
Von optischer 3D-Messtechnik zu lichtfeldbasierter Bildakquisition und -verarbeitung

Bernd JÄHNE

Zusammenfassung

Eine Beschreibung einer 3D-Szene mittels Lichtfeldern führt nicht nur zu einer umfassenden Theorie ALLER möglichen optischen 3D-Aufnahmesysteme, sondern ermöglicht auch neue Ansätze, die auch die Schätzung der optischen Eigenschaften der Oberflächen integrieren. Weiterhin werden damit die Grenzen bisheriger optischer Messtechniken gesprengt. Am Beispiel von drei aktuellen Forschungsarbeiten aus dem HCI wird das Potenzial dieser Techniken aufgezeigt: a) die Bestimmung von Tiefenkarten und Bildaufnahme mit erweiterter Tiefenschärfe mittels hyperchromatischer Bildaufnahme, b) die Gewinnung von Tiefenkarten mittels plenoptischer Kameras nicht durch Korrespondenzanalyse, sondern durch Orientierungsanalyse im Lichtfeldes und c) Stereomessungen an spekulär reflektierenden Oberflächen mittels aktiver Beleuchtung.

1 Limitierung der projektiven Bildaufnahme

„Ein Bild sagt mehr als tausend Worte“, aber es sagt wenig über eine 3D-Szene aus: die 3D-Struktur der Szene geht verloren und wir können nur wenig über die optischen Materialeigenschaften der abgebildeten Objekte aussagen. Sowohl in der optischen 3D-Messtechnik als auch in der Computer Vision wurden in den vergangenen Jahrzehnten enorme Fortschritte gemacht, die 3D-Struktur einer Szene zu erfassen und präzise zu vermessen. Dabei standen generell geometrische Ansätze im Vordergrund. Es wundert daher nicht, dass diese Ansätze versagen bei „nicht kooperativen“ Oberflächen, also solchen mit komplexen optischen Eigenschaften, die nur mit einer bidirektionalen Reflexionsverteilungsfunktion (BRDF) oder noch komplizierteren Funktionen beschrieben werden müssen, und nicht einfach ideal glatt oder matt sind.

Um an diesem Punkt Fortschritte zu erreichen, bedarf es zuerst einer grundlegenden Analyse der Limitierungen der 2D-Bildaufnahme. Zuerst gibt es die gut bekannten geometrischen Limitierungen:

Verlust der Tiefe. Diese elementare Tatsache beruht auf den Gleichungen der projektiven Abbildung. Die Position im Bild gibt uns nur das Verhältnis von Objektgröße zu Entfernung. Ein Objektpunkt könnte an jedem Punkt entlang eines Projektionsstrahls liegen. Wo er tatsächlich liegt, wissen wir nicht.

Verlust der Objektgröße. Mit dem Verlust an Tiefeninformation geht auch der Verlust der Objektgröße einher.

Verlust der Neigung von Oberflächen. Mit der fehlenden Positionsinformation ist auch unbekannt, wie die Oberflächen des Objekts geneigt sind. Es mag verwundern, dass dieser Punkt extra erwähnt wird, aber er hat eine besondere Bedeutung, da die Orientierung einer Oberfläche relativ zur Richtung der Beleuchtung die Helligkeit mit bestimmt, mit der das Objekt von einer Kamera gesehen wird.

Verlust der 3D-Struktur. Erweitert man die Betrachtungen von Oberflächen im Raum zu einer semi-transparenten Welt – wie bei Abbildung mit Röntgenstrahlen –, so werden alle Objekte im Projektionsstrahl überlagert und die 3D-Struktur ist aus einer Projektion gänzlich unzugänglich. Es wird nicht einmal möglich, selbst 2D-Silhouetten der Objekte voneinander zu trennen.

Eine zweite Klasse von Limitierungen wurde bisher wenig beachtet. Diese ist nicht geometrischer, sondern radiometrischer Natur. Die Bestrahlungsstärke auf der Bildebene sagt nämlich nichts aus über die „Helligkeit“ des Objekts, sprich die optischen Eigenschaften des Objekts. Auch hier können wir nur ein mehrfaches Unwissen konstatieren:

Bildhelligkeit ist kein Objektmerkmal. Diese elementare Tatsache drückt eines der Grundprobleme der Bildverarbeitung aus. Die Helligkeit eines Objekts im Bild ist eine komplexe Funktion, die von seinen Reflexionseigenschaften, der Neigung seiner Oberflächen und der Richtung der Beleuchtung abhängt, die seine Oberflächen trifft. Deswegen ist es auch so schwierig, robuste Merkmale für Objekte bei variabler Beleuchtung zu finden.

Optische Eigenschaften des Objekts unbekannt. Aus den vorangegangenen Überlegung ist klar, dass uns ein Bild nichts über die optischen Eigenschaften eines Objekts aussagt. Selbst die Farbe ist nicht eindeutig. Wenn ein Objekt in einem Bild rot aussieht, könnte es tatsächlich ein Objekt roter Farbe sein oder ein weißes Objekt, das von einer roter Lichtquelle beleuchtet wird. Würden wir aber die optischen Eigenschaften aus Bilddaten gewinnen können, so wären dies robuste Merkmale für die weitere Bildverarbeitung.

Bestrahlung des Objekts unbekannt. Aus einem Bild können wir im Allgemeinen nicht auf die Beleuchtung der Szene schließen. Dazu müssten wir neben der Bildhelligkeit die BRDF der Objekte und die Neigung der Oberflächen kennen. Erst dann erschließt sich die Beleuchtungsstärke der Objektoberflächen.

2 Lichtfelder

Woher rühren alle die im letzten Abschnitt besprochenen Limitierungen? Letztlich nur daran, dass EIN Bild viel zu wenig über die Szene aussagt. Was wäre aber, wenn wir ALLE möglichen Bilder von einer Szene aufnehmen könnten?

Damit ist zuerst einmal gemeint, dass wir eine Szene aus allen möglichen Richtungen gleichzeitig aufnehmen. Das führt zu der Größe, die *Lichtfeld* genannt wird. Diese Funktion kann unterschiedlich definiert werden. Einmal könnte man an jedem Punkt im Raum angeben, welche Strahlungsdichte von ihm ausgeht. Die Strahlungsdichte wird im Allgemeinen in jede Richtung eine andere sein, sodass sich eine fünfdimensionale Funktion ergibt:

$$L(X, Y, Z, \theta, \varphi). \tag{1}$$

Diese Definition ist notwendig, wenn im Raum semi-transparenten Objekte vorhanden sind.

Hat man es aber mit undurchsichtigen Oberflächen im Raum zu tun, so weiß man, das auf einem Strahl Licht nur von einer Oberfläche kommen kann, und wir können das Lichtfeld auf eine vierdimensionale Funktion reduzieren. Dazu gibt es zwei Möglichkeiten. Entweder wir legen eine Referenzkugel um die zu beobachtende Szene und betrachten auf jedem Punkt der Kugel aus allen Richtungen die Strahlung. Alternativ können wir auch eine Referenzebene nehmen und erhalten dann

$$L(x, y, \theta, \varphi), \quad (2)$$

wobei x und y die Koordinaten der Referenzebene sind.

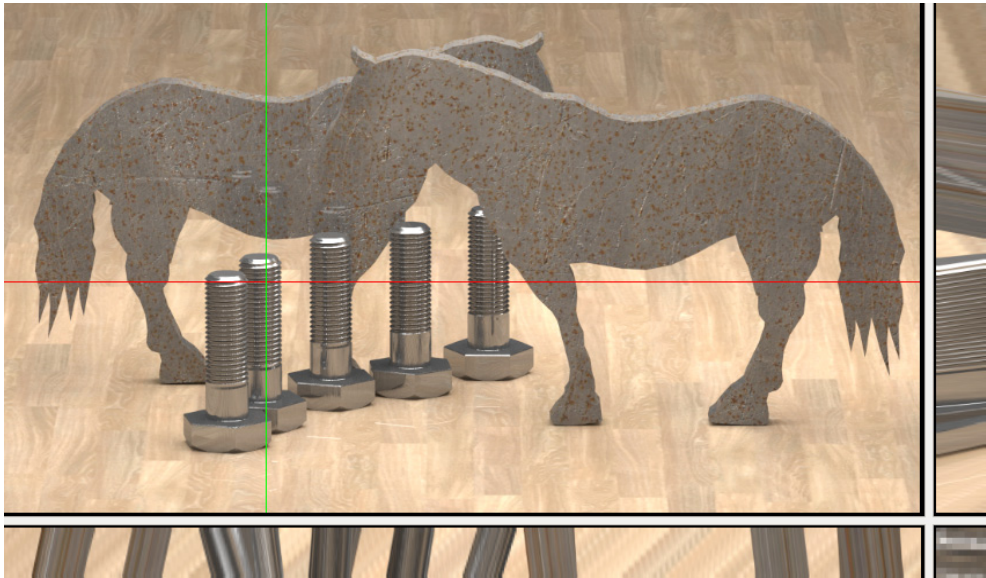


Abb. 1: Präsentation eines kleinen Winkelbereichs eines 4D-Lichtfeldes (aus WANNER et. al. 2012), siehe auch HCI Light Field Toolbox – LiFiT (hci.iwr.uni-heidelberg.de/HCI/Research/LightField/)

In welcher Relation steht jetzt eine klassische projektive Bildaufnahme mit einer Lochkamera zu dem Lichtfeld? Sie ist nichts anderes als eine 2D-Fläche aus dem 4D-Lichtfeld.

Betrachtet man nur einen Punkt im Raum, den man abbildet, dann ist die projektive Bildaufnahme mit einer Lochkamera ein Punkt aus einer Richtung von unendlich vielen möglichen Richtungen.

Damit ist klar, dass die 2D-Bildaufnahme eine viel zu dünne Abtastung des Lichtfeldes ist. Eine reale Kamera mit einer Optik nimmt nicht einen Punkt des Lichtfeldes auf, sondern integriert einen Kegel des Lichtfeldes der der objektseitigen numerischen Apertur des Objektes entspricht. Ein Stereokamerasystem tastet zwei Punkte ab – immer noch viel zu wenig.

Dabei enthält das Lichtfeld die komplette 3D-Information der Szene, weil man durch Betrachtung der Szene aus allen Richtungen die Tiefe der Objekte gewinnen kann. In Abb. 1 ist eine Präsentation eines 4D-Lichtfeldes in drei 2D-Schnitten gezeigt. Das große Bild zeigt die Szene aus einer Richtung, der horizontale/vertikale Streifen die Szene entlang der horizontalen/vertikalen Linie im Bild, wie sie sich bei Änderung der x - bzw. y -Position darstellt. An den Streifen wird deutlich, dass unterschiedlich entfernte Objekte in diesen Richtungen des 4D-Lichtfelds unterschiedlich orientiert sind. Tiefe stellt daher im Lichtfeld Orientierung dar zu kann daraus berechnet werden.

Entlang der orientierten Linien ist die Helligkeit nur dann konstant, wenn ein Objekt eine lambertsche Oberfläche aufweist. In Abb. 1 sieht man viele Linien, deren Helligkeiten sich ändern. Trotzdem bleibt die Orientierung erhalten. Daraus wird schon qualitativ deutlich, dass eine Orientierungsanalyse offensichtlich wesentlich robuster ist als eine Stereokorrespondenzanalyse, die schnell versagt, wenn das Aussehen des Objektes sich mit der Richtung ändert, weil man dann nicht die ganze Linie sieht, sondern nur zwei Punkte auf derselben.

3 BRDF und aktive Lichtfelder

Das bisher beschriebene Lichtfeld liefert immer noch nicht die komplette Information über die Szene. Bisher wurde einfach die Beleuchtung genommen, wie sie ist, und nicht berücksichtigt, die Szene aus unter allen möglichen Richtungen zu beleuchten. Dieser Ansatz erweitert das Lichtfeld um zwei Dimensionen zu

$$L(x, y, \theta, \varphi, \theta_i, \varphi_i), \quad (3)$$

wobei θ_i und φ_i die Richtungen sind, aus der beleuchtet wird. Es ist offensichtlich, dass dieses erweiterte oder aktive Lichtfeld die Information über die BRDF mit einschließt, da jetzt ein Oberflächenpunkt mit bekannter Bestrahlungsstärke aus allen Richtungen beleuchtet und betrachtet wird. Dieses sechsdimensionale Lichtfeld enthält damit die komplette Information über eine Szene, die optisch einsehbar ist. Insbesondere sind damit aussagekräftige robuste und beleuchtungsinvariante Merkmale gegeben, mit denen sich Objekte leicht identifizieren und voneinander unterscheiden lassen.

Das praktische Problem ist allerdings, dass sich diese sechsdimensionalen Lichtfelder nicht vollständig aufnehmen lassen. Bei einer Auflösung von 1.000 Punkten in alle Richtungen beinhaltet es 10^{18} Datenpunkte, weit mehr als heute auf dem größten Rechner heute gespeichert werden kann („Fluch der hohen Dimension“). Damit ist der Schlüssel zum erfolgreichen Einsatz der lichtfeldbasierten Bildakquisition und -verarbeitung eine möglichst dünnbesetzte optimale Abtastung des Lichtfeldes. Welches Potenzial dieser Ansatz ermöglicht, wird in den nächsten drei Beispielen erläutert.

4 Erweiterte Tiefenschärfe

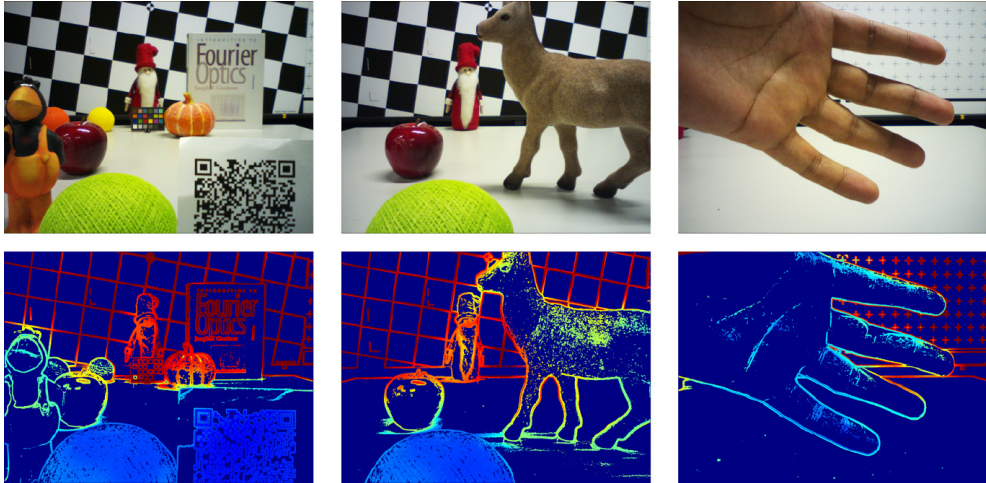


Abb. 2: Bilder mit erweiterter Tiefenschärfe (obere Reihe) und dazugehörige codierte Tiefenkarte (unterer Reihe), gewonnen mittels einer hyperchromatischen Optik und einem Schärfetransportalgorithmus zwischen den Farbkanälen (ATIF & JÄHNE 2012, ATIF 2013).

Das erste Beispiel illustriert die Überwindung einer Limitierung, die für viele Applikationen relevant ist: Aufnahmen mit einer Tiefenschärfe, die über den einer normalen Optik hinausgeht. Mit konventionellen Optiken ist eine hohe Tiefenschärfe nur durch Abblendung möglich, was aber hohe Lichtintensitäten erfordert, wie sie in vielen Applikationen nicht realisiert werden können.

Die Lösung für dieses Problem ist eine „schlechte“ Optik, die unscharf abbildet, dafür aber die Unschärfe über einen weiten Tiefenbereich so begrenzt, dass die Modulationstransferfunktion noch so hohe Werte hat, dass durch inverse Filterung eine optimale Schärfe bei einem akzeptablen Signal/Rausch-Verhältnis erreicht werden kann.

Die einfachste Möglichkeit dafür ist eine hyperchromatische Optik, bei der alle Aberrationen bis auf die chromatische ausreichend gut korrigiert sind. Es ist möglich, Optiken zu konstruieren, deren Brennweite fast linear mit der Wellenlänge von blau nach rot um etwa 10 % variiert. Damit ist für Abbildung von unendlich bis zu einem Abbildungsmaßstab von 0.1 das Bild jeweils für eine Farbe scharf. Durch eine Kombination des Farbwerts mit der Schärfe aus dem am schärfsten abgebildeten Farbkanal („Schärfetransport“) kann sowohl ein Bild mit erweiterter Tiefenschärfe als auch eine Tiefenkarte berechnet werden, bei der sich typischer Weise etwa zehn Tiefenebenen unterscheiden lassen (Abb. 1; ATIF & JÄHNE 2012, ATIF 2013).

5 Plenoptische Tiefenkarten

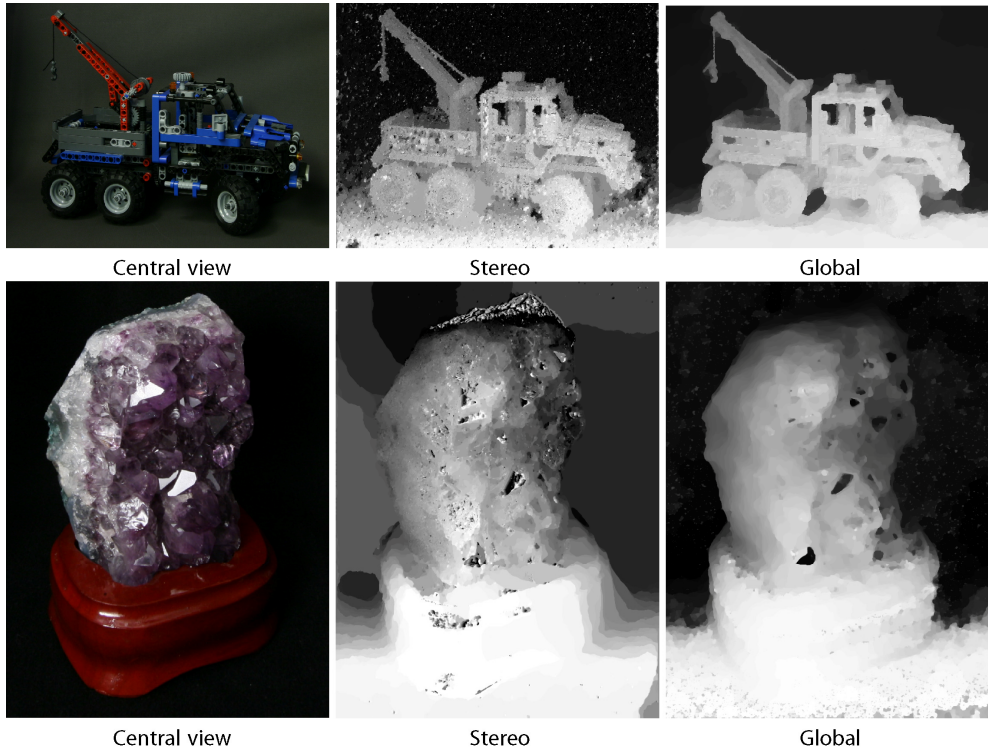


Abb. 3: Tiefenkarten gewonnen aus einer plenoptischen Kamera mit einem konventionellen Stereoansatz und einem globalen Variationsansatz mit Orientierungsschätzung im Lichtfeld (aus WANNER & GOLDLÜCKE 2012)

Bisher ist nur ein Kameratyp kommerziell verfügbar, der Lichtfelder aufnimmt, die sogenannte plenoptische Kamera. Bei einer solchen Kamera wird durch ein Mikrolinsenarray vor dem Bildsensor ein Facettenbild erzeugt, in dem auf Kosten der räumlichen Auflösung, etwa ein Dutzend Strahlenrichtungen durch die Optik unterschieden werden können. Damit wird entsprechend der objektseitigen numerischen Apertur ein kleiner Raumwinkelbereich des Lichtfeldes aufgenommen.

Am interessantesten ist die plenoptische Kamera 2.0, wie sie von der Firma Raytrix realisiert wurde (www.raytrix.de; PERWASS & WIETZKE 2012), da diese durch drei verschiedene Brennweiten der Mikrolinsen einen erweiterten Tiefenschärfebereich hat und gleichzeitig nur einen geringen Auflösungsverlust aufweist.

Mit den so aufgenommen Lichtfeldern ist es möglich, Tiefenkarten durch Schätzung der Orientierung im Lichtfeld (vergl. Abschnitt 2) zu gewinnen, die gegenüber konventioneller Stereokorrespondenzschätzungen wesentlich genauer und bei nicht-lambertschen Oberflächen auch robust sind (Abb. 3; WANNER et al. 2011, WANNER & GOLDLÜCKE 2012a). Mit

ähnlichen Variationssätzen ist es auch möglich, den räumlichen Auflösungsverlust der plenoptischen Kameras weitgehend zu kompensieren (WANNER & GOLDLÜCKE 2012b).

6 Stereo mit spekularen Oberflächen

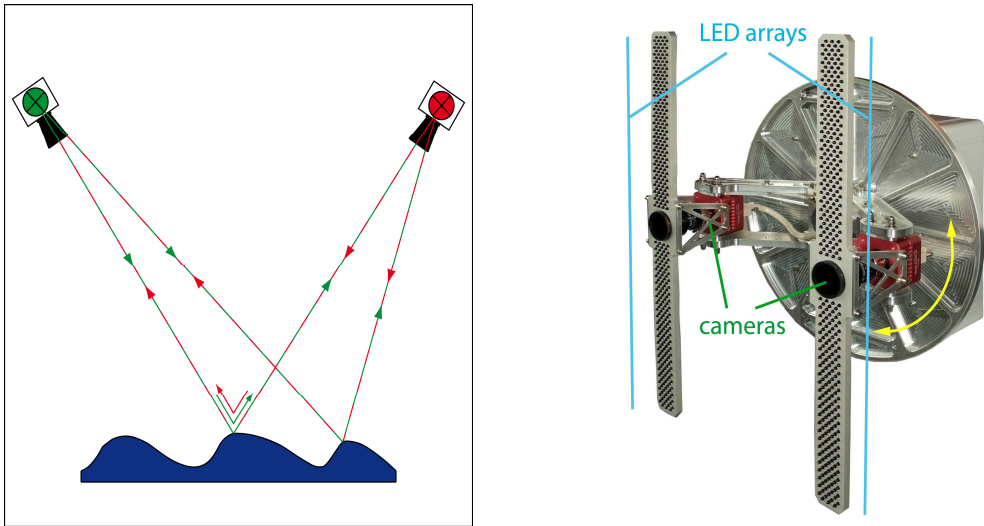


Abb. 4: Prinzip der reziproken Beleuchtung (links) und praktische Realisierung (rechts) eines Stereosystems mit aktiver Beleuchtung für Messungen an spekularen Oberflächen (KIEFHABER et al. 2011)

Das letzte Beispiel illustriert die Möglichkeiten von Akquisitionssystemen mit aktiver Beleuchtung. Konventionelle Stereobilder machen bei spiegelnd reflektierenden Oberflächen keinen Sinn, da sich korrespondierende Punkte nicht finden lassen. Möglich ist dies aber durch ein spezielles Beleuchtungssystem (Abb. 4), das schon von HELMHOLTZ & NAGEL (1906) beschrieben wurde. Die Bilder der rechten bzw. linken Kamera werden jeweils mit einer Lichtquelle von der Position der anderen Kamera aus beleuchtet. Damit ist der Lichtweg bis auf die Richtung für beide Kameras gleich und es werden die gleichen Reflexe gesehen.

Diese Anordnung wurde erstmals von WAAS & JÄHNE (1992) für statistische Messung von Wasserwellen auf der Ozeanoberfläche benutzt und später auch in der Computer Vision benutzt (ZICKLER et al. 2003). KIEFHABER et al. (2011) haben das System von WAAS & JÄHNE (1992) technisch verbessert.

7 Zusammenfassung und Ausblick

Die Einführung von Lichtfeldern führt zu einer vereinheitlichten Theorie der Bildakquisition, mit der alle bisherigen optischen 3D-Aufnahmesysteme als Spezialfälle erscheinen und die auch die Schätzung der optischen Eigenschaften der Oberflächen möglich machen. Mit Lichtfeld-basierten Bildakquisitionssysteme können Defizite konventioneller optischer 3D-Systeme überwunden werden.

Solche Systeme stehen erst am Beginn Ihrer Entwicklung, da mit den plenoptischen Kameras bisher nur eine Art eines solchen Systems kommerziell verfügbar ist. Wie die drei Beispiele in diesem Beitrag zeigen, hat dieser neue Ansatz das Potenzial, die Grenzen konventioneller optischer Systeme zu überwinden und es ist zu erwarten, das in den nächsten Jahren eine Fülle neuartiger lichtfeldbasierter Bildakquisitionssysteme auf den Markt kommen werden, mit Eigenschaften, die bisherige Systeme nicht realisieren konnten.

Literatur

- Atif, M. (2013): *Dissertation*. Heidelberg Collaboratory for Image Processing, Universität Heidelberg (in Vorbereitung).
- Atif, M. & Jähne, B. (2012): *Optimal Depth Estimation from a Single Image by Computational Imaging using Chromatic Aberrations*. Forum Bildverarbeitung, KIT Scientific Publishing.
- von Helmholtz, H. & Nagel, W. A. (1909): *Handbuch der Physiologischen Optik*. 3. Auflage. Verlag von Leopold Voss, Hamburg.
- Kiefhaber, D., Rocholz, R. & Jähne, B. (2011): *Improved optical instrument for the measurement of water wave statistics in the field*. Gas Transfer at Water Surfaces 2010, edited Komori, S., McGillis, W. and Kurose, R., S. 524-534.
- Perwass, C. & Wietzke, L. (2012): *Single lens 3D-camera with extended depth-of-field*. Proc. SPIE 8291, Human Vision and Electronic Imaging XVII, 829108.
- Waas, S. & Jähne, B. (1992): *Combined slope-height measurements of short wind waves: first results from field and laboratory measurements*. Proc. SPIE 1749, S. 295-306.
- Wanner, S., Fehr, J. & Jähne, B. (2011): *Generating EPI Representations of 4D Light Fields with a Single Lens Focused Plenoptic Camera*. Int. Symp. on Visual Comp. (ISVC).
- Wanner, S. & Goldlücke, B. (2012a): *Globally Consistent Depth Labeling of 4D Light-fields*. IEEE Conf. Comp. Vision & Pattern Recogn. (CVPR).
- Wanner, S. & Goldlücke, B. (2012b): *Spatial and Angular Variational Super-Resolution of 4D Light Fields*. Europ. Conf. Comp. Vision (ECCV).
- Zickler, T. E., Ho, J., Kriegman, D. J., Ponce, J. & Belhumeur, P. N. (2003): *Binocula Helmholtz Stereopsy*. Int. Conf. Comp. Vision (ICCV).