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
Caputo, B., M.Sc.	(wiss. Mitarb., Grad.-Stip.)	bis 31.01.02
Deinzer, F., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.01.99
Denzler, J., Dr.-Ing.	(wiss. Assistent)	01.01.93
Deutsch, B., Dipl.-Inf.	(wiss. Mitarb.,SFB 603)	01.10.02
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
Gräßl, C., Dipl.-Inf.	(wiss. Mitarb., EU)	01.08.02
Grzegorzek, M., Dipl.-Inf.	(wiss. Mitarb., Grad.-Stip.)	01.12.02
Hacker, C., Dipl.-Inf.	(wiss. Mitarb.,Vertr. Nöth)	01.10.02 - 31.12.02
Koppe, I.	(Sekretärin, 1/2)	18.12.92
Mattern, F., Dipl.-Inf.	(wiss. Mitarb., DFG)	01.02.02
Montel-Kandy, M.	(Sekretärin, 1/2, SFB 603)	15.11.01
Müller, K.	(Sekretärin,1/2)	01.02.02
Niemann, H., Dr.-Ing.	(Professor)	24.09.75
Nöth, E., Dr.-Ing.	(Akad. Oberrat)	01.02.85, beurlaubt ab 01.10.02
Popp. F.	(Techniker)	01.09.85
Reinhold, M., Dipl.-Ing.	(wiss. Mitarb., Grad.-Stip.)	bis 30.09.02
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
Steidl, S., Dipl.-Inf.	(wiss. Mitarb., EU)	15.11.02
Stemmer, G., Dipl.-Inf.	(wiss. Mitarb.)	13.10.99
Vogt, F., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	15.10.00
Wenhardt, S., Dipl.-Math. (FH)	(wiss.Mitarb., VHB)	01.06.02
Zeissler, V., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.05.01
Zinßer, T., Dipl.-Inf.	(wiss. Mitarb., EU)	15.07.02
Zobel, M., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.01.98

Gäste

Gäste:

Cremers, D., Dr.	(Mannheim, SFB 603)	08.07. – 13.07.2002
Daniildis, K., Prof.	(USA, DSFB 603)	17.11. – 23.11.2002
Gurevich, I., Prof.	(Russland, SFB 603)	17.11. – 01.12.2002
Kittler, J., Prof.	(England, SFB 603)	06.05. – 08.05.2002
Kutulakus, K., Prof.	(Kanada, SFB 603)	18.11. – 22.11.2002
von der Malsburg, C., Prof.	(Bochum, SFB 603)	26.06. – 27.06.2002
Regazzoni, C., Prof.	(Italien, SFB 603)	06.03. – 08.03.2002
Ruffitzkij, V., Dr.	(Russland, SFB 603)	15.04. – 21.04.2002
Salvetti, O., Prof.	(Italien, SFB 603)	06.05. – 08.05.2002
Sarti, A., Prof.	(Italien, SFB 603)	26.05. – 29.05.2002
Weickert, J., Prof.	(Saarbrücken, SFB 603)	17.11. – 23.11.2002
Ilcheva, Z., Dr.	(Bulgarien, DAAD)	05.08. – 30.09.2002
Horita, Y., Prof.	(Japan, Jap. Reg.)	24.06. – 30.09.2002
Novikov, K., Dr.	(Russ. Föderation, DAAD)	01.12. – 28.02.2003
Sokolov, M.	(Russ. Föderation, DAAD/LME)	18.02. – 28.04.2002
Pulakka, H.	(Finnland, DAAD/LME)	01.06. – 31.08.2002
Osipova, E.	(Russ. Föderation, DAAD/LME)	15.07. – 31.08.2002
Levine, K.	(Russ. Föderation, DAAD/LME)	15.07. – 31.08.2002
Krzysztof, C., Prof.	(Polen, Socrates/Erasmus)	11.06. – 17.06.2002

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 Visualisierung komplexer Szenen und Sensordaten“ hergestellt, dessen Sprecher Prof. Niemann ist. Eine Verknüpfung zwischen Bild– und Sprachanalyse wird im Projekt http://smartkom.dfki.de hergestellt, das vom http://www.bmbf.de 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: J. Denzler
(F. Deinzer, B. Deutsch, R. Deventer, J. Drexler, C. Gräßl, M. Grzegorzek, M. Reinhold, J. Schmidt, F. Mattern, I. Scholz, F. Vogt, T. Zinßer, M. Zobel)

Schwerpunkt der Forschungstätigkeiten im Bereich der Bildanalyse am Lehrstuhl ist die statistische Objektmodellierung, –erkennung und Verfolgung, grundlagenorientierte Arbeiten zur optimalen Sensordatenauswahl und –fusion im aktiven Rechnersehen sowie Kamerakalibrierung und 3–D Rekonstruktion mit Anwendungen in der erweiterten und virtuellen Realität. Versuchs– und Anwendungsplattform ist projektübergreifend das autonome, mobile System MOBSY, in dem die verschiedenen Verfahren unter Echtzeitbedingungen und in realer, natürlicher Umgebung ihre Leistungsfähigkeit unter Beweis stellen müssen. 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 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. Der grundlegende Versuchsaufbau für die Projekte der Bildanalyse besteht aus beweglichen rechnergesteuerten Kameras, die beispielsweise an der Hand eines Roboters montiert sind und dadurch im Arbeitsraum des Roboters frei positioniert werden können, oder rechnergesteuerten Multi–Media Farbkameras, welche die Szene durch gezielte Schwenk–/Neigebewegungen überwachen und Details von Objekten durch Änderung der Brennweite betrachten können. Zur kontrollierten Datenaufnahme, die beispielsweise bei der Erstellung von Stichproben erforderlich ist, existieren zwei rechnersteuerbare Aufbauten, die jeweils aus einem Drehteller und einem Schwenkarm bestehen. An dem Schwenkarm ist eine hochwertige Farb–Kamera befestigt, so dass von einem Objekt auf dem Drehteller Ansichten von einer beliebigen Position auf einer Halbkugel um das Objekt aufgenommen werden können.

Für die laufenden Projekte auf dem Gebiet der optimalen Sensordatenauswahl 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 III 850) ermöglichen eine vollständig Autonomie. Zum Rechnercluster des Lehrstuhls besteht eine Verbindung über ein Funkethernet. 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 sowie einem Greifer. Die Plattform wurde im vergangenen Jahr unter Mithilfe der Mechanikwerkstatt der Universität Erlangen–Nürnberg grundlegend erweitert. Der Aufbau wurde um ein Touchscreen LCD–Display ergänzt, damit Interaktion (Starten von Demoprogrammen, Auswahl von Objekten, Debugging) direkt an der ansonsten vollständig autonom agierenden Plattform möglich wird. Desweiteren wurde die Konstruktion des Aufbaus so verändert, dass ohne größere technische Eingriffe, die Position des Stereokopfes, d.h. der Kameras, verändert werden kann. Diese gesteigerte Flexibilität ist in einem Projekt zum sichtbasierten Greifen eines Objekts notwendig, da sichergestellt werden muss, dass die Kamera zu jedem Zeitpunkt den Greifer der Plattform einsehen kann.

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

3.1 Statistische Objekterkennung und Ansichtenplanung

Die Arbeit zur statistischen, erscheinungsbasierten Objekterkennung im Rahmen des von der DFG geförderten Graduiertenkollegs ”Dreidimensionale Bildanalyse und -synthese”wurde fortgesetzt.

In dieser Arbeit kommt ein erscheinungsbasierter Ansatz zum Einsatz, d. h. man führt keinen vorhergehenden Segmentierungsprozess durch, sondern berechnet die Merkmale direkt aus den Bilddaten. Dieser Ansatz verwendet dabei lokale Merkmale, die sich mit Hilfe der Wavelet-Multiskalen- Analyse bestimmen lassen. Die erste Komponente hängt vom Tiefpass-Anteil und die zweite vom Hochpass-Anteil ab.

Ein Objekt lässt sich dann durch die Menge der lokalen Merkmalsvektoren beschreiben, die sich innerhalb des Objektfensters befinden. Dabei ist die Größe und Form dieses Objektfensters variabel, um die Größenänderungen des Objektes im Bild bei Transformationen orthogonal zur Bildebene zu erfassen. Im Folgenden werden diese Transformationen orthogonal zur Bildebene (z. B. Skalierung oder Rotationen senkrecht zur Bildebene) als externe Transformationen bezeichnet. Das variable Objektfenster lässt sich mit Hilfe von Bildern trainieren. Um die Abhängigkeit mathematisch zu formulieren, verwendet man Zugehörigkeitsfunktionen, die sich

mit Hilfe von Summen gewichteter Basisfunktionen als Funktionen der externen Transformationen darstellen lassen.

Die Merkmalsvektoren innerhalb des Objektfensters werden statistisch als normal verteilt modelliert. Auch hier kommen Summen gewichteter Basisfunktionen zum Einsatz, um externe Transformationen zu modellieren.

Um ein Objekt im Bild zu erkennen, schätzt das Objekterkennungssystem zunächst für jede mögliche Objektklasse die beste Lage und wählt dann die Objektklasse mit der besten Bewertung aus.

Da sich dieses einfache Objektmodell nur für unverdeckte Objekte vor homogenem Hintergrund eignet - was bei realen Anwendungen allerdings selten vorkommt -, erfolgt eine Erweiterung um das Szenenmodell. Ein wesentlicher Bestandteil des Szenenmodells ist das statistische Modell für den Hintergrund, für den man eine Gleichverteilung annimmt. Der zweite wichtige Bestandteil des Szenenmodells ist die Zuweisungsfunktion. Diese weist jeden Merkmalsvektor innerhalb des eng umschließenden Objektfensters entweder dem Hintergrund oder dem Objekt zu, so dass die Objektdichte für diese Objekt- und Lagehypothese maximiert wird.

Die zahlreichen Experimente an mehreren Stichproben zeigten, dass sich mit Hilfe der trigonometrischen Basisfunktion selbst drei externe Transformationen (zwei Rotationen und eine Skalierung) gut handhaben lassen. Ebenso ist das System robust bei heterogenem Hintergrund und Verdeckungen. So lagen beispielsweise bei heterogenem Hintergrund und 20% Verdeckung bei zwei externen Transformationen die Lokalisationsrate bei 70% und die Klassifikationsrate bei 55%. Selbst bei drei externen Transformationen erzielte das Objekterkennungssystem eine Lokalisationsrate von 51%.

Eine wesentliche Erweiterung, die im letzten Jahr durchgeführt wurde, ist das Mehrobjektmodell. Während man mit bisherigem Modell immer nur ein Objekt im Bild erkennen konnte, ist es mit dem neuen Mehrobjektmodell auch möglich, mehrere Objekte in einem Bild zu erkennen. Zu diesem Zweck wird eine globale Zuweisungsfunktion eingeführt, die die einzelnen Merkmalsvektoren den Objekten zuordnet (siehe Abbildung 3.1). Dabei kann jeder Merkmalsvektor maximal einem Objekt zugeordnet werden. Die Suche ist seriell implementiert: Zunächst sucht das System nach dem ersten Objekt im Bild und führt die Zuweisungen durch, dann nach dem zweiten Objekt, usw. Da meist die Anzahl der Objekte im Bild vorab nicht bekannt ist, stoppt der Algorithmus selbstständig, wenn alle Objekte im Bild gefunden sind, d. h. keine weiteren "gültigen" Objekthypothesen mehr gibt. Eine Objekthypothese ist bei homogenem Hintergrund gültig, wenn die Objektdichte größer als die entsprechende Hintergrunddichte ist. Bei heterogenem Hintergrund müssen mindestens $S_p\%$ des Objektes sichtbar sein. Das Mehrobjektmodell wurde sowohl bei homogenem als auch heterogenem Hintergrund getestet: Bei homogenem Hintergrund arbeitet es fehlerfrei, während bei heterogenem Hintergrund bisweilen — je nach Wahl der Schwelle S_p — zu wenig oder zu viel Objekte gefunden werden.

Ein Schwerpunkt im Teilprojekt B2 des Sonderforschungsbereichs 603 besteht aus Forschungsarbeiten zur aktiven Objekterkennung in der Bildverarbeitung. Die Erkennungsrate bei der Klassifikation bzw. die Genauigkeit der Lokalisation hängen aufgrund von Mehrdeutigkeiten zwischen Objekten unter Umständen stark von den gewählten Sensordaten ab. Um nur eine minimale Anzahl von Objektansichten aufzunehmen, setzt man eine gezielte *Ansichtenplanung* ein. Deren Aufgabe ist es, für ein gegebenes Objekt die minimal nötige Anzahl von optimalen Ansichten

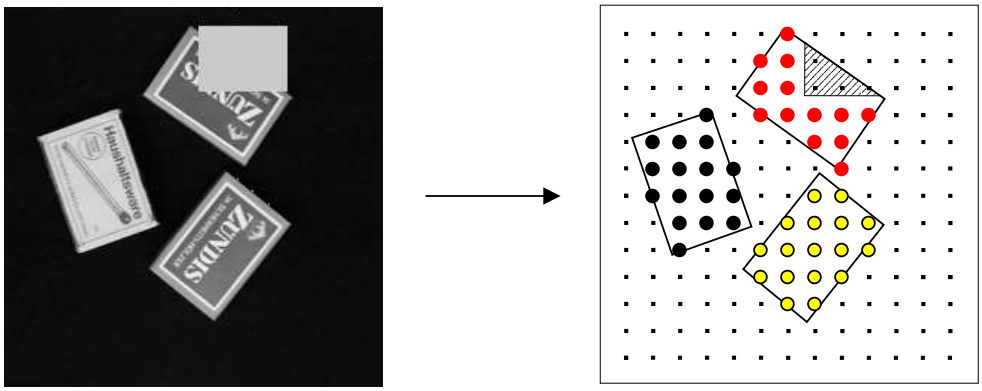


Abbildung 1: Die einzelnen Merkmalsvektoren des Bildes werden durch die globale Zuweisungsfunktion den einzelnen Objekten zugeordnet. Dabei werden beim Szenenmodell auch Verdeckungen berücksichtigt.

auszuwählen, um das Objekt bestmöglich klassifizieren und lokalisieren zu können. Zur Ansichtenplanung existieren bereits seit drei Jahren sehr allgemeine theoretische Grundlagen basierend auf Reinforcement Learning, auf denen weiter aufgebaut wurde.

Ein Schwerpunkt der Arbeiten im vergangenen Jahr waren Untersuchungen zur Planung von *Ansichtensequenzen*. Eine Voraussetzung dafür ist, dass die Objekterkennung in der Lage sein muss, Informationen aller angefahrenen Ansichtspositionen zu fusionieren. Dazu wurden bereits in 2001 ein entsprechendes Verfahren zur Fusion von Dichten über die Zeit entwickelt und vorgestellt [8]. Darauf aufbauend lassen sich wieder die vorhandenen Grundlagen zur Ansichtenplanung auch zur Sequenzplanung einsetzen. Hierbei wird ausgehend von den aktuellen, fusionierten Dichten der Objekterkennung aller bisherigen Ansichten die optimale Aktion (z. B. Änderung der Kameraposition) berechnet. Dabei ist es das Ziel einer optimalen Aktion, dass die ausgewählte Aktion und die daraus resultierende Ansicht die Entropie der aus der Sensordatenfusion resultierende Dichte minimiert. Dieses Entropiemaß eignet sich sehr gut um die Unterscheidbarkeit von Objekten auszudrücken. In Abbildung 2 ist ein Beispiel zur Planung von Ansichtensequenzen dargestellt.

Neben der Planung von Ansichtensequenzen war ein Forschungspunkt, zu zeigen, dass sich die allgemeinen Methoden der Ansichtenplanung auch auf andere Gebiete der Aktionsplanung übertragen. Dies wurde innerhalb des Demosystems im Teilprojekt B2 realisiert, wo es als Greifplanung eingesetzt wird, um der mobilen Plattform ein sicheres Greifen von schwierigen Objekten zu ermöglichen.

Ein weiterer Forschungsbereich waren statistische Verfahren zur Objekterkennung. Viele dieser Verfahren beruhen im Kern auf Gauss-Dichten, mit denen Verteilungen der Merkmale einzelner Ansichten statistisch modelliert werden. Dabei hat man allerdings das Problem, dass die geschätzten Dichten nur für einzelne Ansichten gültig sind. Zwischen benachbarten Ansichten gibt es häufig Lücken im Modell, die meist nur dadurch geschlossen werden können, dass auf bekannte Ansichten zurückgegriffen wird und damit eine Diskretisierung des Bereichs der be-

kannten Lageparameter durchgeführt wird. Um diesem Problem zu begegnen, wurde ein Verfahren entwickelt, das es erlaubt, Gauss-Dichten zu parametrisieren, sofern eine Abstandsbeziehung zwischen einzelnen Ansichtspositionen existiert. Diese Parametrisierung wurde erfolgreich eingesetzt, um den statistischen Eigenraumklassifikator um kontinuierliche Objektmodelle zu erweitern.

Im Bereich der Ansichtenplanung wird das Verfahren das bisher nur anhand von synthetischen Bildern getestet wurde auf komplexe reale Probleme übertragen. Dabei stellen diese Evaluationen besondere Anforderungen an Robustheit, Effizienz und Speicherbedarf. In diesen Bereichen wird das Verfahren weiter optimiert werden.

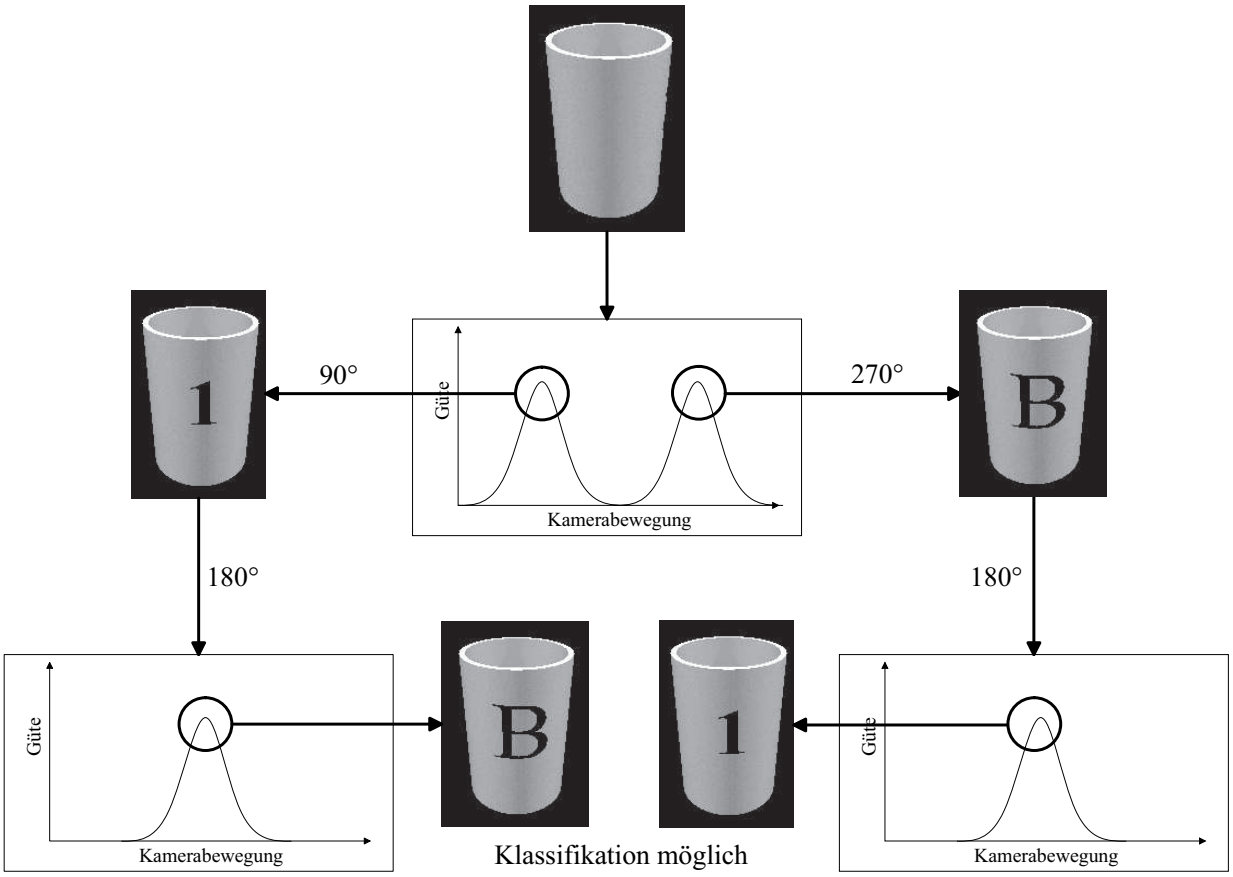


Abbildung 2: Beispiel für Ambiguitäten zwischen Objekten und deren Auflösung mittels Planung von Ansichtensequenzen. Die Becher sind auf einer Seite mit einem Buchstaben (**A** oder **B**) und auf der gegenüberliegenden Seite mit einer Ziffer (**1** oder **2**) gekennzeichnet. In einem ersten Schritt muss man sich entweder um 90° oder um 270° um den Becher bewegen. Ausgehend von der Beobachtung (entweder Ziffer oder Buchstaben sind nun sichtbar) erlaubt nun eine Kamerabewegung um 180° die Klassifikation des Bechers.

3.2 Generische Objektmodellierung und –erkennung

Aufbauend auf das im Projekt DIROKOL bis 2001 entwickelte erscheinungsbasierte 3–D Objekterkennungssystem wurde eine statistische Modellierung und eine hierarchische generische Objekterkennung entwickelt. Das Ziel dabei ist es in einer Szene 3–D Objekte bei heterogenem Hintergrund und teilweiser Verdeckung zu erkennen und dessen Drehlage zu bestimmen. Desweiteren sollen Objekte, die während des Trainings nicht gesehen wurden aber bestimmten Trainingsobjekten sehr ähnlich sind, einer generischen Oberklasse zugeordnet werden können, anstatt sie wie bisher entweder zurückzuweisen oder einer nicht korrekten Objektklasse zuzuweisen.

Im Rahmen dieser Arbeit wurden Verfahren zur unüberwachten und überwachten Oberklassengenerierung untersucht und eine Möglichkeit erarbeitet, die gegenüber Verdeckung, Skalierung und heterogenem Hintergrund robusten Verfahren aus der nichtstatistischen erscheinungsbasierten Objektmodellierung auf die statistischen Modelle zu übertragen.

Das unüberwachte Training der Oberklassen und die Klassifikation erfolgte bisher durch Clusterbildung anhand von Merkmalen die auf der Karhúnen–Loéwe–Transformation basieren. Für die Berechnung der Merkmale wird also keine Klassenzugehörigkeitsinformation verwendet. Dadurch können sich Oberklassen bilden die Objekte beinhalten die sich visuell sehr ähnlich sind aber für einen menschlichen Betrachter keine gemeinsame Klasse bilden [16, 17]. Durch die Berechnung der Merkmale unter Verwendung der linearen Diskriminanzanalyse (LDA) ist es möglich, genau diese Information mit einfließen zu lassen. Nachteilig dabei ist, dass manuell eine Hierarchie auf den Trainingsobjekten definiert werden muss die dann für die Merkmalsberechnung und Bestimmung der Oberklassen verwendet wird. Abbildung 3 zeigt anhand der COIL–20 Stichprobe exemplarisch, wie sich die Merkmale bei Verwendung von PCA und LDA unterschiedlich verteilen.

Mit der Einführung der statistischen Eigenräume für die generische Objekterkennung ist im Vergleich zum klassischen Ansatz die Möglichkeit der robusten Merkmalsberechnung verloren gegangen. Für eine robuste statistische Klassifikation wurde daher ein zweistufiges Verfahren entwickelt. Im ersten Schritt erfolgt eine Merkmalstransformation und –auswahl aufgrund der Karhúnen–Loéwe–Transformation und auf diesen Merkmalen werden die statistischen Eigenräume, entweder überwacht oder unüberwacht, trainiert.

Da der erste Schritt identisch mit dem ursprünglichen Verfahren ist, wenngleich der resultierende Merkmalsraum im Allgemeinen eine höhere Dimension hat, können die robusten Merkmalsberechnungsverfahren aus dem klassischen Ansatz angewendet werden. Die Güteberechnung der so erhaltenen Merkmale erfolgt jedoch mittels der im zweiten Schritt erstellten statistischen Modelle und nicht mehr anhand von Unterraummodellen wie bisher.

Im Oktober 2001 hat das Projekt HABGOR (Hierachical Appearance-Based Generic Object Recognition) begonnen. In diesem, von der DFG geförderten, grundlagenorienterten Projekt werden verschiedene erscheinungsbasierte Verfahren zur generischen 3-D Objekterkennung untersucht und bewertet. Ziel des Projekts ist es einen statistischen Ansatz zu entwickeln, der auch zur hierarchischen Erkennung von Objekten in Mehrobjektszenen Anwendung finden kann und mit dem eine große Anzahl verschiedener Objekte modelliert werden können.

Eine erscheinungsbasierte generische Objekterkennung beinhaltet eine Klassifikation anhand

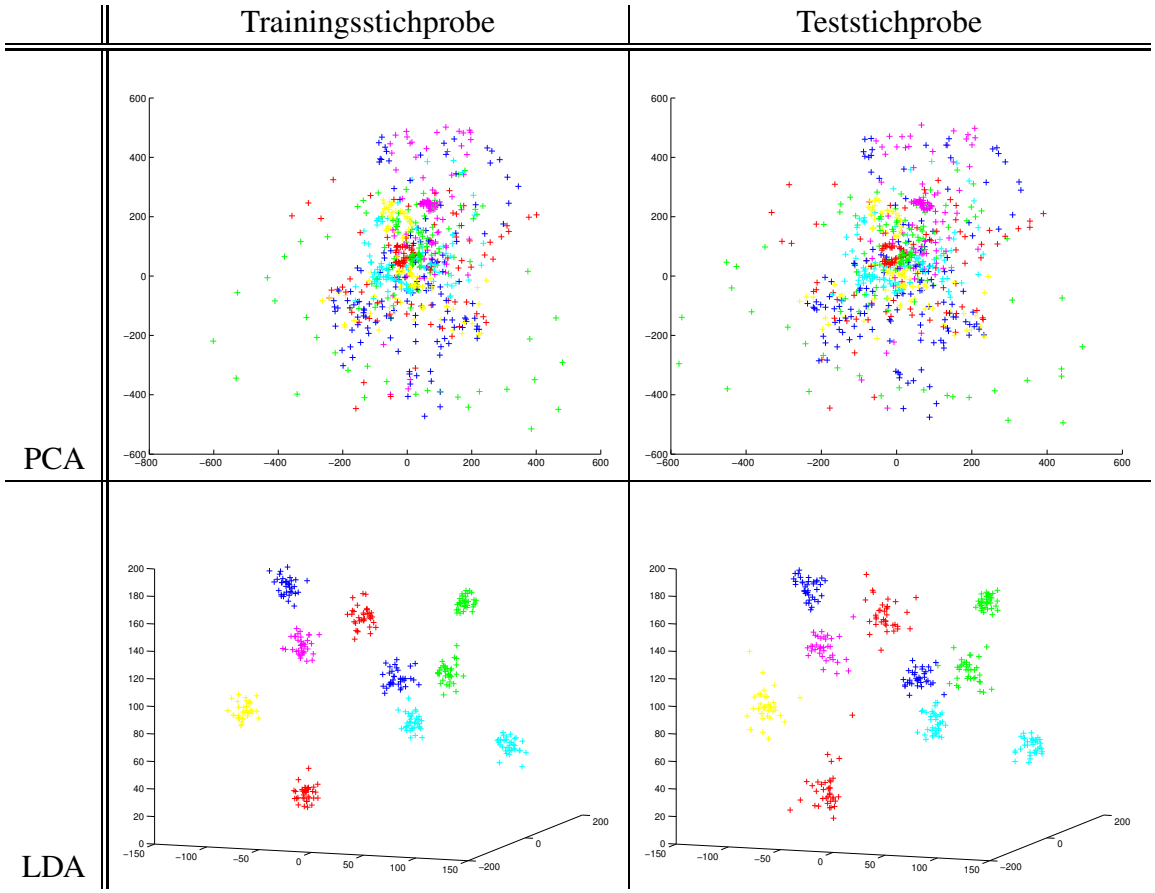


Abbildung 3: Vergleich der Merkmalsgruppierung von PCA und LDA anhand von 10 Klassen aus der COIL–20 Datenbank und disjunkter Trainings– und Teststichprobe.

von Merkmalen, die sich direkt aus Ansichten von Trainingsbildern bestimmen lassen und sich besonders gut zur Klassifikation eignen. Darüberhinaus wird ein hierarchisch strukturierbares Objektmodell benötigt, um eine Einteilung von unbekannten Objekten in Kategorien bekannter Objekte zu ermöglichen. So soll es beispielsweise möglich sein eine unbekannte Tasse in die generische Klasse der Tassen einzuordnen, eine bekannte Tasse soll darüberhinaus als solche erkannt werden.

Im Rahmen dieses Projekts wurden Experimente gestartet um das Verhalten eines erscheinungsbasierten generischen Objekterkennungssystems hinsichtlich der unüberwachten Kategoriesierungsfähigkeit, Hierarchiebildung und der generischen Objekterkennungsfähigkeit zu untersuchen.

Hierzu werden zunächst die Trainingsbilder durch eine PCA in ihren Eigenraum transformiert um eine Datenredukion zu erreichen und anschließend mit PPCA Modellen klassifiziert. Dieses Verfahren kann man für jede entstandene Klasse rekursiv wiederholen, so dass eine Modellhierarchie entsteht. Diese Modelle dienen anschießend zur Klassifikation der Testbilder, die vorher mit der gleichen PCA transformiert werden (siehe Abbildung 4).

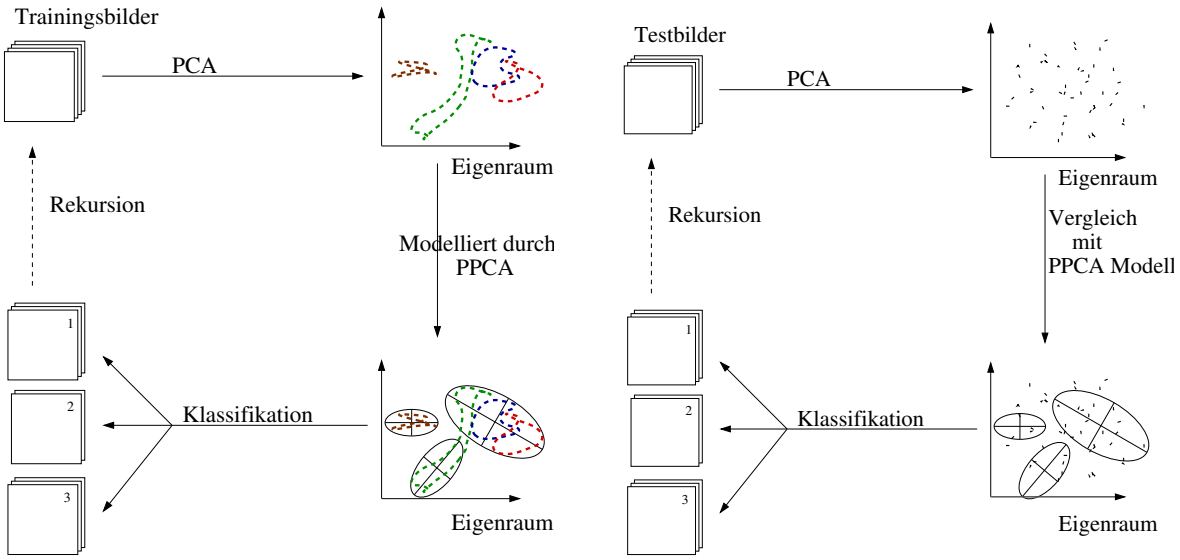


Abbildung 4: Schematische Darstellung einer generischen erscheinungsbasierten Objekterkennung: Training mit Datenreduktion durch PCA, Modellierung mit PPCA und Klassifizierung mit MAP oder ML Schätzung (links), Test durch Transformation mit der gleichen PCA und Vergleich mit trainiertem Modell und Klassifizierung anhand diesem (rechts)

Die Ergebnisse, veröffentlicht in [16] und [17], haben gezeigt, dass die unüberwachte Kategorisierung durchaus sinnvolle Einteilungen der Bilder erstellt, die allerdings nicht immer für den Menschen intuitiv erscheint. In nächsten Jahr sollen deshalb verstärkt überwachte generische Klassifikationsverfahren untersucht werden. Erfolgversprechende Ergebnisse lieferte dabei ein Verfahren, dass nach der PCA eine LDA Transformation durchführt und darauf Modelle generiert.

3.3 Bildanalyse für autonome Systeme

Die optimale Wahl der Kameraparameter für Objektverfolgung und -erkennung ist der zentrale Forschungsgegenstand im Teilprojekt B2 des Sonderforschungsbereichs 603. Durch gezielten Eingriff in den Bildaqusitionsprozess sollen die nachfolgenden Verarbeitungsverfahren mit den bestmöglichsten Sensordaten versorgt werden. Als Optimierungskritierium kommen dabei unterschiedliche Gütemaße zum Einsatz. Beispielsweise ist dies für die Objektverfolgung, die als Zustandsschätzproblem eines dynamischen Systems aufgefasst wird, die bedingte Entropie zwischen Zustand und Beobachtung, die es durch Veränderung der Brennweite vor dem Durchführen der tatsächlichen Bildaufnehme zu minimieren gilt. Durch diese Optimierung erreicht man eine Reduktion der Unsicherheit in der Zustandsschätzung und damit eine Reduktion des Schätzfehlers. Dies wurde bereits für den Spezialfall eines erweiterten Kalman-Filters als Zustandsschätzers demonstriert [10]. Es ist nun gelungen, dies auch für den allgemeinen Fall unter Einsatz eines Partikelfilters zu realisieren. Allerdings schränken die dabei auftretenden und rechenintensiven Monte-Carlo-Schritte zur Approximation von Verteilungsdichten und Integralen

den praktische Einsatz noch ein.

In Zusammenhang einer aktiven Brennweitensteuerung ist es wichtig zu untersuchen, inwieweit die zum Einsatz kommenden Verfolgungsverfahren auf eine Änderung der Brennweite während der Verfolgung reagieren. Für das im Teilprojekt eingesetzte regionenbasierte Verfahren beispielsweise ist es von Bedeutung, bei welcher Brennweite das Verfahren initialisiert, bzw. wie sich relativ dazu die Brennweite während der Verfolgung ändert. Eine Initialisierung bei kleiner Brennweite (also in einem Überblicksbild) und ein anschließendes Heranzoomen an das Objekt erwies sich dabei als vorteilhaft bezogen auf den erzielbaren 3-D-Schätzfehler [39].

Die automatische Nachführung einer Kamera hinter einem sich bewegenden Objekt bedarf im Allgemeinen der expliziten Definition einer entsprechenden Regelstrecke. Das Ziel dabei ist die Aufrechterhaltung der Fähigkeit zur Beobachtung des Objekts durch Fixation, d. h. das bewegte Objekt soll möglichst in der Bildmitte gehalten werden. Eine alternative Vorgehensweise bietet die Übertragung des oben genannte Verfahrens zur Optimierung der Brennweite auf die Steuerung der Kameraparameter Pan und Tilt [40]. Durch Verwendung eines ortsabhängigen Beobachtungsrauschen auf der Bildebene (wachsend mit dem Abstand zum Bildmittelpunkt) erreicht man, dass die Optimierung der bedingten Entropie diejenigen Steuerwinkel für die Kamerasteuerung liefert, die eine fortwährende Fixation auf das Objekt ermöglichen.

Die so genannten Lichtfelder zählt man zu den ansichtenbasierten Bildgenerierungsverfahren. Aus einer Menge von Trainingsansichten einer Szene lassen sich in der Anwendungsphase neue Ansichten der Szene aus beliebigen Blickrichtungen erstellen, und das in fotorealistischer Qualität. In Kooperation der Teilprojekte B2 und C2 des Sonderforschungsbereichs 603 wurden Lichtfelder nun erstmalig als Objektmodell für die 3-D-Objektverfolgung und Lageschätzung erfolgreich eingesetzt [41]. Hierfür wurde eine geeignete Gütefunktion ausgewählt, die durch den Vergleich von Auswertung des Lichtfelds und tatsächlicher Beobachtung die Zustandsschätzung steuert. Abbildung 5 zeigt das Ergebnis einer Verfolgungssequenz anhand der Gegenüberstellung von Originalaufnahme des Objekts und der aus dem Lichtfeld generierten Ansicht, die auf der aktuellen Zustandsschätzung beruht.

Als Anwendungsgebiet für die Evaluation der im Teilprojekt B2 entwickelten Ansätze und Verfahren dient ein Szenario aus dem Bereich der Dienstleistungsrobotik. Mittels der mobilen Plattform MOBSY wird dabei ein Objekt in einer Büroumgebung erkannt, angefahren und gegriffen. Dies bedarf des koordinierten Zusammenspiels von Objekterkennung, Objektverfolgung, Ansichten- und Greifplanung in einem gemeinsamen System. Die Objektverfolgung behält das Objekt bei optimal eingestellten Brennweiten während der Anfahrt der mobilen Plattform im Blickfeld der Kameras. Mit Hilfe der Ansichtenplanung wird sowohl eine robuste Erkennung des zu greifenden Gegenstands sowie eine exakte Ermittlung der Objektlage ermöglicht. Dabei bestimmt eine Greifplanung die optimale Greifposition bezogen auf das Objekt und steuert daraufhin die Plattform und den Greifer entsprechend. In einer weiteren Ausbaustufe werden in dieses Szenario zwei weitere, jedoch statische, Kameras mit einbezogen, die sowohl die Plattform als auch das Objekt beobachten können. Probleme wie beispielsweise Verdeckungen des Objekts aus Sicht der Plattform sollen damit überwunden werden, wobei die Herausforderung in der Fusion der vielfältigen Sensorinformation liegt.

Eine Möglichkeit zur Sensordatenfusion wurde in Kooperation mit der University of San Diego im Rahmen einer BaCaTeC-Förderung entwickelt [11]. Das Verfahren erlaubt die automatische

und dynamische Gewichtung der eintreffenden Sensordaten während der Objektverfolgung, basierend auf dem Verfahren der Demokratischen Integration. Dabei wird die Übereinstimmung der gelieferten Information der einzelnen Sensoren bezogen auf das fusionierte Gesamtergebnisse berücksichtigt und eine schlechte Übereinstimmung mit einem entsprechend niedrigem Gewicht gewertet. Es besteht dabei jedoch die Möglichkeit, dass sich die Sensoren an das globale Ergebnisse durch eine Adaption ihrer internen Parameter anpassen und somit in darauf folgenden Schritten ihr Einfluss bei der Fusion wieder zunimmt.

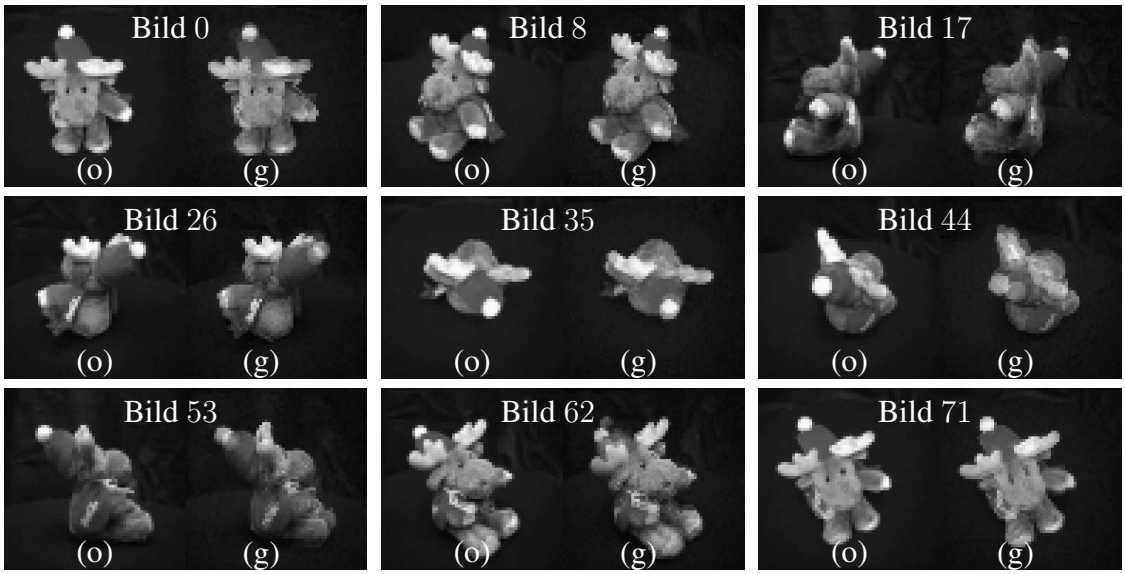


Abbildung 5: Vergleich zwischen Originalbild (o) der Sequenz und dem gerenderten Bild (g) aus dem Lichtfeld basierend auf der aktuellen Zustandsschätzung. Jedes neunte Bild der Sequenz ist angegeben.

3.4 Bildbasierte Modellierung und erweiterte Realität

Das Teilprojekt C2 des Sonderforschungsbereichs 603 beschäftigt sich seit 1998 mit der „Analyse, Codierung und Verarbeitung von Lichtfeldern zur Gewinnung realistischer Modelldaten", und wird in Zusammenarbeit mit dem Lehrstuhl für Graphische Datenverarbeitung (LGDV) behandelt. Das Lichtfeld ist ein Verfahren der bildbasierten Modellierung, das es erlaubt, aus einer Sammlung von Aufnahmen eines Objekts oder einer Szene beliebige neue Ansichten zu generieren. Dazu werden die genauen Parameter der Kamera benötigt, die im Teilprojekt C2 durch Verfahren der „Struktur aus Bewegung" (Structure from Motion) aus dem Bildstrom einer handgeführten Kamera selbst berechnet werden. Dies ist Aufgabe des Lehrstuhls für Mustererkennung, während sich der LGDV um die Visualisierung der Lichtfelder kümmert.

Ein Hauptaugenmerk wurde in 2002 auf die verstärkte Kooperation zwischen den Lehrstühlen gelegt. Anstatt wie vorher den Datenaustausch ausschließlich über eine Dateischnittstelle zu handhaben, wurde eine Klassenbibliothek entwickelt, welche die Verfahren der Rekonstruktion und

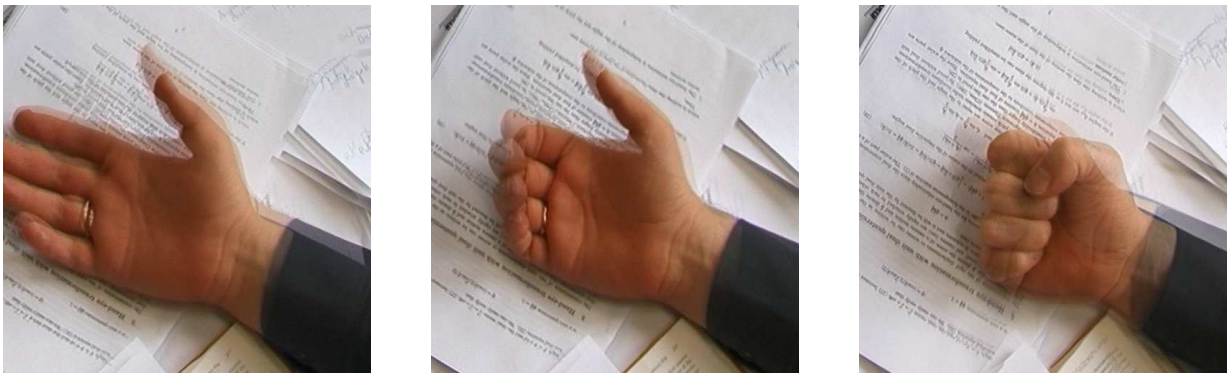


Abbildung 6: Eine Ansicht aus einem dynamischen Lichtfeld zu drei verschiedenen Zeitpunkten visualisiert. Die Position der virtuellen Kamera ist dabei zu jedem Zeitpunkt die gleiche.

Visualisierung integriert und gemeinsame Datenstrukturen zur Verfügung stellt. Diese Plattform, lgf3 [32] genannt, ermöglicht, neue Verfahren mit wenig Aufwand in das Gesamtsystem einzufügen. Die Rückkopplung von Informationen aus der Lichtfeldsynthese, die in Zukunft genutzt werden soll, wird dadurch stark vereinfacht.

Eine der grundlegenden Datenstrukturen in lgf3 ist das Ansichtennetz, das Nachbarschaften zwischen Kamerapositionen definiert. Bei Verwendung von Bildsequenzen ist anfangs nur bekannt, dass zwei aufeinander folgende Bilder ähnliche Ansichten zeigen, also benachbart sind. Daher ist nur zwischen diesen eine Punktverfolgung sinnvoll. Nach einer ersten Kalibrierung können aber weitere ähnliche Ansichten anhand der Entfernung zweier Kameras und deren Blickrichtungen aufgefunden werden. Diese zusätzliche Information kann nun genutzt werden, um weitere Punktkorrespondenzen zwischen vorher unverknüpften Bildern zu finden und so mehr Bilder als vorher auf einmal zu kalibrieren. Dieser zusätzliche Schritt kann die Rekonstruktion deutlich verbessern.

Neben Arbeiten an einer gemeinsamen Softwareplattform, lgf3 [32] genannt, die gemeinsam mit dem LGDV entwickelt und implementiert wurde, konnte mit Untersuchungen am so genannten dynamischen Lichtfeld begonnen werden. Bisher waren Lichtfelder auf die Rekonstruktion von statischen Szenen und Objekten beschränkt. Das dynamische Lichtfeld berücksichtigt allerdings auch zeitliche Veränderungen. Dadurch wird die Dimension der im Lichtfeld parametrisierten Funktion erhöht, was einen größeren Speicher- und Rechenaufwand nach sich zieht, aber vor allem die Rekonstruktion aus einem Bildstrom erschwert, da keine zuverlässigen Punktkorrsepondenzen für die Kalibrierung auf bewegten Objekten zu finden sind.

In [30] wurde ein erstes Modell für dynamische Lichtfelder entwickelt, das verschiedene Zeitschritte als jeweils einzelne, statische Lichtfelder betrachtet. Dabei wird für die Kalibrierung die vereinfachende Annahme gemacht, dass die Zeitschritte als einzelne Bildsequenzen vorliegen. Die Registrierung der daraus resultierenden Lichtfelder erfolgt durch je eine gemeinsame Ansicht in zwei aufeinander folgenden Bildsequenzen, deren Kameraparameter also als gleich angenommen werden können. Abbildung 6 zeigt drei Ansichten zu verschiedenen Zeitpunkten von jeweils der gleichen Kameraposition aus gesehen, die aus einem derartigen dynamischen

Lichtfeld rekonstruiert wurden.

Zwei Studienarbeiten, die in 2002 entstanden, untersuchen mögliche Anwendungsgebiete von Lichtfeldmodellen. In der ersten dieser Arbeiten werden Lichtfelder als Objektmodelle verwendet, um die Lage verschiedener Objekte erkennen und verfolgen zu können [41]. Die zweite Arbeit beschäftigt sich mit Erweiterter Realität (Augmented Reality), also der Ergänzung realer Umgebungen um virtuelle Objekte. In diesem Fall werden von den Köpfen verschiedener Personen Lichtfeldmodelle rekonstruiert, um schließlich die Gesichter anderer Personen in einer Bildsequenz durch entsprechende Ansichten aus diesen Lichtfeldern zu ersetzen.

Die Arbeitsschwerpunkte im Bereich erweiterte Realität (Augmented Reality) lagen im vergangenen Jahr auf der Echtzeitverarbeitung von Stereobildern und auf der Selbstkalibrierung eines Stereokamerasystems aus einer Bildfolge.

Betrachtet wurde die Berechnung von dichten Tiefenkarten aus Stereobildern in Echtzeit für Anwendungen in der erweiterten Realität [28]. Beispielbilder sind auf der Augmented Reality Seite im WWW zu finden. Das in [28] vorgestellte Verfahren kann zudem zur Verbesserung der Bildqualität in Lichtfeldern (siehe SFB 603, Teilprojekt C2) verwendet werden. In Kooperation mit dem Teilprojekt B6 (Abschnitt 3.6) wurde die Tiefenkartenberechnung auf medizinische Bilddaten angewandt, die mit Hilfe eines Endoskopieroboters erzeugt wurden [29].

Da bisher zur Tiefenkartenberechnung ein *kalibriertes* Stereokamerasystem vorhanden sein muss, wird derzeit an Verfahren gearbeitet, die die relative Position und Orientierung der beiden Kameras allein aus Bilddaten berechnen. Voraussetzung hierfür ist die 3-D Rekonstruktion der Szenengeometrie und der Kamerapositionen aus den Bildern der beiden Kameras mit Hilfe von Verfahren, wie sie auch in Teilprojekt C2 des SFB 603 entwickelt wurden. Die Rekonstruktionen werden anschließend registriert und können somit zur Berechnung der relativen Position und Orientierung der beiden Kameras des Stereosystems verwendet werden.

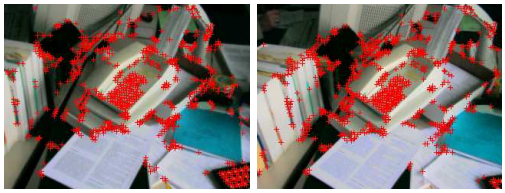
Ein Beispiel für die Registrierung zweier 3-D Rekonstruktionen zeigt Abbildung 7.

3.5 Das Projekt VAMPIRE

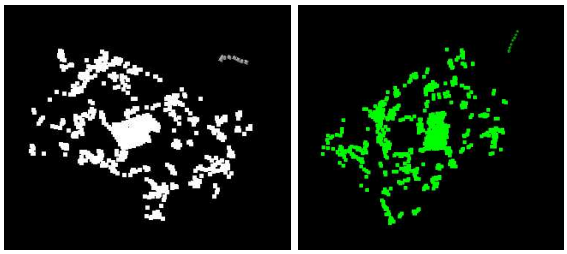
Im Juni 2002 begann das EU-Projekt VAMPIRE (Visual Active Memory Processes and Interactive Retrieval), an dem neben dem Lehrstuhl für Mustererkennung auch die Universität Bielefeld, die Technische Universität Graz und die University of Surrey beteiligt sind. Hauptziel dieses Projekts ist die automatische Analyse von Videosequenzen über einen längeren Zeitraum, bei der neben der Erkennung von Objekten und Bewegungsabläufen besonders das Lernen neuer Objekt- und Bewegungsmodelle im Vordergrund steht.

Als Anwendungsszenario ist eine Büroumgebung vorgesehen, in der sich ein oder mehrere Benutzer bewegen und typische Handlungen durchführen, also zum Beispiel telefonieren oder Bücher aufheben, lesen und ablegen. Die gesamte Szene wird dabei von mehreren Kameras beobachtet, die entweder als Teil einer Augmented-Reality-Ausrüstung (siehe Abbildung 8) von den Benutzern getragen werden, oder an festen Orten in der Szene aufgestellt sind. Nach der Analyse des aufgenommenen Bildmaterials soll das Gesamtsystem in der Lage sein, Anfragen bezüglich der beobachteten Objekte und Handlungen entgegenzunehmen und die Antworten mit Hilfe der Augmented-Reality-Ausrüstung zu visualisieren. Sucht ein Benutzer beispielsweise nach einer Tasse, die durch ein abgelegtes Buch verdeckt wird, so kann die Position der Tasse in

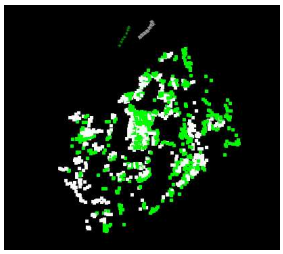
(a) Original Stereobilder



(b) verfolgte Punkte



(c) 3-D Rekonstruktionen



(d) Registrierte 3-D Rekonstruktionen

Abbildung 7: Hier ist die Abfolge der Arbeitsschritte bei der Selbstkalibrierung eines Stereokamerasystems zu sehen: Nach der Aufnahme werden Punkte verfolgt, die zur 3-D Rekonstruktion verwendet werden können. Die beiden getrennten Rekonstruktionen werden anschließend regisitriert.

der Videobrille des Benutzers angezeigt werden.

Ein zentraler Bestandteil der Systemarchitektur beim Projekt VAMPIRE (siehe Abbildung 8) ist das Visual Active Memory. Es speichert neben den Bilddaten auch die aus diesen extrahierten Informationen in mehreren Abstraktionsebenen. Eine Vielzahl von Modulen verarbeitet die Daten der vier Ebenen des Visual Active Memory, mit dem Ziel, das für die Beantwortung von inhaltsbasierten Abfragen notwendige Wissen zu sammeln.

Der Lehrstuhl für Mustererkennung ist im Projekt VAMPIRE verantwortlich für die Objektverfolgung und Segmentierung, sowie beteiligt an den Modulen Lokalisation, Objekterkennung, Bewegungsanalyse und Visualisierung. Einen Schwerpunkt bildet dabei der Themenbereich 3-D Rekonstruktion und bildbasierte Objektmodelle. Die 3-D Rekonstruktion der beobachteten Szene kann einerseits bei der Selbstlokalisation der Augmented-Reality-Ausrüstung verwendet werden. Die bei der Rekonstruktion ermittelten 3-D Daten sind aber auch für Objektverfolgung und Bewegungsanalyse nützlich. Die bildbasierten Objektmodelle kommen bei der modellbasierten Objektverfolgung, der Objekterkennung und der Visualisierung zum Einsatz.

Nach einer Untersuchung der bereits am Lehrstuhl vorhandenen Verfahren für die 3-D Rekonstruktion und die Erzeugung von bildbasierten Objektmodellen (siehe Abschnitt 3.4) wurde mit der Anpassung dieser Verfahren an die Erfordernisse im Projekt VAMPIRE begonnen. So verlangt die Augmented-Reality-Ausrüstung die Verwendung schneller Algorithmen, um auf

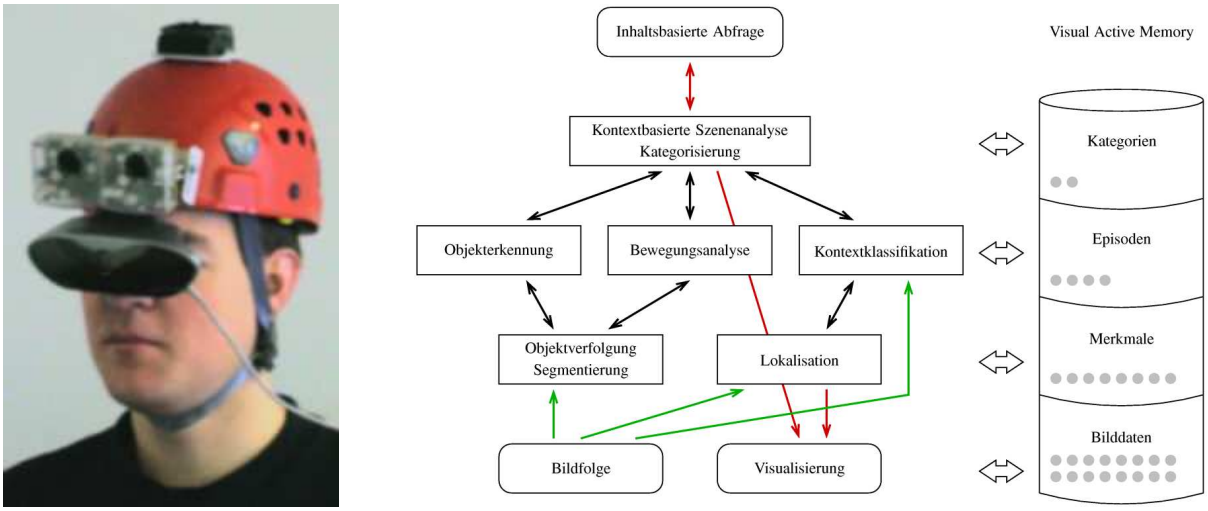


Abbildung 8: Prototyp des Helms der Augmented-Reality-Ausrüstung (links), Systemarchitektur des Projekts VAMPIRE (rechts)

Aktionen und Anfragen des Benutzers in angemessener Zeit reagieren zu können. Die Punktverfolgung, welche als erster Schritt bei der 3-D Rekonstruktion für die weitere Verarbeitung unabdingbar ist, konnte bereits erfolgreich beschleunigt werden.

Neben der 3-D Rekonstruktion ist die Objektverfolgung ein weiterer Schwerpunkt im Projekt VAMPIRE. Dabei ist hier die Objektverfolgung eng mit der Objekterkennung und der Bewegungsanalyse verbunden. Die Verbindung zum Objekterkennungsmodul hängt damit zusammen, dass nach einer Detektion einer Bewegung keine robuste modellbasierte Objektverfolgung möglich ist, da zu diesem Zeitpunkt nicht bekannt ist, um welches Objekt es sich handelt. Dieses Wissen wird von der Objekterkennung geliefert. Bis eine Klassifikation stattgefunden hat, muss bei der Objektverfolgung eine datengetriebene Technik verwendet werden, um das Objekt ohne a priori Wissen zu verfolgen. Dieser Punkt wurde als erstes untersucht.

Wichtige Arbeit auf diesem Gebiet wurde bereits von 3.3 im SFB 602 Teilprojekt B2 geliefert. Prinzipiell bietet sich hierbei ein 2-D Schablonenvergleichsverfahren an, welches die Bewegungsparameter so abschätzt, dass der quadratische Fehler der Bildintensitäten zwischen dem verfolgten Objekt und der Schablone minimal ist. Die Schablone erhält man aus einem initialen Bild bzw. als Ergebnis der Bewegungsdetektion. Da die Verfolgung in Echtzeit (Verarbeitung von 15-30 Bilder pro Sekunde) durchgeführt wird, kann auf optimierungsbasierte Verfahren nicht zurückgegriffen werden. Stattdessen werden die Bewegungsparameter unter Verwendung von Grauwertdifferenzen zwischen dem aktuellen und dem vorherigen Bild durch ein lineares System approximiert. Ein solcher Ansatz wurde von Frederic Jurie und Michel Dhome (Hyperebenen Ansatz) entwickelt und am Lehrstuhl implementiert. Dabei wurde das System so aufgebaut, dass verschiedene Bewegungsarten (in der Bildebene) erkannt werden: reine Translation, Translation und Rotation, Translation, Rotation und Skalierung sowie affine Bewegung. In Abbildung 9 werden Beispiele für diese Bewegungen dargestellt. Experimente haben gezeigt, dass der Berechnungsaufwand sehr gering ist und somit eine Objektverfolgung in Echtzeit durchgeführt

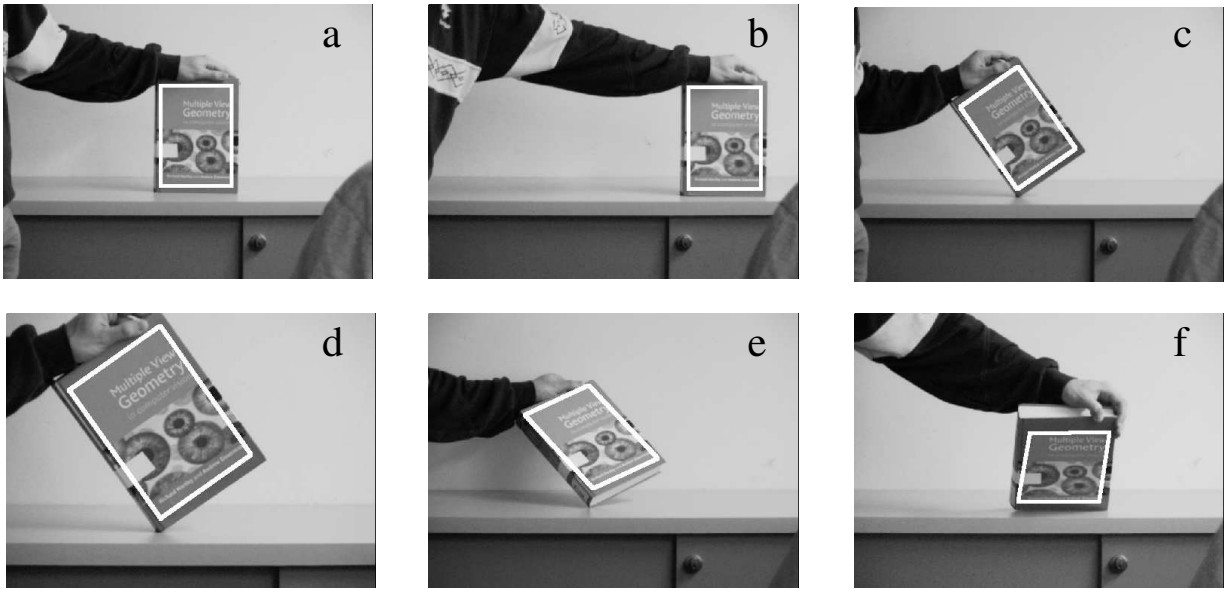


Abbildung 9: Datengetriebene Objektverfolgung in 2-D. (a) Schablone (markiert durch ein Viereck), (b) Translation, (c) Rotation, (d) Skalierung, (e,f) affine Verzerrung

werden kann.

Neben der Objekterkennung ist die Bewegungsanalyse mit der Objektverfolgung eng verbunden. Ziel ist es hierbei, anhand der erkannten Bewegung typische Bewegungsabläufe für Aktionen zu lernen bzw. diese zu erkennen. Beispiele für solche Aktionen sind: das Abheben eines Telefonhörers, das Trinken aus einer Tasse, die Verwendung eines Hefters. Die Entwicklung des Bewegungsanalysemodul wird vom Lehrstuhl für Neuroinformatik der Universität Bielefeld durchgeführt.

3.6 Medizinische Anwendungen

Im Teilprojekt B6, „Rechnergestützte Endoskopie des Bauchraums“, des Sonderforschungsbereichs 603 (SFB 603) wurden in Zusammenarbeit mit der Chirurgischen Universitätsklinik die Arbeiten zur Einführung einer Rechnerunterstützung in der minimal-invasiven Chirurgie weitergeführt. Der Schwerpunkt der Arbeiten im Jahr 2002 bestand im Aufbau des Gesamtsystems, der Kalibrierung des Roboterarms AESOP 3000 und ersten Experimenten mit dem System im Labor.

Die Erweiterung der Arbeiten zur Glanzlichtsubstitution mit Lichtfeldern [35, 33] fand Anfang des Jahres statt. Durch Berechnung von Signal-zu-Rausch Verhältnissen auf synthetischen Daten konnte nachgewiesen werden, dass die Substitution (auf synthetischen Bildern) das Signal-zu-Rausch Verhältnis erhöht [34].

Daran schlossen sich Experimente mit dem Roboterarm AESOP 3000 der Firma Computer Motion Inc. an. Sechs Winkelwerte und ein Längenwert sind aus dem Roboterarm über eine serielle Schnittstelle auslesbar. Die Software zur Berechnung der Position und Orientierung der

Endoskophalterung auf Grund der Kinematik des Roboterarms wurde freundlicherweise von Computer Motion Inc. zur Verfügung gestellt, musste allerdings noch an das Betriebssystem Linux angepasst werden.

Die Transformation von der Endoskophalterung zur Endoskopspitze sowie die Blickrichtung des Endoskops müssen zur Berechnung der Endoskopposition bekannt sein. Der Winkel der Endoskop-Optik wurde zusätzlich zu den genannten Informationen in die Kinematik des Roboterarms integriert, da bei endoskopischen Operationen des Bauchraums vor allem 30 Grad Optiken (d. h. die Optik ist an der Spitze um 30 Grad abgeknickt) zum Einsatz kommen, um die Übersicht zu erhöhen.

Drei unbekannte Faktoren der Kinematik sind bei jeder Operation bzw. jeder Experimentserie zu bestimmen: Länge des Endoskops von der Halterung bis zur Spitze, Winkel zwischen Kamerakopf und Endoskop sowie Einspannwinkel des Endoskops in der Endoskophalterung. Zunächst wird die Länge von der Halterung bis zur Spitze des Endoskops von Hand gemessen. Zur Bestimmung des Winkels zwischen Kamerakopf und Endoskopoptik wird eine Kerbe, die nach „oben“ zeigt, in der Endoskopoptik automatisch detektiert. Den Einspannwinkel erhält man durch zwei Aufnahmen eines Kalibriermusters und Berechnung der relativen Kamerapositionen.

Mehrere Experimente im Labor führten zu dem beschriebenen Ablauf zur Kalibrierung des Roboterarms. Endoskoppositionen konnten unter OP-nahen Bedingungen (im Labor) bestimmt und Lichtfelder daraus erzeugt werden. Durch Aufnahme eines Kalibriermusters konnten außerdem die Fehler bei der Positions- und Orientierungsberechnung bestimmt werden. Überraschend hoch sind die Fehler vor allem an Wendepunkten einer Kamerafahrt. Hier sind die Werte bei der Positionsbestimmung um bis zu 7 mm falsch, bei der Orientierung beträgt die Abweichung bis zu 4 Grad pro Achse. Ursache der Fehler ist die Endoskophalterung des Roboterarms, die nicht im Hinblick auf exaktes Positionieren konstruiert wurde.

Die Darstellungsqualität von Lichtfeldern wird durch Integration von Szenengeometrie (Tiefenkarten) drastisch erhöht. Wegen der fehlerbehafteten Positions- und Orientierungsbestimmung lässt sich Tiefeninformation allerdings nicht direkt über ein Stereoverfahren bestimmen. Die Erstellung von Tiefenkarten wird in der ersten Hälfte des Jahres 2003 bearbeitet.

Um die fehlerbehafteten Positionen und Orientierung des Endoskops zu optimieren, können nachträglich wenige gute Punkte von Bild zu Bild verfolgt werden. Korrespondierende Punkte können dann dazu genutzt werden die Positionen und Orientierung nicht-linear zu optimieren [29]. Die Experimente hierzu wurden zusammen mit Jochen Schmidt (vergleiche section 3.4) durchgeführt.

Die beiden Komponenten des Gesamtsystems (Roboterarm AESOP 3000 und Videoendoskopieturm mit Computer) sowie ein erstes Ergebnisbild sind in Abbildung 10 dargestellt.

3.7 Die Virtuelle Hochschule Bayern

Seit Juni 2002 wird am Lehrstuhl für Mustererkennung eine Vorlesung mit dem Titel „Rechnersehen mit Anwendungen in der Augmented Reality sowie beim bildbasierten Rendering“ für die virtuelle Hochschule Bayern (vhb – http://www.vhb.org) erarbeitet. Das Projekt wird mit Mitteln der High-Tech-Offensive Bayern gefördert.

Fachliche Schwerpunkte des Kurses sind die Themen 3D-Rekonstruktion (3.4), Augmented

Abbildung 10: Das neue Endoskopiesystem inklusive Computer und zweitem Monitor (links), der sprachgesteuerte Roboterarm AESOP 3000 (mitte), ein gerendertes Bild aus einem mit dem Gesamtsystem im Labor erzeugten Lichtfeld (rechts).

Reality (http://www5.informatik.uni-erlangen.de/ar) und Lichtfelder (3.4), an denen aktiv am Lehrstuhl geforscht wird.

Das Lehrangebot beschränkt sich ausschließlich auf Methoden des Internets. Dazu wurde ein didaktisches Konzept entworfen, dass für diese Spezialvorlesung passend zugeschnitten ist. Dabei wurde natürlich auch darauf geachtet, dass das Konzept umsetzbar ist im Bezug sowohl auf technische Möglichkeiten bzw. Beschränkungen als auch auf die zeitlichen Vorgaben bei der Entwicklung.

Die Schwerpunkte des didaktischen Konzepts sind Interaktivität sowie Betreuung und Schaffung eines sozialen Umfeldes beim Lernen.

Die Möglichkeit durch Interaktivität Sachverhalte darzustellen bildet einen erheblichen Vorteil der elektronischen Medien gegenüber Printmedien. Den Studenten und Studentinnen wird die Möglichkeit geboten selbst verschiedene Algorithmen an Bildern zu testen und so den Lernprozess selbst zu kontrollieren. So wird das Verständnis vertieft und der Behaltensgrad des Gelernten erhöht. Der Student, die Studentin lernt Stärken und Schwächen einzelner Verfahren an selbstgewählten Beispielen kennen und soll somit auch Kompetenz bei der häufig schwierigen Wahl von Parametern erwerben. Durch die Interaktivität wird der/die Lernende aktiv am Lernprozess beteiligt. Gleichzeitig soll er/sie motiviert werden und ein Interesse für die Ergebnisse soll geweckt werden.

Auch die Betreuung ist für den Lernprozess sehr wichtig. Der Student/die Studentin darf sich in keinem Fall alleine gelassen fühlen. Denn dann besteht die Gefahr, dass bei auftretenden Problemen die Motivation nachlässt, was sich negativ auf die Lernleistungen auswirkt. Selbst wenn keine Probleme auftreten, so macht Lernen in der Gruppe doch mehr Spaß und fördert

so automatisch die Motivation. Deswegen sieht das Konzept die folgenden Möglichkeiten zur Kommunikation vor:

- E-Mail: Durch verschiedene Verteilerlisten können Studenten mit dem Betreuer, der gesamten Gruppe der Kursteilnehmer oder in selbst festgelegten Untergruppen diskutieren.
- Newsgroups: Um Diskussionen im Internet zu führen, eignen sich Newsgroups am besten, denn dort kann der Verlauf der Diskussion am besten nachvollzogen werden und man kann gezielt auf einzelne Aussagen von anderen Teilnehmern eingehen.
- Chat: Ein Chat bietet die Möglichkeit synchron zu diskutieren und dies fast in Echtzeit. So kann durch einen Dialog ein Problem schneller als bei asynchroner Diskussion erörtert werden.

Das didaktische Konzept sowie die ausgewählten Techniken wurden in [37] auf dem 1. Workshop der Gesellschaft für Informatik Fächergruppe Didaktik der Informatik in Witten-Bommerholz im Oktober vorgestellt.

Nach der Erstellung des didaktischen Konzeptes und der Festlegung der zu verwendeten Techniken wurden die ersten beiden Kapitel von insgesamt acht umgesetzt. Zur Umsetzung eines Kapitels gehört natürlich die Erstellung des entsprechenden Skripts. Des weiteren müssen die interaktiven Elemente für die jeweiligen Abschnitte programmiert werden. Für eine möglichst einfache Nutzung ist das Dokument an möglichst viele Stellen mit Links versehen. Diese haben mehrere Aufgaben: sie dienen zum Springen innerhalb des Dokuments an verschiedene Stellen, z. B. zur Wiederholung von bestimmten Begriffsdefinitionen, zum Erreichen der Experimentierumgebung in der entsprechenden Stelle oder wenn eine Quelle zitiert wird, kann man sich den entsprechenden Eintrag im Literaturverzeichnis einfach mit einem Klick ansehen. Zusätzlich wird an geeigneten Stellen auch auf andere Online-Quellen verwiesen.

Nach diesen Erfahrungen bei der Umsetzung können nun die restlichen Kapitel zügig im Jahr 2003 realisiert, getestet und evaluiert werden.

3.8 Statistische Modellierung von Daten

Der Sonderforschungsbereich SFB396 „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 (BN). 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 standen die Weiterentwicklung des bayesnetzbasierten Reglers, die Steigerung der Robustheit und die Modellierung von nichtlinearen Kennlinien.

Der bayesnetzbasierte Regler, der im Jahresbericht 2001 vorgestellt wird, basiert auf einem Markov-Modell n-ter Ordnung. Die Versuche mit diesem Modell basieren auf einem Workaround, in dem die Werte älterer Zeitscheiben unverändert zur aktuellen Zeitscheibe durchgereicht werden. Dieses Verfahren erfordert jedoch zusätzliche Knoten, die die Fähigkeit des Reglers in Echtzeit zu reagieren reduziert. Deswegen wurde die verwendete BN-Toolbox so erweitert, dass auch Markov-Modelle höherer Ordnung direkt implementiert werden können. Die Erweiterung wurde durch die Regelung einer simulierten Regelstrecke dritter Ordnung getestet.

Um die Performance des neu entwickelten Reglers besser beurteilen zu können wurden Vergleiche mit einem PI-Regler, der nach dem Verfahren nach Ziegler-Nichols eingestellt wurde, und mit einem Dead-Beat Regler, angestellt. Diese beiden Reglertypen wurden deswegen ausgewählt, da die Einstellungen nach Ziegler und Nichols immer noch häufig verwendet werden und der Dead-Beat Controller ein Ausregeln der Störung in minimaler Zeit erlaubt. Es zeigt sich, dass der Bayesregler eine bessere Performance zeigt, als der PI-Regler. Die Performance wurde dabei mit dem quadratischen Fehlersumme beurteilt. Im Vergleich mit dem Dead-Beat Regler werden können fast die gleichen Ausregelzeiten bzw. quadratische Fehlersummen mit dem bayesnetzbasierten Regler erzielt werden. Bei diesem Vergleich sollte allerdings beachtet werden, dass die Parameter des bayesnetzbasierten Reglers durch Training gewonnen werden, während bei einem Dead-Beat Regler die Parameter mathematisch berechnet werden, wobei eine mathematische Beschreibung der Regelstrecke vorausgesetzt wird.

Ein zweiter Schwerpunkt war die Steigerung der Robustheit der Modelle. Die Daten, die uns von kooperierenden Teilprojekten am LFT und LKT zur Verfügung gestellt werden, basieren teilweise auf Versuchsplänen, um Versuchskosten möglichst niedrig zu halten. Bedingt dadurch werden nicht alle Kombinationen von Eingabeparametern gemessen. Dies kann bei ungeschickter Modellierung zu Parametern führen, die auf nie beobachteten Parameterkombinationen beruhen, und deren Training daher unmöglich ist. Andererseits repräsentieren diese diese Parameter auch meist Wechselwirkungen höherer Ordnung, die mit den Versuchsplänen nicht erfasst werden.

Es wurden kritische Modellstrukturen ermittelt und Ideen entwickelt, wie diese Strukturen vermieden werden können. Basierend auf diesen Ideen wurde die Zugkraft, die zum Trennen von zwei verschweißten Blechen erforderlich ist, in Abhängigkeit vom Anstellwinkel und dem Versatz des Laserstrahls modelliert. Bild 11 stellt die Abhängigkeit der Zugkraft vom Nahtversatz dar. Es zeigt sich, dass der Verlauf exakt gelernt wurde und dass auch Vorhersagen für Eingaben möglich sind, die nicht während des Trainings präsentiert wurden.

Zusätzlich wurde angefangen häufig auftretende Nichtlinearitäten zu modellieren, z. B. eine Sättigungskennlinie, bei der die Ausgabe durch ein Maximum begrenzt wird, und eine Hystereseschleife. Dieses Forschungsvorhaben wird einer der Schwerpunkte im Jahr 2003 sein.

4 Sprachverstehen

Leitung: E. Nöth
(J. Adelhardt, W. Fentze, C. Frank, C. Hacker, M. Levit, R. Shi, S. Steidl, G. Stemmer, V. Zeißler)

Die inhaltlichen Schwerpunkte der Forschungsaktivitäten zur Sprachverarbeitung bilden das ma-

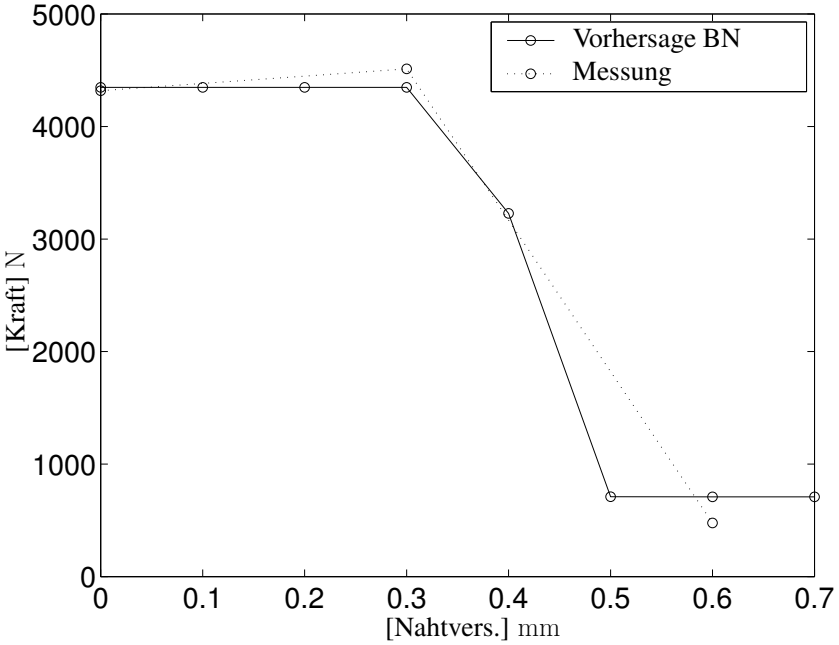


Abbildung 11: Vorhersage der Zugkraft

schinelle Erkennen und Verstehen gesprochener Äußerungen sowie Fragestellungen des multimodalen Mensch-Maschine-Dialogs. Die Arbeiten im Berichtsjahr können zwei anwendungsorientierten Projekten zugeordnet werden: Erkennung von spontaner Sprache und Emotionen im Rahmen des *PF-STAR*–Projekts sowie die Entwicklung des multimodalen Dialogsystems *SmartKom*.

Im von der Europäischen Union geförderten Projekt *PF-STAR* sollen Grundlagen für zukünftige Forschungsaktivitäten im Bereich der Mensch-Technik-Interaktion gelegt werden. Dazu werden vor allem drei Bereiche untersucht: Übersetzung gesprochener Sprache, automatische Erkennung und Ausdruck von Emotionen sowie die Entwicklung von Algorithmen zur Verarbeitung von Kindersprache. Am *PF-STAR*–Projekt sind insgesamt sieben Arbeitsgruppen aus mehreren Universitäten und Forschungseinrichtungen beteiligt. Der Lehrstuhl bearbeitet die Bereiche Emotionserkennung und Erkennung von Kindersprache. Weitere Forschungstätigkeiten, die über den Rahmen des *PF-STAR*–Projekts hinaus gehen, beschäftigen sich mit der Verbesserung der automatischen Erkennung spontaner Sprache auch von nicht-nativen Sprechern.

In dem vom BMBF geförderten Projekt *SmartKom* werden Konzepte für neuartige Formen der Mensch-Technik-Interaktion durch die Entwicklung eines Demonstrationssystems bereits praktisch 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

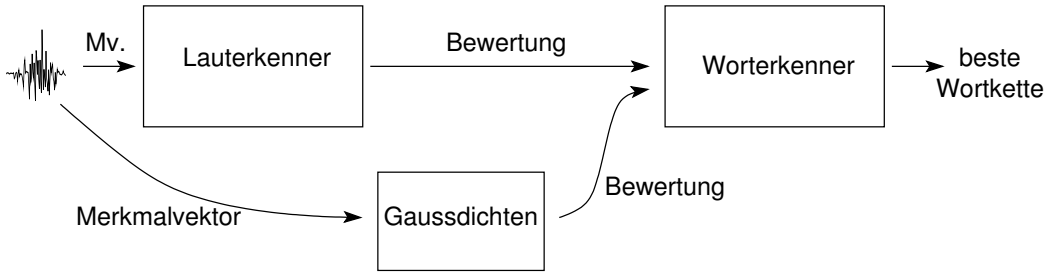


Abbildung 12: Integration der Zustandsbewertungen eines Lauterkenners in einen Worterkenner

Arbeitsgruppen aus mehreren Universitäten, Großforschungseinrichtungen und Firmen beteiligt. Der Lehrstuhl bearbeitet die Bereiche Prosodie–, Mimik– und Gestik–Interpretation.

4.1 Erkennung spontaner Sprache

Das Spracherkennungssystem des Lehrstuhls wurde im Berichtsjahr vor allem im Bereich der akustischen Modellierung, d.h. bei der statistischen Repräsentation der Sprachlaute weiter verbessert.

Das Baum-Welch-Training der akustischen Modelle konnte erheblich beschleunigt werden, indem die Trainingsäußerungen nicht mehr nacheinander, sondern parallel auf mehreren Rechnern eingelesen und verarbeitet werden. Zum Abschluss jeder Trainingsiteration werden die auf den einzelnen Rechnern akkumulierten Bewertungen miteinander verknüpft. Zur Kommunikation zwischen den Prozessen wird die verbreitete PVM (Parallel Virtual Machine) Software eingesetzt.

Eine seit langem bekannte Schwäche der Hidden Markov Modelle (HMM), dem Standardansatz zur akustischen Modellierung in der automatischen Spracherkennung, ist die Annahme der zustandsbedingten statistischen Unabhängigkeit zwischen den Merkmalvektoren. Das hat die unerwünschte Folge, dass ein HMM hohe Bewertungen für Laute haben kann, auf denen es gar nicht trainiert wurde. Viele Ansätze, das Problem zu lösen, führen zu einem drastisch erhöhten Rechenaufwand oder benötigen zuviel Trainingsdaten. Im Berichtsjahr wurde eine Methode entwickelt, die Bewertungen eines sehr einfachen Einzellauterkenners in die akustischen Modelle des Spracherkenners zu integrieren. Dieses Vorgehen ist in Abbildung 12 dargestellt. Dadurch wird die Unabhängigkeitsannahme zwar nicht vollständig aufgehoben, aber den HMM-Zuständen steht Information über die zurückliegenden Merkmalvektoren zur Verfügung und die Akkuratheit des Spracherkenners verbessert sich. Der zusätzliche Rechenaufwand ist relativ gering, da der Dekodierungsalgorithmus für den Einzellauterkenner sehr schnell ist. In naher Zukunft soll evaluiert werden, ob noch Verbesserungen erzielt werden können, wenn statt Lauten andere Wortuntereinheiten eingesetzt werden.

Im Rahmen einer Diplomarbeit wurde die akustische Modellierung für zwei Sprechergruppen näher untersucht, die von den meisten aktuellen Spracherkennungssystemen nur schlecht verstanden werden: nicht-native Sprecher und Kinder. Die schlechte Erkennungsleistung hat meh-

rere Ursachen, die wohl bedeutendste ist die im Vergleich zu Erwachsenen und Muttersprachlern wesentlich geringere Menge von Daten, die für das Training vorhanden sind. Für die Auswertungen benötigte Sprachdaten von Kindern wurden am Ohm-Gymnasium in Erlangen gesammelt. Für die Experimente mit nicht-nativen Sprachdaten standen noch Äußerungen von Deutschen, die Englisch sprechen aus dem VERBMOBIL-Projekt zur Verfügung. Um die negativen Auswirkungen der geringen Datenmengen zu mindern, wurden die akustischen Modelle der beiden Sprechergruppen jeweils mit robust trainierten HMM von Erwachsenen bzw. Muttersprachlern interpoliert. Für die Interpolation wird eine wesentlich kleinere Stichprobe benötigt als für ein komplettes Neutraining der Modelle. In den Auswertungen konnte für beide Sprechergruppen eine Verbesserung der Akkuratheit des Spracherkenners durch die Interpolation gezeigt werden. Weitere Experimente in diesem Gebiet sollen untersuchen, wie gut die automatische Unterscheidung zwischen den Sprechergruppen möglich ist, so dass jede Äußerung mit einem für den jeweiligen Sprecher optimal eingestellten Spracherkennungssystem verarbeitet wird.

4.2 Das SmartKom–Projekt

Ziel des Projektes *SmartKom* (*http://www.smartkom.org/*), das als eines von vier Leitprojekten durch das BMB+F ins Leben gerufen wurde, ist die Entwicklung eines multimodalen multimedialen Dialogsystems. Der Lehrstuhl für Mustererkennung bearbeitet die vier Teilprojekte *Mimikerkennung*, *Erkennung prosodischer Phänomene*, *Emotionserkennung in der Stimme* und *Gestenanalyse & Stifteingabe*.

Das Prosodiemodul, in dem sowohl die Erkennung prosodischer Phänomene als auch die Emotionserkennung integriert ist, wurde im vergangenen Jahr weiterentwickelt. Für bessere Robustheit beim Training der prosodischen Klassifikatoren ist eine Normierung der Merkmale notwendig, die die intrinsischen Faktoren und den individuellen Sprechereinfluss eliminieren soll. Das parametrische Modell, das die verwendete Normierung untermauert, ist in Abbildung 13 dargestellt.

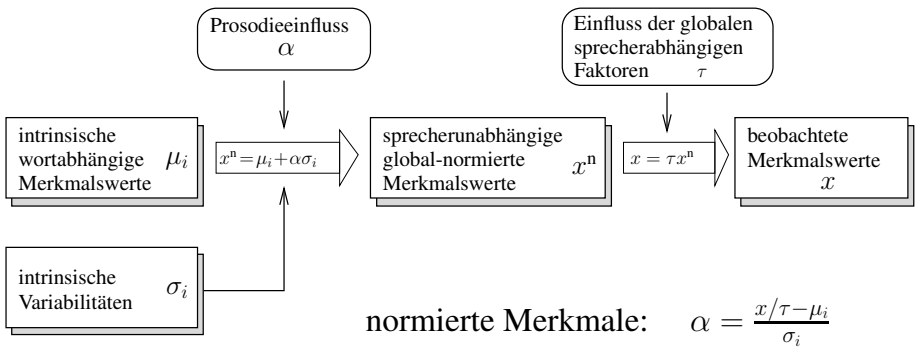


Abbildung 13: parametrisches Modell für die Merkmalsnormierung

Um diese Normierung zu verbessern, wurden mehrere Verteilungsmodelle für die Dauer und

Energiemerkmale untersucht, was ausfürlich in [38] beschrieben ist. Die Tests mit dem optimalen Modell ergaben eine signifikante Verbesserung der Erkennungsraten. Die neuen Normierungsverfahren wurden anschließend in das Prosodiemodul integriert.

Eine weitere Möglichkeit zur Verbesserung der Erkennungsraten stellt die Untersuchung und Integration neuer Merkmale dar, wie z.B. der sogenannten Jitter-Shimmer Merkmale (basierend auf periodensynchroner Grundfrequenz) oder satzglobale prosodische Merkmale. Die im Vorfeld durchgeführten Experimente zeigten ein Verbesserungspotenzial von mehr als 10 % relativ. Um Jitter-Shimmer Merkmale zu berechnen ist allerdings eine Bestimmung der periodensynchronen Grundfrequenz notwendig. Ein dafür verwendetes Programm wurde deswegen angepasst, um seine Integration in das Prosodiemodul zu ermöglichen. Als Klassifikator wird in diesem Programm ein neuronales Netz eingesetzt. Für dessen Training wurde die im Internet veröffentlichte Bagshaw-Stichprobe (*http://www.cstr.ed.ac.uk/~pcb/*) anhand des zeitsynchronen Laryngogramm-Signals mit korrekten Periodenanfängen gelabelt.

Bei den Sprachaufnahmen mit einem Raummikrofon (wie z.B. im *SmartKom Public*-Szenario: *http://www.smartkom.org/public_de.html*) treten in der Regel unerwünschte Verzerrungen und Halleffekte auf, die sich negativ auf die Erkennerleistung bzw. Akkuratheit der F0-Berechnung auswirken, insbesondere dann, wenn das Training auf einer störungsfreien Stichprobe durchgeführt wurde. Eine Möglichkeit, diesen Einfluss zu neutralisieren, ohne den vorhandenen Klassifikator zu ändern, stellt der Einsatz der neuronalen Netze dar, die das verhallte auf das ungestörte Signal abbilden können. Die im Rahmen einer Studienarbeit durchgeführten Experimente haben gezeigt, dass damit die Wortakkuratheit eines Spracherkenners von 18 % auf 29 % (baseline: 48 % – 5 % relative Verbesserung) angehoben werden kann. Bei der Grundfrequenz wurde eine relative Verbesserung von 13 % (Steigerung von 70 % auf 83 %) erzielt.

In das Sprachkorpus wurden im vergangenen Jahr weitere Aufnahmen integriert. Die neuen Dialoge erweitern die Domänen des bisherigen Systems erheblich und erforderten die Überarbeitung des in Sprachmodellen verwendeten Kategoriensystems. Zwei Kategoriensysteme wurden für das Training von Sprachmodellen zur Klassifikation von syntaktisch-prosodischen M-Grenzen entwickelt. Einerseits wurde manuell ein neues Kategoriensystem für das Training eines Sprachmodells entworfen. Andererseits wurde ein automatisches Verfahren zur Sprachmodellberechnung entwickelt, mit dem iterativ ein neues Kategoriensystem für das verwendete kategoriebasierte Sprachmodell berechnet wurde. Beide Kategoriensysteme wurden für das Training von Sprachmodellen verwendet. Die beiden Sprachmodelle wurden mit einem VERBMOBIL-Sprachmodell per rationaler Interpolation verknüpft. Die resultierenden Sprachmodelle lieferten für das Zwei-Klassen-Problem starke-syntaktisch-prosodische Grenze vs. Nicht-Grenze Gesamterkennungsraten von 92% bzw. 93%. Eine Vergrößerung des Vokabulars erfordert bei vergleichbarer Gesamterkennungsrate mit dem automatisierten Verfahren damit ein weniger zeitaufwändiges automatisches Tuning des Klassifikators und kein manuelles Tuning mit aufwändiger Auswahl der Kategorien mit anschließendem Testen des Kategoriensystem-Entwurfs. Sie führt damit zur Verringerung des Zeitaufwands für die Klassifikator-Generierung bei der noch bevorstehenden Ausweitung auf weitere Domänen.

Für die Erkennung von Benutzerzuständen (*UserStates*) wurden die Aufnahmen der drei *SmartKom*-Modalitäten Sprache, Mimik und Gestik analysiert, welche von den Projektpartnern zur Verfügung gestellt wurden. Leider finden sich in den Daten nur sehr wenig Bereiche, die als

nicht-neutraler Zustand markiert sind. Sprachsignale, die prosodisch markiert sind, machen nur 4% der Gesamtdaten aus. Bei der Mimik sind 27% als nicht-neutral etikettiert, davon fallen allerdings 70% unter die Kategorie *Hilflos*. Da nur sehr wenige der Daten als nicht-neutral markiert sind, wurden am LME multimodale Datenaufnahmen durchgeführt. Bei den Aufnahmen wurde den Teilnehmern im Szenario vorgegeben, Benutzerzustände zu simulieren. Jeder Teilnehmer musste in den vier Zuständen *neutral*, *ärgerlich*, *erfreut* und *zögerlich* ihm vorgelegte Sätze sprechen. Vier Szenarien waren zu durchlaufen, dabei hatte der Sprecher erst auf jeweils eine der drei Modalitäten Wert zu legen. Im ersten Szenario wurde die Mimik des Sprechers aufgenommen (die, ohne dass es dem Sprecher bewusst war, auch in den weiteren Szenarien aufgezeichnet wurde). Im zweiten Szenario wurde die Sprache aufgenommen und im dritten die Gestik. Im vierten Szenario musste der Sprecher die drei Modalitäten kombinieren. Für die Darstellung der Benutzerzustände wurde das MMEG (**M**ulti**m**odales **E**mo**g**ramm, in Analogie zum Begriff und zur Darstellungsweise eines Sonagrams) entworfen, das die Modalitäten Sprache (mit Prosodie) und Mimik mit den jeweiligen Benutzerzuständen einander gegenüber stellt.

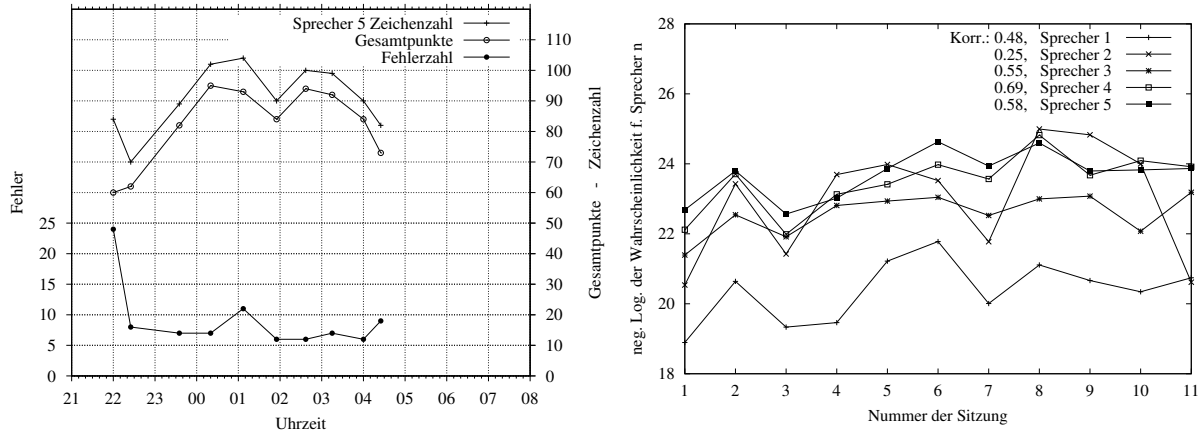


Abbildung 14: Ergebnisse aus dem Müdigkeitsexperiment mit Ergebnissen des Konzentrationstests von Sprecher 5 (links) und sprechergemitteltem Durchschnitt des negativen Logarithmus der Wahrscheinlichkeiten aller Dialoge der Sprecher 1 bis 5 (rechts)

Für die Erkennung von Benutzerzuständen wurde die Arbeit an den Sprachaufnahmen unter Müdigkeit fortgesetzt. Die erfassten Begleitdaten wie z.B. Blutdruck oder Puls und die Ergebnisse der von den Probanden durchzuführenden Aufgaben wie z.B. ein Konzentrationstest wurden nominell wie auch zeitlich analysiert. Die Ergebnisse der erfassten Daten lassen auf bei den Probanden eintretende Müdigkeit schliessen. So zeigen sich z.B. bei den Sprechern im Laufe der Nacht schlechtere Werte für den Konzentrationstest. In Abbildung 14 (links) sind die Ergebnisse von Sprecher 5 gezeigt. Die Skala an der linken Ordinate gibt die Fehler des Sprechers bei den Tests an, die Skala an der rechten Ordinate zeigt die Zahl der Gesamtpunkte und die geschaffte Zeichenzahl bei den Tests. Je höher die Gesamtpunktezahl, desto höher waren Konzentration und Gesamtleistung des Probanden. Mit den Sprachaufnahmen wurde in einer ersten Analyse die Müdigkeitsklassifikation mit Gaußschen Mischungsverteilungen für die ersten fünf Sprecher

Abbildung 15: Im linken Bild sind alle Pixel vorhanden. Zu jedem Bild rechts, werden immer diejenigen Pixel entfernt, die wenig Information für die Benutzerzustandsklassifikation liefern.

durchgeführt. Hier tritt bei allen Probanden eine positive Korrelation (siehe Abbildung 14, rechts) zwischen fortschreitender Zeit und sprechergemitteltem Durchschnitt des negativen Logarithmus der Wahrscheinlichkeiten der Dialoge auf und liefert damit einen Hinweis auf die mit der Zeit zunehmende Veränderung der Stimme eines Sprechers. Abbildung 14 (rechts) zeigt an der Abszisse die fortlaufende Nummer der Sprachaufnahmen (die der zeitlichen Abfolge entspricht). An der Ordinate ist der neg. Logarithmus der Wahrscheinlichkeiten aufgetragen. Je kleiner der Wert im Diagramm von Abbildung 14 (rechts) ist, desto ähnlicher ist der Sprecher zu seiner Referenz-Sprachaufnahme. Im Zusammenhang mit dem Müdigkeitsexperiment sind speziell für die Sprachaufnahmen weitere Analysen und Experimente erforderlich.

Das Mimikmodul in *SmartKom* beachtet den Benutzerzustand Müdigkeit nicht. Die Aufgabe dieses Moduls ist es den internen Zustand des Anwenders zu erkennen um ihm Hilfestellung bei der Bedienung des Systems zu leisten. Wurde ein Benutzerzustand erkannt, so wird diese Information an des Hilfesystem und die Interaktionsmodellierung weitergereicht.

Dadurch ist das *SmartKom*-System in der Lage auf den Benutzerzustand *situationsabhängig* zu reagieren:

- Hilfestellungen zur nächstmöglichen Eingabe zu geben
- oder Gefallen an einer Systemantwort (Darbietung der Fernsehsendungen für den heutigen Abend) als Vorliebe in das Benutzerprofil einzutragen.

Die Benutzerzustände, die das Mimikmodul erkennen kann, sind exakt diejenigen die bei den MMEG-Datenaufnahmen aufgezeichnet wurden (Freude, Neutral, Zögern und Ärger). Dabei repräsentiert Ärger nicht Wut, sondern steht für einen Zustand, der mit Frustration oder Unzufriedenheit umschrieben werden kann.

Für die Mimikerkennung werden SVM und Eigenräume eingesetzt. Experimente haben gezeigt, dass die Klassifikationsleistung der Eigenräume abhängig ist von der Größe des betrachteten Gesichtsausschnittes. Bildbereiche, die keine Information über den Gesichtsausdruck liefern (z.B. Wangen, Hintergrund) Verringern die Erkennungsraten.

In Abbildung 15 ist ein Reihe von Gesichtern mit einem neutralen Gesichtsausdruck dargestellt. Von links nach rechts sind in jedem Bild weniger Pixel vorhanden. Es wurden diejenigen Pixel entfernt, die sich zwischen ärgerlichen und freudigen Gesichtsausdrücken von Personen wenig verändern.

Die SVM wurden zur Klassifikation eines Zwei-Klassenproblems entwickelt. Für die Benutzerzustandsklassifikation war deshalb eine Erweiterung auf vier Klassen durch *one-against-all* und *one-against-one* nötig.

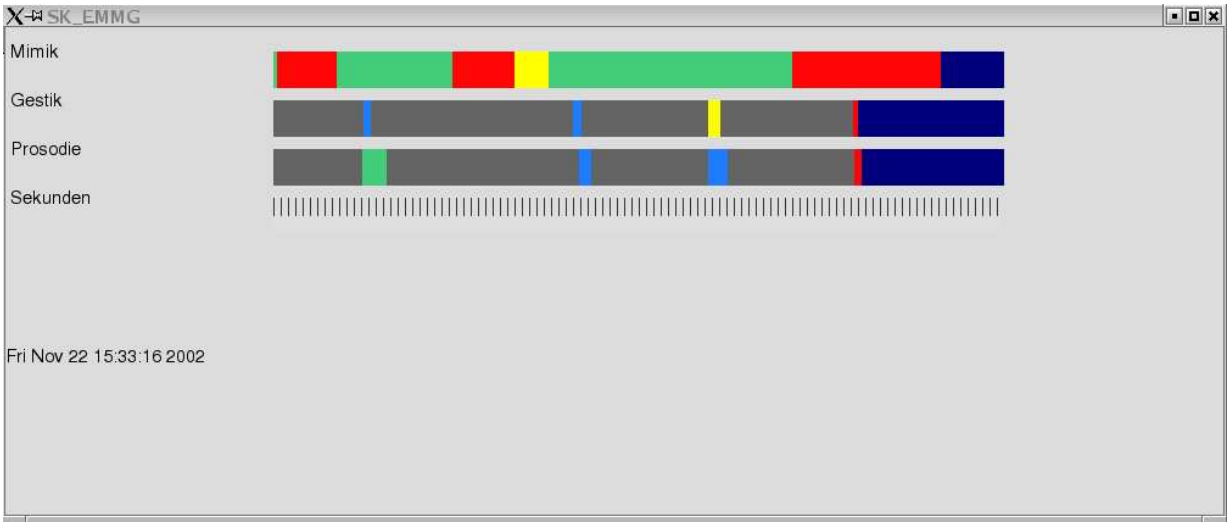


Abbildung 16: Ein Beispiele für die Ausgabe des MMEG-Moduls. Jede Modalität stellt die erzielten Erkennungsergebnisse als Farbbalken dar. Rote Bereiche markieren Ärger, grüne Freude und hellblaue Neutral.

Die Systemreaktionen, die durch die erkannten Benutzerzustände (sowohl mimisch als prosodisch geäußerte) hervorgerufen werden, sind nach Außen hin wenig sichtbar. Sie gehören zu den Systemfunktionen, die unauffällig aber für den Anwender sehr angenehm sind. Sie versuchen den Dialogverlauf auf die Anforderungen und Wünsche des Anwenders anzupassen.
Um die Leistung von Benutzerzustandserkennung durch Mimik, Gestik und Prosodie beobachtbar zu machen, wurde das MMEG-Modul entwickelt. Dies ist eine Art Balkendiagramm, in dem für jede Eingabemodalität die erkannten Benutzerzustände als Farbbalken markiert ist.

In Abbildung 16 ist ein Beispiel für eine Ausgabe dargestellt. Der zeitliche Fortschritt ist abhängig von der Verarbeitungszeit und dem Verarbeitungsmodus. Die Prosodie z. B. liefert Erkennungsergebnisse erst nach dem Ende eines Turns.

Zur Erkennung der Benutzerzustände kann die Gestik auch eine interessante Rolle spielen, denn sie ist neben der Sprache und Mimik ein wichtiger Kanal der Mensch-Maschine-Kommunikation. Die Gestik, die ein Benutzer dem multimodalen Dialogsystem (z.B. *SmartKom*) gegenüber verwendet, wurde zur Ergänzung der bisherigen UserState-Erkennung im Projekt *SmartKom* untersucht. Von Interesse ist dabei, inwiefern die Gestik bei der Erkennung des Benutzerzustandes hilfreich sein kann, dass das System eine benutzerfreundlichere Dialogstrategie führt. Bislang sind nur die zwei Eingabemodalitäten Mimik und Sprache zur Erkennung des Benutzerzustandes eingesetzt worden. Gestik soll nun zur Erkennung von Nachdenken bzw. Zögern, Neutral bzw. Entschlossen sowie Negativ anhand der Dynamik der Bewegung verwendet werden. Dabei entspricht Negativ dem Ärger und Entschlossen der Freude bei den anderen Modalitäten. Zögern ist in allen drei Modalitäten identisch. Insgesamt wurden ca. 50 Versuchspersonen aufgenommen, die die auf dem Display dargestellten Genre-Begriffe von Filmtiteln mit gespielten Emotionen wie z. B. Zögern oder Entschlossenheit auszuwählen hatten. Mit den aufgenomme-

Gestik	Entschlossen(Freude/Neutral)	Zögern	Negativ(Scheibenwischer)
Entschlossen	64%	21%	15%
Zögern	15%	68%	17%
Negativ	10%	10%	80%

Tabelle 1: Erkennungsergebnisse des Benutzerzustandes mit Gesten

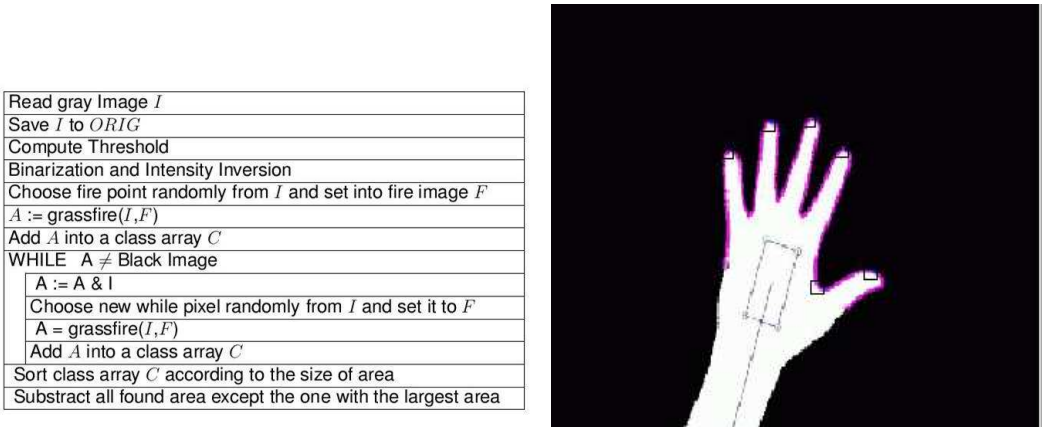


Abbildung 17: Mit dem erweiterten Steppenbrand-Algorithmus geschätzte Fingerspitzen

nen Daten, die die Dynamik (Position, Zeit) der Gesten enthalten, wurde ein diskretes Hidden Markov Modell (HMM) trainiert bzw. getestet. Die Ergebnisse sind in Tabelle 1 veranschaulicht.

An den Ergebnissen ist zu sehen, dass Entschlossenheit manchmal mit Zögern verwechselt wird. Das liegt zum Teil daran, dass die neutrale Gestikeingabe auch als Entschlossenheit angesehen wird, deren Dynamik allerdings zum Teil ähnlich wie die von Zögern ist. Im Falle von Zögern und Negativ passiert das gleiche, weil der Benutzer manchmal beim Zögern seine Hand auch scheibenwischerartig bewegt. Insgesamt ist bei allen drei Benutzerzuständen eine gemittelte Erkennnungsrate von 71% erzielt worden.

Ein anderer Aspekt der Gestenerkennung ist die Lokalization der einzelnen Finger, denn die Erkennung der Finger ist eventuell sehr entscheidend im Falle einer als Kommando zu interpretierenden Gesteneingabe. Ausgehend von dem bekannten Steppenbrand-Algorithmus und der Annahme, dass die Hand in der *SmartKom*-Umgebung meistens das Hauptobjekt ist, wurde ein in Abbildung 17 (*links*) dargestelltes Verfahren entwickelt, das die Hand aus einem komplexen Hintergrund ausschneidet. Um die Fingerspitzen zu finden, wird anschließend durch den Einsatz des *Principal Axis Theorem* die Hauptachse der Hand gesucht. Mit Hilfe dieser wird ein Kettencode der Handkontur generiert, der alle Fingerspitzen enthält. Die Positionen mit der (lokalen) maximalen Krümmungen auf der Konturlinie entsprechen den einzelnen Fingerspitzen, die in Abbildung 17 (*rechts*) mit schwarzen Rechtecken markiert sind. Im Gegensatz zu herkömmlichen Vorgehen wie Template-Matching erfordert dieses Verfahren geringeren Rechenaufwand, da die meisten internen Punkte der Hand durch den Kettencode der Handkontur ignoriert werden.

Weitere Experimente in diesem Gebiet sollen den Fall untersuchen, in dem die Verdeckung von Händen durch andere Objekte vorkommt, so dass die Kontur und damit die einzelnen Fingerspitzen robust erkannt werden können.

4.3 Das PF–Star Projekt

Seit Oktober beteiligt sich der Lehrstuhl an einem neuen EU-Projekt: *PF-STAR* (**P**reparing **f**uture multi**s**ensorial in**t**er**a**ction **r**esearch). *PF-STAR* dient zur Vorbereitung des sechsten Rahmenprogramms der Europäischen Union für Forschung und Entwicklung von Technologien mit dem Ziel die europäische Forschungslandschaft zu bereichern. Ziel von *PF-STAR* ist die Entwicklung und Bewertung von Grundlagen und Referenzsystemen für zukünftige Untersuchungen. Dabei wird die mehrsprachige Kommunikation über verschiedene Kanäle wie z.B. Sprache und Bild analysiert. Das Projekt wird von Partnern aus Italien, Deutschland, England und Schweden bearbeitet und ist in mehrere Teilgebiete (Workpackages) untergliedert:

- Automatische Übersetzung von Sprache
- Emotionserkennung und Emotionssynthese in der Sprache
- Emotionssynthese (Gesichter)
- Erkennung von Kindersprache

Der Lehrstuhl für Mustererkennung der Universität Erlangen ist dabei an den Gebieten Erkennung von Kindersprache und Emotionserkennung beteiligt. Für letzteres ist Erlangen Projektleiter.

Die Erkennung von Emotion ist ein relativ neues und vielversprechendes Gebiet. Zum einen kann eine ausgeprägte Emotion die Spracherkennung verschlechtern, zum anderen kann es wichtig sein, den emotionalen Zustand von Benutzern zu verfolgen, um gegebenenfalls geeignete Maßnahmen ergreifen zu können (z.B. in Dialogsystemen). Abbildung 18 zeigt den Verlauf der Grundfrequenz in Sprachaufnahmen eines verärgerten und eines neutralen Sprechers. In *PF-STAR* wollen wir uns den folgenden Themen widmen:

- Extraktion prosodischer und anderer akustischer (z.B. spektraler) Merkmale, die unterschiedliche Emotionen kennzeichnen
- Klassifikation von emotionaler und nicht-emotionaler Sprache mit diesen Merkmalen

Im Berichtsjahr wurden Daten annotiert, in denen Anfragen an ein Flugbuchungssystem der Firma Sympalog gestellt werden. Es zeigt sich, dass die Benutzer in unterschiedlichen Entwicklungsstufen des Systems verschieden emotional (verärgert) reagieren. Ein erster SYMPAFLY-Erkenner wurde bereits erstellt.

Es soll im Laufe des Projekts auch emotionale und nicht-emotionale Kindersprache aufgenommen und verarbeitet werden. Kindersprache ist ebenfalls ein neues und vielversprechendes Gebiet, mit unterschiedlichen Anwendungsmöglichkeiten in den Bereichen 'Edutainment', Behindertentherapie, Computerspiele, usw. Im Augenblick sind weder die akustische noch die linguistische Modellierung von Kindersprache zur Zufriedenheit gelöst; dies schlägt sich in schlechten

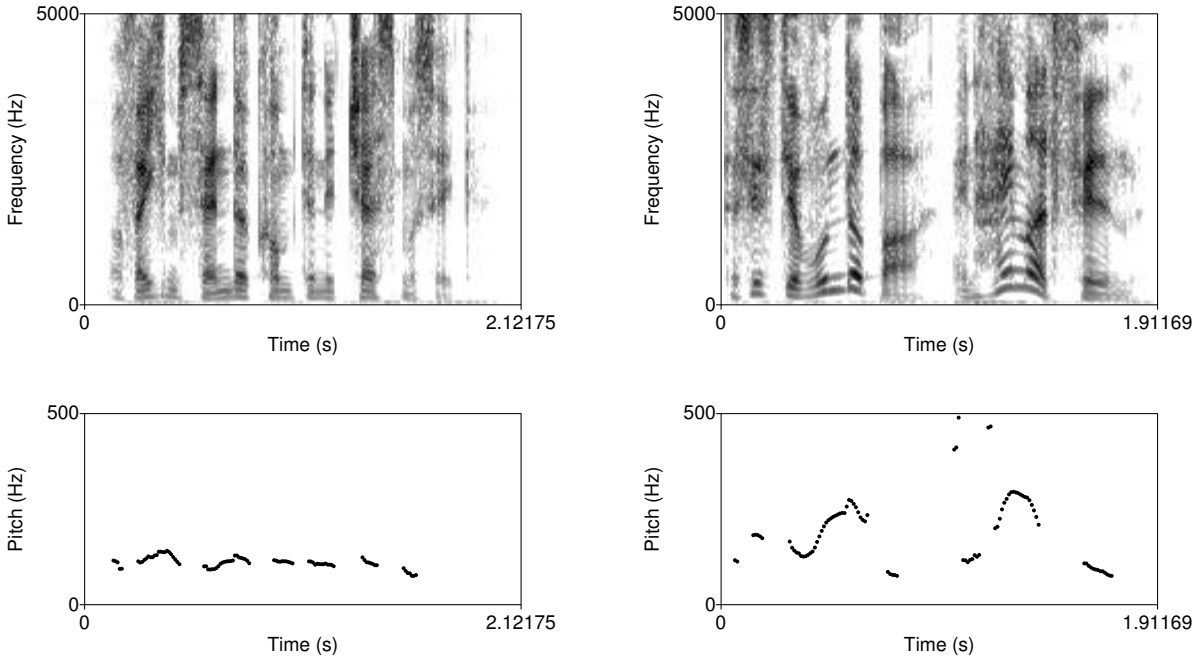


Abbildung 18: Spektrogram und Grundfrequenz einer neutralen Äußerung (links) und einer freudigen Äußerung (rechts) desselben Sprechers. Rechts variiert die Grundfrequenz viel stärker.

Erkennungsraten selbst von gelesener, und noch mehr von spontaner Kindersprache nieder. In *PF-STAR* soll versucht werden, den Stand der Kunst in diesem Gebiet demjenigen der Verarbeitung von Erwachsenensprache anzunähern. Erste Untersuchungen wurden mit Sprachaufnahmen durchgeführt, die im Berichtsjahr am Ohm-Gymnasium aufgezeichnet worden sind, weitere folgen mit dem YOUTH-Korpus der Carnegie Mellon University.

Zur Zeit laufen die Vorbereitungen für Sprachaufnahmen, die im Februar aufgezeichnet werden, auf Hochtouren. Kinder werden beim Spiel mit dem Roboterhund AIBO (Firma Sony) aufgenommen. Unter anderem soll AIBO durch einen Parcour gesteuert werden. AIBO reagiert auf die Umwelt über ein Mikrophon, eine Kamera und Tastsensoren. Dadurch dass AIBO natürlich nicht perfekt funktioniert, hoffen wir emotionale Kindersprache zu gewinnen (Ärger, Freude). Die Aufnahmen finden an zwei Erlanger Schulen statt. Der AIBO-Korpus wird als Grundlagen für viele Untersuchungen in den beiden Teilgebieten Emotionserkennung und Erkennung von Kindersprache dienen, da sowohl Emotionen (kein Schauspielszenario) als auch spontane Kindersprache vorliegen. Parallel dazu wird zusätzlich unser oben genannter Korpus mit gelesener Kindersprache erweitert.

5 Studienarbeiten

1. Fritz, M.: 3-D Objektverfolgung mit Lichtfeldern, September 2002

2. Müller, R.: Anwendung bildbasierter Objektmodelle in der erweiterten Realität, August 2002

3. Weiß, R.: Anwendung von KNN zur Beseitigung der raumbedingten Störungen in einem Sprachsignal, Juli 2002

6 Diplomarbeiten

1. Adler, W.: Evaluation von Verfahren zur Analyse von HRT-Bildern, September 2002

2. Gömmel, G.: Integration von aktiver Objekterkennung und -verfolgung, September 2002

3. Gräßl, C.: Erweiterung des diskret-statistischen Eigenraums auf kontinuierliche Objektmodelle, Juli 2002

4. Grobe, M.: Automatische Segmentierung und Flächenkorrelation von Zellen in Fluoreszenz- und Durchlichtbildern, März 2002

5. Hacker, C.: Semikontinuierliche Hidden-Markov-Modelle mit mehreren Kodebüchern, September 2002

6. Kulicke, A.: Schnelle Sprecheradaption, Mai 2002

7. Steidl, S.: Interpolation von Hidden Markov Modellen, November 2002

8. Zinßer, T.: Robuste Schätzung der Korrespondenzen zwischen 3-D Punktemengen, Juli 2002

7 Master Theses

1. Dorkó, D.: Properties of Ultrametric Spin-Glass Markov Random Fields, Juni 2002

2. Muenzenmayer, C.: Implementation and Evaluation of Colour-Texture-Algorithms, Mai 2002

8 Promotionen

1. Ohler, U.: Computational Promoter Recognition in Eukaryotic Genomic DNA, April 2002

2. Greiffenhagen, M: Engineering, Statistical Modeling and Performance Characterization of a Real-Time Dual Camera Surveillance System, 2002

3. Huber, R.: Prosodisch-linguistische Klassifikation von Emotion, Mai 2002

9 Vorträge

1. Adelhardt, J.: Sprachmodell-Berechnung beim Übergang auf eine neue Anwendung, 13. Konferenz Elektronische Sprachsignalverarbeitung (essv2002), Technische Universität Dresden, Institut für Akustik und Sprachkommunikation, Brandenburgische Technische Universität Cottbus, Lehrstuhl Kommunikationstechnik, 25. -27. Sept. 2002, Dresden, 26.9.2002

2. Batliner, A.: Prosodic Classification of Offtalk: First Experiments, Fifth International Conference on Text, Speech and Dialogue - TSD 2002, Brno, Tschechien, 12.09.02

3. Batliner, A.: WP3 emotions: Speech, Kick-off meeting PF-STAR (Preparing future multisensorial interaction research), 4.-6.11.2003, Trento, Italien, 5.11.2003

4. Deinzer, F.: Improving Object Recognition By Fusion Of Multiple Views, Third Indian Conference on Computer Vision Graphics and Image Processing 2002, Ahmedabad, Indien, 17.12.2002

5. Denzler, J.: On Optimal Active Vision for Object Recognition, INC Seminar Series, Institute for Neural Computation, University of California, San Diego, USA, 09.04.02

6. Denzler, J.: Adaptive Real-Time Segmentation and Sparse 3-D Reconstruction, VAMPIRE Kickoff-Meeting (Visual Active Memory Processes and Interactive Retrieval), Universität Bielefeld, Bielefeld, Deutschland, 06.06.02

7. Denzler, J.: Optimale Sensordatenauswahl und -verarbeitung in intelligenten Systemen, Institut für Informationstechnik, Gerhard Mercator Universität Duisburg, Duisburg, Deutschland, 15.07.02

8. Denzler, J.: Optimale Sensordatenauswahl und -verarbeitung in intelligenten Systemen, Fakultät für Mathematik und Informatik, Universität Passau, Passau, Deutschland. 31.07.02

9. Denzler, J.: On Optimal Camera Parameter Selection in Kalman-Filter Based Object Tracking, Pattern Recognition, 24th annual symposium for Pattern Recognition of the DAGM, Zürich, Schweiz, 16.09.02

10. Deventer, R.: Using Non-Markov Models for the Control of Dynamic Systems, Third International NAISO (Natural and Artificial Intelligence Systems Organization), Symposium on Engineering of Intelligent Systems Malaga, Spain, 25.09.2002

11. Drexler, Ch.: Generic Hierarchic Object Models and Classification based on Probabilistic PCA, IAPR Workshop on Machine Vision Applications (MVA), 11.-13.12.2002, Nara, Japan, 13.12.2002

12. Niemann, H.: 3D Object Recognition and Localization, IAC-CNR, Rom, 6.6.2002

13. Nöth, E., Multimodale Eingabe in der Mensch-Maschine-Kommunikation, Universität Regensburg, Regensburg, 3.5.2002

14. Nöth, E., UserStates and UseCases - Konzepte für den SmartKom-Demonstrator, SmartKom-Projektlekungssitzung, Heidelberg, 10.05.2002

15. Nöth, E., A Tutorial on Decision Strategies Used in Automatic Speaker Recognition/Verification, Bundeskriminalamt, Wiesbaden, 28.5.2002

16. Nöth, E., Herausforderung Mustererkennung, Fachkolloquium der MEDAV GmbH, Uttenreuth, 12.7.2002

17. Nöth, E., Fortschritt durch freien Dialog, ISI 2002 (Internationales Symposium für Informationswissenschaft), Regensburg, 9.10.2002

18. Nöth, E., Ergebnisse zur UserState-Verarbeitung, SmartKom-Projektstandssitzung, Saarbrücken, 18.12.2002

19. Scholz, I.: Teilprojekt C2: Analyse, Codierung und Verarbeitung von Lichtfeldern, Statustreffen des Sonderforschungsbereichs 603 (Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten), Erlangen, 23.01.2002

20. Scholz, I.: Globale Optimierung, Arbeitskreis Optimierung des Sonderforschungsbereichs 603, (Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten), Erlangen, 27.06.2002

21. Scholz, I.: Calibration of Real Scenes for the Reconstruction of Dynamic Light Fields, International Association for Pattern Recognition (IAPR) Workshop on Machine Vision Applications (MVA), Nara, Japan, 11.-13.12.2002, 11.12.2002

22. Stemmer, G.: Multiple Time Resolutions for Dynamic Features of Speech, Laboratoire lorrain de recherche en informatique et ses applications (LORIA), Nancy, Frankreich, 25.02.02

23. Stemmer, G.: Comparison and Combination of Confidence Measures, Fifth International Conference on Text, Speech and Dialogue - TSD 2002, Brno, Tschechien, 09.09.02

24. Vogt, F.: Teilprojekt B6, Rechnergestützte Endoskopie des Bauchraums, Stand und Kooperationen, Statustreffen des Sonderforschungsbereichs 603 (Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten), 23.01.2002

25. Vogt, F.: Unsichtbares wird sichtbar: Glanzlichtsubstitution mit Lichtfeldern, Workshop Bildverarbeitung für die Medizin - Algorithmen, Systeme, Anwendungen, 2002, Lübeck, 11.03.2002

26. Vogt, F.: Sensordatenfusion, Arbeitskreis Visualisierung des Sonderforschungsbereichs 603 (Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten), 25.04.2002 http://sfb-603.uni-erlangen.de,

27. Vogt, F.: Glanzlichtsubstitution mit (endoskopischen) Lichtfeldern, Fraunhofer-Institut für Integrierte Schaltungen, internes Seminar, 26.07.2002

28. Vogt, F.: Highlight Substitution in Light Fields, IEEE 2002 International Conference on Image Processing, ICIP 2002, Rochester, USA, 24.09.2002

29. Vogt, F. und Schick, C.: Moderne Medizin: Computer sehen und helfen beim Behandeln, Collegium Alexandrinum der Universität Erlangen-Nürnberg, Reihe: Computer - das dritte Auge des Arztes, 07.11.2002

30. Wenhardt, S.: Didaktische Aufbereitung von Methoden des Rechnersehens für virtuelle Vorlesungen, 1. Workshop der Gesellschaft für Informatik, Fachgruppe Didaktik der Informatik, 10.-11.10.02, Witten-Bommerholz, 11.10.02

31. Zeissler, V.: Parametrische Modellierung von Dauer und Energie prosodischer Einheiten, 6. Konferenz zur Verarbeitung natürlicher Sprache (KONVENS2002), Saarbrücken, Germany, 30.09.02

32. Zobel, M.: Optimierungsansatz für die Integration von Kamerabildern bei der Klassifikation, Statustreffen des SFB 603 (Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten), Erlangen, 11. Juli 2002

33. Zobel, M.: Binocular 3-D Object Tracking with Varying Focal Lengths, IASTED International Conference on Signal Processing, Pattern Recognition, and Application, 25.-28. Juni 2002, Kreta, Griechenland, 27. Juni 2002

34. Zobel, M.: Entropy Based Camera Control for Visual Object Tracking, IEEE 2002 International Conference on Image Processing, 22.-25. September 2002, Rochester, USA, 25. September 2002

35. Fritz, M.: Object Tracking and Pose Estimation Using Light-Field Object Models, 7th International Fall Workshop Vision, Modelling and Visualization, 20.-22. November 2002, Erlangen, 21. November 2002

Literatur

[1] J. Adelhardt, E. Nöth, G. Stemmer, H. Niemann: *Sprachmodell-Berechnung beim Übergang auf eine neue Anwendung*, in R. Hoffmann (Hrsg.): *Elektronische Sprachverarbeitung*, w.e.b. Universitätsverlag, Dresden, September 2002, S. 302–309.

[2] A. Batliner, V. Zeißler, E. Nöth, H. Niemann: *Prosodic Classification of Offtalk: First Experiments*, in P. Sojka, I. Kopecek, K. Pala (Hrsg.): *Text, Speech and Dialogue, Proceedings of the Fifth International Conference on Text, Speech, Dialogue - TSD 2002* , Bd. 2448 von *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Berlin, August 2002, S. 357–364.

[3] B. Caputo, G. Dorkó, H. Niemann: *An Ultrametric Approach to Object Recognition*, in G. Greiner, H. Niemann, T. Ertl, B. Girod, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2002*, (Akademische Verlagsgesellschaft Aka GmbH, Berlin, Erlangen, 2002, S. 13–20.

[4] B. Caputo, G. Dorko, H. Niemann: *Combining Color and Shape Information for Appearance-based Object Recognition Using Ultrametric Spin Glass-Markov Random Fields*, in S. Lee, A. Verri (Hrsg.): *Proc. of first ICPR Workshop on Pattern Recognition with Support Vector Machines*, (Springer LNCS, Berlin Heidelberg), Niagara Falls, Canada, 2002, S. 97–111.

[5] B. Caputo, H. Niemann: *Storage capacity of kernel associative memories*, in J. Dorronsoro (Hrsg.): *Proc of ICANN 2002*, (Springer LNCS, Berlin Heidelberg), Madrid, Spain, 2002, S. 51–56.

[6] R. Chrastek, M. Wolf, K. Donath, G. Michelson, H. Niemann: *Optic Disc Segmentation in Retinal Images*, in M. Meiler, D. Saupe, F. Kruggel, H. Handels, T. Lehmann (Hrsg.): *Bildverarbeitung für die Medizin 2002: Algorithmen - Systeme - Anwendungen*, (Springer Informatik aktuell, Berlin, Heidelberg, Leipzig, Germany, 2002, S. 263–266.

[7] R. Chrastek, M. Wolf, K. Donath, H. Niemann, G. Michelson: *Automated Calculation of Retinal Arteriovenous Ratio for Detection and Monitoring of Cerebrovascular Disease Based on Assessment of Morphological Changes of Retinal Vascular System*, in *Proc. of IAPR Workshop on Machine Vision Applications*, (ISBN 4-901122-02-9), Nara, Japan, 2002, S. 240–243.

[8] F. Deinzer, J. Denzler, H. Niemann: *Improving Object Recognition By Fusion Of Multiple Views*, in *Proceedings of the Third Indian Conference on Computer Vision Graphics and Image Processing*, Allied Publishers Pvt. Ltd., Ahmedabad, Indien, Dezember 2002, S. 161–166.

[9] J. Denzler, C. Brown: *Information Theoretic Sensor Data Selection for Active Object Recognition and State Estimation*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Bd. 24, Nr. 2, 2002, S. 145–157.

[10] J. Denzler, M. Zobel, H. Niemann: *On Optimal Camera Parameter Selection in Kalman Filter Based Object Tracking*, in L. V. Gool (Hrsg.): *Pattern Recognition*, Springer-Verlag, Zurich, September 2002, S. 17–25.

[11] J. Denzler, M. Zobel, J. Triesch: *Probabilistic Integration of Cues From Multiple Cameras*, in R. Würtz (Hrsg.): *Dynamic Perception*, 2002, S. 309–314.

[12] R. Deventer, J. Denzler, H. Niemann: *Application of Bayesian controllers to dynamic system*, in A. Abraham, M. Koeppen (Hrsg.): *Hybrid Information Systems, Proceedings of the First International Workshop in Hybrid Intelligent Systems, HIS 2001*, Physica, Heidelberg, 2002, S. 555–569.

[13] R. Deventer, J. Denzler, H. Niemann: *Using Non-Markov models for the control of Dynamic Systems*, in *Engineering of Intelligent Systems (EIS)*, ICSC-NAISO Academic Press, September 2002, S. 70 (complete paper on CD–ROM).

[14] R. Deventer, J. Denzler, H. Niemann: *Bayesian Control of Dynamic Systems (to be published)*, in L. C. J. Ajith Abraham (Hrsg.): *Recent Advances in Intelligent Paradigms*, Kap. 3, Springer Verlag, 2002.

[15] S. Di Bona, H. Niemann, O. Salvetti, M. Wolf: *Computational Complexity Analysis of a 3D Neural Network Approach to Volume Matching*, *Pattern Recognition and Image Analysis*, Bd. 12, Nr. 1, 2002, S. 63–69.

[16] C. Drexler, F. Mattern, J. Denzler: *Appearance Based Generic Object Modeling and Recognition Using Probabilistic Principal Component Analysis*, in L. V. Gool (Hrsg.): *Pattern Recognition - 24th DAGM Symposium*, Springer-Verlag, Zurich, September 2002, S. 100–108.

[17] C. Drexler, F. Mattern, J. Denzler: *Generic Hierarchic Object Models and Classification based on Probabilistic PCA*, in K. Ikeuchi (Hrsg.): *Proceedings of MVA 2002*, IAPR MVA Organizing Committee, Dez. 2002, S. 435–438.

[18] G. Greiner, H. Niemann, T. Ertl, B. Girod, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2002*, Akademische Verlagsgesellschaft Aka GmbH, Berlin, Germany, 2002.

[19] T. Haderlein: *Using the ISADORA System for Analyzing Fatigue Symptoms and Robustness of Features against Reverberation*, Lehrstuhl für Multimediakommunikation und Signalverarbeitung, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2002.

[20] J. Hornegger, V. Welker, H. Niemann: *Localization and Classification Based on Projections*, *Pattern Recognition*, Bd. 35, 2002, S. 1225–1235.

[21] Y. Huang, T. Huang, H. Niemann: *Region-Based Method for Model-Free Object Tracking*, in C. S. R Kasturi, D Laurendeau (Hrsg.): *ICPR*, (IEEE Computer Society), Quebec, Canada, 2002, S. III.3.1.

[22] Y. Huang, T. Huang, H. Niemann: *Segmentation Based Object Tracking Using Image Warping and Kalman Filtering*, in A. Tekalp (Hrsg.): *Proc. IEEE Int. Conf. on Image Processing*, (IEEE Computer Society), Rochester, New York, 2002, S. III 601–604.

[23] Y. Huang, T. Huang, H. Niemann: *Two-Handed Gesture Tracking Incorporating Template Warping With Static Segmentation*, in J. O. R. Chellapa, P. Fua (Hrsg.): *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, (IEEE Computer Society), Washington D.C., USA, 2002, S. 275–280.

[24] C. Münzenmayer, H. Volk, D. Paulus, F. Vogt, T. Wittenberg: *Multispectral Statistical Geometrical Features for Texture Analysis and Classification*, in K.-H. Franke (Hrsg.): *8.*

Workshop Farbbildverarbeitung, Schriftenreihe des Zentrums für Bild- und Signalverarbeitung e.V. Ilmenau, 1/2002, Ilmenau, 2002, S. 87–94.

[25] E. Nöth, A. Batliner, V. Warnke, J. Haas, M. Boros, J. Buckow, R. Huber, F. Gallwitz, M. Nutt, H. Niemann: *On the Use of Prosody in Automatic Dialogue Understanding*, *Speech Communication*, Bd. 36, Nr. 1-2, January 2002, S. 45–62.

[26] J. Pösl, H. Niemann: *Erscheinungsbasierte statistische Objekterkennung*, *Informatik Forschung und Entwicklung*, Bd. 17, Nr. 1, 2002, S. 21–40.

[27] C. Schick, T. Horbach, H. Weber, F. Vogt, G. Greiner, D. Paulus, W. Hohenberger: *Rechnergestütze Endoskopie des Bauchraums*, in J. Siewert, W. Hartel (Hrsg.): *Digitale Revolution in der Chirurgie, Deutscher Chirurgenkongreß (DGCH) 2002*, Springer, 2002.

[28] J. Schmidt, H. Niemann, S. Vogt: *Dense Disparity Maps in Real-Time with an Application to Augmented Reality*, in *Proceedings Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002)*, IEEE Computer Society, Orlando, FL USA, Dez. 2002, S. 225–230.

[29] J. Schmidt, F. Vogt, H. Niemann: *Nonlinear Refinement of Camera Parameters using an Endoscopic Surgery Robot*, in K. Ikeuchi (Hrsg.): *Proceedings of MVA 2002*, IAPR MVA Organizing Committee, Dez. 2002, S. 40–43.

[30] I. Scholz, J. Denzler, H. Niemann: *Calibration of Real Scenes for the Reconstruction of Dynamic Light Fields*, in K. Ikeuchi (Hrsg.): *Proceedings of MVA 2002*, IAPR MVA Organizing Committee, Dez. 2002, S. 32–35.

[31] G. Stemmer, S. Steidl, E. Nöth, H. Niemann, A. Batliner: *Comparison and Combination of Confidence Measures*, in P. Sojka, I. Kopecek, K. Pala (Hrsg.): *Text, Speech and Dialogue, Proceedings of the Fifth International Conference on Text, Speech, Dialogue - TSD 2002* , Bd. 2448 von *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Berlin, August 2002, S. 181–188.

[32] C. Vogelgsang, I. Scholz, G. Greiner, H. Niemann: *lgf3 - A Versatile Framework for Vision and Image-Based Rendering Applications*, in G. Greiner, H. Niemann, T. Ertl, B. Girod, H. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2002*, Infix, Erlangen, Germany, November 2002, S. 257–264.

[33] F. Vogt, D. Paulus, B. Heigl, C. Vogelgsang, H. Niemann, G. Greiner, C. Schick: *Making the Invisible Visible: Highlight Substitution by Color Light Fields*, in *Proceedings First European Conference on Colour in Graphics, Imaging, and Vision*, IS&T – The Society for Imaging Science and Technology, Springfield, USA, Poitiers, France, 2002, S. 352–357.

[34] F. Vogt, D. Paulus, H. Niemann: *Highlight Substitution in Light Fields*, in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, IEEE Computer Society Press, Rochester, USA, September 2002, S. 637–640.

[35] F. Vogt, D. Paulus, I. Scholz, H. Niemann, C. Schick: *Glanzlichtsubstitution durch Lichtfelder*, in M. Meiler, H. Handels, F. Kruggel, T. Lehmann, D. Saupe (Hrsg.): *6. Workshop Bildverarbeitung für die Medizin*, Springer Berlin, Heidelberg, New York, Leipzig, 2002, S. 103–106.

[36] 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.

[37] S. Wenhardt, J. Schmidt, H. Niemann: *Didaktische Aufbereitung von Methoden des Rechnersehens für virtuelle Vorlesungen*, in S. Schubert, J. Magenheim, P. Hubwieser, T. Brinda (Hrsg.): *Forschungsbeiträge zur Didaktik der Informatik - Theorie, Praxis, Evaluation*, Köllen Druck und Verlag GmbH, Bonn, Witten-Bommerholz, Germany, Oktober 10-11 2002, S. 87–96, Lecture Notes in Informatics.

[38] V. Zeissler, E. Nöth, G. Stemmer: *Parametrische Modellierung von Dauer und Energie prosodischer Einheiten*, in S. Busemann (Hrsg.): *KONVENS2002*, DFKI GmbH, Saarbrücken, Germany, September 2002, S. 177–183.

[39] M. Zobel, J. Denzler, H. Niemann: *Binocular 3-D Object Tracking with Varying Focal Lengths*, in M. Hamza (Hrsg.): *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Application, Crete, Greece*, ACTA Press, Anaheim, Calgary, Zurich, 2002, S. 325–330.

[40] M. Zobel, J. Denzler, H. Niemann: *Entropy Based Camera Control for Visual Object Tracking*, in *Proceedings of the International Conference on Image Processing (ICIP)*, IEEE Computer Society Press, Rochester, USA, September 2002, S. 901–904.

[41] M. Zobel, M. Fritz, I. Scholz: *Object Tracking and Pose Estimation Using Light-Field Object Models*, in G. Greiner, H. Niemann, T. Ertl, B. Girod, H. Seidel (Hrsg.): *Vision, Modeling, and Visualization 2002*, Infix, Erlangen, Germany, November 2002, S. 371–378.

Download

Dieses Dokument ist in zwei Formen zum Druck verfügbar: als Postscript und als PDF mit Hyperlinks.

Außerdem steht die verkürzte Version aus der gedruckten Fassung des Instituts für Informatik zum Download als PostScript-Datei bereit.