

## **Lehrstuhl für Informatik 5 (Mustererkennung)**

**Anschrift:** Martensstrasse 3, 91058 Erlangen

**Tel.:** +49.9131.85.27775      **Fax.:** +49.9131.303811

**E-Mail:** info@i5.informatik.uni-erlangen.de

### **Leitung:**

Prof. Dr.-Ing. Heinrich Niemann

### **Professor:**

Prof. Dr.-Ing. Joachim Hornegger

### **Sekretariat:**

Iris Koppe

Kristina Müller

Silvia Zettelmaier

### **Sekretariat SFB 603:**

Martina Montel-Kandy

### **Leitung Bildverarbeitung:**

Dipl.-Inf. Jochen Schmidt

### **Bildverarbeitung:**

Ing. Radim Chrastek

Dipl.-Ing. Christian Derichs

Dipl.-Inf. Benjamin Deutsch

Dipl.-Inf. Christoph Gräßl

Dipl.-Inf. Marcin Grzegorzek

Kailash N. Pasumathy, M.S.

Dipl.-Inf. Ingo Scholz

Dipl.-Inf. Florian Vogt

Dipl.-Math. (FH) Stefan Wenhardt

Dipl.-Inf. Timo Zinßer

### **Medizinische Bildverarbeitung:**

Dipl.-Inf. Volker Daum

Soufyane El Allali, M. Sc.

Jingfeng Han, M. Sc.

Dipl.-Ing. (FH) Stefan Hoppe

Dipl.-Inf. Florian Jäger  
Dipl.-Inf. Marcus Prümmer  
Dipl.-Inf. Holger Scherl

**Leitung Sprachverarbeitung:**

Dr.-Ing. Elmar Nöth

**Sprachverarbeitung:**

Dr. phil. Anton Batliner  
Dipl.-Math. Silvia Dennerlein  
Dipl.-Inf. Christian Hacker  
Dipl.-Inf. Tino Haderlein  
Dipl.-Inf. Axel Horndasch  
Dipl.-Inf. Michael Levit  
Dipl.-Inf. Stefan Steidl  
Dipl.-Inf. Viktor Zeissler

**Lehrbeauftragte:**

Dr.-Ing. Thomas Wittenberg

**Nichtwiss. Personal:**

Walter Fentze  
Iris Koppe  
Martina Montel-Kandy  
Kristina Müller  
Friedrich Popp

## **1 Einführung**

Der Lehrstuhl für Informatik 5 (Mustererkennung) ist Teil des Instituts für Informatik der Friedrich-Alexander-Universität Erlangen-Nürnberg. Seit 1975, dem Gründungsjahr des Lehrstuhls, ist Prof. Dr. H. Niemann Lehrstuhlinhaber des LME.

Am Lehrstuhl wird das Problem der “Mustererkennung” untersucht, wobei ganz allgemein die automatische Transformation einer von einem geeigneten Sensor gelieferten Folge von Abtastwerten eines Signals in eine den Anforderungen der Anwendung entsprechende symbolische Beschreibung gesucht wird. In der Bildverarbeitung werden hierfür Sensoren eingesetzt, die unter Umständen vom Rechner gesteuert werden können oder mit spezieller Beleuchtung gekoppelt sind. Sie liefern Informationen in einem oder mehreren Kanälen. Bei der Verarbeitung von zusammenhängend gesprochenen 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 (symbo-

lischen) Wissensverarbeitung. Die Ermittlung einer symbolischen Beschreibung wird auch als Analyse des Musters bezeichnet.

Der Lehrstuhl bearbeitet hauptsächlich zwei Themenkomplexe, nämlich die wissensbasierte Analyse von Bildern und Bildströmen sowie das Verstehen gesprochener Sprache und die Generierung einer Antwort. In der wissensbasierten Bildanalyse werden sowohl grundsätzliche Arbeiten zur Bildverarbeitung und zur Repräsentation und Nutzung problem-spezifischen Wissens als auch spezielle Arbeiten zur Entwicklung eines vollständigen, rückgekoppelten Systems für die schritthaltende Analyse dreidimensionaler Szenen durchgeführt. Eine Brücke zwischen Visualisierung und Analyse wird im Sonderforschungsbereich 603 (<http://sfb-603.uni-erlangen.de>) mit dem Thema "Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten" hergestellt, dessen Sprecher Prof. Niemann ist.

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.

Ein Problem, das in jedem der 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.

## 1.1 Forschungsschwerpunkte

- Bildverarbeitung
- 3D Objekterkennung
- Objektverfolgung
- Aktive Sensordatenverarbeitung
- 3D Rekonstruktion und Kamerakalibrierung
- Plenoptische Modellierung
- Erweiterte Realität
- Autonome, mobile Systeme
- Mimik- und Gestik
- Sprachverarbeitung
- Prosodie
- Dialog
- Benutzerzustandserkennung (von Ärger über Müdigkeit bis Zögern)

## 1.2 Forschungsrelevante apparative Ausstattung

- Mobiles System MOBSY mit aktivem Stereokamera-Kopf <http://www5.informatik.uni-erlangen.de/~mobsy/>
- Drehteller und Schwenkarm zur Bildaufnahme

- Smartkom-Zelle <http://www5.informatik.uni-erlangen.de/SmartKom/page.html>
- Head-Mounted Display mit integriertem Stereokamera-System
- Pan-Tilt Einheit

### **1.3 Kooperationsbeziehungen**

- University of Rochester, Computer Science Department <http://www.cs.rochester.edu>
- University of California, San Diego, Cognitive Science Department <http://cogsci.ucsd.edu/cogsci/>
- Universität Bielefeld I Applied Computer Science <http://www.techfak.uni-bielefeld.de/ags/ai/>
- Universität Bielefeld II Neuroinformatics <http://www.techfak.uni-bielefeld.de/ags/ni/>
- TU Graz <http://www.TUGraz.at/>
- University of Surrey <http://www.ee.surrey.ac.uk/CVSSP/>
- University of Mannheim, Computer Vision, Graphics and Pattern Recognition <http://www.cvgpr.uni-mannheim.de/>

### **1.4 Wissenschaftliche Tagungen**

- 7. Workshop Farbbildverarbeitung 2001
- Vision, Modeling, and Visualization 2002 <http://www9.informatik.uni-erlangen.de/vmv2002>

### **1.5 Veröffentlichungsreihen**

Die Veröffentlichungen des Lehrstuhls befinden sich auf der lehrstuhleigenen Homepage unter <http://www5.informatik.uni-erlangen.de/literature/German/Literature.html>

## 2 Forschungsprojekte

### 2.1 Aktive 3-D Rekonstruktion und Kamerakalibrierung

**Beteiligte:**

Dipl.-Inf. Jochen Schmidt, Dipl.-Math. (FH) Stefan Wenhardt

**Beginn:** 1.1.2004

Die 3-D Rekonstruktion befasst sich mit der Rekonstruktion der 3-D Information einer Szene. Die 3-D Information wird unter anderem für weitere Anwendungen, wie z. B. das bildbasierte Rendering oder zur Erzeugung von erweiterter Realität, benötigt. Dabei werden die Informationen für die Rekonstruktion aus gewöhnlichen Kamerabildern gewonnen - das Wort "aktiv" bedeutet nicht, dass es sich hierbei um aktive Sensoren, wie z. B. Lasersensoren oder strukturiertes Licht, handelt. Vielmehr werden die Kameras aktiv gesteuert, d. h. für die 3-D Rekonstruktion werden die Parameterkonfigurationen der Kameras gesucht, die den zu erwartenden Fehler der Rekonstruktion minimieren.

Der Einfluss der verschiedenen Kameraparameter (Translation, Rotation und Brennweite) wurde dabei an Stereoaufnahmen evaluiert. Die 3-D Rekonstruktion zur Bestimmung der 3-D Koordinaten eines Punktes in der Szene erfolgt dabei durch Triangulation der Sichtstrahlen. Dazu benötigt man die Lage der Kameras im Raum und die Bildkoordinaten des zu rekonstruierenden Punktes im linken und rechten Bild. Um die Bildkoordinaten dieses Punktes zu bestimmen muss man das sogenannte Korrespondenzproblem lösen, d. h. man sucht in beiden Bildern Punkte, die von demselben 3-D Punkt in der Szene stammen. Jedoch ist das Suchen der Korrespondenzen ein Problem, das nicht in vollster Allgemeinheit gelöst ist und deswegen können hier Fehler auftreten. Diese Fehler wurden modelliert und auf Grundlage dieses Modells die Bewertung der Kameraparameter durchgeführt. Dabei wurde der Abstand zwischen realem und trianguliertem Punkt in Abhängigkeit der Kameraparameter berechnet und anschließend minimiert. Dieses Optimierungsproblem wurde durch Monte Carlo Simulationen berechnet, und die gefundenen Lösungen wurden durch reale Experimente verifiziert.

Es wurde über die Kameraparameter Translation, Rotation und Brennweite einzeln und in beliebigen Kombinationen optimiert. Dies ist zum einen deswegen wichtig, da für die Positionierung der Kameras ein Roboterarm verwendet wird und dieser natürlich nur einen eingeschränkten Aktionsbereich hat. Unter Umständen können dann bestimmte Parameter nicht weiter optimiert werden, da der Roboter sie nicht weiter verändern kann. Zum anderen ist auch zu bewerten, wie stark der Einfluss der einzelnen Parameter ist. Denn die Parameter können nicht unabhängig voneinander beliebig verändert werden. So ist z. B. eine große Stereobasis im Allgemeinen gut für die Triangulation. Andererseits verbessert auch eine große Brennweite das Rekonstruktionsergebnis. Es kann aber nicht gleichzeitig die Brennweite und die Stereobasis immer weiter vergrößert werden, da man dann sehr schnell das zu rekonstruierende Objekt aus dem Bild verliert. Deswegen muss eine Konfiguration gefunden werden, die den Einfluss der veränderbaren Parameter in ein ausgewogenes Verhältnis stellt. So konnte mit Hilfe der optimierten Konfiguration der Triangulationsfehler erheblich reduziert werden.

Bei der Verwendung einer auf einem Roboterarm montierten Kamera ist es notwendig, eine sogenannte Hand-Auge Kalibrierung durchzuführen. Diese berechnet die unbekannte starre Transformation von Positionsdaten, die vom Roboter geliefert werden, auf die Position und Orientierung der Kamera. Hierfür werden mit Hilfe eines Kalibrieremusters zunächst zusammenge-

hörende Positionen von Kamera und Roboterarm ermittelt, die anschließend zur Hand-Auge Kalibrierung verwendet werden können. Um bei der Kalibrierung gute Ergebnisse zu erzielen, ist ein zusätzlicher Datenauswahlschritt notwendig. Im letzten Jahr wurden mehrere Datenselektionsalgorithmen entwickelt, wobei der erfolgreichste auf einer Vektorquantisierung der Rotationsachsen der Positionsdaten basiert.

Diese robuste Hand-Auge Kalibrierung kann nicht nur zur Kalibrierung einer auf einem Roboterarm montierten Kamera verwendet werden, sondern auch im Bereich der Selbstkalibrierung von Stereokamerasystemen. Im Teilprojekt B6 des SFB 603 wurde sie zudem erfolgreich für die Kalibrierung eines optischen Trackingsystems für Endoskope eingesetzt.

## **2.2 Erscheinungsbasierte, statistische Objekterkennung**

### **Projektleitung:**

Prof. Dr.-Ing. Heinrich Niemann

### **Beteiligte:**

Dipl.-Inf. Marcin Grzegorzek

### **Stichwörter:**

Objekterkennung; Objektlokalisierung; Objektklassifikation; statistische Modellierung; Wavelets; Farbmodellierung, Kontextmodellierung

**Laufzeit:** 1.6.1999 - 30.11.2005

**Förderer:** DFG

### **Kontakt:**

Dipl.-Inf. Marcin Grzegorzek

Tel.: +49.9131.85.27275, Fax: +49.9131.303811,

E-Mail: grzegorz@informatik.uni-erlangen.de

Ziel dieses Projekts, das im Rahmen des Graduiertenkollegs "Dreidimensionale Bildanalyse und -synthese" durchgeführt wird, ist die Klassifikation und Lokalisierung von Objekten in Bildern. Dazu wird ein erscheinungsbasierter Ansatz eingesetzt. Dementsprechend wird kein vorhergehender Segmentierungsprozess benötigt, der geometrische Merkmale wie Kanten oder Ecken detektiert, sondern die Merkmale werden direkt aus den Intensitäten der Bildpunkte bestimmt. Bei diesem Ansatz werden lokale Merkmale benutzt, die mit Hilfe der Wavelet Multi-Skalen-Analyse berechnet werden. Modelliert werden diese Merkmale statistisch mit Hilfe von Dichtefunktionen. So können auch Beleuchtungsschwankungen und Rauschen gehandhabt werden.

Bei realen Anwendungen sind die Objekte oft partiell verdeckt und können sich vor beliebigem Hintergrund befinden. Deshalb wird auch der Hintergrund explizit modelliert und in der Erkennungsphase für jeden Merkmalsvektor entschieden, ob er zum Objekt oder zum Hintergrund gehört.

Auch mehrere Objekte können im Bild auftreten, die lokalisiert und klassifiziert werden sollten. Im Allgemeinen ist die Anzahl der Objekte, die sich im Bild befinden, nicht bekannt. Ein Ab-

bruchkriterium entscheidet, wann die Suche nach Objekten beendet werden soll. Die Auswahl eines richtigen Abbruchkriteriums ist deshalb von großer Bedeutung.

Um Objektklassen zu trainieren, werden mehrere Bilder von Objekten in bekannten Lagen und Orientierungen gebraucht. Üblicherweise werden die Objekte auf einem Drehteller gestellt, um einen bekannten Winkel gedreht und von einer Kamera, montiert auf einem Roboterarm, aufgenommen. Ein neuer Ansatz wurde entwickelt, der dieses Verfahren vereinfacht. Die Objekte werden mit einer handgeführten Kamera aufgenommen, und ihre Lagen und Orientierungen mit dem "Structure-from-Motion" Algorithmus rekonstruiert.

Zur Evaluierung des Objekterkennungssystems entstand eine neue umfangreiche Stichprobe 3D-REAL-ENV. Bei über 30000 Trainingsbildern und 8000 Testbildern mit realem heterogenem Hintergrund kann man verschiedene Erkennungsalgorithmen ganz objektiv vergleichen. Da die Beleuchtung in den Testbildern anders als in der Trainingsphase ist, konnte auch die Beleuchtungsunabhängigkeit des Systems nachgewiesen werden.

In der Zukunft wird geplant, die Farbinformation der Objekte, die momentan vernachlässigt wird, zu verwenden. Außerdem wird man im Falle der Mehrobjektszenen die gegebenen oder trainierten Kontextzusammenhänge zwischen den Objektklassen ausnutzen.

## **2.3 HUMAINE (Human-Machine-Interaction Network on Emotion)**

### **Projektleitung:**

Dr.-Ing. Elmar Nöth

### **Beteiligte:**

Dr. phil. Anton Batliner, Dipl.-Inf. Stefan Steidl, Dipl.-Inf. Christian Hacker,  
Dipl.-Math Silvia Dennerlein

### **Stichwörter:**

Emotionserkennung; Mensch-Maschine-Interaktion

**Laufzeit:** 1.1.2004 - 31.12.2007

**Förderer:** EU

### **Mitwirkende Institutionen:**

<http://emotion-research.net/Members/>

### **Kontakt:**

Dr.-Ing. Elmar Nöth

Tel.: +49.9131.85.27888, Fax: +49.9131.303811,

E-Mail: [noeth@informatik.uni-erlangen.de](mailto:noeth@informatik.uni-erlangen.de)

HUMAINE (Human-Machine-Interaction Network on Emotion) ist ein NoE (Network of Excellence) im europäischen Rahmenprogramm FP6 und läuft vom 1.1.2004 bis zum 31.12.2007; Vertrags-Nr. 507422. HUMAINE will in der Europäischen Union die Grundlagen schaffen für die Entwicklung von Systemen, die menschliche emotionale oder emotions-ähnliche Zustände und Prozesse erkennen, modellieren und/oder beeinflussen können - sog. "emotionale Systeme".

me". Solche Systeme sind grundlegend für die zukünftige Mensch-Maschine-Kommunikation; allerdings sind ihre Grundlagen noch nicht genügend erforscht. Einer der Gründe dafür ist sicherlich, dass das einschlägige Wissen auf unterschiedliche Disziplinen verteilt ist (Psychologie, Künstliche Intelligenz, Robotik, Spracherkennung, Bildverarbeitung, usw.). In HUMAINE kooperieren Experten aus all diesen Disziplinen. Sechs unterschiedliche thematische Gebiete werden bearbeitet: Emotionstheorie, Signale und Zeichen, emotionale Interaktion, Emotion in Kognition und Aktion, Emotion in Kommunikation und Überzeugung sowie Benutzerfreundlichkeit emotionaler Systeme. Gemeinsame Arbeitsprojekte werden auf Workshops sowie in Plenartreffen abgesprochen; darunter fällt auch die notwendige Infrastruktur (Korpora, die kulturelle und geschlechtsspezifische Unterschiede widerspiegeln, ein ethisches Rahmenwerk sowie ein Web-Portal). Eines der Ziele ist darüber hinaus das Verfassen eines Handbuchs sowie Empfehlungen zur Methodologie.

Die FAU Erlangen-Nürnberg ist Partner in WP4 (Signale und Zeichen) sowie in WP5 (Korpora). Im ersten Jahr konzentrierten sich die Aktivitäten auf das initiale Plenartreffen, eine Sommerschule über Korpora sowie Workshops über Theorie, Signale und Zeichen, und Korpora. Gegen Ende 2004 werden sog. "Exemplars" genauer definiert und beschrieben. Darunter ist ein beispielhaftes Vorgehen zu verstehen, mit dem gezeigt wird, auf welche Weise ein Problem gelöst werden kann, und wie eine interdisziplinäre Herangehensweise auszusehen hat. Die spezielle Form eines solchen Exemplars ist nicht vorgegeben; es kann sich dabei um einen Demonstrator handeln, um ein Korpus, um ein experimentelles Design oder um spezifische Methoden.

## 2.4 Preparing Future Multisensorial Interaction Research (PF-Star)

### **Projektleitung:**

Dr.-Ing. Elmar Nöth, Akad. ORat

### **Beteiligte:**

Dr. phil. Anton Batliner, Dipl.-Inf. Stefan Steidl, Dipl.-Inf. Christian Hacker, Dipl.-Inf. Tino Haderlein, Dipl.-Inf. Viktor Zeissler, Dipl.-Inf. Michael Levit, Dipl.-Math Silvia Dennerlein

### **Stichwörter:**

Spracherkennung; Emotionserkennung; Prosodie; Kindersprache; Bewertung der Aussprache

**Laufzeit:** 1.10.2002 - 30.9.2004

**Förderer:** EU, 5th Framework Programme

### **Mitwirkende Institutionen:**

ITC-irst, Italy, RWTH Aachen, Germany, Universität Karlsruhe, Germany, Kungl Tekniska Högskolan, Sweden, University of Birmingham, UK, Istituto di Scienze e Tecnologia della Cognizione, Sezione di Padova, Italy

Das EU-Projekt PF-Star hatte zum Ziel, technologische Grundlagen, vergleichende Evaluationen und Erfolgsabschätzungen für die Entwicklung verschiedener Schlüsseltechnologien bereitzustellen, die essentiell für die Entwicklung multilingualer und multisensorischer Mensch-Mensch- und Mensch-Maschine-Kommunikationssysteme sind. Das Projekt dauerte zwei Jahre



und wurde Ende September 2004 erfolgreich abgeschlossen. Neben Synthese von emotionaler Sprache und emotionalen Gesichtern sowie maschineller Übersetzung umfasste das Projekt die zwei Arbeitspakete *Erkennung von Emotionen (WP3)* und *Kindersprache (WP5)*, an denen der *Lehrstuhl für Mustererkennung (Informatik 5)* als Leiter bzw. Partner teilnahm. Während im ersten Jahr von PF-Star hauptsächlich neue Korpora aufgezeichnet worden waren und alle Ergebnisse mit existierenden Korpora erzielt worden waren, wurden nun im zweiten Jahr die neuen Korpora verwendet. Das Korpus, das für beide Arbeitspakete, WP3 und WP5, aufgezeichnet worden war, ist das AIBO-Korpus. Es umfasst 9 Stunden Sprache von Kindern, die mit dem AIBO-Roboter spielen (spontane, emotionale Sprache). An der Universität Birmingham wurden vergleichbare Aufnahmen durchgeführt.

Für das Arbeitspaket *Erkennung von Emotionen (WP3)* wurden bis März 2004 zu jedem gesprochenen Wort von fünf Studenten prosodische Eigenheiten und der emotionale Benutzerzustand annotiert. Als dritte Annotation werden zur Zeit die Äußerungen der Kinder mit dem Verhalten des AIBO abgestimmt. Im zweiten Teil des Berichtszeitraumes wurden die englischen Aufnahmen der Uni Birmingham annotiert (emotionaler Benutzerzustand). Beim Labeln der Daten galt es einem zentralen Problem in der Emotionsforschung gerecht zu werden: Ein Klassifikator kann nämlich nur bestenfalls das erkennen, was annotiert ist, und nicht die tatsächlichen Phänomene. Deshalb wurde ein entropiebasiertes Maß entwickelt, um zu zeigen, inwiefern die Erkennung des Klassifikators mit der menschlichen Erkennung übereinstimmt. Es konnte gezeigt werden, dass unser Ansatz, von mehreren Nicht-Experten annotieren zu lassen und anschließend per Mehrheitsentscheidung ein Label zu erhalten, eine gute Alternative zu einem solchen Ansatz ist, in dem Experten-Labeler zuvor trainiert werden.

Um automatisch emotionale Wörter des AIBO-Korpus klassifizieren zu können, wurden verschiedene Merkmalssätze verwendet: prosodische, spektrale und POS-Merkmale (part-of-speech, z.B. "Substantiv"). In verschiedenen Experimenten wurden Neuronale Netze oder LDA-Klassifikatoren trainiert und als Referenz die Mehrheitsentscheidung der Labeler zugrundegelegt. Beim 2-Klassen-Problem "positiv+neutral" versus "negativ" wurde eine klassenweise gemittelte Erkennungsrate (CL) von 76% erzielt. Für sieben Klassen (erfreut, bemutternd, neutral, emphatisch, energisch, verärgert, tadelnd) werden 45% CL erzielt.

Als weiteres Korpus wurde SYMPAFLY - emotionale Sprache von Sprechern, die ein automatisches Flugbuchungssystem anrufen - hauptsächlich im ersten Jahr von PF-Star verwendet. Im Berichtszeitraum wurden alternative Merkmale, wie z.B. MFCC-basierte und HNR-basierte (Harmonicity-to-Noise-Ratio) Merkmale, untersucht. Ferner wurden alle Merkmale sowohl auf Wort- als auch auf Äußerungsebene berechnet. Für das 2-Klassen-Problem konnten bis zu 75% CL erzielt werden, für vier Klassen 57%.

Im Teilgebiet *Kindersprache (WP5)* wurden die neuen Korpora untersucht, wobei Baseline-Erkennungsraten und einige Maße von altersspezifischen Eigenschaften im Vordergrund standen. Am *Lehrstuhl für Mustererkennung* waren vier neue Korpora mit insgesamt etwa 23 Stunden Sprache neu aufgenommen worden: Das AIBO-Korpus umfasst 9 Stunden Spontansprache und OHM8000 (ein Korpus mit einem großen Vokabular von 8000 Wörtern) 9 Stunden gelesener Sprache. Ferner wurden deutsche Kinder, die englische Texte lesen, aufgezeichnet. Vergleichbare Phrasen wurden auch von der *Universität Birmingham* und dem italienischen Forschungsinstitut *ITC-irst* aufgenommen. Das vierte Korpus stellt eine kleine Sammlung von Aufnahmen von Kindern mit Sprachbehinderung dar.

Für all diese Korpora wurden Baseline-Erkenner mit Hilfe verschiedener Trainingsverfahren und basierend auf verschiedenen Merkmalssätzen entwickelt. Zudem wurden Sprechervariabilitäten und die Lage der Formanten im Spektrum untersucht: Intra-Sprecher-Varianzen sind für

junge Kinder größer, Formanten werden in höhere Frequenzbereiche verschoben. Auf den AIBO-Daten konnten höhere Erkennungsraten für Wörter beobachtet werden, die als “verärgert” oder “emphatisch” markiert sind, wohingegen “bemutternd” schlechter erkannt wurde. Erkennungsraten für gelesene Sprache konnten mit Hilfe verschiedener Sprachmodelle verbessert werden. Eine Herausforderung bei nicht-nativer Kindersprache stellt die automatische Bewertung der Aussprache dar. Dazu wurde ein 30-dimensionaler Merkmalsatz entwickelt. Falsch ausgesprochene Wörter konnten so mit 72% Erkennungsrate detektiert werden. Als Anwendungen sind Programme zum computergestützten Erlernen einer Fremdsprache denkbar; für zukünftige Untersuchungen liegen nun auch Referenzbewertungen von mehreren Englischlehrern vor. Das Korpus mit Kindern mit Sprachbehinderung umfasst Aufnahmen von LKG-Kindern (Lippen-, Kiefer-, Gaumenspalte). Der *Lehrstuhl für Mustererkennung* entwickelt ein objektives Maß, um Kinder während einer Sprachtherapie zu bewerten. Erste automatische Messungen des Nasalitätsgrades korrelieren mit den menschlichen Bewertungen zu 0.66. Bei der Analyse wird der Anteil für die Nasalierung typischer Frequenzen in Bezug zur gesamten Sprachenergie gesetzt.

In weiteren Untersuchungen mit dem AIBO-Korpus wurde versucht verhallte und verrauschte Sprache robuster zu erkennen. Dazu wurden TRAP-basierte Merkmale (TempoRAI Pattern) implementiert.

## **2.5 Probabilistisches Szenenmodel**

### **Projektleitung:**

Prof. Dr.-Ing. Heinrich Niemann

### **Beteiligte:**

Kailash N. Pasumarthy, M.S.

### **Stichwörter:**

Bayesnetze; Objekterkennung; probabilistische Szenenmodellierung

**Laufzeit:** 1.11.1999 - 31.12.2005

**Förderer:** DFG

### **Kontakt:**

Kailash N. Pasumarthy, M.S.

Tel.: +49.9131.85.27275, Fax: +49.9131.303811,

E-Mail: kailash.pasumarthy@informatik.uni-erlangen.de

“Scene Exploration Using Bayesian and Neural Networks”, is a project carried out as a part of “3D Image Analysis and Synthesis” at the Graduate Research Center of Chair of Pattern Recognition and is supported by German Research Foundation. The present project may be classified into one of the advanced fields of image processing and finds its application where challenge of the machine perception in complex scenes and work environments is sought.

Exploring scenes using Bayesian nets (BNs) is based on the idea of performing an active knowledge based search on images, unlike conventional visual recognition algorithms. During the indirect search of images, a sample set of training images from different classes is available right

at the onset of an experiment and the nature of the class to be searched is unknown. Usually a recursive search for objects in an image, belonging to all classes is performed using a conventional object recognition system and the Bayesian approach, the goal of the present research work, can obviate this. The search of objects in an image by BNs can be confined only to a specific class or a set of classes. Our initial results have proved that if structural relationships are rightly established between the constituent objects of an image, searching scenes using BNs is quite effective. However the BN structure and the parameters are manually specified in our initial experiments.

Encouraged by the initial results, obtained by manual specification of structure and parameters to the BNs, presently we are applying Gaussian Mixture and QMR Models for object recognition. By applying these models for object recognition obviates the manual specification of parameters and BN Structure. EM Algorithm is employed to compute the A Posteriors for generic object recognition.

## 2.6 Rutschungserkennung mit unterschiedlichen Klassifikationssystemen

### **Projektleitung:**

Maik Hamberger

### **Beteiligte:**

Prof. Dr. Klaus Moser, Dipl.-Inf. Michael Levit, Dr.-Ing. Elmar Nöth,  
Rickli Christian, Hess Josef, Dr.-Ing. Rainer Deventer

### **Stichwörter:**

Rutschungen; Soil Slips; Starkniederschläge; Gefahrenerkennung; Klassifikationsverfahren; LDA; Support Vector Machines

**Laufzeit:** 1.7.2000 - 30.6.2005

### **Mitwirkende Institutionen:**

Lehrstuhl für Angewandte Geologie Erlangen, Eidgenössischen Forschungsanstalt WSL, Amt für Wald und Landschaft des Kanton Obwalden

### **Kontakt:**

Dr.-Ing. Elmar Nöth, Akad. ORat  
Tel.: +49.9131.85.27888, Fax: +49.9131.303811,  
E-Mail: noeth@informatik.uni-erlangen.de

In Zusammenarbeit mit dem Lehrstuhl für Angewandte Geologie wurde in diesem Jahr Forschung im Bereich der automatischen Hangrutschungserkennung betrieben. Die Aufgabe besteht darin, potentiell gefährdete Bereiche von alpinen Hängen zu erkennen, die durch morphologische, geologische, geotechnische und weitere Parameter charakterisiert werden. Diese Parameter sind für das System zugänglich.

Mit Hilfe einer Vielzahl von evaluierten Rutschungen ist es möglich, Zusammenhänge zwischen rutschungsrelevanten Parametern und dem Gefahrenpotential zu ermitteln, die in der Zukunft zur Vorhersage von Rutschungsbereichen herangezogen werden können. Dabei wird das Untersuchungsgebiet in kleine quadratische Zellen (5 x 5 m) aufgeteilt, von welchen jede durch einen Vektor von gemessenen Parametern repräsentiert wird. Insbesondere werden folgende Eigenschaften berücksichtigt: Hangneigung, Hangorientierung, geologische Verhältnisse, Vegetation, Niederschlagsmenge usw. Insgesamt stehen ca. 25K gekennzeichnete Zellen für die Experimente zur Verfügung (die Daten wurden im Rahmen des groß angelegten Projektes "Vegetationswirkungen und Rutschungen" der Eidgenössischen Forschungsanstalt WSL, in zwei Diplomarbeiten des Lehrstuhls für Angewandte Geologie der Universität Erlangen-Nürnberg und bei Geländebegehungen in den Sachler Bergen/Schweiz gesammelt). Nur etwa 5K sind als Rutschungsbereiche markiert, der Rest hat keine Rutschungen vorzuweisen.

Die Vorhersage wird dadurch bewerkstelligt, dass man einen statistischen Klassifikator trainiert, welcher von einem Parametervektor ausgehend auf einen möglichen Erdrutsch in dem betroffenen Quadrat schließen soll. Hierbei ist zu beachten, dass viele der eingesetzten Parameter Wertebereiche haben, die nicht metrisch sind (z.B. Vegetationsart) und von somit mit dem "1 aus N"-Verfahren in Gruppen von binären Subparametern umkodiert werden müssen. Auf diese Weise kommt man in unseren Basisexperimenten auf Parametervektoren, die etwa 60 Elemente

enthalten.

Mit Leave-One-Out Verfahren (LOO) wurden aus den vorhandenen Daten mehrere Trainings- und Testmengen erzeugt, um die Aussagekraft der nachfolgend beschriebenen Experimente zu erhöhen. Zwei Arten von Klassifikatoren wurden zum Vergleich herangezogen: Lineare-Diskriminanzanalyse-Klassifikator (LDA) und Support-Vektor Maschinen. Während die LDA-Klassifikation im Allgemeinen wesentlich schneller ablief (maximal 30 Minuten pro Experiment), lieferten die SVMs in der Regel bessere Klassifikations- bzw. Vorhersageraten.

Eine weitere Verbesserungsmöglichkeit bestand darin, den unmittelbaren Kontext der betroffenen Quadrate als eine zusätzliche Informationsquelle zu verwenden. Im einfachsten Fall nimmt man die lokalen Parametervektoren der Nachbarbereiche mit in den endgültigen Merkmalvektor, der den jeweiligen Kleinbereich beschreibt, wobei auch abgeleitete Größen (wie z.B. Mittelwert, Varianz usw.) verwendet werden können. Sowohl die optimale Nachbarschaftsgeometrie, als auch der optimale Satz von abgeleiteten Größen müssen dabei herausgefunden werden.

In unseren Experimenten wurden die besten Ergebnisse dadurch erzielt, dass man den LDA-Klassifikator die besten 960 Klassifikationsmerkmale (aus den insgesamt mehr als 3200 in Frage kommenden) auswählen ließ, wonach SVMs für die eigentliche Klassifikation eingesetzt wurden. Bei dem Zweiklassenproblem Rutschung/Nichtrutschung konnte man so Klassifikationsraten von über 82% erreichen.

Als nächstes Ziel wird zu diesem Zeitpunkt das Dreiklassenproblem angesehen, wo man bei den Rutschungen auch zwischen Kanten (Rutschungsursprüngen) und Flächen (betroffenen Gebieten) unterscheidet.

## 2.7 SFB 396, C1: Sensor und modellgestützte Optimierung von Prozessketten

### Projektleitung:

Prof. Dr.-Ing. Heinrich Niemann, Prof. Dr.-Ing. Joachim Denzler

### Beteiligte:

Dr.-Ing. Rainer Deventer

### Stichwörter:

Mustererkennung, maschinelles Lernen, wahrscheinlichkeitstheoretische Modell, Bayesnetze

**Laufzeit:** 1.1.1996 - 31.12.2004

**Förderer:** Deutsche Forschungsgemeinschaft

### Kontakt:

Dr.-Ing. Rainer Deventer

Tel.: 09131.85.27825, Fax: 09131.303811,

E-Mail: deventer@informatik.uni-erlangen.de

In der Industrie stellt sich häufig die Frage nach einer angemessenen Prozessregelung. Aufgrund der steigenden Komplexität ist es meist unmöglich, aber zumindest sehr zeitaufwändig, ein mathematisch exaktes Modell zu entwerfen. Zum Zwecke der Regelung ist daher eine selbstadaptive, modellbasierte Regelung von besonderem Interesse.

Ein wichtiger Gesichtspunkt, der einer der Forschungsschwerpunkte im Teilprojekt C1 des Sonderforschungsbereichs 396 bildete, war ein **approximativer Inferenzalgorithmus**, der auf dem Interfacealgorithmus für diskrete Bayesnetze, den K. Murphy in seiner Dissertation vorgestellt hat. Bei der Verwendung für hybride, dynamische Bayesnetze kann nicht länger von der Berechnung eines exakten Mittelwerts bzw. einer exakten Varianz ausgegangen werden. Die Ursache hierfür liegt in der Approximation von gaußschen Mischverteilungen durch Mischverteilungen mit weniger Mischungskomponenten. In der Anwendung in der Regelungstechnik repräsentieren die Mischungskomponenten jedoch verschiedene Arbeitspunkte. Da meist nur ein Arbeitspunkt bzw. bei Übergängen zwei Arbeitspunkte aktiv sind, reicht die erreichte Genauigkeit für Regelungszwecke aus, die Regelungsgüte wird durch das Verwenden des approximativen Inferenzalgorithmus nicht beeinträchtigt. Die Laufzeit des approximativen Algorithmus steigt jedoch nur linear an, der Zeitbedarf des exakten Inferenzalgorithmus steigt exponentiell mit der Anzahl der Zeitscheiben an.

Der zweite Schwerpunkt bildete die **Modellerstellung**, wobei einer der Schwerpunkte das Auffinden einer geeigneten Struktur ist. Betrachtet man die Parameter eines hybriden Bayesnetzes genauer, so fällt auf, dass man jeder Belegungsmöglichkeit der diskreten Väter eines Knotens einen konkreten Parameter zuordnen kann. Falls die durch einen Versuchsplan erhobenen Daten zum Training eines Bayesnetzes dienen sollen, so ist darauf zu achten, dass für jede diskrete Knotenmenge  $\{X_1, \dots, X_n\}$ , die Väter eines anderen Knotens sind, ein vollfaktorieller Versuchsplan durchgeführt wird, oder zumindest alle möglichen Kombinationen  $x_1, \dots, x_n$  beobachtet werden.

Diese Erkenntnisse wurden beim Erstellen eines Versuchsplanes für den Spritzgussprozess angewandt. Die als wichtig erkannten Mehrfachwechselwirkungen können als verborgene Knoten in das Bayesnetz aufgenommen werden. Die Trainierbarkeit und insbesondere die Generalisierbarkeit des Bayesnetzes ist damit sichergestellt.

Selbstverständlich hat der Versuchsplan bzw. die erhobenen Daten auch einen Einfluss auf das automatische Strukturlernen. Falls die Daten direkt bei der Produktion erfasst wurden, so stammen sie meistent aus einem fest vorgegebenen Prozessfenster. Die Eingabegrößen erscheinen dabei stark korreliert, da sie aufeinander abgestimmt werden müssen, um ein optimales Produktionsergebnis zu erzielen. Auf der anderen Seite lassen sich die Eingabewerte meist unabhängig voneinander einstellen, so dass es sinnvoll ist die Eingabegrößen als unabhängig zu modellieren. Auf Bayesnetze bezogen heißt das, dass in diesen Fällen Kanten zwischen den Eingabeknoten überflüssig sind.

## 2.8 SFB 539, A4: Automatisches Glaukom-Screening

### Projektleitung:

Prof. Dr.-Ing. Heinrich Niemann, Prof. Dr. Georg Michelson,  
PD Dr.rer.nat. Berthold Lausen

### Beteiligte:

Ing. Radim Chrastek, Dr. med. Victor Derhartunian, Dipl.-Inf. Werner Adler

### Stichwörter:

Glaukom; optischer Sehnervenkopf; Segmentierung; morphologische Operationen;  
Hugh-Transformation; aktive Konturen

**Laufzeit:** 1.7.2003 - 30.6.2006

### Mitwirkende Institutionen:

Augenklinik mit Poliklinik  
Institut für Medizininformatik, Biometrie und Epidemiologie

Der Lehrstuhl für Mustererkennung befasst sich im Rahmen des SFB 539 (<http://www.sfb539.forschung.uni-erlangen.de>) (Teilprojekt A.4, Automatisches Glaukom -Screening (<http://www.imbe.med.uni-erlangen.de/%7Eteilpa4/>)) mit der Segmentierung des optischen Sehnervenkopfes für eine Glaukomfrüherkennung. Der Sehnervenkopf (der Austritt des Sehnervs aus dem Auge) wird mit dem Heidelberg Retina Tomograph (HRT) aufgenommen. Hierbei handelt es sich um ein Gerät, welches auf dem Prinzip eines konfokalen Mikroskops basiert. Anschließend werden verschiedene Oberflächen- und Volumenparameter (Merkmale) berechnet, die zur Glaukomklassifizierung dienen. Bis jetzt mussten hierfür die Grenzen des optischen Sehnervenkopfes (sog. Elschnigscher Skleraring) von einem Augenarzt aufwändig manuell markiert werden. Es wurden zwei Verfahren für die Segmentierung des optischen Sehnervenkopfes entwickelt.

Das erste Verfahren verwendet die HRT-Reflektivitätsbilder, d.h. Bilder die durch Mittelwertbildung aus den Schichtbildern berechnet werden. Dieses Verfahren basiert auf morphologischen Operationen, Hough -Transformation und aktiven Konturen. Da die HRT -Bilder Beleuchtungsunterschiede aufweisen, werden sie durch Korrekturkoeffizienten verbessert. Die

Korrekturkoeffizienten werden aus der approximierten Hintergrundsbeleuchtung geschätzt. In den nächsten Schritten werden die Lage und der Radius des optischen Sehnervenkopfes geschätzt. Der optische Sehnervenkopf gehört zusammen mit den Gefäßen zu den dunkelsten Stellen im Bild. Sie werden durch einfache Binarisierung extrahiert. Die Regionen des optischen Sehnervenkopfes werden dann anhand der geometrischen Eigenschaften von den Gefäßen getrennt. Außerdem werden relevante anatomische Strukturen des optischen Sehnervenkopfes segmentiert (sog. parapapilläre Atrophie und neuroretinaler Randsaum). Anhand der gewonnenen Informationen werden dann die genauen Grenzen des optischen Sehnervenkopfes mittels der aktiven Konturen ermittelt. Das Verfahren wurde durch Klassifikationsraten von verschiedenen Klassifikatoren anhand der Fallkontrollstudie mit von einem erfahrenen Augenarzt eingezeichneten Grenzen des optischen Sehnervenkopfes validiert. Folgende Ergebnisse hinsichtlich der Glaukomdiagnose (normal/glaukomatös) wurden erzielt: linear discriminant analysis mit **27.7%** der geschätzten Fehlerrate für automatische Segmentierung (*aut*) und **26.8%** der geschätzten Fehlerrate für manuelle Einzeichnung (*man*), classification trees mit **25.2%** (*aut*) und **22.0%** (*man*) und bootstrap aggregation mit **22.2%** (*aut*) und **13.4%** (*man*).

Die oben aufgeführten Ergebnisse zeigten, dass der Algorithmus vergleichbar mit dem Augenarzt war, brauchte allerdings noch Verbesserung. Diese wurde durch die Fusion von HRT - Bildern und Farbfotos erzielt. Die beiden Bildmodalitäten beinhalten komplementäre Informationen des optischen Sehnervenkopfes. Die Segmentierung in HRT - Bildern wird durch Information aus den Farbphotos gesteuert. Zum einen werden die segmentierten Gefäße in Farbphotos zur Eliminierung der falschen Anziehung der aktiven Kontur verwendet. Zum anderen wird der Suchraum der aktiven Konturen durch die Kontur des optischen Sehnervenkopfes, die in den Farbphotos segmentiert wurde, beschränkt. Dieses multimodale Verfahren brachte eine Verbesserung von ursprünglichen ca. 74% (für das monomodale Verfahren) auf 89%, qualitativ beurteilt. Die quantitative Evaluation wird noch durchgeführt werden.

## **2.9 SFB 603, B2: Optimierungsansatz für die Integration von Kamerabildern bei der Klassifikation**

### **Projektleitung:**

Prof. Dr.-Ing. Heinrich Niemann

### **Beteiligte:**

Dipl.-Ing. Christian Derichs, Dipl.-Inf. Benjamin Deutsch

### **Stichwörter:**

Optimierung; Sensordatenfusion; Objektverfolgung; Objektklassifikation

**Laufzeit:** 1.1.1998 - 31.12.2006

**Förderer:** DFG

### **Kontakt:**

Dipl.-Inf. Benjamin Deutsch

Tel.: +49.9131.85.27874, Fax: +49.9131.303811,

E-Mail: deutsch@informatik.uni-erlangen.de



Das langfristige Ziel des Teilprojekts B2 ([http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B2/tpB2\\_home\\_g.html](http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B2/tpB2_home_g.html)) des Sonderforschungsbereiches 603 (<http://sfb-603.uni-erlangen.de/>) ist die Entwicklung von optimierungsbasierten Verfahren für die Integration von Kamerabildern bei der Klassifikation, Lokalisation und Verfolgung von Objekten. In den letzten Jahren konnte die Integration in einem einheitlichen Formalismus probabilistisch modelliert und als ein Optimierungsproblem formuliert werden. In letzter Zeit wurden vermehrt Anstrengungen unternommen, intelligente Komponenten in die Auswahlentscheidung für die nächste auszuführende Aktion des Systemagenten zu integrieren. Im Vordergrund stand jeweils das Problem, die *Kosten* einer Aktion in die Auswahl mit einzubeziehen.

In der Ansichtenplanung zur Klassifikation und Lokalisation wurden Aktionen bisher ausschließlich durch den Informationsgehalt bezüglich ihrer Aussagekraft über Klasse und Lage eines Elementes einer Objektmenge gewertet und mittels geeigneter Reinforcement Learning Techniken trainiert. Da zusätzlich bereits die Möglichkeit bestand, alle eingenommenen Ansichten einer Aktionsfolge mit Hilfe von Partikelfiltern auf eine resultierende Entropieverteilung zu fusionieren, folgte der Agent bei ausreichendem Training im direkten Folgeschritt immer dem absoluten Minimum dieser Verteilung.

Dieses bezüglich einer Einschrittplanung optimale Verhalten, welches bereits zuverlässig gelernt werden konnte, ist jedoch mitunter suboptimal, falls die Aktionen, die dem Agenten zur Verfügung stehen, mit verschiedenen Kosten behaftet sein sollten. Die Wichtigkeit dieser Annahme lässt sich sehr einfach an den Gegebenheiten einer realen Umgebung erkennen, in der der Agent etwa durch den mobilen Serviceroboter MOBSY repräsentiert wird. Offensichtlich muss hier beispielsweise eine kreisförmige Fahrt um das zu klassifizierende Objekt mit relativ größeren Kosten beaufschlagt werden als zum Beispiel ein schlichtes Heranzoomen an selbiges, falls es beispielsweise gelten sollte, die Zeit für die Klassifikation klein zu halten.

Aus diesem Grund wurde das Belohnungsmodell des Reinforcement Learning allgemein durch die Überlagerung beliebiger zusätzlicher Kostenfaktoren erweitert. Insbesondere wurde dort auch implizit die Anzahl der zur Klassifikation notwendigen Ansichten integriert, da die Aufgabe unter der Annahme existierender Kosten nicht mehr zwangsläufig mit der minimalen Anzahl von Ansichten optimal lösbar wäre. Die zu behebende Problematik bestand im Wesentlichen darin, dass in einem beliebigen System prinzipiell keine Information über den Wertebereich der kostenunabhängigen Belohnungen vorliegen. Um aber weder diese entropiebasierte noch die kostenabhängige Belohnung bzw. Bestrafung in der Trainingsphase zu bevorzugen, war es nötig, den einen Kostenfaktor, auf den die Problemstellung reduziert werden konnte, während des Trainings an diesen Wertebereich anzupassen, ihn also mitzulernen. Das besondere Augenmerk war dabei darauf gerichtet, das ursprüngliche Training durch diese Maßnahme nicht entscheidend zu verlängern, was durch eine optimierte Suche des geeigneten Kostenfaktors und unterstützende Maßnahmen, wie die Einführung einer Zustandsqualität zur besseren Vergleichbarkeit von Trainingssituationen, zunächst hinreichend gut gelöst werden konnte.

Des Weiteren konnte die Allgemeingültigkeit der bisher nur zur Objektklassifikation und -lokalisierung verwendeten Lernstrategien der Ansichtenplanung gezeigt werden. Hierzu wurde innerhalb einer Diplomarbeit in Kooperation mit dem Lehrstuhl für Optik (<http://www.optik.uni-erlangen.de/osmin/homepage/>) der FAU dargelegt, dass sich das Verfahren in einfacher Art und Weise auch auf die Problematik der Objektrekonstruktion mittels einer minimalen Anzahl notwendiger Aufnahmen übertragen lässt. Der einzige Unterschied, den es hier neben der angepassten Modellierung einzelner Zustände zu beachten galt, war die durch die andersartige Aufgabenstellung zwangsläufig veränderte und geschickt zu wählende Belohnungsmodellierung für Aktionen des Agenten.

Auch im Bereich der Objektverfolgung konnten die Kosten einer Aktion, im Sinne der Durchführbarkeit einer Sensoraktion in Abhängigkeit von zuvor durchgeführten Aktionen, berücksichtigt werden. Am Beispiel der Brennweitereinstellung ist es schnell ersichtlich, dass eine Kamera zwischen zwei Aufnahmen bei schneller Bildrate nur eine eingeschränkte Brennweitenänderung durchführen kann. Daher ist es notwendig, eine Brennweitenänderung auch dahingehend zu bewerten, welche Änderungen dadurch in Zukunft ermöglicht oder verhindert werden.

Bisher war die zu einem Zeitpunkt gewählte Sensordatenaufnahme bei der Objektverfolgung optimal für einen Blick in die Zukunft, der sich nur auf den nächsten Zeitschritt beschränkt. Aufgrund der Notwendigkeit, auch weiter in der Zukunft liegende Aufnahmen bewerten zu können, musste die Methodik so erweitert werden, dass sie mehrere Zeitschritte in die Zukunft blickt.

Dies wurde erreicht, indem das in der Objektverfolgung verwendete *Kalman Filter* mehrere Zeitschritte in die Zukunft prädiziert wird. Für die Berechnung der erwarteten Entropie, welches die Grundlage für die informationstheoretische Sensoraktionswahl bildet, ist lediglich die Kovarianz der prädizierten probabilistischen Positionsschätzung des Objektes relevant, die im Kalman Filter unabhängig von den (in der Zukunft) gemachten Sensoraufnahmen berechnet werden kann.

Auch die Sichtbarkeit wird in der erweiterten Methodik berücksichtigt. Wie bisher kann eine Kamera, wenn sie in einem Zeitschritt das Objekt nicht sieht, keine Information in die Zustandsschätzung einbringen. Bei der bisherigen Auswertung für einen Zeitschritt ergab sich der Informationsgewinn aus der erwarteten Entropie, gewichtet mit der Wahrscheinlichkeit, das Objekt zu sehen. In der Mehrschritt-Variante für  $t$  Zeitschritte wird die erwartete Entropie für jede der  $2^t$  möglichen Folgen von Sichtbarkeit berechnet und gemäß der Wahrscheinlichkeit dieser Folge gewichtet summiert.

Weiterhin wurde die Objektverfolgung durch die sequentielle Auswertung des Kalman Filters dahingehend erweitert, dass eine beliebige Anzahl von Sensoren ohne Änderung des Algorithmus oder der verwendeten Programme verwendet werden kann. Dadurch ist es möglich, mit  $n$  statischen und  $m$  bewegten Kameras arbeiten zu können.

Durch die sequentielle Auswertung und geschickte Optimierung konnte bereits annähernd Echtzeit erreicht werden, selbst bei der Berücksichtigung mehrerer Zeitschritte. So ist die Objektverfolgung in der Lage, drei Kameras und je drei Zeitschritte in die Zukunft bei 7,5 Bildern in der Sekunde zu verarbeiten. Dies wurde in einer Demonstration mit der Farbhistogramm-Objektverfolgung des EU-Projektes VAMPIRE (<http://www.vampire-project.org/>) gezeigt. Momentane Arbeiten, die Sichtbarkeit enger in die erwartete Kovarianz einbeziehen, werden eine Auswertung ermöglichen, die vom Rechenaufwand linear in der Anzahl der Kameras und linear (statt bisher exponentiell) in den prädizierten Zeitschritten ist.

Weitere Fortschritte konnten beim Teilziel erreicht werden, mit Hilfe eines Agenten ein sich bewegendes Objekt zu greifen. In einer Studienarbeit wurde ein System entwickelt, in dem ein optimaler Bewegungsplan in Abhängigkeit von der erwarteten Objektposition und -lage errechnet wird. Das System ist für verschiedene Agententypen konzipiert und beinhaltet auch probabilistische Ansätze.

Eine weitere Diplomarbeit beschäftigte sich mit der probabilistischen, adaptiven Sensordatenfusion in der 3-D Objektverfolgung. Dabei wurde der *STAPLE-Algorithmus* auf einen kontinuierlichen Raum erweitert, um analog zur *Demokratischen Integration* die Daten aus einer Menge von Hinweisen probabilistisch und adaptiv zu fusionieren.

Zusammen mit dem Teilprojekt B6 ([http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B2/tpB2\\_home\\_g.html](http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B2/tpB2_home_g.html)) des Sonderforschungsbereichs 603 (<http://sfb-603.uni-erlangen.de/>) und dem EU-Projekt VAMPIRE (<http://www.vampire-project.org/>) wurden außerdem mehrere Verfahren untersucht, um existierende Szenenrekonstruktionen durch neue Bildsequenzen zu erweitern.

## 2.10 SFB 603, B6: Rechnergestützte Endoskopie des Bauchraums

### **Projektleitung:**

Prof. Dr. med. Werner Hohenberger, PD Dr. med. Christoph Schick  
Prof. Dr. Günther Greiner, Prof. Dr.-Ing. Heinrich Niemann

### **Beteiligte:**

Dipl.-Inf. Florian Vogt, Dr. med. Sophie Krüger, Dipl.-Inf. Marco Winter

### **Stichwörter:**

Endoskopie; Minimal-invasive Operation; 3-D-Visualisierung; Lichtfeld; Registrierung; Bildverbesserung

**Laufzeit:** 1.1.2000 - 31.12.2006

**Förderer:** DFG

### **Mitwirkende Institutionen:**

Chirurgische Universitätsklinik Erlangen  
Lehrstuhl für Graphische Datenverarbeitung Erlangen

### **Kontakt:**

Dipl.-Inf. Florian Vogt  
Tel.: +49.9131.85.27891, Fax: +49.9131.303811,  
E-Mail: [vogt@immd5.informatik.uni-erlangen.de](mailto:vogt@immd5.informatik.uni-erlangen.de)

Im Teilprojekt B6 ([http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B6/tpB6\\_home\\_g.html](http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B6/tpB6_home_g.html)) "Rechnergestützte Endoskopie des Bauchraums" des Sonderforschungsbereichs 603 ([http://sfb-603.uni-erlangen.de/HTML/sfb603\\_g.html](http://sfb-603.uni-erlangen.de/HTML/sfb603_g.html)) wurde in Zusammenarbeit mit der Chirurgischen Universitätsklinik mit den Arbeiten zu Augmented Reality in der minimal-invasiven Chirurgie begonnen. Hierzu war zunächst die Anschaffung eines genaueren Positionsbestimmungssystems notwendig.

Auf Grund der relativ großen Fehler bei der Positionsbestimmung des Endoskops mit dem Roboterarm AESOP wurde zu Beginn des Jahres 2004 das optische Trackingsystem smARTtrack1 der Firma Advanced Realtime Tracking GmbH (ART) angeschafft und in Betrieb genommen. Dabei war zunächst die Synchronisation mit dem ART-System zu lösen: das S-Video Signal der Endoskop-Kamera dient als Synchronisationssignal. Das heißt die Endoskop-Position wird mit 50 Hz (50 Halbbilder bei S-Video) bestimmt. Für die Positionsbestimmung wird ein Target benötigt. Mehrere Designs wurden untersucht, wobei der Chirurg durch das Target nicht gestört werden darf. In Kooperation mit ART wurde ein Target inklusive Endoskop-Adapter entworfen

und gefertigt, mit dem die Position robust und genau bestimmt werden kann. Die Transformation vom Target zur Endoskopspitze (Hand-Auge-Transformation) wurde mit dem automatischen Hand-Auge-Kalibrierverfahren von Daniilidis bestimmt.

Mit dem optischen Trackingsystem wurden unter OP-realistischen Bedingungen im Labor mehrere Experimente durchgeführt. Die Position des Endoskops kann mit smARTtrack1 ungefähr fünf Mal genauer bestimmt werden als mit AESOP. Insbesondere ist es nun möglich Szenengeometrie sehr exakt zu rekonstruieren. Die Szenengeometrie besteht aus 3-D-Punkten der Oberfläche. Zur Evaluation des Gesamtsystems wurde eine Kugel mit Radius 22,5 mm vermessen. Dabei wurde anhand der rekonstruierten 3-D Punkte der Kugelmittelpunkt und -radius geschätzt. Der Radius wurde bis auf 1 mm (entspricht 4,4% Fehler) genau bestimmt, wobei die mittlere Abweichung der Punkte von der geschätzten Kugeloberfläche 0,7 mm betrug (Formabweichung). Eine Fehlerrate von 5% ergab sich bereits bei der Hand-Auge-Kalibrierung. Die Ergebnisse passen daher sehr gut zusammen. Mit dem System wurden außerdem mehrere Lichtfelder unterschiedlicher Szenen rekonstruiert, beispielsweise Lichtfelder eines Leber-Gallenblasen-Modells aus Silikon.

Es wurde mit der Datensammlung für eine anatomische Datenbank begonnen. Diese soll CT-Daten mehrerer Patienten enthalten, inklusive einer Segmentierung relevanter anatomischer Strukturen. Beispielhaft werden solche Strukturen segmentiert, die für eine laparoskopische Cholecystektomie relevant sind. Mit Hilfe der rekonstruierten Geometrie lässt sich die Registrierung der segmentierten CT-Daten mit smARTtrack1 durchführen. Dies ermöglicht die Einblendung von CT-Daten (Augmented Reality), sowohl in Echtzeit in das endoskopische 2-D live Bild wie auch in das 3-D Lichtfeld.

## **2.11 SFB 603, C2: Analyse, Codierung und Verarbeitung von Lichtfeldern zur Gewinnung realistischer Modelldaten**

### **Projektleitung:**

Prof. Dr.-Ing. Heinrich Niemann, Prof. Dr. Günther Greiner

### **Beteiligte:**

Dipl.-Inf. Ingo Scholz, Dipl.-Inf. Christian Vogelgsang

### **Stichwörter:**

bildbasierte Modellierung; Kalibrierung monokularer Bildsequenzen;

**Laufzeit:** 1.1.1998 - 31.12.2006

**Förderer:** DFG

### **Mitwirkende Institutionen:**

Lehrstuhl für Graphische Datenverarbeitung Erlangen

### **Kontakt:**

Dipl.-Inf. Ingo Scholz

Tel.: +49.9131.85.27891, Fax: +49.9131.303811,

E-Mail: scholz@informatik.uni-erlangen.de

Das Teilprojekt C2 ([http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/C2/tpC2\\_home\\_g.html](http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/C2/tpC2_home_g.html)) des Sonderforschungsbereichs 603 ([http://sfb-603.uni-erlangen.de/HTML/sfb603\\_g.html](http://sfb-603.uni-erlangen.de/HTML/sfb603_g.html)) bearbeitet zusammen mit dem Lehrstuhl für Graphische Datenverarbeitung (<http://www9.informatik.uni-erlangen.de/>) (LGDV) das Thema der “Analyse, Codierung und Verarbeitung von Lichtfeldern zur Gewinnung realistischer Modelldaten”. Die in diesem Teilprojekt zentrale Datenstruktur des Lichtfelds erlaubt es, durch sog. bildbasierte Modellierung beliebige Ansichten einer realen Szene zu generieren, deren Aussehen durch eine Sammlung an Bilddaten bekannt ist. Die benötigten Informationen über Parameter und Positionen der verwendeten Kamera werden über Verfahren der “Struktur aus Bewegung” (Structure from Motion) direkt aus den Bilddaten ermittelt, bei denen es sich meistens um einen Bildstrom einer handgeführten Kamera handelt. Das Teilprojekt C2 wird bereits seit 1998 von der Deutschen Forschungsgemeinschaft (<http://www.dfg.de/>) (DFG) gefördert.

Die bei Lichtfeldern übliche Betrachtung von ausschließlich statischen Szenen wurde in den letzten Jahren immer mehr auch auf bewegte Szenen erweitert. Die bereits in 2003 begonnenen Arbeiten zur Rekonstruktion von Lichtfeldern aus Szenen, die bewegte, aber in sich starre Objekte enthalten, wurden 2004 fortgesetzt. Nach der Trennung von Merkmalen in “Hintergrund” und “bewegtes Objekt” können nun die beiden unabhängig voneinander berechneten 3-D-Rekonstruktionen mit ausreichender Genauigkeit registriert werden. Daraus lässt sich nun die Eigenbewegung des Objekts berechnen. Ähnliche Objektpositionen werden mittels einer Vektorquantisierung jeweils einem Zeitschritt zugeordnet. Die Visualisierung der so entstehenden *dynamischen Lichtfelder* erfolgt durch sog. *Vertrauenskarten* (Confidence Maps), die das Ausblenden des bewegten Objekts zu nicht sichtbaren Zeitschritten erlauben.

Die Berechnung einer global konsistenten Rekonstruktion stellt selbst für statische Lichtfelder nach wie vor ein Problem dar. Ein Verfahren, das wiederkehrende Kamerapositionen ausnutzt um akkumulierte Fehler zu korrigieren, wurde untersucht und führte zur Verringerung des globalen Rekonstruktionsfehlers.

Ein neu entwickeltes Verfahren zur Evaluation der Lichtfeldqualität ermöglicht einen objektiven Vergleich verschiedener Verfahren zur Lichtfeldrekonstruktion und wurde bereits für Untersuchungen zur Berechnung intrinsischer Kameraparameter eingesetzt.

Die Anwendbarkeit der entwickelten Verfahren wurde außerdem in mehreren Kooperationen mit anderen Projekten gezeigt. In Zusammenarbeit mit dem Graduiertenkolleg “Dreidimensionale Bildanalyse und -synthese” (<http://www9.informatik.uni-erlangen.de/Research/gk244/index.html>) wurden 3-D-Rekonstruktionsalgorithmen zum schnellen Training für die Objekterkennung eingesetzt. Zusammen mit dem Teilprojekt B2 ([http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B2/tpB2\\_home\\_g.html](http://sfb-603.uni-erlangen.de/HTML/TEILPROJEKTE/B2/tpB2_home_g.html)) des Sonderforschungsbereichs 603 und dem EU-Projekt Vampire (<http://www5.informatik.uni-erlangen.de/vampire/index.php?lang=de&site=1>) wurden mehrere Verfahren untersucht um existierende Szenenrekonstruktionen durch neue Bildsequenzen zu erweitern.

## 2.12 SmartKom

### **Projektleitung:**

Prof. Dr.-Ing. Heinrich Niemann, Dr.-Ing. Elmar Nöth

### **Beteiligte:**

Dr. phil. Anton Batliner, Dipl.-Inf. Carmen Frank, Dipl.-Inf. Johann Adelhardt, Rui-ping Shi, M.S., Dipl.-Inf. Viktor Zeissler

### **Stichwörter:**

Dialog; Mensch-Maschine-Interaktion; Sprachverarbeitung; Prosodie; Mimik; Gestik

**Laufzeit:** 1.9.1999 - 31.1.2004

**Förderer:** bmb+f

### **Mitwirkende Institutionen:**

DaimlerChrysler AG, European Media Lab, ICSI, Universität Stuttgart, Universität München, MediaInterface Dresden GmbH, Philips GmbH, Siemens AG ZT, Sony International GmbH, Sympalog AG, DFKI

### **Kontakt:**

Dr.-Ing. Elmar Nöth

Tel.: +49.9131.85.27888, Fax: +49.9131.303811,

E-Mail: noeth@informatik.uni-erlangen.de

Eine der wichtigsten Herausforderungen für die Wissensgesellschaft ist derzeit die Schaffung intelligenter multimodaler Benutzerschnittstellen, die den natürlichen Kommunikationsstil von Computerlaien akzeptieren und so unterstützen, dass eine für den Menschen intuitive und benutzergerechte Interaktion entsteht. Computeranwendungen nehmen in allen Lebensbereichen ebenso stark zu wie die Komplexität der Systeme. Hinzu kommt, dass die Benutzer immer weniger Zeit haben, sich in neue Bedienkonzepte einzuarbeiten, und dass die Kosten für die Entwicklung intelligenter Benutzerschnittstellen reduziert werden müssen.

Im Leitprojekt SmartKom wurden Konzepte für die Entwicklung völlig neuartiger Formen der Mensch-Technik-Interaktion erprobt. Solche Konzepte bauen die bestehenden Hemmschwellen von Computerlaien bei der Nutzung der Informationstechnologie ab und leisten somit einen Beitrag zur Nutzerfreundlichkeit - damit auch zur Akzeptanz - und zur Nutzerzentrierung in der Wissensgesellschaft. Die Leitvorstellung des SmartKom-Projekts war es, die Vorteile sprachlich-dialogischer Kommunikation mit den Vorteilen graphischer Bedienoberflächen und taktischer Interaktionsformen zu einem höherwertigen, oft multimodal genannten Bedienparadigma zu verschmelzen.

Der Lehrstuhl für Informatik 5 (Mustererkennung) entwickelte in SmartKom Module zur multimodalen Verarbeitung von Sprache, Mimik und Gestik. Dabei handelt es sich zum einen um die Erkennung semantischer Information bei Sprache und Gestik (Akzent- sowie syntaktisch-prosodische Grenzerkennung bei Sprache, Erkennung von Zeigegesten), zum anderen um die Er-

kennung von emotionalen Benutzerzuständen in allen drei Modalitäten. In Zusammenarbeit mit anderen SmartKom-Partnern werden sog. Use Cases (exemplarische Anwendungs-Konstellationen und -Aufgaben) entwickelt, deren Randbedingungen geeignet sind, den Nutzen der Verarbeitung von emotionalen Benutzerzuständen im Gesamtsystem zu demonstrieren.

## 2.13 SmartWeb

### **Projektleitung:**

Dr.-Ing. Elmar Nöth

### **Beteiligte:**

Dr. phil. Anton Batliner, Dipl.-Math Silvia Dennerlein, Dipl.-Inf. Christian Hacker,  
Dipl.-Inf. Axel Horndasch

### **Stichwörter:**

Spracherkennung; OOV-Verarbeitung; Biosignale; Benutzerzustandsklassifikation;  
multimodale Informationsquellen

**Laufzeit:** 1.4.2004 - 31.3.2007

**Förderer:** Bundesministerium für Bildung und Forschung

### **Kontakt:**

Dr.-Ing. Elmar Nöth  
Tel.: +49.9131.85.27888, Fax: +49.9131.303811,  
E-Mail: noeth@informatik.uni-erlangen.de

## **Leitinnovation SmartWeb**

Das Ziel des BMB+F-geförderten Projekts *SmartWeb* (<http://www.smartweb-projekt.de>) ist der intelligente, multimodale, mobile Zugriff auf das Semantische Web. An der Umsetzung dieser Vision arbeiten seit April 2004 vierzehn Konsortialpartner unter der Führung des Deutschen Forschungsinstituts für künstliche Intelligenz (DFKI). Neben einer Reihe von Universitäten und Forschungseinrichtungen sind auch kleinere und mittlere Unternehmen sowie die Großindustrie beteiligt. Die Planungen sehen eine Laufzeit von drei Jahren bis 2007 vor.

Ein großes Teilprojekt in SmartWeb bildet das *Semantische Web*. Dieses *Internet der nächsten Generation* basiert auf der inhaltlichen Beschreibung digitaler Dokumente mit standardisierten Vokabularien, die eine maschinell verstehbare Semantik haben. In SmartWeb wird vor allem angestrebt, die maschinenlesbaren Inhalte des World Wide Web auch maschinell verstehbar zu machen. Die Arbeitspakete, die am *Lehrstuhl für Mustererkennung (Informatik 5)* bearbeitet werden, sind Teil des zweiten Teilprojekts, das sich mit dem multimodalen, mobilen Zugang zu den Inhalten des Semantischen Webs beschäftigt. Das Architekturmodell von SmartWeb sieht dabei ein Client-Server-Szenario vor, bei dem die Benutzereingaben von einem Engerät (PDA, Smartphone, o.Ä.) zu einem Server geschickt werden, auf dem die Dialogverarbeitung stattfindet. Einen Teil des Dialog-Servers bildet ein multimodaler Erkenner (MME), der neben einem Spracherkennung weitere Software-Module zur Verarbeitung der verschiedenen Eingabemodalitäten (u.a. Sprach-, Videosignale) umfasst. Drei der Module werden am *Lehrstuhl für Mustererkennung (LME)* entwickelt: die Detektion und Verarbeitung von Out-Of-Vocabulary-

Wörtern bei der Spracherkennung (in enger Zusammenarbeit mit der Firma Sympalog), die Klassifikation des Benutzerzustands auf der Basis von Sprach-, Video- und Biosignalen und die Extraktion von multimodalen Informationsquellen zur Verbesserung der Erkennungsergebnisse.

Für alle LME-Module wurden in verschiedenen Arbeitstreffen die proprietären Schnittstellen für die Kommunikation innerhalb des MMEs festgelegt; außerdem wurde eine Spezifikation für den Informationsaustausch zwischen dem MME und den anderen Komponenten des Dialogservers (Dialogmanager, Sprachsynthese-Modul, etc.) erarbeitet. Der Entwurf basiert auf dem XML-Dialekt EMMA (Extensible MultiModal Annotation markup language, <http://www.w3.org/TR/emma>) und wurde um Strukturen zur Darstellung der Ergebnisse der OOV-Verarbeitung erweitert.

### **Erkennung und Verarbeitung von Out-Of-Vocabulary-Wörtern**

Die große Bedeutung der Erkennung und Verarbeitung von OOV-Wörtern in SmartWeb ergibt sich aus dem Projektziel, die Themenbereiche beim Abfragen von Informationen möglichst offen zu gestalten. Für die Architektur des OOV-Moduls wurde in Absprache mit dem Projektpartner Sympalog, der den Spracherkennung für SmartWeb liefert, ein erster Entwurf erarbeitet: Bei der Erkennung von OOV-Wörtern sollen bereits erprobte Methoden für die akustischen Modelle und die Sprachmodellierung eingesetzt und verfeinert werden; mittels eines speziellen, zusätzlichen Erkenners für Wortuntereinheiten werden Hypothesen für die OOV-Bereiche des Sprachsignals generiert. Die Umsetzung der Wortuntereinheiten in eine graphematische Darstellung erfolgt durch ein (sprachabhängiges) Zusatzmodul, dessen Ergebnisse in die Ausgabe des Erkenners integriert werden.

### **Erweiterung und Verbesserung des Spracherkenners**

Um den aktuellen Stand der Erkennung-Forschung am *Lehrstuhl für Mustererkennung*, v.a. im Hinblick auf die Merkmalsberechnung, im Sympalog-Spracherkennung für SmartWeb verfügbar zu machen sowie die bei Sympalog neu entwickelten Verfahren und Architekturansätze für den LME-Erkennung zu übernehmen, wurden Vergleichstests durchgeführt. Erste Ergebnisse zeigten eine leichte Überlegenheit des LME-Erkenners bei der Erkennung von Ziffernfolgen; für Training und Test wurden dabei Telefoniedaten aus unterschiedlichen Anwendungsszenarien verwendet.

Für das mobile SmartWeb-Szenario wird es zwei Erkennung geben, jeweils integriert in ein englisches und ein deutsches System. Das International Computer Science Institute in Berkeley (ICSI) verwendet für das englische System auch alternative Merkmalsätze wie TRAP-Merkmale (TempoRAI Pattern). Es wurde damit begonnen, in Zusammenarbeit mit dem ICSI solche TRAP-basierten Merkmale auch für das deutsche Erkennungssystem umzusetzen. TRAP-Merkmale geben Hinweise über das Verhalten in bestimmten Spektralbändern in einem langen zeitlichen Kontext.

### **Benutzerzustandsklassifikation**

Die Benutzerzustandsklassifikation soll anhand von Sprach-, Video- und physiologischen Daten erfolgen. Dabei können die physiologischen Merkmale als Referenz ('Ground Truth') dienen oder in Situationen, in denen das Messen physiologischer Signale möglich ist, direkt zur Emotionserkennung herangezogen werden. Es wurde beschlossen, in einem geplanten Kfz- und Endgerätszenario lediglich Sprach- und Videodaten zur Benutzerzustandsklassifikation zu verwenden. In einem Motorrad-Szenario soll Stress beim Fahrer anhand physiologischer Merkmale erkannt werden. Dafür wurden in Absprache mit BMW verschiedene Experimente (Fahrsimulator, Windkanal, reale Fahrsituation) geplant. Vorab sind verschiedene Pilotexperimente ange-



setzt.

### **Extraktion von multimodalen Informationsquellen**

In SmartWeb wird automatisch Offtalk, Offview und Offhead aus Sprache und Bild erkannt. Im Falle von Offtalk wird erkannt, ob der Benutzer mit dem System redet, oder etwa mit sich selbst oder einer anderen Person. Erste Experimente wurden mit 95 prosodischen Merkmalen auf dem SmartKom-Korpus durchgeführt. Offhead und Offview werden aus dem Videostrom detektiert. Offhead gibt Auskunft darüber, ob das Gesicht des Benutzers im Kamerabild zu sehen ist, Offview, ob der Benutzer in Richtung des mobilen Endgerätes schaut. Dazu waren erste Absprachen mit den Partnern der Uni Saarbrücken notwendig, die Lippenlesen mit Hilfe des Videostroms vorgeschlagen haben.

## **2.14 Visual Active Memory Processes and Interactive Retrieval (VAMPIRE)**

### **Projektleitung:**

Prof. Dr.-Ing. Heinrich Niemann, Prof. Dr.-Ing. Joachim Denzler

### **Beteiligte:**

Dipl.-Inf. Timo Zinßer, Dipl.-Inf. Christoph Gräßl, Dipl.-Inf. Jochen Schmidt

### **Stichwörter:**

Aktive Speicherprozesse; Kognitives Sehen; Rechnersehen; Erweiterte Realität; Objektverfugung; Objekterkennung; Bildbasierte Modellierung

**Laufzeit:** 1.5.2002 - 30.4.2005

**Förderer:** EU 5. Rahmenprogramm

### **Mitwirkende Institutionen:**

Bielefeld I Applied Computer Science, Bielefeld II Neuroinformatics, TU Graz, Surrey

### **Kontakt:**

Dipl.-Inf. Timo Zinßer

Tel.: +49.9131.85.27799, Fax: +49.9131.303811,

E-Mail: zinsser@informatik.uni-erlangen.de

Im Rahmen des EU-Projekts VAMPIRE (<http://www5.informatik.uni-erlangen.de/vampire>) (Visual Active Memory Processes and Interactive Retrieval) wird seit Juni 2002 Forschung im Bereich der automatischen Analyse von Videosequenzen durchgeführt. Im Vordergrund stehen dabei die Erkennung von Objekten und Bewegungsabläufen, sowie das Lernen neuer Objekt- und Bewegungsmodelle. Neben dem Lehrstuhl für Mustererkennung sind auch die Universität Bielefeld, die Technische Universität Graz und die University of Surrey am Projekt beteiligt.

Ein Schwerpunkt der Arbeit des Lehrstuhls für Mustererkennung innerhalb des Projekts VAMPIRE liegt in den Bereichen *3-D Rekonstruktion* und *bildbasierte Objektmodelle*. Hierzu wird ein *structure-from-motion Ansatz* verwendet, der bereits im Teilprojekt C2 des SFB 603 (<http://sfb-603.uni-erlangen.de>) eingesetzt wird. Ausgangspunkt ist eine Videosequenz, die mit einer

handgeführten Kamera aufgenommen wurde. Der erste Verarbeitungsschritt ist die Bestimmung korrespondierender Merkmalspunkte in aufeinanderfolgenden Bildern der Sequenz, wozu zwei verschiedene Verfahren näher untersucht wurden. Zum einen wurden in jedem Bild Merkmalspunkte mit zugehörigen Merkmalsvektoren extrahiert, um dann das Korrespondenzproblem durch eine Suche des ähnlichsten Merkmalsvektors im jeweils nächsten Bild zu lösen. Zum anderen wurde ein Punktverfolger eingesetzt, der auf dem Ansatz des *Kanade-Lucas-Tomasi Trackers* basiert. Bisher durchgeführte Optimierungen erlauben das Verfolgen von 250 Punkten mit 30 Bildern pro Sekunde. Um das Verfahren besser an eine von der TU Graz entwickelte Hochgeschwindigkeitskamera anzupassen, wurde die Translationsschätzung, welche auf einem iterativen Gradientenabstiegsverfahren mit mehreren Hierarchieebenen basiert, durch ein effizientes Blockvergleichsverfahren ersetzt. Dadurch konnte die Berechnungszeit bei der Verfolgung von zehn Punkten auf 2.8 ms pro Bild verkürzt werden, was einer Verdopplung der vorher erreichten Bildrate entspricht.

Neben der 3-D Rekonstruktion sind die *Objektverfolgung* und *Objekterkennung* ein weiterer Schwerpunkt im Projekt VAMPIRE, der vom Lehrstuhl für Mustererkennung untersucht wird. Bei der Objektverfolgung muss zwischen *datengetriebenen* und *modellbasierten* Verfahren unterschieden werden. Datengetriebene Verfahren besitzen die Eigenschaft, dass sie ohne a priori Wissen (beispielsweise Geometrieinformationen) über das zu verfolgende Objekt arbeiten können. Dies hat den Vorteil, dass sofort nach der Detektion einer Bewegung in einer Szene die Verfolgung eines unbekanntes Objektes starten kann. Nachteilig ist allerdings, dass bei starken externen Rotationen das Objekt verloren wird, weil kein Wissen vorhanden ist, wie das Objekt aus einer anderen Ansicht aussieht. Diesen Nachteil kompensieren Ansätze, die sehr robust gegenüber Erscheinungsänderungen sind, bzw. modellbasierte Verfahren, die während eines Trainings signifikante Informationen über das Objekt sammeln. Nachdem ein Objekt mit Hilfe eines Klassifikationsverfahrens erkannt wurde, kann von einem datengetriebenen auf ein modellbasiertes Verfahren umgeschaltet werden.

Nachdem im Jahre 2002 der *Hyperebenenansatz*, mit welchem sowohl Translation, Rotation, Skalierung und projektive Verzerrungen eines Objekts in der Bildebene geschätzt werden können, implementiert und verbessert wurde, ist dieses Jahr ein weiterer Ansatz zur datengetriebenen Objektverfolgung untersucht worden. Der neue Ansatz basiert auf Farbhistogrammmerkmalen und dem CONDENSATION Algorithmus, der im SFB 603 / B2 (<http://sfb-603.uni-erlangen.de>) bereits eingesetzt wird. Ziel ist es, im aktuellen Bild einer Videosequenz das Rechteck zu finden, das ein möglichst ähnliches HS-I Histogramm besitzt wie das Referenzhistogramm, das in einem initialen Bild bestimmt wurde. Bis jetzt kann nur die Translation des Objekts berechnet werden, doch dieses Verfahren zeichnet sich durch hohe Robustheit bei Beleuchtungsschwankung, Verdeckung, schnellen Bewegungen und anderen Änderungen der Erscheinung aus.

Um eine robuste Objekterkennung in einer natürlichen Szene zu gewährleisten, wurden lokale Merkmale verwendet. Zur Klassifikation wird die Zuordnung der lokalen Merkmale zu den gespeicherten Merkmalen der Objektmodelle ausgewertet. Der Vorteil dieses Verfahrens ist die Invarianz bezüglich Translation, Skalierung und interner Rotation. Des weiteren toleriert dieses Verfahren leichte Beleuchtungsschwankungen, partielle Verdeckungen und externe Rotationen in einem Rahmen von 10-20 Grad.

## 3 Veröffentlichungen und Vorträge

### 3.1 Veröffentlichungen

1. Batliner, Anton; Hacker, Christian; Steidl, Stefan; Nöth, Elmar; Haas, Jürgen: From Emotion to Interaction: Lessons from Real Human-Machine-Dialogues. In: André, E.; Dybkjaer, L.; Minker, W.; Heisterkamp, P. (Hrsg.): *Affective Dialogue Systems, Proceedings of a Tutorial and Research Workshop (Affective Dialogue Systems ADS Kloster Irsee)*. Berlin : Springer, 2004, S. 1-12. (Lecture Notes in Artificial Intelligence Bd. 3068)
2. Batliner, Anton; Hacker, Christian; Steidl, Stefan; Nöth, Elmar; D'Arcy, S.; Russell, M.; Wong, M.: "You stupid tin box" - children interacting with the AIBO robot: A cross-linguistic emotional speech corpus.. In: ELRA (Hrsg.): *Proceedings of the 4th International Conference of Language Resources and Evaluation LREC 2004 (LREC Lisbon)*. 2004, S. 171-174.
3. Bouattour, S.; Heigl, B.; Hornegger, Joachim; Paulus, D.: Intensity-Based 3D-Reconstruction of Non-rigid Moving Stenosis from Many Angiographies. In: Tolxdorff, T.; Braun, J.; Handels, H.; Horsch, A.; Meinzer, H.-P. (Hrsg.): *Bildverarbeitung für die Medizin 2004 (Bildverarbeitung für die Medizin 2004)*. Berlin : Springer, 2004, S. 405-409.
4. Chrástek, Radim; Skokan, M.; Kubecka, L.; Wolf, Matthias; Donath, K.; Jan, J.; Michelson, G.; Niemann, Heinrich: Multimodal retinal image registration for optic disk segmentation. In: *Methods of Information in Medicine* 43 (2004), Nr. 4, S. 336-342
5. Cincarek, Tobias; Gruhn, Rainer; Hacker, Christian; Nöth, Elmar; Nakamura, Satoshi: Pronunciation scoring and extraction of mispronounced words for non-native speech. In: *Acoustical Society of Japan (Veranst.): Proceedings of the Acoustical Society of Japan (Acoustical Society of Japan Okinawa)*. 2004, S. 165-166.
6. Deutsch, Benjamin; Deinzer, Frank; Zobel, Matthias; Denzler, Joachim: Active Sensing Strategies for Robotic Platforms, with an Application in Vision-Based Gripping. In: Araújo, H.; Vieira, A.; Braz, J.; Encarnação, B.; Carvalho, M. (Hrsg.): *INSTICC (Veranst.): Proceedings of the 1st International Conference on Informatics in Control, Automation and Robotics (1st International Conference on Informatics in Control, Automation and Robotics Setúbal)*. Bd. 2. Setúbal, : INSTICC Press, 2004, S. 169-176.
7. Deutsch, Benjamin; Scholz, Ingo; Gräßl, Christoph; Niemann, Heinrich: Extending Light Fields using Object Tracking Techniques. In: Girod, B.; Magnor, M.; Seidel, H.-P. (Hrsg.): *Vision, Modeling, and Visualization 2004 (Vision, Modeling, and Visualization 2004 Stanford, USA)*. Berlin, Amsterdam : Aka / IOS Press, 2004, S. 109-116.
8. Deutsch, Benjamin; Zobel, Matthias; Denzler, Joachim; Niemann, Heinrich: Multi-Step Entropy Based Sensor Control for Visual Object Tracking. In: Rasmussen, C.E.; Bülthoff, H.H.; Giese, M.A.; Schölkopf, B. (Hrsg.): *Pattern Recognition, 26th DAGM Symposium (26th DAGM Symposium Tübingen)*. Berlin, Heidelberg, New York : Springer, 2004, S. 359-366. (Lecture Notes in Computer Science Bd. 3175)
9. Deventer, Rainer; Niemann, Heinrich; Celeghini, M.: Control of a Hydroforming Press with Bayesian Networks. In: Mántaras, R.L.; Saitta, L. (Hrsg.): *ECAI 2004 (ECAI 2004 Valencia)*. Amsterdam : IOS Press, 2004, S. 251-255.
10. Deventer, Rainer: Modeling and Control of Static and Dynamic Systems with Bayesian

Networks. Berlin : Logos, 2004

11. Grzegorzek, Marcin; Scholz, Ingo; Reinhold, Michael; Niemann, Heinrich: Fast training for object recognition with structure-from-motion. In: Geppener, V.V.; Gurevich, I.B.; Ivanova, S.E.; Nemirko, A.P.; Niemann, H.; Puzankov, D.V.; Trusova, Yu.O.; Zhuravlev, Yu.I. (Hrsg.): 7th International Conference on Pattern Recognition and Image Analysis 2004: New Information Technologies (7th International Conference on Pattern Recognition and Image Analysis 2004: New Information Technologies St. Petersburg, Russia,). St. Petersburg : SPbETU, 2004, S. 231-234.
12. Grzegorzek, Marcin; Pasumarthy, Kailash N.; Reinhold, Michael; Niemann, Heinrich: Statistical object recognition for multi-object scenes with heterogeneous background. In: Chanda, B.; Chandran, S.; Davis, L. (Hrsg.): 4th Indian Conference on Computer Vision, Graphics and Image Processing (4th Indian Conference on Computer Vision, Graphics and Image Processing Kolkata). Kolkata : Allied Publishers Private Limited, 2004, S. 222-227.
13. Gräßl, Christoph; Zinßer, Timo; Niemann, Heinrich: A Probabilistic Model-Based Template Matching Approach for Robust Object Tracking in Real-Time. In: Girod, B.; Magnor, M.; Seidel, H.-P. (Hrsg.): Vision, Modeling, and Visualization 2004 (Vision, Modeling, and Visualization 2004 Stanford, USA). Berlin, Amsterdam : Aka / IOS Press, 2004, S. 81-88.
14. Gräßl, Christoph; Zinßer, Timo; Niemann, Heinrich: Efficient Hyperplane Tracking by Intelligent Region Selection. In: IEEE Computer Society (Veranst.): 6th IEEE Southwest Symposium on Image Analysis and Interpretation (6th IEEE Southwest Symposium on Image Analysis and Interpretation Lake Tahoe, USA). Los Alamitos, USA : IEEE Computer Society Press, 2004, S. 51-55.
15. Gräßl, Christoph; Deinzer, Frank; Mattern, Frank; Niemann, Heinrich: Improving Statistical Object Recognition Approaches by a Parameterization of Normal Distributions. In: Pattern Recognition and Image Analysis 14 (2004), Nr. 2, S. 222-230
16. Haderlein, Tino; Steidl, Stefan; Nöth, Elmar; Rosanowski, F.; Schuster, M.: Automatic Recognition and Evaluation of Tracheoesophageal Speech. In: Sojka, P.; Kopecek, I.; Pala, K. (Hrsg.): Proc. Text, Speech and Dialogue; 7th International Conference, TSD 2004 (Proc. Text, Speech and Dialogue; 7th International Conference, TSD 2004 Brno). Berlin, Heidelberg : Springer, 2004, S. 331-338. (Lecture Notes in Artificial Intelligence Bd. 3206)
17. Krüger, S.; Vogt, Florian; Hohenberger, W.; Paulus, Dietrich; Niemann, Heinrich; Schick, C.H.: Evaluation of Computer-assisted Image Enhancement in Minimal Invasive Endoscopic Surgery. In: Methods of Information in Medicine 43 (2004), S. 362-366
18. Levit, Michael; Haffner, P.; Gorin, A.L.; Alshawi, A.L.; Nöth, Elmar: Aspects of Named Entity Processing. In: Kim, S.H.; Youn, D.H. (Hrsg.): Interspeech 2004 ICSLP, 8th International Conference on Spoken Language Processing, Jeju Island, Korea, Proceedings (Interspeech 2004 ICSLP Jeju Island, South Korea). 2004, S. 672-675.
19. Mack, M.; Hornegger, Joachim; Paulus, D.; Galant, A.; Böhm, S.: Echtzeit-Röntgenbildverarbeitung mit Standardhardware. In: Tolxdorff, T.; Braun, J.; Handels, H.; Horsch, A.; Meinzer, H.-P. (Hrsg.): Bildverarbeitung für die Medizin 2004 (Bildverarbeitung für die Medizin 2004). Berlin : Springer, 2004, S. 395-399.
20. Maier, T.; Benz, M.; Schön, N.; Nkenke, E.; Neukam, F.W.; Vogt, Florian; Häusler, G.:

- Automatic coarse registration of 3D surface data in oral and maxillofacial surgery.  
In: Buzug, T.M.; Lueth, T.C. (Hrsg.): Medical Robotics, Navigation and Visualization (MRNV) 2004 (MRNV 2004 Aachen). Singapore : World Scientific, 2004, S. 51-58.
21. Niemann, Heinrich; Denzler, Joachim; Heigl, Benno; Vogt, Florian; Schick, C.; Krüger, S.; Hohenberger, W.: Image- Based Modeling and its Application in Image Processing. In: Pattern Recognition and Image Analysis 14 (2004), Nr. 2, S. 184-189
  22. Nöth, Elmar; Horndasch, Axel; Gallwitz, F.; Haas, J.: Experiences with Commercial Telephone-based Dialogue Systems. In: it - Information Technology 46 (2004), Nr. 6, S. 315-321
  23. Prümmer, Marcus; Hornegger, Joachim; Schnörr, C.: Computergestützte Gefäßanalyse für die interventionelle Anwendung. In: Tolxdorff, T.; Braun, J.; Handels, H.; Horsch, A.; Meinzer, H.-P. (Hrsg.): Bildverarbeitung für die Medizin 2004 (Bildverarbeitung für die Medizin 2004). Berlin : Springer, 2004, S. 314-318.
  24. Rohlfing, T.; Denzler, Joachim; Russakoff, D.; Gräßl, Christoph; Maurer, C.: Markerless Real-Time Target Region Tracking: Application to Frameless Stereotactic Radiosurgery. In: Girod, B.; Magnor, M.; Seidel, H.-P. (Hrsg.): Vision, Modeling, and Visualization 2004 (Vision, Modeling, and Visualization 2004 Stanford, USA). Berlin, Amsterdam : Aka / IOS Press, 2004, S. 5-12.
  25. Schmidt, Jochen; Vogt, Florian; Niemann, Heinrich: Vector Quantization Based Data Selection for Hand- Eye Calibration. In: Girod, B.; Magnor, M.; Seidel, H.-P. (Hrsg.): Vision, Modeling, and Visualization 2004 (Vision, Modeling, and Visualization 2004 Stanford, USA). Berlin, Amsterdam : Aka / IOS Press, 2004, S. 21-28.
  26. Scholz, Ingo; Denzler, Joachim; Niemann, Heinrich: Calibration of Real Scenes for the Reconstruction of Dynamic Light Fields. In: IEICE Transactions on Information & Systems E87-D (2004), Nr. 1, S. 42-49
  27. Scholz, Ingo; Niemann, Heinrich: Globally Consistent 3-D Reconstruction by Utilizing Loops in Camera Movement. In: Rasmussen, C.E.; Bühlhoff, H.H.; Giese, M.A.; Schölkopf, B. (Hrsg.): Pattern Recognition, 26th DAGM Symposium (26th DAGM Symposium Tübingen). Berlin, Heidelberg, New York : Springer, 2004, S. 471-479. (Lecture Notes in Computer Science Bd. 3175)
  28. Stapor, K.; Pawlaczyk, L.; Chrástek, Radim; Michelson, G.: Automatic Detection Of Glaucomatous Changes Using Adaptive Thresholding And Neuronal Networks. In: Bubak, M.; Albada, G.D. van; Sloot, P. M. A.; Dongarra, J. J. (Hrsg.): Proceedings of 4th International Conference on Computational Science, Computational Science - ICCS 2004 (Computational Science - ICCS 2004 Kraków). Berlin, Heidelberg, New York : Springer, 2004, S. 49-55. (LNCS Bd. 3039)
  29. Stapor, K.; Switonski, A.; Chrástek, Radim; Michelson, G.: Automatic Recognition Of Glaucoma Using Morphological Operations And Neuronal Networks. In: Jan, J.; Kozumplik, J.; Provaznik, I. (Hrsg.): Proceedings of 17th Biennial International EURASIP Conference BIOSIGNAL 2004, Analysis of biomedical signals and images (BIOSIGNAL 2004 Brno). Brno : Vutium Press, 2004, S. 278-280.
  30. Stapor, K.; Pawlaczyk, L.; Chrástek, Radim; Niemann, Heinrich; Michelson, G.: Automatic Segmentation and Classification of Fundus Eye Images for Glaucoma Diagnosis. In: Tolxdorff, T.; Braun, J.; Handels, H.; Horsch, A.; Meinzer, H.-P. (Hrsg.): Bildverarbeitung für die Medizin 2004: Algorithmen - Systeme - Anwendungen (Bildverarbei-

- tung für die Medizin 2004: Algorithmen - Systeme - Anwendungen Berlin). Berlin, Heidelberg, New York : Springer, 2004, S. 65-69. (Informatik aktuell)
31. Stapor, K.; Switonski, A.; Chrástek, Radim; Michelson, Georg: Segmentation of Fundus Eye Images Using Methods of Mathematical Morphology for Glaucoma Diagnosis. In: Bubak, M.; Albada, G.D. van; Sloot, P. M. A.; Dongarra, J. J. (Hrsg.): Proceedings of 4th International Conference on Computational Science, Computational Science - ICCS 2004 (Computational Science - ICCS 2004 Kraków). Berlin, Heidelberg, New York : Springer, 2004, S. 41-48. (LNCS Bd. 3039)
  32. Steidl, Stefan; Stemmer, Georg; Hacker, Christian; Nöth, Elmar: Adaption in the Pronunciation Space for Nonnative Speech Recognition. In: Kim, S. H.; Youn, D. H. (Hrsg.): Interspeech 2004 ICSLP, 8th International Conference on Spoken Language Processing, Jeju Island, Korea, Proceedings (Interspeech 2004 Jeju Island, Korea). 2004, S. 318-321.
  33. Steidl, Stefan; Ruff, C.; Batliner, Anton; Nöth, Elmar; Haas, Jürgen: Looking at the Last two Turns, I'd Say this Dialogue is Doomed - Measuring Dialogue Success. In: Sojka, P.; IKopecek, I.; Pala, K. (Hrsg.): Text, Speech and Dialogue, 7th International Conference, TSD 2004, Czech Republic, September 8-11, 2004, Proceedings (TSD 2004 Brno). Berlin, Heidelberg : Springer, 2004, S. 629-636. (Lecture Notes in Artificial Intelligence Bd. 3206)
  34. Streit, M.; Batliner, Anton; Portele, T.: Cognitive-Model-Based Interpretation of Emotions in a Multi-modal Dialog System. In: André, E.; Dybkaer, L.; Minker, W.; Heisterkamp, P. (Hrsg.): Affective Dialogue Systems, Proceedings of a Tutorial and Research Workshop (Affective Dialogue Systems Kloster Irsee). Berlin : Springer, 2004, S. 65-76. (Lecture Notes in Artificial Intelligence Bd. 3068)
  35. Vogt, Florian; Krüger, S.; Zinßer, Timo; Maier, T.; Niemann, Heinrich; Hohenberger, W.; Schick, C.H.: Fusion von Lichtfeldern und CT-Daten für minimal-invasive Operationen. In: Tolxdorff, T.; Braun, J.; Handels, H.; Horsch, A.; Meinzer, H.-P. (Hrsg.): 8. Workshop Bildverarbeitung für die Medizin (8. Workshop Bildverarbeitung für die Medizin Erlangen). Berlin, Heidelberg, New York : Springer, 2004, S. 309-313.
  36. Vogt, Florian; Krüger, S.; Schmidt, Jochen; Paulus, Dietrich; Niemann, Heinrich; Hohenberger, W.; Schick, C.H.: Light Fields for Minimal Invasive Surgery Using an Endoscope Positioning Robot. In: Methods of Information in Medicine 43 (2004), Nr. 4, S. 403-408
  37. Weber, S.; Schüle, Th.; Schnörr, C.; Hornegger, Joachim: A Linear Programming Approach to Limited Angle 3D. In: Methods of Information in Medicine 43 (2004), Nr. 4, S. 320-326
  38. Weber, S.; Schüle, Th.; Schnörr, C.; Hornegger, Joachim: Binary Tomography by Iterating Linear Programs from Noisy Projections. In: University of Auckland (Veranst.): Proceedings of International Workshop on Combinatorial Image Analysis (IWCIA) (International Workshop on Combinatorial Image Analysis (IWCIA) Auckland). 2004, S. 38-51.
  39. Wenhardt, Stefan; Denzler, Joachim; Niemann, Heinrich: On Minimizing Errors in 3-D-Reconstruction for Stereo Camera Systems. In: Geppener, V.V.; Gurevich, I.B.; Ivanova, S.E.; Nemirko, A.P.; Niemann, Heinrich; Puzankov, D.V.; Trusova, Yu.O.; Zhuravlev, Yu.I. (Hrsg.): 7th International Conference on Pattern Recognition and Image

Analysis 2004: New Information Technologies (7th International Conference on Pattern Recognition and Image Analysis 2004: New Information Technologies St. Petersburg, Russia). St. Petersburg : SPbETU, 2004, S. 562-565.

40. Zinßer, Timo; Gräßl, Christoph; Niemann, Heinrich: Efficient Feature Tracking for Long Video Sequences. In: Rasmussen, C. E.; Bülthoff, H. H.; Giese, M. A.; Schölkopf, B. (Hrsg.): Pattern Recognition, 26th DAGM Symposium (Pattern Recognition, 26th DAGM Symposium Tübingen). Berlin, Heidelberg, New York : Springer, 2004, S. 326-333. (Lecture Notes in Computer Science Bd. 3175)

## 3.2 Vorträge

1. Batliner, Anton: "You stupid tin box" - children interacting with the AIBO robot: A cross-linguistic emotional speech corpus. Vortrag: 4th International Conference of Language Resources and Evaluation LREC 2004, Lisbon, 25.5..2004
2. Batliner, Anton: Automatic classification of emotional states: purpose, possibilities, prospects. Vortrag: Slovenian Language Technologies Society, Information Society - IS 2004 Fourth Language Technologies Conference, October 13 - 14, 2004, Ljubljana, 13.10..2004
3. Batliner, Anton: Experience of the PF-STAR project. Vortrag: HUMAINE WP5 Workshop (Databases), Belfast, 1.12..2004
4. Batliner, Anton: From Emotion to Interaction: Lessons from Real Human-Machine Dialogues. Vortrag: ISCA Workshop on Affective Dialogue Systems, Kloster Irsee, 14.6..2004
5. Batliner, Anton: Position paper: Experience with labelling. Vortrag: HUMAINE WP5 Workshop (Databases), Belfast, 2.12..2004
6. Batliner, Anton: Speech and Emotion. Vortrag: PF-STAR Workshop, Regensburg, 22.4..2004
7. Batliner, Anton: The AIBO robot database. Vortrag: HUMAINE Summerschool, QUB, Belfast, 1.9..2004
8. Batliner, Anton: WP3: Speech and Emotion. Vortrag: PF-STAR Workshop, ITC-irst, Trento, 27.9..2004
9. Deutsch, Benjamin: Optimale Kameraaktionen für mehrere Zeitschritte in der aktiven Objektverfolgung. Vortrag: Arbeitskreis Optimierung des SFB 603, 16.12..2004
10. Deutsch, Benjamin: Teilprojekt B2: Optimierungsansatz für die Integration. Vortrag: Berichtskolloquium des SFB 603, 1.7..2004
11. Deutsch, Benjamin: Active Sensing Strategies for Robotic Platforms, with an Application in Vision-Based Gripping. Vortrag: 1st International Conference on Informatics in Control, Automation and Robotics, 26.8..2004
12. Deventer, Rainer: Control of a Hydroforming Press with Bayesian networks. Vortrag: European Conference on Artificial Intelligence, Valencia, 27.8..2004
13. Gräßl, Christoph: Robust Template Matching with Hyperplanes. Vortrag: Invited Talk, Stanford University, Departement of Neurosurgery, Standford, 7.4..2004
14. Grzegorzec, Marcin: Fast Training for Object Recognition with Structure-from-Moti-

- on.Vortrag: 7th International Conference on Pattern Recognition, St. Petersburg, 19.10..2004
15. Grzegorzec, Marcin: Wirkliche Robustheit des Systems fuer statistische Objekterkennung.Vortrag: Graduiertenkolleg 3D-Bildanalyse und -synthese, Universität Erlangen, Erlangen, 9.11..2004
  16. Hacker, Christian: Children's Speech.Vortrag: Pf-Star Workshop, ITC-irst, Trento, 27.9..2004
  17. Hacker, Christian: Children's Speech.Vortrag: Pf-Star Meeting, Regensberg, 22.4..2004
  18. Hacker, Christian: Welche Töne haben Kinder? Vortrag: Tag der offenen Tür, OHM-Gymnasium, Erlangen, 28.7..2004
  19. Haderlein, Tino: Automatic Recognition and Evaluation of Tracheoesophageal Speech.Vortrag: TSD 2004, Brno, 8.9..2004
  20. Haderlein, Tino: Automatische Sprachverarbeitung im medizinischen Umfeld am Lehrstuhl für Mustererkennung.Vortrag: Kieferklinik der Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, 26.7..2004
  21. Hornegger, Joachim: 3D Rekonstruktionsalgorithmen für die Computertomographie, Registrierung und Visualisierung medizinischer Daten.Vortrag: Workshop: Rechnerkonzepte und Beschleunigerhardware für die medizinische Bildrekonstruktion und -visualisierung, Siemens Medical Solutions, Erlangen, 25.10..2004
  22. Hornegger, Joachim: Ausgewählte Aspekte der Medizinische Bildverarbeitung.Vortrag: Informatik-Kolloquium an der Universität Tübingen, Tübingen, 16.1..2004
  23. Hornegger, Joachim: Bildregistrierung mit parametrischen und nicht-parametrischen Abbildungen.Vortrag: Institutskolloquium, Institut für Medizininformatik, Biometrie und Epidemiologie, Erlangen, 20.12..2004
  24. Hornegger, Joachim: Bildverarbeitung in der Medizin: Trends und Perspektiven.Vortrag: SoftMed - Innovationsforum Medizinische Bildverarbeitung, Erlangen, 1.3..2004
  25. Hornegger, Joachim: Excellence in Medical Engineering.Vortrag: IZMP, Erlangen, 30.6..2004
  26. Hornegger, Joachim: Grundlagen der Bildregistrierung.Vortrag: Kolloquium der Diagnostischen Radiologie, Erlangen, 9.12..2004
  27. Hornegger, Joachim: Informatik als Motor in der Medizintechnik.Vortrag: Tag der TechFak, Erlangen, 25.6..2004
  28. Hornegger, Joachim: Medizinische Bildverarbeitung an der FAU.Vortrag: Siemens Medical Solutions, Geschäftsgebiet AX, Forchheim, 12.7..2004
  29. Hornegger, Joachim: Medizinische Bildverarbeitung: Projekte und Visionen.Vortrag: Siemens Medical Solutions, Geschäftsgebiet SP, Erlangen, 26.7..2004
  30. Hornegger, Joachim: Non-rigid Fusion of Morphological and Functional Images Using Anatomical Fix Points and Contours: A New Approach to Overcome the Current Drawbacks of Retrospective Image Fusion.Vortrag: Radiological Society of North America (RSNA), Chicago, 29.11..2004
  31. Hornegger, Joachim: Plattformübergreifende Konzepte in der Medizintechnik.Vortrag: .NET Workshop, 3Soft, Schloss Atzelsberg, 28.10..2004



32. Hornegger, Joachim: Prinzipien und Anwendungen der Registrierung in der Medizinischen Bildverarbeitung. Vortrag: Societas physico-medica Erlangensis e.V., Erlangen, 17.11..2004
33. Hornegger, Joachim: Recent Advances in Rigid and Non-Rigid Image Registration. Vortrag: Siemens Corporate Research, Princeton, NJ, 3.8..2004
34. Hornegger, Joachim: Remarks on regularization methods in medical image processing. Vortrag: First Erlangen Ad-Hoc Registration Day, Erlangen, 11.8..2004
35. Hornegger, Joachim: Technischer Stand und Zukunftsperspektiven der Bildfusion. Vortrag: Qualitätszirkel Nuklearmedizin "Bildfusion", Erlangen, 19.7..2004
36. Hornegger, Joachim: Tumor Therapy Monitoring Based on Monomodal Image Fusion. Vortrag: Radiological Society of North America (RSNA), Chicago, 29.11..2004
37. Hornegger, Joachim: Über die Unverzichtbarkeit der Bildverarbeitung für die moderne Medizin. Vortrag: Antrittsvorlesung, Erlangen, 18.6..2004
38. Hornegger, Joachim: Unüberwachtes Lernen in der Bild- und Sprachverarbeitung. Vortrag: Informatik-Kolloquium außer der Reihe, Erlangen, 29.6..2004
39. Levit, Michael: Aspects of Named Entity Processing. Vortrag: Interspeech 2004 - ICS-LP, Jeju Island, Korea, 5.10..2004
40. Niemann, Heinrich: Towards Automated Diagnostic Evaluation of Retina Images. Vortrag: 7th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-7-2004), St. Petersburg, Russian Federation, 19.10..2004
41. Nöth, Elmar: A Phone Recognizer Helps to Recognize Words Better. Vortrag: ICASSP 2004, Montreal (Kanada), 19.5..2004
42. Nöth, Elmar: Automatische Sprach- und Emotionsverarbeitung am Lehrstuhl für Mustererkennung. Vortrag: BMW, München, 11.10..2004
43. Nöth, Elmar: Benutzerwahrnehmung. Vortrag: DFG-Symposium "Perspektiven kognitiver technischer Systeme", Aachen, 15.9..2004
44. Nöth, Elmar: Looking at the Last Two Turns, I'd Say this Dialogue is Doomed - Measuring Dialogue Success. Vortrag: TSD 2004, Brno, 9.9..2004
45. Nöth, Elmar: Multi-modal Affective User Behavior in Human-Machine-Communication. Vortrag: ISCA Workshop on Affective Dialogue Systems, Kloster Irsee, 15.6..2004
46. Nöth, Elmar: Sprachgesteuerte 3D-Angiographie - Die Software folgt dem Arzt aufs Wort. Vortrag: Design & Elektronik Entwicklerforum Sprachsteuerung, München, 5.10..2004
47. Nöth, Elmar: Voice Activated Analysis of Stenoses. Vortrag: Hipgraphics, Baltimore (USA), 15.5..2004
48. Nöth, Elmar: Voice Activated Analysis of Stenoses. Vortrag: TSD 2004, Brno, 9.9..2004
49. Scholz, Ingo: Optimierung von Tiefenkarten für Lichtfelder durch Rückkopplung aus der Synthese. Vortrag: Arbeitskreis Optimierung des SFB 603, 19.7..2004
50. Scholz, Ingo: Teilprojekt C2: Analyse, Codierung und Verarbeitung von Lichtfeldern. Vortrag: Berichtskolloquium des SFB 603, 2.7..2004

51. Vogt, Florian: Endoskopkalibrierung und 3-D Rekonstruktion. Vortrag: Fraunhofer-Institut f. Integrierte Schaltungen, 23.1..2004
52. Vogt, Florian: Fusion von Lichtfeldern und CT-Daten für minimal-invasive Operationen. Vortrag: Workshop Bildverarbeitung für die Medizin 2004, 11.3..2004
53. Vogt, Florian: Rechnergestützte Endoskopie. Vortrag: MeVis Bremen, 1.12..2004
54. Vogt, Florian: Teilprojekt B6: Rechnergestützte Endoskopie des Bauchraums. Vortrag: Berichtskolloquium des SFB 603, 2.7..2004

## **4 Studien- und Diplomarbeiten, Master Theses**

### **4.1 Studienarbeiten**

- Leipold, Markus: 3-D Rekonstruktion aus Bildfolgen mit differentiellen Verfahren
- Brehm, Joachim: Automatisches Greifen sich bewegender Objekte
- Dennerlein, Frank: Automatische Videoannotation von Tennisspielen
- Kovac, Mikulas: Detektion von Anomalien in mehrlagigen Gewebestrukturen
- Hofmann, Matthias: Extraktion von Schleimhautdeformationen aus Hochgeschwindigkeitsaufnahmen
- Ruff, Christine: Bestimmung des Dialog(schritt)erfolgs mit linguistischen Merkmalen
- Weinlein, Thomas: Methoden zur robusten Faktorisierung
- Rehak, Markus: Reduktion des Kreisartefakts in 3D-Computer-Tomographie Daten

### **4.2 Diplomarbeiten**

- Schlötzer, Stephan: Automatische Auswertung von 3D-Computer-Tomographie Daten
- Fritz, Mario: Categorization by Local Information Using Support Vector Machines
- Dennerlein, Silvia: Modellierung der Kinematik von Objekten im 3-D mit Methoden der geometrischen Algebra
- Kähler, Olaf: Probabilistische, adaptive Sensordatenfusion in der 3-D Objektverfolgung
- Cincarek, Tobias: Pronunciation Scoring for Non-Native Speech

### **4.3 Master Theses**

- Nadtoka, Nataliya: Modeling of the 3D Dynamic Mouth Appearance
- Wei, Ren: Vergleichende Promotorenanalyse von tumor-induzierten Genen

## **Professur für Medizinische Bildverarbeitung**

**Anschrift:** Martensstrasse 3, 91058 Erlangen

**Tel.:** +49.9131.85.27775

**Fax.:** +49.9131.303811

**E-Mail:** info@i5.informatik.uni-erlangen.de

### **Leitung:**

Prof. Dr.-Ing. Joachim Hornegger

### **Sekretariat:**

Iris Koppe

Kristina Müller

### **Wiss. Mitarbeiter**

Soufyane El Allali, M. Sc.

Jingfeng Han, M. Sc.

Dipl.-Ing. (FH) Stefan Hoppe

Dipl.-Inf. Florian Jäger

Dipl.-Inf. Marcus Prümmer

Dipl.-Inf. Holger Scherl

### **Externe Doktoranden:**

Dipl.-Inf. Volker Daum

Dipl.-Inf. Christoph Gütter

Dipl.-Inf. Klaus Kirchberg

## **1 Einführung**

Die Professur für Medizinische Bildverarbeitung wurde 2003 neu geschaffen und ist seit dem 1. Oktober 2003 mit Herrn Joachim Hornegger besetzt. Die Professur ist dem Lehrstuhl für Mustererkennung zugeordnet.

### **1.1 Forschungsschwerpunkte**

Die wissenschaftlichen Schwerpunkte der Arbeitsgruppe liegen auf dem Gebiet der interaktiven medizinischen Bildverarbeitung.

Derzeitige Arbeiten beschäftigen sich mit folgendem Themen:

- nicht-starre Registrierung multimodaler Bilddaten
- monomodale Bildfusion zur Verlaufskontrolle bei der Tumor Therapie
- Verfahren zur Schwächungskorrektur bei der SPECT-Rekonstruktion
- Rekonstruktion bewegter Objekte bei bekannter Projektionsgeometrie

- Berechnung und Visualisierung des Blutflusses in 3D-Angiogrammen
- Segmentierung von CT-Datensätzen
- schnelle Bildverarbeitung auf Standardgrafikkarten
- Diskrete Tomographie
- Sprachsteuerung interventioneller Werkzeuge

Zusätzlich zu den Aufgaben in Forschung und Lehre setzt sich die Arbeitsgruppe auch für die Intensivierung des Austausches zwischen Hochschule und Industrie ein. Dies schließt auch die enge Kooperation mit der Arbeitsgruppe Medizinische Bildverarbeitung des Fraunhofer Instituts für Integrierte Schaltungen ein.

## **1.2 Forschungsrelevante apparative Ausstattung**

Aufgrund der engen Kooperation der Arbeitsgruppe mit den Kliniken und der Industrie besteht Zugriff auf sämtliche Modalitäten, die in der modernen Medizin heute zum Einsatz kommen. Die verfügbare Entwicklungsumgebung erlaubt die schnelle Überführung der neu entwickelten Methoden in den klinischen Test.

## **1.3 Kooperationsbeziehungen**

Klinische Partner der FAU:

- Nuklearmedizinische Klinik (Prof. Kuwert)
- Institut für Diagnostische Radiologie (Prof. Bautz)
- Lehrstuhl für Innere Medizin I (Prof. Hahn)
- Lehrstuhl für Innere Medizin II (Prof. Daniel)

Klinische Partner der Julius-Maximilians-Univ. Würzburg:

- Abteilung für Neuroradiologie (Prof. Solymosi)

Universitäre Kooperationen:

- Lehrstuhl für Bildverarbeitung der Univ. Mannheim (Prof. Schnörr)
- Lehrstuhl für Diskrete Mathematik der Univ. Marburg (Prof. Welker)
- Lehrstuhl für Angewandte Mathematik der Univ. Duisburg (Prof. Rumpf)

Industriepartner:

- Siemens Medical Solutions
- Siemens Corporate Research
- General Electric
- HipGraphics Inc.
- Sympalog

## 2 Forschungsprojekte

### 2.1 3D Bildrekonstruktion aus Kegelstrahl-Projektionsdaten einer Abtastbahn bestehend aus Kreis und Linie

**Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger

**Beteiligte:**

Prof. Dr.-Ing. Joachim Hornegger, Frank Dennerlein, Alexander Katsevich, Günter Lauritsch

**Stichwörter:**

3D Rekonstruktion, Kegelstrahlgeometrie

**Laufzeit:** 1.7.2004 - 31.12.2004

**Förderer:**

Siemens Medical Solutions

**Mitwirkende Institutionen:**

Department of Mathematics, University of Central Florida

Die Kegelstrahlgeometrie spielt in der Computergraphie eine zunehmend wichtige Rolle, da heute Detektoren mit bis zu 1024 Zeilen für die Rekonstruktion eingesetzt werden. Im Rahmen dieses Forschungsprojektes sollen Rekonstruktionsverfahren analysiert werden, welche auf der Kegelstrahlgeometrie basieren.

Die Arbeitspakete umfassen dabei:

- (1) Prüfung verfügbarer Rekonstruktionsalgorithmen auf numerische Stabilität; dies schließt den Vergleich mit der herkömmlichen Feldkamp-Methode bei reiner Kreisabtastung ein.
- (2) Evaluierung der Algorithmen hinsichtlich ihres Potentials zur Vermeidung von Kegelstrahlartefakten, die durch die unvollständige Datennahme einer einfachen Kreisabtastung verursacht werden.
- (3) Klärung des Einflusses der Länge der benötigten Linienbahnkurve sowie Abtastrate entlang der Linie
- (4) Analyse der 3D-Rückprojektion unter Verwendung des Feldkamp-Algorithmus, in dem die Geometrie über Projektionsmatrizen gegeben ist.

Der experimentelle Teil soll die Frage beantworten, ob die direkte 3D--Rückprojektion der gefilterten Projektionsdaten, die entlang geneigter Filterlinien auf dem Detektor vorliegen, besser ist als das Rebinning der gefilterten Projektionsdaten, die entlang geneigter Filterlinien auf dem Detektor vorliegen.

## **2.2 Automatische Extraktion und Klassifikation des Blutgefäßbaumes in retinalen Bildern**

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger, Prof. Dr.-Ing. Heinrich Niemann

### **Beteiligte:**

Ing. Radim Chrastek, Soufyane El Allali, M. Sc.

### **Stichwörter:**

Netzhaut; Gefäßbaum; Segmentierung; Klassifikation; Kardiovaskuläre Erkrankungen; Diabetes; präventive Medizin

**Beginn:** 1.1.2004

**Förderer:** e-EyeCare

### **Mitwirkende Institutionen:**

Augenklinik mit Poliklinik Erlangen, e-EyeCare

### **Kontakt:**

Soufyane El Allali, M. Sc.

Tel.: +49.9131.85 28977, Fax: +49.9131.85.28990,

E-Mail: .soufyane@informatik.uni-erlangen.de

In digital fundus images, detection of the different anatomical structures is an essential task as the first step of an automatic pre-diagnosis procedure of patients. Ophthalmologists use digital fundus images extensively for diagnosing and treating their patients' eyes. These images reflect a patient's eye health. Therefore, diseases such as glaucoma and others can easily be identified by doctors. Yet, these images do not only give a status of the patient's eye health. They are also used as a defining preventative measure that indicates a patient's signs for risk of diabetes, cardiovascular diseases. One measure of these risks can be formulated by detecting the change of size of veins and arteries in retinal images. This is known as the AV-ratio (artery-vein ratio). The expansion of vessels in the body starts initially in small vessels such as the ones in the retina. Thus an early screening of the eye can detect the abnormalities in retina vessels and consequently preventing health problems before they worsen. This can be categorized under the grand theme of preventative medicine.

Retina images are constituted mainly of three major anatomical structures: vessels, categorized into veins and arteries, the optic disk, and the macula. The goal of this research project is to detect these structures automatically given a set of data images. The detection of the optic disk, which has been finished, serves for definition of the measurement zones. The current research aim to achieve complete automated vessel tree detection, and its classification into veins and arteries.

This project is also open to interested students in the field of medical image processing, specifically segmentation. Currently a senior student is undergoing his "Studienarbeit" under title: Atlas based retinal vessels segmentation. In this project the segmentation of the vessel tree is

investigated using atlas based techniques.

This project is jointly done in corporation and sponsoring from e-EyeCare, as part of the TalkingEyes project, an Erlangen based company which function is providing a medical management system, part of preventative medicine, through the data analysis of retina images. Also, close cooperation with doctors and physicians are substantially considered and taken. Prof. Dr. Georg Michelson is an example of such cooperation. Prof. Michelson is a member of the Department of Ophthalmology and Eye Hospital of the Friedrich-Alexander-University of Erlangen.

## **2.3 Entwicklung und Implementierung von Quantifizierungsalgorithmen zur Analyse von Stenosen in Computer-Tomographie-Datensätzen**

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger

### **Beteiligte:**

Dr. med. Michael Lell

Dipl.-Inf. Holger Scherl

### **Stichwörter:**

Computertomographie; 3D-Gefäßanalyse; Segmentierung; Arteria Carotis; Quantifizierung

**Laufzeit:** 1.4.2004 - 1.10.2004

In enger Kooperation mit Dr. Michael Lell vom radiologischen Institut der Universität Erlangen/Nürnberg wurden die Möglichkeiten der automatischen Quantifizierung von Blutgefäßen in Computertomographie-Datensätzen erforscht. Erst die Injektion von Kontrastmittel ermöglicht dem Arzt in den aufgenommenen Bildfolgen die Quantifizierung einer Gefäßverengung (Stenose) der Arteria Carotis Interna. Trotzdem ist es auch für den erfahrenen Mediziner immer noch sehr schwierig den Grad einer Stenose deterministisch zu bestimmen, weil die Öffnung der Arterie für den Blutfluss von Ablagerungen in den Gefäßwänden kaum unterscheidbar ist. Heutzutage gibt es keine zuverlässigen Verfahren, die den Grad einer Stenose aus einem CT-Volumendatensatz berechnen. Im klinischen Umfeld führt der behandelnde Arzt die Vermessung eines stenosierten Gefäßes manuell durch. Diese Vorgehensweise liefert in vielen Fällen nur Schätzwerte. Reproduzierbare Ergebnisse sind daher kaum möglich.

Es wurde ein semi-automatisches Verfahren für die Erkennung von Blutgefäßen entwickelt, das speziell Ablagerungen in der Gefäßwand aus den segmentierten Bereichen ausschloss. Hierbei kam ein Level Set basiertes Verfahren zum Einsatz, in dem nur zwei Startpunkte vor und nach dem stenosierten Bereich innerhalb des Gefäßes vom Mediziner ausgewählt werden mussten. Die Ergebnisse der neu entwickelten Methode wurden anhand von zehn realen Datensätzen ausgewertet und mit sechs verschiedenen Messungen von drei verschiedenen Ärzten verglichen. Ein Resultat der Methode für einen Patienten mit harten Plaqueablagerungen ist unter folgendem Link zu sehen: [scherl-carotis.jpg](http://www5.informatik.uni-erlangen.de/MEDIA/jb2004/scherl-carotis.jpg) (32K) (<http://www5.informatik.uni-erlangen.de/MEDIA/jb2004/scherl-carotis.jpg>).

Das linke Bild zeigt eine VRT-Ansicht (Volume Rendering Technique) des Datensatzes. In der Mitte ist die erzielte dreidimensionale Segmentierung abgebildet. Schließlich zeigt das rechte

Bild das Segmentierungsergebnis in einer Schnittebene durch das Volumen an der Stelle des stenosierte Bereichs (gepunkteter Bereich) und den vom Mediziner eingezeichneten minimalen Durchmesser des Gefäßes (Linie).

## **2.4 Funktionelle Evaluierung von Veränderungen der Atemwege bei COPD**

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger

### **Beteiligte:**

Prof. Dr.-Ing. Joachim Hornegger, Dipl.-Inf. Florian Jäger, Frank Trappe, Daniel Rinck

### **Stichwörter:**

3D-Segmentierung, Atemwegserkrankung, COPD

**Laufzeit:** 1.7.2004 - 31.12.2004

### **Förderer:**

Siemens Medical Solutions

**Mitwirkende Institutionen:** Medizinische Klinik I mit Poliklinik

### **Kontakt:**

Dipl.-Inf. Florian Jäger

Tel.: +49.9131.85.27894, Fax: +49.9131.303811,

E-Mail: jaeger@informatik.uni-erlangen.de

Die Diagnose der Chronischen Obstruktiven Lungenerkrankung (COPD) spielt eine wichtige Rolle im derzeitigen medizinischen Alltag. Dabei beschreibt COPD morphologische Veränderungen der Atemwege, wie z.B. Bronchialwandverdickung, und/oder des Lungengewebes. Für die klinische Erprobung neu entwickelter Medikamente wird ein einfach zu verwendendes Werkzeug benötigt, mit dessen Hilfe die Atemwege zuverlässig und mit hoher Genauigkeit segmentiert werden können. Im Rahmen dieses Projekts wurde ein derartiges Werkzeug entwickelt. Dabei gliedert sich die Verarbeitung der CT Thorax Datensätze in folgende Schritte:

- (1) Separation des Bronchialbaumes vom umgebenden Lungenparenchym
- (2) Skelettierung des Segmentierungsergebnisses und Erstellung eines 3D Modells
- (3) Extraktion eines Pfades aus dem 3D Modell
- (4) Vermessung der Bronchialwand entlang des Pfades und Vergleich mit früheren Messungen (Follow-Up)

Der Schwerpunkt des Projekts liegt auf der 3D Segmentierung des Bronchialbaumes. Hierfür wurde ein neuer Ansatz, die Formbasierte Segmentierung der Atemwege, entwickelt.



## 2.5 Klinische Evaluierungen von Algorithmen der medizinischen Bildverarbeitung

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger

### **Beteiligte:**

Dipl.-Inf. Marcus Prümmer

**Laufzeit:** 1.10.2003 - 31.12.2004

### **Mitwirkende Institutionen:**

Siemens Corporate Research HIP Graphics

### **Kontakt:**

Dipl.-Inf. Marcus Prümmer

Tel.: +49 9131.85.27826, Fax: +49.9131.303811,

E-Mail: pruemmer@informatik.uni-erlangen.de

### Template-Klasse für Registrierungsverfahren

Um die im Rahmen von Studien- und Diplomarbeiten implementierten Registrierungsalgorithmen klinisch evaluieren zu können wurde ein allgemeines Benutzerinterface für 2D-2D, 2D-3D und 3D-3D Registrierungen für die Software InSpace implementiert. Damit kann schnell und ohne zusätzlichen Aufwand für den Arzt eine Workstation in der Klinik aufgebaut werden auf der die verschiedenen Verfahren in einer "Look & Feel - Syngo" Umgebung evaluiert werden können. Für eine Diplomarbeit wurde ein InSpace-Plugin für die Evaluierung der Level-Set-Segmentierung der Carotis implementiert.

### Evaluierung der sprachgesteuerten Gefäßanalyse für die interventionelle Anwendung

Für eine klinische Evaluierung wurde ein Prototyp in Zusammenarbeit mit der Firma Sympalog und Siemens Medical Solutions realisiert, der es ermöglicht eine Stenosen-Selektion und Quantifizierung mittels Sprachkommandos durchzuführen. Ein mit einem Funkmikrofon ausgestatteter Arzt ist damit in der Lage eine Gefäßanalyse durchzuführen, ohne den OP-Tisch verlassen zu müssen. Der Prototyp wurde am Institut für Röntgendiagnostik der Universität Würzburg klinisch evaluiert.

Die Sprachsteuerung wurde von den in die Erprobung eingebundenen Ärzten insgesamt sehr positiv beurteilt. Besonders für die Bedienung von Geräten im sterilen Umfeld sehen sie durch die Sprachbedienung eine spürbare Erleichterung im beruflichen Alltag. Als möglicherweise problematisch für die Akzeptanz eines solchen Systems wird lediglich die Tatsache beurteilt, dass das System das Tragen eines Headsets verlangt.

## 2.6 Koronarangiografie - Rekonstruktion von Herzkranzgefäßen aus Aufnahmen herkömmlicher C-Bogen Systeme

### Projektleitung:

Prof. Dr.-Ing. Joachim Hornegger

### Beteiligte:

Dipl.-Inf. Marcus Prümmer

### Stichwörter:

Koronar Angiography; 3D-Rekonstruktion; SART; CAV; Sprachgesteuerte 3D-Gefäßanalyse; Nicht starre 2D-Registrierung

**Laufzeit:** 1.10.2003 - 31.10.2006

### Kontakt:

Dipl.-Inf. Marcus Prümmer

Tel.: +49 9131.85.27826, Fax: +49.9131.303811,

E-Mail: pruemmer@informatik.uni-erlangen.de

### Problemstellung

Nicht bewegte Gefäße wie beispielsweise das zerebrale Gefäßsystem können durch Anwendung der inversen Radontransformation dreidimensional rekonstruiert werden. Unter der Annahme, dass sich ein Patient während der Durchführung einer Angiographie nicht bewegt, kann aus der Bildfolge mit Verfahren wie der gefilterten Rückprojektion oder ART ein mit Kontrastmittel gefülltes Gefäßsystem mit einer für die Praxis ausreichender Bildqualität rekonstruiert werden. Dieses Projekt beschäftigt sich damit Rekonstruktionsalgorithmen zu entwickeln, die auch für bewegte Objekte wie beispielsweise das Herz eine für den klinischen Einsatz ausreichende Rekonstruktion ermöglichen.

### Stand der Forschung

Für die Evaluierung von Rekonstruktionen wurde ein Modell der Herzkranzgefäße bestehend aus mit Kontrastmittel gefüllten PVC Schläuchen erstellt. Von der Firma Siemens Medical Solution wurde ein realer Datensatz einer Koronarangiographie bereitgestellt. State-of-the-art Algorithmen wie Feldkamp, ART, SART und CAV zur Rekonstruktion aus Cone-Beam Projektionen setzen voraus, dass sich das zu rekonstruierende Objekt während der Aquisition nicht bewegt. Ist dies nicht gewährleistet entstehen starke Rekonstruktionsartefakte. Das Ergebnis ist für eine klinische Nutzung bisher noch nicht brauchbar. Für den realen Datensatz existiert eine EKG-Triggerung anhand derer sich die Bildsequenz in verschiedene Herzphasen einteilen lässt.

Erste Rekonstruktionsversuche wurden aus den Projektionen gleicher Herzphasen durchgeführt. Anhand des Modells und der realer Daten wurden verschiedene Registrierungsalgorithmen implementiert und darauf untersucht ob damit die Bewegung der Koronarien erfasst und kompensiert werden kann. Die Idee dabei ist zu den wenigen Projektionen aus einer bestimmten Herzphase zusätzliche Projektionen aus anderen Herzphasen nach Kompensation der Bewegung hinzuzufügen um das Rekonstruktionsergebnis zu verbessern. Um verschiedene Rekon-

truktionsalgorithmen vergleichen zu können wurde außer Feldkamp noch SART und CAV implementiert, sowie verschiedene Tools für Projektion und Rückprojektion von 3D Volummen.

Für Registrierungsalgorithmen, die auf segmentierten Objekten arbeiten, wurde im Rahmen einer Studienarbeit eine Level-Set-Segmentierung der Koronarien auf realen Daten untersucht. Da Algebraische Rekonstruktionsverfahren aus zeitkritischen Gründen in der Praxis noch nicht eingesetzt werden können wird derzeit im Rahmen einer Studienarbeit SART speziell auf Grafikkartenhardware re-implementiert und optimiert. Alternativ wird auch im Rahmen einer Studienarbeit ein Variationsansatz für die 3D-Rekonstruktion implementiert.

## **2.7 Multimodale Bildregistrierung mit Punkt- und Konturkorrespondenzen**

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger, Dr. med. Wolfgang Römer,  
Prof. Dr. med. Torsten Kuwert

### **Beteiligte:**

Prof. Dr.-Ing. Joachim Hornegger, Prof. Dr. med. Torsten Kuwert,  
Dr. med. Wolfgang Römer, Prof. Dr. med. Werner Bautz, Jingfen Han

### **Stichwörter:**

Bidregistrierung, nicht-starre Registrierung, Level Set Segmentierung

**Laufzeit:** 1.1.2004 - 21.12.2004

### **Mitwirkende Institutionen:**

Institut für Diagnostische Radiologie, Nuklearmedizinische Klinik

## **2.8 Registrierung morphologischer und funktioneller Volumendatensätze**

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger

### **Beteiligte:**

Jingfeng Han, M. Sc.

### **Stichwörter:**

multimodal image registration; mage registration; non-rigid registration; image fusion; CT/PET image fusion

**Laufzeit:** 1.11.2004 - 31.12.2006

### **Kontakt:**

Jingfeng Han, M. Sc.  
Tel.: +49.9131.85.27826, Fax: +49.913.303811,  
E-Mail: jingfeng@informatik.uni-erlangen.de

## Mono-modal Image Registration

Quantification of tumor response to therapy is still an open problem and there exist no software packages that allow the computation, visualization, and quantification of differences in volume data sets. We developed a new algorithm using mono-modal image registration to solve the described problem. According to the idea of digital subtraction angiography, we register CT data acquired at different time points and compute the subtracted volume. In the ideal case, the difference volume exhibits the change of the tumor over time. The voxels of this “difference tumor” can be identified by a binary segmentation algorithm. These segmentation results allow the 3D quantification of the tumor. The registration method for mono-modal image fusion is intensity based and requires no anatomical or fiducial markers. Due to the fact that the volume data originate from the same modality, we can apply a simple similarity measure based on the sum of squared differences. Non-rigid soft tissue deformation, however, will imply a bias to the registration result, because local deformations cannot be modeled by rigid transforms. Anatomical substructures, like bones or teeth, are not affected by these deformations, but follow a rigid transform. This important observation is incorporated in the proposed registration algorithm. The selection of anatomical substructure is done by manual interaction of medical experts adjusting the transfer function of the volume rendering software. The parameters of the transfer function are used to identify the voxels that are considered for registration. A rigid transform is estimated by a quaternion gradient descent algorithm based on the intensity values of the specified tissue classes. Commonly used voxel intensity measures are adjusted to the modified registration algorithm.

Our approach to tumor therapy monitoring provides a powerful visualization and measurement tool. Compared to conventional 2D measurement techniques the proposed method yields deterministic and reproduceable quantitative data. The registration and visualization based on subtracted volumes was incorporated into InSpace3D. Clinicians were used to the user interface and, therefore, an efficient evaluation of the developed algorithms was possible in the clinical environment. The research on mono-modal image registration is done in a joint project with the Nuclear Medicine Clinics (Prof. T. Kuwert, Dr. W. Römer) and the Institute of Diagnostic Radiology (Prof. W. Bautz).

## Multi-modal Image Registration

Retrospective fusion of multimodal data sets is a major task in medical imaging. Commercially available software solutions are mostly limited to rigid transforms. Due to differences in patient position, registration results based on estimates of rigid transforms are often insufficient. In these cases, deformations are required to achieve the needed quality. In a joint research project C10 with the Nuclear Medicine Clinics (Prof. T. Kuwert, Dr. W. Römer) and the Institute of Diagnostic Radiology (Prof. W. Bautz) financed by the SFB 603 a new method for non-rigid registration was developed. In addition to existing approaches two important anatomical feature correspondences are incorporated into non-rigid registration algorithms: corresponding points (landmarks) and contours. These fix points and contour lines are selected interactively on both data sets. The new registration algorithm guarantees the one-to-one mapping of the selected points and possibly deformed contours. This approach allows for the explicit use of expert knowledge and experience to navigate the registration into desired image alignment. By visual inspection, the hybrid registration scheme generates more anatomically reasonable fusion results. The satisfactory performance of feature based multimodality registration has been verified by the software based registration experiments on the target CT and PET image data. (Figure 1 ([http://www5.informatik.uni-erlangen.de/MEDIA/jb2004/Max\\_MultiReg\\_2.jpg](http://www5.informatik.uni-erlangen.de/MEDIA/jb2004/Max_MultiReg_2.jpg)), Figure 2 ([http://www5.informatik.uni-erlangen.de/MEDIA/jb2004/Max\\_MultiReg\\_1.jpg](http://www5.informatik.uni-erlangen.de/MEDIA/jb2004/Max_MultiReg_1.jpg)))

The hybrid registration algorithm and the related level set segmentation algorithm are implemented in C++ library, which is integrated in the medical image processing toolkit, InSpace3D plugin framework.

## **2.9 Segmentierungsalgorithmen für Radiographieaufnahmen**

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger

### **Beteiligte:**

Prof. Dr.-Ing. Joachim Hornegger, Christine Ruff, Stefan Böhm

### **Stichwörter:**

Diagnostische Radiographie, Segmentierung, Direktstrahlungsartefakte, Implantat-erkennung

**Laufzeit:** 15.6.2004 - 15.12.2004

### **Förderer:**

Siemens Medical Solutions

### **Kontakt:**

Prof. Dr.-Ing. Joachim Hornegger

Tel.: +49.9131.85.27883, Fax: +49.9131.303811,

E-Mail: Joachim.Hornegger@informatik.uni-erlangen.de

Die medizinische Bildgebung mit digitalen Flachbilddetektoren ersetzt zunehmend die konventionelle Filmfolientechnologie in der diagnostischen Radiographie. Ein wesentlicher Nachteil der konventionellen Radiographie können Fehlbelichtungen sein, welche die Bilder unbrauchbar machen und zu Wiederholungsaufnahmen führen. Bei der digitalen Bildgebung kann man mit Hilfe der Bildverarbeitung Fehlbelichtungen korrigieren, um die angesprochenen Wiederholungsaufnahmen zu vermeiden. Eine wesentliche Rolle hierbei spielt die Bildsignalnormierung, die gewissen Störeinflüssen unterworfen ist.

Im Rahmen dieses Forschungsprojektes soll ein Segmentierungsverfahren entwickelt werden, mit dem eine hochzuverlässige Erkennung der drei wichtigsten Störeinflüsse möglich ist: Blendstellung, Direktstrahlung und Implantate.

## 2.10 Sprachgesteuerte Gefäßanalyse

### **Projektleitung:**

Prof. Dr.-Ing. Joachim Hornegger

### **Beteiligte:**

Dipl.-Inf. Marcus Prümmer, Dr. J. Regn, Siemens Medical Solutions,  
Dr. F. Gallwitz, Sympalog, Prof. Solzmosy, Univ. Würzburg

### **Stichwörter:**

Quantifizierung; Angiogramm; Stenose; Sprachsteuerung

**Laufzeit:** 1.10.2003 - 30.9.2004

### **Förderer:**

Siemens Medical Solutions, Sympalog

### **Kontakt:**

Prof. Dr.-Ing. Joachim Hornegger  
Tel.: +49.9131.85.27883, Fax: +49.9131.303811,  
E-Mail: Joachim.Hornegger@informatik.uni-erlangen.de

Eine Intervention ist ein medizinischer Eingriff zur Behandlung pathologischer Gefäße. Zur Behandlung einer Stenose wird ein Stent mittels eines Katheters beispielsweise an der Hüfte eingeführt und durch das Gefäßsystem zum krankhaften Gefäß durchgeschoben. Um den Stent möglichst präzise auszuwählen und schnell zu platzieren, ist es erforderlich eine Gefäßkarte zu erstellen und eine Quantifizierung des pathologischen Gefäßabschnittes vorzunehmen. Für die richtige Wahl des Stents wird der Gefäßdurchmesserlauf und die Länge des stenotisierten Gefäßabschnittes benötigt. Die Entwicklung von Algorithmen zur quantitativen Auswertung einer Gefäßaufnahme sowie zur computerunterstützten Gefäßnavigation konzentriert sich zunehmend auf deren Einsatz in 3-D-Angiographie Systemen. Da im Verlauf einer interventionellen Behandlung der Arzt den OP-Tisch und somit den sterilen Bereich verlassen muss um an einer Workstation die Gefäßanalyse durchzuführen, stellt eine sprachgesteuerte Gefäßanalyse im interventionellen Umfeld eine wertvolle Bereicherung dar. Aufbauend auf eine Arbeit zur quantitativen Analyse von Volumendaten wurde ein bestehendes SW-Paket (Syngo-Applikation, Siemens Medical Solutions, Forchheim) mit einem Spracherkennung der Firma Sympalog ([www.sympalog.de](http://www.sympalog.de)) ausgestattet. Für eine klinische Evaluierung wurde ein Prototyp realisiert, der es ermöglicht eine Stenosen-Selektion und Quantifizierung mittels Sprachkommandos durchzuführen. Ein mit einem Funkmikrofon ausgestatteter Arzt ist damit in der Lage eine Gefäßanalyse durchzuführen, ohne den OP-Tisch verlassen zu müssen. Funktionell kann eine semi-automatische Schwellwertsegmentierung (Sprachbefehl: Erhöhe bitte den Schwellwert um 32), die Orientierung und Größe eines 3D-Volumens (Sprachbefehl: Vergrößere/verkleinere das Volumen oder Rotiere das Volumen nach unten) sowie die Selektion einer Stenose mittels eines computerunterstützten 3D-Zeigers via Sprachbefehl durchgeführt werden. Eine Sprachsteuerung bietet den Vorteil, dass alle Steuerbefehle in einer Kommandoebene angeordnet und somit umständliche hierarchische Menüs zu vermeiden werden können. Der entwickelte Prototyp wurde in eine Siemens Leonardo Workstation integriert und erlaubt eine

Sprachsteuerung eines typischen Workflows einer Gefäßanalyse. Der Prototyp wird derzeit am Institut für Röntgendiagnostik der Universität Würzburg klinisch evaluiert und weiterhin untersucht in welchem Umfang die sprachgesteuerte Gefäßanalyse verbessert werden kann.

### **3 Studien- und Diplomarbeiten, Studienarbeiten**

- Hahn, Dieter: Automatic Intra-Modality Rigid 3D/3D Registration using Voxel Similarities

#### **3.1 Diplomarbeiten**

- Scherl, Holger: Entwicklung und Implementierung von Quantifizierungsalgorithmen zur Analyse von Stenosen in Computer-Tomographie-Datensätzen
- Jäger, Florian: Functional Evaluation of Airway Thickening in the Diagnostic Context of Chronic Obstructive Pulmonary Disease
- Weiberle, C. T.: Implementierung und Untersuchung von Verfahren zum objektiven Vergleich von Segmentierungs- und Klassifikationsergebnissen
- Ruff, Christine: Segmentierungsalgorithmen für Radiographieaufnahmen
- Fischer, Frank: Untersuchungen zur Kodierung von Angiographie-Bilddaten

### **4 Bacheor Theses und Master Theses**

#### **4.1 Bacheor Theses**

- Keck, Benjamin: Cone Beam Reconstruction Using Fixed Point Arithmetics
- Keck, Benjamin: Filtered Backprojection using Fixed Point Arithmetics

#### **4.2 Master Theses**

- Schmolla, Frank: A Method to Automatically Assign Optimal Image Processing Operators in Automatic X-Ray Inspection Systems
- Wang, Jing: Blood Flow Visualization in 3-D Angiograms
- Han, Jingfeng: Hybrid Non-Rigid Image Registration

