Bachelorarbeit im Studiengang Audiovisuelle Medien AM7

Light Fields für Virtual Reality

Herausforderungen und Chancen des Einsatzes von Lichtfeld-Technologie in Virtual Reality Systemen

vorgelegt von Johannes Kammerer

an der Hochschule der Medien Stuttgart

am 31.08.2016

zur Erlangung des akademischen Grades eines Bachelor of Engineering

Erstprüferin: Prof. Katja Schmid

Zweitprüfer: Prof. Dr. Andreas Koch

Ehrenwörtliche Erklärung

Hiermit versichere ich, Johannes Kammerer ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Light Fields für Virtual Reality – Herausforderungen und Chancen des Einsatzes von Lichtfeld-Technologie in Virtual Reality Systemen“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

_____

Ort, Datum Unterschrift

Zusammenfassung

Im Rahmen dieser Arbeit werden die Herausforderungen und Chancen untersucht, die durch den Einsatz von Light Field Technologie in Virtual Reality Systeme entstehen. Zunächst werden die theoretischen Grundlagen von Virtual Reality und der visuellen Wahrnehmung behandelt. Aktuelle Virtual Reality Systeme und verschiedene Inhalte werden dargestellt. Außerdem werden die technischen Hintergründe der Lichtfeldtechnik erörtert sowie verschiedene Aufnahmesysteme und Light Field Anwendungen analysiert. Das Ziel der Arbeit ist es, den aktuellen Forschungsstand von Light Field Technologie sowie deren Vor- und Nachteile auf dem Gebiet der virtuellen Realität zu untersuchen.

Abstract

This paper focuses on analyzing the challenges and the opportunities of using light field technology in virtual reality systems. First, the abstract foundations of virtual reality and the visual perception will be discussed. Current virtual reality systems and different kinds of content will be presented. Furthermore, the technical background of light field technology will be debated and a variety of capturing systems and light field applications will be analyzed. The purpose of this paper is to represent the current state of research of light field technology in the field of virtual reality, as well as its assets and drawbacks.

Inhaltsverzeichnis

Abbildungsverzeichnis III
1 Einleitung 1
1.1 Motivation und Problemstellung 1
1.2 Vorgehensweise und Aufbau der Arbeit 2
2 Virtual Reality 3
2.1 Was ist eine virtuelle Realität? Eine Begriffsdefinition 3
2.1.1 Die perfekte virtuelle Realität 3
2.1.2 Reizerzeugung/Simulation 4
2.1.3 Suspension of Disbelief 5
2.2 Merkmale einer virtuellen Realität 6
2.2.1 Präsenz 6
2.2.2 Immersion 6
2.2.3 Technologieorientierte Merkmale 7
2.3 Wahrnehmung im Raum 8
2.3.1 Stereosehen 8
2.3.2 Weitere Tiefenhinweise 10
2.4 Entstehungsgeschichte der virtuellen Realität 11
2.5 Aktuelle VR-Systeme 13
2.5.1 Samsung Gear VR 14
2.5.2 Google Cardboard 15
2.5.3 Oculus Rift 15
2.5.4 HTC Vive 16
2.6 Virtual Reality Inhalte 17
2.6.1 Videospiele: Real-Time Content 18
2.6.2 Live-Action: Pre-Rendered Content 18
2.6.3 Einsatzgebiete von Virtual Reality 19
2.7 Ausblick und aktuelle Probleme 20
2.7.1 360° Video: ist das Virtual Reality? 21
2.7.2 Der Vergence-Fokus Konflikt 22
3 Light Field Technologie 24
3.1 Light Field Rendering 24
3.2 Theoretische Grundlagen der Plenoptik 25

3.2.1 Die plenoptische Funktion ... 25
3.2.2 Das 4D Light Field ... 26
3.3 Light Field Aufnahmegeräte ... 27
3.3.1 Fotokamera ... 28
3.3.2 Kamera-Array ... 29
3.3.3 Plenoptische Kamera ... 30
3.4 Light Field Anwendungen ... 32
3.4.1 View Rendering ... 32
3.4.2 Synthetic Aperture ... 34
3.4.3 Visual Effects und Postproduktion ... 36
4 Light Field Technologie für VR: Aktueller Stand der Forschung ... 38
4.1 OTOY ... 39
4.2 Lytro Immerge ... 40
4.3 Light Stage Technologie ... 42
4.4 Light Field Displays ... 44
4.4.1 Near-Eye Light Field Display ... 45
4.4.2 Light Field Stereoscope ... 47
4.5 Weitere Anwendungsgebiete von Light Fields für VR ... 51
5 Fazit ... 53
5.1 Zusammenfassung ... 53
5.2 Ausblick ... 56
6 Literaturverzeichnis ... 59

Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung eines VR-Systems (Dörner et al. 2013, S. 24) 7

Abbildung 2: Das Stereosehen (Dörner et al. 2013, S. 37) 8

Abbildung 3: Manipulation des Stereosehens durch ein Stereodisplay (Dörner et al. 2013, S. 37)..... 9

Abbildung 4: Liste von Tiefenhinweisen (Dörner et al. 2013, S. 41) 10

Abbildung 5: Ivan Sutherlands Head-Mounted Display (Sutherland 1968) 12

Abbildung 6: Virtual Boy von Nintendo (https://de.wikipedia.org/wiki/Virtual_Boy#/media/File:Virtual-Boy-Set.png) 12

Abbildung 7: Development Kit 1 der Oculus Rift (https://de.wikipedia.org/wiki/Oculus_Rift#/media/File:Oculus_Rift_Dev_Kit.jpg) 13

Abbildung 8: Samsung Gear VR (http://www.samsung.com/de/promotions/galaxynote4/feature/gearvr/) 14

Abbildung 9: Google Cardboard (https://vr.google.com/intl/de_de/cardboard/get-cardboard/) 15

Abbildung 10: Consumer-Version der Oculus Rift (https://static.oculus.com/web/www_static/production/US/fb4c45313ba855fe7505d835b8b017613c12b7ce/images/rift/family.png) 16

Abbildung 11: HTC Vive (https://www.htcvive.com/de/) 17

Abbildung 12: Schematische Darstellung der plenoptischen Funktion (McMillan und Bishop 1995) . 25

Abbildung 13: Parametrisierung des Licht Felds durch einen Light Slab (Levoy und Hanrahan 1996, S. 32)..... 26

Abbildung 14: Abtastung des Licht Felds durch eine Fotokamera (Levoy und Hanrahan 1996, S. 34)26

Abbildung 15: Verschiedene Darstellungen eines Light Fields. Darstellung (a) zeigt eine Sammlung von perspektivisch leicht unterschiedlichen Bildern auf der Ebene (u, v). Darstellung (b) zeigt ein 4D-Array an Pixeln der Ebene (s, t), aus denen sich die einzelnen Perspektiven der Ebene (u, v) zusammensetzen (Levoy und Hanrahan 1996, S. 34) 27

Abbildung 16: Schematische Darstellung des Spherical Gantry (Levoy und Hanrahan 1996, S. 36) .. 28

Abbildung 17: Das Stanford Multi-Camera-Array (http://graphics.stanford.edu/projects/array/) 29

Abbildung 18: Kompaktes Array aus HD-Kameras (Schäfer et al. 2013, S. 2) 30

Abbildung 19: Strahlengang in einer plenoptischen Kamera (http://www.plenoptic.info/pages/derivation.html) 30

Abbildung 20: RAW Light Field Aufnahme einer plenoptischen Kamera. Die Vergrößerungen zeigen deutlich die kreisförmigen Subimages der einzelnen Mikrolinsen (Ng 2006, S. 44) 31

Abbildung 21: Transformation von (u, v) und (s, t) Ebene in Bildkoordinaten (x, y) (Levoy und Hanrahan 1996, S. 38)..... 32

Abbildung 22: View Rendering am Beispiel eines Terracottadrachens (Levoy 2006, S. 48) 33

Abbildung 23: Refokussierung auf eine virtuelle Sensorebene E'_0 auf Höhe des Mikrolinsenarrays. Die Synthese eines Pixels auf E'_0 erfolgt durch die Interpolation aller Pixel des jeweiligen Subimages (http://www.plenoptic.info/pages/refocusing.html) 34

Abbildung 24: Refokussierung einer Light Field Aufnahme auf unterschiedliche Ebenen (Ng 2006, S. 34)..... 35

Abbildung 25: Refokussierung auf eine virtuelle Sensorebene $E´_1$ hinter dem Bildsensor. Die Synthese eines Pixels auf $E´_0$ erfolgt nun durch die Interpolation einzelner Pixel aus unterschiedlichen Subimages (http://www.plenoptic.info/pages/refocusing.html)..... 35

Abbildung 26: Light Field Aufnahme mit einem Panoramastativ (https://home.otoy.com/otoy-demonstrates-first-ever-light-field-capture-for-vr/)..... 39

Abbildung 27: Bildmaterial des 360° Light Fields, dargestellt durch ein VR-System (https://home.otoy.com/wp-content/uploads/2015/05/OTOY-Light-Field-Capture-61.jpg)..... 39

Abbildung 28: Lytro Immerge 360° Light Field Kamera (http://www.wired.com/wp-content/uploads/2015/11/lytro-5-1024x609.jpg)..... 41

Abbildung 29: Produktionssystem der Lytro Immerge(http://ireadopedia.com/wp-content/uploads/2015/11/immerge-system.jpg)..... 42

Abbildung 30: Ganzkörperaufnahme einer Person mit der Light Stage 6 (Debevec 2012, S. 1)..... 43

Abbildung 31: Digitaler Charakter, gerendert aus unterschiedlichen Blickwinkeln im Rahmen des Relighting Human Motion Projekts (Debevec 2012, S. 2)..... 44

Abbildung 32: Aufbau des Near-Eye Light Field Displays (Lanman und Luebke 2013, S. 4)..... 46

Abbildung 33: Light Field Aufnahme, wie sie auf dem Bildsensor des Displays dargestellt wird. (links) Vor dem Sensor ist noch kein Mikrolinsenarray angebracht. Die einzelnen Subimages sind deutlich sichtbar. (rechts) Über der rechten Hälfte des Bildes liegt nun ein Mikrolinsenarray. Der Betrachter nimmt jetzt ein scharfes Bild wahr (Lanman und Luebke 2013, S. 7)..... 46

Abbildung 34: Prototyp des Near-Eye Light Field Displays (Lanman und Luebke 2013, S. 1)..... 47

Abbildung 35: Schematische Darstellung des Light Field Stereoscopes (Huang et al., S. 2)..... 48

Abbildung 36: Technischer Aufbau des Light Field Stereoscopes (Huang et al., S. 4)..... 49

Abbildung 37: "Rank-1" Faktorisierung des Light Fields. Das Bild (links) wird dabei auf das vordere Display gelegt, das Bild (rechts) auf das hintere. Im Auge des Betrachters setzt sich multiplikativ ein Light Field zusammen, in dem auf verschiedene Tiefenebenen fokussiert werden kann (Huang et al., S. 6)..... 50

Abbildung 38: CGI Beispielszene, die mit einer Fotokamera durch das Light Field Sterescope aufgenommen wurde. Die Fotokamera kann frei auf unterschiedliche Bildinhalte fokussieren (Huang et al. 2015, S. 6)..... 50

1 Einleitung

1.1 Motivation und Problemstellung

Durch den großen Hype der letzten Jahre, ausgelöst vor allem durch die Entwicklerversionen der Oculus Rift, ist das Thema Virtual Reality auf dem Vormarsch. Konzerne wie Microsoft, Google oder Facebook investieren Milliarden-Beträge in die Entwicklung von neuen VR-Systemen, Head-Mounted Displays (HMD) und VR-Anwendungen. Bis 2020 soll das Marktvolumen von Virtual Reality Hard- und Software auf bis zu 30 Milliarden US-Dollar anwachsen (Digi-Capital 2016).

Dabei scheint 2016 das Jahr zu sein, in dem Virtual Reality endlich den Massenmarkt erobert. Die ersten Consumer-Versionen von Oculus Rift, HTC Vive, sowie Playstation VR versprechen spannende VR Inhalte und immersive Videospiele für das eigene Wohnzimmer. Andere Systeme wie die Samsung Gear VR bieten mobile VR-Inhalte auf dem Smartphone an.

Allen technischen Entwicklungen zum Trotz, steht die aktuelle Generation von Virtual Reality Systemen noch vor zwei zentralen Problemen:

Bis jetzt gibt es keine Möglichkeit, Positionstracking und Bewegungsparallaxe in Live-Action VR zu realisieren. 360° Videos zeigen lediglich Panoramaaufnahmen und sind auf einen festen Rotationspunkt des Kopfes limitiert. Darüber hinaus unterstützen aktuelle Head-Mounted Displays keine natürliche Akkommodation der Augen. Einzelne Objekte können nicht fokussiert werden und der Betrachter leidet am Vergence-Fokus Konflikt. Dabei gehen wichtige Tiefeninformationen verloren, die Raumwahrnehmung wirkt unnatürlich und es kommt zu Ermüdungserscheinungen und Nebenwirkungen wie Schwindel und Kopfweh (Huang et al. 2015, S. 1–2).

Eine ganze Reihe von Experten (USC Institute for Creative Technologies 2015) sehen die Lösung dieser Probleme in der Anwendung von Light Field Technologie für Virtual Reality. Ziel dieser Arbeit ist es, beide Themen miteinander in Beziehung zu setzten und die Herausforderungen und Chancen zu analysieren, die sich durch den Einsatz von Light Field Technologie in Virtual Reality ergeben.

1.2 Vorgehensweise und Aufbau der Arbeit

Kapitel 2 bietet eine Einführung und allgemeinen Überblick in das Gebiet der virtuellen Realität. Es werden wichtige Begriffe wie Präsenz und Immersion definiert, sowie technologieorientierte Merkmale eines VR-Systems dargestellt. Das Kapitel 2.3 beschäftigt sich außerdem mit wichtigen Aspekten der visuellen Wahrnehmung. Nach einem Rückblick auf die Entstehungsgeschichte von VR-Systemen in Kapitel 2.4, werden in Kapitel 2.5 Systeme der aktuellen Generation vorgestellt und miteinander verglichen. Außerdem werden in Kapitel 2.6 die Unterschiede aktueller VR Inhalte erklärt und deren Einsatzgebiete aufgezeigt. Darüber hinaus werden in Kapitel 2.7 aktuelle Probleme und Herausforderungen von VR-Systemen behandelt.

Kapitel 3 widmet sich der Light Field Technologie im Allgemeinen. Neben den theoretischen Grundlagen der Plenoptik und dem 4D Light Field in Kapitel 3.2, werden in Kapitel 3.3 verschiedene Aufnahmesysteme vorgestellt. Anschließend werden in Kapitel 3.4 wichtige Light Field Anwendungen behandelt.

In Kapitel 4 werden beide Themen zusammengeführt und verschiedene Einsatzgebiete von Light Fields für Virtual Reality aufgezeigt. Es werden verschiedene Unternehmen behandelt, die aktuell an Light Field Technologie für VR arbeiten und den aktuellen Stand der Forschung widerspiegeln. Neben Systemen zur effektiven Aufnahme und dem Rendering von 360° Light Fields, werden in Kapitel 4.4 außerdem Light Field Displays und deren Vor- und Nachteile in Bezug auf Virtual Reality analysiert.

2 Virtual Reality

2.1 Was ist eine virtuelle Realität? Eine Begriffsdefinition

Seit der Veröffentlichung der ersten Entwicklerversion der Oculus Rift 2013 erlebt das Thema Virtual Reality (VR) nun schon seit einigen Jahren einen regelrechten Hype. Mit der Veröffentlichung der ersten Consumer-Versionen der Oculus Rift, der HTC Vive sowie verschiedenen VR-Systemen für Smartphones scheint das Jahr 2016 ganz im Zeichen von Virtual Reality zu stehen. Ein komplett neuer Industriezweig arbeitet an VR-Videospielen und mit immersiven Filminhalten. Facebook und Youtube bieten mittlerweile 360° Video und VR Inhalte an. Doch was genau verbirgt sich hinter dem Begriff Virtual Reality?

2.1.1 Die perfekte virtuelle Realität

Die Utopie einer virtuellen Realität ist bedeutend älter; schon immer haben sich Menschen in andere, fantasievolle Welten geflüchtet, sei es durch die Kunst, das Lesen von Büchern oder mit Hilfe von Filmen oder Videospielen. Die Wachowski Brüder beschreiben in ihrem Science-Fiction-Film „Matrix“ von 1999 eine simulierte Welt, die für ihre Benutzer so täuschend echt ist, dass sie keinen Unterschied mehr zwischen Realität und virtueller Umgebung bemerken.

Dörner et al. (2013, S. 2) bezeichnen diesen Zustand als **perfekte virtuelle Realität**. Die Realität beruht auf der Wahrnehmung von Reizen durch verschiedene Sinnesorgane. So wird beispielsweise Licht von Objekten in der realen Welt reflektiert, gelangt in unser Auge und wird dort von verschiedenen Sinneszellen als Reiz wahrgenommen, über die Sehnerven an das Gehirn weitergeleitet und dort im Sehzentrum verarbeitet. Diese verschiedenen Stationen, von den Augen bis zum Gehirn, bilden das **visuelle System** des Menschen (Dörner et al. 2013, S. 2–3). Weitere Sinne des Menschen sind außerdem das Hören, das Riechen, das Schmecken, das Erfühlen und Ertasten, der Gleichgewichtssinn, die Körperempfindung, das Temperaturgefühl und die Schmerzempfindung. 70% aller Sinneszellen des Menschen befassen sich allerdings mit dem Sehen und machen somit das visuelle System zum wichtigsten Sinnesorgan der menschlichen Wahrnehmung (Dörner et al. 2013, S. 4).

Die Wahrnehmung der realen Welt ist nicht objektiv, sondern von Mensch zu Mensch unterschiedlich. Ca. 9% der Männer und 1% aller Frauen gelten als farbfehlsichtig und können bestimmte Farben, beispielsweise bei einer Rot-Grün-Schwäche nicht oder nur schlecht

unterscheiden. Gleichzeitig ist das visuelle System in der Lage, seine Arbeitsweise zu ändern und sich an neue Sehgewohnheiten anzupassen. Trägt man zum Beispiel eine sogenannte Umkehrbrille, die einfallende Lichtreize vertikal und horizontal spiegelt, nimmt man anfangs seine Umwelt verkehrt herum wahr. Nach einer Eingewöhnungszeit von mehreren Tagen adaptiert sich das visuelle System an die neuen Lichtreize und man kann wieder aufrecht sehen. Nimmt man die Brille ab, gewöhnt sich das visuelle System wieder an den Normalzustand (Dörner et al. 2013, S. 2–3).

Außerdem lässt sich die menschliche Wahrnehmung durch künstliche Reize manipulieren. Bilder von realen Objekten reflektieren Licht auf eine ähnliche Weise wie das Objekt selbst und lösen daher auch ähnliche Reize im visuellen System aus (Dörner et al. 2013, S. 4). Zwar kann ein Mensch in den meisten Fällen leicht zwischen einer Fotografie und einem echten Objekt unterscheiden, allerdings fällt diese Unterscheidung zunehmend schwerer, je glaubwürdiger und realistischer die künstlichen Reize erzeugt werden.

In einer perfekten virtuellen Realität muss also die Wahrnehmung des Menschen perfekt manipuliert werden. Damit ein Mensch zwischen Realität und Simulation nicht mehr unterscheiden kann, müssten nicht nur die visuellen, sondern alle Sinneseindrücke glaubhaft simuliert und direkt in die jeweilige Gehirnregion eingespielt werden (Dörner et al. 2013, S. 4).

Die perfekte virtuelle Realität ist nach dem heutigen Stand der Technik folglich eine Utopie und es ist fraglich, ob sie jemals erreicht wird, geschweige denn, ob es ethisch vertretbar ist eine Welt im Sinne der Matrix zu erschaffen. Gleichzeitig ist durch die Annäherung an eine perfekte Virtualität natürlich eine Verbesserung der heutigen VR-Erfahrungen möglich.

2.1.2 Reizerzeugung/Simulation

Um eine virtuelle Realität zu erschaffen, muss also eine Vielzahl an verschiedenen Reizen erzeugt werden. Visuelle Reize können beispielsweise durch computergenerierte Bilder erzeugt werden. Gleichzeitig will der Benutzer eine VR-Umgebung nicht nur betrachten, sondern sie auch erkunden und mit ihr interagieren. Kann sich ein Nutzer in der virtuellen Realität bewegen, erfordert dies die Erzeugung neuer Reize abhängig von seiner Position im Raum. Will er mit Objekten interagieren und Gegenstände bewegen oder aufheben, erfordert dies ein physikalisches Simulationsmodell der virtuellen Welt (Dörner et al. 2013, S. 6–7).

Die Erzeugung dieser Reize kann sehr komplex sein und stellt hohe Anforderungen an ein Computersystem. Die glaubhafte Simulation eines Menschen in einer VR-Umgebung benötigt ein fotorealistisches Rendering von feinen Details wie Haut, Kleidung oder Haaren, sowie die Simulation von menschlichem Verhalten und Sprache. All diese Reize müssen in Echtzeit berechnet und dargestellt werden. In VR-Anwendungen benötigt man bis zu 60 Bilder pro Sekunde, um Bewegungen als kontinuierlich wahrzunehmen. Haptische Reize müssen sogar bis zu 1000-mal in der Sekunde abgetastet werden (Dörner et al. 2013, S. 6–7).

2.1.3 Suspension of Disbelief

Die realistische Simulation aller Sinneseindrücke ist daher eine hochkomplexe Aufgabe, die eine hohe Rechenleistung sowie eine Vielzahl an Eingabegeräten und Sensoren benötigt. Kann mit heutiger Technik überhaupt eine überzeugende Virtual Reality Erfahrung erzeugt werden?

Laut Dörner et al. haben Menschen *„die Eigenschaft in bestimmten Situationen den augenscheinlichen Widerspruch einer virtuellen oder fiktiven Welt zur Realität auszublenden und dies auch zu wollen.“* (Dörner et al. 2013, S. 8)

Zuschauer von Animationsfilmen und Cartoons hinterfragen zum Beispiel nicht, warum Tiere im Film sprechen können und menschliche Kleidung tragen. Experimente zeigen, dass Nutzer, die in einer virtuellen Umgebung in eine gefährliche Situation gebracht werden, mit Angstschweiß und erhöhtem Blutdruck reagieren, obwohl sie genau wissen, dass die gezeigte Umgebung nicht der Realität entspricht. Diesen Vorgang des willentlichen Ausblendens der Realität bezeichnet man als **Suspension of Disbelief** (Dörner et al. 2013, S. 7–8).

Für eine glaubhafte Simulation einer VR-Umgebung muss also keine perfekte virtuelle Realität geschaffen werden. Die menschliche Wahrnehmung gibt sich schon mit weitaus weniger zufrieden. Die perfekte VR ist also nur ein Extremfall einer virtuellen Umgebung und es ist schwer eine einheitliche Definition zu finden. Eine Annäherung kann über typische Eigenschaften und Merkmale von Virtual Reality gelingen.

2.2 Merkmale einer virtuellen Realität

2.2.1 Präsenz

Das Merkmal der **Präsenz** beschreibt das Gefühl, sich wirklich innerhalb der virtuellen Welt zu befinden und zeichnet sich vor allem durch die Reaktionen des Nutzers aus. Verhält er sich in der VR ähnlich wie in der realen Welt, ist es gelungen ein Gefühl der Präsenz herzustellen (Slater und Wilbur 1997, S. 606–607). Laut Dörner et al. (2013, S. 18–19) basiert das Gefühl der Präsenz auf drei Teilaspekten:

Die **Ortsillusion** beschreibt das Gefühl, sich wirklich an dem simulierten Ort zu befinden. Hierbei ist es wichtig, die Erwartungen des Nutzers zu erfüllen. Bewegt der Nutzer seinen Kopf, müssen in Echtzeit neue Blickwinkel generiert werden. Ist dies nicht der Fall, kommt es zu einem „Break of Presence“ und die Präsenz ist zerstört.

Plausibilität bezieht sich weniger auf die Darstellung, als vielmehr auf die Inhalte einer virtuellen Realität. Verhalten sich Menschen und Objekte wie wir es gewohnt sind? Ist die Welt in sich glaubwürdig? Plausibilität beschreibt demnach die Sinnhaftigkeit der simulierten Welt.

Involviertheit beschreibt den Grad der Aufmerksamkeit, die wir unserer Umgebung schenken. Ist der Inhalt der simulierten Welt spannend gestaltet, neigen wir eher dazu uns länger mit ihr zu beschäftigen. Wenn uns die Umgebung der VR nicht interessiert, verringert dies die Präsenz.

2.2.2 Immersion

Immersion ist ein weit verbreiteter Begriff zur Beschreibung von virtuellen Anwendungen und Umgebungen und beschreibt sehr allgemein die Qualität einer VR Experience.

Slater und Wilbur (1997, S. 605) beschreiben den Begriff Immersion als technologisches Merkmal durch vier Aspekte:

1. Der Grad, in dem eine Anwendung den Nutzer von der realen Welt isoliert
2. Die Anzahl an Sinnesreizen, die von einer Anwendung angesprochen werden
3. Die Größe des Blickfelds (**Field of View**) einer Anwendung
4. Die Lebendigkeit der Darstellung in Bezug auf Bildqualität, Auflösung und Farbtiefe

Die meisten modernen Head-Mounted Displays sind immersive Displays, da sie den Benutzer vollkommen umgeben und von der Außenwelt isolieren. Je besser die technischen

Eigenschaften des Displays, wie z.B. Field of View, Auflösung oder Farbtiefe sind, desto größer ist die Immersion (Dörner et al. 2013, S. 14).

2.2.3 Technologieorientierte Merkmale

Ein wichtiges Merkmal für Virtual Reality Anwendungen sind sicherlich typische Ein- und Ausgabegeräte wie Head-Mounted Displays, Tracking Sensoren und Datenhandschuhe oder Controller. In einem typischen VR-System werden vom Nutzer zuerst Daten wie Position und Lage im Raum über ein Eingabegerät an das System übermittelt. Anhand dieser Daten erfolgt nun die Simulation und Reizerzeugung der virtuellen Welt, die dann wiederum über Ausgabegeräte an den Nutzer übermittelt werden. (Dörner et al. 2013, S. 12–13).

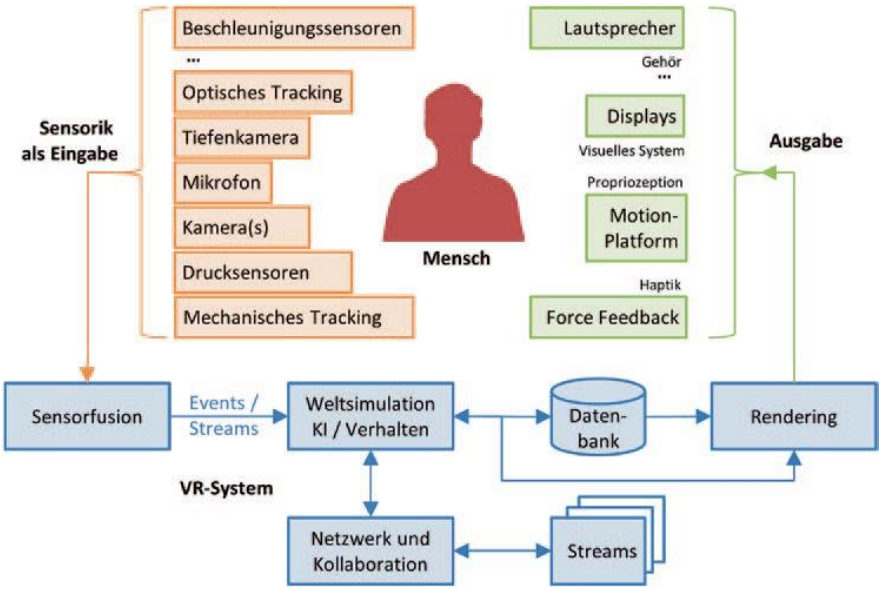

Abbildung 1: Schematische Darstellung eines VR-Systems (Dörner et al. 2013, S. 24)

Bei frühen VR-Systemen waren diese Ein- und Ausgabegeräte meist unpraktische, mechanische Gegenstände wie Datenhelme, Anzüge und Datenhandschuhe. Heutige VR-Systeme bestehen meist aus einem Head-Mounted Display, Trackingsensoren, um die Position und Lage des Benutzers im Raum zu bestimmen sowie zusätzlichen Controllern, die die Navigation des VR-Systems erleichtern. Je nach Anwendung kann das System durch Eingabegeräte wie Eyetracking, Handtracking oder Laufbänder und Gewichtssensoren ergänzt werden (Dörner et al. 2013, S. 22).

2.3 Wahrnehmung im Raum

Um die Anordnung von Objekten im Raum zu bestimmen, nutzt das visuelle System sogenannte **Tiefenhinweise** (Depth Cues). Dabei unterscheidet man zwischen monokularen Tiefenhinweisen, also Hinweisen, die aus der Information eines Auges entstehen, binokularen Tiefenhinweisen, für die das visuelle System die Information beider Augen benötigt sowie dynamischen Tiefenhinweisen, die sich aus der relativen Bewegung zwischen Betrachter und Objekt ergeben (Dörner et al. 2013, S. 37–38). Laut Huang et al. spielen diese Tiefenhinweise eine kritische Rolle für die Erkennung dreidimensionaler Formen sowie für Einschätzung von räumlicher Tiefe und Entfernungen Huang et al. (2015, S. 1).

2.3.1 Stereosehen

Die Wahrnehmung der Umwelt durch zwei Augen, das sogenannte Stereosehen, gehört zu den wichtigsten Instrumenten der menschlichen Wahrnehmung. Das visuelle System nutzt zwar beide Augen des Menschen, nimmt aber nicht zwei Einzelbilder, sondern ein einziges Bild mit räumlicher Tiefe wahr (Dörner et al. 2013, S. 36).

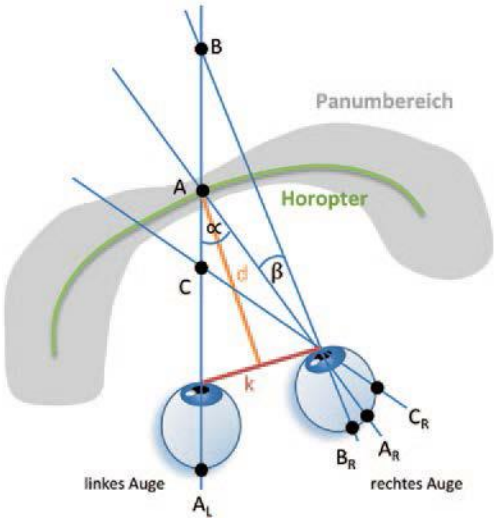

Abbildung 2: Das Stereosehen (Dörner et al. 2013, S. 37)

Abbildung 2 zeigt eine vereinfachte Grafik der Stereoskopie. Betrachten die beiden Augen Punkt A, müssen sie zueinander rotiert werden. Je näher ein Punkt am Betrachter liegt, desto größer ist die Rotation. Diesen Rotationswinkel α bezeichnet man als **Konvergenz**. Das Gehirn kann nun über eine einfache Triangulation und die Konvergenz auf die Entfernung von Punkt A schließen (Dörner et al. 2013, S. 37). Das visuelle System benötigt dazu die Information beider Augen, die Konvergenz ist demzufolge ein binokularer Tiefenhinweis.

Die Punkte A_L und A_R auf der Netzhaut der Augen würden bei einer Überlagerung an derselben Stelle liegen und bilden somit korrespondierende Punkte. Alle Punkte, die auf korrespondierende Punkte auf der Netzhaut abgebildet werden, bilden den sogenannten **Horopter** (Dörner et al. 2013, S. 37).

Betrachtet man nun Punkt B, erhält das rechte Auge zwar Lichtreize von diesem Punkt, in das linke Auge fallen aber noch immer Lichtreize von Punkt A. Auf der Netzhaut werden also keine korrespondierenden Punkte abgebildet, Punkt B liegt nicht auf dem Horopter. Die Abweichung von Punkt A_L zu Punkt B_R bezeichnet man als **Disparität** (Dörner et al. 2013, S. 37). Sie wird vom visuellen System genutzt, um die Entfernung von verschiedenen Punkten im Raum zu bestimmen. Im Gegensatz zur Konvergenz, kann die Lage der Objekte aber nicht absolut, sondern nur relativ zueinander bestimmt werden. Dies gilt sowohl für Objekte, die hinter dem Horopter liegen, als auch für Objekte davor (Punkt C). Wird die Disparität zwischen zwei Punkten zu groß, können die Bilder nicht mehr zusammengeführt werden und wir nehmen zwei Einzelbilder anstatt einem Bild wahr. Alle Punkte, die noch zu einem Bild fusioniert werden können, bilden den **Panumbereich** (Dörner et al. 2013, S. 38).

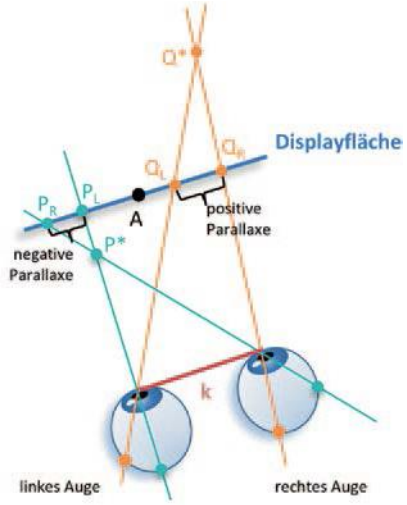

Abbildung 3: Manipulation des Stereosehens durch ein Stereodisplay (Dörner et al. 2013, S. 37)

Heutige Head-Mounted Displays sind in der Regel Stereodisplays und nutzen die Disparität, um einen dreidimensionalen Eindruck einer Szene zu bewirken, obwohl die Bildinhalte von einem 2D-Display stammen (Dörner et al. 2013, S. 39). Dem Benutzer werden dafür im Head-Mounted Display zwei Bilder mit einem leichten räumlichen Versatz, der sogenannten **Parallaxe,** präsentiert. Das visuelle System erklärt sich diesen Versatz damit, dass die Lichtreize von einem gemeinsamen Punkt ausgehen, der räumlich vor oder hinter der Displayfläche liegt. Fusionieren zwei Punkte P_L und P_R in Abbildung 3 zu einem Punkt P* vor der Displayebene, spricht man von einer „negativen Parallaxe“. Die Punkte Q_L und Q_R hingegen erzeugen einen Punkt hinter der Displayebene; man spricht von einer „positiven Parallaxe“ (Dörner et al. 2013, S. 39).

2.3.2 Weitere Tiefenhinweise

Disparität und Konvergenz sind binokulare Tiefenhinweise, es werden also beide Augen benötigt um die jeweilige Tiefeninformation zu erhalten (Dörner et al. 2013, S. 41).

Darüber hinaus gibt es noch eine ganze Reihe weiterer Tiefenhinweise, die vom visuellen System für die Raumwahrnehmung genutzt werden und deshalb auch eine wichtige Rolle für den Grad der Immersion in einer VR-Umgebung spielen.

Tiefenhinweis	Wirkungsbereich	Klassifizierung	Positionsbestimmung
Verdeckung	Kompletter Bereich	Monokular	Relativ
Disparität	Bis 10 m	Binokular	Relativ
Konvergenz	Bis 2 m	Binokular	Absolut
Akkomodation	Bis 2 m	Monokular	Absolut
Image Blur	Kompletter Bereich	Monokular	Relativ
Linearperspektive	Kompletter Bereich	Monokular	Absolut
Texturgradient	Kompletter Bereich	Monokular	Relativ
Relative Größe	Kompletter Bereich	Monokular	Absolut
Bekannte Größe	Kompletter Bereich	Monokular	Absolut
Höhe im Gesichtsfeld	Über 30 m	Monokular	Relativ
Atmosphärische Perspektive	Über 30 m	Monokular	Relativ
Shape from Shading	Kompletter Bereich	Monokular	Relativ
Schattenwurf	Kompletter Bereich	Monokular	Relativ
Bewegungsparallaxe	Über 20 m	Dynamisch	Relativ
Accretion	Kompletter Bereich	Dynamisch	Relativ

Abbildung 4: Liste von Tiefenhinweisen (Dörner et al. 2013, S. 41)

Wird ein Objekt A von einem anderen Objekt B verdeckt, kann davon ausgegangen werden, dass es räumlich hinter dem Objekt B liegt. Die **Verdeckung** von Objekten ist demnach ein weiterer Tiefenhinweis, der für VR-Umgebungen genutzt werden kann (Dörner et al. 2013, S. 39).

Die Muskelkraft, die benötigt wird um ein Objekt mit dem Auge zu fokussieren und scharf auf der Retina abzubilden, bezeichnet man als **Akkommodation**. Um Objekte nah am Betrachter zu fokussieren, muss mehr Kraft angewendet werden als bei Objekten die in der Ferne liegen. Die aufgewendete Muskelkraft wird vom visuellen System als Tiefenhinweis herangezogen (Dörner et al. 2013, S. 40).

Wird ein Objekt nah am Betrachter fokussiert, verschwimmen umliegende Bildhinhalte, die auf einer anderen Schärfeebene liegen. Diese Unschärfe bezeichnet man als **Retinal Blur**. Die Stärke der Unschärfe wird als Tiefenhinweis genutzt (Dörner et al. 2013, S. 40).

Um Tiefenhinweise aus Verdeckung, Akkommodation und Retinal Blur zu generieren, benötigt man nur ein Auge, es sind somit monokulare Tiefenhinweise (Dörner et al. 2013, S. 41).

Der wichtigste Vertreter der dynamischen Tiefenhinweise ist die **Bewegungsparallaxe**. Ist der Betrachter einer Szene in Bewegungen, bewegen sich Lichtreize von Gegenständen, die nah am Betrachter sind, schneller über die Netzhaut als die Lichtreize von Gegenständen in weiter Ferne. Durch die unterschiedliche Bewegungsgeschwindigkeit von Objekten kann auf deren relative Entfernung zueinander geschlossen werden (Dörner et al. 2013, S. 41).

Dabei werden laut Dörner et al. alle Tiefenhinweise für die Raumwahrnehmung *„in Form einer gewichteten Summe berücksichtigt"* (Dörner et al. 2013, S. 42).

Mark Bolas bezeichnet Bewegungsparallaxe, Akkommodation und Konvergenz sowie Disparität als die wichtigsten Aspekte der Raumwahrnehmung im Nahbereich und somit besonders kritisch für Virtual Reality (Bolas 2015, Min. 21). Während sich Tiefenhinweise wie Disparität und Konvergenz mit heutigen Head-Mounted Displays relativ leicht realisieren lassen, fehlt üblichen 360° Videos aufgrund ihrer statischen Natur jegliche Bewegungsparallaxe. Auch die Akkommodation der Augen ist nur begrenzt möglich, da sie immer auf die Entfernung des 2D-Displays und nicht auf die einzelnen Bildinhalte fokussieren können. Die künstliche Trennung von Konvergenz und Akkommodation führt außerdem zum sogenannten **Vergence-Fokus Konflikt**, der in Kapitel 2.7.2 noch genauer erläutert wird.

2.4 Entstehungsgeschichte der virtuellen Realität

Die Ursprünge heutiger Virtual-Reality Systeme liegen in den 60er Jahren des letzten Jahrhunderts. 1965 beschreibt Ivan Sutherland in seinem VR Meilenstein „The Ultimate Display“ erstmals ein futuristisches Display zum interaktiven Betrachten von dreidimensionalen Welten. Schon damals erkennt er die Notwendigkeit von Kopf- und Positionstracking sowie interaktiven Eingabegeräten (Sutherland 1965). 1968 präsentiert Sutherland das erste Head-Mounted Display. Über ein mechanisches Tracking konnten Kopfneigung und Position im Raum bestimmt und einfache 3D-Geometrien gezeigt werden. Das sogenannte „Damoklesschwert“ war so groß und schwer, dass es von Benutzern nicht selbst

getragen werden konnte und durch mechanische Verbindungen an der Raumdecke befestigt war (Sutherland 1968). 1985 arbeiteten Wissenschaftler am NASA Ames Research Center im Rahmen des VIEW Projekts (Virtual Environment Interface Workstations) an einer Multisensorischen Workstation für die Simulation virtueller Weltraumstationen (Dörner et al. 2013, S. 19). 1987 entwickeln Thomas Zimmermann und Jaron Lanier den „Dataglove“, einen Datenhandschuh zur Erfassung von Fingerbewegungen. In diesem Kontext benutzt Lanier den Begriff „Virtual Reality“ erstmals in einem wissenschaftlichen Kontext (Dörner et al. 2013, S. 20). Im Rahmen der SIGGRAPH 1992 präsentieren Cruz-Neira et al. (1992) das Cave Automatic Virtual Environment (CAVE). Dabei handelt es sich um VR-System auf Projektionsbasis, der Benutzer befindet sich in einer „Höhle“ aus Projektionsflächen, die um ihn herum eine virtuelle Welt erzeugen.

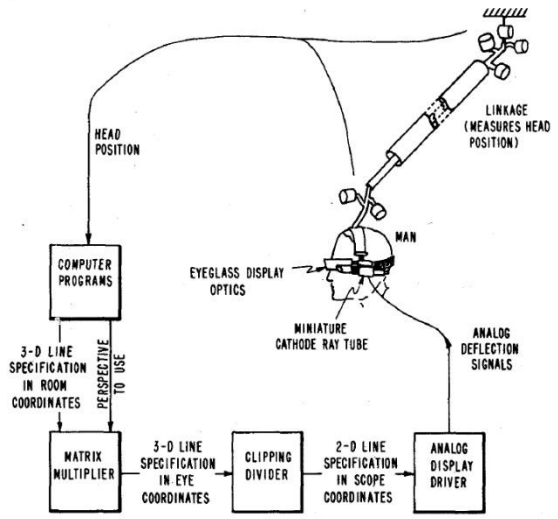

Abbildung 5: Ivan Sutherlands Head-Mounted Display (Sutherland 1968)

Mitte der 90er erscheinen erste VR-Entertainment-Systeme auf dem Massenmarkt. Das Forte VFX 1 ist ein Head-Mounted Display mit Headtracking, eingebauten Kopfhörern und Eingabegeräten und unterstützte damals verschiedene Videospiele am PC. Allerdings war es mit damals 1800 DM für den breiten Markt noch zu teuer und bot lediglich eine Auflösung von 384x224 Pixeln (Hertel 2015).

Auch der Virtual Boy von Nintendo versuchte sich als Entertainment System im Videospielmarkt zu behaupten. Der Headset war nicht portabel, musste fest im Raum installiert werden und scheiterte schließlich ebenfalls an schlechter Auflösung und unzureichender Bildwiederholrate (Hertel 2015). Die Rechen- und Grafikleistung, sowie die Displaytechnologie der 90er Jahre war noch nicht ausgereift genug und die ersten Virtual Reality Systeme ein großer finanzieller Misserfolg (Hertel 2015).

Abbildung 6: Virtual Boy von Nintendo

Erst die großen technischen Fortschritte der letzten Jahre, nicht zuletzt durch die Gaming Industrie, im Bereich Grafikleistung, CPUs, Displaytechnik und Tracking haben den neuen Hype um Virtual Reality überhaupt ermöglicht (Dörner et al. 2016, S. 37). Im Jahr 2012 gründet Virtual Reality Fan Luckey Palmer die Firma Oculus und entwirft zusammen mit John Carmack den ersten Prototypen der Oculus Rift. Innerhalb weniger Stunden sammeln sie über 250.000 Dollar auf der Crowdfunding Plattform Kickstarter und motivieren somit viele junge Start-Ups und große Unternehmen in Virtual Reality zu investieren. Mittlerweile wurde Oculus für über 2 Milliarden Dollar von Facebook aufgekauft (Hertel 2015).

Abbildung 7: Development Kit 1 der Oculus Rift

2014 erscheint mit der Google Cardboard ein low-budget Head-Mounted Display für Smartphones, das Virtual Reality zu einem günstigen Preis erstmals für die breite Öffentlichkeit interessant macht. Mittlerweile gibt es verschiedene VR-Systeme, die Smartphones als Display benutzen. 2016 erscheinen die ersten kommerziellen Versionen der Oculus Rift, der HTC Vive und im späteren Verlauf wohl auch noch die VR-Brille von Playstation und gründen damit die erste Generation der neuen VR-Systeme im Consumer-Bereich (La Orta 2016, Chapter 1).

2.5 Aktuelle VR-Systeme

Aktuelle VR-Systeme lassen sich grob in zwei Kategorien einordnen: Zum einen gibt es mobile VR-Systeme**,** die moderne Smartphones sowohl als Recheneinheit als auch als Display nutzen. Die Sensoren des Smartphones werden zum Tracking von Rotation und Lage im Raum benutzt (Dörner et al. 2016, S. 34). Zum anderen gibt es kabelgebundene VR-Systeme, die einen PC als Rechner und externe Trackingsysteme zur Positionsbestimmung im Raum nutzen. Kabelgebundene Systeme können üblicherweise auf eine höhere Grafik- und Rechenleistung sowie eine größere Bandbreite zurückgreifen als mobile VR-Systeme (Nichols et al. 2015, S. 19). Neben dem **Headtracking** verfügen kabelgebundene Systeme wie Oculus Rift und HTC Vive, auch über die Möglichkeit des **Positionstrackings**. Im Gegensatz zu den mobilen

Systemen, wird also nicht nur die Orientierung, sondern auch die Lage des Kopfes im Raum gemessen.

Wichtige technische Qualitätsmerkmale moderner Head-Mounted Displays sind das vom System abgedeckte Blickfeld, die Bildwiederholrate (Hz), die Auflösung und Farbtiefe des Displays sowie die Qualität des Kopf- und Positionstracking und der vom Tracking berücksichtigte Raum.

Das visuelle System des Menschen hat ein vertikales Sichtfeld von ca. 130° und ein horizontales Sichtfeld von ungefähr 190°, von dem ca. 120° für das Stereosehen genutzt werden. Die kritische Flickerfrequenz des Menschen, die benötigt wird, um Bilder flüssig wahrzunehmen, liegt bei ungefähr 60 Hz (Huang et al. 2015, S. 3). Moderne Head-Mounted Displays sollten dementsprechend einen möglichst großen Raum des menschlichen Sichtfelds abdecken und eine Bildwiederholrate von mindestens 60 Bildern/Sekunde gewährleisten.

2.5.1 Samsung Gear VR

Das Samsung Gear VR Headset ist eine Kooperation von Samsung und Oculus und nutzt im weitesten Sinne die Technik der Oculus Rift. In das Headset wird ein Samsung Smartphone eingesetzt, das gleichzeitig als Rechner und als Display dient. Es benutzt die Sensoren des Smartphones zur Lagebestimmung des Kopfes und gibt blickrichtungsabhängige Bilder an den Nutzer weiter. Inhalte können über eine App direkt aus dem Oculus Store geladen werden (La Orta 2016, Chapter 2: Samsung Gear VR).

Abbildung 8: Samsung Gear VR

Die finale Consumer-Version der Gear VR kostet ca. 100 Euro und ist somit eine der günstigsten VR-Systeme auf dem Markt. Allerdings funktioniert sie aktuell nur mit den Samsung Geräten der Serie S6 und dem Note 5 (Springer 2015).

Die Auflösung des Headsets ist abhängig vom verwendeten Smartphone und erreicht mit den aktuellen Geräten 1280 x 1440 Pixel pro Auge bei einer Bildwiederholfrequenz von 60 Bildern/ Sekunde. Das Sichtfeld Beträgt ca. 96° und ist damit etwas kleiner als bei der HTC Vive oder Oculus Rift (VR Brillen Vergleich 2016).

2.5.2 Google Cardboard

Das Google Cardboard wurde 2014 als low-budget VR-Brille entwickelt und besteht aus einem Kit mit Pappkarton, Linsen und einer Bauanleitung. Das Cardboard wurde als kostengünstige Alternative zu den teuren VR-Systemen entwickelt und soll so das Interesse an Virtual Reality Inhalten steigern. Google ist jedoch nicht der offizielle Hersteller und verschiedene Unternehmen bieten mittlerweile eigene Versionen des Google Cardboards an (La Orta 2016, Chapter 1).

Abbildung 9: Google Cardboard

Eingesetzt werden kann jedes Smartphone, das genug Rechen- und Grafikleistung hat, um VR-Inhalte darzustellen. Sie wird hauptsächlich dazu genutzt, um 360° Videos und Fotos zu betrachten. Technische Daten wie Field of View, Bildwiederholrate und Auflösung sind abhängig vom eingesetzten Gerät (VR Brillen Vergleich 2016).

2.5.3 Oculus Rift

Die Oculus Rift gilt als Vorreiter der modernen Head-Mounted Displays und ist verantwortlich für den aktuellen Hype um Virtual Reality. Nach den Development Kits DK1 und DK2, die als Prototypen für VR-Enthusiasten und Entwickler verkauft wurden, gibt es seit dem ersten Quartal 2016 die erste Consumer-Version der Oculus Rift für 699 Euro zu kaufen. Im Umfang der Rift ist ein eingebautes Mikrofon, eingebaute Kopfhörer sowie ein Xbox 1 Controller und eine Fernbedienung enthalten (La Orta 2016, Chapter 1: Oculus Rift).

Die Oculus Rift hat ein diagonales Sichtfeld von ca. 100° und eine Auflösung von 1200 x 1080 Pixel pro Auge bei einer Bildwiederholrate von bis zu 90 Bildern/Sekunde. Im Vergleich zur Oculus DK2 bietet die Rift ein verbessertes Tracking über einen größeren Bereich mit einer Abtastrate von 90 Hz (VR Brillen Vergleich 2016).

Abbildung 10: Consumer-Version der Oculus Rift

Noch 2016 soll mit dem Oculus Touch System außerdem ein Controller veröffentlicht werden, der neben klassischen Eingabemethoden auch Hand- und Fingergesten erkennen kann und durch 3D-Tracking eine präzise und natürliche Steuerung im virtuellen Raum ermöglichen soll. Im Oculus Touch System enthalten ist außerdem eine weitere Tracking Kamera, die den Trackingbereich vergrößert und somit auch Roomscale-VR, also einen ähnlich großen Bereich wie bei der HTC Vive, ermöglichen soll (Oculus Rift / Facebook-VR 2015). Die Rift läuft wie die Samsung Gear VR über das Programm Oculus Home. Der Oculus Store bietet derzeit die größte Auswahl an Virtual Reality Inhalten und Videospielen an (La Orta 2016, Chapter 1: Oculus Rift).

2.5.4 HTC Vive

2015 präsentierte der Hardwarehersteller HTC in Kooperation mit Videospielentwickler Valve die HTC Vive. Die technischen Daten übertrafen die der Oculus DK2 und die Vive stellte damit erstmals eine wirkliche Konkurrenz zum VR-Flaggschiff Oculus Rift dar (Hertel 2015). Die finale Consumer-Version der HTC Vive wurde im April für 899 Euro veröffentlicht und ist das teuerste VR-System der aktuellen Generation. Das Display besteht aus zwei OLED-Displays mit einer Auflösung von 1080 x 1200 Pixel pro Auge bei einer Bildwiederholrate von 90 Bildern/Sekunde. Das Sichtfeld beträgt 110° und ist damit etwas größer als bei der Oculus Rift (HTC Vive mit Steam VR 2016).

Der große technische Vorsprung gegenüber der Oculus Rift liegt aber in einem deutlich größeren Bereich des Positionstracking. Die sogenannte „Light House“ Technologie der HTC Vive besteht aus zwei Trackingkameras, die im Raum installiert werden und insgesamt eine Fläche von ca. 4,5 m^2 abdecken. Innerhalb dieser Fläche können sich Benutzer frei bewegen (Nichols et al. 2015, S. 4). Im Lieferumfang der Vive sind außerdem zwei Controller enthalten, die ebenfalls durch Sensoren getrackt werden und somit eine interaktive Steuerung im virtuellen Raum erlauben (La Orta 2016, Chapter 1: HTC Vive). Ein System also, das Oculus mit ihrem Touch System gerade erst aufholt.

Abbildung 11: HTC Vive

Nicht zuletzt ist der Kooperationspartner Valve einer der größten Wegbereiter der Videospielindustrie, vor allem bekannt durch seine Videospielserie Half Life. Durch die Einbindung in Valves Distributionsnetzwerk Steam steht der HTC Vive also auch die größte Online Plattform für Videospiele zur Verfügung (La Orta 2016, Chapter 1: HTC Vive).

2.6 Virtual Reality Inhalte

Neben unterschiedlichen VR-Systemen gibt es auch eine Vielzahl an unterschiedlichen Inhalten für Virtual Reality. Der Oculus Store ist aktuell die größte Plattform für Virtual Reality Inhalte und bietet neben immersiven Videospielen und Applikationen zum Betrachten von 360° Videos und Panoramafotos die Möglichkeit, herkömmliche Spielfilme in einem virtuellen Kino anzusehen.

Nichols et al. unterscheidet die Inhalte für VR anhand der Bewegungserkennung von Rotation und Translation. In 360° Videos und Panoramafotos schaut sich der Betrachter nur um und das VR-System benötigt lediglich die Rotationsinformation des Kopfes Nichols et al. (2015, S. 4). Diese Inhalte sind demnach geeignet für mobile VR-Systeme wie Google Cardboard und Samsung Gear VR, da die benötigten Sensoren bereits im Smartphone verbaut sind.

Kabelgebundene Systeme hingegen unterstützen sowohl Rotation als auch Translation in der virtuellen Welt, weil neben dem Headtracking auch die Position im Raum gemessen wird (Nichols et al. 2015, S. 4). Bewegt man sich aber nun in der virtuellen Welt, muss das VR-System nicht nur einen neuen Blickwinkel eines Panoramabildes, sondern eine komplett neue

Perspektive mit Parallaxe und Verdeckungen generieren. Das VR-System benötigt dazu entweder Bilder, die in Echtzeit aus einer 3D-Szene berechnet werden oder ein sogenanntes Deep-Imaging-System, das Informationen über verschiedene Blickwinkel im Raum enthält (Nichols et al. 2015, S. 4).

Infolgedessen unterscheidet man Virtual Reality Inhalte also in Real-Time Content und Pre-Rendered Content.

2.6.1 Videospiele: Real-Time Content

Real-Time Content bezeichnet Inhalte, die das VR-System während der Anwendung in Echtzeit berechnet. Wie bereits in Kapitel 2.5. beschrieben, sind die Anforderungen der Systeme dabei relativ hoch. Es müssen für jedes Auge 60 bis 90 Bilder/Sekunde in einer Auflösung nahe des HD-Standards (1920 x 1080 Pixel) berechnet werden.

Um eine solche Rechenleistung zu gewährleisten, greifen Entwickler auf Render-Engines aus der Videospielbranche zurück. Engines wie die Unreal Engine 4 oder Unity 3D sind weit verbreitet und bieten mittlerweile eigene Software-Implementierungen und Tutorials zum Erstellen von VR-Inhalten an (La Orta 2016, Chapter 4: 360-degree animation). Durch Real-Time Content sich als Videospielinhalte mit Bewegung in Interaktion innerhalb der virtuellen Welt realisieren.

2.6.2 Live-Action: Pre-Rendered Content

Live-Action Content sind Inhalte aus der realen Welt, gefilmt als 360° Panorama. Innerhalb dieser Panoramabilder kann sich der Benutzer einer VR-Anwendung nun umsehen und bekommt in Abhängigkeit seiner Kopfrotation neue Bildinhalte gezeigt.

360° Videos lassen sich beispielsweise durch ein 360° Rig aus einzelnen Videokameras realisieren. Diese müssen dazu aber aufwendig kalibriert und synchronisiert werden. Die aufgenommenen Bildinhalte müssen später außerdem zusammengeführt und an den Kanten zu einem Panoramabild „vernäht“ (**Stitching**) werden. Einfacher ist daher der Einsatz einer 360° Kamera, wie zum Beispiel der Rico Theta. Diese Kamera filmt ihre Umgebung über zwei gegenüberliegende Fisheye-Objektive. Jedes der Objektive deckt einen Bereich von über 180°

ab; Bildinhalte überschneiden sich und lassen sich daher vergleichsweise einfach zu einem Panoramabild zusammenführen (Debevec 2015, Min. 6–8).

Eine weitere Möglichkeit 360° Panoramabilder zu erstellen, ist das Rendering von synthetischen Bildern durch 3D-Software (Nichols et al. 2015). Branchenübliche 3D-Tools wie Cinema 4D oder Maya erlauben hierfür das Rendering von sphärische Panoramabildern. Aufgrund der mathematisch perfekten Kameras in einer 3D-Umgebung entfällt eine aufwendige Kalibrierung und das Zusammenführen einzelner Bilder. Außerdem lassen sich in einer 3D-Umgebung verschiedene Kameras in beliebigem Abstand zueinander platzieren. Auf diese Weise können korrekte Stereobilder für eine VR-Anwendung erzeugt werden.

Da das benötigte Bildmaterial bereits im Vorfeld einer VR-Anwendung aufgezeichnet oder synthetisch erzeugt wird und keine Berechnung neuer Bilder in Echtzeit notwendig ist, bezeichnet man diese Inhalte auch als Pre-Rendered Material. Gleichzeitig enthalten 360° Panoramabilder nur die Bildinformation von einem Standpunkt in der Szene. Sie erlauben somit lediglich die Rotation, nicht aber die Translation des Kopfes in einer VR-Umgebung und unterstützen folglich kein Positionstracking oder die Darstellung von Bewegungsparallaxe (Nichols et al. 2015, S. 4).

2.6.3 Einsatzgebiete von Virtual Reality

360° Videos sind beliebte Virtual Reality Inhalte und dienen meist der Unterhaltung. Sie versetzten den Betrachter an Events, Konzerte oder Sportereignisse, Vergnügungsparks und Achterbahnen. Gleichzeitig ist das Potenzial von Virtual Reality aber größer und es gibt eine Vielzahl an möglichen Anwendungsgebieten.

Virtual Reality kann bei der Visualisierung in Architektur und Industrie von großer Bedeutung sein. Entwürfe können so dreidimensional und interaktiv präsentiert werden. Der Kunde kann sich in geplanten Räumen bewegen und ein besseres Verständnis über den Bauvorgang gewinnen. Prototypen und Modelle können kostengünstiger erstellt und Fehler frühzeitig ausgebessert werden (Dörner et al. 2016, S. 31). Das VR-System dient dabei als perfekte Schnittstelle zwischen Mensch und Maschine. Nutzer können Anwendungen im Idealfall intuitiv bedienen, genau wie sie es aus der echten Welt gewohnt sind (Dörner et al. 2016, S. 31). Diese natürliche Art der Bedienung eröffnet große Chancen im Bereich von Ausbildung und Lehre.

Zum Beispiel können medizinische Daten interaktiv visualisiert werden und Ärzte in ihrer Arbeit unterstützen. VR-Umgebungen können zur Therapie gegen Posttraumatische Belastungsstörungen genutzt werden. Hierfür werden spezielle Szenarios entwickelt, in denen Patienten langsam an Probleme herangeführt werden können, ohne sich einem realen Erlebnis aussetzen zu müssen (La Orta 2016, Chapter 1: Medicine). Gleichzeitig hat auch das Militär Virtual Reality für sich entdeckt und trainiert Soldaten vermehrt in immersiven Umgebungen. Durch Head-Mounted Displays, haptische Datenhandschuhe und VR-Waffen als Eingabegeräte soll eine möglichst realitätsnahe Simulation erzeugt und Soldaten für bestimmte Einsätze ausgebildet werden (La Orta 2016, Chapter 1: Education and Training). Der größte Markt für Virtual Reality aber ist der Entertainment-Bereich. Die aktuellen VR-Systeme sind stark auf die Computerspielbranche ausgerichtet, einen Massenmarkt also, der allein in Deutschland auf ein Marktvolumen von 1,82 Milliarden Euro geschätzt wird (Dörner et al. 2016, S. 30).

2.7 Ausblick und aktuelle Probleme

Neben diesen Einsatzgebieten ergeben sich durch Virtual Reality außerdem neue Möglichkeiten des filmischen Erzählens. Oculus hat zu diesem Zweck eine eigene Filmdivision gegründet, die ausschließlich an narrativen Inhalten arbeitet. Lucasfilm produzierte im Vorfeld zu „Star Wars: The Force Awakens“ eine Reihe von Trailer in Form von 360° Videos (La Orta 2016, Chapter 5: The Future of Virtual Reality).

Andrew Chochrane (2016) vergleicht den aktuellen Stand von Virtual Reality mit der Entstehung des Films Anfang des 20. Jahrhunderts. Seiner Meinung nach ist VR als Medium so neu, dass eventuelle Möglichkeiten und Anwendungen noch gar nicht entdeckt wurden. Das filmische Erzählen hat mehrere Jahrzehnte gebraucht, bis sich gewisse Regeln und Standards in Bezug auf Akquisition, Rhythmus, Schnitt und Narration entwickelt haben.

Heutige 360° Videos und Filme sind eher der Versuch, ein altes Medium neu zu verpacken. Chochrane hingegen sieht Virtual Reality als eine Mischung aus Film, Theater, Freizeitpark und Videospiel (Chochrane 2016, Min. 36–38). Der Begriff Storytelling ist nicht mehr angemessen. Betrachter suchen sich ihren Bildausschnitt selbst aus und wollen mit der Welt interagieren. Passender sei daher der Begriff Storyarchitektur oder Storyworld (Chochrane 2016, Min. 48) (Chochrane 2016, Min 48). Der Regisseur kann nur noch die Rahmenbedingungen festlegen und versuchen, den Blick auf wichtige Inhalte zu leiten

(Chochrane 2016, Min. 22). Die Story aber muss sich aus der Erfahrung ergeben. Der Benutzer schreibt die Geschichte mit und kann im besten Fall auch Einfluss darauf nehmen (Chochrane 2016, Min. 45–48) (Chochrane 2016, S. 45–48).

Andrew Chochrane (2016, Min. 13–15) unterscheidet Virtual Reality Inhalte außerdem in drei Kategorien:

1. 360° Video:
   Live-Action und 360° Videos sind relativ simple Inhalte und benötigen keine Narration. Vielmehr geht es um die „Experience", also um eine besondere Erfahrung in einer virtuellen Umgebung.
2. Immersive (1st Person) Video:
   Diese Inhalte sind zwar ebenso 360° Videos, allerdings mit dem Hauptaugenmerk auf Narration. Der Betrachter findet sich dabei im Mittelpunkt einer Geschichte, die um ihn herum aufgebaut ist.
3. True VR:
   Eine Virtual Reality Umgebung mit vollständiger Immersion, in der alle Sinnes-eindrücke angesprochen werden. Der Benutzer kann sich innerhalb der VR-Umgebung frei bewegen und mit der Welt interagieren.

Chochrane beschreibt damit dieselbe Utopie, die Dörner et al. als „perfekte VR" bezeichnen. True VR, also ein Zustand ähnlich der Matrix, in dem Betrachter nicht mehr zwischen Realität und Simulation unterscheiden können, ist immer noch eine Zukunftsvision und mit aktuellen Systemen nicht realisierbar.

2.7.1 360° Video: ist das Virtual Reality?

Wie nahe aber kommen heutige Inhalte einer perfekten, True VR Erfahrung?

High-End Geräte wie Oculus Rift und HTC Vive ermöglichen in Kombination mit Real-Time Content schon heute sehr immersive VR-Erfahrungen. Benutzer können sich bis zu einem gewissen Maß frei bewegen. Durch Positionstracking werden in Echtzeit Bildinformationen mit korrekter Bewegungsparallaxe und neuem Blickwinkel errechnet. Über 3D-Eingabegeräte kann entsprechend mit der virtuellen Welt interagiert und kommuniziert werden.

360° Videos (Live-Action VR) hingegen stehen vor einem fundamentalen Problem: Sie unterstützen kein Positionstracking und somit keine Bewegungsparallaxe (Debevec 2015, Min.

7–8). Sie erlauben lediglich die Rotation des Kopfes, nicht aber die Darstellung einer neuen Perspektive aus einem anderen Punkt im Raum. Bewegt man den Kopf, bewegt sich die gesamte Umgebung mit und die Welt scheint am Kopf des Betrachters fixiert zu sein. Dieser Effekt steht in einem starken Konflikt mit unseren natürlichen Sehgewohnheiten und verhindert eine angemessene Ortsillusion. So entstehen nicht nur Nebenwirkungen wie Cybersickness und Schwindelgefühle (Debevec 2015, Min. 8), sondern auch ein vermindertes Gefühl von Präsenz und somit eine geringere Immersion in der virtuellen Umgebung.

Darüber hinaus lassen sich stereoskopische Inhalte mit 360° Videoaufnahmen nur schwer realisieren. Während in einer 3D-Umgebung leicht zwei verschiedene Bilder mit einem gewissen Versatz ausgegeben werden können, benötigt man für die Stereoskopie in Live-Action Material zwei 360° Kameras in geringem Abstand zueinander. Das Bildmaterial jeder Kamera enthält nun fälschlicherweise die andere Kamera. Diese muss nun in der Postproduktion aufwendig entfernt werden. Laut Mark Bolas (2015, Min. 8) gehören binokulare Disparität und Bewegungsparallaxe aber zu den wichtigsten Tiefenhinweisen für die Wahrnehmung im Raum. Diese Tiefenhinweise werden durch heutige 360° Videos nicht oder nur sehr schwach unterstützt.

Zu guter Letzt sind 360° Videos Pre-Rendered Content und damit Material, das bereits im Vorfeld einer Anwendung aufgezeichnet oder berechnet wurde. Man kann während der Anwendung also keinen Einfluss auf das Material nehmen. 360° Videos sind in ihrer Natur nicht interaktiv.

Heutigen 360° Videos fehlen also eine ganze Reihe entscheidender Merkmale einer virtuellen Realität und man kann zu Recht behaupten, dass diese Inhalte keine Virtual Reality Anwendung im eigentlichen Sinne sind. Im Gegensatz zu Real-Time VR, gibt es für Live-Action VR noch keine wirklich immersiven Anwendungen.

2.7.2 Der Vergence-Fokus Konflikt

Head-Mounted Displays der aktuellen Generation unterstützen viele der möglichen Tiefenhinweise wie Disparität und Stereosehen, Parallaxe, Verdeckung und Konvergenz. Aufgrund ihrer Bauweise ergibt sich allerdings eine künstliche Trennung von Akkommodation und Konvergenz, der sogenannte Vergence-Fokus Konflikt.

Die Bildinhalte in einem Head-Mounted Display stammen in der Regel von einem 2D-Display, das über eine Vergrößerungslinse betrachtet wird. Um die Bildinhalte scharf auf der Retina des Auges abzubilden, müssen die Augen also auf die Entfernung des Displays fokussieren. Gleichzeitig wird durch die Disparität ein Stereobild mit dreidimensionalem Eindruck erzeugt. Bildinhalte erscheinen in räumlicher Tiefe und die Augen konvergieren auf Objekte vor- oder hinter dem Display. Will man auf die scheinbare Entfernung der Objekte fokussieren, erscheinen sie unscharf, da nun nicht mehr das Display fokussiert wird. Der Punkt für Akkommodation und Konvergenz liegt nicht auf derselben Tiefenebene. Da einzelne Bildinhalte nicht fokussiert werden können, erfolgt außerdem kein Retinal Blur. Objekte, egal in welcher Tiefe, werden mit der gleichen Schärfe abgebildet (Dörner et al. 2013, S. 52).

Diese künstliche Trennung von Konvergenz und Akkommodation steht in Konflikt mit unserer natürlichen Wahrnehmung und kann in VR-Umgebungen zu erheblichen Nebenwirkungen führen. Diese Nebenwirkungen können von Ermüdungserscheinungen und Augenschmerzen, bis zu sichtbaren Doppelbildern, Kopfweh und Schwindelgefühlen in einer VR-Umgebung führen (Huang et al. 2015, S. 1–2).

3 Light Field Technologie

Genau diese Probleme können durch Light Field Technologie für Virtual Reality adressiert werden. Ein Light Field erlaubt die Rekonstruktion aller möglichen Blickpunkte in einem definierten Volumen. Somit können neue, blickpunktabhängige Perspektiven einer Szene in Echtzeit gerendert werden. Der Benutzer kann sich innerhalb des von einem Light Fields abgedeckten Bereichs frei bewegen und neue Positionen einnehmen.

Darüber hinaus lassen sich durch Light Field Technologie Displays realisieren, die eine natürliche Akkommodation der Augen gewährleisten. Der Benutzer kann innerhalb des Displays frei auf unterschiedliche Bildinhalte fokussieren. Light Field Displays adressieren somit direkt den Vergence-Fokus Konflikt und können zu einer verbesserten Raumwahrnehmung in einer VR-Umgebung beitragen.

3.1 Light Field Rendering

Der Begriff Light Field Rendering geht zurück auf die Arbeit von Levoy und Hanrahan (1996) und bezeichnet eine von verschiedenen Techniken des **Image-Based Renderings** (IBR). Image-Based Rendering ist ein Oberbegriff verschiedener Rendertechniken, hauptsächlich aus den 90er Jahren, für die keine 3D-Geometrie, dafür aber eine Sammlung an Bildern einer Szene benötigt wird (Levoy 2006, S. 48). Light Field Rendering benötigt im Vergleich zu herkömmlichen Rendertechniken nur eine sehr geringe Rechenleistung. Außerdem ist die Renderzeit unabhängig von der Komplexität einer Szene (Levoy und Hanrahan 1996, S. 31). Levoy und Hanrahan beschreiben Light Field Rendering schon damals, als effiziente Methode um *„inward and outward looking views“* (Levoy und Hanrahan 1996, S. 31) zu erschaffen, wobei sich gerade Letzteres natürlich für VR Inhalte anbietet. Light Field Rendering ist also eine geeignete Methode, um neue Blickpunkte und Perspektiven in Echtzeit zu berechnen und somit besonders interessant im Hinblick auf Virtual Reality und Live-Action VR.

3.2 Theoretische Grundlagen der Plenoptik

3.2.1 Die plenoptische Funktion

Der Begriff Light Field geht zurück auf die Arbeit des russischen Forschers A. Gershun (1939). Er definiert das Lichtfeld als die Menge an Licht, die sich von jedem Punkt im Raum in jede mögliche Richtung ausbreitet. In ihrer Arbeit „The Plenoptic Function“ beschreiben Adelson und Bergen (1991) erste Versuche das Lichtfeld zu messen und mathematisch zu festzuhalten. Sie definieren eine Funktion, in der sich jeder Blickwinkel zu jeder Zeit und an jedem Ort rekonstruieren lässt.

> *"A true holographic movie would allow reconstruction of every possible view, at every moment, from every position, at every wavelength, within the bounds of the space-time wavelength region under consideration."* (Adelson und Bergen 1991, S. 5)

Die Plenoptische Funktion beschreibt also "die Intensität eines Lichtstrahls mit beliebiger Wellenlänge λ, zu jedem Zeitpunkt t in jedem möglichen Winkel Θ und Ø, der auf eine Kamera an jedem Ort x,y,z fällt" (Schreer 2005, S. 226) und somit das Licht Feld in 7 Dimensionen.

P = P(Θ, Ø, λ, t, x, y, z) (Adelson und Bergen 1991, S. 5)

Laut Schreer (2005, S. 229) kann die plenoptische Funktion auch als Abtasttheorem für optische Strahlen im dreidimensionalen Raum gesehen werden. Gleichzeitig ist die Funktion ein idealisiertes Konzept. Man bräuchte unendlich viele Kameras, um eine Szene von jedem Ort und zu jeder Zeit betrachten zu können (Schreer 2005, S. 227). Die nötige Datenmenge, um die gesamte plenoptische Funktion zu repräsentieren, wäre folglich entsprechend groß. McMillan und Bishop (1995) beschränken deshalb das Lichtfeld auf eine statische Szene, ohne Bewegung und mit konstanter Beleuchtung. Durch die Vernachlässigung der Zeit t und Wellenlänge λ kann die plenoptische Funktion von 7 auf 5 Dimensionen reduziert werden.

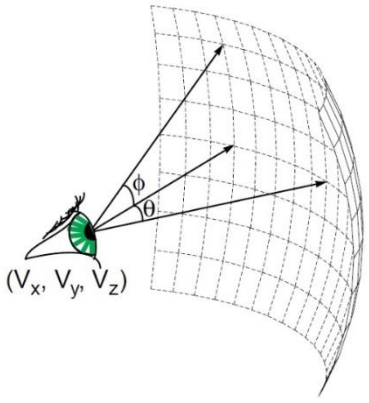

Abbildung 12: Schematische Darstellung der plenoptischen Funktion (McMillan und Bishop 1995)

P = P (Θ, Ø, x, y, z) (Schreer 2005, S. 228)

3.2.2 Das 4D Light Field

Levoy und Hanrahan (1996) greifen die Idee der plenoptischen Funktion auf und entwickeln sie weiter. Hierfür gehen sie von der Annahme aus, dass die Intensität eines Lichtstrahls konstant ist, solange er sich in freiem Raum ohne Abschattungen oder Verdeckungen bewegt. Somit kann die Intensität des Lichtstrahls vernachlässigt werden. Die plenoptische Funktion enthält redundante Information und kann von 5D auf 4D reduziert werden (Levoy 2006, S. 47).

"In free space, the light field is a 4D, not a 5D function." (Levoy und Hanrahan 1996, S. 31)

Außerdem beschreiben sie in ihrer Arbeit eine neue Parametrisierung des Lichtfelds. Die Orientierung eines Lichtstrahls ist eindeutig definiert durch seine Schnittpunkte an zwei virtuellen Ebenen (u, v) und (s, t). Diese Repräsentation bezeichnen sie als **Light Slab**.

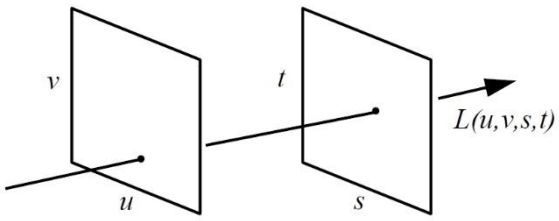

Abbildung 13: Parametrisierung des Licht Felds durch einen Light Slab (Levoy und Hanrahan 1996, S. 32)

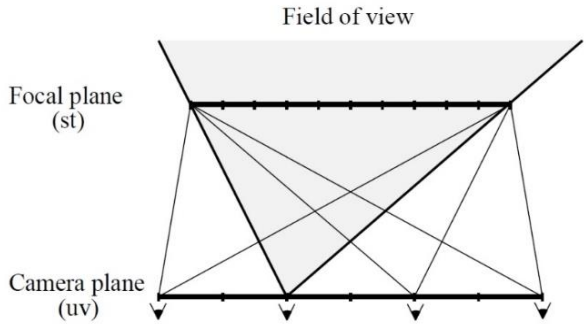

Abbildung 14: Abtastung des Licht Felds durch eine Fotokamera (Levoy und Hanrahan 1996, S. 34)

Ein Light Slab deckt allerdings nicht das gesamte 4D Light Field, sondern nur einen Teil davon ab. Strahlen die parallel zu (u, v) und (s, t) Ebene eintreffen, werden in dieser Parametrisierung nicht berücksichtigt. Für ein 360° Light Field einer Szene benötigt man folglich mehrere Light Slabs aus unterschiedlichen Winkeln der Szene (Levoy und Hanrahan 1996, S. 32–33).

Analog zur Fotografie beschreiben sie die Ebene (u, v) nun als Kameraebene und die Ebene (s, t) als Fokusebene. Das Licht Feld kann also mit einer Fotokamera gemessen werden (Levoy 2006, S. 47). Dazu muss eine Szene nun von vielen unterschiedlichen Standpunkten einer Ebene (u, v) aufgenommen werden. Das Light Field ist nun eine Sammlung perspektivisch leicht versetzter Bilder einer Szene auf der (s, t) Ebene, geschossen von unterschiedlichen Punkten der (u, v) Ebene (Levoy 2006, S. 47–48). Es ist daher eine zweidimensionale Sammlung von 2D-Bildern und ein 4D-Array an Pixeln (Levoy 2006, S. 48).

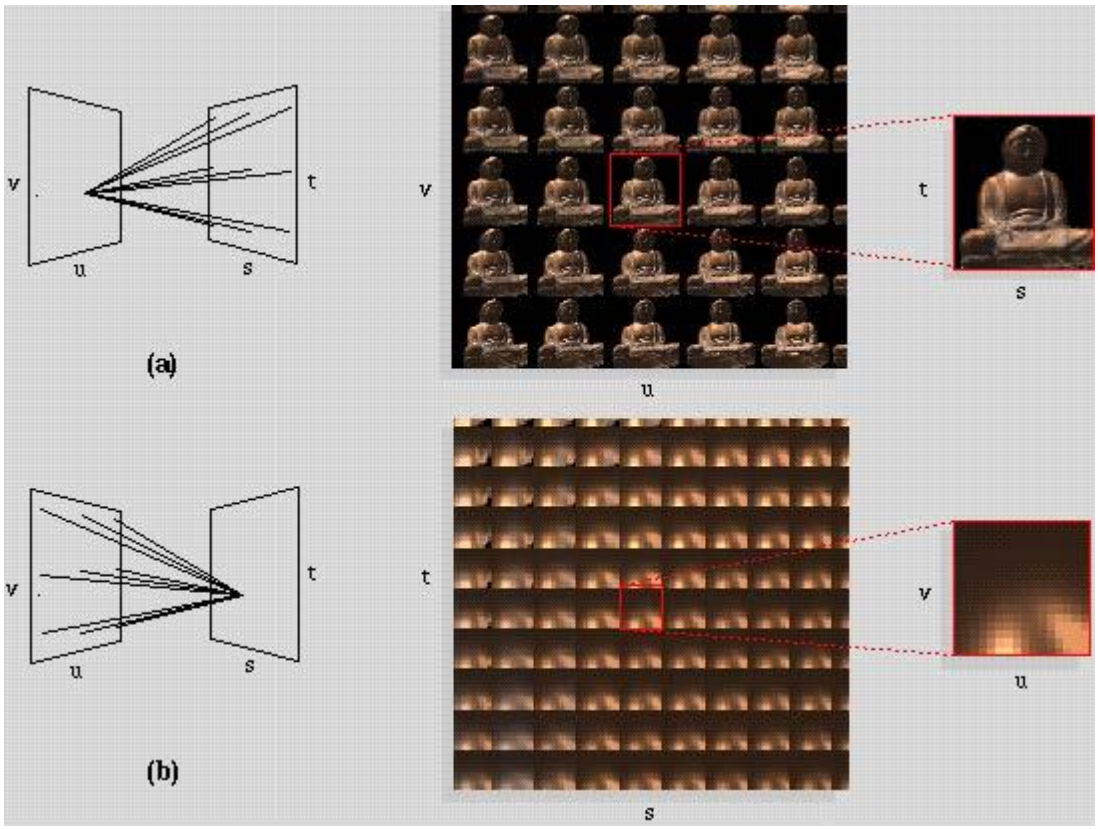

Abbildung 15: Verschiedene Darstellungen eines Light Fields. Darstellung (a) zeigt eine Sammlung von perspektivisch leicht unterschiedlichen Bildern auf der Ebene (u, v). Darstellung (b) zeigt ein 4D-Array an Pixeln der Ebene (s, t), aus denen sich die einzelnen Perspektiven der Ebene (u, v) zusammensetzen (Levoy und Hanrahan 1996, S. 34)

Ein 4D Light Field nach Levoy und Hanrahan ist also eine Sammlung von Bildern mit unterschiedlichen Perspektiven einer Szene und enthält daher Informationen über ihre räumliche Beschaffenheit. Sind diese Informationen dicht genug gesammelt, können nun komplett neue Perspektiven der Szene konstruiert werden, die nicht von der Kamera aufgezeichnet wurden (Levoy 2006, S. 48).

3.3 Light Field Aufnahmegeräte

Wie viele Bilddaten benötigt werden, um ein Light Field zu erzeugen, ist abhängig von der späteren Anwendung. Schon ein einzelnes Bild einer Szene kann als 2D-Slice in einem 4D Light Field gesehen werden. Dieses ist damit allerdings nur sehr schwach abgetastet. Benötigt man nur wenig Parallaxe in einer holographischen Anwendung, genügt ein Light Slab mit einer gewissen Anzahl an Bildern. Für eine 360° Anwendung benötigt man entsprechend mehrere Light Slabs. Sollen Betrachter die Szene auch noch aus einem sehr geringen Abstand betrachten können, benötigt man entsprechend hoch aufgelöste Bilder. Als Beispiel für ein sehr hoch

abgetastetes Light Fields nennt Levoy (2006, S. 48) eine Aufnahme von Michelangelos Skulptur „Nacht“ aus der Medici-Kapelle in Florenz. Im Rahmen eines Projekts der Stanford Universität entstand ein Light Field aus insgesamt 24.000 Bildern mit einer Auflösung von 1,3 Megapixel pro Bild. Dies ist aber ein extremes Beispiel und Light Fields können schon mit weitaus weniger Bildmaterial angemessene Ergebnisse liefern. Dabei gibt es verschiedene Möglichkeiten eine entsprechende Bildersammlung für ein 4D Light Field anzufertigen.

3.3.1 Fotokamera

Der Einsatz einer gewöhnlichen Fotokamera ist wohl die einfachste Möglichkeit ein Light Field zu erstellen. Solange es sich um eine statische Szene ohne Bewegung und mit konstanter Beleuchtung handelt, können die einzelnen Bilder nacheinander und mit nur einer Kamera geschossen werden (Levoy 2006, S. 52). Damit allerdings eine Zuordnung von (u, v) zur (s, t) Ebene getroffen werden kann, muss die Lage und Orientierung der Kamera für jedes Bild rekonstruierbar sein. Um das zu gewährleisten, entwickeln Levoy und Hanrahan 1996 ein eigenes „Spherical Gantry“, mit dem sich 360° Light Fields von Objekten erstellen lassen.

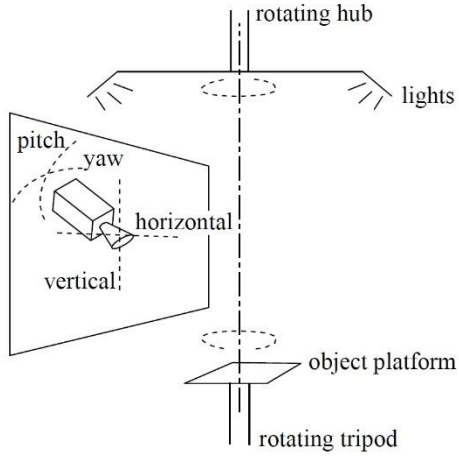

Abbildung 16: Schematische Darstellung des Spherical Gantry (Levoy und Hanrahan 1996, S. 36)

Das Gantry ist ein Motion Control System mit einem Pan-Tilt Motor. Orientierung und Lage der Kamera lassen sich genau kalibrieren. Die Kamera bewegt sich während der Aufnahme ausschließlich auf einer zweidimensionalen Ebene, welche die (u, v) Ebene repräsentiert. Somit wird ein Light Slab mit bis zu 64x64 Bildern erstellt (Levoy und Hanrahan 1996, S. 39). Das Objekt befindet sich auf einer drehbaren Plattform. Um das komplette Objekt abzudecken und ein 360° Light Field zu generieren, wird es nach jedem Light Slab um 90° gedreht. Solange sich die Kameraposition rekonstruieren lässt, können Light Fields auch aus der Hand geschossen werden. Dabei ist nicht gewährleistet, dass alle Fotos genau von einer Ebene (u, v) stammen. Die Kameraposition muss also durch Optical Flow Algorithmen oder ein Trackingsystem abgeschätzt werden (Levoy 2006, S. 52).

Davis et al. (2012) präsentieren in ihrer Arbeit “Unstructured Light Fields” ein System zur interaktiven Aufnahme von Light Fields mit einer Fotokamera und einem Laptop. Der Benutzer fokussiert dabei eine Szene mit einer Kamera. Um das Objekt wird nun eine virtuelle Kugel, die sogenannte “Bounding Sphere” gelegt. Der Benutzer nimmt nun Fotos von unterschiedlichen Positionen um das Objekt auf. Das System zeigt interaktiv an, von welchen Positionen der Bounding Sphere noch Bilder benötigt werden und markiert farblich, ob Position, Orientierung und Entfernung zum Objekt richtig eingestellt sind. Schon während der Aufnahme kann das System Auskunft darüber geben, ob noch weitere Samples benötigt werden. Das System funktioniert auch als mobile Smartphone Applikation und ist daher besonders geeignet für den Consumer-Bereich.

3.3.2 Kamera-Array

Light Fields von nicht-statischen Szenen können durch ein Array von Kameras aufgenommen werden. Dazu müssen verschiedene Kameras in einem Array kalibriert und synchronisiert werden. Mit einem 1D-Array aus Kameras können so spannende Timelapse-Effekte erschaffen werden. Ein schneller Wechsel zwischen den verschiedenen Bildern erzeugt den Eindruck von einer Kamerafahrt in einer “eingefrorenen” Szene. Besondere Bekanntheit erlangte diese Technik als “Bullet-Time” Effekt im Film Matrix (Levoy 2006, S. 52). Mit einem 2D-Array von Kameras kann hingegen ein volles 4D Light Field aufgenommen werden. Das Stanford Multi-Camera-Array besteht aus 128 Kameras und ermöglicht neben statischen Light Fields auch die Aufnahme von Video Light Fields. Durch einen leichten Versatz der Auslösezeitpunkte können auch extreme High-Speed Aufnahmen einer Szene realisiert werden.

Abbildung 17: Das Stanford Multi-Camera-Array

Belichtet man die einzelnen Kameras unterschiedlich, können außerdem **High Dynamic Range** (HDR) Light Fields erzeugt werden (Levoy 2006, S. 52). Das Stanford Kamera Array ist allerdings sehr groß und eignet sich nicht für einen mobilen Einsatz.

Schäfer et al. (2013) präsentieren in ihrer Arbeit ein kompaktes Array aus High Definiton Kameras. Der Prototyp besteht aus einem Array von 2 x 3 Kameras, kann aber aufgrund seiner modularen Bauweise auf 8 x 8, bzw. 16 x 16 Kameras aufgestockt werden. Das aufgenommene Light Field ist durch 2 x 3 Kameras erst einmal nur spärlich abgetastet (Sparse Light Field Sampling). Durch die Disparität der einzelnen Bilder können allerdings implizite geometrische Informationen der Szene berechnet werden. Durch "Depth-Image-Based Rendering" Techniken können so Zwischenbilder der einzelnen Kameras interpoliert werden, die zusammen ein hoch abgetastetes Light Field (Dense Light Field Sampling) ergeben (Schäfer et al. 2013, S. 2).

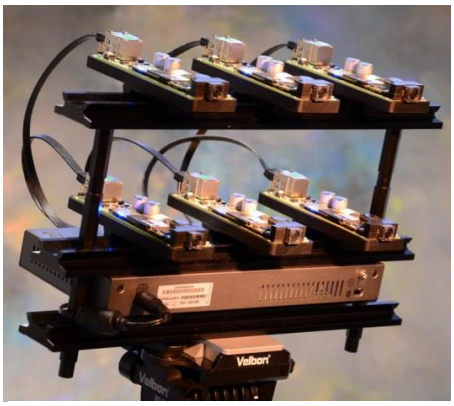

Abbildung 18: Kompaktes Array aus HD-Kameras (Schäfer et al. 2013, S. 2)

3.3.3 Plenoptische Kamera

Der Einsatz einer plenoptischen Kamera ist eine weitere Methode zur Aufzeichnung von nicht-statischen Light Fields. Eine plenoptische Kamera gleicht in Aufbau und Bedienung weitestgehend einer normalen Fotokamera. Vor dem Bildsensor der Kamera ist allerdings ein **Array aus Mikrolinsen** angebracht (Ng et al. 2005, S. 1).

Bei einer regulären Fotokamera werden einfallende Lichtstrahlen eines fokussierten Objekts durch die Hauptlinse auf einen Punkt auf dem Bildsensor gebündelt. Die Richtungsinformationen der einfallenden Lichtstrahlen gehen dabei verloren. In einer plenoptischen Kamera werden die Strahlen eines fokussierten Objekts auf dem Mikrolinsenarray gebündelt. Die jeweilige Mikrolinse separiert nun die Lichtstrahlen basierend auf ihrer Einfallsrichtung und projiziert sie in Form eines **Subimages** auf einem Pixelarray auf dem Sensor. Diese Subimages fangen also die Struktur und Richtungsinformationen des Lichts mit ein (Ng et al. 2005, S. 2). Die so gesammelten Subimages entsprechen einem 4D Light Field. Die Auflösung der Ebene (u, v) entspricht der Anzahl an Mikrolinsen vor dem Sensor. Jedes Subimage zeichnet nun verschiedene

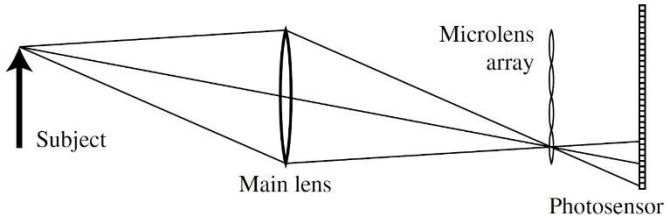

Abbildung 19: Strahlengang in einer plenoptischen Kamera

Richtungsinformationen der einzelnen Bildbereiche von unterschiedlichen Punkten der Ebene (u, v) auf. Die Auflösung der Ebene (s, t) wird bestimmt durch die Anzahl an Pixeln, die von einer Mikrolinse abgedeckt werden (Levoy 2006, S. 52).

Aus dem 4D Light Field können nun durch ein erneutes Abtasten (**Resampling**) der Subimages Bilder synthetisiert werden, als wären sie mit einer konventionellen Kamera geschossen worden. Die Bildsynthese dieser Sub-Aperture Images entspricht der physikalischen Simulation einer Kamera mit virtueller Bildebene und Blende (Ng et al. 2005, S. 3). Das ermöglicht nun auch das Rendering neuer Perspektiven. Allerdings ist die von den Mikrolinsen abgedeckte Parallaxe relativ klein und neue Blickwinkel beschränken sich weitestgehend auf den Bereich des Bildsensors (Levoy 2006, S. 52).

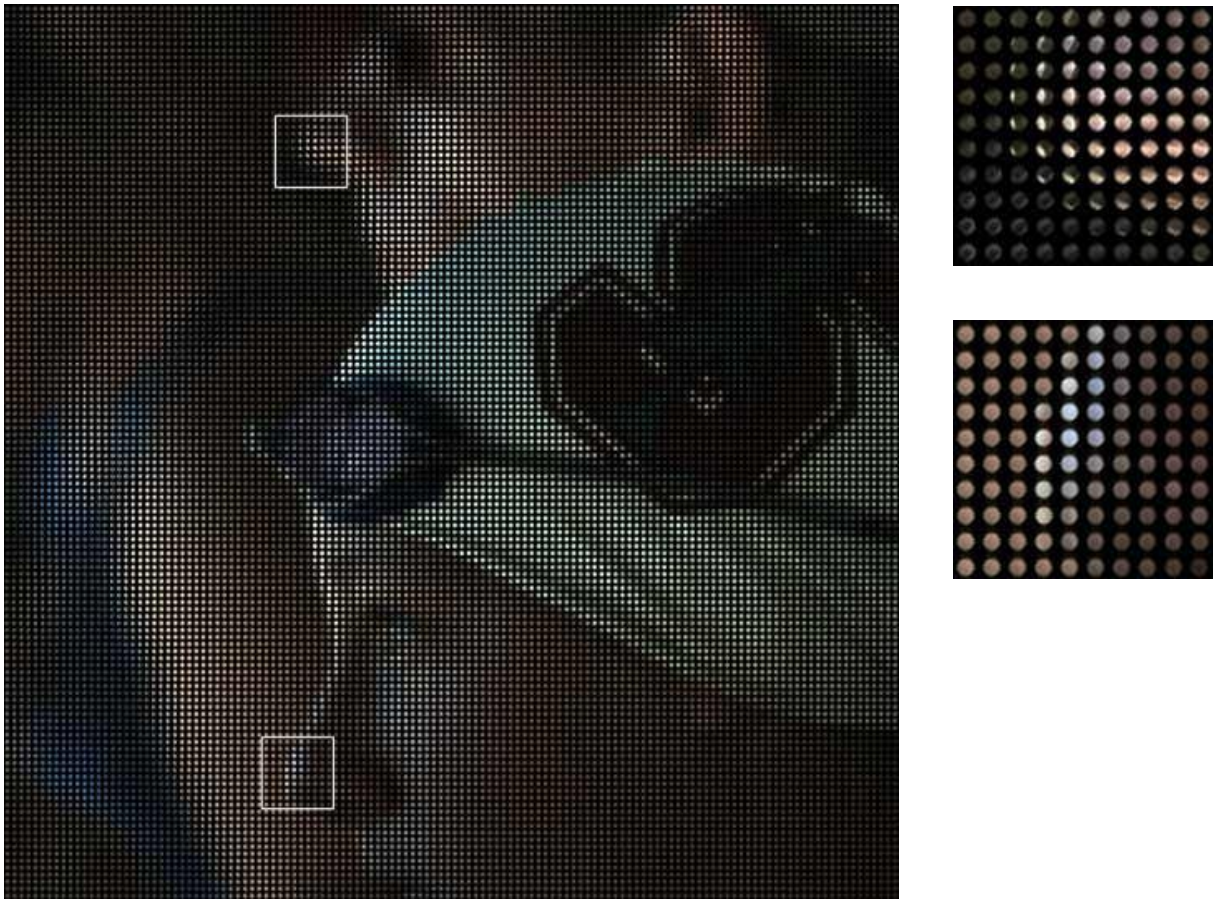

Abbildung 20: RAW Light Field Aufnahme einer plenoptischen Kamera. Die Vergrößerungen zeigen deutlich die kreisförmigen Subimages der einzelnen Mikrolinsen (Ng 2006, S. 44)

Durch die Simulation einer virtuellen Kamera können allerdings Bilder mit verschiedenen Schärfeebenen und Blendenöffnungen synthetisiert werden. Das erlaubt die Veränderung der Tiefenschärfe und das Refokussieren auf verschiedene Bildelemente in der Postproduktion. (Levoy 2006, S. 50–51).

Der große Nachteil einer plenoptischen Kamera ist die vergleichsweise kleine Auflösung der synthetisierten Bilder. Ein Mikrolinsenarray mit P x P Linsen, bei dem jede Linse einen Bereich von N x N Pixel des Bildsensors abdeckt, können Bilder mit einer finalen Auflösung von P x P Pixel synthetisiert werden (Levoy 2006, S. 53). Der Prototyp der plenoptischen Kamera von Ren Ng (2006) besitzt einen Bildsensor mit einer effektiven Auflösung von 4096 x 4096 Pixel und ein Mikrolinsenarray mit 296 x 296 Linsen, die jeweils wieder 12 x 12 Pixel auf dem Sensor abdecken. Synthetisierte Bilder haben also trotz des 16 Megapixel-Sensors lediglich eine Auflösung von 296 x 296 Pixel. Laut Ng Ren kann dieses Problem nur durch den Einsatz größerer Sensoren mit einer höheren Auflösung in Kombination mit feineren Mikrolinsenarrays behoben werden. (Ng 2006, S. 40)

3.4 Light Field Anwendungen

3.4.1 View Rendering

Das View Rendering bezeichnet die Synthese von neuen Blickpunkten auf eine Szene und ist somit eine der wichtigsten Anwendungen von Light Fields, die schon 1996 von Levoy und Hanrahan beschrieben wird. Sie bezeichnen das View Rendering als einfache und robuste Lösung, durch Abtastung und Kombination vorhandener Bildinformation eine neue Perspektive zu generieren (Levoy und Hanrahan 1996, S. 31). Um aus dem Light Field eine neue Perspektive zu erstellen, muss die Orientierung und Position einer virtuellen Kamera innerhalb des Light Slabs definiert werden. Aus dem resultierenden Strahlengang kann nun eine Zuordnung zwischen (u, v) und (s, t) Ebene getroffen werden. Durch Umsortierung und Interpolation der vorhandenen Bildpixel entsteht so eine neue Ansicht (Schreer 2005, S. 230).

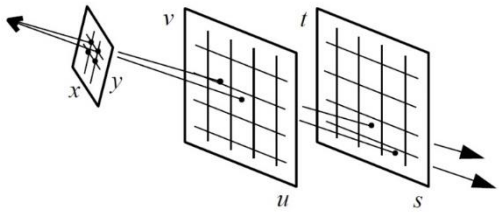

Abbildung 21: Transformation von (u, v) und (s, t) Ebene in Bildkoordinaten (x, y) (Levoy und Hanrahan 1996, S. 38)

Levoy und Hanrahan beschreiben das Mapping von L (u, v, s, t) in die Bildkoordinaten (x, y) als einfache Projektion, die lediglich lineare Rechenoperationen erfordert. Das Rendering von neuen Perspektiven ist somit unabhängig von der Komplexität der Szene und erfordert nur eine geringe Rechenleistung (Levoy und Hanrahan 1996, S. 38).

Levoy (2006) beschreibt in seinem Fachartikel “Light Fields and Computational Imaging” das View Rendering eines Light Fields am Beispiel eines Terracotta-Drachens. Um das Light Field zu erzeugen, wird eine große Anzahl an Fotos von einem Umkreis in einer bestimmten Distanz zum Objekt aufgenommen. Solange die Kamerapositionen nicht mit der konvexen Hülle (Strichlinie um das Objekt in Abbildung 22 b) des Objekts kollidieren, gilt die Voraussetzung eines freien Raums und ein 4D Light Field kann erzeugt werden. Ist das Sampling dicht genug, können nun neue Kameraperspektiven innerhalb des Umkreises synthetisiert werden.

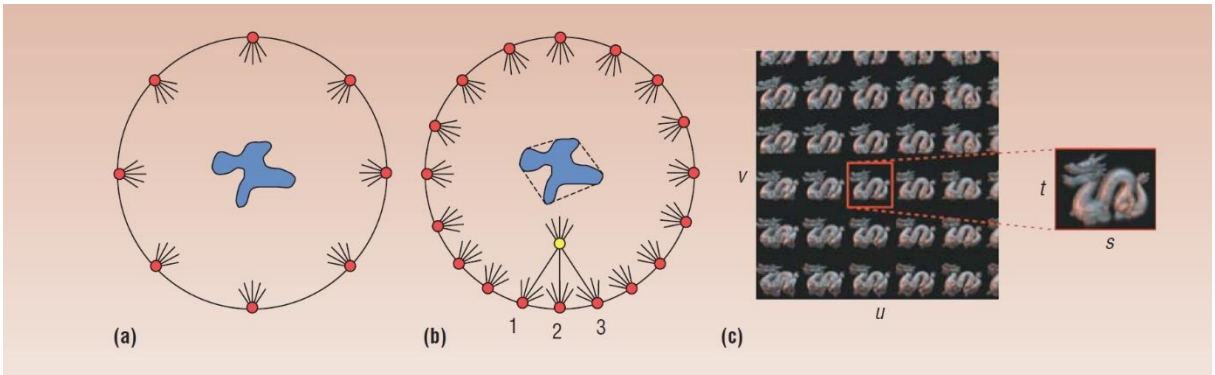

Abbildung 22: View Rendering am Beispiel eines Terracottadrachens (Levoy 2006, S. 48)

Der gelbe Punkt in Abbildung 22(b) ist eine virtuelle Kamera an einer neuen Position, die ursprünglich nicht aufgezeichnet wurde. Der zentrale Lichtstrahl der virtuellen Kamera ist jedoch identisch mit dem von Kamera (2). Lichtstrahlen, die links in die virtuelle Kamera einfallen, wurden bereits von Kamera (3) aufgezeichnet; Lichtstrahlen rechts durch Kamera (1). Ist das Light Field entsprechend abgetastet, so ist bereits alle Bildinformation vorhanden, die für die Rekonstruktion einer virtuellen Kamera benötigt wird (Levoy 2006, S. 48).

Das View Rendering ist letztendlich eine Transformation der (u, v) und (s, t) Koordinaten in Bildkoordinaten (x, y) sowie eine Intensitätsinterpolation der vorhandenen Bildpixel. Das View Rendering ist unabhängig von der Komplexität der Szene, benötigt keine große Rechenleistung und ist daher besonders geeignet für Echtzeitanwendungen.

Das Light Field des Terracotta-Drachens zeigt ein 360° Light Field eines Objekts und somit einen **inward looking view**. Gleichzeitig können Light Fields auch zur Erzeugung von **outward looking views** genutzt werden (Levoy und Hanrahan 1996, S. 31). Dreht man die Kamera an jeder Position um 180°, kann das View Rendering zur Erzeugung von Virtual Reality Inhalten genutzt werden. Innerhalb des Kreises können nun in Echtzeit neue Blickpunkte berechnet werden.

3.4.2 Synthetic Aperture

Synthetic Aperture ist eine beliebte Anwendung der plenoptischen Kamera und bezeichnet die Möglichkeit, Tiefenschärfe und Schärfeebene einer Aufnahme in der Postproduktion zu verändern.

Eine gewöhnliche Fotokamera fokussiert einen Punkt aus der realen Welt in genau einem Punkt auf dem Sensor und hat dadurch nur eine wirkliche Schärfeebene. Punkte, die auf einer anderen Ebene liegen, werden durch einen Unschärfekreis (**Circle of Confusion**) auf dem Sensor abgebildet. Erreicht dieser Unschärfekreis eine bestimmte Größe in Abhängigkeit zum Bildsensor, wird ein Punkt als unscharf wahrgenommen. Durch den Einsatz einer Blende hinter dem Objektiv kann die Menge des einfallenden Lichts auf den Sensor reguliert werden. Wählt man eine kleine Blendenöffnung, fallen die Lichtstrahlen gebündelt auf den Sensor und der resultierende Unschärfekreis ist entsprechend kleiner. Eine kleine Blendenöffnung vergrößert also die Tiefenschärfe und Objekte werden über einen größeren Bereich als scharf wahrgenommen (Levoy 2006, S. 50).

Die Bildsynthese aus einem 4D Light Field entspricht der physikalischen Simulation einer konventionellen Kamera. Durch eine Verschiebung der virtuellen Bildebene kann im Nachhinein auf unterschiedliche Bildinhalte fokussiert werden.1

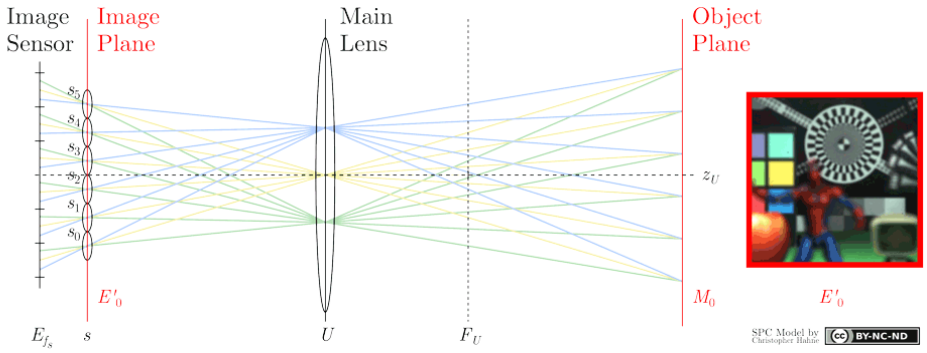

Abbildung 23: Refokussierung auf eine virtuelle Sensorebene E´₀ auf Höhe des Mikrolinsenarrays. Die Synthese eines Pixels auf E´₀ erfolgt durch die Interpolation aller Pixel des jeweiligen Subimages

Abbildung 23 zeigt den Strahlengang in einer plenoptischen Kamera. Lichtstrahlen von Punkten auf M_0 werden durch die Hauptlinse scharf auf dem Mikrolinsenarray abgebildet, durch die Mikrolinsen gebrochen und als Subimage auf dem Sensor gespeichert. Ein Bild mit virtueller Sensorebene $E´_0$ an der Stelle des Mikrolinsenarrays lässt sich nun synthetisieren, indem man die Pixel der einzelnen Subimages interpoliert (Hahne 2012).

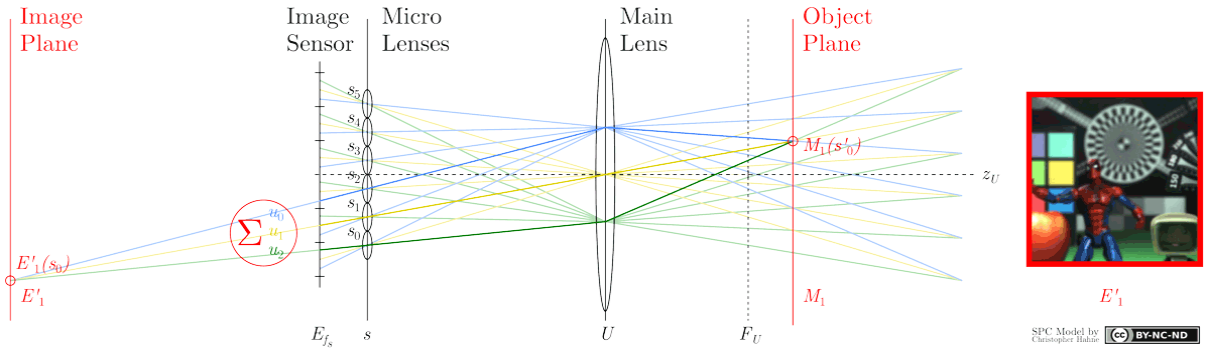

Abbildung 24: Refokussierung auf eine virtuelle Sensorebene E´₁ hinter dem Bildsensor. Die Synthese eines Pixels auf E´₀ erfolgt nun durch die Interpolation einzelner Pixel aus unterschiedlichen Subimages

Genauso kann nun aber auch auf andere Bildinhalte fokussiert werden. Durch Verschiebung der virtuellen Bildebene auf eine Ebene E´₁ hinter dem Sensor werden korrespondierende Lichtstrahlen der Objektebene M_1 fokussiert. Die Intensität eines Pixels $E'_1(s_0)$ kann nun durch Interpolation an den Stellen (u_2, s_0) (u_1, s_1) und (u_0, s_2) berechnet werden. Ein Bild an Ebene E_1 setzt sich also aus korrespondierenden Pixeln der unterschiedlichen Subimages zusammen. Je näher ein Objekt an der Kamera fokussiert werden soll, desto größer ist der Abstand zwischen den einzelnen Subimages, aus denen Pixel interpoliert werden (Hahne 2012).

Abbildung 25: Refokussierung einer Light Field Aufnahme auf unterschiedliche Ebenen (Ng 2006, S. 34)

Inwiefern sich Bildinhalte refokussieren lassen, ist abhängig von der Anzahl an Pixel, die von einer Mikrolinse abgedeckt werden (N x N) sowie dem Blendenwert der Hauptlinse der plenoptischen Kamera f/A. Synthetisierte Bilder können mit einer Schärfe abgebildet werden als wären sie mit einer Kamera mit dem Blendenwert f´(A x N) geschossen worden. Der von Levoy beschriebene Prototyp einer plenoptischen Kamera hat einen Blendenwert f/A = 4 und einen 16 Megapixel-Sensor, auf dem eine Mikrolinse jeweils 14 x 14 Pixel abdeckt. Bilder lassen sich folglich auf einen Schärfebereich fokussieren, der einer Aufnahme mit einem Blendenwert f´/A = 4 x 14 = 56 entspricht (Levoy 2006, S. 53).

3.4.3 Visual Effects und Postproduktion

Darüber hinaus bietet Light Field Technologie die Möglichkeit zu erheblichen Neuerungen in den Bereichen Cinematografie, Postproduktion und Visual-Effects.

Laut Levoy et al. ist das Abschätzen von 3D-Geometrie aus vorhandenem Bildmaterial ein zentrales Problem der Computer Vision. Eine Methode dafür sind „shape-from-stereo“ Algorithmen, die Tiefeninformationen anhand der Disparitäten und Korrespondenzen von Bildern mit unterschiedlichen Blickpunkten gewinnen. Bei „shape-from-focus“ Algorithmen hingegen wird 3D-Information aus zwei oder mehreren Bildern derselben Position, dafür aber mit unterschiedlichen Schärfeeinstellungen, rekonstruiert (Levoy et al. 2006, S. 930). Die Datenstruktur eines Light Fields enthält eine Vielzahl an verschiedenen Perspektiven einer Szene, die sich mit unterschiedlichen Schärfeebenen synthetisieren lassen. Es ist daher besonders geeignet, um 3D-Geometrie aus 2D-Bildmaterial zu generieren. Anhand der Korrespondenzen im Bildmaterial können so texturierte 3D-Modelle einer Szene erstellt werden (Levoy 2006, S. 49). So lassen sich in der Postproduktion digitale Kamerafahrten realisieren und 2D-Szenen nachträglich in einer 3D-Umgebung beleuchten (Tao et al. 2013, S. 679–680).

Die Möglichkeit, 2D-Material in einer 3D-Repräsentation zu bearbeiten, erleichtert außerdem das Zusammenspiel von Realbildaufnahmen und computergenerierten (CGI) Bildmaterial. Durch Light Field Aufnahmen lassen sich Tiefeninformationen für jeden Bildpixel ableiten. Anhand dieser Information können Bildinhalte nun durch Schwellwert-Operationen voneinander getrennt werden. Durch dieses „Depthscreen-Removal“ können beispielsweise Schauspieler effektiv vom Hintergrund einer Szene entfernt werden. Diese Technik könnte in Zukunft das aufwendige Ausleuchten und die Postproduktion von Greenscreen-Aufnahmen ersetzen (Karafin 2015, Min. 19).

Durch das Refokussieren in der Postproduktion lassen sich Bilder erzeugen, die ein Kameraassistent am Set nur schwer reproduzieren kann. Bilder mit extrem kleiner Tiefenschärfe stellen in der Realität ein großes Problem dar. Gerade bei Kamerafahrten oder Bewegung im Bild ist es schwer, eine durchgängige Schärfe zu gewährleisten. Light Field Aufnahmen hingegen können Bild für Bild in der Postproduktion fokussiert werden (Karafin 2015, Min. 12). Darüber hinaus lassen sich aus den Daten eines Light Fields Bewegungsvektoren generieren. Laut Karafin (2015, Min. 13–14) kann durch diese Bewegungsinformation die Belichtungsdauer und somit auch der Motion-Blur von Bildmaterial in der Postproduktion gesteuert werden. Zuletzt enthalten Light Fields bereits wichtige

Informationen zur Darstellung von Stereoskopie. Durch View Rendering lassen sich leicht perspektivisch korrekte Bilder mit gewünschter Disparität realisieren. Dabei kann bei jeder Anwendung der individuelle Augenabstand eines Benutzers berücksichtigt werden (Karafin 2015, Min. 16).

Light Field Technologie kann die Art und Weise, wie wir Bildmaterial aufnehmen und produzieren grundlegend verändern. Die vergleichsweise einfache Integration von 3D-Daten und CGI-Elementen, kann bestehende VFX-Workflows erheblich erleichtern. Viele wichtige Entscheidungen, die heute noch am Set getroffen werden müssen, können in Zukunft auf die Postproduktion ausgelagert werden. Wichtige kreative Merkmale wie Tiefenschärfe und Bewegungsunschärfe können leichter gesteuert und nach Belieben angepasst werden. Gleichzeitig besteht dadurch aber auch die Gefahr, Probleme und schwierige Entscheidungen in der Produktion einfach auf die Postproduktion abzuwälzen.

4 Light Field Technologie für VR: Aktueller Stand der Forschung

Eine ganze Reihe von Experten sehen in Light Field Technologie die Zukunft von Virtual Reality (USC Institute for Creative Technologies 2015). Durch den Einsatz von Light Fields sollen vor allem die zentralen Probleme heutiger VR-Systeme adressiert werden: Positionstracking in Live-Action VR sowie der Vergence-Fokus Konflikt.

Das folgende Kapitel widmet sich exemplarisch verschiedenen Ansätzen, Light Field Technologie für Virtual Reality zu nutzen und spiegelt den aktuellen Stand der Forschung wider. OTOY und Lytro arbeiten unabhängig voneinander an einer kompletten Pipeline zur Light Field Produktion für Virtual Reality. Light Stage Technologie ermöglicht das 360° Light Field Capture von Personen und kann für die Erzeugung digitaler Charaktere in Virtual Reality Umgebungen genutzt werden. Darüber hinaus werden in Kapitel 4.4 verschiedene Ansätze aufgezeigt, Light Field Technologie für die Verbesserung heutiger Head-Mounted Displays zu nutzen.

Wie aber lassen sich Light Fields nun für die Erzeugung von Virtual Reality Inhalten nutzen? Eine wichtige Voraussetzung ist natürlich die komplette Abbildung des Raums, 360° horizontal und vertikal. Die in Kapitel 3.4.1 beschriebene Methode zur Aufnahme von 360° Light Fields eines Objekts eignet sich genauso, um auch „outward looking views“ einer Szene zu generieren. Dazu muss die Kamera lediglich umgedreht werden; sie sammelt nun Fotos des Raums von unterschiedlichen Positionen eines Kreises. Um eine Aufnahme des gesamten Raums, horizontal wie auch vertikal zu gewährleisten, muss man dieses Modell um eine Dimension erweitern. Die Fotos werden nun nicht von Punkten eines Kreises, sondern von unterschiedlichen Punkten auf einer Kugel im Raum gesammelt.

Laut Debevec (2012) ließe sich dieses Modell mit einem kugelförmigen Rig aus Videokameras realisieren. Reduziert man das Light Field auf eine statische Szene, spielt der Zeitpunkt der Aufnahmen keine Rolle mehr. Ähnlich der Aufnahme mit einem „Spherical Gantry“, können die Bilder nun nacheinander und mit nur einer Kamera aufgenommen werden. Für eine 360° Aufnahme eignet sich dazu ein Panoramastativ. Um Aufnahmen von verschiedenen Punkten einer Kugel und somit Parallaxe innerhalb der einzelnen Bilder zu gewährleisten, muss die

Kamera allerdings außerhalb ihres Nodalpunkts mit einem gewissen Abstand auf dem Stativ angebracht werden. Genau dieses Modell eines Panoramastativs hat Paul Debevec in Zusammenarbeit mit der Rendering Company OTOY realisiert und damit das erste Light Field Aufnahmesystem für Virtual Reality entwickelt. Das aufgenommene Light Field erlaubt volle Bewegungsfreiheit des Kopfs innerhalb der Kugel bei voller Parallaxe und realistischen Lichtreflexen (Debevec 2012, Min. 30–32).

Abbildung 26: Light Field Aufnahme mit einem Panoramastativ

4.1 OTOY

Die Rendering Company OTOY wurde 2008 in Kalifornien gegründet und gilt als Vorreiter in den Bereichen Cloud-Computing und Real-Time Rendering. Aktuell sind sie vor allem durch ihre Renderengine Octane bekannt. Octane-Render ist ein physikalisch korrekter GPU-Renderer. Die Engine nutzt hierfür die Rechenleistung von einer oder mehreren Grafikkarten, um physikalisch korrekte, fotorealistische Bilder zu berechnen. Aufgrund der Parallelisierung von Rechenoperationen ist Octane aber bis zu 50 Mal schneller, als herkömmliche CPU-Renderer (Otoy inc. 2015c). Im Mai 2015 präsentieren Paul Debevec und OTOY die erste 360° Light Field Aufnahme für Live-Action VR und somit erstmals eine fotorealistische VR-Umgebung mit Positionstracking, Bewegungsparallaxe, akkuraten Lichtreflexen und Spiegelungen (Otoy inc. 2015d).

Abbildung 27: Bildmaterial des 360° Light Fields, dargestellt durch ein VR-System

Gleichzeitig nutzt OTOY das Potenzial ihrer Renderengine Octane zur Erzeugung von synthetischen Light Fields. Die für ein Light Field benötigte Bildersammlung wird hierfür nicht von einer Fotokamera erstellt, sondern synthetisch aus einer 3D-Szene gerendert. Octane kann nun alle benötigten Lichtstrahlen aus einem Light Field in Echtzeit berechnen und in einem

VR-System darstellen (Otoy inc. 2015b). Die Engine erlaubt allerdings einen viel höheren Grad an Fotorealismus als gewöhnliche Real-Time Renderengines:

> *„Every reflection, refraction, and absorption of that field of light is modelled as it would be in the real world.“* (Otoy inc. 2015b)

Zusammen mit der Beta-Version von Octane VR veröffentlichte OTOY in Kooperation mit Oculus zu Beginn des Jahres 2015 den Wettbewerb „Render The Metaverse“. Nutzer von Octane können mit den neuen Features von Octane VR eigene Virtual Reality Umgebungen erstellen und in den Wettbewerb einreichen. Die monatlichen Gewinner erhalten ihre Szene als vollständig navigierbares Light Field Rendering, die auf der Homepage von OTOY betrachtet werden können (Otoy inc. 2015g).

Die Funktionen der Beta-Version sollen in die nächste Version von Octane integriert werden und ermöglichen Nutzern das Rendering von hyperrealistischen Virtual Reality Erfahrungen (Otoy inc. 2015e). Neben dem Rendering von herkömmlichen Pre-Rendered CGI Content soll Octane VR in der Lage sein, direkt ein Light Field einer Szene zu synthetisieren.

Laut Jules Urbach (2015, Min. 7) arbeitet OTOY außerdem an einem System für „Production Ready Light Field Capture and Playback“, dem 360° Light Field Capture von Personen sowie an einem Kamera-Rig für 360° Light Field Videos.

Ermöglicht werden diese Entwicklungen unter anderem durch OTOYs eigenes Dateiformat. ORBX Files gewährleisten eine effektive Kompression und Speicherung von Light Field Daten und somit das Rendering und Streaming von Light Field Videos in Echtzeit. Verglichen mit dem Industriestandard H.264 Video Codec benötigen ORBX Files lediglich die halbe Bitrate bei gleicher Bildqualität (Otoy inc. 2013). OTOYs eigener ORBX Mediaplayer findet sich mittlerweile im Oculus und Gear VR Store und kann dazu genutzt werden, Light Fields und holographische Videos zu betrachten.

4.2 Lytro Immerge

Die Firma Lytro wurde 2006 von Ren Ng gegründet und ist ein Hersteller von Lichtfeld Kameras für den Consumer-Bereich. Von der Stanford Universität wurde Ren Ng im Vorfeld für die beste Thesis im Gebiet Computer Science ausgezeichnet. In seiner Doktorarbeit entwickelte er den Prototyp einer plenoptischen Kamera, die 2012 als erste kommerzielle Licht

Feld Kamera „Lytro“ auf den Markt kam (Lytro Inc. 2014a). Sie erlaubt eine leichte Bewegungsparallaxe, sowie das Refokussieren von Bildern in der Postproduktion. 2014 erschien mit der Lytro Illum die 2. Generation ihrer Lichtfeldkamera, die sich durch einen größeren Bildsensor, höhere Auflösung und einen größeren Refokus-Bereich auszeichnet (Lytro Inc. 2014a).

Aktuell arbeitet das Unternehmen mit der Lytro Cinema an einer Light Field Kamera für Kino und TV-Produktionen und spricht damit vor allem professionelle Nutzer an. Nach eigenen Angaben filmt die Kamera in einer Auflösung von 755 Megapixel mit bis zu 300 Bildern pro Sekunde und einem Dynamikumfang von bis zu 16 Blenden (Lytro Inc. 2014b). Das aufgenommene Light Field ermöglicht ungeahnte Möglichkeiten in der Postproduktion. Bildinhalte lassen sich nach der Aufnahme refokussieren. Shutter-Angle und Motion-Blur können artifiziell in der Postproduktion verändert werden. Das „Depthscreen-Removal“ (Siehe Kapitel 3.4.3) erleichtert die Trennung von Schauspielern und Hintergrund und kann somit herkömmliche Greenscreen-Verfahren ersetzen (Lytro Inc. 2014b). Die Art und Weise, wie Bilddaten als dreidimensionales Volumen verarbeitet werden, erleichtert die Integration von CGI-Elementen und daher ist die Lytro Cinema besonders interessant für professionelle Visual Effects Produktionen (Karafin 2015, Min. 13–14).

Abbildung 28: Lytro Immerge 360° Light Field Kamera

Mit der Lytro Immerge präsentiert Lytro 2015 die erste 360° Light Field Videokamera für Virtual Reality. Die kugelförmige Kamera besteht aus einem modifizierbaren Array einzelner Videokameras. In einer Pressemitteilung von 2015 verspricht Lytro 360° Light Field Aufnahmen in professioneller Kinoqualität (Lytro Inc. 2015). Die Größe des aufgenommenen Light Field Volumens ist dabei durch die Größe des kugelförmigen Kamera-Arrays definiert.

> *"A fully sampled Light Field Volume includes ray Data from all directions at any location within a given Volume"* (Karafin 2015, Min. 25)

Innerhalb des Light Field Volumens können beliebig neue Perspektiven mit gewünschtem Field of View rekonstruiert werden. Das aufgenommene Light Field erlaubt in einer VR-Umgebung

eine Bewegungsfreiheit um sechs Achsen und somit ein gesteigertes Gefühl von Präsenz in Live-Action VR (Lytro Inc. 2015).

Die Lytro Immerge ist aber nicht nur eine Kamera, sondern vielmehr ein komplettes Produktionssystem für Virtual Reality, von der Aufnahme bis zur Postproduktion (Karafin 2015, Min. 26). Um die großen Datenraten von Light Field Videos in den Griff zu bekommen, wird die Immerge mit einem eigenen Server ausgeliefert. Durch eine große Bandbreite soll eine angemessene Speicherung und Verarbeitung der anfallenden Bilddaten ermöglicht werden.

Abbildung 29: Produktionssystem der Lytro Immerge

Die Bearbeitung der Light Field Videos soll durch eigene Plugins für professionelle Editing- und VFX-Software gewährleistet werden. Die Einbindung in bestehende Tools und Workflows erleichtert die Integration von Live-Action VR und CGI Elementen. Zum Betrachten der Light Field Aufnahmen stellt Lytro einen eigenen Player zur Verfügung, der durch alle VR Headsets und Systeme der aktuellen Generation unterstützt wird. Ein eigener Streaming Server soll außerdem das Streaming von Light Field Videos in Echtzeit ermöglichen (Lytro Inc. 2015).

4.3 Light Stage Technologie

Eine Light Stage ist eine Vorrichtung, in der Objekte durch kontrollierbare Lichtquellen vollständig ausgeleuchtet werden. Sie ermöglichen die Beleuchtung und Simulation von Licht auf einem Objekt mit jeder Farbe, Intensität und Richtung. Light Stage Aufnahmen können in der Postproduktion nach Belieben neu beleuchtet werden (**Relighting**). Mittlerweile gibt es verschiedene Modelle der Light Stage für unterschiedliche Anwendungsbereiche (Debevec 2012, S. 1).

Zielsetzung der ersten Light Stage im Jahr 2000 war das Relighting von realem Bildmaterial durch **Image-Based Lighting** (IBL) Techniken. Image-Based Lighting bezeichnet die

Beleuchtung von 3D-Szenen durch 360° HDR-Panoramas einer realen Umgebung. Dadurch das Licht einer realen Szene in einer CGI-Umgebung nachempfunden werden; feine Details in den Glanzlichtern und Schatten einer Szene gewährleisten einen hohen Grad von Realismus. (Debevec 2012, S. 1)

Die Light Stage 1 bewegt ein gerichtetes Licht kugelförmig um das Gesicht einer Person. Eine Videokamera zeichnet jede mögliche Lichtrichtung auf, aus der eine Person beleuchtet werden kann (Debevec 2012, S. 1). Analog zu einer Light Field Aufnahme werden hier also nicht verschiedene Perspektiven bei einer konstanten Beleuchtung, sondern eine Perspektive mit jeder möglichen Lichtrichtung aufgezeichnet. Diese Sammlung an Bilddaten bezeichnet Debevec als **4D Reflectance Field** (2012, S. 1).

Aufgrund des linearen Verhaltens von Licht können nun einzelne Farbkanäle angesteuert, zwischen den verschiedenen Beleuchtungssituationen überblendet und somit komplett neue Lichtsituationen geschaffen werden. Diese Lichtsituationen sind eine lineare Kombination des vorhandenen Bildmaterials und berücksichtigen eine realitätsgetreue Abbildung von diffusen und spekularen Reflektionen, Subsurface Scattering sowie Lichtreflexen und Verschattungen. Durch Image-Based Lighting und können Aufnahmen nun außerdem durch HDR-Panoramas beleuchtet werden (Debevec 2012, S. 1).

Neuere Generationen der Light Stage bestehen aus einer großen Anzahl, individuell steuerbarer LEDs, angeordnet in einem kugelförmigen Array. Durch sogenannte „Time Multiplexing“ Verfahren kann die Beleuchtung einer Szene in Sekundenbruchteilen mehrmals gewechselt werden. Durch den Einsatz von Highspeed-Kameras können dynamische Reflectance Fields von Personen aufgenommen werden, die selbst feine Details wie Faltenwurf und Bewegung festhalten.

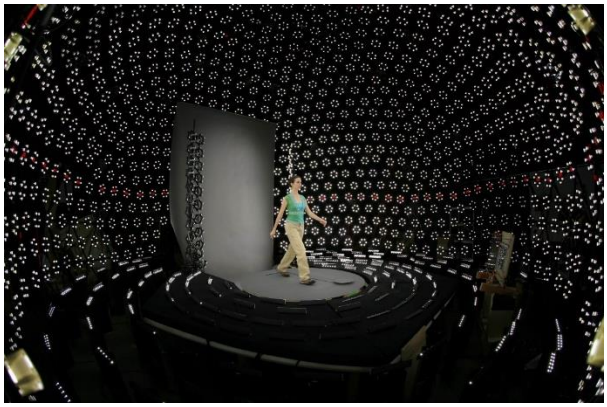

Abbildung 30: Ganzkörperaufnahme einer Person mit der Light Stage 6 (Debevec 2012, S. 1)

In Kombination mit verschiedenen 3D-Scanning Verfahren, können so extrem detaillierte 3D-Rekonstruktionen von Personen angefertigt werden (Debevec 2012, S. 1–2).

Ein 4D Refletance Field ist einem 4D Light Field in vielerlei Hinsicht sehr ähnlich. Tatsächlich lassen sich beide Technologien durch eine Light Stage Aufnahme kombinieren.

Während des „Relighting Human Motion“ Projekts wurden Schauspieler in Bewegung auf einem rotierenden Laufband aufgenommen. Die einzelnen LED-Lichter der Light Stage wiederholen pro Sekunde 33 verschiedene Beleuchtungen der Szene. Die Schauspieler werden durch drei High-Speed Kameras in einem vertikalen Array mit 990 Bildern pro Sekunde aufgenommen. Durch die horizontale Drehung des Laufbands wird so neben dem Reflectance Field auch ein 4D Light Field aufgenommen (Debevec 2012, S. 2).

Abbildung 31: Digitaler Charakter, gerendert aus unterschiedlichen Blickwinkeln im Rahmen des Relighting Human Motion Projekts (Debevec 2012, S. 2)

Das Ergebnis ist nun ein dynamisches 360° Light Field eines Schauspielers, welches zugleich noch Informationen über 33 verschiedene Lichtsituationen enthält. Eine fotorealistische, digitale Version des Schauspielers kann nun in Echtzeit aus jedem Blickwinkel und mit jeder Kamerabewegung in neue Bildinhalte eingesetzt und dabei komplett neu beleuchtet werden (Debevec 2015, Min 26–28). Solche 360° Light Fields von Personen lassen sich natürlich für Virtual Reality Anwendungen nutzen. Fotorealistische Light Stage Aufnahmen können in VR-Anwendungen als digitale Charaktere dienen und sowohl in Live-Action VR als auch in Videospielinhalte eingebaut werden (Siehe Kapitel 4.5). Die Light Stage Technologie wurde 2008 von OTOY lizensiert und weiterentwickelt. Für die Aufnahme von Schauspielern bietet OTOY mittlerweile eine mobile Version der Light Stage an, die bei ihnen gemietet werden kann (Otoy inc. 2015a).

4.4 Light Field Displays

Wie bereits in Kapitel 2.7.2 beschrieben, ist der Vergence-Fokus ein zentrales Problem heutiger Head-Mounted Displays und VR-Systeme. Durch die scheinbare räumliche Tiefe von Stereobildern konvergieren die Augen auf den virtuellen Abstand von Objekten vor oder hinter dem Display. Gleichzeitig müssen die Augen aber auf den Abstand des Displays fokussieren, um die gezeigten Bildinhalte scharf wahrzunehmen. Dabei spielen Akkommodationshinweise vor allem im Nahbereich von 20 bis 120 cm eines Betrachters eine wichtige Rolle (Huang et al. 2015, S. 3). Einem Bereich also, der gerade in VR-Umgebungen für Kopfbewegungen oder die Interaktion der Hände genutzt wird. Diese künstliche Trennung von Konvergenz und

Akkommodation der Augen widerspricht unseren natürlichen Sehgewohnheiten und führt zu Nebenwirkungen, wie dem Auftreten von Doppelbildern, Ermüdungserscheinungen, Kopfschmerzen oder Schwindelgefühlen (Huang et al. 2015, S. 1–2).

Der Akkommodationsabstand des menschlichen Auges führt zu einem weiteren Problem aktueller Head-Mounted Displays. Das Display in herkömmlichen VR-Brillen befindet sich sehr nah an den Augen eines Betrachters und kann erst einmal nicht fokussiert werden. Durch den Einsatz einer Vergrößerungslinse wird der Abstand zwischen Auge und Display künstlich verlängert, um Bildinhalte scharf abbilden zu können. Dieser Einsatz von optischen Elementen erfordert allerdings eine Vergrößerung des gesamten Systems, da nun die Brennweite der Vergrößerungslinse berücksichtigt werden muss. Heutige Head-Mounted Displays sind immer noch relativ groß, schwer und unbequem (Lanman und Luebke 2013, S. 1).

Durch den Einsatz von Light Fields lassen sich wesentlich dünnere und kleinere Head-Mounted Displays realisieren, die in ihrer Erscheinung eher an eine Sonnenbrille erinnern (Lanman und Luebke 2013, S. 1). Im Display wird anstatt einem 2D-Bild ein Light Field synthetisiert. Analog zu einer Light Field Aufnahme erlaubt ein Light Field Display eine natürliche Akkommodation und das Fokussieren auf unterschiedliche Schärfeebenen (Lanman und Luebke 2013, S. 2).

4.4.1 Near-Eye Light Field Display

Lanman und Luebke (2013) präsentieren in ihrer Arbeit ein „Near-Eye Light Field Displays“ den Prototypen eines Light Field Displays, das eine nahezu korrekte Konvergenz und Akkommodation der Augen, sowie binokulare Disparität und Retinal Blur erlaubt.

Ein Light Field Display teilt sich viele technische Eigenschaften mit einer plenoptischen Kamera. Über das Display wird dem Betrachter ein typisches Bild einer Light Field Aufnahme, mit einer Vielzahl an Subimages aus unterschiedlichen Perspektiven gezeigt. Durch den Einsatz eines Mikrolinsenarrays vor dem Display wird ein Light Field im natürlichen Akkommodationsabstand des menschlichen Auges synthetisiert (Lanman und Luebke 2013, S. 1). Der Betrachter sieht nun nicht ein Array aus Subimages, sondern ein Bild mit korrekter Perspektive und Tiefenschärfe. Im Gegensatz zu herkömmlichen Displays, mit nur einer Vergrößerungslinse, erlaubt der Einsatz von Mikrolinsen eine geringere Brennweite bei gleichem Field of View und somit eine leichte und kompakte Bauweise des Head-Mounted Displays (Lanman und Luebke 2013, S. 1).

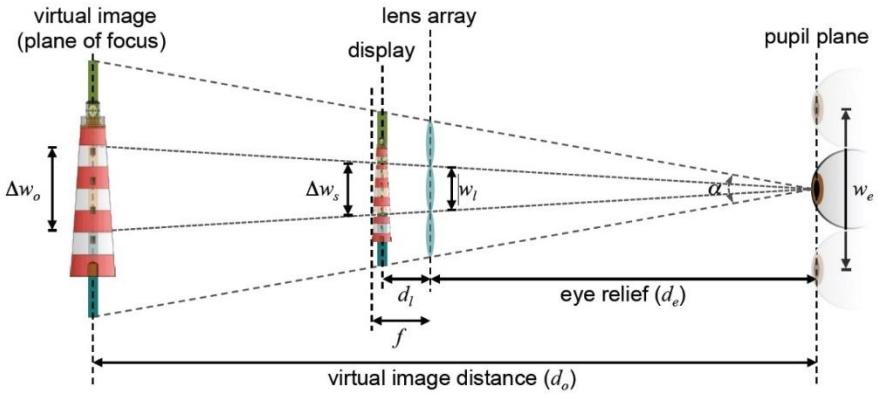

Abbildung 32: Aufbau des Near-Eye Light Field Displays (Lanman und Luebke 2013, S. 4)

Abbildung 32 zeigt den Aufbau des Near-Eye Light Field Displays. In Abstand d_l vor dem Display ist ein Array von Mikrolinsen mit Auflösung N_l angebracht. Jede Mikrolinse hat die breite W_l und Brennweite f. Das Display zeigt nun eine Light Field Aufnahme, bestehend aus einer großen Anzahl an Subimages, mit leicht unterschiedlicher Parallaxe. Wie in einer plenoptischen Kamera wird jedes Subimage von einer Mikrolinse abgedeckt. Jede Mikrolinse verhält sich nun wie eine unabhängige Vergrößerungslinse, die jeweils eine unterschiedliche Perspektive der virtuellen Bildebene in Abstand d_o synthetisiert.

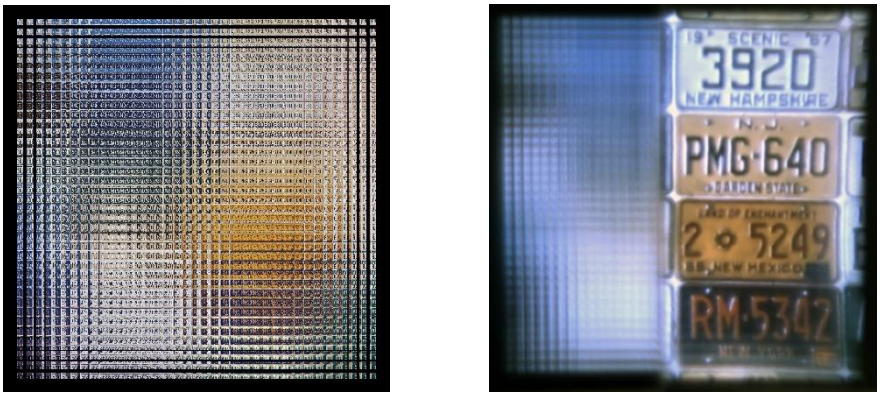

Abbildung 33: Light Field Aufnahme, wie sie auf dem Bildsensor des Displays dargestellt wird. (links) Vor dem Sensor ist noch kein Mikrolinsenarray angebracht. Die einzelnen Subimages sind deutlich sichtbar. (rechts) Über der rechten Hälfte des Bildes liegt nun ein Mikrolinsenarray. Der Betrachter nimmt jetzt ein scharfes Bild wahr (Lanman und Luebke 2013, S. 7)

Somit wird ein 4D Light Field erzeugt, das die gesamte Eyebox mit Breite W_e abdeckt. Solange sich das Auge des Betrachters innerhalb der Eyebox befindet, werden nun in Abhängigkeit von Position und Rotation des Auges perspektivisch korrekte Bilder wahrgenommen und der Betrachter kann auf unterschiedliche Bildinhalte fokussieren (Lanman und Luebke 2013, S. 4–5).

Das Field of View des Displays ist proportional abhängig von der Anzahl der verbauten Mikrolinsen N_l, ohne dabei die Brennweite der einzelnen Linsen und somit die Dicke des Head-Mounted Displays zu beeinflussen. Dadurch kann ein Light Field Display wesentlich dünner und leichter gebaut werden als herkömmliche HMDs (Lanman und Luebke 2013, S. 5).

Analog zu den technischen Eigenschaften einer plenoptischen Kamera geht mit den Vorteilen einer dünnen Bauweise aber auch ein Verlust an finaler Bildauflösung einher. Dieser Verlust ist proportional zu dem Verhältnis der Brennweite der Mikrolinsen f und dem Abstand des Displays zum Auge d_e + f. Der dünne Formfaktor des Light Field Displays steht also im direkten Konflikt mit der Bildauflösung des Displays (Lanman und Luebke 2013, S. 5).

Im Prototyp des Near-Eye Light Field Display ist ein OLED Display mit einer Auflösung von 1280 x 720 Pixel verbaut. Die synthetisierten Bilder des Light Field Displays haben allerdings lediglich eine finale Auflösung von 146 x 78 Pixel bei einem Field of View von 29,2° x 16°. Anhand verschiedener Beispielszenen wurden außerdem Bildwiederholraten von 15-70 Hz gemessen. Innerhalb des Displays kann der Betrachter in einem Bereich von 30,6 cm vor dem Display bis zur optischen Unendlichkeit auf Bildinhalte fokussieren (Lanman und Luebke 2013, S. 6–7).

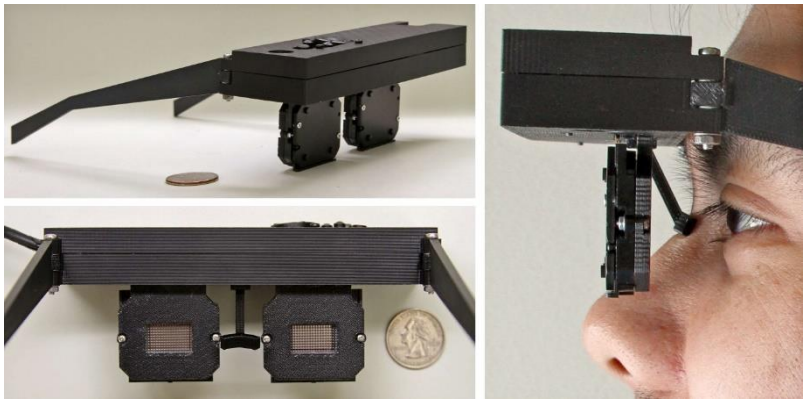

Abbildung 32: Prototyp des Near-Eye Light Field Displays (Lanman und Luebke 2013, S. 1)

Das Near-Eye Light Field Display erlaubt also den Bau von leichten, bequemen Light Field Displays mit einem dünnen Form-Faktor und adressiert gleichzeitig den Vergence-Fokus Konflikt. Aufgrund eines relativ kleinen Field of View, sowie einer relativ geringen Bildauflösung eignet sich das Near-Eye Light Field Display nur bedingt als kommerzielle Lösung für aktuelle VR-Systeme.

4.4.2 Light Field Stereoscope

Huang et al. verzichten in ihrem Display auf den Einsatz eines Mikrolinsenarrays und verfolgen mit dem „Light Field Stereoscope“ einen anderen Ansatz. In ihrer Arbeit präsentieren sie ein kostengünstiges und praktisches Light Field Display, das alle monokularen und binokularen

Tiefenhinweise unterstützt und direkt den Vergence-Fokus Konflikt adressiert. (Huang et al. 2015, S. 2).

Sie bezeichnen ihr Head-Mounted Display als Hybrid aus einem **Multi-Focal-Plane Display** und einem Light Field Display. Herkömmliche Multi-Focal-Plane Displays bestehen aus mehreren halbtransparenten Displays, die in bestimmten Abständen hintereinander in einem HMD verbaut sind. Bildinhalte lassen sich so auf verschiedenen Tiefenebenen innerhalb des Displays staffeln und können einzeln fokussiert werden. Das finale Bild setzt sich im Auge des Betrachters additiv aus den einzelnen Bildinhalten zusammen. Head-Mounted Displays mit bis zu fünf einzelnen Displays erlauben die Akkommodation der Augen auf einen Bereich von 25 cm bis zur optischen Unendlichkeit. Problematisch ist allerdings die erforderliche Bildwiederholrate. 300 Bilder müssten pro Sekunde berechnet werden, um auf jedem der fünf Displays eine Bildwiederholrate von 60 Bildern pro Sekunde zu gewährleisten (Huang et al. 2015, S. 2–3).

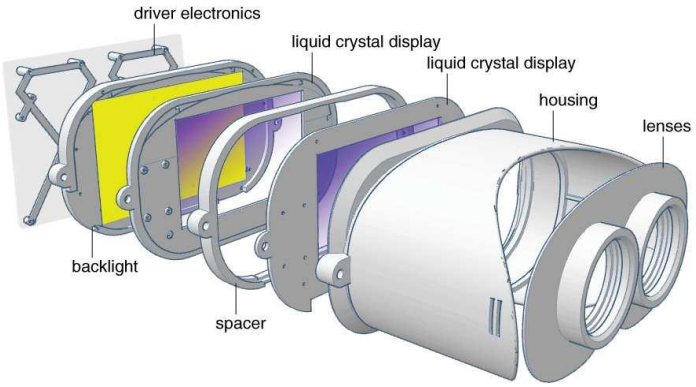

Abbildung 33: Schematische Darstellung des Light Field Stereoscopes (Huang et al., S. 2)

Das Light Field Stereoscope hingegen besteht aus nur zwei halbtransparenten LCD Displays, die hintereinander in einem HMD verbaut sind, dafür aber ein Light Field erzeugen und somit eine korrekte Akkommodation und Retinal Blur ermöglichen. Die Bildsynthese im Auge erfolgt dabei nicht additiv, sondern multiplikativ und soll das Fokussieren innerhalb eines größeren Bereichs als herkömmliche Multi-Focal-Plane Displays ermöglichen (Huang et al. 2015, S. 3).

Abbildung 36 zeigt den technischen Aufbau des Light Field Stereoscopes. Die beiden physikalischen Displays befinden sich auf Distanz d'_1 und d'_2 zur Hauptlinse. Für den Betrachter ergeben sich daraus zwei virtuelle Bilder in den Abständen d_1 und d_2. Das Auge eines Betrachters befindet sich innerhalb der Eyebox mit Breite e. Das View Frustrum ist der Bereich, der pro Frame gerendert werden muss. Dieses ist in seiner vertikalen Darstellung (Side View) zwar symmetrisch und somit für beide Augen gleich, in der horizontalen (Top View) aber aufgrund des Augenabstandes asymmetrisch (Huang et al. 2015, S. 4). Durch die beiden Displays wird nun ein Light Field synthetisiert.

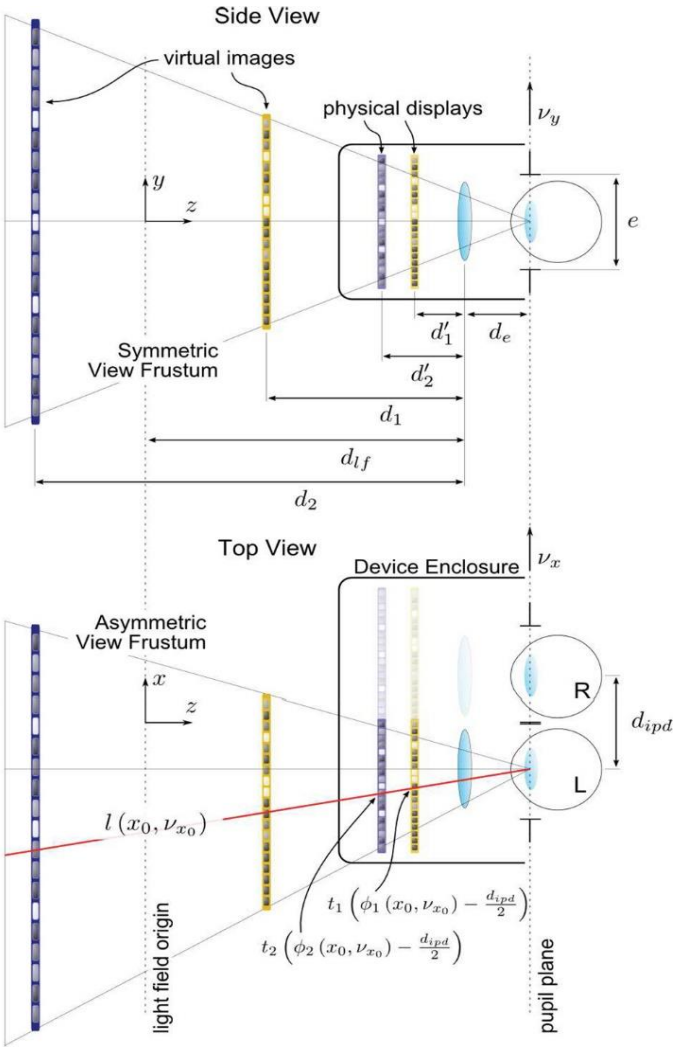

Abbildung 34: Technischer Aufbau des Light Field Stereoscopes (Huang et al., S. 4)

Die Parametrisierung des Light Fields erfolgt analog zu Levoy und Hanrahan (1996) durch die Schnittpunkte eines Lichtstrahls an zwei Ebenen. Ein Lichtstrahl ist in diesem Fall durch seine Schnittpunkte an der Light Field Ebene (x, y) und der Pupillen-ebene (Vx, Vy) definiert. Aufgrund des asymmetrischen View Frustrums ergeben sich unterschiedliche Bilder eines Light Fields für jedes Auge (Huang et al. 2015, S. 4).

An dieser Stelle stellt sich die Frage, wie durch nur zwei Bilder pro Auge ein Light Field erzeugt werden kann. In Kapitel 3.3 wurde bereits festgestellt, dass auch ein einzelnes 2D-Bild ein Light Field repräsentiert, es aber nur sehr schlecht abtastet.

Huang et al. (2015) kommen zu dem Ergebnis, dass innerhalb eines Light Field Displays eine Abtastung durch zwei Bilder pro Auge ausreicht, um ein angemessenes Light Field zu erzeugen. Sie bezeichnen die Aufteilung eines Light Fields auf nur zwei Bilder als **Rank-1 Faktorisierung** (Huang et al. 2015, S. 9). Da sich das Auge eines Betrachters innerhalb der Eyebox befindet und dementsprechend nur einen kleinen Bewegungsspielraum hat, sind die individuellen Blickpunkte auf eine Szene sehr ähnlich. Informationen über die Parallaxe unterschiedlicher Perspektiven sind redundant und können vernachlässigt werden. Aufgrund dieser Tatsache lässt sich ein Light Field nun soweit komprimieren, dass eine Rank-1 Faktorisierung ausreicht, um die Akkommodation der Augen zu gewährleisten. (Huang et al. 2015, S. 9).

Durch komplexe Algorithmen wird die zugrundeliegende Light Field Aufnahme nun faktorisiert. Pro Auge werden zwei Bilder mit unterschiedlicher Tiefenschärfe synthetisiert und auf die beiden Displays gelegt. Bilder auf dem vorderen Display sind auf Bildinhalte im Vordergrund der Szene fokussiert; Bilder auf dem hinteren Display auf Bildinhalte im Hintergrund (Huang et al. 2015, S. 5).

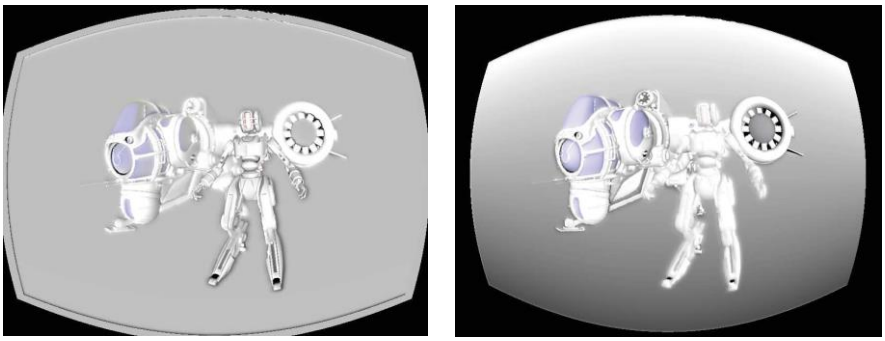

Abbildung 35: "Rank-1" Faktorisierung des Light Fields. Das Bild (links) wird dabei auf das vordere Display gelegt, das Bild (rechts) auf das hintere. Im Auge des Betrachters setzt sich multiplikativ ein Light Field zusammen, in dem auf verschiedene Tiefenebenen fokussiert werden kann (Huang et al., S. 6)

Das Light Field wird nun optisch rekonstruiert. Die Bildsynthese erfolgt dabei multiplikativ im Auge des Betrachters. Da das synthetisierte Light Field die gesamte Eyebox umspannt, ist kein Eyetracking notwendig (Huang et al. 2015, S. 3). Das Light Field Stereoscope erlaubt die Akkommodation der Augen auf einen größeren Bereich und einen natürlicheren Retinal Blur als herkömmliche Multi-Focal-Plane Displays (Huang et al. 2015, S. 9).

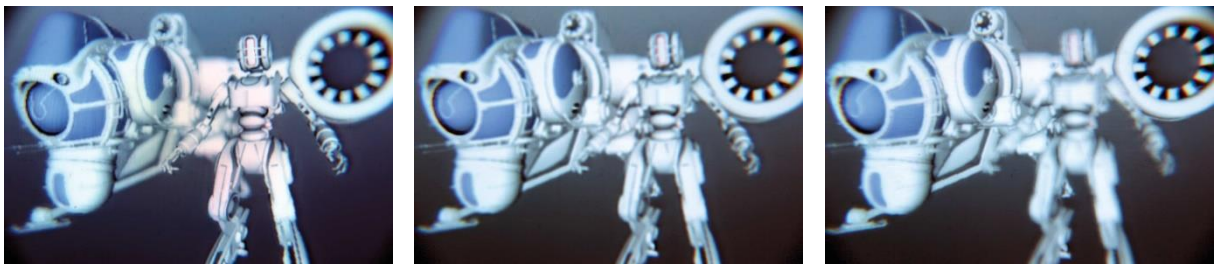

Abbildung 36: CGI Beispielszene, die mit einer Fotokamera durch das Light Field Sterescope aufgenommen wurde. Die Fotokamera kann frei auf unterschiedliche Bildinhalte fokussieren (Huang et al. 2015, S. 6)

Der Prototyp des Light Field Stereoscopes besteht aus zwei LCD-Panels mit einer Auflösung von 1280 x 800 Pixel. Die Panels sind mit einem Abstand von 3,8 cm und 4,8 cm vor der Hauptlinse im Display verbaut. Die virtuellen Bilder und somit der Akkommodationsbereich, in dem auf Bildinhalte fokussiert werden kann, entstehen in einem Abstand von 19 cm und 123 cm vor der Hauptlinse (Huang et al. 2015, S. 5–6). Das Field of View des Light Field Stereoscopes beträgt 87° x 91° (Huang et al. 2015, S. 3).

Ein Nachteil des Light Field Stereoscopes ist die relativ geringe Bildwiederholrate. Für die Faktorisierung eines einzelnen Bilds müssen 5 x 5 x 2 Bilder aus dem Light Field generiert werden (Huang et al. 2015, S. 10). Dadurch ergibt sich für den Prototyp eine End-to-End Latenz von 65 ms und damit eine Bildwiederholrate von lediglich ca. 15 Bildern pro Sekunde. Die kritische Flickerfrequenz der menschlichen Wahrnehmung liegt dagegen bei ca. 60 Hz. Für eine angehnehme VR Erfahrung sollte die Latenz folglich nicht mehr als 17 ms betragen. (Huang et al. 2015, S. 3)

4.5 Weitere Anwendungsgebiete von Light Fields für VR

Verschiedene Anbieter arbeiten also aktuell an effektiven Lösungen für die Aufnahme, Rendering und Streaming von Light Field Daten sowie an einer verbesserten Darstellung von Virtual Reality Inhalten durch Light Field Displays. Andere Unternehmen wiederum haben das Potenzial von Light Fields für Virtual Reality bereits erkannt und integrieren Light Field Technologie in bestehende Produktionsprozesse.

Die Renderingcompany OTOY kündigte im Oktober 2014 in Zusammenarbeit mit Warner Brothers eine Neuverfilmung der Kultserie „Batman: The Animated Series“ für Virtual Reality an (Otoy inc. 2014). Durch den Einsatz von Light Fields soll die Welt von Batman in Form eines holographischen Videos zu neuem Leben erweckt werden. Dabei lassen sich klassische Orte der Originalserie frei erkunden und somit neu erleben. Diese Kooperation markiert den ersten kommerziellen Einsatz von OTOYs Light Field Rendering, um große, komplett navigierbare Welten für VR-Umgebungen zu erzeugen (Otoy inc. 2015f).

> *“Being able to climb inside the Batman: The Animated Series universe, to experience that world as if you’re living in it, is absolutely incredible. Today’s demonstration validates the tremendous storytelling opportunities we have with VR and OTOY’s light field rendering technology“*
>
> (Sam Register (2015) President of Animation and WarnerDigital Series,
> Zitiert nach Otoy inc. (2015f))

Die Firma NextVR ist auf das Live-Streaming von VR Inhalten spezialisiert und Vorreiter in der Übertragung von großen Live-Events, wie Konzerten oder Sportereignissen in Form von 360° Videos. Das Unternehmen gab im März 2015 bekannt, zukünftig das Capturing und Live-

Streaming von Light Fields in ihr bestehendes VR-System integrieren zu wollen. Dadurch soll eine größere Bewegungsfreiheit innerhalb der aufgezeichneten Szenen und demzufolge eine stärkere Immersion ermöglicht werden (Hyaden 2015). Benutzer der Inhalte von NextVR finden sich direkt am Spielfeldrand von großen Sportevents oder in der Menge eines Konzerts wieder. Solche immersiven Live-Inhalte bieten daher eine interessante Alternative zur herkömmlichen Aufzeichnung von Live-Events und eröffnen einen interessanten Markt. NextVR ist es im August 2016 gelungen, Investitionen in der Höhe von 80 Millionen US-Dollar für die Forschung an neuen Technologien und eine Ausweitung ihres Markts zu erhalten (Springer 2016).

Das Unternehmen Uncorporeal verspricht ein „360° Reality Capture“ durch Light Field Technologie. Durch ein eigenes Light Stage System mit 240 synchronisierten Kameras ist das Unternehmen in der Lage, dynamische 360° Light Field Volumen aufzuzeichnen. Aus den Bildern generieren sie nach eigenen Angaben so genannte 10D-VR Light Fields:

> *„3D-Spatial + 1D Temporal + 6D Degrees of Freedom = 10D VR“*
>
> (Uncorporeal Systems Inc. 2016)

Ein 10D-VR Light Field lässt sich nun in 3D-Geometrie und einzelne Texturen decodieren. Durch eine Einbindung in herkömmliche Game-Engines, wie Unity 3D oder Unreal Engine 4, können aufgenommene Personen nun in Videospielumgebungen integriert und dabei in Echtzeit neu beleuchtet und anschließend gerendert werden. (Uncorporeal Systems Inc. 2016)

Diese Integration in eine Game Engine erlaubt nun Real-Time Rendering und somit Positionstracking und Kopfbewegungen bei voller Parallaxe. Live-Action Content kann also beliebig in neue Inhalte und Umgebungen eingefügt und als völlig navigierbare VR-Erfahrung betrachtet werden (Rubin 2015). Uncorporeal ermöglicht auf diese Weise das fotorealistische Rendering von Menschen mit Live-Performance für VR-Umgebungen in Echtzeit.

5 Fazit

5.1 Zusammenfassung

> *„The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked.“*
>
> (Sutherland 1965, S. 508)

Mit diesem Zitat beschreibt Ivan E. Sutherland schon 1965 den Zustand völliger Immersion in einem VR-System. Seither hat sich auf dem Gebiet der virtuellen Realität eine Menge getan. Gerade die technischen Entwicklungen der letzten fünf Jahre haben VR-Systeme hervorgebracht, von denen Sutherland nur träumen konnte. Trotzdem ist der Zustand völliger Immersion noch nicht erreicht und die perfekte virtuelle Realität noch immer eine Utopie. Der Einsatz von Light Field Technologie für Virtual Reality kann dabei helfen, eine virtuelle Realität im Sinne Sutherlands zu erschaffen.

Das zweite Kapitel dieser Arbeit bietet einen allgemeinen Überblick über das Thema Virtual Reality. Wichtige Begriffe wie Präsenz, Immersion und typische Merkmale wie Head-Mounted Displays, Trackingsensoren und Eingabegeräte wurden vorgestellt. Dabei ist es schwer, eine einheitliche Definition für Virtual Reality zu formulieren. VR-Anwendungen stellen unterschiedliche Ansprüche an die Abbildung der Realität und es gibt verschiedene VR-Systeme, mit unterschiedlichen Eingabegeräten in diversen Ausbaustufen. Eine perfekte virtuelle Realität, in der echte und virtuelle Welt nicht mehr voneinander unterschieden werden können, würde die perfekte Simulation aller menschlichen Sinnesreize erfordern.

Wichtige Aspekte der visuellen Wahrnehmung wurden behandelt. Die Wahrnehmung im Raum basiert hauptsächlich auf verschiedenen Tiefenhinweisen. Besonders wichtig für eine VR-Anwendung sind vor allem das Stereosehen, die Konvergenz und Akkommodation der Augen sowie die Bewegungsparallaxe.

Nach einem Rückblick auf die Entstehungsgeschichte des Head-Mounted Displays, werden VR-Systeme der aktuellen Generation behandelt und miteinander verglichen. Die Vorreiter Oculus Rift und HTC Vive sowie Mobile Displays wie die Samsung Gear VR spiegeln den aktuellen Stand der Technik wider, an dem sich zukünftige Displays messen müssen.

Virtual Reality Inhalte lassen sich in Pre-Rendered Content und Realtime-Rendered Content unterscheiden. Pre-Rendered Content, wie 360° Videos und Live-Action VR erlauben dabei lediglich die Rotation des Kopfes; Bewegungsparallaxe kann nicht dargestellt werden. Aufgrund dieser Tatsache lässt sich darüber diskutieren, ob 360° Videos überhaupt eine Virtual Reality Anwendung im eigentlichen Sinne darstellen. Darüber hinaus unterstützen aktuelle Head-Mounted Displays nur eine kleine Anzahl wichtiger Tiefenhinweise. Der Vergence-Fokus Konflikt und fehlender Retinal Blur können bei Benutzern erhebliche Nebenwirkungen hervorrufen. Zusammenfassend lassen sich also zwei zentrale Probleme heutiger Virtual Reality Systeme formulieren:

1. Live-Action VR unterstützt bisher kein Positionstracking und somit keine Bewegungsparallaxe.
2. Aktuelle Head-Mounted Display unterstützen keine Akkommodation der Augen. Der Vergence-Fokus Konflikt führt zu einer unnatürlichen Raumwahrnehmung.

Im dritten Kapitel dieser Arbeit wird aufgezeigt, wie der Einsatz von Light Field Technologie für Virtual Reality helfen kann, genau diese Probleme zu überwinden. Die plenoptische Funktion beschreibt ein Light Field als Volumen, in dem die Intensität und Richtung von jedem Lichtstrahl bekannt ist, in sieben Dimensionen. Durch die Reduzierung auf eine statische Szene in freiem Raum kann die plenoptische Funktion auf vier Dimensionen reduziert werden. Durch die Parametrisierung als Light Slab kann das Lichtfeld nun mit einer Kamera gemessen werden. Verschiedene Aufnahmeverfahren eines Light Fields wurden erörtert und miteinander verglichen. Light Fields können entweder durch eine einzelne Kamera, ein Kamera-Array oder mit einer plenoptischen Kamera aufgezeichnet werden. Daraufhin wurden typische Light Field Anwendungen analysiert. Zusammenfassend kann gesagt werden:

1. Durch View-Rendering können in Echtzeit neue, perspektivisch korrekte Blickpunkte aus einem Light Field synthetisiert werden.
2. Durch Synthetic-Aperture Anwendungen kann sowohl die Tiefenschärfe, als auch die Schärfeebene einer Light Field Aufnahme in der Postproduktion verändert werden.
3. Aufgrund zusätzlicher Tiefeninformationen bieten Light Field Aufnahmen große Vorteile in Bezug auf Visual-Effects und CGI-Integration.

Nutzt man Light Field Technologie nun für Virtual Reality Anwendungen, können dadurch zentrale Probleme heutiger VR-Systeme gelöst werden. Durch View-Rendering in einem 360° Light Field Volumen können perspektivisch korrekte Bilder für die Wiedergabe in einem Head-Mounted Display in Echtzeit erzeugt werden. Für Stereoverfahren können ohne Probleme auch zwei Bilder mit gewünschter Disparität aus einem Light Field synthetisiert werden. Daraus ergeben sich folgende Aussagen:

1. Light Field Technologie ermöglicht Positionstracking und folglich Bewegungsparallaxe in Live-Action VR.
2. Durch Light Field Technologie kann die Stereoskopie in Live-Action VR verbessert werden.

Durch den Einsatz von Light Fields lassen sich außerdem Probleme heutiger Head-Mounted Displays, allen voran der Vergence-Fokus Konflikt, adressieren. Im Rahmen dieser Arbeit wurden verschiedene Light Field Displays vorgestellt und miteinander verglichen. Zusammenfassend können folgende Aussagen getroffen werden:

1. Light Field Displays erlauben eine natürliche Akkommodation und Konvergenz der Augen. Unterschiedliche Bildinhalte können frei fokussiert werden.
2. Light Field Displays sind in der Lage, einen natürlichen Retinal-Blur zu erzeugen.
3. Durch den Einsatz eines Mikrolinsen-Arrays lassen sich leichte Displays mit einem geringen Form-Faktor realisieren.

Darüber hinaus erlaubt der Einsatz von Light Stage Technologie ein detailliertes Scanning und die 3D-Rekonstruktion von Personen. Light Field Aufnahmen von Schauspielern und deren Live-Performance kann in Virtual Reality Anwendungen integriert werden.

Mit OTOY und Lytro arbeiten aktuell gleich zwei Unternehmen an einer kompletten Light Field Produktionspipeline für Virtual Reality. Dabei entwickeln sie neben Aufnahmesystemen für 360° Light Fields auch Lösungen für eine effektive Kompression und Speicherung der Bilddaten, Tools für die Bearbeitung von Light Field Videos sowie eigene Light Field Player und Streaming Plattformen.

Im Rahmen dieser Bachelorarbeit konnte keine dieser Anwendungen einem praktischen Test unterzogen werden. Beide Systeme befinden sich noch in der Entwicklung und sind nicht öffentlich zugänglich. Eine praktische Analyse und Auswertung der einzelnen Systeme ist daher zukünftigen Arbeiten vorbehalten.

Allgemeine Probleme der Licht Feld Technik sind aber vor allem die anfallenden Datenraten, die gerade bei dynamischen Light Field Aufnahmen nochmal um den Faktor Zeit erweitert werden. Damit Light Field Technologie in kommerziellen Anwendungen eingesetzt werden kann, muss eine effektive Kompression und Speicherung der Bilddaten gewährleistet sowie eine hohe Bandbreite zur Verfügung gestellt werden. OTOYs eigenes ORBX Dateiformat, der Einsatz von Cloudcomputing und das integrierte Serversystem der Lytro Immerge sind hierfür sicherlich ein erster wichtiger Schritt.

Light Field Displays sind durch ihre Verbesserungen in Bezug auf den Vergence-Fokus Konflikt und einen geringen Form-Faktor möglicherweise das VR-Display der Zukunft. Aktuell sind sie allerdings noch nicht für einen kommerziellen Einsatz als Head-Mounted Display geeignet. Wichtige technische Aspekte wie Bildauflösung, Field of View und Bildwiederholrate werden den Ansprüchen an ein aktuelles VR-System nicht gerecht. Weder das Light Field Stereoscope, noch das Near-Eye Light Field Display erreichen auch nur ansatzweise die technischen Spezifikationen von Oculus Rift oder HTC Vive. Weitere Forschung im Bereich der Displaytechnik sowie höher aufgelöste Mikrolinsen-Arrays sind nötig, um diese Probleme in Zukunft zu beheben.

5.2 Ausblick

Light Field Technologie hat das Potenzial, die Entwicklung des noch relativ jungen Mediums Virtual Reality auf lange Sicht zu prägen. Ob sich Light Field Technologie durchsetzt, ist hauptsächlich eine Frage der technischen Entwicklungen der nächsten Jahre. 360° Light Field Aufnahmesysteme und Light Field Displays müssen noch einige technische Hürden überwinden, um sich auf dem Markt etablieren zu können und bei Benutzern Anklang zu finden.

Den Consumer Light Field Kameras von Lytro ist dieser Schritt bis heute nicht gelungen. Trotz der technischen Finesse sehen Benutzer in ihnen meist nicht mehr als eine Spielerei. Die Möglichkeit, Bilder im Nachhinein zu fokussieren, kann eine geringe Bildauflösung und eine verminderte Bildqualität nur schwer aufwiegen. Die Lytro Immerge ist somit auch für Lytro als Unternehmen ein Schritt in eine neue Richtung; weg vom Endkunden und hin zu einer professionellen Produktionsanwendung. Im Gegensatz zu den Light Field Kameras nutzt die Immerge aber das Potenzial der Light Field Technologie voll aus (Rosenthal 2016).

Für John Karafin hingegen ist die Frage nicht ob, sondern lediglich wann sich Light Field Technologie für Virtual Reality durchsetzt. Seine eigene Antwort: Wahrscheinlich schneller als

wir denken (Karafin 2015, Min. 3). Er sieht in Light Fields nicht weniger als die Zukunft der Bildaufzeichnung. Seiner Meinung nach hat sich die Fotografie seit ihrer Entdeckung Anfang des 19. Jahrhunderts zwar stetig entwickelt, die zugrundeliegenden Prinzipien aber sind immer noch dieselben. Ein Foto ist eine 2D-Repräsentation der Welt, abgebildet in einem Rechteck. Ein Light Field hingegen ist nicht auf zwei Dimensionen limitiert, viel mehr fängt es das Volumen einer Szene ein und eröffnet eine Vielzahl an technischen Möglichkeiten (Karafin 2015, Min. 2–3). Jules Urbach (2015, Min. 22) ist der Meinung, dass Light Field Anwendungen und Displays schon in den nächsten zwei Jahren massentauglich werden und Light Field Streaming gewöhnliche 2D-Inhalte auf Youtube oder Facebook ersetzten könnten.

Zusammenfassend kann gesagt werden, dass Light Field Technologie die Präsenz und Immersion in einer Virtual Reality Anwendung erheblich verstärken kann. Darüber hinaus eröffnen Positionstracking und Bewegung in Live-Action VR neue Möglichkeiten für die Erstellung narrativer Inhalte. Verglichen mit den Voraussetzungen und Anforderungen an ein VR-System aus dem ersten Kapitel dieser Arbeit, ist die Integration von Light Field Technologie ein wichtiger Schritt in Richtung einer perfekten virtuellen Realität.

6 Literaturverzeichnis

Adelson, Edward H.; Bergen, James R. (1991): The Plenoptic Function and the Elements of Early Vision. In: Michael S. Landy und J. Anthony Movshon (Hg.): Computational models of visual processing. 2. Aufl. Cambridge, Mass.: MIT Press, S. 3–20.

Bolas, Mark (2015): Light Field Imaging: The Future of VR-AR-MR- Part 2 [Videodatei]. In: VES - Visual Effects Society. Los Angeles. USC Institute for Creative Technologies. Online verfügbar unter https://www.youtube.com/watch?v=ftZd6h-RaHE, zuletzt geprüft am 28.08.2016.

Chochrane, Andrew (2016): Creating Narratives For Virtual Reality [Videodatei]. FMX 2016 - Blending Realities. Stuttgart, Germany. Online verfügbar unter http://www.fmx.de/program/fmx-2016-streaming/fmx-2016-livestream-archive/, zuletzt geprüft am 12.07.2016.

Cruz-Neira, Carolina; Sandin, Daniel J.; DeFanti, Thomas A.; Kenyon, Robert V.; Hart, John C. (1992): The CAVE. Audio visual experience automatic virtual environment. In: *Communications of the ACM* 35 (6), S. 64–72. DOI: 10.1145/129888.129892.

Davis, Abe; Levoy, Marc; Durand, Fredo (2012): Unstructured Light Fields. In: *Computer Graphics Forum* 31, S. 305–314. DOI: 10.1111/j.1467-8659.2012.03009.x.

Debevec, Paul (2012): The Light Stages and Their Applications to Photoreal Digital Actors. In: SIGGRAPH Asia 2012 Technical Briefs, Bd. 2. SIGGRAPH Asia 2012. New York, NY: ACM.

Debevec, Paul (2015): Light Field Imaging: The Future of VR-AR-MR- Part 1 [Videodatei]. In: VES - Visual Effects Society. Los Angeles. USC Institute for Creative Technologies. Online verfügbar unter https://www.youtube.com/watch?v=Raw-VVmaXbg, zuletzt geprüft am 28.08.2016.

Digi-Capital (2016): The reality of AR/VR business models. Online verfügbar unter http://www.digi-capital.com/news/2016/04/the-reality-of-120-billion-arvr-business-models/#.V7xMpa1glyU, zuletzt geprüft am 23.08.2016.

Dörner, Ralf; Broll, Wolfgang; Grimm, Paul; Jung, Bernhard (Hg.) (2013): Virtual und Augmented Reality (VR/AR). Grundlagen und Methoden der Virtuellen und Augmentierten Realität. Berlin: Springer Vieweg (eXamen.press).

Dörner, Ralf; Broll, Wolfgang; Grimm, Paul; Jung, Bernhard (2016): Virtual Reality und Augmented Reality (VR/AR). In: *Informatik Spektrum* 39 (1), S. 30–37. DOI: 10.1007/s00287-014-0838-9.

Gershun, A. (1939): The Light Field. In: *Journal of Mathematics and Physics* 18 (1-4), S. 51–151.

Hahne, Christoph (2012): The Standard Plenoptic Camera. Online verfügbar unter http://www.plenoptic.info/pages/refocusing.html, zuletzt geprüft am 03.08.2016.

Hertel, Yannic (2015): Die Geschichte der virtuellen Realität. Online verfügbar unter http://www.vrnerds.de/die-geschichte-der-virtuellen-realitaet/, zuletzt geprüft am 18.07.2016.

HTC Vive mit Steam VR (2016). Online verfügbar unter https://vrodo.de/htc-vive-steam-vr/, zuletzt geprüft am 20.07.2016.

Huang, Fu-Chung; Luebke, David; Wetzstein, Gordon (2015): The Light Field Stereoscope. Immersive Computer Graphics via Factored Near-Eye Light Field Displays with Focus Cues. In: Prooceedings of SIGGRAPH 15 -ACM SIGGRAPH 2015 Emerging Technologies. Los Angeles, California, S. 1.

Hyaden, Scott (2015): NextVR Adds Light Field Tech for Pseudo-positional Tracking in Live-action VR Video. Road to VR. Online verfügbar unter http://www.roadtovr.com/nextvr-light-field-vr-video/, zuletzt geprüft am 22.08.2016.

Karafin, John (2015): Light Field Imaging: The Future of VR-AR-MR- Part 4 [Videodatei]. In: VES - Visual Effects Society. Los Angeles. USC Institute for Creative Technologies. Online verfügbar unter https://www.youtube.com/watch?v=_PVok9nUxME, zuletzt geprüft am 28.08.2016.

La Orta, Mauricio de (2016): Virtual Reality. How to Experience and Create Amazing VR Content. 1. Aufl.: Eduvea.

Lanman, Douglas; Luebke, David (2013): Near-Eye Light Field Displays. In: Proceedings of SIGGRAPH 13 - ACM SIGGRAPH 2013 Emerging Technologies: ACM New York.

Levoy, Marc (2006): Light Fields and Computational Imaging. In: *Computer* 39 (8), S. 46–55. DOI: 10.1109/MC.2006.270.

Levoy, Marc; Hanrahan, Pat (1996): Light Field Rendering. In: John Fujii (Hg.): Proceedings of SIGGRAPH 96. New York, NY, USA: ACM, S. 31–42.

Levoy, Marc; Ng, Ren; Adams, Andrew; Footer, Matthew; Horowitz, Mark (2006): Light Field Microscopy. In: John Finnegan und Julie Dorsey (Hg.): Proceedings of SIGGRAPH 06 - ACM SIGGRAPH 2006 Papers. Boston, Massachusetts, S. 924–934.

Lytro Inc. (2014a): About Lytro. Online verfügbar unter https://www.lytro.com/about, zuletzt geprüft am 22.08.2016.

Lytro Inc. (2014b): Lytro Cinema. The Ultimate Creative Tool. Online verfügbar unter https://www.lytro.com/cinema, zuletzt geprüft am 22.08.2016.

Lytro Inc. (2015): Lytro Immerge: The World's First Professional Light Field Solution for Cinematic VR. Online verfügbar unter https://www.lytro.com/press/releases/lytro-immerge-the-worlds-first-professional-light-field-solution-for-cinematic-vr, zuletzt geprüft am 22.08.2016.

McMillan, Leonard; Bishop, Gary (1995): Plenoptic Modeling. In: Susan G. Mair und Robert Cook (Hg.): Proceedings of SIGGRAPH 95. New York, NY, USA: ACM, S. 39–46.

Ng, Ren (2006): Digital Light Field Photography. Doktorarbeit. Stanford Universität, Stanford, California.

Ng, Ren; Levoy, Marc; Brédif, Mathieu; Duval, Gene; Horowitz, Mark; Hanrahan, Pat (2005): Light Field Photography with a Hand-held Plenoptic Camera. In: Computer Science Technical Report CSTR, Bd. 2, S. 1–11.

Nichols, Christopher; Baumann Larsen, Kim; Denton, Justin (2015): Guide To VR. Understanding and creating pre-rendered Virtual Reality content in V-Ray. 1. Aufl. Chaosgroup. Online verfügbar unter http://www.chaosgroup.com/public_images/VRguide_v1.pdf, zuletzt geprüft am 28.08.2016.

Oculus Rift / Facebook-VR (2015). Online verfügbar unter https://vrodo.de/oculus-rift/, zuletzt geprüft am 19.07.2016.

Otoy inc. (2013): ORBX 2 Technical Introduction. Online verfügbar unter https://aws.otoy.com/docs/ORBX2_Whitepaper.pdf, zuletzt geprüft am 22.08.2016.

Otoy inc. (2014): WB, DC Entertainment and OTOY take fans inside Batman: The Animated Series. Online verfügbar unter https://home.otoy.com/warner-bros-dc-entertainment-and-otoy-take-fans-inside-the-acclaimed-batman-the-animated-series-in-industrys-first-commercial-holographic-video-deployment-for-virtual-reality-and-ligh/, zuletzt geprüft am 22.08.2016.

Otoy inc. (2015a): Lighstage. Online verfügbar unter https://home.otoy.com/capture/lightstage/, zuletzt geprüft am 25.08.2016.

Otoy inc. (2015b): Light-Fields. Online verfügbar unter https://home.otoy.com/render/light-fields/, zuletzt geprüft am 22.08.2016.

Otoy inc. (2015c): Octane Render. Online verfügbar unter https://home.otoy.com/render/octane-render/, zuletzt geprüft am 22.08.2016.

Otoy inc. (2015d): OTOY demonstrates first ever light field capture of a real world environment. Online verfügbar unter https://home.otoy.com/otoy-demonstrates-first-ever-light-field-capture-for-vr/, zuletzt geprüft am 22.08.2016.

Otoy inc. (2015e): OTOY unveils OctaneVR. Online verfügbar unter (https://home.otoy.com/otoy-unveils-octanevr/), zuletzt geprüft am 22.08.2016.

Otoy inc. (2015f): Preview of upcoming Batman: The Animated Series VR experience. Online verfügbar unter https://home.otoy.com/preview-upcoming-batman-animated-series-vr-experience/.

Otoy inc. (2015g): Render The Metaverse. A Virtual Reality Art Competition. Online verfügbar unter https://home.otoy.com/vr-competition/, zuletzt geprüft am 22.08.2016.

Rosenthal, Jason (2016): War Stories: Why I Lit Up Lytro. Lytro Inc. Online verfügbar unter https://backchannel.com/war-stories-why-i-lit-up-lytro-b46124da32a6#.558zvefrn, zuletzt geprüft am 22.08.2016.

Rubin, Peter (2015): These FX Wizards May Have Solved VR Films' Biggest Problem. Wired. Online verfügbar unter http://www.wired.com/2015/10/uncorporeal-vr-movies/, zuletzt geprüft am 22.08.2016.

Schäfer, P.; Foessel, S.; Zilly, F.; Ziegler, M.; Schöberl, M.; Keinert, J. (2013): Light-field acquisition system allowing camera viewpoint and depth of field compositing in post-production. In: International Broadcasting Convention (IBC) 2013 Conference. Amsterdam, Netherlands, S. 1–8.

Schreer, Oliver (2005): Stereoanalyse und Bildsynthese. Berlin, Heidelberg: Springer-Verlag.

Slater, Mel; Wilbur, Sylvia (1997): A Framework for Immersive Virtual Environments (FIVE). Speculations on the Role of Presence in Virtual Environments. In: *Presence: Teleoperators and Virtual Environments* 6 (6), S. 603–616.

Springer, Christoph (2015): Finale Version der Samsung Gear VR im Test. Hg. v. vrnerds.de. VR Nerds Solutions GmbH. Hamburg. Online verfügbar unter http://www.vrnerds.de/finale-version-der-samsung-gear-vr-im-test/, zuletzt geprüft am 19.07.2016.

Springer, Christoph (2016): NextVR erhält 80 Millionen US-Dollar Investitionen. VR Nerds Solutions GmbH. Online verfügbar unter http://www.vrnerds.de/nextvr-erhaelt-80-millionen-us-dollar-investitionen/, zuletzt geprüft am 22.08.2016.

Sutherland, Ivan E. (1965): The Ultimate Display. In: Proceedings of IFIP Congress, S. 506–508.

Sutherland, Ivan E. (1968): A Head-Mounted Three-Dimensional Display. In: Proceeding AFIPS '68 (Fall, part I). San Francisco, California. AFIPS, S. 757–764.

Tao, Michael W.; Hadap, Sunil; Malik, Jitendra; Ramamoorthi, Ravi (2013): Depth from Combining Defocus and Correspondence Using Light-Field Cameras. In: 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia, S. 673–680.

Uncorporeal Systems Inc. (2016): Uncorporeal. Online verfügbar unter www.uncorporeal.com, zuletzt geprüft am 22.08.2016.

Urbach, Jules (2015): Light Field Imaging: The Future of VR-AR-MR- Part 3 [Videodatei]. In: VES - Visual Effects Society. Los Angeles. USC Institute for Creative Technologies. Online verfügbar unter https://www.youtube.com/watch?v=0LLHMpbIJNA, zuletzt geprüft am 28.08.2016.

USC Institute for Creative Technologies (2015): VES - Visual Effects Society. Los Angeles. Online verfügbar unter https://www.visualeffectssociety.com/events/event/event-light-field-imaging-future-vr-ar-mr-los-angeles, zuletzt geprüft am 05.070.2016.

VR Brillen Vergleich (2016). Hamburg. Online verfügbar unter http://www.vrnerds.de/vr-brillen-vergleich/, zuletzt aktualisiert am 2016, zuletzt geprüft am 19.07.2016.