

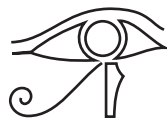
Echtzeit-Akquisition bildbasierter 3D Modelle zur Objekterkennung (PMDLumi)



Prof. Dr.-Ing.
Andreas Kolb

Zentrum für Sensordaten (ZESS)
und
Lehrstuhl für
Computergraphik und
Multimediasysteme

Universität Siegen
Hölderlinstr. 3
57068 Siegen



Prof. Dr.-Ing.
Klaus-Dieter Kuhnert

Zentrum für Sensordaten (ZESS)
und
Institut für
Echtzeit-Lernsysteme

Universität Siegen
Hölderlinstr. 3
57068 Siegen



Dr.-Ing.
Christof Rezk-Salama

Lehrstuhl für
Computergraphik und
Multimediasysteme

Universität Siegen
Hölderlinstr. 3
57068 Siegen

Antragspaket *Dynamisches 3D Sehen*

Prof. Dr.-Ing.
Andreas Kolb

Zentrum für
Sensordaten
(ZESS)

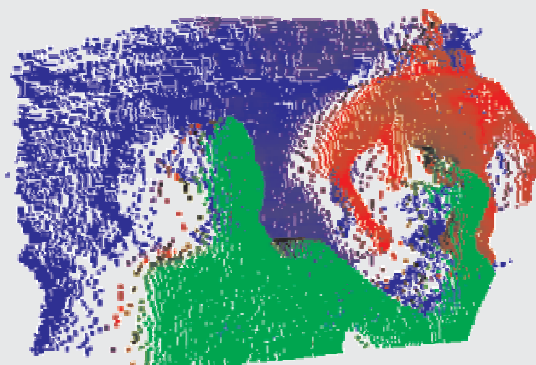
Universität Siegen
Hölderlinstr. 3
57068 Siegen



Prof. Dr.-Ing. habil.
Otmar Loffeld

Zentrum für
Sensordaten
(ZESS)

Universität Siegen
Paul-Bonatz-Str. 9-11
57068 Siegen



Inhaltsverzeichnis

1	Allgemeine Angaben	3
1.1	Antragsteller	3
1.2	Thema	4
1.3	Kennwort	4
1.4	Fachgebiet und Arbeitsrichtung	4
1.5	Voraussichtliche Gesamtdauer	4
1.6	Antragszeitraum	4
1.7	Gewünschter Beginn der Förderung	4
1.8	Zusammenfassung	4
2	Stand der Forschung, eigene Vorarbeiten	5
2.1	Stand der Forschung: Bildbasierte Modellrekonstruktion	5
2.1.1	Bild- und Videobasiertes Rendering	5
2.1.2	3D Modell-Rekonstruktion	6
2.1.3	Lumigraph-Modelle	6
2.1.4	Datenstrukturen	7
2.2	Stand der Forschung: 3D-Szenenerkennung	7
2.2.1	Klassische Analyse durch Synthese	8
2.2.2	Stichprobenvergrößerung durch Ansichtssynthese	9
2.2.3	Ähnlichkeitsmaße für die Modellanpassung	9
2.2.4	Lageschätzung durch Neuronale Netze	10
2.3	Eigene Vorarbeiten	12
3	Ziele und Arbeitsprogramm	16
3.1	Ziele	16
3.1.1	Datenmodell	17
3.1.2	Interaktive Lichtfeld-Akquisition	18
3.1.3	Echtzeit-Bildsynthese	18
3.1.4	Objekterkennung	19
3.1.5	Beleuchtungstoleranz	19
3.2	Arbeitsprogramm	20
3.2.1	Teilschritt A: Demonstration der Durchführbarkeit des Ansatzes	20
3.2.2	Teilschritt B: Anpassung des Systems an statische Beleuchtungssituationen	22
3.2.3	Teilschritt C: Anpassung des Systems an unbekannte Beleuchtungssituationen	23
3.2.4	Zeitliche Einordnung der Aufgaben	23
3.3	Untersuchungen am Menschen	25
3.4	Tierversuche	25
3.5	Gentechnologische Experimente	25
4	Beantragte Mittel	25
4.1	Personalkosten	26
4.2	Wissenschaftliche Geräte	26
4.3	Verbrauchsmaterial	26

4.4	Reisen	26
4.5	Publikationskosten	27
4.6	Sonstige Kosten	27
5	Voraussetzung für die Durchführung des Vorhabens	27
5.1	Zusammensetzung der Arbeitsgruppe	29
5.2	Zusammenarbeit mit anderen Wissenschaftlern	29
5.3	Arbeiten im Ausland und Kooperation mit ausländischen Partnern	30
5.4	Apparative Ausstattung	30
5.5	Laufende Mittel für Sachausgaben	30
5.6	Sonstige Voraussetzungen	30
6	Erklärungen	30
6.1	Max-Planck-Institut	31
7	Unterschriften	
8	Verzeichnis der Anlagen	

1 Allgemeine Angaben

Antrag auf Gewährung einer Sachbeihilfe
Neuantrag

1.1 Antragsteller

Andreas Kolb, Dr.-Ing.
Universitätsprofessor, Projektbereichsleiter
Zentrum für Sensorsysteme (ZESS) und
Lehrstuhl für Computergraphik und Multimediasysteme
geb. 04.06.1965, Deutscher

Dienstadresse:
Fachbereich 12 – Elektrotechnik & Informatik
Universität Siegen
Hölderlinstr. 3, 57068 Siegen
Tel.: 0271 / 740 - 2404
Fax.: 0271 / 740 - 3337

Privatadresse:
Meilerstr. 76, 57074 Siegen
Tel.: 0271 / 741 - 1950

Klaus-Dieter Kuhnert, Dr.-Ing.
Universitätsprofessor, Projektbereichsleiter
Zentrum für Sensorsysteme (ZESS) und
Institut für Echtzeit-Lernsysteme
geb. 20.01.1954, Deutscher

Dienstadresse:
Fachbereich 12 – Elektrotechnik & Informatik
Universität Siegen
Hölderlinstr. 3, 57068 Siegen
Tel.: 0271 / 740 - 4779
Fax.: 0271 / 740 - 4421

Privatadresse:
Vor der Hurth 37, 57250 Netphen,
Tel. 02732-590882

Christof Rezk-Salama, Dr.-Ing.
Wissenschaftlicher Assistent
Lehrstuhl für Computergraphik und Multimediasysteme
geb. 01.05.1970, Deutscher

Dienstadresse:
Fachbereich 12 – Elektrotechnik & Informatik
Universität Siegen
Hölderlinstr. 3, 57068 Siegen
Tel.: 0271 / 740 - 2826
Fax.: 0271 / 740 - 3337

Privatadresse:
Hauptstr. 7, 51598 Friesenhagen
Tel.: 02734 / 479942

1.2 Thema

Echtzeit-Akquisition bildbasierter 3D Modelle zur Objekterkennung.

1.3 Kennwort

PMDLumi

1.4 Fachgebiet und Arbeitsrichtung

Informatik, Computergraphik, Objekterkennung, Bildverarbeitung

1.5 Voraussichtliche Gesamtdauer

4 Jahre

1.6 Antragszeitraum

24 Monate

1.7 Gewünschter Beginn der Förderung

01.01.2006

1.8 Zusammenfassung

Das Ziel des Projekts ist die Entwicklung des Prototypen eines bildbasierten, echtzeitfähigen Systems zur Akquisition dreidimensionaler Modelle und zur Erkennung (Klassifikation und Lageschätzung) von 3D Objekten. Das Gesamtsystem besteht aus der echtzeitfähigen Rekonstruktion von *Lichtfeldern* mit Hilfe der PMD-Technologie und einer bildbasierten Objekterkennung, die nach dem Prinzip der *Analyse durch Synthese* arbeitet.

Sowohl die Objekterkennung als auch die Modellakquisition soll in einem echtzeitfähigen Prozess stattfinden. Photorealistische Modelle realer Objekte sollen mit diesem System interaktiv aufgenommen werden. Die Lichtfeldrekonstruktion bietet den Vorteil, dass sie anschließend eine photorealistische Bildsynthese in Echtzeit ermöglicht. Dieser Vorteil soll für die bildbasierte Objekterkennung genutzt werden. Die Bestimmung von Objektklasse und Raumlage erfolgt dabei durch iterative Annäherung an die optimalen Syntheseparameter durch fortgesetzten Vergleich.

Die Kombination einer herkömmlichen Videokamera mit einer Tiefenkamera basierend auf PMD-Technologie verspricht eine kostengünstige und portable Lösung für die Akquisition photorealistischer 3D Modelle und die 3D Objekterkennung.

2 Stand der Forschung, eigene Vorarbeiten

Die bildbasierte Erkennung und Lageschätzung von Objekten beinhaltet das Problem, dass die Abbildung eines räumlichen Objekts auf eine zweidimensionale Bildebene nicht ohne weiteres umkehrbar ist. Aus diesem Grund ist es einfacher, Bilder dreidimensionaler Objekten zu generieren als umgekehrt, 3D Objekte aus Bildern zu extrahieren. Dies gilt sowohl für Photographien realer Objekte als auch für computergenerierte Bilder virtueller Szenen. Diese bekannte Tatsache nutzen Algorithmen, die nach dem Prinzip der *Analyse durch Synthese* arbeiten. Sie automatisieren das Überprüfen von Hypothesen, erkennen schlechte Hypothesen und korrigierten diese iterativ. Die Effizienz eines Algorithmus hängt natürlich stark von der Qualität und der Performanz der Bildsynthese ab. In den beantragten Projekt sollen Lichtfeld-basierte Verfahren zur Objekterkennung untersucht werden. Voraussetzung dafür ist die handhabbare, interaktive Akquisition photorealistischer Lichtfeld-Modelle. Im Ergebnis soll ein robustes, echtzeitfähiges Objekterkennungssystem erarbeitet werden, das auf Basis photorealistischer Bildsynthese arbeitet.

2.1 Stand der Forschung: Bildbasierte Modellrekonstruktion

Die Grundidee von Lichtfeld [1] und Lumigraph [2] ist die Aufnahme der kompletten Lichtverteilung innerhalb einer begrenzten Region des dreidimensionalen Raumes. Der Vorteil im Vergleich zu traditionellen Verfahren ist hier die Entkoppelung der Komplexität der Bildsynthese von der Szenenkomplexität. Seit der Veröffentlichung der ursprünglichen Ideen wurden die Verfahren sehr stark weiterentwickelt. Eine vollständige Liste der Publikationen würde den Rahmen dieses Antrags sprengen. Wir beschränken uns daher auf einen Überblick der wichtigsten Beiträge im Kontext des Antrags.

Die vollständige Erfassung photorealistischer Lichtfeldmodelle realer Objekte ist zur Zeit mit sehr hohen Kosten und erheblichem Arbeitsaufwand verbunden. Ein integraler Teil dieses Antrags ist die interaktive Aufnahme bildbasierter Modelle statischer Objekte (Lumigraph) mit unmittelbarer Visualisierung des Aufnahmeergebnisses während des Aufnahmeprozesses. Der PMD-Tiefensensor bietet in diesem Zusammenhang eine Lösung zur echtzeitfähigen, vollflächigen Akquisition von Tiefendaten mit einem portablen System.

2.1.1 Bild- und Videobasiertes Rendering

Existierende Verfahren zur Akquisition photorealistischer Lumigraph-Modelle (geometriegestützte Lichtfelder) basieren entweder auf günstigerameratechnik (z.B. [3, 4]) mit aufwändiger Nachverarbeitung zur Ermittlung der Tiefendaten oder auf kostenintensiven

Mehrkamerasystemen für *Video-based Rendering* [5], *Free-Viewpoint Video* [6] und *3D TV* [7, 8]. Die aufwändigeren Systeme sind in der Praxis durch die hohen Kosten und den großen Aufwand für Transport, Aufbau und Handhabung sehr stark eingeschränkt. Viele aktuelle Arbeiten bewegen sich momentan mit aufwändigen Kamera-Arrays und PC-Mini-Clustern in die Richtung dynamische, geometriegestützte Lichtfelder [9]. Der Nachverarbeitungsaufwand aller bekannten Verfahren ist jedoch beträchtlich, sodass das Ergebnis der Akquisition oft erst im Nachhinein beurteilt werden kann. Interaktive, einfach zu handhabende Systeme zur Akquisition geometriegestützte Lichtfelder mit unmittelbarer Darstellung des Aufnahmeergebnisses sind nach dem momentanen Stand der Technik nicht verfügbar.

2.1.2 3D Modell-Rekonstruktion

Unter dem Begriff *3D Photographie* [10] werden allgemeine Techniken zur Rekonstruktion von Szenenmodellen aus statischen Bildfolgen zusammengefasst. In der Literatur gibt es eine ganze Reihe an Ansätzen zur Rekonstruktion von Oberflächengeometrie aus Bildern (z.B. [11, 12]). Einige Verfahren gibt es auch zur Rekonstruktion der Oberfläche aus Tiefenkarten oder Punktwolken, generiert von Laser Range Scannern [13, 14]. Von besonderem Interesse im Fokus dieses Antrags steht ein Ansatz, der die Unsicherheit und Messungenauigkeiten in einem Voxelmodell speichert [15]. Die Autoren verwenden dabei eine kumulative, gewichtete Distanzfunktion und extrahieren das 3D Modell als Isofläche aus den Voxeldaten. Unter bestimmten Einschränkungen wird gezeigt, dass die entstehende Oberfläche optimal im Sinne einer quadratischen Fehlermetrik ist.

Ein weiteres sehr interessantes Verfahren zur bildbasierten 3D Geometrierekonstruktion wurde 2002 in [16] vorgestellt. Die Tiefeninformation wird in diesem Fall in einem Triangulationsverfahren ermittelt, das mit Projektion von strukturiertem Licht arbeitet. Die Besonderheit des Verfahrens liegt in dem echtzeitfähigen, visuellen Feedback während der Aufnahme. Die 3D Registrierung erfolgt dabei durch eine schnelle Implementierung des ICP-Algorithmus (*Iterated Closest Point* [17]). Die unmittelbare Darstellung des Ergebnisses ermöglicht es dem Benutzer während der Aufnahme zu prüfen, ob das 3D Modell vollständig erfasst wurde, oder ob es noch Löcher durch fehlende Aufnahmen gibt.

Die oben aufgelisteten Arbeiten behandeln lediglich das Problem der Geometrierekonstruktion, nicht die Aufnahme photorealistischer, bildbasierter Modelle. Lumigraph-Modelle haben den Vorteil, dass sie auch Mikrogeometrie erfassen, die die visuellen Eigenschaften deutlich beeinflusst, jedoch nicht explizit polygonal dargestellt werden kann (z.B. Oberflächenrauigkeit).

2.1.3 Lumigraph-Modelle

Geometriegestützte Lichtfelder basieren in der Regel auf der Berechnung der visuellen Hülle des aufgenommenen Objekts als (grobe) Geometrie, auf die blickpunktabhängige Texturen projiziert werden. Die meisten Akquisitionsverfahren arbeiten ohne Tiefensensor und rekonstruieren die Geometrie anhand von Silhouetten aus den Farbbildern. Diese visuelle

Hülle wird entweder als volumetrische Repräsentation [18, 19], bildbasiert [3, 20], hardwaregestützt [21] oder als polygonales Objekt [22, 23] gespeichert. Für die Geometrieberechnung können darüber hinaus Tiefendaten aus den Farb-Bildern bestimmt werden [4]. Der bisher einzige praktische Ansatz, der mit aufgenommenen Tiefendaten arbeitet, ist ein aufwändiges, ortsgebundenes Mehrkamarasystem, das in [24] beschrieben wird.

2.1.4 Datenstrukturen

Lichtfeldmodelle stellen eine Diskretisierung der (vereinfachten) plenoptischen Funktion dar. Die einfachste Möglichkeit zur Speicherung der plenoptischen Funktion ist die Zwei-Ebenen-Parametrisierung [1]. Dabei sind die Kamerapositionen auf den Gitterpunkten einer Ebene angeordnet, während eine zweite, parallele Ebene der gemeinsamen Bildebene aller Kameras entspricht. Freiform-Lichtfelder [25] erlauben eine flexiblere Anordnung der Kameras. Die Kamerapositionen sind hier die Punkte eines beliebigen Gitternetzes und die Bildebenen der einzelnen Kameras sind unterschiedlich. Die Datenstruktur beinhaltet in diesem Fall neben den Bildern und den Kalibrierungsdaten keine weiteren Informationen.

Die Hinzunahme einer akkuraten Geometrie führt zu den Ideen des *blickpunktabhängigen Texture-Mappings* [26, 27]. Diese Ansätze verwenden eine Datenstruktur, die für jeden Eckpunkt der Geometrie eine sogenannte *View Map* speichert. Diese erlaubt es, sehr schnell die Texturinformation blickpunktabhängig aus drei benachbarten Kameraperspektiven zu interpolieren. Zwischen den Extremen der geometrieunabhängigen Lichtfelder auf der einen Seite und der blickpunktabhängigen Texturierung möglichst exakter Geometrie auf der anderen, liegt als übergreifendes Framework der unstrukturierte Lumigraph [28]. Dieser Ansatz zur Speicherung geometriegestützter Lichtfelder erlaubt die flexible Anpassung von Bilddaten und Geometrie an die Anforderungen und Randbedingungen der Bildsynthese. Viele der bekannten Datenstrukturen und Darstellungsverfahren wurden im Rahmen des lgf-Projekts [29] implementiert.

2.2 Stand der Forschung: 3D-Szenenerkennung

Die modellgestützte 3D-Szenenerkennung nach dem Prinzip der Analyse durch Synthese ist in der Vergangenheit für einige eng begrenzte Anwendungssituationen untersucht worden. Zu nennen sind hier v.a. das 4D-Modell von Dickmanns et al. [30, 31, 32], in dem neben den drei Raumkoordinaten die Zeit als vierte Dimension modelliert ist, sowie die Arbeiten von Liedtke und Koch [33, 34, 35]. Die untersuchten Verfahren beziehen sich auf die Erkennung von Verkehrsszenen, Anwendungen in der Luftfahrt und die Erkennung von Gebäuden. Als Modelle kommen dreidimensionale Eckenmodelle, allerdings auch detailliertere Gitternetze zum Einsatz [35]. In einem Ansatz von Blanz et al. [36] geschieht die Objekterkennung mit Hilfe von Support Vector Machines anhand zweidimensionaler Objektansichten.

Mit den Fortschritten der Computergraphik und der Graphikhardware hat sich der Detaillierungsgrad der Modelle enorm gesteigert, sodaß inzwischen Erkennungsverfahren existieren, die mit sehr realistischen Modellen arbeiten [37, 38, 39, 40, 41]. Die aktuellen Arbeiten kon-

zentrieren sich v.a. auf die Gesichtserkennung. Die Analyse industrienaher Szenen, bei denen es um die Unterscheidung vieler verschiedener Objekte geht, ist insbesondere für wechselnde Beleuchtung wissenschaftlich noch nicht durchdrungen. Dabei ist der Einsatzbereich des Prinzips 'Analyse durch Synthese' durchaus weit und reicht von Sprach- und Videokodierung bis in die Psychologie. Im Bereich der Mustererkennung wird diese Technik beispielsweise auch zur Bewegungserkennung (vgl. [42] für eine Übersicht), zur Ganganalyse [43] oder zur Erkennung von Szenenüberblendungen in Videoströmen [44] eingesetzt.

Konzentriert man sich auf den für dieses Projekt relevanten Bereich der Objekterkennung, sind in der Forschung zwei Hauptströmungen erkennbar, die sich darin unterscheiden, in welchem Schritt der Bildverarbeitungskette die Synthese eingesetzt wird. Zum einen wird die Synthese während der Analyse eingesetzt, zum anderen während des Trainings des Erkennungssystems. Im zweiten Fall geschieht die eigentliche Objekterkennung mit statistischen Methoden. Die einzelnen Ansätze unterscheiden sich darüber hinaus in der Art des verwendeten Modells. In den meisten Ansätzen werden dreidimensionale Geometriemodelle eingesetzt, die je nach Erfordernissen durch spezielle Zusatzinformationen erweitert werden. Es gibt zwar bereits wenige Ansätze, die Lichtfelder zur Erkennung von Objekten nutzen [45], die Kombination von Lichtfeld-Rendering und Analyse durch Synthese ist derzeit aber eine offene Frage.

2.2.1 Klassische Analyse durch Synthese

Zu den Verfahren, in denen die Synthese integraler Bestandteil der Objekterkennung selbst ist, gehört der Ansatz von Dickmanns [30, 32]. Die zu erkennenden Objekte werden durch ihren Massenschwerpunkt, Bewegungs- und Rotationsgeschwindigkeiten, sowie eine Reihe leicht extrahierbarer Merkmale beschrieben. Die Szenenerkennung geschieht durch den Vergleich der aus den tatsächlichen Sensorwerten extrahierten Merkmale mit einer synthetischen Situationsbeschreibung. Diese wird durch Extrapolation der zeitlich vorangehenden Szenenbeschreibung in die Gegenwart/Zukunft mit Hilfe eines Kalman-Filters erzeugt.

Koch und Liedtke dagegen modellieren die zu erkennenden Objekte durch ein texturiertes 3D-Gitternetz [34, 35, 33, 46]. Dieses wird aus der Analyse von Stereobildfolgen gewonnen, indem Flächen gleicher Disparität und Bewegung zusammengefaßt werden. Die dazu nötige Bestimmung des optischen Flusses dient v.a. der Kompensation der Betrachterbewegung. Die Objektgeometrie wird iterativ durch den Vergleich mit synthetisierten Bildern verbessert. Zur Synthese wird auf die Oberflächentextur, die direkt aus dem Kamerabild gewonnen wird, und den optischen Fluß zurückgegriffen. Der Vergleich zwischen dem tatsächlichen und dem synthetischen Bild geschieht direkt auf den Intensitätswerten. Da Fehler in der Segmentierung ein synthetisches Bild schnell unrealistisch wirken lassen, wird die Bildsegmentierung durch eine Wissensbasis unterstützt, in der der Aufbau der zu erkennenden Objekte grob beschrieben ist.

2.2.2 Stichprobenvergrößerung durch Ansichtssynthese

Blanz et al. stellen ein modellbasiertes Gesichtserkennungssystem vor, das ebenfalls nach dem Prinzip der Analyse durch Synthese arbeitet. Die synthetischen Bilder werden allerdings nicht während des Erkennungsprozesses erzeugt, sondern während der Trainingsphase des Systems [36, 37, 47, 38, 39]. Als Grundlage für die Gesichtsmodellierung dient ein dreidimensionales Geometriemodell, das aus der Überlagerung mehrerer Laserscans verschiedener Gesichter erzeugt wird und deren Mittelwert, sowie die Hauptrichtungen der Variation enthält. Während des Trainings wird dieses Modell mit Hilfe von drei Aufnahmen aus verschiedenen Perspektiven an eine bestimmte Person angepaßt. Für die Anpassung ist eine grobe manuelle Justierung nötig. Die Rechenzeit beträgt wenige Minuten pro Bild. Mit Hilfe des angepaßten Modells werden daraufhin eine große Menge photorealistischer Ansichten unter verschiedenen simulierten Beleuchtungen erzeugt, die als Stichprobe für den eigentlichen Gesichtsklassifikator dient. Der Klassifikator selbst besteht aus einer Hierarchie von Support Vector Machines, die zuerst grob den Hintergrund ausfiltern und das Gesicht anschließend klassifizieren.

2.2.3 Ähnlichkeitsmaße für die Modellanpassung

Ähnlich Blanz et al. gehen sowohl Lu et al. [40] als auch Hu et al. [41] vor. Die eigentliche Gesichtserkennung funktioniert über eine Hauptkomponentenanalyse und ein speziell angepaßtes Abstandsmaß. In [40] geschieht die Modellanpassung über eine Reihe automatisch ermittelter Merkmalspunkte. Die Autoren geben eine Geschwindigkeitssteigerung um den Faktor 15 an im Vergleich zu dem in [47] beschriebenen Verfahren. Noch weiter gehen Krüger et al., die einen vollständig merkmalsbasierten Ansatz verfolgen [48] und eine Objekterkennung auf synthetischen Bildern aus VRML-Modellen demonstrieren. Die beschriebenen Experimente regen den Einsatz komplexerer Merkmale an, wie sie beispielsweise in [49] oder [50] beschrieben sind, und zeigen, dass der Vergleich zwischen echten und synthetischen Bildern nicht nur über Pixeldifferenzen erfolgen muss.

Lee [51] setzt zur Erkennung von Gesichtern ein Abstandsmaß ein, das abgestimmt ist auf ein gemischtes Modell, welches ein 3D-Gitternetz mit Farbflächen und Kantenmerkmalen kombiniert. Die Anpassung des Modells an ein bestimmtes Bild geschieht durch eine rekursive, zunächst grobe Quantisierung des Modell-Parameterraums. Die Bereiche des Parameterraums mit dem geringsten Abstandsmaß werden anschließend durch eine feinere Quantisierung genauer untersucht. Lee dokumentiert deutlich bessere Ergebnisse für dieses Verfahren als für ein auf der Hauptkomponentenanalyse basierendes Verfahren.

Burckhardt wendet das Prinzip der Analyse durch Synthese an, um für eine medizinische Anwendung die Lage eines künstlichen Hüftgelenks im Knochen zu vermessen [52]. Die Öffnung der Hüftpfanne ist in Röntgenbildern als Ellipse erkennbar, was eine initiale Lageschätzung erlaubt. Diese wird iterativ verfeinert, indem künstliche Röntgenbilder mit anderen Positionen der Hüftpfanne erzeugt werden und pixelweise mit der echten Röntgenaufnahme verglichen werden. Die Schätzung der Lageparameter der Hüftpfanne wird über einen kombinierten Gauß-Newton und modifizierten Newton-Algorithmus iterativ verfeinert.

2.2.4 Lageschätzung durch Neuronale Netze

Shashua und Poggio arbeiten im Gegensatz zu den bisher genannten Verfahren nicht mit einem echten 3D-Modell, sondern speichern zu einem Objekt mehrere verschiedene Ansichten [53, 54]. Neue Ansichten werden mit Hilfe des optischen Flusses aus zwei benachbarten Ansichten interpoliert. Die Zuordnung von Lageparametern zu Ansichten geschieht mit Hilfe eines Neuronalen Netzes.

Literatur zum Stand der Forschung

- [1] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. ACM SIGGRAPH*, pages 31–42, 1996.
- [2] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In *Proc. ACM SIGGRAPH*, pages 43–54, 1996.
- [3] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan. Image-based visual hulls. In *Proc. ACM SIGGRAPH*, pages 369–374, 2000.
- [4] H. Schirmacher, M. Li, and H.-P. Seidel. On-the-fly processing of generalized lumigraphs. *Computer Graphics Forum*, 20:165–173, 2001. (Proc. Eurographics).
- [5] M. Gross et al. Blue-C: A spatially immersive display and 3D video portal for telepresence. *ACM Transactions on Graphics*, 22(3), 2003. (Proc. ACM SIGGRAPH '03).
- [6] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. *ACM Transaction on Graphics*, 22(2):569–577, 2003.
- [7] C. Fehn et al. An evolutionary and optimised approach on 3D-TV. In *Proc. International Broadcast Conference*, pages 357–365, 2002.
- [8] W. Matusik and H. Pfister. 3D TV: A scalable system for real-time acquisition, transmission and autostereoscopic display of dynamic scenes. *ACM Transactions on Graphics*, 23(3), 2004. (Proc. ACM SIGGRAPH '04).
- [9] M. Magnor. *Video-Based Rendering*. AK Peters, July 2005.
- [10] B. Curless and S. Seitz. 3D photography course notes. In *ACM SIGGRAPH*, 2000.
- [11] S.M. Seitz and C.R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proc. Computer Vision and Pattern Recognition Conference (CVPR'97)*, pages 1067–1073, 1997.
- [12] J.-Y. Bouguet and P. Perona. 3D photography using shadows in dual-space geometry. *International Journal of Computer Vision*, 35(2):129–149, 1999.
- [13] M.K. Reed and P.K. Allen. 3D modeling from range imagery: An incremental method with a planning component. *Image and Vision Computing*, 17:99–111, 1999.
- [14] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proc. ACM SIGGRAPH*, pages 311–318, 1994.
- [15] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proc. ACM SIGGRAPH*, pages 303–312, 1996.
- [16] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3D model acquisition. In *ACM SIGGRAPH*, pages 438 – 446, 2000.
- [17] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *Journal of Computer Vision*, 13(2), 1994.
- [18] M. Potmesil. Generating octree models of 3D objects from their silhouettes in a sequence of images. *Computer Vision, Graphics, and Image Processing*, 40:1–29, 1987.
- [19] R. Szeliski. Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58(1):23–32, July 1993.
- [20] E. Boyer and J.-S. Franco. A hybrid approach for computing visual hulls of complex objects. *Computer Vision and Pattern Recognition (CVPR'03)*, 1:695–701, June 2003.

- [21] M. Li, M. Magnor, and H.-P. Seidel. A hybrid hardware-accelerated algorithm for high quality rendering of visual hulls. In *Graphics Interface'04*, May 2004.
- [22] W. Matusik, C. Buehler, and L. McMillan. Polyhedral visual hulls for real-time rendering. In *Proc. Eurographics Workshop on Rendering*, 2001.
- [23] J.-S. Franco and E. Boyer. Exact polyhedral visual hulls. In *Proc. British Machine Vision Conference (BMVC'03)*, volume I, pages 329–338, September 2003.
- [24] S. Naemura, J. Tago, and H. Harashima. Real-time video-based modeling and rendering of 3D scenes. *IEEE Computer Graphics and Applications*, pages 66–73, March 2002.
- [25] B. Heigl. *Plenoptic Scene Modelling from Uncalibrated Image Sequences*. Doktorarbeit, University of Erlangen-Nuremberg, 2003.
- [26] Paul Debevec, C. Taylor, and J. Malik. Modeling and rendering architecture from photographs. In *Proc. ACM SIGGRAPH*, pages 11–20, 1996.
- [27] Paul Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture mapping. In *Eurographics Rendering Workshop*, pages 11–20, 2001.
- [28] C. Buehler, M. Bosse, A. B. K. McMillan, S. Gortler, and M. Cohen. Unstructured lumigraph rendering. In *Proc. ACM SIGGRAPH*, pages 425 – 432, 2001.
- [29] C. Vogelgsang. *The Icf3 Project: A Versatile Implementation Framework for Image-Based Modeling and Rendering*. Doktorarbeit, University of Erlangen-Nuremberg, 2005.
- [30] E.D. Dickmanns. Visual dynamic scene understanding exploiting high-levelspatio-temporal models. In *10th Int. Conf. on Pattern Recognition*, volume 2, pages 373–378, 1990.
- [31] Jens Schick. *Gleichzeitige Erkennung von Form und Bewegung durch Rechnersehen*. Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, Institut für Systemdynamik und Flugmechanik, 1992.
- [32] E.D. Dickmanns, M. Maurer, S. Fürst, R. Behringer, and F. Thomanek. A compact vision system for road vehicle guidance. In *Proc. 13th Int. Conf. on Pattern Recognition*, pages 313–317, 1996.
- [33] S. G. C.-E. Liedtke, O. Grau. Use of explicit knowledge for the reconstruction of 3-d object geometry. In *Proc. 6th Int. Conf. Computer Analysis of Images and Patterns (CAIP'95)*, 1995.
- [34] R. Koch. Dynamic 3D scene analysis through synthesis feedback control. In *Proc. IEEE Trans. Patt. Anal. Mach. Intell., Special issue on analysis and synthesis*, volume 15(6), pages 556–568, 1993.
- [35] R. Tönjes, O. Grau, and R. Koch. Analyse durch Synthese Modellierung von 3D Objekten in Stereobildfolgen. In *Workshop of Visual Computing, Darmstadt, Germany*, 1994.
- [36] V. Blanz, B. Schölkopf, H. Bülhoff, C. Burges, V. Vapnik, and T. Vetter. Comparison of view-based object recognition algorithms using realistic 3D models. In *Proc. Int. Conf. on Artificial Neural Networks*, pages 251–256, 1996.
- [37] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. ACM SIGGRAPH '99*, pages 187–194, 1999.
- [38] J. Huang and B. Heisele. Component-based face recognition with 3D morphable models. In *Proc. 4th Conf. on Audio- and Video-Based Biometric Person Authentication*, 2003.
- [39] V. Blanz, P. Grother, J. Phillips, and T. Vetter. Face recognition based on frontal views generated from non-frontal images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition CVPR 2005*, 2005.
- [40] X. Lu, R.-L. Hsu, A.K. Jain, B. Kamgar-Parsi, and B. Kamgar-Parsi. Face recognition with 3D model-based synthesis. In *Proc. Int. Conf. on Biometric Authentication*, pages 139–146, 2004.
- [41] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang. Automatic 3D reconstruction for face recognition. In *Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.
- [42] T. Moeslund. The analysis-by-synthesis approach in human motion capture: A review. In *Proc. 8th Danish Conf. on Pattern Recognition and Image Analysis*, 1999.
- [43] Roman Calow. *Markerlose Ganganalyse mit einem Multikamerasystem*. Doktorarbeit, Otto-von-Guericke-Universität Magdeburg, 2005.
- [44] M. Covell and S. Ahmad. Analysis-by-synthesis dissolve detection. In *Proc. Int. Conf. on Image Processing*, volume 2, pages 425–428, 2002.

- [45] B. Heigl, J. Denzler, and H. Niemann. On the application of lightfield reconstruction for statistical object recognition. In *European Signal Processing Conference (EUSIPCO)*, pages 1101–1104, 1998.
- [46] C.-E. Liedtke and R. Koch. Modelling of 3D scenes from the analysis of stereoscopic image sequences. *Int. Journal on Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications)*, 3(3):346–351, 1993.
- [47] J. Huang, V. Blanz, and B. Heisele. Face recognition using component-based svm classification and morphable models. In *S.-W. Lee and A. Verri (Eds.): SVM 2002, LNCS 2388*, pages 334–341. Springer-Verlag, 2002.
- [48] T. Krüger, J. Wickel, P. Alvarado, and K.-F. Kraiss. Feature extraction from vrml models for view-based object recognition. In *Proc. 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 391–394, 2003.
- [49] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 22(12), pages 1349–1380, 2000.
- [50] C. Zhang and T. Chen. Efficient Feature Extraction for 2D/3D Objects in Mesh Representations. In *Proc. ICIP 2001*, pages 935–938, 2001.
- [51] M. W. Lee and S. Ranganath. Pose-invariant face recognition using a 3D deformable model. In *Pattern Recognition*, volume 36, pages 1835–1846, 2003.
- [52] K. V. Burckhardt. *Locating implants with respect to the bone in diagnostic x-ray images of the pelvis*. Dissertation, Technische Wissenschaften ETH Zürich, 2001. Nr. 14262.
- [53] D. Beymer, A. Shashua, and T. Poggio. Example based image analysis and synthesis. Technical Report AIM-1431, Massachusetts Institute of Technology, November 1993.
- [54] S. Avidan, T. Evgeniou, A. Shashua, and T. Poggio. Image-based view synthesis. Technical Report AIM-1603, Massachusetts Institute of Technology, 1997.

2.3 Eigene Vorarbeiten

Der Lehrstuhl für *Computergraphik und Multimediasysteme* an der Universität Siegen wurde im Jahre 2003 eingerichtet. Prof. Andreas Kolb ist seit 13 Jahren im Bereich der Computergraphik aktiv. Die Mitarbeiter des Lehrstuhls für Computergraphik und Multimediasysteme verfügen über breites Fachwissen im Bereich Computergraphik, Wissenschaftliche Visualisierung und 3D Bildverarbeitung. Neueste Arbeiten beschäftigen sich mit echtzeitfähigen Verfahren unter Nutzung von Graphik-Hardware, beispielsweise zur Beleuchtungsberechnung [1], Partikelsimulation [2], hardwarebeschleunigter Volumenvisualisierung [3, 4] und 3D Bildverarbeitung [5, 6].

Dr. Christof Rezk-Salama, wissenschaftlicher Assistent am Lehrstuhl für Computergraphik und Multimediasysteme, bearbeitet seit ca. 7 Jahren Themen der wissenschaftlichen Visualisierung und Computergraphik. Dr. Rezk-Salama hat als Stipendiat des Graduiertenkollegs “3D Image Analysis and Synthesis” der Universität Erlangen-Nürnberg in enger Zusammenarbeit mit dem Sonderforschungsbereich 603 “Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten” promoviert [7]. In seiner Arbeit hat er sich unter anderem mit 3D Bildverarbeitung und bildbasierten Ansätzen beschäftigt. Der Focus der Arbeiten lag dabei in besonderem Maße auf Visualisierungsverfahren für Medizin [8, 9, 10, 11] und Technik [12].

Das *Institut für Echtzeit Lernsysteme* unter der Leitung von Prof. Klaus-Dieter Kuhnert besitzt langjährige Erfahrung auf dem Gebiet der digitalen Bildverarbeitung und der Mustererkennung. Mit der Szenenanalyse beschäftigen sich die Mitarbeiter des Instituts schon

viele Jahre im Zusammenhang mit der Entwicklung von Fahrerassistenzsystemen. Für ein auf Verstärkungslernen basierendes System wurde eine Bilderkennung für Verkehrsszenen implementiert [13, 14, 15]. Für einen Einparkassistenten wurde ein System entwickelt, daß seine Umgebung über Laserscanner wahrnimmt. Dieses erkennt freie Parklücken während der Fahrt und kann einen Wagen vollkommen selbständig in die Parklücke steuern [16, 17]. Für ein weiteres System zur autonomen Fahrzeugsteuerung, das mit einem neuronalen Netz arbeitet, wurde ein Verfahren zur Erkennung der Fahrbahnränder entwickelt [18, 19, 20, 21].

Im Bereich der wissenschaftlichen Visualisierung existieren umfassende Vorarbeiten zu den Themen Volumenvisualisierung [22, 7], Volume Processing [5], Isoflächenextraktion [8] und Datenfusion [23]. Weitere Schwerpunkte liegen in der Entwicklung nichtlinearer Deformationsmodelle [5, 24] für echtzeitfähige Datenfusion. Diese Arbeiten berücksichtigen Ergebnisse der 3D Bildregistrierung (Matching) mehrerer multimodaler Datensätze (z.B. [25, 23] hier im medizinischen Kontext). Effiziente Verfahren der 3D-Bildfusion, d.h. der Verarbeitung mehrerer Datensätze in einer gemeinsamen, visuellen Darstellung, wurde im Rahmen von Operationsplanung [6, 26] und Augmented Reality [11, 27] entwickelt. Im Fokus des beantragten Projekt sind diese Vorarbeiten wichtige Grundlagen im Bereich 3D Bildverarbeitung und Datenfusion.

Die bildbasierte Geometrierekonstruktion aus Konturen (Visual Hull) wurde als Teil eines BMBF-Projektes zum 3D-Scanning mit 2D Silhouetten [28] untersucht. In diesem Zusammenhang wurden semiautomatische Segmentierungs- und benutzergesteuerte Registrierungsverfahren verwendet. Die Geometrierekonstruktion erfolgte hierbei durch den Aufbau eines Volumenmodells und einer anschließenden Polygonalisierung. Die Optimierung der visuellen Qualität von Oberflächen wurde unter Einsatz verschiedener Qualitätsmaße untersucht [29, 30, 31, 32]. Darüber hinaus wurden Verfahren zur Reparatur polygonaler Oberflächen erarbeitet [33].

Zur Lösung des Problems der Objekterkennung im allgemeinen Fall wurden Verfahren entwickelt, um automatisch abstrakte Konturmerkmale zu erzeugen, die zur Unterscheidbarkeit von Objekten verschiedener Klassen optimal geeignet sind [34, 35]. Diese Verfahren unterscheiden sich deutlich von der traditionellen heuristischen Merkmalsextraktion, bei denen vordefinierte Merkmale aus einem Bild extrahiert und auf ihre Tauglichkeit für die Klassifikation überprüft werden. Für das Problem hochpräziser Konturerkennung wurde ein auf Höhenlinien basierendes Verfahren entwickelt [36]. Ein Objekterkennungsverfahren nach dem derzeit auf relativ großes Forschungsinteresse stoßenden erscheinungsbasierten Ansatz wurde in [37] vorgestellt. Das genannte Verfahren bietet eine Lösung für die automatische Erzeugung von klassenabdeckenden Objektrepräsentanten mit Hilfe eines genetischen Algorithmus.

Als konkrete Vorarbeit für das beantragte Forschungsvorhaben wurde bereits ein einfaches Verfahren zur Generierung synthetischer Bild- und Tiefendaten in Echtzeit implementiert. Dieser Ansatz verwendet ein einfaches Lochkameramodell und erzeugt ideale Datensätze (d.h. ohne Bildfehler und Verzerrungen) für die Evaluierung der bildbasierten Algorithmen. Diese Arbeit ermöglicht es, die Entwicklung von Objekterkennungsverfahren auf Basis synthetischer Daten sofort zu beginnen und Algorithmen anhand idealer Daten zu evaluieren.

Im Rahmen einer Diplomarbeit in Kooperation mit PMDTec, dem Hersteller des Sensorelements, wurden bereits die erforderlichen Schnittstellen für den Datentransfer der Tiefendaten in Echtzeit implementiert [38]. Erste Darstellungsverfahren auf Basis realer Daten wurden in diesem Kontext ebenfalls entwickelt. Die Fusion der Tiefendaten mit Videobildern wurde in einer weiteren Vorarbeit untersucht. Bei dieser Studie handelt es sich bereits um einen binokularen Aufbau, bestehend aus dem PMD-Sensorelement kombiniert mit einer handelsüblichen Videokamera (Webcam) in fester Lage. Die Kalibrierung und Registrierung der Sensordaten erfolgte in diesem Fall durch ein extrem einfaches markerbasiertes Verfahren. Die exakte Evaluierung des Kalibrierungsfehlers und der Datenfusion, sowie die Identifizierung von Sensorfehlern und deren Abhängigkeiten steht noch aus. Die beschriebenen spezifischen Vorarbeiten mit dem PMD-Sensorelement sind viel versprechend und bescheinigen ein hohes Entwicklungs- und Anwendungspotential für das beantragte Projekt.

Literatur zu den eigenen Vorarbeiten

- [1] L. Latta and A. Kolb. Homomorphic factorization of brdf-based lighting computation. In *Proc. ACM SIGGRAPH*, volume 21, pages 509–516, 2002.
- [2] A. Kolb, L. Latta, and C. Rezk-Salama. Hardware-based simulation and collision detection for large particle systems. In *Proc. Graphics Hardware*, 2004.
- [3] C. Rezk-Salama and A-Kolb. A vertex shader for efficient box-plane intersection. In *Proc. Vision Modeling and Visualization (VMV'05)*, 2005.
- [4] K. Engel, M. Hadwiger, J. Kniss, A. Lefohn, C. Rezk-Salama, and D. Weiskopf. Real-time volume graphics. In *ACM SIGGRAPH Course Notes*, 2004.
- [5] C. Rezk-Salama, M. Scheuring, G. Soza, and G. Greiner. Fast volumetric deformation on general purpose hardware. In *Proc. SIGGRAPH/Eurographics Workshop on Graphics Hardware*, 2001.
- [6] P. Hastreiter, C. Rezk-Salama, G. Soza, G. Greiner, R. Fahlbusch, O. Ganslandt, and C. Nimsky. Strategies for brain shift evaluation. *Medical Image Analysis*, 8(4):447–464, 2004.
- [7] C. Rezk-Salama. *Volume Rendering Techniques for General Purpose Graphics Hardware*. Dissertation, University of Erlangen-Nuremberg, 2002.
- [8] C. Rezk-Salama, K. Engel, M-Bauer, G. Greiner, and T. Ertl. Interactive volume rendering on standard pc graphics hardware using multi-textures and multi-stage rasterization. In *Proc. SIGGRAPH/Eurographics Workshop on Graphics Hardware*, 2000.
- [9] C. Rezk-Salama, P. Hastreiter, K. Eberhardt, B. Tomandl, and T. Ertl. Interactive direct volume rendering of dural arteriovenous fistulae in MR-CISS data. In *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 1999.
- [10] P. Hastreiter, C. Rezk-Salama, K. Eberhardt, B. Tomandl, and T. Ertl. Functional analysis of the vertebral column based on MR and direct volume rendering. In *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2000.
- [11] M. Scheuring, C. Rezk-Salama, H. Barfuß, K. Barth, A. Schneider, G. Greiner, G. Wessels, and H. Feussner. Intra-operative augmented reality (ar) with magnetic navigation and multi-texture based volume rendering for minimally invasive surgery. In *Rechner u. Sensorgestützte Chirurgie Heidelberg*, 2001.
- [12] C. Rezk-Salama, P. Hastreiter, C. Teitzel, and T. Ertl. Interactive exploration of volume line integral convolution based on 3D–texture mapping. In *Proc. IEEE Visualization*, 1999.

- [13] K.-D. Kuhnert and M. Krödel. A learning autonomous driver system on the basis of image classification and evolutionary learning. *Conference Machine Learning and Data Mining (MLDM), Leipzig, Germany, 2003*.
- [14] K.-D. Kuhnert and M. Krödel. A learning autonomous driver system on the basis of image classification and evolutionary learning. *Conference Machine Learning and Data Mining (MLDM), 2003*.
- [15] K.-D. Kuhnert and M. Krödel. Behaviour learning by visual gestalt completion. *In-Vehicle Cognitive Computer Vision Systems (IVCCVS), 2003*.
- [16] A. Schanz, A. Spieker, and K.-D. Kuhnert. Autonomous parking in subterranean garages – a look at the position estimation. *IEEE Intelligent Vehicles Symposium, Columbus, Ohio, USA, 2003*.
- [17] A. Schanz, A. Spieker, and K.-D. Kuhnert. Towards automatic parking. *Proc. International Conference on Robotics, Vision, Information and Signal Processing, Penang, Malaysia, 2003*.
- [18] K.-D. Kuhnert and W. Dong. Über die lernende Regelung autonomer Fahrzeuge mit neuronalen Netzen. *18. Fachgespräch Autonome Mobile Systeme (AMS), Karlsruhe, 2003*.
- [19] W. Dong and K.-D. Kuhnert. Robust adaptive control of nonholonomic mobile robots with parameter and nonparameter uncertainties. *IEEE Transactions on Robotics and Automation, 2004*.
- [20] W. Dong and K.-D. Kuhnert. Robust adaptive neural network based control of autonomous vehicles. *IEEE Transaction on Intelligent Transportation Systems, 2004*.
- [21] W. Dong and K.-D. Kuhnert. Robust adaptive control of nonholonomic mobile robots with parameter and nonparameter uncertainties. *IEEE Transactions on Robotics, 21(2):534–543, 2005*.
- [22] J. Kniss, M. Hadwiger, K. Engel, and C. Rezk-Salama. High quality volume graphics on consumer pc hardware. In *ACM SIGGRAPH Course Notes, 2002*.
- [23] C. Rezk-Salama, P. Hastreiter, G. Greiner, and T. Ertl. Non-linear registration of pre- and intraoperative volume data based on piecewise linear transformations. In *Proc. Vision, Modelling, and Visualization (VMV), 1999*.
- [24] R. Westermann and C. Rezk-Salama. Real-time volume deformation. In *Computer Graphics Forum (Eurographics), 2001*.
- [25] P. Hastreiter, C. Rezk-Salama, C. Nimsy, and G. Greiner C. Lürig. Registration techniques for the analysis of the brain shift in neurosurgery. *Computers & Graphics, 2000*.
- [26] G. Soza, P. Hastreiter, M. Bauer, C. Rezk-Salama, C. Nimsy, and G. Greiner. Intraoperative registration on standard pc graphics hardware. In *Bildverarbeitung für die Medizin, 2004*.
- [27] M. Scheuring, C. Rezk-Salama, H. Barfuss, A. Schneider, and G. Greiner. Augmented reality based on fast deformable 2D-3D registration for image guided surgery. In *SPIE Medical Imaging, 2002*.
- [28] A. Kolb. Reparatur und Texturierung von Polygonmodellen für Virtual Reality Anwendungen. <http://vrlab.fh-wedel.de/themen/modellreparatur.html>, 2002.
- [29] A. Kolb and H.-P. Seidel. Interpolating scattered data with C^2 surfaces. *Computer Aided Design, 27:4:277–282, 1995*.
- [30] G. Greiner, A. Kolb, and A. Riepl. Scattered data interpolation using data-dependent optimization techniques. *Graphical Models, (64):1–18, 2002*.
- [31] A. Kolb, H. Pottmann, and H.-P. Seidel. Surface reconstruction based upon minimum norm networks. In M. Dæhlen, T. Lyche, and L.L. Schumaker, editors, *Math. Methods for Curves and Surfaces*, Ulvik, Norway, 1995. Vanderbilt University Press.
- [32] A. Kolb, H. Pottmann, and H.-P. Seidel. Fair surface reconstruction using quadratic functionals. In *Computer Graphics Forum (Proc. Eurographics '95)*, volume 14, pages 469–479. Eurographics, Blackwell Publishers, 1995.

- [33] A. Kolb and L. John. Volumetric model repair for virtual reality applications. In *Proc. Eurographics, Short Papers*, 2001.
- [34] D. Pechtel and K.-D. Kuhnert. Towards feature fusion - the synthesis of contour sections distinguishing contours from different classes. *Proc. of 9th Discrete Geometry for Computer Imagery Conference (DGCI), Uppsala, Sweden*, 2000.
- [35] D. Pechtel and K.-D. Kuhnert. Towards automatically learning an implicit model from 2d-images based on a local similarity analysis of contours. *Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS), Takamatsu*, 2000.
- [36] M. Stommel and K.-D. Kuhnert. Subpixel accurate segmentation of small images using level curves. *Proc. Int'l Conf. on Computer Vision and Graphics (ICCVG), Warsaw, Poland*, 2004.
- [37] M. Stommel and K.-D. Kuhnert. Appearance based recognition of complex objects by genetic prototype-learning. *Proc. 13th Int'l Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG), Plzen, Czech Republic*, 2005.
- [38] M. Profittlich. Entwurf und Implementierung eines Software-Frameworks zur Ansteuerung entfernungs-messender PMD-Kameras und Verarbeitung der gewonnenen Daten. Diplomarbeit, Fachhochschule Wedel/Universität Siegen, 2004.

3 Ziele und Arbeitsprogramm

3.1 Ziele

Im Mittelpunkt des Projekts steht die Prototypentwicklung eines bildbasierten Systems zur Akquisition dreidimensionaler, photorealistischer Modelle und zur Erkennung (Klassifikation und Lageschätzung) von 3D Objekten in Echtzeit. Das Gesamtsystem basiert auf einem binokularen Kameraaufbau, bestehend aus einer hochauflösenden Videokamera und einem PMD-Sensorelement zur vollflächigen Tiefenmessung in niedrigerer Auflösung. Ziel des Projekts ist die Erforschung der erforderlichen Algorithmik für ein derartiges Objekterkennungssystem. Dabei sollen sowohl zur Modellakquisition als auch während der Szenenanalyse die Daten der Tiefenkamera und der hochauflösenden Videokamera synchron ausgewertet werden.

Das Gesamtsystem soll sich zusammensetzen aus der echtzeitfähigen Rekonstruktion geometriegestützter *Lichtfelder* und deren Anwendung zur bildbasierten Objekterkennung nach dem Prinzip der *Analyse durch Synthese*. Modelle realer Objekte sollen mit diesem System interaktiv aufgenommen werden. Die Lichtfeldrekonstruktion bietet dabei den Vorteil, dass anschließend eine photorealistische Bildsynthese in Echtzeit möglich ist. Dieser Vorteil soll für die bildbasierte Objekterkennung genutzt werden. Die Bestimmung von Objektklasse und Raumlage erfolgt dabei durch iterative Annäherung an die optimalen Syntheseparameter durch fortgesetzten Vergleich. Der vorliegende Projektantrag folgt somit sehr stark dem aktuellen, internationalen Trend zur Vereinigung von Forschungsinteressen aus *Computergraphik* und *Computer Vision*.

Die gegenwärtig eingesetzten Fertigungssysteme sind aufgrund der einfachen bzw. oft fehlenden Sensorik durch fest einprogrammierte Abläufe gekennzeichnet. Es wird erwartet,

dass sich das in diesem Antrag skizzierte System aufgrund der spezifischen Sensorik toleranter gegenüber unsystematischen Variationen der Arbeitsumgebung verhält, insbesondere gegenüber Lagevariationen der Bauteile oder zeitlichen Schwankungen innerhalb der Arbeitsprozesse.

Eine besondere Herausforderung bei der Objekterkennung entsteht durch unterschiedliche Beleuchtungssituationen. Der direkte Vergleich photorealistisch synthetisierter Ansichten mit Realaufnahmen funktioniert zunächst nur, wenn die Objekterkennung in exakt der gleichen Beleuchtungsumgebung stattfindet wie die Modellgewinnung. Für industrielle Situationen mag dies schon ausreichend sein. In dem beantragten Projekt soll jedoch auch die Untersuchung von Anwendungsszenarien mit unterschiedlichen Beleuchtungssituationen angegangen werden, evtl. sogar die Integration dynamischer Beleuchtung. Dabei sollen die Lösungsmöglichkeiten sowohl auf der Syntheseseite als auch auf der Erkennungsseite systematisch untersucht werden.

3.1.1 Datenmodell

Die Realisierung des oben genannten Systems erfordert als gemeinsame Basis für Modellgewinnung und Bildsynthese zunächst die Entwicklung eines flexiblen Datenmodells für geometriestützten Lichtfelder (Lumigraphs). Konkrete Anforderungen an das Datenmodell sind:

- Effiziente *Datenspeicherung* in einem progressiven (evtl. hierarchischen) Modell
- Effiziente Integration neuer Daten in die Repräsentation vorhandener Daten (*Akkumulation*)
- Effizienter *Zugriff* und schnelles Rendering für ein schnelles, visuelles Feedback
- Möglichkeiten zur Integration von Beleuchtungsinformation.

Ein möglicher Ansatz zur Akkumulation der zeitlich hochaufgelösten 2D-3D-Sensordaten ist die Verwendung eines Voxelmodells als zentrale Datenstruktur. Die statistische Information über die Genauigkeit des gemessenen Tiefenwerts wird in diesem Voxelmodell akkumuliert, das die aufgenommene Oberfläche als implizite Fläche speichert. Aus den Messdaten des Sensorelements wird dazu für jeden Messpunkt eine Wahrscheinlichkeitsverteilung des Tiefenwerts bestimmt und auf ein dreidimensionales Gitter rückprojiziert. Zusätzlich können optische Eigenschaften aus dem Farbbild in das Voxelmodell integriert werden um möglichst konsistente Oberflächenkorrespondenzen aus den Einzelmessungen zu finden. Für die statistische Analyse der fehlerbehafteten Messdaten bietet das relativ junge und aktive Forschungsgebiet *Independent Component Analysis* stochastische Ansätze um sensorspezifische Fehler in multimodalen Messdaten zu detektieren und zu eliminieren. Hier existieren umfassende Vorarbeiten im Bereich Volumenvisualisierung [1, 2], Volume Processing [3], Isoflächenextraktion [4] und Datenfusion [5]. Alternative Ansätze mit hohem Potential sind die bildbasierten Modellbeschreibungen [6]. Die Herausforderung bei diesen auf Silhouetten basierenden Ansätzen wäre die Integration der gemessenen, teilweise fehlerbehafteten Tiefenkarten in Echtzeit.

In späteren Teilen des Projektes sollen Möglichkeiten untersucht werden, die Objekterkennung in unterschiedlichen Beleuchtungsumgebungen durchzuführen. Bei der Entwicklung des Datenmodells müssen demnach Möglichkeiten zur Integration von Beleuchtungsinformation von Anfang an eingeplant werden.

3.1.2 Interaktive Lichtfeld-Akquisition

In bisherigen Anwendungen wurde das Lumigraph-Modell in einem relativ zeitaufwändigen Nachverarbeitungsschritt offline aus den Bilddaten berechnet. Derartige Verfahren bieten demnach auch keine oder nur sehr stark eingeschränkte visuelle Kontrolle über das Aufnahmeergebnis während, bzw. unmittelbar nach der Aufnahme. Dies ist jedoch für die Handhabbarkeit der Verfahren in der Praxis enorm wichtig um die Qualitätssicherung direkt am Aufnahmeort zu gewährleisten. Für die praktische Modellgewinnung ist direktes visuelles Feedback notwendig um während des Aufnahmeprozesses die Kontrolle über die Qualität des kompletten bildbasierten Modells zu haben und gegebenenfalls Ansichten zu ergänzen, bzw. Aufnahmen zu wiederholen.

Durch die Kombination einer herkömmlichen Videokamera mit einer Tiefenkamera basierend auf PMD-Technologie soll zunächst ein Verfahren zur interaktiven Akquisition des Lumigraph-Modells erarbeitet werden. Bei Verwendung des PMD-Sensors stehen Tiefenkarten ohne Berechnungsaufwand, d.h. ohne zeitliche Verzögerung zur Verfügung. Dies wird genutzt, um direktes visuelles Feedback zu erreichen d.h. echtzeitfähiges Rendering progressiver Lichtfelder während der Aufnahme. Für die pure Oberflächenrekonstruktion wurde solch ein direktes visuelles Feedback vor kurzer Zeit erfolgreich demonstriert [7]. Eine Schwierigkeit bei der Modellakquisition liegt in der begrenzten Auflösung des Tiefensensors. Die Entwicklung echtzeitfähiger Algorithmen zur Kompensation widersprüchlicher Information aus unterschiedlichen Einzelmessungen und zur Kombination sich ergänzender Information des Farb- und Tiefensensors ist hier von großer Bedeutung. Des Weiteren werden echtzeitfähige Fehlermodelle zur Bestimmung der Unsicherheit der Daten, d.h. der Notwendigkeit progressiver Verfeinerung benötigt.

Neben der in diesem Antrag fokussierten Anwendung für die Objekterkennung sind praktikable Lösungen zur Akquisition photorealistischer, bildbasierter 3D-Modelle von enormer Bedeutung für die Digitalen Medien, E-Commerce (Online-Produktpräsentationen, Virtual Showroom), Natur- und Kulturwissenschaften (interaktives, virtuelles Museum) und Kriminalistik (Spurensicherung). Viele der wissenschaftlichen Herausforderungen, die aus der Anwendung in der Objekterkennung stammen, entfallen hier allerdings.

3.1.3 Echtzeit-Bildsynthese

In der ersten Projektphase soll zunächst eine Prototyp-Implementierung des beschriebenen Erkennungssystems als *Proof-of-Concept* implementiert werden. In dieser frühen Phase können existierende Implementierungen zum Lumigraph-Rendering für die Echtzeit-Bildsynthese verwendet werden. Wenn es um die Erweiterung des Prototypen um beleu-

chungstolerante Verfahren geht, müssen eigene Bildsynthese-Module entwickelt bzw. existierende Implementierungen erweitert werden, um Lösungen zu finden die unterschiedliche Umgebungsbeleuchtung berücksichtigen können.

3.1.4 Objekterkennung

Der Kern des Objekterkennungssystems soll ausgehend von einer initialen Schätzung fortwährend Hypothesen über die unbekannte Szene erzeugen. Gemäß dieser Hypothesen werden mit Hilfe des in einer Trainingsphase erzeugten Modells photorealistische Ansichten synthetisiert. Diese Ansichten erlauben die schrittweise Korrektur der Hypothesen durch den hochpräzisen Vergleich mit den tatsächlichen Sensordaten. Vorangegangene sichere Teilhypothesen stellen dabei a-priori Wissen für mehrdeutige Bildteile dar [8].

Aufgrund der Leistungsfähigkeit moderner Graphikhardware und den Fortschritten in der Computergraphik ist die Synthese photorealistischer Ansichten in Echtzeit möglich [9]. Da spätestens nach Beendigung einer Initialisierungsphase nur wenige Schritte für die Analyse notwendig sind und auch der Hypothesentest effizient auf der Graphikarte implementiert werden kann [10], soll auch die Erkennung in Echtzeit laufen.

Das Prinzip der Analyse durch Synthese erlaubt zudem eine einfache Art der Sensorfusion durch die Synthese von Hypothesenrepräsentation für verschiedene Sensortypen, wobei im Gegensatz zu traditionellen merkmalsbasierten Methoden die zur Verfügung stehenden Sensordaten vollständig genutzt werden [9]. Darüber hinaus bietet das Verfahren den Vorteil, dass alle Modellparameter in Form von Hypothesen bekannt sind, was das Problem von Verdeckungen, Beleuchtungs- und Formänderungen, sowie Bewegungen entschärft [11]. Auf künstlich in die Szene eingefügte Hilfsmittel (Marker) kann daher verzichtet werden [12].

3.1.5 Beleuchtungstoleranz

Wie oben erwähnt, sollen im Rahmen des Projektes auch systematisch Möglichkeiten untersucht werden, das System unabhängig von der Umgebungsbeleuchtung bei der Modellakquisition zu machen. Dies erfordert Grundlagenarbeit sowohl auf Seiten der Modellakquisition als auch auf Seiten der Objekterkennung. Hier gibt es unterschiedliche Ansätze, die miteinander kombiniert zu viel versprechenden Lösungsansätzen führen. Der erste Ansatz ist die Integration einer Parametrisierung der Umgebungsbeleuchtung in die Lichtfeld-Daten. Dazu soll zunächst die Beleuchtung bei der Aufnahme auf geeignete Weise vermessen werden. Dies kann durch die Aufnahme von Spiegelkugeln oder durch kontrollierte Aktivierung einzelner Punktlichtquellen während der Akquisition erfolgen, ähnlich dem Ansatz zum *Image Based Lighting* [13].

Mögliche Ansätze auf Seiten der Objekterkennung sind die Entwicklung neuer Ähnlichkeitsmaße für den Vergleich zwischen Realbild und dem Ergebnis der Bildsynthese. Existierende beleuchtungstolerante Ähnlichkeitsmaße basieren vorwiegend auf Kantendetektion oder Texturmerkmalen. Der photorealistische Vergleich hat demgegenüber den Vorteil, Oberflächen- und Materialeigenschaften auf Pixelbasis vergleichen zu können. Die Bewertung der Model-

lanpassung kann in dem Fall anhand eines der direkten Wahrnehmung näheren Kriteriums auf Basis der natürlichen Objekterscheinung geschehen. Es ist daher das Ziel, beleuchtungstolerante Oberflächeneigenschaften zu bestimmen und ihre Eignung für den bildbasierten Vergleich zu untersuchen.

3.2 Arbeitsprogramm

Das Forschungsvorhaben mit der geplanten Gesamtlaufzeit von 4 Jahren kann in 3 Teilschritte mit steigendem Forschungsrisiko gegliedert werden. Die einzelnen Teilschritte unterscheiden sich vor allem durch den Grad, in dem auf die Beleuchtung Einfluß genommen werden kann. Da das vorgeschlagene Thema eines PMD/Lichtfeld-basierten Erkennungssystems nach dem Prinzip der Analyse durch Synthese neuartig ist, wird in den ersten zwei Jahren ein Prototyp erstellt, der die Durchführbarkeit des Ansatzes für eine überschaubare Aufgabenstellung demonstriert. Dies ist unten als Teilschritt A aufgeführt. In Teilschritt B soll die Modellbildung und die Szenenanalyse für eine zeitlich unveränderliche, nicht gestaltbare Beleuchtung untersucht werden. Teilschritt C behandelt den Fall unterschiedlicher Beleuchtung während Modellbildung und Analyse. Aufgrund der Fülle ungelöster Probleme, die mit diesem Teil verbunden sind, ist dieser Schritt wissenschaftlich besonders interessant.

3.2.1 Teilschritt A: Demonstration der Durchführbarkeit des Ansatzes

Ziel dieses Teilschritts ist der Aufbau eines funktionierenden Gesamtsystems. Die Beleuchtungssituation in diesem Teilschritt soll vollständig selbst gestaltet werden können und während der Modellgewinnung und Objekterkennung identisch sein. Dies kann z.B. durch Verwendung eines Ringblitzes als einzige Beleuchtungsquelle sowohl bei der Modellakquisition als auch bei der Objekterkennung erfolgen.

Computergraphik		Echtzeit-Lernsysteme	
1 Grundlagenentwicklung und Schnittstellenspezifikation.			
1.1	Definition der gemeinsamen Aufgabenstellung		
1.2	Systemdesign (Schnittstellen und Datenformate)		
1.3	Aufbau und Kalibrierung des binokularen Systems		
1.4a	Fertigstellung des Simulationsprogramms für synthetische Daten	1.4b	Identifizierung der Modellparameter
1.5	Beschaffung von Beispieldatensätzen		
2 Entwicklung des Prototypen			
2.1a	Entwicklung des Bildsynthese-Moduls	2.1b	Festlegung und Vermessung der Zielfunktion
2.2a	Entwicklung der Modell-Akquisition auf Basis idealer Daten	2.2b	Auswahl des Optimierungsverfahrens
2.3a	Entwicklung der Modell-Akquisition für reelle Daten	2.3b	Implementierung der Objekterkennung
3 Implementierung und Test des Gesamtsystems			
3.1	Auswertung der synthetischen/echten Bilder		
3.2	Versuchsreihen		
3.3	Beurteilung und Strategieplanung für die folgenden Teilschritte		
4 Anpassung und Optimierung des Systems			
4.1a	Kompensation sensorspezifischer Fehler	4.1b	Auswahl geeigneter Verfahren zur Automatisierung der initialen Schätzung
4.2a	Modellanalyse und Postprocessing	4.2b	Implementierung ausgewählter Verfahren
4.3a	Implementierung von Datenkompressionsverfahren	4.3b	Evaluierung und Dokumentation der Erkennungsleistung

zu 1.1 Definition der Aufgabenstellung

Dies bezieht sich auf die Definition der Anforderungen an das Erkennungssystem und das Erstellen von Fallbeispielen. Für den Anfang ist die Definition einer überschaubaren Aufgabenstellung mit wenigen Freiheitsgraden vorteilhaft, z.B. zunächst Erkennung der Geometrie und des Typs starrer, nicht spiegelnder Objekte.

Die für den Analyse-durch-Synthese-Zyklus notwendige initiale Schätzung kann anfangs durch Hilfestellung des Benutzers (manuelle Eingabe) geschehen. Eine gute initiale Schätzung ist wichtig für die schnelle Konvergenz des Erkennungsvorgangs. Die Automatisierung der Schätzung führt schnell auf das Objekterkennungsproblem im allgemeinen Fall. Um die Anzahl der Lösungsansätze zu begrenzen, sollte diese Frage erst in einer späteren Projektphase angegangen werden, in der die Anforderungen an die initiale Schätzung klarer gestellt sind und quantifiziert werden können.

zu 1.2 Systemdesign (Schnittstellen und Datenformate)

Um die reibungslose Zusammenarbeit aller Beteiligten zu gewährleisten, muss am Anfang ein Systemdesign erarbeitet werden, das die notwendigen Komponenten des Systems beschreibt und die erforderlichen Schnittstellen identifiziert.

zu 1.5 Beschaffen von Beispieldatensätzen

Um die Projektabschnitte Lichtfeldrekonstruktion und Objekterkennung parallel bearbeiten zu können, müssen Beispieldatensätze vorliegen, die anstelle echter Sensordaten

verwendet werden können, um den Erkennungsteil während der Implementierung zu testen.

zu 2.1b Festlegung und Vermessung der Zielfunktion

Die Zielfunktion wird im Analyse-durch-Synthese-Zyklus optimiert. Sie ist ein Maß für die Ähnlichkeit zwischen echten und synthetischen Ansichten und kann auf Pixelbasis, auf den Modellparametern, oder aus beiden gemischt berechnet werden.

Die Vermessung der Zielfunktion über dem Modellparameterraum gibt Anschluss über die Größe des Konvergenzradius der Optimierung. Sie legt daher fest, wie gut die initiale Schätzung sein muss und erlaubt die Wahl des Optimierungsverfahrens für den Analyse-durch-Synthese-Zyklus.

zu 2.2b Auswahl des Optimierungsverfahrens

In Frage kommen beispielsweise numerische Verfahren (z.B. Gauß-Newton wie in [14]), die rekursive Vermessung des Parameterraums (vgl. [15]) bei mehreren Maxima der Zielfunktion oder die Untersuchung des optischen Flusses (ähnlich [16]), falls dieser Ansatz zu einer schnelleren Konvergenz führt als ein rein auf der Zielfunktion basiertes Verfahren.

zu 4.1b Automatisierung der initialen Schätzung

Mit den Erfahrungen zur Handhabung des Modells und Kenntnis des Verlaufs der Zielfunktion über dem Modellparameterraum ist es möglich, die initiale Schätzung zu automatisieren.

Während des Analyse-durch-Synthese-Zyklus müssen Modelle verschiedener Objekte an die Szene angepasst werden. Die Bestimmung des Objekttyps ist ein integraler Bestandteil des Analysevorgangs und kein zusätzlicher Klassifikationsschritt am Ende der Geometrieanalyse.

3.2.2 Teilschritt B: Anpassung des Systems an statische Beleuchtungssituationen

Ziel dieses Teilschritts ist die Adaption sowohl der Modellgewinnung als auch der Objekterkennung an statische, unterschiedliche Beleuchtungssituationen. Die Umgebungsbeleuchtung soll in diesem Teilschritt durch Aufnahme von Spiegelkugeln oder 360°-Panoramakameras vermessen werden.

Computergraphik	Echtzeit-Lernsysteme
5 Statische, nicht selbst gestaltbare Beleuchtung	
5.1 Vermessung der Umgebungsbeleuchtung (z.B. durch Spiegelkugel, 360°-Kamera)	
5.2 Evaluierung des Beleuchtungseinflusses auf die Objekterkennung	
5.3a Untersuchung möglicher Ansätze zur Kompensation des Beleuchtungseinflusses durch Anpassung des Modells	5.3b Untersuchung möglicher Ansätze zur Kompensation des Beleuchtungseinflusses durch Anpassung des Optimierungsverfahrens
5.4 Kombination/Integration erfolgreicher Lösungen	
5.5 Versuchsreihen	

zu 5.1 Vermessung der Beleuchtung

Hier soll eine praktikable Methode zur Vermessung der Umgebungsbeleuchtung entwickelt werden (z.B. Aufnahme einer spiegelnden Kugel mit einer Kamera). Dieser Schritt liefert auch Informationen über die Konstanz oder das Rauschen der Beleuchtung.

zu 5.3 Kompensation des Beleuchtungseinflusses

Die neue Beleuchtungsinformation muss logisch in das System integriert werden. Hier werden systematisch Möglichkeiten untersucht, Änderungen der Beleuchtungsumgebung sowohl auf Seiten des Modells als auch auf Seiten der Objekterkennung zu kompensieren. Das Gesamtsystem wird modifiziert und getestet.

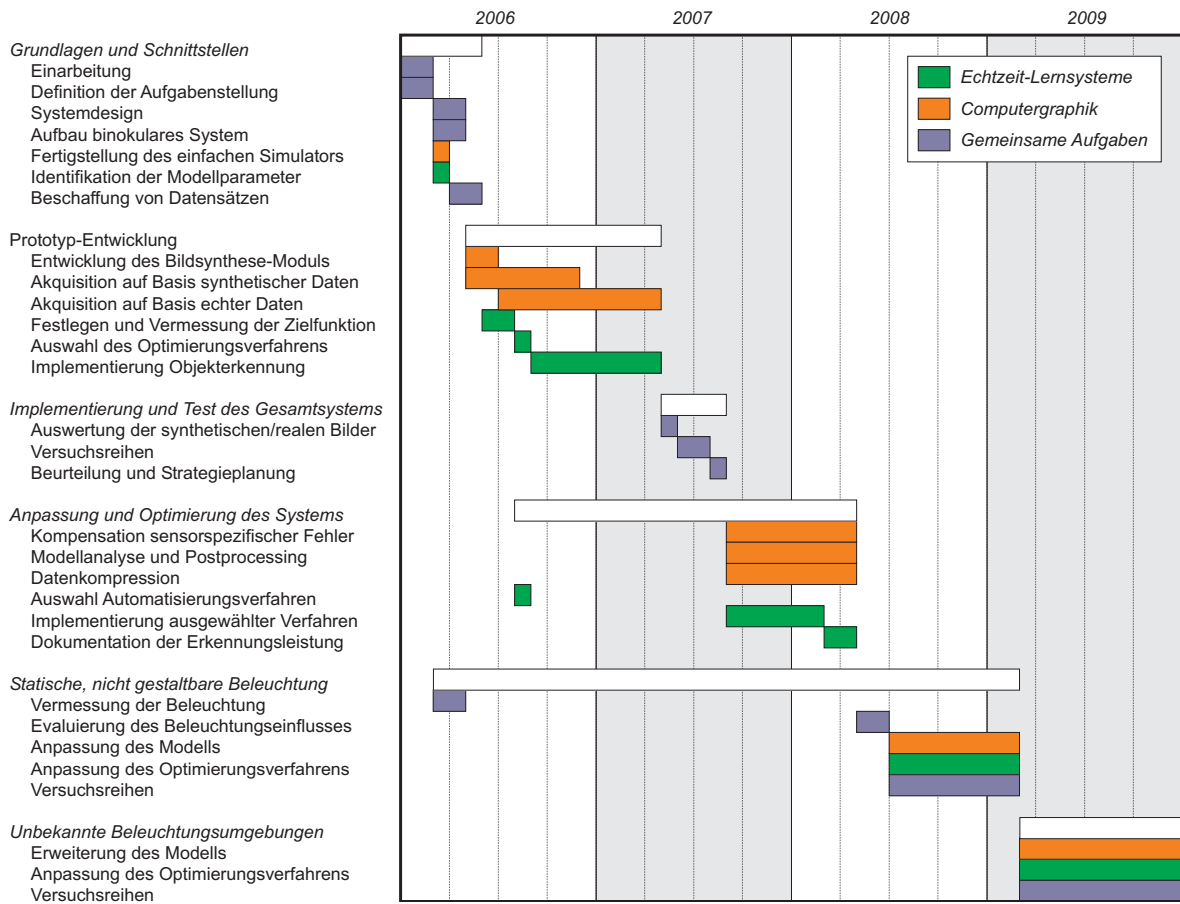
3.2.3 Teilschritt C: Anpassung des Systems an unbekannte Beleuchtungssituationen

Ziel dieses Teilschritts ist die Untersuchung von Möglichkeiten das Objekterkennungssystem an völlig unbekannte Beleuchtungssituationen anzupassen. Aufbauend auf den Erkenntnissen der ersten beiden Teilschritte sollen Möglichkeiten erarbeitet werden, auf die Vermessung der Umgebungsbeleuchtung zumindest während der Objekterkennung zu verzichten. Auf Seiten der Modellakquisition sind Ansätze um beleuchtungsunabhängige Material- und Oberflächeneigenschaften in das Modell zu integrieren. Auf Seiten der Objekterkennung müssen Ähnlichkeitsmaße gefunden werden, die beleuchtungstolerant sind ohne die Vorteile des Photorealismus zu verlieren.

Computergraphik		Echtzeit-Lernsysteme	
6 Unbekannte, evtl. dynamische Beleuchtungssituationen			
6.1a	Untersuchung möglicher Ansätze durch Erweiterung des Modells	6.1b	Untersuchung möglicher Ansätze durch Anpassung des Optimierungsverfahrens
6.2	Kombination/Integration erfolgreicher Lösungen		
6.3	Versuchsreihen		

3.2.4 Zeitliche Einordnung der Aufgaben

Nachstehende Tabelle zeigt die zeitliche Einordnung der Einzelaufgaben in die verschiedenen Quartale im Laufe der Projektbearbeitung:



Um die Entwicklung der Modellakquisition und der Objekterkennung im ersten Teilschritt voneinander zu entkoppeln, wird zunächst auf simulierte Lichtfeld-Daten zurückgegriffen. Eine sehr einfache Simulation ist bereits implementiert (siehe Vorarbeiten). Komplexere Simulationsdaten werden im Rahmen der Teilprojekte *2D3DProc* und *PMD-Sim* erstellt. Die Teilaufgaben der Lichtfeld-Akquisition und der Objekterkennung können somit von Anfang an parallel bearbeitet werden. Dies hat den Vorteil, dass Erkenntnisse und Fortschritte bei der Objekterkennung direkt für die Akquisition genutzt werden können und umgekehrt.

Literatur zu Zielen und Arbeitsprogramm

- [1] J. Kniss, M. Hadwiger, K. Engel, and C. Rezk-Salama. High quality volume graphics on consumer pc hardware. In *ACM SIGGRAPH Course Notes*, 2002.
- [2] C. Rezk-Salama. *Volume Rendering Techniques for General Purpose Graphics Hardware*. Dissertation, University of Erlangen-Nuremberg, 2002.
- [3] C. Rezk-Salama, M. Scheuring, G. Soza, and G. Greiner. Fast volumetric deformation on general purpose hardware. In *Proc. SIGGRAPH/Eurographics Workshop on Graphics Hardware*, 2001.
- [4] C. Rezk-Salama, K. Engel, M-Bauer, G. Greiner, and T. Ertl. Interactive volume rendering on standard pc graphics hardware using multi-textures and multi-stage rasterization. In *Proc. SIGGRAPH/Eurographics Workshop on Graphics Hardware*, 2000.

- [5] C. Rezk-Salama, P. Hastreiter, G. Greiner, and T. Ertl. Non-linear registration of pre- and intraoperative volume data based on piecewise linear transformations. In *Proc. Vision, Modelling, and Visualization (VMV)*, 1999.
- [6] E. Boyer and J.-S. Franco. A hybrid approach for computing visual hulls of complex objects. *Computer Vision and Pattern Recognition (CVPR'03)*, 1:695–701, June 2003.
- [7] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3D model acquisition. In *ACM SIGGRAPH*, pages 438 – 446, 2000.
- [8] R. Tönjes, O. Grau, and R. Koch. Analyse durch Synthese Modellierung von 3D Objekten in Stereobildfolgen. In *Workshop of Visual Computing, Darmstadt, Germany*, 1994.
- [9] Sven Fleck, Peter Biber, and Andreas Schilling. Direct estimation of pose and illumination - an analysis by synthesis approach. In *7th Tübingen Perception Conference (TWK 2004)*, 2004.
- [10] C. Theobalt, J. Carranza, M. Magnor, and H.-P. Seidel. A parallel framework for silhouette-based human motion capture. *Proc. Vision, Modeling, and Visualization (VMV-2003)*, Munich, Germany, pages 207–214, November 2003.
- [11] Roman Calow. *Markerlose Ganganalyse mit einem Multikamerasystem*. Doktorarbeit, Otto-von-Guericke-Universität Magdeburg, 2005.
- [12] Marcus Magnor. 3D-TV - Computergraphik zwischen virtueller und realer Welt. *Informatik Spektrum* 27, pages 497–503, 2004.
- [13] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Proc. ACM SIGGRAPH*, 2000.
- [14] K. V. Burckhardt. *Locating implants with respect to the bone in diagnostic x-ray images of the pelvis*. Dissertation, Technische Wissenschaften ETH Zürich, 2001. Nr. 14262.
- [15] M. W. Lee and S. Ranganath. Pose-invariant face recognition using a 3D deformable model. In *Pattern Recognition*, volume 36, pages 1835–1846, 2003.
- [16] R. Koch. Dynamic 3D scene analysis through synthesis feedback control. In *Proc. IEEE Trans. Patt. Anal. Mach. Intell., Special issue on analysis and synthesis*, volume 15(6), pages 556–568, 1993.

3.3 Untersuchungen am Menschen

3.4 Tierversuche

3.5 Gentechnologische Experimente

finden nicht statt.

4 Beantragte Mittel

Zur Abschätzung des Gesamtvolumens wird auch eine ungefähre Kostenabschätzung für das 3. und 4. Förderungsjahr aufgeführt.

4.1 Personalkosten

1. zwei BAT2a-Stellen für die Dauer von 2 Jahren
2. Zwei studentische Hilfskräfte mit insgesamt 80h/Monat

Am Lehrstuhl für Computergraphik & Multimediasysteme wird ein einzustellender Doktorand mit der Erarbeitung interaktiver Akquisitionstechniken bildbasierter 3D Modelle be-
traut. Am Institut für Echtzeit-Lernsysteme wird ein einzustellender Doktorand bildbasierte
Methoden der Objekterkennung entwickeln. Die Doktoranden werden mit hoher Wahrschein-
lichkeit aus den Reihen der Absolventen des Fachbereichs gewonnen. Dieser Personenkreis
hat bis zum Erwerb des Diploms eine fundierte theoretische und praktische Ausbildung in
den relevanten Anwendungsfeldern erworben.

Geschätzter Aufwand im 3. und 4. Förderungsjahr:

Äquivalent zum ersten Förderzeitraum.

4.2 Wissenschaftliche Geräte

- | | | |
|-------------------------|-------------------------------|--------------------|
| 1. | PMDvision, 19k, 160x120 pixel | 5788,40 EUR |
| Einmalige Summe: | | 5788,40 EUR |

Die PMD-Kamera ist Voraussetzung für das aufzubauende binokulare System, das die
Grundlage für die Forschungsarbeiten des Antrags darstellt.

4.3 Verbrauchsmaterial

Es werden in geringem Umfang Druckkosten pro Jahr anfallen.

- | | | |
|-----------------------|--|----------------|
| 1. | Trockentinte für Farbsublimationsdrucker | 300 EUR |
| 2. | Druckaufträge für Poster u.ä. | 100 EUR |
| Summe pro Jahr | | 400 EUR |

Geschätzter Aufwand im 3. und 4. Förderungsjahr:

Äquivalent zum ersten Förderzeitraum.

4.4 Reisen

Als jährlichen Reisekosten werden in den ersten beiden Jahren veranschlagt¹:

¹Mittlere Kosten für eine der Konferenzen für zwei Teilnehmer

1.	Präsentation von Ergebnissen auf internationalen Tagungen	
1.1	EUROGRAPHICS, PacificGraphics oder SIGGRAPH	4.000 EUR
1.2	ICPR, DAGM oder IEEE PatternRecognition, ICCV	4.000 EUR
2.	Arbeitstreffen mit externen Partnern	1.000 EUR
	Summe pro Jahr	9.000 EUR

Für Publikationen von Forschungsergebnissen der Lumigraph-Akquisition sollten Reisen zu den unter 1.1 genannten Computergraphik-Konferenzen eingeplant werden. Zur Veröffentlichung von Fortschritten im Bereich der Objekterkennung sind die unter 1.2 genannten Konferenzen zur Mustererkennung wichtig.

Geschätzter Aufwand im 3. und 4. Förderungsjahr:

Äquivalent zum ersten Förderzeitraum.

4.5 Publikationskosten

Für Publikationen werden für die ersten beiden Jahre veranschlagt:

1.	Kosten für Farbseiten in Journalen oder Tagungsbänden	750 EUR
	Summe pro Jahr	750 EUR

Geschätzter Aufwand im 3. und 4. Förderungsjahr:

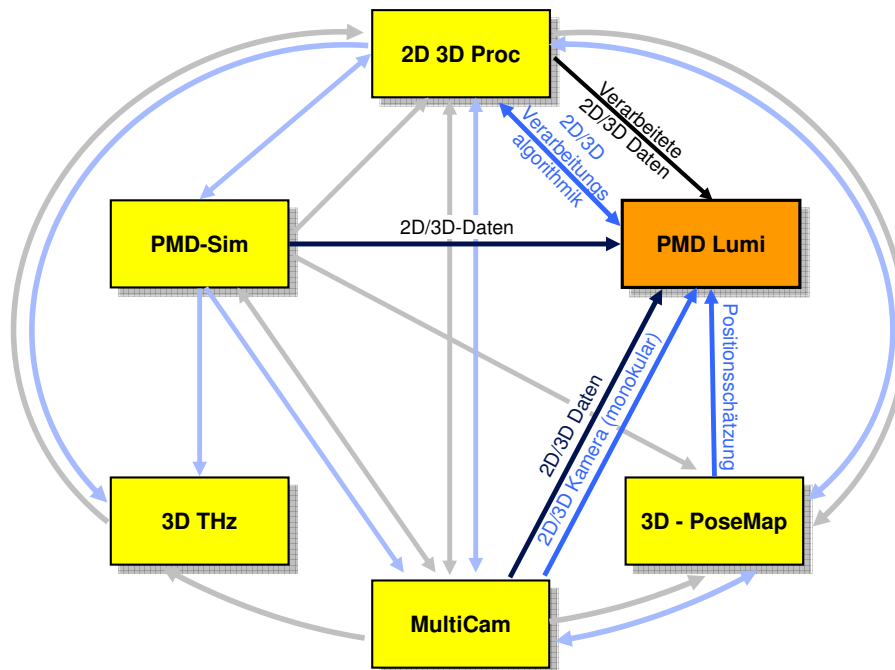
Äquivalent zum ersten Förderzeitraum.

4.6 Sonstige Kosten

Keine

5 Voraussetzung für die Durchführung des Vorhabens

Im Gesamtkonzept des Paketanspruchs *Dynamisches 3D Sehen* fügt sich **PMDLumi** als anwendungsorientiertes Forschungsvorhaben mit starker Vernetzung zu unterschiedlichen Teilprojekten ein. Die Zusammenhänge sind in der folgenden Grafik dargestellt.



Die Teilprojekte **PMD-Sim**, **2D3DProc** und **MultiCam** generieren kombinierte 2D/3D-Daten mit verschiedenen Ansätzen und unter unterschiedlichen Aspekten. **PMD-Sim** liefert aus der Simulation sowohl ideale Kameradaten als auch Bilddaten, die sensorspezifische Fehler enthalten. Derartige Daten sind für das beschriebene Forschungsvorhaben von enormem Wert um Fehlerquellen schnell zu identifizieren und die Ursache von Ungenauigkeiten zu lokalisieren. Das Teilprojekt **2D3DProc** beschäftigt sich darüber hinaus mit der Vorverarbeitung von 2D/3D-Daten, der Koregistrierung und der Kalibrierung der binokularen Systems, sowie mit der Kompensation sensorspezifischer Abbildungsfehler auf Software-Seite. Das Teilprojekt **MultiCam** beschäftigt sich mit allen Aspekten der Kameraentwicklung, insbesondere mit der Kompensation von Sensorfehlern auf Hardware-Seite. Fortschritte in diesen Teilprojekten sind daher für **PMDLumi** von unmittelbarem Nutzen. Neben dem im Antrag beschriebenen binokularen Aufbau kann im späteren Verlauf auch eine monokulare 2D/3D Kamera eingesetzt werden, wie sie im Rahmen von **MultiCam** entwickelt wird. Der Vorteil daraus wäre eine mögliche Präzisionssteigerung durch den Wegfall von Fehlern, die bei der externen Koregistrierung des binokularen Systems entstehen.

Das gemeinsame Problem des dynamischen Trackings zur Positionsschätzung der Kamera wird im Teilprojekt **3D-PoseMap** untersucht. Obwohl **PMDLumi** logisch auf der Positionsschätzung aufbaut, tritt hier jedoch keine zeitliche Abhängigkeit ein. Die im vorliegenden Antrag zu entwickelnden Akquisitionsverfahren können zu Beginn – auch wenn das zunächst eine Einschränkung der Flexibilität bedeutet – mit markerbasiertem Tracking arbeiten, wie in den Vorarbeiten bereits demonstriert wurde, alternativ auch unter Verwendung des in **MultiCam** aufgebauten Messplatzes. In der Zusammenarbeit mit **3DPoseMap** entstehen

Synergien durch den Austausch und die gegenseitige Evaluierung von Teillösungen. Einerseits müssen sowohl die Registrierungsverfahren als auch die Bildsynthese anhand idealer, synthetischer Daten (→**PMD-Sim**) getestet werden, um anschließend die sensorspezifischen Fehler (→**2D3DProc**) identifizieren und kompensieren zu können. Andererseits benötigt man für die Entwicklung der Registrierungsverfahren natürlich Möglichkeiten, die Qualität der Ergebnisse zu visualisieren, was wiederum sehr anschaulich unter Verwendung des Lumigraphen funktioniert. Die im Rahmen von **PMDLumi** zu entwickelnden Verfahren zur Lageschätzung von Objekten sind in diesem Zusammenhang wertvolle Hilfsmittel zur Validierung.

Besonders bei der Kompensation der Beleuchtungsabhängigkeit, wird der im Rahmen des Teilantrags **MultiCam** aufgebaute Messplatz von großer Bedeutung sein. Hier wird es möglich sein, Beleuchtungsbedingungen exakt zu kontrollieren, unterschiedliche Kameraaufbauten zu evaluieren, sowie Bewegungen von Kamera- und Aufnahmeobjekt im Submillimeterbereich exakt zu steuern. Die aus diesem Anwendungskontext gewonnenen Erkenntnisse über Beleuchtungsverhalten und damit verbundene sensorspezifische Fehler sollen in enger Zusammenarbeit mit **MultiCam** und den Entwicklern der Sensorik genutzt werden, um die Genauigkeit des Kamerasystems weiter zu verbessern.

5.1 Zusammensetzung der Arbeitsgruppe

1. Prof. Dr. Andreas Kolb, Lehrstuhlinhaber Computergraphik & Multimediasysteme (CG), ZESS-Mitglied
2. Prof. Dr. Kuhnert, Institutsleiter Echtzeit Lernsysteme
3. Dr. Christof Rezk-Salama, wissenschaftlicher Assistent.
4. zwei Doktoranden aus den beantragten Projektmitteln

5.2 Zusammenarbeit mit anderen Wissenschaftlern

Es wird mit den anderen Arbeitsgruppen im Rahmen dieses Paketantrages in enger Abstimmung kooperiert. Eine Zusammenarbeit besteht zudem mit folgenden anderen Wissenschaftlern bzw. Arbeitsgruppen:

Dr. Strzodka, Ceasar Research Center/Stanford University:

Simulationsalgorithmen auf programmierbarer Graphik-Hardware

Prof. Wiechert, FOMAAS², Universität Siegen: Simulation und Visualisierung von Partikelsystemen und Hochenergiemühlen unter Einsatz programmierbarer Graphikhardware.

Universitätsklinikum Erlangen: Entwicklung intuitiver Benutzerschnittstellen und spezialisierter Darstellungsverfahren für konkrete, klinische Problemstellungen im Bereich Visualisierung

²FOMAAS=Forschungsinstitut für Multidisziplinäre Analysen und Systemoptimierung

5.3 Arbeiten im Ausland und Kooperation mit ausländischen Partnern

Kooperationen besteht mit folgenden ausländischen wissenschaftlichen Einrichtungen:

Siemens Corporate Research, Princeton, USA: Entwicklung von Visualisierungsverfahren mit dem Schwerpunkt der Generierung von Transferfunktionen zur Visualisierung und Analyse tomographischer Bilddaten mit breitem Anwendungsspektrum im Bereich Medizintechnik.

VRVis Research Center, Österreich: Arbeiten im Bereich prototypischer Applikationen für die Segmentierung und Fusion medizinischer Bilddaten und zur Weiterentwicklung echtzeitfähiger Volumenvisualisierung.

5.4 Apparative Ausstattung

Das Institut für Echtzeit-Lernsysteme und der Lehrstuhl für Computergraphik & Multimediasysteme besitzen die räumliche und apparative Ausstattung zur Durchführung des beantragten Projektes. Die Rechner-Ausstattung umfasst neben PC für die Ausbildung mehrere Arbeitsplatzrechner für Mitarbeiter mit Dual-Prozessoren.

Ein HBFGE-Antrag zur Einrichtung eines Labors für Virtual Environments (DFG Antragsnummer HBFGE-013-180) wurde positiv kürzlich entschieden. Das Labor wird im Laufe des Jahres 2006 eingerichtet und in Betrieb genommen.

5.5 Laufende Mittel für Sachausgaben

Weitere Mittel für Sachausgaben, die über die speziellen Verbrauchsmaterialien unter 4.3 hinausgehen, werden von den Arbeitsgruppen selbst getragen.

5.6 Sonstige Voraussetzungen

keine

6 Erklärungen

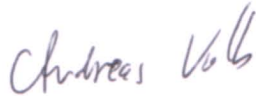
1. Es bestehen keine thematischen Zusammenhänge zwischen dem Vorhaben und Arbeiten eines am Ort befindlichen Sonderforschungsbereiches.
2. Ein Antrag auf Finanzierung dieses Vorhabens wurde bei keiner anderen Stelle eingereicht. Wenn wir einen solchen Antrag stellen, werden wir die Deutsche Forschungsgemeinschaft unverzüglich benachrichtigen.

3. Der Vertrauensdozent der Hochschule wird unverzüglich von dieser Antragstellung unterrichtet.

6.1 Max-Planck-Institut

Nicht anwendbar.

7 Unterschriften



(Prof. Dr. Andreas Kolb)



(Prof. Dr. Klaus-Dieter Kuhnert)



(Dr. Christof Rezk-Salama)

8 Verzeichnis der Anlagen

1. Angebot über eine PMD-Kamera, PMDvision, 19k, 160x120 pixel
2. Tabellarischer Lebenslauf der Antragsteller mit Schriftenverzeichnis
3. Einführender Artikel zum Photonic Mixing Device, PMD:
H. Kraft, J. Frey, T. Moeller, M. Albrecht, M. Grothof, B. Schink, H. Hess and B. Buxbaum: *3D-Camera of High 3D-Frame Rate, Depth-Resolution and Background Light Elimination Based on Improved PMD (Photonic Mixer Device)-Technologies*, OPTO, Nürnberg, Mai 2004.
4. Kurzprofil des Zentrums für Sensorsysteme (ZESS) als zentrale wissenschaftliche Einrichtung der Universität Siegen
5. IPP-Broschüre des DAAD