Bildverstehen in der KI: Konzepte und Probleme[*]

ANDREAS SCHIERWAGEN

Prof. Dr. Andreas Schierwagen, Universität Leipzig, Institut für Informatik, Augustuspl.10, 04109 Leipzig; e-Mail: schierwa@informatik.uni-leipzig.de.

Zusammenfassung

Nach einer Übersicht über die Entwicklung des Fachgebietes Computersehen seit den 50er Jahren erläutern wir den konzeptuellen Rahmen, in dem Bildverstehen als hierarchischer, wissensbasierter Prozeß beschrieben wird. Wir kennzeichnen das Verstehensproblem des Computersehens als Variante der Frage, ob semantische Maschinen möglich sind. Mit dem Symbolverankerungsansatz stellen wir einen Ansatz vor, der einem Symbolsystem den Zugang zu seiner Umwelt ermöglichen soll. Durch die angestrebte Verankerung der internen Repräsentationen in visueller etc. "Erfahrung" kann aber eine Lösung des Verstehensproblems im eigentlichen Sinne nicht erreicht werden. Wir argumentieren dafür, daß die Redeweise von bildverstehenden Systemen in der KI nur metaphorisch sein kann.

Summary

After an outline of the development of the field of Computer Vision since the 50's we explain the conceptual framework within which Image Understanding has been described as hierarchical, knowledge-based process. We characterize the problem of understanding in the field of Computer Vision as a variation of the question whether semantic machines are possible. We discuss the symbol grounding approach which is to enable a symbol system to achieve access to its environment. This approach is to cause the internal representations to be grounded in visual etc. "experience". In this way, a real solution for the understanding problem cannot be obtained, however. We argue that we should speak of image understanding systems in AI only in a metaphorical sense.

[*] In: Vom Realismus der Bilder. Interdisziplinäre Forschungen zur Semantik bildhafter Darstellungsformen. (K. Sachs-Hombach, K. Rehkämper, Hg.). Scriptum Verlag, Magdeburg 1999

1. Einleitung

Bildverstehen (Image Understanding) wird im Rahmen der Forschungen zur Künstlichen Intelligenz (KI) dem Maschinensehen (Computer Vision) zugeordnet. Die Zielstellungen des Bildverstehens bilden einen Schwerpunkt bei den Bemühungen, eine Maschine zur "intelligenten" Interaktion mit ihrer Umwelt zu befähigen. Mit Sensoren gewinnt sie Information aus ihrer 3D-Umwelt, die in Form von natürlicher Sprache, Bildern, Geräuschen u.ä. auftreten können. Diese Information wird weiter verarbeitet, um zu verschiedenen Formen interner Repräsentation zu gelangen, die wiederum die Interaktion mit der Umwelt ermöglichen, sei es in sprachlicher Form oder durch Handlungen eines Roboters. Die internen Repräsentationen stellen das "Wissen" bzw. die "Modelle" des wissens- bzw. modellbasierten Computersehens dar. Die Komplexität dieser Verarbeitungsschritte wird nach der herkömmlichen Methodologie dadurch bewältigt, daß jedes Problem auf drei voneinander unabhängigen Ebenen formuliert und untersucht wird - den Ebenen der Kompetenztheorie, des Algorithmus und der Implementation (MARR 1982).

Wissensbasiertes Computersehen definiert sich als Teil der "traditionellen" KI. Die zentralen Konzepte sind Symbolverarbeitung und Repräsentation; als theoretischer Rahmen dient einerseits die Hypothese des physikalischen Symbolsystems von NEWELL & SIMON (1976), andererseits die Konzeptualisierung des Sehens als Prozeß der Informationsverarbeitung von MARR (1982). Diesem Zugang inhärent ist das sog. Symbolverankerungsproblem, das darin besteht zu erklären, wie ein (natürliches oder künstliches) kognitives System kausal mit seiner Umwelt verknüpft ist / werden kann, so daß sowohl sein Verhalten als auch die zugrunde liegenden Mechanismen, Repräsentationen usw. bedeutsam für es selbst sind und Bedeutung nicht erst durch einen externen Designer / Beobachter zugewiesen wird.

Das Verstehensproblem des Computersehens stellt sich damit als Variante der Frage dar, ob semantische Maschinen möglich sind. Der Beitrag kennzeichnet den Semantikbegriff der klassischen KI als den einer internalistischen Semantik, mit der ein Zugang zur Welt nicht erreicht werden kann. Wir stellen weiter den Symbolverankerungsansatz vor, der einem Symbolsystem den Zugang zu seiner Umwelt ermöglichen soll. Damit wird eine Verankerung der intrinsischen Repräsentationen in sensorischer (visueller etc.) "Erfahrung" angestrebt, eine Lösung des Verstehensproblems im eigentlichen Sinne aber nicht erreicht. Wir argumentieren dafür, daß die Rede von bildverstehenden Systemen in der KI nur metaphorisch sein kann.

2. Entwicklung von Computersehen

Die Wissenschaft vom Computersehen war in den vergangenen vier Jahrzehnten mehreren Wechseln der Perspektive unterworfen (vgl. z.B. NEUMANN 1993; CROWLEY & CHRISTENSEN 1995). Ihre Anfänge rühren aus den 50er Jahren, in denen erste Versuche unternommen wur-

den, die neuen Rechenmaschinen für die Verarbeitung von Bildern einzusetzen. Die Periode 1965-1975 ist dadurch gekennzeichnet, daß Sehen als Problem der ***Mustererkennung*** angesehen wurde. Dabei wird ein Objekt durch einen Satz von Merkmalen beschrieben. Die Ähnlichkeit von Objekten wird definiert durch den quantifizierbaren Grad der Übereinstimmung der Merkmalssätze, die die Objekte beschreiben. Eine Übersicht über Arbeiten aus dieser Zeit geben DUDA & HART (1972).

Der Mustererkennungsansatz stieß bald auf verschiedene grundsätzliche Schwierigkeiten. Ein Hauptproblem war die Segmentierung des Bildes in bedeutungsvolle "Chunks", die klassifiziert werden konnten. Speziell das Segmentierungsproblem erwies sich als unlösbar im allgemeinen Fall. Es wurde offensichtlich, daß Segmentierung mehr als nur Messungen im Bild verlangt. Nur in Bezug auf den intendierten Gebrauch läßt sich die geeignete Segmentierung definieren. Schließlich setzte sich die Erkenntnis durch, daß Maschinensehen ein Verständnis der Welt verlangt, die im Bild dargestellt ist. Damit vollzog sich der Übergang zu der Auffassung, daß Sehen ein Anwendungsfeld für KI-Techniken sei.

Die Einsicht, daß Weltwissen zur Segmentierung erforderlich ist, führte zu dem Ansatz, Sehen als ***Bildverstehen*** zu untersuchen. Diese Umorientierung erfolgte in den 70er Jahren, als in der KI neue Techniken zur Programmierung von Expertensystemen entwickelt wurden, insbesondere Verfahren der Wissensrepräsentation und der Inferenz. Die Erwartung war, daß es mit diesen Verfahren möglich sein würde, das Weltwissen zu erfassen, das für die Analyse und das Verstehen von Bildern benötigt wird. Einen repräsentativen Überblick über Arbeiten aus dieser Periode gibt der Sammelband von HANSON & RISEMAN (1978).

Auch der Bildverstehensansatz stieß bald auf Schranken, die seinen Erfolg stark einschränkten. Vor allem erwies sich die Aufgabe, das nötige Weltwissen zu erfassen und zu formalisieren, nur für sehr eingeschränkte Domänen als durchführbar. Das Segmentierungsproblem kann mit dem Bildverstehensansatz nicht gelöst werden. Ein wesentlicher Grund ist der, daß die meisten KI-Verfahren sehr empfindlich auf Qualtätsmängel bei der Bildsegmentierung reagieren. Die Segmentierung zu Beginn der Bildverarbeitung stellt auch heute ein wichtiges Problem dar, an dem viele zunächst erfolgversprechende Algorithmen scheitern.

Ein anderer Ansatz betrachtet als Voraussetzung für das Verstehen eines Bildes, daß man von den 2-dimensionalen (2D-) Mustern von Grauwerten oder Farbintensitäten auf die 3-dimensionale (3D-) Form der Objekte zurück schließen kann. Dieser ***Rekonstruktionsansatz*** wurde von MARR (1982) und seinen Kollegen am MIT zum bislang einflußreichsten Konzept für das Maschinensehen entwickelt. Auf dieser Grundlage wurden verschiedene Verfahren mit dem Ziel angegeben, ausgehend von Bildmerkmalen wie Schattierung, Textur, Kontur, Bewegung usw. die Form von abgebildeten Objekten zu rekonstruieren. Für viele dieser sog. Shape-from-X-Verfahren konnte nachgewiesen werden, daß sie im mathematischen Sinn inkorrekt gestellte Probleme darstellen. Das bedeutet, daß der Rekonstruktionsansatz i.a. keine eindeutigen Lösungen liefert. Eindeutigkeit bei der Rekonstruktion läßt sich vielfach errei-

chen, wenn kontrollierte Kamerabewegungen eingesetzt werden, d.h. Bilder der Szene von verschiedenen Blickwinkeln aus gewonnen werden (ALOIMONOS et al. 1988).
Dieser Beitrag des ***Aktiven Sehens*** war zunächst noch in den Rahmen des Rekonstruktionsansatzes eingebettet. Seit Beginn der 90er Jahre hat sich die Modellierung eines Sehsystems als aktiv handelnder Agent als eigenständiges Forschungsgebiet herausgebildet. Damit wird Kritik an der Konzeption von KI-Maschinen als wissensbasierte Systeme aufgegriffen. Sehen wird nicht länger als passiver Rekonstruktionsprozeß angesehen, sondern als ein Prozeß der selektiven Datenaufnahme in Raum und Zeit. Eine Theorie des Sehens sollte nach dieser Auffassung das Interface zwischen Wahrnehmung und anderen Kognitionen wie Problemlösen, Planen, Lernen und Handeln bereitstellen. Im Rahmen dieses Ansatzes rücken Aspekte der Aufmerksamkeit, der Zielorientierung und Zweckgebundenheit in den Vordergrund (SOMMER 1995; SCHIERWAGEN 1998).
Parallel dazu gibt es Projekte, die den wissensbasierten Ansatz weiterführen. Ausgangspunkt ist die Annahme, daß Objekterkennung den Vergleich der Objekte mit internen Repräsentationen von Objekten und Szenen im bildverstehenden System (BVS) beinhaltet. Aus einer komputationalen Perspektive (auf der Ebene des Algorithmus und der Repräsentation) ergeben sich verschiedene Möglichkeiten der Realisierung. Während MARR die datengetriebene Rekonstruktion der visuellen Objekte realisieren wollte, schlagen andere Forscher einen ***bildbasierten Zugang*** vor (vgl. die Übersicht in TARR & BÜLTHOFF 1998). Dieser Ansatz benötigt keine Rekonstruktion im Sinne einer Berechnung von 3D-Repräsentationen. Statt dessen repräsentieren bildbasierte Modelle Objekte durch deren Bild von einem bestimmten Blickwinkel aus. Ein solches Vorgehen erfordert robuste Matching-Algorithmen, um die perzeptuelle Ähnlichkeit eines Inputbildes mit einem bekannten Objekt zu bestimmen. TARR & BÜLTHOFF (1998) plädieren zusammenfassend für eine Konzeption der Objekterkennung, die Aspekte von rekonstruktions- und bildbasierten Modellen vereint.

3. Wissensbasiertes Maschinensehen heute

Der beschriebene mehrfache Wandel der Leitideen des Forschungsgebietes Bildverstehen ist nicht von allen Forschern konsequent vollzogen worden, so daß heute verschiedene konzeptuelle Sichten nebeneinander weiter bestehen. Eine weitgehende, verschiedene Ansätze zusammenfassende Definition wurde von NEUMANN (1993, 567) vorgeschlagen:
"Bildverstehen ist die Rekonstruktion und Deutung einer Szene anhand von Bildern, so daß mindestens eine der folgenden operationalen Leistungen erbracht werden kann:

- Ausgabe einer sprachlichen Szenenbeschreibung
- Beantwortung sprachlicher Anfragen bezüglich der Szene
- kollisionsfreies Navigieren eines Roboters in der Szene
- planmäßiges Greifen und Manipulieren von Objekten in der Szene."

Diese Definition schließt Bilddeutung ein und legt damit den Akzent auf das Verstehen. Der gewählte operationale Verstehensbegriff soll sichern, daß nicht etwa der Programmierer des BVS die Verstehensleistung erbringt, sondern tatsächlich das System selbst. Eingaben für das System sind Bilder einer Kamera, aus denen in einem mehrstufigen Prozeß eine Repräsentation der Umweltszene, die die Bilder verursacht hat, gewonnen werden soll. Als Szene wird dabei ein räumlich-zeitlicher Ausschnitt der Umwelt bezeichnet. Statische Szenen sind i.a. 3-dimensional, dynamische Szenen 4-dimensional. Ein Bild ist eine 2D- Projektion einer statischen Szene; dynamische Szenen führen auf Bildfolgen. Die computerinterne Beschreibung der Szene als Ausgabe besteht aus zwei Teilen: (i) Information über die räumlich-zeitlichen Beziehungen der Objekte einer Szene und (ii) Deutung des Szeneninhaltes, speziell Objekterkennung. Für die interne Repräsentation der Szenenbeschreibung werden Wissensrepräsentations-Methoden und Inferenz-Verfahren (insbesondere räumliches Schließen) eingesetzt.

Der konzeptuelle Rahmen, in dem in der kognitivistisch orientierten KI Sehen untersucht wird, ist in Abb. 1 im Überblick dargestellt. Bildverstehen als Prozeß wird durch das Zusammenwirken von vier aufgabenspezifischen Teilprozessen beschrieben, die jeweils spezifische Zwischenrepräsentationen erfordern.

Die ***primäre Bildanalyse*** geht vom digitalen Rasterbild aus, in dem die radiometrischen Eigenschaften (Intensität und ggf. Farbe) jedes Bildpunktes erfaßt sind, und gelangt zur Bestimmung von Bildelementen (Kanten, homogene Bereiche, Textur u.a.).

Die ***niedere Bilddeutung*** hat das Ziel, Bildelemente als Szenenelemente zu interpretieren, d.h. als Resultate der Abbildung von Teilen einer 3D-Szene. Prozesse dieser Stufe sollen eine zentrale Aufgabe des Bildverstehens lösen: die Extraktion von Realwelteigenschaften aus Bildeigenschaften. Dazu zählt insbesondere die Rekonstruktion der 3D-Objektformen mittels sog. Shape-from-X-Verfahren.

Im anschließenden Verarbeitungsschritt, der ***Objekterkennung***, werden Objekte in den bisher extrahierten Bilddaten und auf der Grundlage der Szenenelemente identifiziert. Eine entscheidende Rolle spielt dabei das Vorwissen darüber, welche Ansichten von der Kamera erzeugt werden, wenn Objekte aus unterschiedlichen Blickwinkeln betrachtet werden. Dieses Vorwissen ist in Gestalt von Objektmodellen in der Wissensbasis verfügbar.

Die ***höhere Bilddeutung*** faßt weitere Verarbeitungsschritte zusammen, die das Ziel haben, "objekt- und zeitübergreifende Zusammenhänge zu erkennen, z.B. interessante Objektkonfigurationen, spezielle Situationen, zusammenhängende Bewegungsabläufe u.a. Ähnlich wie bei der Objekterkennung spielt hier modellhaftes Vorwissen über das, was man erkennen will, eine wichtige Rolle" (NEUMANN 1993, 570). Der Inhalt der resultierenden Beschreibung hängt nicht nur von der Szene bzw. dem zugehörigen Bild ab, sondern auch von der Fragestellung bzw. dem Kontext, in dem die Ausgabe benutzt werden soll.

Obwohl aktuelle wissensbasierte Bildverstehens-Systeme keine strikte Unterteilung in hierarchisch organisierte Teilprozesse aufweisen, orientieren sie sich weiter an dem skizzierten

konzeptuellen Rahmen. Sie verfügen über eine interaktiv-hierarchische Architektur, in der Teilergebnisse früherer Verarbeitungsschritte Prozesse auf höheren Ebenen auslösen, deren Ergebnisse auf die Verarbeitungsschritte niederer Ebenen rückwirken. Beispiele dafür sind wissensbasierte Systeme zur Integration von Maschinensehen und Verarbeitung natürlicher Sprache (vgl. z.B. HILDEBRANDT et al. 1995; HERZOG et al. 1996; PAULI et al. 1995 und die dort angegebene Literatur).

Wissensarten	**Repräsentationsebenen**	**Teilprozesse**
Alltagswissen Situationsmodelle Vorgangsmodelle	Vorgänge Situationen Objektkonfigurationen	*höhere Bilddeutung*
Objektmodelle	Objekte, Trajektorien	*Objekterkennung*
projektive Geometrie Photometrie	Szenenelemente: 3D-Oberflächen Volumina, Konturen	*niedere Bilddeutung*
Physik allgemeine Realwelteigenschaften	Bildelemente: Kanten, Bereiche Textur, Bewegungsfluß	*primäre Bildanalyse, Segmentierung*
	digitales Rasterbild (Rohbild)	

Abb. 1: Bildverstehen als hierarchischer, wissensbasierter Prozeß. Dargestellt sind die unterschiedlichen Wissensarten, mit denen der Rückschluß vom Bild auf die Szene realisiert werden soll (links) und die Zwischenrepräsentationen auf den verschiedenen Ebenen (Mitte), die durch die entsprechenden Teilprozesse (rechts) erzeugt werden (nach NEUMANN 1993, 568).

4. "Verstehen" bei KI-Maschinen

Nach der Darstellung des Bildverstehens als Prozeß aus verschiedenen Verarbeitungsschritten wenden wir uns der Frage zu, welcher Verstehensbegriff hier zugrunde gelegt wird und wodurch bzw. an welcher Stelle in diesem Prozeß "Verstehen" stattfindet.
In der o.g. Definition von Bildverstehen (Abschnitt 3) ist es zum einen damit verbunden, daß Objekte einen Namen zugewiesen bekommen. Dies kann durch verschiedene Matching-Verfahren erreicht werden, d.h. Resultate der niederen und mittleren Bildverarbeitung werden auf der hohen Repräsentationsebene (Bild- oder Szenenbeschreibung) mit gespeicherten Objekt-Modellen bezüglich Übereinstimmung verglichen. Im Fall eines aktiven Roboters stellt sich die Frage anders dar. Nicht das "explizite" Verstehen durch Designation ist wichtig, sondern der "implizite" Beweis für Verstehen, indem es in seiner Interaktion mit der Umwelt angepaßtes Verhalten zeigt.
In beiden Fällen dient also ein ***Turing-Test*** zur Einschätzung (durch uns als Beobachter), ob das BVS die Szene versteht: das (sprachliche oder sensomotorische) ***Verhalten*** wird als Kriterium für Verstehen verwendet. Die Aussagefähigkeit von Turing-Tests zur Verstehenskapazität von Symbolsystemen ist in der KI und in der Philosophie des Geistes umstritten. Die Kritik richtet sich gegen die physikalische Symbolsystemhypothese (PSSH) von NEWELL & SIMON (1976), nach der ein physikalisch realisierter Zeichenmanipulator - ein physikalisches Symbolsystem - die hinreichenden und notwendigen Bedingungen für "Intelligenz" besitzt. Für NEWELL & SIMON ist das Konzept des Symbols vollständig innerhalb der Struktur des Symbolsystems definiert, auch wenn eine Verbindung zum designierten Objekt gefordert wird. Die Form der Symbole ist arbiträr, ihre Interpretation erfolgt gemäß sozialer Übereinkunft zwischen Beobachtern des Symbolsystems. Nach dieser Hypothese besteht intelligentes Verhalten aus folgenden Schritten: Erzeugung von Symbolen durch den sensorischen Apparat, die anschließende Manipulation dieser Symbole (etwa mit Inferenz-Techniken oder algorithmischer Suche) mit dem Ziel, ein Symbol oder eine Symbolstruktur als Ausgabe zu erzeugen.
Als Beispiel können wir uns ein geeignet programmiertes System vorstellen, das den Turing-Test für Bildverstehen besteht. Nach dem Anspruch der "starken KI" ***versteht*** ein solches System die Szene und stellt gleichzeitig auch die Erklärung dafür dar, wie Menschen diese Szene verstehen. SEARLE (1980) formulierte eines der inzwischen bekanntesten Argumente gegen die PSSH. Im Kontext des Bildverstehens (Abb. 1) läßt es sich verkürzt etwa so formulieren: auf die frühen, signalnahen Verarbeitungsschritte folgen solche der Symbolmanipulation (Objekterkennung, höhere Bilddeutung), für die sich die Argumentation SEARLES bezüglich des "chinesischen Zimmers" anwenden läßt. Obwohl ein Beobachter des BVS den Eindruck hat, daß es die Szene versteht, kann davon nicht die Rede sein: die Algorithmen, die ein

Programmierer in einer bestimmten Programmiersprache formuliert hat, besitzen von sich aus und für sich selbst keine Bedeutung[1].

Im Anschluß an SEARLE wurde von HARNAD (1990) ein konzeptuelles Modell entwickelt, mit dem Symbole in der Umwelt eines Systems verankert werden sollen. HARNAD kritisiert den Symbolverarbeitungsansatz für die Behauptung, daß bedeutungsvolle Programme durch regelgeleitete Symbolmanipulationen enstehen könnten. Als ***Symbolverankerungsproblem*** (symbol grounding problem) versteht HARNAD die Aufgabe, einem formalen Symbolsystem die Semantik mitzugeben, anstatt daß diese nur "parasitär in unseren Köpfen ist". Als Lösung schlägt er vor, Symbole über nichtsymbolische (ikonische und kategoriale) Zwischenrepräsentationen kausal mit den Objekten zu verbinden, auf die sie verweisen. Neuronale Netze sollen dazu dienen, die Zwischenrepräsentationen zu erzeugen. Ziel ist ein hybrides System, das die Verbindung von sensorischer Erfahrung und Symbol enthält. HARNAD betont, daß mit derart verankerten Symbolen die Kompositionalität des Systems leicht zu erreichen wäre. Ein bildverstehendes System (BVS) könnte danach eine komplexe Szene verstehen, indem Elementarobjekte in sensorischer Erfahrung verankert würden und die inhärente Bedeutung von komplexen Objekten, Objektkonstellationen u.ä. gemäß dem FREGEschen Prinzip sich daraus ergeben würde.

Später hat HARNAD eingeräumt, daß durch Symbolverankerung wohl nur erreicht werden kann, die Interpretationsmöglichkeiten für die Symbole einzuschränken. Es kann nicht sichergestellt werden, daß die Semantik der Symbole intrinsisch, also unabhängig von Interpretation ist (HARNAD 1993). Mit dem Symbolverankerungsansatz läßt sich eine Korrelationssemantik realisieren, die für eine technisch ausgerichtete KI von Vorteil sein kann.[2]

5. Schlußfolgerungen

In diesem Beitrag wurde der konzeptuelle Rahmen dargestellt, in dem Bildverstehen als hierarchischer, wissensbasierter Prozeß beschrieben wird. Wir betrachteten das Verstehensproblem des Computersehens, d.h. die Frage, ob die Leistung eines bildverstehenden Systems auf die Bearbeitung von Zeichen (Signale bzw. Symbole) beschränkt ist oder ob erreicht werden kann, daß die Zeichen eine intrinsische Bedeutung erhalten. Dazu untersuchten wir den Semantiktyp, der beim Bildverstehen Anwendung findet. Herkömmliche bildverstehende Systeme weisen eine internalistische Semantik auf, d.h. die Bedeutung eines Symbols wird in der konzeptuellen Rolle gesehen, die es in Bezug auf die anderen Symbole spielt. Auf diese

[1] Die Argumente SEARLEs werden durch Analysen des Semantikkonzepts der Informatik gestützt (vgl. etwa HESSE 1992, 285 ff; LENZ& MERETZ 1995, 70 ff).

[2] In den Worten HARNADs (1990, 340): "...the fact that our own symbols do have intrinsic meaning whereas the computer's do not, and the fact that we can do things that the computer so far cannot, may be indications that even in AI there are performance gains to be made (especially in robotics and machine vision) from endeavouring to ground symbol systems."

Weise kann das bildverstehende System keinen Zugang zur Welt erlangen; seine Semantik ist "geborgt", eine Interpretation durch Nutzer ist erforderlich. Die Semantik von (hybriden) Symbolverankerungs-Systemen ist korrelativer Natur, d.h. ein Symbol hat Bedeutung durch die Korrelation von Zeichen und Bezeichnetem. Das bedeutet, daß auch in diesen Systemen Interpretation weiter nötig ist, die Symbole aber nicht völlig arbiträr in ihrer Interpretation sind. In beiden Fällen kann nicht erreicht werden, daß Symbole eine intrinsische Bedeutung besitzen.

Wir sind der Auffassung, daß die skizzierten Schwierigkeiten beim Bildverstehen prinzipieller Natur sind. Die PSSH klammert semantische Aspekte bei der Beschreibung intelligenten Verhaltens aus, indem zwischen der syntaktischen und der semantischen Ebene, d.h. den regelgeleiteten Manipulationen von Zeichen und der jeweiligen semantischen Interpretation Unabhängigkeit postuliert wird. HAUGELAND (1981, 23) formuliert dies so: "... if you take care of the syntax, the semantics will take care of itself." Wie wir gesehen haben, können BVS bestenfalls den Anschein erwecken, als würden sie verstehen; in Wirklichkeit sind wir es (als Nutzer, Programmierer usw.), die diesen Systemen Bedeutung borgen. Es ist also unser Dasein, durch das die physikalischen Symbolstrukturen eines BVS usw. semantisch instanziiert werden können. Mit anderen Worten, die Redeweise vom Bildverstehen (wie auch vom Sprachverstehen) bei Maschinen ist ein Kategorienfehler: Maschinen haben keine Subjektivität, und deshalb ist es unsinnig zu erwarten, daß sie je eine irgendwie geartete Verstehenskapazität haben könnten. Die Einsicht greift Raum, daß diese Einschätzung nicht auf "intelligente" Systeme auf der Basis des Symbolverarbeitungsansatzes beschränkt ist, sondern auch für alternative (konnektionistische, enaktive u.a.) Ansätze Gültigkeit hat (z.B. LENZ & MERETZ 1995; ZIEMKE 1999). Ursächlich dafür ist m.E., daß es mit dem Informationsverarbeitungsparadigma - ungeachtet der Betonung der Agent-Umwelt-Interaktion - nicht gelingen kann, die historisch begründete, ganzheitliche Natur von Lebewesen und ihre lebensweltliche Einbindung zu erfassen: Kognitionen sind eben keine Komputationen (SEARLE 1990, HARNAD 1994)! Für das Computersehen (und die KI allgemein) kann das nur bedeuten, die Werkzeugperspektive (LENZ & MERETZ 1995, 82 ff) wiederzugewinnen und ihre Möglichkeiten gezielt und konstruktiv zu nutzen.

Literaturverzeichnis

MARR, D. (1982): Vision, San Francisco: W.H. Freeman.

NEWELL, A. & SIMON, H.A. (1976): Computer science as empirical enquiry: Symbols and search, in: Commun. ACM 19(3), 113-126.

NEUMANN, B. (1993): Bildverstehen - ein Überblick, in: GÖRZ, G. (Hg.): Einführung in die künstliche Intelligenz, Bonn [u.a.]: Addison-Wesley, S. 559-588.

CROWLEY, J.L. & CHRISTENSEN, H.I. (1995): Vision as Process, Berlin [u.a.]: Springer.

DUDA, R. & HART, P. (1973): Pattern Recognition and Scene Analysis, New York: Wiley.
HANSON, A. & RISEMAN, E. (1978): Computer Vision Systems, New York: Academic Press.
ALOIMONOS, Y. & WEISS, I. & BANDOPADHAY, A. (1988): Active vision, in: Int. J. Comp. Vision 7, 333-356.
SOMMER, G. (1995): Verhaltensbasierter Entwurf technischer visueller Systeme, in: KI 3, 42-45.
SCHIERWAGEN, A. (1998): Visuelle Wahrnehmung und Augenbewegungen: Neurale Mechanismen der Sakkadenkontrolle, in: SACHS-HOMBACH, K. & REHKÄMPER, K. (Hg.): Bild - Bildwahrnehmung - Bildverarbeitung, Wiesbaden: Deutscher Universitäts-Verlag 1998, S. 275-284.
TARR, M.J. & BÜLTHOFF, H.H.: Image-based object recognition in man, monkey and machine, in: Cognition 67 (1998) 1-20.
HILDEBRANDT, B., MORATZ, R., RICKHEIT, S. & SAGERER, G. (1995): Integration von Bild- und Sprachverstehen in einer kognitiven Architektur, in: Kognitionswissenschaft 4: 118-128.
HERZOG, G., BLOCHER, A., GAPP, K.-P., STOPP, E. & WAHLSTER, W. (1996): VITRA: Verbalisierung visueller Information, in: Informatik Forschung und Entwicklung 11: 12-19.
PAULI, J., BLÖMER, A., LIEDTKE, C.-E. & RADIG, B.(1995): Zielorientierte Integration und Adaptation von Bildanalyseprozessen, in: KI 3, 30-34.
SEARLE, J.R. (1980): Minds, brains and programs, in: Behav. Brain Sci. 3, 417-457.
HARNAD, S. (1990): The symbol grounding problem, in: Physica D, 42, 335-346.
HARNAD, S. (1993): Symbol grounding is an empirical problem, in: Proc 15th Annual Conference of the Cognitive Science Society, Boulder, CO, pp. 169-174.
HESSE, W. (1992): Können Maschinen denken - eine kritische Auseinandersetzung mit der harten These der KI, in: Kreowski, H.-J. (Hg.) : Informatik zwischen Wissenschaft und Gesellschaft (Informatik-Fachberichte Bd. 309), Berlin Heidelberg: Springer, S. 280 - 289.
LENZ, A. & MERETZ, S. (1995): Neuronale Netze und Subjektivität, Braunschweig/Wiesbaden: Vieweg.
HAUGELAND. J. (1981): Semantic engines: An introduction to mind design, in: HAUGELAND. J. (Hg.): Mind Design. Philosophy Psychology Artificial Intelligence, Cambridge, Mass. / London, England: MIT Press, pp. 1-34.
ZIEMKE, T. (1999): Rethinking grounding, in: Riegler, A., vom Stein, A. & Peschl, M. (eds.): Does Representation Need Reality? New York: Plenum Press, pp. 97-100.
SEARLE, J.R. (1990): Is the brain a digital computer? In: Proc. Adr. Amer. Philos. Assoc. 3, 21-37.
HARNAD, S. (1994): Computation is just interpretable symbol manipulation; cognition isn't, in: Mind and Machines 4, 379-390.