

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

Abschlussbericht



Prof. Dr.-Ing.
Andreas Kolb



Prof. Dr.-Ing.
Klaus-Dieter Kuhnert

Prof. Dr.-Ing. habil.
Christof Rezk Salama

Lehrstuhl für
Computergraphik und
Multimediasysteme

Institut für
Echtzeit-Lernsysteme

Universität Siegen
Hölderlinstr. 3
57068 Siegen

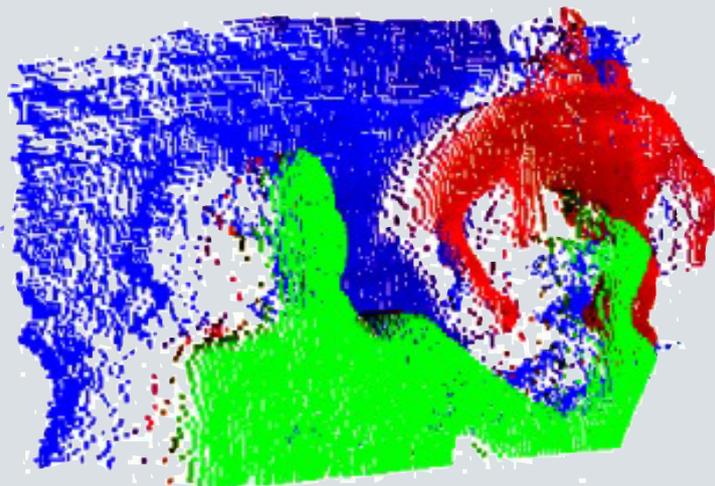
Universität Siegen
Hölderlinstr. 3
57068 Siegen

Dynamisches 3D Sehen

Prof. Dr.-Ing.
Andreas Kolb

Zentrum für
Sensorsysteme
(ZESS)

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



Prof. Dr.-Ing. habil.
Otmar Loffeld

Zentrum für
Sensorsysteme
(ZESS)

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

Inhaltsverzeichnis

1	Allgemeine Angaben	1
1.1	DFG-Geschäftszeichen	1
1.2	Antragsteller	1
1.3	Institut/Lehrstuhl	1
1.4	DFG-finanzierte wissenschaftliche Mitarbeiter	2
1.5	Thema	2
1.6	Berichts- und Förderzeitraum	2
1.7	Fachgebiet, Arbeitsrichtung	2
1.8	Verwertungsfelder	2
1.9	Am Projekt beteiligte Kooperationspartner (Name, Ort)	2
2	Zusammenfassung	2
2.1	Wesentliche Ergebnisse und erzielte Fortschritte	3
2.2	Ausblick	4
3	Arbeits- und Ergebnisbericht	4
3.1	Ausgangslage	4
3.2	Beschreibung der durchgeführten Arbeiten	5
3.3	Darstellung der erzielten Ergebnisse	7
3.4	Ausblick auf zukünftige Arbeiten	10
3.5	Interdisziplinäre Weiterentwicklung	11
3.6	Verwertungspotenzial	12
3.7	Beteiligte Wissenschaftler	13
4	Liste der im Teilprojekt entstandenen Publikationen	13
4.1	Begutachtete Publikationen	13
4.2	Weitere Veröffentlichungen	14
5	Publikationen Dritter	15

1 Allgemeine Angaben

Ergebnisbericht im Rahmen des DFG Forschungspakets PAK-73 „Dynamisches 3D Sehen mit PMD“

1.1 DFG-Geschäftszeichen

KO-2960/6-2, KU-689/6-2

1.2 Antragsteller

Andreas Kolb, Prof. Dr.-Ing.
Klaus-Dieter Kuhnert, Prof. Dr.-Ing.
Christof Rezk-Salama, Prof. Dr.-Ing.

1.3 Institut/Lehrstuhl

Prof. Dr.-Ing. Andreas Kolb, Dr.-Ing.(habil) Christof Rezk-Salama
Lehrstuhl für Computergraphik und Zentrum für Sensorsysteme (ZESS)
Fakultät für Natur- und Ingenieurwissenschaften
Department Elektrotechnik & Informatik
Universität Siegen
Hölderlinstr. 3, 57068 Siegen
Tel.: 0271 / 740 - 2404
Fax.: 0271 / 740 - 3337

Prof. Dr.-Ing. Klaus-Dieter Kuhnert
Lehrstuhl für Echtzeit-Lernsysteme und Zentrum für Sensorsysteme (ZESS)
Fakultät für Natur- und Ingenieurwissenschaften
Department Elektrotechnik & Informatik
Universität Siegen
Hölderlinstr. 3, 57068 Siegen
Tel.: 0271 / 740 - 4779
Fax.: 0271 / 740 - 4421

1.4 DFG-finanzierte wissenschaftliche Mitarbeiter

Name	Zeitraum	TVL-13 Stelle
Severin S. Todt	2/08-6/08,	50%
	7/08	25%
Jens Orthmann	05/08	25%
	06/08-09/09	50%
	05/10-07/10	50%
	10/10-11/10	50%
Lisa Brückbauer	06/08-12/09	50%
Iurie Chiosa	01/10-02/10	50%
Matthias Langer	02/08-08/10	100%
Stefan Thamke	09/10-12/10	Wiss. HK

1.5 Thema

PMDLumi: Echtzeit-Akquisition bildbasierter 3D Modelle zur Objekterkennung

1.6 Berichts- und Förderzeitraum

Förderzeitraum 1.2.2008 – 31.12.2010 (2. Projektphase)

1.7 Fachgebiet, Arbeitsrichtung

Informatik, Computergraphik, Bildverarbeitung, bildbasiertes Rendering, Objekterkennung

1.8 Verwertungsfelder

Bildauswertung, visuelle Objektlageschätzung, visuelle Objektverfolgung in Echtzeit, Objekterkennung im industriellen Umfeld, virtuelle Produktpräsentationen, Entertainment.

1.9 Am Projekt beteiligte Kooperationspartner (Name, Ort)

–entfällt–

2 Zusammenfassung

Das Projekt *PMDLumi – Analyse durch Synthese* ist eines der Teilprojekte des Forschungspaketes *Dynamisches 3D Sehen mit PMD*. Ziel dieses Teilprojektes war die Erforschung eines Analyse-durch-Synthese-Ansatzes zur echtzeitfähigen Objekterkennung und -poseschätzung. Hierbei ermöglicht ein Lichtfeldrenderer auf Basis bildbasierter Renderingmethoden die Echtzeitsynthese beliebiger Objektansichten mittels einer zuvor akquirierten, diskreten Bilddatenbasis. Die Synthese wird durch das Analysemodul angestoßen. Ein Objekt, das sich vor

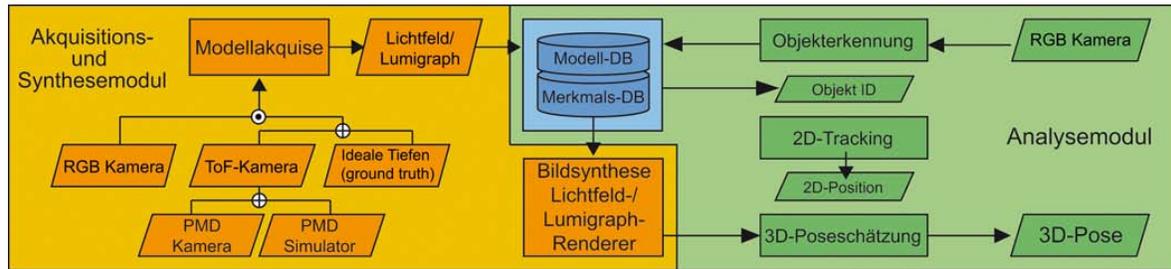


Abbildung 1: Organisationsdiagramm des PMDLumi-Projekts

einer einfachen 2D Kamera befindet, kann durch den Abgleich mit dem Objektmodell des Lichtfeldrenderers identifiziert und dessen Pose vor der 2D Kamera iterativ ermittelt werden.

Zur Generierung von Daten für den Syntheseschritt wird eine RGB-Kamera in Verbindung mit einer Time-of-Flight-Kamera als Eingabesensor benutzt. Hierbei wird auf die PMD-Technologie (Photonic Mixing Device) zurückgegriffen.

Das entwickelte Gesamtsystem lässt sich in einen Offline-Akquisitionsschritt und einen in Echtzeit ablaufenden Bereich unterteilen. In der Offline-Phase werden Modelldaten der Objekte, die später erkannt werden sollen, gesammelt und entsprechende Merkmale in Datenbanken abgelegt. Insbesondere werden in dieser Phase die 3D Modelldaten für den Lichtfeldrenderer generiert. Im Echtzeitschritt finden die Objekterkennung, das Modell-Rendering und die 3D-Poseschätzung statt.

Für die Objekterkennung und Poseschätzung in Echtzeit werden zunächst Objektmerkmale berechnet. Diese werden mit der in der Offline-Phase berechneten Merkmalsdatenbank abgeglichen. Wurde ein Objekt erfolgreich erkannt, wird es von der 2D Kamera verfolgt und das passende Modell des erkannten Objektes in den Lichtfeldrenderer geladen. Daraufhin werden die Syntheseparameter der virtuellen 3D-Kamera iterativ geschätzt. Wurde eine hinreichende Übereinstimmung erzielt, können die Poseparameter des realen Objekts vor der realen Kamera durch die Parameter der virtuellen 3D Kamera ausgedrückt werden. Im Online-Betrieb ist also keine weitere spezielle Sensorik nötig, um Abstand und Lage des Objekts bezüglich der aufnehmenden Kamera zu messen. Das System erreicht bei angelernten Objekten mit einfachen, handelsüblichen Webcams genaue und robuste Ergebnisse.

2.1 Wesentliche Ergebnisse und erzielte Fortschritte

Modellgewinnung und Bildsynthese Zu Beginn der zweiten Projektphase hat sich herausgestellt, dass die Tiefendaten der PMD-Kamera für eine ausreichend genaue Rekonstruktion der 3D-Geometrie qualitativ nicht hinreichend ist. Deshalb wurden die Lichtfeld-Algorithmen zunächst basierend auf idealen Tiefendaten weiter entwickelt. Hieraus hat sich eine neue Forschungsrichtung ergeben, die den Einsatz von bildbasierten Verfahren zur effizienten und hochqualitativen Bildsynthese im Falle von aufwändigen Material- oder Objekteigenschaften ermöglicht. Exemplarisch wurde die Nutzung von Lichtfeldern für schmalbandige Übertragungswege (Internet) und für hoch-performante Anwendungen (Computerspiele) gezeigt.

Hinsichtlich der Beleuchtungstoleranz von Lichtfeldern wurde ein Ansatz zum Relighting mit Punktlichtquellen entwickelt. Der Ansatz basiert auf der Übertragung der Polynomi-

al Texture Mapping-Technik (PTM) [6] auf sphärische Lichtfelder. Hiermit kann ein mit Punktlichtquellen akquiriertes Lichtfeld mit variabler Punktlichtquelle neu ausgeleuchtet werden.

Die Fragestellung, wie sphärische Lichtfelder mit Environment Maps in Beleuchtungsumgebungen integriert werden können, die sich von der Aufnahmesituation unterscheiden, wird gegenwärtig und über das offizielle Ende der DFG-Förderung hinaus noch bearbeitet werden.

Objekterkennung und Lageschätzung In der ersten Projektphase wurde die Beleuchtung konstant gehalten. In der jetzt abgeschlossenen zweiten Phase wurde das System auf variierende Beleuchtung erweitert. Dabei war das System ursprünglich nur für die Verwendung unter Bürobedingungen vorgeschlagen worden. In den Versuchen zeigte sich aber eine unerwartete Robustheit, so dass eine vereinfachte Variante auf den mobilen Außenbereichsroboter AMOR der AG Kuhnert portiert wurde. Damit konnte über die eigentlich geplante Unempfindlichkeit gegen Beleuchtungseinflüsse hinaus sogar die Außenbereichstauglichkeit (d.h. unter deutlichen Beleuchtungsvariationen) des Ansatzes nachgewiesen werden. Plangemäß sollten die Objekte nur vor homogenem Hintergrund platziert werden. Das implementierte Verfahren ist jedoch robust genug, um auch in normaler Büroumgebung zuverlässige Ergebnisse zu liefern (siehe Abb. 4 als typisches Beispiel).

2.2 Ausblick

Die ersten Ergebnisse aus der beleuchtungstoleranten Lichtfeldsynthese sind sehr vielversprechend, weshalb die Forschungsarbeiten an diesem Thema weiter geführt werden. Parallel werden die Optimierungsverfahren für PMD-basierte Lichtfelder fortgeführt. In diesem Zusammenhang wird erwartet, dass sich die Ergebnisse insbesondere in dem Bereich Cultural Heritage als sehr nutzbringend erweisen werden.

Die unerwartet hohe Robustheit und die Unempfindlichkeit gegenüber Beleuchtungseinflüssen machen die Ergebnisse des Analysemoduls für Anwendungen in interaktiven Spielen oder der mobilen Robotik nutzbar. Mit einer qualitativ höherwertigen Kamera und einer Präzisionseichung sollten noch höhere Messgenauigkeiten erreichbar sein, womit sich auch Anwendungen in der 3D Qualitätsprüfung und der Automatisierungstechnik eröffnen würden.

3 Arbeits- und Ergebnisbericht

3.1 Ausgangslage

Zu Beginn der zweiten Projektphase wurde ein Proof-of-Concept in Form eines ersten Komplettsystems (s. Abb. 1) realisiert. Dieses Komplettsystem umfasst sowohl die Offline-Module (Lichtfeld-Akquisitionsmodul, Modelldatenbank, Merkmalsdatenbank) als auch die während der Objekterkennung benötigten Echtzeit-Module (Lichtfeld-Renderer, Objekterkennung, 2D-Tracking, 3D-Poseschätzung). Das Renderingmodul arbeitet sowohl mit synthetischen als auch mit realen Daten. Die Qualität der realen Bilddaten hat sich jedoch für die Feature-Erkennung als nicht ausreichend herausgestellt, weil aufgrund der noch zu geringen Auflösung der PMD-Tiefendaten die Lichtfeldmodelle zu sehr artefaktbehaftet sind

(Ghosting). Durch die Nutzung von synthetischen Renderdaten wurde aber ein Proof-of-Concept erzielt, der von der Güte der real akquirierten Daten zunächst unabhängig ist.

In der ersten Projektphase von *PMDLumi* wurde davon ausgegangen, dass die Beleuchtungssituationen zur Akquisitions- und zur Objekterkennungszeit identisch sind. Das Ziel der zweiten Projektphase war die Anpassung des Gesamtsystems an variierende Beleuchtungssituationen. In diesem Zusammenhang wurden drei aufeinander aufbauende Arbeitspakete definiert:

1. Modellgewinnung und Objekterkennung in konstanter Beleuchtung
2. Objekterkennung in dynamischer Beleuchtung
3. Modellgewinnung und Objekterkennung in dynamischer Beleuchtung

Die Herausforderung auf Seiten der Synthese war es, effiziente Datenstrukturen und Algorithmen zu finden, die den Austausch von Beleuchtungsumgebungen bei Lichtfeldern ermöglichen. Auf Seiten der Analyse stellten sich die Fragen, welche Bildmerkmale für eine zuverlässige Objekterkennung eingesetzt werden sollten, wie Objekte robust und in Echtzeit vor der Kamera verfolgt werden, und was ein gutes Übereinstimmungsmaß bei der 3D Poseschätzung bezüglich Modellparametern und extrahierten Objektmerkmalen sein kann.

Die Antworten auf obige Fragen werden ausführlich in Abschnitt 3.3 besprochen und die erzielten Ergebnisse präsentiert.

3.2 Beschreibung der durchgeführten Arbeiten

3.2.1 Wesentliche Erkenntnisse

Nutzbarmachung der Lichtfeldtechnologie aus Projektphase I

Zusätzlich zum ursprünglichen Konzept haben sich im wesentlichen zwei neue Schwerpunkte ergeben, die jeweils erfolgreich publiziert werden konnten. Es handelt sich dabei um die Nutzbarmachung der sehr effizienten und hochqualitativen sphärischen Lichtfelder für webbasierte Anwendungen [RSTK08], [TRSBK08] (s. Abschnitt 3.3.1), und für die Spieleindustrie [TRSK08] (s. Abschnitt 3.3.2). Beide Ansätze basieren auf einer effizienten Repräsentation und Darstellung von Lichtfeldern für Objekte mit komplexen Materialeigenschaften.

Beleuchtungstolerante Lichtfelder

Im Hinblick auf beleuchtungstolerante Lichtfelder wurde ein Verfahren entwickelt, mit dem es möglich ist, ein sphärisches Lichtfeld mittels einer Punktlichtquelle beliebig neu zu beleuchten (siehe [BRSK10] und Abschnitt 3.3.3). Darüber hinaus ist es möglich, in diesem Verfahren die Oberflächeneigenschaften des akquirierten Modells, insbesondere die Normale und die Reflexionseigenschaften aus den Bilddaten auch auf der Ebene von Meso-Strukturen abzuschätzen.

Objekterkennung

Für die Objekterkennung kamen komplexe Bildmerkmale wie SIFT [5] und später SURF [1] zum Einsatz. Um den Aufwand während der Online-Objekterkennung zu reduzieren, wurden innovative Verfahren entwickelt, wie z.B. eine hierarchische, Entscheidungsbaum-basierte Vorsortierung der in der Datenbank vorhandenen Objekte, sowie ein Binarisierungsverfahren

für SURF Merkmale. Letzteres Verfahren sorgt nicht nur für eine wesentliche Reduktion der Speicheranforderungen der Merkmalsdatenbank (um den Faktor 1:32), sondern auch für eine beschleunigte Korrespondenzanalyse einzelner Merkmale. Der hohe Kompressionsgrad ist dabei nicht mit einer Minderung der Erkennungsleistung verbunden.

3.2.2 Probleme in der Durchführung

Im Gegensatz zu den Darstellungen im Antrag haben sich die Randbedingungen für die zweite Projektphase entscheidend verändert. Einerseits wurden beantragte Geräte nicht genehmigt, was die Bearbeitung des Arbeitspaketes zur Erfassung dynamischer Lichtfelder stark beeinflusst hat. Konkret ergaben sich folgende Probleme:

Messung einer dynamischen Beleuchtungsumgebung: Für die Szenarien mit dynamischer Beleuchtung wurde zum Antragszeitpunkt davon ausgegangen, dass die Beleuchtungsumgebung durch zusätzliche Sensorik (Umgebungs-kamera Point Grey Ladybug2) vermessen werden kann, die jedoch nicht bewilligt wurde.

In diesem Zusammenhang wurde in der zweiten Projektphase zunächst auf synthetische Daten und teilweise auf Daten eines PMD-Simulators zurückgegriffen.

Exakte Ansteuerung von Kamerapositionen: Eine Apparatur zum genauen Ansteuern von Raumpositionen außerhalb der konvexen Hülle eines Objekts (beantragter *Peace River Studio Object Maker*) war nicht verfügbar.

Daraus ergab sich die Notwendigkeit, das Kamerasetup markerbasiert zu tracken. Die Lösung des Trackingproblems war sowohl hardware- als auch softwareseitig zeitaufwändig. Das Ergebnis ließ hinsichtlich der Genauigkeit der vermessenen Positionen im Vergleich zu einem ansteuerbaren Setup zu wünschen übrig. Eine alternative Schätzung der Kamera-Pose mittels der Verfahren aus dem Teilprojekt 3D-PoseMap hat sich aufgrund der hohen Genauigkeitsforderungen bei der Lichtfeldsynthese als nicht praktikabel erwiesen.

Darüber hinaus haben die bereits erwähnten Ungenauigkeiten in den PMD-Tiefendaten, vor allem im Bereich der Objekt-Silhouetten [Tod09], die Erforschung von Optimierungsverfahren zur Verbesserung der Tiefendaten motiviert. Diese Arbeiten basieren auf der Ausnutzung der Redundanz in den Tiefenkarten der sphärischen Lichtfelder und sind noch nicht abgeschlossen.

Im Zusammenspiel der Komponenten für das Gesamtsystem wurde entsprechend auf vergleichsweise einfache Objekte in der Erkennung zurückgegriffen, deren Lichtfelder sich in ausreichender Genauigkeit und Qualität durch eine synthetische Akquisition gewinnen lassen, bzw. in Form von externen Referenzdatenbanken, wie der Coil100-Datenbank [9] zur Verfügung stehen.

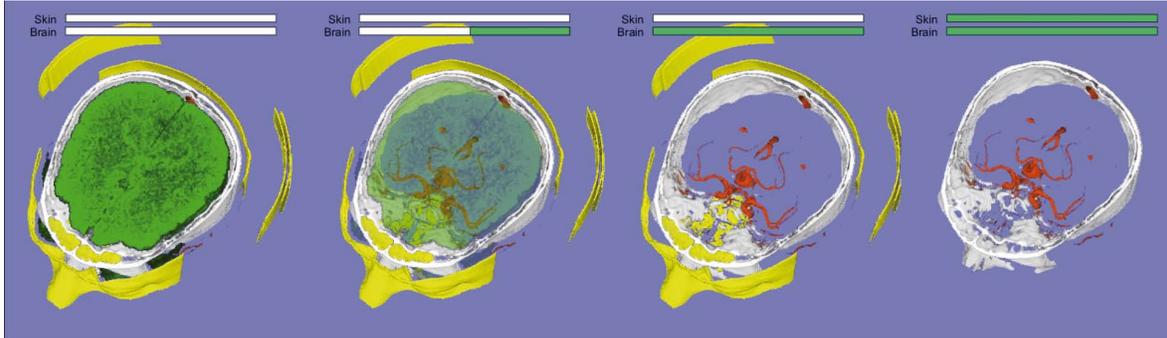


Abbildung 2: Multi-Lightfield, erzeugt aus mehreren Schichten eines Computertomographie-Datensatzes: Separate Lichtfelder für Knochen/Gefäße, Haut und Hirn [RSTK08].

3.3 Darstellung der erzielten Ergebnisse

3.3.1 Nutzung von Lichtfeldern für webbasierte Anwendungen

Light Field Galleries [RSTK08] zielen darauf ab, hochspezialisierte Visualisierungen von z.B. Volumendatensätzen, durch eine leichtgewichtige, interaktive Präsentationsform auch für Laien nutzbar zu machen. Light Field Galleries werden dabei als Ergänzung zu den ursprünglichen (hochspezialisierten und nur für Experten zugänglichen) Visualisierungsmethoden betrachtet und können überall dort zum Einsatz kommen, wo einfache Bedienbarkeit und schnelle Abrufbarkeit gefragt sind. Das Funktionsprinzip basiert dabei auf dem Compositing, Blending und anschließenden Raycasting mehrerer Lichtfelder zum Erzeugen eines sogenannten Multi-Lightfields (s. Abb. 2). Die einzelnen Lichtfelder können zum Beispiel aus hochqualitativen Volumenrenderings von Computertomographie-Datensätzen erzeugt werden [7], die sich in der Einstellung der Transferfunktionen unterscheiden. Die Lichtfeldrepräsentation gibt die hochqualitativen visuellen Effekte der originalen Volumenrenderings wieder, ist dabei jedoch, anders als die originalen Algorithmen, echtzeitfähig. Eine Anwendung wäre z.B. im Bereich der medizinischen Ausbildung denkbar, da sich Lichtfelder hier, unter Beibehaltung der Interaktivität und Explorationsmöglichkeiten, als effizientes Datenaustausch- und/oder Speicherformat für Volumendaten erweisen.

Eine Fortentwicklung stellen *Progressive Light Fields* [TRSBK08] dar, bei denen das bestehende Lichtfeld-Datenformat hierarchisch organisiert wurde um ein Level-of-Detail Rendering zu realisieren. Hierzu wurde ein Client-Server-basiertes Framework entwickelt. Angepasst an den Blickpunkt des Betrachters kann die Datenrate und somit die Qualität des Lichtfelds angepasst werden. Die progressive Verfeinerung verbessert die Performanz deutlich hinsichtlich Speicherbedarf und visuellem Feedback.

3.3.2 Nutzung von Lichtfeldern in Spielen

Für die Anwendung der Lichtfeldtechnologie in Spielen wurde ein einfaches Compositingverfahren basierend auf per-fragment-Tiefe entwickelt, welches das korrekte Compositing eines Lichtfelds und einer polygonalen Szene ermöglicht. Der geringe Speicherbedarf in Verbindung mit der schon in [TRSBK08] vorgestellten Level-of-Detail Strategie ermöglicht es außerdem, mehrere Zustände eines Lichtfelds zu speichern. So ist es möglich, Deformationen und Animationen zu realisieren.



Abbildung 3: Lichtfeld mit drei unterschiedlichen Beleuchtungen (a), 3D-PTM ohne (b, oben) und mit (b, unten) blickabhängigem, spekularem Shading

Ein weiterer Beitrag im Games-Kontext war eine Produktionspipeline, die es Spieleentwicklern erleichtert, aus herkömmlichen 3D Modellen auf einfache Art Lichtfelder zu generieren. Hierzu wurde ein Plugin für die Software Autodesk Maya entwickelt, mit dem - unter Ausnutzung aller in Maya verfügbaren Shading- und Rendering-Algorithmen - automatisiert sowohl einfache Lichtfelder als auch Animationssequenzen erstellt werden können.

Die Publikation *Light Field Rendering for Games* [TRSK08] wurde mit dem Best Student Presentation Award ausgezeichnet.

3.3.3 Beleuchtungstolerante Lichtfelder

Die Neu- oder Wiederbeleuchtung eines Lichtfeldes wurde zunächst durch die Variation der Beleuchtungsrichtung realisiert. Hierbei wurde davon ausgegangen, dass Licht nur aus einer bestimmten Richtung auf das zu beleuchtende Objekt einstrahlt (directional light).

Mit *PTM-Lightfields* [BRSK10] wurde ein Verfahren entwickelt, ein Lichtfeldmodell ohne eine explizite Speicherung mehrerer Beleuchtungszustände wie in [TRSK08], zu gewinnen. Dieses Lichtfeldmodell kann im Syntheseschritt in Echtzeit neu beleuchtet werden. Basis dafür ist die Technik der Polynomial Texture Maps (PTMs, [6]), welche eine beobachtete Beleuchtung als Polynom approximiert, dessen Koeffizienten speichereffizient in Texturen abgelegt werden können. Eine beliebige neue Beleuchtung kann erzeugt werden, indem die neue Lichtposition in das gespeicherte Polynom eingesetzt und die Pixelfarbe in Abhängigkeit der neuen Beleuchtung rekonstruiert wird.

PTM-Lightfields erweitern den ursprünglichen, zweidimensionalen PTM-Ansatz um die Gültigkeit im dreidimensionalen Raum, indem die Texturen blickrichtungsabhängig ausgewertet werden. PTM-Lightfields profitieren von der Stärke der PTM-Technologie, gerade feine Mesostrukturen eines aufgenommenen Objekts bildbasiert rekonstruieren zu können. Durch Abschätzung der Materialkonstanten k_d , k_s aus den Bilddaten und durch Variation des spekularen Exponenten s können außerdem im Modellgewinnungsstadium unterschiedliche Oberflächeneigenschaften eines Modells rekonstruiert werden.



Abbildung 4: Beispiele verwendeter Testobjekte.

3.3.4 Objekterkennung

Zunächst wurden zur Objekterkennung auf Basis von SIFT-Merkmalen [5] auf der Coil100 Bilddatenbank [9] evaluiert, in die sich eigene Testobjekte integrieren lassen (siehe Abb. 4). Die meisten der im Projekt eingesetzten Testobjekte wiesen eine mehr oder minder starke Texturierung auf, auf die SIFT Merkmale besonders ansprechen. Zur Beschleunigung der SIFT-Operatoren und deren Anwendung auf größeren Objektdatenbanken, wurde ein entscheidungsbaum-basiertes Verfahren entwickelt, um den Suchraum zu begrenzen [TLRS⁺08, LK08]. Zur Belehrung wurden einfache Farbmerkmale verwendet, die zwar nur eine Erkennungsleistung von ca. 76% aufwiesen, allerdings auch nur als Vorverarbeitungsschritt dienten, um die nachfolgende Objekterkennung mittels SIFT zu beschleunigen. Die eigentliche Objekterkennung erfolgt durch Berechnung der SIFT-Merkmale des aktuellen Kamerabildes und Abgleich mit den Merkmalen einer Merkmalsdatenbank. Die Merkmalsdatenbank wurde offline aus charakteristischen Objektansichten der zuvor akquirierten 3D-Modelle erstellt.

3.3.5 2D-Tracking

Einen weiteren Baustein des Analyse-durch-Synthese-Systems stellt das 2D-Objekttracking dar. Hierbei werden erkannte Objekte kontinuierlich verfolgt und die ermittelten Daten an die Poseschätzung weitergereicht. Damit kann eine zeitaufwändige, initiale Objektsuche in der Merkmalsdatenbank vermieden werden. SIFT-Merkmale sind auch für die Objektverfolgung geeignet, wobei die berechneten Merkmale mit denen für das erkannte Objekt in der Merkmalsdatenbank abgeglichen werden, womit die Hälfte an Rechenzeit eingespart wird. Da die Laufzeit gegenüber traditionellen Trackingverfahren nicht ausreichend war, wurde zu SURF-Merkmalen übergegangen, da diese bei vergleichbarer Objekterkennungsleistung deutlich schneller zu berechnen sind [5]. Die Merkmalsvektorgroßen wurden durch ein innovatives Binarisierungsverfahren basierend auf [8] um den Faktor 1:32 reduziert, bei nur 2%iger Reduktion der Merkmalsausbeute. Zudem beschleunigt sich der Vergleich von Merkmalsvektoren, da in der binarisierten Version die Hamming-Distanz als Abstandsmaß verwendet wird, die sich im Vergleich zum Euklidischen Abstand hocheffizient berechnen lässt. Zur Stabilisierung des 2D-Objekttrackings wurde zusätzlich ein Camshift-Tracker implementiert [TLKK10, 2]. Hierbei handelt es sich um einen Kerndichteschätzer, der auf HSV-Histogrammen arbeitet und sich bezüglich der SIFT-Merkmale komplementär verhält, da er sehr gut auf Objekte mit homogenen Farbverläufen anspricht. Die Robustheit dieses Trackers konnte auf der Elrob 2010 unter Beweis gestellt werden und wurde dort mit dem

Relighting mittels Environment Maps Im Zentrum der aktuellen Arbeiten steht ein Ansatz, ein Lichtfeld mit einer Beleuchtungsumgebung (Environment Map) neu zu beleuchten. Das Vorhaben gliedert sich in einen Akquisitionsschritt und einen Relighting-Schritt. Voraussetzung für das bildbasierte Relighting sind Informationen über die Normalen in einem Bild. Im Folgenden wird das geplante Vorgehen für das Relighting eines Einzelbildes skizziert:

Im Akquisitionsschritt werden für mehrere Beleuchtungsszenarien fotografische Aufnahmen eines Objekts erstellt. Die Reflektionseigenschaften des Objekts können hierbei über die Oberfläche variieren. Zusätzlich dazu wird eine Environment Map (EM) dieses Beleuchtungsszenarios erfasst. Ziel ist es, die EM geeignet zu filtern, um damit die akquirierten Radiance-Maps mit Hilfe der bekannten Normalen zu approximieren.

Im Relighting-Schritt ist es dann möglich, mit Hilfe der Normaleninformation und der Approximation der gemessenen Radiance jedes Pixels eine beliebige, neue Beleuchtungsumgebung in das Lichtfeld einzufügen.

Optimierung der Merkmalsberechnung während der Analyse Einer der rechenintensivsten Prozesse des Analysesystems sind die hochkomplexen Merkmale, die in Echtzeit berechnet werden müssen. Diese gewährleisten die hohe Toleranz gegenüber unterschiedlichen Beleuchtungseinflüssen. Da SIFT und SURF Merkmale schon kurz nach ihrer Veröffentlichung einen hohen Grad an Popularität erreicht haben, gibt es bereits jetzt diverse Optimierungen (wie die in Abschnitt 3.3.4 vorgestellten) sowie neue Wege zur Berechnung der Keypoints und Deskriptoren, die zusammen die Merkmale bilden. Insbesondere gibt es einen Trend zur Laufzeitoptimierung bei diesen neuen Varianten. Hier könnte als eine weitere Maßnahme evaluiert werden, ob und in wieweit sich die bestehenden Merkmalsberechnungen durch andere Verfahren, die möglicherweise ein günstigeres Laufzeitverhalten aufweisen, ersetzt werden können. In diesem Zusammenhang ist ebenfalls interessant zu klären, ob das Binärsierungsverfahren, das sowohl auf SIFT als auch auf SURF Merkmale anwendbar ist, sich ebenfalls auf weitere Verfahren übertragen lässt.

3.5 Interdisziplinäre Weiterentwicklung

Teile des Projektes (insbesondere die Methoden für das 2D Tracking) wurden bereits erfolgreich auf einen Außenbereichsroboter übertragen. Hier könnte interessant sein, zu klären, ob die 3D Poseschätzung das Potential bietet, als weiteres Model für die Hinderniserkennung eingesetzt zu werden. Robotern mit Greifarmen könnten zudem eine weitere Möglichkeit geboten werden, Objekte zielsicher anzusteuern und zu greifen.

Die Mensch-Maschine Interaktion ist eine weitere Anwendung. Schon jetzt zeigt Microsofts Kinect-Technologie hier eine kommerzielle Applikation als Steuerung von Videospielekonsolen rein über menschliche Gesten. Hier könnte die Präzision der Poseschätzung des Spielers erhöht und die Robustheit der Spielererkennung gesteigert werden. Da die Treiberentwicklung des Kinect-Systems mittlerweile offiziell freigegeben wurde und sich schon jetzt eine aktive Entwicklergemeinschaft rund um Kinect gebildet hat, besteht hier großes Potential das PMDLumi-Projekt um das Kinect-Sensorsystem zu erweitern und entsprechende Algorithmen zu portieren.

Zusätzlich zu den bisher genannten Optimierungs- und Erweiterungsvorschlägen, könn-

te ebenfalls ein Ausbau der Freiheitsgrade bei der 3D Modellierung bzw. des Renderings vollständig neue Anwendungen eröffnen. Bisher können einfach strukturierte und starre Objekte modelliert und für die Objekterkennung/-poseschätzung verwendet werden. Ein Ausbau des Renderers mit der Möglichkeit, verformbare Objekte sowohl zu modellieren als auch in Echtzeit zu rendern, eröffnet auch hier neue Möglichkeiten. Um an das letztgenannte Beispiel anzuknüpfen, könnte so die Mensch-Maschine Interaktion weiter ausgebaut werden und virtuelle Objekte auf gestische Manipulation des Spielers in natürlicher Art und Weise reagieren. Anstatt einzelne, starre Objekte in der virtuellen Welt nur zu drehen und zu verschieben, kann dann ein virtuelles Objekt durch Handbewegungen des Menschen interaktiv verformt werden. Ebenso können sich verformende Objekte wie z.B. Menschen, Tiere, Pflanzen, Wolken durch Analyse-durch-Synthese modelliert werden. Dies stellt eine neue Stufe synthetischer Mustererkennung, aber auch interaktiver virtueller Realität dar.

3.6 Verwertungspotenzial

Das erstellte Analyse-durch-Synthese-System bietet die Möglichkeit verschiedener Einsatzgebiete.

Teile des Systems (insbesondere des Objekttrackings) wurden bereits erfolgreich in der mobilen Außenbereichsrobotik angewandt [TLKK10]. Weitere Anwendungen lassen sich im Bereich der immer mehr in den wissenschaftlichen und öffentlichen Fokus rückenden Augmented Reality finden.

Es existieren bereits Lösungen für interaktive Museumsführer, die dem Besucher auf Sichtgeräten passende virtuelle Ansichten von Kunstobjekten generieren, vor denen sich der Besucher gerade befindet. Dies könnte in ähnlicher Form im Bereich der medizinischen nichtinvasiven Diagnostik weitergeführt werden, indem bspw. aus Tomographie-Bildern 3D Modelle berechnet werden, die dann mittels eines Handheld-Sichtgerätes auf die Stellen des realen Körpers des Patienten projiziert werden, auf welche der behandelnde Arzt sein Sichtgerät ausgerichtet hat. Viel mehr noch als bei einem interaktiven Museumsführer kommt es hierbei auf eine exakte Poseschätzung an, damit das 3D Modell passgenau mit den realen Organen des Patienten übereinstimmt.

Desweiteren besitzt das System hohes Potential, im Bereich der Unterhaltungselektronik eingesetzt zu werden. Projekte wie Microsofts Kinect streben beispielsweise an, Videospiele komplett ohne Controller auskommen zu lassen und die Spielsteuerung durch Gesten des Spielers zu ermöglichen. Systeme zur Objekterkennung, insbesondere für die Erkennung der Gesichter, basierend auf derartiger Tiefensensorik und einer RGB-Kamera zur Akquise von 3D-Raumdaten, könnten mittels des Analyse-durch-Synthese-Systems hochpräzise Abstand und Ausrichtung der Spieler zum Sensorsystem schätzen. Zudem können dem System bekannte Objekte ebenfalls eingebunden werden.

Im Bereich der Synthese wurden direkte wissenschaftliche Ergebnisse erzielt, die als Grundlagen für den Einsatz von Lichtfeldern im Multimedia- und Entertainmentbereich nützlich sind. Multi- oder Volume Light Fields und wiederbeleuchtbare Lichtfelder weisen ein hohes Nutzungspotenzial für z.B. E-Learning- oder künstlerische Anwendungen auf, ebenso für geometrieloze Produktvisualisierungen.

3.7 Beteiligte Wissenschaftler

3.7.1 Projektleiter

Prof. Dr. Andreas Kolb Wissenschaftliche Betreuung der Doktoranden Lisa Brückbauer, Jens Orthmann, Severin Todt am Lehrstuhl für Computergraphik und Projekt-Koordination. Verknüpfung verschiedener Teilprojekte und wissenschaftliche Verwertung in Form von Publikationen (z.B. [KBKL10]) und Workshops.

Prof. Dr. Klaus-Dieter Kuhnert Wissenschaftliche Betreuung des Doktoranden Matthias Langer am Lehrstuhl für Echtzeit-Lernsysteme und Projekt-Koordination.

Prof. Dr. Christof Rezk Salama Weitere wissenschaftliche Unterstützung der Doktoranden Lisa Brückbauer und Severin Todt. Herr Rezk hat die in dem Projekt entstandenen Algorithmen zum Raycasting konzeptionell und teilweise auch algorithmisch auf Volumen-Lichtfelder erweitert [RSTK08].

3.7.2 Wissenschaftliche Mitarbeiter

Dipl.-Inf. Lisa Brückbauer hat sich schwerpunktmäßig mit der Entwicklung und Implementierung von beleuchtungstoleranten Lichtfeldern befasst [BRSK10]. Frau Brückbauer wird den Lichtfeldteil des Projektes weiterführen.

Dipl.-Inf. Matthias Langer hat das echtzeitfähige Erkennungssystem inklusive aller Komponenten wie Feature-Erkennung und Tracking realisiert.

Dipl.-Inf. Jens Orthmann hat sich mit zwei grundlegenden Themen der Implementierung befasst. Einerseits wurde der im BMBF-Verbundprojekt „Lynkeus“¹ entwickelte Simulator [KOKP07] für die Akquisition von tiefengestützten Lichtfeldern erweitert. Zudem hat Herr Orthmann Grundlagen zur Integration von GPU-Funktionalitäten in das Szenegraph-basierte System `OpenSceneGraph` geschaffen, welche auch in anderen Projekten am Lehrstuhl für Computergraphik zum Einsatz kommen [OKK09].

Dr. Severin Todt hat in der zweiten Projektphase die Nutzung der sphärischen Lichtfelder für den Spielbereich weitergeführt [TRSK08]² und seine Promotion abgeschlossen [Tod09].

4 Liste der im Teilprojekt entstandenen Publikationen

4.1 Begutachtete Publikationen

4.1.1 Wissenschaftliche Zeitschriften

[RSTK08] C. Rezk-Salama, S. Todt, and A. Kolb. Raycasting of Light Field Galleries from Volumetric Data. *J. Computer Graphics Forum (EuroVis)*, 27(3):839–846, 2008.

[TRSK08] S. Todt, C. Rezk-Salama, A. Kolb, and K.-D. Kuhnert. GPU-based spherical light field rendering with per-fragment depth correction. *J. Computer Graphics Forum*, 27(8):2081–2095, 2008.

¹Förderkennzeichen: 16SV2296-310

²Die Arbeit wurde mit dem *Best Student Paper Presentation Award* ausgezeichnet.

4.1.2 Peer-Reviewed Fachkonferenzen

- [BRSK10] L. Brückbauer, C. Rezk-Salama, and A. Kolb. Relighting spherical light fields with polynomial texture mapping. In *Proc. Vision, Modeling and Visualization*, pages 73–80, 2010.
- [LK08] M. Langer and K.-D. Kuhnert. A new hierarchical approach in robust real-time image feature detection and matching. In *The 19th Int. Conf. on Pattern Recognition (ICPR)*, pages 1–4, 8–11 December 2008.
- [LKA⁺09] M. Langer, L. Kuhnert, M. Ax, D. N. Van, and K.-D. Kuhnert. 3d object recognition and localization employing an analysis by synthesis system. In *IADIS Int. Conf. on Applied Computing*, pages 132–140, Rome, Italy, 19–21 November 2009.
- [TRSK08] S. Todt, C. Rezk-Salama, and A. Kolb. Light field rendering for games. In *Proc. Theory and Practice of Computer Graphics*, pages 27–33, Best Student Paper Presentation Award, 2008.

4.2 Weitere Veröffentlichungen

- [AKL⁺10] M. Ax, L. Kuhnert, M. Langer, J. Schlemper, and K.-D. Kuhnert. Architecture of an autonomous mini unmanned aerial vehicle based on a commercial platform. In *Proc. of the ISR/ROBOTIK 2010 Conference*, Munich, June 2010.
- [KAL⁺09] L. Kuhnert, M. Ax, M. Langer, D. N. Van, and K.-D. Kuhnert. Absolute high-precision localisation of an unmanned ground vehicle by using real-time aerial video imagery for geo-referenced orthophoto registration. In *Fachgespräche Autonome Mobile Systeme (AMS)*, Karlsruhe, Germany, 2009.
- [KBKL10] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *J. Computer Graphics Forum*, 29(1):141–159, 2010.
- [KOKP07] M. Keller, J. Orthmann, A. Kolb, and V. Peters. A simulation-framework for time-of-flight sensors. In *Int. Sym. on Signals Circuits & Systems (ISSCS), session on Algorithms for 3D TOF-cameras*, pages 125–128. IEEE, 2007.
- [LK08] M. Langer and K.-D. Kuhnert. *New hierarchical approaches in real-time robust image feature detection and matching*, chapter 13, pages 538–554. Computer Vision. I-Tech Education and Publishing, Vienna, Austria, November 2008.
- [OKK09] J. Orthmann, M. Keller, and A. Kolb. Integrating GPGPU functionality into scene graphs. In *Proc. Vision, Modeling and Visualization*, pages 233–242, 2009.
- [TLKK10] S. Thamke, M. Langer, L. Kuhnert, and K.-D. Kuhnert. Robust outdoor tracking by fusion of laser scanner and image processing data. In *Int. Conf. on Computational Intelligence and Vehicular Systems*, Seoul, South Korea, 22–23 September 2010.
- [TLRS⁺08] S. Todt, M. Langer, C. Rezk-Salama, A. Kolb, and K.-D. Kuhnert. Spherical lightfield rendering in application for analysis by synthesis. *Int. J. on Intell. Systems and Techn. and Appl. (IJISTA), Issue on Dynamic 3D Imaging*, 5(1):304–314, 2008.
- [Tod09] Severin S. Todt. *Real-Time Rendering and Acquisition of Spherical Light Fields*. PhD thesis, Universität Siegen, 2009.
- [TRSBK08] S. Todt, C. Rezk-Salama, L. Brückbauer, and A. Kolb. Progressive light field rendering for web based data presentation. In *Workshop on Hyper-Media 3D Internet*, pages 23–32, 2008.

5 Publikationen Dritter

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, 2006.
- [2] G. R. Bradski. Computer video face tracking for use in a perceptual user interface. *Intel Technology Journal*, Q2:705–740, 1998.
- [3] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communication of the ACM*, 24(6):381–395, 1981.
- [4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.
- [5] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 20:91–110, 2003.
- [6] T. Malzbender, D. Gelb, and H. Wolters. Polynomial texture maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 519–528, New York, NY, USA, 2001. ACM.
- [7] C. Rezk Salama. GPU-Based Monte-Carlo Volume Raycasting. In *Proc. Pacific Graphics*, pages 411–414, 2007.
- [8] M. Stommel and O. Herzog. Binarising SIFT-descriptors to reduce the curse of dimensionality in histogram-based object recognition. In Dominik Slezak, Sankar K. Pal, Byeong-Ho Kang, Junzhong Gu, Hideo Kurada, and Tai-Hoon Kim, editors, *Signal Processing, Image Processing and Pattern Recognition*, pages 320–327. Springer, 2009.
- [9] Columbia University. Columbia university image library (coil-100). www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php. Abgerufen am 10.08.2010.