U. Ahlrichs, B. Heigl, D. Paulus, H. Niemann

Wissensbasierte aktive Bildanalyse

Inhaltsverzeichnis

Rev.: 1.15
LaTeX'ed 31. August 1998

Wissensbasierte aktive Bildanalyse

U. Ahlrichs, B. Heigl, D. Paulus, H. Niemann

Lehrstuhl für Mustererkennung (LME, Informatik 5)
Martensstr. 3, Universität Erlangen–Nürnberg, 91058 Erlangen
Tel.: +49 (9131) 85–7894 — Fax: +49 (9131) 303811
{ahlrichs,heigl,paulus}@informatik.uni-erlangen.de
http://www5.informatik.uni-erlangen.de

Zusammenfassung Ein zentrales Problem des Rechnersehens besteht darin, Objekte zu erkennen und zu lokalisieren. Die interessierenden Objekte sind dabei in den zu betrachtenden Szenen unter Umständen nicht oder nur teilweise sichtbar, was eine rechnergesteuerte, aktive Szenenexploration erforderlich macht. Dies bedingt eine Verbindung von signalnahen und wissensbasierten Ansätze zu einem rückgekoppelten Gesamtsystem aus Sensorik, wissensbasierter Analyse und Aktorik.

In diesem Beitrag liegen die Schwerpunkte auf der Vorstellung von Algorithmen zur Bestimmung von Tiefeninformation aus monokularen Bildfolgen sowie auf Untersuchungen zur wissensbasierten aktiven Szenenexploration.

In der exemplarischen Realisierung der Ansätze werden mit einer beweglichen, rechnergesteuerten Kamera in einer Büroumgebung relevante Objekte gesucht und gegebenenfalls die Kameraparameter verändert. Das Wissen über Kameraaktionen ist ebenso wie das Wissen über die Objekte explizit in Form eines semantischen Netzes repräsentiert.

Die wissensbasierte Bildanalyse stellt aufgrund der während der Exploration einzuhaltenden Zeitschranken ein geeignetes Anwendungsgebiet dar, um Parallelisierungs– und Verteilungsstrategien zu entwickeln und zu testen.

1 Einordnung und Motivation

Die Fähigkeiten biologischer visueller Systeme sind beeindruckend. Dem Menschen gelingt es mit wenigen Blicken, sich in einem Raum zu orientieren und Gegenstände zu identifizieren. Ist ein gesuchtes Objekt nicht sichtbar, so werden Augen– oder Kopfbewegungen durchgeführt, so daß es sich anschließend in der Mitte des Blickfeldes befindet. Die meisten Lebewesen haben zwei Augen und verwenden diese, um Information über die dreidimensionale Natur der Szene zu erlangen. Weiterhin dient die Farbe eines Objektes als wesentliches Erkennungsmerkmal. Die Exploration einer Szene folgt einer Strategie, die durch Erfahrung geleitet wird; beispielsweise wird man eine Tasse nicht an der Decke des Raums suchen. Gegenstände werden ebenfalls aufgrund von Vorwissen erkannt.

Diese zur Exploration einer Szene notwendigen Fähigkeiten gilt es mit Maschinen nachzubilden. Das menschliche Gehirn wird als ein hochparalleles System angesehen. Die äußerst komplexen Berechungen, die zur visuellen Exploration mit technischen Systemen notwendig sind, erfordern eine Aufteilung der Aufgaben in kleinere Einheiten, die in einigen Fällen zur Erhöhung der Effizienz gleichzeitig durchgeführt werden.

Dabei ist es nicht das Ziel der Mustererkennung, die menschlichen Fähigkeiten technisch exakt nachzubauen. Es soll stattdessen eine dem Menschen *vergleichbare* Leistung bei der Exploration erzielt werden. Der Aufbau biologischer Systeme dient also als Richtlinie, aber nicht als Vorschrift. Technisch werden mehrere Ansichten aus Farbkameras eingesetzt, um aufgrund von Farbinformation einen ersten Hinweis auf gesuchte zu Objekte bekommen und um aus den verschiedenen Ansichten der Objekte dreidimensionale Information zu ermitteln. In technischen Systemen wird die Erfahrung in einer Wissensbasis kodiert.

Einsatzmöglichkeiten solcher technischer Systeme bieten sich beispielsweise in autonomen Fahrzeugen und Robotern an, die mit visuellen Sensoren ausgestattet sind.

Das Ziel des hier vorgestellten Projekts ist die Demonstration eines Gesamtsystems zur Szenenexploration, das sich aus Modulen für Teilaufgaben auf signalnaher und wissensbasierter Ebene zusammensetzt. In Abs. 2 werden die Untersuchungen im Bereich der signalnahen Bildverarbeitung erläutert, welche sich hauptsächlich auf die Gewinnung und Verwendung von Tiefendaten konzentrieren. Abs. 3 beschreibt den Aufbau der Wissensbasis und der Kontrolle, welche unter Verwendung der Ergebnisse der signalnahen Bildverarbeitung der aktiven Szenenexploration dienen. Die Integration der Module aus Abs. 2 und Abs. 3 wird in Abs. 4 für ein Anwendungsbeispiel dargestellt: Die Aufgabenstellung ist es, Objekte in einem Büroraum zu finden, selbst wenn sie zunächst nicht sichtbar sind. Eine rechnergesteuerte Kamera verändert ihren Blickwinkel und nimmt durch Brennweitenveränderung Nahaufnahmen von Objekten auf. Farbinformation dient zum einen zur groben Lokalisierung von Objekten und zum anderen zum Vergleich der Objekthypothesen mit den in der Wissensbasis enthaltenen Objekten. Eine Zusammenfassung in Abs. 5 beschließt den Beitrag.

2 Signalnahe Bildverarbeitung

Im hier beschriebenen Projekt konzentrieren sich die Bestrebungen zur signalnahen Bildverarbeitung hauptsächlich auf die Ermittlung und Verwendung von Tiefendaten eines Objektes oder einer Szene (Abs. 2.1), wobei die Tiefendaten auch statistisch modelliert werden (Abs. 2.2). Eine notwendige Voraussetzung für die darauf aufbauende Szenenexploration ist das Wissen über die Abbildungsgeometrie der Optik und Kamera sowie die Steuerung der Kameraparameter, was durch eine entsprechende Kalibrierung erreicht wird (Abs. 2.3). Da die verwendeten Algorithmen teilweise sehr zeitaufwendig sind, spielt die Parallelisierung eine entscheidende Rolle, um die entwickelten Methoden für Echtzeitsysteme einsetzen zu können (Abs. 2.4).

2.1 Tiefenermittlung und Modellierung

Zur Tiefenermittlung aus Bildfolgen werden in der Literatur mehrere Verfahren vorgestellt. Eine Übersicht hierzu ist in [Bes89] zu finden. Im folgenden werden hierzu drei Algorithmen herausgegriffen, die jeweils weiterentwickelt und angepaßt wurden. Die Auswahl des jeweiligen Verfahrens findet in Abhängigkeit davon statt, wie dicht Tiefendaten vorhanden sein müssen, wieviel Zeit zur Verarbeitung zur Verfügung steht, und ob die Aufnahmegeometrie zu jedem Einzelbild bekannt ist. Das erste Verfahren liefert

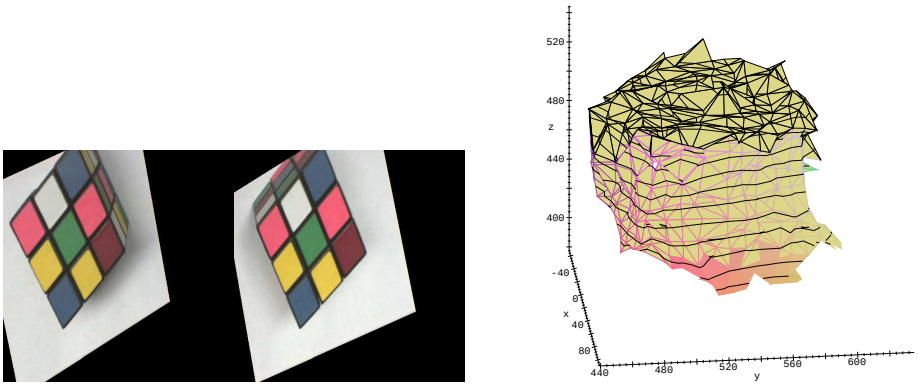


Bild 1. Beispiel für ein normiertes Stereobild und vollständiges Objektmodell eines Würfels aus einem kalibrierten Bildstrom von 40 Einzelbildern.

eine Beschreibung der Oberfläche des Objektes aus einem kalibrierten Bildstrom. Ohne Kalibrierung kommt das deutlich schnellere zweite Verfahren aus, in dem nur einzelne Objektpunkte rekonstruiert werden. Das dritte Verfahren erfordert eine Kalibrierung und eignet sich ebenfalls zur schnellen Vermessung von Szenen, wobei auch hier die Tiefeninformation nur an segmentierten Punkten der Szene verfügbar ist.

Objektmodell aus kalibrierten Bildströmen Die räumliche Struktur eines Objektes liefert neben der Farbinformation wichtige Anhaltspunkte zur Identifizierung und Lagebestimmung. Deshalb wurden Untersuchungen angestellt, um in einer Laborumgebung Objekte aus Bildfolgen zu rekonstruieren. Mittels eines Roboters wird eine Kamera an definierte Positionen gefahren, um verschiedene Ansichten des zu rekonstruierenden Objektes zu betrachten. Die Bilder der Folge werden aufgenommen, wobei die Bahn der Kamera auf einer Halbkugel über dem Objekt liegt und zwischen zwei Aufnahmen ca. 60° Unterschied in der Betrachtungsrichtung gewählt werden. Um aus der daraus resultierenden monokularen Farbbildfolge ein vollständiges Objektmodell zu generieren, wurde ein neues Verfahren entwickelt [Beß97]. Für je zwei aufeinanderfolgende Farbbilder wird ein normiertes Stereobildpaar durch Neuabtastung in einem virtuellen Stereobild mit parallelen optischen Achsen berechnet. In diesem Bild werden mit den Verfahren aus [Har96] Linien und Kreisbögen segmentiert, durch deren Zuordnung erste Anhaltspunkte für die Tiefenverläufe gegeben sind. Um dichte Tiefenkarten zu erhalten, werden diese Ergebnisse als Initialisierung eines Blockvergleichsverfahrens verwendet, für das der Suchbereich eingeschränkt wird. Dabei wird ein mittlerer Zuordnungsfehler eines Pixels im Stereobild von 1.93 Pixeln erreicht. Die einzelnen Tiefenkarten werden kombiniert, um die fehlende Information einzelner Ansichten zu ergänzen und ein Gesamtmodell des beobachteten Objektes zu generieren (Bild 1). Die vollständige Berechnung eines Gesamtmodells benötigt etwa 15 Minuten.

Tiefengewinnung aus unkalibrierten Bildströmen Ein Problem des oben beschriebenen Ansatzes besteht darin, daß die Kamerapositionen jeder Einzelaufnahme genau be-

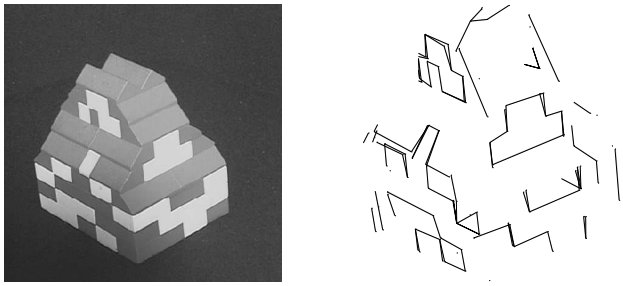


Bild 2. Objektmodell eines Hauses aus einem unkalibrierten Bildstrom von 400 Einzelbildern; die Linien wurden zwischen den segmentierten Punkten zur Visualisierung hinzugefügt.

kannt sein müssen, was zum einen das Vorhandensein eines Roboters mit genügend hoher Positioniergenauigkeit und zum anderen eine exakte Greifer–Kamera–Kalibrierung voraussetzt.

Deshalb wurden die Untersuchungen dahingehend erweitert, Tiefeninformation aus solchen Bildströmen zu errechnen, in denen die Änderungen der Beobachtungsrichtung unbekannt sind. Hierzu wurde das Verfahren [Tom92] angewandt und so modifiziert, daß bei gleichbleibender Güte der Rechenzeit– und Speicherbedarf drastisch reduziert wird [Beß96]. Dabei beträgt der Fehler in der Rekonstruktion relativ zur Objektgröße im Schnitt 13%, maximal 18%. Als Eingabe erwartet das Verfahren die Trajektorien von Punktmerkmalen und liefert die Kameraposition zu jedem Bild und die 3D–Positionen der Punktmerkmale (Bild 2). Die Zeiten für die Tiefenberechnung aus den Trajektorien beträgt 2.3 Sekunden. Hinzu kommen die Zeiten für die Berechnung der Trajektorien, welche vom gewählten Verfahren abhängen. Unter Verwendung der durch dieses Verfahren ermittelten Kamerapositionen kann obiges Verfahren anschließend direkt eingesetzt werden, um dichte Tiefenkarten zu berechnen.

Tiefengewinnung aus Linearbewegung Ein weiterer Ansatz beschäftigt sich mit der Tiefengewinnung mittels einer Schwenk/Neige/Zoom–Kamera, die auf einem Linearschlitten montiert ist. Bei diesem Versuchsaufbau schlagen sich die Fehler der Kalibrierung nur gering nieder, da zur Tiefenberechnung die Schwenk/Neige–Stellung der Kamera beibehalten wird, und die Konstellation so nur einen Freiheitsgrad zuläßt, da die Brennweite konstant gehalten wird. Im Gegensatz dazu akkumulieren sich die Fehler der einzelnen Achsen eines Roboters im ersten Verfahren und erschweren so eine genaue Positionierung.

In der Linearschlitten–Konstellation befinden sich korrespondierende Punkte auf Geraden, die durch die Epipolargeometrie bestimmt sind. Durch Verfolgung farbiger Punkte während einer Linearbewegung der Kamera bei minimaler Zoomeinstellung werden Tiefenwerte ermittelt, die unmittelbar nach der Verfolgung zur Verfügung stehen [Hei97, Hei98]. Die Farbpunktverfolgung ist eine Weiterentwicklung des grauwertbasierten Verfahrens [Tom91]. Zusätzlich wird Wissen über die epipolargeometrischen Zusammenhänge verwendet, um die Verschiebungsrichtung der einzelnen Bildpunkte zu bestimmen und so den Suchraum einzuschränken. Außerdem kann hierdurch die

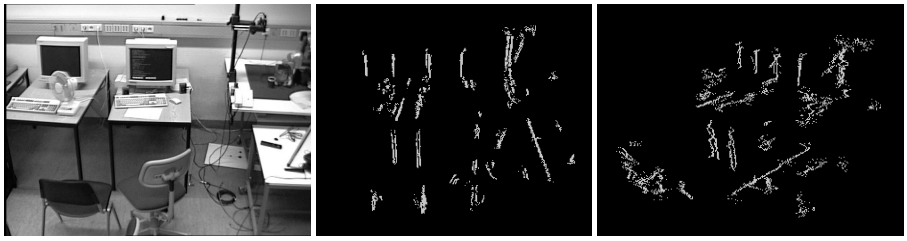


Bild 3. Teilbild einer Bildsequenz aus 100 Bildern einer sich linear bewegenden Kamera und daraus berechnete 3D–Punktemenge aus zwei unterschiedlichen Ansichten

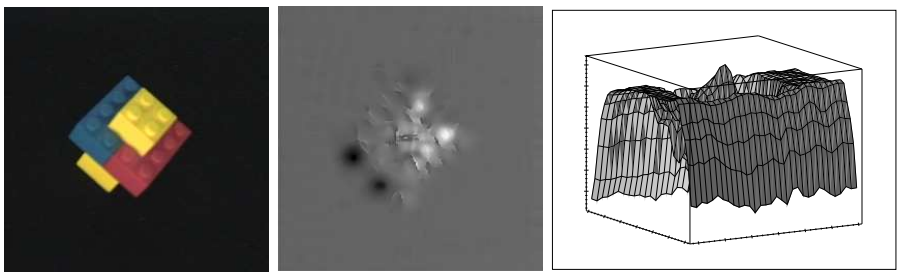
Anzahl der verfolgten Bildpunkte erheblich vergrößert werden (in einem Beispiel auf das zehnfache). Die Dauer für die Vermessung eines Raumes beträgt einschließlich der Kamerabewegungen und Tiefenberechnung etwa eine Minute. Im Vergleich zum ersten Verfahren wird auf die Generierung von normierten Stereopaaren verzichtet, was zum einen Rechenzeit spart und zum anderen Fehler vermeidet, die durch die notwendige Neuabtastung hinzukommen. Bild 3 zeigt ein Beispiel für die ermittelten 3D–Punkte einer Büroszene aus zwei unterschiedlichen Ansichten. Zur Bewertung der Zuverlässigkeit der Verfolgung wurden in dieser Szene 1000 Punktmerkmale über 20 Bilder mit einem Fenster von 7×7 Pixeln verfolgt. Dabei wurde die Grauwertbildfolge aus der Farbbildfolge durch Berechnung des Helligkeitskanals ermittelt. Ein Punktmerkmal wurde als verloren bezeichnet, wenn der mittlere Grauwertunterschied für ein Fenster im Vergleich zur vorherigen Position geringer als ein Schwellwert ist. Dabei wurden beim Farbverfahren im Vergleich zum Grauwertverfahren 8% weniger Punkte verloren (156 gegenüber 170 von 1000).

Durch lineare Regression können die Parameter der linearen Funktion der Bildkoordinate in Abhängigkeit von der Linearschlittenposition für jedes Punktmerkmal ermittelt und so die Lage des entsprechenden Punktes im Raum errechnet werden. Hierbei fällt zusätzlich zum Schätzwert ein Fehlerwert ab, welcher die Unzuverlässigkeit der Verfolgung beschreibt und eine Eliminierung unsicher verfolgter Merkmale erlaubt.

Dieser Ansatz eignet sich dazu, Szenen zu vermessen, um ungefähre Anhaltspunkte für die räumliche Struktur zu erhalten, und beispielsweise die Bildgröße bekannter Objekte zu ermitteln, um die Detektion zu erleichtern. Hierfür ist ein Beispiel in Abs. 4 zu finden.

2.2 Statistische Modellierung von Tiefendaten

In der Mustererkennung spielen statistische Methoden eine immer wichtigere Rolle. Erste beeindruckende Ergebnisse wurden in der Sprachverarbeitung Mitte der 80iger Jahre durch Verwendung von Hidden–Markov–Modellen erzielt. Derzeit beruhen nahezu alle Sprachverarbeitungssysteme auf statistischer Modellierung. Auch in der Bildverarbeitung kommen zunehmend statistische Verfahren zum Einsatz, gerade in Bereichen, die sich nur schwer durch physikalisch motivierte Vorstellungen modellieren lassen. Im Umfeld des hier beschriebenen Projektes wurden Ansätze entwickelt, die zum einen geometrische Objektstrukturen statistisch beschreiben [Hor96a] und zum anderen Objekte



(a) (b) (c)

Bild 4. (a) eines der verwendeten Objekte; (b) zugehörige gwichtete Tiefenkarte; (c) Beispiel für logarithmierte Dichte für zwei Translationsparamter

erscheinungsbasiert modellieren. Das heißt, es wird direkt das Verhalten der Bildinformation in Abhängigkeit von Positionsparametern erlernt [Pös97].

Zur Verwendung von Tiefendaten im Zusammenhang mit statistischen Objektmodellen können direkt Methoden der erscheinungsbasierten statistischen Objektmodellierung verwendet werden [Pös98]. Das Verfahren modelliert das statistische Verhalten von lokalen Bildmerkmalen in Abhängigkeit von der Objektlage. Unter Verwendung mehrerer Auflösungsstufen kann eine 3D–Lokalisierung durchgeführt werden. Als Merkmale werden hier Waveletkoeffizienten verwendet, die sowohl für die Bilder der einzelnen Farbkanäle als auch für eine gewichtete Tiefenkarte berechnet werden. Diese Tiefenkarte wird aus mehreren Einzelkarten kombiniert, wobei jeder Eintrag mit seinem Zuverlässigkeitswert gewichtet wird. Nach Konkatenation der so ermittelten Einzelmerkmale wird ein lokaler Gesamtvektor gebildet, der durch eine Normalverteilung modelliert wird. Ihre Parameter werden in Abhängigkeit von den Positionsparametern unter Verwendung von Basisfunktionen beschrieben. Zum Training werden große Stichproben verwendet, für welche die 3D–Objektlagen bekannt sind (867 Bilder pro Objekt, aufgenommen aus 289 unterschiedlichen Aufnahmepositionen bei drei unterschiedlichen Beleuchtungen). Die Auswertung einer hierzu disjunkten Teststichprobe (289 Aufnahmepositionen bei einer vierten Beleuchtung) zeigt, daß sich durch die Verwendung der Farbinformation im Vergleich zur reinen Grauwertinformation die Erkennungsraten in einem Beispiel von 77% auf 98% und in einem zweiten Beispiel von 45% auf 63% verbessern. Durch Hinzunahme der Tiefeninformation ergeben sich zusätzliche Verbesserungen von 98% auf 100% im ersten Beispiel und von 63% auf 78% im zweiten. Bei diesen Erkennungsraten wird die Lage eines Objektes als richtig erkannt bezeichnet, wenn der Lokalisationsfehler kleiner oder gleich 10° in der Rotation und kleiner als 10 Pixel in der Verschiebung war. Bild 4 zeigt das Objekt des ersten Beispiels zusammen mit der gewichteten Tiefenkarte, die neben der Farbinformation modelliert wird. Außerdem ist ein Beispiel für eine Dichte gezeigt, deren Maximum das Optimum zweier Translationsparamter zeigt.

Diese Untersuchungen zeigen, daß die Tiefeninformation als Zusatz zur reinen Bildinformation die Erkennungsraten signifikant verbessert, was als Indiz dafür gewertet werden kann, daß die räumliche Struktur zur Beschreibung von Objekten eine Notwendigkeit darstellt.

2.3 Kamerakalibrierung und Kamerasteuerung

Greifer–Kamera–Kalibrierung Für Verfahren zur Tiefenermittlung, die kalibrierte Bildströme als Eingabe erwarten (Abs. 2.1), ebenso wie für die Erzeugung von Stichproben zum Training und Test von statistischen Objektmodellen (Abs. 2.2) muß eine Kamera mittels eines Roboters in anzugebende Positionen und Orientierungen gebracht werden. Um dies erreichen zu können, muß zum einen die Robotergeometrie und deren Zusammenhang zu den Steuersignalen bekannt sein (was das Datenblatt des Roboters liefert) und zum anderen die Position der Kamera relativ zum Greifer des Roboters. Letzteres kann nur durch optische Kalibrierung erreicht werden. Hierzu wurde das Verfahren von Tsai und Lenz [Tsa88] angewandt, das die Verwendung einer Vielzahl von Aufnahmen erlaubt und so den Positionierungsfehler der Kamera noch unter die Kalibriergenauigkeit der Kamerakalibrierung senkt. Anstelle des bei vergleichbaren Ansätzen notwendigen Einsatzes von aufwendigen, nichtlinearen Optimierungsverfahren sind dabei lediglich lineare Gleichungssysteme zu lösen. Hierdurch besitzt das Verfahren erhebliche Geschwindigkeitsvorteile gegenüber anderen Methoden, die einen Einsatz bei der Kalibrierung aus Bildfolgen motiviert.

Zoomkalibrierung Um Nahaufnahmen mit definierter Größe eines Objektes im Bild zu erstellen, muß der Zusammenhang zwischen Brennweite und Schrittmotorstellung des Zoomobjektivs bekannt sein. Für die Kalibrierung dieser Funktion wurde ein robustes Verfahren entwickelt, das die Farbpunktverfolgung aus Abs. 2.1 benutzt. Zur Ermittlung der Brennweite in Abhängigkeit von der Zoomeinstellung existiert ein Verfahren von Willson [Wil94], was jedoch zu praktischen Problemen führt, da für alle Zoom– und Fokuseinstellungen ein Kalibriermuster so im Bild plaziert werden muß, daß gleichzeitig eine genügend genaue Segmentierung möglich ist. Um für unseren Anwendungsfall eine Lösung zu erhalten, wurde hier ein anderer Weg eingeschlagen. Anstelle dieses Verfahrens wird hier der Vergrößerungsfaktor zwischen zwei sich nur gering unterscheidenden Zoomeinstellungen für beliebige Linien eines Bildes experimentell bestimmt, welche sich durch Verbindung von je zwei gut verfolgbaren Punkten ergeben. Dabei ist eine beliebige Szene wählbar, die nur genügend strukturiert ist. Um die Vergrößerung von der minimalen Einstellung in eine andere Zoomeinstellung zu ermitteln, müßten alle Zwischenfaktoren aufmultipliziert werden. Dieser Ansatz zeigte jedoch zu große Anfälligkeit für lokale Fehler der Faktoren. Deshalb werden hier die lokalen Vergrößerungsfaktoren logarithmiert und anschließend durch eine lineare Funktion approximiert. Letztlich wird über diese Funktion integriert, was zu einer Reduktion und Elimination lokaler Fehler führt, die sich durch Ausmultiplizieren über mehrere Schritte akkumulieren würden. Nach Anwendung der Exponentialfunktion erhält man daraus den Vergrößerungsfaktor gegenüber der minimalen Zoomeinstellung. Die jeweilige Brennweite verhält sich zur Vergrößerung proportional mit der minimialen Brennweite als Proportionalitätsfaktor [Pau98].

Ansichtenauswahl Die Verwendung einer Schwenk/Neige–Kamera mit Zoomobjektiv ermöglicht es, die Kameraeinstellungen so zu wählen, daß Details eines bei einer kleinen Brennweite aufgenommenen Bildes vergrößert und im Bild zentriert dargestellt

werden. Ist die Lage eines Objektes als Position in einem Bild bei minimaler Brennweite bekannt, so wird der horizontale und vertikale Winkel bestimmt, welcher der dieser Position entsprechende Sichtstrahl mit der optischen Achse der Kamera einschließt. Daraufhin wird die Ausrichtung der Kamera entsprechend angepaßt, so daß nach dieser Bewegung die optische Achse mit diesem Sichtstrahl übereinstimmt. Durch das Wissen über den Zusammenhang von Brennweite zu Zoomeinstellung kann nachfolgend letztere so gewählt werden, daß das Objekt in einer vorgegebenen Größe dargestellt ist.

Weitere Untersuchungen wurden dahingehend angestellt, daß für einen Roboter unter Vorgabe von Anfangsposition und Endposition eine Bahnplanung durchgeführt wird, die eine optimale, kollisionsfreie Bewegung des Roboters unter Berücksichtigung von statischen Hindernissen erlaubt [Wag94].

2.4 Parallelisierung lokaler Bildverarbeitungsalgorithmen

Untersuchungen zur Parallelisierung von aufwendigen Bildverarbeitungsalgorithmen wurden auf dem Multiprozessorsystem MEMSY durchgeführt. Hierbei wird speziell für das Blockvergleichsverfahren im ersten Ansatz aus Abs. 2.1 das normierte Stereobild in horizontale Streifen partitioniert und jedem Rechenknoten je ein solches Teilbild zur Berechnung zugewiesen. Die Experimente zeigen, daß hierbei ein nahezu linearer Speedup erzielt werden kann. Ähnliche Resultate liefert eine vergleichbare Implementierung auf einem Workstationcluster durch Kommunikation über PVM[1].

Motiviert durch diese Ergebnisse lag es nahe, ein System zu schaffen, das derartige Partitionierungs– und Verteilungsaufgaben übernimmt, und es einem Benutzer ermöglicht, lediglich durch Angabe eines Algorithmus und der zu verwendenden Rechenknoten einfache Bildverarbeitungsalgorithmen verteilt auszuführen. Unser System [Sch98] wurde objektorientiert konzipiert und ermöglicht hierdurch den dynamischen Austausch von Verteilungs– und Partitionierungsstrategien; eine einheitliche Schnittstelle für Bildformate und Algorithmen wird festgelegt. Außerdem wurde das Kommunikationsmodul so gestaltet, daß völlig unterschiedliche Kommunikationswerkzeuge verwendet werden können. In unserem Fall wählten wir die Kommunikation über PVM und TCP/IP–Sockets. Neben der oben erwähnten starren Partitionierung des Eingabebildes wurden dynamische Verfahren entwickelt, welche eine der Auslastung der Rechenknoten angepaßte Aufteilung ermöglichen, und so die Gesamtrechendauer reduzieren, falls für die einzelnen Rechenknoten keine vorhersehbare oder konstante Rechenleistung zur Verfügung steht.

3 Modellbasierte Bildanalyse

3.1 Problemstellung

Die Analyse und Interpretation von einzelnen Bildern ist vorrangiges Ziel der klassischen wissensbasierten Bildverarbeitung. Hierzu sind eine Reihe von Systemen wie VISIONS [Han78], SPAM [McK85] oder SIGMA [Mat90] aus der Literatur bekannt,

[1] **P**arallele **V**irtuelle **M**aschine siehe `http://www.epm.ornl.gov/pvm`

deren Aufbau sich am *Marr–Paradigma* [Mar82] orientiert: die wissensbasierte Bildanalyse wird verstanden als ein Problem, eine initiale symbolische Beschreibung eines Bildes mit den in einer Wissensbasis repräsentierten Objekten in Beziehung zu setzen. Durch Segmentierungsfehler und Mehrdeutigkeiten in der Interpretation ergibt sich so ein hochkomplexes Suchproblem.

Mit der Einführung der Verarbeitungsstrategie des *aktiven Sehens* [Alo88, Baj92] steht nicht mehr die optimale Lösung dieses Suchproblems im Vordergrund. Vielmehr gewinnt das Schaffen optimaler Voraussetzungen für die Weiterverarbeitung der Bilddaten zum Beipiel in Form einer optimalen Kameraeinstellung an Bedeutung.

Für die Integration der Ideen des aktiven Sehens in eine Wissensbasis oder für die wissensbasierte Analyse gab es bisher nur erste Ansätze und Einzellösungen [Lev89, Bük96, Rim93]. Im folgenden wird zum einen das Wissen über die Umgebung und die Objekte, die in die zu erfüllende Aufgabe involviert sind, explizit repräsentiert; zum anderen werden die Verarbeitungsschritte, die zur Lösung einer Aufgabe notwendig sind, ebenfalls in die Wissensbasis mit aufgenommen. Als solche Verarbeitungsschritte können sowohl Aufgaben mit hoher Komplexität wie das Umfahren von Hindernissen angesehen werden, als auch einfache Aufgaben, wie Segmentierungsoperationen, Kameraaktionen und Instantiierung von Konzepten, die den Objekten zugeordnet sind. Durch die Kameraaktionen wird die Einstellung der Kamera, zum Beispiel der Fokus, verändert, um optimale Voraussetzungen für die Weiterverarbeitung der Daten zu schaffen. Mit Hilfe der Segmentierungsoperationen erfolgt zum Beispiel die Einteilung eines Bildes in Regionen.

Die Repräsentation dieses Wissens in einem *einheitlichen* Formalismus erlaubt die Verwendung *einer* Kontrolle zur Nutzung des Wissens während der Erfüllung einer Aufgabe. Die Aufgabe der Kontrolle besteht dabei darin, zu jedem Zeitpunkt den nächsten, im Sinne eines Gütekriteriums optimalen Verarbeitungsschritt auszuwählen, um das Ziel — die Erfüllung der Aufgabe auf der Basis der besten Interpretation der Szene — mit geringst möglichem Aufwand und höchst möglicher Sicherheit zu erreichen.

3.2 Grundlagen

Die Wissensrepräsentationssprache ERNEST (**ER**langer semantisches **NE**tzwerk **S**ys**T**em, [Nie90]), erweist sich als adäquate Basis im Bereich der Wissensrepräsentation und Wissensnutzung für die Bild– und Sprachverarbeitung. Der Formalismus für semantische Netze definiert Konzepte zur Repräsentation von Objekten und Fakten sowie verschiedene Kanten: die Bestandteilskante, die Spezialisierungskante und die Konkretisierungskante. Während die ersten beiden Kantentypen im allgemeinen in semantischen Netzen zu finden sind, ist die Konkretisierungskante neu hinzugekommen. Mit dieser Kante können verschiedene Abstraktionsebenen, die im Netz repräsentiert sind, verbunden werden. Eine weitere Besonderheit des Formalismus zur Wissensrepräsentation stellen die Modalitäten dar. Diese liefern einen großen Beitrag zur Strukturierung von Konzepten, indem sie zulässige Kombinationen von Bestandteils– und Konkretisierungskanten festlegen, die unterschiedliche Ausprägungen eines Konzepts repräsentieren. Der Formalismus hat sich in mehreren Anwendungen bewährt [Sal95, Mas93, Sch90]. Zur Nutzung des repräsentierten Wissens stellt ERNEST eine A*–Kontrolle zur Verfügung [Kum90]. Weiterhin wurde eine iterativ–optimierende Kontrolle [Fis95a] entwickelt, die

die parallele Nutzung des Wissens erlaubt. Die Aufgabe dieser Kontrollen besteht darin, diejenige Zuordnung zwischen Objekthypothesen und den entsprechenden Konzepten auszuwählen, die eine maximale Bewertung des Zielkonzepts liefert. Zielkonzepte sind Konzepte, die hinreichende Aussagen über das Anwendungsgebiet beinhalten.

Alle im folgenden beschriebenen Ansätze aus der Bildverarbeitung gliedern sich in das in [Pau97a] beschriebene objektorientierte System zur Mustererkennung ein. Dies gilt ebenso für die im Rahmen des SFB 182 realisierte objektorientierte Wissensbasis, deren Formalismus auf der ERNEST–Philosophie basiert, sowie für die Klassenhierarchie von Kontrollalgorithmen, die die verwendete Kontrolle enthält.

3.3 Die Wissensbasis

Der Schwerpunkt der hier vorgestellten Arbeiten besteht in der Entwicklung einer einheitlichen Repräsentation und Nutzung von Wissen über Objekte und Kameraaktionen. Dazu werden die zu bewältigenden Aufgaben und die dabei durchzuführenden Kameraaktionen *zusammen* mit dem Wissen über die Objekte in *einem* semantischen Netz repräsentiert [Ahl98, Pau98]. Die zum Aufbau des semantischen Netzes verwendeten Strukturen basieren auf der in Abs. 3.2 erwähnten ERNEST–Philosophie. In Bild 5

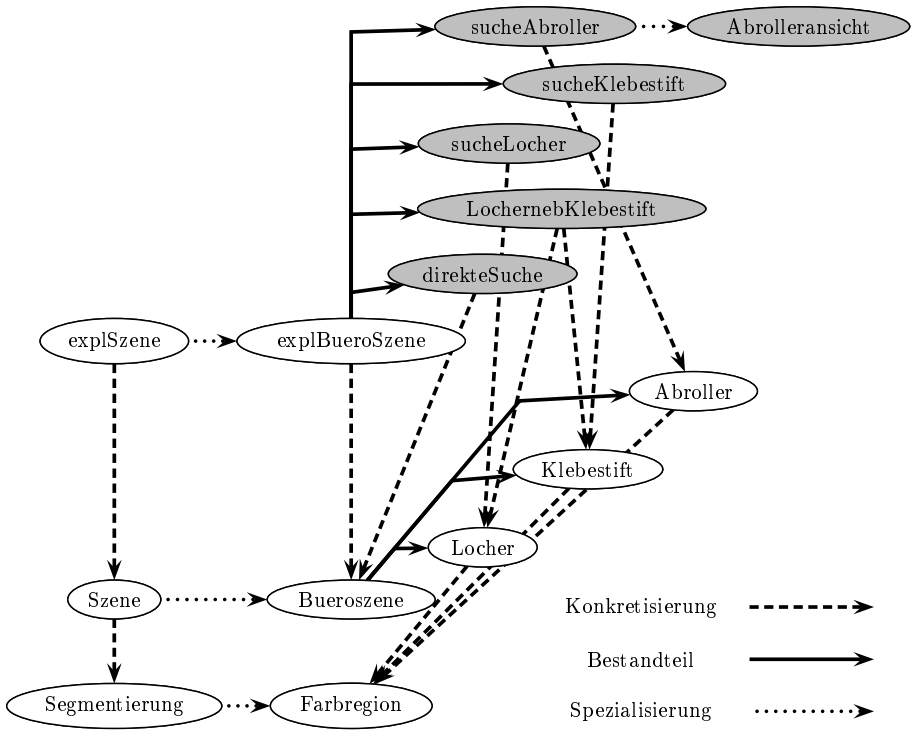


Bild 5. Aufbau der Wissensbasis für Szene und Kameraaktionen

ist ein Auszug aus der konzipierten Wissensbasis dargestellt. Der untere Teil des Bildes enthält die Szenenwissensbasis, die drei verschiedene Objekte, einen Locher, einen Abroller und einen Klebestift, umfaßt. Diese Objekte werden durch farbige Regionen konkretisiert. Für die Bewertung der Zuordnung zwischen den von einer Farbsegmentierung gelieferten Regionen [Den95] und dem Konzept "Farbregion" wird zum einen der mittlere Farbwert herangezogen. Zum anderen werden zum Beispiel die Höhe und Breite der Objekte als Attribute verwendet. Für deren Berechnung fließen die von der signalnahen Bildverarbeitung bestimmten Tiefenwerte in die Attributberechnung mit ein.

Analog zu den Objekten werden auch die Kameraaktionen ebenfalls durch Konzepte repräsentiert (vergleiche die Konzepte "sucheLocher" oder "Abrolleransicht"). Je nach Abhängigkeit von den Konzepten, die der Szenenrepräsentation zugeordnet sind, werden die Kameraaktionen auf verschiedenen Abstraktionsebenen des semantischen Netzes angesiedelt. Das Konzept "direkteSuche" ist hierfür ein Beispiel. Die Verbindung zwischen Konzepten zur Repräsentation von Kameraaktionen und Objekten wird über Konkretisierungskanten hergestellt. Konkurrierende Kameraaktionen wie "direkteSuche", "LochernebKlebestift" oder "sucheAbroller" werden durch unterschiedliche Modalitätsmengen repräsentiert, die in Bild 5 durch die grau unterlegten Konzepte angedeutet sind. Dabei wird durch die Instantiierung des Konzepts "direkteSuche" eine Schwenkbewegung der Kamera ausgelöst, mit der die Szene sukzessiv abgefahren wird, das heißt disjunkte Überblicksbilder mit kleiner Brennweite erzeugt werden. Mit der Instantiierung des Konzepts "LochernebKlebestift" ist hingegen nur eine kleine Schwenkbewegung verbunden, bei der eine kleine Region neben dem Locher in das Blickfeld der Kamera geholt wird.

3.4 Die Kontrolle zur Wissensnutzung

A*–Kontrolle Zur Nutzung des in dem semantischen Netz repräsentierten Wissens wird ein Kontrollalgorithmus benötigt, der aus konkurrierenden Segmentierungsergebnissen, zum Beispiel Regionen, die zu dem repräsentierten Wissen optimal passenden heraussucht. Außerdem müssen Mehrdeutigkeiten in der Wissensbasis, die durch Modalitätsmengen zum Beispiel zur Repräsentation von konkurrierenden Kameraaktionen entstehen, aufgelöst werden. Hierzu wird die in Abs. 3.2 erwähnte, auf dem A*–Algorithmus basierende Kontrolle verwendet. Die Steuerung der Analyse erfolgt durch die modellgetriebene Berechnung eines Pfades, der die Reihenfolge für eine datengetriebene Instantiierung festlegt. Nach der Bindung von Hypothesen (hier: für die Regionen) an die primitiven Knoten des semantischen Netzes (vergleiche Bild 6) werden datengetrieben Instanzen zu den Konzepten berechnet. Für die Ausführung von Kameraaktionen sind Erweiterungen der A*–Kontrolle notwendig, die sich nicht auf die Instantiierung der Konzepte zur Objekt– und Kameraaktionsrepräsentation beziehen, sondern vielmehr für die Verarbeitung der Segmentierungshypothesen notwendig sind, die sich nach einer Kamerabewegung ergeben. So erhält man zum Beispiel durch die Instantiierung des Konzepts "direkteSuche" (Bild 5) ein Überblicksbild der Szene, das andere Information beinhaltet als das zur Instantiierung des Konzepts "Bueroszene" verwendete Überblicksbild. Durch die Analyse dieses neuen Übersichtsbildes erhält

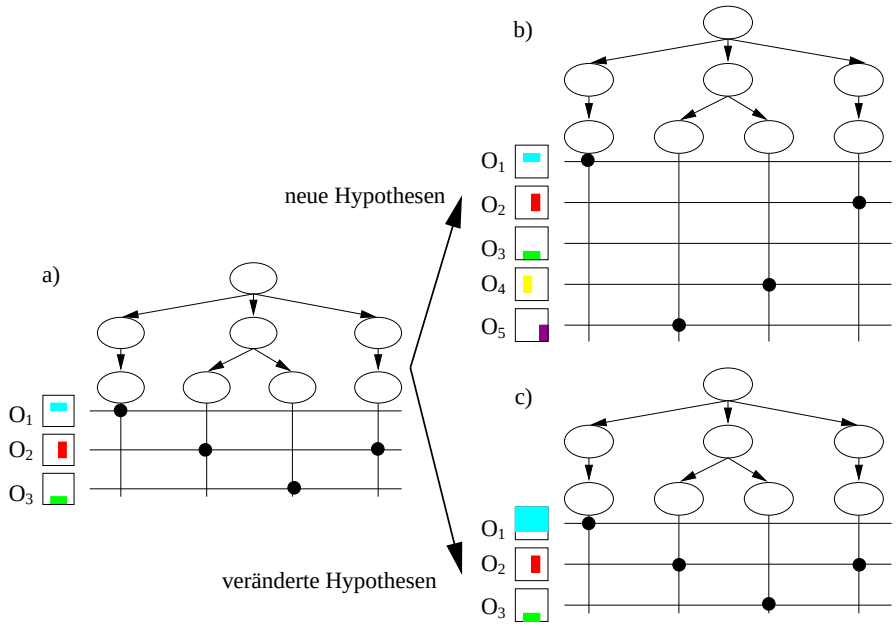


Bild 6. Auswirkungen der Ausführung von Kameraaktionen auf die Kontrolle. Die Kästchen links symbolisieren Ansichten der Objekte $O_1 \ldots O_5$. Die schwarzen Punkte im Gitter markieren die Zuordnung einer Objekthypothese zu dem entsprechenden Konzept; in b) ist die Situation nach der Instantiierung des Konzepts “direkteSuche” dargestellt, in c) die Aktualisierung einer Hypothese nach dem Zoomen auf ein Objekt

man neue Hypothesen, die bei der Analyse mit dem alten Übersichtsbild noch nicht auftraten und in der A*–Suche berücksichtigt werden müssen (vergleiche Bild 6 b). Wird hingegen eine Objekthypothese aufgrund anderer Kameraparameter, zum Beispiel durch eine Fovealisierung, wiederholt zur Analyse herangezogen, erhält man keine weiteren konkurrierenden Hypothesen (Bild 6 c). Stattdessen soll die alte, schon an das entsprechende Objekt gebundene Hypothese durch die nach der Kamerabewegung erzeugte neue Hypothese ersetzt werden.

Parallele iterativ–optimierende Kontrolle In der A*–Kontrolle wird das in Abs. 3.1 dargestellte Suchproblem durch eine Pfadsuche basierend auf der maximalen Bewertung des Zielkonzepts gelöst. Bei der iterativ–optimierenden Kontrolle wird dieses Suchproblem auf ein kombinatorisches Optimierungsproblem reduziert [Fis95b]. Dabei werden die in einem Zustandsvektor repräsentierten Segmentierungshypothesen den primitiven Knoten eines sogenannten Attributflußgraphens zugeordnet. Die Motivation für die Transformation eines semantischen Netzes in diesen Graphen besteht in der guten Parallelisierbarkeit dieser Struktur [Fis95b]. Während ein Konzept mit vielen Attributen und

Relationen einen Engpaß für eine parallele Berechnung darstellen würde, läßt sich die Berechnung komplexer Konzepte durch die feine Granularität des Graphens entkoppeln. Die Kontrolle wird in Bild– und Sprachverarbeitung erfolgreich eingesetzt [Fis97].

3.5 Statistische Objektmodelle

Wie in Abs. 2.2 beschrieben, finden statistische Modelle in der Objekterkennung und –lokalisation Anwendung. Dabei werden die Objekte mit Hilfe von parametrischen Dichtefunktionen modelliert, wobei die Parameter anhand von Beobachtungen gelernt werden [Hor96a, Hor95, Hor96b]. Das Training der Modelle wird im Prinzip mit denselben Schätzverfahren gelöst, die bei der Berechnung der Parameter von Hidden–Markov–Modellen verwendet werden. Die Zuordnung zwischen den im Bild beobachteten Objektmerkmalen und den Modellmerkmalen ist unbekannt. Durch die statistische Modellierung der Zuordnung kann diese mit dem Expectation–Maximization–Algorithmus geschätzt werden. Die Lokalisierung von Objekten wird ebenso wie das Erlernen der Modelle auf ein Parameterschätzproblem zurückgeführt.

Diese Herangehensweise bietet eine Alternative zu der strukturellen Beschreibung von Objekten in Abs. 3.3. In der sogenannten holistischen Objekterkennung [Bük96, Kum97] sind beide Herangehensweisen koexistent, indem Teile des semantischen Netzes durch statistische Objektmodelle oder neuronale Netze ersetzt werden.

4 Anwendungsbeispiel

In dem Gesamtsystem zur wissensbasierten Szenenexploration kommen die oben genannten Arbeiten zum Einsatz [Pau98]. Das Ziel ist es, die Kooperation von signalnaher und modellbasierter Verarbeitung sowie die Durchführung von zielgerichteten Kameraaktionen zu demonstrieren. Das System besteht aus drei Hauptkomponenten. Positionen von Objekten werden datengetrieben aufgrund ihrer Farbe hypothetisiert; Farbe wird auch zur Bildsegmentierung eingesetzt. Eine zweite Komponente, die dem aktiven Sehen zuzuordnen ist, steuert eine Schwenk/Neige–Kamera, die auf einem Linearschlitten montiert ist; hiermit wird zum einen eine Schätzung der Entfernung der Objekte von der Kamera errechnet, zum anderen werden anschließend die Objekthypothesen formatfüllend aufgenommen. Ein drittes Modul zur wissensbasierten Bildanalyse verwendet eine Beschreibung von Objekten und ihren Relationen in Form eines semantischen Netzes um abschließend mittels explizit repräsentierter Kameraaktionen die einzelnen Objekte zu erkennen und die Kamera zu steuern. Das System ist in Bild 7 zusammengefaßt.

Experimente in einer Büroumgebung belegen, daß diese Architektur dazu verwendet werden kann, flexibel Objekte zu erkennen. In 17 Versuchen mit einer Wissensbasis, die nur die Szenenrepräsentation umfaßte, wurden fünf rote Objekte verwendet, von denen drei in der Wissensbasis modelliert waren. Die Objekte waren teilweise in mehrfacher Ausfertigung vorhanden. Durch die datengetriebene Hypothesengenerierung mit Histogrammrückprojektion [Swa91] wurden von 90 Objekten 71 erkannt. 17 Objekte kamen in mehr als einer Hypothese gleichzeitig vor. In 15 weiteren Experimenten mit einer Wissensbasis, die zwei Kameraaktionen (die Konzepte “direkteSuche” und

“LochernebKlebestift”) enthielt, wurden von 45 Objekten 38 gefunden. In 4000 Experimenten wurde belegt, daß eine Farbnormierung nicht unbedingt die Erkennungsrate erhöht [Csi98]; diese Aussage wird auch in [Fun98] getroffen.

Um die Größe des Suchbaums für die A*–Kontrolle auf ungefähr 300–600 Knoten zu reduzieren, wurden in den Experimenten zwei weitere Heuristiken eingeführt, um die roten von den anderen Regionen zu trennen. Die Farbsegmentierung für die Bestimmung der Regionen lieferte für die verschiedenen Versuche zwischen 54 und 152 Regionen, die an die primitiven Knoten des semantischen Netzes gebunden wurden. Durch die Bewertung der Farbe ließ sich die Zahl jedoch drastisch reduzieren, zum Beispiel von 54 auf 8 Regionen. Für beide Wissensbasen wird in 68 % der Fälle die richtige Region einem Objekt zugeordnet. Wenn die Ansicht des Objektes in die Berechnung einbezogen wird, kann sich diese Rate noch deutlich erhöhen, da die in den Experimenten verwendeten Merkmale von der Betrachtungsrichtung abhängig sind und die Objekte nur grob von Hand postiert wurden.

Abhängig von der Anzahl der gefundenen Regionenhypothesen benötigt die Verifikation dieser Hypothesen zwischen 0.14 sec und 0.66 sec auf einer SGI O2 (R10000, 195 MHz). Durch die in der Wissensbasis konzipierten Kameraaktionen ergibt sich die in Bild 6 b) dargestellte Situation, das heißt durch die Schwenkbewegung ergeben sich neue Hypothesen. Werden nach der Analyse des ersten Übersichtsbildes die dabei getroffenen Zuordnungen bei der Analyse des nach der Schwenkbewegung aufgenommenen Bildes übernommen, sinkt die Rechenzeit für das zweite Bild von durchschnittlich 0.76 sec auf 0.47 sec. 0.76 sec werden erreicht, wenn eine komplette Neuanalyse mit allen bis dahin gefundenen Regionen durchgeführt wird. Die Gesamtdauer für einen vollständigen Verarbeitungszyklus liegt bei ca. 2 min. Große Teile dieser Zeit werden zur Bewegung der Motoren, mit Warten auf das Erreichen der Zielposition, zur Medianfilterung der Rückprojektion und zur Segmentierung der Farbregionen verwendet.

5 Zusammenfassung und Ausblick

Dargestellt wurden die Architektur und Teile eines Systems zur modellbasierten Erkennung und Lokalisation von Objekten unter Verwendung einer steuerbaren Kamera. Die Einstellung der Kamera wird dabei durch explizit repräsentierte Kameraaktionen vorgenommen. Die signalnahe Verarbeitung wird teilweise parallel ausgeführt. Bei der Integration der zahlreichen Module zu einem Gesamtsystem war die durchgängige Software–Architektur in objektorientierter Programmierung, wie sie in [Pau97b] beschrieben wird, von großem Nutzen. Auch der Modul zur wissensbasierten Verarbeitung in semantischen Netzen bedient sich einer Klassenhierarchie in C++.

Die Kombination von statistischen Modellen und semantischen Netzen ist Gegenstand aktueller Arbeiten. Im Anwendungsbeispiel der Exploration einer Büroumgebung wurde die Tragfähigkeit des Ansatzes demonstriert. Die Ergebnisse zur Farbnormierung in [Csi98] legen es nahe, die Auswahl des Normierungsverfahrens von den Farben des Objektes abhängig zu machen und diese Auswahl in die Explorationsstrategie zu integrieren. Das Ziel der Arbeiten stellt es dar, alle in Abs. 2 und Abs. 3.3 beschriebenen Module in das Gesamtsystem zu integrieren und damit die Leistungsfähigkeit des Systems weiter zu steigern.

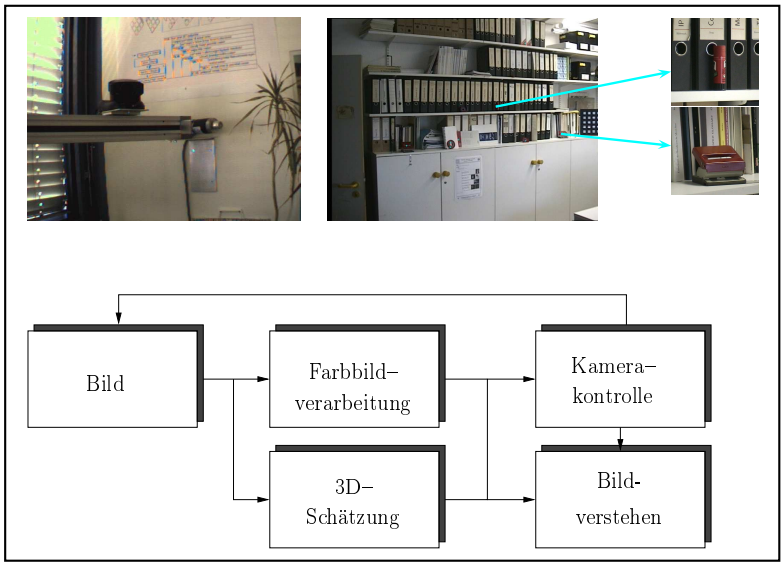


Bild 7. Anwendungsbeispiel im Gesamtsystem. Oben links die Kamera auf dem Linearschlitten, in der Mitte ein Überblicksbild über die Szene, rechts zwei Nahaufnahmen von Objekten, deren errechnete Position in der Szene durch Pfeile angegeben ist; unten schematisch die wesentlichen Module und ihre Abhängigkeiten

Literatur

[Ahl98] U. Ahlrichs: *Semantic Networks in Active Vision Systems – Aspects of Knowledge Representation and Purposive Control*, in H. Christensen, D. Hogg, B. Neumann (Hrsg.): *Knowledge Based Computer Vision, Dagstuhl–Seminar–Report*, Dagstuhl, 1998, S. 42.

[Alo88] J. Aloimonos, I. Weiss, A. Bandyopadhyay: *Active Vision*, *International Journal of Computer Vision*, Bd. 2, Nr. 3, 1988, S. 333–356.

[Baj92] R. Bajcsy, M. Campos: *Active and Exploratory Perception*, *Computer Vision, Graphics, and Image Processing*, Bd. 56, Nr. 1, 1992, S. 31–40.

[Bes89] P. Besl: *Advances in Machine Vision*, Kap. Active Optical Range Imaging Sensors, Springer Verlag, 1989, S. 1–63.

[Beß96] R. Beß: *Schnelle Tiefenberechnung aus monokularen Farbbildfolgen durch Faktorisierung* , in A. Pinz (Hrsg.): *Pattern Recognition 1996* , Bd. 90 von *Schriftenreihe der östereichischen Computergesellschaft*, R. Oldenburg, Wien, 1996, S. 215 – 226.

[Beß97] R. Beß: *Registering Depth Maps from Multiple Views Recorded by Color Image Sequences* , *Pattern Recognition and Image Analysis*, Bd. 7, Nr. 3, 1997.

[Bük96] U. Büker, J. Dunker, G. Hartmann, E. Seidenberg: *Aktives Sehen für die 3-D Objekterkennung in hybrider Architektur*, in B. Mertsching (Hrsg.): *Aktives Sehen in technischen und biologischen Systemen*, Proceedings in Artificial Intelligence, Infix, Sankt Augustin, Dezember 1996, S. 174–181.

[Csi98] L. Csink, D. Paulus, U. Ahlrichs, B. Heigl: *Color Normalization and Object Localization*, in V. Rehrmann (Hrsg.): *Vierter Workshop Farbbildverarbeitung*, Föhringer, Koblenz, 1998, S. to appear.

[Den95] J. Denzler, B. Heigl, D. Paulus: *Farbsegmentierung für aktives Sehen*, in V. Rehrmann (Hrsg.): *Erster Workshop Farbbildverarbeitung*, Bd. 15 von *Fachberichte Informatik*, Universität Koblenz–Landau, 1995, S. 9–12.

[Fis95a] V. Fischer: *Parallelverarbeitung in einem semantischen Netzwerk für die wissensbasierte Musteranalyse*, Infix-Verlag, St. Augustin, 1995.

[Fis95b] V. Fischer: *Parallelverarbeitung in einem semantischen Netzwerk für die wissensbasierte Musteranalyse*, Dissertation, Technische Fakultät, Universität Erlangen–Nürnberg, Erlangen, 1995.

[Fis97] J. Fischer, H. Niemann: *Applying a Parallel Any–Time Control Algorithm to a Real–World Speech Understanding Problem*, in *Proceedings of the 1997 Real World Computing Symposium*, Real World Computing Partnership, Tokyo, 1997, S. 382–389.

[Fun98] B. Funt, K. Barnard, L. Martin: *Is machine colour constancy good enough?*, in H. Burkhard, B. Neumann (Hrsg.): *Computer Vision — ECCV '98*, Nr. 1406 in Lecture Notes in Computer Science, Springer, Heidelberg, 1998, S. I/445–459.

[Han78] A. Hanson, E. Riseman: *VISIONS: A Computer System for Interpreting Scenes*, in A. Hanson, E. Riseman (Hrsg.): *Computer Vision Systems*, Academic Press, Inc., New York, 1978, S. 303–333.

[Har96] M. Harbeck: *Objektorientierte linienbasierte Segmentierung von Bildern*, Shaker Verlag, Aachen, 1996.

[Hei97] B. Heigl, D. Paulus: *Punktverfolgung in Farbbildsequenzen*, in D. Paulus, T. Wagner (Hrsg.): *Dritter Workshop Farbbildverarbeitung*, IRB-Verlag, Stuttgart, 1997, S. 87–92 & 105.

[Hei98] B. Heigl, D. Paulus, H. Niemann: *Tracking Points in Sequences of Color Images*, in *Proceedings 5th German–Russian Workshop on Pattern Analysis*, 1998, angenommen.

[Hor95] J. Hornegger, H. Niemann: *Statistical Learning, Localization, and Identification of Objects*, in *Proceedings of the* 5^{th} *International Conference on Computer Vision (ICCV)*, IEEE Computer Society Press, Boston, Juni 1995, S. 914–919.

[Hor96a] J. Hornegger: *Statistische Modellierung, Klassifikation und Lokalisation von Objekten*, Shaker Verlag, Aachen, 1996.

[Hor96b] J. Hornegger, E. Nöth, V. Fischer, H. Niemann: *Semantic Network Meet Bayesian Classifiers*, in B. Jähne, P. Geißler, H. Haußecker, F. Hering (Hrsg.): *Mustererkennung 1996*, Springer, Berlin, September 1996, S. 260–267.

[Kum90] F. Kummert: *Flexible Steuerung eines Sprachverstehenden Systems mit homogener Wissensbasis*, Dissertation, IMMD 5 (Mustererkennung), Universität Erlangen–Nürnberg, Erlangen, 1990.

[Kum97] F. Kummert, G. Fink, G. Sagerer: *Schritthaltende hybride Objektdetektion*, in E. Paulus, F. Wahl (Hrsg.): *Mustererkennung 1997*, Braunschweig, 1997, S. 137–144.

[Lev89] T. Levitt, T. Binford, G. Ettinger, P. Gelband: *Probalility Based Control for Computer Vision*, in *Proc. of DARPA Image Understanding Workshop*, 1989, S. 355–369.

[Mar82] D. Marr: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Company, San Francisco, 1982.

[Mas93] M. Mast: *Ein Dialogmodul für ein Spracherkennungs- und Dialogsystem*, Bd. 50 von *Dissertationen zur Künstlichen Intelligenz*, Infix, Sankt Augustin, 1993.

[Mat90] T. Matsuyama, V. Hwang: *SIGMA. A Knowledge-Based Aerial Image Understanding System*, Bd. 12 von *Advances in Computer Vision and Machine Intelligence*, Plenum Press, New York and London, 1990.

[McK85] D. McKeown, W. Harvey, J. McDermott: *Rule-Based Interpretation of Aerial Imagery*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 7, Nr. 5, 1985, S. 570–585.

[Nie90] H. Niemann, G. Sagerer, S. Schröder, F. Kummert: *ERNEST: A Semantic Network System for Pattern Analysis*, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Bd. 12, Nr. 9, 1990, S. 883–905.

[Pau97a] D. Paulus, J. Hornegger: *Pattern Recognition of Images and Speech in C++*, Advanced Studies in Computer Science, Vieweg, Braunschweig, 1997.

[Pau97b] D. Paulus, J. Hornegger: *Pattern Recognition of Images and Speech in C++*, Advanced Studies in Computer Science, Vieweg, Braunschweig, 1997.

[Pau98] D. Paulus, U. Ahlrichs, B. Heigl, H. Niemann: *Wissensbasierte aktive Szenenanalyse*, in P. Levi (Hrsg.): *Mustererkennung 1998*, Springer, Heidelberg, September 1998, angenommen.

[Pös97] J. Pösl, H. Niemann: *Wavelet Features for Statistical Object Localization without Segmentation*, in *Proceedings of the International Conference on Image Processing (ICIP)*, Bd. 3, IEEE Computer Society Press, Santa Barbara, Kalifornien, USA, Oktober 1997, S. 170–173.

[Pös98] J. Pösl, B. Heigl, H. Niemann: *Color and Depth in Appearance Based Statistical Object Localization*, in H. Niemann, H.-P. Seidel, B. Girod (Hrsg.): *Image and Multidimensional Digital Signal Processing '98*, Infix, Alpbach, Österreich, Juli 1998, S. 71–74.

[Rim93] R. Rimey: *Control of Selective Perception using Bayes Nets and Decision Theory*, Department of Computer Science, College of Arts and Science, University of Rochester, Rochester, New York, 1993.

[Sal95] R. Salzbrunn: *Wissensbasierte Erkennung und Lokalisierung von Objekten*, Shaker Verlag, Aachen, 1995.

[Sch90] S. Schröder: *Integration einer Wissenserwerbkomponente in eine Systemumgebung für die Musteranalyse*, Reihe 10: Informatik / Kommunikationstechnik, VDI Verlag, Düsseldorf, 1990.

[Sch98] I. Scholz: *A Distributed Image Processing System*, Studienarbeit, IMMD 5 (Mustererkennung), Universität Erlangen–Nürnberg, Erlangen, 1998.

[Swa91] M. J. Swain, D. H. Ballard: *Color Indexing*, *International Journal of Computer Vision*, Bd. 7, Nr. 1, November 1991, S. 11–32.

[Tom91] C. Tomasi, T. Kanade: *Detection and Tracking of Point Features*, CMU-CS-91-132, Carnegie Mellon University, 1991.

[Tom92] C. Tomasi, T. Kanade: *Shape and Motion from Image Streams under Orthography: a Factorization Method*, *International Journal of Computer Vision*, Bd. 9, Nr. 2, Nov. 1992, S. 137–154.

[Tsa88] R. Y. Tsai, R. K. Lenz: *Real Time Versatile Robotics Hand/Eye Calibration Using 3D machine vision*, in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, IEEE Computer Society Press, Philadelphia, April 1988, S. 554–561.

[Wag94] A. Wagner: *Kollisionsfreie Bahnplanung für einen Roboter*, Diplomarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Mustererkennung (Informatik 5), 1994.

[Wil94] R. Willson: *Modeling and Calibration of Automated Zoom Lenses*, PhD thesis, Carnegie Mellon University, Pittsburgh, 1994.