitätspunkte wird ebenfalls der Radius des Radiusbildes gemerkt, aus dem der jeweilige Punkt hervorgegangen ist. Dieser Radius ist korreliert mit der Größe des symmetrischen Objektes und kann genutzt werden, um zum Beispiel den Radius eines Aufmerksamkeitsfensters für einen variablen FOA festzulegen

oder um die Parametrisierung für eine inhomogene Abtastung (Retina) zu übernehmen. Abbildung 4.7 zeigt vier exemplarische Radiusbilder und die größten lokalen Maxima in der Auffälligkeitskarte Symmetrie für eine einfache Spielzeugszene. Die Anzahl der Radien und damit der Radiusbilder sowie die Größe des Summationsbereichs kann eingestellt werden. Zusätzlich kann ein Offset für den kleinsten Radius bestimmt werden. Für das in Abbildung 4.7a dargestellte 256x256 Pixel große Eingangsbild wurden 10 Radiusbilder mit einem Summationsbereich von jeweils vier Pixeln und einem Offset von null Pixeln für den kleinsten Radius berechnet. Es ergeben sich also Radien der Größe $2 + 4 \cdot i,\ i = \{0, 1, 2, \ldots, 9\}$. Nach ausreichenden empirischen Untersuchungen könnte in späteren Programmversionen die Anzahl der Radien direkt mit der Bildgröße gekoppelt werden. Je größer das Eingangsbild und je kleiner die Summationsbereiche, um so mehr Radiusbilder ergeben sich. Die Größe des Summationsbereichs entspricht in etwa der Breite einer Gaborfilterantwort auf eine optimale Grauwertkante (Differenz von 255 Grauwertstufen) in der hier verwendeten Auflösung des Gaborfiltersatzes.

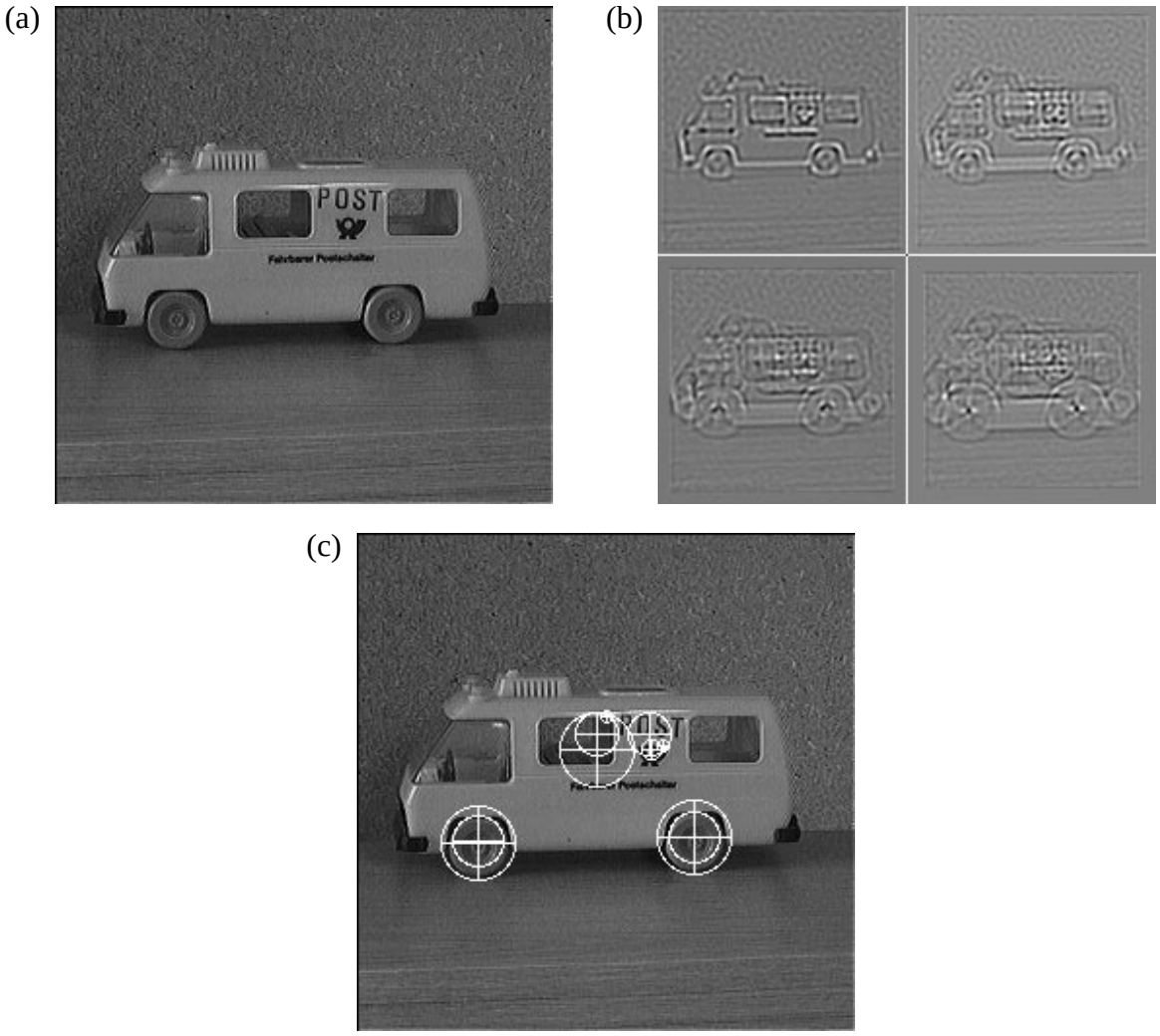


Abbildung 4.7: (a) Eingangsbild, (b) vier exemplarische Radiusbilder (jeweils um den Faktor 2 unterabgetastet dargestellt), (c) Attraktivitätspunkte in der Auffälligkeitskarte Symmetrie (dargestellt sind die 10 Punkte mit der höchsten Wertigkeit)

Die gute Übereinstimmung des Symmetrien bewertenden Teils der Attraktivitätsrepräsentation mit experimentalpsychologischen Untersuchungen soll an dieser Stelle anhand eines aus [Reisfeld 1995] entlehnten Experiments gezeigt werden. Kaufman und Richards [Kaufman 1969] bestimmten die Positionen spontaner Fixationen, wenn Menschen mit sehr einfachen geometrischen Formen konfrontiert werden. Abbildung 4.8 zeigt die Ergebnisse. Die Testpersonen richteten ihre Augen in 86% der Fälle in den gestrichelt eingezeichneten Kreisausschnitt.

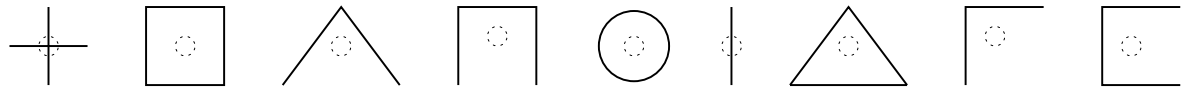


Abbildung 4.8: Fixationsergebnisse in einfachen geometrischen Formen nach Kaufman und Richards

Die in Abbildung 4.9 dargestellten Fixationsergebnisse sind mit der in diesem Kapitel beschriebenen Strategie zur Detektion auffälliger Symmetrien erzeugt worden. Dabei variiert allerdings die Parametrisierung des Verfahrens von Form zu Form. So kann z. B. das geschlossene Quadrat nicht mit der gleichen Summationsbreite wie der schmale senkrechte Strich detektiert werden, wenn die Gesamtzahl der Radiusbilder etwa in der Größenordnung vier bis sechs liegen soll.

Abbildung 4.9: Ergebnisse der symmetriebasierten Komponente der Aufmerksamkeitssteuerung in NAVIS bei Präsentation der Kaufmanschen Formen

In die geometrischen Formen ist jeweils nur der auffälligste Attraktivitätspunkt (bzw. für den senkrechten Strich und das Kreuz die beiden auffälligsten Attraktivitätspunkte) eingezeichnet, um die Ergebnisse möglichst gut mit den Untersuchungen von Kaufman und Richards vergleichen zu können. So werden z. B. für das Quadrat auch Symmetrien in den vier Ecken detektiert, die aber eine geringere Auffälligkeit besitzen als der zentrale Punkt. Die Ergebnisse stimmen für die meisten Formen sehr gut überein. Allerdings werden für das Kreuz die vier ineinanderstoßenden Ecken als Symmetrien detektiert und nicht der Kreuzungspunkt selbst und für den senkrechten Strich liefert das vorgestellte Verfahren nur die Endpunkte und nicht den Mittelpunkt.

Trotz der insgesamt guten Übereinstimmung muß man aber auch sagen, daß diese Ergebnisse nur über eine von Hand eingestellte, optimale Parametrisierung erreicht werden konnten. Ein Ziel bei der Entwicklung von NAVIS ist es dagegen, ein weitgehend autonomes System zu er-

schaffen, das ohne solche Eingriffe von außen auskommt. Anders als in dem dargestellten Experiment ist für die Funktionsfähigkeit von NAVIS und insbesondere für die Erkennung allerdings auch nicht entscheidend, ob die Positionen der Attraktivitätspunkte durch einfaches Betrachten der Ergebnisbilder nachzuvollziehen sind, sondern ob diese Positionen stabil sind, das heißt, ob die gelernten Attraktivitätspunktkonfigurationen von dem Erkennungssystem auch sicher im Bild wiedergefunden werden können (experimentelle Ergebnisse hierzu siehe Abschnitt 5.3). Daher werden die Parameter für die Objekterkennung konstant gehalten, auch wenn so die detektierten Symmetriepunkte dem Betrachter oder der Betrachterin nicht immer optimal erscheinen. Gleiches gilt für die gesamte Parametrisierung der Aufmerksamkeitssteuerung.

4.2.2 Auffälligkeitskarte Exzentrizität

Die Detektion orientierter Kanten ist in der Bildverarbeitung sehr viel weiter verbreitet als die Detektion orientierter Flächen. Ein Grund hierfür könnte die umstrittene biologische Plausibilität der Verarbeitung von Flächenelementen sein. So sind bis in die Schicht V1 des visuellen Cortex bisher keine Zellen gefunden worden, deren Reizmuster geschlossene Flächen repräsentieren. Es existiert in der ‘Computer Vision’-Forschung z. T. sogar die Vorstellung, daß Flächen nur über ihre Grenzen, also die Kanten, definiert sind. Für eine solche Vorstellung spricht zum Beispiel, daß die Farbe einer Fläche, die das gesamte visuelle Feld einnimmt, nicht bestimmt werden kann (in der Fliegerei wird dieser Effekt, der unter blauem, wolkenlosen Himmel auftritt, ‘Graying out’ genannt) [Gouras 1981].

In unserem Modell wird allerdings nicht etwa die Berechnung primitiver Merkmale, wie sie aus der ‘Early Vision’-Theorie bekannt sind, angestrebt; vielmehr sollen die implementierten Merkmalskarten komplexere Merkmale enthalten, auf deren Basis die Selektion eines Targets möglich ist. Eine solche Betrachtungsweise ist konform mit den psychophysischen Ergebnissen, die gezeigt haben, daß die visuelle Suche auf einer relativ späten Repräsentation grundlegender Merkmale basiert (vgl. Abschnitt 3.2.1). Aus den ‘Visual Search’-Experimenten geht auch hervor, daß die Farbe und die Orientierung zu den wesentlichsten Stimuluseigenschaften gehören. Es liegt daher nahe, die Auffälligkeit der orientierten Flächen nach der Ausprägung ihrer Orientierung zu bemessen. Ein solches Maß liefert die Berechnung der Exzentrizität ε [Jähne 1997]:

$$\varepsilon = \frac{(m_{2,0} - m_{0,2})^2 + 4m_{1,1}^2}{(m_{2,0} + m_{0,2})^2} \tag{4.29}$$

Die Momente *m* sind gemäß Gleichung (4.10) definiert.

Abbildung 4.10 zeigt das Ergebnis des Segmentierungprozesses, die mittels Hauptachsentransformation erzeugten Merkmalskarten sowie die Auffälligkeitskarte Exzentrizität für die in Abbildung 4.7a dargestellte Spielzeugszene. Aufgeführt sind hier nur die vier Merkmalskarten, die Segmente enthalten: links oben Orientierung [0^o,15^o), rechts oben Orientierung [15^o,30^o), links unten Orientierung [45^o,60^o), rechts unten Orientierung [90^o,105^o). Die Attraktivitätspunkte in der Auffälligkeitskarte entsprechen den Schwerpunkten der Segmente. Zur Verdeutlichung der Funktionsweise der Hauptachsentransformation sind in Abbildung 4.11 die Transformationsergebnisse für eine künstliche Szene bestehend aus orientierten Balken und einem orientierungsfreien Quadrat dargestellt. In diesem Fall ist jede der 13 Merkmalskarte mit genau einem Objekt besetzt.

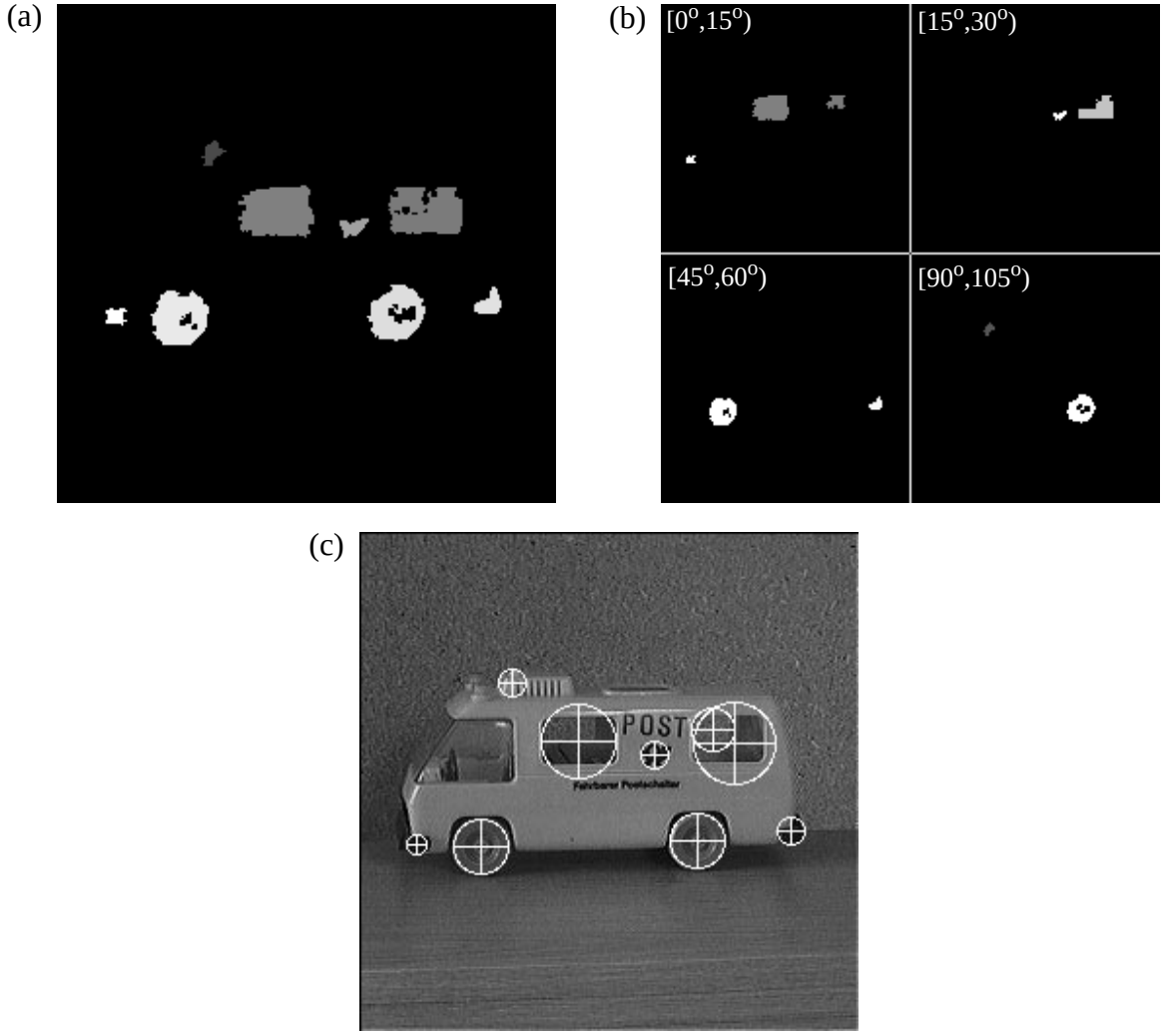


Abbildung 4.10: (a) Segmentierungsergebnis für die Spielzeugszene aus Abbildung 4.7a, (b) Ergebnis der Hauptachsentransformation (dargestellt sind nur die Karten, die Daten enthalten) und (c) Auffälligkeitskarte Exzentrizität

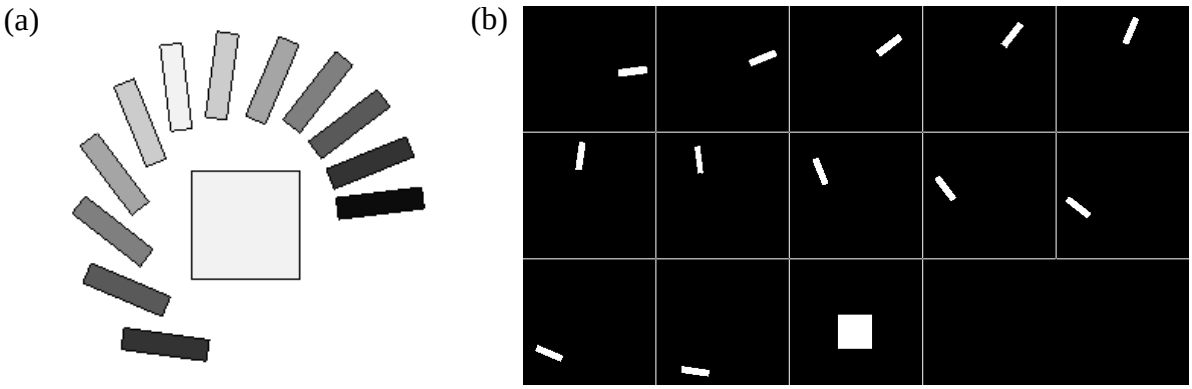


Abbildung 4.11: (a) Eingangsbild, (b) Ergebnis der Hauptachsentransformation

4.2.3 Auffälligkeitskarte Farbkontrast

Diese Auffälligkeitskarte bewertet den Farbkontrast zwischen den aus einer Farbsegmentierung hervorgegangenen Segmenten und ihren Nachbarn und markiert die Schwerpunkte besonders kontrastreicher Flächen als potentielle Aufmerksamkeitspunkte. Die erste Stufe zur Berechnung der Auffälligkeitskarte stellt die Segmentierung des Eingangsbildes dar. Eine solche farbbasierte Segmentierung ermöglicht eine gegenüber der luminanzbasierten Segmentierung verbesserte Objekt-Hintergrund-Trennung. So können z. B. Objekte mit isoluminanten Kanten nur in Farbbildern detektiert werden (siehe auch [Bollmann 1996]). Das Ziel dieser Verarbeitungsstufe ist eine Segmentierung, die mit der menschlichen Farbempfindung möglichst gut übereinstimmt; hierzu ist ein geeigneter Farbraum auszuwählen. Die Generierung der Auffälligkeitskarte basiert auf der Transformation in denselben gleichabständigen Farbraum, der auch die Grundlage für die Generierung der Merkmalskarten bildet (*MTM* oder *CIELAB*). Für die Segmentierung ist eine orthogonale Repräsentation der Farbkoordinaten besser geeignet, so daß auf die Überführung in ein zylindrisches Koordinatensystem hier verzichtet wird.

Bei dem implementierten Farbsegmentierungsalgorithmus handelt es sich um ein auf einer zentroiden Verkettung basierendes Bereichswachstumsverfahren. Die Grundidee der zentroiden Verkettung besteht darin, den (Farb-)Wert des aktuell untersuchten Pixels mit den Mittelwerten der benachbarten Segmente zu vergleichen. Liegt die Farbdifferenz zwischen dem betrachteten Bildpunkt und mehr als einem Mittelwert unterhalb einer dynamischen Schwelle, wird der Bildpunkt dem ähnlichsten Segment zugeordnet und dessen Mittelwert aktualisiert. Erfüllt keines der Segmente das Ähnlichkeitskriterium, so wird ein neues Segment generiert. Als Farbähnlichkeitskriterium dient die Euklidische Distanz. Die Höhe des Schwellwertes hängt ab von der Varianz der Farbvektoren in der unmittelbaren Nachbarschaft des betrachteten Bildpunktes. Eine solche Parametrisierung verhindert eine Übersegmentierung in stark texturierten Bildbereichen, wie sie in Outdoor-Szenen häufig vorkommen. Um Verkettungsfehler zu vermindern, die dadurch entstehen, daß das Verfahren streng sequentiell arbeitet, wird die Richtung,

in der die Bildzeilen bearbeitet werden, alterniert. Alle Parameter, die ein Segment charakterisieren, wie der Mittelwert der Farbkanäle, die Varianz, die Segmentgröße und der Schwerpunkt, werden während des Segmentierungsprozesses berechnet. Ebenso werden die für die Berechnung des Farbkontrastes benötigten Informationen über Nachbarschaftsbeziehungen zwischen Segmenten und die Länge ihrer gemeinsamen Grenze ermittelt. Die Auffälligkeit eines Segments wird berechnet als der mittlere Farbgradient entlang der Grenze zu den Nachbarsegmenten, wobei der Farbgradient zwischen zwei Segmenten als der Euklidische Abstand der Farbmittelwerte modelliert ist. Formal ergibt sich der Farbkontrast F_i des i-ten Segmentes zu:

$$F_i = \frac{1}{U_i} \cdot \sum_{j \in B_i} b_{ij} \cdot d(\langle C_i \rangle, \langle C_j \rangle) \quad , \quad U_i = \sum_{j \in B_i} b_{ij} \tag{4.30}$$

wobei U_i der Umfang des *i*-ten Segmentes, B_i die Menge der Indizes aller Nachbarsegmente des Segments *i*, b_{ij} die Länge der gemeinsamen Grenze zwischen den Segmenten *i* und *j* sowie $d(\langle C_i \rangle, \langle C_j \rangle)$ der Euklidische Abstand zwischen den Farbmittelwerten $\langle C \rangle$ der Segmente *i* und *j* (im ausgewählten Farbraum) ist. Der auf diese Weise ermittelte Wert ist um so größer, je mehr zwei Farbreize im Farbraum voneinander entfernt sind. Somit weisen die Komplementärfarben den größten modellierten Farbkontrast auf; dies ist auch aus psychologischer Sicht plausibel.

Abbildung 4.12b zeigt die farbbasierten Merkmalskarten für eine beliebig ausgewählte Verkehrsszene (Abbildung 4.12a). Dargestellt sind nur die Merkmalskarten, die auch Daten enthalten. Die folgenden Abbildungen 4.12c und 4.12d zeigen das Ergebnis der Farbsegmentierung beziehungsweise die aus dem Farbkontraste bewertenden Verfahren gewonnenen Attraktivitätspunkte. Verwendet wurde hier der *MTM*-Farbraum.

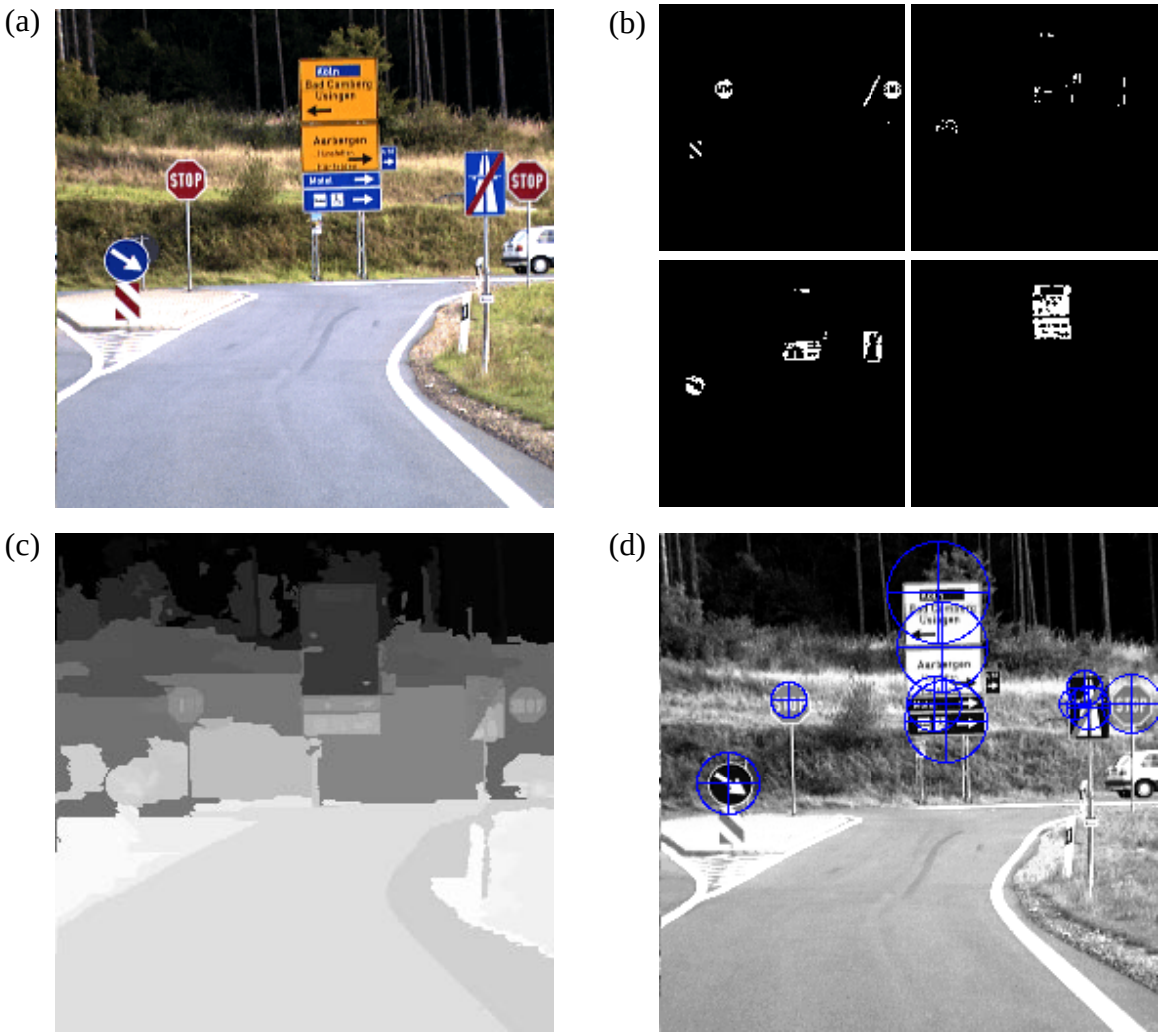


Abbildung 4.12: (a) RGB-Eingangsbild; (b) farbbasierte Merkmalskarten Rot (oben links), Purpur (oben rechts), Blau (unten links) und Orange (unten rechts), die Merkmalskarten Grün und Gelb enthalten keine Daten und sind daher nicht dargestellt; (c) Ergebnis der Segmentierung; (d) Auffälligkeitskarte Farbkontrast

4.3 Das Binding-Problem und konkurrierende Aufmerksamkeit

In der optischen Mustererkennung besteht das *Binding-Problem* in der richtigen Zuordnung separat gewonnener Merkmale zu den im Bild befindlichen Objekten. Unter den restriktiven Annahmen, daß

- ein Merkmal eindeutig ist, d. h. ein Merkmal genau einer Kategorie (Merkmalskarte) zugesprochen werden kann

- ein Objekt homogen ist, d. h. ein Merkmal das gesamte Objekt charakterisiert
- jedes Objekt getrennt fokussiert werden kann, d. h. umfassend vom Hintergrund getrennt werden kann

läßt sich das Binding-Problem einfach durch die jeweilige Fokussierung eines Objekts mit anschließender ODER-Verknüpfung der binarisierten Merkmalskarten lösen. Die Position des Objektes kann dann z. B. durch eine Schwerpunktsberechnung bestimmt werden.

In der in diesem Kapitel vorgestellten Attraktivitätsrepräsentation ist die erste Bedingung insofern durch die Einführung der Auffälligkeitskarten erfüllt, als daß hier die intradimensionalen Merkmalskarten verknüpft werden, so daß letztendlich nur eine Karte je interdimensionalem Merkmal bestehen bleibt. Damit sind Mehrdeutigkeiten auf der Ebene der Auffälligkeitskarten, und die gilt es in diesem speziellen Fall zu verknüpfen, ausgeschlossen. Allerdings steht auf der Ebene der Auffälligkeitskarten nicht mehr die gesamte Information zur Verfügung, die auf der Ebene der Merkmalskarten gewonnen wurde. So läßt sich zum Beispiel aus einem in die entsprechende Auffälligkeitskarte eingetragenen Farbkontrast nicht mehr bestimmen, welche Farbe das zugehörige Objekt besitzt.

Die zweite Bedingung, die Annahme der Homogenität der Objekte, ist in realen Szenen kaum erfüllt. In der Regel besitzen Objekte durchaus eine auffällige Struktur oder verteilte Farbgebung, so daß sie sich in Teilobjekte untergliedern lassen. Insbesondere ist zu beachten, daß die Berücksichtigung von Symmetrien zu Auffälligkeitspunkten zwischen symmetrisch angeordneten Objekten führen kann. Abbildung 4.13 zeigt ein Beispiel, das die unterschiedlichen Charakteristiken des Exzentrizitäten bewertenden und des Symmetrien bewertenden Verfahrens deutlich machen soll. Während das Exzentrizitäten bewertende Verfahren, das ja bekanntlich auf einer Segmentierung im Grauwertbild basiert, Flächenschwerpunkte als Fokuspunkte liefert, ist das Symmetrien bewertende Verfahren in der Lage, Fokuspunkte außerhalb von Objektgrenzen zu finden.

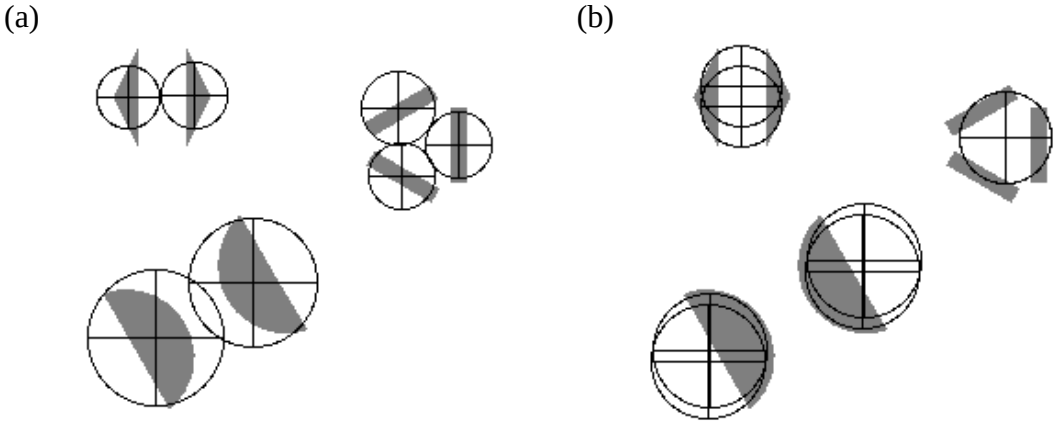


Abbildung 4.13: Attraktivitätspunkte berechnet mit dem (a) Exzentrizitäten bewertenden Verfahren, (b) Symmetrien bewertenden Verfahren

Auch die dritte Bedingung, die einfache Lösung des Objektes vom Hintergrund, kann in realen Szenen nicht garantiert werden. Objekte können durchaus Merkmale mit ihrem Hintergrund teilen, zudem können sich Objekte teilweise gegenseitig verdecken.

Festzustellen bleibt, daß das Binding-Problem für reale Szenen nicht durch einfache ODER-Verknüpfung der Merkmalskarten im 2D-Raum gelöst werden kann. In einem ersten Schritt werden daher in dieser Arbeit die Einträge in den verschiedenen Auffälligkeitskarten nicht miteinander verrechnet, sondern lediglich in der Attraktivitätskarte zu einer gemeinsamen Liste verknüpft (siehe Abbildung 4.1 sowie Abschnitt 4.3.1). Ein intelligenteres 'Feature Binding' auf der Basis von Tiefeninformationen wird in dem Fortsetzungsprojekt ESAB-2 untersucht. Hierin sollen die generierten Attraktivitätspunkte 3D-Raumausschnitten zugeordnet werden und abhängig von ihrer räumlichen Nähe, in der Bildebene und der Tiefe, zusammengefaßt werden.

4.3.1 Die Attraktivitätskarte

Der Sinn der Attraktivitätskarte liegt in der Zusammenführung der drei beschriebenen Auffälligkeitskarten, die vergleichbare Daten unter Berücksichtigung verschiedener auffälliger Szenenmerkmale liefern. Die Datenstrukturen, die von uns Attraktivitätspunkte genannt werden, beinhalten Informationen über:

- die Art der verwendeten Strategie (Typattribut),
- die Größe,
- die horizontale und
- die vertikale Position sowie
- die Wertigkeit der Auffälligkeit (Attraktivität)

eines lokalen Ausschnitts. Jeder Attraktivitätspunkt kann somit durch eine aus fünf Feldern bestehende Datenstruktur repräsentiert werden. Deutlichste Unterscheidungskriterien der Attraktivitätspunkte untereinander stellen das Typattribut und die Größeninformation dar. Die Wertigkeit kann für einen bestimmten Punkt innerhalb mehrerer Aufnahmen etwas variieren, weswegen sie für einen Punkt nicht zu dessen Identifizierung geeignet ist. Letztendlich sind die horizontale und vertikale Position aufgrund möglicher Translationen für einen Attraktivitätspunkt uncharakteristisch. Für eine Menge von Punkten dagegen sind die Distanzen der Punkte untereinander charakteristisch, was die grundlegende Annahme für die in Abschnitt 5.3 beschriebene Erkennungsstrategie ist. Der Aufbau der Attraktivitätskarte konnte gegenüber [Götze 1995] nicht verändert werden, da das Erkennungsmodul der 2D-Objekterkennung eine Karte in der vorgestellten Form erwartet.

Es bleibt festzuhalten, daß die vorgestellten Auffälligkeitskarten sowie die Attraktivitätskarte letztendlich verkettete Punktlisten sind, die aber die Charakteristiken von Karten besitzen, da sie die räumliche Lage der Attraktivitätspunkte enthalten und jederzeit dem Eingangsbild überlagert werden können. Abbildung 4.14 zeigt ein weiteres Beispiel. Das Eingangsbild entspricht Abbildung 4.12a. Die Attraktivitätspunkte sind entsprechend ihres Typattributes unterschiedlich eingefärbt: Cyan markiert das Symmetrien bewertende Verfahren, Magenta das Exzentrizitäten bewertende Verfahren und Gelb das Farbkontraste bewertende Verfahren. Es sei darauf hingewiesen, daß die Größe der Kreuze nicht etwa die Attraktivität der Punkte bezeichnet, sondern die Größe des Szenenausschnitts, auf den sie sich beziehen.

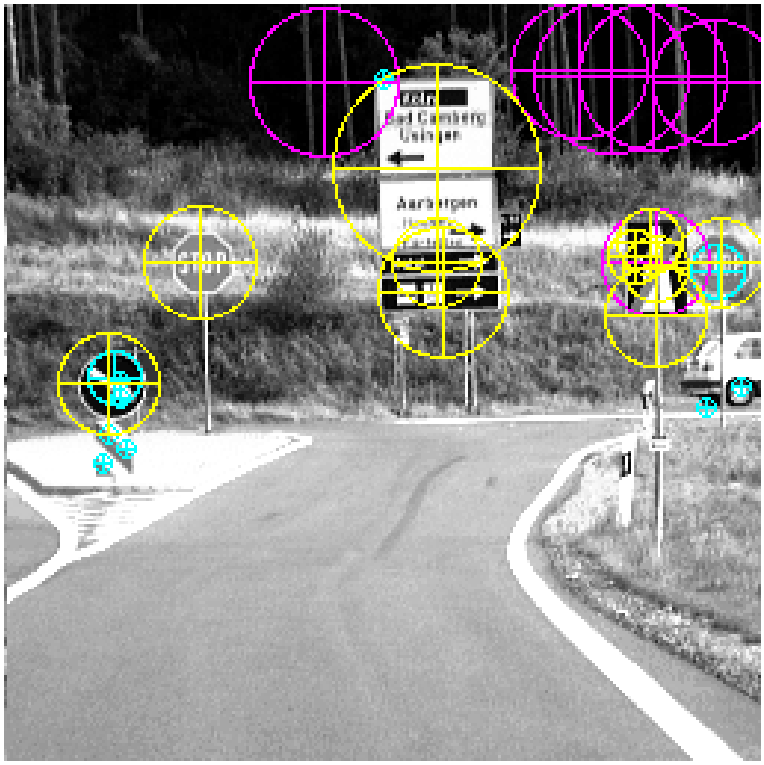


Abbildung 4.14: Attraktivitätskarte bei Präsentation des RGB-Eingangsbildes aus Abbildung 4.12a (dargestellt sind nur die Attraktivitätspunkte mit der höchsten Wertigkeit)

Aus Abbildung 4.14 ist unmittelbar ersichtlich, daß das Farbkontraste und das Exzentrizitäten bewertende Verfahren sehr unterschiedliche Attraktivitätspunkte liefern, obwohl beide auf einem Bereichswachstumsverfahren basieren. Während das erste Verfahren die farbigen Verkehrsschilder markiert, detektiert das zweite besonders die länglichen Strukturen im Hintergrund, die durch die Baumstämme entstehen. Das Symmetrien bewertende Verfahren markiert ebenfalls einige der Verkehrsschilder sowie den symmetrischen Autoreifen, der keine auffällige Farbgebung besitzt.

Vor der Integration der drei Auffälligkeitskarten ist natürlich auch noch eine unterschiedliche Gewichtung dieser Karten möglich. Eine solche Gewichtung macht dann Sinn, wenn die Aufgabe des aktiven Sehsystems z. B. die visuelle Suche nach einem bekannten Target ist, das bestimmte Eigenschaften, z. B. eine auffällige Symmetrie, besitzt.

4.3.2 Die Aufmerksamkeitskarte

Aus der Aufmerksamkeitskarte wird in jedem Zyklus (Zeitraum zwischen zwei Kameraschwenks) genau ein Attraktivitätspunkt als Aufmerksamkeitspunkt ausgewählt und der Motorsteuerung nach der Umrechnung in einen Schwenk- und einen Neigewinkel übergeben. Dieser Punkt dient der Erkennung als Index. Die Aufmerksamkeitskarte enthält aber nicht nur den Aufmerksamkeitspunkt selbst, sondern auch alle anderen Attraktivitätspunkte, die mit den Punkten aus der Hemmungs- und der Verstärkungskarte verrechnet wurden (vgl. Abbildung 4.1). Die Inhibitionskarte dient dazu, bereits untersuchte Punkte in ihrer Wertigkeit zu verringern, so daß diese kein zweites Mal innerhalb weniger Zyklen ausgewählt werden. Hierdurch werden Blickwechsel zu bisher nicht untersuchten Punkten erleichtert. Die Verstärkungskarte sorgt dagegen dafür, daß während der Validierung einer Objekthypothese die zu dem betreffenden Objekt gelernten Punkte bevorzugt als Aufmerksamkeitspunkte ausgewählt werden. Damit wird gewährleistet, daß alle gelernten Teilformen des betreffenden Objektes auf ihre Anwesenheit im Bild untersucht werden. Ein "Überspringen" auf andere Objekte wird dagegen erschwert (Details zum Lernen und Erkennen von Objekten siehe Abschnitt 5.3.).

Für eine weitergehende Betrachtung bietet sich die Formalisierung der verwendeten Begriffe, wie in [Götze 1995] beschrieben, an. Eine Menge von Attraktivitätspunkten wird im folgenden mit *AT* bezeichnet, wobei ein Element a aus *AT* ein Fünf-Tupel aus Typattribut, Größe, horizontaler und vertikaler Position sowie Wertigkeit ist, d. h. $a = (a_t, a_g, a_h, a_v, a_w)$. Die horizontale und die vertikale Position beziehen sich hierbei auf Koordinaten innerhalb der Bilder, die jeweils in die Merkmals- und Auffälligkeitsberechnung eingehen (theoretisch ist es möglich, die verschiedenen Merkmale in verschieden unterabgetasteten Kamerabildern zu berechnen). Es gilt für $a \in AT$: $a_w \in [0,1]$, wobei der Wert 1 eine hohe Attraktivität darstellt. Ebenso werden die Elemente der Inhibitionskarte in der Menge *IK* und die der Verstärkungskarte in *VK* zusammengefaßt. Auch hier sind die Werte im Bereich [0,1], wobei der Wert 0 eine niedrige, der Wert 1 eine hohe Hemmung bzw. Verstärkung darstellt.

Eine Bewertung der Nähe zweier beliebiger Elemente a_k und a_l ist durch eine Distanzfunktion $d_i(a_k,a_l) : AT \cup IK \cup VK \times AT \cup IK \cup VK \rightarrow [0,1]$ gegeben, wobei der Wert 1 eine geringe räumliche Distanz, ähnliche Größe und gleiches Typattribut beschreibt, der Wert 0 entsprechend umgekehrte Verhältnisse. Insbesondere existiert also auch die Möglichkeit, Attraktivitätspunkte mit Elementen aus der Verstärkungskarte oder der Inhibitionskarte zu vergleichen. Es werden verschieden abgestufte Distanzfunktionen d_i verwendet, die aber alle prinzipiell folgende Form haben:

$$d_i(a_k, a_l) = \exp\left(-\frac{\Delta a_h^2 - \Delta a_v^2}{2c_i}\right), \quad i = 1, 2, \ldots \tag{4.31}$$

Der Parameter c_i variiert zwischen den verschiedenen Distanzfunktionen. Δa_h und Δa_v geben den räumlichen Abstand zwischen den Attraktivitätspunkten a_k und a_l in horizontaler bzw. vertikaler Richtung an. Es gilt also $\Delta a_h = a_{hk} - a_{hl}$ bzw. $\Delta a_v = a_{vk} - a_{vl}$.

Wie wird nun aus den gegebenen Karten ein Aufmerksamkeitspunkt bestimmt? Einerseits sollen bereits untersuchte Attraktivitäten durch Punkte der Inhibitionskarte gehemmt werden, andererseits soll eine Verstärkung bestimmter Punkte durch die Verstärkungskarte erfolgen. Aus der Menge AT wird also eine neue Menge AT' gebildet, in dem jedes Element $a \in AT$ durch Änderung der Wertigkeit a_w in das Element $a' \in AT'$ überführt wird:

$$a_w' = a_w \cdot \prod_{b \in IK} (1 - b_w \cdot d_1(a, b)) \cdot \prod_{c \in VK} (1 + c_w \cdot d_2(a, c)) \tag{4.32}$$

Für den zur Distanzfunktion d_1 gehörenden Parameter c_1 (vgl. (4.31)) hat sich ein Wert von 10 als geeignet gezeigt; c_2 kann zu 1,5 gesetzt werden. Aus obiger Gleichung ist ersichtlich, daß der Einfluß eines Inhibitionspunktes $b \in IK$ auf den jeweiligen Attraktivitätspunkt durch Multiplikation mit einer spezifischen Distanzfunktion d_1 von der Entfernung abhängig gemacht wird. Je weiter die Punkte voneinander entfernt sind (und je geringer der Inhibitionswert), desto weniger wird der Attraktivitätswert a_w unterdrückt. Weiterhin ist anzumerken, daß die Distanzfunktion Punkte mit unterschiedlichem Typattribut nicht bewertet. Entsprechendes gilt auch für die Punkte der Verstärkungskarte *VK*, nur daß hier eine Verstärkung statt einer Unterdrükkung mit einer typspezifischen Distanzfunktion d_2 vorgenommen wird.

Zwei nahe zusammenliegende Elemente a_k' und a_l' aus AT' mit unterschiedlichen Typattributen können sich auf dasselbe Szenendetail beziehen, so daß es Sinn macht, die Punkte, die eine dicht mit Attraktivitäten besetzte Region kennzeichnen, in ihrer Wertigkeit zu erhöhen. Dies ist in Gleichung (4.33) durch Bildung einer neuen Menge AT'' aus AT' dargestellt.

$$a_{wk}'' = a_{wk}' + \sum_{l \neq k} a'_{wl} \cdot d_3(a_k', a_l') \tag{4.33}$$

Der Parameter c_3 der Distanzfunktion d_3 ist ebenfalls zu 1,5 gesetzt. Auch hier bestimmt die Entfernung der Punkte den Grad der gegenseitigen Beeinflussung, jedoch ist nun der Zusammenhang additiv. Zudem werden bei der Distanzfunktion d_3 die Typattribute nicht berücksichtigt. Schließlich ergibt das Element aus AT'' mit der höchsten Wertigkeit den gesuchten Aufmerksamkeitspunkt. Abbildung 4.15 zeigt die Werteänderungen in der Aufmerksamkeitskarte und die zugehörigen Einträge in der Inhibitionskarte bei statischer Kamera und unbesetzter Verstärkungskarte. Die Farbigkeit der Punkte in der Aufmerksamkeitskarte ist proportional zu

ihrer Wertigkeit (Schwarz entspricht einer Wertigkeit nahe Null). Cyan repräsentiert erneut das Symmetrien bewertende Verfahren und Magenta das Exzentrizitäten bewertende Verfahren.

Aufmerksamkeitskarte: Inhibitionskarte:

Abbildung 4.15: Werteänderungen in der Aufmerksamkeitskarte und die korrespondierenden Einträge in der Inhibitionskarte (zeilenweise von links nach rechts)

4.4 Kopplung von Aufmerksamkeitssteuerung und Kamerakopf

Dieser Abschnitt beinhaltet die Kopplung der Aufmerksamkeitssteuerung mit dem binokularen Kamerakopf des aktiven Sehsystems NAVIS. Die in dieser Arbeit verwendeten Bilder werden derzeit allerdings nur von einer Kamera des binokularen Kopfes geliefert. Diese Kamera bildet zusammen mit dem Träger das in drei Freiheitsgraden bewegliche Kamerasystem. Ziel der Kamerasteuerung ist es, den jeweils aktuellen Aufmerksamkeitspunkt in die Bildmitte abzubilden. Hierfür wurde in [Götze 1995] ein einfaches Modell mit Soll-Istwert-Vergleich entworfen, das auch in dieser Arbeit eingesetzt wird.

Eine Skizze des Kamerasystems ist in Abbildung 4.16 gegeben. Ausgehend von der dargestellten Initialposition dreht das eingezeichnete Gelenk 1 die Kamera in der Horizontalebene, Gelenk 3 in der Vertikal- und Gelenk 2 wiederum in der Horizontalebene. Die Kamera ist dabei so mit dem Vektor ***vk1*** fest verbunden, daß dieser stets in der optischen Achse liegt. Ausgehend vom Gelenk 1 zeigt die Spitze des Vektors auf die Apertur der Kamera, wobei im zugrundegelegten Lochkameramodell die Apertur mit der Lochöffnung identisch ist. Durch die mechanische Hintereinanderschaltung der Kamerakopfgelenke ist insbesondere die Aperturposition im Raum von allen drei Gelenkwinkelstellungen *wk*1, *wk*2 und *wk*3 abhängig.

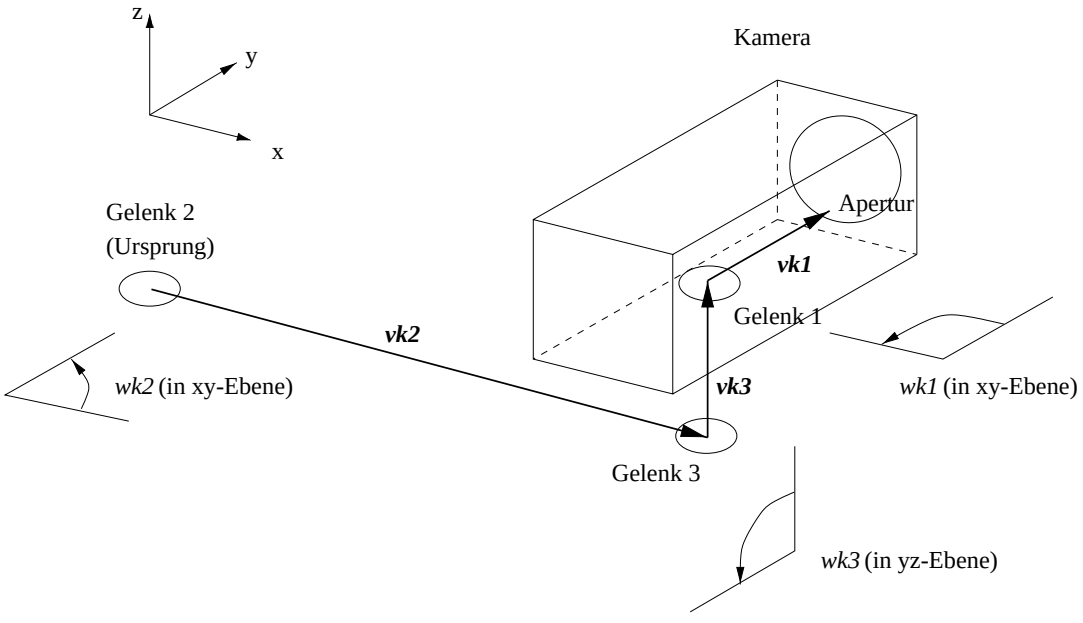


Abbildung 4.16: 3D-Modell des Kamerasystems (aus [Götze 1995])

Für die Beschreibung der Abbildung der realen Szene auf den Kamerasensor sei im folgenden die Helligkeitsverteilung der im Raum vorhandenen Punkte

$$S : (x, y, z) \rightarrow \Re \ , \quad (x, y, z) \in \Re^3 \tag{4.34}$$

bekannt, wobei (*x*,*y*,*z*) einen Punkt der Szene im Raumkoordinatensystem mit Ursprung im Gelenk 2 bezeichnet (siehe Abbildung 4.16). Die Abbildung von *S* auf die Grauwertverteilung der Bildpunkte

$$B : (h, v) \rightarrow H \ , \quad (h, v) \in \aleph^2 \tag{4.35}$$

ergibt sich aus einer Abbildung *Q*, welche die Position der Kamera sowie die Quantisierung der Helligkeit berücksichtigt:

$$B \ = \ Q\,(S, wk1, wk2, wk3) \tag{4.36}$$

Die endlichen, diskreten Mengen *H* und $\aleph^2$ repräsentieren dabei die Menge aller Grauwerte bzw. Positionen im Kamerabild. Das Koordinatensystem ist so gewählt, daß im aufgenommenen Bild die linke, obere Ecke den Punkt (0,0) repräsentiert und die Bildkoordinaten horizontal nach rechts bzw. vertikal nach unten aufsteigen. In Abbildung 4.17 ist eine 2D-Aufsicht auf das Kamerasystem dargestellt; insbesondere sieht man hier die geometrische Abbildung eines Objektpunktes auf das CCD-Element der Kamera im Rahmen eines Lochkameramodells. Aufgrund der endlichen Ausdehnung des Kamerasensors kann nicht die gesamte Szene *S*, sondern nur ein Ausschnitt um die optische Achse (Verlängerung von ***vk1*** in der Abbildung) in die Aufnahme *B* abgebildet werden.

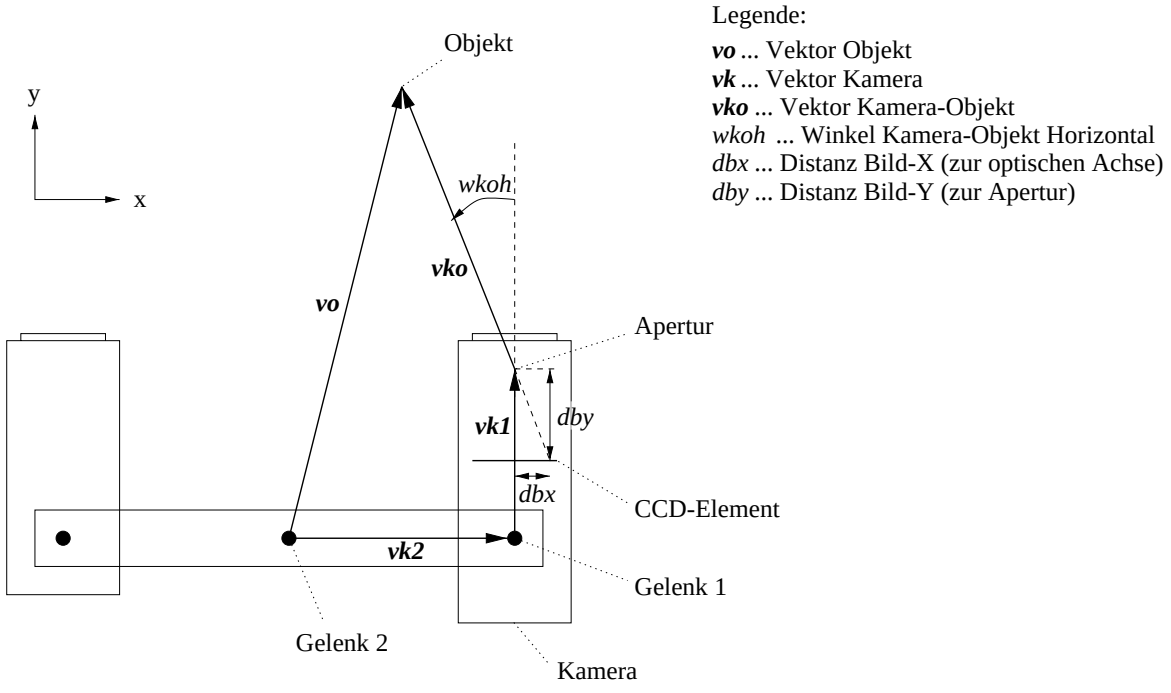


Abbildung 4.17: 2D-Modell des Kamerasystems (aus [Götze 1995])

In Abbildung 4.17 ist gezeigt, wie ein Objektpunkt geometrisch abgebildet wird. Die zentrale Frage ist nun, wie die Kamerakopfwinkel verändert werden müssen, damit ein bestimmter Bildpunkt (*h*,*v*) bei der nächsten Aufnahme in der Bildmitte liegt. Mit Hilfe des Abstandes *dbx*, welcher ja mit der Distanz des Fovealisierungspunktes *f* zum Bildmittelpunkt identisch ist, und des in Bildpunkten angegebenen Abstandes *dby* der Sensorebene zur Apertur, welcher durch Messungen bestimmt werden kann, ergibt sich für den Einfallswinkel *wkoh* des Objektpunktes auf die Kamera:

$$wkoh = \text{atan}\left(\frac{dbx}{dby}\right) \tag{4.37}$$

Das Objekt könnte also theoretisch durch die Drehung der Kamera im Aperturpunkt um den Winkel *wkoh* fovealisiert werden. Tatsächlich ist aber nur eine Drehung in der Horizontalebene um Gelenk 1 bzw. Gelenk 2 möglich. Da diese vom Aperturpunkt entfernt sind, würde eine Drehung um den Winkel *wkoh* beispielsweise im Gelenk 1 zu einer Fehlfovealisierung führen, d. h. daß *f* nach der Drehung nicht exakt im Bildmittelpunkt liegen würde. Das Problem ist in Abbildung 4.18 deutlich zu sehen. Hierbei ist der Kamerakopf in den Gelenken 1 und 2 um den Winkel *wk*1 und *w*k2 gedreht. Der Einfallswinkel *wkoh*, der sich aus Gleichung (4.37) ergibt, ist aber um den Winkel *w*1 größer als der gesamte Stellwinkel *w*0 = *wk*1 + *wk*2.

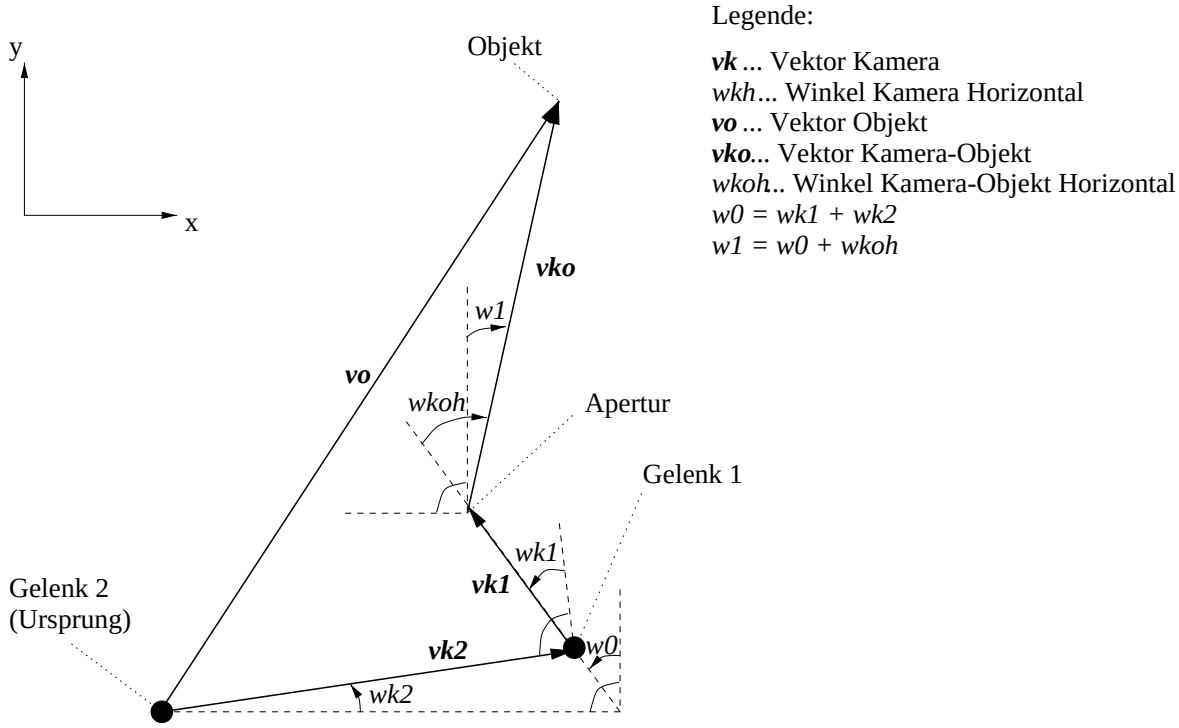


Abbildung 4.18: 2D-Modell zur Erläuterung der Fehlfovealisierung (aus [Götze 1995])

Durch die Verwendung der Einfallswinkel als Stellwinkel kann entweder nur das Gelenk 1 oder nur das Gelenk 2 als Drehachse in der Horizontalebene fungieren. Eine Kombination beider Gelenkbewegungen würde dazu führen, daß die Bewegung um Gelenk 3 nicht mehr ausschließlich in der Vertikalebene stattfände. Die Gelenk-2-Fehler sind aber wegen der in einem eher horizontal ausgedehnten Aktionsraum kleineren Vertikalbewegungen der Kamera kleiner als die Gelenk-1-Fehler. Daher wird in dieser Arbeit Gelenk 2 und nicht Gelenk 1 bewegt.

Feinpositionierung

Die Feinpositionierung dient dazu, die Abweichungen des Fovealisierungspunktes vom Bildmittelpunkt aufgrund der Fehlfovealisierung zu ermitteln und zu korrigieren. Hierzu wird vom Kamerasteuerungsmodul noch vor der Neupositionierung des Kamerakopfes ein Teilbild B_1 um den Fovealisierungspunkt *f* im Zentrum extrahiert. Dieses Teilbild wird im Soll-Bildspeicher abgelegt und ist nach der Kamerabewegung dem Feinpositionierungsmodul zugänglich. Nach der Kamerabewegung erfolgt die Aufnahme des nun aktuellen Bildes *B*. Läge keine Fehlfovealisierung vor, dann müßte das Teilbild B_1 symmetrisch in *B* zentriert sein. Andernfalls kann nun der Positionierungsfehler ermittelt werden, in dem das zu B_1 passende Teilbild in *B* gefunden und dessen Abweichung von der Mittelposition festgestellt wird. Da ein bestimmter maximaler Positionierungsfehler bestimmbar ist, muß nicht das gesamte Bild *B* betrachtet werden, sondern nur ein zentriertes Teilbild B_2. Hierbei ist B_2 größer als B_1. In beiden Teilbildern werden die horizontalen Kanten mittels Sobel-Operators extrahiert. Die gewonnenen Kantenbilder werden anschließend auf die vertikale Achse projiziert und miteinander korreliert. Der Ort des Maximums der Korrelationsfunktion gibt die Verschiebung des Teilbildes B_1 gegenüber dem Teilbild B_2 in vertikaler Richtung an. Analog zum oben beschriebenen Verfahren wird die horizontale Verschiebung ermittelt. Durch eine getrennte Ermittlung in beide Raumrichtungen (statt einer zweidimensionalen Korrelation) bleibt der Rechen- und damit auch der Zeitaufwand verhältnismäßig gering.

Was geschieht nun mit der ermittelten Verschiebung? In den nachgeschalteten Verarbeitungsblöcken wird immer ein kleinerer als der maximal mögliche Bildausschnitt verarbeitet. Somit braucht eine Neupositionierung des Kamerasystems nicht zu erfolgen, stattdessen kann einfach ein Bildausschnitt extrahiert und weitergegeben werden, der um den Fovealisierungsfehler verschoben ist.

Abbildung 4.19 zeigt eine kurze Sequenz aufeinanderfolgender Fixationen in einer statischen Szene, wenn keine ‘Top down’-Information zur Verfügung steht. Das weiße, formatfüllende Kreuz zeigt den aktuellen Aufmerksamkeitspunkt, das gelbe Kreuz das Ziel des nächsten Blickwechsels und die blauen Kreuze verweisen auf Regionen mit einer geringeren Attraktivität als der nachfolgende Aufmerksamkeitspunkt. Zu erkennen ist auch, daß die Attraktivitätspunkte in aufeinanderfolgenden Frames stabil sind, d. h. sich jeweils auf dieselben Bildstrukturen beziehen. So werden z. B. der Kasten zwischen den Rädern des Kranwagens oder die Fenster des gestreiften Busses immer wieder als interessante Regionen eingestuft.

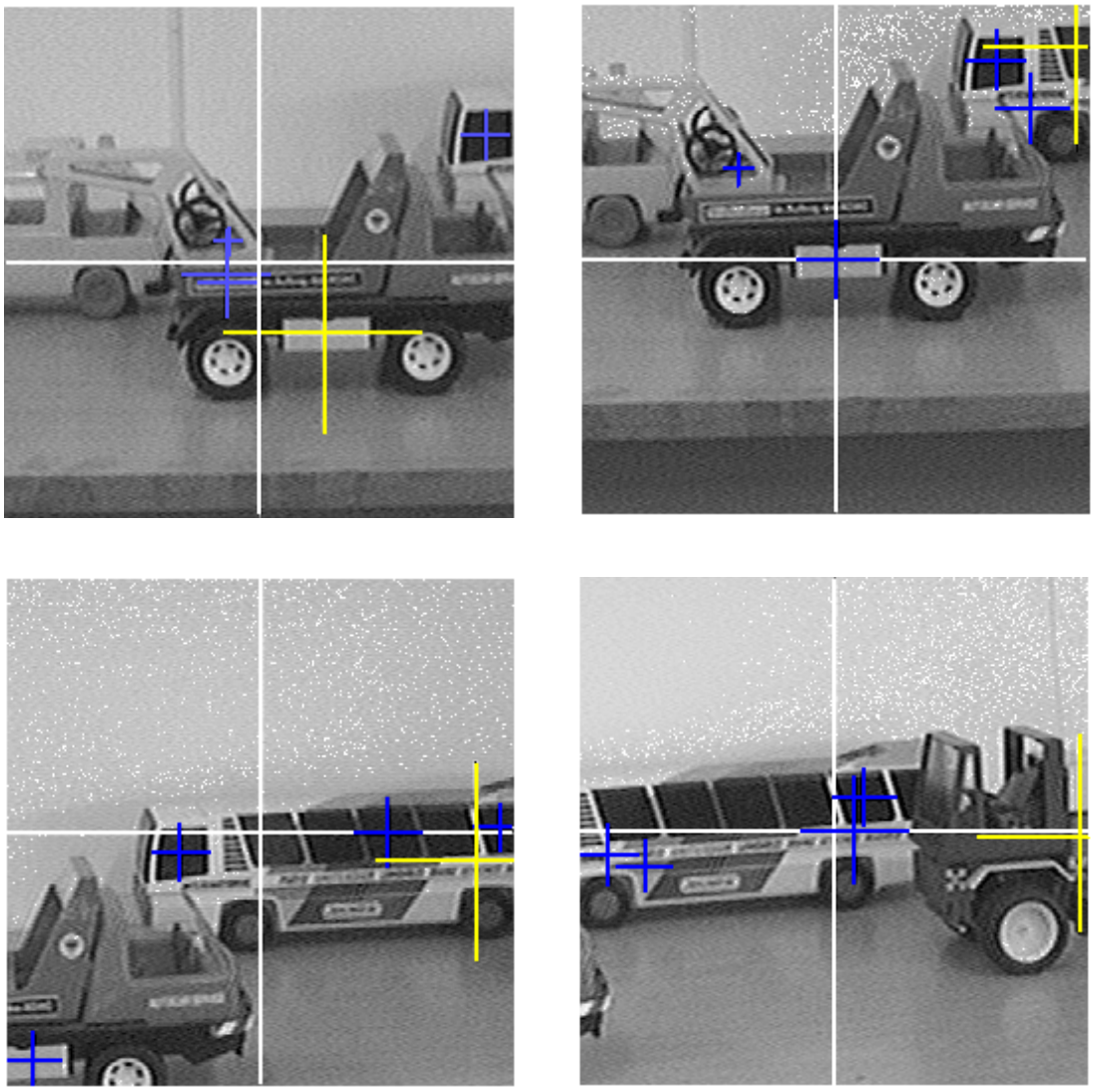


Abbildung 4.19: ‘Bottom up’-gesteuerte Fixationssequenz in einer statischen Szene (das weiße Kreuz markiert den aktuellen Fixationspunkt, das gelbe Kreuz das Ziel der nächsten Fixation und die blauen Kreuze Punkte mit niedrigerer Attraktivität)

4.5 Visuelle Suche mit der datengetriebenen Aufmerksamkeitssteuerung

An dieser Stelle soll gezeigt werden, inwieweit mit der in den vorherigen Abschnitten beschriebenen datengetriebenen Aufmerksamkeitssteuerung auch visuelle Suchaufgaben, wie sie aus der Psychophysik bekannt sind, gelöst werden können. Die Anwendung der Aufmerksamkeitssteuerung auf aus Experimenten der Psychophysik bekannte Displays bietet gegenüber realen Szenen den Vorteil bekannter Rahmenbedingungen. Gleichwohl werden hier nur qualitative

und keine quantitativen Aussagen zu bestimmten Suchergebnissen gemacht. Dadurch unterscheidet sich unsere Herangehensweise etwa von den Arbeiten von Humphreys et al. [Humphreys 1993] und Phaf et al. [Phaf 1990], die konnektionistische Netze entworfen haben, mit denen sich Reaktionszeiten und Antwortverhalten in 'Visual Search'-Experimenten durch die Umrechnung von Netziterationen in Zeitintervalle simulieren lassen. Ein Problem bei dem Entwurf von Funktionen, die den zeitlichen Verlauf einer visuellen Suche beschreiben sollen, besteht in deren geeigneter Parametrisierung. So kann die Allgemeingültigkeit der Funktionen auch für nicht betrachtete Versuchsanordnungen nicht vorausgesetzt werden.

4.5.1 Merkmalssuche

Einfluß der Anzahl der Distraktoren

In einem ersten Experiment soll der Einfluß der Anzahl der Distraktoren auf die Merkmalssuche untersucht werden. Abbildung 4.20 zeigt den verwendeten Satz Displays. Das Target ist das blaue, horizontal ausgerichtete Rechteck. Dieses Element ist auch in allen weiteren in diesem Abschnitt beschriebenen Experimenten, in denen die Farbe die relevante Merkmalsdimension ist, das Target. In diesem ersten Experiment wird die Anzahl der roten, ebenfalls horizontal ausgerichteten Distraktoren von einem Display zum nächsten um jeweils eins erhöht.

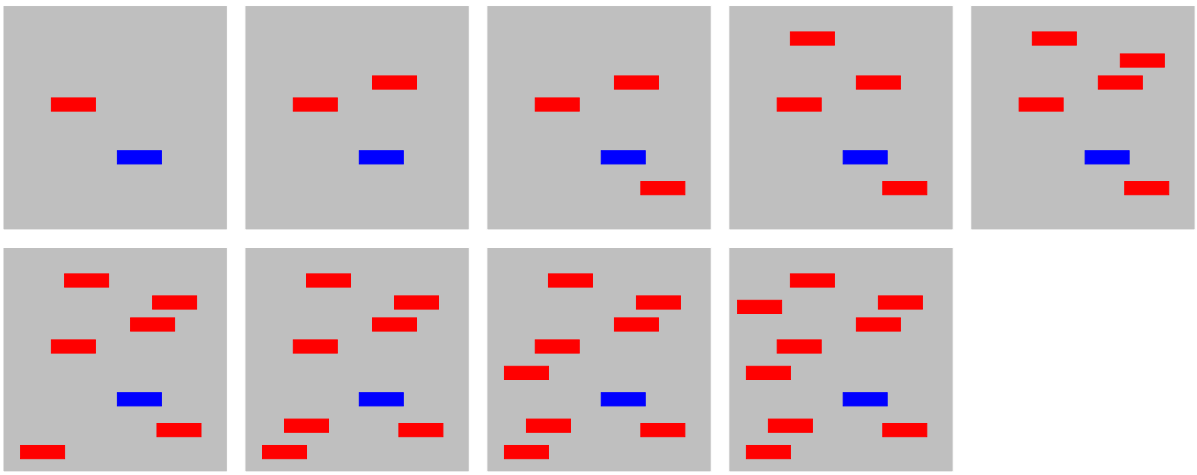


Abbildung 4.20: Displays zur Untersuchung des Einflusses der Anzahl der Distraktoren; relevante Merkmalsdimension Farbe

Wie sieht nun konkret die Ausgabe der bilddatengetriebenen Aufmerksamkeitssteuerung bei Präsentation eines Displays aus? Abbildung 4.21 gibt die von dem System erzeugte ASCII-Ausgabe für das erste Display aus Abbildung 4.20 wieder.

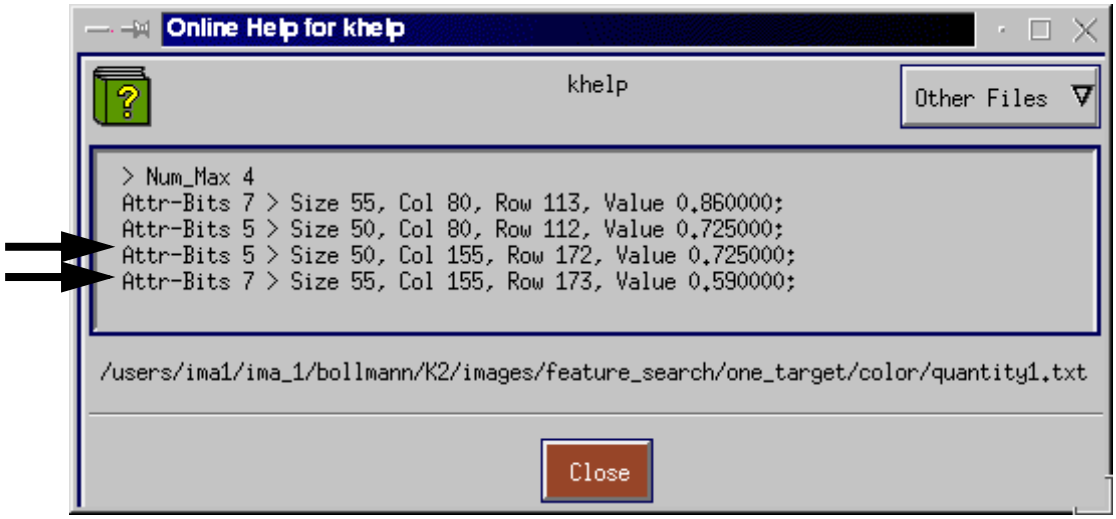


Abbildung 4.21: Attraktivitäten im ersten Display aus Abbildung 4.20

"Attr-Bits 7" bezeichnet das Farbkontraste bewertende Verfahren, wogegen "Attr-Bits 5" entsprechend für das Exzentrizitäten bewertende Verfahren steht. Das Symmetrien bewertende Verfahren besitzt das Typattribut "Attr-Bits 3". Es wird in diesen Experimenten nicht eingesetzt, weil es eben auch Auffälligkeiten außerhalb von Objektgrenzen detektieren kann (vgl. Abschnitt 4.3, insbesondere Abbildung 4.13) und ein einfaches ‘Feature Binding’ aufgrund von Nachbarschaftsbeziehungen schwierig macht. Die weiteren Elemente in obiger ASCII-Datei entsprechen dem in Abschnitt 4.3.1 vorgestellten Fünf-Tupel (Typattribut, Größe, horizontale Position, vertikale Position, Attraktivität). Abbildung 4.21 zeigt, daß das Target (Position (155,172) bzw. (155,173), siehe Pfeile) und der Distraktor (Position (80,112) bzw. (80,113)), wie erwartet, eine gleich auffällige Exzentrizität aufweisen. Der Farbkontrast des roten Distraktors ist dagegen höher als der des blauen Targets. Das Target kann hier also nicht detektiert werden, da es kein Merkmal besitzt, das eine höhere Exklusivität besitzt als die Merkmale des Distraktors. Zudem war das Target nicht als solches vorgegeben, so daß keine ‘Top down’-Information die Suche hätten erleichtern kann.

Da sowohl das Farbkontraste bewertende Verfahren als auch das Exzentrizitäten bewertende Verfahren auf einer Segmentierung basieren und die Positionen der Attraktivitätspunkte damit Flächenschwerpunkten entsprechen, kann hier ein einfaches ‘Feature Binding’ angewendet werden, das die Attraktivitäten addiert und die Größen und Positionen mittelt. Geringe Abweichungen zwischen beiden Verfahren bezüglich der berechneten Größe und Position haben ihre Ursache in den unterschiedlichen Segmentierungsverfahren. Abbildung 4.22 zeigt die für das zweite Display aus Abbildung 4.20 detektierten Attraktivitätspunkte und deren Kombination. Das Target ist erneut durch den Pfeil markiert.

Abbildung 4.22: (a) Attraktivitätspunkte und (b) kombinierte Attraktivitätspunkte im zweiten Display aus Abbildung 4.20

Welche Strategie kann nun angewendet werden, um ein nicht a priori bekanntgegebenes Target zu detektieren? Van de Laar bewertet in dem von ihm vorgestellten Verfahren [van de Laar 1997] die Aktivitäten der Merkmalskarten pixelweise mit der in jeder Karte vorliegenden Gesamtaktivität. Auf diese Weise erhalten exklusive Merkmale eine hohe Wertigkeit und häufig vorkommende Merkmale eine geringe Wertigkeit. Auch wir schlagen hier eine Bewertung der Attraktivitäten entsprechend der Exklusivität der beteiligten Merkmale vor. Da in unseren Merkmalskarten allerdings bereits parametrische Beschreibungen ausgesuchter Attraktivitätspunkte vorliegen, muß nun nicht mehr die Gesamtaktivität pro Merkmalskarte, etwa durch pixelweise Addition, festgestellt werden, sondern es reicht aus, die relative Häufigkeit der sich auf ein bestimmtes Merkmal beziehenden Attraktivitätspunkte festzustellen. Abbildung 4.23 zeigt die Ergebnisse einer solchen Bewertung wiederum für das zweite Display aus Abbildung 4.20. Das Target ist wiederum durch einen Pfeil markiert.

Abbildung 4.23: (a) Bewertete Attraktivitäten im zweiten Display aus Abbildung 4.20 und (b) die zugehörigen verknüpften Attraktivitätspunkte

Die vorgeschlagene Bewertung der Attraktivitätspunkte entsprechend der Exklusivität ihrer Merkmale und das anschließendes Binding der Attraktivitäten führt also zu dem gewünschten Erfolg: Das Target besitzt jetzt die größte Attraktivität und würde von dem System als erstes fixiert und untersucht.

Um die Vielzahl der experimentellen Ergebnisse kompakt darstellen zu können, bietet es sich an dieser Stelle an, ein Maß für die Signifikanz S des Targets in Relation zu seinen Distraktoren einzuführen. Ich definiere die Signifikanz S hier als das Verhältnis der höchsten Attraktivität unter den Distraktoren zur Attraktivität des Targets:

$$S = 1 - \frac{\max(a_t(\text{Distr.}))}{a_t(\text{Target})} \tag{4.38}$$

Ist die Signifikanz größer Null, so wird das Target mit der ersten Fixation detektiert; ist sie kleiner Null, so wird zunächst ein Distraktor fixiert; ist sie gleich Null, ist die Detektion des Targets mit der ersten Fixation zufällig, da mindestens ein Distraktor dieselbe Attraktivität aufweist. Die Signifikanz ist Eins, wenn kein Distraktor im Display vorhanden ist. Je kleiner (aber immer noch positiv) die Signifikanz ist, um so rauschanfälliger ist das Ergebnis der Suche. Für die Displays aus Abbildung 4.20 ergibt sich die in Tabelle 4.4 dargestellte Signifikanz des Targets.

Tabelle 4.4: Signifikanz *S* des Targets in Abhängigkeit von der Anzahl der Distraktoren

Anzahl der Distraktoren	1	2	3	4	5	6	7	8	9
Signifikanz *S*	- 0,284	0,193	0,389	0,517	0,592	0,649	0,691	0,716	0,755

Die Signifikanz des Targets nimmt also mit der Anzahl der Distraktoren zu. Zwei Distraktoren reichen allerdings bereits aus, um das Target als eben solches zu definieren und die Selektion des Targets im ersten Versuch zu ermöglichen.

Ähnliche Ergebnisse liefern Experimente, in denen die Orientierung der Elemente die relevante Merkmalsdimension darstellt. Abbildung 4.24 zeigt die in diesen Experimenten verwendeten Displays, Tabelle 4.5 die Ergebnisse. Auch hier kann das Target detektiert werden, sobald mindestens zwei Distraktoren im Display vorhanden sind. Die Signifikanz des Targets nimmt wiederum mit der Anzahl der Distraktoren zu.

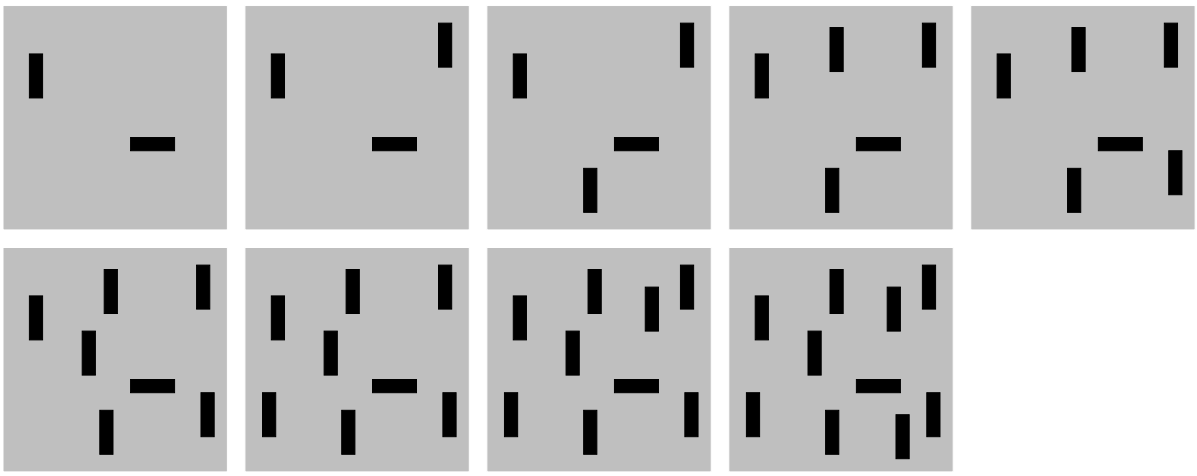


Abbildung 4.24: Displays zur Untersuchung des Einflusses der Anzahl der Distraktoren; relevante Merkmalsdimension Orientierung

Tabelle 4.5: Signifikanz *S* des Targets in Abhängigkeit von der Anzahl der Distraktoren

Anzahl der Distraktoren	1	2	3	4	5	6	7	8	9
Signifikanz *S*	- 0,09	0,308	0,416	0,470	0,502	0,525	0,540	0,552	0,561

Einfluß der Homogenität der Distraktoren

Eine weitere Eigenschaft der Displays, die Einfluß auf die Effizienz der visuellen Suche haben kann, ist die Homogenität der Distraktoren. Diese Homogenität wurde analog zu den vorhergehenden Experimenten in getrennten Versuchsreihen für die beiden relevanten Merkmalsdimensionen sukzessive verringert. Abbildung 4.25 zeigt die Displays dieser Versuchsreihen. Die Ergebnisse sind in Tabelle 4.6 zusammengefaßt.

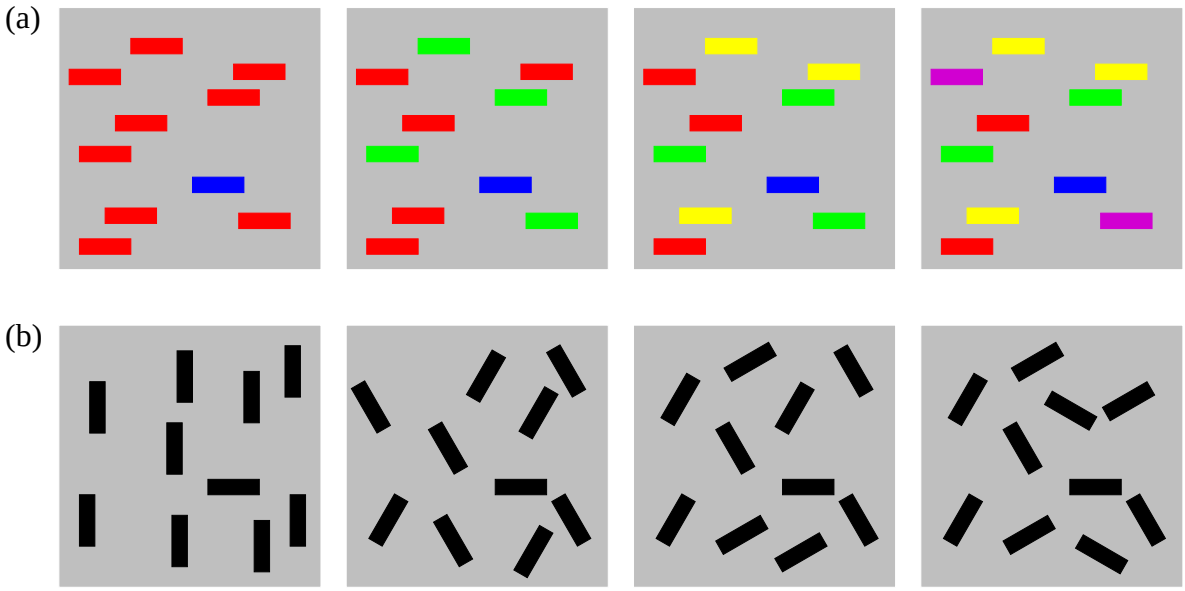


Abbildung 4.25: Displays zur Untersuchung des Einflusses der Homogenität der Distraktoren; relevante Merkmalsdimension (a) Farbe und (b) Orientierung

Tabelle 4.6: Signifikanz *S* des Targets in Abhängigkeit von der Homogenität der Distraktoren

Anzahl der verschiedenartigen Distraktortypen	1	2	3	4
Signifikanz *S*, Farbe variiert	0,735	0,575	0,441	0,103
Signifikanz *S*, Orientierung variiert	0,561	0,475	0,419	0,312

Die Signifikanz des Targets hängt also ebenfalls von der Homogenität der Distraktoren ab; allerdings wird bei den hier durchgeführten Versuchen in jedem Fall augenblicklich, d. h. im ersten Versuch, das Target detektiert.

Einfluß der Ähnlichkeit zwischen Target und Distraktoren

Der Bedeutung der Ähnlichkeit zwischen Target und Distraktoren für die visuelle Suche ist in der Psychophysik umstritten (vgl. Treisman gegenüber Duncan und Humphreys, Abschnitt 3.2). In dem hier vorgestellten technischen System wird der Einfluß der Ähnlichkeit von Kategorisierungseffekten bestimmt. Solange die Distraktoren nicht in dieselbe Merkmalsklasse fallen wie das Target, beeinflußt die Ähnlichkeit zwischen Target und Distraktoren das Ergebnis der visuellen Suche nicht. Abbildung 4.26 enthält die verwendeten Displays und Tabelle 4.7 die berechnete Signifikanz des Targets.

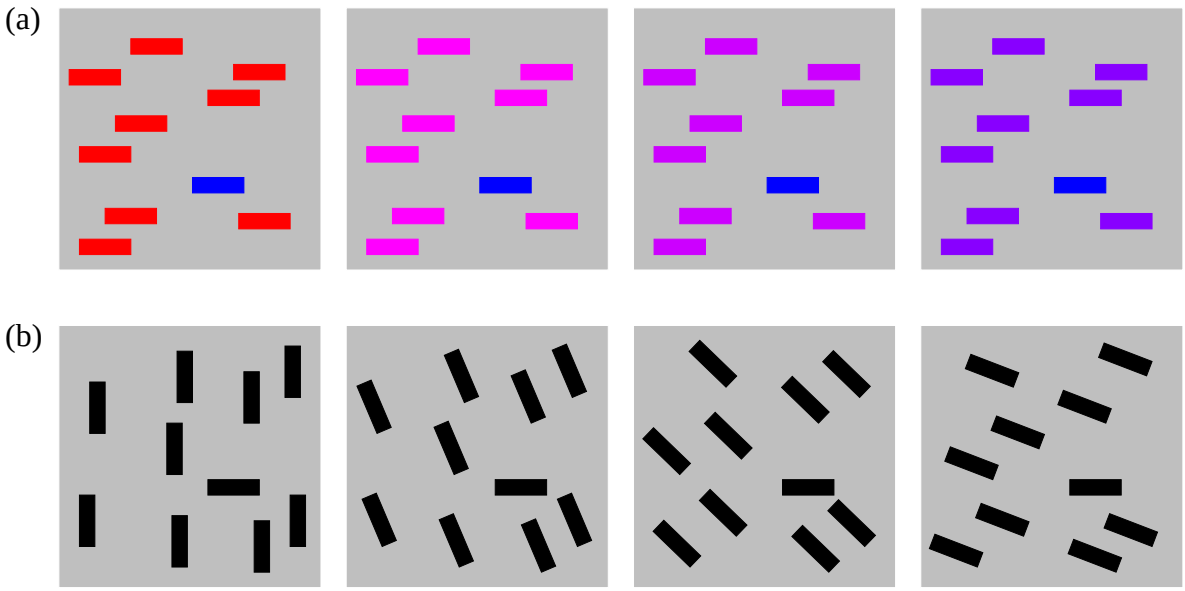


Abbildung 4.26: Displays zur Untersuchung des Einflusses der Ähnlichkeit zwischen Target und Distraktoren; relevante Merkmalsdimension (a) Farbe und (b) Orientierung

Tabelle 4.7: Signifikanz *S* in Abhängigkeit von der Ähnlichkeit des Targets mit den Distraktoren

Ähnlichkeit zwischen Target und Distraktoren	sehr niedrig	niedrig	mittel	hoch
Signifikanz *S*, Farbe variiert	0,755	0,710	0,710	0,710
Signifikanz *S*, Orientierung variiert	0,561	0,567	0,568	0,569

Aus Tabelle 4.7 läßt sich kein evidenter Einfluß der Target-Nontarget-Ähnlichkeit auf das Ergebnis der visuellen Suche feststellen. Die geringen Variationen der Signifikanz *S* hängen von dem variierenden Farbkontrast bzw. der variierenden Exzentrizität der Distraktoren ab. So besitzen z. B. die roten Distraktoren aus dem ersten Display in Abbildung 4.26 einen höheren Farbkontrast zu dem grauen Hintergrund als die violetten Distraktoren in den nachfolgenden Displays.

Einfluß des Zahlenverhältnisses der Distraktoren

Der Einfluß des Zahlenverhältnisses der Distraktoren auf die Merkmalssuche wurde anhand der in Abbildung 4.27 dargestellten Displays untersucht. Das Zahlenverhältnis der Distraktoren variiert von 4:4 über 5:3 und 6:2 bis 7:1; die Gesamtzahl der Distraktoren bleibt innerhalb der Versuchsreihe dagegen konstant. Die Antworten des Systems sind analog zu den vorherigen Experimenten zusammengefaßt (Tabelle 4.8). Das Target wird immer dann im ersten Versuch gefunden, wenn es durch seine Einzigartigkeit aus dem Display hervorgeht, d. h. sobald jeder Distraktortyp wenigstens zweifach im Display vorhanden ist. Die Signifikanz des Targets wächst mit der Ausgeglichenheit des Zahlenverhältnisses der Distraktoren.

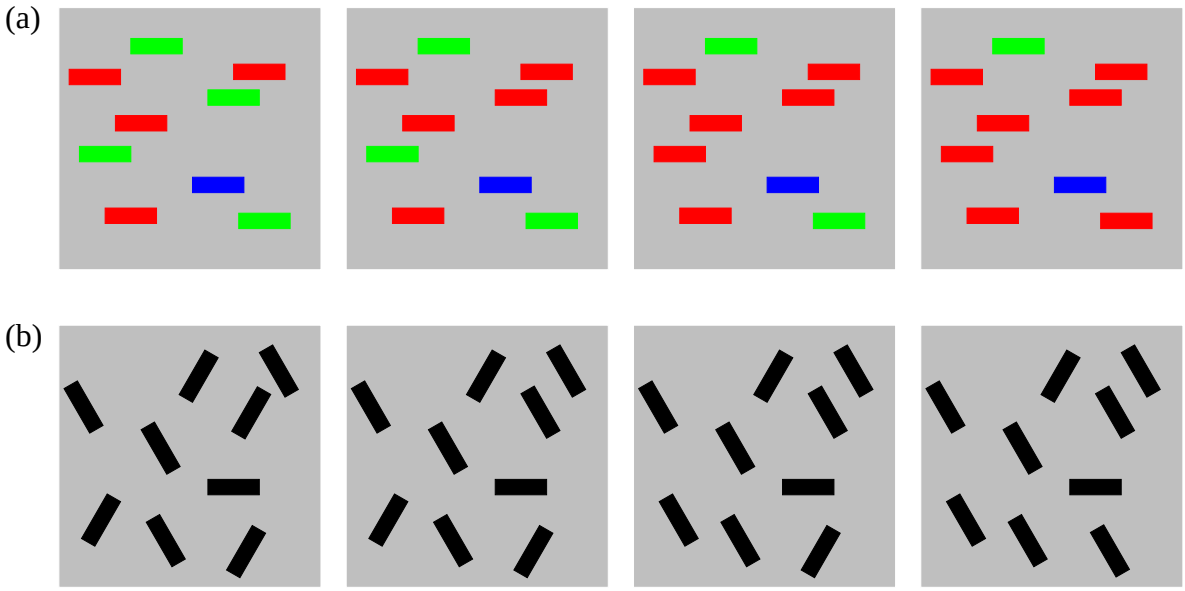


Abbildung 4.27: Displays zur Untersuchung des Einflusses des Zahlenverhältnisses zweier Distraktortypen; relevante Merkmalsdimension (a) Farbe und (b) Orientierung

Tabelle 4.8: Signifikanz *S* des Targets in Abhängigkeit vom Zahlenverhältnis der Distraktoren

Zahlenverhältnis der Distraktoren	4:4	3:5	2:6	1:7
Signifikanz *S*, Farbe variiert	0,552	0,479	0,285	- 0,309
Signifikanz *S*, Orientierung variiert	0,475	0,421	0,309	- 0,022

Zusammenfassend läßt sich feststellen, daß unter der Bedingung, daß die im Display vorhandenen Distraktortypen mit jeweils wenigstens zwei Elementen vertreten sind, mit der angewendeten Suchstrategie jeweils im ersten Versuch das Target gefunden wird. Dies ist vergleichbar mit dem für die Merkmalssuche bekannten Phänomen des ‘Pop out’. Ist dagegen neben dem Target auch ein Distraktor nur einmal im Display vorhanden, ist das Target nicht mehr durch seine Einzigartigkeit definiert und es kann ohne die Hinzunahme von ‘Top down’-Information nicht eindeutig als Target identifiziert werden. In unserem technischen System ist die Signifikanz des Targets abhängig von der Anzahl der Distraktoren, ihrer Homogenität und ihrem Zahlenverhältnis, aber in den durch die Kategorisierung bestimmten Grenzen nicht von der Ähnlichkeit der Distraktoren mit dem Target.

4.5.2 Kombinationssuche

In den Experimenten zur Kombinationssuche besitzt das Target im Gegensatz zu den Experimenten zur Merkmalssuche kein exklusives Merkmal, vielmehr ist das Target durch seine exklusive Merkmalskombination bestimmt. In den von mir zur Kombinationssuche durchgeführten Versuchen wurde wiederum das blaue, horizontal ausgerichtete Element als Target festgelegt. Damit das Target nun kein exklusives Merkmal mehr aufweist, müssen im Display sowohl weitere Elemente vorhanden sein, die ebenfalls horizontal orientiert sind, als auch Elemente, die ebenfalls blau sind. Abbildung 4.28 zeigt die verwendeten Displays. Variiert wurde erneut die Anzahl der Distraktoren, ihre Homogenität und ihr Zahlenverhältnis, nicht aber ihre Ähnlichkeit mit dem Target, da für diese Versuchsreihe keine neuen Ergebnisse gegenüber der Merkmalssuche zu erwarten sind. Auch in der Kombinationssuche besitzt die Ähnlichkeit der Distraktoren untereinander sowie mit dem Target in den jetzt zwei Merkmalsdimensionen so lange keinen Einfluß auf das Ergebnis der Suche, wie die Merkmale nicht in die gleiche Kategorie fallen.

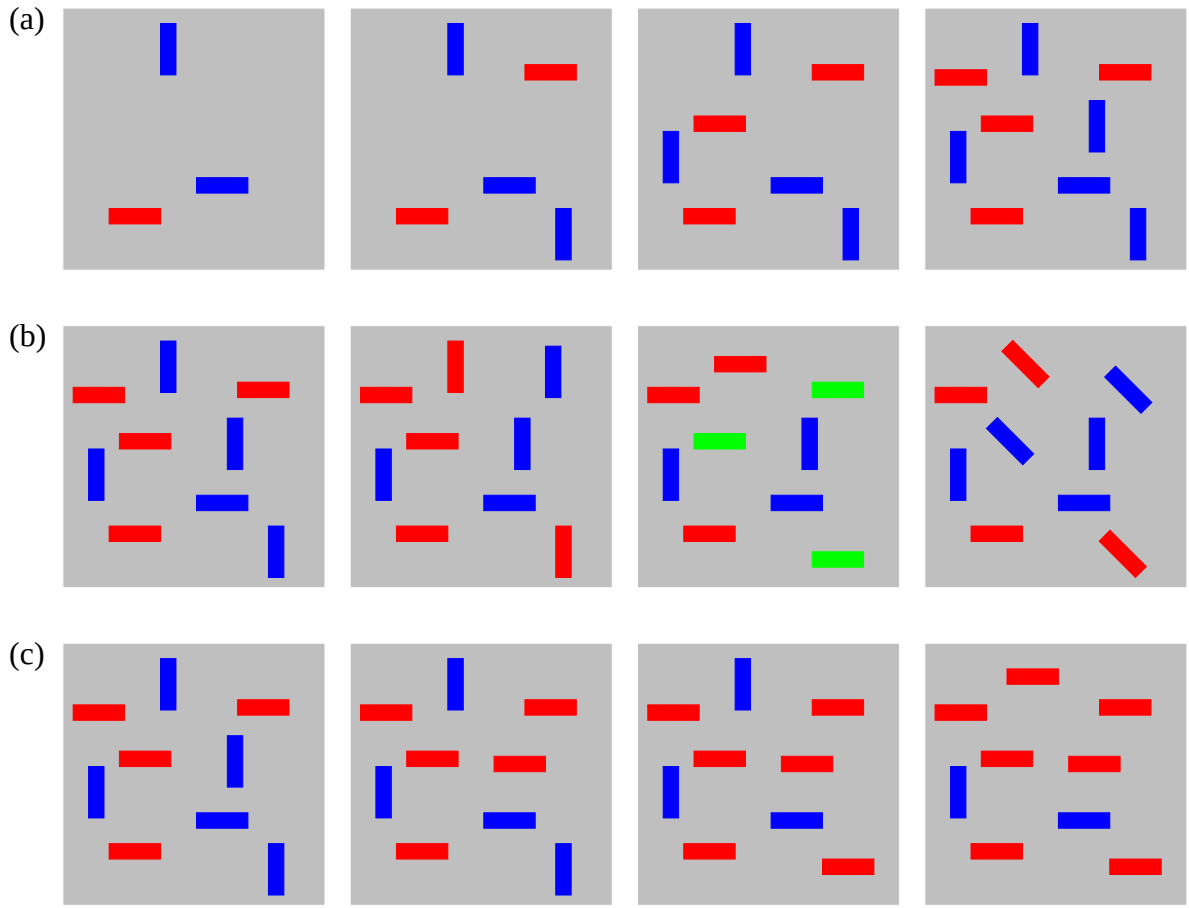


Abbildung 4.28: Displays zur Untersuchung des Einflusses (a) der Anzahl, (b) der Homogenität und (c) des Zahlenverhältnisses der Distraktoren auf die Kombinationssuche

Kann das Target nun ebenso wie in der Merkmalssuche im ersten Versuch fixiert werden? Die folgenden beiden Abbildungen machen deutlich, daß dies nicht der Fall ist. Abbildung 4.29a zeigt die Antwort des ‘Bottom up‘-Aufmerksamkeitssystems auf das zweite Display aus Abbildung 4.28a ohne die Berücksichtigung der Häufigkeit der auftretenden Merkmale. Das Target, in der Abbildung durch einen Pfeil markiert, würde in diesem Beispiel erst an vierter Stelle aufgesucht. Die Bewertung der Auffälligkeiten mit der Häufigkeit des korrespondierenden Merkmals führt in diesem Beispiel, wie Abbildung 4.29b zeigt, sogar dazu, daß die Attraktivität weiterer Distraktoren die des Targets übersteigt. Dieses läßt sich dadurch erklären, daß das Target in dem ausgewählten Display sowohl die häufiger vorkommende Farbe als auch die häufiger vorkommende Orientierung besitzt. Ein Beispiel, in dem die für die Merkmalssuche entworfene Suchstrategie zu einer Erhöhung der Attraktivität des Targets gegenüber den Attraktivitäten einiger Distraktoren führt und damit zu einer schnelleren Detektion des Targets beiträgt, ist dagegen in Abbildung 4.30 dargestellt. Diese Abbildung zeigt das Ergebnis der Suche für das zweite Display aus Abbildung 4.28b. Gleichwohl führt auch hier die angewendete Suchstrategie nicht zur Detektion des Targets mit der ersten Fixation.

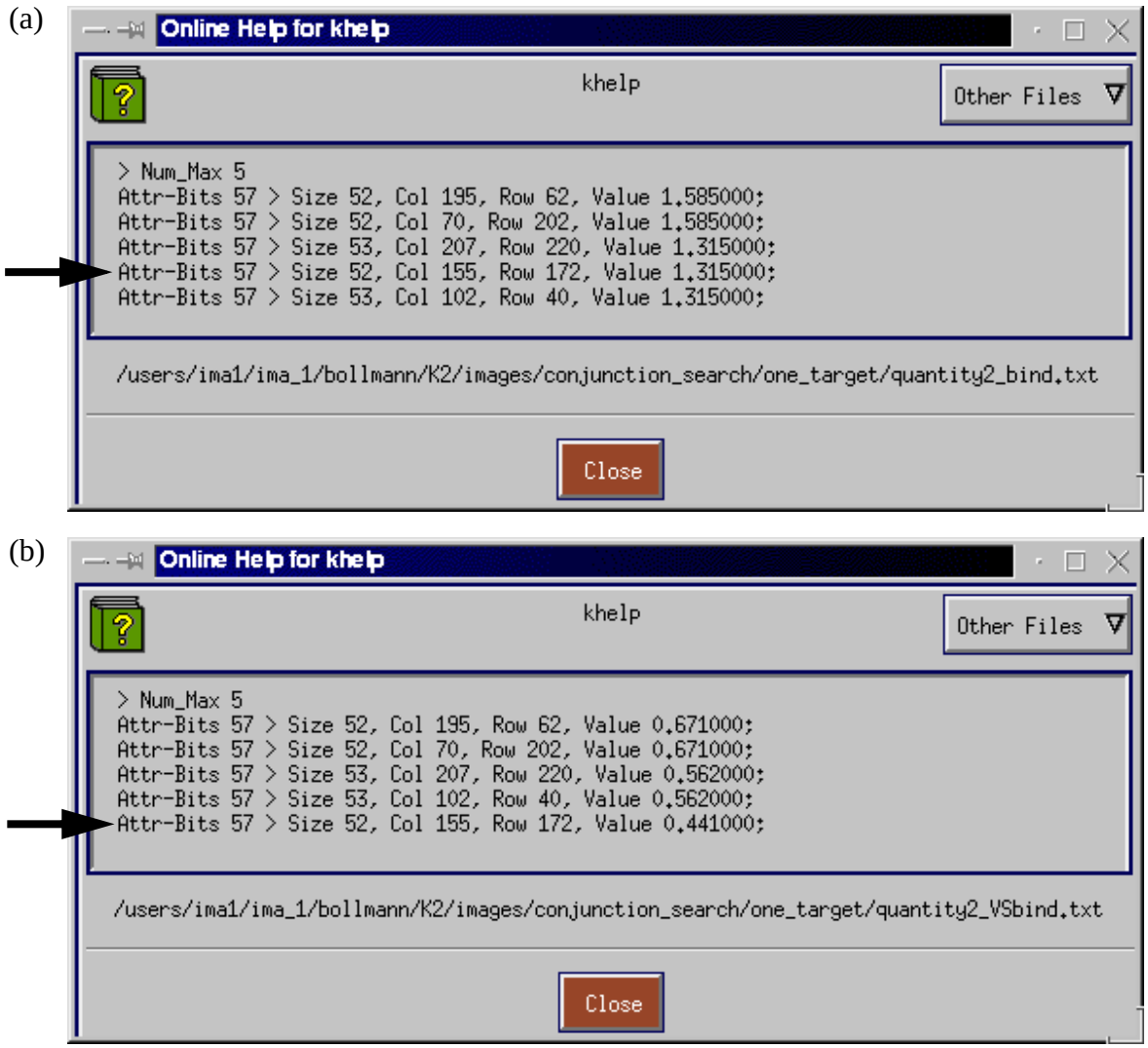


Abbildung 4.29: Antworten des ‘Bottom up’-Aufmerksamkeitssystems auf das zweite Display aus Abbildung 4.28a (a) ohne und (b) unter Berücksichtigung der Häufigkeit der Merkmale

Insgesamt ergibt sich für die Kombinationssuche eine eher zufällige Detektion des Targets. Das heißt, je geringer die Anzahl der Distraktoren im Display, um so größer die Wahrscheinlichkeit, daß das Target früh detektiert wird, oder andersherum ausgedrückt, je größer die Anzahl der Distraktoren im Display, um so geringer die Wahrscheinlichkeit einer schnellen Fixation des Targets. Dieser Zusammenhang läßt sich vergleichen mit Versuchen aus der experimentellen Psychophysik, in denen ebenfalls eine Abhängigkeit der Suchzeit von der Anzahl der Distraktoren festgestellt wurde (vgl. Abschnitt 3.2).

Abbildung 4.30: Antworten des ‘Bottom up’-Aufmerksamkeitssystems auf das zweite Display aus Abbildung 4.28b (a) ohne und (b) unter Berücksichtigung der Häufigkeit der Merkmale

Die Effizienz der Suche kann allerdings in einem maschinellen Aufmerksamkeitsmodell, wie dem hier vorgestellten, durch die Hinzunahme von ‘Top down’-Information wieder erhöht werden. In vielen psychophysischen Experimenten wird den Testpersonen das zu suchende Target a priori vorgegeben. Dementsprechend könnten die zur Detektion des Targets nicht relevanten Merkmalskarten gehemmt werden. Dies würde in den von uns eingesetzten Displays zur Kombinationssuche zur Reduzierung der Auffälligkeiten aller Distraktoren führen, da diese

jeweils wenigstens ein irrelevantes Merkmal besitzen. Für die Suche in realen Szenen wäre zudem eine Hemmung der nicht relevanten Merkmalskarten in Abhängigkeit von der Ähnlichkeit zur Definition des Targets wünschenswert. Das heißt, daß zum Beispiel für die Suche nach einem blauen Target die Merkmalskarte mit den blau-grünen Objekten geringer gehemmt werden könnte als die mit den grünen Objekten. Auf diese Weise könnte man den in realen Szenen vorkommenden Beleuchtungsvariationen Rechnung tragen. Weitere Untersuchungen zu diesen Aspekten sind in dem Fortsetzungsprojekt ESAB-2 geplant.

Kapitel 5

Die modellgetriebene Aufmerksamkeitssteuerung und ihre Integration in NAVIS

Kapitel 5 beschreibt die modellgetriebene Komponente der Aufmerksamkeitssteuerung, die eng verzahnt ist mit weiteren Modulen des *N*euronalen *A*ctive-*Vi*sion-*S*ystems NAVIS. Zunächst werden die experimentellen Plattformen, die NAVIS zur Verfügung stehen, kurz vorgestellt, anschließend wird die Kopplung der Aufmerksamkeitssteuerung mit der Auswertung von Tiefenhinweisen, der Objekterkennung und der Objektverfolgung erläutert. Das Kapitel schließt mit der Beschreibung eines Szenariums, das die Adaption der Aufmerksamkeitssteuerung an eine konkrete Aufgabe zeigt.

5.1 Experimentelle Plattformen

Die NAVIS-Algorithmen können derzeit auf zwei verschiedenen Plattformen evaluiert werden: zum einen auf einem Stereokamerakopf und zum anderen auf einem mobilen Roboter mit monokularem Kamerakopf. Abbildung 5.1 zeigt eine Skizze des Stereokamerakopfes.

Der Stereokamerakopf wurde an der UGH Paderborn unter Berücksichtigung vielfältiger Kalibrierungsmöglichkeiten und geringer Kosten durch die Verwendung von Standardkomponenten entwickelt und in der AG IMA von meinem Kollegen, Herrn Rainer Hoischen, nachgebaut (Abbildung 5.2a). Eine detaillierte Beschreibung der Entwurfsprinzipien und der angewandten Kalibrierungsverfahren findet sich in [Drüe 1996]. Die beiden Kameramodule rotieren von Schrittmotoren angetrieben um voneinander entkoppelte *Vergenzachsen* sowie um eine gemeinsame *Schwenk- und Neigeachse*. Neben diesen vier rotatorischen Freiheitsgraden besitzen die Kameramodule drei translatorische Freiheitsgrade. Diese werden während der Kalibrierung manuell so eingestellt, daß das optische Zentrum der Kameras jeweils im Schnittpunkt zwischen der zugehörigen Vergenzachse und der Neigeachse liegt. Die Kameramodule besitzen

zusätzlich noch drei weitere Einstellvorrichtungen: Jede Kamera kann zyklorotatorisch um ihre optische Achse gedreht werden, bis beide Sensorflächen in einer optischen Ebene liegen, und beide Kameras können in einem kleinen Bereich vertikal und horizontal zu ihrem Kameragehäuse gedreht werden, bis die Sensorebene jeweils senkrecht zur optischen Achse steht.

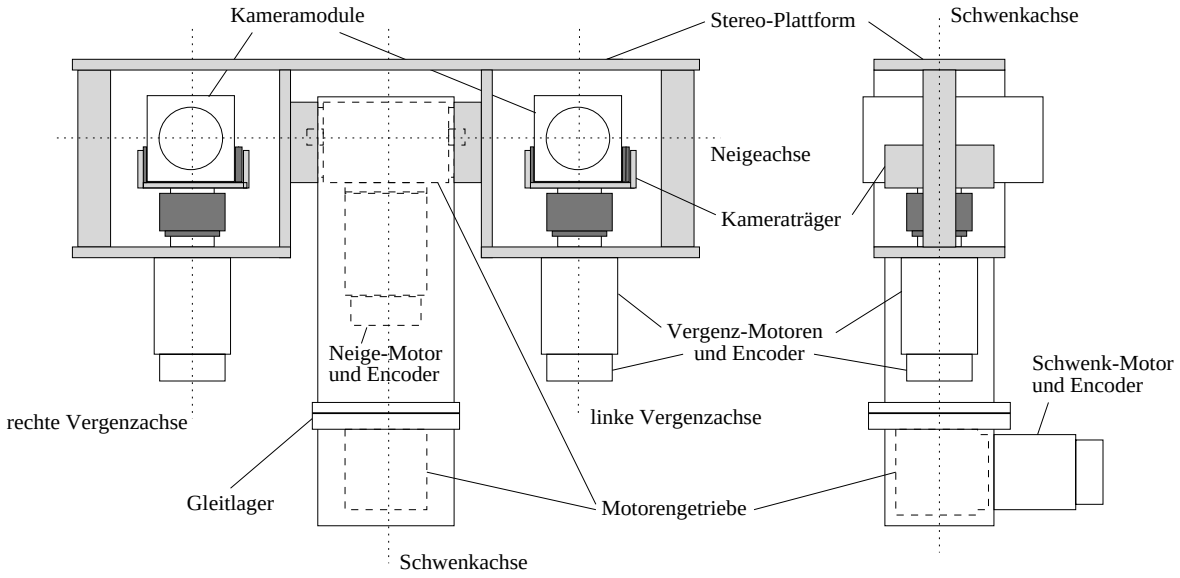


Abbildung 5.1: Skizze des Stereokamerakopfes (nach [Mertsching 1999])

Als mobile experimentelle Plattform steht ein Pioneer1-Roboter zur Verfügung (Abbildung 5.2b). Der Roboter besitzt in seiner Basisversion zwei unabhängig voneinander angetriebene Vorderräder, ein passives Hinterrad in Form einer Möbelrolle, sieben Sonarsensoren und einen Greifer mit einem Freiheitsgrad. Um ihn als experimentelle Plattform nutzen zu können, wurde zusätzlich eine Schwenk-Neige-Einheit und eine CCD-Farbkamera montiert. Auf dem Roboter selbst befindet sich nur das rudimentäre *Betriebssystem PSOS* zum Auslesen der Winkel-Encoder an den Vorderrädern und der Sonardaten. Die *Entwicklungsumgebung SAPHIRA* zur Navigation des Roboters sowie die Bildverarbeitung laufen dagegen auf einer general-purpose Workstation. Die Kommunikation zwischen den verschiedenen Modulen auf der Workstation und dem Roboter wird durch drei Funkübertragungsstrecken gewährleistet, die das analoge Videosignal der Kamera sowie die seriellen Datenströme zur Steuerung der Schwenk-Neige-Einheit und des Roboters übertragen. Weitere Details zum Pioneer1-Roboter und der SAPHIRA-Software können z. B. [Guzzoni 1997], [Konolige 1997] und http://www.activemedia.com/robots/ entnommen werden.

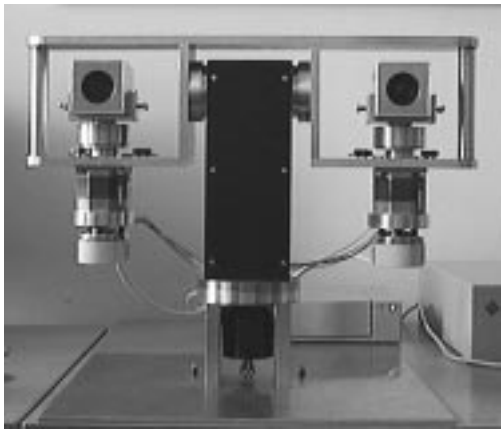
(a)  (b) 

Abbildung 5.2: (a) Stereokamerakopf, (b) mobiler Roboter mit monokularem Kamerakopf

5.2 Stereoskopische Tiefenschätzung

In einer späteren Version der in dieser Arbeit entworfenen und im DFG-Projekt ESAB-2 weiter zu entwickelnden Blicksteuerung soll das ‘Feature Binding’-Problem im 3D-Raum gelöst werden. Durch die Berechnung von *Disparitätskarten* können basierend auf der Hypothese, daß benachbarte Attraktivitätspunkte gleicher Tiefe auf ein identisches Objekt hinweisen, eben solche Punkte zusammengefaßt werden. In diesem Abschnitt wird daher kurz der Stand der NAVIS-Arbeiten betreffend der stereoskopischen Tiefenschätzung vorgestellt.

Die Auswertung der Stereodaten des in Abschnitt 5.1 beschriebenen Kamerakopfes dient derzeit zur Objekt-Hintergrund-Trennung, die die Erkennung von Objekten vor stark strukturierten Hintergründen erleichtern soll. Hierzu wird zunächst das Stereobildpaar *rektifiziert*, d. h. die Bildpunkte werden so abgebildet, daß sie den Aufnahmen parallel ausgerichteter Kameras entsprechen. Anschließend werden Kantenelemente mittels einer Gaborfilterbank (vgl. Abschnitt 4.1.1) aus dem Stereobildpaar extrahiert. Durch ein einfaches Korrelationsverfahren wird anschließend ein relativ dünn besetztes Disparitätsfeld erzeugt, indem jeweils ein Zeilenausschnitt des einen Bildes mit der entsprechenden kompletten Zeile des zweiten Bildes kreuzkorreliert wird. Die Zeilen beider Bilder sind einander über das sogenannte ‘*Epipolar Constraint*’ zugeordnet, welches besagt, daß ein Punkt eines Bildes auf einer Geraden im anderen Bild zu suchen ist. Für rektifizierte Bilder ist diese Gerade eine Waagerechte in der Höhe des betrachteten Bildpunktes. Die anschließende Anwendung eines ‘Continuity Constraints’ weist räumlich benachbarten Bildpunkten gleiche Disparitäten zu. Allerdings bleibt das Disparitätsfeld weiterhin sehr lückenhaft. Daher werden die vorhandenen Disparitäten unter Berücksichtigung eines aus verschmälerten Gaborfilterantworten erzeugten Grauwertkantenbildes in bisher undefinierte Bereiche expandiert. Folgende Heuristiken werden hierzu eingesetzt:

- Die Expansion der Tiefenwerte über Konturelemente hinaus ist nicht möglich, da an diesen Positionen die Wahrscheinlichkeit für das Vorhandensein realer Tiefensprünge hoch ist.
- Die Expansion der Tiefenwerte entlang der Konturelemente wird erleichtert, um möglicherweise vorhandene geschlossene Konturen zu erhalten, die anschließend aufgrund des konservativen Wachstumskriteriums nur nach innen gefüllt, aber nicht nach außen expandiert werden (siehe Abbildung 5.3).
- Benachbarte Bildpunkte, die auch nach der Anwendung des ‘Continuity Constraints’ noch unterschiedliche Disparitäten aufweisen, stellen ebenfalls Barrieren für die Ausbreitung der Tiefenwerte dar.

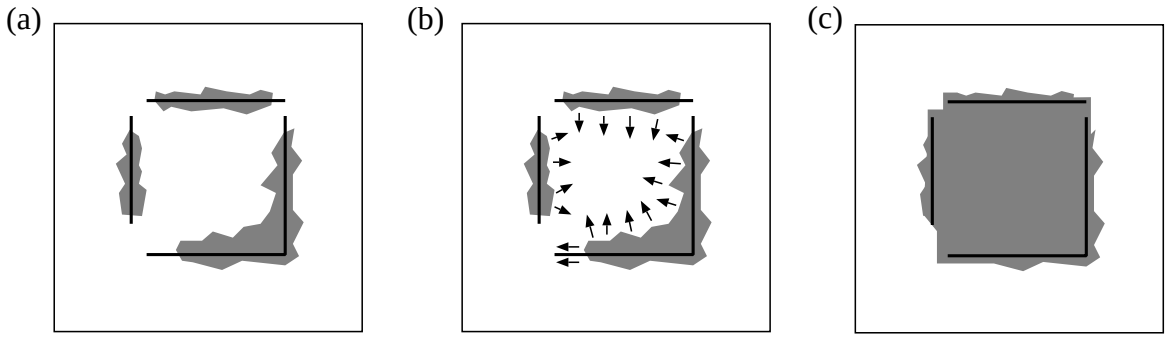


Abbildung 5.3: Prinzip des Expansionsverfahrens: (a) Szene mit Konturelementen und initialen Disparitätswerten, (b) Expansionsrichtungen und (c) Ergebnis

In dem expandierten Disparitätsfeld werden Disparitätsdiskontinuitäten bestimmt, die mit dem Grauwertkantenbild verknüpft werden. Hierbei werden Grauwertkanten, die innerhalb einer Tiefenebene liegen, entfernt. Das neu entstehende Kantenbild, das expandierte Disparitätsfeld und ein vorgegebener Startpunkt werden anschließend in einem Segmentierungsprozeß genutzt, um eine Tiefenebene, die im idealen Fall genau ein Objekt enthält, auszumaskieren. Der Punkt gibt die Startposition des Segmentierungsprozesses vor; er könnte z. B. von der Aufmerksamkeitssteuerung geliefert werden. Die Form des resultierenden Segments hängt von dem Kantenbild und dem expandierten Disparitätsfeld ab.

Details zur Berechnung der Disparitätskarten sowie zum Expansionsalgorithmus und der sich ergebenden Szenensegmentierung können in [Trapp 1995] bzw. [Lieder 1997, 1998] gefunden werden. Exemplarische Ergebnisse des Verfahrens zeigt Abbildung 5.4.

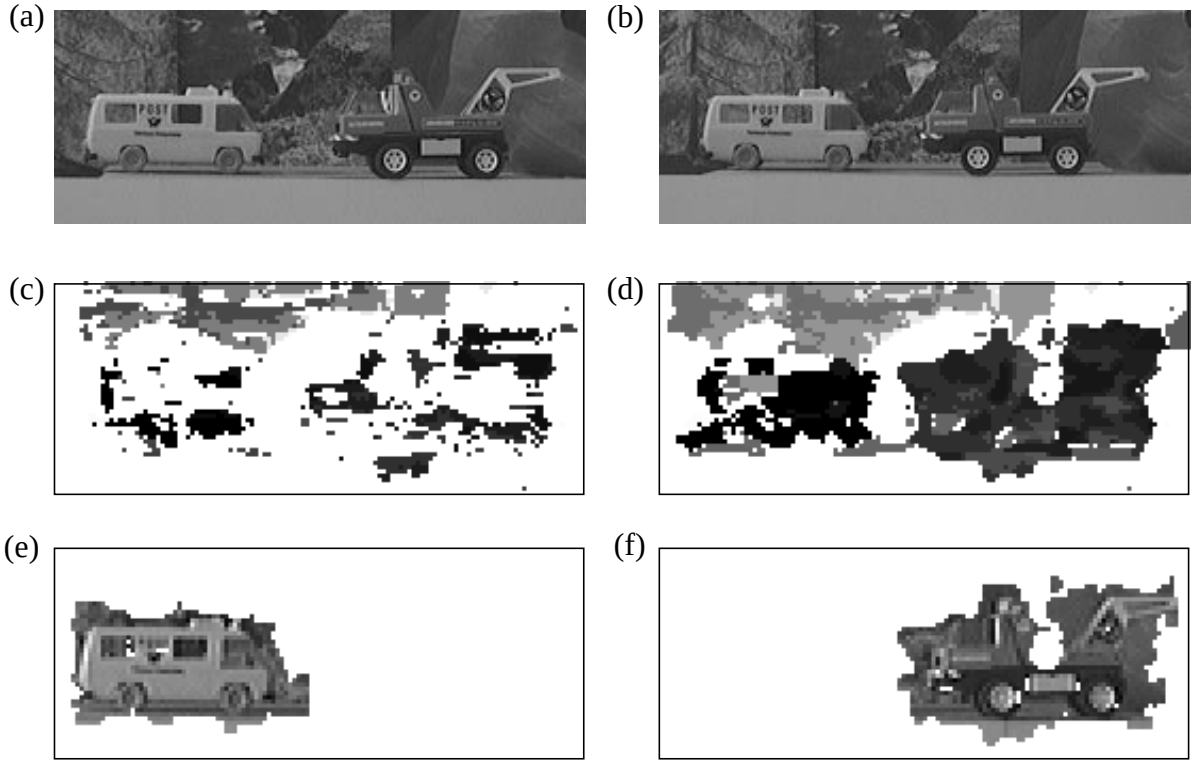


Abbildung 5.4: (a) Linkes, (b) rechtes Bild des Stereobildpaares, (c) Disparitätskarte, (d) expandierte Disparitätskarte, (e) Segmentierungsergebnis für den Fall, daß der bezeichnete Startwert der Segmentierung im Bereich des Busses liegt, und (f) Segmentierungsergebnis, falls der Startwert im Bereich des Kranwagens liegt

5.3 Die 2D-Objekterkennung

Wie bereits in der Einleitung zu Kapitel 4 geschildert, beeinflußt die 2D-Objekterkennung die Auswahl des nächsten Aufmerksamkeitspunktes über die sogenannte Verstärkungskarte. In ihr werden die Punkte eingetragen, die für die Validierung einer aktuellen Objekthypothese relevant sind. Dies sind die Punkte, die in der Lernphase zu bestimmten Teilausschnitten des Objektes gelernt wurden und einen Wechsel der Blicksteuerung zu einem anderen Objekt während der Validierungsphase erschweren sollen. Die folgenden Abschnitte erläutern die Details der 2D-Objekterkennung, die ein Modell der modellgetriebenen Aufmerksamkeitssteuerung darstellt. Dieses Modell kann allerdings auch wiederum in ein Modell zur Generierung einer Objekthypothese und ein Modell zur Validierung einer Objekthypothese unterteilt werden. Unter dem Begriff Modell ist hier einfach eine Ansammlung von Wissen zu verstehen, das in strukturierter Form vorliegt.

5.3.1 Erkennungsstrategie

Für technische Systeme besteht eine Schwierigkeit bei der Erkennung darin, gelernte[1] komplexe Objekte nach einer translatorischen Verschiebung (oder gar Rotationen) überhaupt zu lokalisieren, wenn man nicht davon ausgehen kann, daß sie sich von ihrer Umgebung bereits durch ihre Farbgebung oder Konturverläufe deutlich abheben. Ein Lösungsansatz wäre der Versuch, an jeder Bildposition jedes mögliche Objekt holistisch zu erkennen, jedoch ist dieses für mehrere gespeicherte Objekte viel zu ineffizient und unflexibel. Eine effektivere Suche ergibt sich durch die Reduktion der gelernten Objekte auf ikonische Repräsentationen und den Einsatz einer Auflösungspyramide, in der zunächst nur in der schlechtesten Auflösung gesucht wird (siehe [Rao 1997]). Gleichwohl wird auch in diesem Ansatz das Bild Pixel für Pixel (nach dem minimalen Euklidischen Abstand zwischen dem gesuchten Merkmalsvektor und den aktuellen Merkmalsvektoren) durchsucht. Viel eleganter erscheint es dagegen, die Bildinformation zunächst nach bestimmten Merkmalen zu gruppieren und somit den Suchraum drastisch einzuschränken. Nehmen wir also an, jeder von der Aufmerksamkeitssteuerung gelieferte Attraktivitätspunkt spiegelt die Eigenschaften der Bildpunkte in seiner lokalen Umgebung wider; dann kann ein Objekt mit verschiedenen Merkmalen durch eine kleine Menge von Attraktivitätspunkten unterschiedlichen Typs und an unterschiedlicher Position grob repräsentiert werden, wodurch man es auch effizient in einer Szene lokalisieren kann. Da aber lediglich grobe Strukturen verwendet werden, müssen Details durch ein anderes Verfahren, z. B. ein Template-Matching von Objektteilen, untersucht werden. Durch den Vergleich der Attraktivitätspunktkonfigurationen ergibt sich also lediglich ein Hinweis bezüglich der Präsenz eines Objekts, der die Bildung einer Objekthypothese ermöglicht. Der gesamte Vorgang ist schematisch noch einmal in Abbildung 5.5 dargestellt. Gespeicherte Attraktivitätspunktkonfigurationen der gelernten Objekte werden in der aktuellen Szene gesucht.

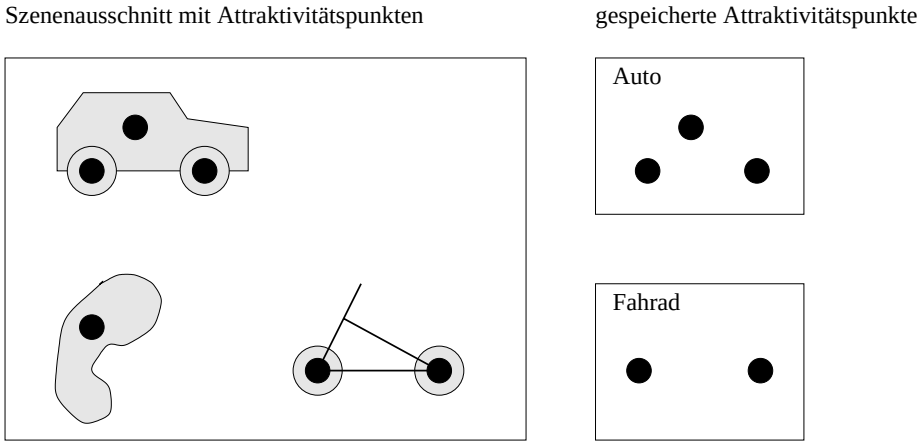


Abbildung 5.5: Hypothesenbildung durch Attraktivitätspunkte (aus [Götze 1995])

[1] In einem technischen System kann ein Objekt als gelernt gelten, wenn Eigenschaften des Objekts im System gespeichert sind, die in einer über dem statistischen Mittel gelegenen Anzahl von Fällen zu einer erneuten Identifizierung des Objekts führen.

Gehen wir nun von der generierten Hypothese aus, daß sich ein Objekt an einer bestimmten Position befindet, dann müßte es ausreichend sein, nicht das gesamte Objekt, sondern nur eindeutige Objektteile zu identifizieren. Diese können besonders effektiv gefunden werden, wenn neben den gelernten Objektteilen auch ihre räumliche Anordnung im System gespeichert ist. Ein solches Vorgehen ist analog zur Scanpath-Theorie von Noton (siehe Abschnitt 3.1.2) und wird z. B. auch in der Gesichtererkennung eingesetzt (vgl. [Yamada 1995]). Eine Objekterkennung basierend auf der Auswertung eines Scanpathes impliziert allerdings eine sehr wesentliche Einschränkung: Es können nur Objekte erkannt werden, die in einer gelernten Ansicht präsentiert werden. Eine Erweiterung der hier beschriebenen Objekterkennung zu einer ansichtenbasierten 3D-Objekterkennung unter Ausnutzung temporaler Assoziationen ist zwar bereits abgeschlossen (vgl. [Massad 1997, 1998a+b]), soll hier aber nicht weiter vorgestellt werden, weil die Ergebnisse nicht die Grundlage der in dieser Arbeit entwickelten Blicksteuerung bilden. Die von dem 2D-Objekterkennungssystem gelernten und wiederzuerkennenden Objektteile werden in NAVIS auch *Formen* genannt.

Wie kann die Untersuchung einer Szene auf bekannte Objekte nun in einen geschlossenen Ablauf gebracht werden? Ein durch die Aufmerksamkeitssteuerung gelieferter Aufmerksamkeitspunkt kann als Referenzpunkt für die Hypothesengenerierung dienen: Wenn der Referenzpunkt eine auffällige Eigenschaft eines bekannten Objekts repräsentiert, dann muß er bereits in der gespeicherten Attraktivitätspunktmenge des Objektes vorhanden sein. Zusätzlich müssen sich die anderen Attraktivitätspunkte des präsentierten Objektes ebenfalls in der gespeicherten Punktmenge befinden. Wird nun eine solche Menge gefunden, dann kann die gebildete Hypothese validiert werden, indem die ebenfalls zu dem bestimmten Objekt gespeicherten Formen mit den im Szenenbild vorhandenen Formen verglichen werden. Stimmen die Formen nicht überein, so muß eine neue Hypothese gebildet werden. Kann keine Hypothese (mehr) gebildet werden, dann entweder, weil keine passende Attraktivitätspunktmenge gefunden wurde, oder, weil die von den passenden Mengen gebildeten Hypothesen aufgrund nicht vorgefundener Details nicht akzeptiert werden konnten. In diesem Fall wird ein anderer Attraktivitätspunkt als neuer Aufmerksamkeitspunkt ausgewählt. Der alte Punkt wird in der Inhibitionskarte gemerkt, damit er nicht noch einmal verwendet wird. Wurde dagegen ein Objekt erfolgreich identifiziert, so werden neben dem Aufmerksamkeitspunkt noch die anderen Attraktivitätspunkte, welche ebenfalls Merkmale des erkannten Objektes darstellen, in die Inhibitionskarte eingetragen.

Um auch Objekte oder Formen außerhalb des jeweils aktuellen Kamerabildes untersuchen zu können, muß die Kamera natürlich entsprechend bewegt werden. So werden während der Überprüfung einer Hypothese ständig neue Szenenausschnitte aufgenommen. Durch die Kamerabewegung findet dann eine Verschiebung der Bildinhalte relativ zum Bildrand statt. Also müssen auch sämtliche Karten mit koordinatenbezogener Information ständig aktualisiert werden. Eine Möglichkeit wäre nun, erst nach der abgeschlossenen Erkennung des gerade untersuchten Objektes neue Attraktivitätspunkte sowie einen neuen Aufmerksamkeitspunkt zu generieren. Interessanter erscheint es aber, während der Hypothesenvalidierung die durch die Kamerabewegung neu ins Bild gerückten Objekte mit zu beachten. Ein solches "offenes" System kann die Untersuchung eines Objekts zu Gunsten eines wichtiger erscheinenden Objekts unter-

brechen. Um dies zu ermöglichen, müssen nach jeder Kamerabewegung die Attraktivitätspunkte und der Aufmerksamkeitspunkt neu berechnet werden. Damit wirklich nur sehr interessante Objekte mit hochwertigen Attraktivitätspunkten den Ablauf der Hypothesenvalidierung unterbrechen können, werden von den neuen Attraktivitätspunkten jene in ihrer Wertigkeit verstärkt (in die Verstärkungskarte eingetragen), die mit den gelernten Punkten übereinstimmen. Insbesondere ist es also für die Erkennung nicht notwendig, daß das zu validierende Objekt in einer einzigen Aufnahme komplett präsentiert wird.

5.3.2 Erkennungssystem

Nachdem im vorherigen Abschnitt die zugrundegelegte Erkennungsstrategie dargestellt wurde, kann nun das implementierte 2D-Erkennungssystem und dessen Kopplung mit den anderen NAVIS-Modulen erläutert werden. NAVIS läßt sich funktional in vier übergeordnete Blöcke gemäß Abbildung 5.2 unterteilen.

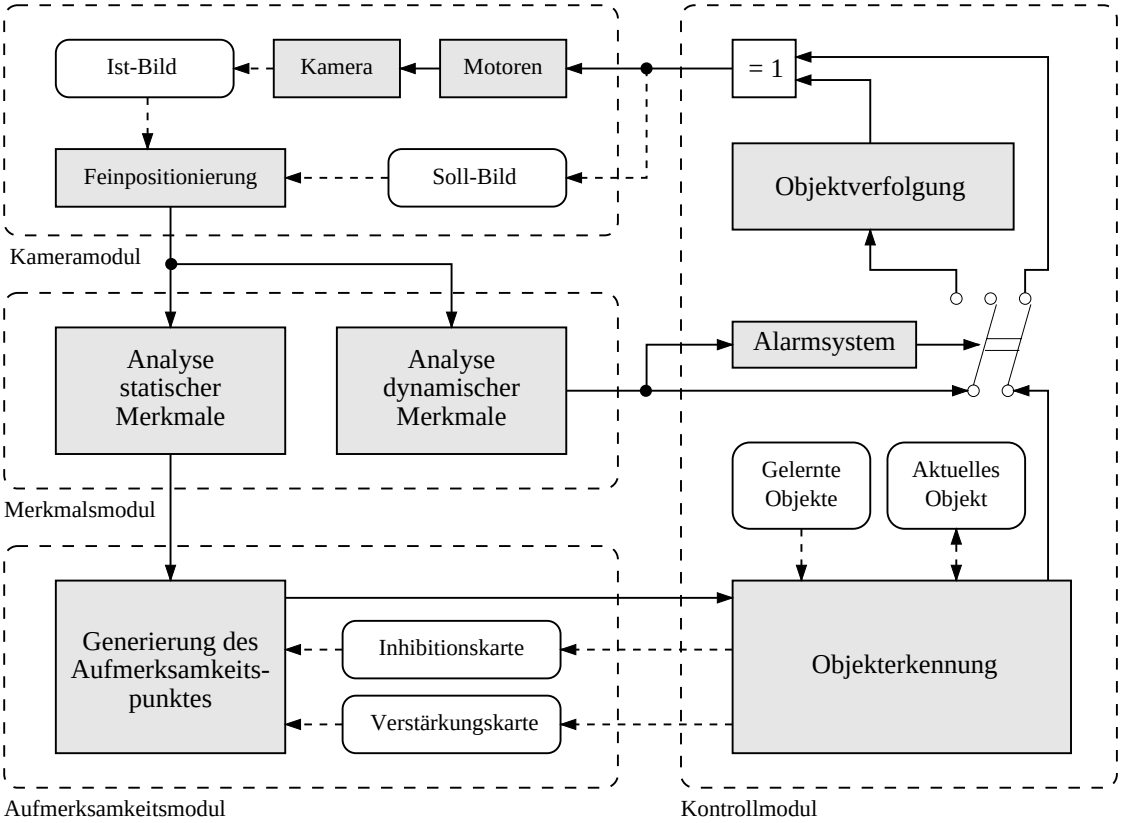


Abbildung 5.6: Übersicht über das Gesamtsystem NAVIS

Das Kameramodul enthält die Bildaufnahme, die Bewegung der Kamera sowie die Feinpositionierung (siehe Abschnitt 4.4). Das Merkmalsmodul generiert ‘bottom up’ die für die Erken-

nung notwendigen Formen (wird im folgenden noch genauer erläutert) sowie die Merkmals- und Auffälligkeitskarten, die neben den bereits vorgestellten statischen Objekteigenschaften (Abschnitte 4.1 und 4.2) auch dynamische Objekteigenschaften berücksichtigen (Abschnitt 5.4). Die für die Formerkennung relevanten Merkmale sind identisch mit den in Abschnitt 4.1 beschriebenen orientierten Kanten. Das Aufmerksamkeitsmodul kombiniert die verschiedenen Auffälligkeitskarten und verrechnet die entstehende Attraktivitätskarte mit der Inhibitions- und der Verstärkungskarte zur Aufmerksamkeitskarte (vgl. Abschnitt 4.3). Das Kontrollmodul dient der Identifizierung gelernter Objekte, so lange sich kein bewegtes Objekt in der Szene befindet (siehe Schalterstellung in Abbildung 5.6), und steuert dann den gesamten Ablauf der Erkennung. Durch den Vergleich der gespeicherten Attraktivitätpunktkonfigurationen mit den um den jeweils aktuellen Aufmerksamkeitspunkt vorhandenen Punkten wird eine Objekthypothese generiert. Diese wird validiert, indem die präsentierte Szene auf die zu einem gelernten Objekt gehörenden Formen sukzessive untersucht wird.

Repräsentation eines Objekts

Ein Objekt setzt sich in der 2D-Objekterkennung von NAVIS aus Formen, das sind spezifische Objektteile, und aus einer Attraktivitätspunktkonfiguration, im folgenden auch Hypothesenpunktmenge *OH* genannt, zusammen.

Das Lernen einer Hypothesenpunktmenge *OH* für ein Objekt zieht sich über eine Sequenz von mehreren Bildern hinweg. Damit nur die zum Objekt gehörenden Punkte gelernt werden, ist es wichtig, daß sich keine zusätzlichen Attraktivitätspunkte einstellen. Daher werden die Objekte in der Lernphase vor homogenem Hintergrund präsentiert. Im allgemeinen wird der Hebbsche Mechanismus zum Lernen der Hypothesenpunkte verwendet, d. h. es werden solche Punkte gelernt, die oftmals an der gleichen Stelle präsentiert werden. Das Objekt darf daher zwischen den Aufnahmen nicht in seiner Position verändert werden.

Formen entstehen im Merkmalsmodul (vgl. Abbildung 5.6) durch die Überlagerung der Gaborfilterantworten (siehe Abschnitt 4.1.1) aller Orientierungen und anschließende Binarisierung. Zu jeder Form gehört ein sich im Zentrum befindender Formpunkt.

In der Lernphase wird die Anzahl, Position und Größe der zu einem Objekt gehörenden Formen durch den Lehrer vorgegeben. Jede Form wird durch das System fixiert und innerhalb eines Zyklus gelernt, d. h. die extrahierte Form wird in einem Objektspeicher gemerkt. Als Fixationspunkt dient der vorgegebene Formpunkt. In der Erkennungsphase werden die gelernten Formen mit aktuell extrahierten und verbreiterten Formen (siehe Abbildung 5.7) verglichen; dabei ist die Übereinstimmung dann am größten, wenn die gelernte Form vollständig in der präsentierten Form enthalten ist. Die Verbreiterung des aktuell extrahierten Kantenbildes erhöht die Invarianz bezüglich geringfügiger Größenänderungen, Translationen und Rotationen.

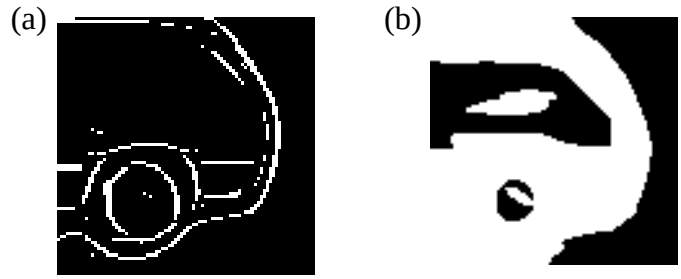


Abbildung 5.7: (a) Gelernte Form, (b) präsentierte Form

An dieser Stelle sollen die zusätzlich benötigten Punktmengen noch formal eingeführt werden: Neben der bereits bekannten Menge *AT* von aktuellen Attraktivitätspunkten wird die zu einem beliebigen Objekt gelernte Attraktivitätspunktmenge *OH* (für *O*bjekt-*H*ypothesenpunkte) benötigt. Die Menge der zu einem einzigen Objekt gehörenden Formpunkte (zu jedem Objekt werden mehrere Formen gelernt) sei mit *OF* bezeichnet. Auch die Elemente in *OH* und *OF* werden durch das bereits in Abschnitt 4.3 vorgestellte Fünf-Tupel beschrieben. Allerdings entspricht der Größe-Parameter a_g jetzt der Ausdehnung der Form und das Typattribut a_t indiziert die Form selbst (z. B. Kotflügel, Reifen, Heckklappe, etc.). Das Wertigkeitsfeld a_w wird in der gelernten Repräsentation nicht benötigt. Der aktuelle Aufmerksamkeitspunkt sei weiterhin mit α referenziert.

Im Erkennungsmodul in Abbildung 5.6 sind die Mengen *OH* und *OF* sowie die Formen der Objektteile durch die Box "gelernte Objekte" repräsentiert. Die Box "aktuelles Objekt" steht für einen Speicher, in dem eine Identifikationsnummer des jeweils hypothesierten Objektes mit seiner vermuteten Bildposition abgelegt wird. Dieser Speicher wird im folgenden auch aktueller Objektspeicher genannt. Ebenso werden hier Formpunkte der gelernten Formen abgelegt, welche zur Hypothesenvalidierung bereits untersucht wurden. Da eine Untersuchung ja durch den Vergleich einer gelernten mit einer aktuellen Form stattfindet, wird mit jedem Formpunkt auch der erzielte Matchwert verknüpft. Soll im darauffolgenden Zyklus eine gelernte Form mit der dann aktuellen Form gematcht werden, so wird der Formpunkt entsprechend markiert ebenfalls in den aktuellen Objektspeicher eingetragen. Dadurch ist dem Erkennungsmodul im nächsten Zyklus bekannt, mit welcher Form die im Merkmalsmodul extrahierte aktuelle Form verglichen werden soll.

Hypothesengenerierung

Zur Bildung einer Objekthypothese werden, wie bereits erwähnt, die Attraktivitätspunktkonfigurationen der gelernten Objekte mit den aktuellen Punkten verglichen. Der Aufmerksamkeitspunkt, der ja ein lediglich in seiner Wertigkeit veränderter Attraktivitätspunkt ist, dient hier als Referenzpunkt, d. h. er dient als Bezugspunkt für den Vergleich zweier Punktkonfigurationen. Wenn nun ein zuvor gelerntes Objekt präsentiert wird, dann müssen auch die gelernten Attrak-

tivitätspunkte (zumindest teilweise) unter allen aktuellen sein, wobei ein gelernter Punkt also mit dem Aufmerksamkeitspunkt übereinfallen und in Typattribut und Größe übereinstimmen muß, damit die Hypothese für das Vorliegen des Objektes gebildet werden kann (vgl. Abbildung 5.8).

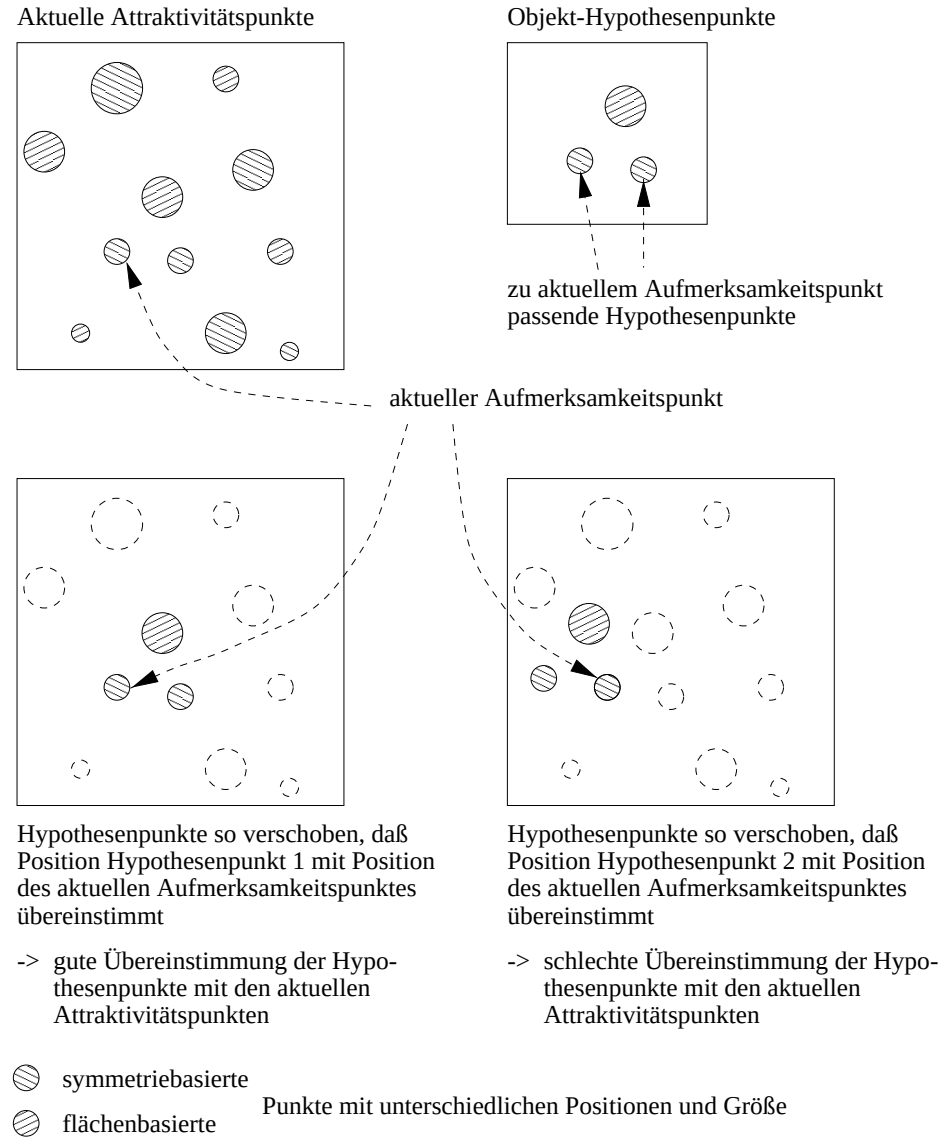


Abbildung 5.8: Punktvergleich zur Hypothesenbildung (aus [Götze 1995])

Den ersten Schritt der Hypothesengenerierung stellt formal die Bildung der folgenden Untermenge $OH' = \{a \in OH \mid a_t = \alpha_t \wedge a_g = \alpha_g\}$ dar, die die zu dem Aufmerksamkeitspunkt α passenden Punkte aus *OH* für ein gelerntes Objekt beinhaltet. Ist die Menge leer, so konnte für das Objekt keine Hypothese gebildet werden. Ansonsten verschieben wir jeweils für einen Punkt aus OH' alle Punkte in *OH* so, daß ersterer Punkt auf dem Aufmerksamkeitspunkt liegt. Anschließend muß festgestellt werden, wieviele der gelernten Attraktivitätspunkte einen kor-

respondierenden, aktuellen Punkt in der näheren Umgebung besitzen. Es sei $a \in OH'$ der Punkt, welcher mit α zur Überdeckung gebracht wurde. Der Hypothesenmatchwert berechnet sich dann aus folgenden Gleichungen:

$$z = \sum_{b \neq a} b_w \cdot c_w \cdot d_4(b, c) \tag{5.1}$$

mit $c \in AT$ so, daß $d_4(b, c) \geq d_4(b, d) \; \forall \; d \in AT$ und

$$n = \sum_{b \neq a} b_w \cdot \begin{cases} c_w \text{ so, daß } 0 \leq d_4(b, c) \geq d_4(b, d) \; \forall \; d \in AT \\ b_w \text{ sonst} \end{cases} . \tag{5.2}$$

Hierin bezeichnet *b* jeden Punkt aus OH', der nicht mit *a* übereinstimmt. Der Quotient *z/n* liefert ein Maß im Intervall [0,1] für die Übereinstimmung zwischen gelernten und aktuellen Attraktivitätspunkten in Abhängigkeit von *a*. Die Distanzfunktion d_4 besitzt ebenfalls einen Verlauf nach Gleichung (4.31).

Es hat sich herausgestellt, daß die Entfernungen der Objekte von der Kamera beim Lernen und späteren Wiedererkennen sehr genau übereinstimmen müssen, um gute Matchwerte zu ergeben. Daher wurde eine entfernungstolerante Erkennung eingeführt, die die Hypothesenpunkte eines Objekts vor dem Match mit den Attraktivitätspunkten innerhalb gewisser Grenzen in ihrer Position und Größe skaliert. Ist z. B. ein Objekt etwas weiter entfernt, so verkleinern sich die Abstände der gebildeten Attraktivitätswerte untereinander. Es muß aber betont werden, daß so nur kleine Variationen in der Entfernung toleriert werden können, da sich ansonsten bei zu großer Skalierung die Hypothesenpunkte der Objekte angleichen.

Das Maximum der Quotienten *z/n* gibt für ein Objekt die beste Verschiebung zur Hypothesenbildung an. Von mehreren gelernten Objekten werden indes die ausgewählt und mit der besten Verschiebung gemerkt, deren jeweiliger Maximalwert einen Schwellwert übersteigt. Ist diese Auswahl leer, so kann zu dem aktuellen Aufmerksamkeitspunkt keine Hypothese gebildet werden. Der Aufmerksamkeitspunkt wird dann mit der Wertigkeit 1 in die Inhibitionskarte übernommen, nachdem alle anderen Punkte in dieser Karte in ihrer Wertigkeit um einen bestimmten Faktor reduziert wurden. Die Verstärkungskarte und der aktuelle Objektspeicher werden gelöscht. Ist die Auswahl dagegen nicht leer, so werden zur Hypothesenbildung eventuell vorhandene, bereits untersuchte Formen herangezogen, deren Matchergebnisse durch die Formpunkte im aktuellen Objektspeicher gemerkt wurden. Sind dort keine Formpunkte enthalten, so wird jenes sich in obiger Auswahl befindliche Objekt als Hypothese ausgewählt, das den höchsten Maximalwert erzielt hat. Sofern sich aber Formpunkte im aktuellen Objektspeicher befinden, wird das Objekt als Hypothese gewählt, deren im Objektspeicher enthaltene Formen den höchsten Durchschnittsmatch erzielt haben. Dieser Durchschnittsmatch muß aber einen Schwellwert überschreiten, ansonsten wird vorherige Hypothese beibehalten.

Zur deutlichen Darlegung des letzten Sachverhaltes sei *OF* die Menge der Formpunkte für ein beliebiges Objekt aus obiger Auswahl und *AF* die Menge der im aktuellen Objektspeicher enthaltenen Formpunkte der bereits untersuchten Formen. Wäre eine Form des Objektes bereits untersucht worden, dann müßte der zugehörige Formpunkt aus *OF* mit einem Formpunkt aus *AF* in seiner vertikalen und horizontalen Position[2], Größe und Typattribut (= Index der Form) übereinstimmen. Die Wertigkeit des passenden Punktes aus *AF* würde dann den Match widerspiegeln, der beim Formvergleich mit der zum damaligen Zyklus aktuellen Form erreicht wurde. Allgemein wird zunächst die Teilmenge von *AF* gebildet, deren Punkte sich auf Formen des Objektes beziehen*:* $AF' = \{a \in AF \mid \exists\, b \in OF$ so, daß $a_t = b_t\}$. Die Mittelwertbildung für den Match *M* ergibt sich dann zu:

$$\begin{aligned} \langle M \rangle &= \frac{M}{AF''} \\ M &= \sum_{a \in OF,\ b \in AF'} b_w \cdot d_5(a,b) \\ AF'' &= \begin{cases} \sum_{a \in OF,\ b \in AF'} d_5(a,b) & \text{für } M > 0 \\ 1 & \text{sonst} \end{cases} \end{aligned} \tag{5.3}$$

Es wird hier zwar die Summe über alle möglichen Kombinationen aus den Mengen *OF* und AF' gebildet, allerdings sorgt die Distanzfunktion d_5 dafür, daß nur Punktkombinationen ungleich Null sind und damit in die Mittelwertbildung eingehen, die im Typattribut exakt und in der Position ungefähr übereinstimmen. Der Nenner sorgt dafür, daß wirklich nur die passenden Punktkombinationen zur Normierung beitragen.

Hypothesenvalidierung

Liegt eine Hypothese im aktuellen Objektspeicher bereits länger als einen Zyklus vor, wird die vom Merkmalsmodul extrahierte, aktuelle Form mit der gelernten Form, welche durch einen markierten Formpunkt im aktuellen Objektspeicher indiziert wurde, verschiebungstolerant gematcht. Das Matchergebnis wird dann im Wertigkeitsfeld des Formpunktes abgelegt. Eine weitergehende Validierung findet nur statt, wenn erstens dieser Matchwert einen Schwellwert überschreitet und sich zweitens der aktuelle Aufmerksamkeitspunkt in der Nähe eines Hypothesenpunktes befindet, da ansonsten davon ausgegangen werden muß, daß sich ein interessanteres Objekt, als das gerade untersuchte, im aktuellen Bild befindet.

Ist die Hypothese im aktuellen oder vorausgegangenen Zyklus generiert worden, werden zunächst alle Punkte der Inhibitionskarte in ihrer Wertigkeit um einen Faktor reduziert, damit ein

[2] Die Formpunkte seien entsprechend der "besten" Verschiebung durch den Hypothesenmatch verschoben

späteres Untersuchen von Objekten in dieser Umgebung wieder möglich wird. Sind alle Formen des hypothesierten Objektes untersucht worden, gilt dieses als erkannt, wenn der durchschnittliche Match der Objektformen größer als der für den Formvergleich vorgegebene Schwellwert ist. Es kann dann der aktuelle Objektspeicher gelöscht werden. Die Wertigkeit der Punkte in der Inhibitionskarte wird verringert und die Hypothesenpunkte mit der Wertigkeit 1 hinzugefügt. Wurden dagegen noch nicht alle Formen des hypothesierten Objektes untersucht, wird mit der Untersuchung fortgefahren. Die nächste noch zu untersuchende Form ist dadurch gekennzeichnet, daß ihr zugehöriger Formpunkt aus *OF* im aktuellen Objektspeicher fehlt. Dieser Punkt wird nun in den aktuellen Objektspeicher geschrieben und für den Match im nächsten Zyklus markiert. Die Position des Formpunktes wird dann an das Kameramodul weitergegeben. Zusätzlich werden die Hypothesenpunkte mit aktualisierten Werten in die Verstärkungskarte eingetragen, damit der im nächsten Zyklus gebildete Aufmerksamkeitspunkt mit einem Hypothesenpunkt des präsentierten Objektes übereinstimmt und so mit der Validierung des gleichen Objektes fortgefahren wird. Abbildung 5.9 zeigt den kompletten Erkennungsvorgang bestehend aus Hypothesengenerierung und -validierung noch einmal schematisch.

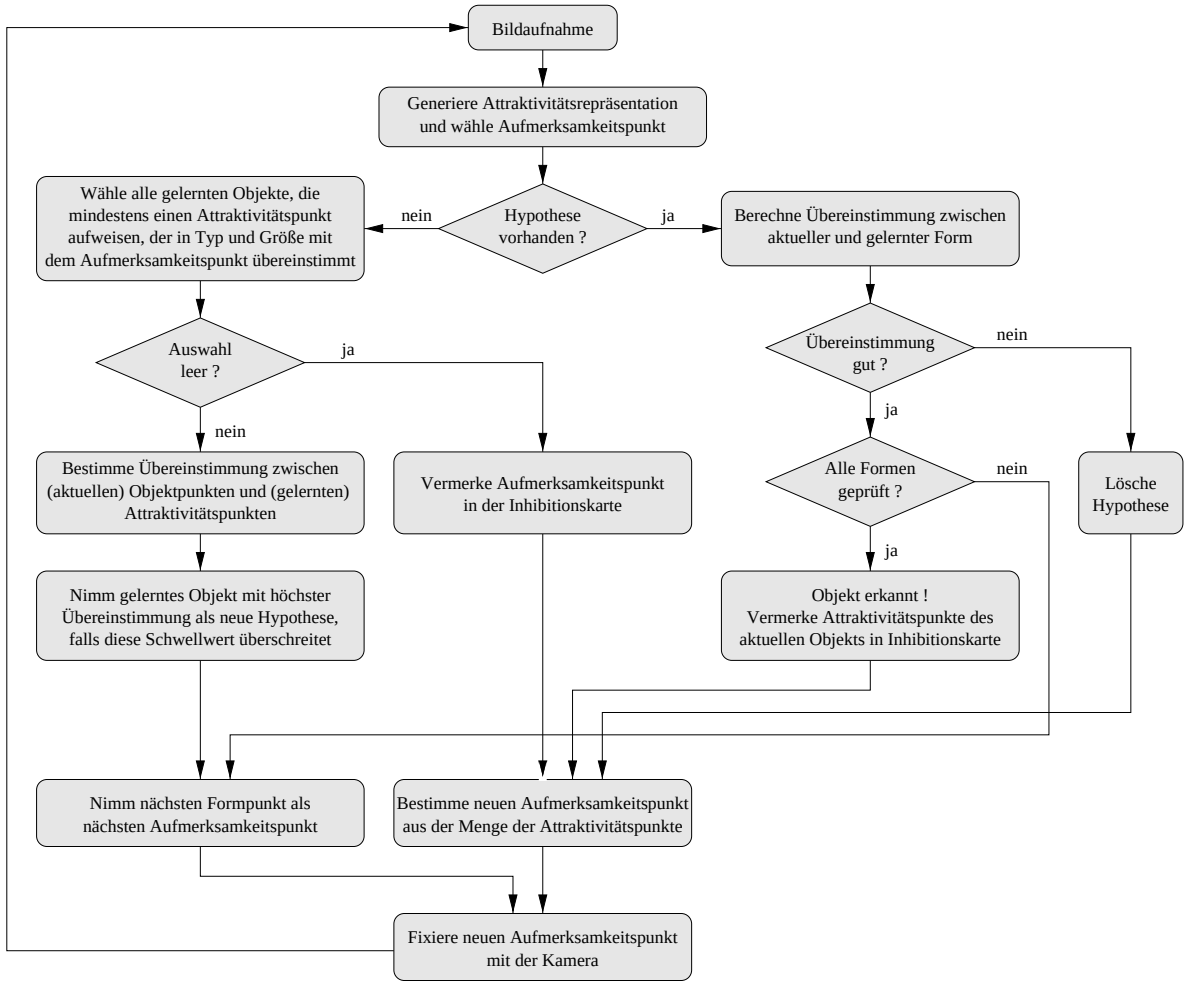


Abbildung 5.9: Erkennungsvorgang

5.3.3 Experimentelle Ergebnisse zur Objekterkennung

Die im folgenden präsentierten Ergebnisse sind [Götze 1996] entnommen. Gelernt wurden insgesamt sechs Objekte, vier Fahrzeuge und zwei von diesen grundsätzlich verschiedene Objekte (Abbildung 5.10).

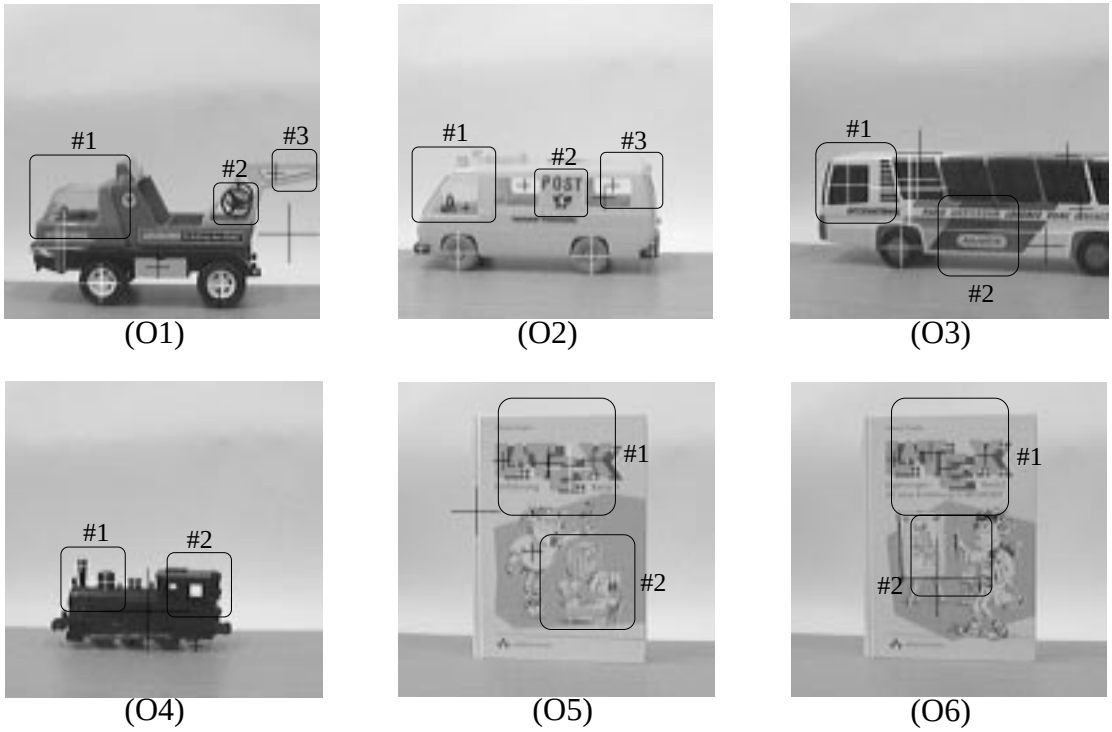


Abbildung 5.10: Gelernte Objekte inklusive gespeicherter Attraktivitätspunkte (schwarze Kreuze repräsentieren flächenbasierte, graue Kreuze symmetriebasierte Punkte) und Formen (Kästen)

Präsentiert wurden anschließend die Objekte O1 und O5 sowohl vor einem homogenen, weißen Hintergrund als auch vor einem strukturierten Hintergrund (Tageszeitung). Der weiße Hintergrund entspricht dabei weitestgehend den Bedingungen während der Lernphase wogegen der strukturierte Hintergrund dem Erkennungssystem unbekannt ist. Das System sucht nach seiner Initialisierung aktiv in seiner Umgebung nach den gelernten Objekten (siehe Abbildung 5.11).

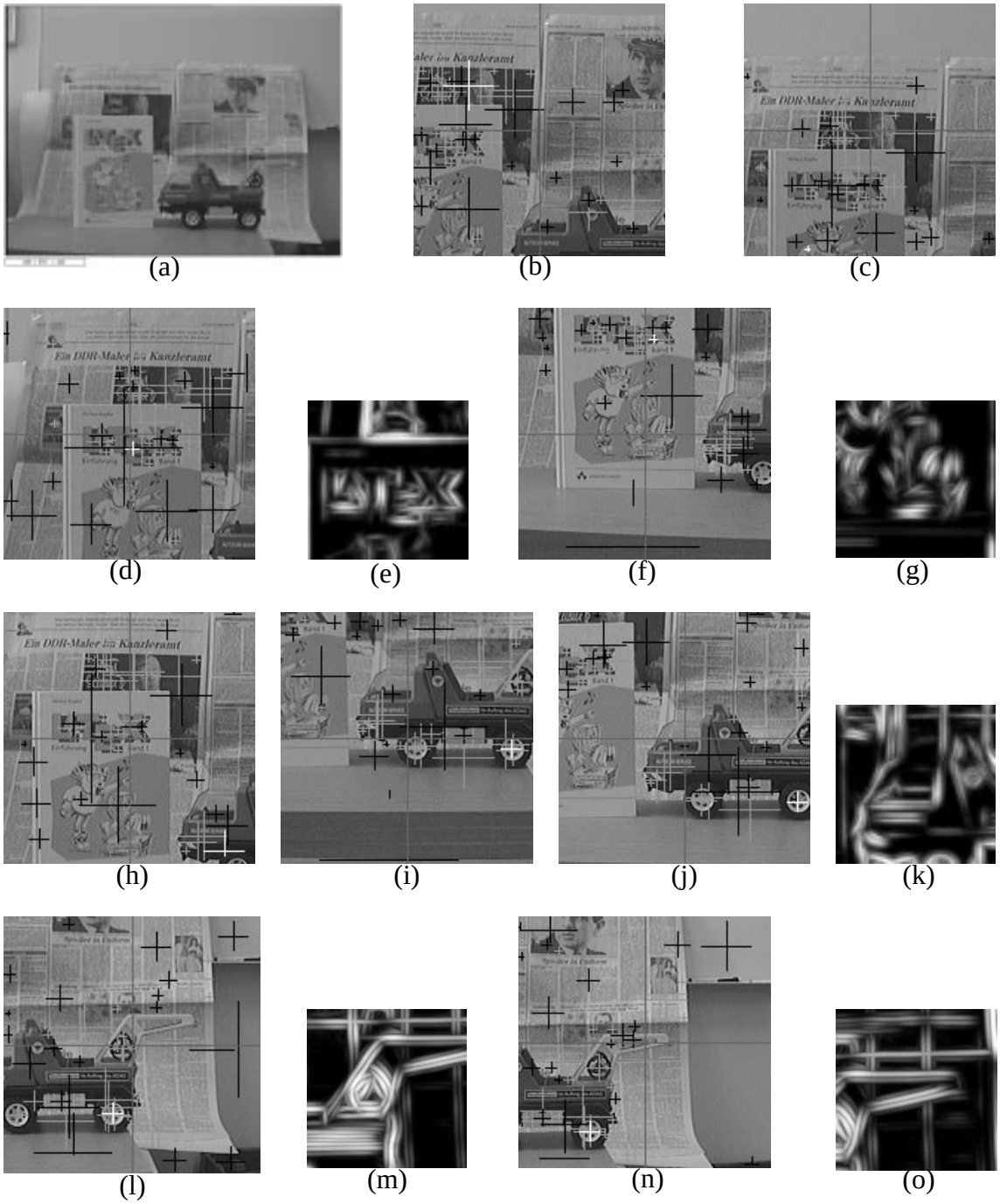


Abbildung 5.11: Experiment E6: (a) gesamte Szene; (b) Attraktivitätspunkte im ersten Bildausschnitt (schwarze Kreuze repräsentieren flächen-, graue Kreuze symmetriebasierte Punkte, das weiße Kreuz den Aufmerksamkeitspunkt), keine Hypothese konnte aufgestellt werden; (c) nach einem Blickwechsel kann die Hypothese O5 aufgestellt werden; (d) die erste vermutete Form wird angefahren und mit der (e) gespeicherten Form gematcht, 97,6%; (f und g) gleicher Vorgang für die zweite Form, Match 94,8%, das Objekt O5 gilt als erkannt; (h) Suche nach neuem Objekt beginnt; (i) Hypothese O1 wird aufgestellt; (j) die erste Form wird angefahren und mit der (k) gespeicherten Form gematcht, 91,8%; (l und m) gleicher Vorgang für die zweite Form, Match 88,8%; (n und o) gleicher Vorgang für die dritte Form, Match 82,9%, O1 gilt als erkannt

Die Tabelle 5.1 faßt die Erkennungsergebnisse zusammen. Aufgeführt sind hier die Schritte bis zur Erkennung, d. h. die Anzahl der Blickwechsel bis eine Objekthypothese bestätigt werden konnte, die Anzahl zuvor aufgestellter und wieder verworfener Hypothesen und ein mittlerer Matchwert für die überprüften Formen gemäß Gleichung (5.3). Wie zu erwarten war, benötigt das System mehr Schritte, bis es ein einzelnes Objekt vor strukturiertem Hintergrund gefunden und verifiziert hat als vor weißem Hintergrund. Ist dagegen mehr als ein Objekt in der Szene, ist die Suche nach einem Objekt und dessen Verifikation selbst bereits sehr viel schwieriger und die Struktur des Hintergrundes beeinflußt den zeitlichen Verlauf des Erkennungsvorgangs sehr viel weniger. Die benötigte Rechenzeit bis zur Generierung einer Objekthypothese liegt auf einer general-purpose Workstation bei etwa einer Minute, die Validierung einer 200x200 Pixel großen Form bei etwa zwei Minuten. Die gesamte Rechenzeit hängt im allgemeinen sowohl von der Größe der Formen als auch der Anzahl der zu einem gelernten Objekt gespeicherten Formen ab.

Tabelle 5.1: Erkennungsergebnisse

Exp.	Beschreibung	Schritte bis zur Erkennung	zurück-gewiesene Hypothesen	Match-wert
E1	Objekt O1 vor weißem Hintergrund	5	0	0,865
E2	Objekt O1 vor einer Tageszeitung	6	0	0,947
E3	Objekt O5 vor weißem Hintergrund	3	0	0,855
E4	Objekt O5 vor einer Tageszeitung	14	0	0,840
E5	Objekte O1 und O5 vor weißem Hintergrund	10	1	0,941 bzw. 0,910
E6	Objekte O1 und O5 vor einer Tageszeitung (siehe Abbildung 5.11)	9	0	0,697 bzw. 0,878

5.4 Bewegungsdetektion und Objektverfolgung

Die Blicksteuerung besteht in der bisher beschriebenen Form aus einer bilddatengetriebenen Komponente, die eine Attraktivitätsrepräsentation erzeugt, und einer modellgetriebenen Komponente, die die zu einer Objekthypothese gelernten Objektteile im Bild sucht und vergleicht. Allerdings ist die Erkennung zweidimensionaler Objekte nicht das einzige in NAVIS inhärente Modell. Ein weiteres Modell ist die in [Springmann 1998] beschriebene Objektverfolgung (vgl. Abbildung 5.6). Die Blicksteuerung benötigt daher eine dritte Komponente, die situationsabhängig jeweils eines der beiden Modelle vorgibt. Das heißt eine Komponente, die abhängig

vom Bildinhalt entscheidet, welches Modell besser zur Analyse der Szene geeignet ist. Diese Komponente bezeichne ich im folgenden als *Verhaltenssteuerung*. Die Verhaltenssteuerung soll z. B. auf unerwartete Ereignisse reagieren, Kollisionen vermeiden helfen oder eventuell auftretende Deadlocks lösen. Hierzu sind die verschiedenen in NAVIS integrierten Modelle unterschiedlich zu priorisieren. In dieser Arbeit werden zunächst nur die beiden Modelle Objekterkennung und Objektverfolgung unterschieden. Die Objektverfolgung soll hier die höhere Priorität besitzen und den auf statischen Merkmalen basierenden Erkennungsvorgang unterbrechen, sobald Bewegung detektiert wird (siehe Alarmsystem in Abbildung 5.6).

Formal wird also durch die Verhaltenssteuerung die Verhaltensklasse "Erkennung" durch die Verhaltensklasse "Tracking" ersetzt. Hierzu ist es nicht notwendig, die Bewegungsparameter zu schätzen, sondern es reicht aus, überhaupt irgendeine Form von Bewegung zu registrieren. Deshalb spricht man bei dieser Form der Bewegungsdetektion auch von einem Alarmsystem. Ein derartiges Verhalten macht für ein aktives Sehsystem Sinn, weil ein neu in die Szene eintretendes Objekt neue Merkmale in diese einbringen kann, die den Erkennungsvorgang stören können, und weil das neue Objekt relevante Merkmale des zu validierenden Objektes verdekken kann.

Abbildung 5.12 zeigt das erweiterte Modell der Aufmerksamkeitssteuerung (vgl. Abbildung 4.1). Aus dieser Abbildung läßt sich ablesen, daß die Attraktivitätsrepräsentation keinen weiteren Einfluß auf die Objektverfolgung hat, sobald diese einmal initialisiert wurde. Damit unterscheidet sich die Verhaltensklasse "Tracking" sehr wesentlich von der Verhaltensklasse "Erkennung", die nach jedem Blickwechsel eine neue Attraktivitätsrepräsentation anfordert.

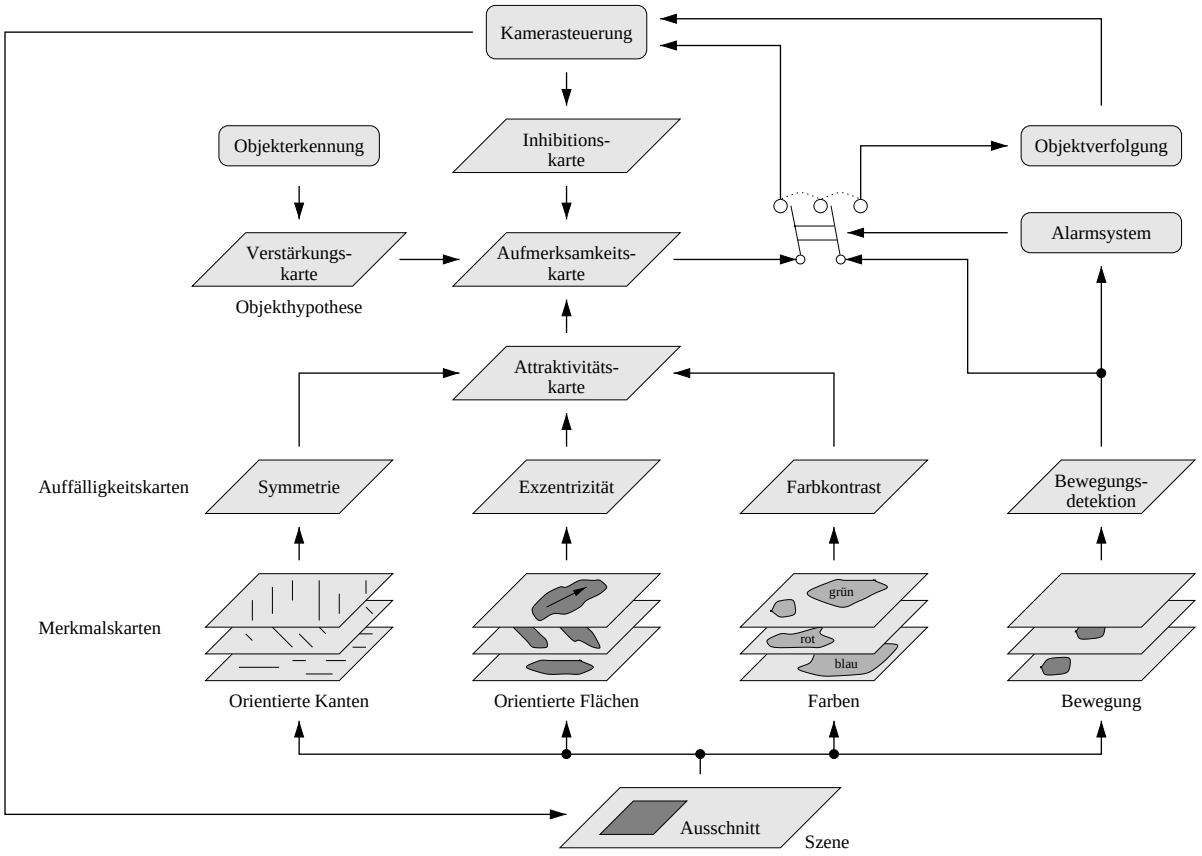


Abbildung 5.12: Erweitertes Schema der Aufmerksamkeitssteuerung unter Berücksichtigung dynamischer Merkmale

Die Bewegungsdetektion muß während des Erkennungsvorgangs ständig im Hintergrund laufen und sollte daher möglichst wenig Ressourcen verbrauchen. Da während des Erkennungsvorgangs natürlich auch der Kamerakopf bewegt wird, z. B. um die zu einer Objekthypothese gehörenden Formen anzufahren, muß während einer solchen Eigenbewegung des Systems die Bewegungsdetektion unterbrochen werden. Versuche hierzu wurden bereits in [Bollmann 1997] beschrieben.

5.4.1 Bewegungsdetektion

Wohin wir unsere visuelle Aufmerksamkeit richten, hängt sehr wesentlich auch von dem Stimulus Bewegung ab. Insbesondere wenn wir Bewegung in der Peripherie unseres Sehfeldes wahrnehmen (extrafoveales Sehen), wird unser Blick nahezu reflektorisch auf das sich bewegende Objekt gerichtet. Dieses wird in der Regel wenigstens so lange verfolgt, bis es von uns erkannt wurde. In technischen Systemen besitzt die Objektverfolgung insbesondere Bedeutung für die Szenenüberwachung, die Kollisionsvermeidung sowie die Stabilisierung und Fixation

eines bewegten Objekts im Bildmittelpunkt. Um allerdings den Tracking-Vorgang überhaupt auslösen zu können, ist in jedem Fall die Allokation von Aufmerksamkeit notwendig. Abbildung 5.13 zeigt das Schema der Bewegungsdetektion in NAVIS.

Da die Bewegungsdetektion bis auf kurze Unterbrechungen während der Blickwechsel ständig im Hintergrund laufen muß, sollte sie nicht sehr rechenintensiv sein; sie ist daher durch eine einfache spatio-temporale Filterung gemäß dem Ansatz von Wiklund et al. [Wiklund 1987] realisiert worden. In einem ersten Schritt werden drei aufeinanderfolgende Bilder $I(x, y, t_i)$, $i = k-1, k, k+1$ durch eine 'Nearest Neighborhood'-Interpolation um den Faktor 2 unterabgetastet und jeweils paarweise subtrahiert:

$$D_i(x, y) = \left| I(x, y, t_{i+1}) - I(x, y, t_i) \right| , \; i = k-1, k \tag{5.4}$$

Die resultierenden Differenzbilder D_k und $D_{k\text{-}1}$ werden anschließend geglättet, um kleinere Lücken zu schließen, und binarisiert, so daß zwei homogene Bewegungsmasken entstehen, die anschließend zu einer einzigen Bewegungsmaske M_k logisch UND-verknüpft werden:

$$M_k(x, y) = \begin{cases} 1 & (D_k(x, y) > \Theta) \wedge (D_{k-1}(x, y) > \Theta) \\ 0 & \text{sonst} \end{cases} \tag{5.5}$$

Die Größe Θ bezeichnet hierin den Schwellwert für die Binarisierung der Differenzbilder D_k und $D_{k\text{-}1}$. Die Wahl des Schwellwertes ist bei einer Binarisierung unkritisch.

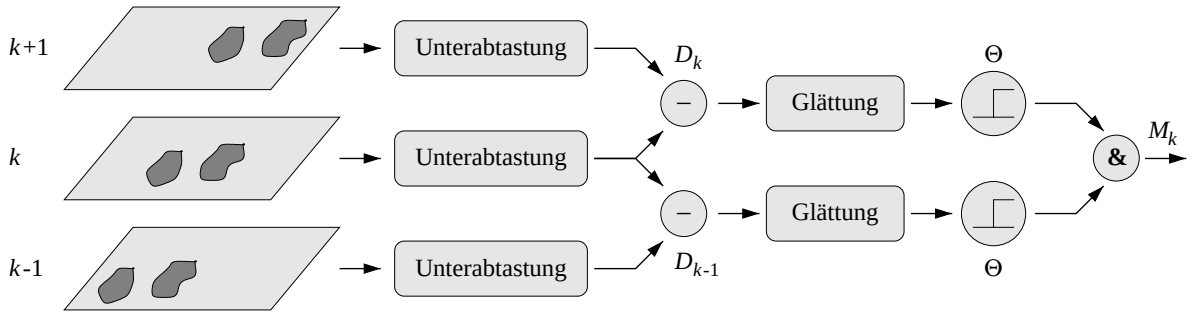


Abbildung 5.13: Schema der Bewegungsdetektion

Das Zentrum der Bewegungsmaske M_k liefert den nächsten Aufmerksamkeitspunkt, der als Startwert für die Objektverfolgung dient. Abbildung 5.14 zeigt ein Beispiel. Der sich rechts in der Szene befindende Spielzeug-Lkw wird während der in Abbildung 4.19 dargestellten Szenenanalyse über eine Fernsteuerung in Bewegung gesetzt.

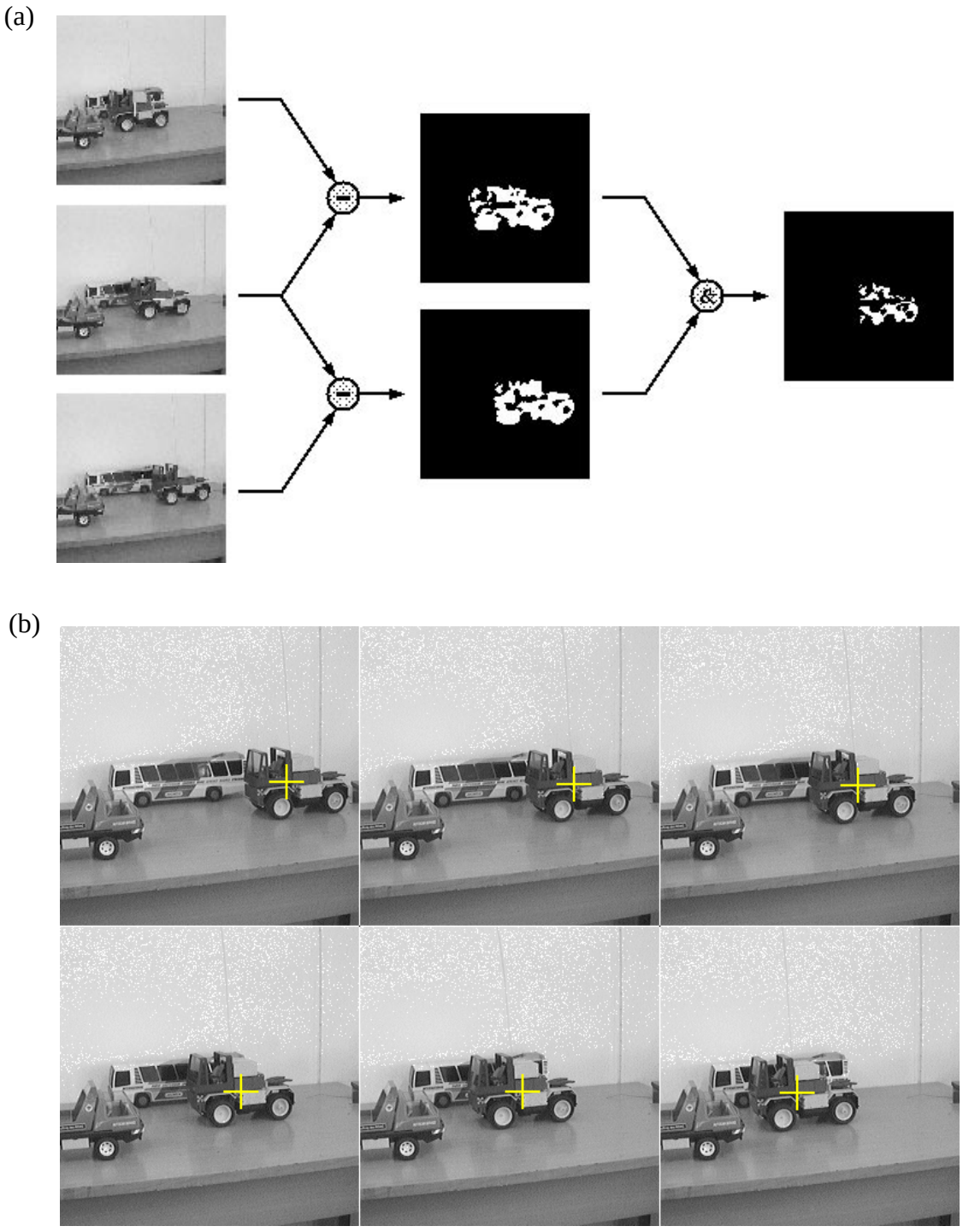


Abbildung 5.14: Beispiel einer Bewegungsdetektion: (a) Prinzip der Berechnung der Bewegungsmaske, (b) Ergebnisse

Der Algorithmus setzt allerdings voraus, daß das sich bewegende Objekt ausreichend stark texturiert ist. Andernfalls ist die Differenz zwischen den aufeinanderfolgenden Bildern im Bereich des Objektes kleiner als der Schwellwert Θ, so daß eine stark zerrissene Bewegungsmaske ent-

steht, die, insbesondere wenn man mehrere sich bewegende Objekte zuläßt, zu Fehlinterpretationen führen kann.

5.4.2 Merkmalskorrespondenz

Die Lösung des *Korrespondenzproblems* ist ein zentraler Punkt in vielen Anwendungsbereichen aktiver Sehsysteme, so z. B. der Stereobildauswertung (Korrespondenz zwischen linkem und rechtem Kamerabild) oder der Objekterkennung (Korrespondenz zwischen gelerntem und präsentiertem Objekt). Für die Objektverfolgung besteht das Korrespondenzproblem in der Zuordnung von Bildstrukturen in aufeinanderfolgenden Bildern. Im allgemeinen ist die Merkmalsgewinnung für die Verfolgung bewegter Objekte so zu gestalten, daß die Lösung des Korrespondenzproblems möglichst einfach wird. Wie bereits erwähnt, basiert die Objektverfolgung in NAVIS auf den Gaborfilterantworten. Einfache Merkmale erzeugen zwar häufig ein dichtes Flußfeld, sie erhöhen aber auch die Wahrscheinlichkeit einer Fehlzuweisung. Lokale, nur eine Pixelposition einschließende Merkmale wie Gaborjets (siehe z. B. [Buhmann 1991]) scheinen daher eher ungeeignet zu sein. Höherwertige Merkmale verringern dagegen die Wahrscheinlichkeit einer Fehlzuweisung, erzeugen aber auch nur ein relativ dünn besetztes Flußfeld. Ein entscheidendes Problem bei der Lösung des Korrespondenzproblems stellt die Objekt-Hintergrund-Trennung dar. Hier bieten Gaborfilter den wichtigen Vorteil, in Bereichen hoher Kontraste, wie sie häufig an Objektkanten auftreten, hohe Aktivitäten zu produzieren. Die entstehenden Aktivitätswolken geben einen Hinweis auf die Struktur der betrachteten Objekte. So entstehen nur in seltenen, ungünstigen Fällen zusammenhängende Gebiete hoher Aktivierung, die unterschiedlichen Objekten zuzuordnen sind. Um eine Aktivitätswolke als Merkmal nutzbar zu machen, muß sie als Ganzes erfaßt werden und in eine parametrische Beschreibung überführt werden. Hier kommt ein einfacher Backtracking-Algorithmus zur Segmentierung des Filterergebnisses zum Einsatz: ein rekursives Bereichswachstum in der Achterumgebung eines Startpixels mit einem benutzerdefinierten Schwellwert. Die Wahl der Höhe des Schwellwertes ist relativ unkritisch, da die Flanken der Aktivitätswolken an gut ausgebildeten Strukturen steil sind. Nach der Segmentierung werden die Aktivitätswolken bezüglich ihrer Orientierung, Größe, maximalen Aktivität, Elongation und des Schwerpunktes analysiert (siehe Abbildung 5.15). Durch die Überführung der ursprünglichen Filterantworten in eine parametrische Beschreibung ergibt sich eine starke Reduktion der Datenmenge sowie eine gute Unterscheidbarkeit der generierten Merkmale. Zusätzlich führt diese Vorgehensweise zur Zusammenfassung solcher Bildbereiche, die eine einzige physikalische Struktur repräsentieren.

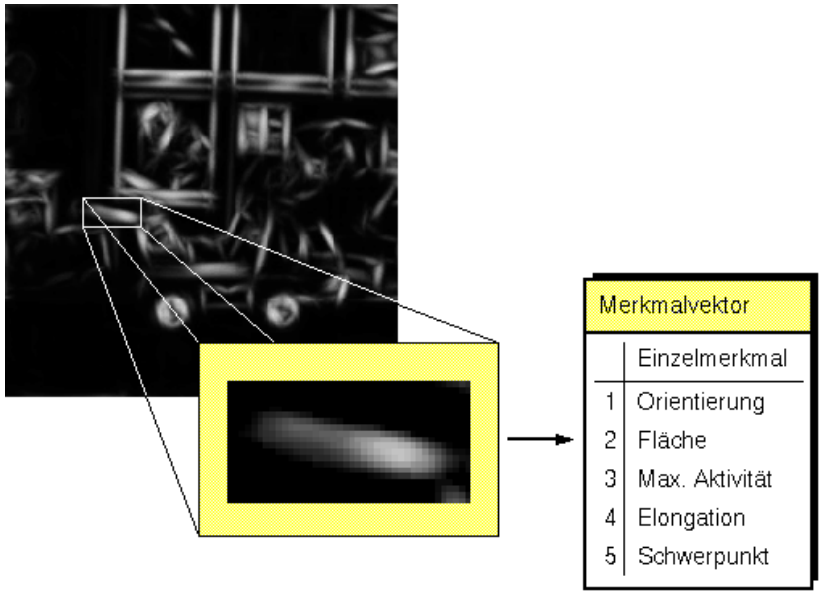


Abbildung 5.15: Gewinnung komplexer Merkmale

Das implementierte Tracking-Verfahren beruht auf der Zuordnung der parametrischen Merkmale in aufeinanderfolgenden Frames mittels eines Ähnlichkeitsmaßes. Unter der Annahme, daß sich die Eigenschaften einzelner Merkmale von Frame zu Frame nicht wesentlich ändern, kann der Suchraum für die Korrespondenz eines Merkmals zusätzlich eingeschränkt werden. Die in dem Objektverfolgungsmodul von NAVIS verwendete Zuordnungsstrategie ist eine Erweiterung der in [Korries 1985] vorgestellten Plausibilitätsbereinigung (Abbildung 5.16):

- Einem Merkmal eines Frames wird höchstens ein korrespondierendes Merkmal im folgenden Frame zugeordnet.
- Ist ein Merkmal bereits mit einem anderen verknüpft und es findet sich zu diesem ein ähnlicheres, wird die ursprüngliche Verknüpfung gelöst und eine neue geschaffen.

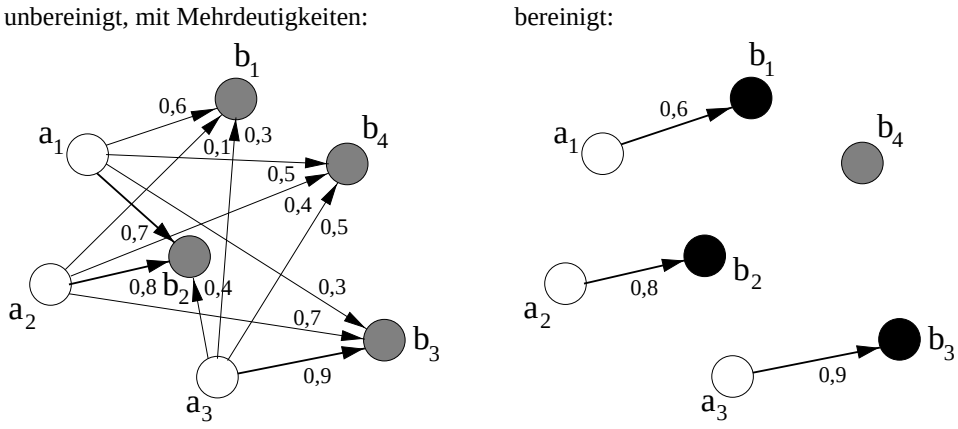


Abbildung 5.16: Korrespondenzanalyse (aus [Hoischen 1999])

Abweichend von der ursprünglichen Definition der Plausibilitätsbereinigung nach Korries werden im hier vorgestellten System im Verlauf der Zuordnung mehrere Iterationen durchgeführt, um Merkmale, deren Zuordnung im Verlauf der Plausibilitätsbereinigung aufgelöst wurden, in einem weiteren Durchlauf den noch verbliebenen Merkmalen neu zuweisen zu können. In der Praxis zeigt sich, daß meist schon nach drei Iterationen keine neuen Verknüpfungen mehr entstehen. Die Zuordnung in dieser Form nutzt kein Vorwissen über die Kamerabewegung oder bereits gewonnene Ergebnisse aus vorherigen Frames. Das Verfahren kann dahingehend erweitert werden, daß solche Verknüpfungen bevorzugt werden, die durch die kamerainduzierte Hintergrundbewegung erklärbar sind.

Eine Menge von Merkmalen eines Frames wird formal mit *AF* bezeichnet. Ein Element *a* dieser Menge besteht aus dem Sieben-Tupel aus Orientierung, Größe, maximaler Pixelwert, Länge und Breite des Segments sowie horizontale und vertikale Position des Schwerpunktes, das heißt $a = (a_o, a_g, a_m, a_l, a_b, a_h, a_v)$. Ein Vergleich zwischen zwei Merkmalen findet nun nur statt, wenn diese in etwa die gleiche Orientierung aufweisen. Als Ähnlichkeitsmaß *M* zwischen zwei Merkmalen *a* und *b* zweier aufeinanderfolgender Frames läßt sich dann das Produkt aus den Verhältnissen der übrigen Parameter definieren:

$$M(a, b) = \begin{cases} \prod_{i \in \{g, m, l, b, h, v\}} \frac{\min(a_i, b_i)}{\max(a_i, b_i)} & , \text{ falls } a_o = b_o \\ 0 & , \text{ sonst} \end{cases} \tag{5.6}$$

Die Ähnlichkeit zweier Merkmale liegt also immer im Intervall zwischen Null und Eins. Die Suche nach Korrespondenzen wird zusätzlich auf einen gewissen Ausschnitt begrenzt. Der beschränkte Suchradius verhindert Zuordnungen über Distanzen, die Objektgeschwindigkeiten entsprechen, die für das System nicht handhabbar sind. Weiterhin verhindert das Ausblenden von dicht am Rand des Suchradius liegenden Merkmalen Fehler, die durch die Analyse nur teilweise erfaßter Merkmalsstrukturen hervorgerufen würden.

5.4.3 Trennung Objekt-Hintergrundbewegung

Sind, wie in Abschnitt 5.4.2 beschrieben, Zuordnungen für die extrahierten Merkmale gefunden, ergeben sich aus den Differenzen ihrer Positionen die gesuchten Verschiebungsvektoren. Die Gesamtheit der Verschiebungsvektoren bildet das *Verschiebungsvektorfeld*. Um vom Verschiebungsvektorfeld auf im Bild vorhandene bewegte Objekte zu schließen, muß eine Trennung der Verschiebungsvektoren in solche des Hintergrundes und solche bewegter Objekte erfolgen. Bei vergleichsweise geringen Kamerabewegungen kann die Hintergrundbewegung durch eine Parallelverschiebung approximiert werden. Auch für die bewegten Objekte soll hier vorausgesetzt werden, daß sich ihre Merkmale annähernd parallel verschieben. Solange diese

Bedingung zumindest für einige Merkmale zutrifft, kann das zugehörige Objekt detektiert werden.

Geht man von einer in etwa gleichartigen Verschiebung der Merkmale des bewegten Objektes einerseits und des Hintergrundes andererseits aus, so bieten sich verschiedene Möglichkeiten an, die zum Objekt gehörenden Verschiebungsvektoren gegenüber denen des Hintergrundes abzugrenzen. Sind die zu detektierenden Objekte kompakt, ergibt sich ein bereichsweise homogenes Verschiebungsvektorfeld, welches durch relaxationsbasierte Segmentierung in Bereiche ähnlicher Bewegung aufgeteilt werden kann. Entstehen so mindestens zwei hinreichend große Bereiche gleichartig verschobener Merkmale, ist von diesen einer mittels Heuristiken als bewegtes Objekt und der andere als Hintergrund zu interpretieren. So enthält der Hintergrund meist mehr Randanteile des Bildes und nimmt eine insgesamt größere Fläche ein. Eine sichere Zuordnung gelingt aber nur, wenn Informationen über die Kamerabewegung vorliegen. Ist das zu detektierende Objekt allerdings zerklüftet oder teilweise durchsichtig, ist ein Relaxationsverfahren kritisch. Hier bietet sich der Einsatz eines histogrammbasierten Ansatzes an. Durch die Berechnung eines zweidimensionalen Histogramms über alle Verschiebungsvektoren lassen sich gleichartig verschobene Merkmale detektieren. Hintergrund und Objekt finden sich als lokale Maxima im Histogramm.

5.4.4 Tracking

Das Ziel einer Objektverfolgung ist die Koordination der Bewegung eines Kamerasystems und der eines bewegten Objekts. Um die Kohärenz zwischen beiden Bewegungen zu gewährleisten, ist es notwendig, die Abbildungseigenschaften des Systems hinsichtlich der Projektion der Szene auf die Sensoroberfläche zu kennen. Jedes Objekt im dreidimensionalen Raum hat sechs Lagefreiheitsgrade (drei Translations- und drei Rotationsfreiheitsgrade), wogegen die Kameraplattform eines aktiven Sehsystems, sofern sie nicht auf eine mobile Einheit montiert ist, nur Rotationsfreiheitsgrade besitzt. Die Rotationsbewegungen der Kameraplattform sind zudem durch die Kinematik eingeschränkt. Die Abbildungseigenschaften des NAVIS-Stereokamerakopfes sind in [Springmann 1998] untersucht worden, worauf an dieser Stelle verwiesen werden soll.

Ein bewegtes Objekt tritt entweder vom Bildrand in den betrachteten Szenenausschnitt ein oder befindet sich bereits dort. Von einem Tracking-System wird gefordert, daß es das bewegte Objekt fixiert und über mehrere Frames, etwa bis zu seiner Erkennung, im Bildzentrum hält. Ist das Objekt an der gewünschten Position abgebildet, wird dessen geschätzte Eigenbewegung durch die Folgebewegung des aktiven Sehsystems ausgeglichen. Korrekturen sind nötig, wenn die Schätzung die Bewegung des Objektes nicht kompensieren konnte. Dann wird die langsame Folgebewegung durch eine Sakkade, d. h. einen "Sprung" zu der aktuellen Objektposition, abgelöst. Nach dem erstmaligen Fixieren eines bewegten Objekts besteht die Aufgabe des Tracking-Systems in der Aufrechterhaltung einer langsamen Folgebewegung. Idealerweise

gleicht diese Bewegung die Objektbewegung vollständig aus. Es wird also eine Bewegungsprädiktion benötigt, die die Position des Objekts im nächsten Frame vorhersagt, da jeder Tracking-Algorithmus bedingt durch seine benötigte Rechenzeit immer verzögert reagiert. Zur Prädiktion der zukünftigen Objektposition wird in NAVIS ein diskretes Kalmanfilter verwendet, das ein explizites Bewegungsmodell voraussetzt. Das zugrundegelegte Bewegungsmodell ist hier eine konstant beschleunigte Bewegung. Das eingesetzte Kalmanfilter ist demnach linear und stellt ein optimales Filter für die Minimierung der Differenz zwischen geschätzter und tatsächlicher Position eines Objekts über der Zeit bei Anwesenheit von Rauschen dar. Messungen der Zustandsvariablen, die ebenfalls von Rauschen beeinflußt werden, werden genutzt, um rekursiv die Zustandsschätzung zu aktualisieren und zu verbessern.

Das in NAVIS implementierte Tracking-Verfahren zeigt Abbildung 5.17. Die Bewegungsdetektion initialisiert die Objektverfolgung und unterbricht so auch eine eventuell bereits stattfindende Erkennung. Das nächste Bild, das erste für die Objektverfolgung, wird mittels Gaborfilterbank mit konstanter Schwerpunktortsfrequenz und verschiedenen Orientierungen gefiltert. Die entstehenden Kantenbilder werden einer Segmentierung mit einem Backtracking-Algorithmus unterzogen und in eine Parameterbeschreibung überführt.

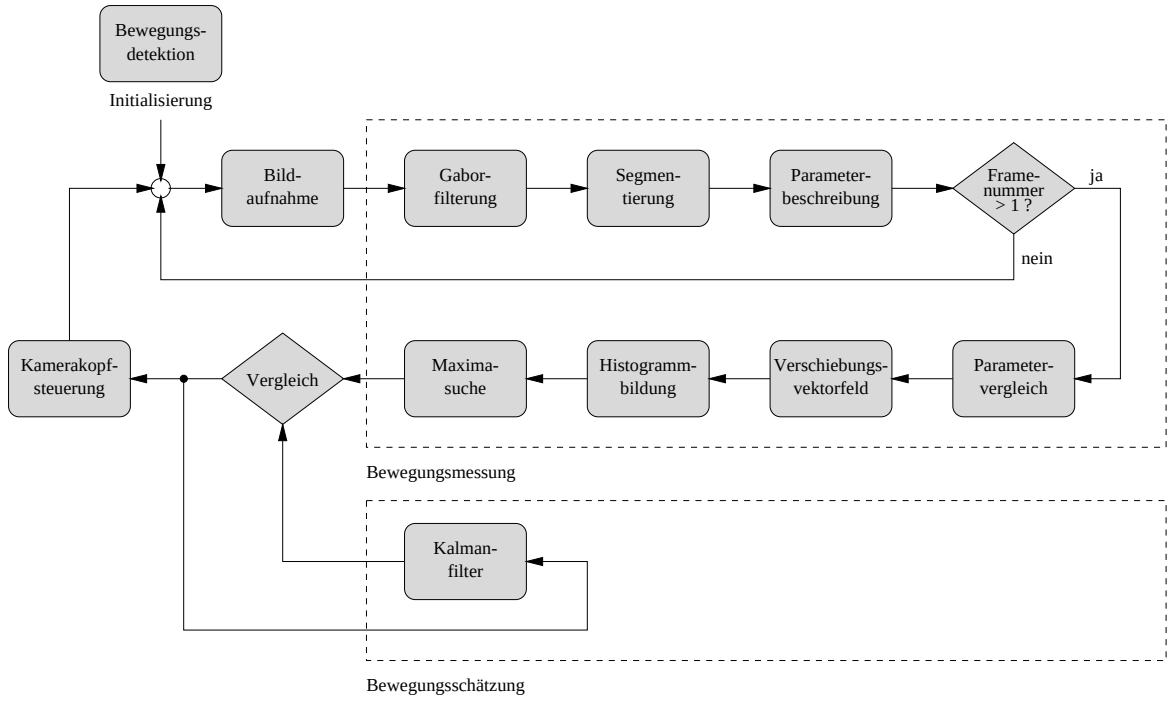


Abbildung 5.17: Blockdiagramm des Tracking-Algorithmus

Befindet sich das System in der Initialisierungsphase, d. h. es wurde bisher erst ein Frame verarbeitet, so wird nun unmittelbar ohne weitere Auswertung ein zweites Bild aufgenommen und ebenfalls in eine Parameterbeschreibung überführt. Stehen dann die Parameterbeschreibungen zweier aufeinanderfolgender Frames zur Verfügung, führt die in Abschnitt 5.5.1 beschriebene

plausibilitätsbereinigte Merkmalszuordnung unmittelbar durch die Differenzen der Schwerpunkte einander zugeordneter Merkmale zu einem Verschiebungsvektorfeld. Um aus dem Verschiebungsvektorfeld die Information sowohl über die durch die Eigenbewegung der Kamera erzeugte Verschiebung des Hintergrundes als auch die Positionsänderung des zu verfolgenden Objektes gewinnen zu können, wird über alle gefundenen Verschiebungsvektoren ein zweidimensionales Histogramm berechnet. Die Verschiebungsvektoren werden hierbei mit dem Ähnlichkeitswert der einander zugeordneten Merkmale bewertet. Lokale Maxima im Verschiebungsvektorhistogramm repräsentieren die Gruppen gleichartiger Verschiebungsvektoren. Nach der Initialisierung findet sich ein lokales Maximum für den Nullvektor. Dieser beschreibt die Hintergrundbewegung, die aufgrund der noch nicht bewegten Kamera gleich Null sein muß. Findet sich ein weiteres lokales Maximum im Histogramm, das deutlich über dem durch Fehlzuweisungen entstehenden Rauschen liegt, wird dieses als Bewegung eines Objekts interpretiert. In den folgenden Durchläufen entsteht eine von Null abweichende Hintergrundverschiebung. Um nun die vom Objekt erzeugte Gruppe von Verschiebungsvektoren von der des Hintergrundes unterscheiden zu können, wird aus der Kamerabewegung und der Kinematik des Kamerakopfes die voraussichtliche Hintergrundbewegung bestimmt. Die Vektorgruppe, deren Schwerpunkt einen geringeren Euklidischen Abstand zur erwarteten Hintergrundverschiebung besitzt, wird als Hintergrund interpretiert. Die verbleibende größte Vektorengruppe beschreibt die Objektbewegung.

Nachdem nun die Verschiebung eines bewegten Objekts über zwei Frames bekannt ist, kann seine Lage bestimmt werden. Hierzu sind die Merkmale, deren Verschiebungsvektoren mit der berechneten Verschiebung übereinstimmen, zu selektieren und einer Schwerpunktsbestimmung zu unterziehen. Die so entstehenden Merkmalsschwerpunkte geben allerdings u. U. aufgrund von Verdeckungen, Schatten oder teilweise geringem Objekt-Hintergrund-Kontrast nur sehr ungenau die Position des bewegten Objektes wieder. Deshalb wird diese Berechnung nur zur Ausführung der Korrektursakkaden genutzt, wenn also eine relevante Differenz zwischen vorausgesagter und tatsächlicher Objektposition besteht. Eine Abweichung gilt dann als relevant, wenn der berechnete Schwerpunkt nicht in einem Fenster um das Bildzentrum mit einer Kantenlänge von einem Drittel der Framegröße liegt. Durch die relativ große Kamerabewegung bei der Ausführung einer Sakkade gehen viele Merkmale verloren, die im vorhergegangenen Frame ermittelt wurden, während zahlreiche neue hinzukommen. Eine Zuordnung von Merkmalen kann dann zu beträchtlichen Fehlern führen. Daher wird nach einer Sakkade zur Steuerung der Folgebewegung nicht das Verschiebungsvektorfeld ausgewertet, sondern das Ergebnis der jetzt stark korrigierten Bewegungsprädiktion verwendet. Korrektursakkaden bilden allerdings die Ausnahme. Im Normalfall erfolgt eine Folgebewegung anhand der gut zutreffenden und daher von Frame zu Frame nur wenig korrigierten Bewegungsprädiktion.

Die vollständigen Handlungszyklen aus Detektion der Objektbewegung, Bewegungsprädiktion und Nachführung der Kamera werden bis zum Erreichen einer Abbruchbedingung wiederholt. Diese Abbruchbedingung muß von der Blicksteuerung vorgegeben werden. Mögliche Abbruchkriterien könnten die Erkennung des Objektes sein oder die Detektion eines weiteren be-

wegten Objekts. Dieses Objekt würde eine Neuigkeit darstellen und würde dadurch womöglich die Aufmerksamkeit auf sich lenken.

Experimentelle Ergebnisse, die den Einfluß der Bewegungsform des Objektes und der Parametrisierung des Tracking-Algorithmus auf das Verfolgungsmodul zeigen, sind in [Springmann 1998] beschrieben.

5.5 Roboter-Szenarium

Mit der wachsenden Forderung nach flexiblen und weitestgehend autonomen Robotern richtet sich das Forschungsinteresse mehr und mehr auf aktive Sehsysteme. Robotersehen ist daher nicht nur als Navigationshilfe geeignet, sondern auch zur Lösung von komplexen Aufgaben aus dem Bereich der Mustererkennung. So eröffnet die Kombination von Seh- und Robotersystemen Möglichkeiten für eine Vielzahl neuer Anwendungen (siehe z. B. [Becker 1995], [Buhmann 1995], [Bayouth 1996], [Vestli 1996], [Maurer 1997], [Oswald 1997], [Vetter 1997]). Zusammengefaßt sollte ein mobiles aktives Sehsystem die folgenden Anforderungen erfüllen:

- Lokalisierung möglicherweise aufgabenrelevanter Objekte
- Identifizierung und Priorisierung der Objekte
- autonome Navigation zu den aufgabenrelevanten Objekten, um eine gewünschte Manipulation durchzuführen

Die Erfüllung dieser Anforderungen in einem integrierten System ist eine schwierige Aufgabe, die für ein spezielles Szenarium, das Dominospiel, in der AG IMA gelöst wurde (Abbildung 5.18). Diese Fallstudie erlaubt erste Untersuchungen, wie das bezüglich Anwendungen universell konzipierte aktive Sehsystem NAVIS und speziell die Blicksteuerung an eine konkrete Anwendung adaptiert werden kann.

Neben mir (Bildverarbeitung und Navigation) haben auch mein Kollege Rainer Hoischen (Navigation und Client/Server-Verbindungen) sowie die Studierenden Christoph Justkowski (Kalibrierung) und Michael Jesikiewicz (Bildverarbeitung) einen Anteil an der Entstehung dieser Fallstudie (siehe auch [Bollmann 1999]).

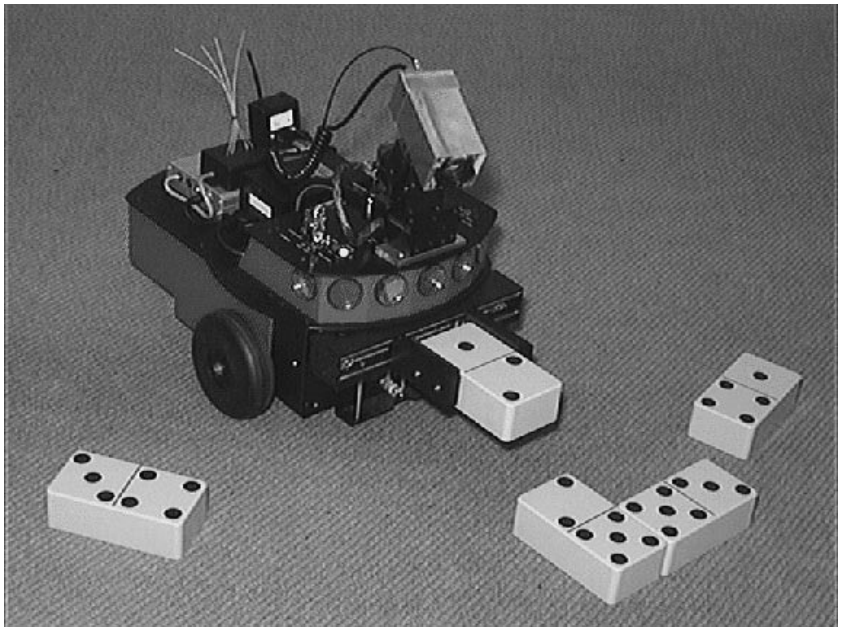


Abbildung 5.18: Ein Domino spielender Roboter als Fallstudie eines aktiven Sehsystems

5.5.1 Systemarchitektur

Das System besteht aus verschiedenen Software-Modulen und dem bereits in Abschnitt 5.1 beschriebenen mobilen Roboter vom Typ Pioneer1. Für das Domino-Szenarium wird der Roboter ausschließlich visuell gesteuert, d. h. insbesondere die sieben Sonarsensoren sind deaktiviert. Die Kommunikation zwischen den verschiedenen Modulen und dem Roboter wird durch drei Funkübertragungsstrecken gewährleistet, die das analoge Videosignal der Kamera sowie die seriellen Datenströme zur Steuerung der Schwenk-Neige-Einheit und des Roboters übertragen (vgl. Abbildung 5.19). Das Videosignal wird mittels Frame-Grabber digitalisiert und zur weiteren Verarbeitung über eine Socket-Verbindung in das Bildverarbeitungs- und Software-Entwicklungswerkzeug KHOROS gegeben. Die zur Lösung der Domino-Aufgabe notwendige Bildauswertung wurde unter KHOROS realisiert, weil diese Software sowohl über eine umfangreiche Bibliothek von Bildverarbeitungsroutinen verfügt als auch die Werkzeuge zur Verfügung stellt, um eigene Routinen einfach in das System zu integrieren (weitere Informationen über KHOROS siehe http://www.khoral.com). Die Navigation des Roboters und Steuerung seines Greifers ist unter der zum Pioneer1-Roboter gehörenden Entwicklungssoftware SAPHIRA implementiert worden. Um einen bidirektionalen Datenaustausch zwischen KHOROS und SAPHIRA zu ermöglichen, wurden zwei unabhängige Client/Server-Verbindungen aufgebaut. Die Bildverarbeitung sendet die Identität, Position und Orientierung aller erkannten Dominosteine oder eines speziellen Dominosteins an die Roboternavigation, wogegen das Navigationsmodul eben diese Dienste in Abhängigkeit von der aktuellen Situation anfordert (eine exakte Beschreibung des Protokolls findet sich in [Bollmann 1999]).

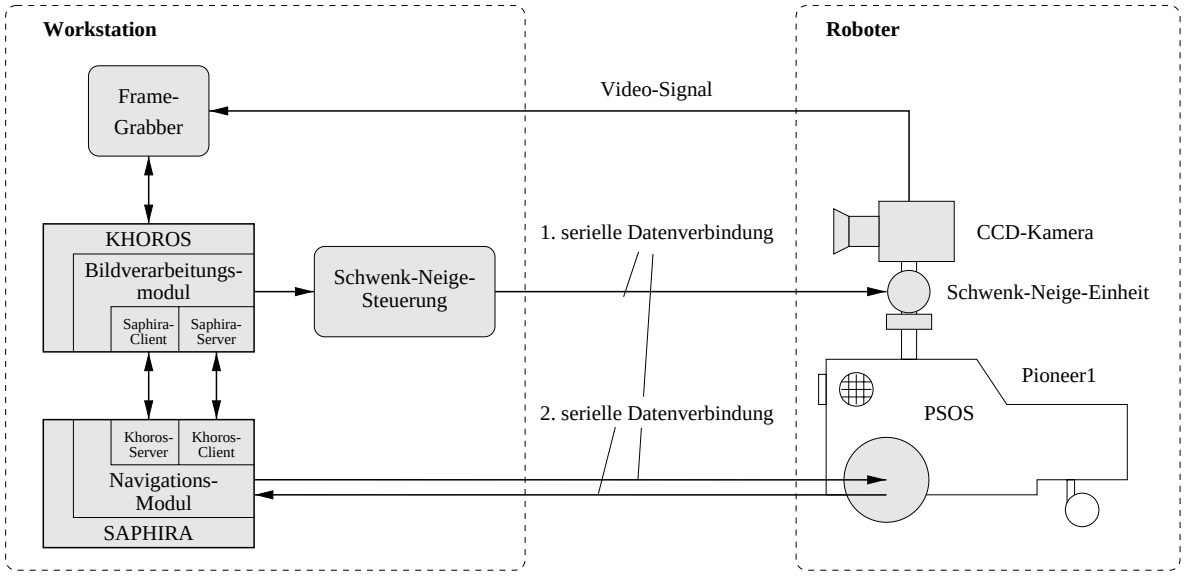


Abbildung 5.19: Systemarchitektur

Aufgrund des unterschiedlichen Task-Schedulings der beiden Software-Module unterscheiden sich die implementierten Client/Server-Verbindungen. Unter KHOROS kann die Synchronisation von Server- und Client-Prozeß dem Betriebssystem überlassen werden. Die SAPHIRA-Umgebung besitzt dagegen ihr eigenes Task-Scheduling, das speziell unter dem Aspekt der Echtzeit-Ausführung entwickelt wurde und bei der Implementation der Client/Server-Verbindung berücksichtigt werden muß.

SAPHIRA basiert auf einem synchronen interrupt-gesteuerten Betriebssystem. Jeder Micro-Task ist als *F*inite-*S*tate *M*aschine (FSM) implementiert und von jeder FSM wird während eines 100ms-Zyklus genau ein Zustand abgearbeitet. Dieser Mechanismus garantiert volle Synchronisation zwischen allen FSMs, weil der Gesamtzustand des Systems bereits aktualisiert ist, bevor eine bestimmte FSM, die von anderen FSMs abhängen kann, ausgeführt wird. Die Server- und Client-Micro-Tasks müssen nun ebenfalls auf FSMs abgebildet werden, damit SAPHIRA sie korrekt ausführen kann. Abbildung 5.20 zeigt die FSMs des Server- und des Client-Prozesses. Die Anzahl der Zustände innerhalb jeder FSM wurde heuristisch ermittelt (siehe auch [Rosenthal 1998]).

Insbesondere ermöglicht die Client/Server-Architektur natürlich auch die Verteilung des Bildverarbeitungs- und des Navigationsmoduls auf verschiedene Workstations. Eine weitere Client/Server-Verbindung besteht zwischen KHOROS und dem Host für den Frame-Grabber.

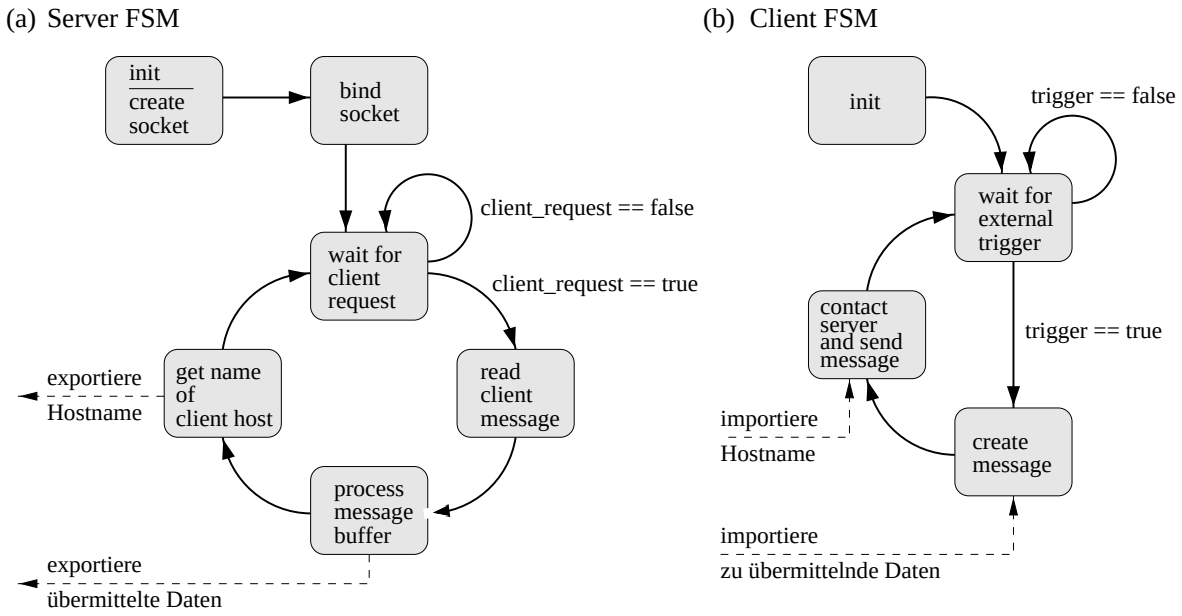


Abbildung 5.20: (a) SAPHIRA Server FSM, (b) SAPHIRA Client FSM

5.5.2 Perzeptions-Aktions-Zyklen

Ziel des Roboter-Dominospiels ist das Formen einer geraden Kette aus Dominosteinen, in der die benachbarten Hälften zweier Dominosteine jeweils die gleiche Punktzahl aufweisen. Das Anlegen in Form einer Kette wurde ausgewählt, um die Navigation zu vereinfachen. Im allgemeinen sind natürlich auch andere Konfigurationen denkbar. Mehrere *Perzeptions-Aktions-Zyklen* (engl. ‘Perception Action Cycle’) der in Abbildung 5.21 dargestellten Form sind notwendig, um diese Aufgabe zu lösen.

Das Spiel beginnt mit einer nahezu beliebigen Verteilung der Dominosteine im Raum und der Vorgabe der Identität des Dominosteines, der den Anfang der Dominokette bilden soll.

Der erste Perzeptions-Aktions-Zyklus startet mit der visuellen Exploration der Umgebung durch den Roboter. Der Roboter bestimmt die Position und Identität aller Dominosteine und wählt einen passenden Dominostein aus. Anschließend fährt der Roboter zu einem sogenannten *virtuellen Punkt* 50 cm vor dem betreffenden Stein. Dabei hängt die Position des virtuellen Punktes natürlich von der Richtung ab, aus der der Roboter den Dominostein greifen muß, um ihn richtig herum an die Dominokette anlegen zu können (vgl. Abbildung 5.22).

Wenn der Roboter den virtuellen Punkt erreicht hat, muß er seine tatsächliche Position relativ zu dem ausgewählten Dominostein bestimmen und seine internen Koordinaten bei einer festgestellten Abweichung angleichen. Seine tatsächliche Position bestimmt der Roboter über die Position des zu greifenden Dominos in einer aktuellen Aufnahme und der bekannten Abbil-

dungsgeometrie. Das Anfahren des Dominosteines über den virtuellen Punkt ist notwendig, weil eine Positionsabweichung (Differenz zwischen internen und externen Koordinaten), verursacht durch Lagerspiel, Unwucht der Räder oder Schlupf, u. U. zu nicht tolerierbaren Fehlern führen kann. Der Abstand des virtuellen Punktes vom endgültigen Ziel muß einerseits so groß gewählt werden, daß der begangene Positionsfehler auf dem Weg zum Ziel noch korrigiert werden kann, und andererseits so gering, daß ein neuerlicher Positionsfehler klein bleibt.

Nach der Anpassung der internen Koordinaten an die Realwelt greift der Roboter den Dominostein, fährt zu dem virtuellen Punkt vor der Dominokette, verifiziert erneut seine Position mit Hilfe des Kamerasystems, legt den Stein nieder und fährt zurück zu seiner initialen Position, die er ebenfalls verifiziert. Mit der Auswahl des nächsten Dominosteines beginnt anschließend ein neuer Perzeptions-Aktions-Zyklus. Anzumerken bleibt, daß der Roboter während der Bildanalyse steht, wogegen er während der Navigation blind ist. Ein anderes Ablaufschema ist aufgrund der Rechenzeiten derzeit nicht möglich. Das komplette Szenarium dauert etwa 15 Minuten, wobei die Bildverarbeitung den größeren Anteil ausmacht.

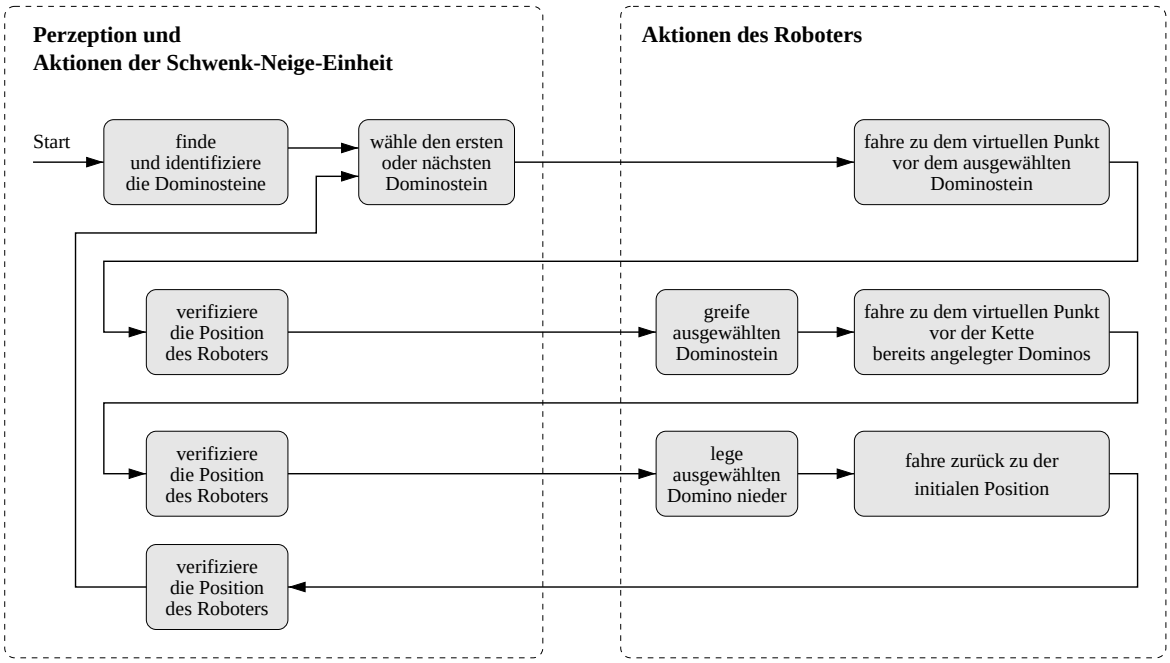


Abbildung 5.21: Perzeptions-Aktions-Zyklus

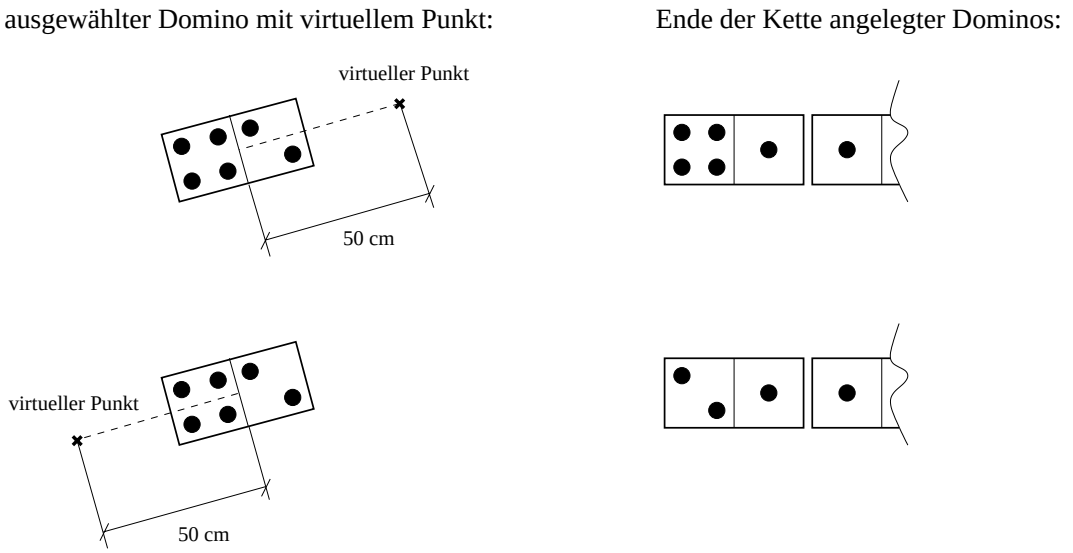


Abbildung 5.22: Position des virtuellen Punktes in Abhängigkeit von dem zuletzt angelegten Dominostein

Die nachfolgenden Abschnitte beschreiben die einzelnen Phasen während der Perzeptions-Aktions-Zyklen ausführlich.

5.5.3 Visuelle Suche und Identifizierung der Dominosteine

Die Suche nach den Dominosteinen ist eine systematische Suche, d. h. der Roboter sucht seine Umgebung mittels einer zu Beginn der Suche bereits festgelegten Reihenfolge von Fixationen vollständig ab. Der Roboter besitzt also a priori nahezu kein Wissen über die Positionen der Dominosteine. Wir setzen lediglich voraus, daß sich die Dominosteine bezüglich der initialen Position des Roboters im vorderen Halbraum befinden, d. h. in einem Schwenkbereich der Kamera von -90^o bis +90^o. Weiterhin erwartet der Roboter die Dominosteine in einer Entfernung von bis zu 2 m. Mit diesen Annahmen kann der Neigewinkel des Kamerakopfes zunächst fest auf -20^o eingestellt werden und es ergeben sich sieben Schwenkwinkel im Abstand von 30^o. Bei einem 48^o-weiten Sehfeld stellen sich somit Überlappungen zwischen aufeinanderfolgenden Bildern ein.

In der ersten Iterationsschleife werden mit dieser systematischen Suche Cluster von Dominopunkten detektiert (siehe Abbildung 5.23). Anschließend werden die Punkt-Cluster nacheinander fovealisiert (der Neigewinkel ist jetzt natürlich variierbar). Konnten alle Punkte in dem aktuell fovealisierten Cluster Dominozahlen zugeordnet werden, wird der nächste Punkt-Cluster fovealisiert, andernfalls soll der Roboter seine Distanz zu dem betreffenden Cluster sukzessive halbieren, bis die Dominopunkte zugeordnet werden konnten. Mit Hilfe der systematischen Suche müssen also lediglich Punkt-Cluster detektiert werden; das bedeutet, daß nicht unbedingt

alle Punkte der Dominosteine zu Beginn der Suche sichtbar sein müssen. Sind alle Cluster untersucht, startet das Einsammeln und Anlegen der Dominosteine.

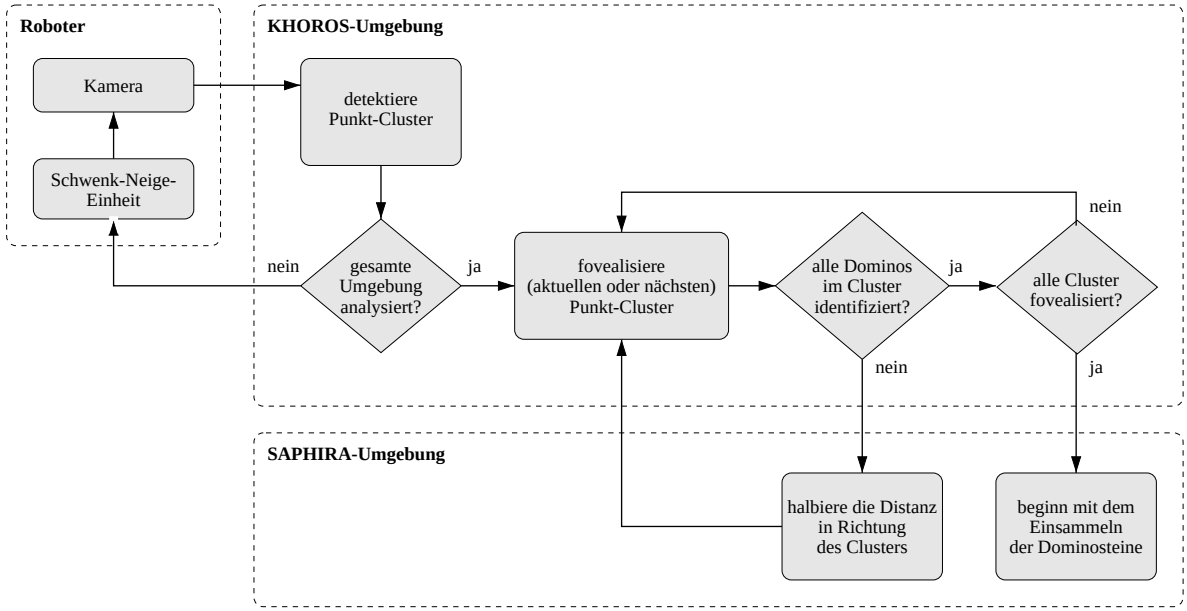


Abbildung 5.23: Suche und Identifizierung der Dominosteine

Bilder werden also insgesamt in drei verschiedenen Phasen analysiert: während der systematischen Suche zu Beginn des Dominospiels, während der anschließenden Fovealisierung der einzelnen Punkt-Cluster und während der Verifikation der aktuellen Roboterposition. Jede dieser Analysen startet mit einer Anfrage aus KHOROS an den Frame-Grabber. Das vom Grabber gelieferte Grauwertbild wird in KHOROS zunächst geglättet, um die durch Reflexionen verursachten Interferenzen zu reduzieren. Anschließend wird das Bild mit dem in Abschnitt 4.1.2 beschriebenen Bereichswachstums- und Verschmelzungsverfahren segmentiert und die geometrischen Eigenschaften der Segmente wie Größe, Schwerpunkt, Orientierung und Exzentrizität berechnet. Ziel dieser Prozedur ist die Detektion der Dominopunkte nicht der Steine selbst. Der Grund hierfür ist der nahezu konstante Kontrast zwischen den Punkten und den Steinen, der die Festlegung einer Segmentierungsschwelle erleichtert. Der Kontrast zwischen den Dominosteinen und ihrer Umgebung hängt dagegen stark vom Untergrund und, wegen der 3D-Charakteristik der Steine, von der Beleuchtungsrichtung ab.

Der Segmentierungsprozeß ist nicht fehlerfrei, so daß einige Dominopunkte nicht detektiert werden, während andere störende Segmente entstehen. Daher ist ein Test notwendig, der die gewünschten Dominopunkte von anderen Segmenten trennt. Wir benutzen hier die Exzentrizität und Orientierung der Segmente als relevante Prüfkriterien. Die ellipsenförmigen Dominopunkte besitzen eine Exzentrizität um 0,5 und eine horizontale Ausrichtung. Beide Größen werden von der Exzentrizitäten bewertenden Komponente der Blicksteuerung für alle Segmente geliefert. Die Segmente, die von den Vorgaben wesentlich abweichen, werden entfernt.

Nach der Segmentierung sind die Schwerpunkte der Dominopunkte in Bildkoordinaten gegeben. Eine solch variable Darstellung der Dominopunkte ist natürlich für ein Template-Matching, das die einfachste Lösung zur Erkennung der Dominozahlen, einer begrenzten Anzahl bekannter Punktkonfigurationen, darstellt, ungeeignet. Zusätzlich benötigen wir die Positionen der Dominosteine in dem egozentrischen Koordinatensystem des Roboters, um die gewünschten Manipulationen durchführen zu können. Der erste Schritt zur Überführung der Schwerpunkte der Dominopunkte in das roboterzentrierte Koordinatensystem ist die Vermessung der Geometrie des Sehsystems. Die grundlegenden Parameter der Geometrie des Sehsystems und der Projektionseigenschaften der Kamera sind in Abbildung 5.24 dargestellt. *OC* bezeichnet den Ursprung des kamerazentrierten Koordinatensystems, *O* den Ursprung des roboterzentrierten Koordinatensystems. Die Transformation zwischen den beiden Koordinatensystemen hängt ab von den Strecken *a*, *b*, *c* und *s* sowie den Schwenk- und Neigewinkeln *h* beziehungsweise *v*. Die Strecken *a*, *b* und *c* sind konstant und müssen lediglich einmal gemessen werden. Die aktuellen Schwenk- und Neigewinkel werden von der Blicksteuerung bestimmt und sind in dem Bildverarbeitungsmodul direkt zugreifbar. Die Projektionseigenschaften der Kamera wurden dagegen mit verschiedenen Testbildern experimentell ermittelt. Die aus diesen Untersuchungen berechneten Kameraparameter *s* und *f* sowie die radialsymmetrische Entzerrung und die Korrektur der Zyklorotation wurden anschließend experimentell validiert. Ein Vergleich zwischen den bekannten Koordinaten in einem Punktraster und den Koordinaten, die mit den ermittelten Parametern berechnet wurden, ergab einen relativen Fehler von weniger als 1,5%.

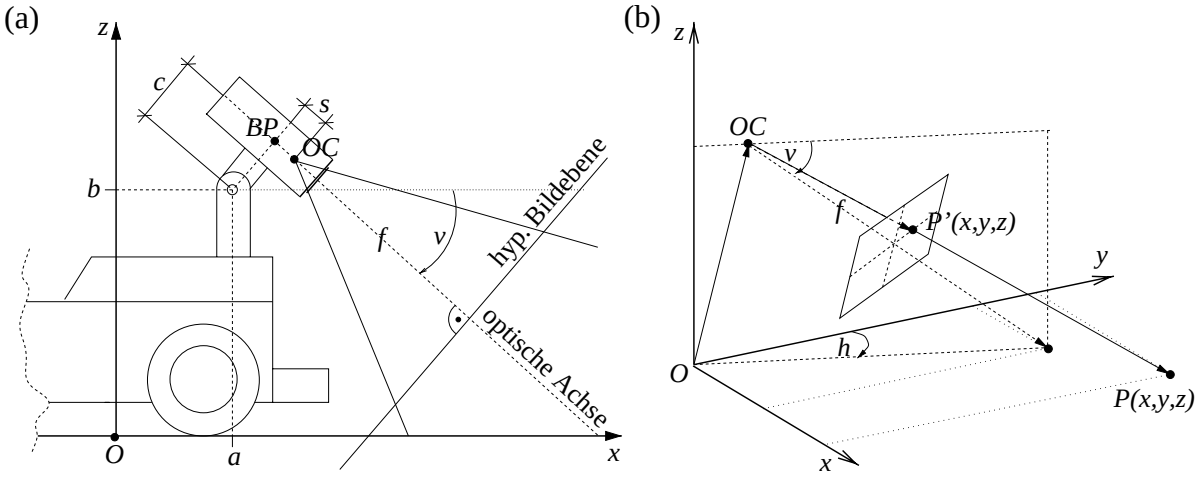


Abbildung 5.24: (a) Geometrie des Sehsystems, (b) Abbildungseigenschaften der Kamera (Erläuterungen siehe Haupttext)

Nach der Transformation der Schwerpunkte der Dominopunkte in das roboterzentrierte Koordinatensystem werden die Dominozahlen aus den Punkt-Clustern extrahiert. Jede Dominozahl ist durch ein Referenzmodell mit idealen Punktabständen repräsentiert. Hierbei sind allerdings nur die Zahlen Eins bis Sechs zugelassen. Im Gegensatz zu den tatsächlichen Dominospielre-

geln muß die Zahl Null entfallen, da sie nicht mit den derzeitig verwendeten Algorithmen detektiert werden kann. Jedes Punkt-Cluster wird mit allen zugelassenen Schablonen sukzessive durchsucht. Da die Koordinatentransformation zwar zu einer größeninvarianten aber orientierungsvarianten Abbildung führt, muß der Vergleich für verschiedene diskrete Orientierungen der Schablonen durchgeführt werden. Das Ergebnis ist eine Rotation der Schablonen um die Randpunkte der Punkt-Cluster (vgl. Abbildung 5.25). Ein Toleranzradius *r* erlaubt geringe Positionsabweichungen zwischen Template und tatsächlicher Punktkonfiguration, die aus der Koordinatentransformation oder den diskreten Rotationswinkeln entstehen können.

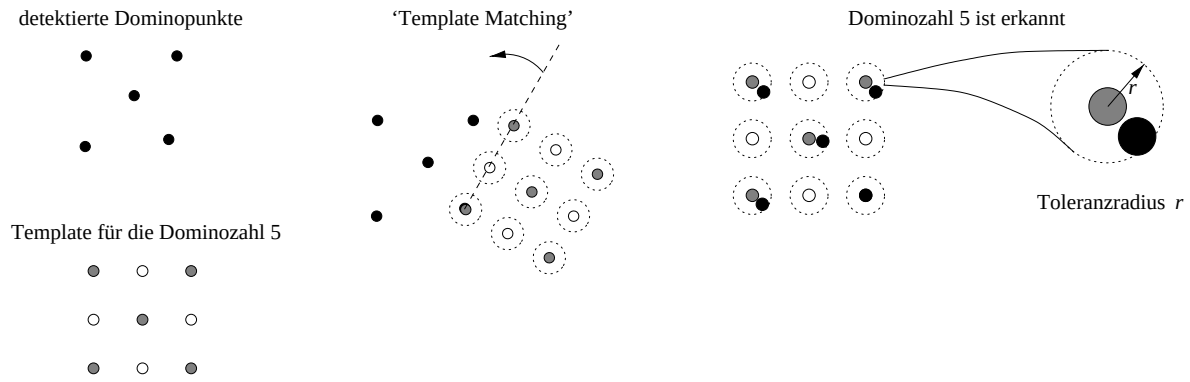


Abbildung 5.25: 'Template Matching'

Jede erkannte Zahl auf einer Dominosteinhälfte ist repräsentiert durch ihren Schwerpunkt. Diese Schwerpunktsberechnung ist notwendig, um die zwei Zahlen, die zu einem Dominostein gehören, zuordnen zu können. Das Zuordnungskriterium ist einfach der bekannte Abstand zwischen zwei Dominohälften (genauer zwischen ihren Schwerpunkten). Natürlich können auch hierbei Fehler auftreten. Zum einen können Dominopunkte durch eine fehlerbehaftete Segmentation verloren gehen, zum anderen können Mißklassifikationen durch ungünstige räumliche Konstellationen der Dominosteine entstehen. Insbesondere kann eine Dominohälfte mehr als einen Nachbarn in einem Abstand besitzen, der auf eine Zusammengehörigkeit der Hälften hinweist. Die meisten dieser Probleme können durch den Abgleich mit einer Wissensbasis beseitigt werden, in der die tatsächlich vorhandenen Dominosteine abgelegt sind (siehe Abbildung 5.26a). Zur Zeit ist die absolute Zahl der Dominosteine mit nur fünf Steinen sehr gering und insbesondere gibt es keine doppelten Steine. Allerdings sollte das Bildverarbeitungssystem auch sehr viel größere Mengen an Dominosteinen handhaben können, da sich die Erkennung auch ohne Abgleich mit der Wissensbasis als sehr robust erwiesen hat.

Nach dem Abgleich der detektierten Dominosteine mit der Wissensbasis werden die Schwerpunkte und Orientierungen der Steine aus den Schwerpunkten der zusammengehörigen Dominohälften berechnet. Die Orientierung der Dominosteine ist hierbei in einem Koordinatensystem definiert, welches übereinstimmt mit dem internen Koordinatensystem des Roboters (siehe Abbildung 5.26b). Hierdurch ist keine weitere Koordinatentransformation auf Seiten des Navigationsmoduls notwendig.

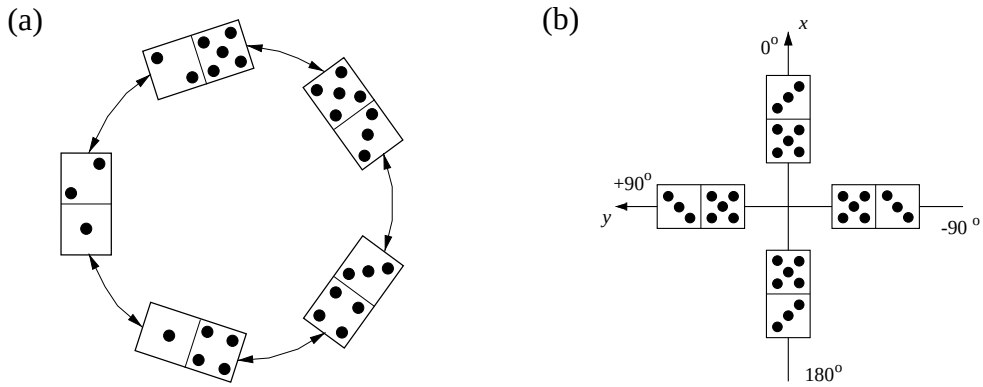


Abbildung 5.26: (a) Wissensbasis, (b) Koordinatensystem bezüglich der Orientierung der Dominos

5.5.4 Navigation

Der aktive Teil des Dominospiels wird von dem Pioneer1-Roboter durchgeführt. Da die Exploration der Welt durch den Roboter nur von der Information abhängt, die die Bildverarbeitung liefert, muß das Interface zwischen dem Bildverarbeitungsmodul und dem Robotersteuerungs- und -navigationsmodul in der Lage sein, die gesamte benötigte Information auf Anfrage und damit situationsabhängig zu liefern.

SAPHIRA stellt eine abstrakte Beschreibung des aktuellen Status des Roboters in einem sogenannten ‘State Reflector’ bereit. Dieser enthält Informationen über verschiedene Statusvariablen wie translatorische und rotatorische Geschwindigkeit, Position und Orientierung, etwaiges Blockieren der Motoren, Meßergebnisse der Sonarsensoren und Kontrollstrukturen, die etwas über die Art aussagen, durch die die Bewegungen des Roboters gesteuert werden. Eine selbstgeschriebene Routine kann nun die gewünschten Statusvariablen des ‘State Reflectors’ abfragen oder setzen. Im letzteren Fall sendet SAPHIRA die geeigneten Kommandos an den Roboter, um den erzwungenen Status zu erzielen.

Es existieren innerhalb von SAPHIRA verschiedene Möglichkeiten, um die Bewegungen des Roboters zu steuern. Der einfachste Weg ist die Verwendung von sogenannten ‘*Direct Motion*’-Befehlen, die direkt die Variablen im ‘State Reflector’ auf die gewünschten Werte setzen, was zu recht abrupten Bewegungen führen kann. Eine zweite Methode ist die Implementation von sogenannten *Behaviors*, die sehr viel komplexere Bewegungsarten dadurch zulassen, daß sie auf Fuzzy-Regelungen basieren. Behaviors besitzen zusätzliche Statusvariablen wie z. B. ein Ziel und eine Priorität, die die Interaktion mit anderen Behaviors ermöglicht. Eine dritte Methode, um den Roboter zu navigieren, besteht in der Erstellung sogenannter *Activities* in der Interpreter-Sprache Colbert, die das Planen und Ausführen von Abläufen zur Laufzeit ermög-

licht. Activities können ‘Child Activities’ und Behaviors starten und Aktionen mit anderen, gleichzeitig laufenden Activities koordinieren.

SAPHIRA besitzt im wesentlichen zwei verschiedene Repräsentationen des Raumes: einen sogenannten ‘*Global Map Space*’ (GMS) mit ortsfestem Bezugspunkt für größere Perspektiven und einen ‘*Local Perceptual Space*’ (LPS), ein egozentrisches Koordinatensystem mit Ursprung auf dem Roboter. In den GMS können nun Strukturen eingetragen werden, die bestimmte Objekte in der Umgebung des Roboters repräsentieren und *Artefakte* genannt werden. SAPHIRA kennt verschiedene Typen von Artefakten wie z. B. Punkte, Linien und Wände. Für die Repräsentation der Dominosteine ist das Punkt-Artefakt ausreichend, da es Variablen enthält für die räumlichen Koordinaten, die Orientierung und einen Bezeichner. Demnach kann die vom Bildverarbeitungsmodul gelieferte Information über jeden Dominostein direkt zur Erzeugung oder Veränderung eines Punktartefakts in dem GMS verwendet werden. Jedes Artefakt in dem GMS wird automatisch auch in den LPS übernommen und während der Bewegungen des Roboters aktualisiert.

Wie Abbildung 5.21 zeigt, beginnt das Domino-Szenarium mit der Aktivierung des Bildverarbeitungsmoduls, dessen erste Aufgabe in der Suche und Identifikation aller Dominosteine in der Umgebung des Roboters besteht. Nach der Übertragung der Ergebnisse werden die Dominosteine als Punktartefakte in den GMS (und damit auch in den LPS) eingetragen. Dieser erste Kontakt des KHOROS-Servers durch den SAPHIRA-Client instanziiert eine übergeordnete Activity, die alle weiteren Aktionen des Roboters durch die Ausführung untergeordneter Activities, Behaviors und ‘Direct Motion’-Befehle steuert. Um den von der Bildverarbeitung bestimmten Dominostein zu greifen, wird der zunächst anzufahrende virtuelle Punkt vor dem Stein als zusätzliches Punkt-Artefakt in die Map eingetragen. Dieses Artefakt gibt das Ziel für das GOTO-Behavior vor. Wenn das Behavior sein Ziel erreicht hat, wird das Punkt-Artefakt wieder aus der Map entfernt und der Roboter fordert den Kontrollblick an. Hierbei wird die vermutete Position und Orientierung des ausgewählten Dominosteines an die Bildverarbeitung übertragen. Die Bildverarbeitung wiederum kann diese Position natürlich nur validieren, wenn die Positionsabweichung des Roboters nicht zu groß ist, wenn also die Kamera beim Kontrollblick den ausgewählten Dominostein überhaupt noch im Bild hat. Allerdings war dies in den von uns durchgeführten Versuchen, in denen der Roboter nie mehr als ca. 3 m Wegstrecke bis zur nächsten Position zurücklegen mußte, an der ein Kontrollblick durchgeführt wurde, immer der Fall. Nach der Aktualisierung der Koordinaten der Artefakte in der LPS durch eine untergeordnete Activity öffnet der Roboter seinen Gripper durch eine weitere untergeordneten Activity, fährt gesteuert durch ‘Direct Motion’-Kontrolle bis zu dem ausgewählten Dominostein und schließt den Greifer (untergeordnete Activity). Anschließend fährt der Roboter gesteuert durch das GOTO-Behavior zu dem virtuellen Punkt vor der bereits angelegten Dominokette, korrigiert erneut seine Koordinaten durch die visuelle Kontrolle, fährt auf die Dominokette zu, legt den Stein durch Öffnen des Grippers ab, fährt eine bestimmte Strecke geradeaus zurück, um einen gewissen Rangierabstand von der Dominokette zu bekommen, und fährt dann zu seiner Ausgangsposition zurück, wo der nächste Zyklus startet.

5.5.5 Experimentelle Ergebnisse des Roboter-Szenariums

Abbildung 5.27 zeigt die Ergebnisse der wesentlichen Schritte bis zur Bestimmung der Dominosteine. Zunächst untersucht der Roboter den kompletten Halbraum vor seiner initialen Position nach Punkt-Clustern. Diese Phase bezeichnen wir als systematische Suche (Abbildung 5.27a). Die Punkt-Cluster ergeben sich durch die Segmentierung und anschließende Überprüfung der Segmente hinsichtlich ihrer Orientierung und Exzentrizität. Sind die Punkt-Cluster detektiert und ihre Schwerpunkte berechnet, so werden sie jeweils einzeln fovealisiert. In dem dargestellten Beispiel ergeben sich drei Punkt-Cluster, von denen zwei aus jeweils zwei eng zusammenliegenden Dominosteinen bestehen (Abbildung 5.27b). Die fovealisierten Punkt-Cluster werden wiederum segmentiert und hinsichtlich ihrer Ähnlichkeit zu horizontal ausgerichteten Ellipsen untersucht (Abbildung 5.27c). Anschließend werden die Schwerpunkte der Segmente unter Ausnutzung der Geometrie und der Abbildungseigenschaften des Kamerasystems in das egozentrische Koordinatensystem des Roboters transformiert. Das Ergebnis zeigt Abbildung 5.27d. In dieser normierten Darstellung werden dann mittels Schablonenvergleich die Dominozahlen einzelner Steinhälften bestimmt. Mehrdeutigkeiten bei der Zuordnung der Dominohälften zu kompletten Steinen aufgrund der räumlichen Nähe benachbarter Dominosteine werden durch den Abgleich mit einer Wissensbasis beseitigt. Abbildung 5.27e zeigt die Liste der identifizierten Dominosteine, die an SAPHIRA übermittelt wird, in einem KHOROS-Fenster.

Das komplette Domino-Szenarium wurde mittels Videokamera aufgenommen. Einige Ausschnitte sind in Abbildung 5.28 dargestellt. Die Kette aus Dominosteinen, an die angelegt werden soll, ist durch zwei hintereinanderliegende Steine vorgegeben (siehe obere rechte Ecke in Abbildung 5.28a und 5.28b). Ziel des Dominospiels ist es, die verbleibenden drei Dominosteine an diese Kette in der richtigen Reihenfolge anzulegen. Die beiden Steine, die den Anfang der Dominokette bilden, dürfen also nicht mehr gegriffen werden.

Die Abbildung 5.28a zeigt den Roboter während der Identifizierung der Dominosteine von seiner initialen Position. Ist ein Dominostein als Ziel der nächsten Manipulation ausgewählt, fährt der Roboter zu dem virtuellen Punkt vor dem Stein. Die Abbildung 5.28b zeigt den Roboter auf dem virtuellen Punkt nach der Verifizierung seiner Position kurz vor dem Anfahren des ausgewählten Dominosteines (Greifer bereits geöffnet). Nachdem der Zielstein gegriffen wurde, fährt der Roboter zu dem virtuellen Punkt vor der bereits angelegten Kette aus Dominosteinen. Die Abbildungen 5.28c und 5.28d zeigen den Roboter auf dem virtuellen Punkt vor der Dominokette bzw. während des Anlegens des nächsten Dominosteines. Deutlich zu sehen ist, daß ein passender Stein ausgewählt wurde und in der korrekten Ausrichtung an die Kette angelegt wird.

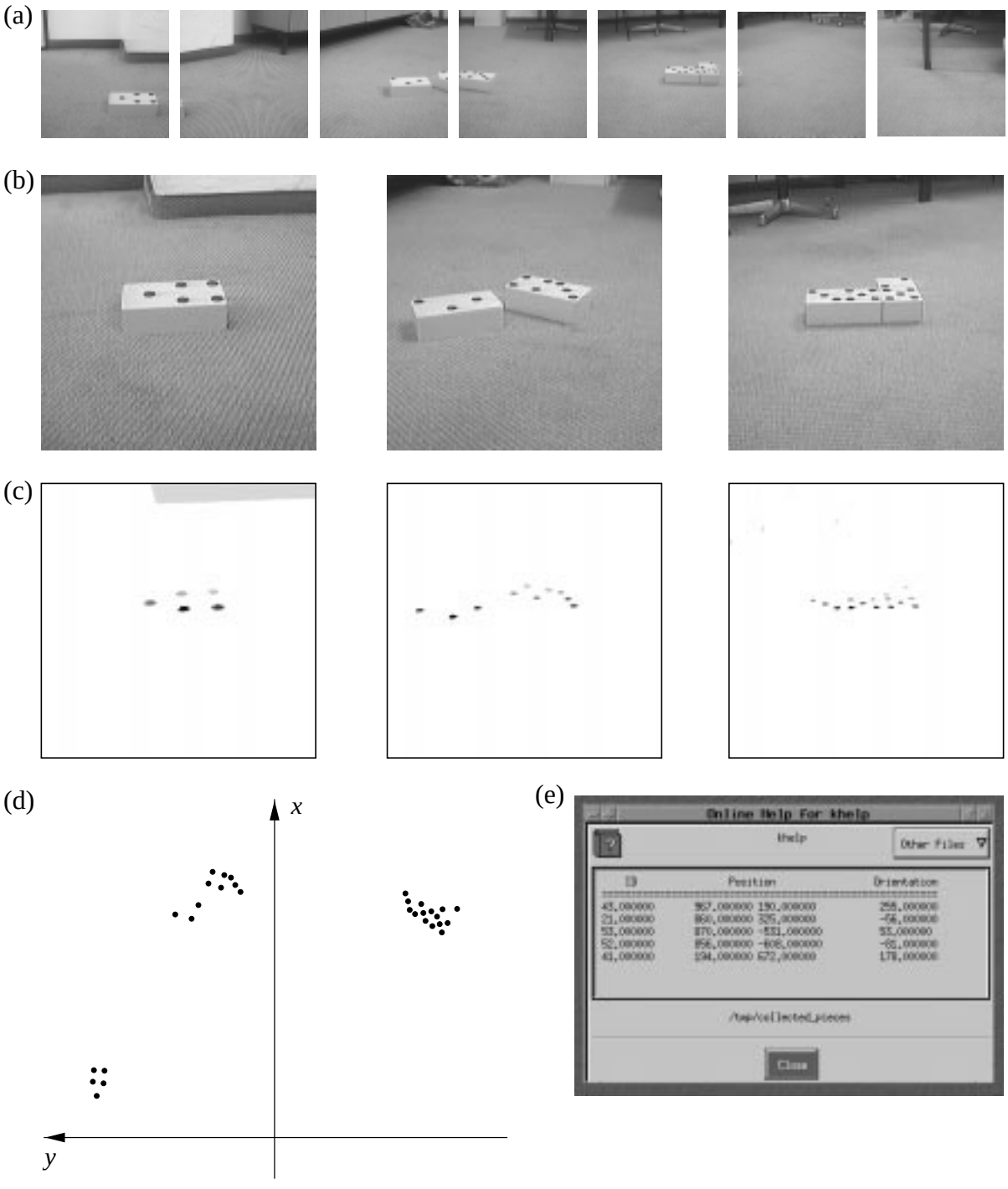


Abbildung 5.27: (a) Systematische Suche, (b) Fovealisierung der Dominopunkt-Cluster, (c) Segmentierungsergebnisse, (d) Rückprojektion der Dominopunkte in das Roboterkoordinatensystem und (e) Liste der erkannten Dominosteine

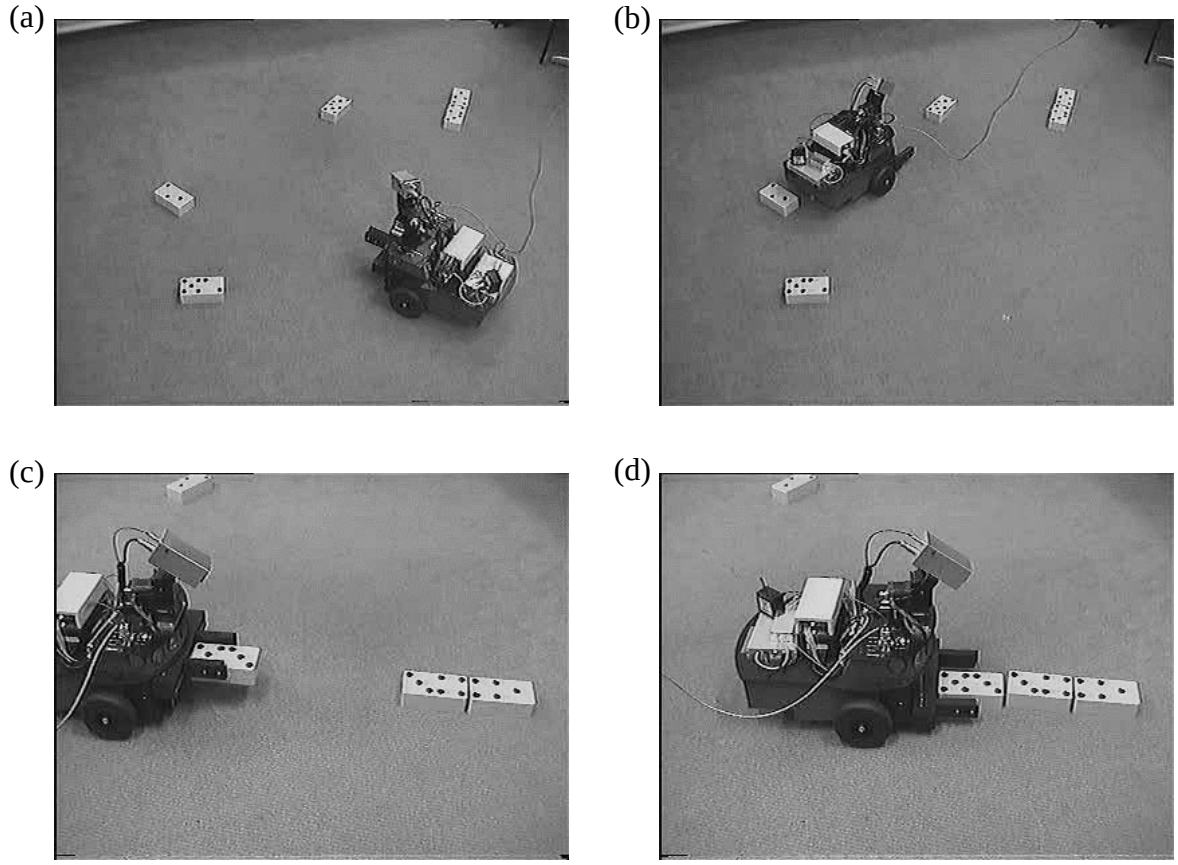


Abbildung 5.28: (a) Roboter in seiner initialen Position, (b) während der Verifizierung seiner Position vor dem zu greifenden Dominostein, (c) während der Verifizierung seiner Position vor der Dominokette und (d) beim Anlegen eines Dominosteins

Aktuell wird untersucht, wie Kollisionen mit nicht ausgewählten Dominosteinen oder der Dominokette selbst vermieden werden können. Der Roboter soll damit über eine Routenplanung verfügen, die die Ansteuerung der virtuellen Punkte auf dem kürzesten kollisionsfreien Weg ermöglicht.

Gleichwohl zeigt das Domino-Szenarium auch in seiner jetzigen Form die Leistungsfähigkeit der implementierten Blicksteuerung. Wobei für das Dominospiel allerdings nur einige wenige Komponenten der Blicksteuerung aus NAVIS zum Einsatz kamen, insbesondere die Exzentrizitäten bewertende Komponente der bilddatengetriebenen Aufmerksamkeitssteuerung. Dies zeigt deutlich, daß die in dieser Arbeit entworfene Blicksteuerung an die jeweilige Anwendung adaptiert werden muß und aufgrund ihrer Modularität auch adaptiert werden kann.

Kapitel 6

Diskussion

6.1 Psychologische und biologische Plausibilität

In dieser Arbeit wurde der Versuch unternommen, aus der Psychophysik und der Neurobiologie gewonnene Erkenntnisse zur visuellen Aufmerksamkeit in die Modellierung der maschinellen Blicksteuerung des aktiven Sehsystems NAVIS einzubeziehen. An dieser Stelle soll deshalb diskutiert werden, wo Analogien zwischen dem maschinellen Aufmerksamkeitsmodell und den psychologischen bzw. biologischen Modellen hergestellt werden konnten.

Eines der wesentlichsten Ergebnisse der Aufmerksamkeitsforschung ist die Feststellung, daß die verdeckte Aufmerksamkeit potentielle Kandidaten für einen Blickwechsel untersucht und Augenbewegungen nur zu den Regionen durchgeführt werden, für die sich eine genauere Untersuchung lohnt. Ein ähnlicher Mechanismus ist auch in unserem Modell realisiert worden. Auffällige Regionen werden durch sogenannte Attraktivitätspunkte gekennzeichnet, wobei eine Kamerabewegung jeweils nur zu dem Attraktivitätspunkt mit dem höchsten Gewicht, dem Aufmerksamkeitspunkt, erfolgt. Der aktuelle Aufmerksamkeitspunkt wird gehemmt, so daß in den darauffolgenden Iterationen andere Attraktivitätspunkte, die zunächst ein niedrigeres Gewicht als der Aufmerksamkeitspunkt aufwiesen, ebenfalls sukzessive angeschaut werden können (vgl. Abschnitt 4.3). Viele Attraktivitätspunkte gewinnen allerdings im Verlauf solcher Blickwechsel nie den WTA-Prozeß und werden somit auch niemals fixiert.

Die Hemmung bereits untersuchter Regionen ist nicht nur in unserem Modell realisiert, sondern ist auch in psychologischen und biologischen Untersuchungen beschrieben worden. Analog zu den psychologischen und biologischen Untersuchungen nimmt auch in unserem System die Hemmung sukzessive wieder ab. Ein Problem bei der konkreten Umsetzung bleibt allerdings die geeignete zeitliche und räumliche Parametrisierung des Hemmungsmechanismus. Diese konnte in unserem Modell lediglich heuristisch ermittelt werden. Der Mensch kann vermutlich die Hemmung einer bereits betrachteten Region durch willentliche Steuerung der Aufmerksamkeit überwinden und so diese Region sofort erneut untersuchen. Auch in unserem Mo-

dell ist neben der automatischen, bilddatengetriebenen Aufmerksamkeitskomponente eine modellgetriebene Aufmerksamkeitskomponente enthalten, die man im weitesten Sinne als "willentlich gesteuert" bezeichnen kann. Liegt eines der beiden Modelle, Validierung einer Objekthypothese oder Verfolgung eines Objekts, vor, so dominiert dieses die Auswahl des Aufmerksamkeitspunktes. Nur wenn das dritte Modell, die Generierung einer Objekthypothese, gegeben ist, haben ‘Top down’-Informationen keinen Einfluß auf die Auswahl des nächsten Aufmerksamkeitspunktes.

Die Auswahl eines Fixationspunktes hängt beim Menschen noch von weiteren nicht visuellen Faktoren, wie Kategorisierungseffekten und der Vertrautheit mit dem Target, ab. Kategorisierungseffekte entstehen in unserem technischen System durch die Aufteilung der Merkmale auf eine begrenzte Anzahl Merkmals- und Auffälligkeitskarten. So werden zum Beispiel die Orientierungen der Kanten und Segmente in 15^o große Intervalle unterteilt und die Farben auf einem Farbkreis in sechs 60^o große Intervalle (vgl. Abschnitt 4.1). Die Vertrautheit mit einem Target spielt in unserer maschinellen Blicksteuerung dagegen keine Rolle. Durch die strikte Trennung des Lern- und des Erkennungsvorgangs in NAVIS ist die Erkennungsleistung des Systems nicht abhängig von der Häufigkeit, mit der ein Objekt bereits in früheren Versuchen erkannt worden ist.

Zu den wesentlichen Ergebnissen der psychologischen Aufmerksamkeitsforschung gehört die Spotlight-Metapher, die ja besagt, daß sich visuelle Aufmerksamkeit in Form eines Scheinwerfers immer nur auf eine Region oder ein Objekt im visuellen Feld konzentriert, also insbesondere nicht geteilt werden kann. Ein ähnlicher Mechanismus ist auch in unserem System berücksichtigt. Allerdings umfaßt der ‘Focus of Attention’ hier immer einen kompletten Bildausschnitt, dessen Größe vor dem Start des Systems eingestellt wird. Dies ist sicherlich eine Einschränkung gegenüber der Spotlight-Metapher, die von einer variablen Größe des FOA mit eher unscharfen Rändern ausgeht.

Während der Kamerabewegungen ist das technische System blind. Es kennt lediglich die Zielkoordinaten, die durch eine Schwenk- und eine Neigebewegung angefahren werden müssen. Somit verhalten sich die Kamerabewegungen ähnlich den ballistischen Sakkaden des Menschen. Eine Sakkade trifft jedoch häufig nicht ihr Ziel, so daß eine nachfolgende kleinere Sakkade zur Korrektur notwendig ist. Auch in unserem System kann ein möglicher Positionsfehler durch die implementierte Korrelation zwischen Soll- und Istbild behoben werden. Yarbus stellte bei der Aufzeichnung von Augenbewegungen fest, daß Menschen zyklisch immer wieder dieselben Fixationspunkte beim Betrachten einfacher Strichzeichnungen aufsuchen und nicht etwa mit der Zeit neue Fixationspunkte hinzunehmen. Lediglich die Verweildauer an den einzelnen Positionen nimmt mit der Zeit zu. Aus eben diesen Beobachtungen entstand Notons Scanpath-Theorie (siehe Abschnitt 3.1.2), auf der auch die Erkennungsstrategie der 2D-Objekterkennung in NAVIS basiert (Abschnitt 5.3.1). Neben den Sakkaden sind auch langsame Folgebewegungen in NAVIS modelliert. Sie werden zur Verfolgung eines Objekts eingesetzt (siehe Abschnitt 5.4). Allerdings wird die Objektverfolgung lediglich von der Blicksteuerung initialisiert, die die Position liefert, an der erstmals Bewegung detektiert werden konnte. Auf

den Verlauf der Objektverfolgung und damit auch auf die langsame Folgebewegung nimmt die Blicksteuerung dagegen keinen Einfluß.

Von den aus der Psychologie bekannten Aufmerksamkeitsmodellen sind die ‘Early Selection’-Theorien, wie z. B. die ‘Feature Integration Theory’ und ‘Guided Search’, unserem Modell am ähnlichsten: Merkmale werden separat, automatisch und früh detektiert und auf intradimensionale Merkmalskarten verteilt. Das heißt, die Merkmale werden insbesondere auch kategorisiert. In unserem Modell fehlt aber bisher die Umsetzung von Gruppierungseffekten, die eine wesentliche Rolle in der ‘Feature Integration Theory’ spielen. Allerdings gestaltet sich die Implementation solcher Effekte für reale Szenen auch sehr viel schwieriger, da sie dort nicht so unmittelbar wiederzufinden sind wie in Displays, die in psychologischen Experimenten zur visuellen Suche eingesetzt werden. In Abschnitt 4.5 konnte gezeigt werden, daß mit der realisierten Blicksteuerung auch typische Experimente zur visuellen Suche nachvollzogen werden können. Eine Integration der Merkmale aus verschiedenen Karten soll in einem Nachfolgeprojekt durch die Indizierung der Merkmale auch in der Tiefe erreicht werden. Die in dieser Arbeit vorgestellte Blicksteuerung ist aber nicht nur bilddatengetrieben, sondern eben auch modellgetrieben. Die Hypothesenvalidierung als Teil der modellgetriebenen Aufmerksamkeitssteuerung weist damit auch Ähnlichkeiten zu Duncans und Humphreys ‘Stimulus Similarity Theory’ auf, die ja unter anderem besagt, daß die strukturellen Einheiten der Szenenrepräsentation mit einem Target-Template gematcht werden müssen. Genau dies geschieht auch während der Hypothesenvalidierung.

Vorschläge zur Architektur eines Aufmerksamkeitssystems (Merkmalskarten, Hemmungskarte, etc.), wie ich sie in dieser Arbeit aus dem Bereich der Experimentalpsychologie vorgestellt habe, fehlen in der Neurobiologie, sofern sie sich mit der visuellen Aufmerksamkeit beschäftigt. Interessante Eigenschaften des menschlichen Gehirns sind generell aber dessen Modularität, Redundanz sowie die parallele und gleichzeitig hierarchische Organisation. Die hier vorgestellte maschinelle Blicksteuerung ist ebenfalls modular aufgebaut. So entsprechen z. B. die verschiedenen Karten der Attraktivitätsrepräsentation einzelnen Software-Modulen, die eine eingegrenzte Aufgabe erfüllen, definierte Schnittstellen besitzen und im Prinzip beliebig kombiniert werden können. Ebenfalls findet sich in dem technischen System innerhalb der Attraktivitätsrepräsentation eine bestimmte Redundanz, die die Anfälligkeit gegenüber einzelnen Fehlern reduziert. So liefern die verschiedenen Auffälligkeitskarten, je nach der Struktur der im Bildausschnitt vorliegenden Gebilde, durchaus gleiche Attraktivitätspunkte. Ein aus dieser Sicht idealer Stimulus wäre z. B. eine homogene Form mit Symmetrieeigenschaften sowie ausgeprägter Exzentrizität und starkem Farbkontrast zur Umgebung. Fehler in einzelnen Karten können also trotzdem zur Analyse der relevanten Regionen führen. Das Modell weist weiterhin sowohl parallele als auch hierarchische Strukturen auf. Die Parallelität beschränkt sich hier auf die Attraktivitätsrepräsentation, in der verschiedene Merkmale gleichzeitig analysiert werden. Hierarchien bestehen zwischen den drei Komponenten der Blicksteuerung, der bilddatengetriebenen Komponente, der modellgetriebenen Komponente und der Verhaltenssteuerung. In der untersten Hierarchie, der bilddatengetriebenen Komponente, reagiert das System automatisch. Dieses Verhalten ist im weitesten Sinne analog zu den reflektorischen Antworten solcher bio-

logischer Zellen, die sich früh in der menschlichen Sehbahn befinden. In der modellgetriebenen Komponente ist die Reaktion des Systems, analog zu den höheren Ebenen im menschlichen Gehirn, abhängig von willentlich gesteuerten Einflüssen. Eine weitere Hierarchie ergibt sich durch die unterschiedliche Priorisierung der verschiedenen Modelle (Generierung einer Objekthypothese, Validierung einer Hypothese, Objektverfolgung) in der verhaltensgesteuerten Komponente.

6.2 Bestandsaufnahme und Abgrenzung

An dieser Stelle möchte ich noch einmal auf die in Abschnitt 2.2 geforderten Eigenschaften für eine maschinelle Blicksteuerung eingehen und diskutieren, inwieweit die Forderungen in dieser Arbeit erfüllt werden konnten. Desweiteren möchte ich die entstandene Blicksteuerung mit den im Abschnitt 3.4 beschriebenen Entwürfen anderer Autoren vergleichen und gegen diese abgrenzen.

Eine sehr wichtige Eigenschaft der Blickwechsel beim Menschen ist deren Abhängigkeit vom Kontext. Blickwechsel hängen ab von der aktuellen Aufgabe und den mit dieser Aufgabe verknüpften Erwartungen. Auch die in NAVIS integrierte Blicksteuerung kann verschiedene Aufgaben unterscheiden. Diese Aufgaben sind:

- das Finden eines Objekts (Generierung einer Objekthypothese)
- das Erkennen eines Objekts (Validierung einer Objekthypothese)
- die Verfolgung eines Objekts

In einige Teile dieser Aufgaben fließen ebenfalls Erwartungen in Form von getroffenen Annahmen ein. So trägt das System zum Beispiel alle zu einer Objekthypothese gelernten Formpunkte in die Verstärkungskarte ein, sobald die zu dem Objekt gespeicherte Konfiguration von Attraktivitätspunkten in der Szene gefunden wurde. Ein zweites Beispiel ist die für die Objektverfolgung getroffene Annahme, daß die aus einer histogrammbasierten Analyse hervorgehende größte Klasse der Verschiebungsvektoren die durch die Kamerabewegung verursachte Bewegung des Hintergrundes darstellt und die zweitgrößte Klasse die Bewegung des Objektes kennzeichnet. Werden die gemachten Annahmen verletzt, kommt es zu Fehlern. Diese sollten allerdings in einem aktiven Sehsystem, das u. a. durch die fortlaufende Auswertung von Bildern gekennzeichnet ist, zeitlich begrenzt bleiben.

Eine Forderung von Abbott [Abbott 1988] für den Entwurf eines maschinellen Aufmerksamkeitsmodells ist die Berücksichtigung der Distanz zwischen aktuellem Fixationspunkt und den Kandidaten für einen Blickwechsel sowie die Berücksichtigung der Blickrichtung. So haben Untersuchungen gezeigt, daß der Mensch Augenbewegungen nach oben solchen nach unten

vorzieht [Levy-Schoen 1981]. Diese Eigenschaft läßt sich auch in die Aufmerksamkeitssteuerung von NAVIS integrieren. Die Auffälligkeit der Attraktivitätspunkte muß hierzu lediglich in Abhängigkeit von der vertikalen Position bewertet werden (vgl. Fünf-Tupel aus Abschnitt 4.3). Auf eine solche Bewertung wurde in dieser Arbeit allerdings zunächst verzichtet, da der Nutzen für eine maschinelle Blicksteuerung unklar ist und der Vergleich zwischen Punkten, die oberhalb bzw. unterhalb des aktuellen Fixationspunktes liegen, erschwert wird. In der Blicksteuerung für NAVIS spielt die räumliche Nähe der Kandidaten für einen Blickwechsel besonders bei der Hemmung des aktuellen Aufmerksamkeitspunktes eine Rolle, denn die umliegenden Attraktivitätspunkte werden in Abhängigkeit von einer Distanzfunktion ebenfalls gehemmt. So wird vermieden, daß das System zu kleine Blickwechsel macht, d. h. Blickwechsel, die dem System kaum neue Information zuführen. Bei der Generierung einer Objekthypothese besitzt die räumliche Nähe in unserem System keinen Einfluß, da sie die Suche weder beschleunigen noch verlangsamen würde. Während der Validierung einer Objekthypothese werden die vermuteten Formen in der gelernten Reihenfolge angefahren. Auch hier hat die räumliche Nähe keinen Einfluß auf das Erkennungsergebnis. In dem dritten Modell, der Verfolgung eines sich bewegenden Objekts, ist die Berücksichtigung der räumlichen Nähe dann sinnvoll sein, wenn man mehrere sich bewegende Objekte in einer Szene zuläßt. Dann spielt neben der Bewegungsrichtung und der Geschwindigkeit der einzelnen Objekte auch deren Distanz zum Beobachter eine Rolle bei der Auswahl des zu verfolgenden Objektes. Ein solches Szenarium wird in dem DFG-Projekt ESAB-2 untersucht.

Bei vorhandener Symmetrie tendieren Menschen dazu, die Symmetrieachsen zu fixieren [Locher 1987]. Ein Aufmerksamkeitsmodell sollte daher nach Abbott auch die Berechnung von Bildcharakteristiken, wie z. B. Symmetrien, beinhalten. Diese wichtige Eigenschaft ist in unserem Modell enthalten. Ihr besonderer Vorteil liegt in der Detektion von Auffälligkeiten, deren Zentrum nicht etwa auf einem Objekt liegt, sondern in einem geometrischen Zentrum zwischen einer Gruppe von Objekten (vgl. Abschnitt 4.2.1). Eine weitere Eigenschaft aus der Klasse der Bildcharakteristiken, die in unserem Modell berücksichtigt ist, ist der Farbkontrast (Abschnitt 4.2.3). Hierbei handelt es sich nicht etwa um eine Objekteigenschaft, da der Farbkontrast maßgeblich vom Objekthintergrund beeinflußt wird. Gleichwohl sind die Objektgrenzen und -eigenschaften insbesondere für die Erkennung wesentlich. Daher ist die Fovealisierung auf Objektzentren nach Abbott ebenfalls bedeutsam und wird auch in unserem Modell berücksichtigt. Zusätzlich werden in unserem Ansatz Objekteigenschaften wie Exzentrizität und Farbe berechnet (vgl. Abschnitt 4.2.2 bzw. 4.1.3). Als ein Objekt wird hierbei allerdings jede homogene Fläche betrachtet; also jedes Segment, welches das Segmentierungsverfahren liefert. Eine verbesserte Abgrenzung der einzelnen Objekte würde sich durch die Ausnutzung von Tiefeninformationen aus der Stereoskopie ergeben. Eine derartige Erweiterung wird ebenfalls in ESAB-2 angestrebt.

Durch die Auswertung von Tiefenhinweisen lassen sich die Attraktivitätspunkte verschiedener Flächen, die zu einem Objekt gehören, aber zum Beispiel eine unterschiedliche Farbe aufweisen, zusammenfassen. Dadurch würde sich, wie von Abbott gefordert, die Anzahl der Kandidaten für einen Blickwechsel minimieren. Auf die Anzahl der Attraktivitätspunkte kann in der

Blicksteuerung für NAVIS aber auch bereits über die Parametrisierung der einzelnen Verfahren Einfluß genommen werden, die die Merkmals- und Auffälligkeitskarten generieren. Eine Minimierung der Attraktivitätspunkte ist somit möglich.

Wie von Abbott gefordert, berücksichtigt die Blicksteuerung auch zeitliche Änderungen. Allerdings wird in NAVIS vorausgesetzt, daß die Bewegung von lediglich einem Objekt stammt. Sobald ein sich bewegendes Objekt in die Szene eintritt, wird die Objektsuche oder Erkennung abgebrochen, um der Bewegung zu folgen. In ESAB-2 werden zusätzlich die sich aus der Bewegungsart ergebenden Merkmale, wie Bewegungsrichtung und Bewegungsgeschwindigkeit, in Form von Merkmals- und Auffälligkeitskarten in die Attraktivitätsrepräsentation integriert. Als besonders auffällig könnten z. B. solche Bewegungen gekennzeichnet werden, die am Bildrand detektiert wurden. Bei diesen Bewegungen handelt es sich entweder um Objekte, die die Szene verlassen und deshalb besonders interessant sind, weil die Gefahr besteht, diese Objekte zu "verlieren", oder es handelt sich um Objekte, die neu in die Szene gelangt sind und deshalb interessant sind, weil sie neue Information darstellen. Eine solche Bewertung stimmt mit neurobiologischen Untersuchungen überein, die gezeigt haben, daß bewegte Objekte bereits im extrafovealen Bereich wahrgenommen werden. Weiterhin sind Objekte mit einer hohen Bewegungsgeschwindigkeit besonders interessant, weil auf sie schneller reagiert werden muß als auf langsame Objekte. Die schnellen Objekte verlassen mit einer höheren Wahrscheinlichkeit das Sehfeld des Kamerasystems, so daß der Kamerakopf entsprechend nachgeführt werden muß. Auch die höhere Wahrscheinlichkeit, daß bei einer bevorstehenden Kollision einem sehr schnellen Objekt nicht mehr ausgewichen werden kann, würde eine Bewertung der Bewegungsgeschwindigkeit rechtfertigen.

Zusammenfassend läßt sich feststellen, daß die in Abschnitt 2.2 aufgestellten Anforderungen an eine maschinelle Blicksteuerung für die hier beschriebene eingeschränkte Welt weitestgehend erfüllt werden konnten und sich somit die von Abbott aufgestellten Forderungen auch in der konkreten Realisierung als sinnvoll erwiesen haben. Allerdings ist die Blicksteuerung für komplexere Szenarien noch auszubauen. Das Roboter-Szenarium aus Abschnitt 5.5 zeigt zudem deutlich, daß die für das universell konzipierte aktive Sehsystem NAVIS entwickelte Blicksteuerung als Bausatz dienen kann, aus dem einzelne Komponenten entnommen und auf eine konkrete Anwendung angepaßt werden können.

Die von Koch und Ullman geforderten Eigenschaften einer maschinellen Aufmerksamkeitssteuerung (siehe Abschnitt 3.4.1) stimmen weitgehend überein mit dem von Treisman aus der ‘Feature Integration Theory’ abgeleiteten psychologischen Aufmerksamkeitsmodell. Folglich sind in meiner Arbeit auch Ähnlichkeiten mit den theoretischen Überlegungen von Koch und Ullman zu finden. So werden hier ebenfalls elementare Merkmale parallel in verschiedenen Merkmalskarten repräsentiert. Im Gegensatz zu Koch und Ullman werden diese Merkmale allerdings nicht direkt in eine nicht-topographische zentrale Karte überführt, sondern zunächst getrennt auf ihre Auffälligkeit untersucht. So existiert für jeden Satz an Merkmalskarten genau eine Auffälligkeitskarte. Solche ‘Conspicuity Maps’ sind auch in dem Modell von Milanese realisiert (Abschnitt 3.4.3). Allerdings existiert dort nur eine Merkmalskarte pro Dimension, so

daß ein willentlich gesteuerter Zugriff auf einzelne (intradimensionale) Merkmale nicht möglich ist. Zudem haben in dem System von Milanese die Auffälligkeitskarten lediglich die Funktion einer Kontrasterhöhung.

Koch und Ullman schlagen weiterhin vor, Nachbarschafts- und Ähnlichkeitsbeziehungen in einem Aufmerksamkeitsmodell zu berücksichtigen. Auch in unser System gehen Nachbarschaftsbeziehungen ein. So werden, wie bereits erwähnt, neben dem aktuellen Aufmerksamkeitspunkt auch die Attraktivitätspunkte abhängig von ihrer räumlichen Distanz zum Aufmerksamkeitspunkt gehemmt. Auf diese Weise können bei geeigneter Parametrisierung zu kleine Blickwechsel, also Blickwechsel, die nur wenig neue Information bereitstellen, erschwert werden. Ähnlichkeitsbeziehungen bestehen durch die Kategorisierung der Merkmale bei der Aufteilung auf die diversen Karten. So kann ‘top down’ auf ähnliche Merkmale zugegriffen werden.

Die Überlegungen von Westelius (Abschnitt 3.4.1), eine effektive Blicksteuerung in eine bilddatengetriebene Komponente, eine modellgetriebene Komponente, eine Verhaltensebene und eine Zufallskomponente zu unterteilen, sind in unserem System weitgehend aufgegriffen worden. Allerdings kann in unserem System die modellgetriebene Komponente verschiedene Modelle enthalten. Entsprechend besteht die Verhaltensstufe nicht nur in der Reduzierung des Einflusses bereits untersuchter Regionen, sondern insbesondere in der Auswahl des jeweils geeignetsten Modells. Der Blicksteuerung des aktiven Sehsystems NAVIS fehlt bisher allerdings eine Zufallskomponente, die immer dann Einfluß auf die Steuerung nehmen müßte, wenn z. B. der Stereokamerakopf an seine mechanischen Grenzen stößt. Diese Grenzen sind durch Endschalter gesichert, die bei Kontakt zur Abschaltung des Kamerakopfes führen. Danach muß das System derzeit praktisch neu gestartet werden.

Die in dieser Arbeit entwickelte Blicksteuerung kann in die Klasse der merkmalsbasierten Aufmerksamkeitsmodelle eingeordnet werden, insofern sie Filteroperationen für reale Szenen enthält. Im Gegensatz zu den Modellen von Milanese und Itti et al. ist unser Modell allerdings auch mit einer ‘Top down’-Komponente versehen, die nicht von den im Bild vorhandenen Strukturen abhängt. Das Modell von Giefing et al. ist daher unter den in Abschnitt 3.4.3 beschriebenen merkmalsbasierten Modellen dasjenige, welches die größte Ähnlichkeit mit unserem System aufweist. Es besitzt sowohl eine ‘Bottom up’- als auch eine ‘Top down’-Komponente und ist ebenfalls mit einem Kamerasystem gekoppelt. Ein Vergleich beider Systeme fällt allerdings schwer, da das Modell von Giefing et al. nicht ausreichend in der Literatur beschrieben ist. Die Blicksteuerung scheint allerdings nur ein Modell zur Objekterkennung zu unterstützen. Die von Maki vorgestellte Blicksteuerung (Abschnitt 3.4.4) besitzt dagegen ähnlich unserem System zwei Betriebsarten: einen Verfolgungsmodus und einen Sakkadenmodus. Im Sakkadenmodus wechselt der FOA allerdings nur von einem bewegten Objekt zum nächsten. Eine Objekterkennung findet nicht statt.

Kapitel 7

Zusammenfassung

Die vorliegende Arbeit gibt einen Überblick über den Stand der Forschung auf dem Gebiet der visuellen Aufmerksamkeit und der Augenbewegungen in den Disziplinen Psychophysik und Neurobiologie. Weiterhin enthält sie eine Auswahl der in der Literatur beschriebenen maschinellen Aufmerksamkeitsmodelle sowie der Blicksteuerungen aktiver Sehsysteme. Die aufgeführten Arbeiten, Modelle und Ergebnisse mögen dem Leser oder der Leserin zu weiteren Studien dienen.

Das Ziel dieser Arbeit war der Entwurf und die Implementation von attentiv gesteuerten Blickwechseln und die Integration einer derartigen Aufmerksamkeitssteuerung in das aktive Sehsystem NAVIS. Die entwickelte Aufmerksamkeitssteuerung sollte zudem biologienah und psychologisch plausibel sein, verschiedene in NAVIS implizite Modelle, wie die Objekterkennung und die Objektverfolgung, initialisieren können sowie geschlossene Abläufe ermöglichen.

Die entstandene Aufmerksamkeitssteuerung unterscheidet drei kontextabhängige Modelle: die Generierung einer Objekthypothese (Suche nach einem Objekt), die Validierung einer Objekthypothese (Identifikation eines Objekts) und die Verfolgung eines Objekts. Die oberste Ebene der Blicksteuerung bildet die sogenannte Verhaltenssteuerung. Sie wählt situationsangepaßt, z. B. durch ein auf Bewegung reagierendes Alarmsystem, das jeweils geeignetste Modell aus. Eine mittlere Ebene, die als modellgetrieben oder ‘Top down’-System bezeichnet wird, nimmt abhängig von dem jeweiligen Modell Einfluß auf die Blicksteuerung. So gibt z. B. das Modell "Validierung" über eine sogenannte Verstärkungskarte die Positionen vor, an denen die zu einem Objekt gelernten und nun zu validierenden Formen zu erwarten sind. Eine Inhibitionskarte ermöglicht Blickwechsel. Die unterste Ebene arbeitet ausschließlich bilddatengetrieben und extrahiert die Merkmale aus einer Szene, die zur Generierung einer Objekthypothese führen können oder für die Suche nach einem bestimmten Objekt genutzt werden können. Sie wird auch als Attraktivitätsrepräsentation bezeichnet und läßt sich wiederum hierarchisch in eine Merkmals- und eine Auffälligkeitsebene unterteilen. Auf der Ebene der Merkmalskarten werden die statischen Merkmale orientierte Kanten und Grauwertflächen sowie Farben analysiert. Gleichzeitig wird die Szene zwischen den Kamerabewegungen ständig auf das dynamische Merkmal Bewegung überprüft. Die aus den Merkmalskarten gewonnenen Auffälligkeiten beziehen sich auf Symmetrien, Exzentrizitäten und Farbkontraste bzw. zeitliche Änderungen.

Die vorgestellte Arbeit ist im Rahmen des Projektes Entwurf von Systembausteinen zur Aktiven Bildanalyse (ESAB-1) zwei Jahre von der Deutschen Forschungsgemeinschaft gefördert worden. In nachfolgenden Arbeiten soll insbesondere das ‘Feature Binding’ verbessert werden, welches die potentiellen Kandidaten für einen Blickwechsel, die aus verschiedenen Merkmals- und Auffälligkeitskarten hervorgehen, integriert. Hierzu ist die Auswertung von Tiefenhinweisen in das System zu integrieren. Weiterhin ist an eine Erhöhung der Anzahl und Komplexität der zur Verfügung stehenden Modelle gedacht. So soll das System zum Beispiel auch mehrere sich gleichzeitig bewegende Objekte in einer Szene handhaben können. Diese Arbeit wird für weitere zwei Jahre von der DFG gefördert (ESAB-2).

Literaturverzeichnis

[Abbott 1988] Abbott, A. L.; Ahuja, N.: Surface reconstruction by dynamic integration of focus, camera vergence, and stereo. In: Int. Conf. on Computer Vision, 1988, S. 532-543

[Adams 1942] Adams, E. Q.: X-Z planes in the 1931 ICI systems of colorimetry. In: J. of the Optical Society of America 32, 1942, S. 168-173

[Ahmad 1991a] Ahmad, S.; Omohundro, S.: Efficient visual search: A connectionist solution. In: Proc. 13th Annual Conf. of the Cognitive Science Society. Hillsdale (Erlbaum) 1991, S. 293-298

[Ahmad 1991b] Ahmad, S.: VISIT: An efficient computational model of human visual attention. PhD Thesis, University of Illinois, Urbana-Champaign, 1991

[Allport 1989] Allport, A.: Visual attention. In: Posner, M. I. (Hrsg.): Foundations of Cognitive Science. Cambridge/Mass. (MIT Press, Bradford Books) 1989, S. 631-682

[Aloimonos 1987] Aloimonos, J. Y.; Weiss, I.; Bandopadhay, A.: Active vision. In: Int. J. on Computer Vision, Bd. 1, Nr. 3, 1987, S. 333-356

[Aloimonos 1990] Aloimonos, J. Y.: Purposive and qualitative active vision. In: Proc. 10th Int. Conf. on Pattern Recognition, 1990, S. 346-360

[Anderson 1987] Anderson, C. H.; Van Essen, D. C.: Shifter circuits: A computational strategy for dynamic aspects of visual processing. In: Proc. National Academy of Sciences 84, 1987, S. 6297-6301

[Atallah 1985] Atallah, M. J.: On symmetry detection. In: IEEE Transactions on Computers, C-34, 1985, S. 663-666

[Bajcsy 1980] Bajcsy, R.; Rosenthal, D. A.: Visual and conceptual focus of attention. In: Tanimoto, S.; Klinger, A. (Hrsg.): Structured Computer Vision. New York (Academic Press) 1980, S. 133-149

[Bajcsy 1988] Bajcsy, R.: Active perception. In: Proc. of the IEEE, Bd. 76, Nr. 8, 1988, S. 996-1005

[Ballard 1988] Ballard, D. H.; Ozcandarli, A.: Eye fixation and early vision: Kinetic depth. In: IEEE 2nd Int. Conf. on Computer Vision, 1988, S. 524-531

[Ballard 1991] Ballard, D. H.: Animate vision. In: Artificial Intelligence, Bd. 48, Nr. 1, 1991, S. 57-86

[Ballard 1995] Ballard, D. H.; Hayhoe, M. M.; Pelz, J. B.: Memory representations in natural tasks. In: J. of Cognitive Neuroscience, Bd. 7, Nr. 1, 1995, S. 66-80

[Ballard 1997] Ballard, D. H.; Hayhoe, M. M.; Pook, P. K.; Rao, R. P. N.: Deictic codes for the embodiment of cognition. In: Behavioral and Brain Sciences, Bd. 20, Nr. 4, 1997

[Barile 1997] Barile, J.; Bishay, M.; Cambron, M.; Watson, R.; Peters, R. A.; Kawamura, K.: Color-based initialisation for human tracking with a trinocular camera system. In: Proc. of the Int. Conf. on Robotics and Manufactoring, 1997

[Bayouth 1996] Bayouth, M.; Thorpe, C.: An AHS concept based on an autonomous vehicle architecture. In: Proc. of the 3rd Annual World Congress on Intelligent Transportation Systems, 1996

[Becker 1995] Becker, C.; Gonzalez-Banos, H.; Latombe, J. C.; Tomasi, C.: An intelligent observer. In: Proc. 4th Int. Symposium on Experimental Robotics, Stanford, 1995, S. 94-99

[Bergener 1997] Bergener, T.; Bruckhoff, C.; Dahm, P.; Janßen, H.; Joublin, F.; Menzner, R.: Arnold: An anthropomorphic autonomous robot for human environments. In: Groß, H.-M. (Hrsg.): SOAVE '97. Selbstorganisation von Adaptivem Verhalten. Düsseldorf (VDI-Verlag) 1997, S. 25-34

[Bigün 1988] Bigün, J.: Pattern recognition by detection of local symmetries. In: Gelsma, E. S.; Kanal, L. N. (Hrsg.): Pattern Recognition and Artificial Intelligence. Amsterdam u. a. (Elsevier) 1988, S. 75-90

[Bollmann 1995] Bollmann, M.; Mertsching, B.; Drüe, S.: Entwicklung eines Gegenfarbenmodells für das Neuronale-Active-Vision-System NAVIS. In: Sagerer, G.; Posch, S.; Kummert, F. (Hrsg.): Mustererkennung 1995. Berlin u. a. (Springer-Verlag) 1995, S. 456-463 + 668

[Bollmann 1996] Bollmann, M.; Hempel, T.; Mertsching, B.: Improved edge detection by the evaluation of color contrast information. In: 2. Workshop Farbbildverarbeitung. Schriftenreihe des Zentrums für Bild- und Signalverarbeitung Ilmenau, Report 1/96, 1996, S. 1-6

[Bollmann 1997] Bollmann, M.; Hoischen, R.; Mertsching, B.: Integration of static and dynamic scene features guiding visual attention. In: Paulus, E.; Wahl, F. M. (Hrsg.): Mustererkennung 1997. Berlin u. a. (Springer-Verlag) 1997, S. 483-490

[Bollmann 1998] Bollmann, M.; Justkowski, C.; Mertsching, B.: Utilizing color information for the gaze control of an active vision system. In: Rehrmann, V. (Hrsg.): 4. Workshop Farbbildverarbeitung. Koblenz (Fölbach) 1998, S. 73-79

[Bollmann 1999] Bollmann, M.; Hoischen, R.; Jesikiewicz, M.; Justkowski, C.; Mertsching, B.: Playing domino: A case study for an active vision system. In: Christensen, H. I. (Hrsg.): Computer Vision Systems. Berlin u. a. (Springer-Verlag) 1999, S. 392-411

[Brady 1984] Brady, M.; Asada, H.: Smooth local symmetries and their implementation. In: The Int. J. of Robotics Research, Bd. 3, Nr. 3, 1984, S. 36-61

[Brand 1971] Brand, J.: Classification without identification in visual search. In: Quarterly J. of Experimental Psychology 23, 1971, S. 178-186

[Bravo 1990] Bravo, M.; Blake, R.: Preattentive vision and perceptual groups. In: Perception 19, 1990, S. 515-522

[Breitmeyer 1986] Breitmeyer, B. G.: Eye movements and visual pattern perception. In: Schwab, E. C.; Nusbaum, H. C. (Hrsg.): Pattern Recognition by Humans and Machines. (Academic Press) 1986, S. 65-86

[Broadbent 1982] Broadbent, D. E.: Task combination and selective intake of information. In: Acta Psychologica 50, 1982, S. 253-290

[Brunnström 1994] Brunnström, K.; Eklundh, J.-O.; Uhlin, T.: Active fixation for scene exploration. In: Int. J. of Computer Vision, Bd. 17, 1994, S. 137-162

[Buhmann 1991] Buhmann, J.; Lange, J.; v. d. Malsburg, C.; Vorbrüggen, J. C.; Würtz, R. P.: Object recognition with Gabor functions in the Dynamic Link Architecture - Parallel implementation on a transputer network. In: Kosko, B. (Hrsg.): Neural Networks for Signal Processing. Englewood Cliffs (Prentice Hall) 1991, S. 121-159

[Buhmann 1995] Buhmann, J.; Burgard, W.; Cremers, A. B.; Fox, D.; Hofmann, T.; Thrun, S.: The mobile robot rhino. In: AI Magazin, Bd. 16, Nr. 1, 1995

[Bundesen 1983] Bundesen, C.; Pedersen, L. F.: Color segregation and visual search. In: Perception and Psychophysics 33, 1983, S. 487-493

[Bundesen 1990] Bundesen, C.: A theory of visual attention. In: Psychological Review, Bd. 97, Nr. 4, 1990, S. 523-547

[Cahn von Seelen 1996] Cahn von Seelen, U. M.; Bajcsy, R.: Model-based gaze control. In: Proc. of Image Understanding Workshop, Palm Springs, 1996, S. 1361-1364

[Campbell 1968] Campbell, F. W.; Robson, J. G.: Applications of Fourier analysis to the visibility of gratings. In: J. Physiology 197, 1968, S. 551-556

[Carpenter 1988] Carpenter, R. H. S.: Movements of the Eyes. London (Pion) 1988

[Cavanagh 1990] Cavanagh, P.; Arguin, M.; Treisman, A.: Effect of surface medium on visual search for orientation and size features. In: J. Experimental Psychology: Human Perception and Performance, Bd. 16, Nr. 3, 1990, S. 479-491

[Cave 1990] Cave, K. R.; Wolfe, J. M.: Modeling the role of parallel processing in visual search. In: Cognitive Psychology 22, 1990, S. 225-271

[Chelazzi 1993] Chelazzi, C.; Miller, E. K.; Duncan, J.; Desimone, R.: A neural basis for search in inferior temporal cortex. In: Nature 363, 1993, S. 345-347

[Chow 1972] Chow, C. K.; Kaneko, T.: Automatic boundary detection of the left-ventricle from cineangiograms. In: Computational Biomedical Research 5, 1972, S. 388-410

[Clark 1988] Clark, J. J.; Ferrier, N. J.: Modal control of an attentive vision system. In: IEEE 2nd Int. Joint Conf. on Computer Vision, 1988, S. 514-523

[Coltheart 1984] Coltheart, M.: Sensory memory: A tutorial review. In: Bouma, H.; Bouwhuis, D. G. (Hrsg.): Attention and Performance X: Control of Language Processes. Hillsdale (Erlbaum) 1984

[Corbetta 1991] Corbetta, M.; Miezin, F. M.; Dobmeyer, S.; Shulman, G. L.; Petersen, S. E.: Selective and divided attention during visual discrimination of shape, color, and speed: Functional anatomy by positron emission tomography. In: J. of Neuroscience 11, 1991, S. 2383-2402

[Corbetta 1993] Corbetta, M.; Miezin, F. M.; Dobmeyer, S.; Shulman, G. L.; Petersen, S. E.: A PET study of visuospatial attention. In: J. of Neuroscience 13, 1993, S. 1202-1226

[Crick 1984] Crick, F.: Function of the thalamic reticular complex: The searchlight hypothesis. In: Proc. of the National Academy of Sciences 81, 1984, S. 4586-4590

[Culhane 1992a] Culhane, S.; Tsotsos, J.: An attentional prototype for early vision. In: Sandini, G. (Hrsg.): Proc. 2nd Europ. Conf. on Computer Vision. Berlin u. a. (Springer-Verlag) 1992, S. 551-560

[Culhane 1992b] Culhane, S.; Tsotsos, J.: A prototype for data-driven visual attention. In: 11th Int. Conf. on Pattern Recognition, 1992, S. 36-40

[Daugman 1985] Daugman, J. G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. In: J. Optical Society of America A, Bd. 2, Nr. 7, 1985, S. 1160-1169

[Davidoff 1995] Davidoff, J.: Color perception. In: Arbib, M. A. (Hrsg.): The Handbook of Brain Theory and Neural Networks. Cambridge/Mass. (MIT Press) 1995, S. 210-215

[Desimone 1990] Desimone, R.; Wessinger, M.; Thomas, L.; Schneider, W.: Attentional control of visual perception: Cortical and subcortical mechanisms. In: Cold Spring Harbor Symposia on Quantitative Biology LV, 1990, S. 963-971

[Desimone 1992] Desimone, R.: Neural circuits for visual attention in the primate brain. In: Carpenter, G. A.; Grossberg, S. (Hrsg.): Neural Networks for Vision and Image Processing. Cambridge/Mass. (MIT Press) 1992, S. 343-364

[de Valois 1993] de Valois, R. L.; de Valois, K. K.: A multi-stage color model. In: Vision Research, Bd. 33, Nr. 8, 1993, S. 1053-1065

[DeYoe 1988] DeYoe, E. A.; Van Essen D. C.: Concurrent processing streams in monkey visual cortex. In: Trends in Neuroscience, Bd. 11, Nr. 5, 1988, S. 219-226

[Drüe 1994] Drüe, S.; Hoischen, R.; Trapp, R.: Tolerante Objekterkennung durch das Neuronale Active-Vision-System NAVIS. In: Bischof, H.; Kropatsch, W. G. (Hrsg.): Mustererkennung 1994. Wien (Technische Universität Wien) 1994, S. 251-264

[Drüe 1996] Drüe, S.; Trapp, R.: Ein flexibles binokulares Sehsystem: Konstruktion und Kalibrierung. In: Mertsching, B. (Hrsg.): Aktives Sehen in technischen und biologischen Systemen. Sankt Augustin (Infix) 1996, S. 32-39

[Duncan 1984] Duncan, J.: Selective attention and the organization of visual information. In: J. Experimental Psychology: General, Bd. 113, Nr. 4, 1984, S. 501-517

[Duncan 1989] Duncan, J.; Humphreys, G.: Visual search and stimulus similarity. In: Psychological Review, Bd. 96, Nr. 3, 1989, S. 433-458

[Duncan 1992] Duncan, J.; Humphreys, G.: Beyond the search surface: Visual search and attentional engagement. In: J. Experimental Psychology: Human Perception and Performance, Bd. 18, Nr. 2, 1992, S. 578-588

[Egeth 1972] Egeth, H.; Jonides, J.; Wall, S.: Parallel processing of multi-element displays. In: Cognitive Psychology 3, 1972, S. 674-698

[Egeth 1984] Egeth, H. E.; Virzi, R. A.; Garbart, H.: Searching for conjunctively defined targets. In: J. Experimental Psychology: Human Perception and Performance, Bd. 10, Nr. 1, 1984, S. 32-39

[Egner 1997] Egner, S.: Zur Rolle der Aufmerksamkeit für die Objekterkennung: Modellierung, Simulation, Empirie. Dissertation, Fachbereich Informatik, Universität Hamburg, 1997

[Eklundh 1996] Eklundh, J.-O.; Nordlund, P.; Uhlin, T.: Issues in active vision: Attention and cue integration/selection. In: Proc. British Machine Vision Conference, 1996, S. 1-12

[Enns 1991] Enns, J. T.; Rensink, R. A.: Preattentive recovery of three-dimensional orientation from line drawings. In: Psychological Review, Bd. 98, Nr. 3, 1991, S. 335-351

[Enns 1992] Enns, J. T.; Rensink, R. A.: An object completion process in early vision. In: Investigative Ophthalmol. Vis. Sci., Bd. 33, Nr. 4, 1992, S. 1263 (Abstract-Nr. 2852)

[Epstein 1990] Epstein, W.; Babler, T.: In search of depth. In: Perception and Psychophysics, Bd. 48, Nr. 1, 1990, S. 68-76

[Eriksen 1985] Eriksen, C. W.; Yeh, Y.-Y.: Allocation of attention in the visual field. In: J. Experimental Psychology: Human Perception and Performance, Bd. 11, Nr. 5, 1985, S. 583-597

[Eriksen 1986] Eriksen, C. W.; St. James, J. D.: Visual attention within and around the field of focal attention: A zoom lens model. In: Perception and Psychophysics, Bd. 40, Nr. 4, 1986, S. 225-240

[Feyrer 1997] Feyrer, S.; Schimmel, O.; Zell, A.: Adaptive farbbasierte Objektverfolgung mit dem mobilen Roboter ROBIN. In: Groß, H.-M. (Hrsg.): SOAVE ’97. Selbstorganisation von Adaptivem Verhalten. Düsseldorf (VDI-Verlag) 1997, S. 176-185

[Fischer 1993] Fischer, B.; Weber, H.: Express saccades and visual attention. In: Behavioral and Brain Sciences, Bd. 16, Nr. 3, 1993, S. 553-610

[Furneaux 1997] Furneaux, S.; Land, M. F.: The role of eye movements during music reading. Proc. of 3rd ESCOM Conference, Uppsala, 1997, S. 210-214

[Gabor 1946] Gabor, D.: Theory of communication. In: J. IEE, Bd. 93, 1946, S. 429-457

[Giefing 1991] Giefing, G.-J.; Janßen, H.; Mallot, H. A.: A saccadic camera movement system for object recognition. In: Kohonen, T. (Hrsg.): Artificial Neural Networks. (Elsevier Science) 1991, S. 63-68

[Giefing 1992] Giefing, G.-J.; Janßen, H.; Mallot, H.: Saccadic object recognition with an active vision system. In: Proc. Int. Conf. on Pattern Recognition, 1992, S. 664-667

[Goldberg 1991] Goldberg, M. E.; Eggers, H. M.; Gouras, P.: The ocular motor system. In: Kandel, E. R.; Schwartz, J. H.; Jessell, T. M. (Hrsg.): Principles of Neural Science. Third Edition. Englewood Cliffs (Prentice Hall) 1991, S. 660-678

[Gonzales 1987] Gonzales, R. C.; Wintz, P.: Digital Image Processing. Reading/ Mass. (Addison-Wesley Publishing Company) 1987

[Goodale 1992] Goodale, M. A.; Milner, A. D.: Separate visual pathways for perception and action. In: Trends in Neuroscience, Bd. 15, Nr. 1, 1992, S. 20-25

[Götze 1995] Götze, N.: Die mehrstufige Erkennung komplexer Objekte unter Ausnutzung verschiedener Fovealisierungsstrategien mit einem Active-Vision System. Diplomarbeit, Fachbereich Elektrotechnik, Universität-GH Paderborn, 1995

[Götze 1996] Götze, N.; Mertsching, B.; Schmalz, S.; Drüe, S.: Multistage recognition of complex objects with the active vision system NAVIS. In: Mertsching, B. (Hrsg.): Aktives Sehen in technischen und biologischen Systemen. Sankt Augustin (Infix) 1996, S. 186-193

[Gouras 1981] Gouras, P.: Visual System 4: Color Vision. In: Kandel, E. R.; Schwartz, J. H. (Hrsg.): Principles of Neural Science. New York u. a. (Elsevier) 1981, S. 249-257

[Granlund 1994] Granlund, G. H.; Knutsson, H.; Westelius, C.-J.; Wiklund, J.: Issues in robot vision. In: Image and Vision Computing, Bd. 12, Nr. 3, 1994, S. 131-148

[Groner 1989] Groner, R.; Groner, M. T.: Attention and eye movement control: An overview. In: Europ. Archives of Psychiatry and Neurological Science 239, 1989, S. 9-16

[Groß 1997] Groß, H.-M.; Böhme, H.-J.: Verhaltensorganisation in interaktiven und lernfähigen mobilen Systemen - das MILVA-Projekt. In: Groß, H.-M. (Hrsg.): SOAVE ’97. Selbstorganisation von Adaptivem Verhalten. Düsseldorf (VDI-Verlag) 1997, S. 1-13

[Guzzoni 1997] Guzzoni, D.; Cheyer, A.; Julia, L.; Konolige, K.: Many robots make short work. In: AI Magazine, Bd. 18, Nr. 1, 1997, S. 55-64

[Hamker 1997] Hamker, F. H.; Pomierski, T.; Groß, H.-M.; Debes, K.: Ein visuomotorisches Sortiersystem auf der Basis von Farbmerkmalen. In: Groß, H.-M. (Hrsg.): SOAVE ’97. Selbstorganisation von Adaptivem Verhalten. Düsseldorf (VDI-Verlag) 1997, S. 232-238

[Hansen 1992] Hansen, O.; Bigün, J.: Local symmetry modeling in multi-dimensional images. In: Pattern Recognition Letters, Bd. 13, Nr. 4, 1992, S. 253-262

[Heidemann 1998] Heidemann, G.; Nattkemper, T.; Ritter, H.: Farbe und Symmetrie für die datengetriebene Generierung prägnanter Fokuspunkte. In: Rehrmann, V. (Hrsg.): 4. Workshop Farbbildverarbeitung. Koblenz (Fölbach) 1998, S. 65-72

[Henik 1994] Henik, A.; Rafal, R.; Rhodes, D.: Endogenously generated and visually guided saccades after lesions of the human frontal eye fields. In: J. Cognitive Neuroscience, Bd. 6, Nr. 4, 1994, S. 400-411

[Hoffman 1986] Hoffman, J. E.: Spatial attention in vision: Evidence for early selection. In: Psychological Research 48, 1986, S. 221-229

[Hoischen 1999] Hoischen, R.; Springmann, S.; Mertsching, B.: Object tracking in image sequences based on parametric features. Erscheint in: ÖVE Verbandszeitschrift Elektrotechnik und Informationstechnik (e & i)

[Hubel 1982] Hubel, D. H.: Exploration of the primary visual cortex, 1955-78 (Nobel lecture). In: Nature 299, 1982, S. 515-524

[Humphreys 1989] Humphreys, G. W.; Quinlan, P. T.; Riddock, M. J.: Grouping processes in visual search: Effects with single- and combined-feature targets. In: J. Experimental Psychology: General, Bd. 118, Nr. 3, 1989, S. 258-279

[Humphreys 1993] Humphreys, G. W.; Müller, H.: Search via recursive rejection (SERR): A connectionist model of visual search. In: Cognitive Psychology 25, 1993, S. 43-110

[Humphreys 1996] Humphreys, G. W.; Gilchrist, I.; Free, L.: Search and selection in human vision: Psychological evidence and computational implications. In: Zangemeister, W. H.; Stiehl, H. S.; Freksa, C. (Hrsg.): Visual Attention and Cognition. Amsterdam u. a. (Elsevier) 1996, S. 79-93

[Hunt 1987] Hunt, R. W. G.: Measuring Colour. Chichester (Ellis Horwood) 1987

[Itti 1998] Itti, L.; Koch, C.; Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, Bd. 20, Nr. 11, 1998, S. 1254-1259

[Jähne 1997] Jähne, B.: Digitale Bildverarbeitung, 4. Auflage. Berlin u. a. (Springer-Verlag) 1997

[Jenning 1997] Jenning, C.; Murray, D.: Gesture recognition for robot control. In: Proc. of the Conf. on Computer Vision and Pattern Recognition, 1997

[Johnston 1985] Johnston, W. A.; Dark, V. J.: Dissociable domains of selective processing. In: Posner, M. I.; Marin, O. S. M. (Hrsg.): Attention and Performance XI: Mechanisms of Attention. Hillsdale (Erlbaum) 1985

[Johnston 1986] Johnston, W. A.; Dark, V. J.: Selective attention. In: Annual Review of Psychology 37, 1986, S. 43-75

[Jones 1987a] Jones, J. P.; Palmer, L. A.: The two-dimensional spectral structure of simple receptive fields in cat striate cortex. In: J. Neurophysiology, Bd. 58, Nr. 6, 1987, S. 1187-1211

[Jones 1987b] Jones, J. P.; Stepnoski, A.; Palmer, L. A.: The two-dimensional spectral structure of simple receptive fields in cat striate cortex. In: J. Neurophysiology, Bd. 58, Nr. 6, 1987, S. 1212-1232

[Jones 1987c] Jones, J. P.; Palmer, L. A.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. In: J. Neurophysiology, Bd. 58, Nr. 6, 1987, S. 1233-1258

[Jonides 1972] Jonides, J.; Gleitman, H.: A conceptual category effect in visual search: O as a letter or a digit. In: Perception and Psychophysics 12, 1972, S. 457-460

[Jonides 1981] Jonides, J.: Voluntary versus automatic control over the mind's eye's movement. In: Long, J.; Baddeley, A. (Hrsg.): Attention and Performance IX. Hillsdale (Erlbaum) 1981, S. 187-203

[Jonides 1983] Jonides, J.: Further toward a model of the mind's eye's movement. In: Bulletin of the Psychonomic Society 21, 1983, S. 247-250

[Justkowski 1998] Justkowski, C.: Integration des Merkmals Farbe in die Aufmerksamkeitssteuerung des aktiven Sehsystems NAVIS. Studienarbeit, Fachbereich Informatik, Universität Hamburg, 1998

[Kandel 1991] Kandel, E. R.; Schwartz, J. H.; Jessell, T. M.: Principles of Neural Science. Third Edition. Englewood Cliffs (Prentice Hall) 1991

[Kaufman 1969] Kaufman, L.; Richards, W.: Spontaneous fixation tendencies for visual forms. In: Perception and Psychophysics, Bd. 5, Nr. 2, 1969, S. 85-88

[Koch 1985] Koch, C.; Ullman, S.: Shifts in selective visual attention: Towards the underlying neural circuitry. In: Human Neurobiology 4, 1985, S. 219-227

[Kolb 1996] Kolb, B.; Whishaw, I. Q.: Neuropsychologie. 2. Auflage. Heidelberg u. a. (Spektrum) 1996

[Konolige 1997] Konolige, K.; Myers, K.; Saffiotti, A.; Ruspini, E.: The Saphira architecture: A design for autonomy. In: J. of Experimental and Theoretical Artificial Intelligence 9, 1997, S. 215-235

[Korries 1985] Korries, R. R.: Maschinelles Bewegungssehen in natürlichen Szenen: Die Auswertung von Bildfolgen gestützt auf Bildmerkmale. Dissertation, Fachbereich Biologie, Johannes-Gutenberg-Universität Mainz, 1985

[Krotkov 1989] Krotkov, E. P.: Active Computer Vision by Cooperative Focus and Stereo. Berlin u. a. (Springer-Verlag) 1989

[LaBerge 1990] LaBerge, D.: Thalamic and cortical mechanism of attention suggested by recent positron emission tomographic experiments. In: J. Cognitive Neuroscience, Bd. 2, Nr. 4, 1990, S. 358-372

[LaBerge 1992] LaBerge, D.; Carter, M.; Brown, V.: A network simulation of thalamic circuit operations in selective attention: In: Neural Computation 4, 1992, S. 318-331

[Levy-Schoen 1981] Levy-Schoen, A.: Flexible and/or rigid control of oculomotor scanning behavior. In: Senders, J. W. (Hrsg.): Eye Movements: Cognition and Visual Perception. Hillsdale (Erlbaum) 1981, S. 299-314

[Lieder 1997] Lieder, T.: Objekterkennung unter Ausnutzung von Tiefenhinweisen. Studienarbeit, Fachbereich Informatik, Universität Hamburg, 1997

[Lieder 1998] Lieder, T.; Mertsching, B.; Schmalz, S.: Using depth information for invariant object recognition. In: Posch, S.; Ritter, H. (Hrsg.): Dynamische Perzeption. St. Augustin (Infix) 1998, S. 9-16

[Livingstone 1987] Livingstone, M. S.; Hubel, D. H.: Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. In: The J. of Neuroscience, Bd. 7, Nr. 11, 1987, S. 3416-3468

[Locher 1987] Locher, P. J.; Nodine, C. F.: Symmetry catches the eye. In: Levy-Schoen, A. (Hrsg.): Eye Movements: From Physiology to Cognition. Amsterdam u. a. (Elsevier) 1987, S. 353-361

[Maki 1996] Maki, A.; Nordlund, P.; Eklundh, J.-O.: A computational model of depth-based attention. In: Proc. 13th Int. Conf. on Pattern Recognition, 1996, S. 734-738 in Bd. 4

[Mangun 1988] Mangun, G. R.; Hillyard, S. A.: Spatial gradients of visual attention: Behavioral and electrophysiological evidence. In: Electroencephalography and Clinical Neurophysiology 70, 1988, S. 417-428

[Marcelja 1980] Marcelja, S.: Mathematical description of the responses of simple cortical cells. In: J. Optical Society of America, Bd. 70, Nr. 11, 1980, S. 1297-1300

[Marola 1989] Marola, G.: On the detection of the axis of symmetry of symmetric and almost symmetric plannar images. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, Bd. 11, Nr. 1, 1989, S. 104-108

[Massad 1997] Massad, A.: Entwicklung eines Bildverarbeitungssystems zur Erkennung von dreidimensionalen Objekten anhand von Ansichtensequenzen. Diplomarbeit, Fachbereich Informatik, Universität Hamburg, 1997

[Massad 1998a] Massad, A.; Mertsching, B.; Schmalz, S.: Combining multiple views and temporal associations for 3-D object recognition. In: Burkhardt, H.: Neumann, B. (Hrsg.): Computer Vision - ECCV'98. Volume 2. Berlin u. a. (Springer-Verlag) 1998, S. 699-715

[Massad 1998b] Massad, A.; Mertsching, B.; Schmalz, S.: Utilizing temporal associations for view-based 3-D object recognition. In: Proc. of the 24th Annual Conference of the IEEE Industrial Electronics Society (IECON'98), Volume 4, 1998, S. 2074-2078

[Maurer 1997] Maurer, M.; Dickmanns, E.: A system architecture for autonomous visual road vehicle guidance. In: IEEE Conf. on Intelligent Transportation Systems, Boston, 1997

[Maylor 1985] Maylor, E. A.; Hockey, R.: Inhibitory component of externally controlled covert orienting in visual space. In: J. Experimental Psychology: Human Perception and Performance 11, 1985, S. 777-787

[Maylor 1987] Maylor, E. A.; Hockey, R.: Effects of repetition on the facilitatory and inhibitory components of orienting in visual space. In: Neuropsychologia 25, 1987, S. 41-54

[Mertsching 1997] Mertsching, B.; Bollmann, M.: Visual attention and gaze control for an active vision system. In: Kasabov, N.; Kozma, R.; Ko, K.; O'Shea, R.; Coghill, G.; Gedeon, T. (Hrsg.): Progress in Connectionist-Based Information Systems. Volume 1. Singapur u. a. (Springer-Verlag) 1997, S. 76-79

[Mertsching 1998] Mertsching, B.; Bollmann, M.; Massad, A.; Schmalz, S.: Recognition of complex objects with an active vision system. In: Proc. of Int. ICSC/IFAC Symposium on Neural Computation (NC'98). Kanada/ Schweiz (ICSC Academic Press) 1998, S. 469-475

[Mertsching 1999] Mertsching, B.; Bollmann, M.; Hoischen, R.; Schmalz, S.: The neural active vision system NAVIS. In: Jähne, B.; Haußecker, H.; Geißler, P. (Hrsg.): Handbook of Computer Vision and Applications. Vol. 3 (Systems and Applications). San Diego (Academic Press) 1999, S. 543-568

[Mesulam 1985] Mesulam, M. M.: Attention, confusional states, and neglect. In: Mesulam, M. M. (Hrsg.): Principles of Behavioral Neurology. Philadelphia (F. A. Davis) 1985, S. 125-168

[Mewhort 1984] Mewhort, D. J. K.; Marchetti, F. M.; Gurnsey, R.; Campbell, A. J.: Information persistence: A dual-buffer model for initial visual processing. In: Bouma, H.; Bouwhuis, D. G. (Hrsg.): Attention and Performance X: Control of Language Processes. Hillsdale (Erlbaum) 1984

[Milanese 1991] Milanese, R.: Detection of salient features for focus of attention. In: Proc. 3rd Meeting of the Swiss Group for Artificial Intelligence and Cognitive Science, 1991

[Milanese 1993] Milanese, R.: Detecting salient regions in an image: From biological evidence to computer implementation. Dissertation, Department of Computer Science, University of Geneva, 1993

[Milanese 1994] Milanese, R.; Wechsler, H.; Gil, S.; Bost, J.-M.; Pun, T.: Integration of bottom-up and top-down cues for visual attention using non-linear relaxation. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1994, S. 781-785

[Miller 1991] Miller, E. K.; Li, L.; Desimone, R.: A neural mechanism for working and recognition memory in inferior temporal cortex. In: Science 254, 1991, S. 1377-1379

[Milner 1993] Milner, A. D.; Goodale, M. A.: Visual pathways to perception and action. In: Hicks, T. P.; Molotchnikoff, S.; Ono, T. (Hrsg.): Progress in Brain Research 95. Amsterdam u. a. (Elsevier) 1993, S. 317-337

[Mishkin 1983] Mishkin, M.; Ungerleider, L. G.; Macko, K. A.: Object vision and spatial vision: Two cortical pathways. In: Trends in Neurosciences 6, 1983, S. 414-417

[Miyahara 1988] Miyahara, M.; Yoshida, Y.: Mathematical transform of (R,G,B) color data to Munsell (H,V,C) color data. In: Visual Communication and Image Processing 1001, 1988, S. 650-657

[Moran 1985] Moran, J.; Desimone, R.: Selective attention gates visual processing in the extrastriate cortex. In: Science 229, 1985, S. 782-784

[Mordkoff 1990] Mordkoff, J. T.; Yantis, S.; Egeth, H. E.: Detecting conjunctions of color and form in parallel. In: Perception and Psychophysics, Bd. 48, Nr. 2, 1990, S. 157-168

[Motter 1993] Motter, B. C.: Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. In: J. of Neurophysiology, Bd. 70, Nr. 3, 1993, S. 909-919

[Nakayama 1986] Nakayama, K; Silverman, G.: Serial and parallel processing of visual feature conjunctions. In: Nature 320, 1986, S. 264-265

[Neill 1987] Neill, W. T.; Westberry, R. L.: Selective attention and the suppression of cognitive noise. In: J. of Experimental Psychology: Learning, Memory, and Cognition 13, 1987, S. 327-334

[Norman 1985] Norman, D. A.; Shallice, T.: Attention to action: Willed and automatic control of behavior. In: Davidson, R. J.; Schwartz, G. E.; Shapiro, D. (Hrsg.): Consciousness and Self Regulation: Advances in Research, Bd. IV. New York (Plenum Press) 1985, S. 1-18

[Noton 1970] Noton, D.: A theory of visual pattern perception. In: IEEE Trans. on Systems, Sci., and Cybernetics 6, 1970, S. 349-357

[Noton 1971a] Noton, D.; Stark, L.: Eye movements and visual perception. In: Scientific American, Bd. 224, Nr. 6, 1971, S. 34-43

[Noton 1971b] Noton, D.; Stark, L.: Scanpaths in eye movements during pattern perception. In: Science 171, 1971, S. 308-311

[Olshausen 1993] Olshausen, B.; Anderson, C.; Van Essen, D.: A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. In: The J. Neuroscience, Bd. 13, Nr. 11, 1993, S. 4700-4719

[Olshausen 1995] Olshausen, B. A.; Anderson, C. H.; Van Essen, D. C.: A multiscale dynamic routing circuit for forming size- and position-invariant object representations. In: J. of Computational Neuroscience, Bd. 2, Nr. 1, 1995, S. 45-62

[O'Regan 1990] O'Regan, J. K.: Eye movements and reading. In: Kowler, E. (Hrsg.): Eye Movements and Their Role in Visual and Cognitive Processes. New York u. a. (Elsevier) 1990, S. 455-477

[Oswald 1997] Oswald, N.; Levi, P.: Cooperative vision in a multi-agent architecture. In: Image Analysis and Processing. Berlin u. a. (Springer-Verlag) 1997, S. 709-716

[Pahlavan 1996] Pahlavan, K.: Designing an anthromorphic head-eye system. In: Zangemeister, W. H.; Stiehl, H. S.; Freksa, C. (Hrsg.): Visual Attention and Vision. Amsterdam u. a. (Elsevier) 1996, S. 269-292

[Pashler 1987] Pashler, H.: Detecting conjunctions of color and form: Reassessing the serial search hypothesis. In: Perception and Psychophysics 41, 1987, S. 191-201

[Phaf 1990] Phaf, R. H.; van der Heijden, A. H. C.; Hudson. P. T. W.: SLAM: A connectionist model for attention in visual selection tasks. In: Cognitive Psychology 22, 1990, S. 273-341

[Podgorny 1983] Podgorny, P.; Shepard, R. N.: The distribution of visual attention over space. In: J. of Experimental Psychology: Human Perception and Performance 9, 1983, S. 380-393

[Poisson 1992] Poisson, M. E.; Wilkinson, F.: Distractor ratio and grouping processes in visual conjunction search. In: Perception 21, 1992, S. 21-38

[Pollen 1981] Pollen, D. A.; Ronner, S. F.: Phase relationships between adjacent cells in the visual cortex. In: Science 212, 1981, S. 1409-1411

[Pomierski 1996] Pomierski, T.: Neurophysiologisch motivierte Architektur zur Erzeugung stabiler Farb- und Texturrepräsentationen. Dissertation, Fakultät für Informatik und Automatisierung, Technische Universität Ilmenau, 1996

[Posner 1980a] Posner, M. I.: Orienting of attention. In: Quarterly J. of Experimental Psychology 32, 1980, S. 3-25

[Posner 1980b] Posner, M. I.; Snyder, C. C. R.; Davidson, B. J.: Attention and the detection of signals. In: J. of Experimental Psychology: General 109, 1980, S. 160-174

[Posner 1982] Posner, M. I.; Cohen, Y.; Rafal, R. D.: Neural systems control of spatial orienting. In: Phil. Trans. Royal Society London B 298, 1982, S. 187-198

[Posner 1984a] Posner, M. I.; Cohen, Y.; Choate, L. S.; Hockey, R.; Maylor, E.: Sustained concentration: Passive filtering or active orienting? In: Kornblum, S.; Requin, J. (Hrsg.): Preparatory States and Processes. Hillsdale (Erlbaum) 1984, S. 49-65

[Posner 1984b] Posner, M. I.; Walker, J. A.; Friedrich, F. J.; Rafal, R. D.: Effects of parietal injury on covert orienting of attention. In: J. Neuroscience, Bd. 4, Nr. 7, 1984, S. 1863-1874

[Posner 1987] Posner, M. I.; Presti, D. E.: Selective attention and cognitive control. In: Trends in Neuroscience 10, 1987, S. 13-17

[Posner 1988] Posner, M. I.; Petersen, S. E.; Fox, P. T.; Raichle, M. E.: Localization of cognitive operations in the human brain. In: Science 240, 1988, S. 1627-1631

[Posner 1994] Posner, M. I.; Dehaene, S.: Attentional networks. In: Trends in Neurosciences, Bd. 17, Nr. 2, 1994, S. 75-79

[Postma 1994] Postma, E. O.: SCAN: A neural model of covert attention. PhD Thesis, Rijksuniversiteit Limburg, Wageningen, 1994

[Pylyshyn 1994a] Pylyshyn, Z.: Some primitive mechanisms of spatial attention. In: Cognition 50, 1994, S. 363-384

[Pylyshyn 1994b] Pylyshyn, Z.; Burkell, J.; Fisher, B.; Sears, C.; Schmidt, W.; Trick, L.: Multiple parallel access in visual attention. In: Canadian J. of Experimental Psychology, Bd. 48, Nr. 2, 1994, S. 260-283

[Quinlan 1987] Quinlan, P. T.; Humphreys, G. W.: Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. In: Perception and Psychophysics, Bd. 41, Nr. 5, 1987, S. 455-472

[Rafal 1988] Rafal, R. D.; Posner, M. I.; Friedman, J. H.; Inhoff, A. W.; Bernstein, E.: Orienting of visual attention in progressive supranuclear palsy. In: Brain 111, 1988, S. 267-280

[Rao 1997] Rao, R. P. N.; Zelinsky, G. J.; Hayhoe, M. M.; Ballard, D. H.: Eye movements in visual cognition: A computational study. In: Technical Report 97.1, National Resource Laboratory for the Study of Brain and Behavior, University of Rochester, 1997

[Reisfeld 1995] Reisfeld, D.; Wolfson, H.; Yeshurun, Y.: Context free attentional operators: The generalized symmetry transform. In: Int. J. Computer Vision 14, 1995, S. 119-130

[Remington 1980] Remington, R. W.: Attention and saccadic eye movements. In: J. of Experimental Psychology 6, 1980, S. 726-744

[Remington 1984] Remington, R.; Pierce, L.: Moving attention: Evidence for time-invariant shifts of visual selective attention. In: Perception and Psychophysics, Bd. 35, Nr. 4, 1984, S. 393-399

[Rosenfeld 1982] Rosenfeld, A.; Kak, A. C.: Digital Picture Processing. Bd. 1 und 2. New York (Academic Press) 1982

[Rosenthal 1998] Rosenthal, M.: Kollisionsfreie Navigation eines mobilen Roboters durch Auswertung von Sonar- und Bilddaten. Diplomarbeit, Fachbereich Informatik, Universität Hamburg, 1998

[Sandon 1990] Sandon, P. A.: Simulating visual attention. In: J. Cognitive Neuroscience, Bd. 2, Nr. 3, 1990, S. 213-231

[Schmidt 1995] Schmidt, B.: Erweiterung des Ahmad'schen Attentionsmechanismus. Diplomarbeit, Lehrstuhl für Kognitive Systeme, Christian-Albrechts-Universität zu Kiel, 1995

[Schneider 1977] Schneider, W.; Shiffrin, R. M.: Controlled and automatic human information processing: I. Detection, search, and attention. In: Psychological Review 84, 1977, S. 1-66

[Schneider 1984] Schneider, W.; Dumais, S. T.; Shiffrin, R. M.: Automatic and control processing and attention. In: Parasuraman, R.; Davies, D. R. (Hrsg.): Varieties of Attention. New York (Academic Press) 1984, S. 1-27

[Sekuler 1990] Sekuler, R.; Blake, R.: Perception. 2. Auflage. (McGraw-Hill) 1990

[Sheperd 1986] Sheperd, M.; Findlay, J. M.; Hockey, R. J.: The relationship between eye movements and spatial attention. In: The Quarterly J. Experimental Psychology, Bd. 38A, 1986, S. 475-491

[Shiffrin 1977] Shiffrin, R. M.; Schneider, W.: Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. In: Psychological Review 84, 1977, S. 127-190

[Shulman 1979] Shulman, G. L.; Remington, R. W.; McLean, J. P.: Moving attention through visual space. In: J. Experimental Psychology: Human Perception and Performance, Bd. 5, Nr. 3, 1979, S. 522-526

[Sparks 1986] Sparks, D. L.: Translation of sensory signals into commands for control of saccadic eye movements: Role of primate superior colliculus. In: Physiological Review, Bd. 66, Nr. 1, 1986, S. 118-171

[Spitzer 1988] Spitzer, H.; Desimone, R.; Moran, J.: Increased attention enhances both behavioral and neuronal performance. In: Science 240, 1988, S. 338-340

[Springmann 1998] Springmann, S. O.: Merkmalskorrespondenz in realen Bildsequenzen zur Bewegungsschätzung und Verfolgung von Objekten. Diplomarbeit, Fachbereich Informatik, Universität Hamburg, 1998

[Takacs 1996] Takacs, B.; Wechsler, H.: Attention and pattern detection using sensory and reactive control mechanisms. In: Proc. 13th Int. Conf. on Pattern Recognition, 1996, S. 19-23 in Bd. 4

[Tanenhaus 1995] Tanenhaus, M. K.; Spivey-Knowlton, M. J.; Eberhard, K. M.; Sedivy, J. E.: Integration of visual and linguistic information in spoken language comprehension. In: Science 268, 1995, S. 632-634

[Tassinari 1987] Tassinari, G.; Aglioti, S.; Chelazzi, L.; Marci, C. A.; Berlucchi, G.: Distribution in the visual field of the costs of voluntarily allocated attention and of the inhibitory after-effects of covert orienting. In: Neuropsychologia 25, 1987, S. 55-71

[Terzopoulos 1995] Terzopoulos, P.; Rabie, T. F.: Animate vision: Active vision in artificial animals. In: Proc. Int. Conf. on Computer Vision, 1995, S. 801-808

[Trapp 1995] Trapp, R.; Drüe, S.; Mertsching, B.: Korrespondenz in der Stereoskopie bei räumlich verteilten Merkmalsrepräsentationen im Neuronalen-Active-Vision-System NAVIS. In: Sagerer, G.; Posch, S.; Kummert, F. (Hrsg.): Mustererkennung 1995. Berlin u. a. (Springer-Verlag) 1995, S. 494-499

[Trapp 1996] Trapp, R.: Entwurf einer Filterbank auf der Basis neurophysiologischer Erkenntnisse zur orientierungs- und frequenzselektiven Dekomposition von Bilddaten. Technischer Bericht, Heinz Nixdorf Institut und Universität-GH Paderborn, 1996

[Treisman 1980] Treisman, A. M.; Gelade, G.: A feature-integration theory of attention. In: Cognitive Psychology 12, 1980, S. 97-136

[Treisman 1984] Treisman, A.; Paterson, R.: Emergent features, attention and object perception. In: J. Experimental Psychology: Human Perception and Performance, Bd. 10, Nr. 1, 1984, S. 12-31

[Treisman 1985] Treisman, A. M.: Preattentive processing in vision. In: Computer Vision, Graphics, and Image Processing 31, 1985, S. 156-177

[Treisman 1988] Treisman, A.; Gormican, S.: Feature analysis in early vision: Evidence from search asymmetries. In: Psychological Review, Bd. 95, Nr. 1, 1988, S. 15-48

[Treisman 1990] Treisman, A.; Sato, S.: Conjunction search revisited. In: J. Experimental Psychology: Human Perception and Performance, Bd. 16, Nr. 3, 1990, S. 459-478

[Treisman 1991] Treisman, A.: Search, similarity, and integration of features between and within dimensions. In: J. Experimental Psychology: Human Perception and Performance, Bd. 17, Nr. 3, 1991, S. 652-676

[Treisman 1992] Treisman, A.: Spreading suppression or feature integration: A reply to Duncan and Humphreys (1992). In: J. Experimental Psychology: Human Perception and Performance, Bd. 18, Nr. 2, 1992, S. 589-593

[Tsal 1983] Tsal, Y.: Movements of attention across the visual field. In: J. Experimental Psychology: Human Perception and Performance, Bd. 9, Nr. 4, 1983, S. 523-530

[Tsotsos 1993] Tsotsos, J. K.: An inhibitory beam for attentional selection. In: Harris, L.; Jenkin, M. (Hrsg.): Spatial visions in humans and robots. Cambridge (Cambridge University Press) 1993, S. 313-331

[Tsotsos 1995] Tsotsos, J. K.; Culhane, S. M.; Wai, W. Y. K.; Lai, Y. H.; Davis, N.; Nuflo, F.: Modeling visual-attention via selective tuning. In: Artificial Intelligence, Bd. 78, Nr. 1-2, 1995, S. 507-545

[Ullman 1984] Ullman, S.: Visual routines. In: Cognition 18, 1984, S. 97-159

[Umilta 1988] Umilta, C.: Orienting of attention. In: Boller, F.; Grafman, J. (Hrsg.): Handbook of Neuropsychology. Volume 1. Amsterdam u. a. (Elsevier) 1988, S. 175-193

[Ungerleider 1982] Ungerleider, L. G.; Mishkin, M.: Two cortical visual systems. In: Ingle, D. J.; Goodale, M. A.; Mansfield, R. J. W. (Hrsg.): The Analysis of Visual Behavior. Cambridge/Mass. (MIT Press) 1982, S. 549-586

[van de Laar 1997] van de Laar, P.; Heskes, T.; Gielen, S.: Task-dependent learning of attention. In: Neural Networks, Bd. 10, Nr. 1, 1997, S. 1-21

[Vestli 1996] Vestli, S.; Tschichold, N.: MoPS, a system for mail distribution in office-type buildings. In: Service Robot: An International Journal, Bd. 2, Nr. 2, 1996, S. 29-35

[Vetter 1997] Vetter, V.; Zielke, T.; von Seelen, W.: Integrating face recognition into security systems. In: Bigün, J.; Chollet, G.; Borgefors, G. (Hrsg.): Audio- and Video-based Biometric Person Authentication. Berlin u. a. (Springer-Verlag) 1997, S. 439-448

[Vieville 1995] Vieville, T.; Clergue, E.; Enriso, R.; Mathieu, H.: Experimenting with 3-D vision on a robotic head. In: Robotics and Autonomous Systems, Bd. 14, Nr. 1, 1995, S. 1-27

[von der Heydt 1984] von der Heydt, R.; Peterhans, E.; Baumgartner, G.: Illusory contours and cortical neuron response. In: Science 224, 1984, S. 1260-1262

[von Seelen 1996] von Seelen, W.; Janßen, H.: Visual navigation for mobile robots. In: Mertsching, B. (Hrsg.): Aktives Sehen in technischen und biologischen Systemen. Sankt Augustin (Infix) 1996, S. 127-136

[Wahl 1984] Wahl, F.: Digitale Bildsignalverarbeitung. Berlin u. a. (Springer-Verlag) 1984

[Wang 1994] Wang, Q.; Cavanagh, P.; Green, M.: Familiarity and pop-out in visual search. In: Perception and Psychophysics, Bd. 56, Nr. 5, 1994, S. 495-500

[Weber 1994] Weber, H.; Fischer, B.: Differential effects of non-target stimuli on the occurrence of express saccades in man. In: Vision Research 34, 1994, S. 1883-1891

[Westelius 1991] Westelius, C.-J.; Knutsson, H.; Granlund, H. G.: Focus of attention control. In: Proc. 7th Scandinavian Conf. on Image Analysis, 1991, S. 667-674

[Westelius 1995] Westelius, C.-J.: Focus of attention and gaze control for robot vision. Dissertations No. 379, Department of Electrical Engineering, Linköping University, 1995

[Wiklund 1987] Wiklund, J.; Granlund, G.: Tracking of multiple moving objects. In: Cappellini, V. (Hrsg.): Time-Varying Image Processing and Moving Object Recognition. Amsterdam u. a. (Elsevier) 1987, S. 241-251

[Wolfe 1989] Wolfe, J. M.; Cave, K. R.; Franzel, S. L.: Guided search: An alternative to the feature integration model for visual search. In: J. Experimental Psychology: Human Perception and Performance, Bd. 15, Nr. 3, 1989, S. 419-433

[Wolfe 1993] Wolfe, J. M.: Guided seach 2.0: The upgrade. In: Proc. Human Factors and Ergonomics Society 37th Annual Meeting, 1993, S. 1295-1299

[Wolfe 1994] Wolfe, J. M.: Guided search 2.0: A revised model of visual search. In: Psychonomic Bulletin and Review, Bd. 1, Nr. 2, 1994, S. 202-238

[Yamada 1995] Yamada, K.; Cottrell, G. W.: A model of scan paths applied to face recognition. In: Proc. 17th Annual Conf. of the Cognitive Science Society, 1995, S. 55-60

[Yamamoto 1996] Yamamoto, H.; Yeshurun, Y.; Levine, M. D.: An active foveated vision system: Attentional mechanisms and scan path convergence measures. In: Computer Vision and Image Understanding, Bd. 63, Nr. 1, 1996, S. 50-65

[Yantis 1990] Yantis, S.; Johnston, J. C.: On the locus of visual selection: Evidence from focused attention tasks. In: J. Experimental Psychology: Human Perception and Performance, Bd. 16, Nr. 1, 1990, S. 135-149

[Yarbus 1969] Yarbus, A. L.: Eye Movements and Vision. New York (Plenum Press) 1969

[Young 1986] Young, R. A.: The Gaussian derivative model for machine vision: Visual cortex simulation. Technical Report GMR-5323, General Motors Research Laboratories, 1986

[Zabrodsky 1995] Zabrodsky, H.; Peleg, S.; Avnir, D.: Symmetry as a continuous feature. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, Bd. 17, Nr. 12, 1995, S. 1154-1166

[Zeki 1993] Zeki, S.: A Vision of the Brain. Oxford (Blackwell Scientific) 1993

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, diese Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen benutzt zu haben. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Hamburg, den 26. April 2000

(Maik Bollmann)

Lebenslauf

Ich wurde am 9. November 1968 in Herford, Westfalen geboren. Von 1979 bis 1988 besuchte ich das Städtische Gymnasium Löhne, Kreis Herford. 1988 beendete ich meine Schullaufbahn mit der Allgemeinen Hochschulreife und begann im Oktober das Studium der Elektrotechnik an der Universität-GH Paderborn mit dem Schwerpunkt Nachrichtentechnik. Das erforderliche dreimonatige Fachpraktikum absolvierte ich 1992 an der Zentralen Forschungseinrichtung der Robert Bosch GmbH in Hildesheim. Meine Diplomarbeit mit dem Titel "Untersuchung eines Gegenfarbenmodells für ein Bildanalysesystem auf der Basis künstlicher neuronaler Netzwerke" schloß ich 1994 ab. Im gleichen Jahr verließ ich die Universität-GH Paderborn mit dem Abschluß Dipl.-Ing. Elektrotechnik und wechselte an die Universität Hamburg, Fachbereich Informatik. Dort arbeitete ich von Oktober 1994 bis März 1999 an der hier vorliegenden Dissertation "Entwicklung einer Aufmerksamkeitssteuerung für ein aktives Sehsystem". Während dieser Zeit wurde ich zwei Jahre von der Deutschen Forschungsgemeinschaft gefördert. In der Lehre war ich insbesondere beteiligt an den Projektseminaren "Farbbildverarbeitung" und "Aktive Sehsysteme".

Meine Interessen umfassen die digitale Signal- und Bildverarbeitung, die Telekommunikation und Software Engineering.

Einige Teile des vorliegenden Werkes sind bereits in Tagungsbänden oder Zeitschriften veröffentlicht worden. Im einzelnen war ich bisher an folgenden Veröffentlichungen beteiligt:

Bollmann, M.; Mertsching, B.; Drüe, S.: Entwicklung eines Gegenfarbenmodells für das Neuronale-Active-Vision-System NAVIS. In: Sagerer, G.; Posch, S.; Kummert, F. (Hrsg.): Mustererkennung 1995. Berlin u. a. (Springer-Verlag) 1995, S. 456-463 + 668

Bollmann, M.; Mertsching, B.: Vergleich zweier Farbkonstanzalgorithmen für die Bildanalyse. In: Rehrmann, V. (Hrsg.): 1. Workshop Farbbildverarbeitung. Fachberichte Informatik 15/95, Universität Koblenz-Landau, 1995, S. 52-55

Bollmann, M.; Hempel, T.; Mertsching, B.: Improved Edge Detection by the Evaluation of Color Contrast Information. In: Franke, K.-H. (Hrsg.): 2. Workshop Farbbildverarbeitung. Schriftenreihe des Zentrums für Bild- und Signalverarbeitung Ilmenau, Report 1/96, Technische Universität Ilmenau, 1996, S. 1-6

Bollmann, M.; Mertsching, B.: Opponent Color Processing Based on Neural Models. In: Perner, P.; Wang, P.; Rosenfeld, A. (Hrsg.): Advances in Structural and Syntactical Pattern Recognition. Berlin u. a. (Springer-Verlag) 1996, S. 198-207

Bollmann, M.; Hoischen, R.; Mertsching, B.: Integration of Static and Dynamic Scene Features Guiding Visual Attention. In: Paulus, E.; Wahl. F. M. (Hrsg.): Mustererkennung 1997. Berlin u. a. (Springer-Verlag) 1997, S. 483-490

Mertsching, B.; Bollmann, M.: Visual Attention and Gaze Control for an Active Vision System. In: Kasabov, N.; Kozma, R.; Ko, K.; O'Shea, R.; Coghill, G.; Gedeon, T. (Hrsg.): Progress in Connectionist-Based Information Systems. Volume 1. Singapur u. a. (Springer-Verlag) 1997, S. 76-79

Bollmann, M.; Justkowski, C.; Mertsching, B.: Utilizing Color Information for the Gaze Control of an Active Vision System. In: Rehrmann, V. (Hrsg.): 4. Workshop Farbbildverarbeitung. Koblenz (Fölbach) 1998, S. 73-79

Mertsching, B.; Bollmann, M.; Massad, A.; Schmalz, S.: Recognition of Complex Objects with an Active Vision System. In: Proc. of Int. ICSC/IFAC Symposium on Neural Computation (NC'98). Kanada/Schweiz (ICSC Academic Press) 1998, S. 469-475

Bollmann, M.; Hoischen, R.; Jesikiewicz, M.; Justkowski, C.; Mertsching, B.: Playing Domino: A Case Study for an Active Vision System. In: Christensen, H. I. (Hrsg.): Computer Vision Systems. Berlin u. a. (Springer-Verlag) 1999, S. 392-411

Bollmann, M.; Hoischen, R.; Jesikiewicz, M.; Mertsching, B.: Ein Roboter spielt Domino. Erscheint in: KI - Künstliche Intelligenz, Bd. 1, 1999

Mertsching, B.; Bollmann, M.; Hoischen, R.; Schmalz, S.: The Neural Active Vision System NAVIS. Erscheint in: Jähne, B.; Haußecker, H.; Geißler, P. (Hrsg.): Handbook on Computer Vision and Applications. San Diego (Academic Press) 1999

Folgende Studienarbeiten habe ich während meiner Tätigkeit am Fachbereich Informatik der Universität Hamburg direkt betreut:

Thorsten Hempel: Erweiterte Kantendetektion durch biologisch motivierte Auswertung von Farbinformationen für das Neuronale-Active-Vision-System NAVIS. Studienarbeit, Fachbereich Informatik, Universität Hamburg, 1996

Christoph Justkowski: Integration des Merkmals Farbe in die Aufmerksamkeitssteuerung des aktiven Sehsystems NAVIS. Studienarbeit, Fachbereich Informatik, Universität Hamburg, 1998