

Visuell basierte Robotik

Stand und Perspektiven einer Technologie

Prof. Dr. G. Sommer

Dr. J. Pauli

Dr. K. Daniilidis

Dipl.-Inform. J. Bruske

Christian-Albrechts-Universität Kiel

im Auftrag der

Technologiestiftung Schleswig-Holstein

März 1996

Inhaltsverzeichnis

1	Einleitung	3
2	Gegenwärtiger Stand	12
2.1	Anwendungsgebiete	12
2.1.1	Gebote für den Einsatz von Robotern	12
2.1.2	Visuell basiertes Manipulieren	15
2.1.3	Visuell basiertes Fortbewegen	17
2.1.4	Aktive visuelle Wahrnehmung	20
2.2	Funktionsprinzipien	22
2.2.1	Bestandteile von Robotersystemen	22
2.2.2	Kategorien der Roboterautonomie	29
2.2.3	Funktionale Modularisierung von Robotern	31
2.2.4	Roboterrelevantes Maschinensehen	33
2.2.5	Roboterrelevantes Maschinenlernen	42
2.3	Bewertung des Entwicklungsstandes	50
2.3.1	Entwicklungsstand der visuell basierten Robotik	50
2.3.2	Entwicklungsstand des Robotersehens	51
2.3.3	Entwicklungsstand des Roboterlernens	52
2.4	Kommerzielle und wissenschaftliche Produkte	55
3	Sehende Robotersysteme als Perspektive	57
3.1	Einige Szenarien	59
3.1.1	Automatisierung von Teleoperationen	59
3.1.2	Automatisierte Unterstützung von Behinderten	60
3.1.3	Automatisiertes Befördern von Gegenständen	62
3.2	Fusion von Robotik, Maschinensehen und Neuroinformatik	63
3.2.1	Konzeption der verhaltensbasierten Modularisierung	63
3.2.2	Rechentechnische Voraussetzungen	68

3.3	Visuell befähigte Robotersysteme	71
3.3.1	Perspektive für das Robotersehen	71
3.3.2	Binokulare Kameraplattformen als Perspektive	74
3.3.3	Mobile sehende Roboter als Perspektive	76
3.3.4	Manipulierende sehende Roboter als Perspektive	77
3.3.5	Autonomes Sehen, Fortbewegen und Manipulieren	78
3.4	Visuelle Informations–Management–Systeme	79
3.4.1	Visuelle Informationssysteme und Bild–Datenbanken	79
3.4.2	Videobasierte Informationssysteme	80
4	Realisierung sehender Robotersysteme	82
4.1	Realisierungschancen und Erfordernisse	82
4.2	Situation und Chancen in Schleswig–Holstein	87
	Index	91
	Literatur	93

1 Einleitung

Wahrscheinlich träumt der Mensch seit jeher davon, sich ein derartig komplexes Werkzeug zu schaffen, daß es an seiner Stelle unangenehme oder gefährliche Arbeiten oder Tätigkeiten verrichten könne. Eine jüdische Sage berichtet, daß der in Prag lebende Rabbi Löw im Jahre 1580 eine *Golem* genannte Tonfigur schuf. Es war der tschechische Schriftsteller Karel Capek, der in der ersten Hälfte dieses Jahrhunderts mit dem Theaterstück *Rossum's Universal Robots* diesem Menschheitstraum seinen heute gebräuchlichen Namen gab: *Roboter*. Dieser Begriff leitet sich aus dem slawischen Verb *robotá* ab, das für frönen, schufteten oder schwer arbeiten steht. Motivierend für Capek war die Überwindung individueller Belastungen, die der Mensch, indem er die moderne Industriegesellschaft dieses Jahrhunderts schuf, zu tragen hat.

Die *Robotik* ist in ihrer Entwicklung einerseits dadurch geprägt, daß sie, stets im Spannungsfeld von *Science Fiction* und Stand der Erkenntnis beziehungsweise Stand des Machbaren stehend, in diesem Jahrhundert verschiedene Entwicklungsstufen der Realisierung erreichte. Andererseits hatten und haben diese Entwicklungsstufen stets auch enorme sozialgesellschaftliche und gesellschaftspolitische Rückwirkungen [Friebe, 1996].

Unser gegenwärtiges Wissen über diese Zusammenhänge erfordert deshalb nicht nur, diese Konsequenzen zu konstatieren, sondern sie aktiv zur Steuerung der gesellschaftlichen Entwicklung einzusetzen. Das heißt, wir sollten erkennen, was machbar ist, aber nicht allein danach streben, das heute Machbare zu tun. Vielmehr sollte unsere Maxime dadurch geprägt sein, unsere rasant gewachsenen wissenschaftlichen und technologischen Kenntnisse auf dem Gebiet Robotik zur Harmonisierung inhärenter Konflikte der modernen Industriegesellschaft einzusetzen.

Robotik ist stets mit *Automatisierung* verbunden.

Automatisierung heißt, einen Vorgang mit technischen Mitteln so einzurichten, daß der Mensch weder ständig noch in einem erzwungenen Rhythmus für den Ablauf des Vorgangs tätig zu werden braucht.

Ein zeitgemäßes Verständnis dieses Begriffs bedeutet nicht das völlige Herauslösen des Menschen als Individuum aus dem zu realisierenden Vorgang, sondern dem Menschen die Chance zur Verwirklichung seiner enormen intellektuellen aber deutlich begrenzten physischen Fähigkeiten zu geben. Die hiermit verbundenen gesellschaftlichen Konsequenzen bedeuten sowohl das Ausnutzen des technologischen Innovationsschubes als auch weitreichende Änderungen der Berufs- und Tätigkeitsprofile. Die eigentliche Problematik liegt darin begründet, daß der wissenschaftlich-technische Fortschritt wesentlich rasanter wächst, als den mit seiner Nutzbarmachung bewirkten gesellschaftlichen Konsequenzen entsprochen werden kann.

Die Robotik befindet sich in einer Phase des Umbruchs beziehungsweise der konzeptio-

nellen Erweiterung, die frühzeitig von der Industrie und der Gesellschaft erkannt werden sollte, um sowohl den technischen Fortschritt maximal nutzen als auch die gesellschaftlichen Folgen aktiv kontrollieren zu können. Diese sich gegenwärtig noch vorwiegend in den Forschungslabors abspielende Umwälzung ist mit wenigen Schlagworten charakterisiert:

- Roboter werden **intelligent**.
- Roboter können **lernen**.
- Roboter werden **mobil**.
- Roboter können **sehen**.

Die aktuelle Robotikforschung hat die Entwicklung von Robotersystemen zum Ziel, die nicht nur in einer modellierbaren sondern auch in einer *unstrukturierten Umgebung* selbständig Aufgaben erledigen können. Das selbständige Agieren in einer unstrukturierten Umgebung schließt eine Roboterprogrammierung im obigen Sinne aus. Stattdessen muß das Robotersystem aus der Wechselwirkung mit der Umwelt fortwährend hinzulernen und dadurch neue Fertigkeiten erwerben beziehungsweise vorhandene Fähigkeiten verbessern. Die hierzu erforderliche Wahrnehmung der Umwelt soll selektiv auf die Aufgabe bezogen sein. Angelehnt an die Bedeutung des menschlichen Sehsystems spielt die *visuelle Wahrnehmung* auch für Robotersysteme eine besonders wichtige Rolle.

Der in der Robotik erlebte Qualitätssprung ist im wesentlichen auf die Fortschritte auf dem Gebiet *Künstliche Intelligenz* im weitesten Sinne zurückzuführen [Dillmann und Bocionek, 1994]. In den sechziger Jahren dieses Jahrhunderts lautete eine Definition für Roboter beispielsweise noch folgendermaßen:

Ein Roboter ist ein programmierbarer, multifunktionaler Manipulator, der Material, Teile, Werkzeuge oder spezielle Geräte bewegen kann, und zwar mittels veränderbar zu programmierender Bewegungen, zur Erfüllung verschiedener Aufgaben.

Aus heutiger Sicht vermißt man in derartigen Definitionen insbesondere jegliche Eigenschaft der *Intelligenz* oder *Kompetenz*. Grundlage für kompetentes Roboter-Verhalten ist die Wahrnehmung der Umwelt und die rechentechnische Umsetzbarkeit der Umweltdaten in Aktionen unter Realzeitbedingungen. Diese Forderungen können heutzutage in hohem Maße erfüllt werden. Die Voraussetzungen wurden auf technologischem Gebiet geschaffen, insbesondere mit der in den letzten dreißig Jahren rasant fortgeschrittenen Miniaturisierung der Halbleitertechnik und dem damit verbundenen Leistungszuwachs der Sensor- und Rechentechnik. Vor allem aber entwickelten sich, relativ wenig von der breiten Öffentlichkeit beachtet, neue Wissenschaftsdisziplinen, die von unmittelbarem Einfluß auf unser heutiges Verständnis der Robotik sind. Einige hiervon, die zum Teil als Unterdisziplinen der Künstlichen Intelligenz zu verstehen sind, stellt die nachfolgende Tabelle in einer Übersicht dar.

Fachgebiet	Gegenstand/Inhalt
Künstliche Intelligenz Artificial Intelligence	Verstehen intelligenter Leistungen biologischer Systeme und Modellierung intelligenter Leistungen technischer Systeme.
Maschinensehen Computer Vision	Verstehen der visuellen Wahrnehmung der Welt durch biologische Systeme und Modellierung visuell wahrnehmender technischer Systeme.
Neuroinformatik	Modellierung informationsverarbeitender Prozesse nach dem Paradigma künstlicher neuronaler Netze.
Computational Neuroscience	Modellierung informationsverarbeitender Prozesse biologischer (neuronaler) Systeme, insbesondere die Wechselwirkung von Struktur und Funktion unter dem Paradigma, daß berechenbare Aufgaben zu lösen sind.
Mustererkennung Pattern Recognition	Modellierung der Wahrnehmung oder des Verstehens (mehrdimensionaler) sensorischer Daten nach strukturbildenden Gesichtspunkten.
Maschinenlernen	Modellierung des Lernens von Aufgaben, insbesondere der Mustererkennung und der Prozeßkontrolle (z.B. Kontrolle von Bewegungen), z.B. auf der Grundlage der Neuroinformatik, evolutionärer Strategien oder genetischer Algorithmen.
Evolutionäre Optimierung	Modellierung von Optimierungsprozessen und des Lernens auf der Grundlage von Evolutionsstrategien beziehungsweise genetischen Algorithmen.
Künstliches Leben Artificial Life	Modellierung technischer oder simulierter Systeme, die auf der Grundlage erworbener Kompetenz in der Lage sind, Verhalten zu zeigen, wie es für biologische Systeme charakteristisch ist.
Mechatronik	Entwicklung komplexer heterogener technischer Systeme, unter den Gesichtspunkten der funktionalen Integration, Reduktion der Zahl der Komponenten und des Volumens.

Tabelle 1.1: Wissenschaftsdisziplinen mit Einfluß auf Robotik.

Diese Studie stellt sich zum Ziel, den gegenwärtig international in den Labors vorhandenen Wissensstand auf dem Gebiet der visuell basierten Robotik darzustellen. Unter dem Begriff *visuell basierte Robotik* werden Robotersysteme verstanden, die entweder alle oder einige der oben erwähnten Fähigkeiten (intelligent, lernend, mobil, sehend) haben. Insbesondere nutzen sie stets visuelle Informationen zur Realisierung ihrer Aktionen. Mit der Studie wird frühzeitig darauf aufmerksam gemacht, daß in naher Zukunft sehende Robotersysteme realisierbar sind, die im Vergleich zu den derzeit eingesetzten Systemen einen wesentlich breiteren Anwendungsbereich haben.

Diese neue Kategorie von Robotern hat zum Teil nur noch sehr wenig zu tun mit den heute im Einsatz befindlichen. Visuell basierte Roboter realisieren in sehr vielfältiger Weise einen *Wahrnehmungs–Aktions–Zyklus* und erreichen hierbei einen hohen Grad an *Autonomie*. Dies können beispielsweise Fahrzeuge oder intelligente Kameras, Baumaschinen oder Dienstleistungsmaschinen sein. Zeitgemäß könnte die Definition für einen Roboter heute folgendermaßen lauten:

Ein Roboter ist ein mit Kompetenzen der Wahrnehmung und Handlung ausgestattetes technisches System, das sich als Ganzes oder in Teilen aufgabenorientiert bewegen und seine Kompetenz durch Lernen verbessern beziehungsweise adaptieren kann, wobei hierfür die für die Ausführung seiner Aufgabe wesentlichen Aspekte der Umwelt herangezogen werden.

Mit den zentralen Bestandteilen — Wahrnehmung, Lernen und Handlung — verwundert es nicht, daß in der visuell basierten Robotik bisher getrennt verfolgte Fachgebiete wie Maschinensehen, Neuroinformatik und Robotik verschmelzen (Abbildung 1).

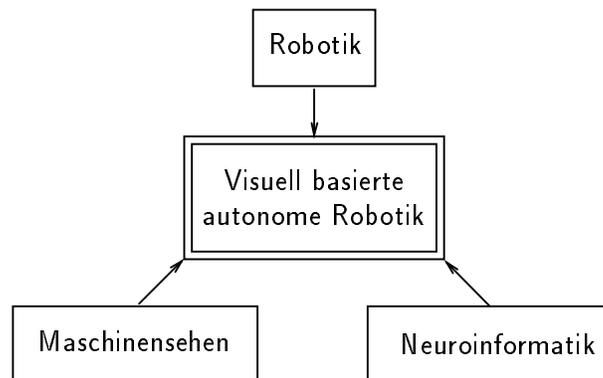
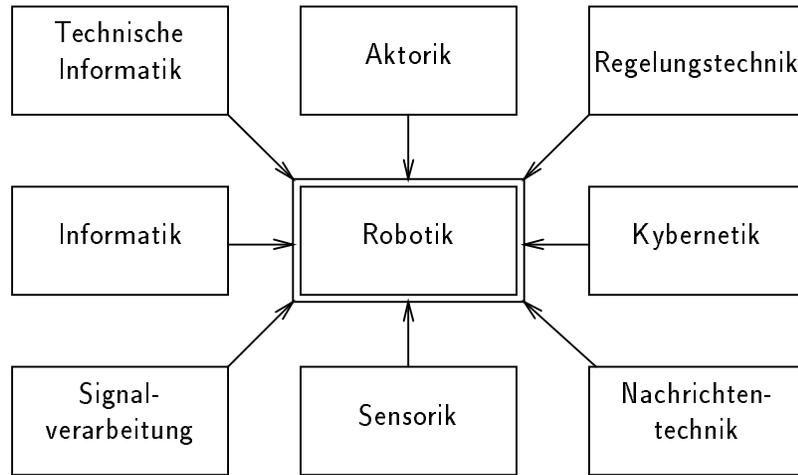
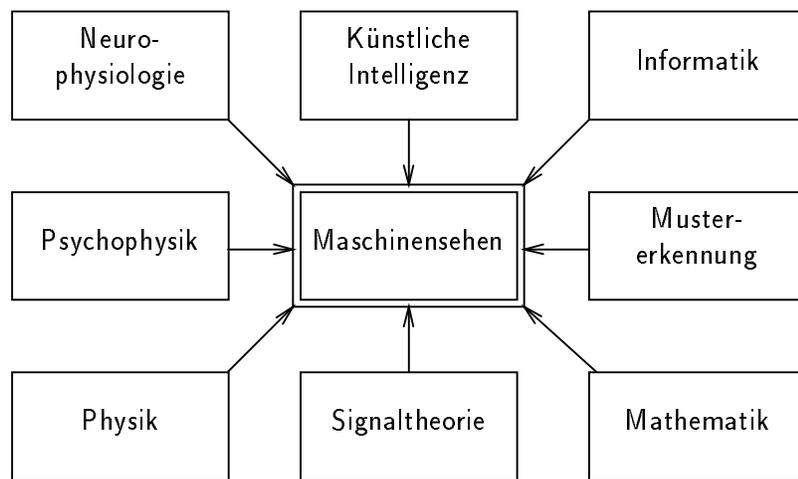


Abbildung 1: Die Hauptquellen visuell basierter Robotik.

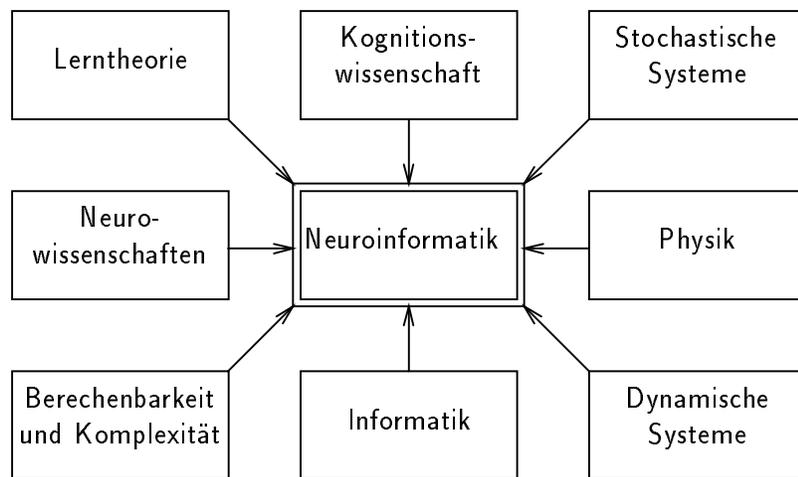
Jede dieser Komponenten ist selbst nur durch stark verzweigte interdisziplinäre Forschung realisierbar. Um hiervon einen Eindruck zu vermitteln, zeigt Abbildung 2 die Einflußsphären der Robotik (a), des Maschinensehens (b) und der Neuroinformatik (c). Dabei sind verwandte Fachgebiete in den jeweiligen Graphiken benachbart dargestellt.



(a)



(b)



(c)

Abbildung 2: Interdisziplinarität der Wurzeln visuell basierter Robotik; (a) Robotik, (b) Maschinensehen, (c) Neuroinformatik.

Gleichzeitig ist gegenwärtig unter dem Begriff *Mechatronik* das Entstehen einer neuen technischen Disziplin beobachtbar [Lückel, 1995]. Mechatronik hat die Fusion elektronischer und mechanischer Komponenten eines technischen Systems zum Inhalt, wobei funktionale Integration zu räumlicher Integration und Miniaturisierung führt, und der Systementwurf so zu realisieren ist, daß die innere Heterogenität äußerlich nicht wahrnehmbar ist (Abbildung 4a).

Die Stoßrichtungen von Mechatronik und visuell basierter Robotik sind gleich:

Die technische Realisierung des Wahrnehmungs–Aktions–Zyklus.

Lebewesen oder, wie die Techniker sagen, biologische Systeme begründen ihre Existenz darauf, diesen Wahrnehmungs–Aktions–Zyklus (in sehr unterschiedlicher Weise) realisieren zu können. Dieser Zyklus ist das Charakteristikum *autonomen Verhaltens*.

In starker Abstraktion nennt man ein derartig befähigtes System in der Künstlichen Intelligenz einen *Agenten*. Ein Agent strebt stets nach dem Erreichen seiner Ziele oder der Erledigung seiner Aufgaben. Dies kann er aber nur, indem er die in seiner Umwelt wirksamen Gesetzmäßigkeiten erkennt und im Rahmen ihrer Gültigkeit handelt. Wie Abbildung 3 verdeutlicht, ist der hieraus erwachsende Konflikt nur lösbar, indem der Agent sowohl die Wahrnehmung der Umwelt verbessert, als auch seine Aktionen entsprechend koordiniert, indem er also lernt, den Konflikt zu reduzieren. Deshalb scheint das Dreibein — Robotik, Maschinensehen, Neuroinformatik — wesentlich für die technische Realisierung des Wahrnehmungs–Aktions–Zyklus zu sein.

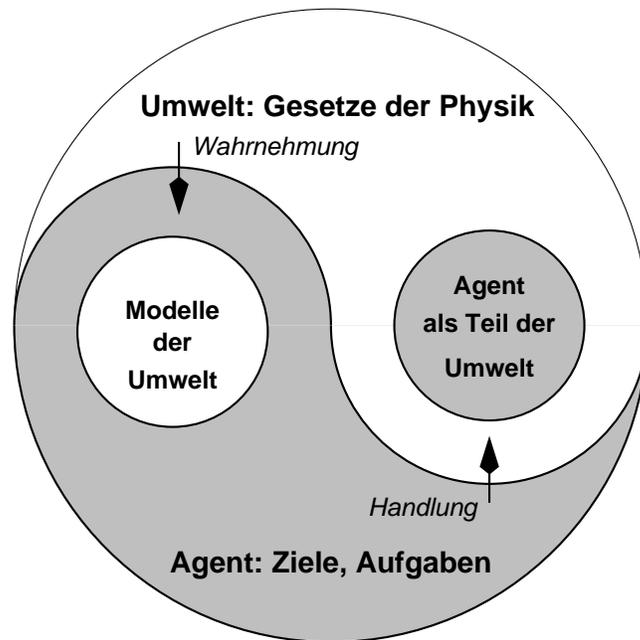
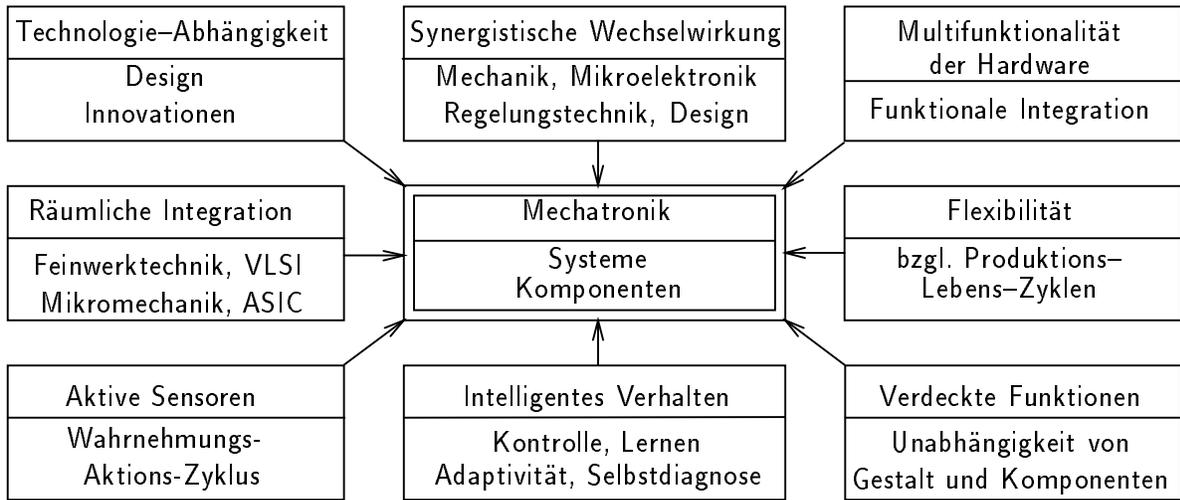
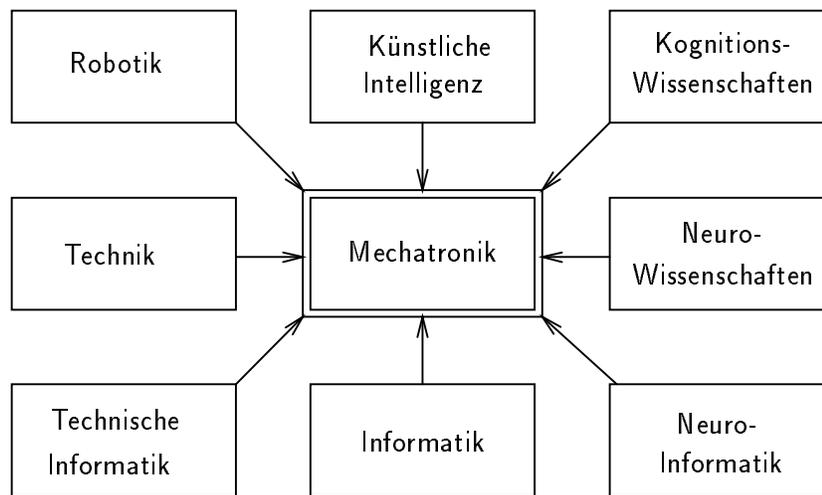


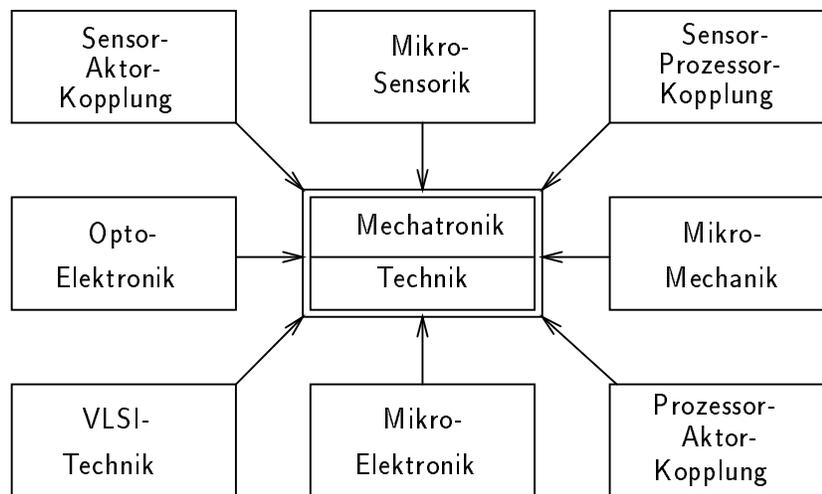
Abbildung 3: Wahrnehmungs–Aktions–Zyklus des autonomen Verhaltens.



(a)



(b)



(c)

Abbildung 4: (a) Entwurfsprinzipien mechatronischer Systeme bzw. Komponenten, (b) Interdisziplinarität der Mechatronik, (c) technische Aspekte der Mechatronik.

Mechatronische Systeme sind sehr komplex bezüglich ihrer Hardware-Komponenten und Software-Strukturen. Die Konzepte der Mechatronik dienen dazu, die *Körperlichkeit der Systeme* an die von ihnen zu bewältigenden Aufgaben anzupassen. Auch in der Biologie kennen wir die enge Beziehung von Struktur und Funktion. Was nötig ist, um dies zu erreichen, verdeutlichen die mittlere (b) und die untere (c) Graphik von Abbildung 4. In (b) werden die interdisziplinären Einflüsse dargestellt und in (c) die technischen Aspekte.

Die Perspektive der gegenwärtig in Entwicklung begriffenen Robotik ist durch ein Architekturkonzept charakterisiert, das stärker an biologische Vorbilder angelehnt ist als frühere Konzepte. Dieses als *verhaltensbasierter Entwurf* bezeichnete Architekturkonzept bedeutet eine Systemmodularisierung nach Verhaltensweisen.

Verhalten ist eine beobachtbare und im Sinne der Realisierung der Ziele eines Systems adäquate Handlung, die Ausdruck einer von der Körperlichkeit und Situiertheit des Systems geprägten spezifischen Kompetenz ist.

Wir erwarten also eine starke Diversifikation der Gestalten und Einsatzfelder künftiger Roboter. Gleichzeitig wird das heute dominierende Universalitätsprinzip informationsverarbeitender Systeme (z.B. in Form von Personalcomputern oder Workstations) zumindest für diesen Anwendungsfall abgelöst durch eine Palette spezifischer Komponenten, die sich im Sinne sowohl des verhaltensbasierten Entwurfs als auch der Mechatronik integrieren lassen.

Die enorme Innovationskraft dieser Bündelung wissenschaftlich-technischer Erkenntnisse führte national und international bereits zu langfristigen Förderprogrammen der Forschung. Stellvertretend seien genannt für:

Deutschland:

- BMBF-Verbundprojekt: “Elektronisches Auge” (seit 1994 gefördert)
- BMBF-Leitprojekte (seit 1995 gefördert) für:
 - Retina-Implantat
 - Greif-Implantat

Europa:

- ESPRIT-Projekt: “NEUROBOT — Neural Network-Based Robots for the Disassembly and Recycling of Automotive Products” (seit 1994 gefördert)
- ESPRIT-Projekt: “CONNY — Robot Control Based on Neural Network Systems” (1992 bis 1994 gefördert)

- ESPRIT-Projekt: “BLEARN — Behavior Learning: Combining Sensing and Action” (1992 bis 1994 gefördert)
- ESPRIT-Projekt: “PANORAMA — Perception and Navigation System for Autonomous Mobile Applications” (1989 bis 1993 gefördert)

USA:

- US Navy: Verschiedene Robotik Projekte
- NASA: “Space Telerobotics Program”
- TRC-NASA: “Real World Adaptive Autonomous Systems”

Japan:

- MITI: “Real World Computing Program” (seit 1993 gefördert)
- Sanyo Electric: “Vision-based Autonomous System”

Zielstellung dieser Projekte ist die Entwicklung von Systemen beziehungsweise Systemkomponenten, die Wahrnehmung und/oder Handlungen in unbekanntem, d.h. nicht vorher modellmäßig erfaßbaren Umgebungen ausführen können.

Einen Aspekt der Anwendung hiervon decken die derzeit popularisierten *Service-Roboter* ab. In einer vom BMBF beauftragten Studie des Fraunhofer-Instituts für Produktionstechnik und Automatisierung, Stuttgart, wird hierfür allein für Deutschland ein Marktpotential zwischen 15 und 40 Mrd. DM im Jahre 2010 prognostiziert [IPA, 1994].

Anliegen der Studie ist es, für eine innovative Technologie im Land Schleswig-Holstein die Aufmerksamkeit zu schärfen und zur Bündelung von Aktivitäten die erforderlichen Informationen bereitzustellen. In Schleswig-Holstein wird an unterschiedlichen Stellen auf robotik-relevanten Gebieten gearbeitet, d.h. produziert und geforscht. Es muß möglich sein, dieses Reservoir zum Zwecke der Entwicklung und Nutzbarmachung einer Hochtechnologie zu vereinen.

2 Gegenwärtiger Stand

Dieses Kapitel beschreibt den gegenwärtig international in den Forschungslabors vorhandenen Wissensstand auf dem Gebiet der visuell basierten Robotik. Zunächst wird das breite Spektrum von möglichen Anwendungsgebieten dargestellt und zugleich die Beschränktheit der zur Zeit typischerweise eingesetzten Robotersysteme aufgezeigt. Der Hauptteil behandelt die grundlegenden Funktionsprinzipien von visuell basierten Robotersystemen. Hierbei wird nach einer Beschreibung der typischen Bestandteile derartiger Roboter ganz wesentlich auf das roboterrelevante Maschinensehen und Maschinlernen eingegangen. Anschließend erfolgt eine Bewertung des Entwicklungsstandes im Hinblick auf die Umsetzbarkeit in praxistaugliche Produkte. Mit einer Auflistung und Kurzbeschreibung von Produkten (visuell basierte Roboter oder relevante Bestandteile), die von Firmen schon angeboten, beziehungsweise in wissenschaftlichen Instituten realisiert werden, schließt das Kapitel.

2.1 Anwendungsgebiete

Bezugnehmend auf die in der Einleitung diskutierte Problematik hinsichtlich der *Akzeptanz* der neuen Robotertechnologie durch die Gesellschaft werden hier zunächst

- allgemeingültige **Gebote**

für den Robotereinsatz angeführt. Anschließend folgen dann Anwendungsgebiete für das

- visuell basierte **Manipulieren**,
- visuell basierte **Fortbewegen** und
- aktive visuelle **Wahrnehmen**.

2.1.1 Gebote für den Einsatz von Robotern

Der Einsatz von Robotern hat sich an einigen Normen zu orientieren:

1. Ein Roboter unterstützt als intelligentes Werkzeug die Tätigkeiten und Aufgaben des Menschen.
2. Ein Roboter wird niemals zum Schaden des Menschen eingesetzt (als Individuum oder Gesellschaft).
3. Ein Roboter arbeitet zuverlässig und stets unter der Kontrolle des Menschen.

Diese Forderungen gelten sowohl für die klassischen als auch für die visuell basierten autonomen Roboter. Sie sind Voraussetzung für die Akzeptanz, sich solcher Maschinen zu bedienen.

Die Einsatzgebiete für Roboter, welche uneingeschränkte Akzeptanz finden können, lassen sich zum größten Teil daraus ableiten, bestehende Widersprüche zwischen dem evolutionär erreichten Entwicklungsstand von Geist und Körper des Menschen einerseits und dem Charakter der von ihm auszuführenden Tätigkeiten andererseits auszugleichen.

Der menschliche Körper stellt einen verletzlichen Organismus mit relativ begrenzten und kaum erweiterbaren körperlichen Fähigkeiten aber sehr flexiblen und leistungsfähigen geistigen Fähigkeiten dar. In diesem Spannungsfeld entwickelte er Produktionsweisen und Produktionsverhältnisse, die sehr ambivalent bezüglich seines evolutionären Entwicklungsniveaus zu bewerten sind. Einerseits erleichtern die erreichten gesellschaftlichen Fortschritte das Leben des Einzelnen. Andererseits führte die Arbeitsteilung zu Überforderung der körperlichen und Unterforderung der geistigen Fähigkeiten. Denn im Ergebnis der Evolution ist der Mensch eigentlich befähigt, seine körperliche und geistige Befähigungen in einem ausgewogenen Verhältnis wechselnd zu beanspruchen.

Hieraus ergeben sich folgende *Einsatzfelder* für Roboter:

1. **Ausführung lebens- oder gesundheitsgefährdender Arbeiten.**
Beispiele zur Lebensgefährdung: Brandbekämpfung, Havariebeseitigung.
Beispiele zur Gesundheitsgefährdung: Umgang mit chemischen oder radioaktiven Substanzen, Tätigkeiten unter Rauch-, Staub- oder Gasbelastung.
2. **Ausführung schwerer körperlicher Arbeiten.**
Beispiele: Verschiedene Arbeiten beim Hoch- und Tiefbau, Beladen von LKWs.
3. **Ausführung gleichförmiger, geistig wenig beanspruchender Tätigkeiten.**
Beispiele: Sortieren, Bestücken, Verpacken, Bewachen.
4. **Ausführung gleichförmiger schmutziger Tätigkeiten.**
Beispiele: Reinigen, Sortieren.
5. **Ausführung langanhaltender Präzisionsarbeiten.**
Beispiele: Bestücken, Kontrollieren.
6. **Ausführung periodisch wiederkehrender oder unperiodischer Handlungen in unphysiologischem Takt oder mit hoher Geschwindigkeit.**
Beispiele: Fließbandarbeit, Steuerung komplexer Maschinen.
7. **Ausführung von Arbeiten, für die mehr als zwei Hände nötig sind.**
Beispiele: Chirurgische Eingriffe, Montagearbeiten.
8. **Ausführung von Tätigkeiten, zu deren Kontrolle die menschlichen Sinnesorgane nicht geeignet sind.**
Beispiele: Navigieren im Dunkeln, Tätigkeiten in kontaminierten Umgebungen.

Die Abbildungen 5 und 6 zeigen Beispiele für zweckmäßige Einsatzgebiete von Robotersystemen.

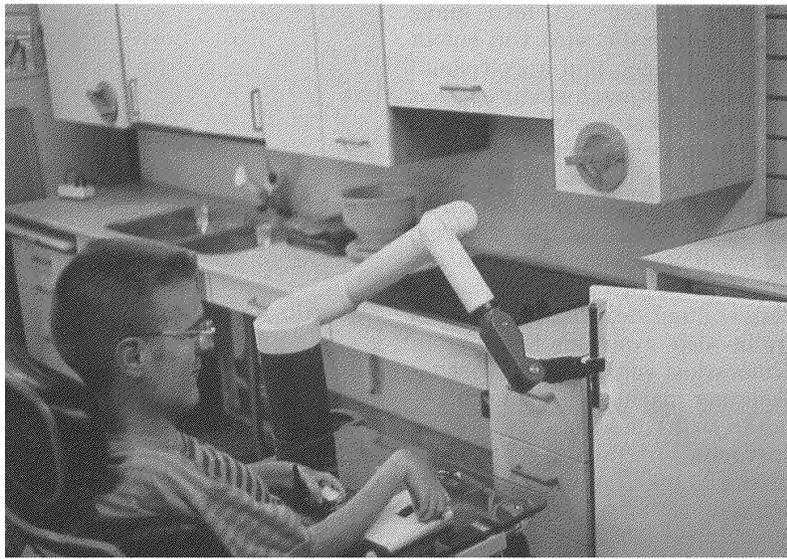


Abbildung 5: Unterstützung von Behinderten (aus [Reuber, 1994]).

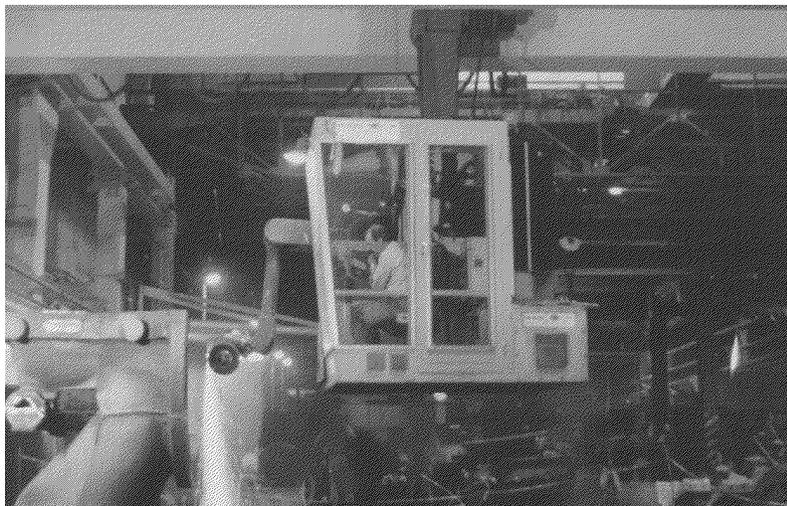


Abbildung 6: Ausführung eines gefährlichen Schleifvorgangs (aus Prospekt der Firma TELEROB, Kiel).

Viele Arbeitsplätze der modernen Industriegesellschaft tragen gleichzeitig mehrere Attribute, die es erforderlich machen, den Einsatz von Robotern zu ermöglichen. Dabei darf das Spannungsfeld, welches sich auf sozialökonomischem Sektor der menschlichen Gesellschaft entwickelt, nicht unbeachtet bleiben. Viele Roboter bedeuten weniger Arbeitsplätze. Auch hierfür muß eine menschenwürdige Lösung angeboten werden.

Niemals dürfen Roboter entwickelt werden, welche die Verletzung oder die Tötung von Menschen zum Ziel haben.

Dies ist auch zur Verbrechensbekämpfung oder zur Kriegsführung nicht gerechtfertigt, weil die Gefahr des Mißbrauchs zu groß ist. Werden die Roboter zukünftig unter Einhaltung der genannten Gebote entwickelt, dann ist eine Akzeptanz durch die Gesellschaft sehr wahrscheinlich. Die Roboter als Errungenschaften von Wissenschaft und Technik werden dann nicht nur als Diener angesehen, sondern darüber hinaus als Hilfsmittel für die Weiterentwicklung der menschlichen Lebensgemeinschaft.

2.1.2 Visuell basiertes Manipulieren

Unter *Manipulation* versteht man in der Robotik die Veränderung der Lage oder der Form eines Gegenstandes. Bei der *Lageveränderung* wird ein Gegenstand an eine andere Stelle gesetzt oder anders ausgerichtet. Derartige Tätigkeiten sind beispielsweise bei der Lagerhaltung zu erledigen (Abbildung 7).



Abbildung 7: Manipulator deponiert Kartons in ein Regal (aus [Nise, 1995]).

Es gibt viele weitere Beispiele. In der Automobilfertigung müssen etwa Karosserieteile entlang einer Montagelinie richtig angeordnet werden, um sie dann durch Punktschweißen zusammenzufügen. Möglich wäre beispielsweise eine visuelle Überwachung des Transportvorgangs, um Abweichungen von vorgegebenen Zielpositionen zu erkennen. Ein weiteres Beispiel zur Lageveränderung von Gegenständen ist das Entfernen von bestimmten fehlerhaften Werkstücken von einem Förderband oder allgemeiner das Sortieren von Gegenständen. So ist etwa die Sortierung von Müll, der über ein Förderband transportiert wird, eine wichtige Aufgabe. Anhand visueller Informationen ließen sich die For-

men der Gegenstände und darauf basierend die geeigneten Greifwerkzeuge ermitteln, um sie im Online-Betrieb richtig einzusetzen. Im Dienstleistungsbereich ist weiterhin die Unterstützung von Behinderten zu nennen. Beispielsweise ist Personen mit bestimmten Lähmungen sehr geholfen, wenn Eßbestecke oder andere Hilfsmittel auf Wunsch automatisch gereicht werden.

Alternativ zur Lage muß in zahlreichen Anwendungen die *Form* von Gegenständen *verändert* werden. Dazu gehören etwa im industriellen Bereich das Biegen, Schneiden, Schleifen, Fräsen oder Bohren von Werkstücken, sowie auch die Formveränderung durch das Montieren mehrerer Teile. Die Abbildung 8 zeigt das Fräsen von Metallteilen mit einem Manipulator, welcher durch eine Video-Kamera und Ultraschall-Sensoren teilautomatisiert arbeitet. Die Abbildung 9 zeigt eine Demontagezelle für alte PKWs. Dabei verrichtet ein Manipulator die eigentliche Arbeit, hier beispielsweise das Zerlegen eines Motors, und ein zweiter Roboterarm trägt ein Stereo-Kamerasystem. Durch visuelle Überwachung des Demontagevorgangs können frühzeitig fehlerhafte Abläufe korrigiert werden.

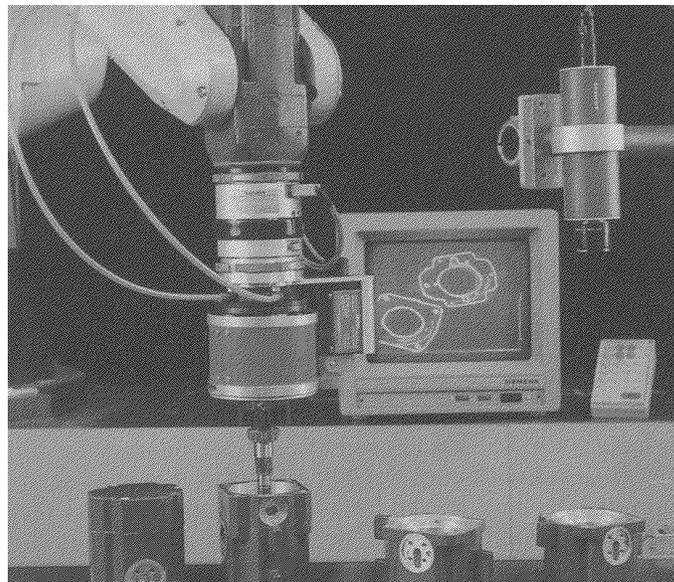


Abbildung 8: Manipulator fräst Metallteile, basierend auf Video und Ultraschall (aus [Marguerre, 1993]).

Realistisch betrachtet muß man festhalten, daß gegenwärtig in der industriellen Fertigung allenfalls eine visuell basierte Beobachtung und Überwachung von Manipulationsvorgängen erfolgt (siehe Abbildungen 8 und 9). Eine visuell basierte Kontrolle der Manipulation ist derzeit kaum möglich, obwohl hierzu von mittelständigen Firmen erste vielversprechende Produkte angeboten werden (siehe Abschnitt 2.4). Es wird aber nicht sinnvoll sein, visuell basierte Kontrolle dort einzuführen, wo die Produktionsabläufe in der geforderten Effizienz und Stabilität auch ohne diese garantiert werden können. Die situative Gleichförmigkeit der Produktion in großen Stückzahlen läßt es zu, kostengünstigere Vorkehrungen hierfür zu treffen. Visuell basierte Manipulation bedarf anderer Szenarien, die mit Handeln in unstrukturierter Umgebung im Zusammenhang stehen.

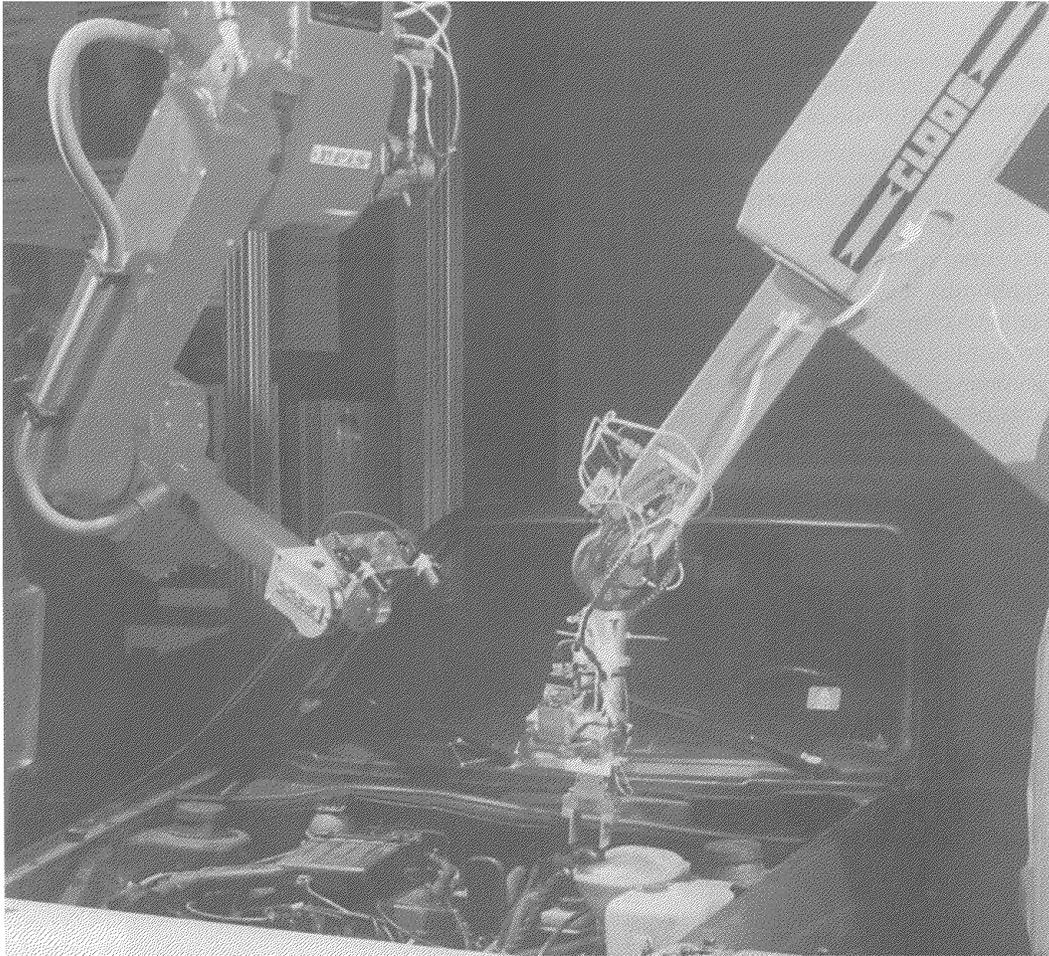


Abbildung 9: Manipulator demontiert Auto, basierend auf Stereo-Kamera (aus Broschüre der Fraunhofer-Gesellschaft, IITB, Karlsruhe).

2.1.3 Visuell basiertes Fortbewegen

Visuell basiertes *Fortbewegen* bedeutet das *autonome Gehen oder Fahren* eines Roboters, bei dem die Erfassung der Umgebung mit Hilfe bildgebender Sensoren stattfindet. Dieser Anwendungsbereich taucht auch unter den Begriffen *visuelle oder bildgestützte fahrerlose Navigation* auf. Die zwei Grundaufgaben des autonomen Fortbewegens sind das lokale Manövrieren und die globale Navigation.

Unter *lokalem Manövrieren* versteht man das Einhalten eines stabilen Kurses, einschließlich der Reaktion auf statische Hindernisse oder gefährliche dynamische Ereignisse auf der Fahrbahn.

Typische Beispiele sind voll autonome oder fahrerunterstützende Systeme für das Autofahren [Nagel, 1995]. Die Auswertung von Bildsignalen ermöglicht eine stabile Fahrtrichtung und Abbiegemanöver, die Erkennung von Gefahren aus statischen Hindernissen oder anderen Verkehrsteilnehmern und das Einhalten eines ausreichenden Abstandes zu vorausfahrenden Autos (Abbildung 10).

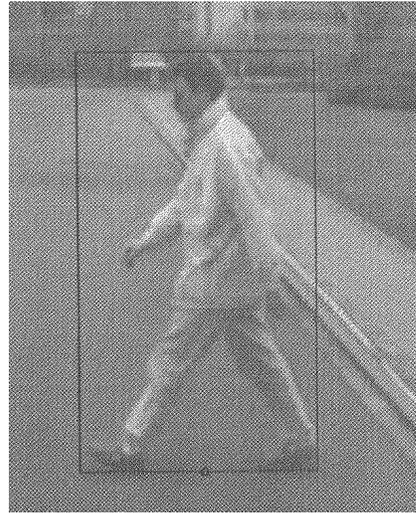
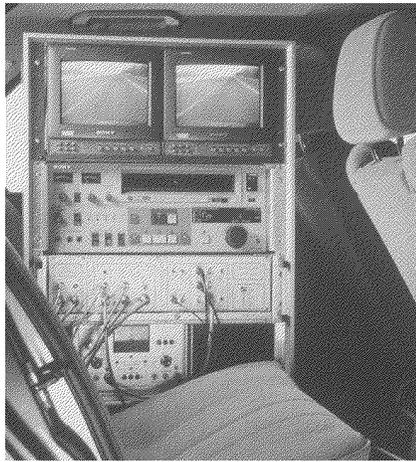


Abbildung 10: Fahrerunterstützung im Verkehr (BMW-Prospekt und [Nagel, 1995]).

Die unumstritten erfolgreichste Entwicklungsarbeit hierzu wird in Deutschland seit Jahren an der Universität der Bundeswehr in München in Zusammenarbeit mit Mercedes-Benz in Stuttgart durchgeführt [Dickmanns und Graefe, 1988]. Ein mit einem Stereo-Kamerasystem und zugehöriger Spezial-Hardware zur Bildverarbeitung ausgestatteter LKW kann sich autonom auf üblichen Bundesstraßen bewegen und erreicht hierbei eine maximale Geschwindigkeit von etwa 100 Stundenkilometer.

Globale Navigation heißt das Anfahren von Zielorten von einem bestimmten Startpunkt aus mit Hilfe von Landmarken und/oder Funksystemen. Dabei sind beispielsweise im öffentlichen Straßenverkehr bestimmte Regeln zu beachten, die durch Ampeln oder Verkehrszeichen angezeigt werden. In den Forschungslabors der großen Automobilkonzerne wird deshalb an der visuellen Erkennung von Ampeln und Verkehrszeichen sowie von Wegweisern gearbeitet.

Andere Anwendungsbeispiele für Außenszenen sind das autonome Fahren in unzugänglichen Gebieten, wie nuklear oder chemisch verseuchten Anlagen oder Minengebieten, sowie die Exploration des Meeresbodens oder der Oberfläche von Planeten. Die Fahrzeuge für derartige Einsatzorte werden gegenwärtig vorrangig durch manuelle *Teleoperation* gesteuert, ohne hierbei maschinell visuelle Informationen zu verarbeiten. Zu den Fortbewegungsmaschinen gehören nicht nur die Fahrzeuge, sondern auch mehrbeinige *Laufmaschinen*, die für die Erkundung auf unregelmäßigen Oberflächen oder in Kanalrohren eingesetzt werden.

Die zweite große Anwendungsgruppe bilden fahrerlose Systeme in Räumen [IPA, 1994]. In Industriehallen werden mobile Plattformen zum Transport und Palettieren eingesetzt. Eine *visuell basierte Kursregelung* würde keine zusätzlichen Infrastrukturkomponenten wie Transportschienen erfordern. Deutlich zeichnet sich der Einsatz von autonomen Fahrzeugen im Dienstleistungsbereich ab.

Fahrerlose Systeme können für den Transport der Post in Bürogebäuden oder für

die Verteilung von Medikamenten oder Eßgeschirr in Krankenhäusern eingesetzt werden (Abbildung 11, links oben). Weiterhin werden Staubsauger gebaut, die ohne menschliche Führung die zu reinigende Fläche erfassen (Abbildung 11, unten). Von größter Bedeutung werden künftig auch Hilfsmittel für kranke und behinderte Personen sein. Hierzu zählen selbstnavigierende Rollstühle, welche die Unabhängigkeit gehunfähiger Personen erhöhen. Aber auch tragbare Navigationsgeräte oder selbstfahrende Führersysteme zur Unterstützung von Blinden sind im Stadium der Entwicklung (Abbildung 11, rechts oben).



Abbildung 11: Navigierende mobile Roboter, für Transportaufgaben im Krankenhaus (oben links), Unterstützung von Blinden (oben rechts), autonome Bodenreinigung (unten), aus [IPA, 1994].

Zu visuell basiertem Fortbewegen befähigte Systeme befinden sich im Stadium der Laborentwicklung. Weiter in den Bereich praktischer Anwendungen vorgedrungen sind Systeme, die sich auf der Basis anderer Sensorik (z.B. Ultraschall) fortbewegen können.

2.1.4 Aktive visuelle Wahrnehmung

Unter *aktiver visueller Wahrnehmung* versteht man die Regelung der visuellen Aufmerksamkeit, die sowohl durch die mechanische Steuerung der Blickrichtung der bildgebenden Sensoren als auch durch die selektive Verarbeitung der Bilddaten erfolgt.

Motivation der *Aufmerksamkeitssteuerung* sind sowohl das eingeschränkte Gesichtsfeld bei üblichen Kameras als auch die uneffiziente Verarbeitung der visuellen Information bei statischen Kameras. Trotz der stetigen Leistungserhöhung der Rechentechnik kann nicht das ganze, von einer Kamera erfaßte, visuelle Feld mit höchster Detailtreue in Echtzeit ausgewertet werden. Die Steuerung der Kameras ermöglicht rechenintensive detaillierte Berechnungen nur für den fixierten Bereich des Gesichtsfeldes.

Bildgebende Sensoren, einzeln oder als Paar auf Schwenk–Neige–Mechanismen montiert, werden für die Erkundung und Überwachung von Szenen eingesetzt, damit der Einsatz mehrerer stationärer Kameras zur Erfassung der gesamten Szene vermieden wird. Die Abbildung 12 zeigt eine neigbare Kameraanordnung auf einem Fahrzeug zur Überwachung von Museumsräumen.

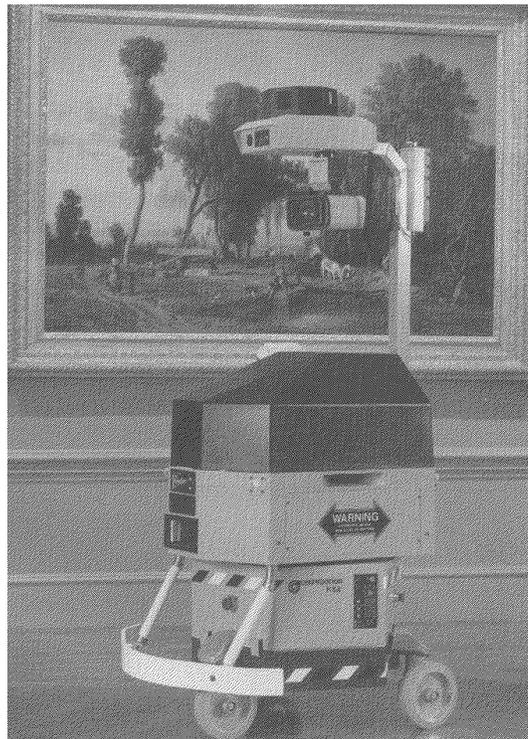


Abbildung 12: Überwachung von Museumsräumen (aus [Nise, 1995]).

Die *Kameranachführung* ermöglicht das visuelle Verfolgen von Personen oder Fahrzeugen in Überwachungsanwendungen. Gegenwärtig erfolgt die Überwachung üblicherweise durch Betrachtung der Kamerabilder durch aufmerksames Personal, wobei die Kameras häufig wegen ihrer starren Befestigung einen eingeschränkten Sichtbereich haben und somit zahlreich aufgestellt sein müssen. Im Kommunikationsbereich können künstliche *aktive Beobachtungssysteme* den natürlichen Bewegungen der aufgenommenen Personen autonom folgen. Dies hat den Vorteil, daß sich beispielsweise Personen vor einem Bildtelefon oder Teilnehmer einer Videokonferenz freier bewegen dürfen.

Weiterhin können schwierige Videoaufnahmen automatisiert werden, beispielsweise von Sportveranstaltungen oder von schwer zugänglichen Gebieten. Industriell verfügbar sind derartige Systeme noch nicht, jedoch wurde beispielsweise auf der Hannover-Messe 1995 ein aktives Stereo-Kamerasystem vorgestellt, das in Echtzeit Bewegungen verfolgen kann. Bildgebende Sensoren mit mechanischen Freiheitsgraden werden auch in Manipulations- und Navigationsaufgaben (siehe 2.1.2 und 2.1.3) eingesetzt.

Stereoskopische Systeme mit mechanischen Freiheitsgraden erlauben die Ermittlung genauerer Abstandsinformation zum Greifen und zuverlässigere Wahrnehmung der Bewegung bei Manövrieraufgaben. Robotersysteme, die basierend auf visueller Information Gegenstände manipulieren oder sich fortbewegen und hierzu auch noch aktiv die Freiheitsgrade der Kameraanordnung verändern, sind gegenwärtig nur in Labors verfügbar.

2.2 Funktionsprinzipien

Nachdem das breit gestreute Spektrum von möglichen Anwendungsgebieten der visuell basierten Robotik dargelegt ist, werden nun die grundlegenden Funktionsprinzipien, die der Realisierung zugrundeliegen, erklärt. Der erste Abschnitt beschreibt die Hauptkomponenten von wahrnehmenden Robotern. Dann folgt eine Kategorisierung von Robotersystemen — die auch die Industrieroboter einschließt — im Hinblick auf den Grad der erreichbaren Autonomie. Schließlich wird im dritten Abschnitt die bis zum Ende der achtziger Jahre vorherrschende Konzeption von wahrnehmenden Robotern vorgestellt. Die letzten beiden Abschnitte behandeln dann als wesentliche Funktionsprinzipien das roboterrelevante Maschinensehen und Maschinenlernen.

2.2.1 Bestandteile von Robotersystemen

Roboter können anhand der grundlegenden physischen Bestandteile charakterisiert werden. Diese Bestandteile sind:

- das **Aktuator-System**,
- das **Sensor-System**,
- das **Antriebs-System** und
- das **Rechner-System**.

Das Aktuator-System

Das *Aktuator-System* umfaßt alle beweglichen Komponenten eines Roboters. Damit kann sowohl die Form des Roboters als auch die Position und Orientierung verändert werden. Aktuator-Systeme werden eingesetzt

- zur **Fortbewegung von Robotern**,
- zur **Manipulation von Gegenständen** und
- zur **Positionierung und Orientierung von Sensoren**.

Diejenigen Komponenten des Aktuator-Systems, die unmittelbar mit der Umwelt in physischen Kontakt kommen, heißen *Effektoren*. Betrachtet man den Roboter als Bestandteil der Umwelt, so wird diese einerseits durch die Bewegung des Aktuator-Systems und andererseits durch eventuell von den Effektoren manipulierte Gegenstände verändert.

Die Effektoren bei mobilen Robotern sind die Fortbewegungsmittel wie beispielsweise Räder (Abbildung 12), Kettenlaufwerke (Abbildung 13) oder Füße (Abbildung 14). Bei ebenem Terrain, etwa in Fabrikhallen, Gebäudefluren oder geteerten Straßen, werden Räder verwendet. Unwegsames Gelände macht statt dessen Kettenlaufwerke oder auch

Beine mit Füßen erforderlich. Kettenlaufwerke werden bei Lasttransport bevorzugt, die Beine beispielsweise bei ökologisch wertvollem Gelände, das nicht zerstört werden darf.



Abbildung 13: Roboter mit Ketten (aus Prospekt der Firma TELEROB, Kiel).

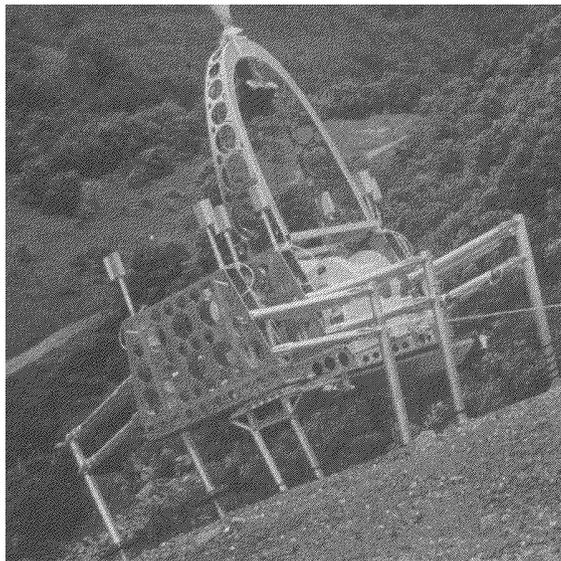


Abbildung 14: Roboter mit Beinen (aus [Nise, 1995]).

Die mit einem Effektor an einem Gegenstand verrichtete Tätigkeit wird als *Manipulation* bezeichnet. Dies kann, wie im Abschnitt 2.1.2 bereits erwähnt, entweder eine formverhaltende Bewegung oder eine physische Veränderung eines Gegenstandes sein. Zur Bewegung von einfachen Gegenständen mit flacher Oberflächengestalt kann als Effektor eine Vakuum-Saugglocke verwendet werden. Bei Gegenständen mit komplexer Form werden Greifer mit zwei oder mehreren starren Backen verwendet. In zunehmendem Maße

kommen auch gelenkige Finger zum Einsatz, eventuell sogar mit elastischer Beschichtung. Die adäquate Positionierung und Orientierung des Effektors erfolgt über einen *Roboterarm*.

Der Roboterarm zusammen mit dem Effektor wird häufig als *Manipulator* bezeichnet. Die Charakterisierung eines Roboterarmes erfolgt durch die Art und die Anzahl von Gelenken. Bei einem Lineargelenk werden zwei benachbarte Komponenten parallel oder orthogonal zueinander verschoben. Bei einem Drehgelenk werden zwei benachbarte Komponenten durch Rotation oder Torsion gegeneinander verdreht.

Wichtige Typen von Roboterarmen sind:

- der **Gantry–Roboterarm**,
- der **Scara–Roboterarm** und
- der **Artikulation–Roboterarm**.

Der *Gantry–Roboterarm* hat eine Folge von Lineargelenken und erstreckt sich häufig über ganze Fabrikhallen, um Gegenstände zu befördern. Der *Scara–Roboterarm* hat dagegen bis zu drei aufeinanderfolgende Drehgelenke, um vorrangig Gegenstände auf einer ebenen Arbeitsfläche zu manipulieren. Der *Artikulation–Roboterarm* ist beispielsweise durch 6 aufeinanderfolgende Drehgelenke so konfiguriert, daß der Effektor in einem bestimmten Arbeitsraum beliebig orientiert werden kann (Abbildung 15).

Ergänzend zu den obigen beiden Anwendungen des Aktuator-Systems, nämlich der Fortbewegung von Robotern und der Manipulation von Gegenständen, werden Aktuatoren neuerdings auch für die geeignete Positionierung und Orientierung von Sensoren eingesetzt. Im einfachsten Fall ist beispielsweise eine Infrarot- oder eine Laser-Kamera auf einem vertikal gerichteten Drehgelenk (*Pan–Bewegung*) montiert, so daß eine Rundum-Beobachtung möglich wird. Mit einem zusätzlichen horizontal gerichteten Drehgelenk (*Tilt–Bewegung*) ist auch eine Beobachtung möglich, die aus der horizontalen Ebene herausführt.

Als *Roboterkopf* wird ein Aktuator bezeichnet, der zusätzlich zur Pan- und Tilt-Bewegung noch Freiheitsgrade zur *Vergenz–Regelung* hat. Damit können zwei Kameras einer Stereo-Anordnung unabhängig voneinander auf einen gewünschten Gegenstand gerichtet werden (Abbildung 16). Der Roboterkopf kann nun selbst auf einer sogenannten Schwenkvorrichtung montiert sein. Damit lassen sich Gegenstände von unterschiedlichen Blickpositionen beobachten, ohne den Roboter insgesamt fortbewegen zu müssen.



Abbildung 15: Artikulation-Roboter mit sechs Drehgelenken.

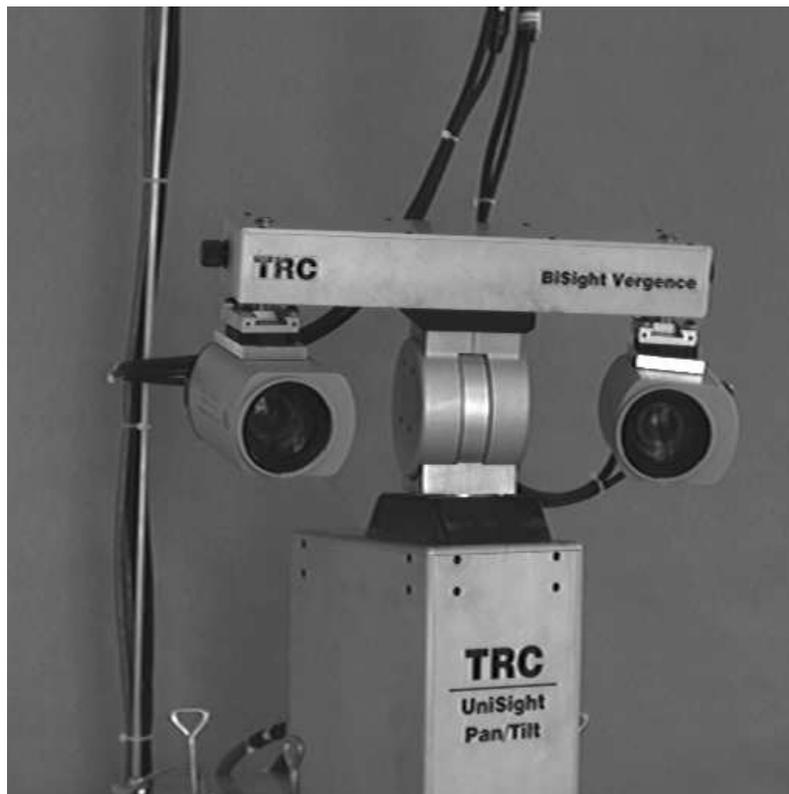


Abbildung 16: Roboterkopf mit Pan-, Tilt- und Vergenz-Freiheitsgraden.

Das Sensor-System

Mit dem *Sensor-System* werden Informationen für den gewünschten Einsatz der Effektoren bereitgestellt. Zu unterscheiden ist zwischen interner und externer Sensorik.

Interne Sensoren messen und erkennen systembezogene Zustände wie Stellgrößen von Motoren oder Fehlersituationen. Darüber hinaus können interne Sensoren auch grobe Informationen über die Umwelt liefern. Beispielsweise ist bei Fahrzeugen durch Messen der Neigung dann auch unmittelbar die Geländeneigung gegeben. Auch die Fahrzeugposition im Gelände läßt sich unter der Annahme einer Geradeaus-Bewegung mit internen Sensoren ungefähr bestimmen. Dabei wird die zurückgelegte Strecke indirekt anhand der Umlaufgeschwindigkeit der Räder in einem bestimmten Zeitintervall (*Odometrie*) berechnet. Eine weiterer Typ von internen Sensoren, der vorrangig bei Roboterarmen vorkommt, sind die *Drehmoment-Sensoren*. Verfügt beispielsweise die Hand des Roboterarms über einen auslenkbaren Finger, so kann damit die Berandung eines Objektes abgefahren und somit das Objekt vermessen werden. Die Steuerung erfolgt hierbei mit Hilfe des Drehmoments, das am Fingergelenk durch die Fingerauslenkung hervorgerufen wird.

Interne Sensoren können die Beschaffenheit der Umwelt nur ungenau und wenig detailliert wiedergeben. Ausführlichere Informationen erhält man mit *externen Sensoren*. Eine plausible Charakterisierung kann nach der Art des Mediums erfolgen, das zwischen Sensor und Umwelt wahrgenommen wird. Man unterscheidet die Wahrnehmung

- durch **Kontakt**,
- durch **Luft** und
- durch **Strahlung**.

Kontakt-Sensoren (auch *taktile Sensoren* genannt) haben eine elastische Oberfläche, die sich bei Berührung mit einem Gegenstand verformt. Dadurch werden bestimmte elektrische Schaltkreise geschlossen, die dem Antrieb der Effektoren neue Steuersignale übermitteln. Bei einer ausreichend gewählten Pufferzone zwischen dem Kontakt-Sensor und den übrigen Komponenten kann der Roboter rechtzeitig geeignet reagieren. Bei mobilen Robotern dienen sie zum Schutz vor Eigenzerstörung sowie der Zerstörung von Hindernissen.

Um Hindernisse ohne Kontaktierung zu erkennen, lassen sich sogenannte *Luft-Sensoren* verwenden. Dabei wird Luft ausgesendet und die eventuell von einem Hindernis verursachte Reflexion ausgewertet. Pneumatische Sensoren sind für den Fall geeignet, daß sich ein Hindernis unmittelbar vor dem Roboter befindet, so daß die Behinderung der ausströmenden Luft gemessen werden kann. *Ultraschall-Sensoren* modulieren Luft mit Schallwellen im Ultraschallbereich und berechnen aus der Hin- und Rücklaufzeit und der Schallgeschwindigkeit die Distanz zu Hindernissen.

Die dritte Klasse der externen Sensoren, *Strahlungs-Sensoren*, nimmt einen Gegenstand durch eine von ihm ausgehende Strahlung wahr. Beispielsweise kann das durch eine stromdurchflossene Hochspannungsleitung verursachte Magnetfeld mit sogenannten Magnetfeld-Sensoren gemessen werden. Dies erfolgt dadurch, daß das Magnetfeld selbst im Sensor wiederum eine Spannung, die sogenannte Hall-Spannung, induziert.

Die weiteren strahlungsbasierten Sensoren können in das Gesamtspektrum der elektromagnetischen Wellen eingebettet werden. Die *Infrarot-Sensoren* und *Infrarot-Kameras* nehmen Strahlung im Wellenlängenbereich von einigen Mikrometern wahr. Derartige Strahlung wird von Gegenständen ausgesendet, die Temperaturen zwischen -20 und $+1500$ Grad Celsius haben. Typische Anwendungsfelder sind die Inspektion von industriellen Anlagen etwa im Hinblick auf Wärmeverlust. Ebenso können beispielsweise fehlerhafte Oberflächenbeschichtungen durch unterschiedliche Wärmeleitfähigkeit erkannt werden.

Wählt man im Gesamtspektrum kürzere Wellenlängen als bei Infrarotstrahlung, so kommt man zur *Lichtstrahlung* (Wellenlängen im Bereich zwischen 0.4 und 0.8 Mikrometern). *Laser-Distanz-Sensoren* und *Laser-Kameras* arbeiten beispielsweise mit einer künstlich intensivierten Lichtstrahlung, die sie in bestimmte Richtungen aussenden und als Reflexion empfangen. Zur Entfernungsmessung existieren mehrere Methoden, wie

- **Triangulationsverfahren,**
- **Laufzeitverfahren,**
- **Phasenverfahren** und
- **Unschärfeverfahren.**

Beispielsweise wird beim Laufzeitverfahren die Hin- und Rücklaufzeit gemessen und durch Verrechnung mit der Lichtgeschwindigkeit die Distanz zu Hindernisobjekten ermittelt. Durch die enge Bündelung der Lichtstrahlung können die angestrahlten Gegenstände sogar dreidimensional vermessen werden.

Die übliche Wahrnehmung der Umwelt durch Licht ist jedoch zweidimensional. Dabei wird die von einem Gegenstand abgestrahlte Lichtmenge über ein Linsensystem gesammelt und auf einen Video-CCD-Chip weitergeleitet. Der *CCD-Chip* (Charge Coupled Device) ist eine Matrix von lichtempfindlichen Rezeptoren, die jeweils durch das lokal auftreffende Licht — vergleichbar mit einem Kondensator — aufgeladen werden. Bei *Monochrom-Kameras* hat man je Matrixelement einen Rezeptor, der über alle Wellenlängen des Lichts integriert. *Farb-Kameras* verfügen über drei Typen von Rezeptoren, die jeweils empfindlich im blauen, im grünen und im roten Wellenlängenbereich des Lichts sind. Ein einzelner Rezeptor hat typischerweise eine Seitenlänge von 11 Mikrometer. Die Abmessungen von kommerziellen CCD-Chips betragen beispielsweise 6.6 Millimeter in der Höhe und 8.8 Millimeter in der Breite. Das Belichten und Entladen der Rezeptoren erfolgt häufig im sogenannten *Interline-Modus*. Hierbei werden jeweils in zwei aufeinanderfolgenden Zeitabschnitten die Rezeptoren mit ungerader oder gerader Zeilennummer entweder belichtet oder entladen. Der Zeitabschnitt für das Entladen (Auslesen des analogen Videosignals)

der ersten Serie von Rezeptoren wird ausgenutzt zum Belichten der zweiten Serie von Rezeptoren. Bei der europäischen PAL-Norm ist eine Aufnahme­rate von 50 Halbbildern je Sekunde erforderlich, bei der amerikanischen NTSC-Norm sind es 60 Halbbilder. Der einer Video-CCD-Kamera nachgeschaltete *Frame-Grabber* baut die beiden Halbbilder jeweils zu einem Vollbild zusammen und führt eine Digitalisierung durch. Es entstehen digitale Bilder mit diskreter Anzahl von Pixeln und diskreter Anzahl von Helligkeitsstufen (Grauwerten).

Für Zwecke der anwendungsorientierten Forschung sind heutzutage Roboter mit verschiedenen Typen von externen Sensoren erhältlich. Üblich ist dabei eine Kombination von unterschiedlichen Sensoren, wie beispielsweise Photo-Dioden, Infrarot-Sensoren, Laser-Distanz-Sensoren und Ultraschall-Sensoren. Dabei hat man aber prinzipiell das Problem, daß die Sensoren nur punktuell die Distanzen zu Szenengegenständen ermitteln, somit also keinen zusammenhängenden Szenenausschnitt abbilden. Hinzu kommt beispielsweise bei Ultraschall-Sensoren eine mögliche Verfälschung der Distanzmessung durch sogenannte *Doppelreflexion*, falls die Szenengegenstände glatte Oberflächen haben. Wegen dieser Nachteile werden zunehmend Video-CCD-Kameras eingesetzt, die die einfachen Sensoren ergänzen oder teilweise ersetzen. Dieser neue Trend wird dadurch gefördert, daß zum einen die zur Verarbeitung der großen Datenmenge erforderliche Rechenkapazität zunehmend verfügbar ist. Zum anderen wurden die zugehörigen Rechner und die Video-CCD-Kameras in ausreichendem Maße miniaturisiert. Allerdings fehlen derzeit noch grundlegende Konzepte zur Handhabung des hochdimensionalen Zustandsraumes, welcher durch die hochaufgelöste Abbildung der Szene entsteht. Die effektive und bedeutungserhaltende Reduzierung dieses Zustandsraumes stellt die Hauptaufgabe der Forschung auf dem Gebiet des Maschinensehens dar.

Das Antriebs-System

Alle Gelenke des Aktuator-Systems müssen individuell angetrieben werden, um die Freiheitsgrade in gewünschter Weise einstellen zu können. Das *Antriebs-System* für ein Gelenk besteht jeweils aus dem Motor und dem zugehörigen Getriebe. Hinzu kommen insbesondere noch Bremsen und Massenausgleichskomponenten. Haben die einzelnen Getriebe jeweils geringe Fehlertoleranzen, so erreicht man mit dem Effektor eine hohe Wiederholgenauigkeit. Bei den Motoren werden bevorzugt *Schrittmotoren* eingesetzt, die eine schnelle und präzise Positionierung oder Vorschubbewegung ermöglichen.

Das Rechner-System

Mit dem *Rechner-System* erhält der Benutzer die Möglichkeit, dem Roboter Befehle zu übermitteln. Diese Roboterbefehle müssen in Signale für die Motorregelung transformiert werden.

Im Falle sensor- oder visuell basierter Robotik wird die zur Erledigung einer Aufgabe erforderliche Sequenz von Roboterbefehlen nicht explizit vorprogrammiert. Vielmehr wird das Roboterprogramm aus einer, besonders im Falle visueller Sensorik, aufwendigen

gen Analyse der sensorischen Daten abgeleitet. Zum Einhalten der Realzeiterfordernisse sind hierfür besondere Rechenkapazitäten bereitzustellen, die im Falle visueller Sensorik in hohem Maße von den zu lösenden Aufgaben des Bilderkennens abhängen.

Als Hardware-Komponenten sind insbesondere ein *Hauptprozessor* vorhanden, der häufig mit speziellen *Co-Prozessoren* und *Signalprozessoren* ergänzt wird. Weitere Hardware-Komponenten des Rechner-Systems sind der Arbeits- und der Hintergrundspeicher, um beispielsweise Programme mit Roboterbefehlen und Daten zur Initialisierung des Programms speichern zu können. Die Software eines Roboter-Rechners enthält im wesentlichen einen Interpreter für eine Roboter-Programmiersprache, beispielsweise VAL von der Firma Stäubli. Hinzu kommen noch Editoren zur Programmerstellung und Systemkommandos zur Dateiverwaltung.

Der Hauptprozessor des Rechner-Systems übersetzt die einzelnen Roboterbefehle in Mikroprogramme, die auf Co-Prozessoren effizient ablaufen. Ein Beispiel hierfür ist die Berechnung von notwendigen Gelenkstellungen, wenn der Effektor eine bestimmte Position und Orientierung einnehmen soll (*Inverse Kinematik*). Die Veränderung der Gelenkstellungen muß dabei exakt zwischen den verschiedenen Gelenken koordiniert werden. Dies ist insbesondere dann erforderlich, wenn der Effektor eine bestimmte Trajektorie abfahren soll, eventuell sogar mit einer bestimmten Geschwindigkeit (*Inverse Dynamik*). Die von den Co-Prozessoren gelieferte Sequenz von gewünschten Gelenkstellungen wird dann von nachgeschalteten Signalprozessoren bearbeitet. Ziel ist die Erzeugung von Servosignalen für die Motorregelung, um schrittweise die gewünschten Gelenkstellungen zu erreichen.

Die für visuelle Wahrnehmung erforderlichen Rechner werden in Abschnitt 3.2.2 charakterisiert.

2.2.2 Kategorien der Roboterautonomie

Voraussetzung für die Entwicklung von autonomen Robotersystemen ist, daß die von der externen Sensorik bereitgestellten Daten einen gewissen Genauigkeitsgrad haben und zweckdienliche Informationen daraus extrahiert werden können. Hinsichtlich der Rolle der externen Sensorik, und somit der Verwendung von Umweltdaten, kann unterschieden werden zwischen:

- **Programm-Roboter,**
- **Teleoperation-Roboter** und
- **Passiv-wahrnehmende-Roboter.**

Programm-Roboter

Programm-Roboter arbeiten gemäß einer klar definierten Abfolge von Befehlen und sind typischerweise in der großindustriellen Serienfertigung im Einsatz. Die Umwelt wird

nicht wahrgenommen oder nur sehr selektiv anhand von Kontakt- oder Näherungs-Sensoren. Alle peripheren Gegenstände und Geräte, mit denen der Roboter zu tun hat, müssen vorbestimmte Formen und Positionen haben, wobei nur kleine Toleranzen zugelassen sind. Die gesamte *Bewegungstrajektorie* des Roboters wird detailliert programmiert. Dies erfolgt beispielsweise durch sogenannte *Teach-in-Programmierung*. Der Roboter wird hierbei manuell von Bahnpunkt zu Bahnpunkt gesteuert, und ein Rechner speichert die einzelnen Schritte.

Eine Alternative ist die sogenannte *Offline-Programmierung*. Hierbei wird die Umgebung des Roboters fernab vom tatsächlichen Einsatzbereich simuliert, und die beabsichtigte Tätigkeit mit Hilfe einer Workstation am Bildschirm entworfen und optimiert. Ein Vorteil dieser Vorgehensweise gegenüber der Teach-in-Programmierung ist, daß bei Umstellungen die Ausfallzeiten der Produktion verkürzt werden. Allerdings muß ein möglichst genaues Modell der Arbeitsumgebung existieren.

Teleoperation-Roboter

Bei unstrukturierter Arbeitsumgebung, die eventuell sogar fortwährenden Änderungen unterworfen ist, sind herkömmliche Programm-Roboter ungeeignet. Falls diese Arbeiten für die Menschen zu schädlich oder gefährlich sind, so kommen manuell bedienbare Roboter zum Einsatz, wobei das Bedienungspersonal beispielsweise durch eine Glaswand geschützt ist (siehe Abbildung 6). Ist der Blick auf das Werkobjekt wegen Unzugänglichkeit oder großer Distanz behindert, so muß die Bedienung über sogenannte *Teleoperation* erfolgen.

Ein *Teleoperation-Roboter* besteht aus dem vor Ort handelnden Aktuatorsystem zusammen mit dem Sensorsystem und dem in sicherer Entfernung befindlichem Kontrollstand. Der handelnde Aktuator besteht typischerweise aus einer mobilen Plattform und einem Manipulator. Mit Hilfe von Sensoren, beispielsweise Ultraschall-, Infrarot- oder Kraft-Sensoren, sowie Video- oder Laser-Kameras, wird die Umgebung wahrgenommen, und die Information per Funk oder Kabel an den Kontrollstand übertragen. Auf dieser Grundlage steuert das Personal den Aktuator, der schritthaltend vor Ort eine bestimmte Aufgabe erledigt. Bei Teleoperation-Robotern ist also der Mensch unmittelbar in den Wahrnehmungs-Aktions-Zyklus integriert.

Es können unterschiedliche Grade der Unterstützung durch das Rechner-System unterschieden werden. Eine minimale Unterstützung liegt dann vor, wenn die Stellgrößen für den handelnden Aktuator unmittelbar von den Bedienern ermittelt werden. Eine Verbesserung bieten sogenannte *Master/Slave-Systeme*. Der im Einsatz befindliche Aktuator ist der Slave, und im Kontrollstand befindet zusätzlich eine vereinfachte Kopie davon, der Master. Durch die Bedienung des Masters wird der Slave geführt, wobei aufgrund der Sensoren der Eindruck entsteht, als würde direkt der Slave bedient. Da Teleoperation-Roboter mühsam zu bedienen sind, und Programm-Roboter nur in genau spezifizierten Umgebungen arbeiten können, sind zunehmend autonom handelnde Roboter erforderlich.

Passiv-wahrnehmende-Roboter

Diesen Robotertyp steuert man vorrangig anhand der Szenendaten, die von externen Sensoren oder Video-CCD-Kameras geliefert werden. Sie nehmen die Umwelt aber nur passiv wahr, d.h. eine Regelung der Blickrichtung zur aktiven Veränderung der Aufmerksamkeit findet nicht statt.

Unter industriellen Bedingungen sind bislang lediglich mobile Roboter imstande, auf der Grundlage von externen Sensoren autonom zu handeln. So können mobile Roboter an Fahrbahnmarkierungen entlangfahren und somit beispielsweise im Fabrikgelände Transport- und Lagerarbeiten durchführen. Die notwendigen Markierungen werden mit einer Video-Kamera aufgenommen und durch einfache Bildverarbeitung erkannt. Im wissenschaftlichen Bereich haben mobile Roboter darüber hinausgehende Fähigkeiten. Beispielsweise ist das Entlangfahren an einer Wand oder die Annäherung an eine Wand möglich, ohne daß Fahrbahn- oder Wandmarkierungen erforderlich sind. Mit Methoden der Odometrie (Integration der Geschwindigkeit), zum Teil in Verbindung mit der Erkennung von Landmarken, kann damit ein grober Plan über die Freigebiete von Räumen ermittelt werden.

Autonom handelnde Manipulatoren, beispielsweise auf der Grundlage von Video-CCD-Kameras, sind bislang in der Industrie nicht im Einsatz, jedoch sind sie Gegenstand intensiver wissenschaftlicher Forschung. Die dominierende Konfiguration besteht darin, daß eine oder zwei Kameras direkt am Effektor befestigt sind. Beispielsweise können bei einem Greifer die Kameras entweder zwischen den Fingern (*Eye-in-Hand*) oder auf dem Handrücken (*Eye-on-Hand*) befestigt sein. Das Greifobjekt muß sich schon in unmittelbarer Nähe des Greifers und somit im Sichtbereich der Kameras befinden. Unter dieser Voraussetzung wird dann der Greifer so bewegt, daß das Greifobjekt in eine bestimmte relative Stellung zu den Greiffingern kommt. Die Greiffinger können dabei selbst im Sichtbereich der Kameras sein. In diesem Fall läßt sich unmittelbar visuell eine stabile Greifstellung ermitteln. Andernfalls wird die a priori bekannte Position, Orientierung und Form der Greiffinger in Beziehung gesetzt zur dreidimensionalen Lage des Greifobjekts, welche aus dem Bildmaterial rekonstruiert werden muß.

Eine alternative Konfiguration besteht darin, die Kameras unabhängig vom Manipulator an einer distanzierten Stelle zu positionieren. Unter Verwendung von Weitwinkelobjektiven kann somit ein größerer Szenenbereich eingesehen werden. Damit läßt sich auch ein größerer Abstand zwischen Greifer und Greifobjekt überbrücken. Allerdings bleiben wegen der größeren Entfernung zwischen Kamera und Greifobjekt manche Details verborgen, so daß unstabile Greifstellungen wahrscheinlicher werden. Eine geeignete Kombination der beiden Konfigurationen wurde bislang noch nicht untersucht.

2.2.3 Funktionale Modularisierung von Robotern

Bis Ende der achtziger Jahre war die Konzeption für passiv wahrnehmende Robotersysteme einheitlich durch eine strenge *funktionale Modularisierung* charakterisiert. Dabei hatte man im wesentlichen drei Funktionen unterschieden, die sequentiell durchlaufen wer-

den (Abbildung 17):

- **Umweltwahrnehmung,**
- **Planung** und
- **Robotersteuerung.**

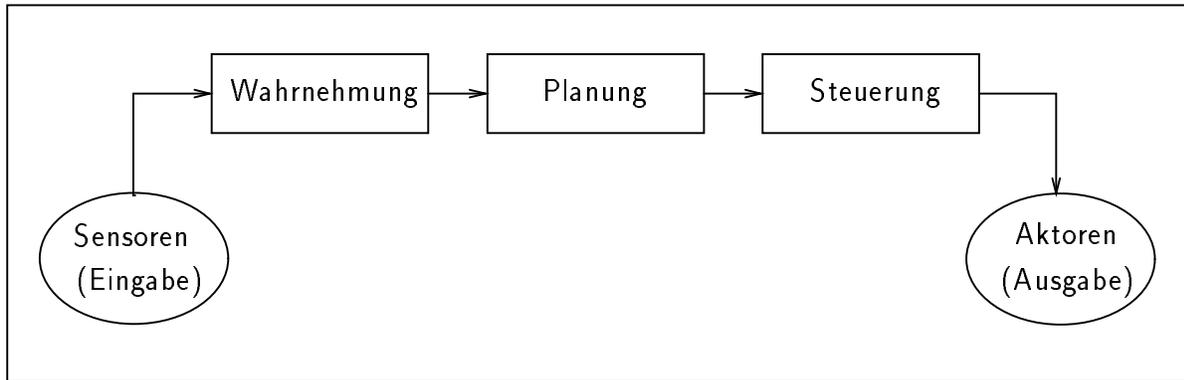


Abbildung 17: Funktionale Modularisierung eines Robotersystems.

Die erste Komponente hat die Funktion, die Umwelt wahrzunehmen, die zweite Komponente ist für die Planung von Bewegungstrajektorien verantwortlich, und die dritte Komponente soll dann die Steuerung der Freiheitsgrade des Robotersystems übernehmen. Man ging von folgenden Hypothesen aus:

- Die Umwelt ist in ihrer vielfältigen Dynamik explizit modellierbar.
- Die drei funktionellen Komponenten sind streng sequentiell durchführbar.
- Jede Komponente gewährleistet eine universelle Funktionalität, ist also für unterschiedliche Probleme und Situationen geeignet.
- Echtzeitfähiges Verhalten folgt irgendwie aus der Gesamtfunktion und bildet nicht ein primäres Entwurfsprinzip.

Basierend auf diesen Annahmen wurden dann Teilaufgaben aus den drei genannten Bereichen zunächst unabhängig voneinander bearbeitet, mit der Intention, die Abhängigkeiten dann anschließend einzubeziehen. Für die Implementierung der Planungskomponente bedeutet dies beispielsweise, daß zunächst die voraussichtlichen Ergebnisse der Umweltwahrnehmung vorhergesagt werden müssen, weil die Planung signifikant auf Umweltdaten basiert. Das Robotersystem sollte also seine Funktionalität durch die explizite Modellierung der zu interpretierenden und zu manipulierenden Welt erlangen.

Die Umwelt und der Wahrnehmungsprozeß sind aber im allgemeinen so komplex, daß die realen Gegebenheiten nur sehr unvollständig und ungenau modelliert werden können

(Problem der Robustheit). Die Konsequenz davon ist, daß eine auf der Grundlage von expliziten Umweltmodellen implementierte Planungskomponente lediglich in einer stark eingeschränkten Umwelt erfolgreich sein kann. Weiterhin kann für real einsetzbare Robotersysteme auch nicht angenommen werden, daß die Abhängigkeiten zwischen Umweltwahrnehmung, Planung und Robotersteuerung unidirektional sind. Vielmehr liegen komplizierte bidirektionale Wechselwirkungen vor, die sich explizit häufig gar nicht beschreiben lassen. Der sequentielle Ablauf der genannten Funktionen schließt darüber hinaus auch ein echtzeitfähiges Verhalten aus, weil zum Zeitpunkt der Planung schon wieder neue Umweltdaten entstehen, die simultan mit einbezogen werden müßten. Der Ausfall oder Mißerfolg einer der drei Komponenten führt zu einem totalen Verlust der Funktionalität des Gesamtsystems (Problem der Stabilität).

Wegen dieser Nachteile werden seit dem Beginn der neunziger Jahre bevorzugt auch Architekturkonzepte entwickelt, gemäß deren ein Robotersystem jeweils durch ein Repertoire von Verhaltensmodulen aufgebaut ist. Die Prinzipien dieser *verhaltensbasierten Modularisierung* werden in Abschnitt 3.2 erläutert.

2.2.4 Roboterrelevantes Maschinensehen

Die zentrale Problemstellung der visuell basierten Robotik ist das Maschinensehen. In diesem Abschnitt werden somit zunächst allgemein das Maschinensehen erläutert und die damit in Beziehung stehenden Fachgebiete diskutiert. Davon ausgehend läßt sich das konkrete Ziel des roboterrelevanten Maschinensehens formulieren. Der Hauptteil behandelt schließlich anhand eines Beispiels die grundlegenden Prinzipien und abschließend einige aktuelle Weiterentwicklungen.

Maschinensehen und verwandte Fachgebiete

Unter dem allgemeinen Begriff *Maschinensehen* soll die Verfolgung des Zieles verstanden werden, aus Bildsignalen eine Interpretation der abgebildeten dreidimensionalen (3D-) Welt zu liefern. Zur weiteren Klärung des Begriffs müssen hier auch Fachgebiete erwähnt werden, die sich mit dem Maschinensehen überschneiden.

Die *Bildverarbeitung* befaßt sich mit den Transformationen von Bildern in Bilder oder Merkmalsvektoren und bietet nur in den Bereichen der *Bildverbesserung*, *Bildkodierung* und *Bildkompression* Endprodukte an. Der in multimedialen Umgebungen bekannte MPEG-Standard zur Kompression von Videosequenzen ist ein solches Beispiel. Ein Beispiel für Bildverbesserung ist die Rekonstruktion guter Bilder aus stark verrauschten oder verzerrten Bildern wie im Fall der unscharfen Aufnahmen von Kennzeichen-Schildern bei schnell fahrenden Fahrzeugen.

Die Interpretation von Bildern, bekannt auch als *Bildverstehen* oder *Bildauswertung*, ist die zweite Stufe des Maschinensehens. Sie befaßt sich mit der dreidimensionalen Rekonstruktion der abgebildeten Szene und der Erkennung ihrer Bestandteile. Die hierfür erforderliche Objekterkennung setzt Methoden aus der klassischen *Mustererkennung* ein,

die sich mit der Klassifikation von Mustern aus Signalen verschiedener Formen befaßt. Typische Beispiele sind hierfür das Lesen von Strichcodes oder von Maschinenschrift, welche bereits breite Anwendung finden. Die Bildauswertung ist erfolgreich in der Medizin und in den verschiedensten Produktionsbranchen im Einsatz. Ihre Verfahren werden nicht nur auf Videobilder und -bildfolgen, sondern auch auf Infrarot-, Ultraschall-, Röntgen-, Tomographie- und Lasersignale angewendet. Beispielhaft seien hier einige roboterfremde Einsatzgebiete aufgelistet:

1. Die **Überwachung** von Müllverbrennungsanlagen und die Erkennung von wiederverwertbaren Stoffen.
2. Die **Inspektion** von Oberflächen aus Metall, Holz, Textil oder chemischer Substanzen.
3. Die qualitative und quantitative **Qualitätsbewertung** von Leiterplatten, Wafern und Chips.
4. Die **3D-Vermessung** von Werkstücken.
5. Die **Analyse** von Luftbildern in der Fernerkundung und die photogrammetrische Auswertung von Gebäudeabbildungen.

Weil man mit dem Maschinensehen visuelle Fähigkeiten zu schaffen versucht, die den visuellen Fähigkeiten der Lebewesen ähnlich sind, ist Maschinensehen auch zur *Wahrnehmungspsychologie* und *Neurobiologie* verwandt. Beide Disziplinen versuchen, anhand psychologischer Experimente, physiologischer Messungen und bildgebender Verfahren zu analysieren, auf welchen Funktionsprinzipien das Sehen beim Menschen und anderen Lebewesen beruht. Die Plausibilität der Natur kann als Hinweis für den Aufbau eines technischen visuellen Systems dienen. Andererseits können die Berechnungsmodelle, welche für die Konstruktion künstlicher visueller Systeme entwickelt werden, das Verstehen des Sehens der biologischen Systeme unterstützen.

Worin die wesentlichen Mechanismen visueller Wahrnehmung biologischer Systeme tatsächlich beruhen, ist allerdings gegenwärtig nicht geklärt, obwohl in den letzten zehn Jahren hierfür eine Reihe wertvoller Erkenntnisse gesammelt und brauchbare Hypothesen aufgestellt wurden. Demzufolge ist es auch für die Forschungsgemeinde, die sich der Entwicklung technischer sehender Systeme zuwendet, äußerst schwierig, tragfähige Strategien für den Architektorentwurf zu finden. Ein Schwerpunkt der Diskussionen der letzten Jahre gruppiert sich um die Frage, ob visuelle Wahrnehmung die vollständige *Rekonstruktion* der wahrgenommenen Szene bedeutet. Diese Frage kann ohne Bezug zu einer Anwendung nicht sinnvoll beantwortet werden. Das Ziel der visuellen Wahrnehmung wird deshalb nachfolgend für die visuell basierte Robotik diskutiert.

Ziel des roboterrelevanten Maschinensehens

Manche Verfahren des *roboterrelevanten Maschinensehens* (im folgenden kurz *Robotersehen* genannt) basieren im Grunde auf denselben Funktionsprinzipien wie Verfahren des allgemeinen Maschinensehens. In den siebziger und achtziger Jahren wurde als Ziel des allgemeinen Maschinensehens die Ermittlung von vollständigen dreidimensionalen Szenenbeschreibungen gesehen. Nach der vorherrschenden Theorie von David Marr sollte sich der Aufbau eines künstlichen visuellen Systems in drei Stufen vollziehen:

1. Aufstellen einer **Berechnungstheorie**,
2. Umsetzung der Theorie in ein **Verfahren**, d.h. Entwurf von Algorithmen und Datenstrukturen,
3. **Implementierung** des Verfahrens auf einer Architektur.

Weil die zweite und dritte Stufe als reine Anwendungssache betrachtet wurden, konzentrierte sich die Forschung auf allgemeine Theorien einer vollständigen Rekonstruktion.

Im Hinblick auf das Robotersehen bedeutet dies, daß der Roboter jede visuelle Aufgabe erfüllen kann, falls jederzeit eine genaue 3D-Beschreibung der Szene vorhanden ist. So könnte er Hindernisse vermeiden, weil ihr Abstand und ihre 3D-Gestalt von der Szenenbeschreibung bekannt sind. Er könnte auch Werkstücke greifen, weil ihre Lage und Abmessungen in der genauen Szenenrekonstruktion enthalten sind.

Tatsächlich hat sich aber die Rekonstruktion der Welt aus den sensorischen Bilddaten als ein sehr schwieriges Problem erwiesen, weil die Abbildung der Welt auf ein Bild eine nicht invertierbare Abbildung ist. Lediglich durch Einbeziehung zusätzlicher Daten, die durch die Abbildung einer Szene in Stereoaufnahmen oder Bildfolgen bereitgestellt werden und durch Nutzen von zusätzlichem Wissen über die Welt, kann diese Invertierung der Abbildung näherungsweise erfolgen.

Die Forschungen auf dem Gebiet der 3D-Rekonstruktion von Szenen hat das Repertoire an Grundlagenwissen des Robotersehens bereichert. Die Ergebnisse konnten aber kaum Hinweise dafür geben, wie ein künstliches visuelles System für eine Robotikaufgabe gebaut werden soll. Einig ist man sich inzwischen immerhin darin, eine Szenenrekonstruktion nur so detailliert durchzuführen, wie dies die Aufgabe unabdingbar verlangt [Jain und Binford, 1991].

Ziel des roboterrelevanten Maschinensehens ist somit die Entwicklung von technischen Systemen, die in der Lage sind, aus bildgebenden Sensoren diejenige Information über Form und Bewegung der abgebildeten Szenenobjekte zu erfassen, welche für die Ausführung spezifischer Aufträge des Roboters notwendig ist.

Erläuterung des Prinzips an einem Beispiel

Die Prinzipien eines klassischen Ablaufs nach dem wissensbasierten Zugang werden hier an dem Beispiel der Beschreibung der Bewegungen und Formen von Fahrzeugen erläutert (Abbildung 18). Mit einer statischen Kamera wird eine Verkehrskreuzung aufgenommen. Ziel ist es, die Bewegungen und die Formen der Fahrzeuge zu erkennen. Eine ähnliche Problemstellung taucht auch bei der visuellen Navigation auf, falls die aufnehmende Kamera auf einem Fahrzeug montiert ist.

Die Eingabe ist eine Folge von digitalen Grauwertbildern, wie sie von einer Grauwert-Kamera erfaßt und von einem A/D-Wandler digitalisiert werden. Ein digitalisiertes Grauwertbild (links unten in der Abbildung wird ein Ausschnitt dargestellt) ist eine Matrix aus Grauwerten, deren Wertebereich durch ganzzahlige Werte zwischen 0 (dunkel) und 255 (hell) aufgespannt wird. Die Matrix bildet sich aus 512x512 Bildelementen (Pixels) und eine solche Matrix wird 25 mal pro Sekunde gebildet.

In einem Grauwertbild sollen erstens die Bereiche gefunden werden, wo ein Objekt von Interesse — hier Fahrzeug — liegt. Diese Figur-Hintergrund-Diskriminierung gehört zum Problemkreis der sogenannten *Bildsegmentierung*. Im Falle einer Bildfolge bieten sich hierfür Ansätze an, die sich auf eine Bewegungsdetektion stützen. In einer Bildfolge versucht man zuerst, die Verschiebungen der Grauwertstrukturen (siehe Abbildung unten rechts) zu berechnen, den sogenannten *optischen Fluß*. Um ein Objekt zu detektieren, sucht man nach den Regionen, wo diese Verschiebungsvektoren ähnlich sind. Gruppiert man örtlich und zeitlich kohärente Verschiebungen, so erhält man eine Region, die als Objektkandidat bezeichnet wird. Das Bild ist in ein Objekt und den uninteressanten Hintergrund segmentiert.

An diesem Punkt wäre es theoretisch möglich, sowohl die Bewegungsrichtung und die Geschwindigkeit des Objektes als auch die Form des Fahrzeugs allein aus den Verschiebungsvektoren zu schätzen. Da die Verschiebungsvektoren aber nur mit einem nicht vernachlässigbaren Fehler gemessen werden können, erhält man die Bewegung des Fahrzeuges nur mit einem beachtlichen Schätzfehler. Deshalb verzichtet man auf diese Art der Rekonstruktion des Objektes. Statt dessen wird a priori Wissen über die Form der Fahrzeuge hinzugenommen, das sich in weiteren Stufen auch zur Klassifikation des Fahrzeugtyps nutzen läßt.

Für die Anpassung der Projektion des 3D-Drahtmodells eines Fahrzeugs benötigt man die Kanteninformation im Grauwertbild. Dies erfolgt durch die sogenannte *Kantendetektion*, die steile Grauwertübergänge extrahiert. Auf die extrahierten Datenkanten (zweite Reihe unten links) wird ein angenommenes Modell projiziert und angepaßt. Unter den angenommenen Modellen wird das am besten passende ausgewählt. Diese Modellausprägung ermöglicht nicht nur die Klassifikation des Fahrzeugs, sondern auch die stabile Berechnung seiner Lage auf der Straße. Die Berechnung der Bewegungsrichtung und Geschwindigkeit erfolgt durch die zeitliche Fortschreibung des Formmodells mit einem Bewegungsmodell.

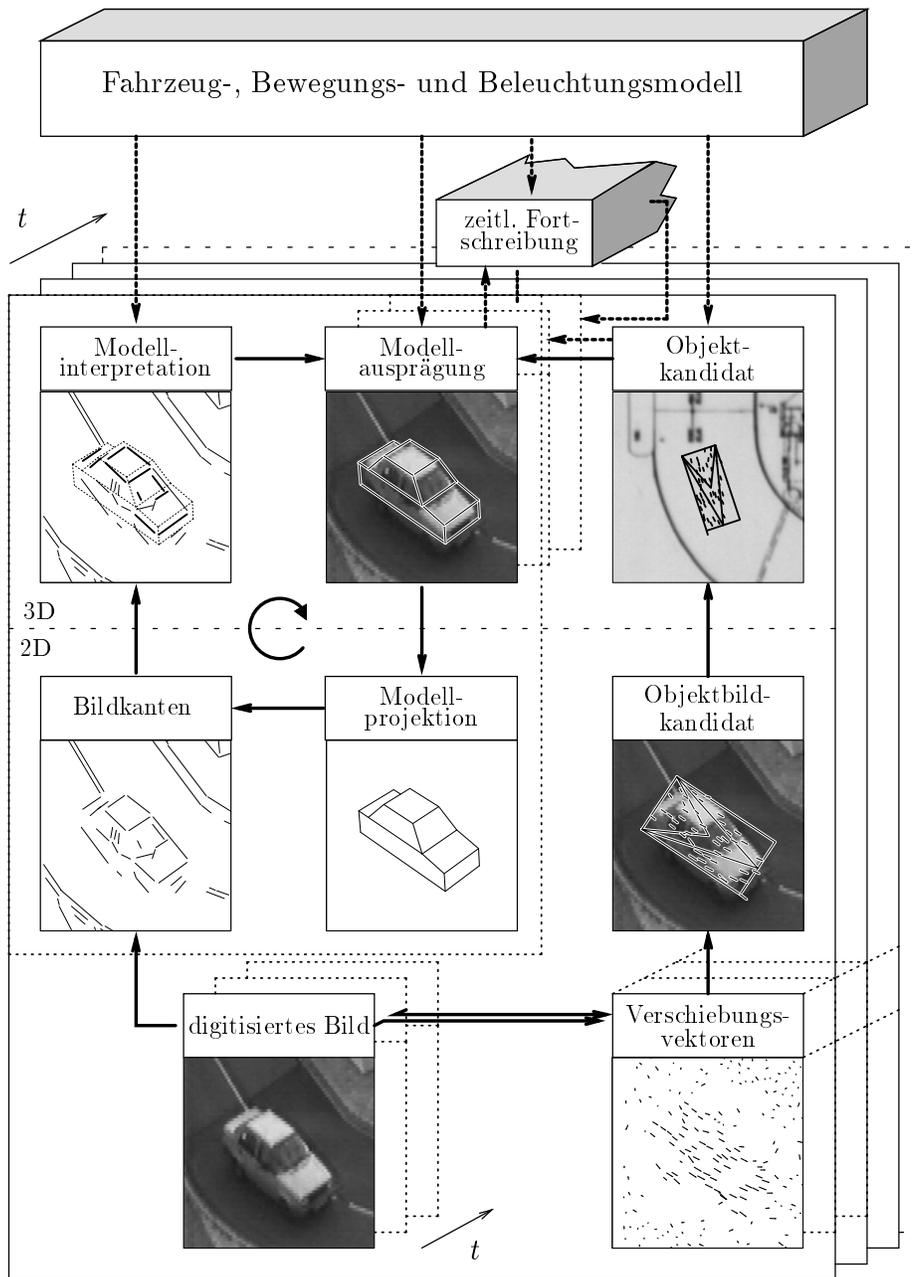


Abbildung 18: Funktionsablauf eines Systems zur Detektion und Verfolgung von Fahrzeugen (aus [Koller, 1992]).

Die Weiterentwicklung der Prinzipien

Das Beispiel der Fahrzeugerkennung zeigt, daß die signalnahe Verarbeitung nur stark verrauschte Fluß- und Kantenbilder liefert. Auf den ersten Blick einfache Probleme wie die Kantendetektion verursachen große Schwierigkeiten in nicht gut strukturierten Umgebungen. Auch der Einsatz eines stereoskopischen Systems zur 3D-Rekonstruktion der Fahrzeugsformen wäre unter solchen Aufnahmebedingungen nutzlos, wenn es keine Modellinformation gäbe. Deshalb ist man gezwungen, Wissen über die Welt einzusetzen. Diese Einschränkung der Freiheitsgrade der Entscheidungen auf eine Menge bekannter Objekte ist mit einem weiteren kritischen Problem verkoppelt, der Recheneffizienz. Die Identifikation der gemessenen Daten mit einem der möglichen Modelle, das *Matchen*, wird als sequentielle Suche im Raum der Modelle realisiert. Dieser Modellraum hat so viele Freiheitsgrade der Variation, wie uns die gleiche Szene unter verschiedenen Bedingungen begegnen kann (Orientierung und Entfernung der Objekte, Beleuchtungsverhältnisse, etc.). Es ist unmöglich, eine Modellbank über alle erwarteten Objektformen zu haben, geschweige denn über die Beleuchtungs- und Bewegungsverhältnisse. Auch im Fall einer eingeschränkten Variation der Umgebung bleiben solche Ansätze extrem rechenintensiv.

Begründet auf der biologischen Evidenz, daß Lebewesen sehr wenig a priori Information über explizite Modelle zur Wahrnehmung der Umwelt einsetzen, wurde ein weiterer Ansatz entwickelt. Für eine stabile und effiziente 3D-Rekonstruktion sollen verschiedene Bildhinweise kombiniert werden.

Hinweise über die Entfernung von Objekten (die Tiefe der Szene), die Orientierung im Raum oder die 3D-Gestalt lassen sich aus einer Reihe unterschiedlicher Effekte bei der zweidimensionalen Abbildung ermitteln:

1. **Stereoskopie:** Aus der Abweichung der Position (*Disparität*) der selben Struktur in der linken und rechten Kamera kann man die Tiefe erschließen, nachdem man die entsprechenden Grauwertstrukturen zugeordnet hat.
2. **Eigenbewegung:** Im Falle einer sich bewegenden Kamera und einer statischen Szene bewegen sich scheinbar alle Objekte relativ zur Kamera, wobei das Ausmaß von der Entfernung zur Kamera abhängt.
3. **Texturgradient:** Das Abbildung einer texturierten Oberfläche hängt vom Neigungswinkel der Oberfläche relativ zur Beobachtungsrichtung ab.
4. **Oberflächenschattierung:** Für eine bestimmte Beleuchtungsgeometrie hängt die Schattierung einer Oberfläche von ihrer 3D-Gestalt ab.

Seit den achtziger Jahren hat man parallel zum modellbasierten Ansatz mathematische Theorien entwickelt, wie diese Hinweise zur Rekonstruktion ausgewertet werden können. Diese Forschung hat zu grundlegenden theoretischen Ergebnissen geführt, auf welche die heutigen Verfahren immer wieder zugreifen.

In der signalnahen Verarbeitung der Bilddaten zeigt sich, daß man markante Grauwertstrukturen nur bei gleichzeitiger Verarbeitung auf verschiedenen Skalen (Maß für die

Detailabstufung von fein zu grob) extrahieren kann. Wendet man Bildoperatoren (Filter) nur auf einer feinen Skala an, so detektiert man auch jedes rauschbedingte Detail. Filtriert man ganz grob, so übersieht man wichtige Hinweise für die Szenenbeschreibung. Eine Kombination der Filterung mit einer Unterabtastung auf groben Skalen führt zu einer Kodierung eines Bildes in Gestalt einer Bildpyramide (Abbildung 19). Diese eher intuitiv getriebene Ansätze finden ihre Fundierung zum einen durch die Theorie der *Wavelets* und zum anderen durch die Theorie der *Diffusionsprozesse*.



Abbildung 19: Bildfilterung auf unterschiedlichen Skalen.

Wavelets [Daubechies, 1992] sind parametrische Familien von auf Bildern angewandten Operatoren, deren Parameter den Aufpunkt, die Skala, und die Orientierung der Funktion angeben. Diese, erst seit etwa zehn Jahren bekannten mathematischen Funktionen, haben außerordentlich interessante Eigenschaften. Ihre Kenntnis hat zu einer Revolution bei der Kodierung von Signalstrukturen geführt. Auf der anderen Seite wurde gezeigt, daß die Anwendung eines bestimmten Operators — die Gaußfunktion — auf das Bild äquivalent zur Lösung der Diffusionsgleichung der Wärmeleitung ist, wobei das Bild die Anfangsbedingung darstellt, und die Zeit der Fein–Grob–Skala entspricht. Um die Glättung von steilen Grauwertübergängen zu vermeiden, wurden Methoden zur anisotropen Diffusion entwickelt, die den Diffusionsprozeß unter den Zwängen der Bildstruktur steuern.

Von großer Bedeutung für das Maschinensehen ist das Prinzip der *Invarianz*. *Photometrische Invarianten* kennzeichnen Signaleigenschaften, die sehr wohl von der spezifischen Gestalt und Oberflächenbeschaffenheit der Objekte abhängen, aber bei Änderung der Beleuchtungsintensität oder der Beleuchtungsgeometrie unverändert bleiben. Zu ihnen findet man Zugang, indem die Projektion eines Signals auf generische Basissysteme berechnet wird (Abbildung 20). Diese Basissysteme sind Klassen von mathematischen Funktionen, die in engem Zusammenhang mit den oben erwähnten Wavelets stehen.



Abbildung 20: Originalbild (links oben), ungerade Basisfunktion (rechts oben), gerade Basisfunktion (links unten), Faltung mit ungerader Basisfunktion (unten rechts).

Im Fall von Stereobildern kennt man eine Vielzahl von Methoden zum Auffinden der Korrespondenzen im linken und rechten Bild. Diese basieren entweder auf der direkten Korrelation der Grauwerte oder auf der Anwendung von Zuordnungsheuristiken. Entsprechende Heuristiken werden auch für die Zuordnung zwischen Daten und Modellteilen beim Matchingprozess eingesetzt, um die kombinatorische Explosion in der Anzahl möglicher Zuordnungen einzuschränken.

Die Berechnung des optischen Flusses aus Bildfolgen stellt sich als ein unterbestimmtes Problem dar, das nur mit Hilfe von a priori Annahmen über seinen Verlauf gelöst werden kann. Diese Einführung von Nebenbedingungen taucht unter dem Begriff der

Regularisierung auf. Ähnlich zur Diffusion führt Regularisierung des optischen Flusses zur Verschmierung von Bewegungsgrenzen.

Bei gegebenen Stereokorrespondenzen oder Verschiebungsvektoren stellt sich die Frage, wie man die Abstände in der Szene triangulieren kann, und im Fall der Bildfolgen, wie die eigene Bewegung aussieht (Abbildung 21). Beide Fragestellungen betreffen direkt die Anwendungsaufgaben von Robotersehen, d.h. die Objekterkennung, die Manipulation und das lokale Manövrieren.

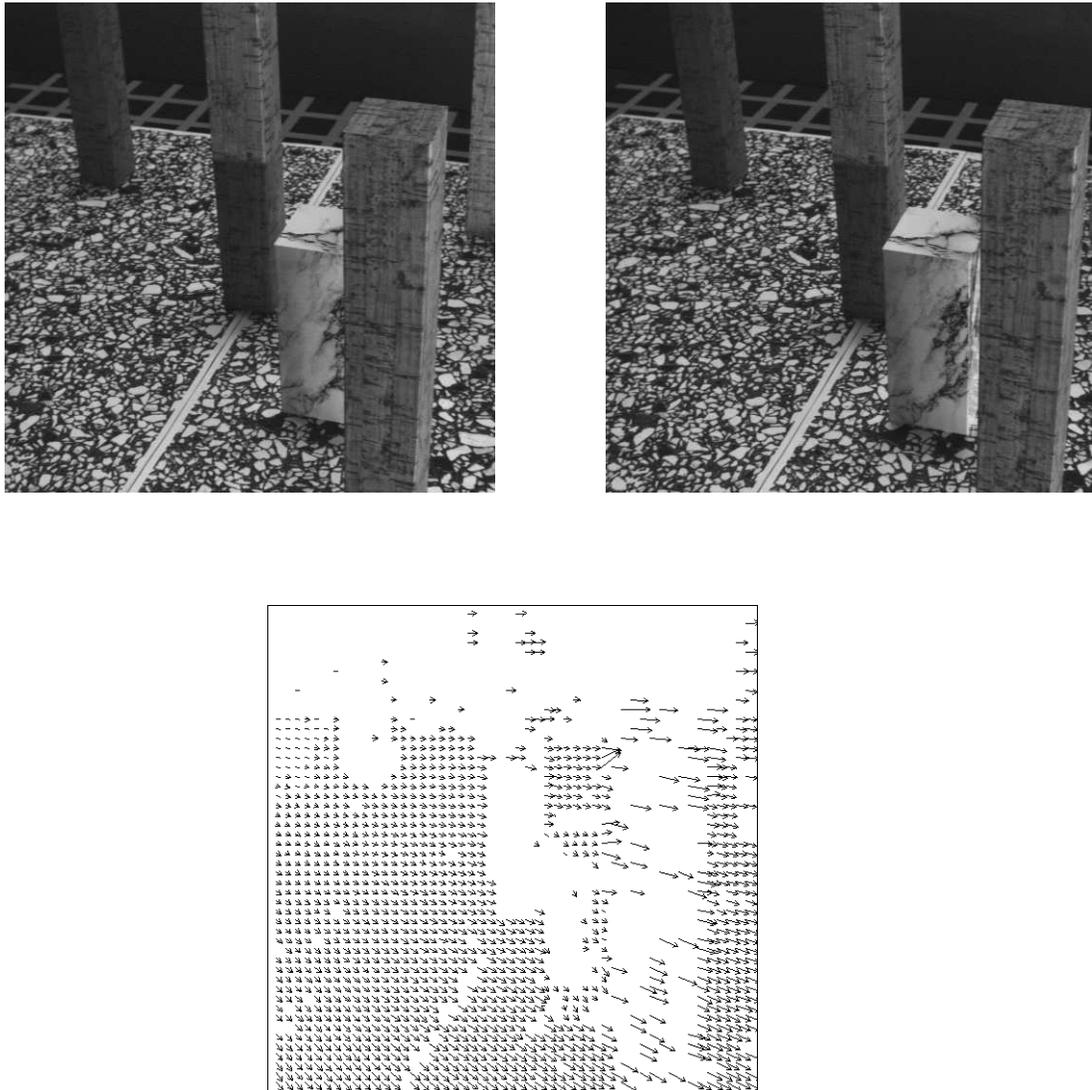


Abbildung 21: Zwei Einzelbilder einer Bildfolge und das Vektorfeld.

Voraussetzung für die Erfüllung metrischer Aufgaben ist eine *Kalibrierung* der Kameras, d.h. die Abbildung einer Abstandseinheit in den Bildern (Pixels) auf Winkeleinheiten im Raum, und die Ermittlung der relativen Lage der Kameras in Stereoanordnungen. Diese Fragen werden durch ein Repertoire von eleganten Lösungen mit Hilfe der *projektiven Geometrie* gelöst, deren Wurzeln in der photogrammetrischen Forschung der ersten Hälfte dieses Jahrhunderts liegen.

Eine bedeutende Rolle bei der Beschreibung der geometrischen Eigenschaften von abgebildeten Objekten spielt die *geometrische Invarianz*. Eine projektiv-geometrische Invariante ist eine objekt-charakteristische Kennzahl, die unter der Wirkung eines Projektionsoperators erhalten bleibt. Kennt man diese Kennzahl eines Objektes, so braucht man nur diese Größe aus den Bilddaten zu berechnen, um das Objekt zu identifizieren. Die einzelne Anpassung von Modellteilen und Datenhinweisen wie im oben beschriebenen Fallbeispiel wird vermieden. Ein Katalog von Objektmodellen kann mit Hilfe solcher Invarianter indexiert werden, so daß die Bilderkennung erheblich effizienter wird. Weiterhin erlaubt die geometrische Invarianz die qualitative Rekonstruktion von Szenen unter Beibehaltung der Tiefenreihenfolge, was den Schritt der metrischen Kalibrierung erspart und Schätzfehler in der Rekonstruktion reduziert.

Damit soll die Beschreibung der bis heute wichtigsten Grundlagen des Robotersehens abgeschlossen werden. Es gibt sicher eine Anzahl bedeutsamer weiterer Punkte, auf deren Darstellung aus Platzgründen verzichtet wird. Der Leser wird schon festgestellt haben, daß bis jetzt auf den letzten Nebensatz der Definition des Robotersehens ("*..., die für die Ausführung spezifischer Aufträge vom Roboter notwendig ist.*") überhaupt nicht eingegangen wurde. Das ist auch charakteristisch für den Entwicklungsstand Anfang der neunziger Jahre. In Abschnitt 2.3.2 wird hierzu der Entwicklungsstand bewertet, manche Andeutungen wurden schon hier gemacht. Darauf folgend wird in Abschnitt 3.3.1 eine Perspektive für das Robotersehen dargestellt.

2.2.5 Roboterrelevantes Maschinenlernen

Als zentraler Bestandteil der visuell basierten Robotik wurde oben das Maschinensehen erklärt. Die zweite Grundlage der neuen Roboter-Technologie ist das Maschinenlernen.

Ziel des roboterrelevanten Maschinenlernens

Im Vorfeld einer Beschäftigung mit dem Maschinenlernen in der Robotik (im folgenden kurz *Roboterlernen* genannt) ist zu klären, welche potentiellen Vorteile Lernen gegenüber einer Vorkodierung von Problemlösestrategien bieten kann. Ein Blick in die Biologie verrät [Brooks und Mataric, 1993], daß Lernen dort im wesentlichen zwei Funktionen erfüllt:

1. die Möglichkeit der **Anpassung an unterschiedliche Umgebungen** und
2. eine **Reduzierung des genetischen Materials**, welches zur Kodierung eines funktionsfähigen Lebewesens sonst notwendig wäre.

Hierzu ist zu bemerken, daß Lernen eine spezifische Form des Erwerbs von Kompetenz durch ein Individuum darstellt. Es befähigt ein biologisches System, in spezifischen Situationen adäquates Verhalten zu zeigen. Diese Form des individuellen Kompetenzerwerbs ist oberhalb einer kritischen evolutionären Entwicklungsstufe nötig, wo zum Ausführen von Handlungen nur noch Teile des Körpers eingesetzt werden müssen.

Bei niederen Entwicklungsstufen ist der Kode weniger umfangreich. Demzufolge dominiert die genetische Kodierung von Kompetenz und deren Vererbung. Dies dient weniger dem Überleben des Individuums als vielmehr der Existenz der Art. Die Auseinandersetzung dieser biologischen Systeme (z.B. der Insekten) mit ihrer Umwelt erfordert den ganzen Körper und ist weit undifferenzierter als diejenige höher entwickelter biologischer Systeme (z.B. Wirbeltiere).

Das zweite Argument hängt also auch eng mit der im Abschnitt 2.2.4 beschriebenen Problematik zusammen, daß die genetische Kodierung von Problemlösungen einen über alle Grenzen wachsenden Aufwand erfordern müßte, um allen möglichen Situationen der Auseinandersetzung des biologischen Systems mit seiner Umwelt zu entsprechen.

Hieraus kann für die Rolle des Lernens in der Robotik folgendes Resümee gezogen werden.

Lernen in der Robotik macht dann Sinn, wenn ein Roboter sich ändernden externen oder internen Umständen anpassen oder sogar gänzlich unvorhergesehene Aufgaben lösen können muß.

Hierzu kann bereits die Aufgabe gehören, das Greifen eines ungleichförmigen Objektes in unbestimmter Lage zu realisieren. Viele Roboteraufgaben, die visuelle Wahrnehmung erfordern, sind von einer Komplexität, welche die Fähigkeit zum Lernen voraussetzt.

Grundlagen künstlicher neuronale Netze

Voraussetzung für solches Lernen ist ein leistungsfähiges Spektrum an Repräsentationsformen, um der Vielfalt des zu lernenden Wissens gerecht werden zu können.

Klassische, symbolische KI-Systeme bewegen sich innerhalb fest vorgegebener *Repräsentations schemata*, beispielsweise sogenannte

- **Frames** oder
- **Semantische Netze.**

Lernen beschränkt sich dabei oft nur auf das Umordnen bekannten Wissens, wie etwa beim sogenannten *Erklärungsbasierten Lernen*. Die meisten der interessanten Roboteraufgaben lassen sich aber wegen ihrer Komplexität nicht explizit symbolisch in derartigen Schemata kodieren. Statt dessen bieten sich *implizite Repräsentationsformen* des Wissens an, wie sie von biologischen Systemen genutzt werden.

Künstliche Neuronale Netze (KNN) sind weitgehend abstrahierende Modelle biologischer neuronaler Systeme. Sie eignen sich für die Repräsentation komplexer Sachverhalte und für die Modellierung des Erwerbs von Wissen durch Lernen. Dies geschieht durch dynamische Entfaltung des Netzes und Veränderung seiner Parameter.

Der Zwang, das Wissen explizit repräsentieren zu müssen, entfällt dadurch, daß es implizit in Form von Verbindungsgewichten, die zwischen den Neuronen über das ganze Netz verteilt sind, kodiert wird. Die KNN versprechen eine Lösung des grundlegenden Problems symbolischer KI-Systeme, die sogenannte *Symbolverankerung* [Neumann, 1995]. Hierunter versteht man die Schwierigkeit, den Elementen eines algorithmischen Verfahrens (den Symbolen) eine Bindung zur Welt zu geben.

Mit Künstlichen Neuronalen Netzen können erstmals Systeme gebaut werden, die lernen, Symbole mit Repräsentationen zu verbinden, die in den eigenen Erfahrungen verankert sind.

Dies ist Ausdruck einer “praktischen Intelligenz” oder Kompetenz. Die Bedeutung der Symbole wird hierbei in Bezug auf die eigene Erfahrungsmöglichkeit erfaßt. Dieses Potential beruht auf der Fähigkeit von KNN, sensorische Eingangssignale mit nichtsensorischen Aktivierungen zu verbinden, Aktivierungsmuster zu kategorisieren und Verbindungen abhängig von der Erfahrung des Systems adaptieren zu können. Wissen wird in KNN nicht als objektiv festlegbare Struktur angesehen, sondern als Struktur, die sowohl von den konkreten Wissensinhalten als auch von der Architektur der KNN abhängt. Damit bieten KNN den besten Ansatz zu roboterrelevantem Maschinenlernen im Sinne des vorhergehenden Abschnitts.

KNN bestehen aus

- einer beziehungsweise mehreren Gruppen von Knoten und
- gewichteten Verbindungen, welche die Knoten miteinander verknüpfen.

Die Knoten werden synonym auch als Einheiten oder Verarbeitungselemente bezeichnet.

Unmittelbaren biologischen Bezug erhält diese Struktur durch die folgenden Äquivalenzen:

- Ein **Knoten** steht für ein *Neuron (Nervenzelle)*, das Signale empfängt und neue von ihm erzeugte Signale weiterleitet.
- Die **Verbindungen** simulieren die Neuronen verbindende *Axone*.
- Die **Gewichte** der Verbindungen beschreiben die Kontaktstellen an den Neuronen, die sogenannten *Synapsen*.

Ein KNN kann als Berechnungsvorschrift interpretiert werden, die ein bestimmtes Eingabemuster auf ein Ausgabemuster abbildet (Abbildung 22). Hierzu werden bestimmte Knoten als *Eingabeknoten* beziehungsweise *Ausgabeknoten* definiert, die übrigen sind die *verdeckten Knoten*. Ein wesentlicher Unterschied zwischen KNN und herkömmlichen Algorithmen ist jedoch, daß es relativ einfacher ist, für neuronale Netze ein geeignetes Lernverfahren anzugeben. Deshalb muß nicht für jedes einzelne Problem eine spezielle Rechenvorschrift implementiert zu werden. Statt dessen findet das Netz selbständig in einem Lernprozeß die zur Lösung des Problems richtige Netz-Konfiguration. Der Lernvorgang eines neuronalen Netzes besteht darin, daß einerseits die Verbindungsgewichte adaptiert werden und andererseits neue Knoten mit Verbindungen zu den alten Strukturen ergänzt beziehungsweise vorhandene Teilstrukturen gelöscht werden.

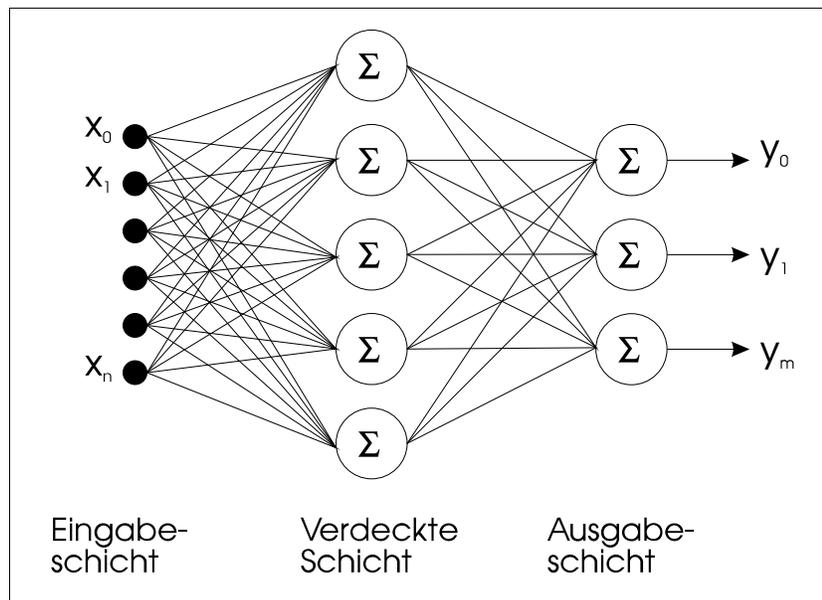


Abbildung 22: KNN als Berechnungsvorschrift.

Lernverfahren für KNN unterscheiden sich nach der Art des zum Lernen genutzten *Fehlersignals*. Man unterscheidet folgende Fälle der Verwendung des Fehlersignals:

- **Quantitatives Fehlersignal,**
- **Qualitatives Fehlersignal** oder
- **Kein Fehlersignal.**

Zu einem bestimmten Eingabemuster berechnet das KNN ein Ausgabemuster. Die Differenz dieses Musters zu einem gewünschten Ausgabemuster ist das Fehlersignal.

Ist dieses Fehlersignal quantitativ verfügbar, so spricht man von

- **Fehlerminimierungsverfahren** (Abbildung 23a).

Ein Beispiel hierfür ist die wichtige *Back-Propagation-Regel*, die lediglich die Verbindungsgewichte adaptiert. Hierbei wird der Fehler rückwärts durch das Netz weitergereicht und die Gewichtung der Verbindungen sukzessiv so modifiziert, daß unter dieser neuen Belegung die Verarbeitung des Eingabemusters zu einem geringeren Fehler führt.

Steht hingegen nur ein qualitatives Fehlersignal zur Verfügung, wie Erfolg oder Mißerfolg, so wird ein sogenanntes

- **Reinforcement-Lernverfahren** (Abbildung 23b)

benötigt. Dieses kann beispielsweise für das Lernen von hindernisvermeidenden Roboterbewegungen herangezogen werden. Durch fortwährendes Agieren in der Umwelt erhält das Robotersystem ständig Belohnungen oder Bestrafungen und erwirbt darauf basierend schrittweise eine optimale Bewegungsstrategie.

Falls gar kein Fehlersignal vorliegt, so kann eine Adaption nur das Ziel haben, in einer Menge von Eingabemustern bestimmte Gruppen oder Korrelationen zu erkennen (z.B. durch Energieminimierung). Man spricht dann von

- **unüberwachtem Lernen** (Abbildung 23c).

Ein Beispiel hierfür ist das sogenannte *Wettbewerbs-Lernen*. Das Eingabemuster wird zu den Ausgabeknoten vorwärts geleitet, und diese konkurrieren dann miteinander bis ein Ausgabeknoten den Wettbewerb gewinnt. Dies wird beispielsweise im Fall der *Hebb-Lernregel* dadurch erreicht, daß bei benachbarten Knoten, die ähnlich aktiv sind, die siegreiche Verbindung ein höheres Gewicht erhält. Der aktive Ausgabeknoten repräsentiert dann eine bestimmte Gruppenzugehörigkeit des Eingabemusters.

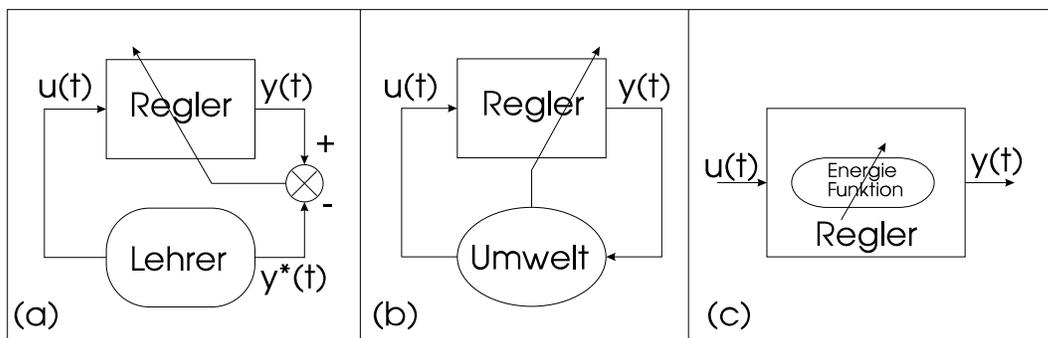


Abbildung 23: (a) Fehlerminimierungsverfahren, (b) Reinforcement-Lernverfahren, (c) unüberwachtes Lernen.

Im Zusammenhang mit neuronaler Regelung von Robotern sind insbesondere die ersten beiden Lernverfahren von Interesse.

Neuronale Ansätze des Roboterlernens

Ein besonders wichtiger Teilbereich des Roboterlernens ist die *Adaptive Regelung*. Klassische Regelungsmethoden benutzen ein Referenzmodell, dessen Diskrepanz zu dem wirklichen System zu großen Fehlern führen kann. Gängige Modellschätzungsverfahren gehen von einer bekannten Struktur des Systems aus und beschränken sich auf Parameteradaptation. Adaptive Regelungsverfahren hingegen sind nicht-modellbasierte Verfahren. Sie waren bislang nur im Falle linearer Systeme erfolgreich. Auch hier eröffnen sich mit dem Einsatz von KNN neue Perspektiven, da KNN als nichtlineare *Funktionsapproximatoren* besonders zur adaptiven nichtlinearen Regelung geeignet sind.

Neuronale adaptive Regler beruhen auf einem der drei folgenden Funktionsprinzipien, die hier als Schlagworte genannt seien:

- **Direkte inverse Modellierung,**
- **Vorwärtsmodellierung** oder
- **Fehlerrückkopplung.**

In allen drei Fällen ist das Ziel, ein inverses Modell des zu regelnden Prozesses zu erlernen (Abbildung 24). Ist das inverse Modell gelernt, so wird es mit der Eingabe des gewünschten Sollwertes diejenige Steuergröße erzeugen, durch die das zu regelnde System eben diesen Sollwert annimmt. *Direkte inverse Modellierung* (Abbildung 24a) bietet sich dann an, falls die zu lernende Abbildung einfach abzutasten ist und diese das zu regelnde System eindeutig abbildet. Im Falle von nichteindeutigen Abbildungen des zu regelnden Systems erweist sich die *Vorwärtsmodellierung* (Abbildung 24b) als sinnvoll. Steht schließlich ein Maß des Regelungsfehlers selbst zur Verfügung, d.h. die Abweichung zwischen Soll- und Istwert, so ist *Fehlerrückkopplung* (Abbildung 24c) die beste Wahl. Letztere scheint gerade in biologischen Systemen eine dominante Rolle zu spielen.

Beispiele für *nichtlineare sensomotorische Abbildungen*, die nach den oben angesprochenen Prinzipien der inversen Modellierung mit KNN erfolgreich gelernt werden, sind:

- die **inverse Kinematik von Roboterarmen,**
- die **inverse Dynamik von Roboterarmen,**
- die **Kraft–Motor–Abbildungen** und
- die **Ultraschall/Infrarot–Motor–Abbildungen,** sowie ganz allgemein
- die **visuomotorische Abbildungen.**

Dabei beschreibt die *inverse Kinematik* eines Roboterarms das Problem, zu einer vorgegebenen Position im kartesischen Raum die erforderlichen Gelenkwinkel zu finden. Bei der *inversen Dynamik* müssen zu einer vorgegebenen Trajektorie die Kräfte und Drehmomente generiert werden, die den Arm der Trajektorie folgen lassen. Beim Lernen von

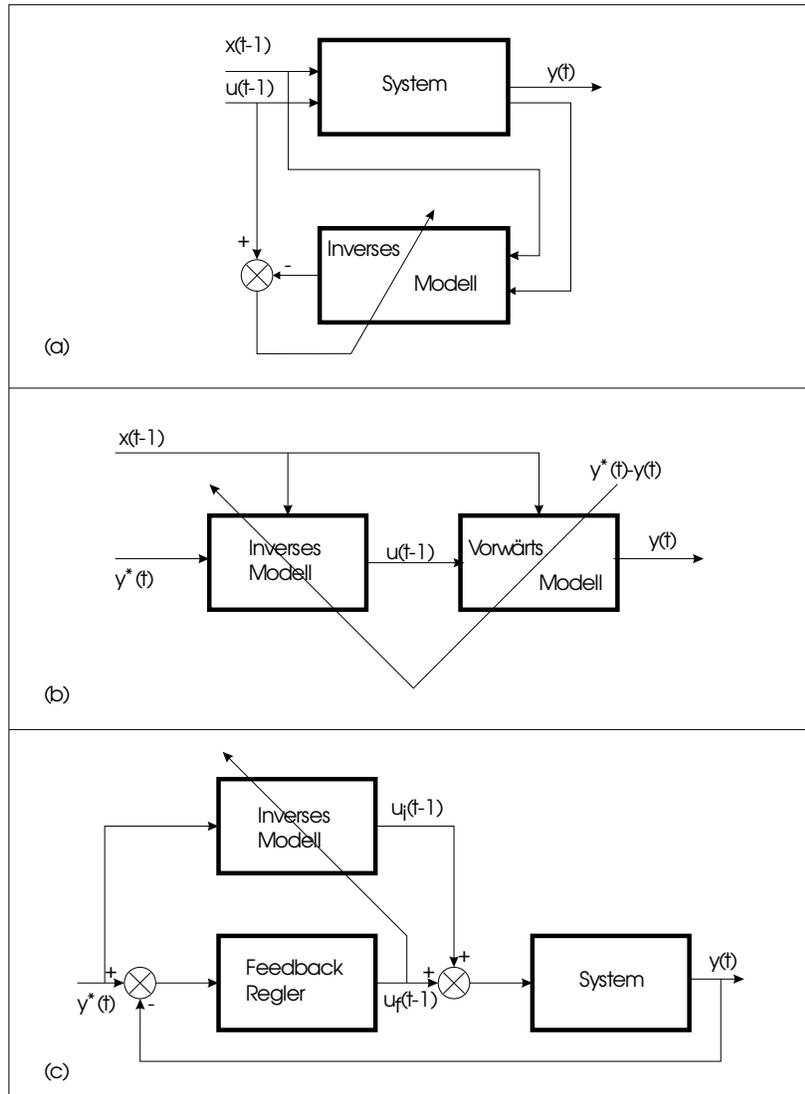


Abbildung 24: (a) Direkte inverse Modellierung, (b) Vorwärtsmodellierung, (c) Fehler rückkopplung (teilweise übernommen aus [Torrás, 1995]).

Kraft-Motor-Abbildungen und *Ultraschall/Infrarot-Motor-Abbildungen* werden motorische Ausgangsgrößen direkt aus den Sensorsignalen abgeleitet, mit dem Ziel, sinnvolle, zielorientierte Aktionen in der Umwelt zu verrichten. *Visuomotorischen Abbildungen* dienen dazu, motorische Aktionen des Systems direkt mit Bildinformationen zu assoziieren, ohne dabei eine modellbasierte dreidimensionale Rekonstruktion durchführen zu müssen.

Für elementare Tätigkeiten von Robotern, beispielsweise das Vermeiden von Hindernissen beim Umherfahren, ist die exakte Form von Gegenständen, etwa der Hindernisobjekte, irrelevant. Sensomotorische Abbildungen sind bezüglich Echtzeitverhalten und Robustheit gegenüber Modellierungsfehlern den modellbasierten Rekonstruktionsverfahren überlegen. Allerdings sind sie im allgemeinen aufgrund hoher Dimensionalität des Eingaberaums und hochgradiger Nichtlinearität analytisch schwer beschreibbar. Daher ist es lohnenswert, lernende KNN gerade in diesem Bereich einzusetzen.

Erläuterung des Prinzips an einem Beispiel

Ein repräsentatives Beispiel für das Erlernen einer visuomotorischen Abbildung sind die Versuche von [Ritter *und andere*, 1991] zum Erlernen der Hand–Auge–Koordination eines Industrie–Roboterarms und eines den Arbeitsbereich beobachtenden Stereokamerasystems. In diesen Versuchen lernt das System allein aus der Beobachtung des Roboterarmmanipulators durch das Kamerasystem, einen vierdimensionalen Vektor — der die kartesische Position der Projektion eines Punktes im 3D–Arbeitsbereich auf die Bildebenen der beiden Kameras repräsentiert — mit dem zum Anfahren dieses Punktes nötigen Gelenkstellwinkeln zu assoziieren.

Das Lernverfahren beruht auf dem Prinzip der Fehlerrückkopplung. Das System benutzt den beim Anfahren zufälliger Zielpunkte wahrgenommenen Fehler in Kamera–Bildkoordinaten, um sich zu verbessern. Zur Fehlerminimierung wird die *Least–Mean–Square (LMS) Methode* eingesetzt, gleichzeitig organisieren sich die verwendeten neuronalen Einheiten auf einem dreidimensionalen Gitter nach der *selbstorganisierenden Lernregel* von T. Kohonen.

Trotz fehlenden Vorwissens in Form eines Roboterarmmodells oder der Abbildungsgleichungen der Kameras und trotz ungünstiger Randbedingungen, wie zufälliger Initialisierung des Neuronalen Netzes und zufälligem Herumstochern im Arbeitsraum, lernt das System in deutlich weniger als 10000 Bewegungen, mit Subpixel–Genauigkeit Punkte im Arbeitsraum anzufahren.

Desweiteren wurde in diesen Versuchen — die für viele weitergehende Experimente anderer Forschungsgruppen als Grundlage dienen — gezeigt, daß das System in der Lage ist, sich schnell an Veränderungen der Kinematik des Roboterarms, z.B. einer Verkürzung eines Armsegmentes, anzupassen. Damit ist dieses Verfahren zum Einsatz in Umgebungen prädestiniert, in denen häufige Änderungen der Betriebsparameter mehrmaliges Rekalibrieren erfordern würden — das System kalibriert sich selbst — oder eine externe Rekalibrierung aufgrund der Betriebsbedingungen (z.B. Unterwasser) ganz ausgeschlossen ist.

Im Unterschied dazu bildet der klassische Ansatz zuerst die Bildkoordinaten des Punktes im Arbeitsraum auf seine 3D–Koordinaten ab. Hierzu notwendig sind Kenntnisse über die genaue Position und Orientierung der Kameras sowie deren Abbildungseigenschaften (Modell der Kamera). Die 3D–Koordinaten werden schließlich über ein Modell der inversen Kinematik des Roboterarms in Gelenkwinkelstellungen umgerechnet. Sämtliche Fehler im Kamera- und Robotermodell gehen in die Ansteuerung des Punktes ein, und unter veränderten Betriebsbedingungen ist eine Neubestimmung der Modellparameter (Rekalibrierung) unumgänglich.

2.3 Bewertung des Entwicklungsstandes

Nachdem die Funktionsprinzipien der visuell basierten Robotik vorgestellt wurden, erfolgt nun eine Einschätzung des derzeitigen Entwicklungsstandes. Dabei wird zuerst übergreifend auf die visuell basierte Robotik im allgemeinen eingegangen und anschließend auf das Robotersehen sowie das Roboterlernen im besonderen.

2.3.1 Entwicklungsstand der visuell basierten Robotik

Als Einsatzfelder für Roboter sind natürliche Umgebungen, die nicht an den Roboter angepaßt sind, und künstliche, angepaßte Umgebungen zu unterscheiden. Die großen Automatisierungserfolge der letzten Jahre beschränken sich dabei auf die künstlichen Umgebungen. In der industriellen Großserienfertigung kommen üblicherweise Robotersysteme zum Einsatz, die visuelle Umweltinformationen nicht benötigen, weil zuvor die Umgebung exakt strukturiert wurde. Die Roboter benötigen nur ein Minimum an nichtvisueller, sensorischer Rückkopplung. In diesem Bereich wird es auch in Zukunft nur geringe Chancen und Notwendigkeiten für visuell basierte, autonome Roboter geben, da die Anpassung der Umgebung noch lange die kostengünstigere und effizientere Alternative sein wird.

Das Anwendungsspektrum der *blinden Roboter* bei der Großserienfertigung ist sehr eingeschränkt. Insbesondere in kleinen und mittelständigen Unternehmen ist statt dessen ein hoher Grad an Flexibilität erforderlich. Wünschenswert ist der Einsatz von visuell basierten, autonomen Robotersystemen in natürlichen Szenen. Zu diesen zählen insbesondere die natürliche Außenwelt oder belebte Umgebungen, die primär auf die Bedürfnisse der Menschen, nicht aber auf Randbedingungen für den Robotereinsatz abgestimmt sind. Von diesen Robotern wird ein hohes Maß an Autonomie gefordert, denn sie müssen jederzeit auf unerwartete Situationen reagieren können, beispielsweise bei drohenden Kollisionen mit Menschen. Beim Einsatz in der natürlichen Außenwelt stellen ständig wechselnde Situationen den Normalfall dar. Hierbei müssen die Roboter durch andauerndes Überprüfen und Korrigieren ihrer Position, Orientierung und Bewegungsgeschwindigkeit relativ zu Objekten der Umgebung reagieren können. Außerdem muß der Eigenbewegung von Zielobjekten und dem Zwang zu möglichen Änderungen der Bewegungstrajektorien zum Erreichen des Zielobjektes Rechnung getragen werden.

Für den Einsatz in natürlichen Umgebungen befinden sich Produktions- und Service-Roboter noch zwischen Entwicklungs- und Prototypenphase. Vielversprechende Ansätze, die sich an der Schwelle zur Produktreife befinden, gibt es in teilstrukturierten innenräumlichen Umgebungen, beispielsweise Bürogebäuden oder Krankenhäusern.

Eher im Bereich der Grundlagenforschung muß man Roboter ansiedeln, die in natürlichen, außerräumlichen Umgebungen eingesetzt werden sollen. Ein wichtiger Erfolg wurde jedoch bei der autonomen Fahrzeugführung und der Fahrerunterstützung erzielt. Die Kooperation der Automobilindustrie mit Forschungsinstituten im Rahmen des europäischen PROMETHEUS-Projekts von 1986 bis 1993 hat zu prototypischen Implementierungen geführt, die erfolgreich demonstriert worden sind. Es wurden Geschwindigkeiten von über

100km/h beim autonomen Fahren auf der Autobahn erreicht [Dickmanns und Graefe, 1988]. Allerdings haben diese Systeme noch nicht das erforderliche Maß an Zuverlässigkeit erreicht, um als vollautonome Fahrzeuge tauglich zu sein. Aber immerhin könnten sie demnächst als sogenannte *maschinelle Kopiloten* zum Einsatz kommen.

Zu diesem Erfolg hat wesentlich die gute Strukturierung der Autobahnen in Europa beigetragen. Dadurch war es möglich, Modelle über den Straßenverlauf, die Verkehrsschilder und die Geometrie der Kreuzungen einzusetzen. Ein weiterer wichtiger Punkt war die Eingliederung des Maschinensehens in den Regelkreis der Fahrsteuerung. Zusätzlich zu den odometrischen Sensoren wurde die Abweichung zwischen gemessener und erwarteter Bildposition der Straßenberandung zur Quer- und Längsregelung ausgenutzt. Ansätze zur Berechnung des Abstands zu den anderen Fahrzeugen und zur Vorausplanung von Überholmanövern hat man in München, Karlsruhe, Bochum und Paris entwickelt [Nagel, 1995].

Während in Europa das Gewicht der Entwicklung autonomer Fahrzeuge im Autobahn- und Stadtverkehr lag, arbeiteten in Amerika während desselben Zeitraums mehrere Gruppen an militärisch orientierten Anwendungen, die das Navigieren auf Landstraßen und Feldern gestatten.

Die hier erwähnten Entwicklungen betreffen eher den Aufgabenbereich des lokalen Manövrierens. Systeme zur globalen Navigation wie der Travel-Pilot der Robert-Bosch GmbH, die bis jetzt mit Hilfe der Odometrie und der satellitengestützten Ortung (GPS) arbeiten, könnten zukünftig auch von der visuellen Erkennung der Ortsbeschilderung auf Straßen profitieren.

2.3.2 Entwicklungsstand des Robotersehens

In innenräumlichen Umgebungen sind beim Robotersehen enorme Fortschritte festzustellen. Die Verfahren können zunehmend auch schwierige Szenen erkennen. Deshalb sind beispielsweise Manipulatoren auch in unstrukturierten Laborumgebungen imstande, einfache bewegliche Objekte mit dem Effektor zu verfolgen und zu greifen. Mit *Differenzbildtechniken* oder durch *Berechnung des optischen Flusses* können Objekte und Objektbewegungen in der Bildfolge erkannt und bei Bedarf dreidimensional rekonstruiert werden.

In außenräumlichen Umgebungen sind die Probleme ungleich größer, weil natürliche Gegebenheiten, wie veränderliche Helligkeit und Witterungen oder wechselnde Objekt-Hintergrund-Relationen inhärent vorhanden sind. Wegen der hier anzutreffenden Komplexität bereitet es gegenwärtig noch Schwierigkeiten, aus dem Bildmaterial zuverlässige Szeneninformation zu extrahieren. Die Forschungen zur Ermittlung von Szenenbeschreibungen hat bislang zu keinem Durchbruch in den Anwendungen geführt. Aber neben der Konsolidierung der mathematischen Grundlagen wurden immerhin einige positive Nebeneffekte erzielt. Methoden der signalnahen Verarbeitung, die von Forschungsgruppen des Robotersehens entwickelt worden sind, fanden Anwendung in verschiedenen Bereichen, von der Qualitätsüberprüfung bis zur Bildkodierung in der Nachrichtentechnik.

Der momentane Stand des Maschinensehens erlaubt den Einsatz von Robotern nur in genau modellierbaren Umgebungen. Die modellbasierten Ansätze hat man überwiegend dadurch vorangebracht, daß in den letzten 15 Jahren die geometrischen Abbildungsprozesse gut beschreibbar wurden. Außerdem wird inzwischen formal gut verstanden, warum die Einführung von Modellwissen über die Welt oder die künstliche Eliminierung von Freiheitsgraden zu stabilen und einsatzfähigen Systemen führt. Ein Werkstück auf einem Fließband, das zwangsmäßig flach liegt, reduziert die Dimension des Problems und entsprechend die Wahrscheinlichkeit von Schätzfehlern und Fehldetektionen. Steuerbare Beleuchtungen erleichtern die signalnahe Kantendetektion. So werden neuerdings zusammen mit Roboter manipulatoren Sichtsysteme angeboten, die vordefinierte Werkstücksmodelle erkennen und ihre Lage berechnen.

Viele Einsatzgebiete unterliegen jedoch nicht der vollständigen Modellierbarkeit, auch wenn prototypische Vorführungen möglich sind (siehe Abschnitt 2.4). Lebensgefährliche Umgebungen, wie Nuklearanlagen oder Situationen, wo die zu manipulierenden Objekte nicht mehr ihre Ursprungsform besitzen, erfordern künstliche visuelle Systeme, die mit einem Minimum von a priori Wissen Aufgaben erledigen können. Deshalb gewinnen zunehmend Verfahren des Maschinensehens an Bedeutung, die unter Verzicht auf die Rekonstruktion als Voraussetzung zum Erkennen der Objekte nur ein Minimum an explizitem Modellwissen erfordern. Worin dieses Minimum besteht, kann nur im Kontext der Zielstellung des Systems entschieden werden. Dieses Wissen muß nicht notwendigerweise explizit vorgegeben sein. Vielmehr wurden Verfahren entwickelt (insbesondere im Bereich Robotersehen), wie derartiges Wissen implizit durch neuronales Lernen erworben werden kann. Dieses *Appearance-Wissen* kann offline gelernt werden, um es zu schneller Objekterkennung anwenden zu können.

Durch Verzicht auf die vollständige Rekonstruktion und durch die Anwendung implizit repräsentierten Wissens über mögliche Situationen werden drei bedeutsame Effekte erzielt:

1. Der Widerspruch zwischen der Zeit zur Objekt- beziehungsweise Situationserkennung und der geforderten Reaktionszeit des Systems wird erheblich reduziert.
2. Die interpolierende Wirkung von KNN vermag neue Situationen aus trainierten Situationen zu rekonstruieren.
3. Der Verzicht auf explizite Modellierung des Situationsraumes hebt die obere Grenze der Komplexität möglicher Situationen im Prinzip auf.

Damit wird ein Weg der Verschmelzung von Maschinensehen und Neuroinformatik gewiesen, der von weitreichender Bedeutung für die Entwicklung künftiger Roboter-Generationen sein wird.

2.3.3 Entwicklungsstand des Roboterlernens

In der Industrie sind auf Künstlichen Neuronalen Netzen basierende Problemlösungen im prototypischen Entwicklungsstadium. Insbesondere in Japan und den USA haben sie die

Schwelle zur Anwendung teilweise schon überschritten. Speziell in Deutschland werden insbesondere bei den großen Industrie-Konzernen, wie Daimler-Benz oder Siemens, KNN für zahlreiche technische Anwendungen erprobt und teilweise eingesetzt. Insbesondere in Süddeutschland gibt es aber auch eine Reihe von kleinen und mittelständigen Unternehmen, die ihre sensor- und visuell basierten Meß- und Prüfaufgaben (außerhalb der Robotik) durch den Einsatz von KNN realisieren. Da bei der visuell basierten Robotik ein breites Spektrum von Fachgebieten ineinander übergreifen (siehe erstes Kapitel), liegt es nahe, KNN gerade auch hier anzuwenden.

Für Deutschland sei hier repräsentativ der wissenschaftlicher Leiter des Arbeitsgebiets *Parallelverarbeitung und Neuronale Netze* bei der Daimler-Benz Forschung und Systemtechnik, Frieder Lohnert, aus seinem Beitrag in [Moeller und Pass, 1994] zitiert:

KNN bieten ein breites Anwendungs- und Innovationspotential in der Robotik. Ansätze wie inverse Kinematik/Dynamik sowie sensorbasierte Steuerung mit KNN werden in Forschungs- und Entwicklungslabors innerhalb des Daimler-Benz Konzerns untersucht und weiterentwickelt. Der Einsatz dieser Verfahren in industriellen Robotern beziehungsweise in künftigen Service-Robotern steht noch aus. Langfristiges Forschungsziel ist hier der Einsatz von KNN in allen Ebenen einer Robotersteuerung.

Zur visuell basierten Robotersteuerung gehören insbesondere folgende Problemklassen:

- **Systemmodellierung:** Die Beschreibung der inversen Kinematik und Dynamik von Manipulatoren ist wegen der inhärenten Nichtlinearität sehr kompliziert. Mit KNN, wie beispielsweise dem bereits erwähnten Kohonen-Netz, ist es gelungen, diese Beschreibung für konkrete Manipulatoren automatisch zu erlernen, somit also das jeweilige System zu modellieren.
- **Nichtlineare Regelung:** KNN bieten prinzipiell einen generischen Ansatz zur Regelung der Freiheitsgrade von Manipulatoren, mobilen Systemen oder Kameratelevisoren. Obwohl zahlreiche Institute weltweit daran arbeiten, ist eine KNN basierte, nichtlineare Regelung, die auch nachweislich stabil ist, noch nicht industriell einsetzbar.
- **Nichtlineare adaptive Filterung:** Lineare nichtadaptive Filter zur Bildanalyse können trivialerweise durch KNN implementiert werden. Dies führt aber letztlich auf die klassischen, einfach formulierbaren Ansätze und erbringt somit kaum Vorteile. Dagegen ist eine nichtlineare Filterung, die zusätzlich auf unterschiedlichen Bildregionen adaptiv modifizierbar ist, ungleich schwerer zu beschreiben. Neuartige Ansätze mit KNN, die kurz mit den Schlagworten *Boltzmann-Maschine* und *RBF-Netz* benannt seien, könnten hierzu einheitliche Lösungskonzepte bieten.
- **Objektlokalisierung:** Für die Annäherung an ein Zielobjekt oder das Manipulieren

muß im Bildmaterial das Objekt erkannt werden. Dies ist letztlich eine auf dem Bild wirkende Funktion, die für jeden Ausschnitt feststellt, ob das gesuchte Objekt darin enthalten ist. KNN haben sich als besonders wertvoll bei der Approximation von Funktionen in hochdimensionalen Räumen herausgestellt. Wesentlich ist hierbei die Adaption von problemangepaßten Basisfunktionen (realisiert durch verdeckte Knoten eines KNN), die es KNN erlauben, hochgenau zu interpolieren.

- **Sensor–Daten–Fusion:** Die von den Sensoren oder Video–CCD–Kameras gelieferten Daten müssen im Hinblick auf eine zielorientierte Robotersteuerung geeignet kombiniert werden. KNN basierte Ansätze sind hierzu in verschiedenen Labors realisiert worden.
- **Echtzeitbetrieb:** Grundlegend wichtig ist in der allgemeinen wie auch in der visuell basierten Robotik die Einhaltung von Echtzeitbedingungen. Im Gegensatz zu Simulationen ist die Zeit in realen Anwendungen keine skalierbare Größe. KNN–Systeme, die den Aspekt der Zeitbeschränkung in der Realität aktiv zu kontrollieren gestatten, sind bislang nicht vorhanden.

Zusammenfassend läßt sich feststellen, daß KNN in zahlreichen Teilbereichen der visuell basierten Robotik untersucht und teilweise eingesetzt werden. Die bisher bekannten Lernverfahren sind allerdings noch nicht zufriedenstellend.

- Das Lernen **konvergiert nur langsam**, so daß ein Lernen im Online–Betrieb derzeit kaum möglich ist. Um Roboterlernen im Echtzeitbetrieb zu realisieren, sind noch große Anstrengungen erforderlich.
- Die Lernverfahren von KNN sind noch zu sehr von **speziellen Netztopologien** und der **spezifischen Kodierung** der Eingangsdaten abhängig. Das tatsächlich lernbare Wissen wird durch die Art der Repräsentation bestimmt.

Vorteile bringen die derzeit vorhandenen KNN durch die

- mögliche hohe **Komplexität des Gelernten**,
- einfache **Portierung des Gelernten** und die
- schnelle **Anwendung des Gelernten**.

Für die angeführten Probleme der visuell basierten Robotik sind neue innovative Netztypen mit zugehörigen Lernverfahren erforderlich, die auch ein Echtzeit–Lernen und eine bedeutungsmodulierte Lernrate gestatten. Unbestritten ist, daß KNN bereits jetzt eine wichtige Schrittmachertechnologie für die Realisierung von visuell basierten Robotersystemen sind.

2.4 Kommerzielle und wissenschaftliche Produkte

In diesem Abschnitt werden zunächst einige visuell basierte Robotersysteme angeführt und anschließend für das Robotersehen und Roboterlernen geeignete Spezialrechner genannt. Die Produkte sind im deutschen Raum kommerziell zu erwerben und können in der Industrie oder für wissenschaftliche Zwecke eingesetzt werden.

Es existiert eine Vielzahl von Systemen des Maschinensehens, die mit Robotern oder anderen Handhabungsmechanismen zusammenarbeiten. Dabei geht es um die Erkennung und Lokalisierung von Werkstücken anhand vordefinierter Modelle zum Zwecke der Manipulation oder Vermessung.

1. Das System *ROBOT-VISION* von ISRA-Systemtechnik GmbH dient zur 3D-Lagebestimmung für die Entpalettierung und Beladung von Maschinen.
2. Das System *CAPMES* von ISRA-Systemtechnik GmbH ist für 3D-Lageerkennung und Vermessung von Karosserien vorgesehen.
3. Das System *VIRO* von VITRONIC Dr.-Ing. Stein GmbH kann sowohl zur Positions- und Drehlagebestimmung als auch zur Identifikation von Werkstücken verwendet werden.
4. Das System *ROBOT-GUIDANCE* von QuISS GmbH kann die Lage von Bauteilen erkennen und wird zur Entpalettierung von Zylinderköpfen und anschließender Montage auf Motorblöcken eingesetzt.

Im Bereich der visuell basierten Navigation in Räumen arbeiten weltweit zahlreiche Universtitätsgruppen. Jedoch sind die eingesetzten Systeme meist mit Ultraschallsensoren zur Abstandsmessung ausgestattet, mit entsprechend eingeschränkten Fähigkeiten. Zum lokalen Manövrieren wird von der C-VIS GmbH in Bochum das System *VISIONBUMPER* angeboten, das mit Hilfe einer Stereokamera-Anordnung Hindernisobjekte detektiert. Zur globalen Navigation bietet VITRONIC Dr.-Ing. Stein GmbH das System *VICAR*, das bei fahrerlosen Fördersystemen in Industriehallen eingesetzt wird. Das System arbeitet weitestgehend unabhängig von künstlich geschaffenen Umgebungsbedingungen.

Viele Verfahren des Maschinensehens basieren auf dem Prinzip *Single Instruction Multiple Data (SIMD)*, d.h. dieselbe Operation wird auf unterschiedliche Daten angewendet. Dies eröffnet die Möglichkeit der hochgradig parallelen Verarbeitung von Bildern. Die MasPar Computer GmbH bei München bietet hierzu beispielsweise den aus 1024 Prozessoren bestehenden MasPar MP-2101A Rechner an. Er erbringt für Integer- und Floating-Point-Operationen Leistungsdaten, die von allgemeinen Workstations noch nicht erreicht werden.

Zur Implementierung der neuronalen Funktionsprinzipien ist ebenfalls eine massiv-parallele Verarbeitung von Signalen erforderlich. Die Siemens Nixdorf AG bietet hierfür

den *General Purpose Neurocomputer* SYNAPSE-1 an. Er bietet einerseits große Flexibilität, weil sich eine Vielzahl der üblichen neuronalen Algorithmen implementieren läßt. Andererseits erreicht er auch eine extrem hohe Rechenleistung, wobei über 50 Millionen Synapsen bearbeitet werden können.

Mit einer kurzen Zusammenfassung der Schwächen der vorgestellten Produkte soll die Diskussion des gegenwärtigen Standes der visuell basierten Robotik beendet werden. Die oben aufgeführten, visuell basierten Robotersysteme basieren zu sehr auf explizit bereitzustellenden Modellen und somit auf künstlich zu schaffenden Randbedingungen. Insbesondere sind diese Robotersysteme auch nicht imstande, das autonome Handeln unter Echtzeitbedingungen zu erlernen. Das Problem der beiden genannten Rechner ist einerseits das beträchtliche Volumen, welches die Entwicklung von mechatronischen Systemen ausschließt, und andererseits die zum Teil schwierige Implementierung der parallelen Algorithmen. Darüber hinaus sind diese SIMD-Rechner nicht allgemein genug verwendbar, um alle notwendigen Verfahren des Robotersehens oder des Roboterlernens darauf implementieren zu können.

3 Sehende Robotersysteme als Perspektive

Robotersysteme, die visuell gestützt Gegenstände manipulieren oder sich in Umgebungen fortbewegen können, haben sehr vielfältige Anwendungsmöglichkeiten. Sie könnten deshalb in den kommenden Jahren vielfältig im Dienstleistungsbereich und teilweise auch im Produktionsbereich zum Einsatz kommen. Video-CCD-Kameras bieten gegenüber Distanzmeßgeräten, wie Laser- oder Ultraschall-Sensoren, signifikante Vorteile:

- Die Kameras haben einen weitaus **größeren Sichtbarkeitsbereich**, wodurch man bereits mit einer Aufnahme detaillierte Information über die Szene erhält. Mit Laser-Sensoren kann man dies beispielsweise nur sukzessiv durch Veränderung der Projektionsrichtung erreichen, weil diese nur punktuell abbilden.
- Die Kameras registrieren die natürlichen Lichtstrahlen der Szenenobjekte, die **Umwelt selbst wird nicht verändert**. Im Gegensatz dazu wird beispielsweise bei Ultraschall-Sensoren die Luft mit Schallwellen moduliert. Da diese Sensoren einen breiten Streuungsbereich haben, kommt es bei mehreren parallel arbeitenden Sensoren zu komplizierten Interferenzen.

Video-CCD-Kameras sind die geeignete Wahrnehmungskomponente von Robotersystemen. Mit der zusätzlichen Möglichkeit, die Blickrichtung der Kameras unabhängig von der eigentlichen Roboterbewegung zu verändern, erhält man ausreichend Datenmaterial über die Umwelt. Bei Echtzeitbetrieb kann aber nur eine eingeschränkte Datenmenge in den vorgegebenen Zeiteinheiten verarbeitet werden. Abhängig von der Aufgabenstellung soll deshalb

- die Blickrichtung auf **interessierende Szenenausschnitte** gerichtet und
- das Bildmaterial nur in **ausgewählten Bereichen detailliert** bearbeitet werden (Abbildung 25).

Die Datenmenge ist so weit einzuschränken, daß eine Echtzeitverarbeitung möglich ist, jedoch die relevante Information noch extrahierbar bleibt.

Robotersysteme müssen befähigt werden, selbständig das zur Erfüllung einer Aufgabe richtige Verhalten zu erlernen.

- Dadurch können sie sich auch in **unstrukturierten, nicht modellierbaren Umgebungen** zurechtfinden. Die explizite Modellierung von statischen Szenenobjekten oder von dynamischen Vorgängen entfällt.
- Das richtige Verhalten schließt auch mit ein, die Umwelt durch Kameraregelung nur **selektiv wahrzunehmen**, um einen Echtzeitbetrieb zu ermöglichen.

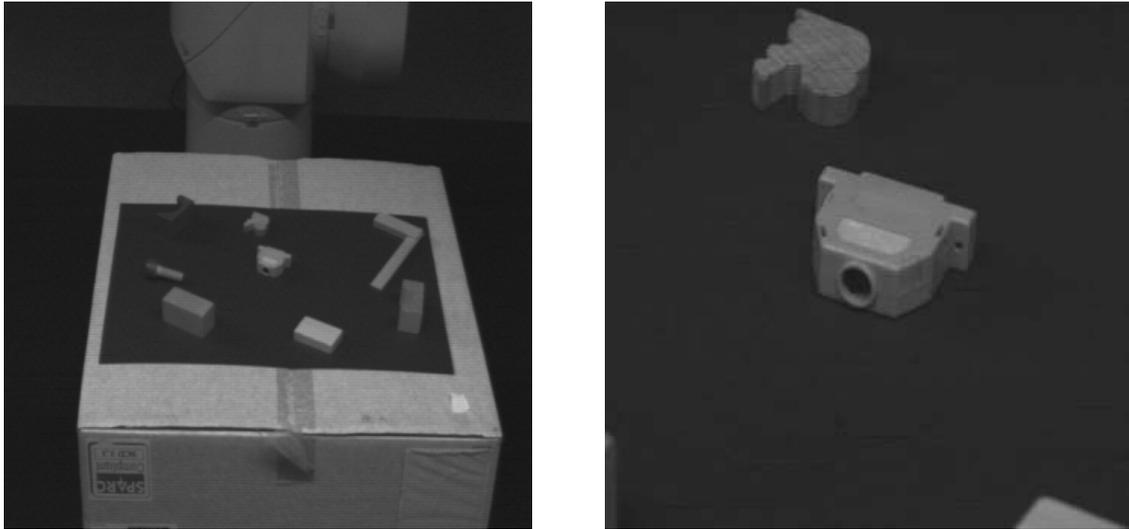


Abbildung 25: Fokussierung auf einen relevanten Szenenausschnitt.

In diesem Kapitel werden zunächst drei reale Szenarien von visuell basierten Robotersystemen skizziert. Der zweite Abschnitt stellt dann eine neue Architektur für ein visuell basiertes, lernendes Robotersystem vor, das *verhaltensbasierte Robotersystem*. Es handelt sich dabei um eine grundlegende Abkehr vom Prinzip der funktionalen Modularisierung, das im Abschnitt 2.2.3 erklärt wurde. Danach werden dann konkrete Zukunftsperspektiven für aktiv wahrnehmende Roboter, für manipulierende sehende Roboter und für mobile sehende Roboter vorgestellt. Der letzte Abschnitt des Kapitels beschreibt den *Spin-off* der visuell basierten Robotik im Hinblick auf die Entwicklung von sogenannten *visuellen Informations-Management-Systemen*. Diskutiert werden gemeinsame beziehungsweise sich ergänzende Aufgabenstellungen und Lösungen.

3.1 Einige Szenarien

Es sollen nun drei potentielle Anwendungen für sehende Robotersysteme ausführlicher diskutiert werden. Unmittelbares Ziel ist dabei jeweils die Navigation zwischen einem Start- und einem Zielort sowie die Manipulation von Gegenständen. Die Wahrnehmung der Umwelt erfolgt durch autonomes, aktives Beobachten.

3.1.1 Automatisierung von Teleoperationen

Teleoperation-Roboter sind mühsam zu handhaben, weil das Bedienungspersonal entweder eine größere Distanz zum Werkobjekt hat und/oder das Werkobjekt nur indirekt über Kameras beobachtet wird. Dabei müssen durch das Abbildungssystem verursachte teilweise starke Verzerrungen der Metrik und der Perspektive vom Bedienungspersonal korrigiert werden. Ein konkretes Anwendungsziel kann darin bestehen, das Bedienungspersonal der über Television gesteuerten Manipulatoren zu unterstützen. Diese Unterstützung wird durch visuell basierte, teilweise autonom handelnde Manipulatoren erreicht. Infolgedessen kommt dem Menschen beim Trainieren eine zentrale, bei der Anwendung des Trainierten aber nur noch eine überwachende Rolle zu.

Der Manipulator kann gesundheitsgefährdende Aufgaben durchführen, wie beispielsweise das Schleifen von bestimmten Materialien (siehe Abbildung 6). Die Gefährdung entsteht durch Funkenflug bei Metall, durch giftige Dämpfe bei Beschichtungen oder durch Bruch des Materials. Notwendig für das autonome Handeln ist insbesondere, daß durch Bildverarbeitung die Wirkung des Schleifvorgangs erkennbar und der zu erreichende Endzustand der Bearbeitung (nicht notwendigerweise explizit) modellierbar ist. Das Personal zur Überwachung des Manipulators befindet sich hinter einer schützenden Glaswand und kann die Aktionen direkt beobachten.

Ein visuell basierter Manipulator hat alternativ dazu auch in einer Nuklear- oder einer Chemieanlage eine wichtige Verwendung. Bei Störfällen können beispielsweise Schweißnähte inspiziert, Abschirmsteine gestapelt oder verseuchtes Material abgebaut und wegtransportiert werden. Die Überwachung des autonom handelnden Manipulators erfolgt in sicherer Entfernung vom Einsatzort mittels Bildübertragung.

Ein drittes Anwendungsbeispiel zur Automatisierung von Teleoperationen ist die Entschärfung von Minen in ehemaligen Kriegsgebieten. Dabei muß zunächst die Mine gefunden, freigelegt, der Zündmechanismus erkannt und schließlich das geeignete Werkzeug zum Entschärfen der Zündung eingesetzt werden.

Aus den angeführten Beispielen wird ersichtlich, daß der Manipulationsvorgang prinzipiell darin besteht,

1. den Effektor an ein Szenenobjekt heranzuführen,
2. eine geeignete Manipulationsstellung einzunehmen und

3. den Manipulationsgegenstand hinsichtlich Lage und/oder Form zu verändern.

Im **ersten Schritt** muß eine hindernisfreie Trajektorie für den Effektor hin zum Manipulationsgegenstand gefunden werden. Das wesentliche Problem besteht hierbei darin, daß sich der gesamte Roboterarm dynamisch durch die Bewegung der einzelnen Gelenke verändert. Dadurch betrifft die Hindernisvermeidung nicht nur den Effektor sondern den gesamten Roboterarm.

Im **zweiten Schritt** nimmt der Effektor eine bestimmte räumliche Beziehung zum Zielobjekt ein. Beispielsweise muß für einen Greifvorgang eine stabile Greifstellung visuell ermittelt werden, indem zwischen den Greifbacken und dem Greifobjekt möglichst große Auflageflächen gesucht werden.

Im **dritten Schritt** wird schließlich der Gegenstand verändert. Für den Fall eines Greifvorgangs muß das Objekt ohne Kontaktierung mit Hindernissen an einer anderen Position platziert werden. Handelt es sich alternativ dazu beispielsweise um Schweißvorgänge, so muß das Schweißgerät fortwährend in einer konstanten geometrischen Relation zum Schweißgegenstand gehalten werden. Eine visuell basierte Regelung dieses Vorgangs bedeutet etwa, daß relevante Punkte oder Konturen aus dem Bildmaterial extrahiert und dreidimensional rekonstruiert werden, um davon ausgehend Schweißnähte an den richtigen Stellen des Gegenstandes anzubringen.

Prinzipiell sind die angeführten Beispiele der Telerobotik dadurch charakterisiert, daß die Manipulationsgegenstände aufgrund der möglichen Vielfalt nicht a priori modelliert und somit exakte Manipulationstrajektorien nicht vorweg programmiert werden können. Statt dessen muß das System in der Lage sein, anhand der visuellen Daten von Szenenaufnahmen ein situationsgerechtes Manipulationsverhalten zu erlernen. Der Nutzen derartiger Anwendungen besteht offensichtlich darin, daß Menschen nicht gefährlichen oder schädlichen Umgebungen ausgesetzt sind.

3.1.2 Automatisierte Unterstützung von Behinderten

In dem hier vorgestellten Szenario handelt es sich um einen (teil-) autonomen Rollstuhl zur Unterstützung von behinderten und allgemein pflegebedürftigen Menschen (siehe Abbildung 5). Existierende motorisierte Rollstühle stellen an den Menschen hohe Anforderungen bei der Ausführung von Fahrmanövern. Sie sind nicht bei hohem Behinderungsgrad einsetzbar, weil der Nutzer mit ihnen nur durch Manipulation kommunizieren kann. Weiterhin existieren keine Hilfsgeräte für Behinderungen der oberen Gliedmaßen. Der Autonomiegrad des Rollstuhls muß an den Behinderungsgrad des Menschen anpaßbar sein, der sich von einer Behinderung eines Beines bis zur vollständigen Querschnittslähmung erstrecken kann, welche möglicherweise auch die Arme betrifft.

Aufgabe des Rollstuhls ist es, teilweise oder vollständig autonom zu navigieren. Darüber hinaus kann der Rollstuhl mit einem Manipulator ausgestattet sein oder von einem Manipulator auf getrenntem Fahrzeug begleitet werden, um die behinderte Person bei

Handhabungsaufgaben zu unterstützen. Im zweiten Fall handelt es sich um eine Fahrhilfe für Transportaufgaben, die auch in Dienstleistungsbranchen Anwendung findet.

Als kommerzielles Produkt wird hier der Rollstuhlarm MANUS der holländischen Firma Exact Dynamics genannt, der mit der Kommunikationsschnittstelle BASCO ausgestattet geliefert wird. Mittels dieser Schnittstelle kann man den Roboterarm mit Kopf-, Kinn-, Hand- oder Zehnbewegungen steuern. Der Roboterarm hat einen Arbeitsradius von 85 Zentimeter und eine Traglast von 2 Kilogramm. Weitere prototypische Entwicklungen erfolgen an der Universität Bremen, an der Fernuniversität Hagen und am Forschungs-Institut Technologie-Behindertenhilfe in Wetter.

Als erstes Ziel des Szenarios wird die Funktion eines Rollstuhls in einer Innenraumumgebung festgelegt. Der *Roboterrollstuhl* muß in der Lage sein, symbolisch eingegebene Aufgaben in Handlungsfolgen umzuwandeln. Beispielsweise:

Fahre mich zum Bücherregal und hole mir das Buch X.

Dieselbe Aufgabe kann auch interaktiv eingegeben werden, indem das Endziel durch das inkrementelle Eingeben der Zwischenziele in nichtsymbolischer Form ermittelt wird. Solche inkrementelle Aufgaben können durch das Zeigen von Türen auf einem Touchscreen oder mit Augenbewegungen erfolgen. In beiden Fällen sind die Kommandos durch ein Bildauswertesystem zu interpretieren.

Im Fall der symbolischen Eingabe steht das Problem der globalen Navigation im Vordergrund. Voraussetzung ist das interaktive Erlernen einer *topographischen Karte* der Umgebung, was durch die inkrementelle visuelle Erkundung erfolgen kann. Diese Karte wird durch einen *relationalen Graphen* repräsentiert, in dem alle möglichen Eingabeziele als Knoten auftreten. Neuere Studien zeigen [Mallot und Schölkopf, 1995], wie man einen solchen Ortsgraphen durch einen aus der visuellen Erkundung erworbenen Ansichtsgraphen ermitteln kann. Die globale Navigation basiert dann auf charakteristischen Ansichten, mit denen geeignete Bewegungen zu assoziieren sind. Zur Eliminierung von Mehrdeutigkeiten können eindeutige künstliche Landmarken benutzt werden.

Im Fall der interaktiven Eingabe wird ein anzufahrendes Zwischenziel im Bild bestimmt. Das im Englischen als *Visual Homing* bekannte Problem besteht aus einer *Bahnplanung*, die mittels eines Referenzsignals im Koordinatensystem der Kamera geregelt wird. Diese Regelung interagiert mit den Verhaltensweisen der lokalen Navigation, wie der Hindernisdetektion und -vermeidung und dem Fahren entlang einer Wand. Diese Koexistenz von konkurrierenden Verhaltensweisen benötigt ein Organisationsprinzip, das Gegenstand einer nichtklassischen Regelungstheorie ist und beispielweise nach dem Prinzip von *diskreten Ereignissystemen* [Košicka und andere, 1995] behandelt werden kann. Insbesondere im Fall eines Rollstuhls ist die Bahnplanung, die der visuellen Szeneninterpretation folgt, eine anspruchsvolle geometrische Aufgabe (Problem des Klaviertransports), weil die lokale Manövrierfreiheit durch die relativ zur Umgebung großen Dimensionen des Rollstuhls erheblich eingeschränkt wird.

In mehreren Rollstuhlfunktionen wird die Bedeutung eines Kamerasystems mit *Blickrichtungssteuerung* ersichtlich. Neben den Vorteilen bei der Lösung von Aufgaben der lokalen Navigation benötigt man die freie Blickwahl zusätzlich zum visuellen Abgleich mit dem Zielort, während das System selbst in Lokalmanöver verwickelt ist. Nicht zuletzt ist die visuelle Erkundung zum Erwerb einer topographischen Karte mit einer fest montierten Kamera praktisch unmöglich.

3.1.3 Automatisiertes Befördern von Gegenständen

Ein typischer *Transportroboter* für eine innenräumliche Umgebung besteht aus einer mobilen Plattform, einem oder mehreren Manipulatoren, einer Ladefläche, der Sensorik und den Onboard-Rechnern. Seine Hauptaufgabe ist die autonome Navigation zwischen verschiedenen Orten. Desweiteren werden eine Vielzahl von autonom zu bewältigende Manipulationsaufgaben anfallen, wie beispielsweise

- das **Öffnen und Schließen** von Türen,
- das **Identifizieren von Zielobjekten**,
- ihr **Aufgreifen, Verstauen und Ablegen**, sowie
- das **Liftfahren** und
- das **Anfahren von Aufladestationen** im Falle leerer Batterien.

Während für Kollisionsvermeidung weitgehend auf Video-Kameras zugunsten von Ultraschall-, Infrarot- und Laser-Sensoren verzichtet werden kann, und zur Selbstlokalisierung externe Signalgeber eingesetzt werden können, erfordert spätestens die Identifikation und Manipulation von Zielobjekten visuelle Unterstützung.

Zwecks größerer Flexibilität sollte der Serviceroboter darüber hinaus seine Zielobjekte und deren Manipulation erlernen können. Ist der Roboter aber mit der nötigen Hardware zur Bildverarbeitung in Video-Echtzeit ausgerüstet, so kann diese auch effektiv zur hindernisvermeidenden Navigation und Selbstlokalisierung eingesetzt werden. Die *Selbstlokalisierung* an visuellen Landmarken bietet dabei die Möglichkeit, auf eine eventuell teure Präparierung der Umgebung mit Signalgebern zu verzichten. Visuelle Hindernisvermeidung macht den Laser-Sensor überflüssig, der aus Sicherheitsgründen nur dort verwendet werden kann, wo nicht die Gefahr besteht, auf Menschen zu treffen. Visuell basierte Hindernisvermeidung und Selbstlokalisierung sind Kondensationspunkte internationaler Forschungsaktivität.

Ein konkretes Beispiel betrifft etwa das Befördern von Mahlzeiten, Arzneien und medizinischen Geräten in Krankenhäusern. Hierzu wird beispielsweise von der amerikanischen Firma TRC der sogenannte Helpmate vertrieben (siehe Abbildung 11, oben links), der allerdings lediglich auf der Grundlage von Ultraschall navigiert.

3.2 Fusion von Robotik, Maschinensehen und Neuroinformatik

Anhand der drei Szenarien aus dem vorigen Abschnitt wird deutlich, daß bei zukünftigen Robotertechnologien die Wissenschaftsdisziplinen Robotik, Maschinensehen und Neuroinformatik verschmelzen müssen (siehe auch erstes Kapitel). Die im Abschnitt 2.2.3 besprochene funktionale Modularisierung von Robotern schließt das enge Zusammenwirken dieser Disziplinen von vornherein aus. Nur mit der nachfolgend dargestellten *verhaltensbasierten Modularisierung* erhält man Robotersysteme, die die erforderliche Integration der drei Fachgebiete realisieren. Die rechentechnischen Voraussetzungen und Erfordernisse zur Umsetzung in marktfähige Produkte werden ebenfalls erläutert.

3.2.1 Konzeption der verhaltensbasierten Modularisierung

Die Modularisierung zukünftiger Robotersysteme erfolgt nach Verhaltensweisen und nicht nach Funktionen. In dieses Konzept läßt sich auch das Robotersehen einbeziehen, so daß man von einem verhaltensbasierten Robotersehen sprechen kann.

Visuell basiertes Roboterverhalten

Seit dem Beginn der neunziger Jahre werden Architekturkonzepte entwickelt, gemäß derer ein Robotersystem jeweils durch ein Repertoire von Verhaltensmodulen aufgebaut ist. Ein Verhaltensmodul realisiert eine bestimmte Verhaltensweise, beispielsweise die Vermeidung von Hindernissen durch einen mobilen Roboter. Mit Hilfe von Kameras wird eine räumliche Beziehung zwischen dem Roboter und einem möglichen Hindernis festgestellt, und diese Beziehung zieht eine Aktion zur Vergrößerung der Distanz nach sich.

Aus mehreren solchen elementaren Verhaltensweisen emergiert ein komplexes Verhalten. Man spricht auch von einer *heterarchischen Beziehung* zwischen Verhaltensweisen oder von der sogenannten *Subsumtions-Architektur* [Brooks, 1986]. Die Abbildung 26 zeigt beispielhaft das Schema eines verhaltensbasierten mobilen Roboters.

Es gibt nicht mehr Funktionsebenen, sondern Kompetenzebenen (oder Verhaltens-ebenen), die aus Verhaltensmodulen bestehen. Die Verhaltensmodule einer höheren Ebene setzen teilweise auf den Modulen der niedrigeren Ebenen auf. Prinzipiell sind simultan (parallel) die Verhaltensmodule aller Kompetenzebenen einsatzbereit. Ob sie parallel arbeiten oder *ereignisgetriggert* oder *aufgabengesteuert* beziehungsweise abgeschaltet werden, hängt sowohl von der jeweiligen spezifischen Funktion als auch von der Architektur der Kontrolle des Gesamtsystems ab. Module können *kooperieren*, also in gegenseitiger Abhängigkeit arbeiten, oder *konkurrieren*. Im letzten Fall müssen Strategien der Konfliktauflösung vorhanden sein.

Erfolgt die Funktion eines Moduls datengetrieben, so kann auf eine übergeordnete Kontrolle verzichtet werden. Man spricht von *Bottom-Up-Kontrolle*. Diese Funktionweise bezeichnet man als *reaktiv*. Reaktive Systeme sind zu unmittelbarer und schneller Reaktion auf Ereignisse befähigt. Würden aber alle Verhaltensmodule reaktiv sein, wäre das

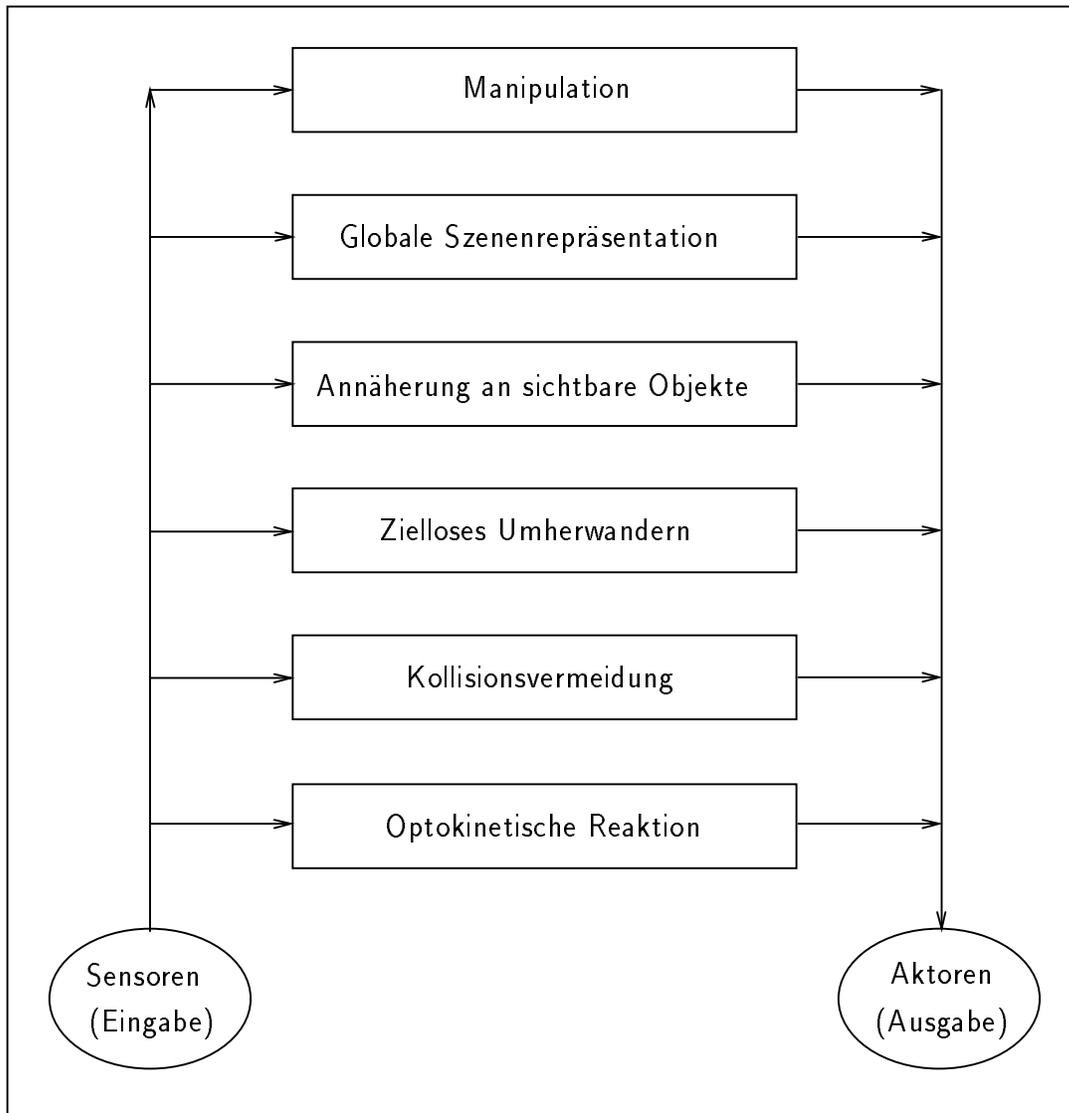


Abbildung 26: Verhaltensbasierte Modularisierung von Robotersystemen.

Gesamtsystem in seinem globalen Verhalten schwer zu kontrollieren. Es wäre auch nicht möglich, anspruchsvolle Aufgaben dem System zu überlassen.

Deshalb sind auch Module mit aufgabengetriebener Funktion erforderlich, die von einer Kontrolleinheit (*Top-Down-Kontrolle*) überwacht werden. Diese Funktionsweise bezeichnet man als *zielorientiert* oder *aktiv*. Die Folge sind weniger unmittelbar wirkende, langsamere Reaktionen. Aber erst das Vorhandensein dieser Module macht es möglich, dem System eine anspruchsvolle Aufgabe zu übertragen und die Funktionalität des Gesamtsystems unter der Kontrolle des Anwenders zu halten.

Reaktive Module können ihre Funktionalität erlernen oder fest programmiert sein. Letztere bezeichnet man als *instinktbasiert*. Diese Module realisieren einfache aber "lebensnotwendige" Verhaltensweisen. Zielorientierte Module sollten lernfähig sein, um sie der Variabilität möglicher situativer Einbettungen der Problemlösung anpassen zu können.

Verhaltensbasierte Systemarchitekturen sind *robust*, wenn sie mit einem Minimum an explizitem Modellwissen über die Umwelt ihre Funktionalität erreichen. Sie sind aber auch *stabil* in dem Sinne, als der Ausfall eines Moduls nicht den Ausfall des Gesamtsystems zwangsläufig zur Folge hat.

Wenn ein Verhaltensmodul einer höheren Ebene ausfällt, dann können die Verhaltensmodule der unteren Kompetenzebenen noch autonom weiterarbeiten. Falls beispielsweise ein Robotersystem mit der visuell basierten Erstellung einer Umweltkarte beschäftigt ist, und dann unerwartet das Licht im Raum ausfällt, dann könnte er sich aufgrund von Ultraschall-Sensoren immer noch hindernisvermeidend bewegen. Darin liegt ein wesentlicher Unterschied zur funktionalen Modularisierung, bei denen das gesamte System zum Stillstand kommt, sobald eine Funktion ausfällt.

Gemäß der Abbildung 26 lassen sich z.B. folgende Verhaltensebenen unterscheiden.

1. **Optokinetische Reaktion:** Auf der **ersten Ebene** nimmt der Roboter lediglich Helligkeitsunterschiede aus der Umgebung wahr und reagiert darauf in primitiver Weise mit einfachen reflexartigen Bewegungen. Beispielsweise kann er sich einer Lichtquelle annähern oder sich von dieser entfernen.
2. **Kollisionsvermeidung:** Auf der **zweiten Ebene** geht es ebenfalls noch um eine einfache, reaktive Verhaltensweise, nämlich dem Ausweichen von starren oder beweglichen Objekten. Erforderlich ist eine minimale Szenendeutung, die nur zwischen Hindernissen und Nicht-Hindernissen unterscheidet. Dadurch werden der Roboter selbst sowie auch die vorhandenen Szenenobjekte vor Zerstörung geschützt.
3. **Zielloses Umherwandern:** Auf der **dritten Ebene** zeigt der Roboter erstmalig aktives Verhalten. Er versucht, fortwährend in Bewegung zu bleiben, ohne dabei mit Hindernissen zu kollidieren. Dieses Umherwandern ist allerdings noch ziellos.
4. **Annäherung an sichtbare Objekte:** Auf der **vierten Ebene** wird ein zielorientiertes lokales Manövrieren realisiert. Dabei geht es darum, sich den in der lokalen Umgebung sichtbaren Objekten gezielt anzunähern, um sie beispielsweise anschließend vermessen oder identifizieren zu können. Ziel ist hierbei, eine genaue Beschreibung der lokalen Umgebung zu erstellen.
5. **Globale Szenenrepräsentation:** Auf der **fünften Ebene** wird durch lokales Manövrieren und globales Navigieren eine vollständige Beschreibung eines bestimmten Raumes oder eines Freigeländes erzeugt. Der Roboter muß auch weiter entfernt liegende Objekte finden und aufsuchen, von deren Existenz zu Beginn keine Informationen vorliegen, und die in der Ausgangslage noch gar nicht sichtbar sind. Die hierbei entstehende Beschreibung der Szene erfolgt global in einem beobachterunabhängigen Koordinatensystem und führt zu einer sogenannten *Karte*.
6. **Manipulation:** Die **sechste Verhaltensebene** realisiert die Manipulation von Szenenobjekten. Notwendig ist eine genaue Formwahrnehmung, um z.B. einen Gegenstand greifen und seine Lage in gewünschter Weise verändern zu können.

Alle Verhaltensmodule haben direkten Zugriff zu den Sensoren beziehungsweise Kameras und zu den Effektoren. Dabei sind die Module der unteren Ebenen (in der Abbildung die Ebenen eins und zwei) ausschließlich datengetrieben. Wahrnehmung und Roboterregelung finden in der Realität statt, eine Planung ist nicht erforderlich. Die Module der oberen Ebenen realisieren aktives Verhalten (in der Abbildung die Ebenen drei bis sechs), welches von der jeweils gegebenen Aufgabenstellung kontrolliert wird. Wahrnehmung, Planung und Roboterregelung sind individuell auf ein Verhaltensmodul zugeschnitten und nicht mehr globale Funktionen wie bei der funktionalen Realisierung.

Typischerweise befinden sich mehrere Verhaltensmodule in Konkurrenz, wobei sie entweder die gleichen oder verschiedene Ziele verfolgen. Bei einer Greifaufgabe durch ein Robotersystem kann beispielsweise simultan zur Annäherung an das Greifobjekt das Umgehen von Hindernissen gefordert sein. Auf jeder Verhaltensebene können mehrere Module installiert sein, die ihre Zuständigkeit für die Realisierung einer bestimmten Aufgabe in unterschiedlichen Situationen unter Beweis stellen. So ist bekannt, daß im menschlichen visuellen System fünf verschiedene Zugänge zur Wahrnehmung räumlicher Tiefe existieren, die sich gegenseitig ergänzen. Jedes Verhaltensmodul strebt danach, Aktionen so zu wählen, daß der individuelle Nutzen optimiert wird.

Ein optimal angepaßtes Gesamtverhalten erhält man schließlich dadurch, daß in jeder Situation die einzelnen Verhaltensmodule miteinander konkurrieren und dieser Konflikt im Hinblick auf die Maximierung des Gesamtnutzens aufgelöst wird.

Charakteristisch ist weiterhin, daß ein Gesamtziel nicht explizit formuliert sein muß, sondern sich implizit aus der Maximierung des Gesamtnutzens ergeben kann. Die Forschung hierzu steht allerdings noch weit in den Grundlagen. Deshalb wird in nächster Zeit auf eine zentrale Kontrolleinheit, der das Gesamtziel des Systems explizit mitzuteilen ist, nicht verzichtet werden können.

Verhaltensbasiertes Robotersehen

Die neunziger Jahre sind durch eine Änderung der Forschungsrichtung des Maschinensehens gekennzeichnet, die das Robotersehen näher an die im Abschnitt 2.2.4 genannte Definition bringen soll. Demnach soll nur diejenige Information erfaßt werden, welche für die Ausführung spezifischer Aufträge des Roboters notwendig ist. Es ist naheliegend, daß man für viele Aufgaben weder eine ausführliche Rekonstruktion noch eine vollständige Modellanpassung benötigt, damit der Roboter eine Aktionen ausführen kann. Der tiefere Hintergrund dieser trivialen Feststellung ist die Erkenntnis, daß die Ermittlung kompletter Beschreibungen in Echtzeit trotz der stürmischen Entwicklung der Rechentechnik unmöglich ist.

Wenn beispielsweise ein Roboter manipulator ein Objekt greifen soll, braucht das System ohnehin nicht die komplette 3D-Form des Objektes zu kennen. Vielmehr ist es ausreichend, die Distanz zwischen zwei greifbaren Flächen zu berechnen. Wenn ein Fahrzeug bremsen soll, braucht das System nicht den Typ des vorausfahrenden Fahrzeuges in Erfahrung zu bringen. Es reicht völlig, den Abstand, die Geländeneigung und die Relativ-

geschwindigkeit zu kennen. Wenn ein Innenraumfahrzeug zu einem Ziel gelangen soll, ist die Rekonstruktion der 3D-Gestalt der Flure nicht notwendig. Die Berechnung eines *Freiraumtunnels* zum Manövrieren ist ausreichend.

Das Handeln als Ziel des Robotersehens führt zu selektiver Verarbeitung. In Abhängigkeit von der Aufgabenstellung des Roboters werden nur diejenigen Größen der Szene berechnet, von denen der Erfolg der Aktionen des Roboters abhängt.

In manchen Fällen, wie beispielsweise auf der untersten Verhaltensebene der optokinetischen Reaktion (siehe Abbildung 26), werden sogar die Stellgrößen des Roboters direkt aus den Bildhinweisen berechnet. Das andere Extrem besteht darin, daß beispielsweise auf der obersten Verhaltensebene der Manipulation (siehe Abbildung 26), ein unbekannter Gegenstand erst aus verschiedenen Blickwinkeln aufgenommen und analysiert werden muß, um ihn anschließend geeignet behandeln zu können. Ist hingegen ein Modell des Gegenstandes vorhanden (er ist also bekannt), dann werden andere effektivere Verhaltensweisen aktiviert. Die mögliche Bewegung der Kamera ist hier ein notwendiges Mittel, um zunächst den Gegenstand geeignet wahrnehmen zu können. Im Gegensatz dazu ist bei der optokinetischen Reaktion die Aktion schon der eigentliche Zweck.

Das Paradigma der aktiven Wahrnehmung beinhaltet nicht nur die erwähnte Veränderung der Beobachtungsrichtung, d.h. der äußeren Freiheitsgrade. Darüber hinaus können auch innere Freiheitsgrade der Wahrnehmung modifiziert werden, etwa die Skala oder die Orientierung von Filtern. In Ergänzung zu den beispielhaft in Abbildung 26 aufgeführten *visuell basierten Verhaltensweisen* wird visuelle Wahrnehmung ebenfalls durch Verhaltensmodule realisiert. Man bezeichnet diese als *visuelle Verhaltensweisen*.

Beispielsweise lassen sich folgende Aufgaben des Maschinensehens als visuelle Verhaltensweisen behandeln:

- Schätzung der **Eigenbewegung**,
- Schätzung **unabhängiger Objektbewegung**,
- Schätzung der **Kollisionszeit**,
- **Objekterkennung**,
- **Attention auf relevante Objekte**,
- **Blickverfolgung bewegter Objekte**,
- Schätzung der **räumlichen Tiefe**,
- Schätzung der **räumlichen Gestalt**.

Eine Verhaltensweise der aktiven Wahrnehmung bezieht sich auf die adäquate Modifikation der Freiheitsgrade von Bildaufnahme und Bildanalyse, um eine bestimmte Information aus dem Bildmaterial zu erhalten. Eine visuelle Verhaltensweise wie “Schätzung der räumlichen Tiefe” kann andere Verhaltensweisen voraussetzen, wie “Attention auf relevante Objekte” oder

- **Regelung der Vergenz zweier Kameras** zur Fixierung eines Objektpunktes.

Das roboterrelevante Maschinensehen bettet sich also auf natürliche Weise in den verhaltenbasierten Entwurf von Robotersystemen ein [Sommer, 1995].

Dieses neue Paradigma der Realisierung visueller Wahrnehmung in technischen Systemen begründet sich auf der Erkenntnis, daß biologische Systeme, die zu Wahrnehmung befähigt sind, diese Kompetenz nur in enger Rückkopplung mit auszuführenden Aktionen erweitern können. Die Kompetenzen von Wahrnehmung und Handlung können nicht isoliert erworben werden. Sie stellen zwei Seiten einer Medaille dar, die im Wahrnehmungs–Aktions–Zyklus erfaßt werden.

In Abbildung 27 wird eine zeitgemäße Vorstellung der Verarbeitung visueller und auditorischer Information im Gehirn des Menschen nach [Kosslyn, 1994] dargestellt. Deutlich erkennbar sind die Module der *Aufmerksamkeitssteuerung* in ihrer Vernetzung mit Modulen der Verarbeitung und unterschiedlichen Realisierungen des Gedächtnisses.

3.2.2 Rechentechnische Voraussetzungen

Für jede Verhaltensweise wird idealerweise ein eigener Rechner vorgesehen. Dieser Rechner realisiert den speziell für diese Verhaltensweise notwendigen Regelkreis aus Bildaufnahme, Bildanalyse und Robotersteuerung. Es ist offensichtlich, daß die verschiedenen Verhaltensweisen auch Rechner mit unterschiedlichen Eigenschaften benötigen.

Für die vorwiegend datengetriebenen Verhaltensweisen der unteren Ebenen würden einfache *Signalprozessoren* genügen, die einfache Regelkreise und elementare Signalverarbeitung erlauben. Primitive Ausführungen von Prozessrechnern, wie sie bei großen technischen Anlagen eingesetzt werden, wären hierfür ausreichend.

Für Verhaltensweisen der mittleren Ebenen spielt die Bildanalyse eine deutlich stärkere Rolle. Notwendig sind insbesondere Spezialrechner zur signalnahen Bildverarbeitung in Echtzeit. *Pipeline–Rechner*, wie beispielsweise die DATACUBE von der gleichnamigen amerikanischen Firma, sind hier der aktuelle Stand der Technik. Zur Realisierung von neuronalen Netzen wäre der Einsatz von spezieller *neuronaler Hardware* notwendig. Diese sollte einerseits eine schnelle Matrix–Vektor–Multiplikation ermöglichen, wie beispielsweise die SYNAPSE–1 von Siemens. Ebenso wichtig ist aber auch, daß diese Rechner die dynamische Entwicklung von neuronalen Netzen gestatten und somit ein fortwährendes Lernen und Verbessern der Verhaltensweisen erlauben.

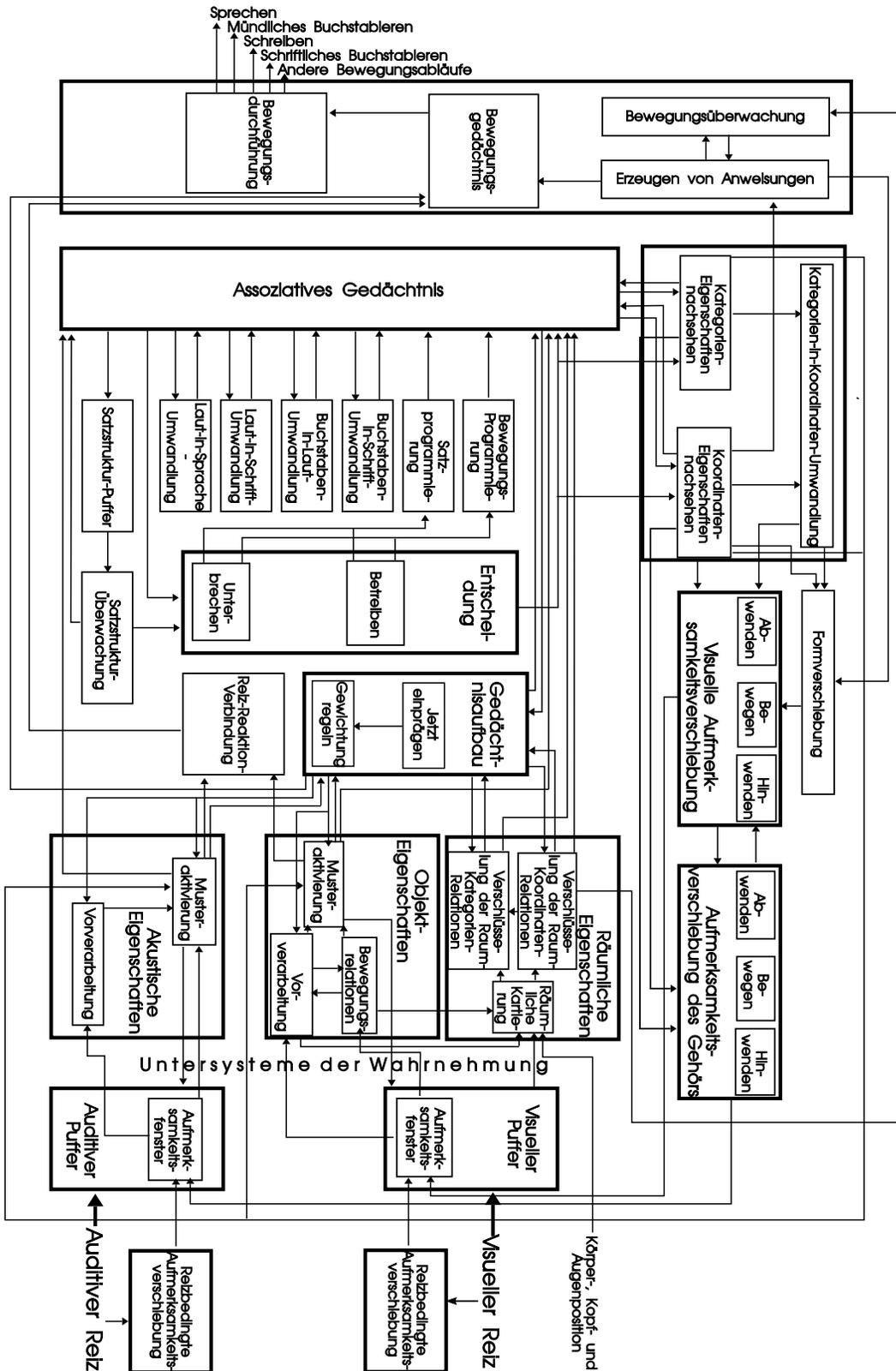


Abbildung 27: Schema der Verarbeitung visueller und auditorischer Information im Gehirn des Menschen (aus [Kosslyn, 1994]).

Die Verhaltensweisen der höheren Kompetenzebenen arbeiten einerseits datengetrieben und andererseits aber vor allem auch aufgabengetrieben. Es muß also möglich sein, Zielvorgaben, Planungsstrategien, Objektbeschreibungen und sonstige Aspekte der Aufgabenstellung zu formulieren. Für die Beschreibung der Aufgabenstellung eignet sich ein *Workstation-Rechner mit Allzweckprozessoren*, der aber zusätzlich über Echtzeitfähigkeiten verfügen muß. Für das eigentliche Robotersehen und das Roboterlernen sollte er mit den oben erwähnten Spezialrechnern gekoppelt werden.

Bei der Kopplung der verschiedenen Rechner stellen allerdings die Kommunikationsanforderungen und die geringen Transportraten der Daten-Busse den *Flaschenhals* für Robotik-Anwendungen dar. Diese Erfahrung und der derzeitige Trend zu Allzweckprozessoren sprechen aus heutiger Sicht dafür, auf die Signalprozessoren zu verzichten und die zusätzliche Rechenleistung einem entsprechend ausgebauten *grobkörnig parallelen Rechner* aus leistungsfähigen Allzweckprozessoren abzufordern. Derartige Rechner heißen auch *MIMD-Rechner*, für *Multiple Instructions Multiple Data*. Diese These wird dadurch gestützt, daß die Fließkomma-Rechenleistung heutiger Allzweckprozessoren aufgrund hoher Taktraten bereits die Leistung von Signalprozessoren übertrifft. Ein solcher MIMD-Rechner bietet darüberhinaus den Vorteil, auf einzelnen Prozessoren auch Anwendungen von neuronalen Netzen implementieren zu können, da geeignete Neurocomputer noch nicht in der gewünschten Spezifikation verfügbar sind. Desweiteren ergibt sich der Vorteil größerer Flexibilität und Rechnerauslastung gerade im Entwicklungsstadium. Darüber hinaus werden die Entwickler von einem weiteren zusätzlichen Programmiermodell entlastet, da die Rechnerausstattung nur noch aus einem Spezialrechner für Bildverarbeitung und dem allgemein verwendbaren MIMD-Rechner besteht.

Hinsichtlich des Einsatzes auf (teil-) autonomen Robotern spielen natürlich Platzbedarf, Gewicht und Energieverbrauch der Rechner eine entscheidende Rolle. Soweit es die Einsatzumgebung zuläßt, kann hier Funkübertragung der Sensor- und Steuersignale, verbunden mit stationärer Rechentechnik, zumindest in der Entwicklungsphase verwendet werden. So muß man nur sicherheitsrelevante Verhalten direkt auf dem Roboter installieren, während Bilddaten und andere sensorische Information über Videofunk zur stationären Rechenanlage und dortigen Weiterverarbeitung in Echtzeit übermittelt werden. Die stürmische Entwicklung auf dem Rechnermarkt, verbunden mit der fortschreitenden Miniaturisierung, läßt hoffen, in absehbarer Zeit eine für visuelle Robotik besser geeignete Rechentechnik verfügbar zu haben, als dies bisher der Fall ist.

3.3 Visuell befähigte Robotersysteme

Im vorigen Abschnitt wurde dargestellt, wie bei der verhaltensbasierten Modularisierung von Robotersystemen die Fachgebiete Robotik, Maschinensehen und Neuroinformatik auf natürliche Weise verschmelzen. Dieser Abschnitt führt in die in naher Zukunft realisierbaren, visuell befähigten Robotersystemen ein. Besprochen werden aktiv wahrnehmende Roboter, manipulierende sehende Roboter und mobile sehende Roboter.

3.3.1 Perspektive für das Robotersehen

Die visuelle Informationsverarbeitung beim Robotersehen hat einen verhaltensbezogenen Charakter. Diese Vorgehensweise ist deutlich praxisrelevanter und somit besser zur Realisierung von Robotersehen geeignet als bisherige Ansätze. Die direkte Abbildung von Bildhinweisen auf verhaltensbezogene Größen eröffnet nun aber wieder grundlegend neue Fragen bezüglich der theoretischen Fundierung des Robotersehens. Ein verhaltensbasiertes visuelles System kann seinen Erfolg nur in einem funktionierenden System beweisen, weil der Wahrnehmungs–Aktions–Zyklus als externer Regelkreis nur in Experimenten in der Realwelt getestet werden kann. Außerdem nimmt die Theorie Rücksicht auf die physikalischen Gegebenheiten und insbesondere auf die eingeschränkten Rechenressourcen. So kommen zum Beispiel numerische Verfahren nicht zur Anwendung, deren Konvergenz nicht garantiert innerhalb des Reaktionszyklus erfolgt.

Diese Leitlinie prägt die Vorgehensweise bei der Entwicklung:

- Es ist **nicht mehr möglich**, Algorithmen für Probleme des Robotersehens zu entwickeln, ohne sie direkt auf einem Roboter zu testen.
- Es ist **nicht mehr möglich**, die Algorithmen offline in Simulationen mit längeren Zykluszeiten zu testen, und darauf zu hoffen, daß die Entwicklung der Prozessoren den Echtzeit–Engpaß automatisch auflösen wird.

Die Bedeutung der Repräsentation

Die Abbildung von Bilddaten auf Aktionen oder aktionsbezogene geometrische Größen bringt das Problem der Repräsentation dieser Daten und ihrer Zwischentransformationen in den Vordergrund. Die Frage der Repräsentation betrifft das ganze Spektrum des Übergangs vom wahrgenommenen Signal zur auszuführenden Aktion. Für jedes Verhalten muß eine geeignete Repräsentation gefunden werden, die

1. eine **direkte Abbildung** auf die beabsichtigte Aktion erlaubt und
2. **invariant** gegenüber Faktoren ist, die irrelevant in Bezug auf diese Aktion sind.

Wie in der Zusammenfassung der Funktionsprinzipien beschrieben wurde, bieten sich auf der signalnahen Ebene Funktionensysteme an, die, angewandt auf das Originalbild,

einen Satz von Basisantworten liefern, welche invariant gegenüber bestimmten Bilddeformationen sind. Diese Basisantworten können zur Anpassung an den lokalen Signalverlauf linear kombiniert werden.

Es sollen einige Beispiele aus dem Stereo- und Bewegungssehen angeführt werden. Man kennt aus der Akustik die *Kurzzeitspektrogramme*, welche die lokalen Frequenzanteile eines Signals vergleichbar der Notenschreibweise in der Musik darstellen. Entsprechend existieren *lokale-spektrale Repräsentationen* für Bilder, welche die Frequenzverteilung von Betrag und Phase des Signals eines kleinen Bildabschnitts angeben.

Filtert man mit entsprechenden Operatoren das linke und rechte Bild eines Stereopaars, so kann man direkt aus dem Phasenunterschied der Antworten die Positionsabweichung bestimmen und dadurch auch die Tiefe der abgebildeten Szenenstruktur. In [Hansen und Sommer, 1996] wird gezeigt, daß dies in Video-Echtzeit möglich ist.

Transformiert man die Bilder mittels einer komplex-logarithmischen Abbildung, so lassen sich in der lokal-spektralen Repräsentation nicht nur Positionsabweichungen sondern auch Vergrößerungen und Rotationen darstellen. Weil das Abbild eines sich annähernden Objektes mit der Zeit größer wird, kann man Warnsignale direkt aus einer solchen Repräsentation extrahieren.

In höheren Verarbeitungsebenen, wo schon Bildhinweise wie Geraden und Punkte extrahiert worden sind, ist es möglich, mittels geometrischer Überlegungen eine qualitative Repräsentation der Umgebung herzuleiten, in der die Anordnung der Szenenkomponenten erhalten bleibt, ohne daß metrisches Wissen über die Aufnahmeparameter benutzt wird. Der Verzicht auf dieses Wissen bietet sich an, weil Aufnahmeparameter in der hier konzipierten Robotik entweder unbekannt sind oder nur fehlerhaft ermittelt werden können.

Zur Problematik der Repräsentation visueller Information im Cortex der Primaten wurden in den letzten Jahren bedeutsame Fortschritte erzielt, die Anregungen für die Realisierung technischer visuell wahrnehmender Systeme lieferten. Die Abbildung 28 zeigt ein Schema der Datenströme im visuellen Cortex von Affen nach [DeYoe und Essen, 1988]. Deutlich zu erkennen ist eine vertikale und horizontale Modularisierung, die durchaus an das Funktionsschema der verhaltensbasierten Modularisierung in Abbildung 26 erinnert. Die Symbole innerhalb der Blöcke weisen auf Beiträge zur Verarbeitung unterschiedlicher Information wie räumliche Tiefe, Orientierung von Mustern, Richtung von Bewegung oder Farbe von Objekten hin. Zu erkennen ist aber auch, daß die funktionalen Blöcke durchaus Beiträge zu mehreren Problemlösungen liefern. Deshalb spricht man auch von einer *Labyrintharchitektur* und nicht von einer Subsumtionsarchitektur.

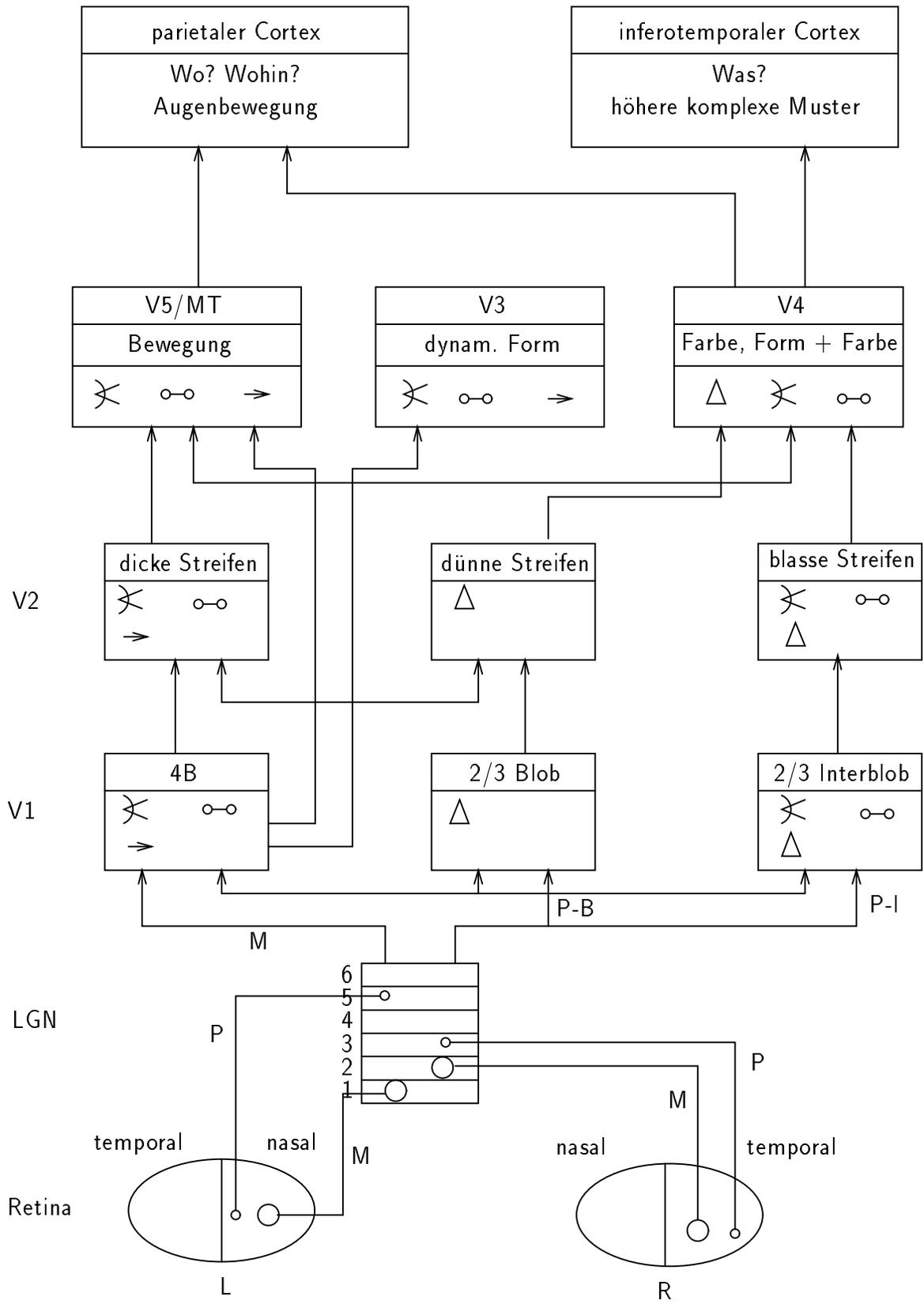


Abbildung 28: Datenströme visueller Informationsverarbeitung im Cortex von Affen (aus [DeYoe und Essen, 1988]).

Aktives und ortsvariantes Sehen

Die Netzhaut (Retina) der Primaten, einschließlich des Menschen, besitzt eine *ortsvariante Auflösung*. Fixiert man auf einen Punkt, so kann man ein Wort am Rande des Gesichtsfeldes nicht lesen. Will man dennoch das Wort lesen, so muß man die Blickrichtung auf dieses Wort richten. Aktives und ortsvariantes Sehen sind also eng verkoppelt. Übliche Kameras besitzen anders als die Retina im ganzen Gesichtsfeld dieselbe Auflösung. Das hindert aber nicht an einer ortsvarianten Verarbeitung. Weshalb die Anwendung von ortsvariantem Sehen für Roboter interessant ist, soll ein Beispiel verdeutlichen.

Damit der Roboter seine Fahrt planen kann, muß er bewegte Objekte an der Peripherie des Gesichtsfeldes detektieren. Diese Detektion braucht aber nicht gleichzeitig mit der Angabe der Geschwindigkeit gekoppelt zu sein. Ein binäres Signal bezüglich des Vorhandenseins ist hinreichend und kann mit einer sehr groben Auflösung abgeleitet werden. Möchte man die Geschwindigkeit des Objekts schätzen, so muß man die betreffende Bildregion im Detail verarbeiten. Das könnte, ohne die Kamera zu bewegen, durch hochauflösende Fenster in verschiedenen Bereichen des Bildes erfolgen.

Dennoch gibt es weitere Gründe, die für Beweglichkeit der Kameras sprechen. Das Gesichtsfeld kommerzieller Kameras ist so eingeschränkt, daß entweder mehrere Kameras oder bewegliche Kameras nötig sind, um die Umgebung zu explorieren und um bewegte interessante Objekte innerhalb des Gesichtsfeldes zu behalten. Darüber hinaus bringt ein bewegliches Kamerapaar erhebliche Vorteile beim stereoskopischen Sehen. Die Positionsabweichung der Bildstrukturen im linken und rechten Bild bleibt beschränkt, so daß die Komplexität des Zuordnungsproblems reduziert wird.

3.3.2 Binokulare Kameraplattformen als Perspektive

Für das aktive Sehen wurden *Kameraplattformen* mit mechanischen und optischen Freiheitsgraden gebaut, die anhand der durch die visuelle Verarbeitung ermittelten Ziele geregelt werden.

In Gestalt der Okulomotorik wird das Sehen selbst eine Roboteraufgabe, bei der das Handeln aus der Steuerung von Blickbewegungen besteht.

Über zehn verschiedene Typen mono- und binokularer Kameraplattformen wurden bisher in Forschungslabors gebaut. Inzwischen werden auch Geräte kommerziell angeboten. Eine binokulare Kameraplattform besitzt mindestens vier *mechanische Freiheitsgrade*:

- Eine Schwenkung um den **Hals**,
- eine Neigung des **Kopfes** und
- je Kamera einen Freiheitsgrad für die **Vergenzbewegungen** der Kameras.

Jeder Freiheitsgrad unterliegt einer Achsenregelung, die in höherem Takt als die visuelle Regelung erfolgt. Jede Kamera besitzt außerdem drei optische Freiheitsgrade für die Regelung der Blende, des Fokus und der Brennweite.

Die Vorteile eines solchen Systems, das nicht als Selbstzweck betrachtet wird, wurden im letzten Abschnitt unter dem Gesichtspunkt der Reduzierung der Komplexität der visuellen Aufgaben beschrieben. Danach eröffnet die Lösung der damit zusammenhängenden visuellen Aufgaben neue Forschungs- und Entwicklungsfragen. Blickbewegungen müssen möglichst unmittelbar nach dem Empfang des Auslösereizes erfolgen. Diese Reaktionszeit hängt von der Verarbeitungszeit ab, welche die effektive Sensorabtastrate bestimmt, und von der Latenzzeit, welche die Zeit zwischen Reiz und Befehlstart beschreibt. Nach der europäischen Norm (CCITT) erhält man von einer CCD-Kamera 25 Vollbilder pro Sekunde. Das Abtastintervall von 40 Millisekunden ist auch für die Regelungsrate wünschenswert, damit sich schnell bewegende Objekte verfolgt werden können. Demzufolge ist der Anspruch an Algorithmen niedriger Komplexität und an schnelle dedizierte Rechner sehr hoch.

Auf einer binokularen Plattform werden verschiedene visuelle Fähigkeiten realisiert, deren Koordination von einem übergeordneten Zustandsautomaten erfolgen soll. Diese Fähigkeiten sind entweder direkt anwendungsbezogen, wie beispielsweise die Verfolgung einer Person, oder tragen zu den Aufgaben der Manipulation und Navigation bei. In der folgenden Auflistung möglicher *okulomotorischer Aufgaben* werden auch Verhalten berücksichtigt, welche den Fall einer Kameraplattform auf einem Fahrzeug betreffen.

1. Die datengetriebene **Zuwendung der Blickrichtung** auf bewegte Objekte oder auf Objekte, deren Grauwert bzw. Farbe einen hohen Auffälligkeitsgrad aufweist.
2. Die modellgetriebene **Änderung der Blickrichtung**, um ein Objekt nach einem Suchplan zu finden.
3. Die **glatte Blicknachführung**, um ein bewegtes Objekt im Zentrum des Bildes zu halten.
4. Die **glatte Blicknachführung**, um einen stationären Punkt zu fixieren, während sich die Kameraplattform vorwärts bewegt.
5. Die **Stabilisierung der Blickrichtung**, während die Plattform vorwärts fährt oder während die Plattform um den Hals rotiert.
6. Die **Regelung des Vergenzwinkels** der linken und der rechten Kamera, um einen Punkt zu fixieren, dessen Tiefe zu ermitteln ist.
7. Die **Zoombewegung**, um eine Struktur in höherer Auflösung zu bearbeiten, ohne sich ihr anzunähern.
8. Die Scharfeinstellung durch **Regelung des Fokus**, um scharfe Bilder in beliebigem Abstand im Nahbereich aufnehmen zu können.

Binokulare Kameraplattformen stellen also eine neuartige Kategorie von Robotersystemen dar, die allein der visuellen Wahrnehmung nach dem Paradigma des aktiven Sehens dienen. Hiermit lassen sich völlig neuartige Aufgaben lösen, die auf die anthropomorphe Realisierung des Sehens angewiesen sind. Da diese Systeme aber bisher die Labors noch nicht verlassen haben, wird erst deren Popularisierung zu der erwarteten Breite der Definition von Einsatzfeldern führen.

Die Aufgabenklassen lassen sich bereits jetzt erkennen:

- **Überwachung** von Szenen und Prozeßabläufen,
- **Exploration** räumlicher Tiefe an schwer zugänglichen Einsatzorten,
- **visuelle Kommunikation** von Mensch und Maschine.

Eine beträchtliche Diversifizierung der Einsatzfelder ergibt sich durch die Kopplung mit mobilen Plattformen.

3.3.3 Mobile sehende Roboter als Perspektive

Wie zu Beginn des dritten Kapitels in den dargestellten Szenarien beschrieben wurde, haben mobile sehende Roboter einen sehr breiten Anwendungsbereich. Allerdings müssen Aufgaben des *lokalen Manövrierens* visuell gestützt gelöst werden. Beispiele hierfür sind:

1. Hindernisse in Fahrrichtung detektieren, abbremsen und Hindernisse vermeiden,
2. entlang einer 3D-Struktur wie einer Wand oder innerhalb eines Flures fahren,
3. ein im Bild definiertes Ziel anfahren,
4. einem anderen Fahrzeug oder einer Person im bestimmten Abstand folgen oder
5. Andockmanöver durchführen, um zu parken oder eine zum Greifen günstige Lage einzunehmen.

Die Aufgaben der *globalen Navigation* hängen davon ab, ob der Roboter in Außen- oder Innenszenen eingesetzt wird. Der Roboter muß über die Fähigkeit verfügen,

1. mittels visueller Erkundung die Umgebung zu kartographieren,
2. ohne Karte ein besuchtes Ziel wiederzufinden,
3. selbst aufgebaute Karten mit eindeutigen Landmarken zu kombinieren oder
4. den kürzesten Pfad zum Ziel zu berechnen.

Die Mehrheit dieser Aufgaben wird parallel aber asynchron abgearbeitet, wobei jede Aufgabe einen eigenen Takt erfordern kann. Die Koordination nach bestimmten Optimalitätskriterien gehört zum Gebiet der diskreten Regelung.

3.3.4 Manipulierende sehende Roboter als Perspektive

Zukünftig sollen Systeme realisiert werden, die auf visueller Wahrnehmung der Umwelt basierend die Kompetenz erwerben, Manipulationsaufgaben autonom durchzuführen. Die Szenarien der Manipulation sind im allgemeinen nur teilweise vorhersehbar. Die Formen und die Posen der Gegenstände sind nicht immer bekannt. Deshalb kann die Regelung des Roboterarms und des Effektors nicht ausschließlich in der Weise vorprogrammiert werden, daß bei Vorliegen einer gewissen Form und Pose eines Gegenstandes eine bestimmte Manipulationsstellung anzuwenden ist. Vielmehr muß sich das Robotersystem autonom der Situation anpassen und explorativ für bislang unbekannte Gegenstände jeweils Manipulationsstellungen ermitteln (Abbildung 29). Die fortwährende Bereitschaft und Fähigkeit zum Lernen ist eine zentrale Voraussetzung dafür, daß das Robotersystem situationsgemäß adäquat handelt.

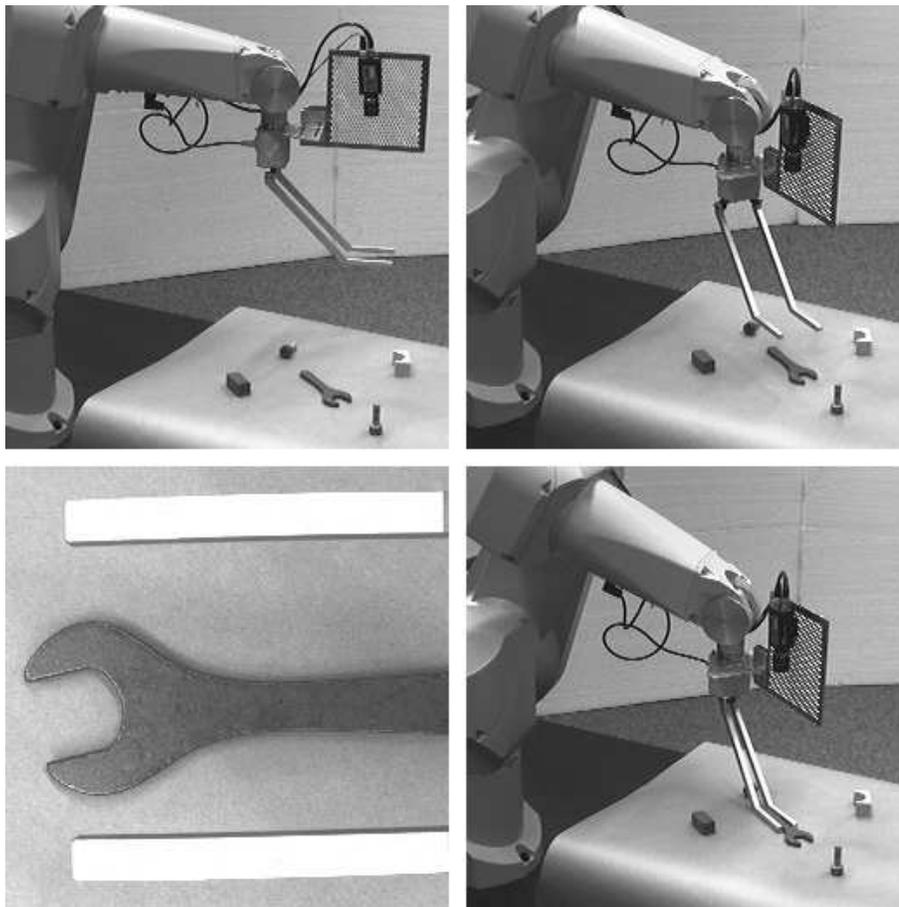


Abbildung 29: Visuell basiertes Erkennen einer möglichen Greifstellung und entsprechende Steuerung des Manipulators.

Als technische Zielsetzung ist insbesondere die Ausführung der Manipulationsaufgabe in einem bestimmten Zeitintervall zu gewährleisten. Entscheidend hierfür ist eine schnelle Verarbeitung des Bildmaterials. Durch die konsequente Anwendung des Paradigmas

der aktiven visuellen Wahrnehmung erreicht man die nötige Effizienz. Beispielsweise kann die Aufmerksamkeit unter grober Auflösung schnell auf interessierende Bildbereiche gelenkt werden und unter detaillierter Auflösung dann selektiv die geforderte Interpretation erfolgen.

Eine andere technische Zielsetzung ist das Manipulieren mit unterschiedlichen Effektoren. Bei Greifern unterscheidet man beispielsweise Zweibacken- und Dreibacken-Greifer. Einerseits ist zu erwarten, daß wichtige Komponenten des Systems weitgehend unabhängig von der Art des Greifers sind, beispielsweise die Repräsentation der Wahrnehmungs-Aktions-Assoziation als neuronales Netz. Andererseits ist es wünschenswert, das System so zu realisieren, daß sich durch neuronales Lernen das Greifverhalten autonom auf den verwendeten Greifer einstellt. Beides ist realisierbar.

3.3.5 Autonomes Sehen, Fortbewegen und Manipulieren

Erst die Integration von visuell basierter Manipulation und Navigation wird die Entwicklung wirklich nützlicher Serviceroboter ermöglichen, wie sie zur Unterstützung Behinderter oder zur Realisierung eines Helfers im Haushalt benötigt werden.

Das Hauptproblem der Integration wird wiederum in der Koordination manipulations- und navigationsbezogener Verhaltensmodule liegen. Die Navigation muß die Manipulation unterstützen, kann also nicht unabhängig von ihr konzipiert werden. Darüber hinaus sind Zugriffskonflikte auf den aktiven Sensor (das Kamerasystem) zu lösen, soll die Steuerung des Robotersystems nicht allein datengetrieben, sondern durch aufmerksamkeitsgesteuertes Aktivieren der Module realisiert werden. Spätestens hier wird deutlich, daß eine in den Anfängen der verhaltensbasierten Robotik vorgeschlagene strikte Trennung der Module und Handlungsebenen nicht aufrecht zu halten ist. Ein mehr globaler, aufgabengesteuerter Mechanismus der Konfliktlösung wird notwendig, der darüber hinaus höchstwahrscheinlich adaptiv sein muß.

Ein weiteres Problem der Integration besteht darin, einen handlungsbezogenen Repräsentationsrahmen zu finden, der sowohl die Manipulation als auch die Navigation umfaßt. Nur so können die Verhalten sinnvoll interagieren, beispielsweise ein weiteres Annähern an das Zielobjekt ausgelöst werden, falls der Manipulator noch nicht eingesetzt werden kann.

Die Integration von Manipulation und Navigation wirft nicht nur Probleme auf, sondern bietet auch neue Problemlösungen für die einzelnen Teilgebiete an. Ein Beispiel betrifft die Kalibrierung eines Kamerasystems zur hindernisvermeidenden Navigation auf (ebenen) Fluren. Da diese Kalibrierung mit einer Menge von auf der Ebene liegenden Punkten durchgeführt werden kann, ist es möglich, den Manipulator aktiv einzusetzen, um eine solche Menge zu markieren. Ein anderes, sehr naheliegendes Beispiel ist der Einsatz des Manipulators zur Beseitigung von Hindernissen beziehungsweise allgemein zur aktiven Präparierung der Umgebung durch den Manipulator zu Navigationszwecken. Umgekehrt kann ein geschicktes Anfahren von Zielobjekten deren Manipulation wesentlich vereinfachen.

3.4 Visuelle Informations–Management–Systeme

Die visuell basierte Robotik geht aus der klassischen Robotik durch eine Verschmelzung mit den mit visueller Wahrnehmung beziehungsweise Lernen betrauten Disziplinen hervor. Ein typisches Charakteristikum der visuell basierten Robotik besteht in der Fähigkeit, Handlungen aus einer Echtzeitanforderungen gerecht werdenden Analyse visueller Daten abzuleiten. Es ist leicht vorstellbar, daß derartige Systeme die Handlungen nur noch rudimentär ausprägen. Außerdem müssen die Daten nicht mehr nur unmittelbar aus der Umwelt bereitgestellt werden, sondern werden aus einer Bibliothek digital oder analog abgespeicherter Bilddaten entnommen. Ein derartiges System nennt man ein *Visuelles Informations–Management–System (VIMS)*. Seine Fähigkeiten sind im wesentlichen auf die intelligente Leistung reduziert, große Mengen an Bilddaten gezielt durchmustern zu können, Assoziationen herzustellen und Aussagen abzuleiten. Diese Fähigkeiten stellen aber einen zentralen Kern visuell basierter Robotik dar.

Im Zeitalter, da Videotechnik jedermann verfügbar ist, da von Daten–Highways gesprochen wird und Internet sowie World–Wide–Web (WWW) nicht mehr nur leere Begriffe sind, spielen Bilddaten eine zunehmend bedeutende Rolle für die Kommunikation und für die Präsentation komplexer Sachverhalte. Auf einem im Juni 1995 in Cambridge/Massachusetts organisierten NSF–ARPA–Workshop über VIMS [Jain *und andere*, 1995] wurde auf die Notwendigkeit hingewiesen, in absehbarer Zeit bisher unvorstellbare Sammlungen von Bildern (in der Größenordnung Hunderttausend bis Millionen) effektiv nach bestimmten Inhalten durchmustern zu können.

Das dort herausgearbeitete Problemverständnis soll hier kurz dargelegt werden, weil es in vielen Aspekten Ähnlichkeiten zu dem in dieser Studie dargelegten Problemkreis aufweist.

3.4.1 Visuelle Informationssysteme und Bild–Datenbanken

Klassische Informations–Management–Systeme haben die moderne Industriegesellschaft revolutioniert. Sie arbeiten aber nur bei alphanumerischen Daten effektiv. Es ist vergleichsweise viel einfacher, nach dem Schlüssel “gelber Hund” in einem Text zu suchen, als ein entsprechendes Objekt in einem Bild oder gar einer Bilddatenbank zu finden.

Ähnlich wie in der visuell basierten Robotik werden aktive Mechanismen des Navigierens in den Daten, des Stöberns und Suchens nötig sein, die viel mehr semantisch orientiert sind als dies klassische Datenbanktechniken erfordern.

In der Folge werden einige zentrale Probleme benannt, für die es noch keine fertigen Lösungen gibt. An ihrer Bearbeitung wird aber sehr anwendungsorientiert, zum Teil unter der Federführung der Industrie, vor allem in den USA gearbeitet.

1. **Semantischer Flaschenhals:** Nutzeranfragen an visuelle Informationssysteme werden auf einem semantischen Niveau formuliert, welches im Widerspruch zur Repräsentationsform visueller Daten steht. Es sind Transformationsmechanismen erforderlich, die auch für die Übermittlung von Aufgaben an autonome Robotersysteme benötigt werden.
2. **Quellen- und Repräsentationsunabhängigkeit:** Die Daten derartiger Systeme zur Auskunft stammen aus unterschiedlichsten Quellen (z.B. Video, Audio) und Aufnahmetechniken. Sie sind normalerweise in den für diese Quellen entwickelten Formen repräsentiert. Es sind invariante Schnittstellen auf höherer Abstraktionsebene (kognitive Ebene) erforderlich (siehe Abbildung 27), um Zugriff zu der den Daten aufgeprägten Information zu erhalten.
3. **Kontext- und Zeitabhängigkeit der Information:** Wie in der visuell basierten Robotik interessieren zu verschiedenen Zeiten unterschiedliche Aspekte der visuellen Daten. Bei VIMS kommt die Subjektivität des Nutzers hinzu.
4. **Nutzer als Bestandteil des Systemkonzepts:** Anders als in der visuell basierten Robotik ist der Mensch als Nutzer stets in den Wahrnehmungs–Aktions–Zyklus integriert. An seine speziellen Bedürfnisse und Möglichkeiten ist die Funktionalität der Systeme anzupassen. Dies betrifft
 - schnelle und adäquate Reaktion und
 - Formulierung der Anfragen und Aufträge in einer *visuellen Sprache*, die der Umgangssprache semantisch und syntaktisch im wesentlichen entspricht.
5. **Bildadressierung, Kommentierung und Indizierung:** Die Anfragen an Bilddatenbanken sind semantisch kodiert, und entsprechend ist auch auf diesem Niveau die Problemlösung angesiedelt. Inhaltsbasierte Suche wird im Gegensatz zu konventioneller Datenbankanalyse die Anwendung von Methoden der Mustererkennung erfordern. Die Attributierung und Adressierung von Bildern sollte diesem Niveau der Nutzer–System–Kommunikation entsprechen.

3.4.2 Videobasierte Informationssysteme

Videobasierte Informationssysteme unterscheiden sich von Bilddatenbank–Systemen in folgenden Aspekten:

- Die Datenmenge der Video–Files ist noch viel größer.
- Zusätzlich zur statischen Semantik ist auch zeitliche Semantik zu beachten.
- Zeitliche Semantik trägt dazu bei, die Komplexität zu reduzieren.

Folgende Probleme sind zu lösen:

1. **Einbeziehen der Vertonung:** Oft sind Videofilme vertont. Soll dieser Aspekt in die Informationssysteme einbezogen werden, ist die Forschung, welche bisher zum automatischen Sprachverstehen geleistet wurde, auf das Verstehen anderer Geräusche und Musik zu erweitern. Für eine vereinheitlichte Repräsentation audio-visueller Information existieren gegenwärtig keine Ansätze.
2. **Semantik-erhaltende Kompressionstechniken:** Die gegenwärtig verfügbaren Kompressionsstandards für visuelle Signale (MPEG2) sind nicht ausreichend für Aufgaben des Informationsmanagements von Bilddaten. Mit dem Standard MPEG4 ist in Kürze eine modellbasierte Kompressionstechnik verfügbar. Es wird erforderlich sein, hier noch mehr Flexibilität auf semantischem Niveau zu erreichen.
3. **Verstehen komprimierter Bilder:** Im Sinne der effektiven Beantwortung von Aufträgen sind Bildanalysetechniken zu entwickeln, die auf komprimierte visuelle Signale anwendbar sind. Dafür sind auch spezielle Kompressionstechniken zu entwickeln, die mehr der Spezifik von Anfragen entsprechen als, wie bisher gefordert, eine Rekonstruktion der Daten unterstützen.
4. **Smart Kameras:** Zur Unterstützung von Editionsarbeiten sind Videodaten mit Metadaten zu verbinden, die Aussagen über die Umstände der Filmaufnahme enthalten. Hierfür sind künftige Kameras konzeptionell vorzubereiten und auszustatten.

Es ist zu erwarten, daß die in der nächsten Zeit zu entwickelnden Konzepte für Videobasierte Informationssysteme einen Zugriff auf abgebildete Aktivitäten von Personen gestatten. Beliebige Anfragen werden hingegen noch lange Zeit nicht möglich sein.

Da anders als in der visuell basierten Robotik der Mensch stets im Wahrnehmungs-Aktions-Zyklus integriert ist, werden auch einfachere Zugänge zum Trainieren und Anpassen der Systeme an unterschiedliche Nutzer und Aufgaben erwartet. Der Nutzer kann durch Rückkopplung unmittelbar zur Relevanz der Aktionen des Systems Stellung nehmen. Dadurch wird überwacht Trainieren impliziter neuronaler Verfahren für so komplexe Aufgaben wie Durchblättern der Datenbank nach verbal beschriebenen Sachverhalten möglich.

Der Anspruch, nutzeradaptive Informationssysteme zu entwickeln, ist von vergleichbarer Komplexität, wie autonome visuell basierte Robotersysteme für unstrukturierte Umgebungen zu schaffen. Deshalb dürften die dort entwickelten neuronalen Abbildungsverfahren auch geeignet sein, Ähnlichkeiten zwischen Anfragen und Bildstrukturen zu repräsentieren.

Zusammenfassend ist festzustellen, daß die visuelle Robotik zur Zeit weiter entwickelt ist als der Problembereich der visuellen Informations-Management-Systeme. Beide Richtungen können sich gegenseitig ergänzen. Es ist zu erwarten, daß wegen der breiten potentiellen Anwendungsbasis von VIMS in den nächsten Jahren gezielt Produkte auf dem Markt erscheinen.

4 Realisierung sehender Robotersysteme

In diesem Kapitel soll der erreichte Entwicklungsstand sehender Robotersysteme zusammenfassend eingeschätzt und hinsichtlich seiner Umsetzung in konkrete Anwendungen bewertet werden. Außerdem wird, so weit es den Autoren möglich ist, die Ausgangssituation in Schleswig-Holstein dargestellt, um Anregungen zu einer Bündelung der Aktivitäten zu geben.

4.1 Realisierungschancen und Erfordernisse

In dieser Studie wurde die Perspektive einer neuen Kategorie von Robotersystemen dargestellt, die aus der konzeptionellen Umorientierung des Verständnisses intelligenter Systeme folgt. Diese Systeme können *Sehen, Handeln und Lernen*. Erst die Interpretation des Begriffs Intelligenz im Sinne einer “praktischen Intelligenz”, die zwar in ihrer Beschreibung mit mathematischen Mitteln (z.B. der Logik) etwas Abstraktes ist, aber in Gestalt der Kompetenz — ein Verhalten des Systems bezüglich seiner Umwelt zu entwickeln — etwas sehr konkret Entwerfbares darstellt, machte diese Entwicklung möglich.

Die Architektur- und Funktionskonzepte von technischen Systemen, die mittels Eigenbewegung, Wahrnehmung und Handeln in einer nicht komplett vorstrukturierten Umwelt agieren können, werden heute in den akademischen Labors entwickelt. Dies bedeutet aber keinesfalls, daß industrielle Anwender solcher Systeme warten sollten, bis ihnen von der wissenschaftlichen und technischen Forschung ein Komplettsystem zur Verfügung gestellt wird. Ein solches System wird es nie geben. Vielmehr sollte die frühzeitige Kopplung akademischer Forschung an potentielle Anwendungsfelder dazu beitragen, Irrwege der Wissenschaft zu vermeiden und Teillösungen gezielt nutzbar zu machen.

Die Realisierung wahrnehmender und autonom handelnder Systeme gehört zu den großen Herausforderungen der Künstlichen Intelligenz [Reddy, 1995]. Gleichzeitig wird mit ihr ein enormer Innovationsschub für die Etablierung der Informationstechnologie in der menschlichen Gesellschaft erwartet. Worauf beruht dieser Innovationsschub?

Systeme des Maschinensehens, die wir heute in der industriellen Anwendung finden, werden meist für quantitative Qualitätskontrolle oder Prozeßüberwachung eingesetzt. Sie haben in mehr oder weniger weitem Sinne Meßaufgaben zu erledigen. Ihr Einsatz erfordert die genaue Modellierung sowohl der Aufgabe als auch der visuell zu erfassenden Aspekte der Welt. Lediglich diese Modellierung gestattet auch, die erforderlichen visuellen Erkennungsaufgaben zu lösen. Insbesondere unter dem Gesichtspunkt, einen Prozeßtakt einhalten zu müssen, können diese Modelle und folglich die Erkennungsleistungen nur sehr bescheiden sein.

Gegenwärtiges industrielles Maschinensehen löst ein spezielles Problem unter sehr spezifischen Annahmen.

Jede Anwendung erfordert meist eine Neuentwicklung der kompletten Problemlösung. Die Systeme sind nicht flexibel bezüglich Portierung in andere Anwendungsfelder, und sie sind auch sensibel bezüglich innerer und äußerer Störungen. Die Überwindung dieser Probleme wurde in der Vergangenheit (siehe zweites Kapitel) dadurch angestrebt, daß visuelle Wahrnehmung als ein allgemein lösbares Problem verstanden wurde, für welches eine entsprechende Theorie zu entwickeln sei.

Ein visuelles System, das nach der "Computational Theory of Vision" entworfen wird, löst ein allgemeines Wahrnehmungsproblem unter Einbeziehung spezifischer Annahmen über die Einsatzbedingungen.

Dieses Supersystem gibt es aber nicht unter den Lebewesen und wird es auch nicht als technische Variante geben. Hierfür gibt es einen wesentlichen Grund:

Generelle Wahrnehmung heißt auch, unter allen möglichen Umständen den sensorischen Daten eine Bedeutung zuweisen zu können, also ein allgemeines Erkennungsproblem zu lösen. Dies können auch wir Menschen nicht. Und es wird keinen Systemdesigner geben, der die Zuordnung $\text{Signal} \longleftrightarrow \text{Symbol}$ für alle Eventualitäten vorprogrammieren könnte.

Vielmehr müssen Systeme entwickelt werden, die in der Lage sind, für jeweils spezifische Situationen, in Abhängigkeit von der Spezifik ihrer rechentechnischen, mechanischen und sonstigen Ressourcen, die Transformation $\text{Signal} \rightarrow \text{Symbol}$ selbst zu erlernen. Dies können sie aber nur, wenn sie auch befähigt sind, die inverse Transformation $\text{Symbol} \rightarrow \text{Signal}$ zu erlernen. Der Wahrnehmungs-Aktions-Zyklus stellt den Rahmen für diese Transformationskette dar, die, wenn sie vom System beherrscht wird, Kompetenz darstellt, Verhalten zu zeigen.

Ein verhaltensbasiertes visuelles System hat spezifische Wahrnehmungsaufgaben unter allgemeinen Annahmen über die Umwelt zu lösen.

Der eigentliche Zweck von Verhalten ist Handeln. Also muß nicht nur Maschinensehen mit Robotik verbunden werden, um durch Erfahrung die Kluft zwischen Signal und Symbol zu schließen. Umgekehrt bedarf Robotik auch der sensorischen Verkopplung mit der Umgebung (speziell durch visuelle Wahrnehmung), um von einem vorprogrammierten Steuerprozeß auf einen wahrnehmungsbasierten Regelprozeß erweitert werden zu können.

Wenn nur allgemeine Annahmen über die Welt zugelassen sind, bedeutet dies auch ein Minimum an notwendigem expliziten Wissen über die Umwelt. Lernen nimmt statt dessen eine zentrale Stellung für die Repräsentation der allgemeinen Zusammenhänge zwischen Wahrnehmung und Handeln ein.

Der mit der Einführung visuell basierter Robotik erwartete Innovationsschub wird sich in verschiedenen Aspekten ausdrücken:

- Roboter werden nicht mehr lediglich für gleichförmig wiederkehrende Tätigkeiten eingesetzt. Vielmehr werden sie in der Lage sein, sich an unterschiedliche Gegebenheiten zu adaptieren. Die Folge der Handlungsabläufe wird komplexer werden.
- Roboter sind nicht mehr auf den Einsatz in Werkhallen, also von Menschen voll auf die Erfordernisse der Roboter abgestimmten Umgebungsbedingungen, angewiesen. Vielmehr werden sie Einsatzfelder erobern, die bisher nur dem Einsatz der menschlichen Intelligenz vorbehalten waren.
- Erstmals bedeutet technischer Fortschritt, falls die weitgesteckten Ziele erreicht werden, daß nicht die Menschen ihre Arbeits- und Lebensbedingungen den Maschinen anpassen müssen. Vielmehr kann die natürliche Umgebung des Menschen als Randbedingung für die Funktionsfähigkeit technischer Systeme definiert werden.

Bis solche Ziele in großer Breite und Konsequenz realisierbar sind, ist noch ein erhebliches Pensum an Forschungsarbeit auf unterschiedlichsten Gebieten zu leisten.

Einige Probleme seien hier genannt:

1. Die postulierte Emergenz komplexer Verhaltensweisen aus einer Fülle primitiver oder einfacher Verhaltensweisen konnte bisher kaum realisiert werden. Die Entwicklung verhaltensbasierter Systeme erfolgt erst seit etwa 4 Jahren mit unterschiedlichen Konzepten und Zielstellungen. Es muß das Ziel sein, die Entwicklung komplexer Systeme prototypisch zu realisieren. Erst an solchen Systemen können deren konkret erreichbare Leistungen studiert werden. Die zu solchen Arbeiten erforderlichen finanziellen Mittel sind erheblich und überschreiten die Etats der Hochschulforschung. Heutige Systeme werden bezüglich konkret zu erreichender bescheidener Funktionalität explizit entworfen. Der Einfluß des Designers (Systemingenieurs) bestimmt völlig die Funktion des Produktes. Emergente Systeme hingegen entwickeln unterschiedliche Funktionalität in unterschiedlichen Situationen.
2. Verhaltensbasierter Entwurf ist die Voraussetzung, daß die Systeme in unterschiedlichen Situationen auch verschiedene Reaktionen zeigen. Die Verkopplung zielorientierten (vom Nutzer geforderten) Verhaltens mit ereignis- oder datengetriebenem Verhalten ist bisher nicht zufriedenstellend gelöst. Diese Verkopplung berührt sehr stark das heute noch als Forschungsgegenstand gesehene Problem der Systemarchitektur. Ziel muß es sein, die Systeme zu befähigen, die Bearbeitung einer Aufgabe konsequent zu verfolgen, auch wenn ihre Realisierung durch äußere Einflüsse behindert wird.
3. Die Ereignis-triggerung des Verhaltens eines Systems stellt einen Verlust an Determiniertheit dar, die wir von den gegenwärtig genutzten "unintelligenten" Maschinen gewohnt sind und die wir aus Gründen unserer Sicherheit auch zum Teil berechtigt

fordern. Allerdings ist dies nur eine scheinbare Nichtdeterminiertheit. Vom Standpunkt des Erlebens der Umwelt durch das technische System bleibt die Reaktion wohl determiniert. Erst das Studium der Verhaltensweisen von kompletten Systemen wird Aufschluß über Einsatzfelder geben, die sich verbieten.

4. Die Verwendung neuronaler Methoden hat auch einen Verlust an Kausalität in dem Sinne zur Folge, als nicht jeder interne Berechnungsschritt des Systems nachvollzogen werden kann. Hierzu ist eine Umgewöhnung erforderlich. Die Wissenschaftskultur seit Aristoteles und Descartes hat bisher nur Produkte hervorgebracht, die in ihrer Funktionalität völliger Kausalität unterworfen waren, insofern als wir Ursache–Wirkungs–Beziehungen explizit modellieren und aufschreiben können. Diese Beziehungen sind aber auf lineare Funktionen beschränkt. Was wir im Zusammenhang mit visueller Robotik beherrschen müssen, sind aber hoch–nichtlineare Abbildungen, die sich einer expliziten Behandlung entziehen. Wir müssen uns entscheiden, was wir wollen — einfache Systeme, die wir voll durchschauen können oder Systeme mit komplexen Leistungen, die uns in ihrer detaillierten Arbeitsweise in gewissen Grenzen verborgen bleiben.
5. Der Grad der Realisierung autonomer visueller Robotik hängt von der Überwindung der Widersprüche zwischen eingeforderter und verfügbarer Rechnerleistung ab. Es ist zu erwarten, daß die Miniaturisierung und der damit verbundene Leistungsanstieg weiter voranschreiten. Heutige Spezialrechner bieten bereits Leistungen von einigen Gigaoperationen pro Sekunde. Es wird eingeschätzt, daß in wenigen Jahren die Rechnerleistungen Teraoperationen pro Sekunde erreichen. Mit dem Anstieg der Rechnerleistung geht die Miniaturisierung einher. Deshalb werden die heute oft noch hemmend wirkenden Zwänge für stationäre Rechnerleistung (nicht auf dem Roboter untergebracht) aufgehoben werden.
6. Der Grad der Realisierung autonomer visueller Robotik hängt auch von der Verfügbarkeit spezieller Rechentechnik ab, welche ein hoch–effektives Umsetzen von Multiagentenkonzepten unterstützt. Der bisher anhaltende Trend, lediglich Universal–Rechnerarchitekturen immer leistungsfähiger und kleiner zu machen, ist zwar sinnvoll für den globalen Ausbau der Informationsgesellschaft. Er unterstützt aber nicht die spezifischen Erfordernisse der Mechatronik und der visuell basierten Robotik. Erst wenn es gelingt, den Bedarf in hinreichender Stückzahl durch überzeugende prototypische Entwicklungen nachzuweisen, werden auch die erforderlichen Spezialrechner entwickelt werden.

Es ist dem verhaltensbasierten Entwurfskonzept visueller Robotik immanent, daß es eine Großfertigung universell einsetzbarer Roboter (wie im Falle der gegenwärtig verfügbaren) nicht geben wird. Vielmehr ist zu erwarten, daß Robotersysteme mit sehr unterschiedlicher Gestalt für sehr unterschiedliche Aufgaben entwickelt werden. Roboter werden als Spezialisten entwickelt. Dies erfordert ein hohes Maß an spezifischem anwendungsorientierten Wissen und große Flexibilität der Produzenten. Dabei ergibt sich eine hervorragende Chance insbesondere für kleine und mittelständige Unternehmen.

Die Stückzahlen können dabei sehr weit aufgefächert sein, je nach Anwendungsfall [IPA, 1994]. Für den Heimgebrauch wird diese Technik noch lange nicht reif sein. Hauptsächlich liegt dies daran, daß die für Maschinensehen erforderliche Rechentechnik teuer ist. Dieser Zustand wird auch gegenwärtig bestimmen, wofür diese Technologie zum Einsatz kommt. Zur Zeit sind prototypische Entwicklungen für die Anwendungen möglich, wo die neuartigen Fähigkeiten zwingend erforderlich sind und die durch Alternativlösungen entstehenden Kosten die relativ hohen Entwicklungs-/Anschaffungskosten der Roboter übersteigen. Diese Situation wird sich in den nächsten Jahren ändern.

Es ist zu erwarten, daß die vor allem in Japan forcierten Serviceroboter für Anwendungen im Bau- und Dienstleistungsbereich bald ein vertretbares Preis-Leistungs-Verhältnis erreichen. Insbesondere auf dem Bausektor werden visuell basierte Roboter dazu beitragen, den längst überfälligen Innovationsschub zur Erhöhung der Arbeitsproduktivität zu initiieren.

Die Chancen und Erfordernisse zur Realisierung von visuell basierten Robotern können folgendermaßen zusammengefaßt werden.

- **Realisierungschancen:** Für lernende, visuell basierte Roboter ist ein breites Spektrum von Einsatzgebieten vorhanden. Der Stand der Technik in den relevanten Gebieten, Robotik, Maschinensehen und Neuroinformatik hat ein Niveau erreicht, welches die Realisierung von spezialisierten, visuell basierten Robotern in naher Zukunft prinzipiell ermöglicht. Beide Aspekte, Bedarf und Realisierbarkeit, werden in den nächsten Jahren und Jahrzehnten dazu führen, daß man weltweit die Herstellung von visuell basierten Robotern mit großer Anstrengung vorantreibt. Autonome oder teilautonome Roboter werden deshalb schon bald in zahlreichen Bereichen der Gesellschaft einsetzbar sein.
- **Erfordernisse:** Zur Entwicklung und Herstellung der neuen Roboter-Technologie sind allerdings noch wichtige Innovationen nötig, wie oben beschrieben, die nur durch enge Kooperation zwischen wissenschaftlichen Einrichtungen und Industrieunternehmen erzielt werden. Während Hochschulen oder andere Forschungsinstitute den nötigen wissenschaftlichen Hintergrund zum Entwurf von visuell basierten, lernenden Robotersystemen bieten, kann nur aufbauend auf der Aufgabenspezifikation und dem Know-How des industriellen Partners eine Realisierung erfolgen. Durch Technologie-Transfer zwischen Industrie und Hochschulen können die Forschungsarbeiten problembezogen gelenkt und frühzeitig Zwischenprodukte genutzt werden. Notwendig ist hierbei die Bereitschaft zu interdisziplinärer Zusammenarbeit. Dies betrifft erstens die relevanten Fachbereiche innerhalb der wissenschaftlichen Einrichtungen, zweitens die relevanten Firmen oder Teilbereiche innerhalb von Industrieunternehmen und drittens die Zusammenarbeit von Industrie und Hochschulen.

4.2 Situation und Chancen in Schleswig–Holstein

Anliegen der Studie war es, für die innovative Technologie der visuell basierten Roboter im Land Schleswig–Holstein die Aufmerksamkeit zu schärfen und zur Bündelung von Aktivitäten die erforderlichen Informationen bereitzustellen. Die Chancen für Industriebetriebe und Hochschulen in Schleswig–Holstein, an der Realisierung dieser Hochtechnologie teilzunehmen, müssen in Abhängigkeit von der aktuell gegebenen Situation und Infrastruktur beurteilt werden.

Folgende drei Aspekte sind wesentlich:

- Sind an den Hochschulen und Forschungsinstituten des Landes relevante Fachbereiche vorhanden, sowie die Bereitschaft zu interdisziplinärer und industriebezogener Zusammenarbeit?
- Gibt es Industriebetriebe im Bundesland, welche bereits jetzt an der Herstellung von Robotern oder Produkten des Robotersehens oder Roboterlernens arbeiten?
- Gibt es im Land Schleswig–Holstein Industrieunternehmen, die Roboter bereits jetzt einsetzen und sich durch die Einführung von visuell basierten Robotern Vorteile versprechen?

Alle drei Fragen können eindeutig positiv beantwortet werden. Die nachfolgenden Tabellen (4.1, 4.2 und 4.3) belegen, daß landesweit an vielen Stellen bereits jetzt Konzepte entwickelt sowie Produkte hergestellt und/oder eingesetzt werden, welche für die neue Robotertechnologie von zentraler Bedeutung sind.

Die erste Tabelle führt wissenschaftliche Hochschulen oder Fachhochschulen sowie eigenständige Institute an, welche an Verfahren der Robotik, des Maschinensehens oder des Maschinenlernens arbeiten. Die zweite Tabelle benennt Hersteller von relevanten Produkten der visuell basierten Robotik, insbesondere Roboter oder Roboterkomponenten, Systeme zum Maschinensehen sowie Rechentechnik. Die dritte Tabelle gibt einen Überblick über kleine und mittelständige Industriebetriebe sowie größere Unternehmen, welche Roboter oder Systeme des Maschinensehens im Einsatz haben. Diese Firmen sind auch die potentiellen Anwender für visuell basierte Robotik.

Hochschulen, Institute	Forschungsrichtung Ausstattung
Christian-Albrechts-Universität (CAU), Kiel, Technische Fakultät, Lehrstuhl für Kognitive Systeme	Verhaltensbasierte Roboter und technische visuelle Systeme, Fahrzeugnavigation und Hand-Auge-Koordination; Mobile Roboter mit Roboterkopf inklusive Stereokamera, Artikulation-Manipulator.
CAU, Kiel, Technische Fakultät, Lehrstuhl für Automatisierungs- und Regelungstechnik	Nichtlineare Regelung, Fuzzy-Regelung, künstliche neuronale Netze; Artikulation-Manipulator, weitere Anlagen zur Regelung.
Fachhochschule, Kiel, Institut für CIM Technologie Transfer, Fachgebiet Robotik	Simulation von Bewegungsabläufen, Simulation und Tests von Beanspruchungen; Mehrere Industrieroboter mit Sensorik.
Fachhochschule, Kiel, Fachbereich Technik, Labor für Getriebe, Robotik und Sensorik	Automatisierung bei der Fertigung, Getriebe-Design, Schnittstellenprobleme; Roboter, Automatisierter Meßplatz, Hochgeschwindigkeitskameras.
Fachhochschule, Kiel (Standort Eckernförde), Fachbereich Bauwesen, Institute für Bauchemie und Materialprüfung	Automatisierung am Bau, Photometrische Aufnahme von Bauschäden, Prüfverfahren für Betone; Laborgeräte für photometrische und chemische Analysen.
Fachhochschule, Lübeck, Fachbereich Elektrotechnik, Labor für Mikromechanik und Labor für Mikroprozessortechnik	Leiterplattenentwurf, Drahtbonden, programmierbare Steuerungen, Sensor-Signalverarbeitung; CAD-Systeme, Bestückungsgeräte, Laserschneidegeräte.
Fachhochschule, Wedel, Fachbereich Technische Informatik, Institut für Robotik, Bildverarbeitung und CIM-Anwendungen	Industrielle Bildverarbeitung, neuronale Netze, Fuzzy Logik, Robotik, intelligente Sensorik, CIM-Technologien; Fertigungszellen mit Scara- und Artikulation-Roboter, CNC-Fräse.
Fraunhofer-Institut für Siliziumtechnologie (ISIT), Itzehoe	Miniaturisierte Systeme in Silizium, anwendungsspezifische Bauelemente, Integration von Sensoren und Effektoren.
GEOMAR Technologie GmbH, Kiel, Forschungszentrum für marine Geowissenschaften	Entwicklung neuer Systeme für die Meeresbodenforschung, Unterwasserrobotik, Bildauswerteverfahren; Verschiedene Großgeräte, Elektronik und Feinmechanik-Werkstatt.

Tabelle 4.1: Hochschulen (Institute) mit relevanter Forschungsrichtung.

Hersteller	Produktpalette
Anschütz & Co GmbH, Kiel	Automatisierte Reinigungsmaschinen, Kreiselkom- passe zur Selbstlokalisierung und -orientierung.
Computer & Motion, Raisdorf	Verfahren zur Kompression und Dekompression von Videodaten.
Deutsche Systemtechnik – In- dustriautomation GmbH (DST- DIA), Kiel	Roboter zur Holzbearbeitung, automatisierte Her- stellung von Möbeln, visuelle Überwachung der Fertigung.
Hako-Werke, Bad Oldeslohe	Automatisierte Reinigungsmaschinen für Flure, Putzautomaten mit Ultraschall-Sensoren.
Ingenieurbüro Atlas Kiel GmbH & Co KG (IBAK), Kiel	Spezialfernsehanlagen zur Inspektion von Kanal- rohren, Brunnen, Reaktoranlagen.
ISATEC Software- und Hardware GmbH (ISATEC), Kiel	Spezial-Parallelrechner für Bildverarbeitung und neuronale Netze.
Maschinenbau und Konstruktion GmbH (MUK), Elmshorn	Handhabungssysteme aller Art.
Telerobotik Gesellschaft für Fern- hantierungstechnik (TELEROB), Kiel	Master/Slave-Manipulatorsysteme, ferngesteuerte Kettenfahrzeuge und Manipulatoren.

Tabelle 4.2: Hersteller von relevanten Produkten.

In Schleswig-Holstein sind die grundlegenden, strukturellen Voraussetzungen für die Realisierung von visuell basierten Robotern sehr wohl gegeben. Allerdings fehlt in den möglichen Entwicklungs- und Einsatzstätten häufig noch das nötige Know-How in Robotik, Maschinensehen oder Maschinenlernen. Dies liegt auch daran, daß eine eventuell gewünschte Ausstattung, beispielsweise Spezialrechner zur Bildauswertung, in der erforderlichen Spezifikation nicht verfügbar oder zu teuer ist. In diesem Zusammenhang muß man bedenken, daß sich die Realisierung der neuen Robotersysteme weltweit erst in den Anfängen befindet.

Letzteres eröffnet aber gerade die Chance, von Beginn an bei der Umsetzung einer zukünftigen Schlüsseltechnologie in die industrielle Praxis eine gewichtige Rolle zu spielen. Das in Schleswig-Holstein zweifelsfrei vorhandene Interesse an visuell basierter Robotik sollte genutzt werden, um durch gezielte Förderung dieser Zukunftstechnologie die industrielle Fortentwicklung des Landes weiter voranzubringen.

Anwender	Einsatzgebiete
Dräger–Werke, Lübeck	Manipulatoren zur Fertigung von medizinischen Geräten, Bildauswertesysteme zur Unterstützung der Diagnose, Verwaltung von Bild–Datenbanken.
Ethicon, Norderstedt	Fertigung von Instrumenten der minimal–invasiven Chirurgie, Roboterarm zur autonomen Kameraführung.
Hagenuk AG, Kiel	Manipulatoren zur Fertigung von Mobilfunktelefonen.
Howaldtswerke Deutsche Werft AG (HDW), Kiel	Handhabung beim Schiffsbau, Transportaufgaben, Kamerasysteme zur Inspektion von Schäden.
Kuhnke GmbH, Malente	Manipulatoren zur Fertigung von Steuerungen, Bestückung von Schaltungen, Kamerasysteme zur Qualitätskontrolle der Produkte.
Linotype–Hell AG, Kiel	Greiftechnik zur Handhabung von Druckmaschinen, Papierbestückung, Austausch von Farbbehältern.
Quarzglastechnik, Geesthacht	Manipulatoren zur Herstellung von Glas.
Roditec, Heikendorf	Sanierung maroder Kanalrohre mit mobilen Robotern, Rohrabdichtung mit Manipulatoren.

Tabelle 4.3: Mögliche Anwender und Einsatzgebiete.

Die Chancen der neuen Roboter–Technologie sind vielfältig. Das Spektrum reicht von der Erschließung neuer gewinnbringender Märkte bis zur Schaffung neuer hoch qualifizierter Arbeitsplätze.

Index

- Adaptive Regelung, 47
- Agent, 8
- Aktive visuelle Wahrnehmung, 20
- Aktives Beobachtungssystem, 21
- Aktuator-System, 22
- Akzeptanz, 12
- Antriebs-System, 28
- Appearance-Wissen, 52
- Artikulation-Roboterarm, 24
- Aufmerksamkeitssteuerung, 20, 68
- Automatisierung, 3
- Autonomes Verhalten, 8
- Autonomie, 6

- Back-Propagation-Regel, 46
- Bahnplanung, 61
- Bewegungstrajektorie, 30
- Bildauswertung, 33
- Bildkodierung, 33
- Bildkompression, 33
- Bildsegmentierung, 36
- Bildverarbeitung, 33
- Bildverbesserung, 33
- Bildverstehen, 33
- Blickrichtungssteuerung, 62
- Blinder Roboter, 50
- Boltzmann-Maschine, 53
- Bottom-Up-Kontrolle, 63

- CCD-Chip, 27

- Differenzbildtechnik, 51
- Diffusionsprozeß, 39
- Direkte inverse Modellierung, 47
- Diskretes Ereignissystem, 61
- Drehmoment-Sensor, 26

- Effektor, 22
- Erklärungsbasiertes Lernen, 43
- Externe Sensoren, 26
- Eye-in-Hand, 31
- Eye-on-Hand, 31
- Farb-Kamera, 27

- Fehlerminimierungsverfahren, 45
- Fehlerrückkopplung, 47
- Fehlersignal, 45
- Formveränderung, 16
- Frame-Grabber, 28
- Freiraumtunnel, 67
- Funktionale Modularisierung, 31
- Funktionsapproximation, 47

- Gantry-Roboterarm, 24
- Geometrische Invariante, 42
- Globale Navigation, 18
- Grobkörnig paralleler Rechner, 70

- Hauptprozessor, 29
- Hebb-Lernregel, 46
- Heterarchische Beziehung, 63

- Implizite Repräsentationsform, 43
- Infrarot-Kamera, 27
- Infrarot-Sensor, 27
- Intelligenz, 4
- Interline-Modus, 27
- Interner Sensor, 26
- Invarianz, 40
- Inverse Dynamik, 29, 47
- Inverse Kinematik, 29, 47

- Kalibrierung, 42
- Kameranachführung, 21
- Kameraplattform, 74
- Kantendetektion, 36
- Kompetenz, 4
- Kontakt-Sensor, 26
- Kraft-Motor-Abbildung, 48
- Körperlichkeit der Systeme, 10
- Künstliche Intelligenz, 4
- Künstliche Neuronale Netze, 44

- Labyrintharchitektur, 72
- Lageveränderung, 15
- Laser-Distanz-Sensor, 27
- Laser-Kamera, 27
- Laufmaschine, 18

Least-Mean-Square Methode, 49
 Lichtstrahlung, 27
 Lokale-spektrale Repräsentation, 72
 Lokales Manövrieren, 17

 Manipulation, 15, 23
 Manipulator, 24
 Maschinelles Kopilot, 51
 Maschinensehen, 33
 Master/Slave-System, 30
 Matchen, 38
 Mechatronik, 8
 MIMD-Rechner, 70
 Monochrom-Kamera, 27
 Mustererkennung, 33

 Neurobiologie, 34
 Neuron, 44
 Neuronale Hardware, 68

 Odometrie, 26
 Offline-Programmierung, 30
 Okulomotorik, 75
 Optischer Fluß, 51
 Ortsvariante Auflösung, 74

 Pan-Bewegung, 24
 Passiv-wahrnehmender-Roboter, 31
 Photometrische Invariante, 40
 Pipeline-Rechner, 68
 Programm-Roboter, 29
 Projektive Geometrie, 42

 RBF-Netz, 53
 Rechner-System, 28
 Regularisierung, 41
 Reinforcement-Lernverfahren, 46
 Rekonstruktion der Szene, 34
 Relationaler Graph, 61
 Repräsentationsschema, 43
 Roboter, 3
 Roboterarm, 24
 Roboterkopf, 24
 Roboterlernen, 42
 Roboterrelevantes Maschinenlernen, 42
 Roboterrelevantes Maschinensehen, 35
 Robotersehen, 35

 Robotik, 3

 Scara-Roboterarm, 24
 Schrittmotor, 28
 Selbstlokalisierung, 62
 Selbstorganisierendes Lernen, 49
 Sensor-System, 26
 Service-Roboter, 11
 Signalprozessor, 68
 SIMD-Rechner, 55
 Stereoskopisches System, 21
 Strahlungs-Sensor, 27
 Subsumtions-Architektur, 63
 Symbolverankerung, 44

 Taktile Sensor, 26
 Teach-in-Programmierung, 30
 Teleoperation, 18
 Teleoperation-Roboter, 30
 Tilt-Bewegung, 24
 Top-Down-Kontrolle, 64
 Topographische Karte, 61
 Transportroboter, 62

 Ultraschall-Sensor, 26
 Ultraschall/Infrarot-Motor-Abbildung, 48
 Unstrukturierte Umgebung, 4
 Unüberwachtes Lernen, 46

 Vergenz-Regelung, 24
 Verhaltenbasierte Modularisierung, 63
 Verhaltensbasiertes Robotersystem, 58
 Visual Homing, 61
 Visuell basierte Kursregelung, 18
 Visuell basierte Robotik, 6
 Visuell basiertes Fortbewegen, 17
 Visuelle Verhaltensweise, 67
 Visuelle Wahrnehmung, 4
 Visuelles Informations-Management, 58
 Visuomotorische Abbildung, 48
 Vorwärtsmodellierung, 47

 Wahrnehmungs-Aktions-Zyklus, 6
 Wahrnehmungspsychologie, 34
 Wavelet, 39
 Wettbewerbs-Lernen, 46

Literatur

- R. Brooks und M. Mataric. *Robot Learning*, Kapitel: Real Robots, Real Learning Problems, Seiten 193–214. Kluwer Academic Press, 1993.
- R. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2:14–23, 1986.
- I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.
- E. DeYoe und D. Van Essen. Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience*, 11:219–226, 1988.
- E. Dickmanns und V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1:241–261, 1988.
- R. Dillmann und S. Bocionek (Hrsg.). *KI und Robotik*. Nummer 3 in Künstliche Intelligenz, Journal von GI-Fachbereich 1. Interdata-Verlag, 1994.
- K. Friebe. Alltagskultur des Wandels — Wandel der Alltagskultur. *Technologiedialog, Magazin der Technologiestiftung Schleswig-Holstein*, 20:3–4, 1996.
- M. Hansen und G. Sommer. Real time vergence control using local phase differences. In: *Fourth International Conference on Computer Graphics and Image Processing*. Machocice, Polen, Mai 20–24, 1996.
- IPA (Hrsg.). *Serviceroboter — ein Beitrag zur Innovation im Dienstleistungswesen*. Fraunhofer-Institut für Produktionstechnik und Automatisierung (IPA), Projektleitung M. Hägele, Stuttgart, 1994.
- R. Jain, A. Pentland und D. Petkovic. Workshop report: NSF-ARPA Workshop on Visual Information-Management-Systems. World Wide Web: <http://www.virage.com/vir-res/reports>, 1995.
- R. Jain und T. Binford. Ignorance, myopia and naiveté in computer vision systems. *Computer Vision, Graphics, and Image Processing – Image Understanding*, 53:112–117, 1991.
- D. Koller. *Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Straßenverkehrsszenen*. Dissertationen zur Künstlichen Intelligenz, Band 13, Infix-Verlag, Sankt Augustin, 1992.
- S. Kosslyn. *Image and Brain*. MIT Press, 1994.
- J. Košecka, H. Christensen und R. Bajcsy. Discrete event modelling of visually guided behaviors. *International Journal of Computer Vision*, 14:179–191, 1995.
- J. Lückel (Hrsg.). *Third Conference on Mechatronics and Robotics*. Paderborn, Deutschland, Oktober 4–6, Teubner-Verlag, Stuttgart, 1995.
- H.-P. Mallot und B. Schölkopf. Learning of cognitive maps from sequences of views. In: *Third European Symposium on Artificial Neural Networks*. Brüssel, Belgien, April 19–21, 1995.
- H. Marguerre. Echoortende Fledermäuse und sensorgeführte Roboter. *Elektronik*, (18):38–41, 1993.

- K. Moeller und G. Pass. Künstliche Neuronale Netze: eine Bestandsaufnahme. *Künstliche Intelligenz*, (4):37–61, 1994.
- H.-H. Nagel (Hrsg.). *Sichtsystemgestützte Fahrzeugführung und Fahrer–Fahrzeug Unterstützung*. Infix–Verlag, Sankt Augustin, 1995.
- H. Neumann. Neuroinformatische Modellbildung zur visuellen Informationsverarbeitung. *Künstliche Intelligenz*, (6):27–31, 1995.
- N. Nise. *Control Systems Engineering*. Benjamin/Cummings Publishing Company, Redwood City, California, 1995.
- R. Reddy. Grand Challenges in AI. *ACM Computing Surveys*, 27:301–365, 1995.
- C. Reuber. Vision wird Wirklichkeit. *Elektronik*, (25):48–53, 1994.
- H. Ritter, T. Martinetz und K. Schulten. *Neuronale Netze*. Addison–Wesley, 1991.
- G. Sommer. Verhaltenbasierter Entwurf technischer visueller Systeme. *Künstliche Intelligenz*, (3):42–45, 1995.
- C. Torras. Robot adaptivity. *Robotics and Autonomous Systems*, 15:11–23, 1995.