

2D/3D Datenverarbeitung und -fusion auf Basis der PMD-Technologie (2D3DProc)

Abschlussbericht



Prof. Dr.-Ing.
Andreas Kolb

Lehrstuhl für
Computergraphik und
Multimediasysteme

Universität Siegen
Hölderlinstr. 3
57068 Siegen



Prof. Dr.-Ing. habil.
Otmar Loffeld

Zentrum für
Sensorsysteme
(ZESS)

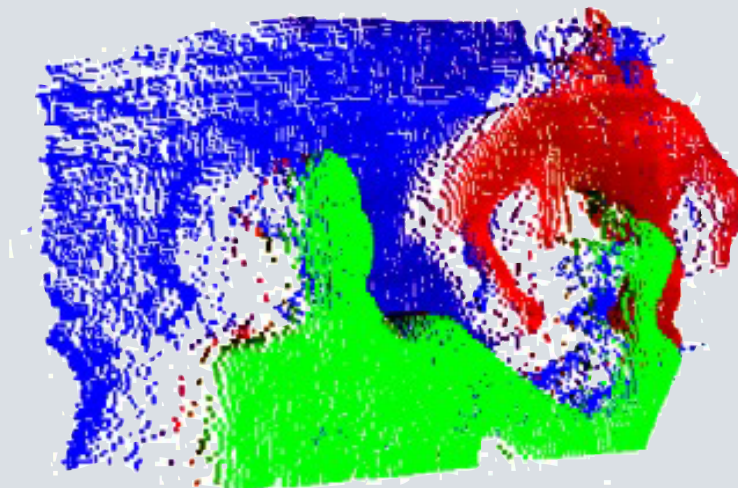
Universität Siegen
Paul-Bonatz-Str. 9-11
57068 Siegen

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	1
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	2
2	Zusammenfassung	2
2.1	Wesentliche Ergebnisse und erzielte Fortschritte	2
2.2	Ausblick	3
3	Arbeits- und Ergebnisbericht	4
3.1	Ausgangslage	4
3.2	Beschreibung der durchgeführten Arbeiten	4
3.3	Darstellung der erzielten Ergebnisse	6
3.4	Ausblick auf zukünftige Arbeiten	10
3.5	Interdisziplinäre Weiterentwicklung	10
3.6	Verwertungspotenzial	10
3.7	Beteiligte Wissenschaftler	10
4	Liste der im Teilprojekt entstandenen Publikationen	11
4.1	Begutachtete Publikationen	11
4.2	Weitere projektrelevante Veröffentlichungen	11
5	Publikationen Dritter	12

1 Allgemeine Angaben

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

1.1 DFG-Geschäftszeichen

KO 2960/5-2

1.2 Antragsteller

Andreas Kolb, Prof. Dr.-Ing.
Otmar Loffeld, Prof. Dr.-Ing. habil.

1.3 Institut/Lehrstuhl

Andreas Kolb, Prof. Dr.-Ing.
Zentrum für Sensorsysteme (ZESS) und
Lehrstuhl für Computergraphik und Multimediasysteme
Universität Siegen
Hölderlinstr.3
57068 Siegen

Tel.: 0 271 / 7 40 - 24 04
Email: andreas.kolb@uni.siegen.de

Otmar Loffeld, Prof. Dr.-Ing. habil.
Zentrum für Sensorsysteme (ZESS) und
Institut für Nachrichtenverarbeitung
Universität Siegen
Paul-Bonatz-Str. 9-11
57068 Siegen

Tel.: 0 271 / 7 40 - 31 25
Email: loffeld@zess.uni.siegen.de

1.4 DFG-finanzierte wissenschaftliche Mitarbeiter

Marvin Lindner 02. 2008 - 09. 2009
Iurie Chiosa 12. 2009
Christian Köhler 01., 07. und 09. - 12. 2010 (50%)

1.5 Thema

2D3DProc: 2D/3D Datenverarbeitung und -fusion auf Basis der PMD-Technologie

1.6 Berichts- und Förderzeitraum

Förderzeitraum 1.2.2008 – 31.12.2010

1.7 Fachgebiet, Arbeitsrichtung

Informatik, Informationsverarbeitung, Signalverarbeitung, Bildverarbeitung, Datenstrukturen, Sensordatenverarbeitung, Sensorfusion

1.8 Verwertungsfelder

Anwendungen in Industrie- und Entertainmentbereich mit Einsatz von PMD-Tiefenkameras u.A. zur akkuraten Objekterkennung und Umgebungs-Akquisition.

1.9 Am Projekt beteiligte Kooperationspartner

Reinhard Koch, Prof. Dr.-Ing.

Arbeitsgruppe Multimediale Systeme zur Informationsverarbeitung (MIP),
Institut für Informatik

Christian-Albrechts-Universität zu Kiel

Mit der Arbeitsgruppe von Prof. Koch wurden vielfältige Kooperationen im Bereich der Kalibrierung durchgeführt.

2 Zusammenfassung

2.1 Wesentliche Ergebnisse und erzielte Fortschritte

In der zweiten Förderphase wurden wesentliche Ergebnisse im Bereich der Fehlerkorrektur und der Kalibrierung, sowie in der Tiefenverfeinerung mit impliziter Ausreißerkorrektur erzielt.

Im Bereich der Fehlerkorrektur und Kalibrierung wurden drei wesentliche Fortschritte erzielt, welche eine deutliche Verbesserung der PMD-Tiefendaten für nachgelagerte Verarbeitungsprozesse liefern (siehe Abschnitt 3.2.1). Einerseits wurde ein alternatives Demodulationsschema basierend auf der Annahme eines Rechteck-Signals entwickelt, welches ohne zusätzlichen Samplingaufwand für die Korrelationsfunktion zwischen dem Referenzsignal der PMD-Kamera und dem reflektierten optischen Signal auskommt. Als Ergebnis konnte ein kombiniertes Demodulationsschema entwickelt werden, welches mit einer minimalen Anzahl von zwei Referenzmessungen ein verbessertes Demodulationsergebnis liefert. Ein zweites,

wichtiges Ergebnis ist die Entwicklung eines effizienten Kalibrierschemas für die kombinierte Korrektur des intensitätsabhängigen und des systematischen Fehlers durch Entkopplung beider Effekte. Dadurch konnte die Zahl der notwendigen Referenzmessungen drastisch reduziert werden. Als drittes zentrales Ergebnis ist die Behandlung von Bewegungsartefakten zu nennen, welche sich durch die sequentielle Aufnahme der vier Phasenbilder in der PMD-Kamera ergeben. Hier wurde ein Verfahren entwickelt, welches sowohl laterale, als auch axiale Bewegungen kompensiert. Zur Korrektur der lateralen Bewegung werden Objektpunkte zwischen den Phasenbildern durch optischen Fluss getrackt, was eine zusätzliche Normierung der Phasenbilder erfordert. Die axiale Bewegung wird durch ein Bewegungsmodell auf Basis der gemessenen Entfernungsdaten modelliert.

Im Bereich der Tiefenverfeinerung wurde ein echtzeitnahes, GPU-basiertes Verfahren zur lokalen Oberflächenrekonstruktion entwickelt, welches auf einem Ansatz zum Raycasting von Moving-Least-Squares (MLS) Approximationen basiert. Dieser Ansatz erlaubt eine implizite Ausreißer-Unterdrückung (siehe Abschnitt 3.2.2).

Zur Segmentierung von Tiefenkarten wurde im Gegensatz zu den meisten anderen Verfahren kein k-mean-Clustering, sondern ein hierarchisches Verfahren eingesetzt. Erste Ergebnisse wurden in einem Hand-Tracking-Verfahren eingesetzt. Ein neu entwickeltes Hybrid-Verfahren erzeugt qualitativ sehr hochwertige Ergebnisse, ist allerdings für eine Echtzeitverarbeitung von PMD-Tiefenkarten nicht schnell genug (siehe Abschnitt 3.2.3).

Die wissenschaftlichen Ergebnisse dieses Projektes im Bereich Fehlerkorrektur, Kalibrierung und Tiefenverfeinerung sind in die Dissertation von Marvin Lindner [Lin10] eingeflossen. Die Ergebnisse der Segmentierung sind Teil der Dissertation von Iurie Chiosa [Chi10], deren Schwerpunkt allerdings im Bereich des Mesh-Clusterings liegt.

2.2 Ausblick

Die Ergebnisse dieses Teilprojektes werden aktuell in weiterführenden Projekten als Grundlage genutzt (siehe Abschnitt 3.6). Die wesentliche Stoßrichtung ist hierbei die Online-Fusionierung und Modellrekonstruktion der durch die Arbeiten in diesem Teilprojekt verbesserten, unregistriert erfassten PMD-Tiefendaten in einem Multikamera-Aufbau. Diese Fragestellung wird im Rahmen eines BMWi-Projektes im Kontext von Waschanlagen für Fahrzeuge praxisnah erforscht.

Weitere interessante Fragestellungen ergeben sich etwa in der Nutzung von Multikamera Aufbauten für Anwendungen in der Szenenüberwachung, wie sie im DFG Graduiertenkolleg 1564 untersucht werden.

Im Bereich der lokalen Oberflächenrekonstruktion mittels MLS sind weitere Verbesserungen zur Reduktion des Contour-Bleedings notwendig.

3 Arbeits- und Ergebnisbericht

3.1 Ausgangslage

Ausgehend von der sensornahen Datenverarbeitung von 3D-Information und deren Fusion mit 2D-Bildern, lagen die geplanten Ziele des Projektantrags in der Erforschung grundlegender Fragestellungen, welche sich beim Einsatz einer oder mehrerer, mono- oder binokularer 2D/3D-Kamerakombinationen ergeben. Die wesentlichen Ziele waren:

1. die Verfeinerung von Tiefendaten,
2. die Klassifikation und Segmentierung von 2D/3D-Daten sowie
3. die Koregistrierung dynamischer Kamerapositionen.

Maßgeblich für die Ziele des Projekts waren ferner die grundsätzlichen Fragestellungen bzgl.

- der Behandlung variabler Reflexionseigenschaften von Objektoberflächen sowie
- der Problematik an Objektkanten, insbesondere der Verbesserung der Messgenauigkeit bei inhomogenen Abständen innerhalb eines Pixels und bei bewegten Objekten (aufgrund der Mehrphasenmessung)

3.2 Beschreibung der durchgeführten Arbeiten

Gemäß der Hinweise der Gutachter und der aktuellen Entwicklungen zu Beginn des Projektes, wurde der Schwerpunkt des Förderzeitraums auf die sensornahen Datenverarbeitung und Fusion von 2D- und 3D-Informationen gelegt.

Die im Projektantrag zuvor beschriebene Vorgehensweise zur bifokalen Fusion zeigte sich stark abhängig von der Genauigkeit der Tiefeninformationen, so dass eine Verbesserung und Verfeinerung von Tiefeninformationen angestrebt wurde. Als besonders einflussreich haben sich hierbei, neben dem systematischen Demodulationsfehler, ebenfalls Messabweichungen auf Grund von Reflektionsänderungen der Objektmaterialien sowie Bewegungsartefakte innerhalb dynamischer Szenen gezeigt, weshalb diese tiefergehend untersucht wurden.

3.2.1 Erhöhung der Messgenauigkeit

Demodulations-Fehler Auch auf Grund der Vorschläge der Gutachter wurden zunächst weitere Untersuchungen zur tieferen Modellierung des systematischen Fehlers durchgeführt. Ausgangspunkt waren Veröffentlichungsergebnisse, die besagen, dass das Korrelationssignal einer Mischung aus Sinus- und Rechteckfunktion entspricht [6]. Basierend auf dieser Beobachtung wurde eine alternative Demodulationsmethode für Rechtecksignale umgesetzt und analysiert [LKR08].

Intensitätsabhängige Fehler Das zuvor im Projektantrag beschriebene Verfahren zur Reduktion intensitätsabhängiger Fehler liefert zwar gute Ergebnisse, setzt jedoch einen zeitlich aufwendigen Kalibrierungsprozess voraus. Aus diesem Grund wurden im bewilligten Förderungszeitraum mögliche Ansätze zur Entkoppelung beider Modellparameter untersucht, um die Komplexität zu reduzieren [Lin10, LSKK10].

Bewegungsartefakte Neben den systematischen Fehlerquellen haben Bewegungsartefakte ebenfalls signifikanten Einfluss auf die Tiefeninformationen dynamischer Szenen. Als Ursache lassen sich folgende Fehlerquellen identifizieren:

- *Laterale Bewegungen*, welche zur Mischungen von Phasensamplern verschiedener Oberflächen führen,
- *Axiale Bewegungen*, welche in der Mischung von Phasensamplern verschiedener Tiefen resultieren und
- *Texturänderungen*, welche auf Grund unterschiedlicher Reflektivitätseigenschaften eines Objekts konstanter Entfernung auftreten.

Zur Reduktion von Bewegungsartefakten wurde ein Kompensationsmodell umgesetzt, welches das starre Samplingschema aktueller PMD-Kameras aufbricht und gezielt zusammengehörige Phasensamples eines Oberflächenpunkts zur Demodulation identifiziert [LK09]. Darüber hinaus wurde ein Modell für axiale Bewegungen entwickelt, welches die theoretischen Einflüsse auf die Messgenauigkeit beschreibt und zur Korrektur genutzt werden kann.

3.2.2 Verfeinerung

Zur Verbesserung des Fusionsergebnisses wurde neben der Reduktion von systematischen Messfehlern zusätzlich die Verfeinerung von Tiefeninformationen einzelner PMD-Bilder betrachtet. Ausgangspunkt war die Anwendung einer expliziten Oberflächenrekonstruktion auf Basis eines lokalen Polynom-Fittings [Lin10]. Diese bietet im Gegensatz zum gradientenbasierten Verfahren der ersten Projektphase den Vorteil, dass sie eine Fusion mehrerer Tiefenbilder unter Ausnutzung redundanter Informationen sowie eine implizite Ausreißerkennung ermöglicht.

Die im Antrag für die zweite Projektphase beschriebene monokulare Verfeinerung durch Modellierung eines Markov-Prozesses wurde zu Beginn des Förderungszeitraums bereits durch andere Arbeitsgruppen untersucht [5], so dass eine zusätzliche Vertiefung dieses Themas als wissenschaftlich wenig erfolgversprechend eingestuft wurde.

Im Kontext der binokularen Verfeinerung von Tiefendaten mittels *Mutual Information* wurde angestrebt, den sehr vielversprechenden Ansatz auf Grundlage von *shading constraints* [1] durch Reformulierung zu parallelisieren, um durch eine GPU-Implementierung Echtzeitfähigkeit zu erreichen. Grundlage des Verfahrens war die Beobachtung, dass Tiefen- und Intensitätsinformationen durch das Lambertsche Gesetz, ähnlich zu einem *shape-from-shading*-Ansatz, voneinander abhängen. Das in [1] beschriebene Verfahren löst hierzu ein sehr aufwändiges, globales Optimierungsproblem und liefert sehr gute Ergebnisse. Es wurden verschiedene Parallelisierungsansätze untersucht und testweise implementiert, aber es ist

leider nicht gelungen, das globale Optimierungsproblem für eine effiziente GPU-basierte Lösung so umzuformulieren, dass die Qualität des ursprünglichen Ansatzes in etwa erreicht wurde.

3.2.3 Segmentierung von Tiefendaten

Nach Absprache mit den Projektpartnern wurde entschieden, dass das Thema der Segmentierung nicht auf dieses Teilprojekt allein beschränkt bleiben soll, sondern auch in den Teilprojekten **MultiCam** [2] und bei dem Projektpartner Koch im Teilprojekt **3DPoseMap** [7] verankert werden muss, da das Thema sehr stark an applikationsspezifischen Fragestellungen ausgerichtet ist.

In diesem Teilprojekt **2D3DProc** wurden neue Segmentierungsansätze für Tiefendaten untersucht. Im Unterschied zu dem im Teilprojekt **MultiCam** genutzten k-means Clustering wurden in diesem Teilprojekt hierarchische Verfahren betrachtet [CKL11] und neue, hybride Methoden entwickelt [CKCL09].

3.2.4 Koregistrierung und Fusion

Zu Beginn des Förderungszeitraums wurde festgestellt, dass bereits erste Verfahren zur *unregistrierten Fusion* von PMD-Bildern unter Ausnutzung von Standardansätzen wie *Iterative Closest Point (ICP)* untersucht wurden [3, 4]. In den jeweiligen Ergebnissen hat sich bereits gezeigt, dass sich die beschriebenen Ansätze auf Grund der Datenqualität und der daraus resultierenden Komplexität kaum zur Echtzeit-Fusion eignen.

Aus diesem Grunde wurde von einer weiterführenden Behandlung der unregistrierten Fusion zunächst abgesehen und der Fokus auf eine weitere Verbesserung der Eingangsdaten gelegt. Aktuell wird mit den Ergebnissen in einem BMWi-Projekt die Fusion unregistrierter PMD-Kameradaten fortgesetzt (siehe Abschnitt 3.6).

Die Fusion *registrierter PMD-Bilder* gehörte im Rahmen der Generierung einer Umgebungskarte primär zum Forschungsschwerpunkt des Teilprojektes **3DPoseMap**, weshalb eine parallele Bearbeitung des Themas in diesem Teilprojekt als nicht sinnvoll eingestuft wurde.

3.3 Darstellung der erzielten Ergebnisse

Im Hinblick auf eine akkurate Datenfusion wurden im Verlauf des Bewilligungszeitraums folgende Ergebnisse zur Verbesserung der Tiefeninformationen erzielt:

3.3.1 Erhöhung der Messgenauigkeit

Demodulations-Fehler Die Untersuchungen eines alternativen Demodulationsverfahrens auf Basis einer Rechteck-Modellierung haben ähnliche, jedoch gegensätzliche Charakteristika der Messgenauigkeit gezeigt (s. Abb. 1). Die gegensätzlichen Messabweichungen wurden ausgenutzt, um mittels einer einfachen Linearkombination eine *light-weight Kalibrierung* zu ermöglichen, welche ohne Look-Up-Tabellen (LUT) oder funktionale Approximationen (z.B. B-Splines) zur Korrektur der Messabweichungen auskommt. Im Gegensatz zu gängigen,

phänomenologischen Modellen benötigt diese lediglich zwei Referenzbilder zur Bestimmung der optimalen Linearkombination, wodurch der Aufwand der Akquisition von Referenzdaten erheblich reduziert werden kann. Vergleiche mit LUT- und B-Spline-basierten Modellen zeigen jedoch, dass die neue Light-Weight Kalibrierung weniger genaue Korrekturergebnisse erzielt. Die Vorteile der Light-Weight Variante liegen daher primär in einer schnellen Kalibrierung und dem Verzicht auf teures Equipment wie Verfahrsschienen.

Intensitätsabhängiger Fehler Durch eine weiterführende Untersuchung des intensitätsabhängigen Fehlers wurde ein Ansatz zur Entkopplung der beiden Fehlerquellen, des systematischen und des intensitätsabhängigen Fehlers, gefunden und in ein Korrekturverfahren integriert.

Das entkoppelte Korrekturverfahren basiert auf der Beobachtung, dass wiggling-korrigierte Messwerte bzgl. verschiedener Objektentfernungen gleiche Charakteristika in dem durch die Intensitätsabhängigkeit verursachten Restfehler aufweisen. Diese sind jedoch vertikal zueinander verschoben (vgl. Abb. 2). Der gegebene Sachverhalt lässt sich dadurch erklären, dass die Referenzdaten für eine Demodulationskorrektur auf Grund des distanzabhängigen Lichtabfalles bereits Intensitätsänderungen berücksichtigen. Die daraus resultierenden, intensitätsabhängigen Fehler werden somit vorab bereits teilweise durch die Demodulationskorrektur ausgeglichen.

Ist der vertikale Offset bekannt, lassen sich die intensitätsabhängigen Fehler auf Grund der gleichen Charakteristik durch eine eindimensionale, lediglich von der Intensität abhängigen Funktion korrigieren. Vorteilhafterweise lässt sich jener Offset ebenfalls auf der Korrekturfunktion bestimmen, da dieser für eine gegebene Distanz dem Wert der Korrekturfunktion an der Stelle der für die Wiggling-Korrektur verwendeten Intensität h_{wig} entspricht (vgl. Abb. 2). Die Fehlerkorrektur ist somit durch

$$m_c(p) = m_w + \delta(h_{wig}(m_w)) - \delta(h) \quad (1)$$

gegeben, wobei m_w der wiggling-korrigierte Messwert, $\delta(h_{wig}(m_w))$ der Offset-bedingten Vorkorrektur bzw. und $\delta(h_p)$ der eigentlichen intensitätsabhängigen Korrektur ist.

Durch die Berücksichtigung der vertikalen Verschiebungen lassen sich Demodulations- und intensitätsabhängige Fehler letztlich voneinander unabhängig behandeln, wodurch die Anzahl notwendiger Referenzmessungen signifikant reduziert wird. Zur Bestimmung der Korrekturfunktion genügt es, ein Set von Referenz-Reflektivitäten für eine feste Entfernung zu bestimmen. Diese sollte im Hinblick auf das Intensitätsintervall möglichst nah zur Kamera gewählt werden. Die Kalibrierungsdaten zur Bestimmung von $h_{wig}(d)$ entsprechen den Referenzbildern der Demodulationskorrektur, so dass keine neuen Daten aufgenommen werden müssen.

Ein Vergleich zeigt, dass das entkoppelte Modell die gleiche Genauigkeit bei weniger Referenzdaten erzielt, wie das zuvor beschriebene, gekoppelte Modell (vgl. Tab. 1).

Bewegungsartefakte Neben den systematischen Fehlerquellen spielen Bewegungsartefakte eine erhebliche Rolle im Bezug auf die Gültigkeit von Tiefeninformationen und somit auf die Qualität von Fusionsergebnissen dynamischer Szenen. Zur Reduktion von Bewegungsartefakten wurde daher ein Modell entwickelt, welches eine Abschätzung des Optischen

Flusses zur Oberflächenverfolgung nutzt und zusammengehörige Phasensamples identifiziert.

Auf Grund der verwendeten Variationsmethode ergeben sich Annahmen an die Vergleichbarkeit von Intensitätswerten zwischen den einzelnen Phasenbilder, welche zunächst nicht erfüllt werden (s. Abb. 3 oben). Berücksichtigt man jedoch die Aufteilung der durch das einfallende Licht erzeugten Elektronen auf die beiden Auslesedioden eines PMD-Pixels, lässt sich die Gesamtlichtmenge durch eine Addition beider Auslesedioden bestimmen. Diese liefert die gewünschten Intensitätswerte pro Phasenbild (s. Abb. 3 unten).

Neben samplingbedingten Intensitätsunterschieden kommt es zusätzlich zu radialen Licht- bzw. Intensitätsabschwächungen zum Randbereich. Zur Reduktion der Abschwächungseinflüsse wurde eine Intensitätsnormierung umgesetzt, welche einzelne Pixel-Charakteristiken ausgleicht und für eine homogene Ausgangslage zur Flussschätzung sorgt (s. Abb. 4).

Durch das Abbilden der einzelnen Phasenbilder aufeinander und das Aufbrechen des starren Demodulationsschemas, lassen sich die in Abb. 5 exemplarisch dargestellten Ergebnisse für laterale Bewegungen und Texturänderungen in Echtzeit erzielen.

Zusätzlich wurden Untersuchungen durchgeführt, die Erkenntnisse über den theoretischen Einfluss axialer Bewegungen entlang eines Sehstrahls liefern. Diese besagen, dass der theoretische Tiefenmesswert bei einer uniformen Geschwindigkeit um die Mitte der zurückgelegten Distanz pendelt. Ist die Geschwindigkeit eines Objekts bekannt, lässt sich die theoretisch aufkommende Abweichung durch die Umkehrfunktion des Modells korrigieren. Auf Grund der Abweichungscharakteristik und der zugrunde gelegten Geschwindigkeiten lässt sich jedoch behaupten, dass axiale Bewegungsartefakte im Vergleich zu lateral und textuell induzierten Artefakten in den gängigsten Szenarien vernachlässigt werden können.

3.3.2 Verfeinerung

Zur kantenerhaltenden Tiefenverfeinerung wurde die Möglichkeit der lokalen, polynomiellen Approximation von Objektoberflächen auf Basis von *Moving Least Square Surfaces* untersucht. Diese bietet, neben einer impliziten Behandlung von Ausreißern und Diskontinuitäten, ebenfalls eine höherstetige Interpolation.

Zur Approximation wurde ein iteratives Raycastingverfahren für MLS-Oberflächen umgesetzt [8]. Dieses erzielte allgemein gute Ergebnisse bei rund 7 fps. Wie sich jedoch gezeigt hat, kommt es bei Tiefenbildern auf Grund von Tiefendiskontinuitäten und der gewählten polynomiellen Repräsentation vereinzelt zu Contour-Bleeding. Generell hängt die Qualität der Rekonstruktion / Verfeinerung von der Wahl des verwendeten Verfahrens ab. So gehen die meisten Verfahren von stetigen Oberflächen aus, so dass es an Oberflächenkanten meist zur Feature-Glättung kommt. Robustere Verfahren sind auf Grund ihrer Komplexität kaum echtzeitfähig.

3.3.3 Segmentierung von Tiefendaten

In dem Teilprojekt **2D3DProc** wurden einerseits hierarchische Verfahren zur Segmentierung von Tiefendaten verwendet, andererseits wurde eine neue hybride Segmentierungstechnik entwickelt.

Der grundsätzliche Ansatz eines hierarchischen Clusterings geht von der initialen Situation aus, dass jedes Pixel ein Cluster darstellt und verschmilzt benachbarte Cluster in der Reihenfolge ihrer Ähnlichkeit bzgl. eines definierten Ähnlichkeitsmaßes, sofern die Abweichung nicht oberhalb eines gegebenen Schwellwertes liegt.

Dieser hierarchische Ansatz wurde für eine GPU-Implementierung parallelisiert, indem Pixel unabhängig voneinander betrachtet werden. Zur Vermeidung von Inkonsistenzen prüft jedes Cluster den besten Merge-Partner in seiner Nachbarschaft und es werden nur Cluster verschmolzen, sofern sie sich gegenseitig als besten Merge-Partner identifiziert haben. Deadlock-Situationen werden durch eine lokale Priorisierung über Cluster-IDs vermieden. Hierzu wurde ein neues Homogenitätskriterium eingeführt, welches den Zusammenhang zwischen Abfall der Intensität aufgrund der aktiven Beleuchtung und der Entfernung berücksichtigt.

Das hierarchische Clustering hat den Vorteil, dass es schnell ausgeführt werden kann und ein datengetriebenes Abbruchkriterium hat, wodurch sich die Anzahl der Cluster der Datenkomplexität anpasst. Der Nachteil des Verfahrens liegt in der Tatsache, dass ungeschickte Verschmelzungen in einer frühen Phase nicht korrigiert werden können. Ein erster Einsatz des Verfahrens zum Tracking parallelen von zwei Händen zeigt ein sehr robustes Verhalten, auch bei Extremsituationen wie verdeckten Händen oder zusätzliche Hände in der Szene (siehe Abb. 7 und [CKL11]). Das Hand-Tracking-System benötigt ca. 60 ms pro Bild für ein Tiefenbild mit einer Auflösung von 204^2 Pixeln.

Entsprechend wurde eine neue, hybride Clusteringtechnik, das *Multi-Level-Clustering*, entwickelt, welche das hierarchische Verfahren mit einem k-means-artigen Optimierungsverfahren kombiniert. Hierbei werden Merging- und Clusteroptimierung abwechselnd durchgeführt. Durch Ausnutzung von räumlichen und (bzgl. der Optimierung) zeitlichen Kohärenzen wird ein vergleichsweise sehr effizientes und hochgenaues Clustering möglich. Hierbei wird in der Optimierung ausgenutzt, dass sich Cluster durch die Optimierung nicht beliebig verschieben (räumliche Kohärenz). Somit werden nicht alle Tiefenpunkte mit allen Clustern bzgl. des besten Fittings getestet, sondern nur Pixel am Clusterrand mit dem jeweils benachbarten Cluster. Dieses Verfahren lässt sich wiederum für eine GPU-Implementierung parallelisieren. Es hat sich jedoch gezeigt, dass dieses hochgenaue hybride Clustering für eine Echtzeitanwendung zu langsam ist. Entsprechend wurde der Ansatz in das Anwendungsfeld des Mesh-Clusterings gebracht [CKCL09] und auch sehr erfolgreich auf allgemeines Datenclustering erweitert [CK11].

3.3.4 Verarbeitungs-Framework

Im Rahmen des geförderten Teilprojekts wurde ein generisches Framework zur effizienten Datenverarbeitung entwickelt und für die Implementierung der o.g. Verfahren genutzt. Dieses basiert auf dem Pipeline-Konzept und erlaubt neben einer effizienten Berechnung und Wiederverwendung von Zwischenergebnissen, ebenfalls eine flexible Kombination von Algorithmen zur Laufzeit. Darüber hinaus umfasst das Framework eine automatische Speicherverwaltung in Form von Objekt-Managern und Smart-Pointern, sowie einen automatischen Datenabgleich zwischen Arbeits- und Grafikkartenspeicher. Letzterer erlaubt einen einfachen Austausch von Berechnungsergebnissen zwischen CPU und GPU, wodurch eine effiziente Implementierung gemischter Systeme ermöglicht wird. Durch seine Modularität und seinen generischen Unterbau ist das Framework problemlos um neue Filteralgorithmen erweiterbar.

3.4 Ausblick auf zukünftige Arbeiten

Grundsätzlich ergibt sich auf Grundlage der Ergebnisse dieses Projektes die Fragestellung der effizienten Online-Fusionierung und Übertragung auf PMD-Tiefeninformationen. Hierbei gilt es, Repräsentationsformen zu finden, welche eine effiziente Online-Integration und Fusionierung von PMD-Tiefendaten für Multi-Kameraaufbauten unter Ausnutzung vorhandener Redundanzen ermöglicht.

Diese Zielrichtung wird aktuell in Teilen im Graduierten-Kolleg 1564 „Imaging New Modalities“ und in einem BMWi-ZIM-Projekt aufgegriffen und behandelt (siehe Abschnitt 3.6).

3.5 Interdisziplinäre Weiterentwicklung

Die in diesem Teilprojekt erworbenen Kenntnisse haben unmittelbare Bedeutung für alle Anwendungsfelder von PMD-basierten Tiefendaten. Beispielsweise sei die Industrie-Robotik und die Automatisierungstechnik zu nennen. In diesem Anwendungskontext wurde auch in dem zeitlich parallel laufenden BMBF Verbundprojekt „Lynkeus“ gearbeitet [FHK⁺10]. In diesem Zusammenhang haben sich wichtige Synergien, z.B. bei der Entwicklung von PMD-Kameramodellen für die Online-Simulation von PMD-Sensoren ergeben [KK09].

3.6 Verwertungspotenzial

Grundlegende Forschungsergebnisse zur Verbesserung der Messgenauigkeit, der Bewegungskompensation und Verfeinerung wurden, bzw. werden, aktuell in den nachstehend aufgeführten Folgeprojekten genutzt:

- Lynkeus BMBF-Verbundprojekt „Lynkeus“¹ (2006-2009)
- DFG Graduierten-Kolleg 1564 „Imaging New Modalities“ (1. Projektphase 2009-2013)
- BMWi ZIM-Kooperationsprojekt „Entwicklung einer berührungslosen Erfassung von Fahrzeugkontouren zur Steuerung von Kfz-Waschanlagen auf Basis von PMD-Sensoren“²

3.7 Beteiligte Wissenschaftler

3.7.1 Projektleiter

Prof. Dr. Andreas Kolb Wissenschaftliche Betreuung der Doktoranden Marvin Lindner und Iurie Chiosa am Lehrstuhl für Computergraphik sowie Projekt-Koordination. Verknüpfung verschiedener Teilprojekte und wissenschaftliche Verwertung in Form von Workshops.

Prof. Dr. Otmar Loffeld Wissenschaftliche Zweitbetreuung der Doktoranden in diesem Projekt.

¹Förderkennzeichen 16SV2296-310

²Förderkennzeichen KF2383701RR9

3.7.2 Wissenschaftliche Mitarbeiter

Marvin Lindner Marvin Lindner hat sich schwerpunktmäßig mit der Verbesserung der Messgenauigkeit [LKR08, LK09, LSKK10] und der Verfeinerung von Tiefeninformationen befasst [Lin10].

Iurie Chiosa Iurie Chiosa hat sich im Rahmen des Projekts mit der Entwicklung von Methoden zum GPU-basierten Mesh-Clustering und deren Anwendung für PMD-Tiefeninformationen befasst.

Das in diesem Zusammenhang entwickelte hybride Clusterverfahren wurde im Weiteren sehr erfolgreich für das Mesh-Clustering [CKCL09] und das Datenclustering [CK11] eingesetzt.

Christian Köhler Christian Köhler hat maßgeblich bei der Implementierung und Weiterentwicklung des in der ersten Projektphase entworfenen Frameworks zur Echtzeit-Verarbeitung von Tiefenbildsequenzen mitgewirkt.

4 Liste der im Teilprojekt entstandenen Publikationen

4.1 Begutachtete Publikationen

4.1.1 Wissenschaftliche Zeitschriften

[LSKK10] M. Lindner, I. Schiller, A. Kolb, and R. Koch. Time-of-Flight Sensor Calibration for Accurate Range Sensing. *Computer Vision and Image Understanding*, 114(12):1318 – 1328, 2010.

4.1.2 Peer-Reviewed Fachkonferenzen

[CKCL09] I. Chiosa, A. Kolb, N. Cuntz, and M. Lindner. Parallel Mesh Clustering. In *Proc. Eurographics Symp. on Parallel Graphics and Visualization (EGPGV)*, pages 33–40, 2009.

[LK09] M. Lindner and A. Kolb. Compensation of Motion Artifacts for Time-of-Flight Cameras. In *Proc. Dynamic 3D Imaging*, LNCS, pages 16–27. Springer, 2009.

[LKR08] M. Lindner, A. Kolb, and T. Ringbeck. New Insights into the Calibration of TOF Sensors. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Workshop on ToF Camera based Computer Vision (TOF-CV)*, pages 1–5, 2008. DOI 10.1109/CVPRW.2008.4563172.

4.2 Weitere projektrelevante Veröffentlichungen

[Chi10] Iurie Chiosa. *Efficient and High Quality Clustering*. PhD thesis, University of Siegen, 2010.

- [CK11] I. Chiosa and A. Kolb. GPU-based multilevel clustering. *IEEE Trans. on Visualization and Computer Graphics*, 17(2):132–145, 2011.
- [CKL11] R. Cespi, A. Kolb, and M. Lindner. Hand tracking based on hierarchical clustering of range data. In *IEEE Proc. Int. Conf. on Computer Vision; Workshop on Consumer Depth Cameras for Computer Vision*, 2011. submitted.
- [FHK⁺10] S. Fuchs, S. Haddadin, M. Keller, S. Parusel, A. Kolb, and M. Suppa. Cooperative bin-picking with time-of-flight camera and impedance controlled dlr lightweight robot iii. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 4862–4867, 2010.
- [KBKL10] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *Computer Graphics Forum*, 29(1):141–159, 2010.
- [KK09] M. Keller and A. Kolb. Real-time simulation of time-of-flight sensors. *J. Simulation Practice and Theory*, 17:967–978, 2009.
- [Lin10] Marvin Lindner. *Calibration and Real-Time Processing of Time-of-Flight Range Data*. PhD thesis, University of Siegen, 2010.

5 Publikationen Dritter

- [1] M. Böhme, M. Haker, T. Martinetz, and E. Barth. Shading constraint improves accuracy of time-of-flight measurements. *Computer Vision and Image Understanding*, 114:1329–1335, 2010.
- [2] S. Ghobadi, O. Loepprich, O. Lottner, K. Hartmann, O. Loffeld, and W. Weihs. Improved object segmentation based on 2d/3d images. In O. Sablatnig, R.;Scherzer, editor, *The Fifth IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA 2008)*, pages 42–47. ACTA Press, 2008.
- [3] B. Huhle, P. Jenke, and W. Straßer. On-the-fly scene acquisition with a handy multisensor-system. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 5(3/4):255–263, 2008.
- [4] Y.M. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and calibration of a multi-view ToF sensor fusionsystem. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision*, 2008.
- [5] A. Rajagopalan, Arnav Bhavsar, Frank Wallhoff, and Gerhard Rigoll. Resolution enhancement of pmd range maps. In Gerhard Rigoll, editor, *Pattern Recognition*, volume 5096 of *Lecture Notes in Computer Science*, pages 304–313. Springer Berlin / Heidelberg, 2008.
- [6] H. Rapp. Experimental and theoretical investigation of correlating TOF-camera systems. Master’s thesis, University of Heidelberg, Germany, 2007.
- [7] I. Schiller, B. Bartczak, F. Kellner, and R. Koch. Increasing realism and supporting content planning for dynamic scenes in a mixed reality system incorporating a time-of-flight camera. *Journal of Virtual Reality and Broadcasting*, 7(4), 2010.

-
- [8] E. Tejada, J.P. Gois, L. G. Nonato, A. Castelo, and T. Ertl. Hardware-accelerated extraction and rendering of point set surfaces. In *Proc. of Eurographics/IEEE VGTC Symposium on Visualization*, pages 21–28, 2006.

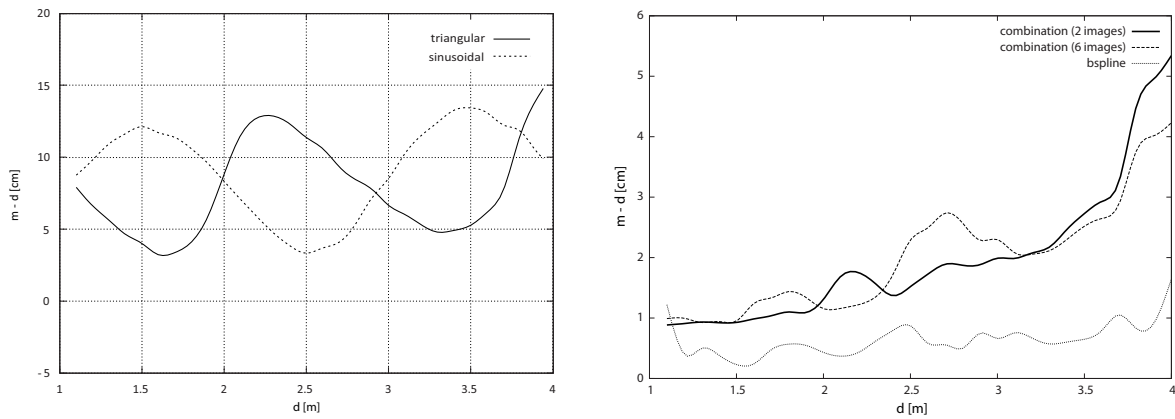


Abbildung 1: Links: Durchschnittlicher Restfehler für die ursprüngliche (gestrichelt) und die alternative Demodulation (durchgezogen). Rechts: Durchschnittlicher Restfehler für den kombinierten Ansatz unter Berücksichtigung von zwei (durchgezogen) bzw. sechs Referenzmessungen (gestrichelt) im Vergleich zu einem B-Spline-basierten Ansatz basierend auf ca. 30 Referenzmessungen.

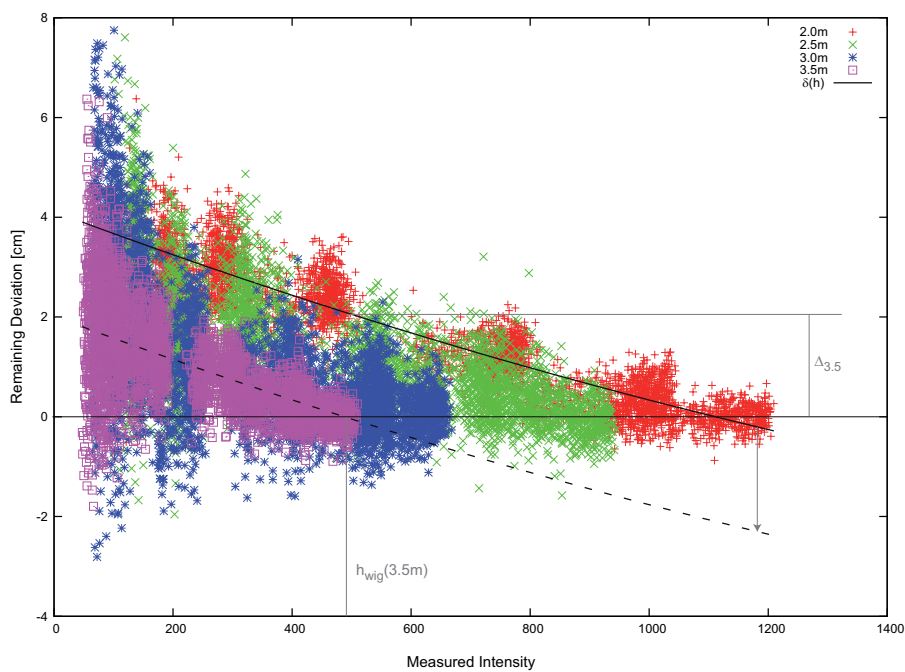


Abbildung 2: Restfehler nach einer Demodulationskorrektur für sechs verschiedene Reflektivitäten mit vier unterschiedlichen Entfernungen.

		100%	80%	60%	40%	20%	0%
1.0 m	Wiggling	0.1073	0.8905	1.6932	2.4432	2.7929	2.7218
	Coupled	0.1973	0.1918	0.2714	0.2838	0.3498	0.7179
	Decoupled	0.1325	0.1239	0.1470	0.2032	0.3627	0.7572
1.4 m	Wiggling	0.1088	0.7991	1.4507	1.6025	2.0964	2.5058
	Coupled	0.3619	0.3220	0.3809	0.3715	0.6800	1.4711
	Decoupled	0.1095	0.1974	0.3848	0.4257	0.8659	1.7188
1.8 m	Wiggling	0.2170	0.5018	0.7423	1.12713	2.1967	3.5027
	Coupled	0.6699	0.5972	0.6131	0.7908	1.5728	3.4599
	Decoupled	0.2483	0.3338	0.4809	0.9292	1.8140	3.3225
2.2 m	Wiggling	0.2438	0.6284	1.1121	1.7288	2.9523	5.0360
	Coupled	0.9506	0.9824	0.8875	1.2031	2.0513	3.3471
	Decoupled	0.2257	0.5347	0.9104	1.5165	2.8597	4.9302

Tabelle 1: Durchschnittlicher Messfehler in [cm] für verschiedene Reflektivitäten und Entfernungen bzgl. einer Wiggling-Korrektur (\varnothing 1.62 cm), einer gekoppelten (\varnothing 0.97 cm) sowie einer entkoppelten intensitätsabhängigen Korrektur (\varnothing 0.98 cm).

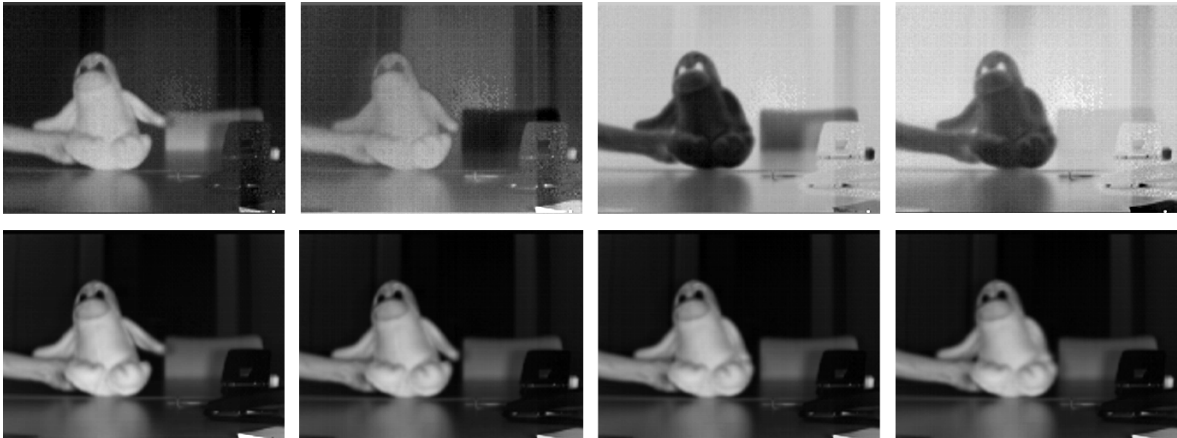


Abbildung 3: Visueller Unterschied zwischen den eigentlichen Phasenbildern $I_i = A_i - B_i$ (obere Reihe) und den entsprechenden Intensitätsbildern $I_i^+ = A_i + B_i$ (untere Reihe).

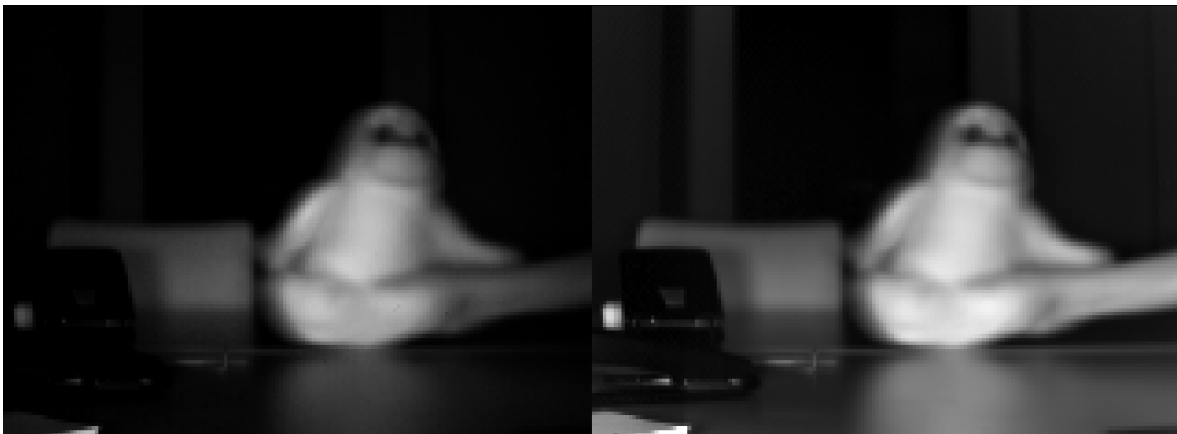


Abbildung 4: PMD Intensitätsbild vor (links) und nach (rechts) der Homogenitätsanpassung.

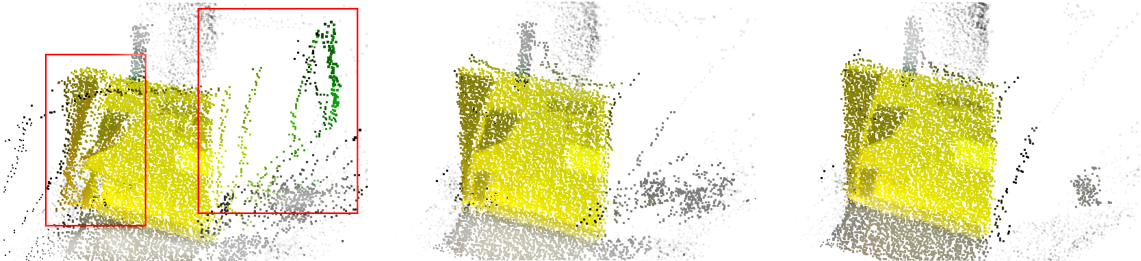


Abbildung 5: Beispielszene vor (links) und nach einer Bewegungskompensation (Mitte) im Vergleich zur statischen Referenzszenen (rechts).

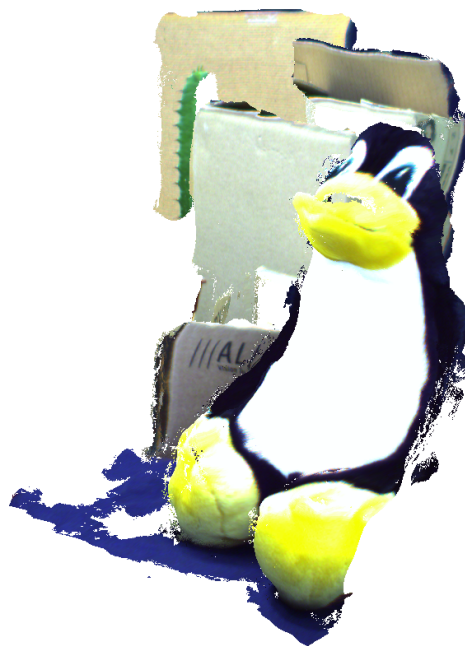


Abbildung 6: Fusionsergebnis auf Basis einer Moving Least Square Rekonstruktion / Verfeinerung.

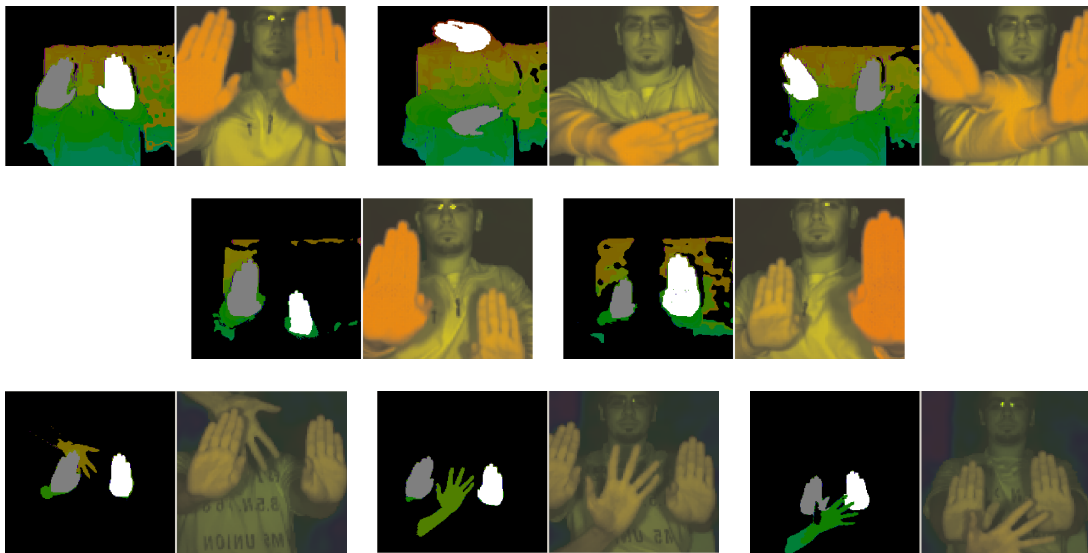


Abbildung 7: Ergebnisse des hierarchischen Clusteringverfahrens (die weiße und die graue Hand werden als linke bzw. rechte Hand des Nutzers erkannt): Dynamische Erfassung der Hände (obere Reihe), bei Vor- und Zurückbewegungen (mittlere Reihe) und bei Auftreten einer weiteren, störenden Hand (untere Reihe).