

Jahresbericht LME 1999

D. Paulus / E. Nöth

Inhaltsverzeichnis

1 Mitglieder der Arbeitsgruppe	2
2 Einführung	3
3 Modellierung von Prozessketten und DNA	4
4 Bildanalyse	5
5 Sprachverstehen	13
6 Promotionen	24
7 Diplomarbeiten	24
8 Studienarbeiten	25

1 Mitglieder der Arbeitsgruppe

Lehrstuhl für Mustererkennung (Informatik 5)

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

Mitarbeiter:

Ahlrichs, U., Dipl.-Inf.	(wiss. Mitarb., DFG)	01.06.96
Batliner, A., Dr.-Phil.	(wiss. Mitarb., BMBF)	01.01.97
Buckow, J., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.04.97
Caputo, B., M.Sc.	(wiss. Mitarb., Grad.-Stip.)	01.11.99
Deinzer, F., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.01.99
Denzler, J., Dr.-Ing.	(wiss. Assistent)	01.01.93
Deventer, R., Dipl.-Inf.	(wiss. Mitarb., SFB 396)	01.02.99
Drexler, Ch., Dipl.-Inf.	(wiss. Mitarb., DIROKOL)	01.02.98
Endres, B.	(Sekretärin, 1/2)	15.01.86
Fentze, W.	(Programmierer)	05.12.88
Fischer, J., Dipl.-Inf.	(wiss. Mitarb., SFB)	16.01.95
Frank, C.M., Dipl.-Inf.	(wiss. Mitarb., SMARTCOM)	01.09.99
Gallwitz, F., Dipl.-Inf.	(wiss. Mitarb., DFG)	16.12.94
Gebhard, A., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.04.98
Haas, J., Dipl.-Inf.	(wiss. Mitarb.)	01.04.95
Heigl, B., Dipl.-Inf.	(wiss. Mitarb., SFB)	01.04.97
Hornegger, J., Dr.-Ing.	(wiss. Assistent)	01.05.92
Huber, R., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.03.97
Karadag, C.	(Sekretärin, 1/2, SFB 603)	01.07.98
Koppe, I.	(Sekretärin, 1/2)	18.12.92
Merz, T., Dipl.-Inf.	(wiss. Mitarb., DFG)	15.09.95
Meyer, D., Dipl.-Ing.	(wiss. Mitarb., Grad.-Stip.)	15.08.96
Niemann, H., Dr.-Ing.	(Professor)	24.09.75
Nöth, E., Dr.-Ing.	(Akad. Oberrat)	01.02.85
Obermayer, W.	(Programmierer)	01.07.76
Ohler, U., Dipl.-Inf.	(wiss. Mitarb.)	01.11.97
Paulus, D., Dr.-Ing.	(Akad. Oberrat)	01.03.87
Popp, F.	(Techniker)	01.09.85
Reinhold, M., Dipl.-Ing.	(wiss. Mitarb., Grad.-Stip.)	09.06.99
Stemmer, G., Dipl.-Inf.	(wiss. Mitarb.)	13.10.99
Warnke, V., Dipl.-Inf.	(wiss. Mitarb., BMBF)	01.08.96
Zobel, M., Dipl.-Inf.	(wiss. Mitarb., SFB 603)	01.01.98

Gäste:

Yuan, C. (China, KAS) 03.04.97
Huang, Yu, Dr. (China, AvH) 01.08.99

2 Einführung

Unter dem Arbeitsgebiet „Mustererkennung“ des Lehrstuhls wird 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 verstanden. 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 wissensbasierte Analyse von Bildern und Bildströmen sowie Verstehen gesprochener Sprache und Generierung einer Antwort. In der wissensbasierten Bildanalyse werden sowohl grundsätzliche Arbeiten zur Bildverarbeitung und zur Repräsentation und Nutzung problemspezifischen Wissens als auch spezielle Arbeiten zur Entwicklung eines vollständigen, rückgekoppelten Systems für die schritt haltende 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 wurde im neu begonnenen Projekt Smartkom hergestellt, das vom BMBF als Leitprojekt gefördert wird.

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

Die Modellierung von Sensordaten und Mustern mit statistischen Mitteln spielt in Bild und Sprachverarbeitung eine bedeutende Rolle. Dieser Weg wird auch in zwei Projekten zur Genomanalyse und zur Modellierung von Prozessketten eingeschlagen.

Ein Problem, das in jeddem 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. 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 Modellierung von Prozessketten und Genomanalyse

(U. Ohler, R. Deventer)

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

Anschaulich ist ein Bayes-Netz ein gerichteter azyklischer Graph, dessen Knoten die physikalischen Mess- und Einstellgrößen der Prozesskette, sowie daraus abgeleitete Qualitätsbewertungen als Zufallsvariablen modellieren und dessen Kanten die Abhängigkeitsstruktur der involvierten Größen repräsentieren. In einem ersten Versuch wurden die verschiedenen Phasen von Spritzgiesen modelliert. Dabei wurden kontinuierliche Größen quantisiert. Es stellte sich heraus, dass eine sehr grosse Parameteranzahl für die Modellierung benötigt wird.

Zur Vermeidung der Quantisierungsfehler wurde im vergangenen Jahr das Verhalten von Bayesnetzen untersucht, die sowohl diskrete als auch kontinuierliche Knoten verwenden. Dabei wird davon ausgegangen, dass die Werte eines kontinuierlichen Knotens normalverteilt sind, wobei sich deren Mittelwert als Linearkombination aus den Werten der Vaterknoten ergibt. Diese hybriden Bayesnetze wurden verwendet, um den Zusammenhang zwischen Volumen und Druck beim Innenhochdruckumformen in den beiden Phasen Vorformen und Kalibrieren zu modellieren. Beide Kurven sind nichtlinear, die Kurve beim Kalibrieren, die Abbildung 1 zeigt, weist zusätzlich noch Unstetigkeiten auf.

Zur Approximation der Kurven wurden die Knoten eines hybriden Bayesnetzes so angeordnet, dass sie eine lineare Approximation der Kurve mit mehreren Stützstellen berechnen. Hierbei gibt ein kontinuierliches - diskretes Wertepaar jeweils den Originalwert V und eine dazu benachbarte Stützstelle V_d an. Ein weiterer Knoten Q_e berechnet den Quantisierungsfehler $V_d - V$. Durch diese Architektur ist der Knoten P , bei entsprechend gesetzten oder gelernten Gewichten fähig eine lineare Approximation durchzuführen, deren Stützstelle durch den diskreten Knoten V_d ausgewählt wird. Die dabei verwendete Architektur des Bayesnetzes zeigt Abbildung 2:

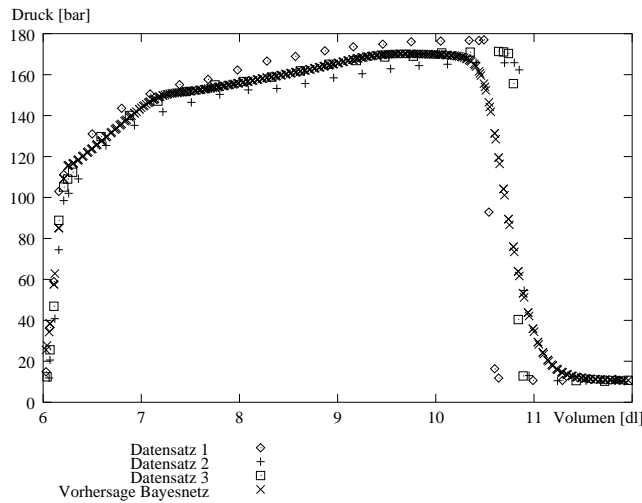


Bild 1: Druck - Volumenzusammenhang beim Kalibrieren

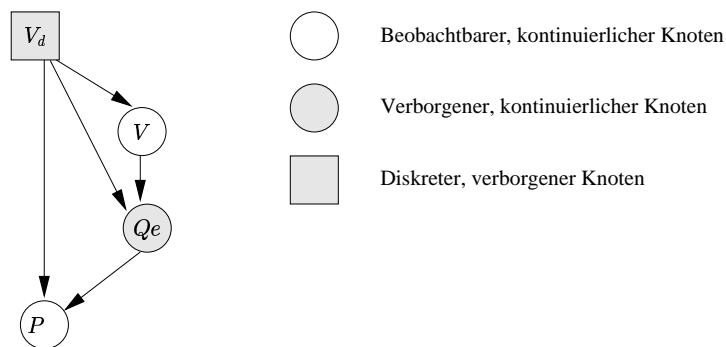


Bild 2: Vorhersage des Bayesnetzes beim Kalibrieren

Verwendet man die soeben vorgestellte Architektur zur Vorhersage des Druckes bei gegebenem Volumen, so ergibt sich ein relativer Fehler von 20,58%, bei Vorhersage des Volumens ein Fehler von 4,3%. Der relative Fehler bei der Vorhersage des Druckes beim Kalibrieren entsteht dabei in der Nähe der Sprungstelle, wobei die Tatsache, dass die Sprungstelle nicht immer an der gleichen Stelle liegt, eine der Ursachen ist. Wesentlich bessere Ergebnisse lassen sich beim Vorformen erzielen. Die Vorhersage des Druckes ist hier mit einer relativem Fehler von 1,95% möglich, beim Volumen ergibt sich ein relativer Fehler von 3,76%. Die Ergebnisse zeigen klar, dass es mit dem angegebenen Verfahren möglich ist auch nichtlineare Bereiche zu modellieren.

Im neuen Forschungsgebiet der Bioinformatik, der rechnergestützten Analyse molekularbiologischer Daten wie etwa des Erbmaterials DNA, wurde das im Jahr zuvor begonnene Projekt „Statistische Modellierung, Lokalisierung und Analyse regulatorischer DNA-Sequenzen“ fortgesetzt. Das Projekt wird vom Boehringer Ingelheim Fonds gefördert und in Zusammenarbeit mit dem Institut für Klinische und Molekulare Virologie der Universität Erlangen und dem Berkeley

Drosophila Genome Project an der University of California durchgeführt. Ziel des Projektes sind Verfahren der Mustererkennung, die spezifische Teilsequenzen, sogenannte Promotoren, im Erbgut von Organismen identifizieren können. Promotoren sind den proteincodierenden Abschnitten der DNA vorgelagert und weisen eine komplexe, oft sehr variable Struktur auf. Sie sind wesentlich an der differenzierten Regulation von Genen beteiligt.

Das entwickelte System stützt sich auf interpolierte Markovketten [38] und wurde in zwei Richtungen erweitert: Zum einen wurden neue, diskriminative Trainingsverfahren implementiert [37], die es ermöglichen, auch negative Beispiele in das Training eines Modells mit einzubeziehen. Zum anderen wurde ein stochastisches Segmentmodell für Promoterregionen entwickelt. Dieses Modell repräsentiert eine Promotorsequenz als Folge von fünf Segmenten, die unüberwacht gelernte Teilmodelle enthalten.

Mit unserem Algorithmus nahmen wir am internationalen „Genome Annotation Assessment Project“ teil, das die Beurteilung der Leistungsfähigkeit aktueller Verfahren zur automatischen Annotierung der Genome komplexer Organismen zum Ziel hatte.

4 Bildanalyse

(U. Ahlrichs, B. Caputo, F. Deinzer, J. Denzler, J. Drexler, A. Gebhard, B. Heigl, Y. Huang, T. Merz, D. Meyer, W. Obermayer, D. Paulus, M. Reinhold, C. Yuan, M. Zobel)

Ausstattung

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 einem rechnergesteuerten Stereogestell mit hochwertigen Farbkameras, welche die Szene überwachen. Zum Roboter zählen zudem ein Drehteller und ein Linearschlitten. Speziell ausgestattete Bildeinzugsrechner vom Typ SGI Onyx gestatten die Digitalisierung zweier Bildkanäle mit 25 Bildern pro Sekunde und voller S-VHS Auflösung. Die bewegliche Kamera ermöglicht es, ein Objekt aus verschiedenen Blickwinkeln aufzunehmen, sodass dessen Oberfläche vollständig erfasst werden kann. Dies ist insbesondere dann notwendig, wenn ein dreidimensionales Modell des Objektes gebildet werden soll, oder eine einzelne Ansicht nicht ausreicht, um ein komplexes Objekt zu erkennen. Bildverarbeitung wird verteilt an HP-Workstations, SGI-Workstations (O2), und an hochwertigen PC's durchgeführt.

Für die laufenden Projekte auf dem Gebiet der probabilistischen Folgenmodellierung sowie auf dem Gebiet des Rechnersehens für autonome mobile Systeme steht seit Anfang 1998 die Plattform XR4000 der Firma Nomadic zur Verfügung, die im Projekt MOBSY (s.u.) eingesetzt wird. Die beiden auf der Plattform installierten Rechnersysteme (Pentium Pro und Dual Pentium II 300) ermöglichen eine vollständige Autonomie; die Verbindung zum Rechnercluster des Lehrstuhls wird über ein Funkethernet sichergestellt. Die Plattform verfügt neben Infrarot-, Ultraschall- und mechanischen Sensoren über einen Stereo-Kopf mit Schwenk-Neige-Vergenz-Steuerung und Farbkameras zur visuellen Wahrnehmung der Umwelt. Die Plattform wurde

von Konzeptberechnungen am schnellsten zur Lösung der Szenenexplorationsaufgabe führt.

Zur Szenenexploration ist Information über die dreidimensionale Struktur der Szene notwendig. In den Arbeiten des SFB 182 stellte sich heraus, dass die in Software realisierten Verfahren noch zu ungenau oder nicht effizient genug sind. Aus diesem Grund wurde in dem DFG-Projekt eine 3D-Laserkamera angeschafft, die diese 3D-Information zusammen mit einem Grauwertbild direkt liefert.

Neben den Arbeiten zur Integration der Lernverfahren wurde das System zur Szenenexploration [1] weiter vervollständigt. So wurden die heuristischen Bewertungsfunktionen zur Auswahl von Kameraaktionen durch ein theoretisch fundiertes Nützlichkeitskalkül ersetzt [3] und das durch die Integration notwendige gewordene Schließen des Kreislaufs aus Sensorik und Aktorik realisiert [2]. Experimente in zwei verschiedenen Büroumgebungen, in denen nach drei Objekten, einem Locher, einem Abroller und einem Klebestift, gesucht wurde, belegen durch eine Erkennungsrate von 93 % beziehungsweise 80 % die Funktionsfähigkeit des Systems.

Zur Objektlokalisierung wurden verschiedene Verfahren untersucht, die auf Farbhistogrammen basieren. Die interessierenden Objekte werden dabei mit einem Zoom-Objektiv formatfüllend aufgenommen und das Farbhistogramm in unterschiedlichen Farbräumen und wählbarer Quantisierung gespeichert. Zur Lokalisierung eines Objekts wird die Szene zunächst mit einer geringen Brennweite aufgenommen. Anschließend wird die Schwenk-Neigeposition der Kamera und die Brennweite so verändert, dass das hypothetisierte Objekt sich wieder formatfüllend in der Bildmitte befindet. Durch den Vergleich von Histogrammen des Objekts und von Histogrammen, die aus Teilbildern des Szenenbilds gewonnen werden, können nun Hypothesen für die Position des gesuchten Objekts im Bild gemacht werden. Zum Vergleich von Histogrammen wurden zahlreiche Verfahren untersucht; in [23] wird dargestellt, dass sich das neuartige Verfahren der so genannten "earth movers's distance" als Maß für den Vergleich gut eignet. Allerdings ist im praktischen Einsatz die hohe Laufzeitkomplexität oft ein Grund, ein schnelleres Verfahren — beispielsweise den statistischen χ^2 -Test zu verwenden.

Bild 4 zeigt die Ergebnisse von Histogrammvergleichen mittels χ^2 Test in Abhängigkeit der gewählten Farbräume und der Quantisierung der Farbhistogramme.

Bildbasierte Szenenmodellierung

Im Teilprojekt C2 des Sonderforschungsbereichs 603 werden bildbasierte Verfahren zur Visualisierung komplexer Szenen verwendet. Das Szenenmodell wird generiert, indem mittels einer handgeführten Kamera Bilder einer beliebigen Szene von vielen (mehrere hundert) verschiedenen Seiten aufgenommen werden (Akquisition), die entsprechenden Positionen der Kamera zu jedem Bild direkt aus den Bilddaten errechnet wird (Kalibrierung), eine geometrische Approximation der Szene erfolgt (geometrische Rekonstruktion) und diese Information dazu verwendet wird, neue Ansichten der Szene aus beliebigen Positionen zu erzeugen (plenoptische Modellierung).

Die Untersuchungen zur Kalibrierung gingen dahingehend, die speziellen Eigenschaften von Bildströmen auszunutzen, die mit einer handgeführten Kamera aufgenommen wurden. Dabei ist speziell das Auftreten von Überkreuzungen des Bewegungspfades der Kamera zu erkennen und in der Schätzung zu berücksichtigen. Hierfür wurde in Kooperation mit PSI-VISICS

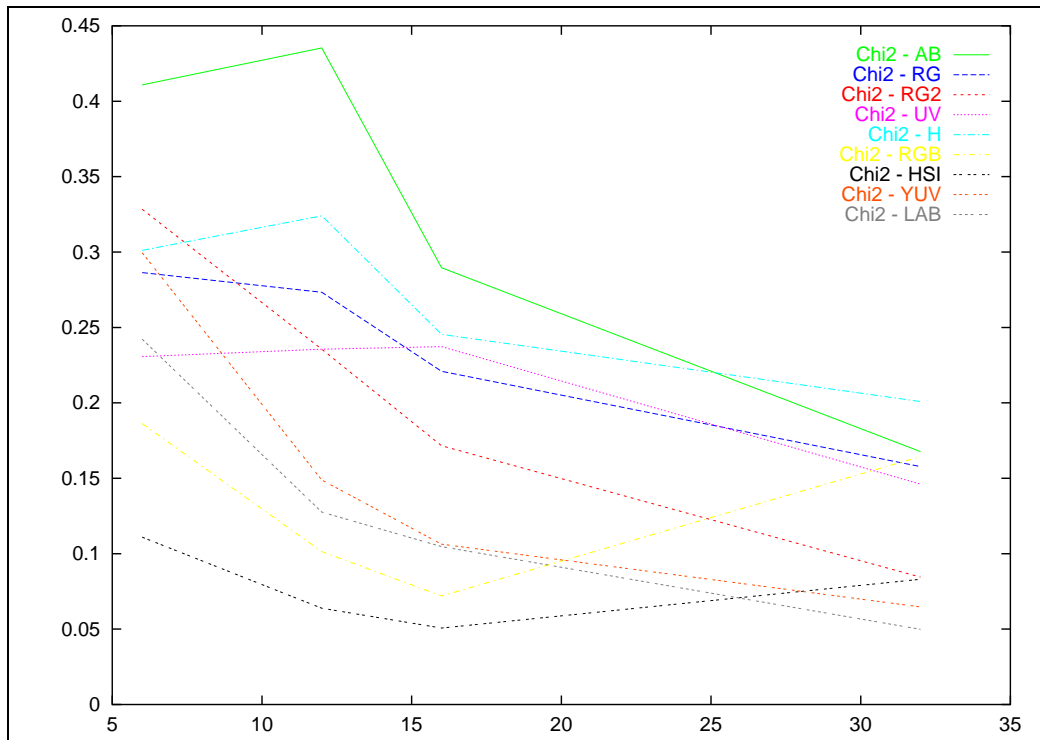


Bild 4: Histogrammvergleich mittels χ^2 Test

Leuven ein Ansatz entwickelt, der sämtliche Nachbarschaftsbeziehungen zwischen beliebigen Bildern der Sequenz berücksichtigt [27]. Um speziell aus langen Punkttrajektorien über viele Bilder hinweg eine stabile Rekonstruktion zu erzeugen, ist aus der Literatur die so genannte Faktorisierungsmethode bekannt. Sie setzt eine vollständige Sichtbarkeit aller verwendeter Szenenpunkte während der gesamten Sequenz voraus und liefert eine projektive Rekonstruktion. Um diese Nachteile zu beheben, wurde diese Methode auf Teilsequenzen angewandt, die Teilrekonstruktionen fusioniert und mittels Selbstkalibrierung eine metrische Rekonstruktion ermittelt [22]. Bild 5 zeigt das Rekonstruktionsergebnis einer realen Szene.

In der bildbasierten Szenenmodellierung wurden Erweiterungen der bekannten Verfahren entwickelt, die speziell bei der Verwendung von Bildern einer handgeführten Kamera notwendig wurden. Die bisher bekannten Verfahren verwendeten die geometrisch regulär angeordnete Zwei-Ebenen-Parametrisierung des Lichtfeldes, welche bei Verwendung realer Kamerabilder mit nichtregulären Aufnahmepositionen interpoliert werden musste. Um diesen fehlerbehafteten Interpolationsschritt zu vermeiden, wurde eine Möglichkeit entwickelt, direkt aus den Kamerabildern neue virtuelle Szenenansichten zu erzeugen, wobei die rekonstruierte Szenengeometrie adaptiv in wählbarer Genauigkeit verwendet wird. [21, 26].

Plenoptische Szenenmodelle fanden außerdem Anwendung in der automatischen Selbstlokalisierung mobiler Systeme. Hierbei wurde versucht, mittels Optimierung diejenigen Positionsparameter zu finden, für welche eine virtuell erzeugte Szenenansicht mit der realen, von der Kamera gelieferten Ansicht übereinstimmt. Dabei werden für vom probabilistischen Bewegungsmodell

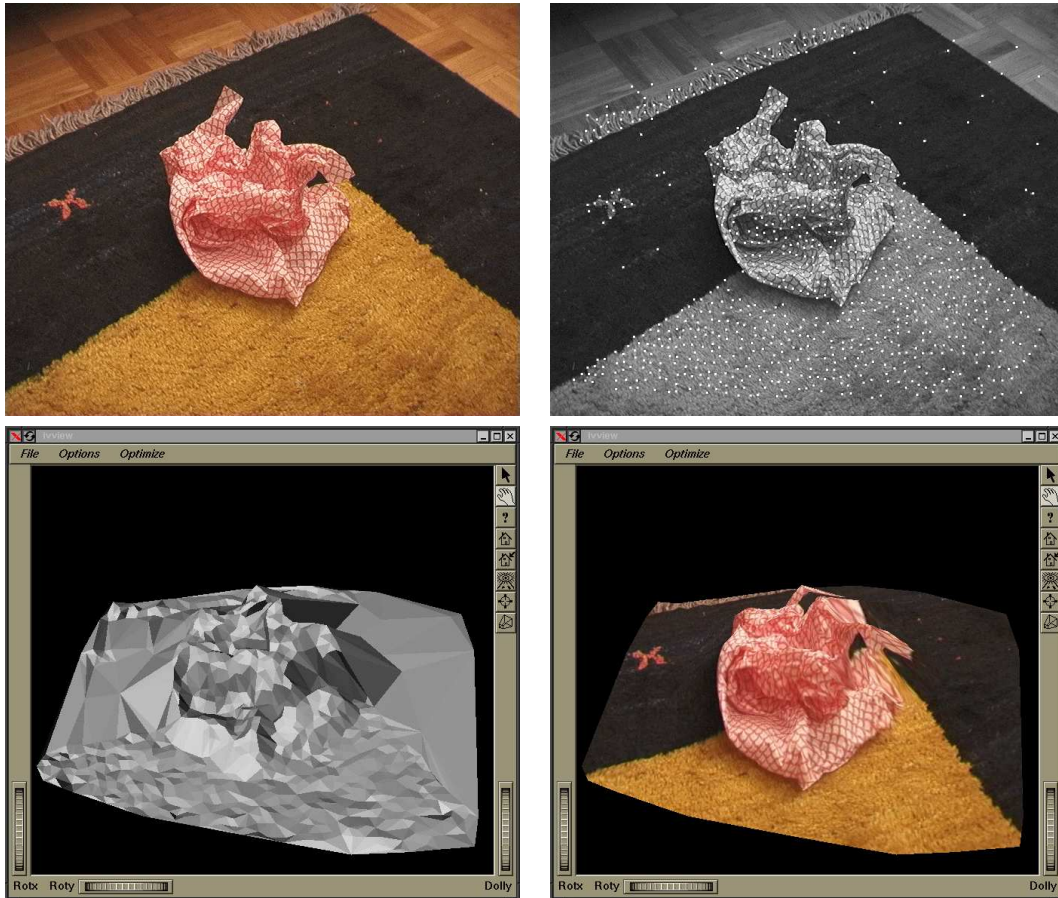


Bild 5: Oben links: Bild einer Sequenz von 100 Einzelaufnahmen. Oben rechts: extrahierte Punktmerkmale, welche über alle 100 Aufnahmen hinweg verfolgt werden konnten. Unten links: die sich nach der Kalibrierung ergebenden rekonstruierten 3D-Punkte als Oberfläche visualisiert. Unten rechts: die rekonstruierte Oberfläche mit Textur.

gelieferte Hypothesen virtuelle Bilder erzeugt und mit dem realen Kamerabild verglichen. In Experimenten in realer Umgebung konnte unter Berücksichtigung einer vom mobilen System ausgeführten Bewegung in wenigen Optimierungsschritten die tatsächliche Position ermittelt werden [?].

Sichtprüfung

Die Arbeiten zur automatischen Sichtprüfung technischer Objekte auf interferometrischer Basis wurden im Rahmen des gemeinsam mit dem Bremer Institut für Angewandte Strahltechnik (BIAS) durchgeführten DFG Projekts abgeschlossen. Das Ziel des Projekts war die Entwicklung eines Prüfsystems zur zerstörungsfreien Identifikation verborgener Materialfehler [28]. Dies sollte durch den Vergleich eines elastisch verformten Prüfobjekts mit der simulierten Verformung

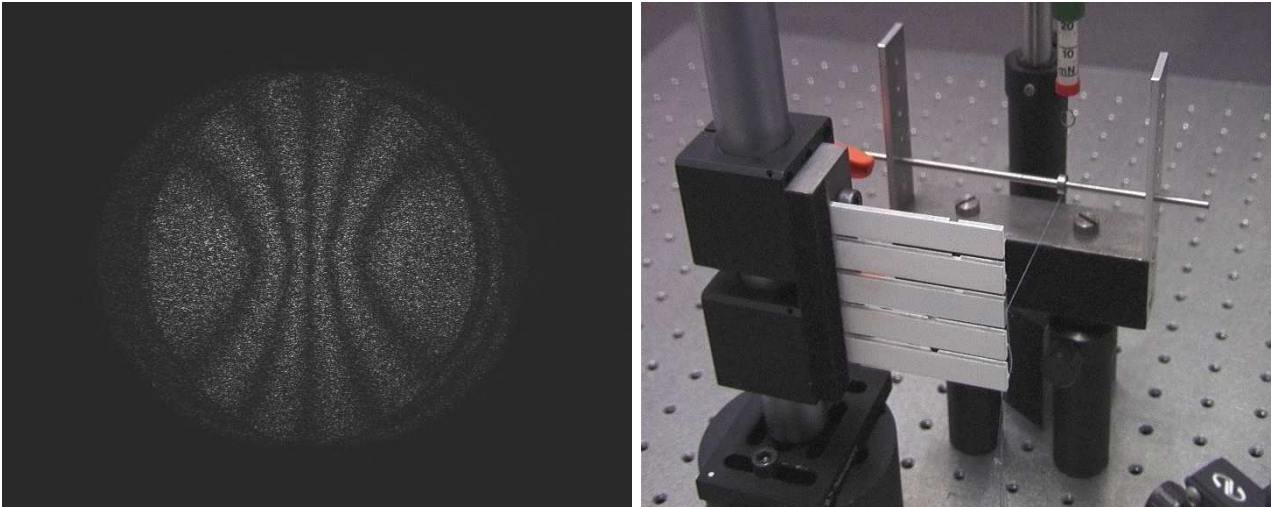


Bild 6: a) Scherogramm einer Kreisplatte b) Versuchsobjekt Balken mit Nut

eines Objektmodells mit geschätztem Materialfehler erzielt werden. Die realen Verformungen wurden mittels interferometrischer Messverfahren gemessen und die simulierten Verformungen analytisch bzw. mittels finiter Elemente berechnet. Der Schwerpunkt der Arbeiten des Lehrstuhls lag in der Entwicklung von Verfahren zur Merkmalsberechnung aus interferometrischen Messungen, der Registrierung der Messobjekte und der Lösung des Optimierungsproblems.

Die prinzipielle Funktion eines solchen Systems wurde am Lehrstuhl anhand einfacher Messobjekte demonstriert. So wurde u.a. die Querkontraktionszahl aus der Biegung einer Kreisplatte (Bild 6a) ermittelt und die Lage und Tiefe verschiedener Nute in Balken geschätzt (Bild 6b). Die Verformung erfolgte durch eine punktförmige Zugbelastung. Zur Lösung des nicht-linearen Optimierungsproblems wurde zunächst eine Parameterhypothese generiert und dann die Maximum Likelihood Funktion mit dem Levenberg-Marquardt Verfahren maximiert. Als Messverfahren wurde neben der holografischen Interferometrie im vergangenen Jahr auch die Speckle Interferometrie und die in der Industrie verbreitete Scherografie eingesetzt (Bild 7).

Zur Ermittlung der Genauigkeit der Parameterschätzung wurden Versuche mit allen Mess- und Auswerteverfahren zur Messung des Rauschens der Eingangsdaten des Parameterschätzers durchgeführt. Mittels Monte Carlo Methoden konnte somit auch ohne Verwendung einer großen Stichprobe die Streuung der Parameter ermittelt werden. Es hat sich am Beispiel der Balken gezeigt, dass selbst im Falle der Anwendung des Systems zur Identifikation von Wanddickenschwächungen einfacher Objekte, das Schätzproblem schlecht konditioniert ist und erst ausgeprägte Schwächungen sicher gemessen werden können. Zudem hat sich herausgestellt, dass die Simulation realer Verformungen weitaus schwieriger ist, als angenommen wurde und die dabei gemachten Fehler berücksichtigt werden müssen. Inwiefern dieses Prüfverfahren für industriell relevante Prüfobjekte, wie z.B. Satellitentanks geeignet ist, soll in Zukunft am BIAS gezeigt werden.

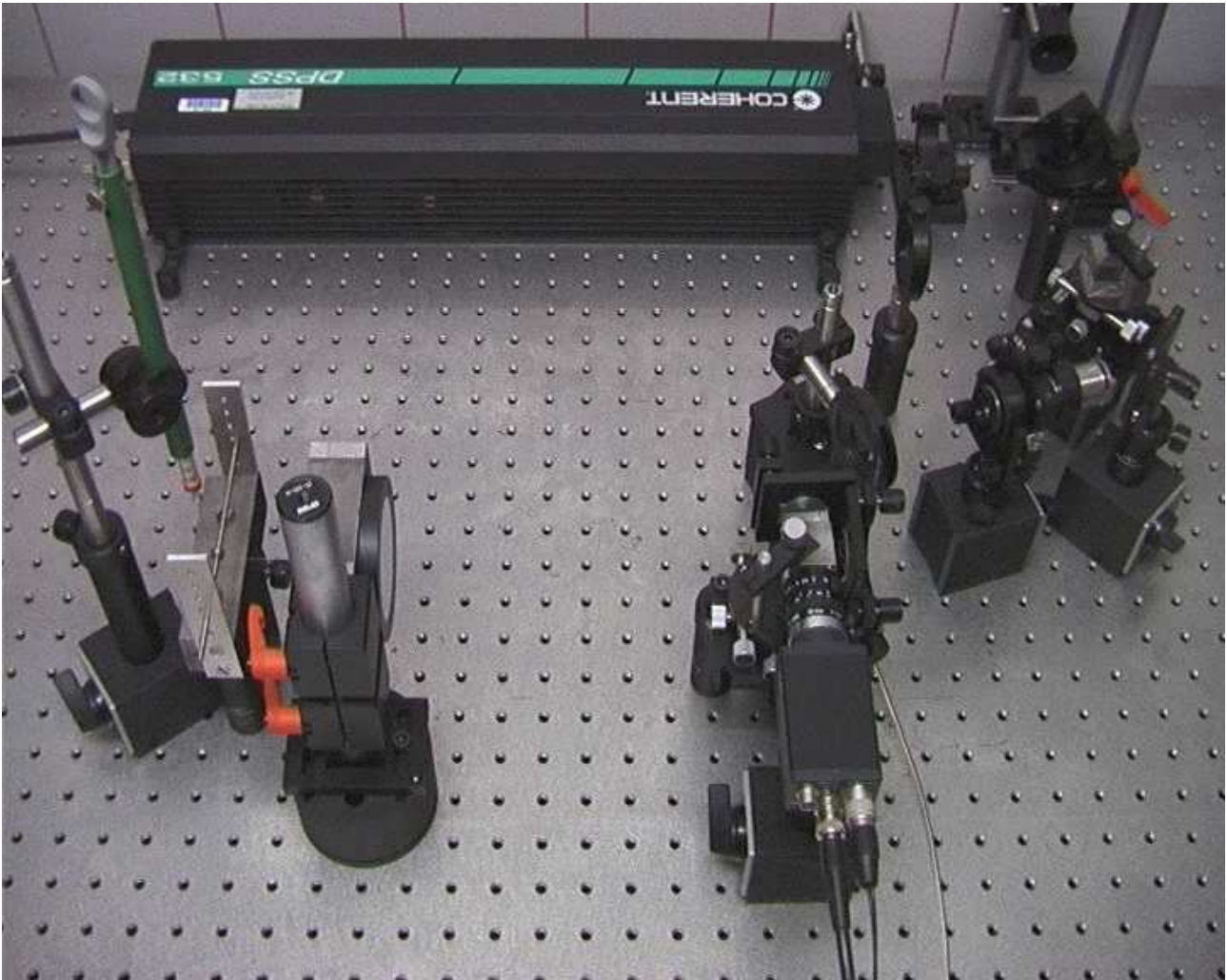


Bild 7: Versuchsaufbau zur Scherografie mit Kreisplatte als Messobjekt

Modellierung und Erkennung von Objekten

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

Im Gegensatz zu üblichen neuronalen Verfahren werden hier segmentierungsfreie Methoden verwendet, um unerwünschte Unter- bzw. Übersegmentierung zu vermeiden. Dabei wird versucht, die Intensitätswerte der Bildpunkte direkt als ursprüngliche Merkmale zu verwenden. Mit drei unterschiedlich strukturierte neuronale Netze werden Objekte detektiert, klassifiziert, und Objektlage geschätzt.

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

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

Um bessere Ergebnisse bei der Lokalisation von ähnlichen Objekten zu erzielen, wurde speziell ein auf Kohonens SOFM basierendes Verfahren entwickelt [47]. Dieses Verfahren gilt für 2D-Objekte und hat den Vorteil, gleichzeitig ein Objekt zu lokalisieren und zu erkennen. Im Vergleich zu dem holistischen Verfahren erhöht sich die Erkennungsrate um 6.4% auf 95.7% bei gleichen Stichproben mit der Bildgröße von 512 x 512. Außerdem steigert sich die Lokalisationspräzision um 2 Pixel auf 3 Pixel bei der Translation und um 1.5 Grad auf 3.5 Grad bei der Rotation.

In der letzten Zeit wurde eine Strategie zur translationsinvarianten Objekterkennung entwickelt. Auf dieser Strategie aufbauend wurde ein Verfahren zur Mehr-Objekterkennung entwickelt und implementiert. Dabei wurden Experimente mit Verdeckungen und bei komplexem Hintergrund durchgeführt.

Die in Experimenten verwendeten Objekte sind in Bild 8 dargestellt. Für die Szene mit einem Objekt beträgt die Erkennungsrate 97.3% bei homogenem Hintergrund ohne Verdeckung und 80.3% mit Verdeckungen. Für die Szene mit mehreren Objekten wurde eine Erkennungsrate von 70% bei homogenem Hintergrund ohne Verdeckung erzielt. Bei komplexem Hintergrund wurde 72% der Objekte richtig erkannt und 16% davon falsch erkannt. Die Lokalisationspräzision bei 512 x 512 2D-Bildern beträgt durchschnittlich 2.3 Pixel bei der Translation und 5 Grad bei der Rotation mit dem Samplingintervall von 10 Grad. Für die 3D experimente mit der Bildgröße von 256 x 256 Pixel liegt die Lokalisationsgenauigkeit im Mittel bei 1 Pixel bezüglich der Translation, 2 Grad bezüglich der internen Rotation, und 3 Grad bezüglich der externen Rotation.

Im Rahmen des von der DFG geförderten Graduiertenkollegs „Dreidimensionale Bildanaly-

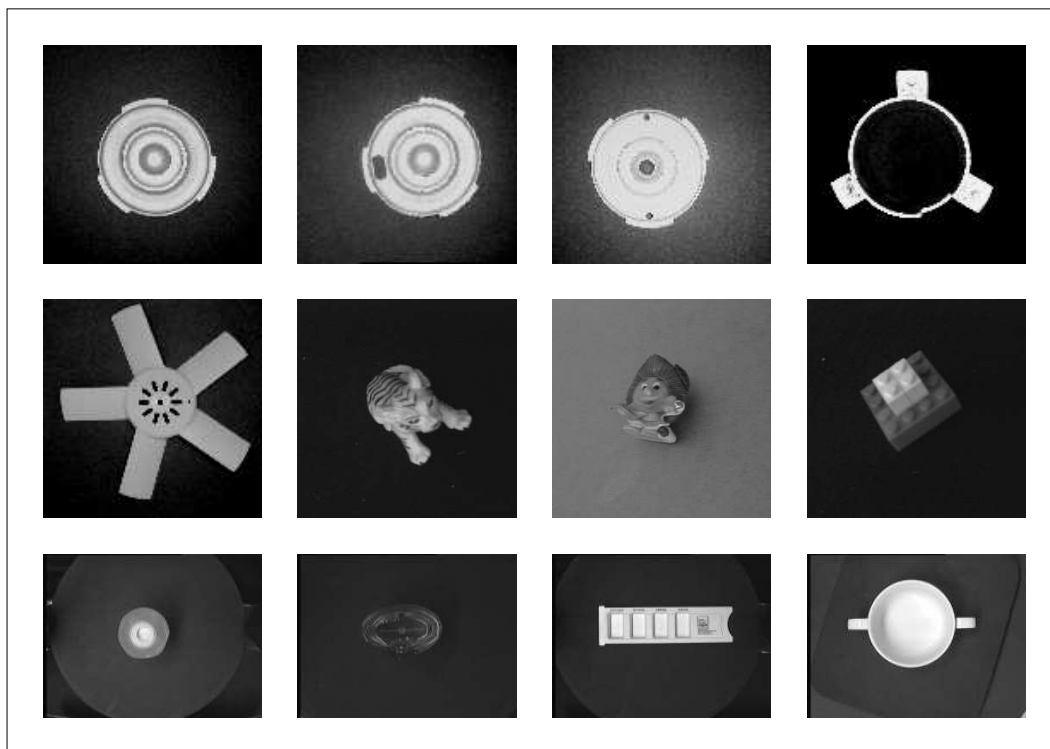


Bild 8: Verwendete 2D und 3D Objekte

se und -synthese“ wurde eine Arbeit zur erscheinungsbasierten statistischen Objekterkennung fortgesetzt. Ziel ist die Lokalisation und Erkennung von 3D-Objekten in 2D-Aufnahmen.

Zahlreiche Ansätze in der Bildanalyse führen dafür zunächst eine Segmentierung der Aufnahmedaten durch und verwenden erst die Segmentierungsergebnisse, wie Kanten und Ecken, für den eigentlichen Lokalisations- und Erkennungsprozess. Im Gegensatz dazu werden bei dem hier verwendeten Ansatz die Intensitätswerte des Bildes direkt ausgewertet.

Um die Berechnungskomplexität zu reduzieren, wird eine Auflösungshierarchie benutzt. Mit Hilfe der diskreten Wavelettransformation werden aus den Bilddaten für jede Auflösungsstufe punktweise lokale Merkmale berechnet. Experimente zeigten, dass dabei mit Johnston-Wavelets die besten Merkmale generiert werden können. Auf der Basis der lokalen Merkmalswerte einer Auflösungsebene lassen sich statistische Dichtefunktionen definieren, die die Wahrscheinlichkeit einer Beobachtung unter der Bedingung eines Lageparametersatzes beschreiben. Die Objektlokalisation kann dann durch eine Maximum-Likelihood-Schätzung durchgeführt werden. Die resultierenden Dichtewerte ermöglichen eine Bayes-Klassifikation.

Die Lokalisationsgenauigkeit liegt bei den in den Experimenten verwendeten 2D-Objekten im Mittel bei einem Pixel bezüglich der Translation und einem Grad bezüglich der Rotation. Die Klassifikation erfolgt nahezu fehlerfrei. Bei den 3D-Experimenten ergab sich bei einem zusätzlichen Freiheitsgrad für die Rotation außerhalb der Bildebene ein mittlerer Rotationsfehler von fünf Grad. Die Lokalisierungsergebnisse können durch Einbeziehung lokaler Abhängigkeiten verbessert werden. Zur besseren Modellierung von heterogenem Hintergrund oder Mehrobjektszenen eignen sich Mischungsverteilungen, die Einzelobjektdichten und eine Hintergrunddichte miteinander kombinieren. Damit lassen sich auch bei Objektverdeckungen Erkennungsraten von 75% bei 2D-Objekten erreichen.

Bewegung und Verfolgung

Im Rahmen des von der DFG geförderten Graduiertenkollegs für Dreidimensionale Bildanalyse und -synthese

wurde die Arbeit über Gangartenklassifikation abgeschlossen. Das Ziel dieser Arbeit ist, mit Hilfe eines statistischen Ansatzes verschiedene menschliche Gangarten, z.B. Gehen, Laufen, Hüpfen und Humpeln zu unterscheiden. Im medizinischen Bereich gibt es weitere Anwendungen bei der Erkennung von pathologischen Gangarten zur Unterstützung der Diagnosestellung.

Aus den Differenzbildern zweier aufeinander folgender Bilder wird der Bereich des Bildes bestimmt, in dem sich eine sich bewegende Person befindet. Anhand dieser beiden Bilder wird das Verschiebungsvektorfeld bestimmt, welches die Bewegungen der Personen in der Bildebene beschreibt. Das Bild wird in verschiedene Bereiche eingeteilt, hier z.B. ein Gitter mit drei Reihen und Spalten. Durch Mittelwertbildung in den einzelnen Bereichen, erhält man die Bewegungen z.B. im Bereich der Füße oder des Kopfes. Diese Mittelwerte beschreiben über die Bildfolge hinweg die Bewegung des Menschen. Die Differenz der Mittelwerte zweier aufeinander folgender Verschiebungsvektorfelder entsprechen der Beschleunigung der Körperteile. Sie werden ebenfalls als Merkmale benutzt. In jedem der neun Bereiche werden also Mittelwerte und Beschleunigungen in x - bzw. y -Richtung bestimmt, es ergibt sich ein 36-dimensionalen Merkmalsvektor.

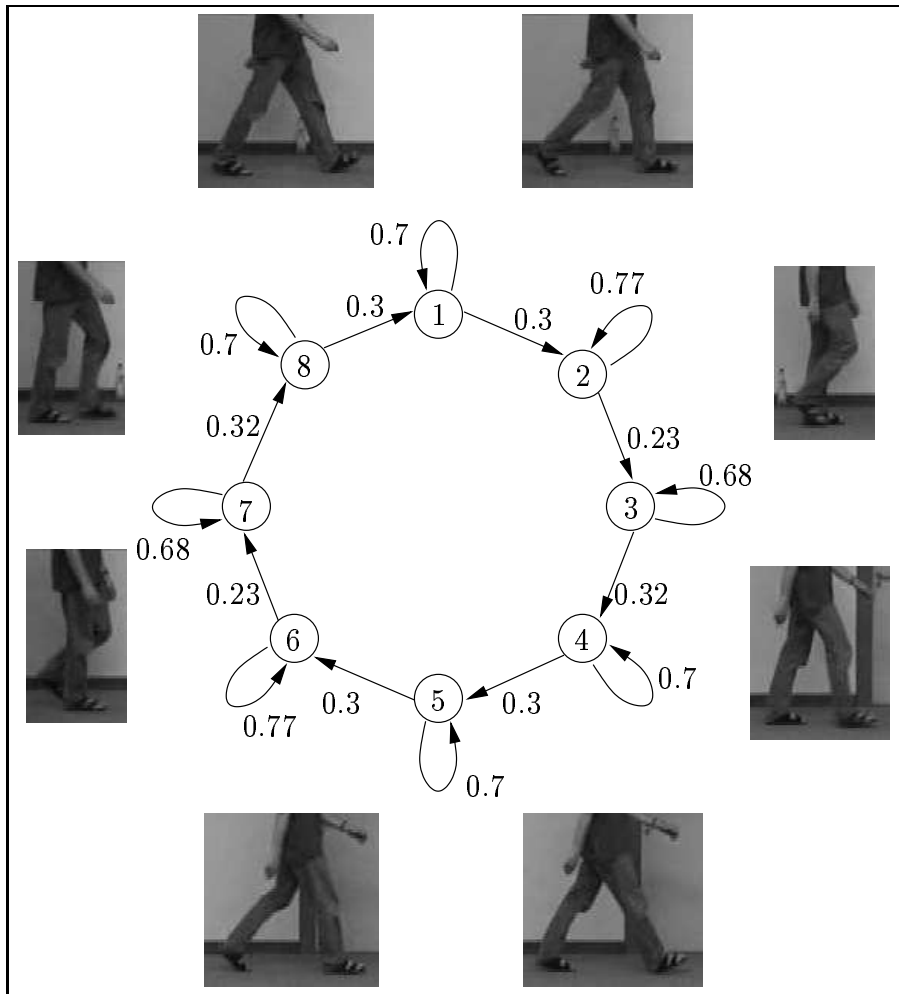


Bild 9: Gehen. Der Bewegungsablauf wird in acht Phasen unterteilt, die den Zuständen eines HMMs entsprechen. Die möglichen Übergänge sind eingezeichnet, die Bewegung ist periodisch. Die Bildausschnitte zeigen die Beine in den Positionen, in denen ein Übergang von einer Phase in die nächste geschieht.

Der Gang von Personen ist ein zyklischer Vorgang. In dieser Arbeit wird er durch ein Hidden–Markov–Modell (HMM) beschrieben. Jede Gangart besteht demnach aus acht Zuständen, die zyklisch durchlaufen werden. Eine Periode umfasst dabei einen rechten und einen linken Schritt. Die Zustände der HMMs entsprechen bestimmten Phasen des Ganges. Dieses ist im Bild am Beispiel des Gehens dargestellt. Die eingetragenen Übergangswahrscheinlichkeiten ergeben sich aus einer Menge manuell initialisierter Trainingsdaten. Der erste Zustand (1) bezeichnet die Phase, in der beide Füße den Boden berühren, also den ersten Doppelkontakt. Das linke Bein ist dabei vorne. In der zweiten Phase wird das rechte Bein nach vorne gezogen, bis zu dem Zeitpunkt, an dem sich beide Füße nebeneinander befinden. Dann beginnt die dritte Phase. In diesem Zustand befindet sich die Person so lange, bis der rechte Unterschenkel senkrecht zum Boden ist. Die vierte Phase beinhaltet das Strecken des rechten Beines nach vorne, bis zum Aufsetzen des Fußes. Aufgrund der Periodizität und der Symmetrie des Ganges sind die nächsten vier Zustände analog zu finden, die Doppelstandphase mit dem rechten Fuß vorne (5), das Anziehen des linken Fußes bis zur Höhe des rechten Beines (6), das Schwungholen bis der linke Unterschenkel senkrecht ist (7), und schließlich das Strecken des linken Beines bis zum Aufsetzen (8). Nach dieser Phase beginnt die Bewegung wieder im ersten Zustand.

Wenn die Personen sich schräg auf die Kamera zu oder von ihr weg bewegen, so wird ebenfalls die Richtung bestimmt, in der sie gehen. Für jede Gangart in jeder Bewegungsrichtung wird dabei ein eigenes HMM trainiert. Eine Person, die auf die Kamera zu läuft, erscheint im Bild erst klein, dann immer größer. Anhand der Größenänderung kann eine Aussage über die Bewegungsrichtung gemacht werden, die als a priori Wahrscheinlichkeit bei der Klassifikation berücksichtigt wird.

Die Erkennungsrate für die Klassifikation der vier Gangarten liegt bei 92.4 %. Im klinischen Umfeld wurden erste Experimente mit pathologischen Gangarten durchgeführt. Dabei wurde der Gang von Patienten mit Morbus Perthes und Arthrose den Gangarten zugeordnet. Um eigene Modelle für jede Gangart zu trainieren, sind allerdings umfassendere Untersuchungen mit viel Datenmaterial nötig.

Analyse von Gesichtsbildern

Der Einsatz der Bildanalyse zur Rehabilitation von Patienten mit Gesichtslähmungen wird im Teilprojekt B3 des Sonderforschungsbereichs 603 (SFB 603) untersucht. Ein Teil des Gesamtvorhabens ist ein System zur Diagnoseunterstützung von Gesichtsparesen. Eine zu untersuchende Person blickt in eine Kamera, die an einem Rechner angeschlossen ist. Von der Person werden fünf Portraitaufnahmen gemacht. In der ersten hält er das Gesicht entspannt. In den nächsten vier führt er mimische Übungen aus, die die Gesichtsparese auf kanonische Weise verstärken. Das sind die Übungen Augen schließen, die Stirn runzeln, den Mund spitzen und die Zähne zeigen.

Die Diagnoseaufgabe wurde unterteilt in die Bereiche Lokalisation von Gesichtern und Gesichtsmerkmalen, Analyse der Gesichtsasymmetrie im Augen- und Mundbereich und die Klassifikation der extrahierten Asymmetrieparameter zur Erkennung einer eventuell vorliegenden Parese. Für die drei Teilprobleme wurden Lösungen angegeben.

Zur Lokalisation von Gesichtern und Gesichtsmerkmalen wurden folgende Einschränkun-

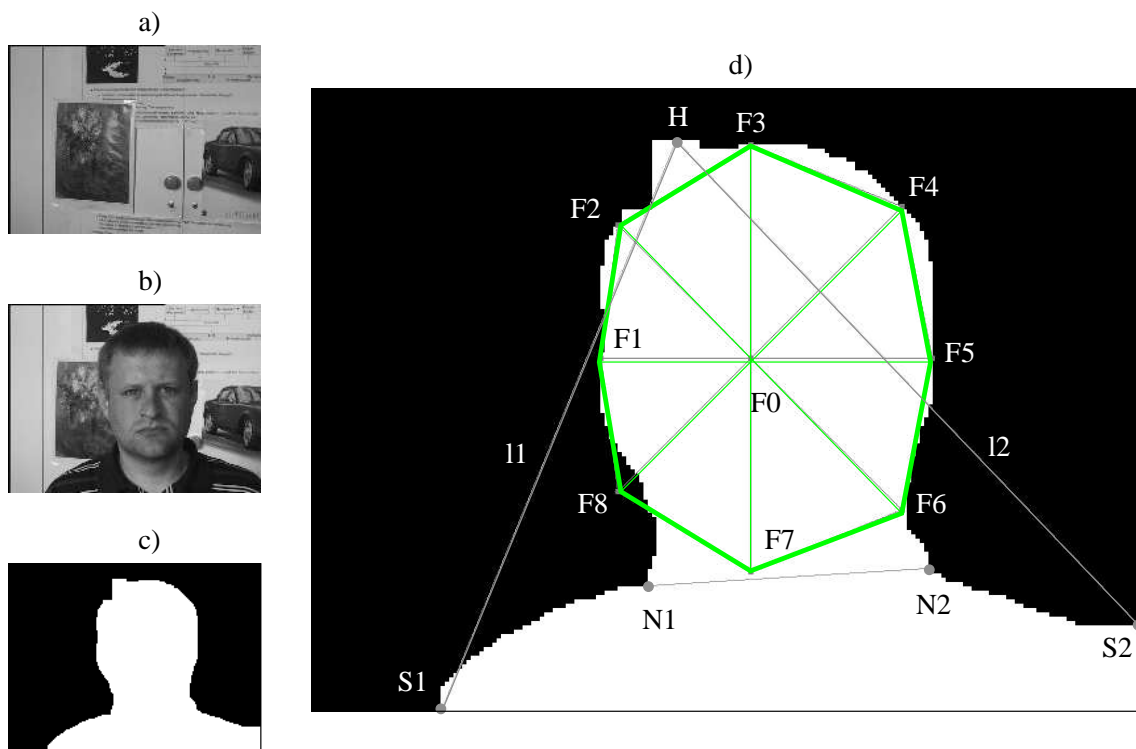


Bild 10: Lokalisation eines Gesichts

gen getroffen: Die zu untersuchende Person sitzt vor einem heterogenen, statischen, bekannten Hintergrund (Bild 10a). Die Kamera ist derart konfiguriert, dass von dieser Person Portrait-aufnahmen (Kopf-/Schulteraufnahmen) generiert werden können (Bild 10b).

Zur Lokalisation wird das Differenzbild des Hintergrunds mit dem Protraitbild berechnet. Im binarisiertem Differenzbild (Bild 10c) erscheint die Kontur des Patienten in der die Kopfregion gefunden wird. Innerhalb der Kopfregion werden modellgetrieben Augen- und Mundwinkel bestimmt.

Die lokale Umgebung dieser Gesichtspositionen wird mit mittelnden Keilfiltern (d.h. speziellen steuerbaren Filtern) bzgl. der Orientierungsinformation analysiert. In der Orientierungsinformation sind u. a. die Öffnungswinkel von Augen und Mund enthalten. Diese Winkel können sich bei Patienten mit Gesichtslähmungen stark unterscheiden und somit zu Asymmetrien zwischen den beiden Gesichtshälften führen. Ein Vergleich der Orientierungsinformation aus der linken und rechten Gesichtshälfte liefert Merkmale über die Gesichtsasymmetrie.

Auch durch die Beleuchtung oder sonstige anatomische Gegebenheiten können Asymmetrien im Gesicht auftreten, die nicht bei der Bewertung der Gesichtslähmung berücksichtigt werden dürfen. Daher werden die Merkmale, die extrahiert wurden während der Patient mimische Übungen ausführte, mit denen normiert (durch Division), die aus dem entspannten Gesicht extrahiert wurden. Liegen die normierten Merkmale des Mundbereichs über einem Schwellwert, so gehen wir davon aus, dass aufgrund mimischer Übungen starke Asymmetrien im Gesicht auftreten und eine Gesichtslähmung in der entsprechenden Gesichtsregion vorliegt.

In einem ersten Experiment wurden 19 Personen vom System untersucht. Von diesen hatten 10 Lähmungen im Augenbereich und 12 Lähmungen im Mundbereich. 7 der 10 Augenlähmungen und 8 der 9 gesunden Augen wurden richtig erkannt. Bei den Gesichtslähmungen im Mundbereich wurden 10 der 12 gelähmten Münder und alle gesunden Münder korrekt erkannt.

Autonome Systeme

4.0.1 Modellierung von Sensordaten- und Aktionsfolgen

Ein zweiter Forschungsschwerpunkt besteht auf dem Gebiet der probabilistischen Modellierung von Sensordaten- und Aktionsfolgen für das aktive Rechnersehen. Dabei werden vor allem im Teilprojekt B2 des Sonderforschungsbereichs 603 Verfahren zur optimalen Einstellung der Brennweite zweier Kameras zur Verfolgung bewegter Objekte sowie zur optimalen Positionierung der Kamera und Brennweiteinstellung für die Objekterkennung entwickelt. In beiden Fällen werden dabei auch hierarchische Modelle untersucht, die auf dem im Teilprojekt C2 des Sonderforschungsbereichs 603 untersuchten Lichtfeldes beruhen. Ein ideales Anwendungsgebiet für die Verifikation der entwickelten Methoden befindet sich im Bereich der autonomen mobilen Systeme. Dort wurden neben Verfahren zur Erkennung von Objekten im Krankenhausbereich mit Anwendung für Serviceroboter (DIROKOL) ein Ansatz zur sichtbasierten, globalen Lokalisation der autonomen Plattform entwickelt. Ein wissensbasierter Ansatz zur Modellierung dynamischer, stochastischer Systeme mittels Bayes-Netze wird im Teilprojekt C1 des Sonderfor-

schungsbereichs 396 entwickelt, mit dem Ziel einer sensor- und modellgestützten Modellierung von Prozessketten.

Auf dem Gebiet der probabilistischen Folgenmodellierung wurden die Untersuchungen zur Zustandsschätzung stochastischer, dynamischer Systeme fortgesetzt. Mittels so genannter Particle Filter, einem im Rechnersehens jüngst entdeckten Vorgehens aus dem Bereich der Monte Carlo Simulation, ist es möglich eine Zustandsschätzung eines stochastischen, dynamischen Systems (Position der Plattform in der Ebene, Rotationswinkel) auch im Falle multimodaler Wahrscheinlichkeitsverteilungen über den Zustandsraum durchzuführen. Damit wird das klassische Kalman-Filter, dem eine Normalverteilung des Zustands als Annahme zugrunde liegt, verallgemeinert.

Wichtigstes Prinzip der Particle Filter ist die Verwendung einer so genannten Zustandsmenge für die diskrete Repräsentation der im Allgemeinen multimodalen Verteilung über den Zustandsraum. Jedes einzelne Teilchen (Particle) in der Zustandsmenge umfasst einen möglichen Zustand des Systems, dessen Plausibilität durch ein Wahrscheinlichkeitsmaß ausgedrückt wird. Nach dem Zustandsübergang des Systems vom Zeitpunkt t nach $t + 1$ werden mittels Sampling Mechanismen (Importance Sampling, Factored Sampling) Teilchen aus der Zustandsmenge gezogen, mit einer Wahrscheinlichkeit proportional zur Plausibilität des repräsentierten Zustands. Das Teilchen wird gemäss der Verteilung über den Zustandsübergang zum nächsten Zeitpunkt fortgeschaltet. Die Ermittlung der Plausibilität zum Zeitpunkt $t + 1$ erfolgt dann durch Auswertung einer Likelihood-Funktion für die Beobachtung (im Falle des Rechnersehens, die Bildinformation) gegeben den Zustand. Zahlreiche theoretische Aussagen existieren, die unter anderem beweisen, dass im Grenzübergang die durch die diskrete Teilchenmenge repräsentierte Verteilung gegen die wahre Zustandsverteilung konvergiert.

Der Ansatz der Particle Filter wurde in einer Arbeit zur visuellen Selbstlokalisierung mit dem Ansatz der Lichtfelder als Szenenmodell gekoppelt. Die Zustandsschätzung umfasst dabei die Schätzung der Position der Plattform in der Ebene des Stockwerks sowie die Drehlage. Das stochastische Bewegungsmodell wird mittels lokaler Bewegungsinformation, die von der Odometrie geliefert wird, ermittelt. Die Auswertung der Likelihood für die Bildinformation, gegeben die Position und Drehlage im Raum, erfolgt über das Lichtfeld, mittels dem es möglich ist, sehr effizient neue, photorealistische Ansichten der Szene synthetisch zu erzeugen. Diese Erwartung aus der Zustandsschätzung über die nächste aufgenommene Bildinformation wird im folgenden mit der tatsächlichen Bildinformation, die von der Kamera geliefert wird, verglichen und liefert so ein Wahrscheinlichkeitsmass für die benötigte Likelihood.

Der vorgeschlagene Ansatz konnte auch experimentell auf seine Güte hin verifiziert werden. In Bild 11 ist die real aufgenommen Bildinformation, und sowohl die erwartete Bildinformation auf Grund der Odometrie als auch die Bildinformation, die sich auf Grund der Zustandsschätzung ergeben müsste, nach einer Fahrt der mobilen Plattform von ca. 3 Meter zu sehen. Die Akkumulation des Fehlers in der Odometrie über die Zeit hinweg verhindert eine genau Positionsschätzung (siehe Abweichung der erwarteten Bildinformation auf Grund der Odometrie, Bild 11, mitte), während die probabilistische Zustandsschätzung mittels Particle Filter eine sehr genaue globale Lokalisation ermöglicht (Bild 11, rechts).

Neben der Selbstlokalisierung von autonomen Dienstleistungsrobotern sind vor allem auch die Lokalisation und Klassifikation von Objekten Schlüsseltechnologien für die Durchführung der



Bild 11: Reales Kamerabild (links), geschätzte Ansicht auf Grund der Odometrie (mitte) und geschätzte Ansicht mittels Particle Filter und Lichtfelder als Szenenmodell (rechts). Die geschätzten Ansichten wurden mittels Lichtfelder synthetisch generiert.

vielfältigen Hol- und Bringdienste und müssen robust gegenüber wechselnden Umgebungsbedingungen und Objektverdeckungen sein. Algorithmen hierfür werden im Rahmen des Projekts DIROKOL entwickelt.

Von besonderer Bedeutung ist dabei, dass die Verfahren robust gegenüber Rauschen, Objektverdeckung und unterschiedlicher Beleuchtung sind. Gerade in dem geplanten Einsatzgebiet kann nicht von einheitlichen Umgebungsbedingungen ausgegangen werden, wie sie beispielsweise in vielen industriellen Anwendungen vorausgesetzt werden.

Standardmethoden, die zur Modellierung von Objekten geometrische Primitiva, wie Kanten oder Eckpunkte, benutzen, werfen in dieser Hinsicht verschiedene Probleme auf. Die Objekterkennungsverfahren basieren auf der korrekten Zuordnung zwischen den aus den Bildern extrahierten Merkmalen zu denen der Objektmodelle. Durch stark strukturierte Oberflächen oder Hintergrund kann es zu einer grossen Zahl von extrahierten Merkmalen kommen, die einen komplexen Zuordnungsprozess nach sich ziehen. Legt man sich auf wenige, sehr signifikante Merkmale fest, kann es bei Verdeckungen vorkommen, dass sich keine Objektmerkmale mehr extrahieren lassen.

Diesen Problemen wird in DIROKOL mit dem Einsatz von erscheinungsbasierten Verfahren begegnet, mit denen sich komplexe Objekte auch bei unterschiedlicher Beleuchtung handhaben lassen. Dabei erfolgt kein Informationsverlust durch die Segmentierung, vielmehr werden Merkmale aufgrund des Farbwerts eines Bildpunktes zur Erkennung herangezogen. Als Objektmodell dienen Bilder aus jeder möglichen Ansicht eines Objekts bei unterschiedlicher Beleuchtung. Um die Informationsflut zu beschränken, werden die Ansichten in einem Eigenraum, basierend auf den grössten Eigenvektoren aller Ansichten, kodiert.

Zu klassifizierende Bilder werden in diesen Eigenraum projiziert und der minimale Abstand zu den Trainingsvektoren sowie der bei der Rückprojektion entstehende Fehler als Gütemass für die Klassenzugehörigkeit verwendet. Dabei findet jedoch nicht das klassische Verfahren mittels Orthogonalprojektion Verwendung. Dies würde nur bei unverrauschten und verdeckungsfreien Objektaufnahmen gute Ergebnisse liefern. Stattdessen erfolgt die Koeffizientenberechnung im Eigenraum dadurch, dass nur eine Untermenge aller Bildpunkte zur Berechnung herangezogen werden. Diese wird anhand von Objektwissen, wie Farbe, und zufallsgesteuert ausgewählt.

Stellt man die zugehörigen Gleichungssysteme für die Darstellung dieser Bildpunkte durch die Eigenvektorwerte auf kann dieses überbestimmte Gleichungssystem mit numerischen Verfahren nach den unbekanntem Koeffizienten gelöst werden.

Durch die Generierung mehrerer Sätze von Objektpunkthypothesen und einer iterative Anpassung können Punkte aus verdeckten Objektbereichen und Rauschen aus dem Berechnungsprozess eliminiert werden und man erhält gute Näherungen für die tatsächlichen Koeffizienten. Bild 12 zeigt ein Objektmodell für eine Tasse, sowie das Ergebnis einer korrekten Klassifikation bei heterogenem Hintergrund.

Für die Objektlokalisierung innerhalb einer Szene wird mit einer Auflösungshierarchie gearbeitet. Auf der niedrigsten Auflösungsstufe werden Hypothesen für Objektposition generiert, die auf den höheren Stufen verifiziert werden.

Im Weiteren wird an der Verbesserung der Erkennungsleistung bei unterschiedlicher Beleuchtung und skalierten Objekten gearbeitet. Um den Trainingsaufwand zu beschränken kann die Beleuchtungsabhängigkeit der Ansichten teilweise algorithmisch erfasst werden. Ähnliches gilt auch für die Skalierung. In die Berechnung der Koeffizienten wird ein Optimierungsprozess hinsichtlich eines Skalierungsfaktors integriert. Dadurch müssen weit weniger explizite Eigenräume für unterschiedliche Skalierungen berechnet werden als das bislang der Fall ist.

4.0.2 Rechnersehen durch Fusion mehrerer Kamerabilder

Ein Schwerpunkt der Forschungsarbeiten im Teilprojekt B2 des Sonderforschungsbereichs 603 liegt in der Erforschung von Verfahren zur aktiven synchronen Verfolgung eines sich bewegendem Objekts mit mehreren Kameras, wobei eine optimale Steuerung der Kameraparameter, z. B. der Brennweiten, angestrebt wird.

Im Bereich des Rechnersehens werden Objekte häufig durch einen einzigen Satz von gleichartigen Merkmalen beschrieben. Beispielsweise durch eine Menge von Kanten, Ecken, Linien, etc. Die Art der Merkmale wird im Allgemeinen a priori festgelegt und ist daher nicht objektspezifisch. Eine Steigerung der Leistung von objekterkennenden oder -verfolgenden Systemen erzielt man durch den Einsatz von objektabhängigen Merkmalen und darauf spezialisierten Sensoren. Man denke beispielsweise bei der Verfolgung von Autos an deren Heckansicht, für die charakteristische Merkmale das weiße Nummernschild und die beiden Rückleuchten sind. Ebenso entscheidend ist dabei die räumliche Anordnung der Merkmale.

Zur ebensolchen Modellierung von Objekten durch die räumliche Beziehung von beliebigen Merkmalen wurde das aktive Objektkonturmodell basierend auf „aktiven Strahlen“ dahingehend verallgemeinert, was zu dem „gekoppelte Struktur“ genannten Objektmodell führte [48]. Dabei wurde konsequent eine probabilistische Formulierung verwendet, durch die sowohl Unsicherheiten in den Daten als auch in den Parametern des verwendeten Objektmodells repräsentierbar werden. In Kooperation mit dem Teilprojekt B3 konnte die Eignung des Objektmodells am Beispiel der Lokalisierung von Gesichtsmerkmalen demonstriert werden [48]. Bild 13 zeigt die Robustheit des Ansatzes bei der Lokalisation im Falle von hochgradigen Störungen und Mehrdeutigkeiten.

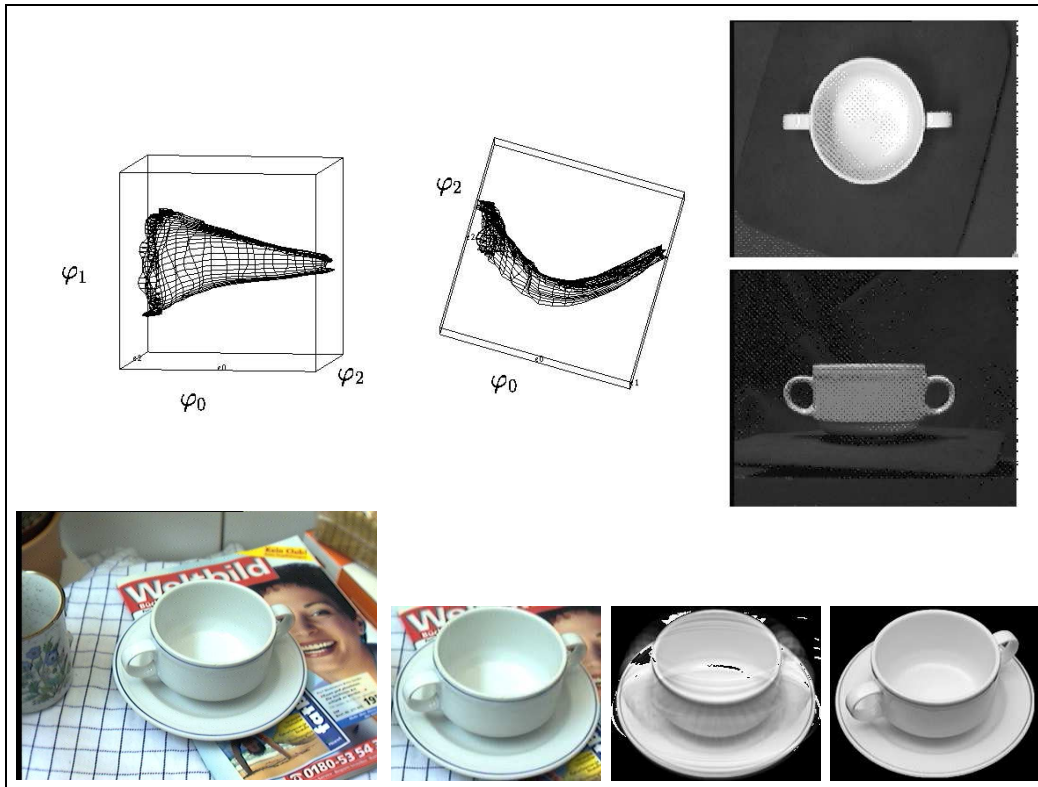


Bild 12: Beispiel für ein Objektmodell im Eigenraum (oben) und die Schritte während der Lokalisations- und Erkennungsphase (unten, v.l.): gesamte Szene, extrahierte Objektposition, Rekonstruktion nach der Koeffizientenberechnung, Visualisierung der geschätzten Lageparameter.

Nach dem statischen Fall der Lokalisierung wurde auch das dynamische Problem der sequentiellen Zustandsschätzung von gekoppelten Strukturen, wie es für die Objektverfolgung erforderlich ist, untersucht. Dabei wurden Verfahren aus dem Bereich der „Particle Filter“ eingesetzt, im Speziellen der so genannte CONDENSATION Algorithmus. Er stellt eine Verallgemeinerung des klassischen Kalman Filters dar, wodurch die Verwendung multimodaler Verteilungen für den Zustandsübergang und das Beobachtungsrauschen ermöglicht wird. Auch hier wurden die Untersuchungen am Beispiel der Gesichtsverfolgung untersucht [49].

Neben den gekoppelten Strukturen wurden auch weiterhin konturbasierte Verfahren verfolgt. Es wurde ein System entwickelt, das es der autonomen, mobilen Plattform MOBSY erlaubt, Personen visuell zu verfolgen und mit konstantem Abstand zu eskortieren. Die zu verfolgende Person wurde dabei durch ihre Kontur des Kopf-Schulterbereichs charakterisiert. Die zeitliche Zustandsschätzung erfolgte wiederum mit dem CONDENSATION Algorithmus, die Bewertung der Güte der geschätzten Kontur wurde über eine probabilistische Diskriminante ermittelt.

Für das Problem der Fusion von Information bei der Verwendung mehrerer Kameras so-

wie die damit verbundene optimale aktive Kamerasteuerung wurden erste theoretische Ansätze angedacht. Diese Problematik wird im folgenden Jahr den Hauptgegenstand der wissenschaftlichen Arbeit darstellen.



Bild 13: Demonstration der Robustheit von gekoppelten Strukturen gegenüber hochgradigen Störungen und Mehrdeutigkeiten. Im Beispiel werden die räumlichen Abhängigkeiten der Gesichtsmarkale Augen und Mund durch drei koppelnde Strahlen modelliert.

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

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

In Bild 14 ist ein Beispiel für derartige Mehrdeutigkeiten zwischen einzelnen Objekten dargestellt. Die gezeigten Tassen unterscheiden sich nur anhand einer Ziffer auf der Vorderseite. Um nun die einzelnen Tassen unterscheiden zu können, muss eine Aktion gewählt werden, aus der eine neue Ansichtsposition resultiert, von der aus die Ziffer sichtbar ist.

Die Suche nach einer neuen optimalen Ansicht ist in dem entwickelten Verfahren als Optimierungsproblem für die Position mit der maximal zu erwartenden Signifikanz, d.h. maximalen

Unterscheidbarkeit der Objekte, formuliert. Dazu wird aus dem während des Trainings gesammelten Wissen — den evaluierten Beispielsichten — eine reelwertige, kontinuierliche Funktion für die erwartete Signifikanz beliebiger Ansichten approximiert. Das Ansichtentraining läuft dabei so ab, dass die Güte einzelner Aktionen von verschiedenen Ansichtspositionen ausgewertet wird. Mit gängigen Optimierungsverfahren wird in dieser Funktion nun nach Maxima gesucht, aus deren Lage daraufhin eine neue Ansichtsposition mit maximaler Signifikanz berechnet wird.

Im Bereich der Ansichtenplanung wird das Verfahren im folgenden Jahr so erweitert, dass multiple Ambiguitäten aufgelöst werden können. Aktuell ist das Verfahren in der Lage, die global beste Ansicht eines Objektes zu finden. Das Verfahren soll nun so erweitert werden, dass weitere „gute“ Ansichtspositionen in Abhängigkeit der bisherigen Aufnahmen berechnet werden können. Das bedeutet, dass Positionen nicht wieder ausgewählt werden dürfen, von denen aus bereits Aufnahmen gemacht wurden.

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

5 Sprachverstehen

(A. Batliner, J. Buckow, W. Fentze, J. Fischer, C. Frank, F. Gallwitz, J. Haas, S. Harbeck, R. Huber, E. Nöth, U. Ohler, G. Stemmer, V. Warnke)

Die inhaltlichen Schwerpunkte der Forschungsaktivitäten zur Sprachverarbeitung bilden das maschinelle Erkennen und Verstehen gesprochener Äußerungen sowie Fragestellungen der Dialogkontrolle [?]. Die Arbeiten im Berichtsjahr konzentrierten sich auf die (Weiter-)entwicklung prototypischer Sprachdialogsysteme. Neben den zwei „traditionellen“ Anwendungsbereichen der Sprachverarbeitungsforschung am Lehrstuhl (InterCity-Fahrplanauskunft mit dem System EVAR und multilinguale Terminabsprache mit dem System VERBMOBIL) wurde ein Dialogsystem für Kinoauskunft entwickelt. Weiterhin wurde mit Arbeiten an der Entwicklung eines multimodalen Dialogsystems (SMARTKOM) und Arbeiten zur Sprecheridentifikation/-verifikation begonnen.

Der für die Zugauskunftsdomäne entwickelte Laborprototyp eines sprecherunabhängigen Systems zur Mensch-Maschine-Kommunikation wurde in wesentlichen Teilen der Verstehensphase (Interpretation und Dialogführung) neu implementiert. Um Erfahrungen mit der schnellen Änderung von Anwendungsdomänen zu sammeln, wurde das Dialogsystem nach der Neuimplementierung auf die neue Domäne „Kinoauskunft“ portiert. FRÄNKI, das *Fränkische Kin*osystem, ist unter 09131/16287 erreichbar und kann Auskunft geben über das Kinoprogramm im mittelfränkischen Raum. Auf der Systems '99 wurde FRÄNKI erfolgreich vorgeführt.

Eine Erweiterung der Forschungsaktivitäten auf dem Gebiet der informationsabfragenden Mensch-Maschine-Dialoge ist die Entwicklung eines Dialogsystems mit Any-Time-Fähigkeit und verbessertem Echtzeitverhalten.

Das zweite Anwendungsfeld (multilinguale Terminabsprache) ist durch die Einbindung des

Lehrstuhls (4 Mitarbeiter/innen) in das BMBF-geförderte VERBMOBIL-Vorhaben bedingt. Projektpartner sind ca. 20 Arbeitsgruppen aus mehreren Universitäten, Großforschungseinrichtungen und Firmen. Im VERBMOBIL-Vorhaben ist ein portables Übersetzungsgerät zu entwickeln, welches auf Konferenzen mit Teilnehmern unterschiedlicher Muttersprachen die Dolmetschfunktion übernimmt. Zur Zeit kann Deutsch, Englisch und Japanisch verarbeitet werden. Das Thema der multilingualen Verhandlung bewegt sich im Rahmen geschäftlicher Terminabsprachen.

Die Dialogsysteme EVAR und FränKi

Im Rahmen des DFG-Projekts „Statistische Modelle für Spontane Sprache“ wurde an der gemeinsamen Verarbeitung von prosodischer Information (z.B. Intonation, Dehnungen) und lautlicher Information gearbeitet. Prosodische Information dient unter anderem dazu, gesprochene Äußerungen in ähnlicher Weise zu untergliedern, wie dies mit Hilfe von Satzzeichen in geschriebener Sprache geschieht. Die zentrale Bedeutung prosodischer Information auch für die automatische Sprachverarbeitung ist allgemein anerkannt. Die Verarbeitung prosodischer Information geschieht jedoch bisher weitgehend getrennt von der Worterkennung. So benötigt das im Rahmen des VERBMOBIL-Projekts am Lehrstuhl für Mustererkennung entwickelte Prosodie-Modul das Ergebnis des Worterkennungsmoduls, um prosodische Grenzen klassifizieren zu können. Auf diese Weise kann detaillierte Information, beispielsweise über die Dehnung oder Betonung benachbarter Wörter, für die Klassifikation herangezogen werden. Der Nachteil dieses Ansatzes ist jedoch, dass die eigentliche Worterkennung *ohne* prosodische Information stattfinden muß; eine Äußerung wird hierbei als unstrukturierte Folge von Wörtern betrachtet. Im Rahmen dieses Projekts wurde nun das weltweit erste Verfahren realisiert, in dem prosodische und lautliche Information gemeinsam verarbeitet und optimiert werden (siehe Bild 15). Dabei wurde zum einen die Erkennung der Wörter verbessert, zum anderen wurden bei der Grenzklassifikation Ergebnisse erzielt, die gleichwertig mit denen sind, die das eigenständige Prosodie-Modul liefert. Gegenüber der bisher praktizierten sequenziellen Verarbeitung bedeutet das integrierte Verfahren vor allem eine deutliche Effizienzsteigerung.

Nachdem sich in den vergangenen Jahren der Einsatz statistischer Modelle auf der Ebene der linguistischen Analyse bewährt hat und die Erkennungsraten in einem akzeptablen Bereich lagen, wurde nun hauptsächlich daran gearbeitet, diese Methoden in EVAR einzubauen. Dabei wird der aktuell verwendete Parser zur linguistischen Analyse ersetzt. Dieser Parser bestand bisher aus einer Grammatik, die alle verstehensrelevanten Teile einer Äußerung vollständig abdeckt. Das neue Modul (vgl. 16) soll dabei aus mehreren Teilen bestehen und die Funktionalität der bereits im Einsatz bewährten Analyse zumindest beibehalten u.U. die Leistungsfähigkeit sogar erhöhen und zudem deutliche Geschwindigkeitsvorteile bieten. Dazu werden die entwickelten statistischen Modelle eingesetzt um für eingehende Sprachsignal zu entscheiden, welche verstehensrelevanten Konzepte (z.B. Zeit-, Datums- oder Ortsangabe) in diesem Signal enthalten sind. Auf Grund der Vorhersage soll anschliessend nicht mehr eine vollständige Grammatik zur weiteren Analyse herangezogen werden, sondern man hat unterschiedliche Grammatikfragmente zur Auswahl, die exakt auf diese verstehensrelevanten Konzepte spezialisiert sind. Durch Verwendung der Fragmente, die von den statistischen Modellen vorhergesagt werden, wird verhindert,

das zu viele unwichtige Teile der Grammatik bei der Analyse mit durchsucht werden müssen. Eine weitere Informationsquelle, die bei diesem Ansatz zur syntaktisch/semantischen Analyse gewinnbringend eingesetzt werden kann, ist die Prosodie. Da mittlerweile die Verarbeitung von Worthypothesengraphen (WHG) Standard ist und diese auch vom Parser gehandhabt werden können, kann die große Anzahl unterschiedlicher Startpunkte für die linguistische Analyse, die sich in einem solchen WHG ergibt, mit prosodischen Markierungen versehen werden, die aussichtsreiche Startpunkt (z.B. die Ortsnamen in den Ortsangaben oder Zahlen in Zeitangaben) bevorzugt.

Ansätze zur statistischen Modellierung von semantischem Wissen werden erst seit kurzem untersucht. Traditionell wird auf diesem Gebiet mit expliziter Wissensrepräsentation gearbeitet. Dies gilt auch für die auf einem semantischen Netz basierende Version des EVAR-Systems, welche weiterentwickelt und an eine parallel iterativ-optimierende Kontrolle angepasst wird. Langfristiges Ziel ist dabei, einen integrierten, multimodalen Ansatz zur Bild- und Sprachanalyse mit Any-Time-Fähigkeiten und Echtzeitverhalten bereitzustellen. Als Grundlagen für die Entwicklung des Dialogsystems dienen der am Lehrstuhl entwickelte parallel iterativ-optimierende Kontrollalgorithmus für die wissensbasierte Musteranalyse und die auf semantischen Netzen basierende Version des am Lehrstuhl für den Aufgabenbereich der Intercity-Auskunft entwickelte Dialogsystem EVAR. Ausgehend von der Zuweisung von Worthypothesen an primitive Attributknoten und von Modalitäten an Konzeptknoten des Attributflußgraphen werden Instanzen für die Zielkonzepte der Analyse, welche einzelne Dialogschritte modellieren, berechnet. Dies wird mittels iterativer Optimierung durchgeführt, wobei nach jedem Iterationsschritt eine partielle Lösung der Analyse, d.h. eine bewertete Instanz zu einem Zielkonzept, vorliegt (Any-Time Fähigkeit). Die Echtzeitfähigkeit des Systems wird durch die potentiell vorhandenen Parallelitätsmöglichkeiten unterstützt.

Das linguistische Wissen des Dialogsystems wird mittels eines semantischen Netzwerkformalismus repräsentiert. Die ursprünglich von demselben Formalismus bereitgestellte A*-Kontrolle wird in dem neuen System durch den oben genannten parallelen Kontrollalgorithmus ersetzt. Dazu sind Erweiterungen bzw. Änderungen sowohl des prozeduralen als auch des deklarativen Wissens der Wissensbasis notwendig. Ausgehend von dem im vorherigen Jahr erstellten Schema für die gesamte Dialogführung mit paralleler Kontrolle wurde im Berichterstattungsjahr ein Systemprototyp fertiggestellt, welcher folgende Dialogschritte verarbeiten kann: Benutzer-Erstanfrage, System-Nachfrage, Benutzer-Ergänzung, System-Präzisierungsfrage, System-Antwort, System-Abschied und System-Wiederholungsfrage. Der Dialogschritt System-Präzisierungsfrage ist noch weiter zu perfektionieren.

Um wissensbasierte Dialogsysteme an neue Domänen anzupassen ist ein erheblicher Aufwand notwendig, da anwendungsspezifisches Wissen in der Regel manuell in die Wissensbasis eingetragen werden muß. Solche anwendungsspezifischen „Restriktionen“ sind unbedingt notwendig um den Suchraum einzuschränken und eine linguistische Analyse in realistischer Zeit zu ermöglichen. Im Berichtserstattungsjahr wurden Verfahren zum Lernen eben solcher Restriktionen in semantischen Netzen untersucht und implementiert. In Abbildung 17 ist das Lernen von Restriktionen mit EVAR-Parallel schematisch dargestellt. Als Methoden wurden zu diesem Zwecke Maximum-Likelihood, Generalisierte Hidden Markov Modelle und ein Varianzverfahren eingesetzt. Für jedes Wort in der Lernstichprobe wird für jeden primitiven Knoten des Attribut-

flußgraphen eine Bewertung gelernt. Diese gibt an in welchem Maße ein bestimmtes Wort, wenn zur Instantiierung für ein bestimmtes Konzept verwendet, zum Erfolg der gesamten Analyse beiträgt. Gelernt und getestet wurde mit Hilfe einer vorhandenen gelabelten Stichprobe spontan gesprochener Äußerungen, welche am Lehrstuhl selbst, durch das am öffentlichen Telefonnetz angeschlossene automatische Intercity Auskunftssystem, aufgenommen wurden ([?], Kap. 7). Die gelernten Bewertungen, und dadurch auch der Ansatz selbst, wurden durch Experimente mit dem kompletten System (d.h., mit manuell eingetragenen Restriktionen) bestätigt. Es wurde z.B. eine Fehlerreduktion von 20% für das „finden“ von Ankunftsorten erreicht (Erhöhung der Erfolgsrate von 94.6% auf 95.6%). Experimente mit dem System ohne manuell eingetragene Restriktionen haben allerdings gezeigt, dass in diesem Bereich noch erhöhter Forschungsbedarf besteht.

Erste Ergebnisse bezüglich Frage–Antwort Dialogen zeigten einen Beschleunigungsfaktor von ca. 10 gegenüber dem A^* -basierten System. Die Zeit für eine Iteration beträgt momentan ca. 0,2 Sekunden, was bei fünf Iterationen einen Echtzeitfaktor zur Analyse einer Benutzer–Erstäußerung von ca. 0,7 ergibt. Dabei wurde eine parallele Suche mit bis zu fünf Prozessoren auf einem einzigen Prozessor simuliert.

Künftige Arbeiten beschäftigen sich weiterhin mit dem Lernen von Restriktionen in semantischen Netzen. Außerdem soll ein Systemprototyp mit PVM (Parallel Virtual Machine) zur parallelen Suche auf mehreren Prozessoren bereitgestellt werden.

Das VERBMOBIL–Projekt

In VERBMOBIL Phase 2 sollen spontansprachliche Äußerungen in den Domänen Terminabsprache, Reiseplanung und PC-Fernwartung in den Sprachen Deutsch, Englisch und Japanisch übersetzt werden. Im Gegensatz zu VERBMOBIL Phase 1 sollen sprachliche Eingaben nun frei erfolgen, d.h. ohne dass der Benutzer einen Knopf drücken muß, wenn er sprechen will.

Dies hat zur Folge, dass das Verbmobil-System selbst erkennen muß, wer gerade spricht und wo Anfang und Ende einer Äußerung sind. Außerdem muß das System die Eingabe auch segmentieren, falls eine sehr lange Äußerungen zu bearbeiten ist. Dies ist einerseits notwendig, um die Antwortzeit des Systems niedrig zu halten. Andererseits dient dies auch der Reduzierung der Komplexität: Eine kurze Äußerung ist z.B. weitaus einfacher syntaktisch zu parsen als eine sehr lange.

Um Segmentierungsfehler zu vermeiden, soll bei der Segmentierungsentscheidung möglichst viel Wissen eingesetzt werden. Hier ist eine Zusammenarbeit von Worterkennung und Prosodie geplant. Pausendetektion, Worterkennung und prosodische Bewertung sollen zu einem segmentierten Strom von Worthypothesengraphen führen, die von nachfolgenden Modulen weiterverarbeitet werden können.

Die Struktur des Prosodie-Moduls, das diese Segmentierung in Zusammenarbeit mit der Worterkennung durchführen soll, ist in Abbildung 18 dargestellt. Die Segmentierung muß erfolgen, bevor die Worterkennung mit höheren Wissensquellen das Worterkennungsergebnis erneut bewertet und dabei den Worthypothesengraphen ausdünnert, weil die Neubewertung und Ausdünnung erst erfolgen kann, wenn ein vollständiger Graph für ein Segment existiert.

Im Zusammenhang mit der Vorsegmentierung werden folgenden Fragen untersucht:

- Wie kann die Berechnung prosodischer Merkmale effizienter gestaltet werden, da die Segmentierung auf noch unausgedünnten durchgeführt werden muß (was einen sehr viel höheren Rechenaufwand bedeutet)?
- Kann auf Silben- und Silbenkernmerkmale verzichtet werden?
- Die Klassifikation der akustischen Merkmale erfolgt mit Neuronalen Netzen. Hier wird untersucht, welche Trainingsalgorithmen und Netztopologien am geeignetsten sind.
- Kann eine Segmentierung allein durch die Prosodie-Komponente durchgeführt werden oder ist eine Vorsegmentierung auf Basis einer Pause-Detektion notwendig?

Ein weiterer Schwerpunkt der Arbeit ist der Einsatz von Prosodie nicht nur für das Deutsche, sondern auch für Englisch und Japanisch. Hier ist von besonderem Interesse, eine Architektur zu entwickeln, die es erlaubt, ein multilinguales Modul (statt dreier monolingualer Module) für die Prosodie-Erkennung einzusetzen. Der Vorteil eines solchen Moduls besteht in einer leichteren Wartbarkeit und vor allem in einer höheren Effizienz.

In der zweiten Phase des VERBMOBIL Projektes werden Dialogakte zum einen als Transfer-Einheiten der gesprochenen Sprache zur flachen Übersetzung und zum anderen als elementare Einheiten zur Plan-Erkennung eines Dialoges eingesetzt. Um eine Segmentierung eines Dialoges in Dialogaktklassen vornehmen zu können, ist es erforderlich sinnvolle Dialogaktgrenzen von normalen Wort-grenzen zu trennen. Weiterhin müssen die Abschnitte zwischen zwei Dialogakt-grenzen einer der vorgegebenen Dialogaktklassen zugeordnet werden.

Im Zeitraum des vergangenen Jahres wurde deshalb an einem Verfahren gearbeitet, das automatisch aus einer Menge von vorgegebenen Trainings-Phrasen diejenigen „günstigen“ Phrasen auswählt, welche die Erkennungsrate erhöhen und somit mögliche Transkriptionsfehler und Sprecher-Abweichungen automatisch aus der Trainings-Menge entfernen [?]. Es entsteht somit ein konsistenter Klassifikator der zum Beispiel für die Klassifikation von Dialogakten verwendet werden kann.

Sind einer Äußerung-Hypothese ihre Dialogaktklassen zugeordnet, dann können diese mit der Hilfe von vordefinierten Mechanismen übersetzt werden. Weiterhin ist eine Verfolgung des Dialoges auf einer abstrakten Ebene, den Dialogaktklassen möglich, und die Planung des Dialoges kann beschrieben werden.

Ein weiterer Schwerpunkt innerhalb des VERBMOBIL Projektes bestand in der Verarbeitung und Erkennung von Emotion. Dafür wurde am Lehrstuhl ein Emotionsklassifikator entwickelt und in das Prodiemodul integriert. Bei der Klassifikation wird dabei nur zwischen Emotion und keiner Emotion entschieden, wobei hier Emotion als Ärger oder Wut aufgefaßt wird. Es werden vor allem keine unterschiedlichen Emotionen klassifiziert, etwa Ärger vs. Freude.

Emotion kann innerhalb gesprochener Sprache auf mehrere Arten markiert sein:

- durch Bedeutung des Wortes bei normaler Sprechweise, z.B. **so eine Drecksbox**
- durch Dehnung

- durch starke Veränderungen in der Energie und/oder F_0
- oder durch Kombinationen

Der am Lehrstuhl entwickelte Klassifikator verwendet dabei ausschließlich prosodische Merkmale wie z.B. F_0 -Verläufe oder die Dehnung von Wörtern und Silben. Reine lexikalische Information wie z.B. das Wissen, dass das Auftreten eines Schimpfwortes wie **DrecksKiste** auf eine Verärgerung des Benutzers schließen lässt wird momentan noch nicht verwendet. Als Klassifikatoren wurden Multi-Layer-Perzeptronen trainiert.

Bild 20 zeigt eine vereinfachte Arbeitsweise des VERBMOBIL-Prototypen mit integrierter Emotionserkennung.

Das Prosodiemodul erhält dabei als Eingabe ein Sprachsignal und einen **Worthypothesengraphen** (WHG), welcher durch eine Koppelung des Worterkenners und einer prosodischen Vorverarbeitung erzeugt wird. Im weiteren Verlauf bestimmt das Prosodiemodul innerhalb dieses WHG prosodische Phrasengrenzen, Akzente, Satzmodus und die Emotion. Wird auf Emotion entschieden, so wird die weitere Verarbeitung der Äußerung unterbrochen. Da im momentanen VERBMOBI Prototypen die Erkennung der Emotion nur demonstriert wird, folgt keine direkte Weiterverarbeitung der emotionalen Äußerung. Dem Benutzer wird lediglich angezeigt, dass eine emotionale Äußerung erkannt wurde. Bei einer echten Weiterverarbeitung der Emotion wäre die Einführung eines Klärungsdialoges seitens des Systems denkbar. Wird hingegen keine Emotion klassifiziert, so leitet das Prosodiemodul dann den prosodisch annotierten WHG an die entsprechenden Module weiter (Syntax, Semantik, Transfer, Dialog) und das System fährt mit der Verarbeitung der Daten fort.

Das SMARTKOM-Projekt

Im September 99 startete das BMBF-Leitprojekt *SmartKom*. In diesem Projekt werden Konzepte für die Entwicklung völlig neuartiger Formen der Mensch-Technik-Interaktion erprobt. Diese Konzepte werden die bestehenden Hemmschwellen von Computerlaien bei der Nutzung der Informationstechnologie abbauen und so einen Beitrag zur Nutzerfreundlichkeit und Nutzerzentrierung 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.

Als Modalitäten stehen dem Benutzer hierbei, neben der bisher im Dialog am häufigsten verwendeten Spracheingabe, auch ein Gestik und Mimik zur Verfügung.

Ein Skizze dieses Systems, wie es am Lehrstuhl für Mustererkennung aufgebaut wird, ist in Abbildung ?? zu sehen.

Die Gestik wird durch ein Grafiktablett registriert, das zur Erfassung von Biometriedaten (Unterschrift für die Personenverifikation) und Zeigegesten auf projizierten Daten dient. Die Mimik wird durch eine, auf den Kopf des Benutzers gerichtet, Kamera erfasst. Aus der Mimik soll Information über den emotionalen Zustand des Benutzers gewonnen werden, um dadurch

die Dialogführung entsprechend anzupassen. Bekanntlich erschweren emotionale Äußerungen die Worterkennung und beeinflussen somit den Dialogerfolg negativ. Die Blickrichtung, und somit das Objekt auf dem das Benutzerinterresse ruht, soll ebenfalls durch die Bilddaten der Mimikkamera bestimmt werden.

Sprechererkennung

In Zusammenarbeit mit den Firmen Dialog Communication Systems AG (DCS, Berlin/Tennenlohe) und MEDAV (Uttenreuth) begann zum 15. Dezember 1999 ein Projekt zur Sprechererkennung. Während sich die DCS AG mit Sprecheridentifikation und -verifikation im Hinblick auf die Weiterentwicklung des biometrischen Authentisierungssystems „BioID“ beschäftigt, stehen bei der Firma MEDAV forensische Anwendungen im Vordergrund. Anforderungen und geplante Schwerpunkte sind insbesondere Textunabhängigkeit, begrenzter Umfang an Trainings- und Testmaterial, Robustheit im Hinblick auf wechselnde und gestörte Kanäle, Landessprachenunabhängigkeit und Untersuchung von zeitlichen Veränderungen der Stimme.

6 Veranstaltungen

In Zusammenarbeit des Sonderforschungsbereichs 603 und des Graduiertenkollegs für Dreidimensionale Bildanalyse und -synthese wurde im Herbst des Jahres 1999 ein Workshop mit dem Titel VISION, MODELING, AND VISUALIZATION in Erlangen durchgeführt, der von der IEEE Signal Processing Society und der Gesellschaft für Informatik GI (FG 1.0.4, FG 4.1.2) unterstützt wurde. Ein international besetztes Publikum mit über 60 Teilnehmern aus 10 Ländern erfuhr in fünf eingeladenen und ca. 50 eingereichten Vorträgen, wie der Stand der Kunst in Visualisierung und Analyse in Wissenschaft und Industrie umgesetzt wird. Die Tagungsunterlagen wurden in einem Buch zusammengefasst [17].

Der Lehrstuhl war ebenfalls mit der Durchführung des Praktikums „Rechner lernen Hören und Sehen“ an der Veranstaltung Mädchen und Technik 1999 beteiligt. Eine Woche lang waren Schülerinnen der 08. bis 10. Jahrgangsstufe eingeladen, ein System zum Verstehen einfacher Kommandowörter wie „oben“ oder „rechts“ sowie ein System zur automatischen Erkennung von Komikfiguren aufzubauen. Mit Hilfe der Kommandowörter wurde eine Kamera so gesteuert, dass sie die im Raum verteilten Komikfiguren aufnimmt und erkennt. Das Praktikum stieß bei den Schülerinnen auf sehr positive Resonanz.

Liste der Veröffentlichungen

- [1] U. Ahlrichs, J. Fischer, J. Denzler, C. Drexler, H. Niemann, E. Nöth, D. Paulus: *Knowledge Based Image and Speech Analysis for Service Robots*, in *Proceedings Integration of Speech and Image Understanding*, IEE Computer Society, Korfu, Griechenland, 1999, S. 21–47.
- [2] U. Ahlrichs, J. Fischer, D. Paulus, H. Niemann: *Approach to Active Knowledge-Based Scene Exploration*, in M. Bramer, A. Macintosh, F. Coenen (Hrsg.): *Research and Deve-*

lopment in Intelligent Systems XVI – Proc. of the 19th SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence (ES99), BCS Conference Series, Springer, Cambridge, England, 1999, S. 289–301.

- [3] U. Ahlrichs, D. Paulus, H. Niemann: *Auswahl von Kameraaktionen zur wissensbasierten Szenenanalyse*, in W. Förstner, J. Buhmann, A. Faber, P. Faber (Hrsg.): *Mustererkennung 1999*, Springer, Heidelberg, September 1999, S. 44–53.
- [4] A. Batliner, J. Buckow, R. Huber, V. Warnke, E. Nöth, H. Niemann: *Prosodic Feature Evaluation: Brute Force or Well Designed?*, in *Proc. of the 14th Int. Congress of Phonetic Sciences*, Bd. 3, San Francisco, CA, Aug. 1999, S. 2315–2318.
- [5] A. Batliner, M. Nutt, V. Warnke, E. Nöth, J. Buckow, R. Huber, H. Niemann: *Automatic Annotation and Classification of Phrase Accents in Spontaneous Speech*, in *Proc. European Conf. on Speech Communication and Technology*, Bd. 1, Budapest, Ungarn, 1999, S. 519–522.
- [6] J. Buckow, V. Warnke, R. Huber, A. Batliner, E. Nöth, H. Niemann: *Fast and Robust Features for Prosodic Classification*, in V. Matoušek, P. Mautner, J. Ocelíková, P. Sojka (Hrsg.): *Proc. of the Workshop on TEXT, SPEECH and DIALOG (TSD'99)*, Bd. 1692 von *Lecture Notes for Artificial Intelligence*, Springer-Verlag, Berlin, September 1999, S. 193–198.
- [7] J. Buckow, V. Warnke, R. Huber, A. Batliner, E. Nöth, H. Niemann: *Multi-Lingual Prosodic Processing*, in *Proc. ESCA Workshop on Dialogue and Prosody*, Eindhoven, Niederlande, September 1999, S. 157–162.
- [8] F. Deinzer, J. Fischer, U. Ahlrichs, E. Nöth: *Learning of Domain Dependent Knowledge in Semantic Networks*, in *Proc. European Conf. on Speech Communication and Technology*, Bd. 5, Budapest, Ungarn, 1999, S. 1987–1990.
- [9] J. Denzler: *Radial Representation of Active Contours*, in R. Fischer (Hrsg.): *CVonline: The Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision*, 1999, <http://www.dai.ed.ac.uk/CVonline/>.
- [10] J. Denzler, H. Niemann: *Active Rays: Polar-transformed Active Contours for Real-Time Contour Tracking*, *Journal on Real-Time Imaging*, Bd. 5, Nr. 3, 1999, S. 203–213.
- [11] S. Di Bona, S. Huwer, H. Niemann, O. Salvetti: *The Brain Matcher*, in B. Radig, H. Niemann, Y. Zhuravlev, I. Gourevitch, I. Laptev (Hrsg.): *Pattern Recognition and Image Understanding*, (infix, Sankt Augustin, ISBN 3-89601-016-6), Herrsching, 1999, S. 219–226.
- [12] C. Drexler, C. Frank, J. Denzler, H. Niemann: *Probabilistisch modellierte Blicksteuerung zur Selbstlokalisierung anhand natürlicher Landmarken*, in G. Schmidt, U. Hanebeck, F. Freyberger (Hrsg.): *Autonome Mobile Systeme 1999*, Springer Verlag, 1999, S. 221–230.

- [13] F. Gallwitz, H. Niemann, E. Nöth: *Spracherkennung — Stand der Technik, Einsatzmöglichkeiten und Perspektiven*, *Wirtschaftsinformatik*, Bd. 41, Nr. 6, 1999, S. 538–547.
- [14] F. Gallwitz, H. Niemann, E. Nöth, V. Warnke: *Prosodic Information for Integrated Word- and-Boundary Recognition*, in *Proc. ESCA Workshop on Dialogue and Prosody*, Eindhoven, Niederlande, September 1999, S. 163–168.
- [15] A. Gebhard, D. Paulus: *Active System to Generate Views of Facial Features with Selectable Resolution*, in B. Girod, H. Niemann, H.-P. Seidel (Hrsg.): *Vision Modeling and Visualization 99*, Infix, Erlangen, Nov. 1999, S. 179–186.
- [16] A. Gebhard, D. Paulus, U. Ahlrichs: *Wissensbasierte Gesichtsmodellierung und Kamera-steuerung zur Analyse von Patientengesichtern*, in H. Evers, G. Glombitza, T. Lehmann, H.-P. Meinzer (Hrsg.): *3. Workshop Bildverarbeitung für die Medizin*, Springer, 1999, S. 117–121.
- [17] B. Girod, H. Niemann, H.-P. Seidel (Hrsg.): *Vision, Modeling, and Visualization '99*, infix (ISBN 3-89601-015-8), Sankt Augustin, 1999.
- [18] J. Haas, V. Warnke, H. Niemann, M. Cettolo, A. Corazza, D. Falavigna, G. Lazzari: *Semantic Boundaries in Multiple Languages*, in *Proc. European Conf. on Speech Communication and Technology*, Bd. 1, Budapest, Ungarn, 1999, S. 535–538.
- [19] S. Harbeck, U. Ohler: *Multigrams for Language Identification*, in *Proc. European Conf. on Speech Communication and Technology*, Budapest, Ungarn, 1999, S. 375–378.
- [20] S. Harbeck, U. Ohler, E. Nöth, H. Niemann: *Information Theoretic Based Segments for Language Identification*, in V. Matoušek, P. Mautner, J. Ocelíková, P. Sojka (Hrsg.): *Proc. of the Workshop on TEXT, SPEECH and DIALOG (TSD'99)*, Bd. 1692 von *Lecture Notes for Artificial Intelligence*, Springer-Verlag, Berlin, September 1999, S. 187–202.
- [21] B. Heigl, R. Koch, M. Pollefeys, J. Denzler, L. V. Gool: *Plenoptic Modeling and Rendering from Image Sequences Taken by a Hand-Held Camera*, in W. Förstner, J. Buhmann, A. Faber, P. Faber (Hrsg.): *Mustererkennung 1999*, Springer, Heidelberg, September 1999, S. 94–101.
- [22] B. Heigl, H. Niemann: *Camera Calibration from Extended Image Sequences for Lightfield Reconstruction*, in *Workshop Vision, Modeling and Visualization*, Erlangen, Nov. 1999, S. 43–50.
- [23] K. Horecki, D. Paulus, K. Wojciechowski: *Object Localization Using Color Histograms*, in K.-H. Franke (Hrsg.): *5. Workshop Farbbildverarbeitung*, Schriftenreihe des Zentrums für Bild- und Signalverarbeitung e.V. Ilmenau, Ilmenau, 1999, S. 59–66.
- [24] J. Hornegger, D. Paulus, H. Niemann: *Probabilistic Modeling in Computer Vision*, in B. Jähne, P. Geißler, H. Haußecker (Hrsg.): *Handbook of Computer Vision and Applications*, Bd. 2, Academic Press, San Diego, 1999, S. 817–854.

- [25] J. Fischer and E. Nöth and H. Niemann: *Combining Statistics with Semantic Networks in a Real-Time Dialogue System*, in *Proc. International Workshop Speech and Computer (SPECOM'99)*, Moskau, Rußland, 1999, S. 75–78.
- [26] R. Koch, B. Heigl, M. Pollefeys, L. V. Gool, H. Niemann: *A Geometric Approach to Lightfield Calibration*, in F. Solina, A. Leonardis (Hrsg.): *Computer Analysis of Images and Patterns — CAIP '99*, Nr. 1689 in *Lecture Notes in Computer Science*, Springer, Heidelberg, 1999, S. 596–603.
- [27] R. Koch, M. Pollefeys, B. Heigl, L. V. Gool, H. Niemann: *Calibration of Hand-held Camera Sequences for Plenoptic Modeling*, in *Proceedings of the 7th International Conference on Computer Vision (ICCV)*, IEEE Computer Society Press, Korfu, Griechenland, September 1999, S. 585–591.
- [28] T. Merz, F. Elandaloussi, D. Paulus, W. Osten: *Active approach for holographic non-destructive testing of satellite fuel tanks*, in M. Kujawinska, W. Osten (Hrsg.): *Optical Measurement Systems for Industrial Inspection*, Bd. 3824, SPIE, 1999, S. 8–19.
- [29] P. Mössner, M. Wolf, P. Weierich, H. Niemann: *A Control Structure for an Observation System for Service Robots*, in B. Radig, H. Niemann, Y. Zhuravlev, I. Gourevitch, I. Laptev (Hrsg.): *Pattern Recognition and Image Understanding*, (infix, Sankt Augustin, ISBN 3-89601-016-6), Herrsching, 1999, S. 211–218.
- [30] H. Niemann: *Knowledge-Based Interpretation of Images*, in B. Jähne, H. Haußecker, P. Geißler (Hrsg.): *Handbook of Computer Vision and Applications*, Vol. 2, Academic Press, San Diego, 1999, S. 855–874.
- [31] E. Nöth, A. Batliner, A. Kießling, R. Kompe, H. Niemann: *Suprasegmental Modelling*, in K. Ponting (Hrsg.): *Computational Models of Speech Pattern Processing*, Bd. 169 von *NATO ASI Series F*, Springer-Verlag, Berlin, 1999, S. 182–199.
- [32] 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*, in *Proc. ESCA Workshop on Dialogue and Prosody*, Eindhoven, Niederlande, September 1999, S. 25–34.
- [33] E. Nöth, F. Gallwitz, M. Aretoulaki, M. Boros, J. Haas, S. Harbeck, R. Huber, H. Niemann: *Research Issues for the Next Generation Spoken Dialogue Systems*, in V. Matoušek, P. Mautner, J. Ocelíková, P. Sojka (Hrsg.): *Proc. of the Workshop on TEXT, SPEECH and DIALOG (TSD'99)*, Bd. 1692 von *Lecture Notes for Artificial Intelligence*, Springer-Verlag, Berlin, September 1999, S. 1–9.
- [34] E. Nöth, J. Haas: *Sprachverarbeitung jenseits der Spracherkennung*, in F. Lehner, G. Braungart, L. Hitzenberger (Hrsg.): *Multimedia — Informationssysteme zwischen Bild und Sprache*, Gabler Edition Wissenschaft, Deutscher Universitäts Verlag, Wiesbaden, 1999, S. 117–125.

- [35] E. Nöth, J. Haas, V. Warnke, F. Gallwitz, M. Boros: *A Hybrid Approach to Spoken Dialogue Understanding: Prosody, Statistics and Partial Parsing*, in *Proc. European Conf. on Speech Communication and Technology*, Bd. 5, Budapest, Ungarn, 1999, S. 2019–2022.
- [36] E. Nöth, S. Harbeck, H. Niemann: *Multilingual Speech Recognition*, in K. Ponting (Hrsg.): *Computational Models of Speech Pattern Processing*, Bd. 169 von *NATO ASI Series F*, Springer–Verlag, Berlin, 1999, S. 363–375.
- [37] U. Ohler, S. Harbeck, H. Niemann: *Discriminative estimation of language model classifiers*, in *Proc. European Conf. on Speech Communication and Technology*, Budapest, Ungarn, 1999.
- [38] U. Ohler, S. Harbeck, H. Niemann, E. Nöth, M. G. Reese: *Interpolated Markov chains for eukaryotic promoter recognition*, *Bioinformatics*, Bd. 15, Nr. 5, 1999, S. 362–369.
- [39] I. Pál, H. Niemann, G. Michelson: *Experimente mit mehrschichtigen Perzeptron–Netzen zur Vorverarbeitung und Merkmalgewinnung auf den SLDF–Perfusionsbildern der Netzhaut*, in E. H., G. Glombitza, T. Lehmann, H.-P. Meinzer (Hrsg.): *Bildverarbeitung für die Medizin 1999*, (Springer, Berlin, Heidelberg, ISBN 3-540-65627-8), Heidelberg, 1999, S. 122–126.
- [40] D. Paulus, U. Ahlrichs, B. Heigl, J. Denzler, J. Hornegger, H. Niemann: *Active Knowledge Based Scene Analysis*, in H. Christensen (Hrsg.): *Computer Vision Systems*, Springer, Heidelberg, Jan. 1999, S. 180–199.
- [41] D. Paulus, J. Hornegger, H. Hiemann: *Software Engineering for Image Processing and Analysis*, in B. Jähne, P. Geißler, H. Haußecker (Hrsg.): *Handbook of Computer Vision and Applications*, Bd. 3, Academic Press, San Diego, 1999, S. 77–103.
- [42] D. Paulus, M. Wolf, S. Meller, H. Niemann: *3D Computer Vision for Tooth Reconstruction, Medical Image Analysis*, Bd. 3, Nr. 11, 1999, S. 1–19.
- [43] B. Radig, H. Niemann, Y. Zhuravlev, I. Gourevitch, I. Laptev (Hrsg.): *Pattern Recognition and Image Understanding — 5th Open German–Russian Workshop, Herrsching*, infix (ISBN 3-89601-016-6), Sankt Augustin, 1999.
- [44] V. Warnke, F. Gallwitz, A. Batliner, J. Buckow, R. Huber, E. Nöth, A. Höthker: *Integrating Multiple Knowledge Sources for Word Hypotheses Graph Interpretation*, in *Proc. European Conf. on Speech Communication and Technology*, Bd. 1, Budapest, Ungarn, 1999, S. 235–238.
- [45] V. Warnke, S. Harbeck, E. Nöth, H. Niemann, M. Levit: *Discriminative Estimation of Interpolation Parameters for Language Model Classifiers*, in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Bd. 1, IEEE Computer Society Press, Phoenix, Arizona, März 1999, S. 525 – 528.

- [46] V. Warnke, E. Nöth, H. Niemann, G. Stemmer: *A Segment Based Approach for Prosodic Boundary Detection*, in V. Matoušek, P. Mautner, J. Ocelíková, P. Sojka (Hrsg.): *Proc. of the Workshop on TEXT, SPEECH and DIALOG (TSD'99)*, Bd. 1692 von *Lecture Notes for Artificial Intelligence*, Springer-Verlag, Berlin, September 1999, S. 199–202.
- [47] C. Yuan, H. Niemann: *Object Localization in 2D Images Based on Kohonen's Self-Organization Feature Maps*, in *1999 International Joint Conference on Neural Networks*, IEEE, 1999, S. 158.
- [48] M. Zobel, J. Denzler, H. Niemann: *Coupling Rays – Probabilistic Modeling of Spatial Dependencies*, in *International Conference on Imaging Science, Systems, and Technology (CISST'99)*, Las Vegas, Nevada, 1999, S. 416–422.
- [49] M. Zobel, J. Denzler, H. Niemann: *Tracking of Probabilistically Coupled Features*, in B. Girod, H. Niemann, H.-P. Seidel (Hrsg.): *Vision Modeling and Visualization 99*, Infix, Erlangen, Nov. 1999, S. 133–140.

7 Diplomarbeiten

- B. Blank: Detektion, Verfolgung und Analyse des Mundes in (Farb-)Bildfolgen zur Beurteilung von Gesichtslähmungen, März 1999
- C. M. Frank: Vergleich probabilistischer Ansätze zur passiven und aktiven Selbstlokalisierung, März 1999
- H. Hertlein: Textunabhängige Sprecheridentifikation zur Zugangskontrolle, September 1999
- A. Horndasch: Design, Implementierung und Evaluierung eines Sprachdialog-Servers, Dezember 1999
- A. Höthker: Vom Wortgraphen zum Interpretationsgraphen - Zusammenführung unterschiedlicher Wissensquellen, 1999.
- M. Scheuring: Semiautomatische Erzeugung von Transferfunktionen für Volume Rendering, Oktober 1999
- R. Schug: Probabilistische Personenlokalisierung und -eskortierung mittels einer autonomen mobilen Plattform, Dezember 1999
- G. Stemmer: Segmentmodelle zur Promotorendetektion und Landessprachenidentifikation, April 1999
- T. Vogel: Automatische Identifikation von Hirnarealen in prä- und intraoperativen MR-Aufnahmen des Gehirns, Mai 1999

8 Vorträge

Ahlrichs, U., Wissensbasierte aktive Szenenexploration, Fachgespräch: Modellbasierte Bildanalyse, FGAN/FOM, Karlsruhe, 29.09.1999

Ahlrichs, U., An Approach to Active Knowledge-Based Scene Exploration, 19th SGES International Conference on Knowledge-Based Systems and Applied Artificial Intelligence, Cambridge, England, 15.12.1999

Batliner, A.: Querschnittproblem: Segmentierung, Verbmobil Projektlenkungssitzung, Stuttgart, 12.05.99

Buckow, J.: Neue wortbasierte prosodische Merkmale, VERBMOBIL Akustik-Workshop, Dresden, 25.05.1999

Buckow, J.: Fast and Robust Features for Prosodic Classification, Marienbad, Tschechien, 13.09.1999

Buckow, J.: Multi-lingual Prosodic Processing, Veldhoven, Niederlande, 02.09.1999

Denzler, J.: Aktives Sehen zur Echtzeitobjektverfolgung, Vision Group, Fraunhofer Gesellschaft, Erlangen, 22.03.99

Denzler, J.: Coupling Rays — Probabilistic Modeling and Tracking of Multi-part Objects, Seminar des SFB 527, Universität Ulm, 14.05.99

Denzler, J.: Aktives Sehen für die industrielle Bildverarbeitung, Fachunterausschuss VDE/VDI „Anforderungen an industrielle Bildverarbeitungssysteme“, Kaiserslautern, 11.06.99

Denzler, J.: Knowledge Based Image and Speech Analysis for Service Robots, ICCV Workshop on Integration of Speech and Image Analysis, Korfu, Griechenland, 21.09.99

Denzler, J.: Probabilistic Sequence Modeling with Application to visual ego-localization, Bayesian Group Seminar, Brain and Cognitive Science Department, University of Rochester, Rochester, USA, 19.11.99

Denzler, J.: Combining Computer Graphics and Computer Vision: Visual Self-Localization by Particle Filters, Vision and Robotics Seminar, Computer Science Department, University of Rochester, Rochester, USA, 01.12.99

Denzler, J.: Optimierungsansatz für die Integration von Kamerabildern bei der Klassifikation, SFB 603 Statusseminar, Pommersfelden, 07.05.99

Denzler, J.: Analyse, Codierung und Verarbeitung von Lichtfeldern zur Gewinnung realistischer Modelldaten, SFB 603 Statusseminar, Pommersfelden, 07.05.99

Fischer, J.: Combining Statistics with Semantic Networks in a Real-Time Dialogue System, SPECOM'99, Moskau, Russland, 05.10.1999

Gallwitz, F.: Speech Recognition and Prosody, Workshop on Recent Advances in Speech Recognition, Atami, Japan, 06.02.1999

Gallwitz, F.: The Spoken Dialogue System EVAR, Advanced Telecommunication Research Institute International (ATR), Kyoto, Japan, 10.02.1999

Gallwitz, F.: Speech Recognition and Prosody, Advanced Telecommunication Research Institute International (ATR), Kyoto, Japan, 10.02.1999

Gallwitz, F.: Prosodic Information for Integrated Word-and-Boundary Recognition, ESCA Workshop on Dialogue and Prosody, Veldhoven, Niederlande, 02.09.99

Huber, R.: Emotion in VERBMOBIL, VERBMOBIL Prosodie-Workshop, Erlangen,

25.03.1999

Huber, R.: VERBMOBIL – Demonstration, Tag der Informatik, Erlangen, 30.04.1999

Huber, R.: Emotionserkennung – Aktueller Stand, VERBMOBIL Akustik-Workshop, München, 25.11.1999

Gebhard, A.: Wissensbasierte Gesichtsmodellierung und Kamerasteuerung zur Analyse von Patientengesichtern, Bildverarbeitung für die Medizin, Heidelberg, 05.03.1999

Gebhard, A.: Erstes System zur Diagnoseunterstützung von Gesichtsparesen, SFB 603–Arbeitstagung, Pommersfelden, 06.05.1999

Gebhard, A.: Active System to Generate Views of Facial Features with Selectable Resolution, Vision Modelling and Visualization '99, Erlangen, 18.11.1999

Heigl, B.: Camera Calibration from Extended Image Sequences for Lightfield Reconstruction, VMV-Workshop, Erlangen, 17.11.1999

Heigl, B.: Analyse, Codierung und Verarbeitung von Lichtfeldern, Arbeitstagung des SFB 603, Pommersfelden, 06.05.1999

Heigl, B.: Optimierungsprobleme im Teilprojekt C2, Arbeitskreis Optimierung, Erlangen-Tennenlohe, 01.06.1999

Merz, T.: Active approach for holographic non-destructive testing of satellite fuel tanks, SPIE International Symposia on Industrial Lasers & Inspection, München, 16.06.99

Niemann, H.: Statistical Approach to Intensity Based Object Recognition. Dagstuhl Seminar 99491 on Content-Based Image Retrieval, Schloss Dagstuhl, Wadern, 08.12.1999

Niemann, H.: Analyse, Kodierung und Verarbeitung von Lichtfeldern, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen (IWR), Fakultät für Physik und Astronomie, Universität Heidelberg, 22.06.1999

Nöth, E., Acoustic and Information Theoretic Based Sub Word Units for Language Identification, DGA Workshop, Paris, Frankreich, 19.02.1999

Nöth, E., Prosodische Information: Begriffsbestimmung und Nutzen für das Sprachverstehen, Universität Potsdam, 08.03.1999

Nöth, E., Spracherkennung, Universität Potsdam, 08.03.1999

Nöth, E., Acoustic and Information Theoretic Based Sub Word Units for Language Identification, IST/TG-01 Workshop, Boulder, USA, 12.03.1999

Nöth, E., Using Prosodic Information in Speech Understanding Systems, AT&T, Florham Park, USA, 13.04.1999

Nöth, E., Acoustic Features for Prosodic Information, AT&T, Florham Park, USA, 15.04.1999

Nöth, E., Prosodic (?) Classes, AT&T, Florham Park, USA, 20.04.1999

Nöth, E., Prosody in Automatic Speech Understanding: Who are its Partners, AT&T, Florham Park, USA, 22.04.1999

Nöth, E., Using Prosodic Information in Speech Understanding Systems, IBM, Yorktown Heights, USA, 26.04.1999

Nöth, E., Acoustic and Information Theoretic Based Sub Word Units for Language Identification, Lincoln Lab, Lexington, USA, 29.04.1999

Nöth, E., The Erlangen Spoken Dialogue System EVAR: A State-of-the-Art Information Retrieval System, MIT, Boston, USA, 30.04.1999

- Nöth, E., Die Erkennungsphase in einem sprachverstehenden System, Arbeitsgruppenseminar der HNO-Klinik, Uni Erlangen-Nürnberg, 07.07.1999
- Nöth, E., The Erlangen Spoken Dialogue System EVAR: A State-of-the-Art Information Retrieval System, IBM, Beijing, China, 19.08.1999
- Nöth, E., On the Use of Prosody in Automatic Dialogue Understanding, ESCA Workshop on Dialogue and Prosody, Eindhoven, Niederlande, 03.09.1999
- Nöth, E., Research Issues for the Next Generation Spoken Dialogue Systems, TSD '99, Mariánské Lázně, Tschechische Republik, 13.09.1999
- Nöth, E., Information Theoretic Based Segments for Language Identification, TSD '99, Mariánské Lázně, Tschechische Republik, 13.09.1999
- Nöth, E., On the Use of Prosody in Automatic Dialogue Understanding, AIIA Workshop, Trento, 17.12.1999
- Ohler, U.: Interpolierte und diskriminative Markovketten zur funktionellen Annotation von DNA-Sequenzen, Oberseminar der LFE Theoretische Informatik, LMU München, 12.02.1999
- Ohler, U.: Stochastic Segment Models to Identify Promoter Regions in the Drosophila Genome, Drosophila Genome Project, Berkeley, 17.12.1999
- Paulus, D.: Active Knowledge Based Exploration, ICVS '99, Las Palmas de Gran Canaria, 14.1.99
- Paulus, D.: Verfolgung farbiger Punkte mit Anwendungen zur Szenenexploration, Heidelberg, Colloquium der DFG Forschergruppe Bewegungsanalyse, 02.02.99
- Paulus, D.: Selected Applications of Modelling, SFB 603 und GK 3-D Bildanalyse, Ringvorlesung, Erlangen, 13.07.99
- Paulus, D.: Object localization using color histograms, 5. Workshop Farbbildverarbeitung, Ilmenau, 7.10.99
- Paulus, D.: Histogrammvergleich zur Erkennung farbiger Objekte in realen Umgebungen, Vision Interest Group, FhG Erlangen, 23.11.99
- Paulus, D.: Aktives Bildverstehen, TU Berlin, Informatik Kolloquium, 07.12.1999
- Warnke, V.: Discriminative Estimation of Interpolation Parameters for Language Model Classifiers, ICASSP 99, Phoenix, 18.03.1999
- Warnke, V.: Integrating Multiple Knowledge Sources for Word Hypotheses Graph Interpretation, EUROSPEECH 99, Budapest, 06.09.1999
- Warnke, V.: A Segment Based Approach for Prosodic Boundary Detection, TSD 99, Marienbad, 16.09.1999
- Yuan, C.: Object Localization in 2D Images Based on Kohonen's Self-Organization Feature Maps, 1999 International Joint Conference on Neural Networks, Washington DC, USA, 12.07.1999
- Zobel, M.: Fortschrittskriterien im Teilprojekt B2, Arbeitskreis Qualitätskriterien des SFB 603, Erlangen, 16.03.1999
- Zobel, M.: Optimierungsbasierte Echtzeit-Verfolgung von Gesichtern und ihren Bestandteilen, Arbeitskreis Optimierung des SFB 603, Erlangen, 04.05.1999
- Zobel, M.: Optimierungsbasierte Objektklassifikation und Objektlokalisation mittels Ansichtenplanung, Seminar des SFB 603, Pommersfelden, 06.05.1999
- Zobel, M.: Coupling Rays – Probabilistic Modeling of Spatial Dependencies, International

Conference on Imaging Science, Systems, and Technology, Las Vegas, 28.06.1999

Zobel, M.: Tracking of Probabilistically Coupled Features, Vision, Modeling, and Visualization '99, Erlangen, 17.11.1999

9 Anmerkungen zum Dokument / About this document

Dieses Dokument wurde von den Mitgliedern des Lehrstuhls für Mustererkennung in \LaTeX erstellt.

This document was created by the members of the *Chair for pattern recognition* in \LaTeX .

Es wurde von Dietrich Paulus und Elmar Nöth zusammengestellt und editiert.

Dietrich Paulus and Elmar Nöth are the editors.

Unter Verwendung von `tex4ht` wurde es nach `html` convertiert.

It was converted to `html` with `tex4ht`.

5. Mai 2000

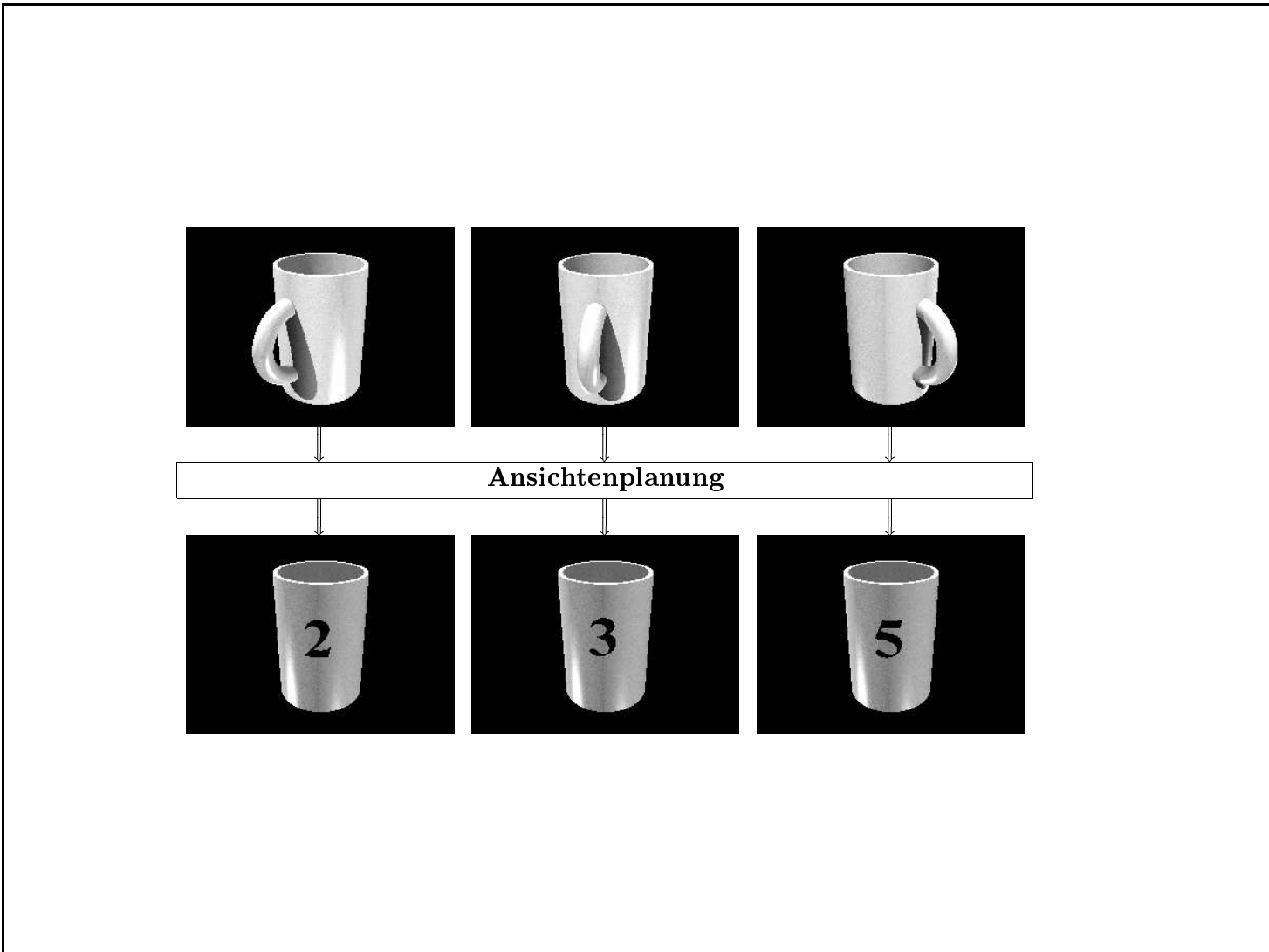


Bild 14: Beispiel für Ambiguitäten zwischen Objekten und deren Auflösung mittels Ansichtenplanung. Die Tassen sind nur unterscheidbar, wenn die Ziffer auf der Vorderseite sichtbar ist.

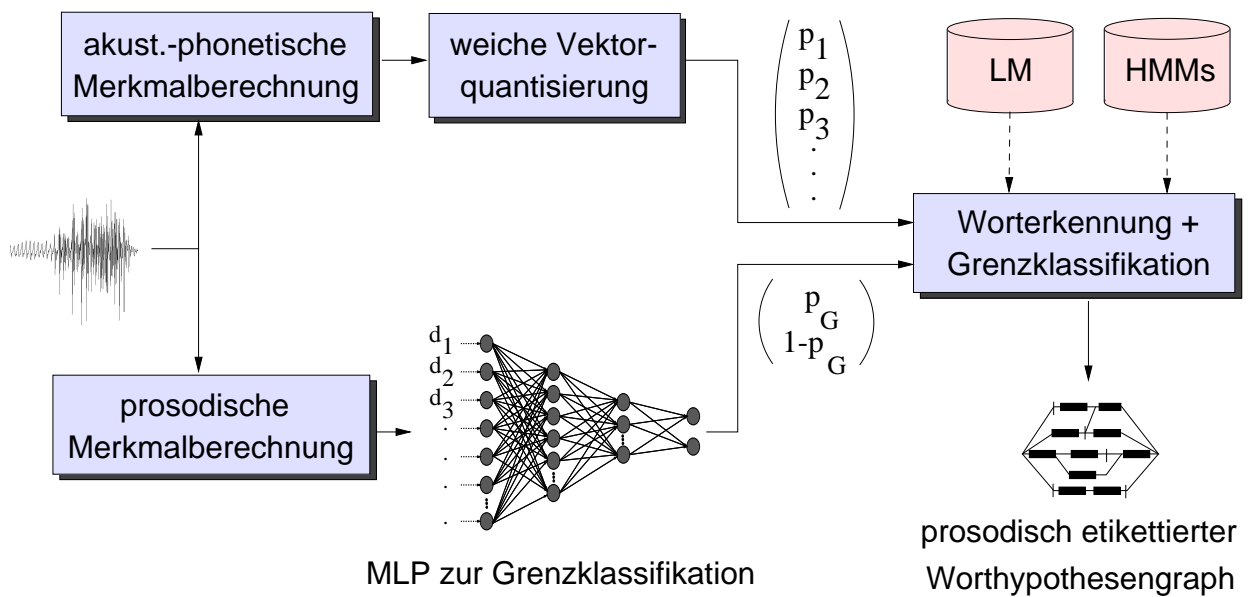


Bild 15: Integrierte Verarbeitung lautlicher und prosodischer Information

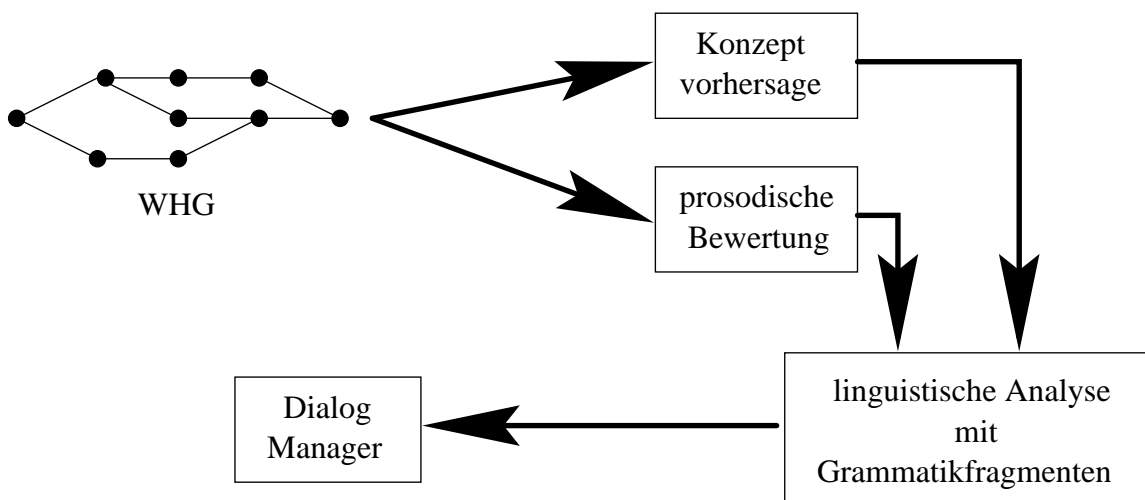


Bild 16: Architektur des Moduls zur linguistischen Analyse mit statistischer Konzeptvorhersage, prosodischer Bewertung und Grammatikfragmenten

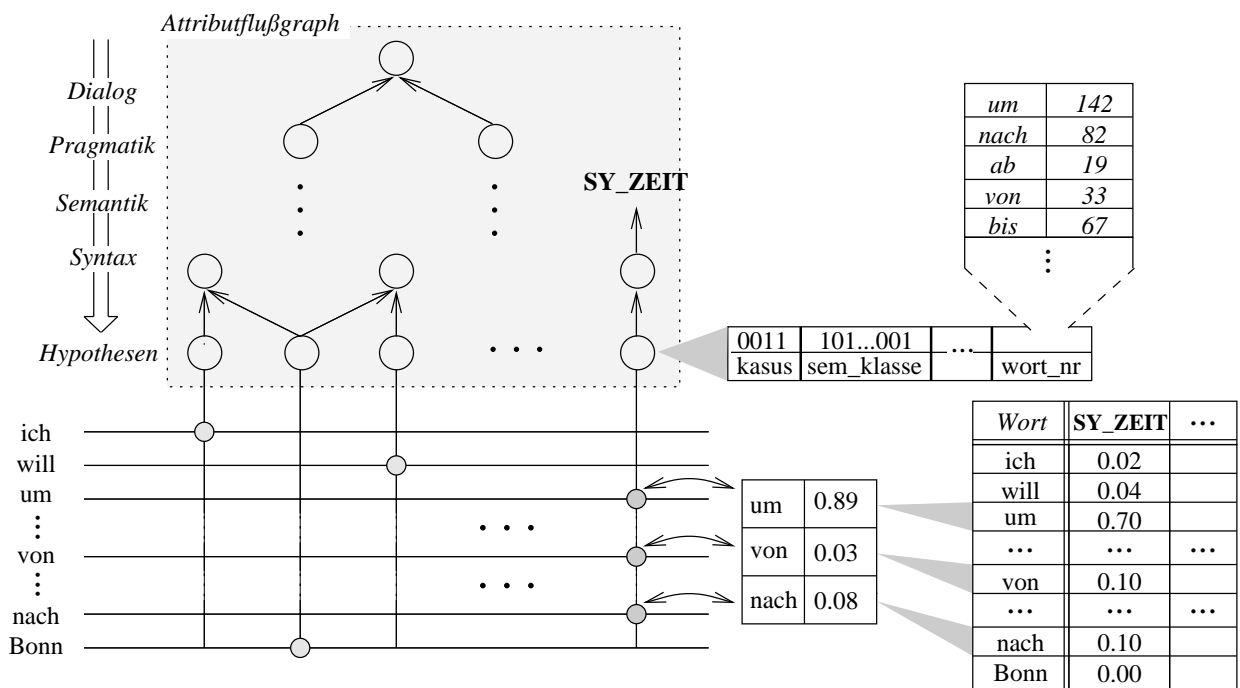


Bild 17: Lernen von Restriktionen mit EVAR-Parallel.

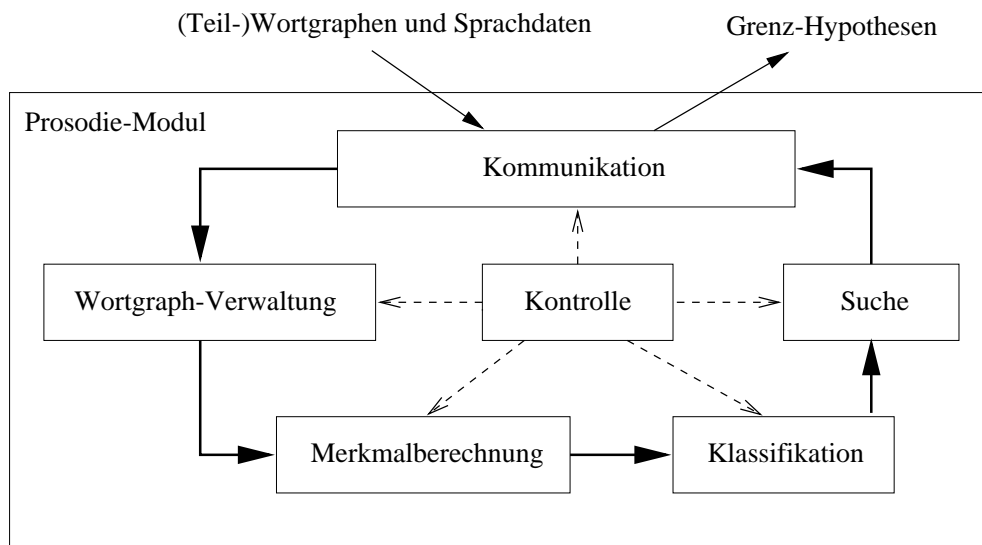


Bild 18: Modulstruktur des Vorsegmentierungsmoduls

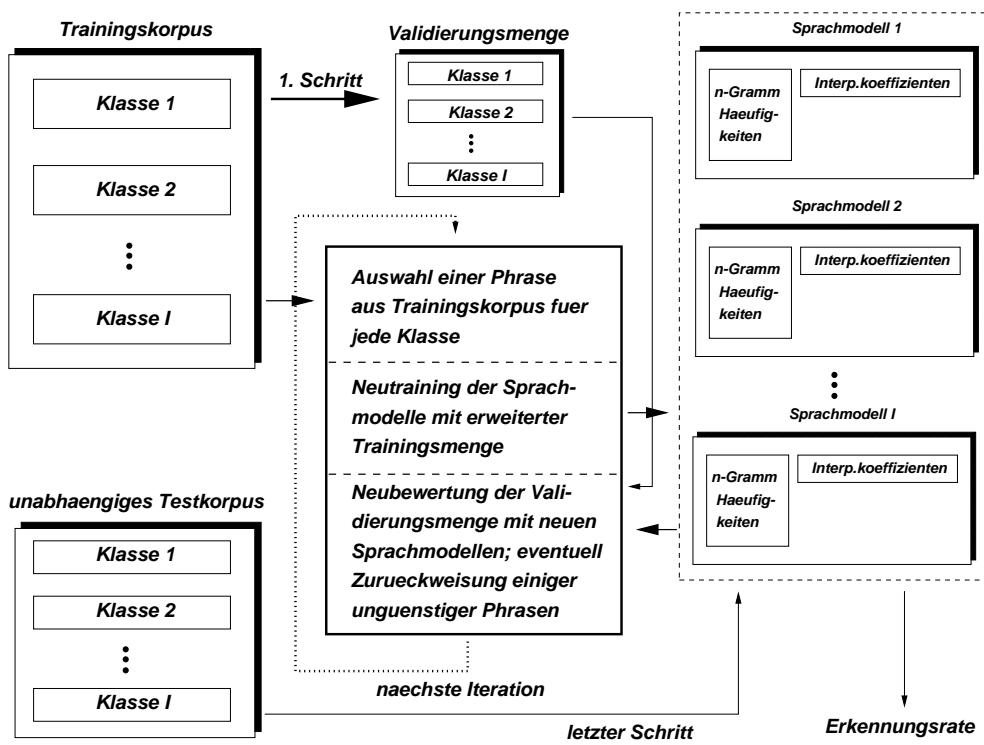


Bild 19: Ein Verfahren zur Auswahl günstiger Trainings-Phrasen

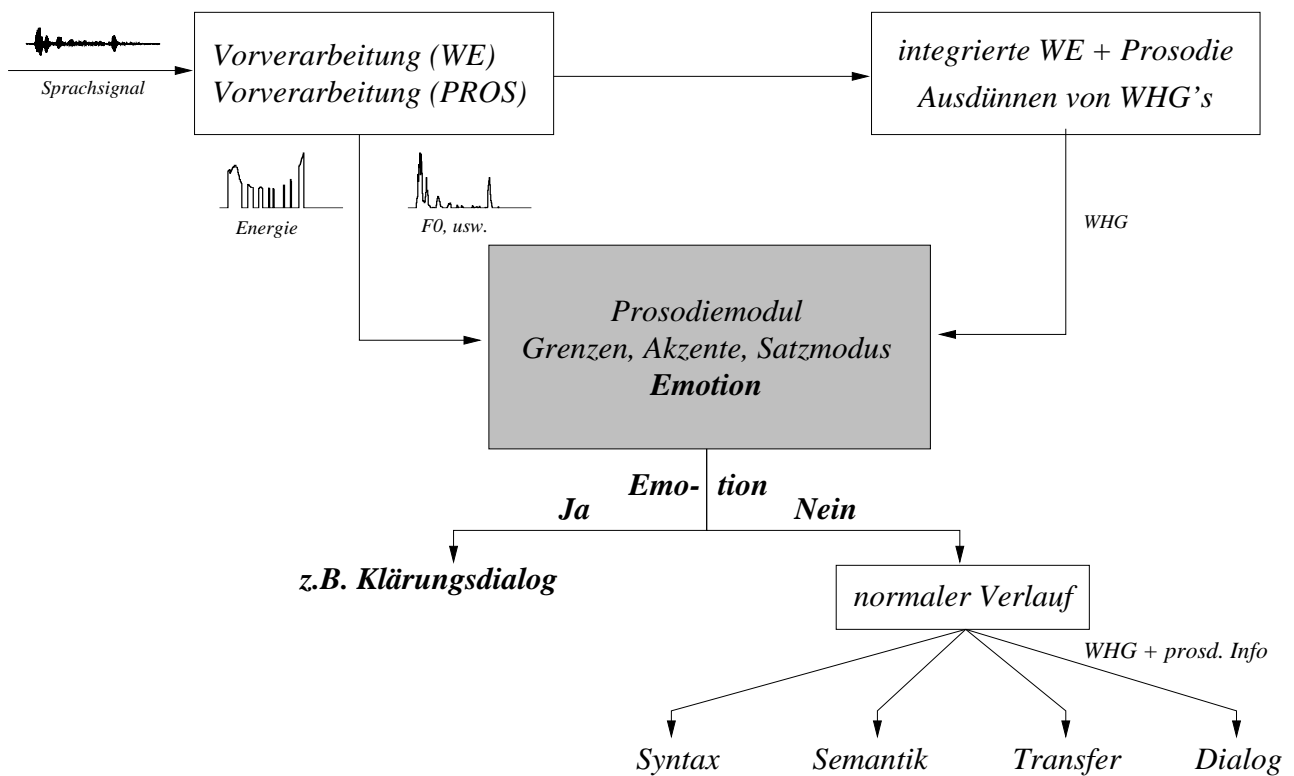


Bild 20: Arbeitsweise des Prototypen bei eingeschalteter Emotionserkennung.