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)	bis 30.06.01
Batliner, A., Dr.-Phil.	(wiss. Mitarb., BMBF)	bis 31.12.01
Buckow, J., Dipl.-Inf.	(wiss. Mitarb., BMBF)	bis 28.02.01
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
Frank, C.M., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.09.99
Gebhard, A., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	bis 30.06.01
Heigl, B., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	bis 30.04.01
Hertlein, H., Dipl.-Inf.	(wiss. Mitarb., MEDAV/DCS)	bis 31.12.01
Huber, R., Dipl.-Inf.	(wiss. Mitarb., BMBF)	bis 28.02.01
Karadag, C.	(Sekretärin, 1/2, SFB 603)	bis 31.10.01
Koppe, I.	(Sekretärin, 1/2)	18.12.92
Levit, M., Dipl.-Inf.	(wiss. Mitarb.)	01.02.99
Montel-Kandy, M.	(Sekretärin, 1/2, SFB 603)	15.11.01
Niemann, H., Dr.-Ing.	(Professor)	24.09.75
Nöth, E., Dr.-Ing.	(Akad. Oberrat)	01.02.85
Obermayer, W.	(Programmierer)	bis 31.10.01
Ohler, U., Dipl.-Inf.	(Stip., Boehringer)	bis 31.10.01
Pal, I.	(wiss. Mitarb, SFB 539)	bis 31.12.01
Paulus, D., Dr.-Ing. habil	(Akad. Oberrat)	01.03.87, beurlaubt ab 01.10.01
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
Scholz, I., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.03.01
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)	bis 28.02.01
Zeissler, V., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.05.01
Zobel, M., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.01.98

Gäste

Gäste:

Eliseeva, Olga	(Russland, xxxx)	05.07. – 20.08.2001
Horsch, Alexander, PD Dr.	(München, xxxx)	27.11. – 27.11.2001
Lahajnar, Fanci	(Slowenien, DAAD)	05.10.2001
Pinz, Axel, Prof. Dr.	(TU Graz, xxxx)	21.11. – 23.11.2001
Pauli, PD Dr. habil.	(xxxx, xxxx)	11.06. – 12.06.2001
Pomplun,Marc,Dr.	(xxxx, xxxx)	17.09. – 18.09.2001
Ribaric, Slobodan, Prof. Dr. sc.	(Zagreb, xxxx)	22.04. – 25.04.2001
Siepmann, Rolf, Dr.	(LMU München, xxxx)	08.10. – 08.10.2001
Sushkova, Lyudmuila T., Prof.	(Vladimir, xxxx)	08.10. – 08.10.2001
Schnoerr, Chr., Prof. Dr.	(xxxx, xxxx)	12.12. – 13.12.2001
Sokolov, Mikhail	(Russland, xxxx)	05.07. – 20.08.2001
Triesch, Jochen, Prof. Dr.	(Univ. of Rochester, xxxx)	18.06. – 08.07.2001
Vetter, Thomas, Prof. Dr.	(Freiburg, xxxx)	30.07. – 01.08.2001
Yuan, C.	(China, KAS)	bis xx.xx.xx
Yuille, Alan L., Prof. Dr.	(xxxx, xxxx)	14.05. – 19.05.2001

2 Einleitung

Seit über 25 Jahren wird am Lehrstuhl das Problem der „Mustererkennung“ untersucht, wobei 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 die 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 Visuali
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 (System Fränki) sowie auf die Entwicklung eines multimodalen Dialogsystems im Rahmen des Verbundprojektes SmartKom. Der Benutzer kann mit diesem System sowohl per Spracheingabe als auch über Zeigegesten kommunizieren. Darüberhinaus interpretiert das System die Mimik des Benutzers in Bezug darauf, ob der Benutzer zufrieden oder verärgert ist, und verwendet diese Information zur Steuerung des weiteren Dialogverlaufs.

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, B. Caputo, F. Deinzer, J. Denzler, J. Drexler, A. Gebhard, B. Heigl, I. Pal, M. Reinhold, J. Schmidt, I. Scholz, F. Vogt, M. Zobel)

Schwerpunkt der Forschungstätigkeiten im Bereich der Bildanalyse am Lehrstuhl ist die Objekterkennung. Arbeiten zur wissensbasierten Bildanalyse auf der Basis von semantischen Netzen wurden zu einem Abschluss gebracht. 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 sowie Eigenraumverfahren zur 3D–Objektlokalisation und Klassifikation. 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. Im vergangenen Jahr wurde der Hauptrechner für die Bild– und Sprachverarbeitung auf einen Dual Pentium III 850 MHz aufgerüstet. Die damit gestiegene Rechenleistung kommt vor allem den Modulen Objektverfolgung und Aktionsauswahl zu Gute.

Das anlässlich des 25–jährigen Bestehens des Lehrstuhls für Mustererkennung enwickelte System Mobsy wurde weiterhin gewartet und bei zahlreichen Anlässen (Tag der Informatik, Erstsemestereinführung, Mädchenpraktkum) vorgeführt: Mobsy wartete im 9. Stock vor den Aufzügen, erkannte ankommende Gäste und nahme diese in Empfang. Danach gab er einen kurzen Überblick über angebotene Demos. Außerdem gab Mobsy bei Fragen Auskünfte über laufende Arbeiten am Lehrstuhl. Das System läuft ohne Eingriff von außen robust und fehlertolerant und zeigt die erfolgreiche Integration von Sprach– und Bildverarbeitung in einem Serviceroboter Szenario. Die Akzeptanz bei den Benutzer macht deutlich, dass natürliche Sprache und Dialog als Schnittstelle zum System sowie aktive Kamerasteuerung zur Gesichtsverfolgung wichtige Aspekte in einem solchen Anwendungsgebiet darstellen. Regelmäßige, automatische Rekalibrierung mit-

tels visueller Information sowie Hinderniserkennung mittels Infrarotsensorik stellt den robusten Betrieb auch bei zahlreichen Besuchern im 9. Stock sicher.

3.1 Objekterkennung und Szenenanalyse

Heutige Kamerasystem sind meistens mit einer Vielzahl, rechnersteuerbarer Freiheitsgrade ausgestattet (z.B. Schwenk–/Neigebewegungen der Kamera, Änderung der Brennweite). Diese Freiheitsgrade gilt es im Bereich des Rechnersehens auszunutzen, um somit die Verarbeitung robuster zu gestalten und neue Funktionalität in bestehende Systeme zu integrieren. In der Praxis ermöglichen diese Freiheitsgrade, die Sensordaten aufzunehmen, die zur Lösung eines gegebenen Problems am besten geeignet scheinen.

Schwerpunkte der vergangenen Jahre waren die Entwicklung einer Ansichtenplanung bei der Objekterkennung mittels Reinforcement Learning sowie die Entwicklung eines theoretisch wohlfundierten Ansatzes zur optimalen Sensordatenauswahl bei der Schätzung des Zustands statischer Systeme. Doch auch im Bereich der Objektverfolgung, d. h. der Schätzung des Zustands eines dynamischen Systems, kann die Zustandsschätzung von optimal gewählten Sensordaten profitieren. Dazu wurde der informationstheoretische Ansatz vom statischen auf den dynamischen Fall übertragen. Das zu optimierende Gütekriterium besteht in diesem Fall aus der bedingten Entropie des Zustands gegeben die Beobachtung. Das entscheidende Problem bei der Übertragung auf den dynamischen Fall ergibt sich aus der prinzipiellen Abhängigkeit der bedingten Entropie von der Beobachtung. Die Beobachtung liegt allerdings erst a posteriori vor, d. h. nach Einstellung der Parameter der Kamera. Es gelang jedoch für eine große Klasse von Schätzverfahren, denen auch das Kalman–Filter angehört, d. h. für den Fall Gaußverteilter Zustandsgrößen, die bedingte Entropie a priori zu berechnen. Das wird möglich, da die bedingte Entropie nur von der Kovarianzmatrix des Schätzfehlers abhängt, die sich wiederum unabhängig von der Beobachtung berechnen lässt. Dadurch kann die bedingte Entropie a priori minimiert werden, bevor die Parameter des Sensor konkret eingestellt wurden. Bild 1 zeigt ein Beispiel für die bedingte Entropie in Abhängigkeit der Brennweiten zweier Kameras während der Objektverfolgung. Man erhält somit wiederum einen geschlossenen Kreislauf aus Sensorparameterauswahl, Sensordatenaufnahme und Zustandsschätzung. Im Gegensatz zum statischen Fall erfolgt von einem Zeitschritt zum nächsten ein Zustandsübergang gemäß der Beschreibung der Dynamik des Systems.

Der entwickelte Formalismus wurde im Teilprojekt B2 des Sonderforschungsbereichs 603 erfolgreich im Bereich der aktiven Objektverfolgung implementiert und getestet. Eine genauere Beschreibung findet man bei der Zusammenfassung des Teilprojekts.

Die Arbeit zur erscheinungsbasierten, statistischen Objekterkennung im Rahmen des von der DFG geförderten Graduiertenkollegs „Dreidimensionale Bildanalyse und -synthese“ wurde weiter fortgeführt. Bei dem hierbei verwendeten Ansatz werden zur Objekterkennung lokale Merkmale benutzt, die mit Hilfe der Wavelet–Multiskalen–Analyse aus den Bildintensitäten berechnet werden. Um das Verfahren robust gegenüber Kamerarauschen und Beleuchtungsänderungen zu machen, werden die Merkmale statistisch modelliert.

Ein wesentlicher Punkt bei der Erkennung, d. h. der Klassifikation und Lokalisation, von dreidimensionalen Objekten ist, dass sich bei wechselndem Blickwinkel meist sowohl das Aussehen

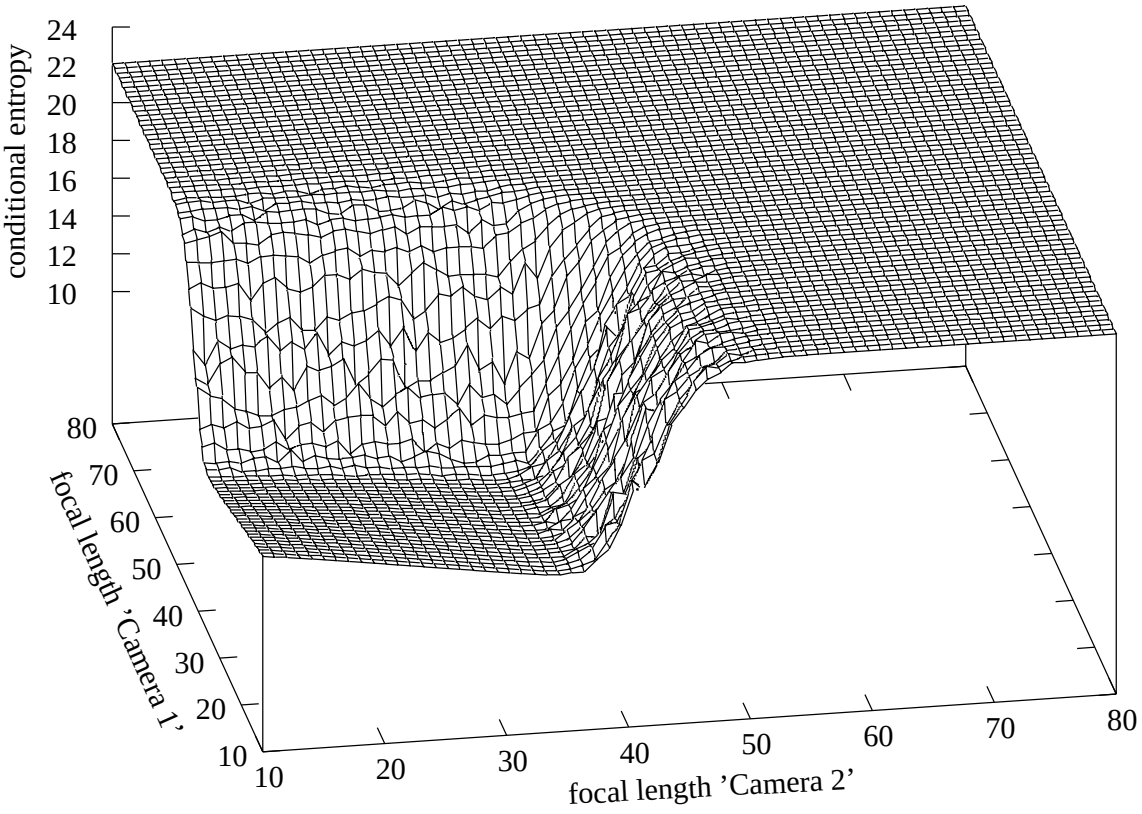

Bild 1: Bedingte Entropie in Abhängigkeit der Brennweiten der linken (x–axis) und der rechten Kamera (y–axis).

als auch die Größe des Objekts im Bild ändern. Ein Beispiel dafür ist in Bild **??** zu sehen. Bei den meisten Verfahren wird diese Größenänderung nicht berücksichtigt: es wird ein Objektfenster fester Größe definiert, und alle Bildpunkte/Merkmale innerhalb dieses Objektfenster werden dem Objekt zugewiesen. Dies hat zur Folge, dass je nach Größe des Objektfensters und gewähltem Blickwinkel Teile des Objektes nicht im Objektfenster liegen bzw. Hintergrundbereiche dem Objekt zugewiesen werden (siehe Bild **??**), was die Erkennung deutlich erschwert, zum Teil sogar unmöglich macht. Um diesen Fehler zu vermeiden, wird bei diesem Ansatz die Größe und Form des Objektfensters im Training gelernt und durch kontinuierliche Funktionen in Abhängigkeit vom Blickwinkel modelliert [29]. So gibt das Objektfenster selbst für Blickwinkel zwischen den Trainings-Blickpunkten die Größe des Objektes im Bild nahezu richtig wieder (siehe Bild **??**). Auf ähnliche Weise werden die Objekt-Merkmale selbst, d. h. das Aussehen des Objekts, als kontinuierliche Funktionen in Abhängigkeit vom Blickwinkel modelliert.

Bei realen Aufgabenstellungen befinden sich die Objekte meist nicht wie beim Training vor homogenen Hintergrund, sondern vor beliebigem, heterogenem Hintergrund; zudem sind sie oft partiell verdeckt. Deshalb wurden ein explizites Hintergrundmodell sowie eine Zuweisungsfunktion entwickelt, die in Abhängigkeit von den Dichtewerten die einzelnen Merkmalsvektoren in-

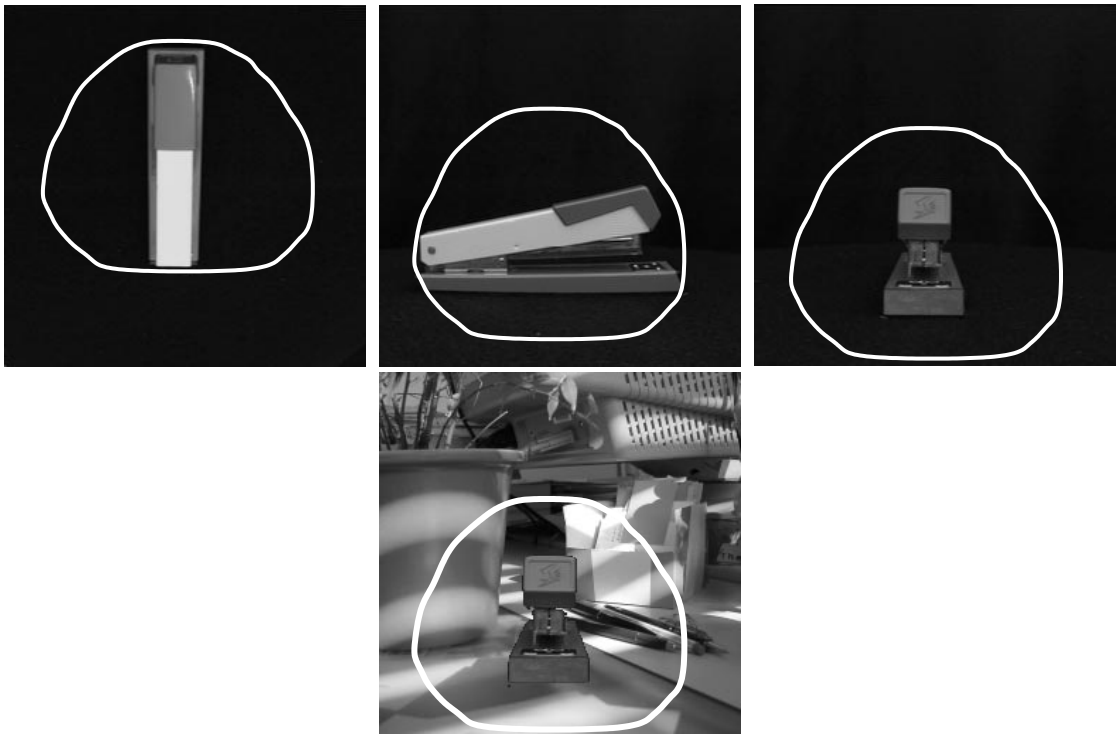

Bild 2: Verschiedene Blickwinkel für einen Hefter: sowohl Aussehen als auch Größe des Hefters im Bild variieren. Die Größe des Objektfensters, das den Hefter für alle Blickwinkel auf einer Halbkugel einschließt, ist eingezeichnet. Wie man sieht, ist das Objektfenster deutlich größer als das Objekt und schließt viele Hintergrundmerkmale ein

nerhalb des Objektfensters dem Objekt bzw. dem Hintergrund zuweist [28], wie z. B. Beispiel in Bild **??** zu sehen ist. Diese Zuweisungsfunktion ermöglicht auch die Erkennung mehrerer Objekte in einem Bild, wobei dann die einzelnen Merkmalsvektoren den verschiedenen Objekten bzw. dem Hintergrund zugeordnet werden.

Getestet wurde dieser statistische Ansatz beispielsweise auf der DIROKOL-Stichprobe [27] (von 13 Objekten wurden jeweils 3720 Bilder gleichmäßig verteilt über eine Halbkugel aufgenommen, wobei die Hälfte der Bilder zum Training verwendet wurde): selbst bei heterogenem Hintergrund lag die durchschnittliche Klassifikation- und Lokalisationsrate über 80%.

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 [**?**] belegen die Tragfähigkeit dieses Ansatzes.

In Abs. 4 findet sich eine ausführliche Darstellung in englischer Sprache. Im vergangenen Jahr wurden die Arbeiten auf dem Gebiet der wissensbasierten Bildverarbeitung im DFG–Projekt Strategie und Aktion als Lernziel der visuellen Exploration zu einem Abschluss gebracht. Der

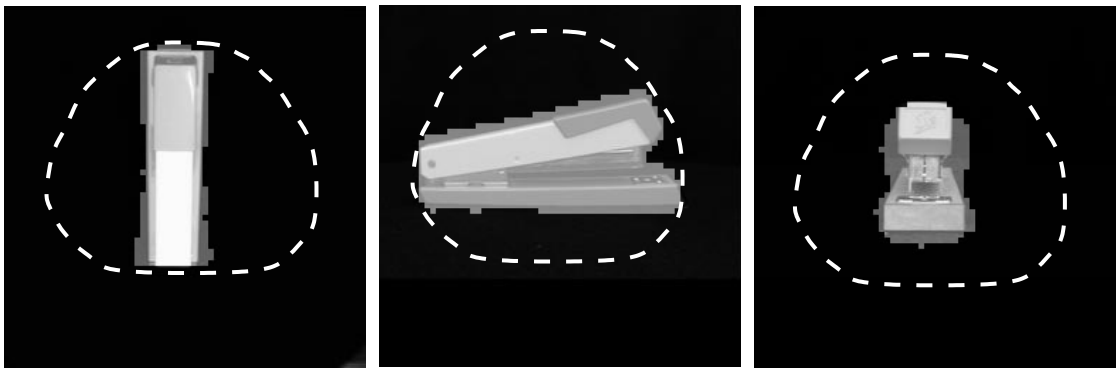

Bild 3: Die gleichen Blickwinkel für den Hefter wie in Bild **??**. Das feste Objektfenster aus Bild **??** ist gestrichelt eingezeichnet. Das trainierte, variable Objektfenster ist grau unterlegt eingezeichnet und umschließt das Objekt sehr eng.

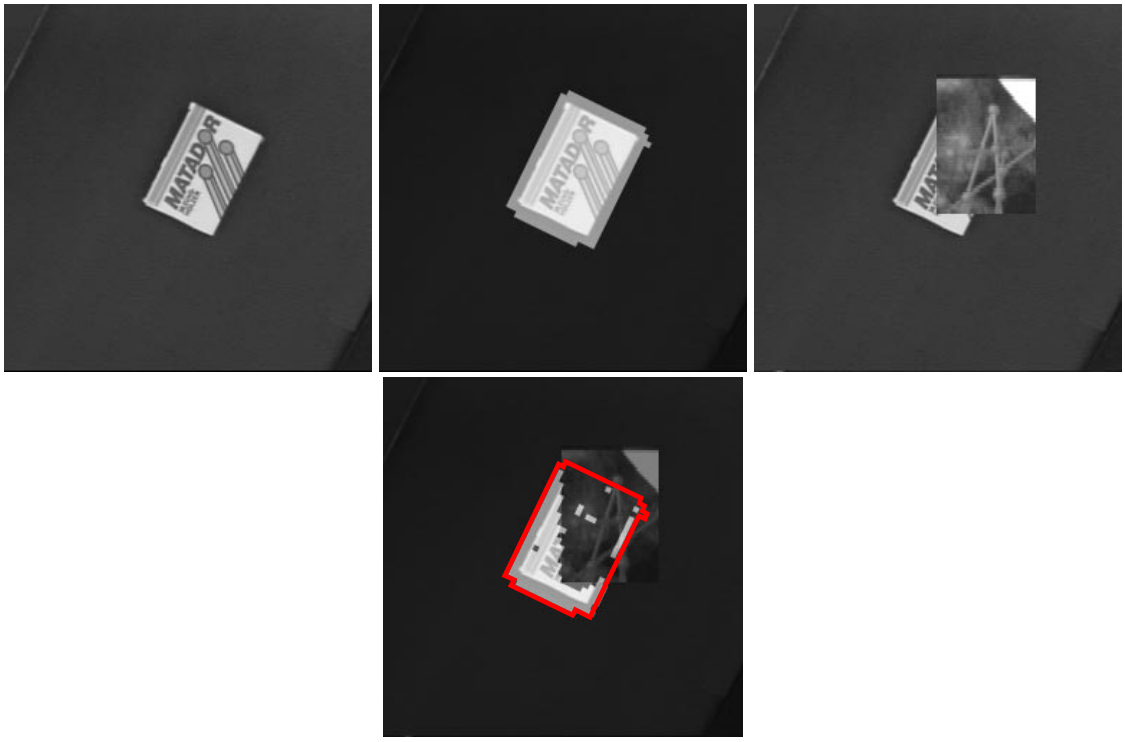

Bild 4: In den beiden Bildern links ist eine Streichholzschachtel und das trainierte, dazugehörige Objektfenster abgebildet. Wird ein Teil der Streichholzschachtel verdeckt (zweites Bild von rechts), so werden fast nur die Merkmalsvektoren der Streichholzschachtel zugeordnet, die nicht verdeckt sind (grau unterlegter Bereich im rechten Bild)

Schwerpunkt der Arbeiten in diesem Projekt lagen auf dem Lernen von Verarbeitungsstrategien zur Nutzung von Wissen, das in Form eines semantischen Netzes repräsentiert wird. Als Lernverfahren wurden Methoden aus dem Reinforcement Learning verwendet.

In Experimenten mit der eigens für das Projekt angeschafften 3-D-Laserkamera, die durch Projektion von Laserpulsen die Entfernung zwischen Kamera und einer Szene bestimmt, wurde die Messunsicherheit der Kamera ermittelt.

Ein weiterer Schwerpunkt der Forschungsarbeiten im Teilprojekt B2 des Sonderforschungsbereichs 603 sind Verfahren zur Objekterkennung in der Bildverarbeitung. Die Forschung dazu konzentriert sich in den letzten Jahren mehr und mehr auf aktive Verfahren. Im Gegensatz zu den passiven Ansätzen der letzten Jahrzehnte, wo eine Entscheidung auf einem einzelnen Bild basierte, verwenden aktive Techniken mehrere Bilder eines Objektes aus verschiedenen Ansichten zur Klassifikation und Lokalisation. In diesem Kontext müssen mehrere Aufgaben gelöst werden. Zum Einen, welche Blickrichtungen auf ein Objekt ausgewählt werden. Hierzu wurden bereits in den letzten Jahren erfolgreich neue Verfahren zur Ansichtenauswahl und Ansichtenplanung am Lehrstuhl für Mustererkennung entwickelt. Zum Anderen müssen natürlich auch alle erhaltenen Aufnahmen aus verschiedenen Blickrichtungen geeignet fusioniert werden, um ein Gesamtergebnis zu erhalten. Dieser Punkt wurde in diesem Jahr ausführlich untersucht und ein Verfahren entwickelt, das in der Lage ist, eine theoretisch fundierte Fusion beliebig vieler Bilder durchzuführen. Dieses Verfahren zur Sensordatenfusion baut auf dem bekannten Condensation-Algorithmus auf, der sich für die Fusion aus drei Gründen besonders eignet. Erstens erlaubt es der Condensation, multimodale Dichten zu behandeln, wie es für beliebige Objekte notwendig ist. Zweitens entsteht durch die Bewegung einer Kamera zwischen zwei Bildern Unsicherheit, da diese Bewegung immer rauschbehaftet ist. Diese Unsicherheiten kann der Condensation Algorithmus handhaben. Drittens muss für den Condensation keine Verteilung in geschlossener Form angegeben werden, was in unserem Fall nicht möglich wäre.

In Bild 3 ist die Arbeitsweise des Fusionsansatzes illustriert. Aufgabe hierbei ist es, vier Tassen zu unterscheiden, die auf einer Seite eine Ziffer “1” bzw. “2”, auf der anderen Seite einen Buchstaben “A” oder “B” tragen. Die farbigen Kreise zeigen die aktuell wahrscheinlichsten Lagen je Tasse an. Wie man erkennt, schränkt sich die mögliche Lage und Klasse nach jedem Bild anschaulich und nachvollziehbar ein. Vor der Aufnahme des ersten Bildes kann über das Objekt keine Aussage gemacht werden. Das erste Bild schränkt die Lage auf zwei Bereiche ein; eine Klassifikation ist noch nicht möglich. Die zweite Aufnahme erlaubt eine eindeutige Lageschätzung und die möglichen Objekte werden eingeschränkt. Das dritte Bild ermöglicht schließlich eine Klassifikation als Tasse “1A”.

Der Einsatzbereich der Fusion von Bildern ist aber nicht – wie man vielleicht vermuten könnte – auf mehrdeutige Objekte beschränkt. Auch bei Folgen von Aufnahmen beliebiger Objekte ist das Ergebnis eine sichere Klassifikation und genauere Lokalisation.

Im folgenden Jahr wird der Schwerpunkt der Forschungsarbeiten auf der Kombination der Sensordatenfusion mit der bereits realisierten Ansichtenplanung liegen. Das Ziel dabei ist, beliebig lange Ansichtenfolgen zu planen und die Ergebnisse Schritthaltend zu fusionieren.

Die Arbeit Untersuchung von neuronalen Netzen zur Objekterkennung und –lokalisation wurden zu einem Abschluss gebracht. Erkennung und Lokalisierung von 3D–Objekten durch einzelne 2D–Grauwertbilder wurde erfolgreich demonstriert.

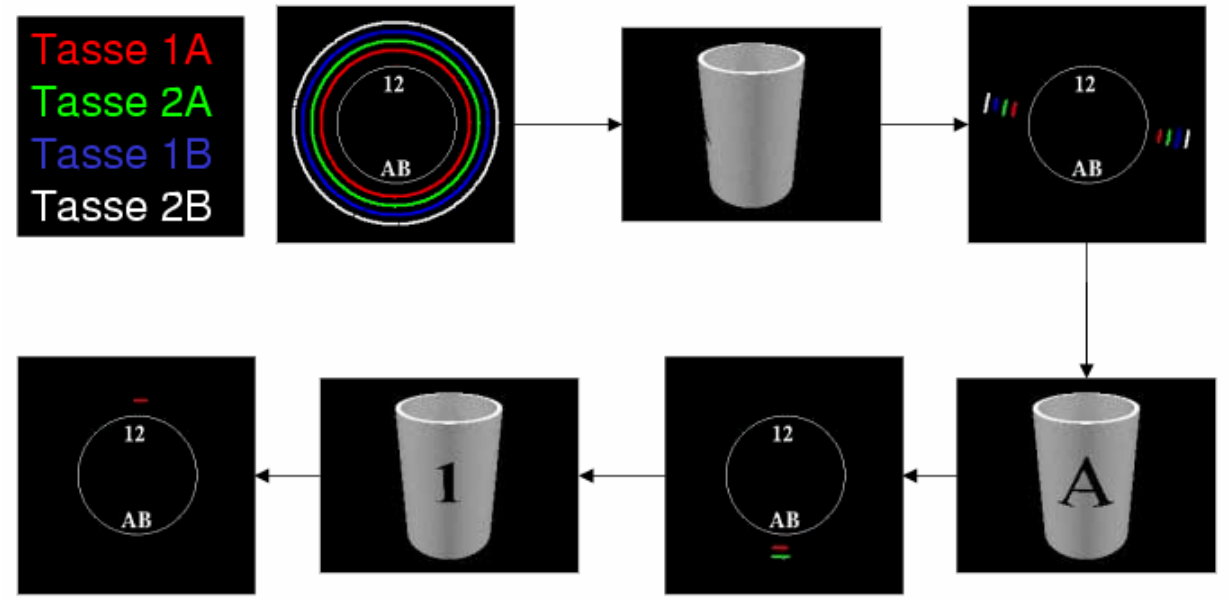

Bild 5: Beispiel zur Sensordatenfusion. Die farbigen Kreise zeigen die möglichen Lagen jeder der vier Tassen an. Durch Hinzunahme neuer Bilder verbessern sich Klassifikations- und Lokalisationsergebnisse.

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) war 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 4).

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.

Im Rechnersehen ist es für die Lösung von Schätzproblemen unabdingbar, dass man den entsprechenden Algorithmen die bestmöglichen Sensordaten zur Verfügung stellt. Die Qualität der Sensordaten wird durch die eingestellten Kameraparameter maßgeblich beeinflusst. So existiert beispielsweise bei der Objektverfolgung, d. h. bei der dynamischen Zustandsschätzung, ein Konflikt bei der Wahl der Brennweite. Eine große Brennweite ermöglicht eine genauere Lokalisation des Objekts, da es größer auf die Bildebene projiziert wird. Dafür nimmt man in Kauf, dass das Objekt durch eine nicht vorhergesehene Bewegung aus dem Sichtfeld verschwindet. Dies würde durch Wahl einer kleineren Brennweite wieder kompensiert werden, wodurch allerdings die Akkuratheit der Lokalisierung geringer wäre. Die optimale, an die jeweilige Verfolgungssituation angepasste Auflösung dieses Konflikts ist eine Aufgabe des Teilprojekts B2 im SFB 603. Im Rahmen von informationstheoretischen Untersuchungen wurde herausgefunden, dass die zu erwartende Unsicherheit in der Zustandsschätzung, ausgedrückt durch die bedingte

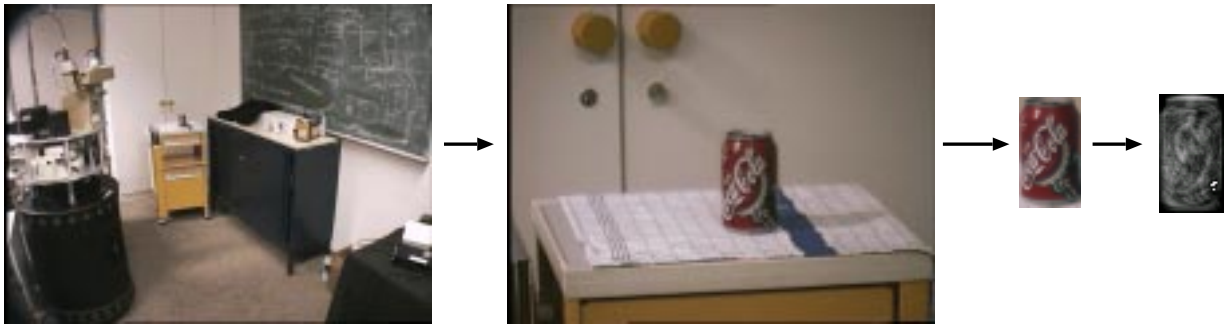

Bild 6: Objektmodellgenerierung, Szenenexploration und Erkenunnungsprozess: Modellgenerierung aus Trainingsdaten und resultierende Eigenraumdarstellung (unten); Tischsuche im Übersichtsbild, Objektlokalisation auf dem Tisch, extrahiertes Objekt, Rekonstruktion des erkannten Objekts (oben).

Entropie zwischen Zustand und Beobachtung, ein geeignetes Gütemaß für die Kameraparameterwahl darstellt. Für jeden Zeitschritt gilt es daher, vor dem Durchführen der Beobachtung, diejenigen Kameraparameter zu bestimmen, die das Gütemaß minimieren. Bei der Optimierung muss berücksichtigt werden, dass nicht für jede denkbare Parameterwahl eine Beobachtung erfolgen kann. Für den Einsatz eines Erweiterten Kalman Filters zur Zustandsschätzung konnte eine Lösung für dieses Problem gefunden werden. In Simulationen und Echtzeitexperimenten wurde durch aktive Auswahl der Brennweiten eine Reduktion des mittleren Schätzfehlers von bis zu 43% erreicht, verglichen mit den Ergebnissen bei fest eingestellten Brennweiten. Beispielhaft zeigt Bild 5 Bilder einer Verfolgungssequenz einer Dose bei bewegter Kamera mit aktiver automatischer Brennweitensteuerung. Die Übertragung des entwickelten Lösungsansatzes auf den allgemeinen Fall multimodaler Verteilungen, handhabbar beispielsweise mit dem Condensation-Algorithmus, ist Gegenstand weiterer Forschung.

Zur Merkmalsgewinnung für die Objektverfolgung wurde auf ein von Hager und Belhumeur vorgeschlagenes regionenbasiertes Verfahren zurückgegriffen, dass sich für die Echtzeitverfolgung unterschiedlichster Objekte eignet. Zur Steigerung der Leistungsfähigkeit des Verfahrens wurde es um eine hierarchische Komponente erweitert. Zusätzlich zu diesem Ansatz wird in Zusammenarbeit mit dem Teilprojekt C2 der Einsatz von Lichtfeldern als Objektmodell in der Objektverfolgung untersucht werden.

3.3 Bildbasierte Modellierung und erweiterte Realität

Das in Zusammenarbeit mit dem Lehrstuhl für Graphische Datenverarbeitung laufende Teilprojekt C2 des Sonderforschungsbereiches 603 (SFB 603) beschäftigt sich mit der Rekonstruktion von Szenenmodellen aus Videosequenzen und der anschließenden Generierung von Lichtfeldern der Szenen. Nachdem seit dem Beginn des Teilprojektes bereits zufriedenstellende Lösungen für die einzelnen Prozessschritte der Punktverfolgung, Kamerakalibrierung und Visualisierung der gewonnenen Lichtfelder erarbeitet wurden, wurde im vergangenen Jahr das Hauptaugen-

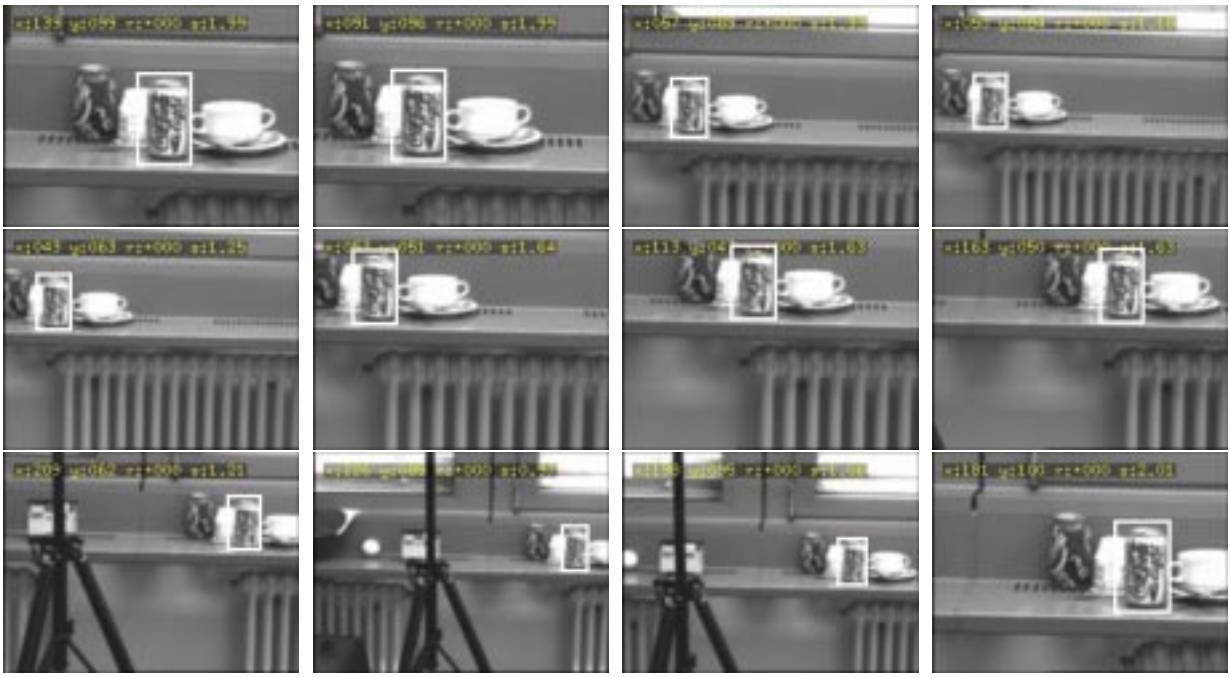

Bild 7: Bilder einer Verfolgungssequenz bei kreisförmig bewegter Kamera mit automatischer dynamischer Brennweitenadaption.

merk auf die Verbesserung der Punktkorrespondenzerstellung und die Synthese der einzelnen Bearbeitungsschritte gerichtet. Außerdem wurden verschiedene Anwendungen von Lichtfeldern betrachtet.

Eine Vergrößerung der Anzahl der Bilder, in denen ein Merkmal verfolgt werden konnte, wurde erreicht, indem die Bildsequenzen nicht mehr nur sequenziell in einer Richtung bearbeitet wurden, sondern Punkte auch rückwärts durch die Sequenz verfolgt wurden. Nach einer ersten Kalibrierung wurde anschließend ein Ansichtennetz erstellt und die bereits gefundenen Merkmale nochmals über Ansichten aus ähnlichen Blickwinkeln verfolgt. Bild 6 zeigt links, dargestellt als rote Verbindungslinien zwischen den Kamerazentren, die normalen Bildnachbarschaften in einer Videosequenz, und rechts die Erweiterung dieser Nachbarschaften auf nahe aneinander grenzende Kameras. Die Position der Kameras im Raum wird hier durch graue Pyramiden mit jeweils dem Zentrum der Kamera in der Spitze verdeutlicht.

Ein wichtiger Schritt zur Verbesserung von Lichtfeldern ist die Rückführung von Informationen aus dem Rendering, z. B. über die Güte der Darstellung oder die Dichte der Tiefenkarten, in frühere Schritte der Bearbeitung. Aus diesem Grund wurde begonnen, die Arbeiten des Lehrstuhls für Graphische Datenverarbeitung und des Lehrstuhls für Mustererkennung in diesem Bereich in einem gemeinsamen Softwarepaket zu integrieren, das auf einer gemeinsamen Datenbasis aufsetzt. Diese umfangreiche Umstellung ist noch nicht beendet und führte bei einzelnen Werkzeugen aber schon zu einer deutlichen Vereinfachung des Programmieraufwandes.

Des Weiteren wurde untersucht, wie Lichtfelder etwa in der Objekterkennung oder der Aug-

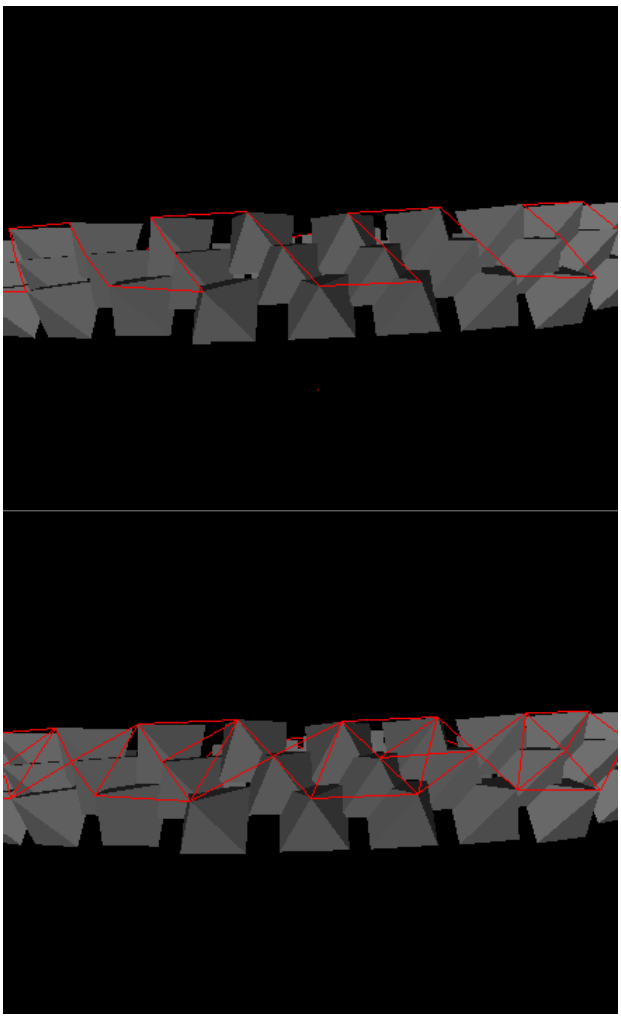

Bild 8: Normale Bildnachbarschaften in einer Bildsequenz (links), Erweiterung der Nachbarschaften (rechts).

mented Reality eingesetzt werden können. Dazu wurde zum einen untersucht, ob aus einem Lichtfeld gewonnene Ansichten zum Training eines Objekterkenners und zur Lageschätzung geeignet sind. Laufende Arbeiten beschäftigen sich mit dem Einsatz von Lichtfeldern in der Augmented Reality.

Im Bereich der erweiterten Realität (Augmented Reality) wurde im vergangenen Jahr an der Platzierung von rechnergenerierten Objekten in realen Szenen mit Hilfe eines echten Würfels mit bekannter Geometrie und Farbe gearbeitet [31, 32]. Durch den Einsatz von nichtlinearen Optimierungsmethoden konnte die Genauigkeit der Kalibrierung und damit der Positionierung eines virtuellen Objekts verbessert werden. Zur Optimierung der extrinisischen Kameraparameter wurde eine Parametrisierung der Kamerarotation vorgestellt, die auf der Verwendung von Quaternionen beruht [30]. Weiterhin wurde an der Detektion und Eliminierung von Ausreißern

in den Kalibrierergebnissen gearbeitet.

Die weitere Arbeit soll im Bereich der Echtzeit-Stereobildverarbeitung liegen; erste Ergebnisse hierzu liegen bereits vor und können zusammen mit Videosequenzen aus dem o. g. System von unserer WWW-Seite heruntergeladen werden.

Ein Beispiel hierfür zeigt Bild 7.

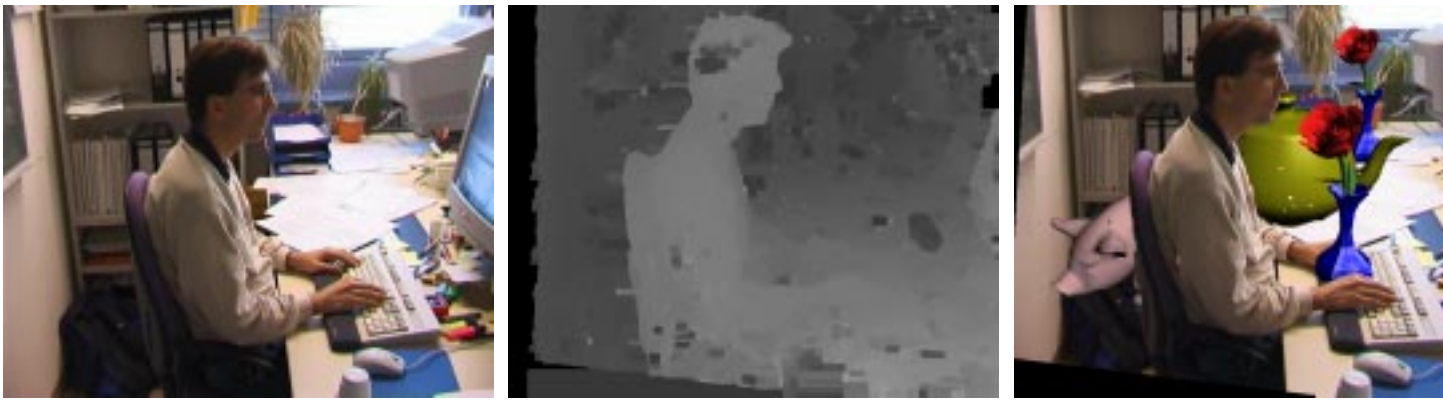

Bild 9: Original (links), Disparitätskarte (Mitte) und augmentierte Szene (rechts)

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 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 wird 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. Dies stellt die Grundlage dar für eine automatische Klassifikation von Retina-Tomograph-Bildern hinsichtlich der Verdachtsdiagnose, was an über ca. 400 Bildern aus dem Erlanger Glaukomregister demonstriert wurde.

Gegenstand von Teilprojekt B3 des Sonderforschungsbereichs 603 (SFB 603). ist die automatische Diagnose von Gesichtslähmungen. Untersuchungen, die hierzu durchgeführt wurden, bezogen sich hauptsächlich auf die Bereiche Lokalisation und Verfolgung von Gesichtern und Gesichtsmerkmalen und der Extraktion von Information zur Diagnose und Graduierung einer Gesichtslähmung.

Zur Lokalisation und Verfolgung von Gesichtern und Gesichtsmerkmalen werden *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

Bild 10: a) Lokalisation und Verfolgung von Gesichtern und Gesichtsmerkmalen; b) Differenzbild während der Ausführung einer symmetrischen Gesichtsbewegung; c) Differenzbild während der Ausführung einer asymmetrischen Gesichtsbewegung.

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-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 Bild8a 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 bearbeitet wurde, ist die Extraktion und die Nutzung der Information aus den aufgenommenen Patientenbildern. Zwischen aufeinanderfolgenden Bildern der aufgenommenen Sequenzen werden Differenzbilder bestimmt. In Bild8b und Bild8c sind Beispiele für diese Differenzbilder von Gesichtern bei symmetrischer (b) und asymmetrischen (c) Bewegungen. Durch Integration der absoluten Differenzen und durch Vergleich zwischen linker und rechter Gesichtshälfte wird ein 12-dimensionaler Merkmalsvektor aus den Bildsequenzen extrahiert, der schließlich zur Graduierung eingesetzt wird. Das Gesamtsytem wurde an der HNO-Klinik Erlangen an mehr als 100 Personen getestet und validiert.

Im Teilprojekt B6 „Rechnergestützte Endoskopie des Bauchraums“, des Sonderforschungsbereichs 603 wurden im Jahr 2001 in Zusammenarbeit mit der Chirurgischen Universitätsklinik die grundlegende Arbeiten zur Einführung einer Rechnerunterstützung in der mikroinvasiven Chirugie durchgeführt.

Bisher wurde der Lichtfeldansatz aus Teilprojekt C2 benutzt um medizinische Lichtfelder aus endoskopischen Bildern zu erzeugen. Erste Lichtfelder konnten auch ohne den vorgesehenen Sensor zur Positionsbestimmung hergestellt werden. Um allerdings die Qualität dieser Lichtfelder, die bei weitem nicht die Qualität der Lichtfelder aus Teilprojekt C2 erreichen, zu erhöhen, ist ein Positionssensor unabdingbar. Dazu steht seit kurzem ein Roboterarm (AESOP 3000) der Chirurgischen Universitätsklinik zur Verfügung. Dieser führt das Endoskop und ermöglicht dadurch unter anderem die Positions- und Orientierungsbestimmung des Endoskops.

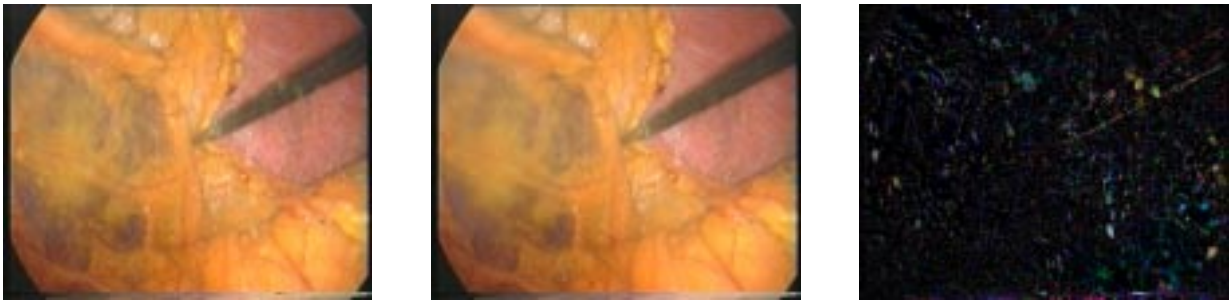

Bild 11: Originalbild (links), zeitlich gefiltertes Bild (mitte), Differenzbild | links - mitte | (rechts).

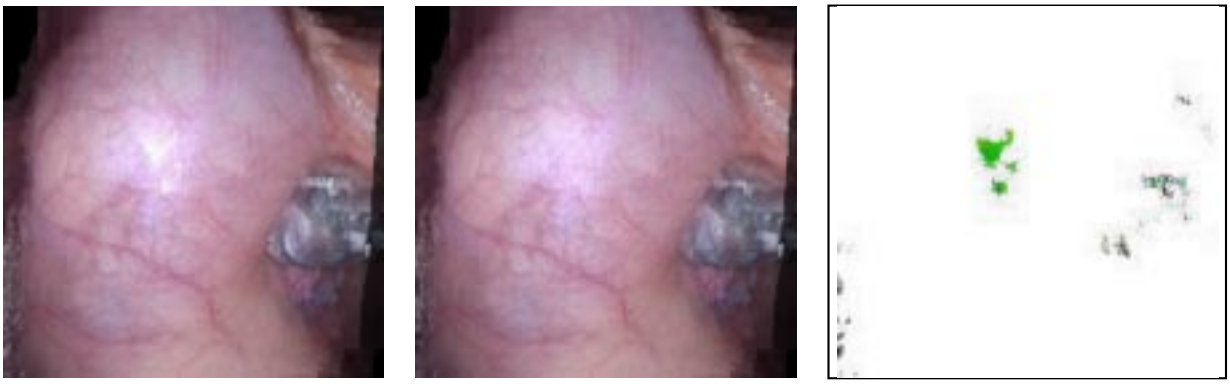

Bild 12: Gerendertes Bild eines Lichtfeldes ohne Glanzlichtsubstitution (links), Gerendertes Bild eines Lichtfeldes mit Glanzlichtsubstitution (mitte), Differenzbild |links - mitte|, invertiert und neu skaliert: Pixelwerte > 100 wurden auf 255 gesetzt und Pixelwerte $\in [0, 100]$ wurden linear in den Bereich $[0, 255]$ transformiert (rechts).

Zur Bildverbesserung in Echtzeit wurden verschiedene Verfahren untersucht: zeitliche und räumliche Filterung, Entzerrung und Farbrotation [39]. Bis auf die räumliche Filterung führen alle genannten Verfahren zu einer Verbesserung der Bildqualität. Zeitliche Medianfilterung bis zur Größe 5 ließ sich bereits in Echtzeit realisieren [40]. Die anderen Filtermethoden liegen im Bereich von ca. einer halben Sekunde pro Bild. Es hat sich herausgestellt, dass vor allem bei der zeitlichen Filterung ganze Bildsequenzen evaluiert werden sollten und keine Einzelbilder. Dies ist natürlich auch für alle anderen Verfahren zu überlegen, da letztendlich üblicherweise eine Sequenz betrachtet wird und nur selten Einzelbilder.

Unter Zuhilfenahme von Lichtfeldern konnten außerdem Glanzlichter (Glanzlichtpixel) durch „echte“ Farbpixel ersetzt werden.

Beispielbilder sind in Bild 9 und Bild 10 dargestellt.

4 Statistische Modellierung von Daten

(R. Deventer, U. Ohler)

Der Sonderforschungsbereich SFB 396 „Robuste, verkürzte Prozessketten für flächige Leichtbauteile" beschäftigt sich im Wesentlichen mit der Optimierung von Prozessketten. Eine Optimierung der Prozessketten kann dabei z. B. durch Zusammenlegen mehrerer Teilprozesse oder durch eine Vergrößerung des Prozessfensters erreicht werden. Der Beitrag des Teilprojektes C1 besteht in der stochastischen Modellierung von Prozessketten, die unter anderem im Qualitätsmanagement und in der Regelung angewendet wird. Die stochastische Modellierung basiert dabei auf Bayesnetzen. Anschaulich ist ein Bayesnetz 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.

Der Kernpunkt der disjährigen Arbeiten stellt der Entwurf eines bayesnetzbasierten Reglers dar. Dank der vorhandenen Lernalgorithmen für Bayesnetze kann sich dieser Regler selbständig an eine wechselnde Umgebung anpassen.

Um eine möglichst breite Anwendbarkeit zu garantieren, wurde dabei von linearen, dynamischen Prozessen ausgegangen. Diese lassen sich durch Differentialgleichungen n-ter Ordnung beschreiben. Durch Einführen sogenannter Zustandsgrößen kann diese Differentialgleichung n-ter Ordnung in n Differentialgleichungen erster Ordnung überführt werden. Diese Beschreibungsform, die sogenannte Zustandsraumbeschreibung, führt dabei zu einem Netz, deren Parameter einerseits durch Training oder aus einem Vergleich mit einem Kalman–Filter gewonnen werden können. Dieses Modell kann nun zur Berechnung geeigneter Eingabegrößen verwendet werden, indem alte Ein- und Ausgaben und der Sollwert w als Evidenz verwendet werden und die geeignete Eingabe per Marginalisierung berechnet wird.

Alternativ kann von einer Differenzengleichung als Modellierung ausgegangen werden. Dabei werden die Informationen, die in den Zuständen gespeichert sind, durch Rückgriff auf alte Eingaben u und alte Modellausgaben y ersetzt. Die beobachtete Ausgabe q ergibt sich als Summe aus der Störgröße z und y. Das Modell, das in Bild 11 dargestellt ist, verzicht auf die Markov–Annahme, hat aber den Vorteil, dass wesentlich weniger verborgenen Knoten verwendet werden. Dies führt einerseits zu einem effizienteren Training, z. B. werden wesentlich weniger Iterationen benötigt, und andererseits zu einer verbesserten Regelungsgüte- bzw. Stabilität. Dieser Ansatz wurde bezüglich seines Führungs- und Störverhaltens mit Regelstrecken zweiter Ordnung getestet. Dabei wurde ein relativer Fehler in Bezug auf den Sollwert von weniger als einem Prozent erreicht.

Im Jahr 2002 soll das geschilderte Prinzip auch auf nichtlineare Regelstrecken angewendet werden. Um auch hierbei eine allgemeine Anwendbarkeit zu garantieren werden zuerst prototypisch nichtlineare Bauelemente modelliert, die danach zu komplexen Prozessenmodellen zusammengesetzt werden können.

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

Das vom Boehringer Ingelheim Fonds geförderte Bioinformatik-Projekt „Statistische Model-

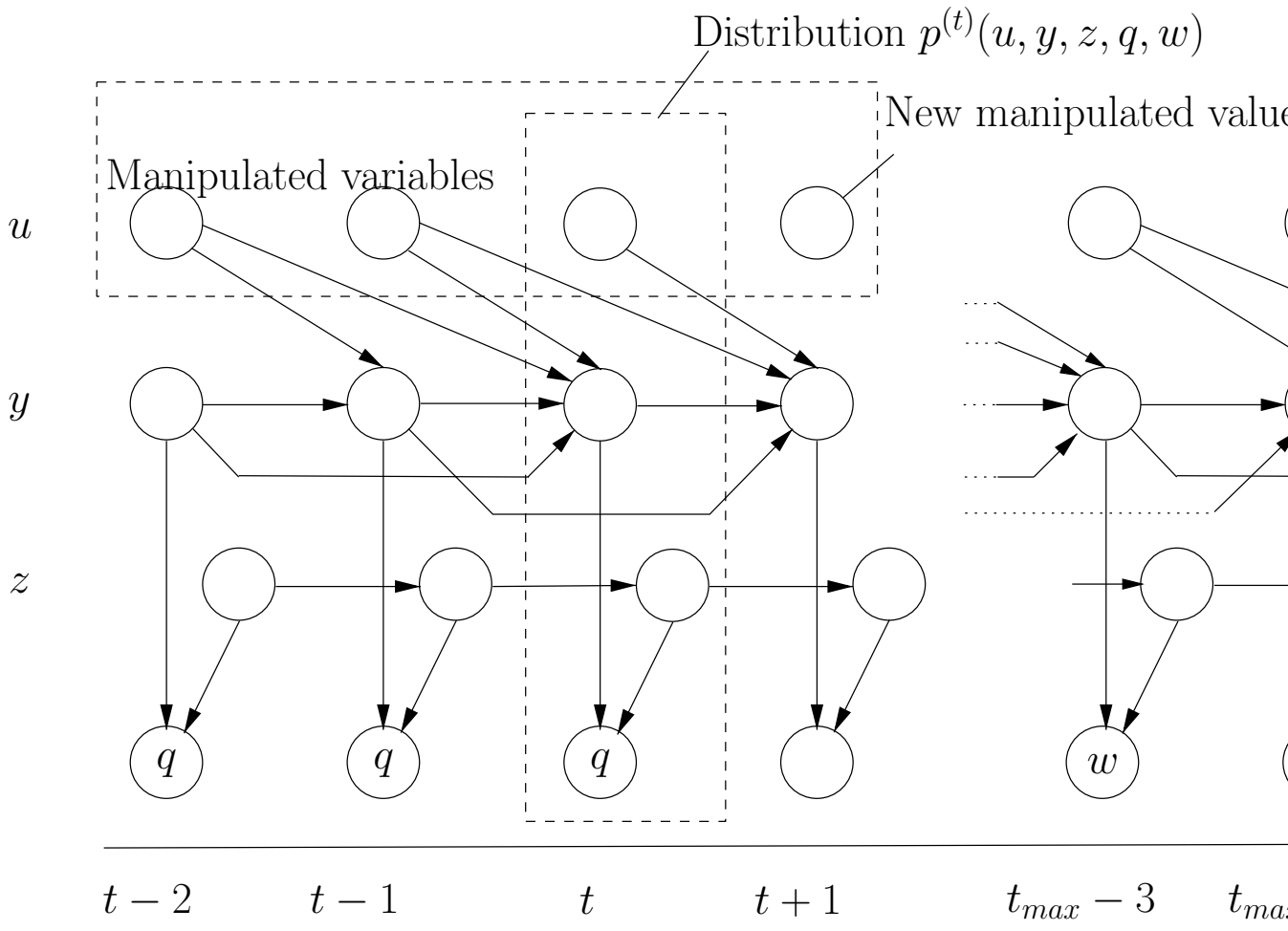

Bild 13: Aufbau eines bayesnetzbasierten Reglers

lierung, Lokalisierung und Analyse regulatorischer DNA–Sequenzen“ wurde im Jahr 2001 abgeschlossen. Im Rahmen des Projektes wurde McPromoter entwickelt, ein System zur Annotierung von Promotoren in DNA-Sequenzen.

Promotoren sind den gencodierenden Abschnitten der DNA vorgelagert und dienen gewissermaßen als Schalter der Gene, um deren differenzierte Regulation zu ermöglichen. Sie weisen daher eine komplexe und oft sehr variable Struktur auf.

Bislang wurden Promotorsequenzen durch ein stochastisches Segmentmodell mit sechs Zuständen repräsentiert. Ein neuronales Netz verarbeitete die Ausgabe dieses Modells zusammen mit der Bewertung eines Hintergrundmodells und traf eine Entscheidung, ob eine vorliegende Teilsequenz einem Promotor entsprach oder nicht. Als neue Komponente ist nun die Modellierung der Struktur der DNA in Promotorregionen hinzu gekommen [25]. Verschiedene Eigenschaften

wie Biegsamkeit oder Winkel werden mit Gauß-Dichteverteilungen gemäß den sechs Zuständen modelliert. Ein erweitertes neuronales Netz nimmt schließlich diese Wahrscheinlichkeiten parallel zu den bisherigen Sequenzwahrscheinlichkeiten als Eingabe. Eine Kombination mehrerer Struktureigenschaften erfolgt mittels Hauptachsentransformation (principal component analysis, PCA).

Das abschließende Gesamtsystem ist in Bild 12 dargestellt. McPromoter wird vor allem zur Annotierung des gesamten Drosophila-Genomes (Kooperation mit dem Berkeley Drosophila Genome Project) wie auch einiger viraler Genome eingesetzt. Ein eigener Server wurde unterdessen eingerichtet und dient zur Bearbeitung von etwa 20 Anfragen pro Tag.

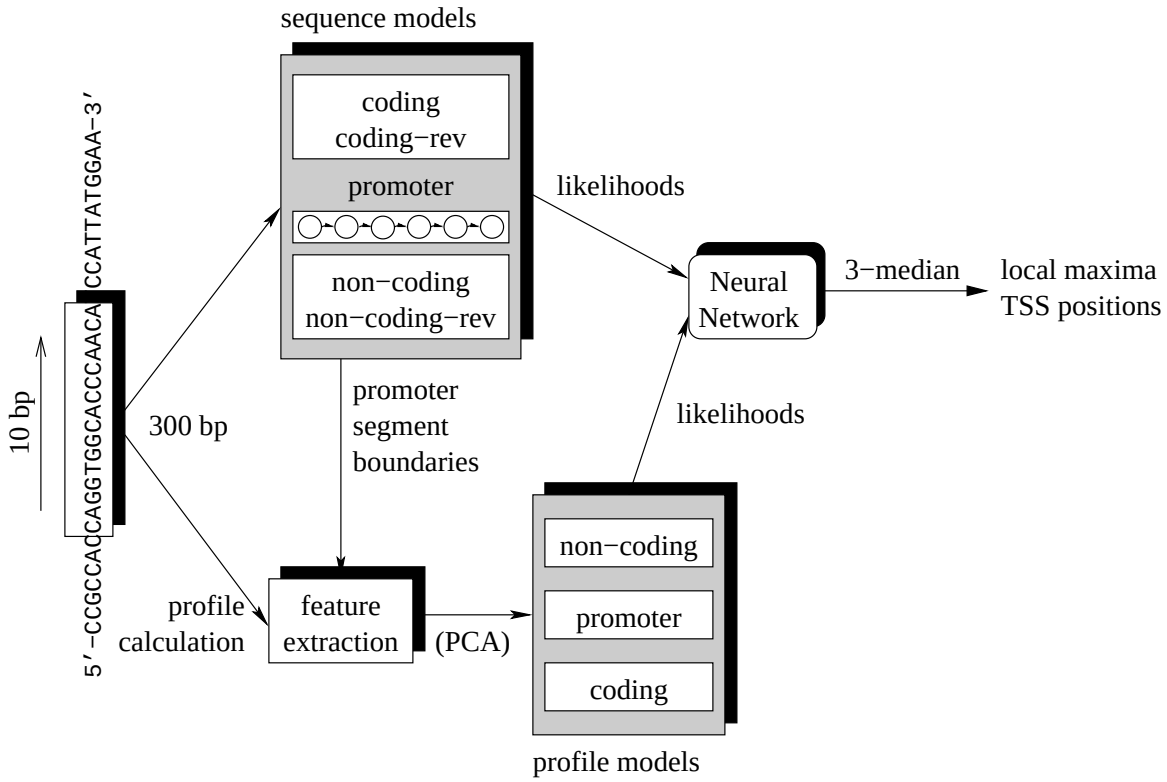

Bild 14: Das McPromoter-System zur Identifizierung eukaryontischer Promotoren in DNA-Sequenzen. Ein Fenster der Größe 300 dient als Eingabe und wird von einem Promotoren- und einem Hintergrundmodell auf Sequenz- und Struktueigenschaften der DNA untersucht. Deren Bewertungen werden schließlich mit einem künstlichen neuronalen Netz kombiniert.

4.1 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 Probability criterion. Such a model should take into account the contextual information provided by the scene where the object is to be found, as well as the informations characterizing the individual object.

We recently developed a new strategy for probabilistic modeling, consisting of a fully connected Markov Random field that integrates results of statistical physics of disordered systems with Gibbs probability distributions via nonlinear kernel mapping. The resulting model, that we call Spin Glass-Markov Random Field, (SG-MRF) overcomes the classical modeling problem of Markov Random Fields for irregular sites. Moreover, it allows to use the power of kernel functions in a probabilistic framework.

We first check the correctness of our model for appearance-based object recognition: we ran experiments on several databases, using different kind of features, and comparing the recognition results with other classifiers. In all the experiments we ran, SG-MRF gave the best performance. We then moved on, testing the robustness of SG-MRF to noise, partial occlusion and heterogeneous background. Again, the results confirmed the effectiveness of our new model.

First results on scene recognition show that the model is extensible to the probabilistic model of a scene, taking in account the contextual information that every scene gives regarding the object contained in it. The approach we take for scene modeling is holistic, as to say we consider a scene as a whole and not as composed as the sum of all its components elements.

Future work will be concentrated on further experiments for scene modeling and recognition of objects in different environments.

5 Sprachverstehen

Leitung: E. Nöth
(H. Adelhardt, A. Batliner, W. Fentze, C. Frank, M. Levit, R. Shi, G. Stemmer, V. Zeißler)

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 mit den zwei Anwendungsbereichen der Sprachverarbeitungsforschung am Lehrstuhl: Kinoauskunft mit dem System *Fränki* und Entwicklung des multimodalen Dialogsystems *SmartKom*. Darüberhinaus wurden Arbeiten zur flachen Inhaltsanalyse von sprachlichen Äußerungen durchgeführt.

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 *Fränk*ische *Ki*nosystem, 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. Die Firma Sympalog erhielt den von der EU geförderten IST-Preis 2001 für die am Lehrstuhl im Laufe der letzten 15 Jahre entwickelte Dialog–Technologie. Seit sechs Jahren wird der Preis von der Europäischen Kommission und Euro-CASE (European Council of Applied Sciences and Engineering) an innovationsstarke Unternehmen verliehen. In den Vorjahren zählten Unternehmen wie Nokia zu den Preisträgern.

Im vom BMBF geförderten Projekt *SmartKom* werden Konzepte für die Entwicklung völlig neuartiger Formen der Mensch-Technik-Interaktion erprobt. Diese Konzepte sollen 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

Das Spracherkennungssystem des Lehrstuhls wurde im Berichtsjahr an mehreren Punkten erweitert und verbessert.

Wie wohl fast alle Mustererkennungssysteme ist es modular aufgebaut. Aus einem Sprachsignal werden zuerst Merkmalvektoren extrahiert, die möglichst viel von der für die Erkennung wesentlichen Information enthalten sollen. Die akustischen Eigenschaften der gesprochenen Laute werden durch statistische Modelle, die Hidden Markov Modelle (HMM) repräsentiert. In der Erkennungsphase sucht ein Dekodierungsalgorithmus mit den HMM und einem linguistischen Modell nach der Wortfolge, die am wahrscheinlichsten gesprochen wurde. An allen genannten Bereichen des Spracherkenners wurden im Berichtsjahr Veränderungen vorgenommen.

Für das Modul zur Merkmalsextraktion wurden im Rahmen einer Studienarbeit neue dynamische Merkmale entwickelt, die im Gegensatz zu bisher nicht auf einem Zeitfenster bestimmter Größe, sondern mit mehreren unterschiedlichen Zeitauflösungen berechnet werden. Damit die Anzahl der Merkmale konstant bleibt, werden die Merkmale aus den unterschiedlichen Auflösungen mit einer Karhunen-Loève-Transformation (KLT bzw. PCA) in der Dimension reduziert. Dieses Vorgehen ist in Bild 5.1 dargestellt. Der Ansatz reduzierte die Wortfehlerrate signifikant. Eine weitere Verbesserung konnte erzielt werden, in dem die Merkmalstransformation in die Ausgabedichten der HMM integriert wurde. Das wurde durch die Verwendung von probabilistischen PCA (PPCA) Dichten ermöglicht, die die bisher verwendeten Normalverteilungsdichten ersetzen.

Auf der Modellebene gab es bisher die Schwierigkeit, dass bei der Erkennung von fremdsprachlichen Wörtern, wie z.B. englische Filmtitel in der Kinoauskunft keine adäquaten Modelle für die fremdsprachlichen Laute existierten. Einfaches Trainieren von HMM für alle englischen Laute mit den zugehörigen Kontexten ist nicht möglich, da nicht genug Aussprachebeispiele englischer Laute von deutschen Muttersprachlern vorhanden sind. Abhilfe schafft ein Ansatz, bei dem gemessen wird, welche englischen Laute von den deutschen Muttersprachlern in ähnlicher Weise wie bestimmte deutsche Laute ausgesprochen werden. Ein Verfahren dazu basiert auf dem datengetriebenen Vereinigen ähnlicher Laute, ein anderes auf den Eigenschaften des standardisierten phonetischen Alphabets IPA. Das Ergebnis beider Methoden ist ein reduziertes Lautinventar, das fremdsprachliche Laute enthält und auch bei wenig Trainingsdaten robust geschätzt werden kann.

Jedes Mustererkennungssystem macht Fehler. Für zahlreiche zukünftige Erweiterungen wie z.B. die unüberwachte Adaption der Modellparameter an einen Sprecher, oder spezielle Dialogstrategien ist es eine wichtige Voraussetzung, dass zusammen mit der am wahrscheinlichsten gesprochenen Wortkette zusätzlich noch eine Konfidenzbewertung der Worthypothesen mit ausgegeben wird, die angibt, mit welcher Wahrscheinlichkeit das Ergebnis der Dekodierung fehlerhaft ist. In einer Studienarbeit konnte ein Modul zur Konfidenzberechnung entwickelt werden, das Worthypothesen in eine der beiden Klassen 'korrekt' oder 'falsch' einordnet. In ca. 75% aller Fälle ist diese Klassifikation richtig. Eine erste Anwendung des Moduls ist für eine noch laufende Diplomarbeit geplant, die sich mit der schnellen Adaption der Ausgabedichten von HMM an den aktuellen Sprecher beschäftigt.

5.2 Das SmartKom–Projekt

Im Bereich der Spracherkennung wird am Lehrstuhl im Rahmen des Verbundprojekts *SmartKom* an multimodalen Dialogsystemen geforscht. Eine der Eingabemodalitäten des zu entwickelnden Systems, die Sprache, wird verarbeitet, um prosodische Information zu gewinnen. Mit Hilfe der prosodischen Information werden in der Benutzeräußerung z.B. prosodische Grenzen zur Bestimmung der Satzstruktur gewonnen. Außerdem wird die Prosodie verwendet, um auf den Zustand des Benutzers, der mit dem System kommuniziert, zu schließen (UserState). Neben den "klassischen" emotionalen Zuständen wie Ärger/Wut und Freude sind auch Zustände wie "Verwirrtheit/Ratlosigkeit" in der Mensch-Maschine-Kommunikation von Bedeutung.

Die Bild 5.2 zeigt die Einbindung des Prosodie-Moduls in die Systemumgebung und Datenströme zwischen den Modulen, welche die Prosodie beeinflussen.

Das Prosodie-Modul wurde zur Erkennung der syntaktischen Grenzen erweitert. Die zur Kommunikation nötige Schnittstelle wurde entworfen und implementiert. Die Ergebnisse der Klassifikationsexperimente wurden mit den bekannten Resultaten aus dem Vorgängerprojekt Verbmobil verglichen. Die dort erzielten Erkennungsraten beim Zwei-Klassen-Problem von 95 % wurden mit Erkennungsraten von 91 % hier noch nicht erreicht. Die Ursache dafür wird im geringeren Umfang der vorhandenen Datenmenge vermutet. Durch eine Erweiterung der Stichprobe mit Verbmobil-Daten konnte die Erkennungsrate jedoch nicht verbessert werden: Leave-One-Out-Experimente mit unterschiedlichen Anteilen von *SmartKom*-Daten in den Trainings- und Testdaten zeigten, dass bei höherem Anteil von *SmartKom*-Daten die Erkennungsrate steigt.

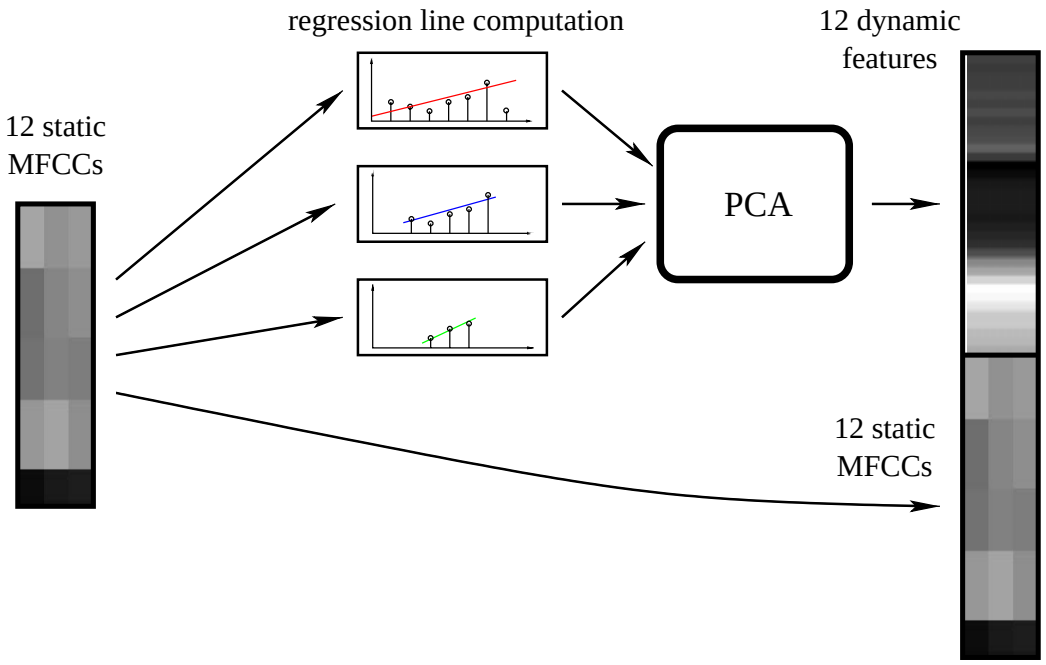

Bild 15: Merkmalsberechnung mit unterschiedlichen Zeitauflösungen für die dynamischen Merkmale.

Als ein mögliches Problem bei der turnbasierten Klassifikation des Benutzerzustands kann der Fall auftreten, dass nur ein Teil der betroffenen Äußerung emotional markiert ist. Dadurch wird dann die gesamte Turnklassifikation in hohem Maße unzuverlässig. Eine Lösung des Problems besteht darin, dass man künftig die Äußerung bereits vor der Klassifikation in Phrasierungseinheiten bzw. Dialogaktabschnitte aufteilt und danach die einzelnen Segmente klassifiziert.

Das Modul wurde bereits zur Erkennung der Benutzerzustände *Ärger* und *Nicht-Ärger* erweitert, weitere Benutzerzustände werden in der nahen Zukunft berücksichtigt. In das Modul wurde auch ein dynamisches Wortlexikon integriert. Die notwendigen Schnittstellen wurden entworfen und implementiert.

Die sich ständig weiterentwickelnde Systemumgebung erforderte die wiederholte Installation der jeweils neuesten Version des Gesamtsystems. Die Tatsache, dass das Gesamtsystem hier am LME in einer lauffähigen Version vorliegt, ist auch den zeitaufwendigen Feineinstellungen (Tuning) zu verdanken. Wiederholt durchgeführte Testläufe des Gesamtsystems auf dem Demonstrator trugen zu einer höheren Stabilität des Systems bei.

Hinsichtlich der Erweiterung der Fähigkeiten zur Erkennung des Benutzerzustands wurden Experimente zur Müdigkeit mit mehreren Probanden durchgeführt. Die Sprachaufnahmen wurden jedoch noch nicht ausgewertet. Die während des Experiments erfassten Daten weisen auf einen Zusammenhang zwischen der akustischen Realisierung der Äußerungen und der Müdigkeit des Sprechers hin.

Eine weiter Eingabemodalität, die im *SmartKom*–System verarbeitet wird, ist die Mimik. Sie

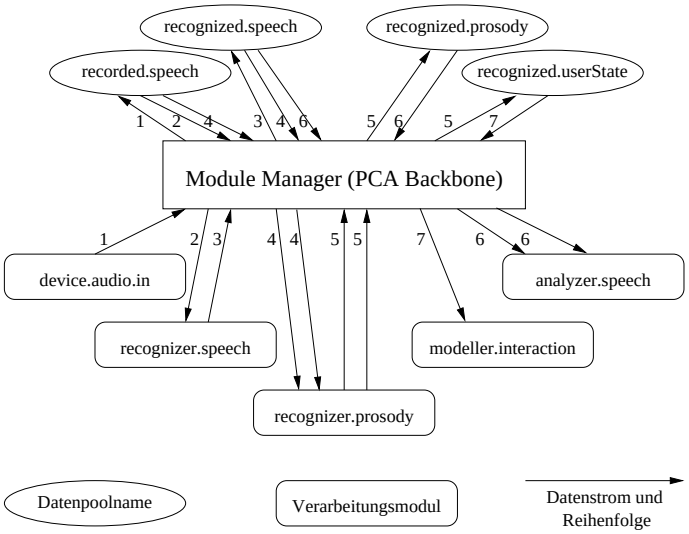

Bild 16: Zusammenspiel des Prosodie-Moduls im System

dient ebenso wie die Emotionserkennung aus der Prosodie dazu, den Zustand eines Benutzers (z. B. Hilflosigkeit) festzustellen und damit eine Änderung der Dialogstrategie des Gesamtsystems zu initiieren.

Das Mimikmodul, das die Analyse des Gesichtsausdruck im *SmartKom*–System durchführt, arbeitet in zwei Schritten. Zuerst wird im Videobild nach dem Gesicht eines Benutzers gesucht. Falls eine Person vor dem *SmartKom*–System steht und deren Gesicht lokalisiert werden konnte, wird der mimische Ausdruck des Gesichts im zweiten Schritt klassifiziert.
Für die Lokalisation des Gesichtes werden alle nicht hautfarbenen Bereiche ausgeblendet. In den übrigen Bereichen wird mit einem, auf Gesichter trainierten, Eigenraumklassifikator die Position mit der größten Übereinstimmung gesucht. Die Vorteile der Hautfarbensegmentierung sind zum einen die Reduzierung des Suchraums, zum anderen werden gesichtsähnliche Texturen im Hintergrund eleminiert.
Die sich anschließende Klassifikation des Gesichtsausdrucks wird ebenfalls von Eigenraumklassifikatoren durchgeführt. Im Gegensatz zur Gesichtslokalisation existiert hier für jede mögliche Klasse an Gesichtsausdrücken ein angepasster Eigenraum. Dieser bietet ein Maß dafür, wie gut ein zu klassifizierendes Bild von diesem Eigenraum modelliert werden kann.

In Bild 5.2 ist ein Gesicht und die daraus modellierten Gesichter des *angry*–, *neutral*–, und *scream*–Eigenraums dargestellt. Es ist deutlich zu erkennen, dass der geöffneten Mund nur vom *scream*–Eigenraum modelliert werden kann. In den beiden anderen Eigenräumen wirkt der Mundbereich verschwommen.

Die Mimikanalyse, basierende auf der hier beschriebenen Methode, wurde im Herbst 2001 zum ersten Mal in das *SmartKom*–Gesamtsystem integriert und bei der MTI–Statustagung (Mensch–Technik–Interaktion) erfolgreich demonstriert.

Nach dem erfolgreichen Aufbau des *SmartKom*-Demonstrators V.1.0 wurden die Architektur des Gesamtsystems weiter entwickelt. Es wurde unter anderem die folgende Änderung beschlos-

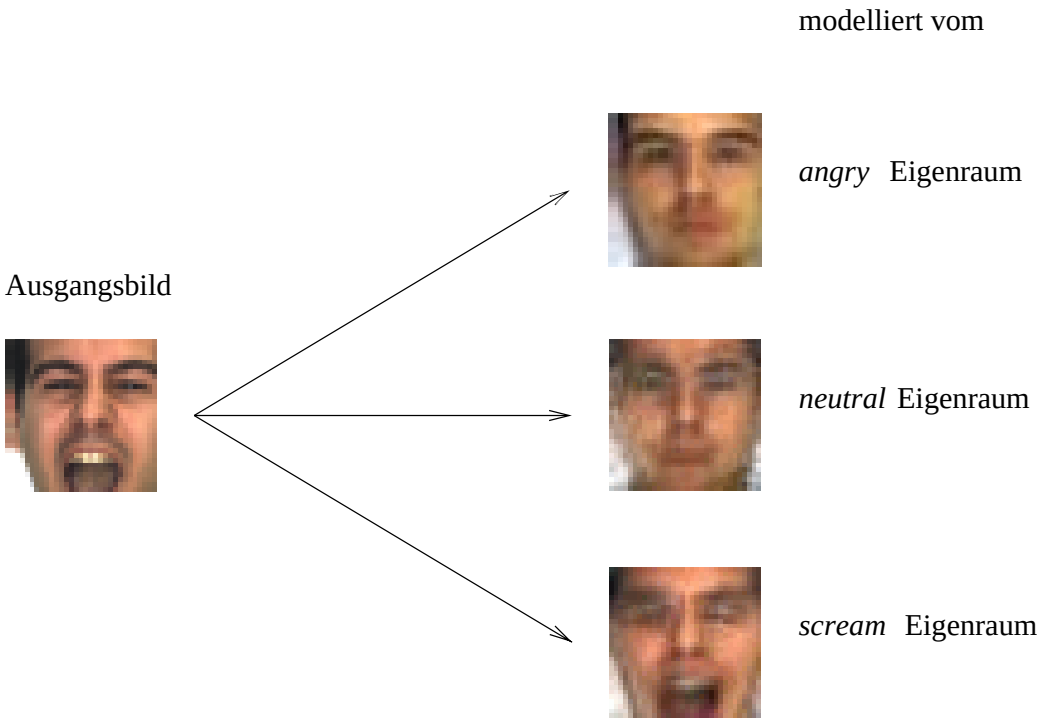

Bild 17: Ein schreiendes Gesicht, das von einem *angry*–, einem *neutral*– und einem *scream*–Eigenraum modelliert wird.

sen und implementiert: Statt eines getrennten Pools für Handgesten und Stiftgesten wird ein einheitlicher Pool für alle Gesteneingaben verwendet.

Im Lauf der an der LMU stattfindenden Untersuchung mit mehreren Benutzern des Systems hat man festgestellt, dass die Benutzer praktisch nur Zeige- und Einkreisgesten verwenden. Daher wurden Vorschläge über mögliche Szenarien gemacht, wie z.B. Zwei-Hand-Gesten als Verstärkung einer negative Äußerung, die die Unzufriedenheit des Benutzers zeigen und Einkreisgesten, die etwa ein Teil eines Stadtplans im Tourismusszenario auf Wunsch vergrössern sollen. Eine Realisierung dieser neuen Möglichkeiten erfordert allerdings eine Kooperation der *SmartKom*-Modulen Display-Management, Aktionsplanung und Medienfusion, die in dieser Form noch nicht existiert. Die entsprechenden Funktionalitäten sind im Modul Gestenanalyse aber bereits implementiert und in das gesamte System integriert.

Ein weiterer Vorschlag bezüglich der Gestenanalyse ist die Verschiebung eines Objekts auf dem Benutzeroberfläche; die ist z.B. notwendig, wenn man sich von einem Stadtplan mehrere Richtungen darstellen läßt. Die erfordert allerdings eine zusätzliche Erweiterung der Schnittstellen des Moduls Gestenanalyse und eine dementsprechende Änderung in anderen Modulen des *SmartKom*-Systems.

Das *SmartKom*-Modul Stifteingabe ist bereit lauffähig, es läßt sich jedoch z.Zt. aus technischen Gründen nicht im das Gesamtsystem demonstrieren: Die Zusammenarbeit führt durch die Komplexität anderer Modulen zu einer asynchronen Ausgabe des Audiosignals und der Grafik. Das Modul hat jedoch bereits in Stand-Alone-Test gut funktioniert. In Bild 16 ist der neue Ablauf der Gestenanalyse in SmartKom zu sehen.

5.3 Flache semantische Analyse

In Zusammenarbeit mit der AT&T Corporation wurde in diesem Jahr ein Forschungsprojekt im Bereich des Automatischen Sprachverstehens betrieben, dessen Besonderheiten darin bestehen, dass der Trainingsvorgang keine Transkription der Daten auf der Wortebene erfordert. Ist eine Menge der aufgezeichneten spontanen Trainingsäußerungen in Form von automatisch erstellten Phontranskriptionen samt ihrer Bedeutung gegeben, so extrahiert man daraus mit Hilfe statistischer Methoden stabile Phonkombinationen, die auch hohe semantische Relevanz im Rahmen der betrachteten Anwendung aufweisen (*akustische Morpheme*).

In der Klassifikationsphase wird nach diesen akustischen Morphemen gesucht, wobei deren semantische Assoziationen als ausschlaggebend für die Klassifikation der ganzen Äußerungen gilt.

Den Ablauf des Klassifikationsprozesses eingebunden in den Mensch-Maschine-Dialog sieht man im Bild 5.3.

Die Methode wurde im Rahmen des *HMIHY (How May I Help You?)* Projektes von AT&T erprobt und hat positive Ergebnisse gebracht. Bei dem HMIHY-Projekt handelt es sich um ein automatisches Telefonauskunftssystem, das Anfragen von Klienten des Telefondienstanbieters annimmt und diese entsprechend behandelt, d.h. beantwortet bzw. weiterleitet.

Für das Jahr 2002 ist eine Fortsetzung der Kooperation geplant, wobei der Schwerpunkt der Forschung auf die Einbindung der *Approximate Matching*-Strategie in Training und Klassifikation gelegt werden soll.

6 Studienarbeiten

1. Adler, W.: Oberflächenapproximation der Papillenregion, August 2001
2. Bäumler, S.: Monokulare Posturschätzung von Personen, Januar 2001
3. Gömmel, G.: Optische Erkennung der Unterlaufaustragsform, September 2001
4. Gräßl, C.: Optimierungsbasiertes Auflösen multipler Ambiguitäten in der Ansichtenplanung mittels Reinforcement-Learning, Februar 2001
5. Grobe, M.: Farbkalibrierung mit normiertem Farbmuster, März 2001
6. Hacker, C.: Optimierung der Merkmalberechnung für die automatische Spracherkennung, November 2001

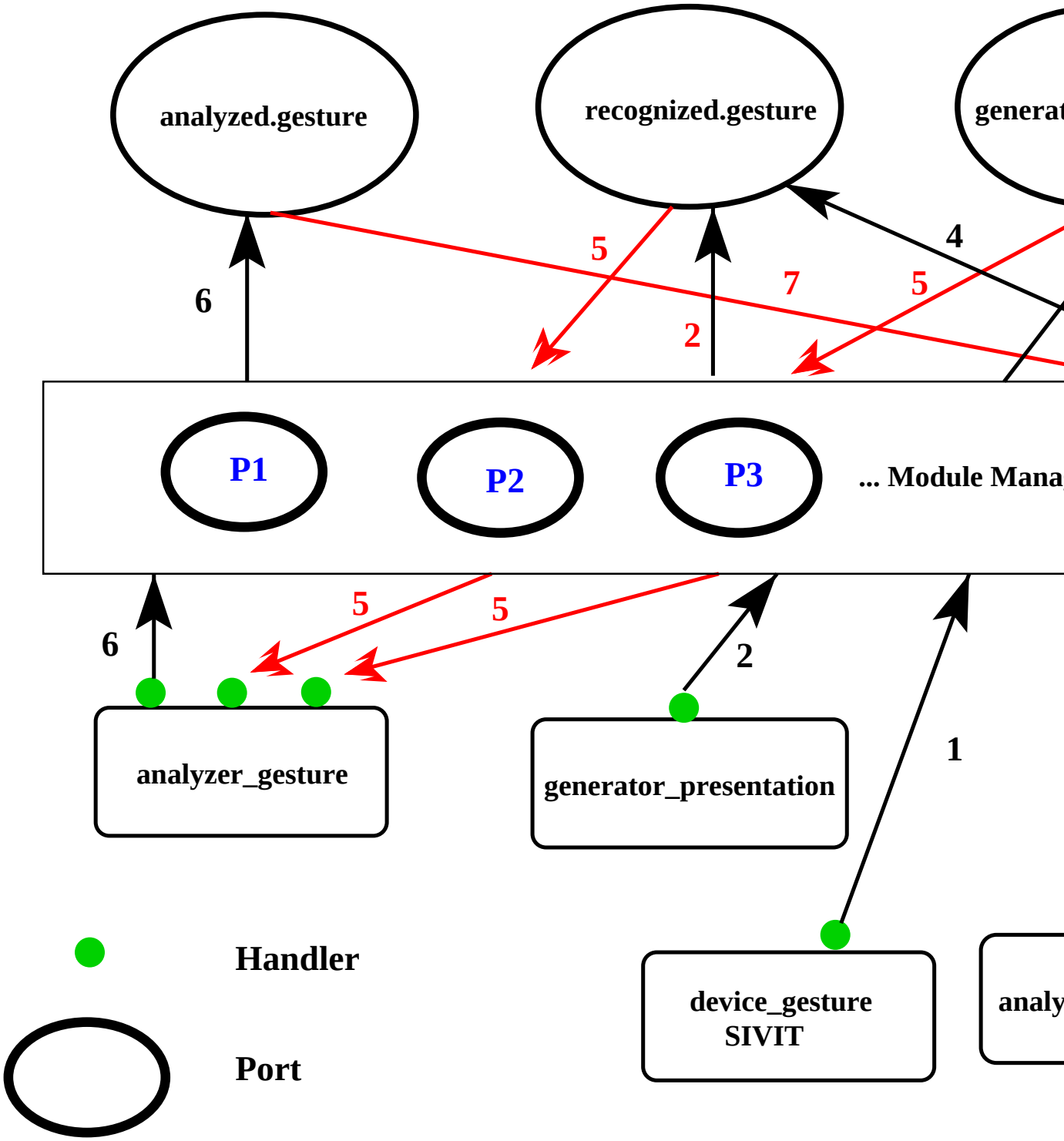

Bild 18: Die neue Gestenanalyse Ablauf in SmartKom

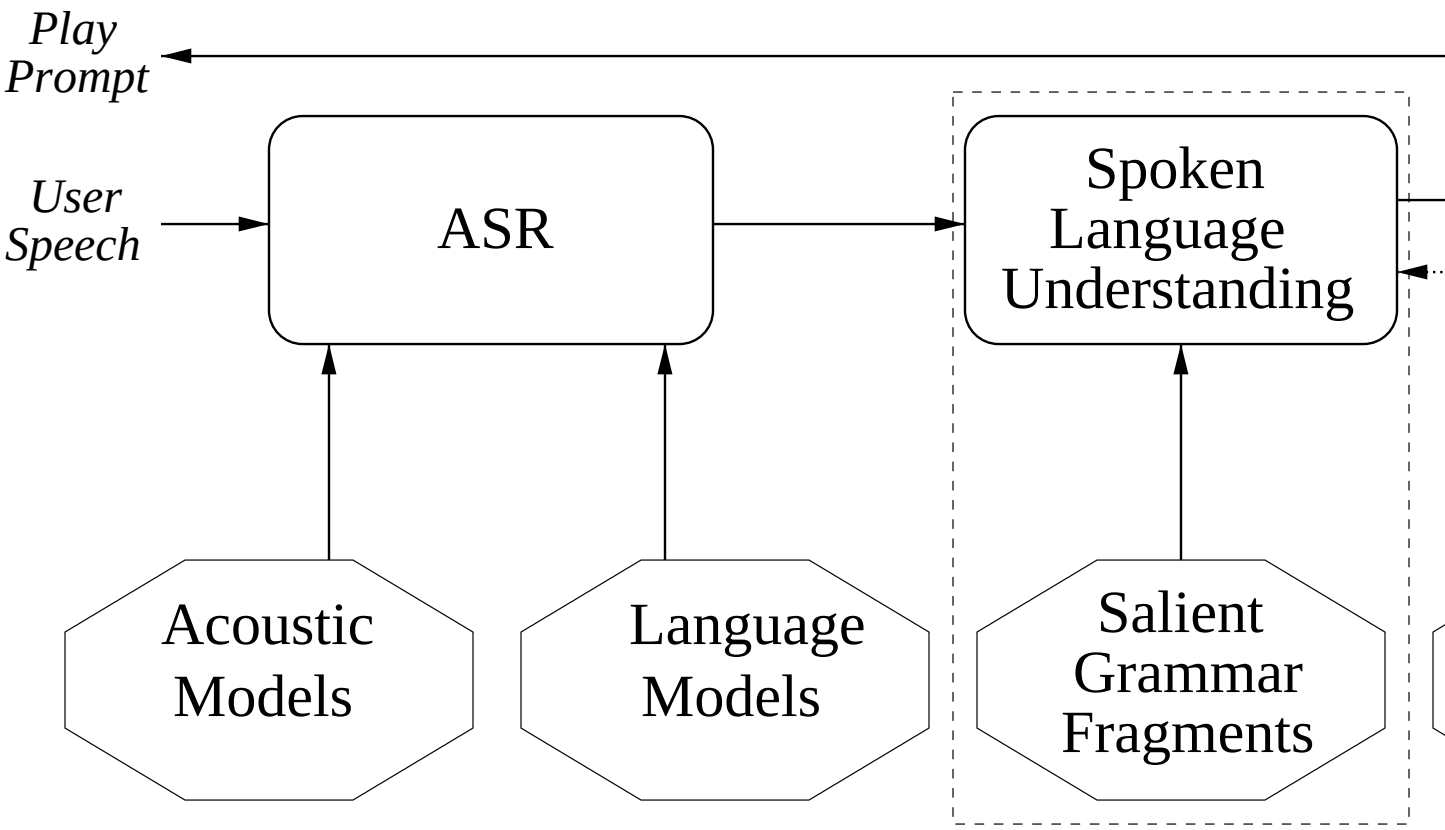

Bild 19: Der Mensch-Maschine-Dialog und die Rolle der akustischen Morpheme.

7. Koestler, H.: Analyse von eukaryotischen Promotorregionen mit exhaustiver Suche, Oktober 2001
8. Steidl, S.: Konfidenzbewertung von Worthypothesen, November 2001
9. Zinßer, T.: Oberflächensegmentierung in Tiefenbildern, März 2001

7 Diplomarbeiten

1. Glomann, B.: Model Generation for Inspection of Electronic Components, Januar 2001
2. Haderlein, T.: Erstellung und Verifikation von möglichen Aussprachealternativen aus textueller Repräsentation, Juli 2001
3. Klimowicz, C.: Lichtfelderzeugung aus endoskopischen Bildfolgen des Bauchraums, Januar 2001
4. Mattern, F.: Probabilistische Hauptachsentransformation zur generischen Objekterkennung, Dezember 2001

5. Vogt, S.: Anwendung von Stereoverfahren zur Verdeckungsberechnung im Bereich erweiterte Realität, Mai 2001

6. Zeißler, V.: Verbesserte linguistische Gewichtung in einem Spracherkenner, März 2001

8 Promotionen

1. Harbeck, S.: Automatische Verfahren zur Sprachdetektion, Landessprachenerkennung und Themendetektion, Juni 2001

2. Ahlrichs, U.: Wissensbasierte Szenenexploration auf der Basis erlernter Analysestrategien, November 2001

3. Fischer, J. M.: Ein echtzeitfähiges Dialogsystem mit iterativer Ergebnisoptimierung, November 2001

4. Gallwitz, F.: Integrierte stochastische Modelle für die Erkennung von Spontansprache, November 2001

9 Vorträge

1. Batliner, A.: How to Repair Speech Repairs in an End-to-End System, ISCA Workshop on Disfluency in Spontaneous Speech (DISS01), Edinburgh, 31.8.2001

2. Batliner, A.: Prosodic Models, Automatic Speech Understanding, and Speech Synthesis: towards the Common Ground, Eurospeech01, Aalborg, 7.9.2001

3. Batliner, A.: Boiling down Prosody for the Classification of Boundaries and Accents in German and English, Eurospeech01, Aalborg, 7.9.2001

4. Batliner, A.: Duration Features in Prosodic Classification: Why Normalization comes Second, and what they Really Encode, Prosody 2001, Red Bank, NJ, 22.10.2001

5. Batliner, A.: Whence and Whither Prosody in Automatic Speech Understanding: A Case Study, Prosody 2001, Red Bank, NJ, 23.10.2001

6. Batliner, A.: Use of prosodic speech characteristics for automated detection of alcohol intoxination, Prosody 2001, Red Bank, NJ, 23.10.2001

7. Deinzer, F.: On Fusion of Multiple Views for Active Object Recognition, Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), 2001, München, 13.09.2001

8. Denzler, J.: MOBSY: Integration of Vision and Dialogue in Service Robots, Computer Vision Systems, Vancouver, Kanada, 08.07.01.

9. Denzler, J.: Optimal Camera Parameter Selection for State Estimation with Applications in Object Recognition, Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM) 2001, München, 13.09.01.

10. Denzler, J.: Optimale Sensordatenauswahl für die Zustandsschätzung im Rechnersehen, Interdisziplinäres Bildverarbeitungs–Kolloquium der Universitäten Heidelberg und Mannheim, Mannheim, 08.05.01.

11. Denzler, J.: Rekonstruktion bildbasierter Szenenmodelle und deren Anwendung im Rechnersehen, Vortragskolloquium, Interdisziplinäres Zentrum für wissenschaftliches Rechnen (IWR) und Institut für Umweltphysik der Universität Heidelberg, Heidelberg, 23.05.01.

12. Denzler, J.: Optimale Sensordatenauswahl fuer die Zustandsschätzung im Rechnersehen, Vortragsreihe Bildverarbeitung und Computergraphik, Universität Leipzig, Leipzig, 19.11.01.

13. Denzler, J.: Computer Vision meets Computer Graphics: Reconstruction and Applications of Image Based Object and Scene Models, Workshop Virtuelles Prototyping und virtuelle Realität für die Bildverarbeitung, Stuttgart, 20.11.01.

14. Nöth, E.: Sprachdialogsysteme, Bosch–Siemens Hausgeräte, Regensburg, 20.2.2001

15. Nöth, E.: Topic Spotting, 1. SKAT Workshop, Uttenreuth, 21.3.2001

16. Nöth, E.: Telephony Based Information Retrieval Systems, Instituto per la Ricerca Scientifica e Tecnologica (IRST), Trento (Italien), 19.4.2001

17. Nöth, E.: Sprachdialogsysteme, Universität Stuttgart, Stuttgart, 11.5.2001

18. Nöth, E.: Erkennung von Benutzerzuständen - Prosodie und Mimik, Projektlekungssitzung, Heidelberg, 22.05.2001

19. Nöth, E.: Neue Trends in der Sprecheridentifikation, 2. SKAT Workshop, Uttenreuth, 4.6.2001

20. Nöth, E.: Spoken Dialogue Systems, International Workshop on Text, Speech and Dialogue (TSD 2001), Zelezna Ruda (Tschechien), 10.9.2001

21. Nöth, E.: Sympalog, International Workshop on Text, Speech and Dialogue (TSD 2001), Zelezna Ruda (Tschechien), 12.9.2001

22. Nöth, E.: Research Issues for the Next Generation Spoken Dialogue Systems Revisited, International Workshop on Text, Speech and Dialogue (TSD 2001), Zelezna Ruda (Tschechien), 13.9.2001

23. Nöth, E.: Erkennung spontaner Sprache, IDEA–LAB „Speech Technologies“, Wissenschaftliche Hochschule fuer Unternehmensfuehrung, Koblenz, 20.10.2001

24. Nöth, E.: Multimodal Interaction and Modeling, MTI-Statustagung, Saarbrücken, 26.10.2001

25. Nöth, E.: Modeling of User State — especially of Emotions, Dagstuhl Seminar „Coordination and Fusion in Multimodal Interaction“, Dagstuhl, 31.10.2001

26. Nöth, E.: Language Models beyond Word Strings, Automatic Speech Recognition and Understanding Workshop (ASRU 2001), Madonna di Campiglio, Italien, 11.12.2001

27. Ohler, U., Statistical Models for the Detection and Analysis of Eukaryotic Promoters, Serono SA, Genf, 10.1.2001

28. Ohler, U., Annotation of Promoters of Protein Encoding Genes in the Complete Drosophila Genome, Jahrestagung der Gesellschaft für Klassifikation, Ludwig–Maximilians–Universität München, 15.3.2001.

29. Ohler, U., Models for the Genome-Wide Detection and Analysis of Eukaryotic Promoters, Valigen SA, Paris, 20.3.2001

30. Ohler, U., A Computer Model for the Genome-Wide Detection of Eukaryotic Promoters, Max-Planck Institut für Chemische Ökologie, Jena, 23.5.2001.

31. Ohler, U., Joint Modeling of DNA Sequence and Physical Properties to Improve Eukaryotic Promoter Recognition, 9th International Conference on Intelligent Systems for Molecular Biology, Kopenhagen, 24.7.2001

32. Ohler, U., Annotation and Analysis of Regulatory Regions in Eukaryotic Genomes, Institut für Molekularbiologie und Biochemie, FU Berlin, 2.8.2001.

33. Ohler, U., Annotation and Analysis of Regulatory Regions in Eukaryotic Genomes, Institut für Bioinformatik, GSF Forschungszentrum für Umwelt und Gesundheit Neuherberg, 6.11.2001.

34. Ohler, U., Annotation and Analysis of Regulatory Regions in Eukaryotic Genomes, Department of Biology, Massachusetts Insitute of Technology, 26.11.2001.

35. Paulus, D., Applications of Structure from Motion (Improwizacja na dany temat), Akademie der Wissenschaften, Gliwice, Polen, 10.05.01

36. Paulus, D., Active image understanding and 3D Reconstruction, Technische Hochschule, Gliwice, Polen, 10.05.01

37. Paulus, D., Segmentierung von Tiefenbildern, Workshop 3D Nord-Ost, Gesellschaft für angewandte Informatik, (GFAI), Berlin, 07.12.2001

38. Paulus, D., Segmentierung von Tiefenbildern, Gastvortrag im Oberseminar Bildverstehen, Universität Koblenz, 11.12.2001

39. Paulus, D., Bildverarbeitung zur Augenheilkunde, Augenärztliche Fortbildung, Augenklinik Erlangen, 12.01.2002

40. Paulus, D., Softwarearchitekturen zur Bildverarbeitung, Gastvortrag im Oberseminar Bildverstehen, Universität Koblenz, 29.01.2002

41. Reinhold, M.: Robuste, erscheinungsbasierte 3-D-Objekterkennung, Graduiertenkolleg 3-D Bildanalyse und -synthese, Universität Erlangen, Erlangen, 24.07.2001

42. Reinhold, M.: Appearance-Based Statistical Object Recognition by Heterogenous Background and Occlusions, Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM) 2001, München, 13.09.2001

43. Reinhold, M.: Improved Appearance-Based 3-D Object Recognition Using Wavelet Features, Vision, Modeling, and Visualization (VMV), 2001, Stuttgart, 23.11.2001

44. Schmidt, J.: Augmented Reality, Bildverarbeitungsforum, Fraunhofer Institut für integrierte Schaltungen, Erlangen-Tennenlohe, 09.03.2001

45. Schmidt, J.: Structure from Motion, Silesian University of Technology, Gliwice, Polen, 08.05.2001

46. Schmidt, J.: Augmented Reality, Silesian University of Technology, Gliwice, Polen, 10.05.2001

47. Schmidt, J.: Structure from Motion, Polish Academy of Sciences, Institute of Theoretical and Applied Computer Science, Gliwice, Polen, 10.05.2001

48. Schmidt, J.: Placing Arbitrary Objects in a Real Scene Using a Color Cube for Pose Estimation, Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), München, 14.09.2001

49. Schmidt, J.: Using Quaternions for Parametrizing 3-D Rotations in Unconstrained Nonlinear Optimization, Vision, Modeling, and Visualization (VMV), 2001, Stuttgart, 23.11.2001

50. Vogt, F.: Teilprojekt B6, Rechnergestützte Endoskopie des Bauchraums, SFB 603 Berichtskolloquium, Erlangen, 26.01.2001

51. Vogt, F.: Bildverarbeitung in der Endoskopie des Bauchraums, Bildverarbeitung für die Medizin (BVM) 2001, Lübeck, 06.03.2001

52. Vogt, F.: Bildverarbeitung in der Endoskopie des Bauchraums, Fraunhofer-Institut f. Intergrierte Schaltungen, Erlangen, 18.05.2001

53. Vogt, F.: Lichtfelder - Theorie und Praxis, Richard Wolf GmbH, Knittlingen, 24.07.2001

54. Vogt, F.: Teilprojekt B6, Rechnergestützte Endoskopie des Bauchraums, Stand und Kooperationen, SFB 603 Berichtskolloquium, Erlangen, 25.07.2001

55. Yuan, C.: Appearance–based Neural Object Recognition and Localization, Int. Conf. Vision, Imaging, and Image Processing (VIIP) 2001, Marbella, Spain, 05.09.2001.

56. Yuan, C.: Web–based 3D Interface for PDM System, Int. Conf. Vision, Imaging, and Image Processing (VIIP) 2001, Marbella, Spain, 03.09.2001.

57. Zobel, M.: Demonstration von Bildverarbeitung und Sprachverstehen in der Dienstleistungsrobotik, 17. Fachgespräch Autonome Mobile Systeme 2001, 12.10.2001

58. Zobel, M.: Optimierung nichtlinearer Least-Squares-Probleme, Arbeitskreis Optimierung, Erlangen, 19.07.2001

59. Zobel, M.: Optimierungsansatz für die Integration von Kamerabildern bei der Klassifikation, SFB-Statustreffen, Erlangen, 26.07.2001

Literatur

[1] U. Ahlrichs, D. Paulus, T. Zinßer, H. Niemann: *Evaluation of Laser Range Sensor*, in S. Donati (Hrsg.): *Optoelectronic Distance / Displacement Measurements and Applications*, Leos, Sept. 2001.

[2] A. Batliner, J. Buckow, R. Huber, V. Warnke, E. Nöth, H. Niemann: *Boiling down Prosody for the Classification of Boundaries and Accents in German and English*, in *EUROSPEECH*, Bd. 4, Aalborg, September 2001, S. 2781–2784.

[3] A. Batliner, B. Möbius, G. Möhler, A. Schweitzer, E. Nöth: *Prosodic models, automatic speech understanding, and speech synthesis: towards the common ground*, in *EUROSPEECH*, Bd. 4, Aalborg, September 2001, S. 2285–2288.

[4] A. Batliner, B. Möbius, A. Schweitzer, S. Goronzy, S. Rapp, P. Regel-Brietzmann: *Guidelines for 'Text-to-Phone' (TTP) Conversion, i.e., the SAMPA - Inventory plus some other Conventions for the SmartKom Lexikon*, SmartKom Technisches Dokument 16, Lehrstuhl für Mustererkennung, Universität Erlangen, Mai 2001.

[5] A. Batliner, E. Nöth, J. Buckow, R. Huber, V. Warnke, H. Niemann: *Duration Features in Prosodic Classification: Why Normalization comes Second, and what they Really Encode*, in M. Bacchiani, J. Hirschberg, D. Litman, M. Ostendorf (Hrsg.): *Proc. of the Workshop on Prosody and Speech Recognition 2001*, Red Bank, NJ, 2001, S. 23–28.

[6] A. Batliner, E. Nöth, J. Buckow, R. Huber, V. Warnke, H. Niemann: *Whence and Whither Prosody in Automatic Speech Understanding: A Case Study*, in M. Bacchiani, J. Hirschberg, D. Litman, M. Ostendorf (Hrsg.): *Proc. of the Workshop on Prosody and Speech Recognition 2001*, Red Bank, NJ, 2001, S. 3–12.

[7] R. Chrastek, M. Wolf, K. Donath, G. Michelson, H. Niemann: *Automatic Optic Disc Segmentation for Analysis of the Optic Nerve Head,* in H. Lemke, M. Vannier, K. Inamura, A. Farman (Hrsg.): *Proc. 15th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS'01)*, Elsevier, Berlin, Germany, 2001, S. 1119.

[8] F. Deinzer, J. Denzler, H. Niemann: *On Fusion of Multiple Views for Active Object Recognition,* in B. Radig, S. Florczyk (Hrsg.): *Pattern Recognition — 23rd DAGM Symposium*, Springer, Berlin, September 2001, S. 239–245.

[9] J. Denzler, C. Brown, H. Niemann: *Optimal Camera Parameter Selection for State Estimation with Applications in Object Recognition,* in B. Radig, S. Florczyk (Hrsg.): *Pattern Recognition — 23rd DAGM Symposium*, Springer, Berlin, September 2001, S. 305–312.

[10] B. A. Draper, U. Ahlrichs, D. Paulus: *Adapting Object Recognition Across Domains: A Demonstration*, S. to appear.

[11] T. Ertl, B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2001 (Proceedings of the International Workshop, Stuttgart, Germany)*, Akademische Verlagsgesellschaft, Berlin, 2001.

[12] A. Gebhard, D. Paulus, B. Suchy, S. Wolf, H. Niemann: *Automatische Graduierung von Gesichtsparesen*, in H. Pöppl (Hrsg.): *5. Workshop Bildverarbeitung für die Medizin*, Springer, Heidelberg, 2001, S. 352–356.

[13] M. Greiffenhagen, D. Comaniciu, H. Niemann, V. Ramesh: *Design, Analysis and Engineering of Video Monitoring Systems: An Approach and a Case Study, Proc. IEEE*, 2001.

[14] M. Greiffenhagen, V. Ramesh, H. Niemann: *The Systematic Design and Analysis Cycle of a Vision System: A Case Study in Video Surveillance*, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, USA, 2001.

[15] T. Haderlein, T. Wittenberg, M. Decher, E. Nöth: *Automatische Stottererkennung und Klassifikation mit Hilfe von Hidden–Markov–Modellen,* in M. Gross, E. Kruse (Hrsg.): *Aktuelle phoniatrisch–pädaudiologische Aspekte 2000/2001*, Median–Verlag, Heidelberg, 2001, S. 124–130.

[16] J. Hornegger, H. Niemann: *An New Probabilistic Model for Object Recognition and Pose Estimation, Int. Journal of Pattern Recognition and Artificial Intelligence*, Bd. 15, 2001, S. 241–253.

[17] Y. Huang, D. Paulus, H. Niemann: *Background-Forground Segmentation Based on Dominant Motion Estimation and Static Segmentation, Journal of Computing and Information Technology*, Bd. 8, Nr. 4, 2000, S. 349–354.

[18] A. Krzyzak, H. Niemann: *Convergence and rates of convergence of radial basis functions networks in function learning, Nonlinear Analysis*, Bd. 47, 2001, S. 281–292.

[19] M. Levit, A. Gorin, J. Wright: *Multipass algorithm for acquisition of salient acoustic morphemes*, in *EUROSPEECH*, Aalborg, September 2001, S. 1645–1648.

[20] M. Levit, R. Huber, A. Batliner, E. Nöth: *Use of prosodic speech characteristics for automated detection of alcohol intoxination*, in M. Bacchiani, J. Hirschberg, D. Litman, M. Ostendorf (Hrsg.): *Proc. of the Workshop on Prosody and Speech Recognition 2001*, Red Bank, NJ, 2001, S. 103–106.

[21] E. Michaelsen, U. Ahlrichs, U. Stilla, D. Paulus: *Where is the punch?*, in B. Radig, S. Florczyk (Hrsg.): *Pattern Recognition — 23rd DAGM Symposium*, Springer, Berlin, September 2001, S. 337–344.

[22] E. Nöth, A. Batliner, H. Niemann, G. Stemmer, F. Gallwitz, J. Spilker: *Language Models beyond Word Strings*, in *Proceedings of the Automatic Speech Recognition and Understanding Workshop (ASRU'01)*, 2001.

[23] E. Nöth, M. Boros, J. Fischer, F. Gallwitz, J. Haas, R. Huber, H. N. G. Stemmer, V. Warnke: *Research Issues for the Next Generation Spoken Dialogue Systems Revisited*, in V. Matoušek, P. Mautner, R. Mouček, K. Tašer (Hrsg.): *Proc. 4th International Conference on Text, Speech and Dialogue (TSD 2001)*, Bd. 2166 von *Lecture Notes for Artificial Intelligence*, Springer–Verlag, Berlin, September 2001, S. 341–348.

[24] U. Ohler, H. Niemann: *Identification and analysis of eukaryotic promoters: recent computational approaches*, *Trends Genet.*, Bd. 17, 2001, S. 56–60.

[25] U. Ohler, H. Niemann, G. Liao, G. M. Rubin: *Joint modeling of DNA sequence and physical properties to improve eukaryotic promoter recognition*, *Bioinformatics*, Bd. 17, 2001, S. S199–S206. 19

[26] D. Paulus: *Aktives Bildverstehen*, Der andere Verlag, Osnabrück, 2001, Habilitationsschrift in der Praktischen Informatik, Universität Erlangen-Nürnberg, Mai 2000.

[27] M. Reinhold, C. Drexler, H. Niemann: *Image Database for 3-D Object Recognition*, LME-TR-2001-02, Informatik 5 (Lehrstuhl für Mustererkennung), Universität Erlangen-Nürnberg, Mai 2001.

[28] M. Reinhold, D. Paulus, H. Niemann: *Appearance-Based Statistical Object Recognition by Heterogenous Background and Occlusions*, in B. Radig, S. Florczyk (Hrsg.): *Pattern Recognition, 23rd DAGM Symposium*, Springer-Verlag, Berlin, Heidelberg, New York, München, September 2001, S. 254–261, Lecture Notes in Computer Science 2191.

[29] M. Reinhold, D. Paulus, H. Niemann: *Improved Appearance-Based 3-D Object Recognition Using Wavelet Features*, in T. Ertl, B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2001*, AKA/IOS Press, Berlin, Amsterdam, Stuttgart, November 2001, S. 473–480.

[30] J. Schmidt, H. Niemann: *Using Quaternions for Parametrizing 3–D Rotations in Unconstrained Nonlinear Optimization*, in T. Ertl, B. Girod, G. Greiner, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2001*, AKA/IOS Press, Berlin, Amsterdam, Stuttgart, Germany, November 2001, S. 399–406. 14

[31] J. Schmidt, I. Scholz, H. Niemann: *Placing Arbitrary Objects in a Real Scene Using a Color Cube for Pose Estimation*, in B. Radig, S. Florczyk (Hrsg.): *Pattern Recognition, 23rd DAGM Symposium*, Springer-Verlag, Berlin, Heidelberg, New York, Munich, Germany, September 12–14 2001, S. 421–428, Lecture Notes in Computer Science 2191. 14

[32] I. Scholz, J. Schmidt, H. Niemann: *Farbbildverarbeitung unter Echtzeitbedingungen in der Augmented Reality*, in D. Paulus, J. Denzler (Hrsg.): *7. Workshop Farbbildverarbeitung*, Universität Erlangen-Nürnberg, Institut für Informatik, Erlangen, Germany, 4.-5. Oktober 2001, S. 59–65, Arbeitsberichte des Instituts für Informatik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Band 34, Nr. 15. 14

[33] R. Siepmann, A. Batliner, D. Oppermann: *Using Prosodic Features to Characterize Off-Talk in Human-Computer-Interaction*, in M. Bacchiani, J. Hirschberg, D. Litman, M. Ostendorf (Hrsg.): *Proc. of the Workshop on Prosody and Speech Recognition 2001*, Red Bank, NJ, 2001, S. 147–150.

[34] J. Spilker, A. Batliner, E. Nöth: *How to Repair Speech Repairs in an End-to-End System*, in R. Lickley, L. Shriberg (Hrsg.): *Proc. ISCA Workshop on Disfluency in Spontaneous Speech*, Edinburgh, September 2001, S. 73–76.

[35] G. Stemmer, C. Hacker, E. Nöth, H. Niemann: *Multiple Time Resolutions for Derivatives of Mel-Frequency Cepstral Coefficients*, in *Proceedings of the Automatic Speech Recognition and Understanding Workshop (ASRU'01)*, 2001.

[36] G. Stemmer, E. Nöth, H. Niemann: *Acoustic Modeling of Foreign Words in a German Speech Recognition System*, in *EUROSPEECH*, Bd. 4, Aalborg, September 2001, S. 2745–2748.

[37] G. Stemmer, V. Zeißler, E. Nöth, H. Niemann: *Towards a Dynamic Adjustment of the Language Weight*, in V. Matoušek, P. Mautner, R. Mouček, K. Taušer (Hrsg.): *Proc. 4th International Conference on Text, Speech and Dialogue (TSD 2001)*, Bd. 2166 von *Lecture Notes for Artificial Intelligence*, Springer–Verlag, Berlin, September 2001, S. 323–328.

[38] J. Sun, C. Yuan: *Web–based 3D Interface for Product Data Management*, in M. Hamza (Hrsg.): *Proc. IASTED Int. Conf. Visualization, Imaging and Image Processing (VIIP 2001)*, ACTA Press, Anaheim–Calgary–Zurich, September 2001, S. 33–38.

[39] F. Vogt, C. Klimowicz, D. Paulus, W. Hohenberger, H. Niemann, C. H. Schick: *Bildverarbeitung in der Endoskopie des Bauchraums*, in H. Pöppl (Hrsg.): *5. Workshop Bildverarbeitung für die Medizin*, Springer, Heidelberg, 2001, S. 320–324. 17

[40] F. Vogt, D. Paulus, C. Schick: *Fast Implementations of Temporal Color Image Filtering*, in D. Paulus, J. Denzler (Hrsg.): *Siebter Workshop Farbbildverarbeitung*, Gruner Druck GmbH, Erlangen, 2001, S. 89–98. 17

[41] A. Weckenmann, R. Bettin, V. Stöber, H. Niemann, R. Deventer: *Modellierungsverfahren zur Optimierung und Regelung verkürzter Prozessketten*, in G. Redeker (Hrsg.): *Qualitätsmanagement für die Zukunft – Business Excellence als Ziel*, Shaker Verlag, Aachen, 2001, S. 109–124.

[42] C. Yuan, H. Niemann: *Appearance based neural object recognition and localization*, in M. Hamza (Hrsg.): *Proc. IASTED Int. Conf. Visualization, Imaging and Image Processing (VIIP 2001)*, ACTA Press, Anaheim–Calgary–Zurich, September 2001, S. 600–603.

[43] C. Yuan, H. Niemann: *Neural networks for the recognition and pose estimation of 3–D objects from a single 2–D perspective view*, *International Journal of Image and Vision Computing*, Bd. 19, August 2001, S. 585–592.

[44] C. Yuan, H. Niemann: *Wavelet features for appearance–based image analysis*, in *Third International Conference on Information, Communications, and Signal Processing (ICICS 2001)*, Singapore, October 2001, S. 3D1.7.

[45] M. Zobel, J. Denzler, B. Heigl, E. Nöth, D. Paulus, J. Schmidt, G. Stemmer: *Demonstration von Bildverarbeitung und Sprachverstehen in der Dienstleistungsrobotik*, in P. Levi, M. Schanz (Hrsg.): *Autonome Mobile Systeme 2001, 17. Fachgespräch, Stuttgart, 11./12. Oktober 2001*, Springer, Berlin, Heidelberg, New York, 2001, S. 141–147.

[46] M. Zobel, J. Denzler, B. Heigl, E. Nöth, D. Paulus, J. Schmidt, G. Stemmer: *MOBSY: Integration of Vision and Dialogue in Service Robots*, in B. Schiele, G. Sagerer (Hrsg.): *Computer Vision Systems, Proceedings Second International Workshop, ICVS 2001 Vancouver, Canada, July 7-8, 2001*, Springer, 2001, S. 50–62, Lecture Notes in Computer Science.