Inhaltsverzeichnis

1 Mitarbeiter am LME

Lehrstuhl für Mustererkennung (Informatik 5)

Leiter: **Prof. Dr.-Ing. H. Niemann**

Mitarbeiter:

Adelhardt, J., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.04.00
Ahlrichs, U., Dipl.-Inf.	(wiss. Mitarb., DFG)	01.06.96
Batliner, A., Dr.-Phil.	(wiss. Mitarb., BMBF)	01.01.97
Buckow, J., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.04.97
Caputo, B., M.Sc.	(wiss. Mitarb., Grad.-Stip.)	01.11.99
Deinzer, F., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.01.99
Denzler, J., Dr.-Ing.	(wiss. Assistent)	01.01.93
Deventer, R., Dipl.-Inf.	(wiss. Mitarb., SFB 396)	01.02.99
Drexler, Ch., Dipl.-Inf.	(wiss. Mitarb., DIROKOL)	01.02.98
Fentze, W.	(Programmierer)	05.12.88
Fischer, J., Dipl.-Inf.	(wiss. Mitarb.)	bis 29.02.00
Frank, C.M., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.09.99
Gallwitz, F., Dipl.-Inf.	(wiss. Mitarb., DFG)	bis 29.02.00
Gebhard, A., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.04.98
Haas, J., Dipl.-Inf.	(wiss. Mitarb.)	bis 29.02.00
Heigl, B., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.04.97
Hertlein, H., Dipl.-Inf.	(wiss. Mitarb., MEDAV/DCS)	01.01.00
Huber, R., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.03.97
Karadag, C.	(Sekretärin, 1/2, SFB 603)	01.07.98
Koppe, I.	(Sekretärin, 1/2)	18.12.92
Niemann, H., Dr.-Ing.	(Professor)	24.09.75
Nöth, E., Dr.-Ing.	(Akad. Oberrat)	01.02.85
Obermayer, W.	(Programmierer)	bis 31.12.00
Ohler, U., Dipl.-Inf.	(Stip., Boehringer)	01.01.97
Pal, I.	(wiss. Mitarb, SFB 539)	01.07.00
Paulus, D., PD Dr.-Ing.	(Akad. Oberrat)	01.03.87
Popp. F.	(Techniker)	01.09.85
Reinhold, M., Dipl.-Ing.	(wiss. Mitarb., Grad.-Stip.)	09.06.99
Schmidt, J., Dipl.-Inf.	(wiss. Mitarb.)	01.05.00
Shi, R., MS Sci	(wiss. Mitarb., BMBF)	01.02.00
Stemmer, G., Dipl.-Inf.	(wiss. Mitarb.)	13.10.99
Vogt, F., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	15.10.00
Warnke, V., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.08.96
Zobel, M., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.01.98

Gäste

Gäste:

Baggenstoss, P. Dr.	(USA)	01.02.2000
Bunke, H., Prof. Dr.	(Schweiz)	28.10.2000
Dalbelo, B., Dr.	(Kroatien, DAAD)	09.05. – 03.06.2000
De Mori, R., Prof. Dr.	(Frankreich)	26.10.–29.10.2000
Draper, B., Prof. Dr.	(USA)	11.11. – 25.11.2000
Iatsko, V., Dr.	(Russ. F., DAAD)	01.09. – 30.11.2000
Huang, Yu, Dr.	(China, AvH)	01.09.99 – 30.08.2000
Yuan, C.	(China, KAS)	03.04.97

2 Einleitung

Seit 25 Jahren wird am Lehrstuhl das Problem der „Mustererkennung“ untersucht, in dem ganz allgemein die automatische Transformation einer von einem geeigneten Sensor gelieferten Folge von Abtastwerten eines Signals in eine den Anforderungen der Anwendung entsprechende symbolische Beschreibung gesucht wird. In der Bildverarbeitung werden hierfür Sensoren eingesetzt, die unter Umständen vom Rechner gesteuert werden können oder mit spezieller Beleuchtung gekoppelt sind. Sie liefern Informationen in einem oder mehreren Kanälen. Bei der Verarbeitung von zusammenhängend gesprochener Sprache werden Mikrophone als Sensoren verwendet.

Eine symbolische Beschreibung kann zum Beispiel eine diagnostische Bewertung einer Bildfolge aus dem medizinischen Bereich enthalten, die Ermittlung, Benennung und Lokalisation eines erforderlichen Montageteils für einen Handhabungsautomaten umfassen oder aus der Repräsentation der Bedeutung eines gesprochenen Satzes bestehen. Die Lösung dieser Aufgaben erfordert sowohl Verfahren aus der (numerischen) Signalverarbeitung als auch aus der (symbolischen) Wissensverarbeitung. Die Ermittlung einer symbolischen Beschreibung wird auch als Analyse des Musters bezeichnet.

Der Lehrstuhl bearbeitet hauptsächlich zwei Themenkomplexe, nämlich die wissensbasierte Analyse von Bildern und Bildströmen sowie das Verstehen gesprochener Sprache und Generierung einer Antwort. In der wissensbasierten Bildanalyse werden sowohl grundsätzliche Arbeiten zur Bildverarbeitung und zur Repräsentation und Nutzung problemspezifischen Wissens als auch spezielle Arbeiten zur Entwicklung eines vollständigen, rückgekoppelten Systems für die schritthaltende Analyse dreidimensionaler Szenen durchgeführt. Eine Brücke zwischen Visualisierung und Analyse wird im Sonderforschungsbereich 603 mit dem Thema „Modellbasierte Analyse und Visualisierung
hergestellt, dessen Sprecher Prof. Niemann ist. Eine Verknüpfung zwischen Bild– und Sprachanalyse wird im Projekt Smartkom hergestellt, das vom BMBF als Leitprojekt gefördert wird.

In der Spracherkennung konzentrierten sich die Arbeiten auf die Entwicklung eines Systems, das über einen begrenzten Aufgabenbereich einen Dialog mit einem Benutzer führen kann, wobei gesprochene Sprache für die Ein– und Ausgabe verwendet wird. Hierbei fand eine komplette Neuimplementierung des der Verstehens– und Dialogphase sowie die Portierung auf eine komplett neue Anwendung statt (von Auskünften über InterCity-Züge zu Kinoauskunft). Als weitere Anwendung werden in der Spracherkennung seit 1993 Teilprobleme im Rahmen des *Verbmobil*-Vorhabens untersucht. Ziel des Gesamtvorhabens, das 2000 erfolgreich abgeschlossen wurde, war die Entwicklung eines portablen Übersetzungsgerätes.

Ein Problem, das in jedem der drei Themenkomplexe eine Rolle spielt, ist die Akquisition, Repräsentation und Nutzung des Wissens, das zur Analyse von Bildern, Sprache und Sensordaten bzw. zum Verstehen der Bedeutung erforderlich ist. In diesem Zusammenhang spielen heute statistische Sprach- und Objektmodelle eine wichtige Rolle. Dieser Weg wird auch in zwei Projekten zur Genomanalyse und zur Modellierung von Prozessketten eingeschlagen. Es ist unter Umständen erforderlich, dass zusätzlich zum Verstehen der Bedeutung auch noch eine sinnvolle Systemreaktion geliefert wird, zum Beispiel auf die Anfrage eines Benutzers eine richtige Auskunft des Systems oder eine Bewegung des Montageroboters oder der Kameramotoren aufgrund der Ergebnisse der Bildanalyse.

3 Bildanalyse

Leitung: D. Paulus
(U. Ahlrichs, P. Baggenstoss, B. Caputo, F. Deinzer, J. Denzler, J. Drexler, A. Gebhard, B. Heigl, Y. Huang, W. Obermayer, Pal, I., M. Reinhold, J. Schmidt, F. Vogt, M. Zobel,)

Schwerpunkt der Arbeiten im Bereich der Bildanalyse am Lehrstuhl ist die Objekterkennung. Arbeiten zur wissensbasierten Bildanalyse auf der Basis von semantischen Netzen wurden hierzu fortgesetzt. Ebenso wurden die Aktivitäten im Bereich der statistischen Objektmodellierung und -erkennung ausgebaut. Objekterkennung ist auch Teil der neu aufgenommenen Arbeiten zur erweiterten und virtuellen Realität.

Als weiterer Forschungsschwerpunkt hat sich der Bereich Rechnersehen für autonome mobile Systeme etabliert. Darunter fallen grundlagenorientierte Arbeiten auf dem Gebiet der probabilistischen Modellierung von Sensordaten– und Aktionsfolgen für das aktive Rechnersehen, optimale Kameraparameterauswahl für die Objekterkennung und –verfolgung des Sonderforschungsbereichs 603, sowie Eigenraumverfahren zur 3D–Objektlokalisation und Klassifikation (Teilprojekt B2). Bildbasierte Modelle, wie der Lumigraph oder das Lichtfeld, die im Teilprojekt C2 des Sonderforschungsbereichs 603 entwickelt und erweitert werden, fließen in allen Bereichen als eine Alternative zu geometriebasierten Objekt– und Umgebungsmodellen ein. Als Anwendungsszenario dient der Bereich der Service– und Dienstleistungsroboter. Dort wurde sowohl eine Objekterkennungskomponente für Pflegeroboter im Krankenhaus (Projekt DIROKOL) als auch – in enger Kooperation mit der Sprachverarbeitung – das mobile System MOBSY entwickelt, das während der 25–Jahr–Feier den Gästen als „Empfangsdame“ zur Verfügung stand. Für die laufenden Projekte auf dem Gebiet der probabilistischen Folgenmodellierung sowie auf auf dem Gebiet des Rechnersehens für autonome mobile Systeme steht seit Anfang 1998 das auf der Plattform XR4000 der Firma Nomadic basierende System Mobsy zur Verfügung. Die beiden auf der Plattform installierten Rechnersysteme (Pentium Pro und Dual Pentium II 300) ermöglichen eine vollständig Autonomie; die Verbindung zum Rechnercluster des Lehrstuhls wird über ein Funkethernet sichergestellt. Die Plattform verfügt neben Infrarot–, Ultraschall– und mechanischen Sensoren über einen Stereo–Kopf mit Schwenk–Neige–Vergenz–Steuerung und Farbkameras zur visuellen Wahrnehmung der Umwelt. Die Hardwarekonstruktion der Plattform wurde im vergangenen Jahr um zahlreiche Aspekte erweitert, wie beispielsweise zahlreiche Lüfter, um einen Dauerbetrieb sicherzustellen, oder LCD–Bildschirme zur Überwachung des Kamerabildes der beiden Kameras.

Während der Feier anlässlich 25–jährigen Bestehens des Lehrstuhls für Mustererkennung wurde eine erste Anwendung für das mobile System Mobsy realisiert und dem Publikum vorgestellt: Mobsy wartete im 9. Stock vor den Aufzügen, erkannte ankommende Gäste und nahm diese in Empfang. Danach gab er einen kurzen Überblick über angebotene Vorführungen. Außerdem gab Mobsy bei Fragen Auskünfte über laufende Arbeiten am Lehrstuhl. In dem System, das während einer zweistündigen Laufzeit ohne Eingriff von außen robust und fehlertolerant lief, wurde eine erfolgreiche Integration zahlreicher Module zur Sprach– und Bildverarbeitung in einem Serviceroboter Szenario demonstriert. Die Akzeptanz bei den Besuchern der 25–Jahr–Feier machte deutlich, dass natürliche Sprache und Dialog als Schnittstelle zum System sowie aktive

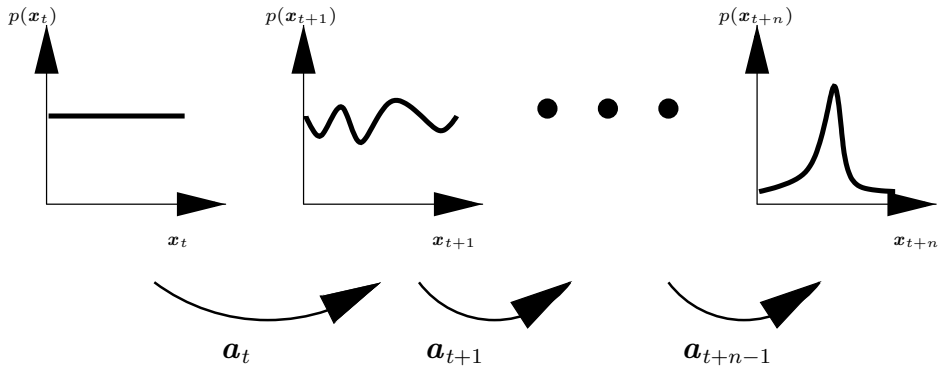


Bild 1: Reduktion der Unsicherheit in der Zustandschätzung durch gezielte Auswahl der Aktionen

Kamerasteuerung zur Gesichtsverfolgung wichtige Aspekte in einem solchen Anwendungsgebiet darstellen. Regelmäßige, automatische Rekalibrierung mittels visueller Information sowie Hinderniserkennung mittels Infrarotsensorik stellte den robusten Betrieb trotz der zahlreichen Besucher im 9. Stock sicher.

3.1 Objekterkennung und Szenenanalyse

Im Rahmen eines DFG Forschungsstipendiums wurden im vergangenen Jahr schwerpunktmäßig Ansätze zur optimalen Sensordatenaquisition untersucht. In vielen Anwendungen im dynamischen Rechnersehen beobachtet man ein Durchlaufen des sogenannten Sensordaten– und Aktionszyklus. Das System nimmt über eine oder mehrere Kameras Sensordaten auf und reagiert nach einer Analyse der wahrgenommenen Umgebung — allgemein gesprochen nach einer Zustandsschätzung — mit einer Aktion. Aktionen sind beispielsweise das Bewegen der Kamera zur Verfolgung bewegter Objekte, die Einstellung optimaler Kameraparameter (Zoom, Fokus) bei der Klassifikation, oder auch die Auswahl geeigneter Algorithmen und Strategien im wissensbasierten Rechnersehen. Das Problem besteht nun in der Ermittlung einer — bezüglich eines definierten Ziels — optimalen Aktionsfolge.

Ausgehend von dieser Problembeschreibung wurde eine probabilistische Modellierung des Zyklus aus Sensordaten– und Aktionsfolgen untersucht. Schwerpunkt war die Entwicklung eines allgemeinen, informationstheoretischen Prinzips zur Aktionsauswahl. Das Prinzip besteht anschaulich in der Minimierung der Unsicherheit in der Zustandsschätzung, und kann somit auf eine Vielzahl von Anwendungen im Bereich des Rechnersehens angewendet werden (siehe Bild 1).

Um die Unsicherheit in der Zustandsschätzung zu messen wurde ein Formalismus basierend auf Shannons Informationstheorie eingeführt. Das Ziel ist, die Unsicherheit und Mehrdeutigkeit in der a priori Wahrscheinlichkeit Schritt für Schritt zu reduzieren. Fanos Ungleichung liefert dazu die theoretische Rechtfertigung.

Die wichtigste Größe in dem sequentiellen Entscheidungsprozess ist die von den ausgewählten

Kameraparametern abhängige Transinformation. Die Transinformation zwischen den Verteilungen über den Zustandsraum und den Beobachtungen gibt an, wieviel Information man über den Zustand gewinnt, wenn man die Umgebung oder das Objekt unter Verwendung der gewählten Kameraparameter beobachtet. Eine Auswahl der Kameraparametern aufgrund des Maximums der Transinformation führt somit zu Kameraparametern, unter deren Verwendung die Unsicherheit in der Zustandsschätzung im Mittel am meisten reduziert wird. Erfolgt nach Aufnahme der Sensordaten eine Maximum a posteriori Schätzung und verwendet man diese a posteriori Verteilung über den Zustandsraum als a priori Verteilung für den nächsten Zeitschritt, so erhält man einen sequentiellen Entscheidungsprozess. Als eine der wichtigsten Eigenschaften dieses Entscheidungsprozesses konnte gezeigt werden, dass dieser gegen eine Verteilung über den Zustandsraum konvergiert.

Um die Relevanz des entwickelten Ansatzes für sequentielle Entscheidungsprozesse für praktische Bildverarbeitungsproblem aufzuzeigen, wurde das Problem der optimalen Blickrichtungskontrolle bei der Objekterkennung untersucht, d.h. der Einstellung der Schwenk–/Neige– und Zoomparameter einer aktiven Kamera. Eine Menge ähnlicher Objekte, die jeweils nur durch gezielte Betrachtung bestimmter Objektbereiche unterschieden werden konnte, diente als Stichprobe. Ohne eine aktiven Blickkontrolle waren die Objekte nicht zu unterscheiden. Eine zufällige Blickkontrolle führte zu einer Erkennungsrate von 81.4% bei durchschnittlich 7.2 Ansichten pro Objekt und Klassifikation. Mittels der entwickelten Methodik zur optimalen Sensordatenaquisition wurde eine Erkennunsrate von 99.8% bei durchschnittlich 2.5 Ansichten erreicht. Als Klassifikator wurde ein statistischer Eigenraumklassifikator verwendet.

In Zukunft wird diese Methodik auf die Zustandsschätzung dynamischer Systeme übertragen, und auf das Problem der optimalen Brennweitenauswahl mehrere Kameras bei der Objektverfolgung angewendet.

Im Rahmen des von der DFG geförderten Graduiertenkollegs „Dreidimensionale Bildanalyse und -synthese“ wurde die Arbeit zur erscheinungbasierten, statistischen Objekterkennung fortgeführt.

Bei diesem Ansatz werden die Merkmale direkt ohne vorhergehende Segmentierung aus den Bilddaten bestimmt. Verwendet werden dabei lokale Merkmale, die aus den Koeffizienten der Wavelet–Multiskalen–Analyse berechnet werden. In Bild 2 ist die prinzipielle Funktionsweise des Objekterkennungssystems dargestellt. In der Trainingsphase wird das Objekt aus verschiedenen Blickwinkeln aufgenommen, und die Parameter der Modelldichtefunktion werden geschätzt. In der Erkennungsphase können dann die Klasse und Lage des Objektes mit Hilfe von Maximum-Likelihood-Schätzungen aus der Modelldichtefunktion ermittelt werden.

Im vergangenen Jahr wurde dieser Ansatz in folgenden Punkten verändert: durch eine verbesserte Berechnung der Merkmale konnte die Schätzung der Parameter der Dichtefunktion verbessert werden. Beschleunigt wurde das Verfahren durch Ersetzen des bisherigen Interpolationsschemas, das einer linearen Interpolation nachempfunden wurde, durch eine bilineare Interpolation. Außerdem konnte durch Modifikationen bei den Geometrieberechnungen die numerische Stabilität erhöht werden.

Getestet wurde dieser Ansatz auf einer Stichprobe, die aus 13 Objekten aus dem Krankenhaus- und Bürobereich besteht. Aufgenommen wurde je Objekt 3720 Bilder gleichmäßig verteilt über eine Halbkugel. Die Hälfte der Bilder wurde für das Training verwendet, die andere Hälfte für die

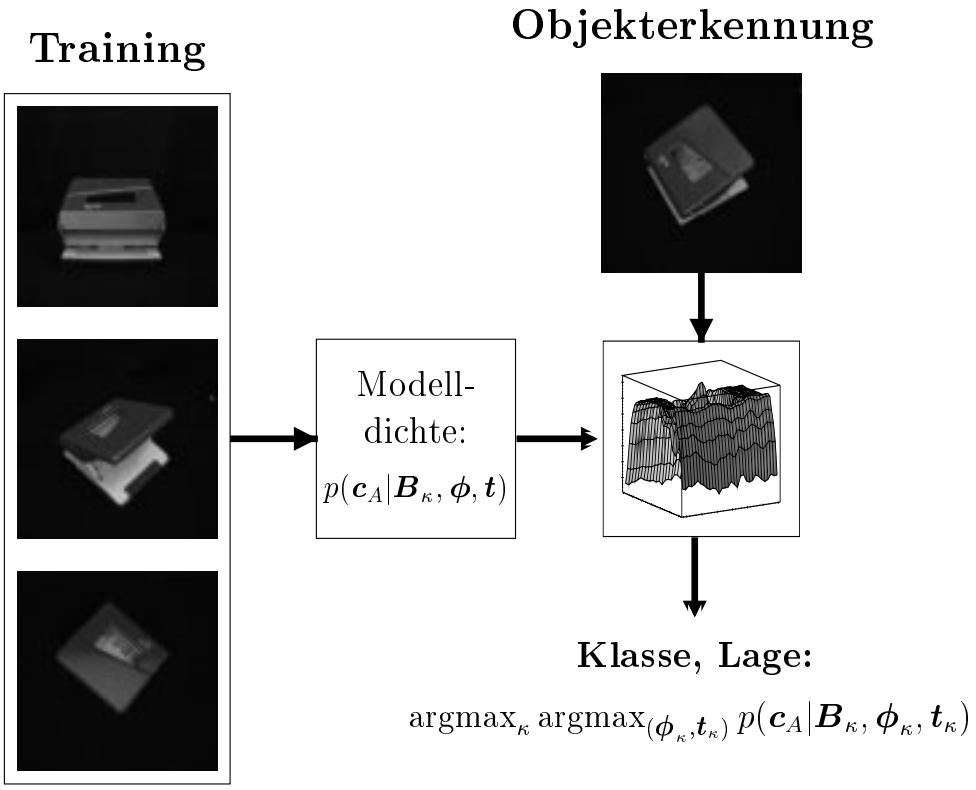


Bild 2: Prinzipielle Funktionsweise des Objekterkennungssystems

Experimente. In der Erkennungsphase lag die durchschnittliche Lokalisationsrate bei homogenen Hintergrund bei 91% und die durchschnittliche Klassifikationssrate bei 98%.

Um auch Objekte, die von mehreren Seiten gleich aussehen und sich nur von einer Seite unterscheiden, zuverlässig zu klassifizieren, wurde das Objekterkennungssystem mit einer Ansichtenplanung [48] kombiniert. Diese bestimmt ausgehend von einem ersten, von einem zufällig gewählten Blickwinkel aufgenommenen Bild die optimale Aktion (z. B. Drehtellerbewegung), um dann die zweite Ansicht zur Klassifikation zu benutzen. Bei vier Testobjekten, wobei jeweils zwei nahezu identisch waren, konnte so die Erkennungsrate von 72% auf 99,5% erhöht werden.

Letztlich wurde mit Arbeiten begonnen, um die Objekterkennung bei heterogenem Hintergrund und Verdeckungen zu verbessern. Dazu werden zusätzlich zum Objekt eine eigene Hintergrundklasse und eine Zuweisungsfunktion definiert. Die Zuweisungsfunktion weist jeden Merkmalsvektor entweder dem Objekt oder dem Hintergrund zu.

Ein besonders schwieriges Problem ist es, mehrere Objekte in einer Szene zu erkennen. Aus Sicht der Mustererkennung kann hierzu ein Objekt als Kontext eines anderen angesehen werden. Hierzu bietet es sich an, ein statistisches Modell zu verwenden. Ein so genanntes Maximum A Posteriori-Markov Zufallsfeld (MAP-MRF) und mit der physikalischen Theorie – der Spin-Glass Theory – motiviert. Ergebnisse in [9] belegen die Tragfähigkeit dieses Ansatzes.

In Abs. 4 findet sich eine ausführliche Darstellung in englischer Sprache. Im vergangenen

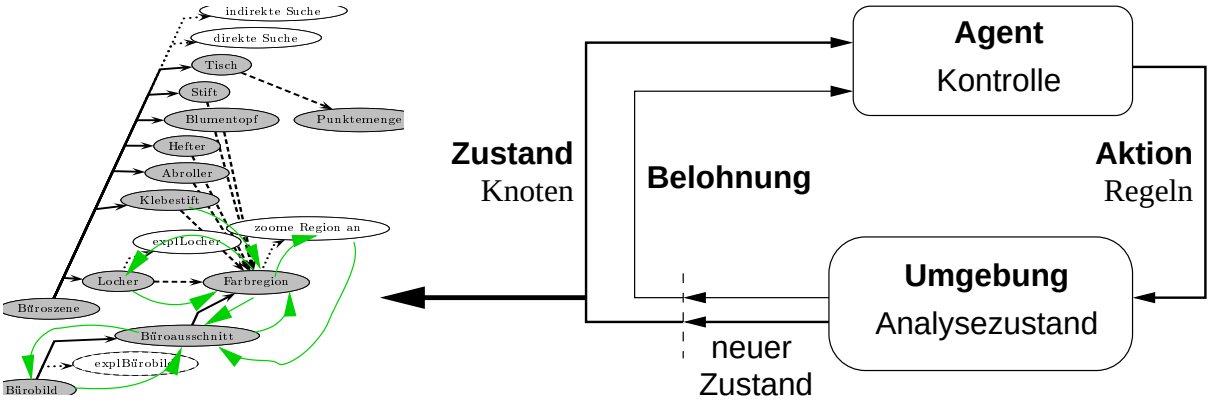


Bild 3: Reinforcement Learning zur Instantiierung eines semantischen Netzes.

Jahr wurden die Arbeiten auf dem Gebiet der wissensbasierten Bildverarbeitung im DFG–Projekt Strategie und Aktion als Lernziel der visuellen Exploration fortgeführt. Der Schwerpunkt der Arbeiten in diesem Projekt liegt auf dem Lernen von Verarbeitungsstrategien zur Nutzung von Wissen, das in Form eines semantischen Netzes repräsentiert wird. Die Strategien legen fest, in welcher Reihenfolge die Knoten im Netz während einer visuell gestützten Exploration berechnet werden. Als Lernverfahren werden Methoden aus dem Reinforcement Learning verwendet, deren prinzipielle Vorgehensweise sich leicht auf die im Rahmen des Projekt eingesetzte Kontrolle übertragen lässt. Der Zusammenhang zwischen den Lernverfahren und dem im Projekt verfolgten wissensbasierten Ansatz ist in Bild 3 dargestellt. Im Reinforcement Learning wählt ein Agent basierend auf einer erworbenen Entscheidungstaktik in jedem besuchten Zustand eine Aktion aus und erhält für diese Aktion eine Belohnung, die entweder gut oder schlecht ausfallen kann. Die Aktionen entsprechen hier Regeln zur Berechnung von Knoten im Netz, während die Knoten selbst als Zustände fungieren. Bei der Anwendung der Regeln auf die Knoten des Netzes entstehen so genannte Suchbaumknoten, die den Zustand der Analyse repräsentieren. Ein Analysezustand bildet die Umgebung für die Lernverfahren. Bild 3 zeigt im linken Teil zusätzlich ein semantisches Netz, bei dem im unteren Teil der Wissensbasis mit Hilfe von Pfeilen eine Folge von Berechnungen der Knoten gezeigt wird. Damit wird die gelernte Entscheidungstaktik verdeutlicht, die nicht nur die Folge einer Anwendung von Regeln beschreibt, sondern auch die Reihenfolge festlegt, in der die Knoten des Netzes besucht werden.

Als Bewertung für eine Folge aus Aktions-Zustandspaaren wird die Anzahl generierter Suchbaumknoten herangezogen. Die Auswertung von Monte Carlo Lernverfahren in dem beschriebenen Anwendungsgebiet ergab, dass für eine Wissensbasis, die Information über drei zu suchende Objekte enthält, die Anzahl der generierten Suchbaumknoten bei einer geeigneten Parametereinstellung nach 10000 Iterationen 25 Knoten betrug. Hierzu wurden 59 Aktionen ausgeführt. Dies entspricht einer Rechenzeit von 9.5 sec. Experimente mit der eigens für das Projekt angeschafften 3-D-Laserkamera, die durch Projektion von Laserpulsen die Entfernung zwischen Kamera und einer Szene bestimmt, zeigten, dass die Tiefenbestimmung auf mehrfarbige oder aus verschiedenen Materialien bestehende Objekte sehr empfindlich reagiert.

Ein weiterer Schwerpunkt der Forschungsarbeiten im Teilprojekt B2 des Sonderforschungsbereichs 603 sind Verfahren zur aktiven Objekterkennung in der Bildverarbeitung.

Die Erkennungsrate bei der Klassifikation bzw. die Genauigkeit der Lokalisation hängen aufgrund von Ambiguitäten zwischen einzelnen Objekten entscheidend von den gewählten Sensordaten ab. Um nur eine minimale Anzahl von Objektaufnahmen zu erzeugen, setzt man eine gezielte *Ansichtenplanung* ein. Deren Aufgabe ist es, für ein gegebenes Objekt, z. B. durch eine gezielte Wahl der Blickrichtung, automatisch signifikante Ansichten zu generieren, welche die Objekterkennung robuster und zuverlässiger gestalten. Zu diesem Zweck wurde am Lehrstuhl ein Verfahren entwickelt, das — basierend auf den Methoden des Reinforcement Learning — automatisch signifikante Ansichten für jedes Objekt in der Objektdatenbank lernt [11, 10, 48]. Ein großer Vorteil des entwickelten Verfahrens liegt darin, dass es klassifikatorunabhängig arbeitet. Das bedeutet, dass jederzeit ein neues Verfahren zur Klassifikation und Lokalisation verwendet werden kann, ohne Änderungen an den Methoden zur Ansichtenplanung durchführen zu müssen.

In Bild 4 ist ein Beispiel für derartige Mehrdeutigkeiten zwischen einzelnen Objekten dargestellt. Die Playmobil Männchen unterscheiden nur anhand kleiner Accessoires, die auch nur aus bestimmen Blickrichtungen sichtbar sind. Um nun die einzelnen Männchen unterscheiden zu können, muss eine Aktion – eine Änderung der Blickrichtung – gewählt werden, aus der eine neue Ansichtsposition resultiert, von der aus die Unterscheidungsmerkmale sichtbar sind.

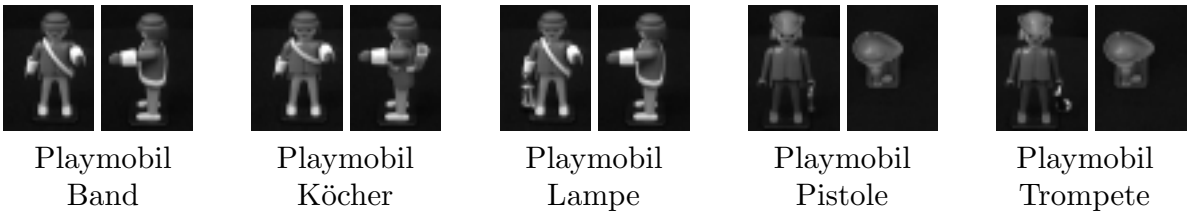
Die Suche nach einer neuen optimalen Ansicht ist in dem entwickelten Verfahren als Optimierungsproblem für die Position mit der maximal zu erwartenden Signifikanz, d. h. maximalen Unterscheidbarkeit der Objekte, formuliert. Dazu wird aus dem während des Trainings gesammelten Wissen — den evaluierten Beispielansichten — eine reelwertige, kontinuierliche Funktion für die erwartete Signifikanz beliebiger Ansichten approximiert. Während des Ansichtentrainings wird die Güte einzelner Aktionen von verschiedenen Ansichtspositionen ausgewertet. Mit gängigen Optimierungsverfahren wird in dieser Funktion nun nach Maxima gesucht, aus deren Lage daraufhin eine neue Ansichtsposition mit maximaler Signifikanz berechnet wird.

Ein Schwerpunkt in den Forschungen zur Ansichtenplanung in diesem Jahr war die Erweiterung des Verfahrens auf zwei Freiheitsgrade für die Kameraposition und die detaillierte Evaluation der Leistungsfähigkeit auf realen Bildern. Dabei hat sich gezeigt, dass das entwickelte Verfahren sehr gut in der Lage ist, auch unterschiedliche Objektmehrdeutigkeiten aufzulösen und neue Ansichtspositionen auszuwählen, die eine sichere Klassifikation erlauben.

Im Bereich der Ansichtenplanung wird das Verfahren im folgenden Jahr so erweitert, dass komplexe Mehrdeutigkeiten aufgelöst werden können. Aktuell ist das Verfahren in der Lage, die global beste Ansicht eines Objektes zu finden. Das Verfahren soll nun so erweitert werden, dass weitere „gute“ Ansichtspositionen in Abhängigkeit der bisherigen Aufnahmen berechnet werden können.

Ein weiterer Schwerpunkt der Forschungsarbeiten wird die Fusion mehrerer Kamerabilder sein. Die durch die Ansichtenplanung generierten neuen Bilder sollen nicht isoliert betrachtet werden, sondern die bisherigen Klassifikation- und Lokalisationsergebnisse verbessern. Dazu werden bisherige Klassifikationsverfahren um eine statistische Komponente erweitert, die eine wahrscheinlichkeitstheoretisch fundierte Behandlung des Fusionsproblems zulassen.

Zur Objektlokalisierung wurden verschiedene Verfahren untersucht und erweitert, die auf Farbhistogrammen basieren. Die Stichprobe ist im WWW zu finden und ermöglicht es, andere Ver-



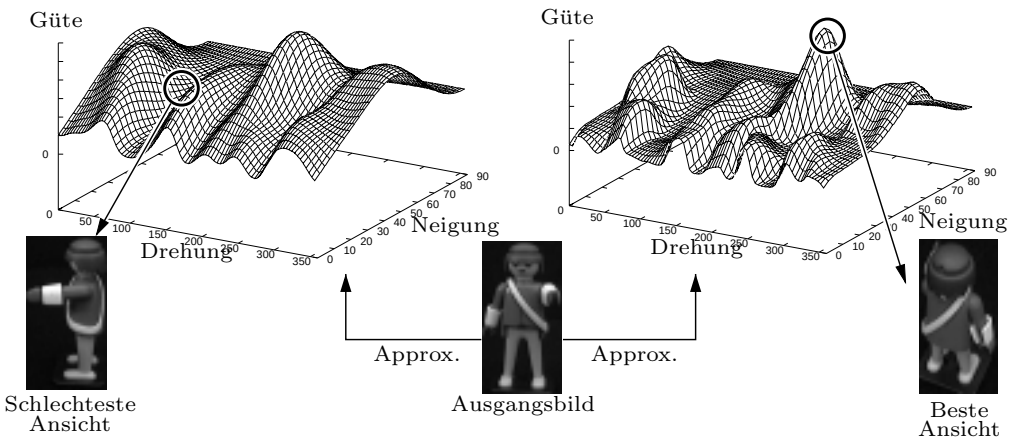


Bild 4: Beispiel für Ambiguitäten zwischen Objekten und deren Auflösung mittels Ansichtenplanung. Die Playmobil-Männchen sind nur aus bestimmten Blickrichtungen unterscheidbar

fahren zu testen. Die interessierenden Objekte werden dabei mit einem Zoom–Objektiv formatfüllend aufgenommen und das Farbhistogramm in unterschiedlichen Farbräumen und wählbarer Quantisierung gespeichert. Zur Lokalisierung eines Objekts wird die Szene zunächst formatfüllend mit einer geringen Brennweite aufgenommen. Anschließend wird die Schwenk-Neigeposition der Kamera und die Brennweite so verändert, dass das hypothetisierte Objekt sich formatfüllend in der Bildmitte befindet. Für dieses Verfahren ist es erforderlich, die Größe des gesuchten Objekts im Bild ungefähr zu kennen. Dies kann durch die Kenntnis der Entfernungen der Objekte zur Kamera errechnet werden. Hierzu wird durch eine gezielte Beweung der Kamera, die auf einem Linearschlitten montiert ist, aus einer Bildfolge eine Schätzung der Tiefenwerte an den Punkten durchgeführt, die sich in Farbbildern gut detektieren lassen. Die Nahaufaufnahmen werden dann segmentiert. In [44] wurden zwei Regionensegmentierungsverfahren verglichen, wobei das Ziel verfolgt wurde, eine Szenenexploration bestmöglich zu gestalten. In laufenden Arbeiten wird in Zusammenarbeit mit der Colorado State University verglichen, wie sich die Parameter der Segmentierungsverfahren automatisch so bestimmen lassen, dass die Objekterkennungsrate in der Szenenexploration maximiert wird.

Bild 5 zeigt die Ergebnisse der am Lehrstuhl bislang verwendeten Regionensegmentierung sowie des in Koblenz entwickelten so genannten Color Structure Codes.

Bild 5: Regionensegmentierung: links das Eingabebild, Split & Merge Segmentierung, Ergebnis der CSC-Segmentierung

Die Arbeit Untersuchung von neuronalen Netzen zur Objekterkennung und –lokalisation wurde fortgesetzt. Ziel ist die Erkennung und Lokalisierung von 3D–Objekten durch einzelne 2D–Grauwertbilder. Die Motivation eines neuronalen Ansatzes liegt in ihren vielen positiven Eigenschaften wie z.B Lernfähigkeit, höhere Fehlertoleranz usw.

Im Gegensatz zu üblichen neuronalen Verfahren werden hier segmentierungsfreie Methoden verwendet, um unerwünschte Unter– bzw. Übersegmentierung zu vermeiden. Dabei wird versucht, die Intensitätswerte der Bildpunkte direkt als ursprüngliche Merkmale zu verwenden. Mit drei unterschiedlich strukturierte neuronale Nezte werden Objekte detektiert, klassifiziet, und Objektlage geschätzt (siehe Bild 6).

Bei Verwendung der Intensitätswerte ist die zu bearbeitende Datenmenge sehr groß, was das Training der neuronalen Netze unerwünscht verlängern und u. U. unrealistisch machen kann. Auf diesem Grund werden zunächst ein Merkmalsauswahl durchgefürt.

Nachdem die Merkmale gewonnen sind, wird zuerst durch ein Detektionsnetz bestimmt, ob Objekte vorhanden sind. Falls ja, werden dann diese Objekte durch ein dreischichtiges Erkennungsnetz klassifiziert, wobei ein Rprop Algorithmus verwendet wird. In weiterem Schritt wird Objektlage durch klassespezifisches Nezt geschätzt. Dabei werden zwei Translationsparameter (x– und y–Koordinate von Objektzentrum), ein interner Rotationsparameter (Rotationswinkel innerhalb der Bildebene) und ein externer Rotationsparameter (Rotationswinkel außer der Bildebene) durch entsprechende Netze berechnet.

Um bessere Ergebnisse bei der Lokalisation von ähnlichen Objekten zu erzielen, wurde speziell ein auf Kohonens SOFM basierendes Verfahren entwickelt. Dieses Verfahren gilt für 2D–Objekte und hat den Vorteil, gleichzeitig ein Objekt zu lokalisieren und zu erkennen. Im Vergleich zu dem holistischen Verfahren erhöht sich die Erkennungsrate um 6.4% auf 95.7% bei gleichen Stichproben mit der Bildgröße von 512 x 512. Außerdem steigert sich die Lokalisationspräzision um 2 Pixel auf 3 Pixel bei der Translation und um 1.5 Grad auf 3.5 Grad bei der Rotation.

In der letzten Zeit wurden weitere Experimente zum Testen der neuronalen Verfahren durchgeführt. Die auf Principal Component Network (PCN) basierende Merkmale wurde mit einer Reihe Wavelet–Merkmale verglichen. Die Ergebnisse wurden in [57] zusammengefasst. Außer-

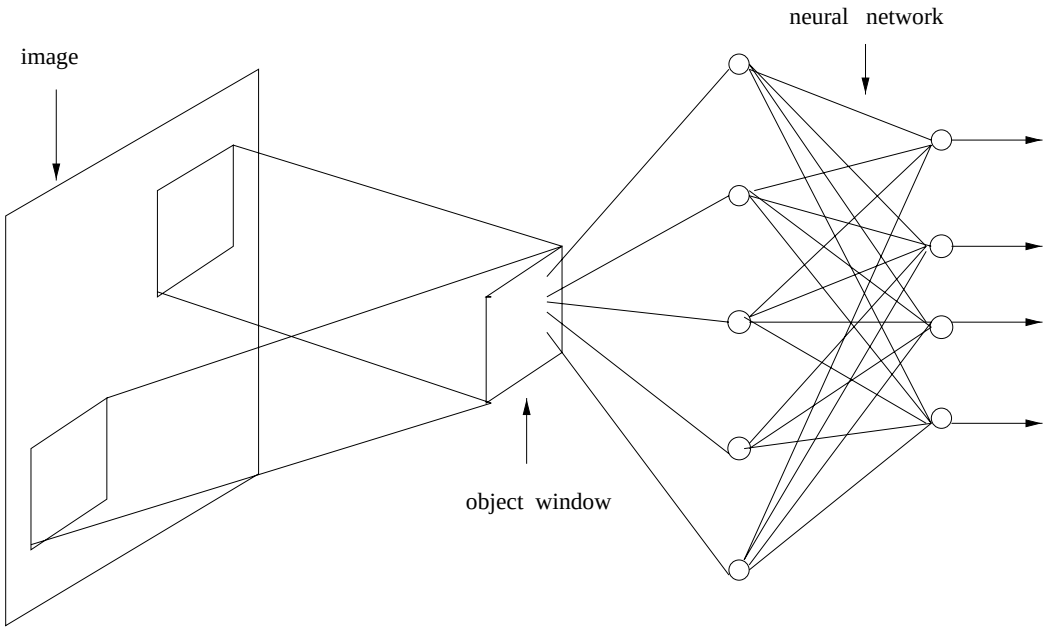


Bild 6: Neuronales Verfahren für die Objekterkennung

dem sind Untersuchungen anhand der neuen Stichprobe, die insgesammt 14 Objekte zusammenfasst, durchgeführt. Darüber hinaus wurde eine Strategie zur translationsinvarianten Objekterkennung entwickelt. Auf dieser Strategie aufbauend wurde ein Verfahren zur Mehr–Objekterkennung entwickelt und realisiert. Dabei wurden Experimente mit Verdeckungen und bei komplexem Hintergrund durchgeführt.

3.2 Bildanalyse für autonome Systeme

Die Aufgabe des Lehrstuhls für Mustererkennung im Projekt http://www5.informatik.uni-erlangen.de/HTML/Ger
(DienstleistungsRoboter in KOstengünstiger Leichtbauweise) ist die Verarbeitung von visueller Information zur Szenenexploration sowie Objektlokalisation und -erkennung. Die visuelle Information spielt bei der Bewältigung der vielfältigen Aufgaben eines Dienstleistungsroboters eine wichtige Rolle. Erst durch sie ist eine nahtlose Einbindung in die alltägliche Umgebung im Einsatzgebiet möglich.

Im Rahmen des Projekts wurde ein Explorations- und Erkennungssystem entwickelt, das in einem Trainingsschritt zuerst die Objektmodellgenerierung durchführt, indem anhand von Trainingsbilddaten objektspezifische Information extrahiert und in den Modellen gespeichert wird. Während der Bildanalyse wird diese Information genutzt um das Objekt in einer Explorationsphase zu lokalisieren und im Anschluss eine Klassifikation durchzuführen. Neben der reinen Klassifikation wird bei der Erkennung auch die für eine Manipulation benötigte Lageschätzung des Objekts durchgeführt (Bild 7).

In dem System finden erscheinungsbasierte Modelle, sogenannte Eigenraummodelle, Anwendung, die auf der direkten Modellierung von 2-D Ansichten, also den Grau- bzw. Farbwerten, basieren und keine Segmentierung hinsichtlich geometrischer Primitiven benötigen. Die Basis des Merkmalsraumes bilden die Eigenwerte der Kovarianzmatrix, die aus Trainingsbildern, die eine gute Repräsentation aller Objektansichten bilden, erstellt wird. Die Merkmalsvektoren der Trainingsbilder stellen im Merkmalsraum eine Teilmenge der Projektion aller möglichen Objektansichten dar. Dadurch Interpolation zwischen den Vektoren ist es möglich, Merkmale von im Training nicht beobachteten Ansichten zu approximieren. Man erhält so eine Repräsentation des Unterraums innerhalb des Eigenraums, der die Merkmalsvektoren aller mglichen Objektansichten darstellt. Der Merkmalsraum und der approximierte Ansichtenunterraum ergeben zusammen mit den Objektlageparametern, die während der Aufnahme protokolliert werden, das Objektmodell.

Die Klassifikation eines Testbildes erfolgt aufgrund einer Untermenge aller Bildpunkte und der Berechnung eines Merkmalsvektors anhand dieser Daten und dem Objektmerkmalsraum. Anhand der selektierten Bildpunkte und den zugehörigen Elementen der Basisvektoren des Merkmalsraumes kann ein überbestimmtes Gleichungssystem mit den Eigenraumkoeffizienten des Testbildes, welche den Merkmalsvektor bilden, aufgestellt werden. Über numerische Verfahren, aktuell findet die Singulärwertzerlegung Anwendung, können die Unbekannten bestimmt werden. Durch einen Nächster–Nachbar–Klassifikator erfolgt die Klassenzuweisung. Wissensbasierte Auswahlverfahren und eine iterative Verfeinerung für die Selektion ermöglichen es, Bildpunkte zu detektieren die dem zu erkennenden Objekt zugehörig sind und Punkte, die Hintergrund, Verdeckung oder Rauschen repräsentieren, nicht in die Berechnung mit einzubeziehen. Damit ist eine Klassifikation auch bei heterogenem Hintergrund, Teilobjektverdeckungen und Sensorrauschen möglich. Mittels der im Modell gespeicherten Objektlageparamter der Trainingsdaten kann zusätzlich eine Schätzung der Objektlage im Testbild durchgeführt werden.

Eine affine Transformationsschätzung ermöglicht eine Berechnung der Objektskalierung gegenüber der Trainingsgröße und eine Verfeinerung der Positionsschätzung. Hierfür wird aufgrund der Objektlageparameter ein Rückprojektion aus dem Merkmalsraum in den Bildraum durchgeführt. Die Transformationsparameter für Skalierung und Translation in x– und y–Richtung werden so bestimmt, dass die Rückprojektion mit dem Testbild zur Deckung gebracht wird.

Objektaufenthaltshypothesen für die Lokalisation von Objekten innerhalb von Übersichtaufnahmen liefert die Szenenexploration. Diese Hypothesen dienen als Grundlage für eine Verifikation mittels des Klassifikationsalgorithmus. Die Szenenexploration verwendet zusätzliche Objektinformation, beispielsweise in Form von Farbhistogrammen, und bietet die Möglichkeit einer indirekten Suche. Dabei erfolgt eine Suchraumeinschränkung für das Zielobjekt anhand eines leichter zu lokalisierenden, größeren Objekts, welches in einem räumlichen Zusammenhang zu dem gesuchten Objekt steht. Diese Zusammenhänge können durch eine semantisches Netz oder durch feste Regeln repräsentiert werden.

Die Verfahren zur visuellen 3D Objektverfolgung benötigen im Allgemeinen einen Abgleich zwischen einem Objektmodell und den Beobachtungen der Kamera. Die Wahl eines geeigneten Objektmodells ist daher entscheidend. Im Teilprojekt B2 des Sonderforschungsbereichs 603 ist ein allgemeines, probabilistisches Objektmodell erarbeitet worden, die so genannten *gekoppelten Strukturen*. Als Weiterführung dieser Arbeiten wurde in Kooperation mit dem Teilprojekt B3 des

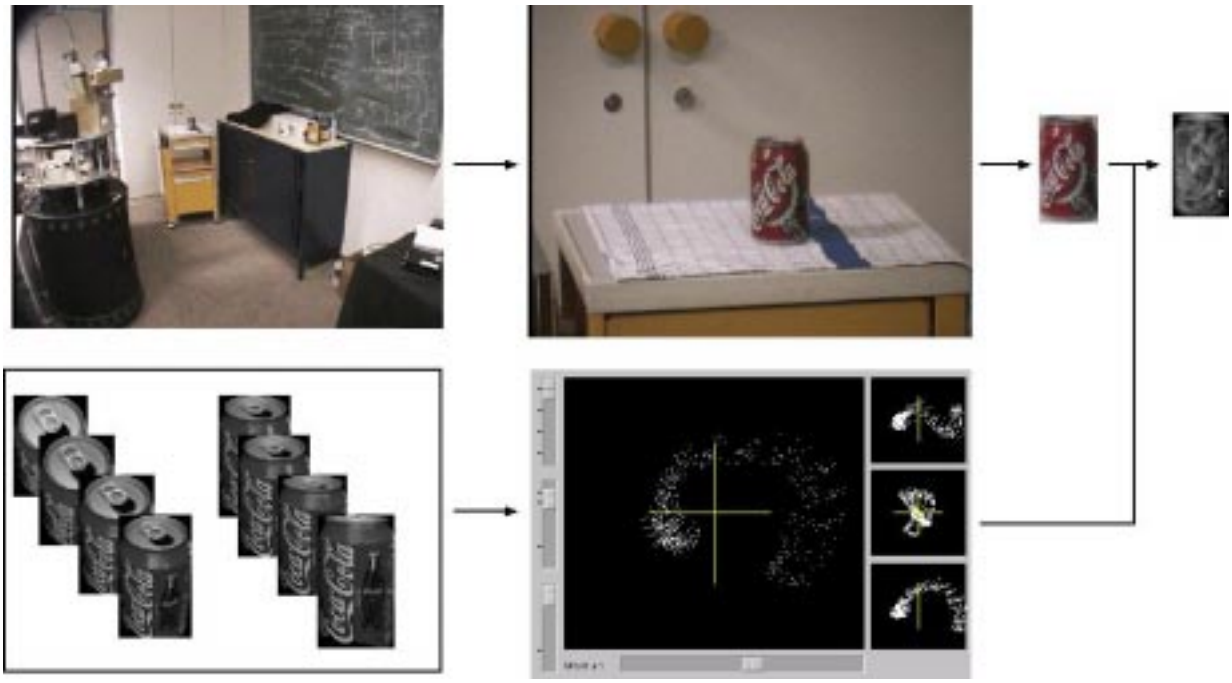


Bild 7: Objektmodellgenerierung, Szenenexploration und Erkenunnungsprozess: Modellgenerierung aus Trainingsdaten und resultierende Eigenraumdarstellung (unten); Tischsuche im Übersichtsbild, Objektlokalisation auf dem Tisch, extrahiertes Objekt, Rekonstruktion des erkannten Objekts (oben).

SFB 603 die Eignung des entworfenen Modells am Beispiel der Lokalisierung von Gesichtern in DCT kodierten Bildern demonstriert [58].

Eine weiteres Forschungsziel ist die Erarbeitung der theoretischen Grundlagen für die optimale Auswahl von Kameraparametern für verschiedene Probleme aus dem Bereich des Rechnersehens. Die Kameraparameter umfassen dabei sowohl solche mit denen der Abbildungsprozess beeinflusst wird, z. B. die Brennweite, als auch Parameter für die Position und Ausrichtung der Kameras, z. B. der Schwenk-Winkel. Ein besonderes Augenmerk liegt dabei auf der Problematik der Sensordatenfusion, die durch die Verwendung mehrerer Kameras aufgeworfen wird. Als Beispiel hierzu ist die multiokulare 3D Objektverfolgung zu nennen, bei der versucht wird, den unbekannten Zustand des verfolgten Objekts, meist seine Position im Raum und einige kinematische Größen, aus den Beobachtungen der einzelnen Kameras zu schätzen.

Ein neueres Verfahren zur Objektverfolgung ist der CONDENSATION Algorithmus. Dieses aus dem Bereich der Partikel Filter stammende Verfahren erlaubt, im Gegensatz zum Kalman Filter, die dynamische Schätzung der *multimodalen* a posteriori Wahrscheinlichkeit für den Objektzustand, gegeben die Folge der bisherigen Beobachtungen. Dazu wird die a posteriori Wahrscheinlichkeitsdichte über den Zustand nicht funktional beschrieben, sondern durch eine Menge von Teilchen repräsentiert, die über die Zeit hinweg propagiert und entsprechend den Beobachtungen gewichtet werden. Für die Verfolgung mit mehreren Kameras konnte dieses Verfahren dahinge-

hend erweitert werden, dass nicht nur die Beobachtungen von einer Kamera, sondern auch von mehreren Kameras zur Gewichtung herangezogen werden können. Das entwickelte Sensordatenfusionsschema fügt sich dabei nahtlos in den probabilistischen Rahmen des Verfahren ein (vgl. Bild 8).

Mit der beschriebenen Erweiterung des CONDENSATION Algorithmus war es möglich, erste praktische Untersuchungen zur Brennweitenwahl bei der multiokularen 3D Objektverfolgung durchzuführen. Ziel dabei war die Klärung der Frage nach dem Einfluss der Brennweite auf die Güte der Objektverfolgung. Mit Hilfe zweier Zoom-Kameras wurde dazu ein Objekt über einen längeren Zeitraum verfolgt. Die Brennweiten der einzelnen Kameras wurden dabei definiert von klein (Weitwinkel) bis groß (Zoom) variiert. Es zeigte sich, dass bei gegebenem Objekt und Verfolgungsverfahren, die Güte der Verfolgung in Abhängigkeit zur eingestellten Brennweitenkombinationen steht.

Die automatische Adaption der Brennweiten während der Verfolgung, abhängig von der gerade aktuellen Verfolgungsgüte, ist Gegenstand weiterer Forschungen. Erste Ansätze, basierend auf Konzepten der Informationstheorie, wurden dazu schon diskutiert.

Im Teilprojekt B2 steht die, zusammen mit dem Projekt DIROKOL finanzierte, autonome mobile Plattform MOBSY zur Verfügung. Auf dieser Plattform sollen die entwickelten Verfahren auch im Bereich der Service Robotik zum Einsatz kommen. Eine essentielle Fähigkeit eines mobilen Service Roboters ist die selbstständige Navigation im Aufgabenbereich. Für diesen Zweck wurde für MOBSY aufbauend auf so genannten *Belegtheitsgittern* ein Modul zur Erkundung und Kartierung der Umgebung bei der Verwendung von Ultraschallsensoren entwickelt. Mit Hilfe der daraus resultierenden „Landkarten“ ist MOBSY in der Lage, beliebige Punkte, z. B. Büros, selbstständig anzufahren, wie es beispielsweise für Hol- und Bringdienste charakteristisch ist. Für die dazu notwendige Selbstlokalisierung wurde auch hier der oben beschriebene CONDENSATION Algorithmus verwendet.

3.3 Bildbasierte Modellierung und erweiterte Realität

Am Lehrstuhl für Mustererkennung werden in Zusammenarbeit mit dem Lehrstuhl für Graphische Datenverarbeitung im Rahmen des Teilprojektes C2 des Sonderforschungsbereichs 603 Verfahren zur automatischen Analyse von Bildströmen zur Generierung von Lichtfeldern und zu ihrer Visualisierung entwickelt. Im Vergangenen Jahr wurden im Bereich der Analyse hauptsächlich zwei Bereiche vorangetrieben.

Zum einen war dies die Entwicklung geeigneter Werkzeuge, um die aus den Bilddaten extrahierte Information für die Visualisierung in geeigneter Form verfügbar zu machen Neben Kameraparametern wurden auch Bilder geliefert, welche Tiefenwerte und deren Zuverlässigkeit codieren. Das dadurch ermöglichte Gesamtsystem ist in der Lage, zwischen unterschiedlichen Lichtfeld–Repräsentationen zu wählen und je nach Verfügbarkeit bei der Echtzeitvisualisierung unterschiedlich skalierbare Datenmengen zu verwenden [52].

Zum anderen konnten Verbesserungen in der Kalibrierung der monokularen Bildströme erreicht werden. Die automatisch extrahierten Punktkorrespondenzen werden auf ihre Konsistenz hin untersucht und so auftretende Ausreißer erkannt und eliminiert. Es zeigte sich, dass durch Kombination verschiedener Verfahren die Robustheit der Selbstkalibrierung erheblich gestei-

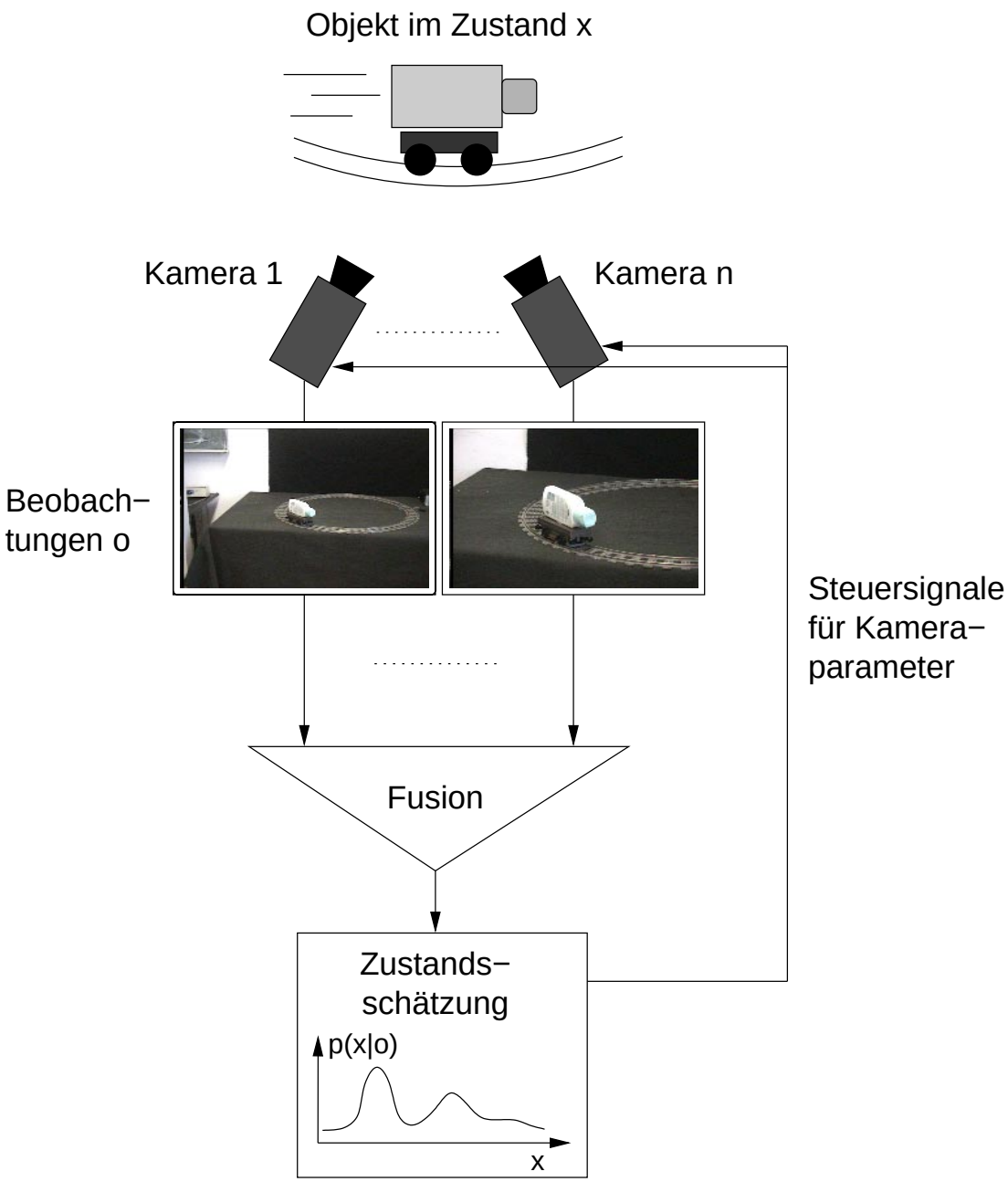


Bild 8: Systemüberblick: Objektverfolgung mit mehreren Kameras

gert werden kann. Hierbei wird für eine erste Approximation das vereinfachende Modell der schwach–perspektivischen Projektion angenommen und durch Anwendung linearer und nicht–linearer Optimierungsverfahren eine stabile Rekonstruktion für das genauere perspektivische Kameramodell berechnet. Ein Beispiel für eine kalibrierte Sequenz so wie ihre Visualisierung ist in Bild 3.3 zu sehen.

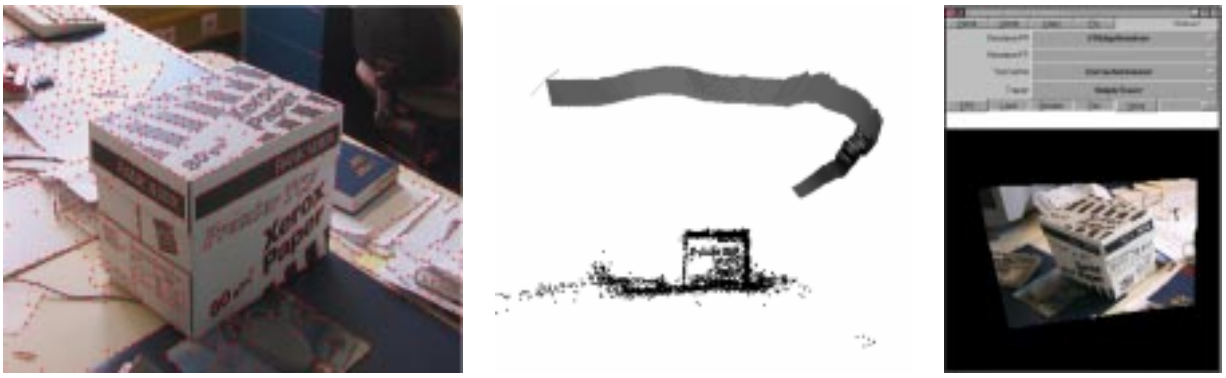


Bild 9: Lichtfeldakquisition, Rekonstruktion und Visualisierung. Links: eines der 180 Eingabebilder mit extrahierten Punktmerkmalen. Mitte: rekonstruierte Kamerapositionen. Rechts: interaktive Visualisierung.

Als Anwendung des Lichtfeldes wurde die Selbstlokalisation eines autonomen mobilen Systems weiter untersucht. Die auf dem Roboter montierte Kamera nimmt ein Bild der Umgebung auf, das mit einem virtuellen, mittels des Lichtfeldes generierten Bildes verglichen wird Die hierfür notwendigen Hypothesen für die aktuelle Position und Orientierung des mobilen Systems werden mittels probabilistischer Monte–Carlo–Methoden generiert [25].

Seit Mitte des Jahres wird am Lehrstuhl für Mustererkennung im Bereich der erweiterten Realität (Augmented Reality) geforscht. Unter diesem Begriff versteht man die Ergänzung der Umwelt, wie sie von einem Beobachter wahrgenommen wird, durch vom Rechner generierte Daten oder Objekte. Ermöglicht wird dies durch die Verwendung eines sog. Head-Mounted Displays (HMD), durch das der Benutzer mit Hilfe von Kameras sowohl seine Umgebung als auch die virtuellen Objekte wahrnehmen kann.

In einer Studienarbeit wurde bisher die Bestimmung der Kopfposition eines Benutzers betrachtet. Dies geschieht durch Kalibrierung der Position und Orientierung einer Infrarot-Kamera, die auf einem Helm montiert ist. So wird die Betrachtung eines mit OpenGL gerenderten Objekts von mehreren Seiten alleine durch die Bewegung des Kopfes möglich.

Eine Diplomarbeit befasste sich mit dem Thema der Bestimmung von Position und Orientierung eines (realen) Würfels mit bekannter Geometrie und Farbe. Im weiteren Verlauf soll im HMD an Stelle des Würfels ein rechnergeneriertes Objekt erscheinen, dessen Orientierung durch Bewegung des Würfels verändert werden kann. Ein Beispiel hierfür zeigt Bild 10.

Wichtige Punkte für die Zukunft sind die Echtzeitfähigkeit sowie die realistische Darstellung der virtuellen Objekte. Berücksichtigt werden soll insbesondere die konsistente Darstellung von

- Schattenwurf,
- Beleuchtung,
- Verdeckungen des virtuellen Objekts duch reale Gegenstände.

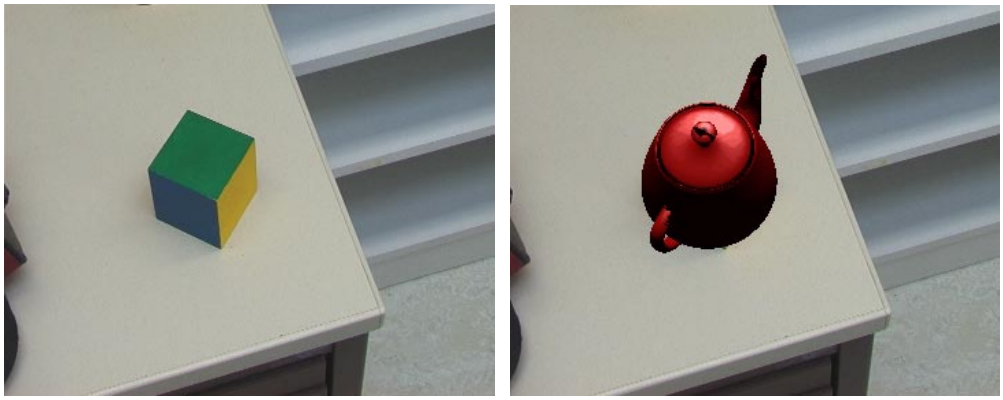


Bild 10: Ersetzen eines realen Würfels durch eine rechnergenerierte Teekanne in. gleicher Position und Orientierung

3.4 Medizinische Anwendungen

An der Augenklinik der Universität–Erlangen–Nürnberg werden in Zusammenarbeit mit dem Lehrstuhl für Mustererkennung und Institut für Medizinische Biometrie und Epidemologie (IMBE) im Rahmen des SFB539 Teilprojekts A4 Verfahren ab 01.07.2000 zum automatischen Glaukom-Screening entwickelt. Ziel dieses Projekts ist die Glaukompatienten vor Beginn der subjektiven Sehstörungen zu identizifieren und damit eine frühzeitige, ärztliche Therapie zuführen zu können. Dazu soll eine automatische 3D-Analyse der Papillentomographibilder (Heidelberg Retina Tomograph, HRT) mit Hilfe von automatischen Markierung des Papillenrandes des HRT-Bildes, von mathematischen Beschreibung der Papillenfläche und von automatischen morphologischen Segmentierung und Analyse der juxtapapillären Gefäße duchgeführt werden. Anschließend wird untersucht wie und ob sich die Position des Skleralrings aus den dreidimensionalen Schnittbildern ableiten lässt. Dies stellt die Grundlage dar für eine automatische Klassifikation von Retina-Tomograph-Bildern hinsichtlich der Verdachtsdiagnose. Bislang wurden die Merkmale der Papillenflächenaproximation über ca. 400 Bilder für die Klassifikation vorbereitet, die durch IMBE durchgefüchrt wird.

Ein typisches Bild dieser Situation ist in Bild 11 gezeigt.

Die automatische Diagnose von Gesichtslähmungen, die im Rahmen des Teilprojekt B3 des Sonderforschungsbereichs 603 (SFB 603) untersucht wird, wurde im vergangenem Jahr an zwei zentralen Systemmodulen entscheidend verändert. Diese beiden Punkte sollen im folgenden erläutert werden.

Zur Lokalisation und Verfolgung von Gesichtern und Gesichtsmerkmalen werden nun *Support Vector Machines* (SVM) zusammen mit einem Kalman-Filter eingesetzt. Die Verfolgung wird mit einer Lokalisation des Patientengesichts im Bild gestartet. In einer niedrigen Auflösungsstufen des Originalbilds (Bildgröße Originalbild 384×288 Bildpunkte, niedrige Bildauflösung 24×16 Bildpunkte) zunächst Hypothesen für ein Gesicht im Bild bestimmt. Die Hypothesen werden dann bei höherer Bildauflösung (96×72 Bildpunkte) nochmals genauer untersucht und die tatsächliche Gesichtsposition bestimmt. Die gefundene Position wird jeweis einem Kalman-

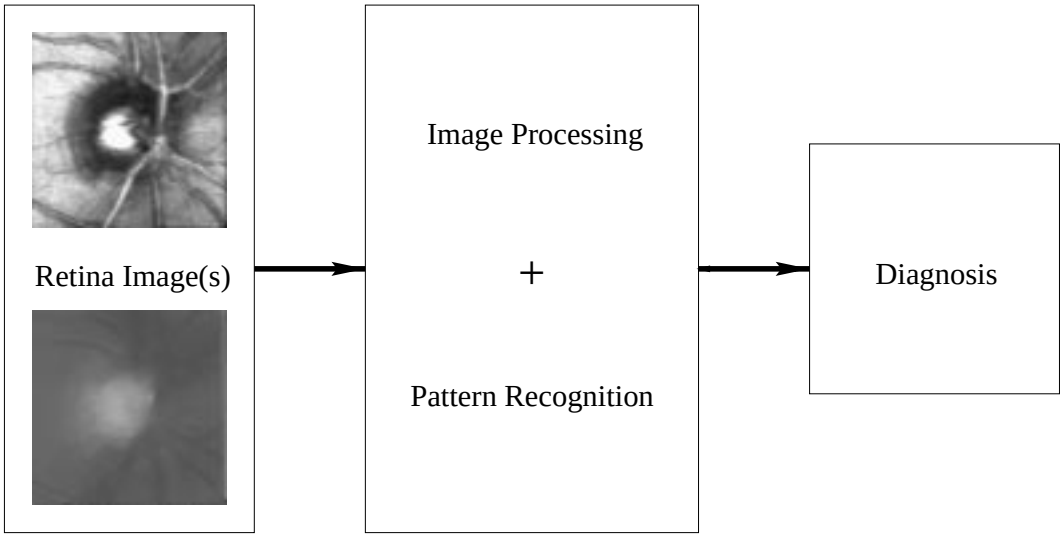


Bild 11: Augenhintergrund und Diagnose

Bild 12: a) Lokalisation und Verfolgung von Gesichtern und Gesichtsmerkmalen; b) Differenzbild während der Ausführung einer symmetrischen Gesichtsbewegung; c) Differenzbild während der Ausführung einer asymmetrischen Gesichtsbewegung.

Filter (dynamisches Modell: konstante Geschwindigkeiten) als Beobachtung übergeben. Nach einer Initialisierungsphase wird dann vom Kalman-Filter eine Voraussage bezüglich der nächsten Position des Patientengesichts getroffen, die den möglichen Hypothesenraum einschränkt.

Ist die Gesichtsposition bestimmt, lokalisiert eine weitere SVM das Augenpaar. Die Position des Gesichts (der Nasenspitze) und des Augenzwischenpunktes liefert anschließen die Position des Mundes. In Bild12a ist das Ergebnis einer Gesichtsverfolgung dargestellt. Das äußere Quadrat umschließt das gefundene Gesicht (Mittelpunkt Nasenspitze). Das innere gelbe Quadrat ist der Bereich der vom Kalman-Filter als möglichen Hypothesenraum vorgibt.

Der zweite Punkt, der starke Änderungen erfuhr, ist die Extraktion und die Nutzung der Information aus den aufgenommenen Patientenbildern. Waren es zunächst noch statische Gesichtsbilder, die durch rotierende Keilfilter analysiert wurden, so sind es nun Bildsequenzen des Patientengesichts. Zwischen aufeinanderfolgenden Bildern der aufgenommenen Sequenzen werden Differenzbilder bestimmt. In Bild12b und Bild12c sind Beispiele für diese Differenzbilder

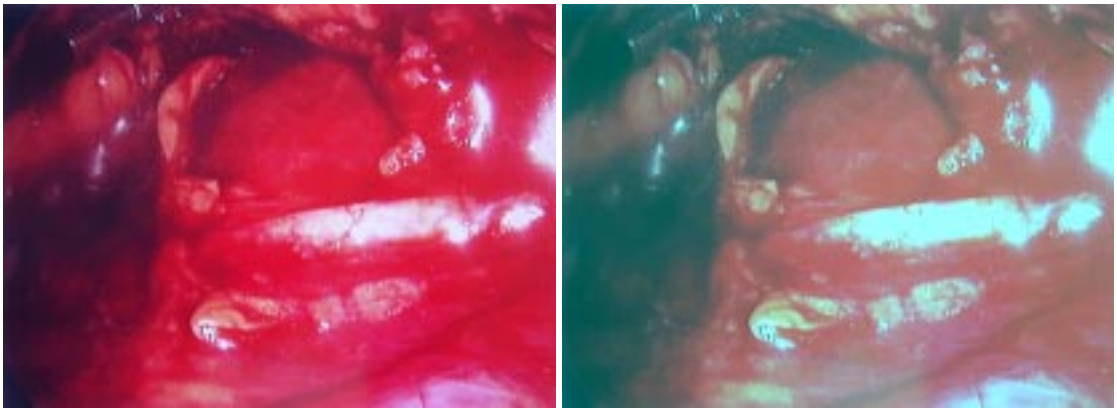


Bild 13: Endoskopisches Bild einer Galle: links Original, rechts Bildverbesserung durch Farbrotation

von Gesichtern bei symmetrischer (b) und asymmetrischen (c) Bewegungen. Durch Integration der absoluten Differenzen und durch Vergleich zwischen linker und rechter Gesichtshälfte kann ein Merkmalsvektor aus den Bildsequenzen extrahiert werden, der zur Graduierung der Gesichtslähmung herangezogen werden kann.

Im Teilprojekt B6 „Rechnergestütze Endoskopie des Bauchraums“, des Sonderforschungsbereichs 603 (SFB 603) sollen ab dem Jahr 2001 in Zusammenarbeit mit der Chirurgischen Universitätsklinik die grundlegenden Arbeiten zur Einführung einer Rechnerunterstützung in der mikroinvasiven Chrirugie durchgeführt werden. Neben dem Ziel der Bildverbesserung in Echtzeit soll der Lichtfeldansatz aus Teilprojekt C2 weiterentwickelt und auf die endoskopischen Bildsequenzen angewandt werden. Zu diesem Zweck soll auch ein Sensor zur Positionsbestimmung der Kamera (des Endoskops) aufgebaut werden. Die ersten Arbeiten zu diesem Problem wurden aufgenommen. Es gelang mit einfachen Verfahren, eine von den Ärzten als bedeutsam eingeschätzte Bildverbesserung durchzuführen. Ein typisches Bild dieser Situation ist in Bild 13[1] gezeigt.

4 Statistische Modellierung von Daten

(R. Deventer, U. Ohler,)

Im Rahmen des Sonderforschungsbereiches “Robuste, verkürzte Prozessketten für flächige Leichtbauteile„ (SFB 3
wurden die Arbeiten zur sensor- und modellgestützten Optimierung von Prozessketten weitergeführt. Zusammen mit 6 Lehrstühlen des Instituts für Fertigungstechnik werden neue Herstellungsverfahren für Leichtbauteile mit dem Ziel untersucht, die Prozessketten so auszulegen, dass die Anfälligkeit gegenüber äußeren Störungen und die Anzahl der notwendigen Einzelschritte minimiert wird. Hierzu wird im Rahmen des Teilprojektes C1 am Lehrstuhl für Mustererken-

[1] Mit Dank an die Chirurgische Universitätsklinik (Direktor Prof. Dr. W. Hohenberger)

nung eine stochastische Modellierung von Prozessketten mit dem Formalismus der Bayesnetze verfolgt.

Anschaulich ist ein Bayes–Netz ein gerichteter azyklischer Graph, dessen Knoten die physikalischen Mess- und Einstellgrößen der Prozesskette, sowie daraus abgeleitete Qualitätsbewertungen als Zufallsvariablen modellieren und dessen Kanten die Abhängigkeitsstruktur der involvierten Größen repräsentieren. Hierbei bieten Bayesnetze eine Reihe von Vorteilen, insbesondere sind Trainingsalgorithmen zur Adaption der Struktur und Parameter vorhanden, die auch mit unvollständigen Daten arbeiten. Außerdem repräsentiert das Bayesnetz eine Verbundverteilung, d. h. es wird kein Unterschied gemacht zwischen Ein- und Ausgabedaten. Der Rückschluß von Ausgabedaten auf die Eingabedaten ist daher möglich.

Im Jahr 2000 stand die Entwicklung einer Regelung im Vordergrund unserer Arbeit. Hierfür muss die Echtzeitfähigkeit eines Bayesnetzes unter Beweis gestellt werden. Dies geschieht durch Aufbau einer Regelung des Kraftflusses beim Innenhochdruckumformen. Hierbei soll die Kraft, die auf das umzuformende Blech wirkt, mit dem Ziel einer konstanten Verteilung geregelt werden. Hierfür wurde der Zusammenhang zwischen den Eingangskräften der Pressenzylinder und den Kräften auf das umzuformende Blech modelliert. Das Bayesnetz wurde dabei mit 69 Datensätzen trainiert und die Güte der Modellierung wurde durch Vorhersage der Kraft an 4 Punkten mit jeweils 14 verschiedenen Eingabekräften ermittelt. Hierbei ergab sich ein mittlere Fehler von 5.66%. Die Anbindung der Presse an das Bayesnetz ist über AD/DA-Wandler geplant. Tests in diesem Bereich stehen noch aus.

Ein weiterer Bereich der Forschung war der prinzipielle Aufbau eines modellbasierten Reglers. Anfängliche Überlegungen beschränkten sich auf lineare, zeitinvariante dynamische Systeme. Diese können durch Kalman-Filter beschrieben werden. Da diese den dynamischen Bayesnetzen sehr ähnlich sind, können durch Analogieschluss Struktur und Parameter des dynamischen Bayesnetzes direkt aus der analytischen Beschreibung des dynamischen Systems abgeleitet werden [14]. Die dabei verwendete Architektur des Bayesnetzes zeigt Bild 14. In diesem Bayesnetz

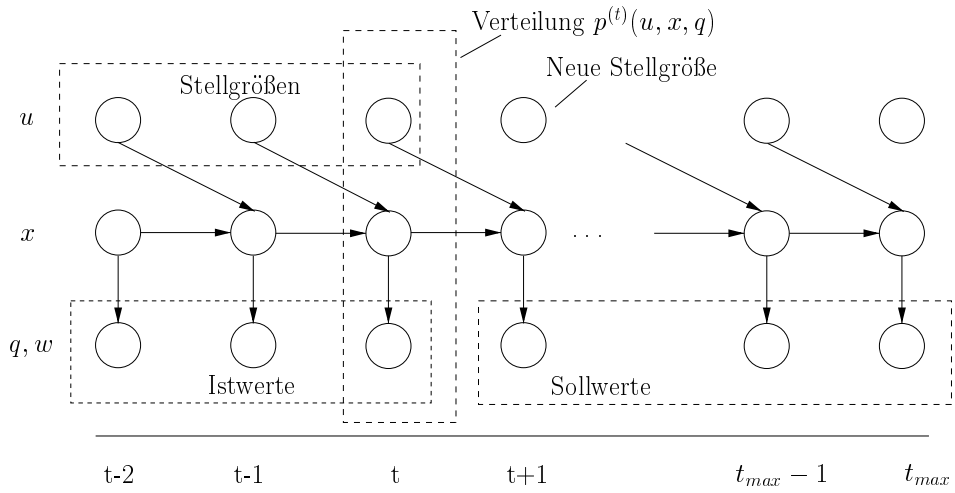


Bild 14: Aufbei eines modellbasierten Reglers

wird die Eingabe, der Zustand und die Ausgabe des Systems zu verschiedenen Zeitpunkten in der

Vergangenheit und der Zukunft modelliert. Gibt man für alle Zeitpunkte der Vergangenheit Ein- und Ausgabe u bzw. q und für Ausgabeknoten in der Zukunft den Sollwert w als Evidenz an, so ist es möglich via Marginalisierung auf geeignete Eingabewerte zu schliessen. Der Zustand x kann dabei nicht beobachtet werden. Als Test wurde das Führungsverhalten für dynamische Systeme zweiter Ordnung herangezogen. Hierbei zeigte sich, dass Systeme mit hoher Dämpfung problemlos geregelt werden können. Bei Systemen mit einer geringen Dämpfung, die ein schwingendes Verhalten zeigen, tritt bei diesem Ansatz auch ein schwingendes Eingangssignal auf. Falls dieses gedämpft wird, können auch Systeme mit geringer Dämpfung geregelt werden. In der Zukunft folgen noch Tests zur Systemidentifikation. Außerdem soll dieses System durch die Verwendung hybrider Bayesnetze für die Regelung nichtlinearer Systeme erweitert werden.

Weitere Anwendungen der statistischen Modellierung, die sich aus den Erfahrungen der Sprachanalyse ergaben, wurden in dem Projekt zur Analyse von Genomdaten eingebracht.

Im Bioinformatik-Projekt „Statistische Modellierung, Lokalisierung und Analyse regulatorischer DNA–Sequenzen“, das vom Boehringer Ingelheim Fonds gefördert wird, wurde das im Jahr zuvor entwickelte System zur Annotierung von Promotoren in DNA-Sequenzen McPromoter erweitert. Promotoren sind den proteincodierenden Abschnitten der DNA vorgelagert und weisen eine komplexe, oft sehr variable Struktur auf. Sie sind wesentlich an der differenzierten Regulation von Genen beteiligt.

Promotorregionen werden nun als stochastisches Segmentmodell mit sechs Zuständen repräsentiert [38]. Als Teilmodelle dienen dabei die bereits zuvor zur Promotorerkennung eingesetzten interpolierten Markovketten. Ein neuronales Netz verarbeitet die Ausgabe dieses Modells zusammen mit der Bewertung eines Hintergrundmodells und trifft eine Entscheidung, ob eine vorliegende Teilsequenz einem Promotor entspricht oder nicht (siehe Bild 15).

Das System wird zur Zeit zur Annotierung des gesamten Drosophila-Genomes eingesetzt (Kooperation mit dem Berkeley Drosophila Genome Project). Bei der umfangreichen Auswertung des internationalen „Genome Annotation Assessment Project“ [47, 37] zeigte sich, dass das Systems zur automatischen Annotierung der Genome komplexer Organismen geeignet ist. McPromoter wurde im Rahmen des Innovationswettbewerbs Bioinformed mit dem Bioinformatik-Nachwuchspreis der Jury ausgezeichnet.

5 Sprachverstehen

Leitung: E. Nöth
(H. Adelhardt, A. Batliner, J. Buckow, W. Fentze, C. Frank, R. Hertlein, R. Huber, R. Shi, G. Stemmer, V. Warnke)

Die inhaltlichen Schwerpunkte der Forschungsaktivitäten zur Sprachverarbeitung bilden das maschinelle Erkennen und Verstehen gesprochener Äußerungen sowie Fragestellungen des multimodalen Mensch–Maschine–Dialogs. Die Arbeiten im Berichtsjahr konzentrierten sich auf die Weiterentwicklung prototypischer Sprachdialogsysteme. Neben den zwei „tradionellen“ Anwendungsbereichen der Sprachverarbeitungsforschung am Lehrstuhl (Kinoauskunft mit dem System *Fränki* und multilinguale Terminabsprache mit dem System *Verbmobil*) wurde mit Arbeiten an der Entwicklung eines multimodalen Dialogsystems (*SmartKom*) und Arbeiten zur

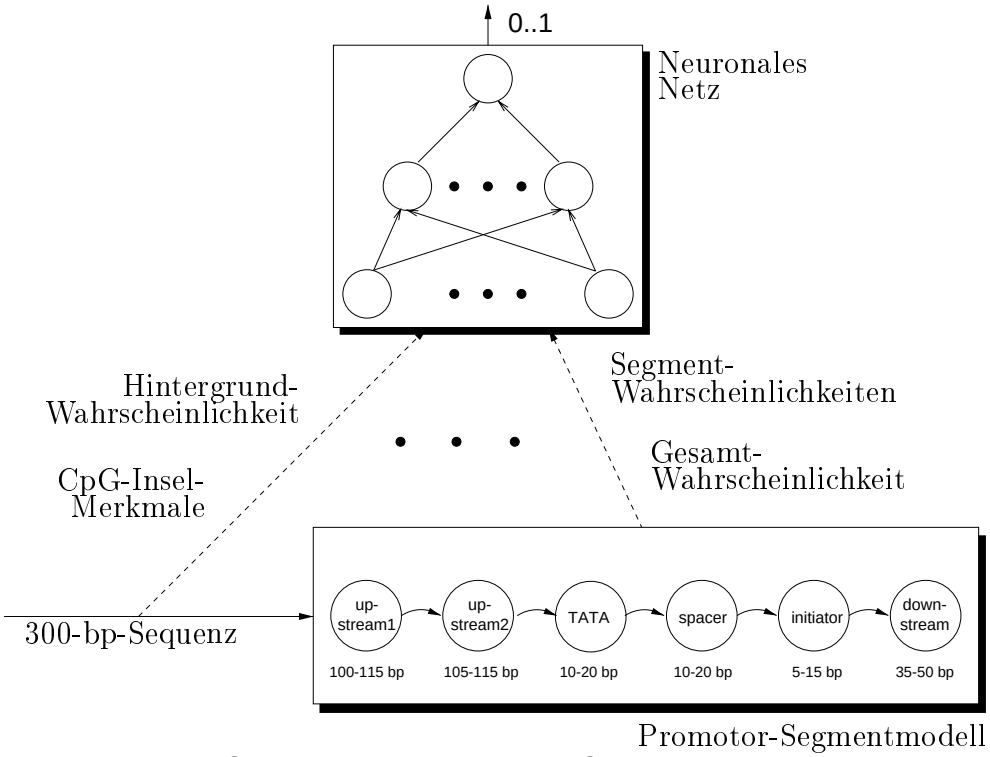


Bild 15: Der Klassifikator im System zur Identifizierung eukaryontischer Promotoren in DNA-Sequenzen. Ein Fenster der Größe 300 dient als Eingabe und wird von einem Promotoren- und einem Hintergrundmodell bewertet. Deren Bewertungen werden schließlich mit einem künstlichen neuronalen Netz kombiniert.

Sprecheridentifikation/–verifikation begonnen.

Der für die Zugauskunftsdomäne entwickelte Laborprototyp eines sprecherunabhängigen Systems zur Mensch–Maschine–Kommunikation wurde 1999 in wesentlichen Teilen der Verstehensphase (Interpretation und Dialogführung) neu implementiert. Um Erfahrungen mit der schnellen Änderung von Anwendungsdomänen zu sammeln, wurde das Dialogsystem nach der Neuimplementierung auf die neue Domäne „Kinoauskunft“ portiert. *Fränki*, das *Fränkische Kinosystem*, ist unter 09131/16287 erreichbar und kann Auskunft geben über das Kinoprogramm im mittelfränkischen Raum. Die schnelle Portierung auf neue Anwendungsbereiche sowie die eleganten Möglichkeiten, einen freien Mensch–Maschine–Dialog zu entwerfen, führten dazu, daß sich seit 1.3.2000 mit der Sympalog AG eine Ausgründung der Sprachverarbeitungsgruppe mit der kommerziellen Vermarktung der am Lehrstuhl entwickelten Technologie beschäftigt. Auf der CeBit 2000 wurde auf einem Gemeinschaftsstand des Lehrstuhls und von Sympalog ein Kinoreservierungssystem vorgestellt.

Das zweite Anwendungsfeld (multilinguale Terminabsprache) ist durch die Einbindung des Lehrstuhls (4 Mitarbeiter/innen) in das BMBF-geförderte *Verbmobil*–Vorhaben bedingt, das im Berichtsjahr erfolgreich abgeschlossen wurde. Projektpartner waren ca. 20 Arbeitsgruppen aus

mehreren Universitäten, Großforschungseinrichtungen und Firmen. Im *Verbmobil*-Vorhaben war ein portables Übersetzungsgerät zu entwickeln, welches auf Konferenzen mit Teilnehmern unterschiedlicher Muttersprachen die Dolmetschfunktion übernimmt. Zur Zeit kann Deutsch, Englisch und Japanisch verarbeitet werden. Das Thema der multilingualen Verhandlung bewegt sich im Rahmen geschäftlicher Terminabsprachen.

Im Nachfolgeprojekt des *Verbmobil*-Vorhabens, dem ebenfalls vom BMBF geförderten *SmartKom*-Vorhaben, werden Konzepte für die Entwicklung völlig neuartiger Formen der Mensch-Technik-Interaktion erprobt. Diese Konzepte werden die bestehenden Hemmschwellen von Computerlaien bei der Nutzung der Informationstechnologie abbauen und so einen Beitrag zur Benutzerfreundlichkeit und Benutzerzentrierung der Technik in der Wissensgesellschaft liefern. Das Ziel von *SmartKom* ist die Erforschung und Entwicklung einer selbsterklärenden, benutzeradaptiven Schnittstelle für die Interaktion von Mensch und Technik im Dialog. Am *SmartKom*–Projekt sind insgesamt 12 Arbeitsgruppen aus mehreren Universitäten, Großforschungseinrichtungen und Firmen beteiligt. Der Lehrstuhl bearbeitet die Bereiche Prosodie–, Mimik– und Gestik–Interpretation.

5.1 Das Dialogsystem FränKi

Im Bereich der Spracherkennung wird am Lehrstuhl traditionell an der Verarbeitung von spontaner Sprache geforscht, wie sie im Mensch–Maschine–Dialog auftritt. Dabei wurde im Berichtsjahr vor allem an der Verbesserung des Dialogsystems *Fränki* gearbeitet. Der modulare Aufbau von *Fränki* ist in Bild 5.1 zu sehen. Das Dialogsystem *Fränki* gibt AnruferInnen Informationen über das aktuelle Kinoprogramm im mittel*Fränki* schen Raum. Da das Projekt *Fränki* noch sehr neu ist und deshalb noch nicht genügend Anrufe aufgezeichnet werden konnten, wurden alle im folgenden beschriebenen Experimente auf den Daten der Zugauskunft-Stichprobe, die seit 1993 gesammelt wurde, durchgeführt. Die Untersuchungen konzentrieren sich darauf, das Spracherkennungsmodul besser in das gesamte Dialogsystem zu integrieren. Beispielsweise soll das Spracherkennungsmodul Informationen besser nutzen, die mit dem Dialogverlauf zusammenhängen, und vom Modul zur Dialogsteuerung zur Verfügung gestellt werden können. Dazu wurden statistische Sprachmodelle entwickelt, die vom aktuellen Zustand des Dialogs abhängen und dennoch robust geschätzt werden können [51]. In den Experimenten bewirken sie eine signifikante Verbesserung des Erkennerverhaltens. In Zukunft soll dieser Ansatz hin zu einer dynamischen Anpassung der Sprachmodelle an die aktuelle Dialogsituation verbessert werden. Dabei ist bei einem System zur Kinoauskunft beispielsweise auch zu berücksichtigen, dass bestimmte, erfolgreiche Filme von AnruferInnen häufiger nachgefragt werden als andere. Eine Diplomarbeit beschäftigt sich mit der Frage, ob eine solche dynamische Anpassung auch bei einem anderen Erkennerparameter, dem linguistischen Gewicht, erfolgreich sein kann.

Weiterhin muss das Dialogsystem jede Woche an das aktuelle Kinoprogramm angepasst werden. Das erfordert zur Zeit die manuelle Verschriftung der Filmtitel in Lautfolgen. Eine Diplomarbeit hat zum Thema, Möglichkeiten zu untersuchen, wie die Umsetzung der Orthografie in die Lautschrift automatisiert werden kann. Das schließt auch die Verwendung von Aussprachebeispielen der neuen Wörter mit ein. Eine automatische Verschriftung neuer Wörter ist eine wichtige Voraussetzung für viele denkbare Erweiterungen des *Fränki* -Auskunftssystems, wie z.B. die Integration von Wissen aus Datenbanken im Internet.

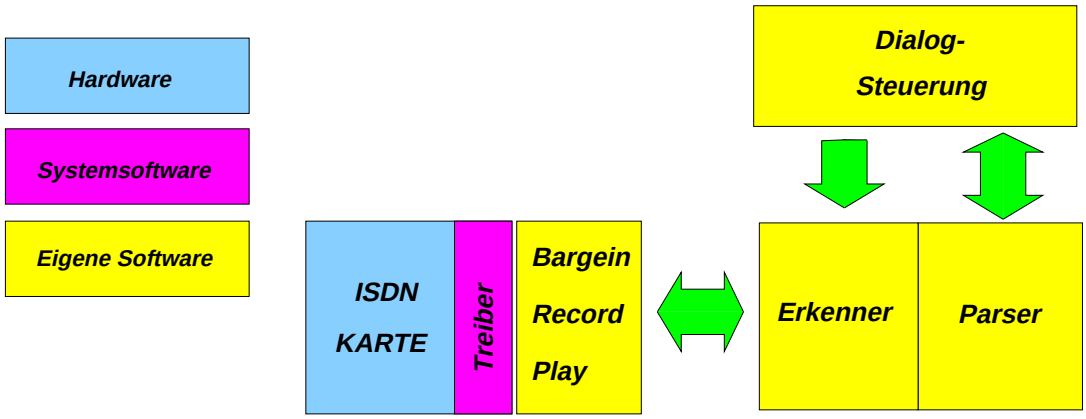


Bild 16: Modularer Aufbau des Dialogsystems *Fränki* .

5.2 Das VERBMOBIL–Projekt

Während die Forschungsarbeit in den vorausgegangenen drei Jahren vor allem dadurch bestimmt war, die Prosodie-Komponente des *Verbmobil*-Systems an die Anforderungen in *Verbmobil* Phase 2 anzupassen, lag der Schwerpunkt im letzten Jahr wieder auf der Verbesserung der Klassifikationsergebnisse bei der Klassifikation prosodischer Ereignisse.

In den ersten drei Jahren des *Verbmobil*-II-Projekts wurde eine Architektur für ein multilinguales Prosodiemodul geschaffen. Synergien durch die Verwendung eines multilingualen Moduls statt dreier monolingualer Module wurden erfolgreich genutzt. Durch einen neuen (wortbasierten) Merkmalsatz, der wesentlich effizienter als der alte (silbenbasierte) berechnet werden kann, wurde die Erfüllbarkeit der Echtzeitanforderungen auch im multilingualen *Verbmobil*-II-Szenario sichergestellt. Durch die wortbasierten Merkmale konnte nicht nur die Berechnung der Merkmale stark beschleunigt und der Speicherbedarf erheblich reduziert werden, die Klassifikation prosodischer Ereignisse wurde ausserdem in vielen Fällen deutlich verbessert. Im letzten Jahr des *Verbmobil*-II-Projekts war unser Ziel nun, durch Einbeziehung weiterer Wissensquellen die Klassifikationsgüte noch zu steigern.

In der Vergangenheit wurde aus dem Sprachsignal auf Basis der von einem Worterkenner gelieferten Worthypothesen ein hoch-dimensionaler Merkmalsatz berechnet, der jedoch rein akustisch-prosodische Merkmale enthielt. Das lexikalische Wissen (Wortidentitäten) und das syntaktische Wissen (Abfolge der Worte) wurde erst in einem späteren Schritt nach der Klassifikation der akustischen Merkmale mit Neuronalen Netzen durch Verknüpfung mit Sprachmodellen eingebracht.

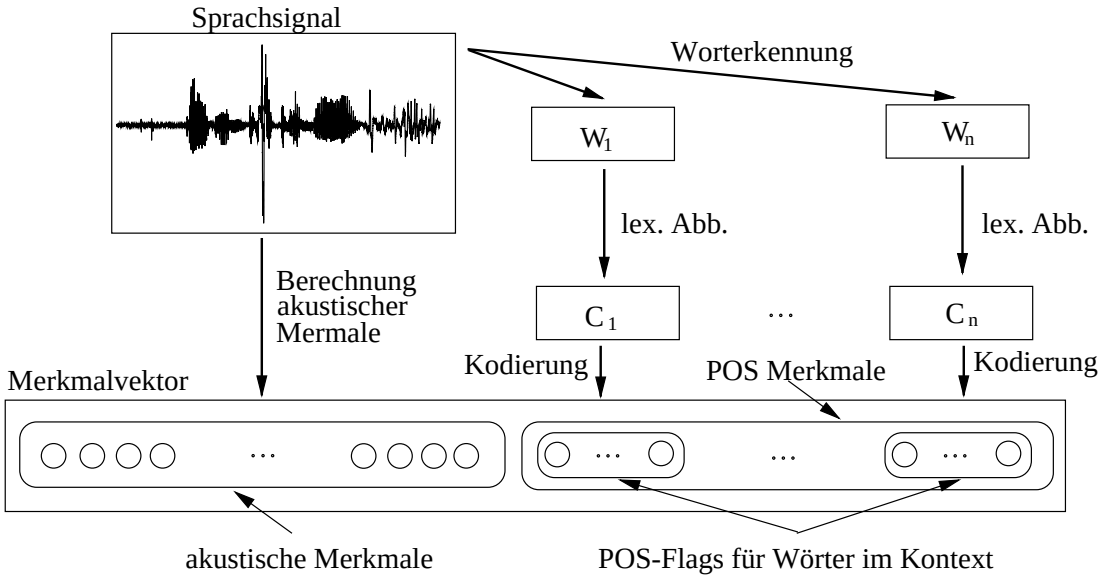
Als wichtige Neuerung erwies sich die Einbeziehung so genannten *Part-Of-Speech* (POS) Merkmale. Für einen Kontext von bis zu +/- 3 Wörtern wurden die Wortidentitäten der erkannten Wörter auf Wortklassen (POS-Klassen) abgebildet. Das verwendete Klassensystem war hierbei syntaktisch motiviert (daher der Name *Part-Of-Speech*). Um eine effiziente Berechnung der POS-Merkmale zu gewährleisten, wurde die Wortklasseninformation in das Lexikon eingebunden, d.h. *nicht* durch so genannten Parsing ermittelt. Notgedrungen war das resultierende Klassensystem unterspezifiziert, d.h. Wörter, die je nach Kontext verschiedenen Klassen angehören

können, wurden nach bestimmten Kriterien eindeutig auf eine POS-Klasse abgebildet.

Zusätzlich zu den akustischen Merkmalen wurde nun die POS-Information in Form von Flags in den Merkmalvektor integriert. Bei einer Kodierung als Flag gibt es für jedes Wort des betrachteten Kontexts soviele Merkmale, wie es unterschiedliche Klassen gibt; ein Flag nimmt den Wert 1 an, falls das Wort zur entsprechenden Klasse gehört, andernfalls den Wert 0. Das in *Verbmobil* verwendete hierarchische POS-Klassensystem hat im Deutschen 15 Klassen auf der untersten Hierarchie-Ebene. Auf der nächsthöheren Hierarchieebene sind diese 15 Klassen zu sechs Oberklassen zusammengefaßt. Eine noch gröbere Klasseneinteilung mit nur noch zwei Klassen existierte zwar, wurde aber nicht eingesetzt. Durch die Verwendung von POS-Information wird dem Neuronalen Netz ermöglicht, bei einem Kontext von +/- 3 Wörtern ein einfaches 7-gramm Sprachmodell zu erlernen. In akustisch ambigen Situationen kann dieses syntaktische Wissen verwendet werden, um die Ambiguitäten aufzulösen. Eine nachträgliche Einbeziehung von syntaktischem Wissen durch Sprachmodelle kann das nicht leisten.

Tatsächlich konnte durch Einbeziehung von syntaktischen Wissen in Form von POS-Merkmalen die Klassifikation der prosodischen Ereignisse erheblich verbessert werden, und zwar für alle der drei untersuchten prosodischen Phänomene (Grenzen, Akzente und Fragen) in den Sprachen Englisch und Deutsch.

Die um POS-Merkmale erweiterte Merkmalberechnung ist im folgenden Bild skizziert. Die Worterkennung liefert die Worthypothesen $W_1 \ldots W_n$ (für einen Kontext von n Wörtern). Diese Wörter werden mit Hilfe der im Lexikon kodierten POS-Abbildung auf die POS-Klassen $C_1 \ldots C_n$ abgebildet. Durch eine Kodierung als Flag wird diese Klasseninformation in binäre Merkmale umgewandelt. Weitere Informationen zu diesem Forschungskomplex sind in [6] zu finden.



Die so berechnete prosodische Information wird für die sogenannte „flache Analyse“ verwendet. Hierbei wird der Redebeitrag des Benutzers in Dialogakte (z.B. Begrüßung, Vorschlag) zer-

legt. Diese Dialogakte werden zum einen als Transfer-Einheiten der gesprochenen Sprache zur flachen Übersetzung und zum anderen als elementare Einheiten zur Plan-Erkennung eines Dialoges eingesetzt. Die Zerlegung eines Dialoges in einzelne Benutzeräußerungen, denen dann jeweils ein Dialogakt zugeordnet werden kann, wird traditionell als sequentieller Prozess realisiert.

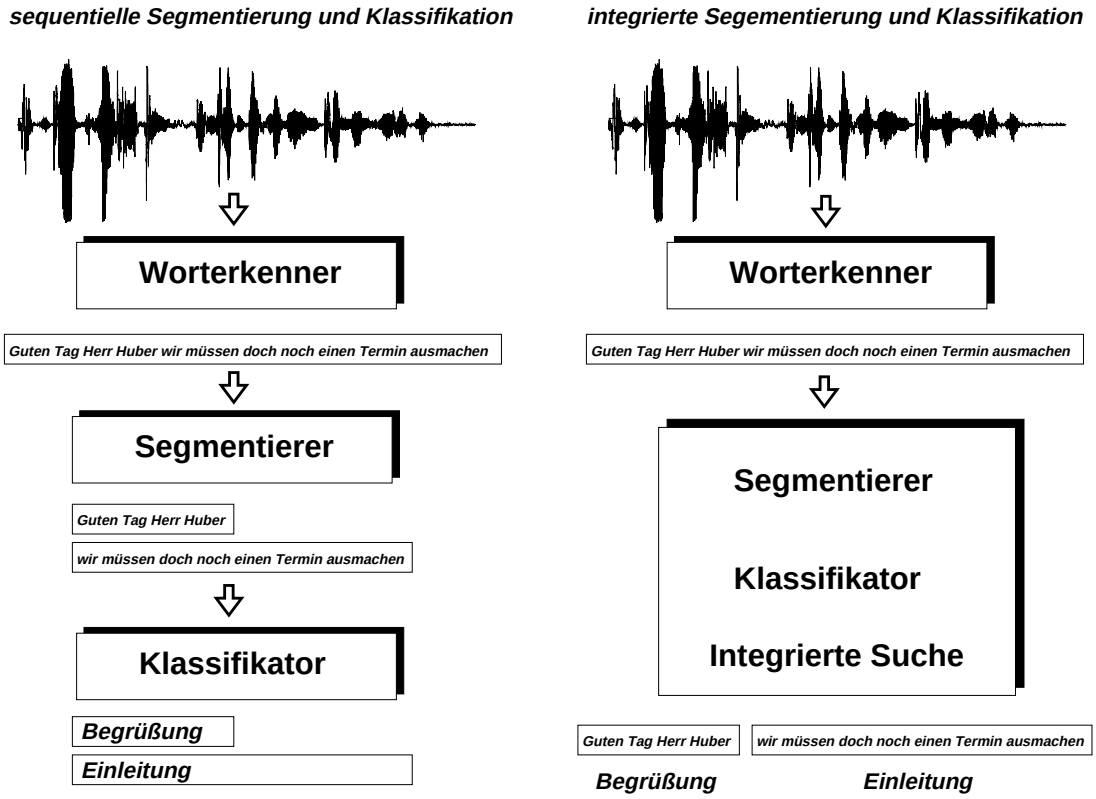


Bild 17: Sequentielle und integrierte Segmentierung und Klassifikation von Dialogakten

Ein wesentlich eleganteres Verfahren führt die Segmentierung gleichzeitig mit der Klassifikation der Dialogakte durch. Für diese integrierte Verarbeitung wird ein effizienter Suchalgorithmus benötigt. Im Rahmen der Arbeiten des *Verbmobil*–Projektes wurde hierfür der A^*–Algorithmus verwendet, der es erlaubt, gleichzeitig verschiedene Informationsquellen für diese Aufgabenstellung zu verwenden. Im Gegensatz zur sequentiellen Verarbeitung kann die integrierte Verarbeitung durch den gleichzeitigen Einsatz mehrerer Informationsquellen Fehlentscheidungen eines Klassifikators korrigieren. Dieses hat eine Verbesserung der Erkennungsrate sowohl bei der Segmentierung als auch bei der Dialogakterkennung zur Folge [**?**].

In automatischen Dialogsystemen kann es leicht vorkommen, dass der Benutzer, z.B. ein Anrufer bei einem vollautomatischen Call–Center, vom System falsch verstanden wird. Geschieht dies mehrmals während des Dialogs oder mehrfach hintereinander, so ist ein automatisches Weiterleiten zu einem Angestellten durchaus sinnvoll, um zu vermeiden, dass der Anrufer genervt auflegt und nie wieder anruft. Dabei ist es wichtig, dass das System in der Lage ist, die Veränderung

des Gemütszustands des Anrufers von einem neutralen Ausgangszustand in einen verärgerten Zustand zu erkennen. Ab einem gewissen Punkt muss dann das System auf den Mitarbeiter des Call–Centers weiterleiten.

Für die automatische Klassifikation des Gemütszustandes ist die Prosodie eine mögliche Wissensquelle. Je verärgerter ein Anrufer wird, desto mehr Veränderungen im Vergleich zur neutralen Ausgangslage werden in der Energie und der Grundfrequenz des Gesprochenen auftreten [3, 31]. jedoch die gesprochenen Wörter einer Äußerung, so werden die Veränderung, auch wenn der Sprecher sichtlich genervt ist, nicht in allen Wörtern auftreten. Eine vorausgehende syntaktisch-prosodische Segmentierung der Äußerung in einzelne syntaktisch-prosodische Einheiten und die Klassifikation der einzelnen Einheiten mit prosodischen Merkmalen ist daher notwendig. Würde man immer nur prosodische Merkmale wie z.B. den Mittelwert der Grundfrequenz über die kompletten Äußerungen berechnen, so kann es durchaus vorkommen, dass sich die prosodischen Veränderungen einzelner Teile von der gesamten Äußerung nicht abheben und eine Klassifikation fehl schlagen würde. Bild 18 zeigt eine Segmentierung einer Äußerung nach syntaktisch-prosodischen Einheiten (senkrechte Striche) und die Annotation der einzelnen Einheiten mit prosodischen Merkmalen (grau unterlegte Wörter).

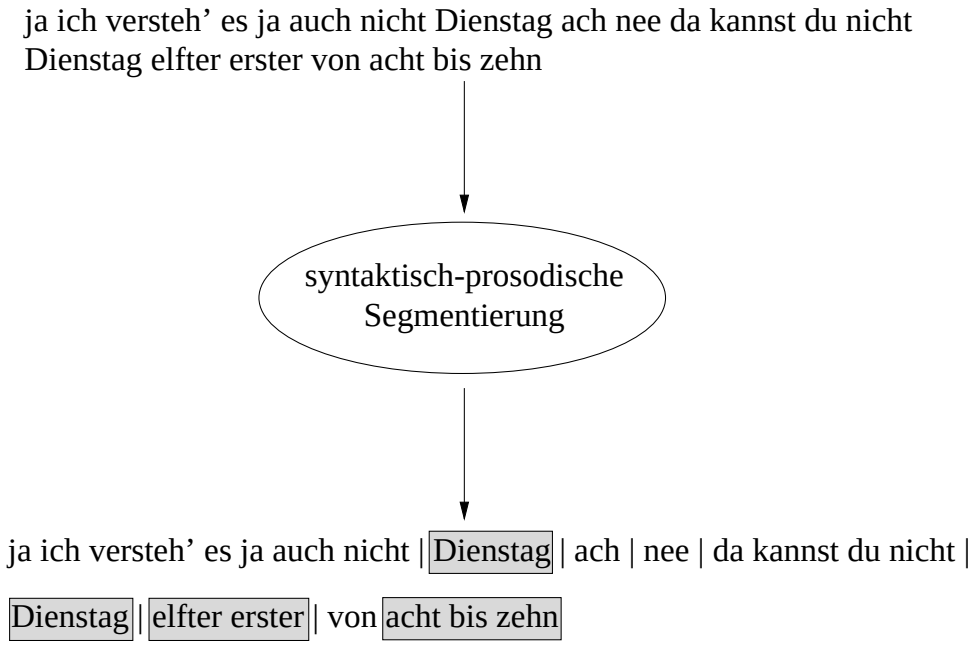


Bild 18: Syntaktisch-prosodische Segmentierung einer Äußerung. Die senkrechten Striche markieren die syntaktisch-prosodischen Grenzen und die grau unterlegten Wörter wurden prosodisch auffällig gesprochen.

Als Konsequenz aus den Untersuchungen in *Verbmobil*werden im *SmartKom*-Projekt (s.u.) die Arbeiten zur Emotion dahingehend erweitert, dass ein Modul entwickelt wird, welches generell nach Kommunikationsproblemen bei automatischen Dialogsystemen sucht. Dieses Modul verwendet linguistische und visuelle Information und entscheidet mit Hilfe von unterschiedlichen Klassifikatoren, ob eine kritische Phase im Dialog zu erkennen ist, d.h., ob der Benutzer

erkennbar verärgert ist; ist dies der Fall, so wird eine Aktion des Dialogmanagers initiiert, die das Funktionieren des weiteren Dialogs gewährleisten soll.

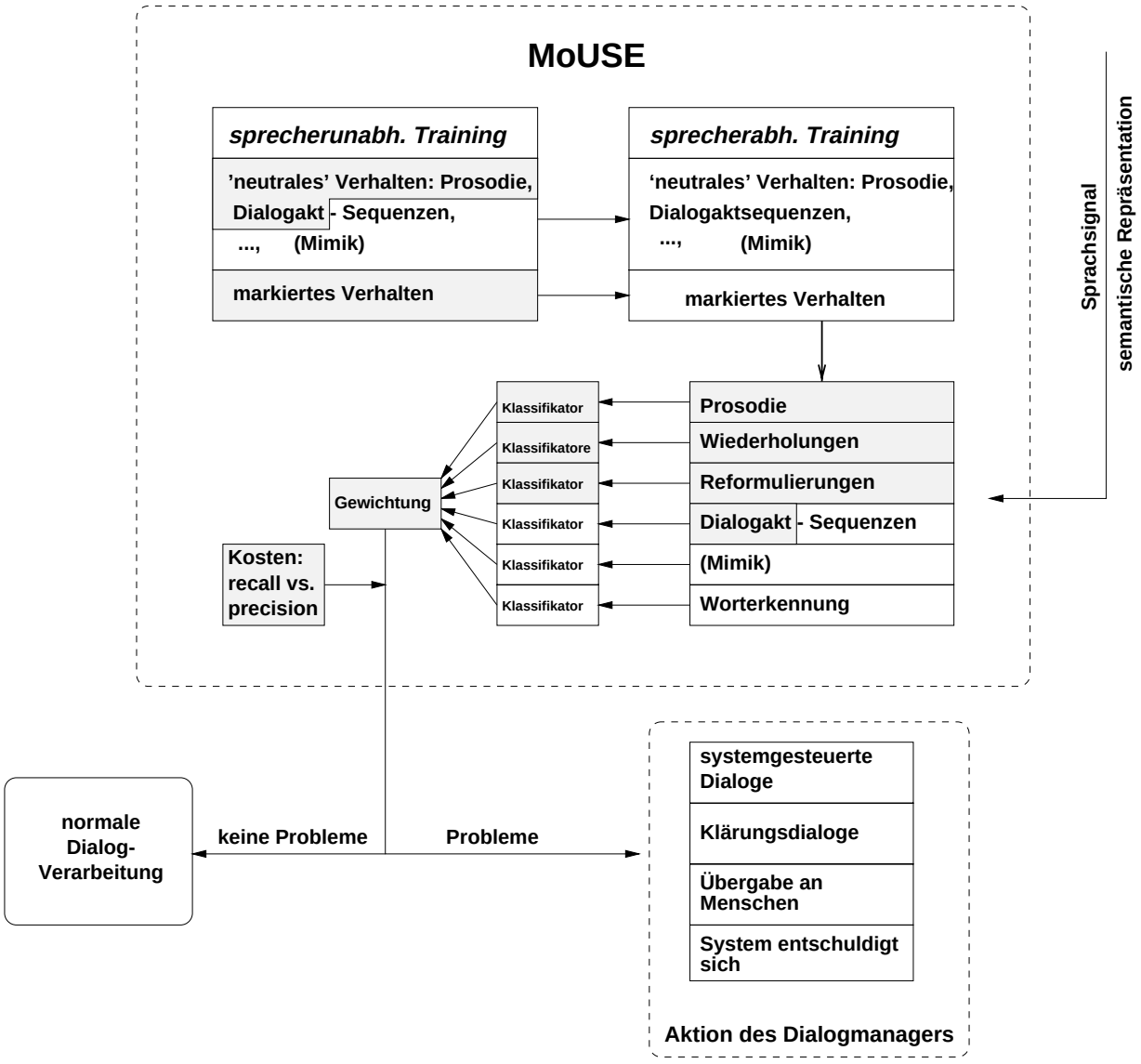


Bild 19: **M***onitoring* **o***f* **U***ser* **S***tate [especially of]* **E***motion* **MoUSE** – ein Überblick über die Architektur; schon implementierte Komponenten sind schraffiert

5.3 Das SmartKom–Projekt

In *SmartKom* ist ein Ziel, dass der Benutzer multimodal mit dem System kommunizieren kann. Einer der Modi ist die Sprache und in der Sprache ist ein Aspekt die Prosodie. Für die automatische Sprachverarbeitung ist die Verwendung von prosodischer Information allgemein anerkannt. Mit Hilfe der prosodischen Information wie Intonation oder Lautheit werden im Gesprochenen Grenzen gesucht, die Rückschlüsse auf die Satzstruktur des Gesprochenen zulassen.

Das Prosodie-Modul arbeitet eng mit der Worterkennung zusammen. Die Ergebnisse der Worterkennung werden vom Prosodie-Modul z.B. für die Segmentierung der Phrasen verwendet. Schwache und starke Phrasengrenze sowie irreguläre Grenze und keine Grenze werden klassifiziert. Das Prosodie-Modul verwendet das Sprachsignal ebenso als Eingabe, um Fragen oder Satzgrenzen zu detektieren. Beim Satzmodus wird Frage und Nichtfrage unterschieden, bei der Akzentuierung kein Akzent, sekundärer und prominenter Akzent.

Im Rahmen von *SmartKom* wurde das Prosodie-Modul an die neue Umgebung angepasst. Dabei waren zum einen Änderungen auf Grund neuer Transkriptionsregeln und bei der Verarbeitung der Lexikoneinträge sowie neuer Regeln im Lexikon nötig. Zum anderen waren an den Schnittstellen weitreichende Änderungen erforderlich. Die bisherige Schnittstelle im reinen Textformat wurde durch eine neue ersetzt. Als neues Datenaustausch-Format wird die aktuelle *Markup*-Sprache XML (Extensible Markup Language) verwendet, welche mehr Flexibilität und Robustheit bietet und gleichzeitig auch leicht die Dokumentation gestattet. Dazu war es einerseits erforderlich, das neue Format zu definieren, andererseits mussten die neuen Ein- und Ausgabeformate für die Verarbeitung in das System integriert werden.

Für die Detektion der prosodischen Merkmale werden in der Demonstratorversion des Prosodiemoduls derzeit neuronale Netze verwendet, welche mit VERBMOBIL-Daten trainiert worden sind. Für die nächste Version sollen hier vermehrt *SmartKom*-Korpora verwendet werden, welche bisher aber noch nicht in ausreichendem Maß zur Verfügung standen.

Die Mimikanalyse, die im Bereich der Dialogsysteme eine völlig neue Komponente darstellt, kann die Kommunikation zwischen Mensch und Maschine durch Erkennung unterschiedlicher Phänomene unterstützen. Dazu gehört unter anderem die Erkennung

- von Kommunikationsproblemen (Ärger, Langeweile, Lächeln, . . .) und
- des Aufmerksamkeitsfokus.

Solche Benutzerzustände werden durch die Verfolgung und Auswertung des Gesichts derjenigen Person bestimmt, die gerade mit dem *SmartKom*-System arbeitet.
Da sich der Hintergrund in allen drei *SmartKom*-Szenarien (Home, Public, Mobil) dynamisch ändern kann, können Verfahren die nur Differenzbildanalyse basieren nicht eingesetzt werden. Deshalb wurde im ersten Schritt versucht, das Gesicht des Anwenders durch Hautfarbensegmentierung zu Lokalisieren. Experimente haben gezeigt, dass dies in einem beleuchtungsunabhängigem Farbraum (YC_rC_b) sehr gut durch Modellierung mit einer Mischungsverteilung erreicht werden kann.

Für die Klassifikation von Mimik wurde vorerst eine Internetdatenbank mit statischen Bildern verwendet. Die Bilder einer Person sind in Bild20 zu sehen. Diese Portraitaufnahmen von un-

Bild 20: Eine Person mit den vier verschiedenen Gesichtsausdrücken *Neutral*, *Lächeln*, *Ärger* und *Schreien*.

terschiedlichen Personen sind in elf Kategorien eingeteilt, von denen für erste Versuche *Neutral*, *Lächeln*, *Ärger* und *Schreien* verwendet wurden.

Für die Klassifikation wurden Neuronale Netze und Support Vector Machines auf den Grauwerten und ein Bayesklassifikator auf Waveletmerkmalen getestet. Es zeigt sich, dass der neutrale Gesichtsausdruck häufig mit dem ärgerlichen verwechselt wird, wohingegen *Lächeln* und *Schreien* sehr sicher unterschieden werden können.

Die Gestenanalyse hat im Projekt *SmartKom* eine andere Bedeutung im Sinne der klassischen Gestenerkennung. Sie soll zusammen mit den anderen Kanälen wie Sprache und Mimik die Kommunikation zwischen Maschine und Mensch unterstützen. Insgesamt sind vier Gesten definiert:

- Klicken
- Löschen
- Hervorheben
- Wählen

Diese Gesten werden durch Verfolgung und Auswertung der Hand- bzw. Stiftbewegung auf einem Grafiktableau erkannt und analysiert. Von Bedeutung sind die Gestennamen: nicht die formalen, sondern die funktionalen Gesten bzw. die Semantik der Gesten sind voneinander zu unterscheiden.

Der Benutzer soll sich frei verhalten können. Deswegen sollen viele Gesten vorkommen können und es ist deshalb sinnvoll, nur die Semantik zu definieren. Beliebige Gestenformen können so im Verlauf des Tranierens des Modells festgestellt bzw. eingeführt werden.

In der ersten Phase ist es gelungen ein Demosystem aufzubauen. Dazu wurde auch die Schnittstelle in XML definiert. Weil *SmartKom* ein multimodales System ist, sind die Koordination bzw. Kommunikation der einzelnen Komponenten sehr kritisch. Diese ist durch die Implementierung

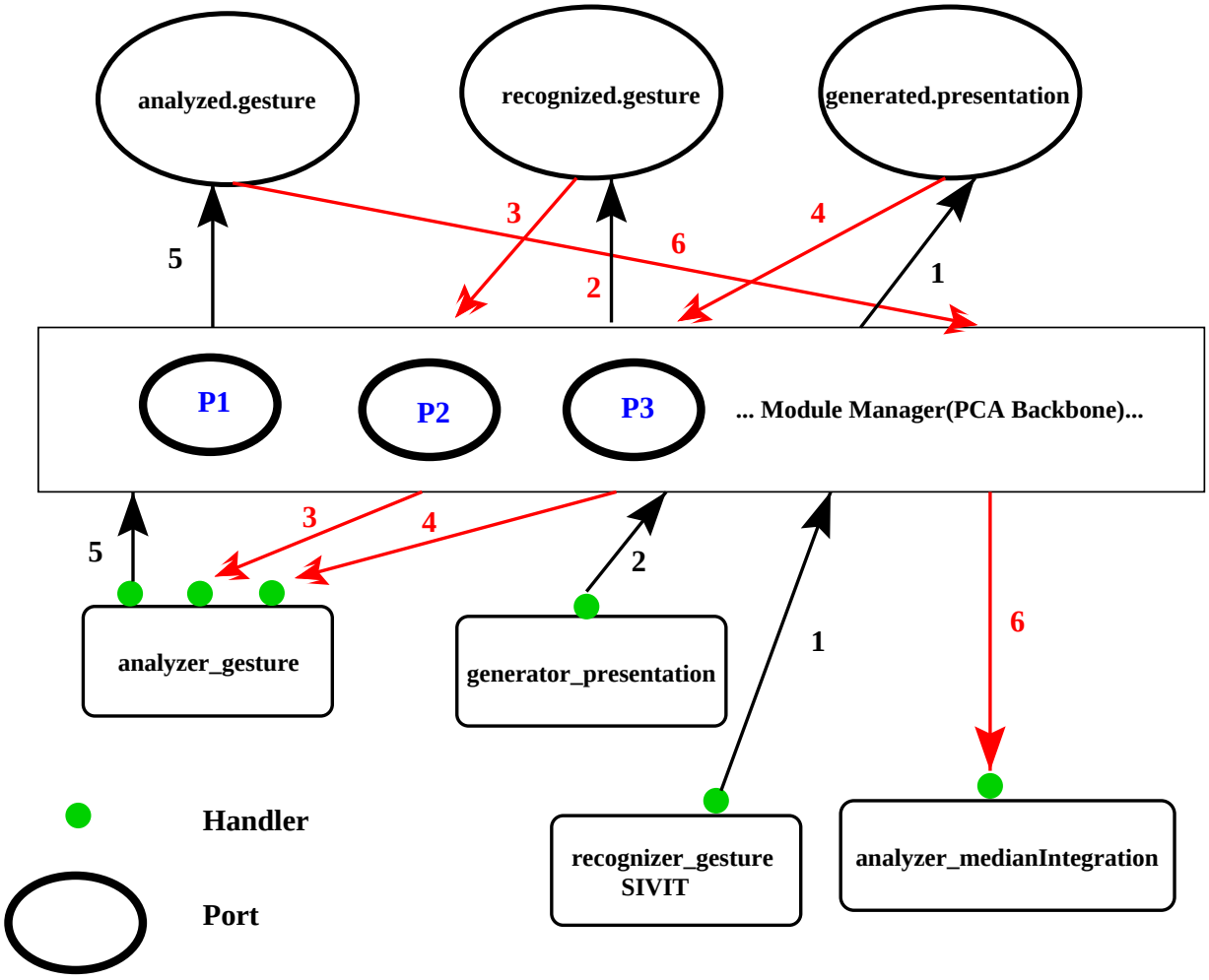


Bild 21: Der Gestenanalyseablauf in *SmartKom*

des Pool-Modul-Datenaustauchs erreicht worden, die auf PCA (*Pool Communication Architecture*) basiert. Dies erwies sich als sehr großer Programmierungsaufwand.

Als Merkmal zur Gestenerkennung wurde der Winkel zwischen zwei nacheinander folgenden Punkten der Gestenspur ausgesucht, damit die relative Position der Gesten vernachlässigt werden kann. Dabei sind sowohl absolute als auch relative Winkel möglich. Es wird getestet, ob eine von beiden oder eine Kombination von beiden das beste Merkmal liefert.

Im Bild 21 ist der Ablauf der Gestenanalyse in *SmartKom* zu sehen.

Für das Tranieren und die Klassifikation der Gesten wird geplant, HMMs einzusetzen. Diese sind aber im ersten System nicht verwendet worden, weil der Zusammenhang zwischen den obigen zwei Gestenklassen aufgrund mangelnder Testdaten noch nicht geklärt ist. Im Demosystem wurde eine *Klicken*-Geste getestet. Die weitere Untersuchung erfolgt in der zweiten Projektphase.

5.4 Sprechererkennung

In Zusammenarbeit mit den Firmen Dialog Communication Systems AG (D.C.S., Berlin/Tennenlohe) und MEDAV (Uttenreuth) wird ein Projekt zur Sprechererkennung durchgeführt. Während sich die DCS AG mit Sprecheridentifikation und –verifikation im Hinblick auf die Weiterentwicklung des biometrischen Authentisierungssystems „BioID“ beschäftigt, steht bei der Firma MEDAV die Überwachung von Funk- und Telefonkanälen im Vordergrund. Besonderes Interesse gilt hierbei den Aspekten Textunabhängigkeit, begrenzter Umfang an Trainings– und Testmaterial, Robustheit im Hinblick auf wechselnde und gestörte Kanäle, Landessprachenunabhängigkeit sowie der Untersuchung von zeitlichen Veränderungen der Stimme.

Ein auf der Klassifikation mit Gaußschen Mischverteilungen basierendes Sprecherverifikationssystem wurde auf der NIST-Stichprobe 1999 der jährlichen Sprechererkennungs-Evaluation bewertet. Es wird hierbei ein statistischer Ansatz verfolgt, wobei die Verteilung der Merkmale der trainierten Sprecher als gewichtete Summe von Gaußverteilungen modelliert wird. Wie in Bild 22 zu sehen, wird die Trainingsäußerung eines Zielsprechers zur Parameterschätzung einer Gaußschen Mischverteilungsdichte benutzt. Für eine zu verifizierende Testäußerung kann daraufhin ein Ähnlichkeitsmaß bezüglich des trainierten Sprechermodells berechnet werden. Zur Normierung dieses Ähnlichkeitsmaßes wird ein Weltmodell (ebenfalls ein GMM) herangezogen. Die eigentliche Verifikationsentscheidung, nämlich die Akzeptierung oder Rückweisung der Testäußerung, stellt eine Schwellwertentscheidung anhand des normierten Ähnlichkeitsmaßes dar.

Object recognition with statistical methods

The problem to identify several objects in a scene is a high-level vision task. From a pattern recognition point of view, an individual object can be considered as a pattern of contextual constrained features; at a higher level, a scene can be interpreted on the basis of contextual constraints between objects features.

A possible solution is to develop a statistically optimal model for recognition of objects in a scene, using a Maximum A Posteriori-Markov Random Field (MAP-MRF) approach; this model has been tested on a scene containing 2-D objects. The extension of this approach to scenes containing 3-D objects is not straightforward at all, and it is very difficult even for isolated 3-D objects, due to the problem of defining a neighborhood system for irregular sites.

Generally speaking, two major tasks in MRF modeling are how to define the neighborhood system, and how to choose the energy function for a proper encoding of constraints. How to define the neighbor relations between sites is related to their regularity; in the irregular case the neighborhood system must be defined by means of a heuristic distance that will be feature-dependent. When the considered application is 3-D object recognition, we have additional problems: if the chosen features are not invariant to pose, we should incorporate the pose parameters into the energy formulation and in the neighbor relations definition, with a dramatical increase in complexity; moreover, due to mutual occlusion, neighborhood changes with pose parameters. The energy function is a quantitative cost measure of the quality of a solution which defines the best solution as its minimum. In the case of irregular sites, the energy function’s formulation can be-

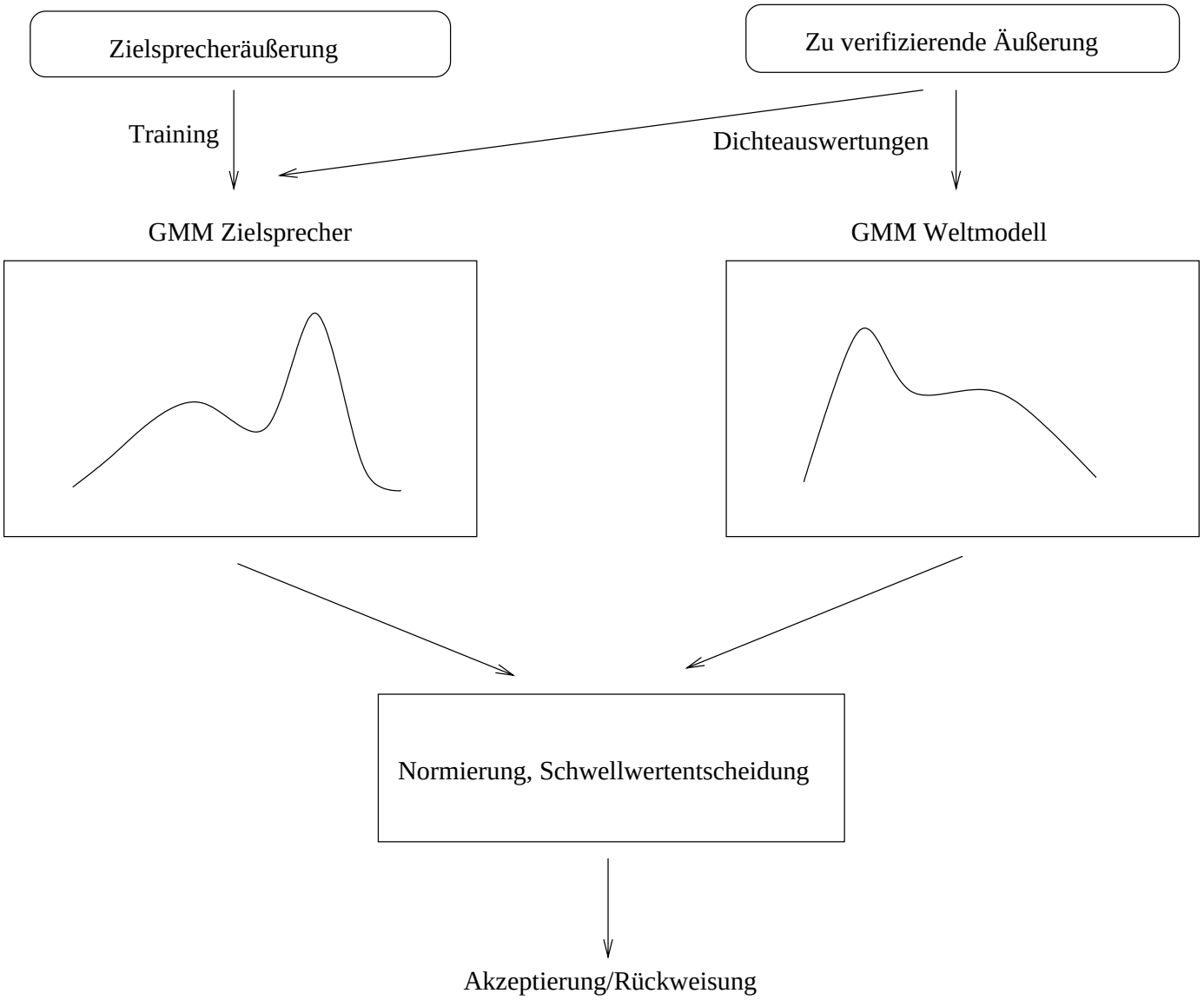


Bild 22: Sprecherverifikation

come something of an art, as it is generally done manually. These problems are so relevant that until now MRF models have never been used for probabilistic 3-D object recognition.

These problems can be solved using the tools of statistical mechanics; one of the greatest success in this research field has been the development of Spin Glass Theory (SGT). SGT provides sophisticated techniques and knowledge which can be used to deal with MRFs modeling problems in an elegant manner: full connectivity makes the neighborhood definition irrelevant, and the energy function is defined independently from the considered application; this makes it possible to find the analytical properties of the minima and may make it unnecessary to construct fast algorithms for seaching the absolute minima. To our knowledge, there are no previous works

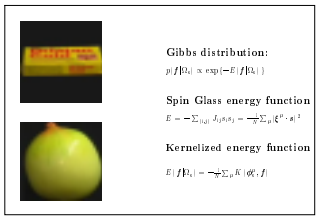


Bild 23: Object recognition strategy with SG-MRF

attempting to integrate SGT results in a MRF-MAP framework.

Bild 23 shows a scheme of the object recognition strategy based on SG-MRF.

In order to test the correctness of SG-MRF model, experiments were successfully performed on texture classification [9] and 3-D object recognition problems.The excellent experimental results obtained show the correctness of the new proposed model.

Future work will be focused on a comparison of the recognition performances of SG-MRF with those of other algorithms, like Nearest Neighbor Classifier, multi-layer perceptron and Support Vector Machines; an important issue will be also the investigation of possible connections between SG-MRF and Support Vector Machines. Once these points will be answered, it is our belief that it will be possible to use SG-MRF for the probabilistic model of a scene.

6 Studienarbeiten

1. Bouattour, S.: Objekterkennung und Lageschätzung mit geometrischen Modellen, Januar 2000

2. Kulicke, A.: Variable Length Markov Chains, November 2000

3. Haderlein, T.: Quantitative Stotteranalyse: Erstellung eines Diagnoseprogramms zur Fest-

stellung von Unflüssigkeiten in gesprochener Sprache, Februar 2000

4. Mattern, F.: Automatische Umgebungskartenerstellung durch probabilistische Fusion von Sensordaten mit einem autonomen mobilen System, September 2000

5. Vogt, S.: Optische Bestimmung der Kopfposition und -orientierung zur Visualisierung eines Lichtfeldes über ein Head-Mounted-Display, August 2000

7 Diplomarbeiten

1. Jaeger, M.: Geometrische modell-basierte Objektlokalisation, Erkennung und Inspektion, Mai 2000

2. Levit, M.: Benutzung von Sprachcharakteristika zur Klassifikation von Sprachvarietäten und Sprecherzuständen, Juli 2000

3. Schmidt, J.: Erarbeitung geeigneter Optimierungskriterien zur Berechnung von Kameraparametern und Szenengeometrie aus Bildfolgen, Mai 2000

4. Scholz, I.: A system for the visualization of virtual objects using a head-mounted display by localization of a real object of known geometry and color, Dezember 2000

5. Zhang, W.: A Recognizer for Inquiries in the Chinese Language, Februar 2000

8 Promotionen

1. Haas, J.: Probalistic Methods in Linguistic Analysis, Juni 2000

2. Merz, T.: Ein System zur modellbasierten Analyse von Interferenzmustern technischer Objekte, Dezember 2000

9 Habilitationen

1. Paulus, D.: Aktives Bildverstehen, Mai 2000

10 Vorträge

1. Batliner, A.: What Makes Speakers Angry in Human–Computer–Conversation, Third Workshop on Human–Computer–Conversation, Bellagio, Italien, 4.7.2000

2. Batliner, A.: Desperately Seeking Emotions: Actors, Wizards, and Human Beings, ISCA Workshop on Speech and Emotion, Newcastle, Nordirland, 5.9.2000

3. Batliner, A.: Prosodic models and speech recognition: towards the common ground, Prosody 2000, Krakau, Polen, 5.10.2000

4. Buckow, J.: Erkennung prosodischer Ereignisse in Verbmobil-II, VERBMOBIL Akustik-Workshop, Ulm, 29.06.2000

5. Buckow, J.: Detection of Prosodic Events Using Acoustic-Prosodic Features and Part-Of-Speech Tags, St. Petersburg, 26.09.2000

6. Caputo, B.: Texure Analysis in Mammographic Images: a Comparative Study, Master of Science's thesis discussion, University of Rome "La Sapienza", Rome, 3.02.2000

7. Caputo, B.: A Spin Glass Markov Random Field, Graduiertenkolleg 3D-Bildanalyse und -synthese, Erlangen, 11.07.2000

8. Caputo, B.: Analysis of Periapical Lesion Using Statistical Textural Features, Medical Infobahn for Europe 2000 (MIE2000), Hannover, 29.08.2000

9. Caputo, B.: Digital Mammography: Gabor Filter for Detection of Microcalcifications, Vision, Modeling and Visualization 2000, Saarbrücken, 24.11.2000

10. Caputo, B.: A Spin Glass Model of a Markov Random Field, NIPS2000 Workshop on New Perspective in Kernel Based Learning Methods , Breckenridge, CO (USA),01.12.2000

11. Deinzer, F.: Viewpoint Selection – A Classifier Independent Learning Approach, IEEE Southwest Symposium on Image Analysis and Interpretation, Austin/Texas, 4.4.2000

12. Deinzer, F.: Classifier Independent Viewpoint Selection for 3-D Object Recognition DAGM 2000, Kiel, 14.9.2000

13. Denzler, J.: Combining Computer Graphics and Computer Vision: Visual Self-Localization Using Particle Filters, Robotics Laboratory, Carnegie Mellon University, Pittburgh, USA, 21.01.00

14. Denzler, J.: An Information Theoretic Approach for Optimal Sensor Data Acquisition in State Estimation, Vision and Robotics Seminar, University of Rochester, Rochester, USA, 12.04.00

15. Denzler, J.: Probabilistische Modellierung von Sensordaten– und Aktionsfolgen im dynamischen Rechnersehen, Informatik Kolloqium, Universiät Erlangen, Erlangen, 26.06.00

16. Denzler, J.: Combining Computer Graphics and Computer Vision: Visual Self-Localization Using Particle Filters, Graduiertenkolleg 3D-Bildanalyse und -synthese, Universität Erlangen, Erlangen, 27.06.00

17. Denzler, J.: An Information Theoretic Approach to Optimal Sensor Data Selection for State Estimation of Static Systems, Centre for Vision Research, York University Toronto, Toronto, USA, 14.09.00

18. Deventer, R.: Non-linear modelling of a production process by hybrid Bayesian Networks, European Conference on Artificial Intelligence, Berlin, 25.08.2000

19. Deventer, R.: Control of Dynamic Systems Using Bayesian Networks, 1st Workshop on Probabilistic Reasoning, IBERAMIA/SBIA 2000, Atibaia, 21.11.2000

20. Gebhard, A.: System zur Diagnoseunterstützung von Patienten mit Gesichtslähmungen, Workshop BVM 2000, München, 14.03.2000

21. Gebhard, A.: Systemdemonstration: Diagnoseunterstützung von Patienten mit Gesichtslähmungen, Workshop BVM 2000, München, 14.03.2000

22. Gebhard, A.: Robust Facial Feature Localization by Coupled Features, FG 2000, Grenoble, 29.03.2000

23. Gebhard, A.: Lokalisation von Gesichtern und Gesichtsmerkmalen, Internes Bildverarbeitungsforum Fraunhofer IIS, Erlangen, 04.04.2000

24. Gebhard, A.: Echtzeit-Verfolgung von Gesichtern und Gesichtsmerkmalen, Fachzentrum Mensch-Maschine-Interaktion Siemens, München, 07.12.2000

25. Gebhard, A.: Lokalisation und Verfolgung von Gesichtern und Gesichtsmerkmalen mit Support Vektor Machines, Oberseminar am Institut für Kognitive Systeme der Universität Kiel, Kiel, 14.12.2000

26. Heigl, B.: Combining Computer Graphics and Computer Vision for Probabilistic Visual Robot Navigation, SPIE AeroSense 2000, Enhanced and Synthetic Vision, Orlando, USA, 25.04.2000

27. Heigl, B.: Acquisition and Applications of Image-Based Scene Models, Weekly Vision Seminar, Vision and Robotics Group, Computer Science Department, University of Rochester, Rochester, USA, 27.04.2000

28. Niemann, H.: Unterstützung der Diagnose und Rehabilitation von Gesichtslähmungen mit Methoden der Bildanalyse, Tag der Informatik, Erlangen, 5.5.2000

29. Nöth, E.: Textunabhängige Sprecheridentifikation zur Zugangskontrolle Arbeitsgruppenseminar der HNO–Klinik, Uni Erlangen-Nürnberg, 9.2.2000

30. Nöth, E.: Benutzung von Sprachcharakteristika zur Klassifikation von Sprachvarietäten und Sprecherzuständen, 2. SKAT Workshop, Uttenreuth, 9.6.2000

31. Nöth, E.; Prosodic Analysis, Verbmobil-Abschlusssymposium, Saarbrücken, 30.7.2000

32. Nöth, E.: Telefonische Auskunftsysteme, BMBF Veranstaltung Sprachtechnologie, Bonn, 29.9.2000

33. Nöth, E.: A Multilingual Prosody Module in a Speech–to–Speech Translation System, Workshop on Multi-Lingual Speech Communication, Kyoto, Japan, 13.10.2000

34. Nöth, E.: Telephony Based Information Retrieval Systems, ATR, Kyoto, Japan, 13.10.2000

35. Nöth, E.: Automatic Stuttering Recognition using Hidden Markov Models, International Conference on Spoken Language Processing, Beijing, China, 20.10.2000

36. Nöth, E.: Sprachtechnologie: Der Markt der Anbieter, EUROMAP Informationstag Sprachtechnologie: Forschung und Anwendungen, Berlin, 14.12.2000

37. Ohler, U.: Stochastic Segment Models of Eukaryotic Promoter Regions, Pacific Symposium on Biocomputing, Honolulu, USA, 8.01.2000

38. Ohler, U.: Localization of Eukaryotic Promoters and Transcription Start Sites, Round Table Discussions at the Innovationskolleg Theoretische Biologie, HU Berlin, 30.06.2000

39. Ohler, U.: A Statistical Model for the Detection of Eukaryotic Promoters in Genomic DNA, Summer Seminar of the Boehringer Ingelheim Fonds, Hirschegg/Austria, 12.10.2000

40. Ohler, U.: Bioinformatik am Lehrstuhl für Mustererkennung, 1. Bioinformed-Innovationsforum, TGZ Würzburg, 14.11.2000

41. Paulus D.: Die virtuelle TEX-Maschine, Habilitationsvortrag, Universität Erlangen–Nürnberg, 07.06.2000

42. Paulus D.: Localizing colored objects, Farbworkshop 2000, Berlin, 05.10.2000

43. Paulus D.: Rechnersehen, Kolloquium des SFB 535, Augenklinik Erlangen, 08.11.2000

44. Paulus D.: Aktive Objekterkennung mit Farbe, Informatik-Kolloquium, Universität Koblenz-Landau, 13.12.2000

45. Reinhold, M.: Erscheinungsbasierte, statistische Objekterkennung, Graduiertenkolleg 3D-Bildanalyse und -synthese, Universität Erlangen, Erlangen, 18.07.2000.

46. Reinhold, M.: Active Appearance–Based Object Recognition Using Viewpoint Selection, Workshop Vision, Modeling, and Visualization 2000, Saarbrücken, 22.11.2000.

47. Stemmer, G.: The Utility of Semantic-Pragmatic Information and Dialogue-State for Speech Recognition in Spoken Dialogue Systems, TSD 2000, Brno, Tschechien, 15.09.2000

Literatur

[1] U. Ahlrichs, D. Paulus, H. Niemann: *Integrating Aspects of Active Vision into a Knowledge-Based System*, in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR)*, IEEE Computer Society Press, Barcelona, Spain, 2000, S. IV:579–582.

[2] A. Batliner, A. Buckow, H. Niemann, E. Nöth, V. Warnke: *The Prosody Module*, in W. Wahlster (Hrsg.): *Verbmobil: Foundations of Speech-to-Speech Translations*, Springer, New York, Berlin, 2000, S. 106–121.

[3] A. Batliner, K. Fischer, R. Huber, J. Spilker, E. Nöth: *Desperately Seeking Emotions: Actors, Wizards, and Human Beings*, in R. Cowie, E. Douglas-Cowie, M. Schröder (Hrsg.): *Proc. ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, Newcastle, Northern Ireland, September 2000, S. 195–200. 29

[4] A. Batliner, R. Huber, H. Niemann, E. Nöth, J. Spilker, K. Fischer: *The Recognition of Emotion*, in W. Wahlster (Hrsg.): *Verbmobil: Foundations of Speech-to-Speech Translations*, Springer, New York, Berlin, 2000, S. 122–130.

[5] A. Batliner, E. Nöth, B. Möbius, G. Möhler: *Prosodic models and speech recognition: towards the common ground*, in *Proc. of Prosody 2000*, Krakau, Polen, September 2000, (to appear).

[6] J. Buckow, A. Batliner, R. Huber, H. Niemann, E. Nöth, V. Warnke: *Detection of Prosodic Events Using Acoustic-Prosodic Features and Part-Of-Speech Tags*, in *Proc. of the International Workshop SPEECH AND COMPUTER (SPECOM'00)*, St. Petersburg, 2000, S. 63–66. 27

[7] B. Caputo, G. E. Gigante: *Analysis of Periapical Lesion Using Statistical Textural Features*, in *Medical Infobahn for Europe: Proceedings of MIE2000 and GMDS2000*, IOS Press, Hannover, Germany, 2000, S. 1231–1234.

[8] B. Caputo, G. E. Gigante: *Digital Mammography: Gabor Filter for Detection of Microcalcifications*, in B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2000*, infix, Berlin, Saarbrücken, November 2000, S. 375–381.

[9] B. Caputo, J. Hornegger, D. Paulus, H. Niemann: *A Spin Glass Model of a Markov Random Field*, in *Proceedings of the NIPS 2000 Workshop on New Perspective in Kernel Based Learning Methods*, available at http://www.dcs.rhbnc.ac.uk/colt/nips2000.html, Breckenridge, CO (USA), 2000. 8, 36

[10] F. Deinzer, J. Denzler, H. Niemann: *Classifier Independent Viewpoint Selection for 3-D Object Recognition*, in G. Sommer (Hrsg.): *Mustererkennung 2000*, Springer, Heidelberg, September 2000, S. 237–244. 10

[11] F. Deinzer, J. Denzler, H. Niemann: *Viewpoint Selection - A Classifier Independent Learning Approach*, in *IEEE Southwest Symposium on Image Analysis and Interpretation*, IEEE Computer Society, California, Los Alamitos, Austin, Texas, USA, 2000, S. 209–213. 10

[12] J. Denzler: *Active Vision*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 131–140.

[13] J. Denzler, M. Zobel: *Object Tracking in Image Sequences*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 57–66.

[14] R. Deventer, J. Denzler, H. Niemann: *Control of Dynamic Systems Using Bayesian Networks*, in L. N. de Barros et. al (Hrsg.): *Proceedings of the IBERAMIA/SBIA 2000 Workshops*, Tec Art Editora, São Paulo, Atibaia, São Paulo, Brazil, November 2000, S. 33–39. 22

[15] R. Deventer, J. Denzler, H. Niemann: *Non-linear modeling of a production process by hybrid Bayesian Networks*, in W. Horn (Hrsg.): *ECAI 2000 Proceedings of the 14th European Conference on Artificial Intelligence*, IOS Press, Amsterdam, Berlin, August 2000, S. 576–580.

[16] M. Eitschberger, S. Enzelberger, K. Steuss, M. Roth, C. Frank, I. Fischer: *Mädchen + Technik Praktikum 2000*, Berichte der Frauenhofer Gesellschaft, Fraunhofer IRB Verlag, Stuttgart, Deutschland, 2000.

[17] K. Fischer, A. Batliner: *What Makes Speakers Angry in Human-Computer Conversation.*, in *Proc. of the Third Workshop on Human–Computer–Conversation*, Bellagio, Italien, July 2000, S. 62–67.

[18] A. Gebhard, D. Paulus: *Diagnosis support of patients with facial paresis*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 353–364.

[19] A. Gebhard, D. Paulus, B. Suchy, S. Wolf: *A System for Diagnosis Support of Patients with Facialis Paresis*, *KI*, Bd. 3/2000, 2000, S. 40–42.

[20] A. Gebhard, D. Paulus, B. Suchy, S. Wolf, H. Niemann: *System zur Diagnoseunterstützung von Patienten mit Gesichtslähmungen*, in *4. Workshop Bildverarbeitung für die Medizin*, Springer, 2000, S. 249–253.

[21] B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000.

[22] B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2000*, infix, Berlin, November 2000.

[23] M. Greiffenhagen, V. Ramesh, D. Comaniciu, H. Niemann: *Modeling and Performance Characterization of a Real-Time Dual Camera Surveillance System*, in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, IEEE Computer Society, Hilton Head Island, South Carolina, USA, 2000, S. 335–342.

[24] T. Haderlein, T. Wittenberg, M. Decher, E. Nöth: *Automatische Stottererkennung und Klassifikation mit Hilfe von Hidden Markov Modellen*, in *Proc. der 17. Wissenschaftlichen Jahrestagung der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie*, Tübingen, 2000, S. (to appear).

[25] B. Heigl, J. Denzler, H. Niemann: *Combining Computer Graphics and Computer Vision for Probabilistic Visual Robot Navigation*, in J. G. Verly (Hrsg.): *Enhanced and Synthetic Vision 2000*, Bd. 4023 von *Proceedings of SPIE*, April 2000, S. 226–235. 18

[26] B. Heigl, P. Eisert: *Coordinate Systems and Camera Models*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 26–30.

[27] J. Hornegger, H. Niemann, R. Risack: *Appearance–Based Object Recognition Using Optimal Feature Transforms*, *Pattern Recognition*, Bd. 33, 2000, S. 209–224.

[28] J. Hornegger, D. Paulus, H. Niemann: *Probabilistic Modeling in Computer Vision*, in B. Jähne, H. Haussecker (Hrsg.): *Handbook of Computer Vision and Applications* , Academic Press, London, 2000, S. 517–540.

[29] Y. Huang, D. Paulus, H. Niemann: *Background-Forground Segmentation Based on Dominant Motion Estimation and Static Segmentation*, in S. Loncaric (Hrsg.): *IWISPA 2000*, University Computer Center, Zagreb, Croatia, 2000, S. 69–74.

[30] Y. Huang, D. Paulus, H. Niemann: *Dynamic Gesture Analysis and Tracking Based on Dominant Motion Estimation and Kalman Filter*, in G. Sommer (Hrsg.): *Mustererkennung 2000*, Springer, Heidelberg, September 2000, S. 397–404.

[31] R. Huber, A. Batliner, J. Buckow, E. Nöth, V. Warnke, H. Niemann: *Recognition of Emotion in a Realistic Dialogue Scenario*, in *Proc. Int. Conf. on Spoken Language Processing*, Bd. 1, Beijing, China, Oktober 2000, S. 665–668. 29

[32] B. Möbius, G. Möhler, A. Schweitzer, A. Batliner, E. Nöth: *Prosodic models and speech synthesis: towards the common ground*, in *Proc. of Prosody 2000*, Krakau, Polen, September 2000, (to appear).

[33] H. Niemann: *Knowledge Based Processing of Medical Images*, *Künstliche Intelligenz*, , Nr. 3, 2000, S. 24–29.

[34] E. Nöth, A. Batliner, J. Buckow, R. Huber, V. Warnke, H. Niemann: *A Multilingual Prosody Module in a Speech–to–Speech Translation System*, in *Proc. of the Workshop on Multi-Lingual Speech Communication*, Kyoto, Japan, 2000, S. 110–115.

[35] E. Nöth, A. Batliner, A. Kießling, R. Kompe, H. Niemann: *Verbmobil: The Use of Prosody in the Linguistic Components of a Speech Understanding System*, *IEEE Trans. on Speech and Audio Processing*, Bd. 8, Nr. 5, 2000, S. 519–532.

[36] E. Nöth, H. Niemann, T. Haderlein, M. Decher, U. Eysholdt, F. Rosanowski, T. Wittenberg: *Automatic Stuttering Recognition using Hidden Markov Models*, in *Proc. Int. Conf. on Spoken Language Processing*, Beijing, China, 2000, S. 65–68.

[37] U. Ohler: *Promoter prediction on a genomic scale — the Adh experience*, *Genome Res.*, Bd. 10, Nr. 4, 2000, S. 539–542. 23

[38] U. Ohler, S. Harbeck, G. Stemmer, H. Niemann: *Stochastic segment models of eukaryotic promoter regions*, in R. B. Altman, K. Lauderdale, A. K. Dunker, L. Hunter, T. E (Hrsg.): *Pacific Symposium on Biocomputing*, Bd. 5, World Scientific, Singapore, 2000, S. 377–388. 23

[39] D. Paulus: *Aktives Bildverstehen*, 2000, Habilitationsschrift in der Praktischen Informatik, Universität Erlangen-Nürnberg.

[40] D. Paulus: *Object Recognition*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 108–121.

[41] D. Paulus: *Segmentation*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 80–92.

[42] D. Paulus: *Selected Applications*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 335–397.

[43] D. Paulus, U. Ahrlichs, B. Heigl, J. Denzler, J. Hornegger, M. Zobel, H. Niemann: *Active Knowledge–Based Scene Analysis*, *videre*, Bd. 1, Nr. 4, 2000, online-journal.

[44] D. Paulus, G. Dorkó, U. Ahlrichs: *Color segmentation for scene exploration*, in G. Stanke, M. Pochanke (Hrsg.): *6. Workshop Farbbildverarbeitung*, GFAI, Berlin, 2000, S. 13–20. 11

[45] D. Paulus, C. Drexler, M. Reinhold, M. Zobel, J. Denzler: *Active Computer Vision System*, in V. Cantoni, C. Guerra (Hrsg.): *Computer Architectures for Machine Perception*, IEEE Computer Society, Los Alamitos, California, USA, 2000, S. 18–27.

[46] D. Paulus, K. Horecki, K. Wojciechowski: *Localization of Colored Objects*, in *Proceedings of the International Conference on Image Processing (ICIP)*, IEEE Computer Society Press, Vancouver, BC, September 2000, S. III:492–495.

[47] M. G. Reese, G. Hartzell, N. L. Harris, U. Ohler, J. F. Abril, S. E. Lewis: *Genome annotation assessment in Drosophila melanogaster*, *Genome Res.*, Bd. 10, Nr. 4, 2000, S. 483–501. 23

[48] M. Reinhold, F. Deinzer, J. Denzler, D. Paulus, J. Pösl: *Active Appearance–Based Object Recognition Using Viewpoint Selection*, in B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2000*, infix, Berlin, Saarbrücken, November 2000, S. 105–112. 8, 10

[49] R. Schug, M. Zobel, J. Denzler, H. Niemann: *Sichtbasierte Personeneskortierung mittels einer autonomen mobilen Plattform*, in *Robotik 2000: Leistungsstand – Anwendungen – Visionen – Trends*, VDI-Berichte 1552, VDI Verlag, Düsseldorf, 2000, S. 459–464.

[50] E. Steinbach, B. Heigl: *Structure from Multiple Views*, in B. Girod, G. Greiner, H. Niemann (Hrsg.): *Principles of 3D Image Analysis and Synthesis*, Kluwer Academic Publishers, Boston–Dordrecht–London, 2000, S. 38–56.

[51] G. Stemmer, E. Nöth, H. Niemann: *The Utility of Semantic-Pragmatic Information and Dialogue-State for Speech Recognition in Spoken Dialogue Systems*, in K. P. Petr Sojka, Ivan Kopecek (Hrsg.): *Proc. of the Third Workshop on Text, Speech, Dialogue,- TSD 2000*, Bd. 1902 von *Lecture Notes in Artificial Intelligence*, Springer–Verlag, Berlin, September 2000, S. 439–444. 25

[52] C. Vogelgsang, B. Heigl, G. Greiner, H. Niemann: *Automatic Image–Based Scene Model Acquisition and Visualization*, in B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2000*, infix, Berlin, Saarbrücken, November 2000, S. 189–198. 16

[53] A. Weckenmann, V. Bettin, R. Stöber, H. Niemann, R. Deventer: *Modellierungsverfahren zur Regelung und Qualitätsoptimierung verkürzter Prozessketten*, in M. Geiger, W. Ehrenstein (Hrsg.): *Robuste, verkürzte Prozessketten fuer flächige Leichtbauteile, Tagungsband zum 1. Berichtskolloqium des SFB 396*, Meisenbach-Verlag, 2000, S. 51–68.

[54] C. W. Wightman, A. K. Syrdal, G. Stemmer, A. Conkie, M. Beutnagel: *Perceptually Based Automatic Prosody Labeling and Prosodically Enriched Unit Selection Improve Concatenative Text-To-Speech Synthesis*, in *Proc. Int. Conf. on Spoken Language Processing*, Bd. 2, Bejing, China, October 2000, S. 71–74.

[55] M. Wolf, T. Vogel, P. Weierich, C. Nimsky, H. Niemann: *Automatic Transfer of Pre–Operation fMRI Markers Into Intra-Operation MR–Images*, *Journal of Computer Aided Surgery*, Bd. 5, Nr. 1, 2000.

[56] M. Wolf, T. Vogel, P. Weierich, C. Nimsky, H. Niemann: *Automatische Übertragung von präoperativen fMRI–Markern in intraoperative MR–Datensätze*, in A. Horsch, T. Lehmann (Hrsg.): *Bildverarbeitung für die Medizin 2000*, Springer, Berlin, Heidelberg, München, 2000, S. 23–27.

[57] C. Yuan, H. Niemann: *An appearance based neural image processing algorithm for 3–D object recognition*, in *2000 International Conference on Image Processing (ICIP2000)*, IEEE Signal Processing Society, Vancouver, BC, Canada, 2000, S. 344–347. 12

[58] M. Zobel, A. Gebhard, D. Paulus, J. Denzler, H. Niemann: *Robust Facial Feature Localization by Coupled Features*, in *Proceeding of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Computer Society, California, Los Alamitos, Grenoble, France, 2000, S. 2–7. 15