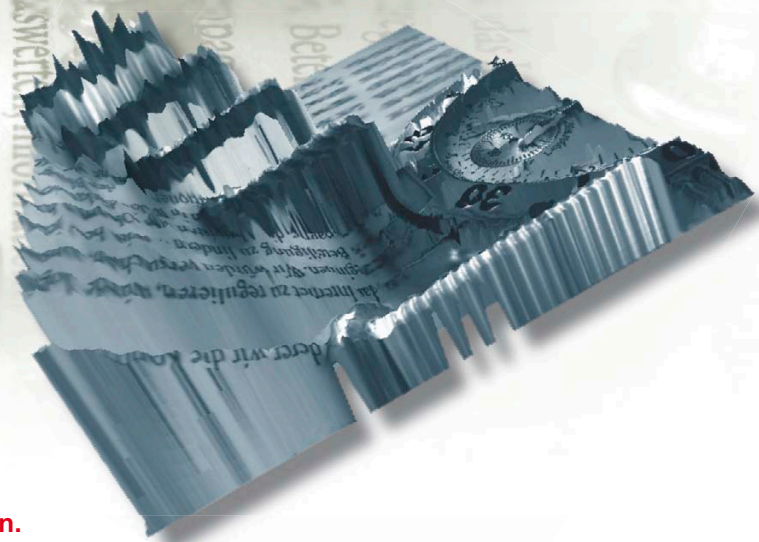


Mikrolinsen-basierte 4D-Lichtfeldkameras zur räumlichen Bilderfassung durch ein einziges Objektiv in einer Aufnahme

Christian Perwaß, Lennart Wietzke, Raytrix GmbH, Kiel

Die zunehmende Leistungsfähigkeit der digitalen Bildverarbeitung erschließt immer weitere Anwendungsbereiche. Besonders die Industrie fordert von Kameraherstellern eine stetige Steigerung der Auflösung und Bildraten sowie einen möglichst hohen Informationsgehalt der erfassten Bilder. Dies bedeutet u.a. Farbe und ganz allgemein eine optimale „Bildqualität“, aber auch der Mehrwert von 3D-Bildern gewinnt an Bedeutung. Wir präsentieren hier eine innovative Technik, die es ermöglicht, mit nur einer Kamera in einer einzigen Aufnahme ein dreidimensionales Bild zu erfassen.



Das Aufnahmeprinzip, das der hier vorgestellten Kameratechnologie zugrunde liegt, wurde schon 1908 bei der französischen Akademie der Wissenschaften unter dem Begriff „Integrale Photographie“ von M. G. Lippmann vorgestellt [1], der später den Nobelpreis für Farbfotographie erhielt. Seine Idee war es, eine Reihe kleiner Linsen so vor einer Fotoplate anzuordnen, dass diese im Brennpunkt der Linsen liegt, wodurch die Richtungen parallel eintreffender Lichtbündel in Positionen auf der Fotoplate umgesetzt werden. Bei einer Rückprojektion eines so aufgenommenen Bildes durch das gleiche Linsenarray, ergibt sich für den Betrachter dann ein dreidimensionaler Eindruck der aufgenommenen Szene. Nach einem ähnlichen Prinzip funktionieren auch aktuelle autostereoskopische Monitore und die „Wackelbilder“-Postkarten. Vermutlich aufgrund der Schwierigkeiten, ausreichend kleine Linsen in der notwendigen Qualität zu produzieren, hat sich Lippmanns Integrale Photographie Anfang des 20. Jahrhunderts nicht durchgesetzt. Die Technologie wurde dann in regelmäßigen Abständen immer wieder neu entdeckt und neu patentiert. Aber erst jetzt, zu Beginn des 21. Jahrhunderts, sind die technischen Voraussetzungen für einen

wirtschaftlichen Einsatz von Lippmanns Idee gegeben: Digitale Bildsensoren mit Auflösungen jenseits von 5 Megapixeln, Mikrolinsenarrays in hoher Qualität sowie eine sehr hohe Rechenleistung sind verhältnismäßig günstig verfügbar. Neben dem Begriff „Integrale Photographie“ gibt es auch den Begriff „plenoptische Kamera“ zur Beschreibung einer Kamera, die ein Linsenarray verwendet. Die verallgemeinerte theoretische Beschreibung dessen, was solche Kame-

treffen. Wegen der vierdimensionalen Koordinatensätze oder „4-Tupel“ (s, t, u, v) spricht man auch von einem 4D-Lichtfeld. Alternativ kann der 4D-Raum auch als zweidimensionale Position eines auftreffenden Lichtstrahls plus zwei Dimensionen zur Beschreibung seiner Richtung betrachtet werden.

1 Optisches Funktionsprinzip

Das grundlegende optische Funktionsprinzip ist in **Bild 1** dargestellt. Das Licht, das von einem Objektpunkt rechts von dem Objektiv ausgeht, wird durch das Objektiv gebündelt und auf einen (virtuellen) Punkt hinter der Bildebene abgebildet. Die Mikrolinsen, die vor dem Bildsensor in den optischen Pfad eingebracht sind, bilden jeweils nur einen Teil des gesamten Lichtbündels auf jeweils einen Punkt auf dem Bildsensor ab. Auf diese Weise wird nicht

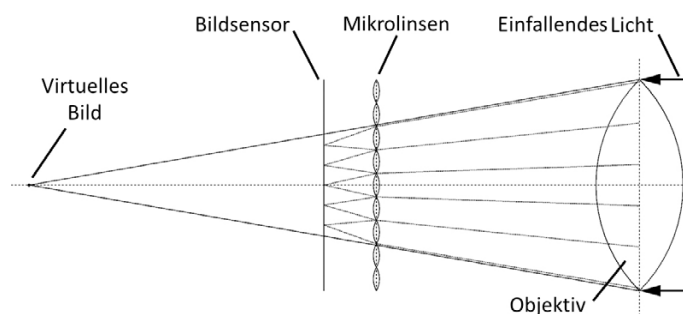


Bild 1: Abbildungsprinzip einer Lichtfeldkamera

das gesamte zum Objektpunkt gehörende Licht wieder in einem Punkt vereint, sondern je nach Abstrahlrichtung auf verschiedene Punkte auf dem Bildsensor verteilt. Dieser erfasst somit die Lichtintensität in Abhängigkeit von Position und Richtung des einfallenden Lichts. Eine andere Betrachtungsweise des Abbil-

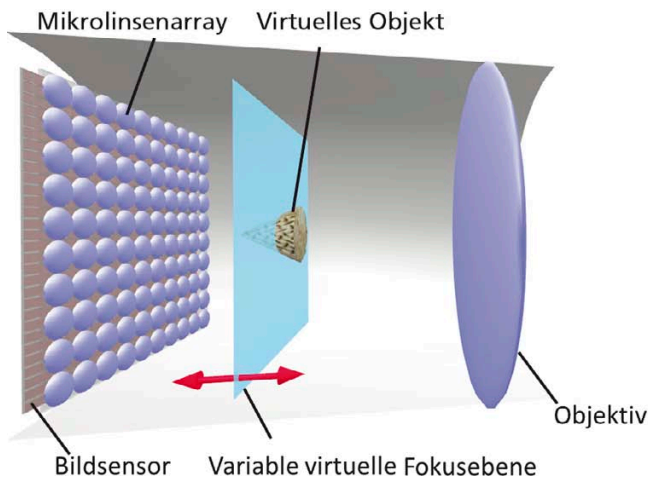


Bild 2:
Betrachtungsweise
des Mikrolinsenarrays
als Kamera-Array



Bildinhalt abhängigen Entfaltung. Bei den Ansätzen von Ng und Georgiev wird die Schärfentiefe einer Kamera nur minimal erhöht, und die Refokussierung ist lediglich eine Reduzierung der Schärfentiefe auf eine bestimmte Ebene.

Die effektive Auflösung einer plenoptischen Kamera, also die Auflösung, mit der sich Bilder aus dem Rohbild synthetisieren lassen, setzt sich aus zwei Komponenten zusammen: Der Anzahl der Mikrolinsen, die einen Punkt des virtuellen Bildes abbilden, und der Schärfentiefe der Mikrolinsen. Je mehr Mikrolinsen einen Punkt abbilden, umso geringer die effektive Auflösung. Wenn die Schärfentiefe der Mikrolinsen außer Acht gelassen wird, so verhält sich die effektive Auflösung umgekehrt proportional zum Abstand eines virtuellen

dungsprinzips zeigt **Bild 2**. Hier erzeugt das Objektiv das virtuelle Bild eines Objektes vor dem Mikrolinsenarray. Es ist jetzt einfach, sich die Mikrolinsen wie kleine Kameras vorzustellen, die aus verschiedenen Perspektiven auf das virtuelle Objekt „schauen“. Die Tiefeninformation kann nun aus den Bildern der Mikrolinsen berechnet werden, da benachbarte Mikrolinsen jeweils ein Stereokamerasystem darstellen. Anstatt also ein Kamera-Array aufzubauen und direkt auf die Szene zu schauen, wird ein Objektiv verwendet, um die Szene verkleinert vor dem Mikrolinsenarray abzubilden. Der geringe metrische Abstand zwischen den Mikrolinsen entspricht hierbei einem wesentlich größeren Abstand im Gegenstandsraum, wodurch ausreichend Parallaxe für eine Tiefenberechnung verfügbar ist.

Jüngere Arbeiten, die die Grundidee von Lippmann wieder aufgreifen, sind die von Yi-Ren Ng, Marc Levoy und Pat Hanrahan aus Stanford [3] sowie von Todor Georgiev und Andrew Lumsdaine von Adobe Systems [4]. Der Ansatz der Stanford-Gruppe platziert das Mikrolinsenarray so, dass die Brennpunkte der Mikrolinsen auf der Bildsensorebene liegen. Wie schon eingangs erwähnt ergibt sich hieraus eine direkte Umsetzung der Richtung des einfallenden Lichts in eine Position auf der Bildebene. Möchte man nun aber aus dem so erzeugten Rohbild ein für einen Betrachter „normales“ Bild synthetisieren, so entspricht die Auflösung des synthetisierten (oder refokussierten) Bildes der Anzahl der Mikrolinsen. Bei einem Mikrolinsenarray mit 40 000 Linsen und einem 16 Megapixel-Sensor beträgt die Auflösung der resultierenden Bilder also lediglich 40 000 Pixel, was den Anwendungsbereich einer solchen Kamera stark einschränkt.

Todor Georgiev hat in seinen Veröffentlichungen vorgeschlagen, das Mikrolin-

senarray so zu platzieren, dass die Brennpunkte der Mikrolinsen nicht auf der Bildebene liegen. Mit diesem Ansatz konnte er zeigen, dass sich aus dem Rohbild einer 16 Megapixel-Lichtfeldkamera, Bilder mit bis zu 1 Megapixel Auflösung synthetisieren lassen. Georgiev gibt in seinen Arbeiten allerdings nicht an, wie genau das Mikrolinsenarray zu platzieren ist, um eine möglichst hohe synthetische Auflösung zu erhalten. Ausgehend von diesen Arbeiten haben wir zwei grundsätzliche Verbesserungen entwickelt:

1. Eine detaillierte Theorie, die den Zusammenhang zwischen sämtlichen Mikrolinsenarrayparametern und der effektiven Auflösung der synthetisierten Bilder beschreibt.
2. Ein Mikrolinsenarray-Design, bei dem Mikrolinsen mit verschiedenen Brennweiten verwendet werden, um die Schärfentiefe einer Kamera zu erweitern.

Damit kann eine effektive Auflösung von bis zu einem Viertel der Sensorauflösung erreicht werden. Bei einem 16 Megapixel-Sensor können somit synthetisierte Bilder mit bis zu 4 Megapixeln erzeugt werden. Durch die Verwendung von Mikrolinsen mit verschiedenen Brennweiten lässt sich weiterhin die Schärfentiefe um bis zu einen Faktor 6 erhöhen, ohne den Einsatz einer mathematisch aufwendigen und vom

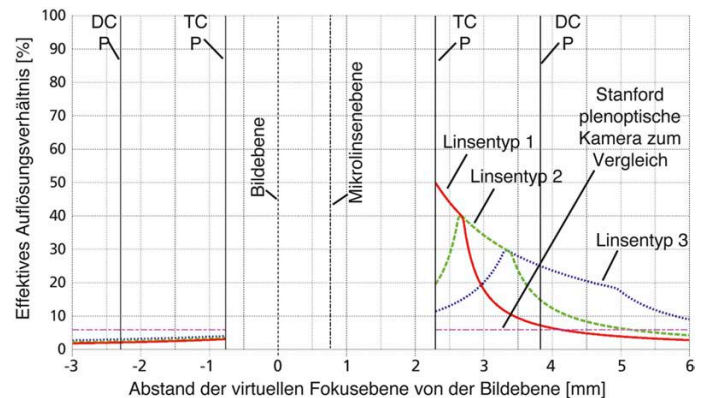


Bild 3: **Abhängigkeit der effektiven Auflösung synthetisierter Bilder vom Abstand der virtuellen Fokusebene vom Bildsensor, berechnet für eine Bilddimension**

Punktes von der Mikrolinsenebene. Dieser starke Abfall in der effektiven Auflösung durch Mehrfachabbildung wird dann noch weiter reduziert durch die Schärfentiefe der Mikrolinsen.

Bild 3 veranschaulicht den Zusammenhang der effektiven Auflösung synthetisierter Bilder mit dem Abstand eines virtuellen Bildes von der Bildebene. In diesem Beispiel wird ein Mikrolinsenarray mit drei verschiedenen Fokallängen angenommen, die genau so berechnet sind, dass ihre Schärfentiefebereiche aneinander anschließen. Zum Vergleich ist die effektive Auflösung des Stanford-Ansatzes dargestellt. Die Schärfentiefe der Mikrolinsen verhält sich hier zur Anzahl der Mikrolinsen, in denen ein Punkt abgebildet wird, gerade so, dass die effektive Auflösung über einen großen Bereich konstant ist. Die mit TCP bezeichnete Ebene beschreibt die "Total Covering Plane", ab der jeder Objektpunkt in mindestens einem Mikrolinsenbild gleichen Linstertyps erfasst wird. Die DCP-Ebene ist analog die "Double Covering Plane", ab der jeder Objektpunkt

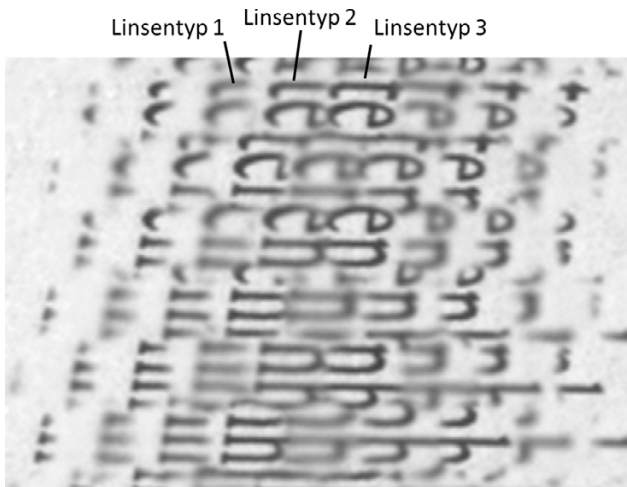


Bild 4: Ausschnitt eines Beispiel-Rohbildes

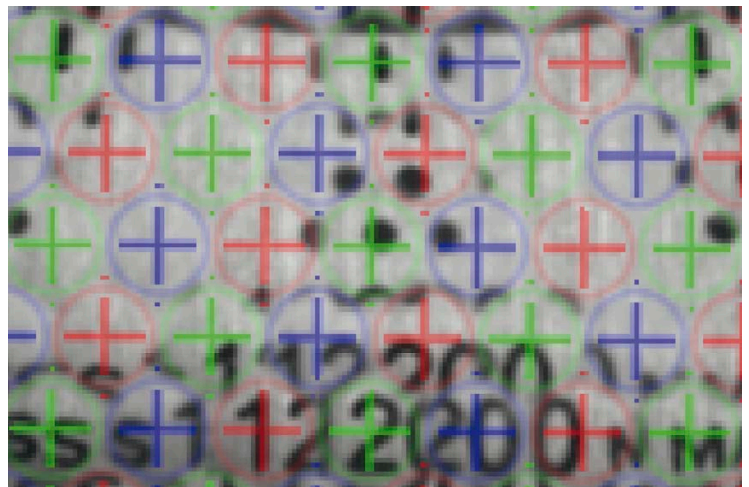


Bild 6: Linsenzentren verschiedener Linsentypen auf dem Rohbild

in mindestens zwei Mikrolinsenbildern gleichen Typs zu sehen ist.

Bild 4 zeigt einen Ausschnitt eines Rohbildes, das durch ein Mikrolinsenarray mit drei verschiedenen Fokallängen aufgenommen wurde.

2 Algorithmik

Das Verfahren zur Synthetisierung oder Refokussierung aus einem Rohbild ist in **Bild 5** schematisch dargestellt. Die geschwungene Refokussierungsoberfläche soll hier verdeutlichen, dass eine Refokussierung nicht notwendigerweise auf einer Ebene stattfinden muss, sondern auf einer beliebigen Oberfläche möglich ist. Um nun den Grau- oder Farbwert des virtuellen Punktes (x,y,z) zu berechnen, wobei die z -Koordinate die Fokusebene beschreibt, müssen lediglich die Projektionsstrahlen durch die Zentren der Mikrolinsen des für diesen Fokusbereich zuständigen Mikrolinsentyps berechnet werden. Die Schnittpunkte der Projektionsstrahlen mit der Sensorebene ergeben dann die Positionen der Bildpunkte, aus denen der resultierende Farbwert berechnet wird, z.B. über den Mittelwert. Dieses Verfahren hat eine geringe mathematische Komplexität und kann für jeden resultierenden Bildpunkt parallel berechnet werden. Dadurch bietet die Implementierung auf einem Grafikprozessor (GPU) eine sehr hohe Rechengeschwindigkeit

gegenüber einer CPU. Bei einer NVIDIA GeForce GTX480 dauert die Refokussierung eines 10,7 Megapixel-Bildes auf ein 3 Megapixel-Bild ca. 5–10 ms.

Zur Berechnung des Abstandes eines Punktes von der Kamera mittels Standard-Stereoalgorithmen betrachtet man die verschiedenen Mikrolinsen wie schon oben erwähnt als kleine Kameras. Aufgrund der präzisen Fertigung der Mikrolinsenarrays mit lithographischen Verfahren kann man davon ausgehen, dass die optischen Achsen aller Mikrolinsen parallel sind und sie exakt den definierten Abstand haben.

Bild 6 zeigt den Ausschnitt eines Rohbildes mit eingezeichneten Mikrolinsenzentren.

Der Tiefenbereich, der zur Zeit maximal abgebildet werden kann, reicht von der Ebene, auf die das Kameraobjektiv fokussiert ist, bis zum äußeren Fokalkpunkt des Objektivs. Ein Punkt, der im Zentrum eines Mikrolinsenbildes und ganz am Rand eines benachbarten Mikrolinsenbildes abgebildet wird, liegt im Gegenstandsraum auf der Fokusebene des Objektivs. Wird ein Objektpunkt hingegen auf die Zentren von benachbarten Mikrolinsenbildern abgebildet, so war das einfallende Licht parallel und somit lag der Objektpunkt auf dem äußeren Fokalkpunkt des Objektivs. Betrachtet man nur benachbarte Mikrolinsenbilder, so wird dieser gesamte Tiefenbereich über

den Radius einer Mikrolinse kodiert. Da aber näher an der Kamera liegende Punkte in mehr Mikrolinsenbildern abgebildet werden, können nicht nur direkt benachbarte Mikrolinsenbilder zur Tiefenberechnung verwendet werden, sondern auch weiter voneinander entfernte, wodurch sich die Tiefenauflösung erhöht. So lassen sich mindestens 20 verschiedene Tiefenebenen unterscheiden. Werden die Radien der Mikrolinsen größer gewählt, erhöht sich entsprechend auch die Tiefenauflösung.

Zur Berechnung der Tiefe aus dem Rohbild wird für jeden Punkt in einem Mikrolinsenbild der korrespondierende Punkt in dem nächsten Linsennachbarn gesucht. Dies kann z.B. über die Summe der absoluten Differenzen (SAD) oder eine Korrelation für einen kleinen Bereich um

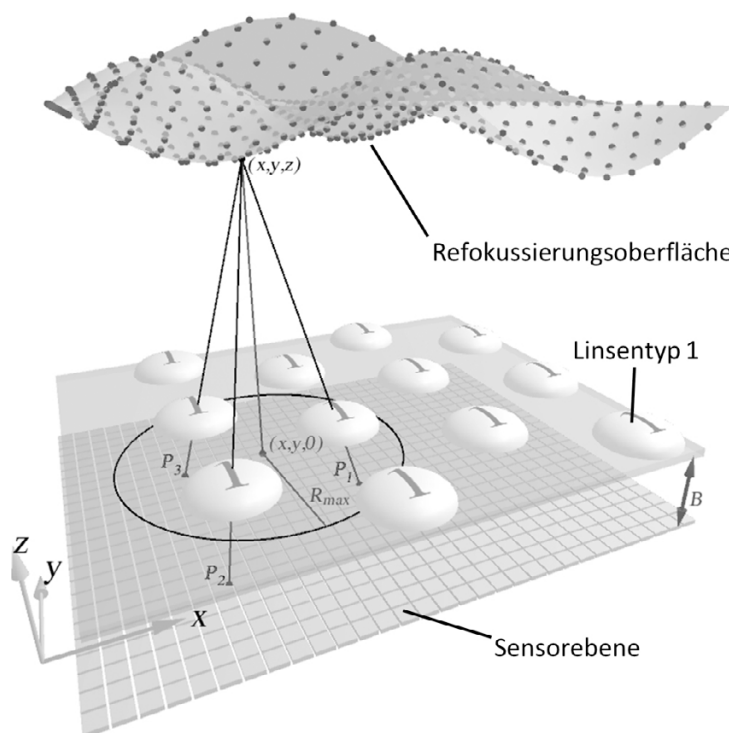


Bild 5: Geometrische Konstruktion für die Refokussierung des Bildes



Bild 7: Refokussiertes Bild

den Punkt geschehen. Da die optischen Achsen der Mikrolinsen parallel sind, verlaufen die Epipolarlinien durch die Zentren benachbarter Linien, so dass der Suchraum nur eindimensional ist.

Dies ist allerdings nur der erste Schritt. Um im refokussierten Bild für jeden Gegenstandspunkt den Tiefenwert zu erhalten, muss noch eine Fusion der Tiefenwerte stattfinden. Dafür verwendet man wieder den Ansatz aus Bild 5, wobei als Rohbild die Tiefenberechnung aus Schritt 1 verwendet wird und man für jeden Punkt (x,y) über alle Fokusebenen in der z -Koordinate sucht, wann die zugehörigen Tiefenwerte aus dem Tiefenbild der aktuellen z -Koordinate möglichst nahe kommen. Dies ist dann das Tiefenergebnis für den Punkt (x,y) . Durch diese Kombination mehrerer Tiefenwerte erhält man eine robuste Tiefenberechnung. Generell ist die Berechnung der Tiefe nur an solchen Stellen im Rohbild möglich wo ausreichender Kontrast vorhanden ist, wie an Kanten und Ecken. Der Rechenaufwand für die Tiefe ist erheblich höher

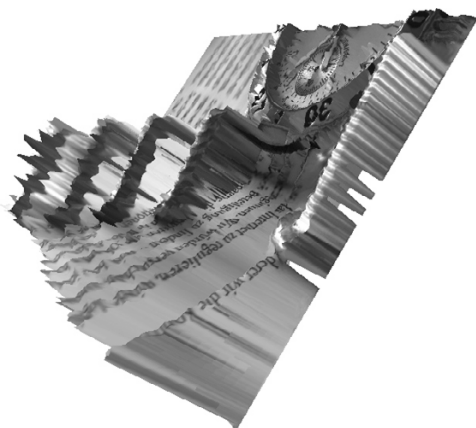


Bild 8: Darstellung der Tiefenberechnung aus einem Rohbild als 3D-Objekt. Der Bereich oben links im Bild liegt hinter der Fokusebene des Objektivs und ist dadurch nicht refokussierbar

als für die Refokussierung. Auf einer NVIDIA GeForce GTX480 dauert die Tiefenberechnung ausgehend von einem 10,7 Megapixel-Rohbild auf eine 1 Megapixel-Tiefenkarte ca. 200 ms.

Wenn ein Tiefenbild berechnet wurde, so kann auch der Fokus für jeden Bildpunkt eines refokussierten Bildes richtig gesetzt

werden, wodurch ein über den gesamten erweiterten Schärfentiefebereich scharfes Bild synthetisiert werden kann. **Bild 7** zeigt eine solche gesamt fokussierte Aufnahme und **Bild 8** die dazugehörige Tiefenberechnung als 3D-Objekt.

3 Anwendungsbereiche

Das hier vorgestellte Verfahren zur Bildaufnahme mit erhöhter Schärfentiefe und der Möglichkeit zur Berechnung der 3D-Form einer Szene hat vielfältige Einsatzgebiete. So bietet es eine einfache Lösung zur Aufnahme der 3D-Form mikroskopischer Objekte, wo ein Stereokamerasystem keinen Platz findet oder der Kameraabstand zu groß wäre. Letzteres ließe sich zwar mit zwei Kameras beheben, die ein Objekt unter verschiedenen Winkeln erfassen. Die Tiefenberechnung ist in diesem Fall aber nur in der Schnittlinie der beiden Schärfentiefebene möglich, was den Einsatzbereich stark einschränkt. Außerdem müssen bei einem Stereosystem zwei Kameras – mit nie ganz identischen Objektiven – zueinander kalibriert werden.

Beim vorgestellten Ansatz hingegen wird die Szene durch nur ein Objektiv gesehen, der Schärfentiefebereich wird erweitert, und die kritischen Kalibrierparameter sind durch die starre Anordnung der Mikrolinsen nicht veränderbar. Wenn nur relative und keine metrische Tiefeninformation gebraucht wird, ist keine Kalibrierung erforderlich. Weiterhin wird nur ein Bild benötigt, so dass auch bewegte Objekte ohne Synchronisierungsprobleme aufgenommen werden können.

Die Tiefenberechnung ist aber auch im makroskopischen Bereich möglich und hängt nur vom verwendeten Objektiv ab. So ist auch der Einsatz mit Teleobjektiven im Sicherheitsbereich denkbar.

Die Grenzen der Technologie liegen bei der Verwendung von Weitwinkelobjek-

tiven. Dort liegen die meisten berechenbaren Tiefenebenen innerhalb des Objekts, wodurch eine Tiefenberechnung von Objekten vor dem Objektiv nicht mehr in ausreichender Genauigkeit möglich ist.

4 Zusammenfassung und Ausblick

Das vorgestellte Verfahren eröffnet viele neue Möglichkeiten in der Bildaufnahme und Bildverarbeitung. Das Rohbild einer plenoptischen Kamera enthält mehr Informationen als nur die Verteilung der Lichtintensitäten. Die einfallende Lichtinformation wird detaillierter abgetastet, wodurch sich die Tiefenberechnung und Refokussierung als offensichtlichste Anwendungen ergeben. Zukünftige Kameras könnten nicht nur die Fokallängen der Mikrolinsen über dem Array verändern, sondern auch die

Lichttransmission, die Spektraleigenschaften oder die Polarisation, wodurch ein noch detaillierteres Abtasten der Lichtinformation möglich würde.

Literaturhinweise:

- [1] G. Lippmann, *Epreuves reversibles donnant la sensation du relief*, Journal of Physics, 7(4):821-825, 1908
- [2] M. Levoy, P. Hanrahan, *Light field rendering*, SIGGRAPH 96, 31-42, 1996
- [3] Ren Ng et al., *Light Field Photography with a Hand-held Plenoptic Camera*, Stanford Tech Report CTSR 2005-02, 2005
- [4] A. Lumsdaine, T. Georgiev, *Full Resolution Light-field Rendering*, Adobe Tech Report, 2008

In eigener Sache

Die Raytrix GmbH sucht derzeit dringend nach neuen Mitarbeitern. Wir bitten interessierte Fachleute um Kontaktaufnahme!

Ansprechpartner:

Dr. habil.
Christian Perwaß
Geschäftsführender
Gesellschafter
Raytrix GmbH
Schauenburgerstr. 116
D-24118 Kiel
Tel. 0431/5606240
Fax 0431/26090065
eMail: christian.perwass@raytrix.de
Internet: www.raytrix.de



*) Das obige Portraitbild ist eine 2D-Projektion einer 3D-Schätzung zur weiteren Veranschaulichung der Möglichkeiten der hier vorgestellten Technik.

www.photonik.de ▶ Webcode 6003