Aktives Sehen in der Robotik

12. Januar 2002

Inhaltsverzeichnis

1 Einführung

In der Künstlichen Inteligenz (KI) kennt man den Begriff *Sehen* als passives, statisches und unwissendes Sehen. Unter dem Paradigma "Aktives Sehen" wird das Sehen im Verhaltenskontext verstanden. Dazu gehörten die Ausnutzung von Wissen über die Welt und vor allem die gezielte Objektverfolgung. Aktive Sehsysteme (ASS) weisen Merkmale auf, die zum Teil der Biologie entstammen. Beispiele sind hier Binokularität, Foveae, schnelle Augenbewegung und Objetverfolgung.

2 Grundlagen

2.1 Automome Systeme

Ein Automomes System wird auch als Argent bezeichnet, der sein Verhalten in einer dynamischen unvorhersagbaren Welt mit eigener Sensorik selbst steuert.

2.2 Räumliches Verhalten

Mobile Automome Systeme zeigen ein räumliches Verhalten. Man unterscheidet in Basisverhalten und Kompexes Raumverhalten, wobei ersteres die Grundlage für das komplexe Verhalten ist.

Basisverhalten	**Komplexes Raumverhalten**
Orientierungsstabilisierung: Beibehaltung einer räumlichen Orientierung bzgl. der Umgebung	**Navigation:** Bestimmung und Aufrechterhaltung einer Trajektorie zu einem ortsfesten Ziel
Kursstabiliesierung: Kompensation von Verdriftungen (Schräglage)	**Verfolgung:** Bestimmung und Aufrechterhaltung einer Trajektorie zu einem beweglichen Ziel
Kollosionsvermeidung	**Exploration:** Durchsuchen eines Raumgebietes nach releavanten Informationen oder Recourcen

3 Aktive Sehsysteme (ASS)

3.1 klassische KI und Sehen

Das Rechnersehen dient der Wahrnehmung der Umwelt. Eine Möglichkeit besteht darin, aus den aufgenommenen Bildern eine vollständige Darstellung der externen Welt im Rechner zu erzeugen. Nach (*Marr '82*).

- In der KI wird das Verhalten des Systems weitgehend ignoriert.
- Die Berechnungstheorie nach *Marr* behandelt nur passives Sehen.
- Jedes Bild oder jede Szene wird unabhängig betrachtet.

Da die Welt 3-dimensional ist und die Kamera aber 2-dimensionale Bilder aufnimmt entsteht ein Informationsverlust. Der Erkennungsprozess muß die fehlende Dimension wieder ausgleichen, was aber im Allgemeinen nicht eindeutig möglich ist. Dazu kann man sich leicht überlegen, dass

bei der Überdeckung eines Objekt von einem anderen die Information über das sich dahinter befindliche Objekt nicht vorliegen. Das Konzept des aktiven Sehens soll da weiterhelfen. Man hat eingesehen, dass man das biologische Vorbild nicht ausser Acht lassen sollte und sich stattdessen diverse Eigenschaften abzuschauen und für künstliche Sehsysteme zu nutzen.

3.2 Aktives Sehen

Von dem Konzept eine Szene vollständig zu erkennen wird zunehmend Abstand gewonnen und zur gezielten Objektexploration und dem aufgabengerichteten Sehen tendiert. Folgende Punkte können für Aktive Sehsysteme festgehalten werden:

- Kann nur mit Wissen über das Verhalten durchgeführt werden.
- Systeme haben biologische Eigenschaften wie Binokularität und Foveae usw.
- Der visuelle Sensor muss mit hoher Geschwindigkeit präzise ausgelenkt werden können.
- Im Gesamten betrachtet hat man einen geringerer Rechenaufwand verglichen mit passiven Sehsystemen.

Bevor wir uns den einzelnen Punkten im Detail nähern möchte ich zunächst auf das biologische Vorbild eingehen.

3.3 Menschliches Sehen

Wollen wir nun einen Blick auf das menschliche Sehen werfen und einige interessante Aspekte für ASS gewinnen.

3.3.1 biologischer Aufbau

In Abbildung 1 ist der Aufbau eines Wirbeltierauges zu sehen. Für die weiteren Ausführungen sind folgende Details interessant:

Linse:	ist für die Schärfe zuständig
Pupille:	regelt den Lichteinfall (Blende)
Retina:	die Netzhaut welche die Lichtrezeptoren (Zapfen und Stäbchen) enthält
Fovea centralis:	Zentrum stärksten Sehens. Im Folgenden nur Fovea genannt.
Nervus opticus:	ist der Sehnerv, der die visuellen Reizinformationen zum Gehirn leitet.

Damit das Licht an die lichtempfindlichen Rezeptoren gelangt muß es durch mehrere fast transparente Zellschichten in der Retina dringen(Abbildung 2). Zur Illustrierung der Linsentätigkeit sind in der Abbildung 3 die beiden Extremfälle der Linsenkrümmung zu sehen. Die Veränderung der Brechkraft der Linse durch deren Streckung oder Stauchung nennt man Akkomodation

3.3.2 interesante Eigenschaften

Folgende Eigenschaften sind für ASS geeignet:

- Blickkontrolle und schnelle Augennachführung
- sehr schnelle Augenbewegung (Sakkaden)
- Fovea als Zentrum schärften Sehens

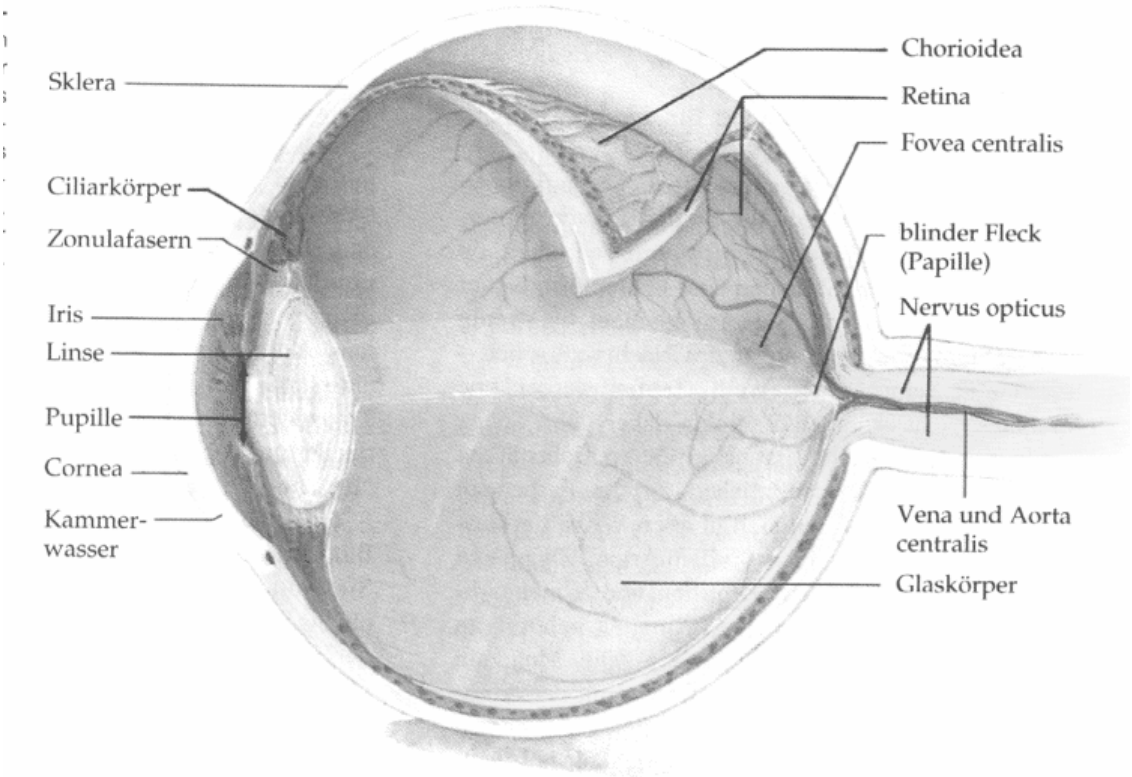

Abbildung 1: Aufbau des Wirbeltierauges

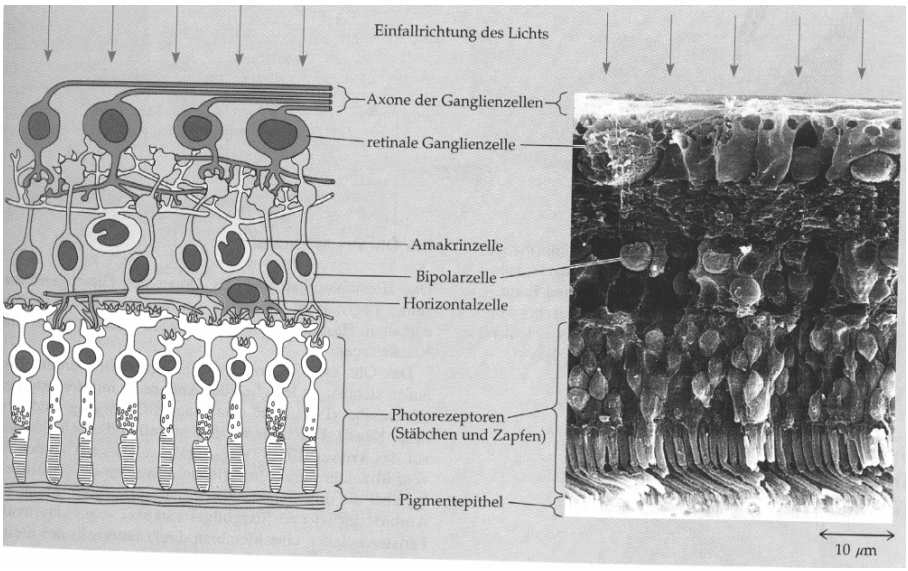

Abbildung 2: Aufbau der Wirbeltierretina

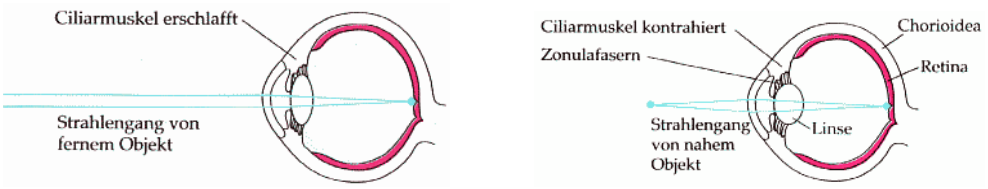

Abbildung 3: Fern- und Nahsehen

4 Konzepte im Detail

Nachdem wir im vorherigen Kapitel den generelle Aufbau und die Eigenschaften von künstlichen ASS gesehen haben wollen wir nun diese Konzepte etwas genauer betrachten.

4.1 Sehen und Erkennen als Vorgang

Das aktiven Sehen wird als Prozess verstanden. Da die Bildverarbeitung bei einem Roboter in Echtzeit erfolgen sollte, kann man nicht ständig das komplette Gesichtsfeld unabhängig von den vorherigen Erkenntnissen durchführen. Vielmehr versucht man das Erkennen von Objekten als Prozess zu betrachten und nicht zu fordern, daß das ASS sofort eine Szene bis in das Detail erkennt. Dazu sammelt man erst bewußt Informationen (Bildfolgen) und analysiert diese sukzessive. Dadurch wird die Genauigkeit der Erkennung im Laufe der Zeit immer genauer. Auf den ersten Blick scheint die Aussage, daß eine Bildfolge im Gegensatz zu einem einzigen Bild die Komplexität des Erkennungsproblemes deutlich veringert Paradox, denn die Datenmange wird enorm vergrößert. Die Ursache liegt an der Bestimmtheit des Problems der Objekterkennung im Falle der stetigen Bildfolge. Im passiven Sehsystem ist dieses Problem hoffnungslos unterbestimmt. Bei ASS kann durch bewusste Kamerabewegung ausreichend Information gesammelt werden. Es müssen dadurch weniger heuritische Annahmen über die Szene gemacht werden. Eine besondere Rolle hat hier die Bildfolge, die durch stetige Sensorbewegung entsteht. Da hier die Abweichungen der Einzelbilder nur verhältnismäßig klein sind, ist das Korrespondenzproblem wesentlich einfacher und robuster. Damit ist die Zuordnung der schon bekannten signifikanten Punkte in der Szene gemeint, die dann auf den einzelnen Bildern wiedergefunden respektive zugeordnet werden müssen.

4.2 Aufmerksamkeitspunkte

Obwohl die Idee des Aktiven Sehens erst seit kurzem beim künstlichen Sehen eingesetzt wird, ist sie schon lange bekannt. Schon im 19. Jahrhundert ist erkannt wurden, das die Bewegung bedeutend für die Wahrnehmung ist. Ebenso alt ist das Konzept der Aufmerksamkeitssteuerung, bei dem es darum geht nur bestimmte Bereiche im Bild zu beachten. Wie Sie sicher auch selber beobachtet haben macht das der Mensch nicht anders. Zum Beispiel bannen bewegte Objekte wie die bekannten Unruhen in einer sonnst stillen Umgebung ihren Blick. Wir müssen also Faktoren finden die dazu führen, daß das ASS sich auf sinnvolle Stellen in der Szene konzentriert. Ein ganz einfacher Faktor ist sicher das bei einer Objektexploration die Aufmerksamkeit auf dieses Objekt gerichtet ist. Außerdem sollten bewegte Objekte Aufmerksamkeit bekommen, da diese zur Veränderung der Umwelt führen oder sogar zu Gefährdung des Systems (Kollisionsvermeidung). Es lassen sich sicher beliegige Aufmerksamkeitsregeln finden, die dann von der momentanen Tätigkeit des Roboters abhängt. In einfachen Systemen könnte man besonders auffällige Stellen wählen, die sich durch einfache Gewichtung von Merkmalen wie Helligkeit oder Kontrast gefunden werden. Die Vorteile die dieses Konzept mit sich bringt sind die Reduzierung des Datenstroms und die Verminderung der Rechenzeit. Der Nachteil der hier in Kauf genommen werden muß ist, dass

unter Umständen wichtige Veränderungen außerhalb das "region of interest" nicht wahrgenommen werden.

4.3 ortsabhängige Auflösung des visuellen Sensors

Um dem eben erwähnten Konzept auch physikalisch gerecht zu werden gibt es beim Mensch keine homogene Auflösung im Auge. Wie schon aus dem biologischen Aufbau ersichtlich wurde hat das Auge eine Fovea, die nahe der optischen Achse liegt und mit 150 000 Rezeptoren(Zapfen) pro Quadratmillimeter die größte Auflösung besitzt. Sie nimmt weniger als 1% der Retina ein und entspricht einem Öffnungswinkel von rund einem Grad. Bussarde und andere Raubvögel haben eine Rezeptordichte von mehr als eine Million pro Quadratmillimeter. Der restliche Bereich der Retina wird als Peripherie bezeichnet. Die Rezeptorendichte ist nach außen hin kontinuerlich abnehmend. Diese Eigenschaft kann auch in ASS sinnvoll eingesetzt werden. Die technische Realisierung kann auf verschiedene Weise realisiert werden. Zum einen kann man spezielle CCDs entwickeln, die eine ortsabhängige Auflösung haben. Der Nachteil ist, daß es sich nicht um Standart Kameramodule handelt und diese somit sehr teuer sind und auch nicht in allen Bauformen verfügbar sind. Ein anderer Ansatz ist rein optisch und benutzt sogenannte Fischaugenobjektive. Diese extremen Weitwinkelobjektive haben die Eigenschaft, das sie nahe der optischen Achse sehr scharf und genau und am Rand verzehrt projezieren. Dadurch wird das Gesichtsfeld (Bereich der mit einem Bild gesehen wird) vergrößert, aber nur im Zentrum ist das Bild hochauflösend. Der Nachteil ist, dass man die Verzerrungen am Rand zurückrechnen muß. Als letztes möchte ich noch die Idee vorstellen einfach zwei Kameras nahe nebeneinander zu montieren wobei die Eine ein Teleobjektiv besitzt und für die genaue Objekterkennung zuständig ist und die Andere nur weniger hoch aufgelöste Bilder erzeugen muß.

Bei der Benutzung dieses Konzepts kann man das Aufmerksamkeitsgerichtete Sehen noch besser nutzten und kann auch ohne großen Rechenaufwand die Randregionen überwachen.

4.4 Fixpunktansatz

Bei bewegtem Beobachter ist es sinnvoll einen raumfesten Punkt eines Objekts als Fixpunkt zu betrachten. Da sich die Bildpunkte der Objekte die vom Betrachter aus vor dem Objekt liegen sich entgegengesetzt zu dem Bildpunkten der Objekte die hinter dem Fixpunkt liegen bewegen, kann man einfach eine Tiefenkarte berechnen. Zusätzlich vereinfacht sich die mathematische Berechnung, falls das Objekt in der Bildmitte liegt, so dass eine orthographische Projektion angewendet werden kann. Die Vorraussetung für diesen Ansatz ist allerdings eine statische Szene.

4.5 Blickkontrolle

Im folgenden Abschnitt werden wir die verschiedenen Bewegungsmechanismen behandeln die unser menschliches Auge besitzt. Außerdem soll klar werden wozu diese dienen und welche dieser Konzepte in künstlichen Systemen genutzt werden können.

4.5.1 Allgemeine Vorgänge

Um optimal die Umgebung erfassen zu können, werden Fixpunkte zur Bildanalyse im peripheren Bereich gesucht und diese dann mittels fovealen Sehens analysiert. Diese Punkte werden nur durch grobe Merkmale wie Farbe und andere von der Orientierung unabhängige Merkmale, sogenannten Invarianten festgelegt. Besteht eine relative Bewegung zwischen Objekt und Auge, so sorgen Kompensationsbewegung für eine stabile Abbildung auf der Netzhaut, so dass keine Bewegungsunschärfe entsteht, die eine Bildverarbeitung unmöglich machen würde.

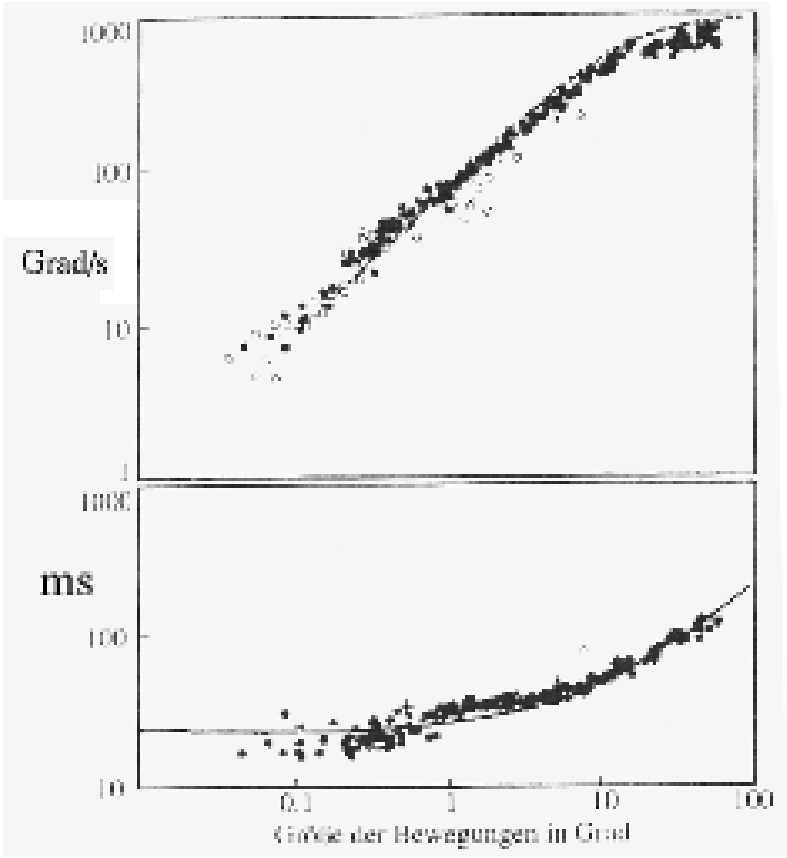

Abbildung 4: Maximalgeschwindigkeit(oben) und Dauer(unten) von horizontalen Sakkaden

4.5.2 Sakkadensteuerung

Eine schnelle Änderung der Blickrichtung wird nach dem französischen Wort für Ruck als Sakkade bezeichnet. Wird während der Augenfolgebewegung die Lageabweichung so groß, dass das betreffende Objekt aus dem Bereich der Fovea herauswandert, so kommt es zu einer Korrektur-Sakkade. Normalerweise laufen Sakkaden unbewusst ab, sie können aber auch bewußt ausgeführt werden. Der Mensch führt rund zwei bis drei Sakkaden pro Sekunde aus. Die Dauer einer solchen dauert zwischen $0,02$ und $0,1$ s und nimmt nahezu proportional zu der Amplitude zu. Siehe Abbildung 4. Die mittlere Geschwindigkeit liegt bei 200 bis 400 Grad/s und ist maximal 600 Grad/s. Vertikale Sakkaden sind etwas langsamer als horizontale. Die Verzögerung, die auf einen visuellen Reiz hin auftritt liegt bei etwa $0,2$ s. Während der Ausführung der Sakkade kann auf Grund der Bewegungsunschärfe keine Bildverarbeitung erfolgen. Nach der Beendigung einer Sakkade können wieder Messungen im Bild vorgenommen werden und es kann eine neue Sakkade gestartet werden. Das System ist eine Art Abtastungssystem mit Totzeit. Es gibt jedoch auch Erkenntnisse, daß auch während der Sakkade Bilder teilweise verarbeitet werden können. Zum Beispiel kann ein erkannter Lagefehler am Anfang der Sakkade noch deren Ziel ändern. Die Augenbewegung ist nicht routinemäßig, sondern von der gestellten Aufgabe abhängig.

4.5.3 Augenfolgebewegung

Augenfolgebewegungen sind eine Antwort auf einen visuellen oder akkustischen Reiz und sind deshalb auch nicht willentlich beeinflußbar. Die Hauptaufgabe der Augenfolgebewegung ist es

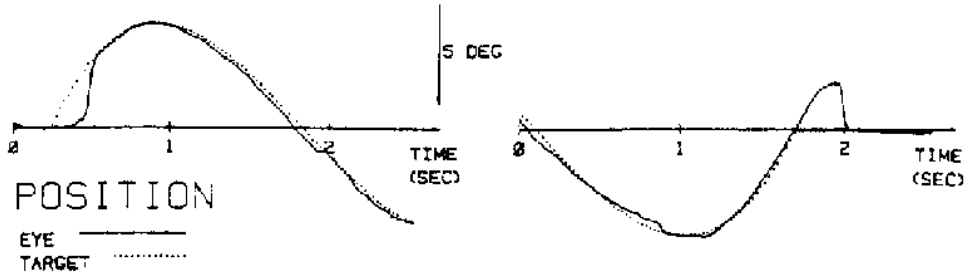

Abbildung 5: Typische Anfand und Ende des menschlichen Verfolgens

das Bild des Objektes ruhig und damit scharf auf der Netzhaut abzubilden. Außerdem sollte der fokusierte Bereich in der Nähe der Foveae gehalten werden. (im Bereich von 2 Grad).

Die Augefolgebewegung ist vorallem vom Geschwindigkeitsfehler abhängig. Zusätzlich nimmt auch der Lagefehler Einfluß auf die Bewegung. Kleine Lagefehler werden durch kleine Veränderungen der Augengeschwindigkeit beseitigt. Grosse Lagefehler müssen statt dessen mit Sakkaden reduziert werden, da eine starke Veränderung der Augengeschwindigkeit bezüglich der Objektgeschwindigkeit zu einer zu großen Bewegungsunschärfe führen würde.

Der den Augefolgebewegungen zugrundeliegende Regelkreis kann als kontinuierlich angesehen werden. Dieser Vorgang ist reflexartiger als die Sakkadensteuerung, da dieser eine geringere Todzeit aufweist (0.31 s). Die Folgebewegungen sind auf Geschwindigkeiten von 60 bis 90 Grad/s begrenzt. Um das Verhalten zu untersuchen wurde mittels periodischer Bewegung von Objekten die Augenbewegung gemessen. Bei einer Sinusschwingung von 1 Hz und einer Phase maximal 10 Grad konnte das Auge dieser Bewegung direkt folgen. Siehe Abb.(5) Allgemein kann man festhalten, daß diese Bewegungsformen verfolgt werden können, die stückchenweise durch konstante Beschleunigungen angenähert werden können.

4.5.4 Kopfbewegungskompensation (Vestibulär-Okularer-Reflex)

Die Bewegung des Kopfes werden über Gleichgewichtsorgane erfaßt. Morphologisch unterscheidet man zwei verschiedene Typen. Die Bodengangsorgane werden durch Winkelbeschleunigungen angeregt und die Maculaorgane werden durch Translationbeschleunigungen angeregt. Diese Messungen werden zur Stabilisierung der Blickrichtung bei Eigenbewegungen verwendet. Im Gegensatz zu den Augefolgebewegungen, die auf der Bildmessung aufbauen, geschieht hier die Kompensationbewegung ohne Zeitverzug. Außerdem können auch relativ schnelle Kopfbewegungen ausgeglichen werden, die das Verfolgungssystem nicht mehr schafft.

Da die Kompensation ohne eine aktuelle Rückkopplung arbeitet, handelt es sich um eine reine Steuerung. Die Parameter wie zum Beispiel die Verstärkung werden erlernt, indem man die Bewegung statischer Objekte bei der Bewegung des Kopfes analysiert. Da Kopfdrehungen durch Augendrehungen direkt kompensiert werden, wird diese Reaktion bei der so genannten aktiven Augen-Kopf Koordination abgeschalten. Es handelt sich um eine Kopfnachführung, wenn Sakkaden mit mehr als 10 bis 15 Grad ausgeführt werden.

4.5.5 Vergenzbewegung

Hierbei handelt es sich um die abgestimmte Bewegung der beiden Augen.

4.5.6 Nystagmen

Nystagmen sind periodische reflexartige Augenbewegung, die abwechselnd aus schnellen Phasen, den Sakkaden und langsamen Phasen, dem Verfolgen bestehen. Man unterscheidet in vestibulärokularen Nystagmus (VON) und optokinetischen Nystagmus (OKN)

Der VON tritt sowohl bei aktiver als auch bei passiver Kopfbewegung auf. Hierbei werden in einer statischen Szene die Kopfbewegungen durch Augenbewegungen kompensiert, indem ein räumlich stationärer Punkt fixiert wird.

Der OKN tritt auf, wenn sich alle Objekte bezüglich des Beobachters bewegen. Zum Beispiel dem Blick aus dem Zug. Die glatten Augenbewegungen werden durch die kurzen in die entgegengesetzte Richtung laufenden Sakkaden unterbrochen.

4.5.7 Mikrobewegung

Die Mikrobewegung wird auch als physiologischer Nystagmus bezeichnet. Die Augen führen auch im scheinbar stillen Zustand ein Zittern(Augentremor) durch, deren Amplitude nur wenige Winkelminuten betragen. Erstaunlicher Weise ergeben sich Frequenzen von 30 bis 60 Hz. Eine technische Realisierung ist aber nicht denkbar und auch aus heutiger Sicht nicht sinnvoll.

4.5.8 Modelle für die Augenbewegung

Da die Differenz der Objektbewegung und der Augenbewegung gemessen wird, besitzt dieses System eine Rückführung, so daß die Regelabweichung errechnet wird. Auf Grund der Zeitverzögerung durch die Bildverarbeitung und der Totzeit des Steuersystems ist die Reglung zu einem stabilen System schwierig. Ein einfacher Regler würde mit diesen Vorraussetzungen nur ein instabiles und langesames Verhalten realisieren. Rein regeltechnisch ist das folgendermaßen zu verstehen: Die externe negative Rückkopplung wird duch eine intern positive Rückkopplung kompensiert. Dadurch ist das System für beliebige Totzeiten stabil. Eine solche Anordnung wird auch als Efferenzkopie der motorischen Kommandos definiert.

5 Objekterkennung und Szenenanalyse

Ich möchte in dieser Arbeit nur einen Überblick über die Methoden und Vorgänge der Objekterkennung geben.

5.1 Architekturen

Zuerst stellt sich die Frage der Architektur, die benutzt werden soll um das Weltmodell zu modellieren. In Abbildung **??** und **??** sind drei verschiedene Architekturen miteinander verglichen. Besonders interessant sind hybride Modelle, die sich auf mehrere Architekturen beziehen.

5.2 Bildverarbeitungsmethoden

5.2.1 datengetriebenes vs. modellgetriebenes Vorgehen

Unabhangig von der Architektur der letztlichen Weltmodellierung müssen immer Bildbearbeitungsschritte durchlaufen werden. Die Bildverarbeitungsmethoden zur Umgebungserkennung bestehen aus der Bildvorverarbeitung, mit der die Merkmale im Bild extrahiert werden, und der

Architektur	Dynamik	Genese
Symbolverarbeitungssystem (Computer-Metapher)	Steuerung	Fremdorganisation
konnektionistisches System (Gehirn-Metapher)	Regelung	Selbstorganisation
interaktionistisches System (Ökosystem-Metapher)	Handlung	kooperatives Problemlösen

Abbildung 6: Architekturen im Vergleich

Architektur	Computermodell	konnektionistisches Modell	Interaktionistisches Modell
Komponenten	Mentale Symbole	Units, Neurone	Dualität von Information und Prozessor
Struktur	Systematizität	Netzwerk	Intentionalität
Umwelt und Funktion	auf Symbole beschränkt Frame-Problem	Eingeschränkte Umwelt (Vorgabe des Programmierers) Robustheit	Situiertheit

Abbildung 7: Architekturen im Vergleich

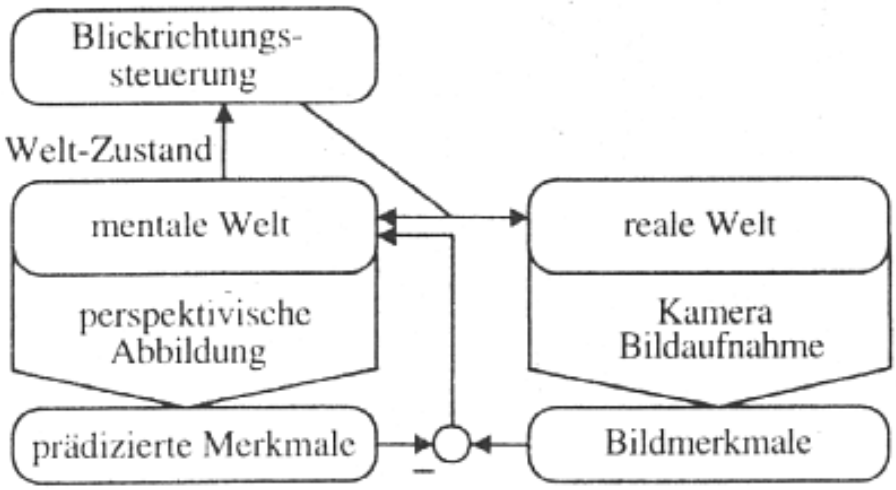

Abbildung 8: Prinzip des dynamischen Sehens

Bildinterpretation oder Objekterkennung. Hierbei unterscheidet man die datengetriebene und die modellgetriebene Vorgehensweise. Beim modellgetriebenen Vorgehen wird ausgehend von speziellen Objektmodellen im Bild nach einfachen Merkmalen gesucht. Für die datengetriebene Erkennung werden zuerst aus dem Bild allgemeine Regionen-, Form- oder Abstanskarten erzeugt. Diese Umrechnung basiert meist auf räumlichen und zeitlichen Ableitungen der Bildhelligkeit (bzw. auf dem optischen Fluß), weshalb sie besonders für gut strukturierte Bilder geeignet ist. Anschließend müssen die höherwertigen Merkmale mit denen einer Objektdatenbank verglichen werden, um zu einer Objekterkennung zu kommen. Einfache reflexartige Operationen lassen sich jedoch auch direkt aus den Merkmalen ableiten, wie die Schärfenregelung und Bildstabilisierung.

Um in der realen Welt rechtzeitig und richtig reagieren zu können, muß die Bildverarbeitung mit den Vorgängen in der Welt mithalten und sich an die Umgebungsbedingungen anpassen können. Beim Aktiven Sehen werden die Ergebnisse der Bildverarbeitung im besonderen zur Regelung der Blickrichtung benötigt, so daß die Bildverarbeitung echtzeitfähig sein muß. Daraus ergibt sich die Forderung nach einer hohen Auswerterate und einer kleinen Totzeit.

Damit die Bildverarbeitung schnell ist, können nur wenige Bildbereiche mit besonderen Merkmalen untersucht werden, da diese Operationen sehr viel Rechenleistung beanspruchen. Eine merkmalsbasierte Bildauswertung entspricht den Erfordernissen des Aktiven Sehens, da im Gegensatz zum datengetriebenen Ansatz die Bildvorverarbeitung auf dem ganzen Bild entfällt.

Eine weitere wichtige Bedingung für die Bildverarbeitung ist, daß sie auch bei bewegter Kamera brauchbare Ergebnisse liefert. Hierbei ist von Vorteil, daß die Welt im wesentlichen statisch ist und sich in ihr nur wenige Objekte bewegen. Bei der Bildinterpretation ist also vor allem die Eigenbewegung der Kamera zu berücksichtigen. Da die Objekte entsprechend den physikalischen Gesetzen durch Kräfte beschleunigt werden wird die Beschreibung der Objektbewegung im inertialen Raum verhältnismässig einfach.

Für diese Anforderungen des Aktiven Sehens ist das Konzept des dynamischen Sehens besonders gut geeignet, wie im folgenden gezeigt wird.

5.2.2 Dynamisches Sehen

Das dynamische Sehen nutzt eine merkmalsbasierte Bilauswertung und rekursive Schätzmethode zur Bestimmung des Zustands von Objekten in der Welt.

Die Grundlage dieses Ansatzes ist eine vollständige interne Darstellung der Welt in Raum und Zeit, was als mentale Welt bezeichnet wird. Sie setzt sich aus räumlich-zeitlichen (4D) dynamischen Modellen von Objekten aus der realen Welt und einer Beschreibung des aktuellen Zustands durch Zustandsgrößen zusammen. Die Kamera tastet die Szene in äquidistanten Zeitschritten ab. Die Lage der Kanten im Bild (2D) wird gemessen. Diese Messungen werden mit dem prädizierten Merkmalslagen verglichen und die Differenz wird zur Korrektur des geschätzten Zustands der Objekte benutzt. Unter Verwendung der vollständigen 4D-Zustandsbeschreibung und der dynamischen Modelle werden die Zustansgrößen im nächsten Abtastzyklus vorhergesagt. Die prädizierten Merkmale erhält man mit Hilfe der perspektivischen Abbildung unter Beachtung der Aspektbedingungen. Der geschätzte Weltzustand kann zur Bestimmung von Steuergrößen benutzt werden, um die reale Welt auf ein gewünschtes Ziel hin zu beeinflussen.

Ein effektiver Algorithmus zur Schätzung der Zustandsgrößen mit kleinstem Fehlerquadrat ist das Kalman-Filter und davon abgeleitete Algorithmen.

Dieser Ansatz hat folgende Vorteile:

- Es wird nur die perspektivische Abbildung (3D-Welt zu 2D-Bild) und nicht die Inverse (2D zu 3D) benutzt, da die perspektivische Abbildung im Gegensatz zu der Inversen immer eindeutig ist.
- Der physikalische Zustand der Objekte der Szene wird geschätzt. Die symbolische Beschreibung kann direkt zur Roboter- und Blickrichtungssteuerung benutzt werden.
- Es wird keine Speicherung alter Bilder benötigt, da die gesamte Information in der Zustandsbeschreibung enthalten ist
- Aufgrund der vorhergesagten Merkmalslagen und Orientierungen müssen die zu messenden Merkmale nur in einem kleinen Bildfenster ('region of interest') mit einer bekannten Maskenorientierung gesucht werden. Die Auswertung kann auf diese kleinen Gebiete begrenzt werden und Echtzeitanwendungen werden möglich.

5.3 Erkennungsprozess

Bei der Betrachtung des Erkennungsprozesses als hierarischer wissensbasierter und nichtlinearen Prozess haben sich viele Vorteile ergeben. (siehe Abbildung 9) Zum Einen kann eine beliebige Schicht mit der ihr untergeorneten Schicht kommunizieren um Korrekturen zu ermöglichen. Und zum Anderen können die Informationen aus den einzelnen Schichten schon für die Kamerabewegung genutzt werden. Zum Beispiel kann die Object-recognition-Schicht der Low-Lewel-Vision-Schicht eine andere Zelegung der Kanten vorschlagen, da sie mehr Wissen hat oder es können die Informationen für die Sensorfolgebewegung gleich aus der Object-recognition-Schicht erwerben. Welcher Ansatz am besten für das Verstehen geeignet ist kann und soll hier nicht diskutiert werden. Wenn man von Verstehen eines künstlichen Systems redet, dann ist man bis heute nur so weit, daß man sagt: "der Programmierer oder Benutzer borgt dem System die Bedeutung bzw. Interpretation". Ein interessanter Ansatz sagt, daß ohne Subjekt mit Bewusstsein kein Verstehen möglich ist.

6 Eigenbewegungsschätzung

Hier soll ein Teilproblem in ASS kurz erläutert werden, der elementar für die Funktionsweise der Kamerasteuerung ist, der Berechnung der Eigenbewegung. Im Klassischen Ansatz ist die Kamera fest auf dem Roboter fixiert. Daraus folgt: Die Eigenbewegung kann einfach aus den singulären Punkten im Flußdiagramm berechnet werden. Bei ASS haben wir folgende Probleme: Die Kamera

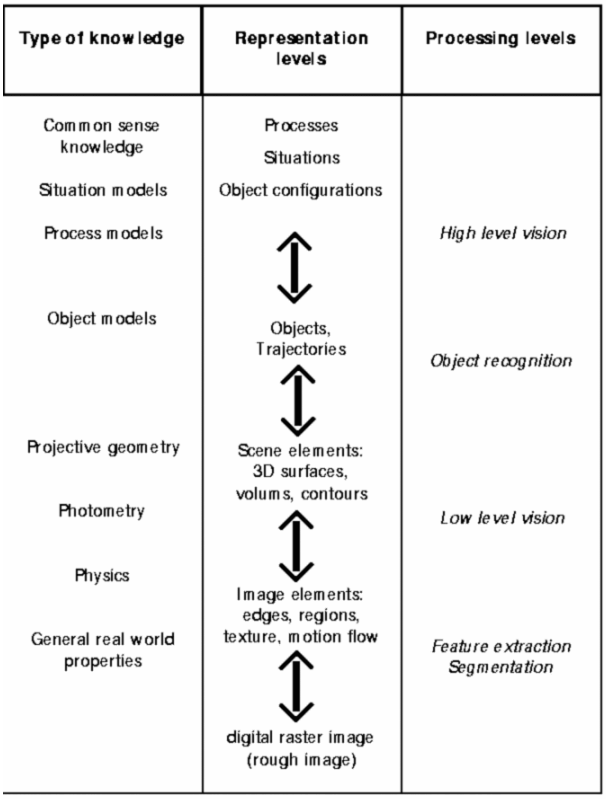

Image understanding as hierarchical, knowledge-based process.

Abbildung 9:

hat zwei oder mehr zusätzliche Achsen: Daraus folgt, dass die singulären Punkte keine direkte Aussage über die Eigenbewegung machen Ein Beispiel ist die Fixierung eines Punktes eines bewegenden Objektes während der Eigenwegung. Welche Aspekte sind zu beachten?

- Kamerastellung
- Kamerabewegung
- Tiefenstruktur
- Eigenbewegung

Ideal wäre ein Neuronales Netz (NN) für die Modellierung der Eigenbewegung aus dem visuellen Eingang.

Über mathematische Rechnung lasst sich die Translationsvektor $\vec{T}$ ohne Wissen über die Rotation und Tiefenverteilung abschätzen. Es handelt sich um eine Kostenfunktion in Abhängigkeit von $\vec{T}$, die man mittels eines NN minimieren kann.

6.1 Multimodale sensorische Integration

Man sollte mehrere unabhängige Quellen zur Berechnung einbeziehen. Beim Mensch bekommt man aus dem Gleichgewichtszentrum im Kleinhirn eine Translationsänderung, die dann integriert die Geschwindigkeit ergibt. Bei ASS kann man das durch Kreisel erreichen. Dazu kommt noch das Muskelfeedback welches in ASS durch Motorfeedback ersetzt wird. Durch das rationale Denken und die Erfahrungen mit den physikalische Grundgesetzen können wir Abschätzungen über Form und mögliches Verhalten von Objekten machen. Dieser als ethologisch bezeichnete Einfluß muss in künstlichen Systemen aus der Weltmodellierung kommen.

Die Informationsgewinnung über die Rotation kann zum Einen aus der Disparität des Flußfeldes erreicht werden und zum Anderen aus dem Motorfeedback. Da die Rotation aber unabhängig von der Entfernung eines Objektes ist reicht der Rotationsparameter aus um jeden Vektor im Flußdiagramm transformieren zu können. Wegen der Ungenauigkeit des Motorfeedbacks muß ein NN nachgeschalten werden, der den Fehler minimiert. Schlussfolgernd können wir festhalten, daß durch die Kombination von verschiedenen Informationsquellen die Eigenbewegungsschätzung robuster und leistungsfähiger wird.

7 Shape from X nach Aloimonos

Die Bezeichnung "Shape from X" steht für eine Reihe von Verfahren, die die Form eines Objektes aus dessen zwei-dimensionaler Projektion berechnet. Das "X" steht eben gerade für die unterschiedlichen Arten der möglichen Eigenschaften, die ausgenutzt werden um die Form zu gewinnen. Die Abbildung 10 zeigt zum einen die Berechnung der Form aus der Schattierung durch Lichteinfall und im zweiten die Berechnung der Form aus der Textur.

Für die Berechnung der Form aus der Schattierung wurde das folgende mathematische Modell benutzt. Die Intensität eines Bildpunktes ist nach dem Labertianischen Relflextionsmodell:

$$I(x,y) = p\frac{p \cdot p_s + q \cdot q_s + 1}{\sqrt{1+p^2+q^2}\sqrt{1+p_s^2+q_s^2}} \equiv R(p,q)$$

Um eine eindeutige Lösung zu erreichen wird die folgende Gleichung minimiert.

$$\iint_{image} \left\{(I-R)^2 + \lambda(p_x^2 + p_y^2 + q_x^2 + q_y^2)\right\} \mathrm{d}x\mathrm{d}y$$

Man erhält eine Fläche, die so glatt wie möglich unter der gegeben Bedingung I=R ist. Der Parameter λ gewichtet die relative Bedeutung der zwei Terme der Funktion.

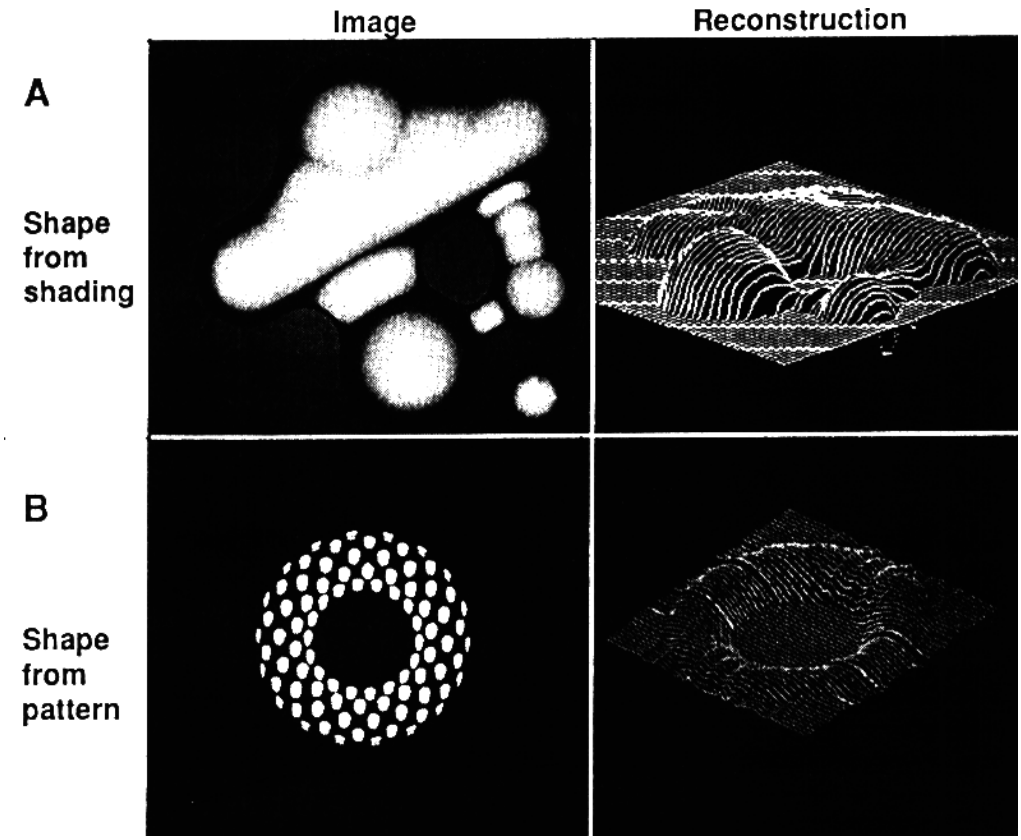

Abbildung 10: Shape from X

8 Literaturangabe

Buecher

Nr.	Titel	Autor	Verlag	Jahr
1	Integration of Visual Modules	Aloimonos	Academic Press	1989
2	Active Perception	Aloimonos	LEA	1993
3	Lecture Notes in artificial Intelligence	Roberto	Springer	1991
4	Fortschrittsbericht	Schieber	VDI-Verlag	Reihe 8, Nr 415, 1995
5	Biologie	Campbell	Spektrum	1998

Internet

http://lmb.informatik.uni-freiburg.de/lectures/praktika/BVPraktikum-II/aktivesSehen.html
http://www.isi2000.de/ahrns.htm
http://dol.uni-leipzig.de/pub/1995-4
http://ima-www.informatik.uni-hamburg.de/Docs/Workshop/workshopDt.html