Stuttgart Media University

BACHELOR'S THESIS

Moving Light Fields

Applications and Methods to integrate Light Fields in Film and TV Production Workflows

Submitted to the department of Electronic Media
in partial fulfillment of the requirements for the degree Bachelor of Engineering

by

Andreas Engelhardt

Student ID: 23956

Supervision by **Prof. Katja Schmid**

and **Daniele Siragusano**

Stefan Müller

Student ID: 23988

Supervision by **Prof. Katja Schmid**

and **Ron Ganbar**

Stuttgart, June 2015

Eidesstattliche Erklärung

Name:	Vorname:
Engelhardt	Andreas
Matrikel-Nr.: 23956	Studiengang: AMB

Hiermit versichere ich, Andreas Engelhardt, an Eides statt, dass ich die vorliegende Bachelorarbeit mit dem Titel *Moving Light Fields - Applications and Methods to integrate Light Fields in Film and TV Production Workflows* selbständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinne 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.

Ort, Datum Unterschrift

Eidesstattliche Erklärung

Name:	Vorname:
Müller	Stefan
Matrikel-Nr.: 23988	Studiengang: AMB

Hiermit versichere ich, Stefan Müller, an Eides statt, dass ich die vorliegende Bachelorarbeit mit dem Titel *Moving Light Fields - Applications and Methods to integrate Light Fields in Film and TV Production Workflows* selbständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinne 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.

Ort, Datum   Unterschrift

Abstract English

Light field technology is increasing in popularity due to end user availability, emerging technologies and demand for life-like content. Light fields are not only allowing to move viewpoints, but furthermore have the potential to start a movement in content production and presentation. The objective of this work is to clarify light field backgrounds and possible applications as well as resulting opportunities in the context of todays film and TV productions. Mainly underpinned by scientific research papers and literature, the first part is devoted to the theoretical approach, development and application of light field technology. In the second part of this work, initially requirements are phrased, insights on workflows are given and test scenarios are conducted and described. Further on, models for integration of light fields in current processes and proposition for procedures are presented. Concluding, the possibilities and limitations are analysed and future trends are portrayed.

Abstract German

Lichtfeldtechnologie erfreut sich dank zunehmender Verfügbarkeit und immer ausgereifteren Technologien steigender Beliebtheit. Lichtfelder bieten nicht nur die Möglichkeit Bilder im Nachhinein zu bewegen, wie etwa durch Perspektivänderung, sondern haben das Potential, Grundlage einer neuen Bewegung hinsichtlich Inhaltserzeugung und -präsentation zu werden. Das Ziel dieser Arbeit ist es, die technischen Hintergründe und Anwendungsmöglichkeiten der Lichtfeldtechnologie im Kontext aktueller Film- und TV-Produktionen aufzuzeigen. Der erste Teil widmet sich der Entwicklung und der Anwendung von Lichtfeldern, die anhand von Forschungsarbeiten und Literatur dargestellt werden. Im zweiten Teil der Arbeit werden Anforderungen formuliert, Einsichten in vorhandene Arbeitsweisen gegeben und die umgesetzten Testszenarien beschrieben. Des Weiteren werden Ideen zu Integrationsmöglichkeiten von Lichtfeldern und mögliche Arbeitsweisen mit Lichtfelddaten dargestellt. Zusammenfassend widmet sich diese Arbeit der Analyse von möglichen Gewinnen und Einschränkungen durch die Technologie, sowie zukünftigen Entwicklungsmöglichkeiten.

About the Works Structure and Partitioning

Due to the wide range of topics that are related to the idea of a workflow description this work is a joint project of two authors. Andreas Engelhardt worked on the content in Chapters 3, 4, 5 and 7, Stefan Müller contributed chapters 2, 6 and 8. In Andreas Engelhardt's chapter 4, Stefan Müller contributed the sections 4.1 and 4.4, also section 7.3 for the 7th chapter. Andreas Engelhardt supported Stefan Müller with the subsections 6.2.3, 6.3.3, 6.3.4, 6.4.2, 8.3.3, 8.3.4 and 8.3.5 in chapters 6 and 8. The responsibility for chapters 1 and 9 was equally divided, Andreas Engelhardt contributed section 1.2, 1.3 and 9.1, whereas Stefan Müller contributed sections 1.1 and 9.2.

Section is defined as a subchapter of the main chapter, formatwise 1.1, whereas *subsection* is defined as the subchapter of a section, formatwise 1.1.1.

The following table shows the responsibilities for the main chapters, as well as the contributed sections and subsections.

Chapter	**Author**
Abstract	Stefan Müller
1.1	Stefan Müller
1.2	Andreas Engelhardt
1.3	Andreas Engelhardt
2	Stefan Müller
3	Andreas Engelhardt
4	Andreas Engelhardt
4.1	Stefan Müller
4.4	Stefan Müller
5	Andreas Engelhardt
6	Stefan Müller
6.2.3	Andreas Engelhardt
6.3.3	Andreas Engelhardt
6.3.4	Andreas Engelhardt
6.4.2	Andreas Engelhardt
7	Andreas Engelhardt
7.3	Stefan Müller
8	Stefan Müller
8.3.3	Andreas Engelhardt
8.3.4	Andreas Engelhardt
8.3.5	Andreas Engelhardt
9.1	Andreas Engelhardt
9.2	Stefan Müller

Content

Abbreviations

AC	Assistant Camera
AOV	Arbitrary Output Vectors
API	Application Programming Interface
CAD	Computer Aided Design
CCD	Charge Coupled Device
CG	Computer Graphics
CMOS	Complementary Metal Oxide Semiconductor
CPU	Central processing unit (processor)
DBD	Displaced Block Difference
DCP	Digital Cinema Package
DI	Digital Intermediate
DP	Director of Photography
DPD	Displaced pixel difference
EPI	Epipolar Plane Image
fps	Frames per second
Fraunhofer-IIS	Fraunhofer Institute for Integrated Circuits
GPU	Graphics Processing Unit
HdM	Hochschule der Medien Stuttgart (Stuttgart Media University)
HDR	High Dynamic Range
HFR	Higher Frame Rates
HMD	Head Mounted Display
IBC	International Broadcasting Convention, Amsterdam
IBL	image based lighting
Lidar	Light Detection and Ranging
LUT	Look-Up Table
MPEG	Motion Pictures Expert Group
NAB	National Association of Broadcasters (USA)
NLE	Non-Linear Editing
Pel/pixel	Picture Element
PoE	Power over Ethernet

Previs	Previsualization
RE	Requirements Engineering
RGB(a)	Red, Green, Blue, Alpha (Channel Information)
Roto	short form for Rotoscope, see glossary
S3D	Stereo 3D
SNR	Signal-to-Noise Ratio
TD	Technical Director
Techvis	Technical Previsualization
ToF	Time of Flight
VFX	Visual Effects
Voxel	Volume Element

1 Introduction

1.1 Motivation

Limited budgets, tight timetables and ever higher quality demands put lots of pressure on today's film and TV productions. Visual Effects processes are an integral part of almost any film production nowadays requiring special shooting practices to capture the data needed. The transition to digital media enabled the transfer of an increasing number of creative decisions to postproduction. For example, the creative color grading process at the finishing stage of a project grew strongly in importance enabling for ever more complex edits and image manipulation. Still the postproduction involves a lot of manual work, which makes high quality results time intensive and therefore expensive. More than ever productions face the challenge of getting more for less as complexity of production workflows increases.

Against this background, light field photography is a technology that promises new workflow possibilities as well as creative possibilities and aesthetics. As technology develops and algorithms advance, light fields could be integrated in production workflows in the near future. This thesis presents possible applications and methods to integrate light field technology in film or TV production workflows of current standards.

Current cinematic cameras only capture a 2D projection of the moving scene. With our attempt we are showing the highly practical use of having a richer scene representation while keeping in mind the quality requirements of cinematic visuals. By applying methods researched in the field of computational photography to a set of multiple views of a live-action scene, a variety of new possibilities emerge that used to be the domain of synthetic computer graphics only. For instance capturing depth-data, normal-data, the ability to set focus after the shoot, realistic relighting of scenes and the opportunity to perform minor corrections of viewing-angles, camera position and even simulating different optics and camera backs after the scene has been shot would mean huge benefits for postproduction flexibility.

Our goal is to bridge the theoretical roots of computational photography and practical every day production practices by presenting some applications in the context of selected scenarios that might ultimately increase efficiency and image quality. By doing this we hope to justify the use of multiple cameras and computational photography in order to capture a partial "lumigraph"[1] or "light field"[2] in future productions. The latter term is used here. Finally, the introduction of light fields also creates new creative possibilities and aesthetics that will be exciting to explore in the context of storytelling. An

[1] Gortler et al. (1996)

[2] Levoy and Hanrahan (1996)

unconventional focus setup that can be set independently from the performance is just one aspect of the emerging field of computational cinematography.[3]

1.2 Scope of this Work and Development Stage

This work will be limited to the integration of some basic and promising applications of light field data only in the context of current demands and practices of film and TV productions. This includes commercial spots and cinematic shorts and other short forms that aim for a high level of visual quality. Fraunhofer IIS is one of very few that specifically do research and development concerning light field with film and TV in mind. On IBC 2013 and NAB 2014 Fraunhofer IIS already presented their approach on light field with a Lego stop motion short[4] and a static scene with Playmobil toys. At these events they were requested to show real-world examples of live-action footage with human actors. The feasibility of successfully applying light field technology to the movie making process is yet to be proved. Therefore, the aim of the underlying project stage of this thesis was to deliver some live action shots that make use of several light field applications while learning about the requirements to integrate light field data in a production workflow. The central questions that guided the explorative study and creation of this work are:

What are promising applications for light field data in film and TV and how can they be integrated in a modern production pipeline incorporating VFX? What are requirements for successful integration in the context of the three scenarios backlot, packshot and portrait?

After an introduction to the basic idea of the light field and an overview about existing research and acquisition systems the first part of the thesis will describe the preproduction of the test shoot. The results of a meeting with industry experts are evaluated and overall requirements as well as possible applications of light field technology in the domain of filmmaking will be collected. The most promising applications will be selected and further defined with use cases as part of three scenarios. In the second part of the work these will be discussed in more detail as todays industry practices get described and theoretical fundamentals on workflow and pipeline development will be presented. The description of the test production will then lead to an updated set of requirements, which will be taken into account when presenting some ideas and considerations on the integration of light fields in future production workflows in the final chapters. A short

[3] the term computational cinematography as used by Rogmans (2014) refers to the use of computational photography techniques in the context of cinematography

[4] a 2D rendered sequence can be found as part of the accompanying data package

review about the project and an outlook considering future trends and possible next steps will be given at the end.

In our tests we had the chance to use the prototype of a light field camera array that was developed by the Fraunhofer IIS. Although all of our tests described in chapter 6 are based on data from this camera system, we try to present ideas that are mostly camera independent as there is not a clear standard design for acquisition systems at the time of writing. This will be discussed further as part of chapter one.

The projects development stage can be described using a classification system. "Technology Readiness Levels (TRL) are a type of measurement system used to assess the maturity level of a particular technology."[5] It describes nine stages of development from idea to market-ready. In terms of the TRL, the goal of this cooperation between IIS and HdM is to take the project from TRL 2, basic research, to 3, proving of feasibility, and prepare for level 4, technology development. The project team at HdM consisted of students from the bachelor and master program with a background in digital film technology and film language. As the video-on-demand market is expanding and new viewing devices evolve, the borders between interactive applications and the conventional cinematic narrative will definitely more and more blur or dissolve into new formats.[6] As we are limiting ourselves to applications for TV and film, the final output format in the considered workflows will be 2D or stereo 3D images. For future applications, a full 4D light field output could also be an option if there are appropriate playback devices like auto stereoscopic displays. Yang et al. presented a light field viewing environment using a cluster of video projectors in 2008, for example.[7]

Besides, Moores Law is supporting light field technology.[8] While Wilburn was limited to 80 MB/s write speed per PC when building his camera array in 2005[9], we were now limited to 1 GB/s, for example, and we will probably worry about another order of magnitude five years from now. As technological development advances and microprocessors get faster and camera sensor resolution and sensitivity increases, some problems of the early stages of light field photography are already solved.[10] Only now it seems possible to pursue the ambitious goal of bringing light field to the movies. Some problems still exist and we will only be able to present possible solutions limited by the state of technology of late 2014. We will discuss the detailed requirements shortly.

---

[5] "Technology Readiness Level" (2013), Mankins (1995) 1-5
[6] cf. McDowell, Hamdy, and Hanisch (2015), Hunter (2015)
[7] Yang et al. (2008) 8-12
[8] "Moore's Law" (2015)
[9] Wilburn (2004) 38
[10] Wilburn (2004) 42

1.3 Methodology

This chapter introduces the methods that were used to approach the subject of this work. Some definitions of important terms and concepts will be given and possible purposes of the work in the context of the overall research explained.

The cooperation in research between HdM Stuttgart and the moving pictures technology department of the Fraunhofer IIS is called a study of technology. The current goal is to find out whether it is possible to match light field technology to the demands of the film and TV industry. First of all, we tried to work out the requirements and some promising applications. Then feasibility has been evaluated through experimental testing in the form of small test shoots. Based on the outcome of the tests concepts for new workflows and equipment are developed.[11]

The aim of this thesis is to formulate detailed requirements for the integration of light field technology into media production workflows. "A *workflow* is a specific set of procedures and deliverables that defines a goal."[12] In the process of creating a pipeline high-level workflow specifications and requirements are taken to design a detailed plan, which is then implemented as the pipeline itself. "A pipeline is, therefore, a workflow specification plus a set of tools that are to be used to achieve the goals defined therein."[13] Some ideas for the design of a pipeline in terms of tools and interfaces are presented as models in chapter seven.

To describe valid requirements and low-level workflow specifications that are hopefully sufficient for system design we rely on requirements analysis.

> *Requirements analysis [or requirements engineering] is a system engineering term for the process of determining the conditions that a system needs to meet and the goals users must be able to achieve using it. The process involves taking into account the potentially conflicting needs of multiple sets of users and beneficiaries (the people who receive the deliverables from the system).*[14]

Ebert defines requirements engineering (RE) as a "[…] disciplined and systematic approach to determination, documentation, analysis, coordination, examination and management of requirements". [15]

---

[11] IIFOP 2.0 Projektskizze Technologiestudie (2014) 1-2
[12] Bugaj (2010) 784
[13] Bugaj (2010) 786-787
[14] Bugaj (2010) 787
[15] Ebert (2012) 35, translation by the author, german original: „[...] ein diszipliniertes und systematisches Vorgehen zur Ermittlung, Dokumentation, Analyse, Abstimmung, Prüfung und Verwaltung von Anforderungen“

We won't be able to complete a requirements document for a whole movie production workflow, instead we will focus on certain applications that will be specified in chapter two and three. We will characterize every included workflow step with regard to the tasks' definition, data input and output, functional description, interface design and performance expectations as well as practicability.[16] Our considerations are based on the methods and toolsets of requirement analysis but not every requirement will be in the form of the IEEE-standards[17]. Still we hope that the requirements mentioned would be defined to an extent that is needed for evaluation and implementation.

Some quality-requirements, that we found talking to industry experts, provide basic guidelines.[18] Based on the external input, personal experience and heuristic techniques as well as literature we chose several applications of light field technology. A limited number of scenarios help to describe these applications. Scenarios in general summarize requirements or can be part of requirements and usually consist of a script-like description of the initial state, the normal course of events, irregular procedures, processes that run in parallel and an end state.[19] Use Cases are employed to describe the scenarios and help defining requirements. A Use Case is a request of a system service from outside the system by means of interaction.[20] In the context of RE the test shoot can be seen as an early iteration cycle and part of a prototyping kind of process from which we derived new and generalized requirements.[21]

Flow graphs and diagrams have been used to model workflow specifications, use cases and the data flow.[22]

The next steps would be to bring our results back to a group of stakeholders or potential clients and to iterate on their feedback.[23] Some first results could be communicated as papers or conference talks. Finally, alpha versions of the proposed software and hardware could be created. These should be tested by a selected group of potential users in the respective field.

---

[16] Bugaj (2010) 787, Ebert (2012) 92-97
[17] Ebert (2012) 96
[18] Ebert (2012) 35, 72
[19] Ebert (2012) 108
[20] Ebert (2012) 109
[21] cf. Ebert (2012) 120
[22] Ebert (2012) 109, 138-146
[23] cf. Ebert (2012) 228

2 The Light Field Concept

2.1 Theory

In order to understand the complex meaning behind the abstract term *light field* a short summary about the theory and physics behind a light field is given. The „light field" is based on the plenoptic function first defined by Adelson and Bergen [1991] and Levoy and Hanrahan [1996]. Adelson and Bergen traced back to examples in Leonardo da Vinci's notebooks, while Levoy and Hanrahan followed the mathematical formulations of the total distribution back to Gershun [1939] whose early work consisted on illumination engineering in Moscow.[24] Gabriel Lippmann, professor of experimental physics at the Sorbonne, also chased the idea of capturing light field data in order to create 3D photographs in his 1908 Paper "*Photographies Integrales*" at the French Academy of Sciences.[25]

The plenoptic function describes the intensity of a light ray with arbitrary wavelength λ, for every instant of time t in every possible angle θ and φ, falling on a camera sensor or eye at any point x, y, z.[26]

$$p=P(\theta, \varphi, \lambda, x, y, z, t)$$

Therefore, the general description of a light ray in space contains seven dimensions. The cameras Position C depends on the 3D Coordinates x, y, z, the change in the scene is defined by the temporal parameter t; the angles θ, φdetermine the light rays angle of entry, as seen demonstrated in Figure 1, while the wavelength λ describes the color of the reflected light. In order to view a scene at any time from any location you would therefore need an infinite amount of cameras.[27]

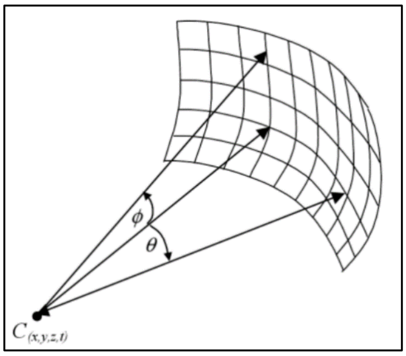


Figure 1: Representation of the plenoptic functions parameters.

[24] cf. Ng (2006) 11

[25] cf. Lippmann (1908) 446

[26] cf. Schreer (2005) 226-227

[27] cf. Schreer (2005) 226-227

A. Gershun [1936] defined the term "*light field*", describing a volume in which all light rays in all directions are known. By limiting the space to a defined time, a defined volume and the transformation of the spherical coordinates in planar image coordinates, a reduction of three dimensions is possible, leaving only 4 Dimensions.

The transformation of the image plane coordinates to sphere coordinates reads as follows:

$$\varphi=\tan(v/u),\ \theta=\tan(u/v)$$

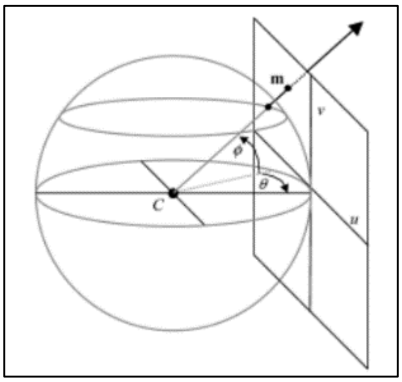


Figure 2: Transformation of the image sensors plane coordinates to the plenoptic functions sphere coordinates.

The plenoptic function can also be considered as sampling theorem for optical rays in a 3-dimensional space because the relationship of the sensors image coordinates and the sphere coordinates of the plenoptic function is simultaneously the required target image resolution of the sensor.[28]

The reduction of the plenoptic function takes place by restricting the focus only on rays travelling through free space, free of occlusions and scattering objects, eliminating the possibility of an inconstant length of light travelling along a ray.

The parametrization can take place by defining a light ray through a point on 2 planes *pq* and *rs*. The volume between both planes is defined as *light field* by Levoy and Hanrahan [1996] and as *Lumigraph* by Gortler et al. [1996]. Levoy and Hanrahan label them *uv* and *st.*[29]

A specific ray in space is hereby defined by a pair of points in the *pq* plane and in the *rs* plane, but other rays may also pass through each of the points, just not the same pair of points at once. For example a ray located

[28] cf. Schreer (2005) 229

[29] cf. Ng (2006) 12

at (2,2) in the *pq* plane and (1,2) in the *rs* plane is explicitly defined by the corresponding pair of points, but any other ray may also pass through (2,2) in the *pq* plane and through (1,2) in the *rs* plane.[30]

Because the optical rays are defined by the positions in the entrance plane (*p,q*) and the exit plane (*r,s*), a 4-dimensional plenoptic function results:

$$P_{4D} = P(p,q,r,s)$$

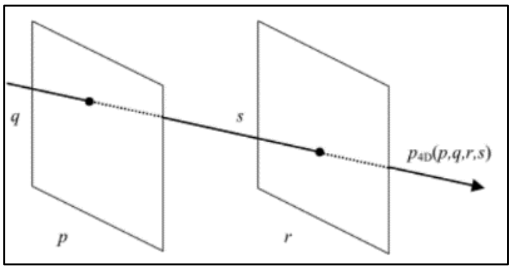


Figure 3: Optical ray through entrance plane(p,q) and exit plane(r,s).

The main reason of reducing the 7-dimensional plenoptic function to a 4-dimensional "light field" is making it possible to measure values without the need of having infinite values.[31] By reducing the plenoptic function to only 4 dimensions, it gets much easier because of the limitation of prior said factors and thus making it possible to capture light fields in an easier way.

The 4D space is sampled by taking multiple pictures with cameras mounted in the *p,q*-plane, while the virtual cameras are positioned in the *r,s*-plane allowing to compute synthetic views of every viewpoint within the planes, and therefore creating a light field.[32]

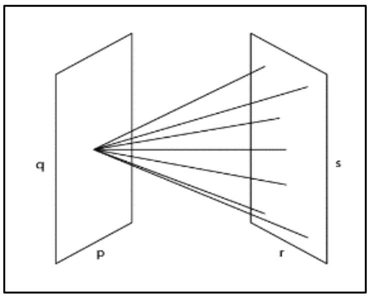


Figure 4: Optical rays from p,q-plane view, containing the cameras, every partial image is a camera view.

[30] cf. Levoy (1996) 32-34

[31] cf. Ng (2006) 13

[32] cf. Schreer (2005) 229

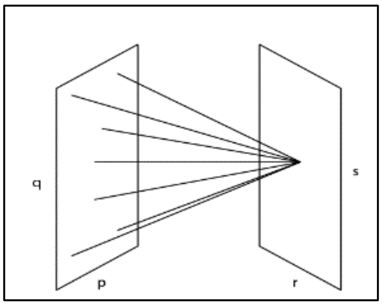


Figure 5: Optical rays from r,s-plane view, combining the partial images of the optical rays of all cameras for one point in the scene.

To generate new views from the light field all you have to do is to set the position and orientation of the camera. Through the resulting optical rays between the *pq* plane and the *rs* plane the new view can be created by copying or interpolating surrounding pixels.[33]

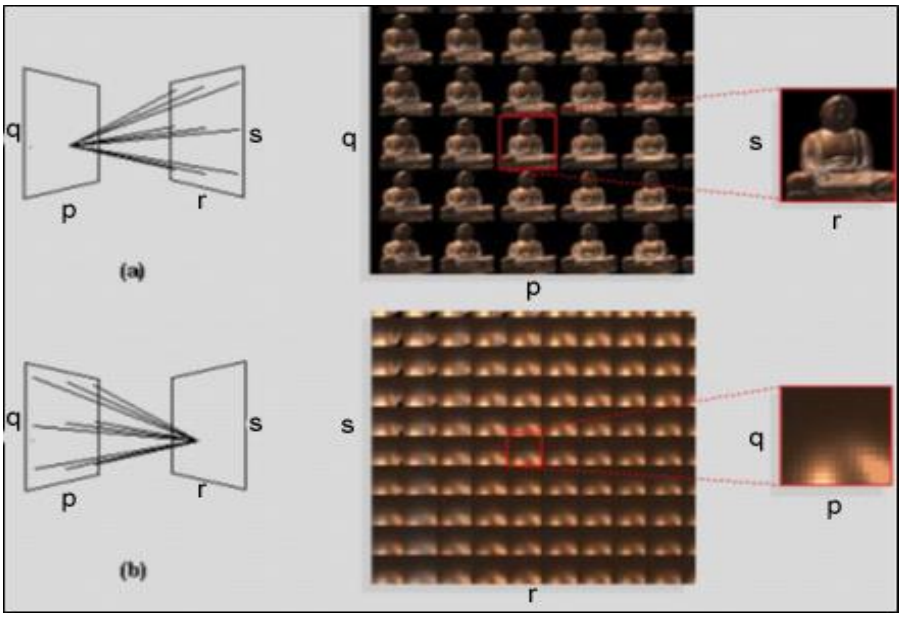


Figure 6: Two visualizations of a light field. (a) Each image in the array represents the rays arriving at one point on the pq plane from all points on the rs plane, as shown at left. (b) Each image represents the rays leaving one point on the rs plane bound for all points on the pq plane. The images in (a) are off axis (i.e. sheared) perspective views of the scene, while the images in (b) look like reflectance maps. The latter occurs because the object has been placed astride the focal plane, making sets of rays leaving points on the focal plane similar in character to sets of rays leaving points on the object.

[33] cf. Schreer (2005) 231

2.2 Limitation and Definition of the Term Light Field in this Thesis

As based on the definition in 2.1., we are not working with a complete light field of a scene, but a 4D-light field with the boundaries of the capturing camera system.

Definition by Fraunhofer IIS:

> *A light-field is defined by the number of light rays within a specific area. If all light rays are captured within a scene, this makes it possible to generate a perspective from any position, including the reconstruction of all depth of field information. Although it's not possible to capture a complete light field with current technologies, a large portion can still be acquired under special conditions. […] Researchers rely on a planar camera arrangement with up to 16 HD cameras that capture slightly different scene perspectives and from which information such as the depth of field can be extracted. This makes it possible to shift the perspective and implement special effects such as virtual camera panning and zooming, all with a single light field recording.*[34] *[Fraunhofer IIS, 2014]*

This statement by the Fraunhofer IIS sums up the basic functionality of a light field, as well as use cases and limitations of current technology.

So in general, a light field in our context can be understood as an array of images of a scene. Each image in the scene slightly differing in perspective from the next one, all of them located on a two-dimensional plane. The partial light field is extracted by comparing the information in the acquired images and therefore gathering information of the light rays leaving each point of the scene by matching light rays among all the images and their position in all the images. This data allows a four-dimensional reconstruction of the scene. By interpolating sampled rays from the 4D- light field rays for new views on the two-dimensional plane, as well as in front or behind the capturing cameras can be generated. The 4D light field therefore can be considered as a collection of two-dimensional images of a scene with the focal points of the different acquisition-cameras in another two-dimensional plane.

[34] Fraunhofer IIS (2014)

2.3 Prior Work

Prior work on image based rendering includes different technologies to capture light fields, mostly developed for still image photography. In this chapter we will give an overview about existing technologies, applications of such, advantages and disadvantages of each technology, as well as limitations.

2.3.1 Moving Cameras

The easiest and most accessible way to capture a light field is by using a common photography camera and moving it through the scene and capturing the object from many different angles, as displayed in Figure 7. The cameras position needs to be estimated in order to align all the images in the correct way. One way to do so would be mounting sensors to the camera in order to know the location at any time. Another way of estimating the cameras pose is the use of algorithms from the computer vision literature, especially optical flow algorithms.[35]

One of those Systems is the Lumigraph system by Gortler et al., it consists only of a regular digital camera which the operator uses to take pictures of an object in different angles.

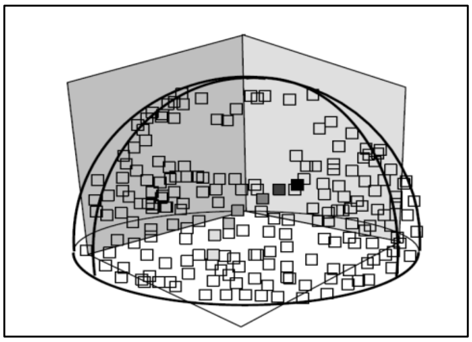


Figure 7: The user interface for the image capture stage displays the current and previous camera positions on a viewing sphere.

Since the images aren't taken along a defined path, the distance to the object always varies. Because of said variation in distance, a rebinning algorithm has to fill in missing information by down-sampling the data (pull) and then building high resolution images from the low resolution ones (push). One of the major drawbacks of this technique is that the process alters the original data, so even if you choose a perspective from a known camera position, afterwards the resulting pixel would differ from the original. Other drawbacks are a relatively low picture quality (depending on the taken images) and a huge amount of images

[35] cf. Levoy (2006) 51

are necessary in order to cover all the angles and achieve a useable result, as well as no option to capture reflective properties of the object and no possibility to capture a moving object/video. Apart from the drawbacks nothing but a standard handheld digital camera, computation of the approximate geometry and a rendering function based on the plenoptic function is necessary.[36]

2.3.1.1 Spherical Gantry

Another Moving Camera System worth mentioning is Stanford's Spherical Gantry, displayed in Figure 8, trying to reduce limitations and problems of the Lumigraph system.

Figure 8: Stanford Spherical Gantry

The Stanford Spherical Gantry was developed by the Stanford Institute (Levoy and Hanrahan) and consists of a single camera mounted on a robot arm moving in a two-dimensional plane as it captures images of an object, therefore no rebinning algorithms (opposing to Gortler et al.'s lumigraph system) are necessary as the camera only moves on a defined path. It is possible to mount various light sources on the other robot arm of the gantry to illuminate the captured object in different ways and from different directions. The object of which a light field is created sits on a central rotating platform, making it possible to capture the object from every direction.

The camera capturing the light field is basically a standard 3-CCD digital still – or video camera. Drawbacks of the gantry systems are that you are limited to small static objects and can't capture outdoor-scenes because they are simply too big, as well as the amount of time necessary to capture a dataset for an object, according to Levoy it takes an average of four hours. The main application for both moving camera systems is to capture a whole light field of an object and it's also possible to capture a detailed geometry based on image

[36] cf. Gortler (1996)

based rendering algorithms. The major drawback of both systems is the lack of application for moving objects (as stated before).[37] [38]

2.3.2 Plenoptic Cameras

Meanwhile the most commonly used technique for light field photography is the plenoptic camera. The basic principle of the plenoptic camera relies on Gabriel Lippmann's 1908 invention of integral photography. A single camera approach capturing the light field on its image plane in a single exposure. The difference to a regular camera capturing photographic images is a microlens-array placed in front of the sensor. Each microlens covers multiple sensor pixels and separates the light rays into small sub-aperture images (seen in Figure 9) with varying viewing angles captured by the sensor. The size of the individual microlenses hereby sets the sampling resolution.

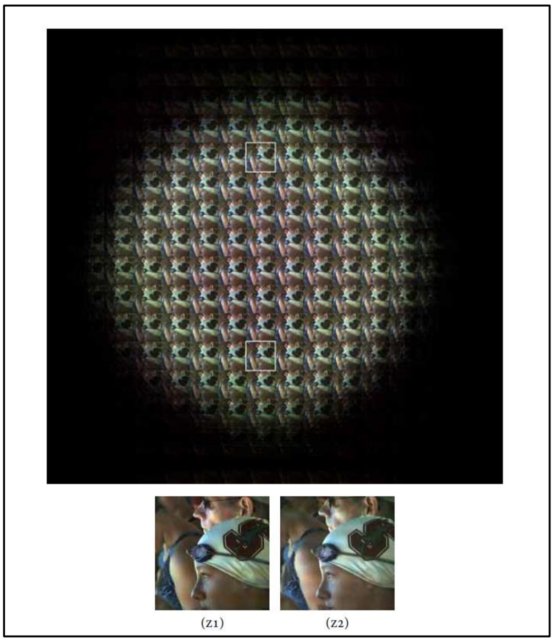


Figure 9: Sub-aperture images of the light field photograph. Images Z1 and Z2 are close-ups of the indicated regions at the top and bottom of the array, respectively.

This additional layer of microlenses makes it possible to record more information about the incoming light rays, including the light from different distances. This step is actually the key to creating a light field and to use refocussing and view rendering.[39] The light field therefore, is basically created inside the lens.

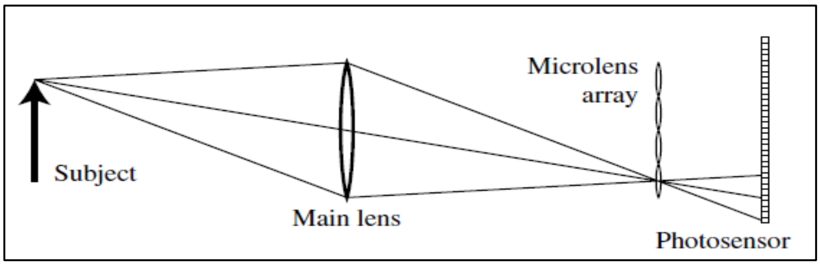


Figure 10: Conceptual schematic of a camera, which is composed of a main lens, microlens array and a photo sensor. The main lens focuses the subject onto the microlens array. The microlens array separates the converging rays into an image on the photo sensor behind it.

---

[37] cf. Levoy (2002)
[38] cf. Levoy (2006) 51
[39] cf. Griggs (2011)

The biggest concern of plenoptic cameras is its limitation in resolution. Because to capture more information about the light field, it simultaneously trades off its resolution of the output-image. “It's an inherent trade-off: Imagine the pixels of a flat image taken by a standard 10-megapixel camera spread out into a 3D light field - you capture information about a larger area, but the resolution must go down” [40]

One way out might be the use of bigger / higher resolution sensors, but bigger sensors also means it’s harder to keep the whole image in focus from the beginning, because as soon as sensor size increases, the depth of field decreases for a given aperture when filling the frame with a subject of the same size and distance. Furthermore the larger the sensor, the higher the production cost of such will be. And reading speed of sensors is currently limited, the bigger the sensor, the longer it takes to read out and process the data.

Even though there are major limitations, the industry concentrates on this capturing-technique because it’s a one camera solution and thereby user friendly. The German company Raytrix based in Kiel was the first company to offer a complete light field camera, mainly to scientific customers in 2010.[41] [42] Since 2012 Raytrix is even offering a Camera capable of capturing up to 180fps high speed video light fields, called Raytrix R5.[43]

While Raytrix is focussing on a scientific client base and specially fitted cameras for certain applications, Ren Ng, former Stanford University and one of the light field pioneers, founder of Lytro followed another mission:

“Ng set out to put his research to use by starting a company that would produce light field cameras that everyone could enjoy “[44]

The first camera by Lytro, called “LYTRO” was introduced in 2011 and for the first time offered the option of capturing light fields in an easy and affordable way for everyone.

Figure 11: First Generation of Lytro’s Light Field Camera “LYTRO”, 11 Megarays, output resolution 1080x1080 px or approx. 1 Megapixel (10% of sensor resolution).

[40] Griggs (2011)
[41] cf. Zhang (2010)
[42] cf. Sinko
[43] cf. Raytrix (2014) 3
[44] cf. Bonnington (2011)

Or as he stated in an interview:

"It's targeted to creative pioneers, the people who embraced color technology when it came out, when many people felt photography was about black and white." [45]

The cameras resolution is 11 Megarays opposing to Raytrix's R29 offering 29Megarays.[46] One Megaray describes one million light rays captured by the camera. The relation between final output resolution and Megaray resolution depends on algorithms to restore original sensor resolution and currently varies between 1/4 and 1/10 of the Megaray resolution.

Figure 12: Raytrix R5, 4.2 Megarays, lateral resolution of max, 1 Megapixel (25% of original sensor resolution).

Figure 13: Raytrix R29, 29 Megarays, lateral resolution of 7.25 Megapixel (25% of original sensor resolution).

Lytro's second generation camera, called the Lytro Illum, was introduced in 2014 and fixes that low resolution problem by outputting 40 Megarays and offering an 8x optical Zoom for more flexibility while capturing the images. The capability of video capturing is still missing on the Illum.[47] [48]

[45] Shih (2014)
[46] cf. Sinko
[47] cf. Chip.de mwa/mil (2014)
[48] cf. Wikipedia Contributors – Lytro (2015)

*Figure 14: Lytro Illum, 40 Megarays, output resolution 2450*1634 px or approximately 4 Megapixel (10% of sensor resolution).*

With Raytrix and Lytro being the major players in the sector, other companies were also working on prototypes and solutions for special applications. Companies like Adobe, Cafadis, Pixar and Toshiba[49] were also experimenting with light field acquisition systems featuring microlens arrays. At the Beginning of Adobes early prototypes in 2004, they experimented with compound lenses made of 19 sub-lenses arranged in a hexagonal array, each facing an individually configured prism set at a unique angle, therefore causing 19 varying focal points. By 2010 Adobe discarded the idea of the compound lens and continued to use a microlens array of approximately 7000 microlenses placed in front of the cameras sensor for their latest publicly known prototype.[50] The University of La Laguna, located in Spain, followed a similar approach like Adobe in the beginning, but still differing quite a bit. The *CAFADIS* light field lens prototype consists of a microlens system that can turn any ordinary 2D camera into a light field camera and is based on a matrix of 510000 microlenses accommodated inside a main optic.[51] Pixar's capturing approach was quite similar to the CAFADIS prototype, based on a plenoptic lens that can be combined with conventional cameras, as well as a patented system called "SUPER LIGHT-FIELD LENS WITH DOUBLET LENSLET ARRAY ELEMENT", with an additional lens collecting the actual light field data, independently of the sensor used.[52] In the final days of 2012, Toshiba announced a light field product for the consumer market supposed

[49] cf. Nolf - LichtFeld Kamera Prototypen
[50] cf. Nolf - Adobe LichtFeld Kamera Prototypen
[51] cf. Nolf - CAFADIS LichtFeld Objektiv Prototyp
[52] cf. DiFrancesco (2011)

to be commercialized by the end of 2013, packing a microlens array of 500000 lenses in a module measuring only 0.8 cm x 0.8 cm, making it suitable for mobile technology. [53] [54] [55]

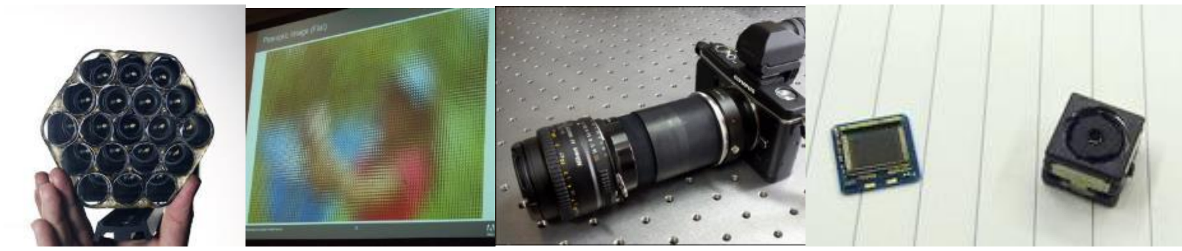


Figure 15: From left to right: Adobe compound Lens, Adobe microlensarray, Cafadis Prototype, Toshiba Light Field Camera Module.

2.3.2.1 Multifocus Plenoptic Cameras

The biggest Problem with Plenoptic Cameras still is the resolution, Lytro's Illum, for example, offering 40 Megarays, is comparable to a traditional 4Megapixel camera, just a tenth of the sensors resolution is used for the final picture.

To overcome the limitation in resolution, Raytrix came up with the Multifocus-Plenotic-camera (MFPC), which through a combination of microlenses with varying focal lengths made it possible to get more depth of field, while keeping the information necessary for 3D-information. With this technique, Raytrix reached a Megapixel Resolution using 25% of the Sensor.[56] [57]

2.3.2.2 Deconvolution Approach

José Manuel Rodríguez-Ramos, Juan Manuel Trujillo Sevilla and Luis Fernando Rodríguez-Ramos are taking the approach even further. They say it is possible to obtain the full optical resolution by using a deconvolution method (by S. A. Shroff, K. Berkner, Image formation analysis and high resolution image reconstruction for plenoptic imaging systems, *Appl. Opt.* 52, p. D22-D31, 2013.) and a "…*restoration much closer to the theoretical maximum can thus be achieved with this technique*".[58]

[53] cf. Owano (2012)
[54] cf. Kamiguri (2012)
[55] cf. Nolf - Toshiba LightField Camera Module
[56] cf. Perwaß & Wietzke (2012) 9-10
[57] cf. Nolf - Raytrix: Neue LichtFeld Kamera mit Bildauflösung von 25 % der Sensor-Auflösung (2012)
[58] Rodríguez-Ramos (2015)

What can't be recovered is the missing information due to the spacing between the round microlenses, a workaround for that will be the use of square shaped microlenses, which they are currently working on.[59]

2.3.3 Multi Camera Arrays

A totally different approach to light field acquisition is the use of camera arrays. Depending on the desired output, it's possible to use digital still cameras or video cameras. By arranging multiple cameras in a 2-dimensional array it is possible to capture light fields and even video light fields when using an array of video cameras.[60]

Stanford University's Wilburn et al. followed the path of developing a light field camera array with over 100 video cameras (seen in Figure 15). Their requirement was to be able to capture progressive video at a resolution of 640x480 pixels and with 30 frames per second. Their goals were to design a camera able to capture video without a huge amount of additional hardware/special equipment, suitable for different configurations and easily scalable at a low per-camera-cost. These goals were followed by the decision to use CMOS sensors and capture the footage in MPEG2 instead of raw-video. CMOS technology became the sensor of choice because it's possible to easily manually set exposure, color gains, gamma, white balance, auto exposure and more, making them easier to measure and control.[61]

Figure 16: Stanford Multi-Camera Array. On the left: arranged 2 inches apart.
On the right: arranged in an arc by mounting them on separate panels.

[59] cf. Rodríguez-Ramos (2015)
[60] cf. Ng (2006) 45
[61] cf. Wilburn et al. (2012) 2

The Fraunhofer Institute for Integrated Circuits IIS concerning Wilburn's large-camera-array:

"*The former approach is complex in calibration and operation due to the high number of cameras resulting data volume.*"[62]

And concerning their statement regarding plenoptic cameras:

"*The latter suffers from a limited resolution per view, as the total resolution of the imaging sensor needs to be shared between all sub-images captured by the individual micro-lenses of the plenoptic camera.*"[63]

Their approach is to use a compact 2D array of high definition cameras (down to 9 cameras, opposed to over 100 (Wilburn et al.)) and the use of disparity-estimation-techniques. The use of disparity estimation gives them implicit geometric information of a scene and with the help of sparse angular sampling, the original sparse light field is converted into a dense light field – making it possible to then use traditional light field rendering techniques.[64]

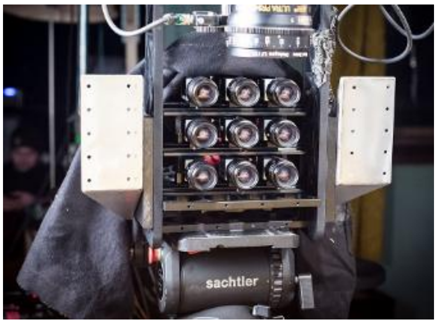


Figure 17: Fraunhofer IIS Multi Camera Array

2.3.3.1 Other noteworthy Camera Array Prototypes

Pelican Imaging is one of many manufacturers concentrating on a camera array compact enough for smartphones. In 2013 they presented their first product called PiCam, consisting of 4x4 fixed focus cameras with each a resolution of 0.75 Megapixels, at a low cost of just 20 US-Dollars.[65] The Images are converted into one 8 Megapixel JPEG-File with an embedded depth-map. For the Acquisition of moving pictures a Full-HD Resolution of 1920x1080 pixels and 30fps is possible. Their examples of application are software refocus, selective depth of field, 3D Range images and intelligent picture editing (for example fore-/background extraction).[66] [67]

It's not only a compact solution compared to the other array camera systems but also follows the unique approach of monolithic sensors. Here each sensor is only responsible for one certain color (RGB) and gain,

[62] Zilly et al. (2013) 1-2
[63] Zilly et al. (2013) 1-2
[64] cf. Zilly et al. (2013) 1-3
[65] cf. Cardinal (2013)
[66] cf. Nolf - Pelican Imaging Array-Kamera
[67] cf. Venkataraman (2013)

exposure and trigger times can be controlled independently for each sensor to maximize the quality-output for each color.

"[…]PiCam does not use a Bayer color filter pattern. In this PiCam instantiation, the filters are on the optical stack. Each camera in the PiCam array is optically isolated and sensitive to just one color of light: red, green, or blue."[68]

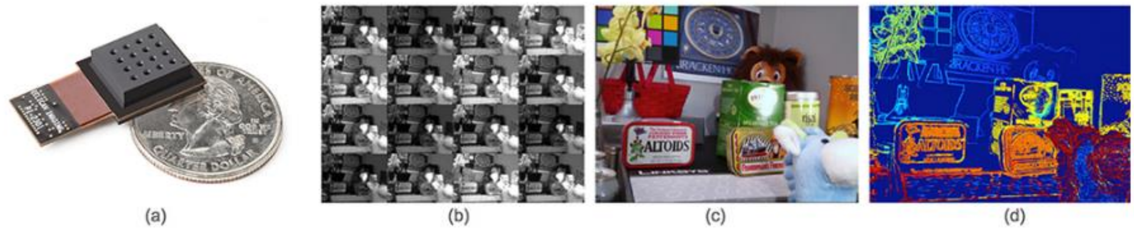


Figure 18: From left to right - (a) The PiCam Camera Array Module (b) Raw 4x4 array images (c) Parallax corrected and superresolved high resolution Image (d) A high resolution filtered depth map

2.3.4 Light Field Displays

When thinking about capturing light fields at some point the thought about displaying light fields also comes to mind. Displaying a light field actually means displaying an object in a similar way we are used to see everyday objects with our eyes.

Different approaches were followed, mainly a glasses free autostereoscopic approach and a glasses based approach. In our everyday vision we can divide stereo perception into two kinds of parallax, one being stereo parallax due to the distance between our eyes and the other being movement parallax, seen when the head is moved. In order to display images for a glasses free approach, a display must be able to view different images to each eye or furthermore each viewing position.

Since displaying technologies are not scope of this thesis, we won't explain how the technologies work, but give an overview on existing technologies and research because of the topics relation to our work:

- Nvidia Near-Eye Light Field Display, offering depth and refocus capability.[69]
- Berkeley Computer Science Division, Eyeglasses-free Display[70]
- USC Graphics Lab, Interactive 360° Light Field Display[71]

[68] Venkataraman et al. (2013) 3
[69] cf. Lanman and Luebke (2013)
[70] cf. Huang et al. (2014)
[71] cf. Jones et al. (2007)

- Disney Research, Multi-Perspective Stereoscopy from Light Fields[72]
- MIT Media Lab, A Compressive Light Field Projection System[73]
- MIT Media Lab, Polarization Fields[74]

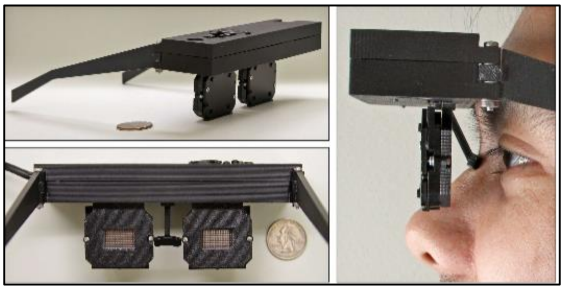


Figure 19: Nvidia Near-Eye Light Field Display Prototype

3 Light Field Applications

3.1 Available Products and Applications

Looking at the first commercial light field camera systems and the existing research literature as described in chapter one, three major application fields can be found: view rendering, synthetic aperture photography and extraction of 3D scene data.[75] In the following chapters the process of transferring them to the context of today's film and TV productions and ultimately selecting the most promising applications is presented. Based on this selection three scenarios will be introduced to be part of the test shoot. This practice represents the preproduction phase of the test shoot.

3.1.1 View Rendering

Historically view rendering was the first goal that has been documented by Levoy and Hanrahan in 1996. Against a background of various techniques of image based rendering that were based on image warping and

[72] cf. Kim et al (2011)
[73] cf. Hirsch, Wetzstein and Raskar (2014)
[74] cf. Lanman et al. (2011)
[75] "LYTRO ILLUM" (2015), "Raytrix Light Field Camera - Nvidia GTC 2012" (2012) 8-34, Venkataraman et al. (2013), Greengard (2014)

interpolation[76], they describe "a simple and robust method for generating new views from arbitrary camera positions without depth information or feature matching"[77]. "Conceptually, rendering an image from a light field amounts to taking a 2D slice of the 4D data […]"[78] This can be thought of as a simple table lookup of light rays in the light field and interpolation of non-existing rays from nearby ones.

Naturally, view rendering can be used as input data of Stereo 3D and holographic displays[79] and in e.g. interactive media, using head-up displays like the oculus rift or the near-eye light field displays presented by Nvidia in 2013[80]. The latter is restricted to stereoscopic display with the focus to present "nearly correct convergence, accommodation, binocular disparity, and retinal defocus depth cues"[81], whereas the former includes sensors that allow for interactively changing the point of view. At the moment both systems are still lacking in resolution and viewer comfort. Until autostereoscopic or holographic devices are found inside of our living rooms, displaying for example recordings of concerts, theatre plays or sports events with options to interactively change view points or stereo depth, they could be incorporated in different event and installation works as well as in the area of immersive environments and gaming. Horn and Chen already showed the incorporation of light field rendering in a 3D game in 2007 to include a high quality "scan" of an object.[82] Web viewers like the one that comes with the Lytro software bring view rendering to web applications and advertising[83]. The startup company Magic Leap, that is partly funded by Google, plans to use light field rendering for high quality augmented reality content.[84] Furthermore, view rendering could also be employed for educational purposes and in a cultural heritage context.[85] Advanced surveillance systems based on light field technology are of relevance, too. The most popular application in film to this date has probably been the effect of orbiting around a scene that has been frozen in time, also known as "bullet time effect" in the movie "The Matrix", Wachowski brothers 1999. It was achieved using still cameras arranged along a 1D path. Feeding this data into a light field viewer would let the observer not only change the position along the camera's path, but also to move the virtual camera back and forth in the direction orthogonal to the camera path.[86] Still it's not a full 4D light field that would need a two-dimensional camera

---

[76] cf. Kurachi (2007) 147
[77] Levoy and Hanrahan (1996) 1
[78] Kurachi (2007) 151
[79] Levoy and Hanrahan (1996) 32, cf. Lanman et al. (2011); Jones et al. (2007); Adelson and Wang (1992)
[80] cf. Lanman and Luebke (2013)
[81] Lanman and Luebke (2013) 1
[82] Horn and Chen (2007) 6
[83] example: http://www.praxis-dollhausen.de/bildergalerie/
[84] Anthony; "Magic Leap"
[85] expert meeting records 1, Levoy and Anderson (2002)
[86] Levoy (2006) 52

array. These camera setups usually need to be specifically designed for a certain shot or application. Wilburn proposes also an advanced application of view rendering as the "spatiotemporal view interpolation system"[87], that stacks trigger times over the camera array and uses optical flow algorithms[88] to extend performance to the axes view position and time.[89]

3.1.2 Synthetic Aperture

Synthetic aperture photography or "digital refocusing" refers to the simulation of the focusing effect of a lens while resampling a light field, in particular simulating different aperture sizes - limited by the size of the camera array - and focal planes by recombining and summing of light rays.[90] Since the introduction of the Lytro camera in 2011 refocusing seems to be the most popular application of the light field data. It has already been proposed in the original paper from 1996 on light field rendering and was first demonstrated in 2000 by Isaksen, McMillan, and Gortler[91]. In his 2006 phd thesis *Digital Light Field Photography* Ren Ng concentrates on applications in every day photography to correct focus after the fact and to maximize light gathering power and flexibility while improving lens quality.[92] This is especially useful when taking pictures of fast moving objects or unpredictable scenes that occur in sports, for example. He describes a "sensation of discovery"[93] that he noticed when people were watching videos of focus sweeps or could control the focus inside a light field viewer application. The latter also falls in the domain of interactive applications and can be connected to the concept of agency, which Janet Murray explains in her book on the future of narratives as "the satisfying power to take meaningful action and see the results of our decisions and choices."[94]

This aspect of interactive storytelling can be found in the advertising of the new Lytro camera Illum, for example.[95] When uploaded onto "pictures.lytro.com" those light fields become means of communication, too. Ng's approach to light field rendering is theoretically close to the image formation process inside a

[87] Wilburn (2004) 98
[88] cf. glossary "optical flow"
[89] Wilburn (2004) 81-93
[90] Levoy (2006) 50
[91] Isaksen, McMillan, and Gortler (2000)
[92] Ng (2006) 1-5
[93] Ng (2006) 88
[94] Murray (1997) 126
[95] ("LYTRO ILLUM")

conventional camera, therefore getting realistic results in terms of image aesthetics. But non-photorealistic focus situations can be simulated as well.

"Vaish et al. [2005] consider the interesting variation of a tilted focal plane, such as one might achieve in a view camera where the film plane may be set at an angle to the main lens [Adams 1995]"[96]

Furthermore, Billy Chen, for instance, proposes the use of multiple focus planes that are achieved by using several camera models and merging image segments after the focusing step[97]. „This is useful when there are several interesting objects at different depths, but unwanted objects between them."[98] This might be the case in sports photography. An extended depth of field compared to a conventional camera of the same light sensitivity is possible, too.[99] Extended focus is nothing entirely new to the language of cinema[100] but the high availability and easy realization when using light field photography maybe has the potential to bring new focus related storytelling tools to the digital visual arts. Another application of synthetic aperture photography is the light field microscope developed at the Stanford Computer Graphics Institute that allows to capture a whole focal stack of images in one light field as well as to view microscopic images with changing perspective views. At the time of writing this project is still in further development.[101]

3.1.3 Extraction of 3D Information

Raytrix claims to be the first company to have built a commercially available light field camera. They focus on industry applications that need, amongst others, the extraction of 3D information of a scene.[102] "Estimating the 3D shape of objects from images of them is a central problem in computer vision."[103] Since the light field holds lots of different observer positions of a scene, shape-from stereo algorithms can be used to calculate 3D information of a scene in forms of disparity or depth maps. In conjunction with digital refocusing shape-from focus algorithms can also be used or even combined for superior results.[104] This 3D data has been used for additional editing steps in the post processing of still images like adding haze or depth

[96] Ng (2006) 51
[97] Chen (2006) 6, multifocus planes are also described in Isaksen, McMillan, and Gortler (2000) 297-306
[98] Chen (2006) 5
[99] Ng (2006) 52
[100] "Citizen Kane" (1941)
[101] Levoy et al. (2006) 1, 5
[102] "Raytrix Light Field Camera - Nvidia GTC 2012" (2012)
[103] Levoy et al. (2006) 7
[104] Tao et al. (2013) 1-2,7

based color correction as well as masking and extraction procedures. More advanced applications are relighting – adding or removing light in a scene – and image based modeling, creating points and surfaces in a 3D coordinate system that match to the shot geometry.[105] These methods can be applied to the scanning of cultural heritage, industry and research e.g. for automated material and surface checks as well as visual effects in television or film. The latter already being explored by Debevec et al. at the University of Southern California (USC) Centers for Creative Technologies. The Lytro SDK released in 2014 addresses a greater demand for light field technology outside the commercial photography market. Clients are amongst others NASA and a division of the US department of defense.[106]

3.2 Overall Requirements

In our initial brainstorming sessions at HdM and IIS we found a variety of possible applications of light field technology suitable for moving picture media based on the known applications described above. Examples are the use of view rendering in documentary shoots or traditional motion control domains, better quality reframing, focus adjustments during grading or simulation of camera models and lenses during previs or postproduction, improved matchmoving and the employment of 3D data in relighting and other VFX processes.

To be able to further refine this list of possible applications we decided to invite a number of industry experts to discuss future applications of light field photography as a tool of cinematography.[107] The goal was to learn about the actual industry's demands in terms of technological development and workflow efficiency and to find some promising applications for light fields in a day-to-day production environment. The meeting took place in November 2014 at the HdM Stuttgart. Six external experts with diverse backgrounds, three HdM professors, three academic associates and eight students attended the meeting.[108] Two more experts, who couldn't come to Stuttgart, joined the project as peers and were available for questions since then.

---

[105] Abbot (2013) 9-22, Venkataraman et al. (2013) 12

[106] "Lytro Releases Dev Kit for Light Field Camera Technology" (2015)

[107] a workshop with stakeholders can be seen as part of our requirements analysis approach as described in Ebert (2012) 70

[108] a detailed list of the attendees with short biographies can be found in the appendix

Six external experts with diverse backgrounds, three HdM professors, three academic associates and eight students attended the meeting.[109] Two more experts, who couldn't come to Stuttgart, joined the project as peers and were available for questions since then.

Since previous knowledge about the light field concept and the particular research project at Fraunhofer IIS varied between the members of the group, the Fraunhofer IIS organized a live webinar one week prior to the meeting. The meeting day itself started off with a brainstorming session in several small groups whose constitution had been done depending on the persons' interests and backgrounds. In the afternoon the ideas were presented and discussed with the whole group. A host helped organizing the process of collecting and evaluating applications.

In the course of the meeting several basic requirements were mentioned that hold for the integration of light field (or to a certain degree, any new technology in the area of film production) and apply to more or less all applications. These will be described now.

3.2.1 Camera and Hardware

3.2.1.1 Handling

The light field camera head should be of the same or only slightly bigger size and weight as a standard production camera or an up-to-date stereo 3D rig.[110] Standard production camera means a camera that would be used to shoot the major part of a motion picture. Popular digital models at the time are the Arri Alexa, Red Epic or Sony F55, which can reach a weight of 12kg when fully rigged and have dimensions of up to 50 x 20 x 35 cm. This requirement might not apply to special applications like in the context of a dedicated light field capture stage or second unit (plate[111]) photography. Additionally, the handling and feel of a light field camera should be comparable to their single camera brothers and sisters. For on set use it's critical to have an easy and intuitive user interface that can be understood immediately by a DP or AC, who is familiar with conventional camera models. Alignment and calibration on set has to be as fast as possible; it should definitely take no longer than the alignment of a current stereo rig. Ideally, it takes the same time as a standard 2D camera rig.[112]

---

[109] a detailed list of the attendees with short biographies can be found in the appendix

[110] Ganbar, Kirchhoff, Neuhaus during the course of the meeting

[111] cf. glossary "plate"

[112] Neuhaus meeting records 5

3.2.1.2 Multi-Purpose Acquisition System and Calibration

It would be best, if one capturing device could deliver to several or all light field applications.[113] As mentioned by Wilburn, different use cases require special array configurations and calibration. For example, synthetic aperture needs less complex calibration if only a set of parallel focal planes have to be achieved in post.[114] More obviously, view rendering requires special configurations depending on the angles and degree of freedom needed in postproduction. While the cameras can be further apart for view rendering, they need to be as dense as possible for applications like the extraction of surface properties.[115] Günter Neuhaus demands a camera array that can be easily configured and calibrated. He emphasizes the importance of solid rig construction and the availability of skilled people on set. These demands refer to the lessons learned from the advent of digital stereo 3D. [116] Especially for bigger arrays Wilburn demands a self-configuring array like the motorized prototype by Zhang and Chen that allows automatic calibration.[117] To a certain degree calibration can be corrected comfortably in post [118] but will usually cost time and spatial image resolution.

3.2.1.3 Characterization for Postproduction

As Siragusano points out, these camera arrays also need to be characterized and calibrated for postproduction.[119] In order to successfully simulate or create the camera effects with light field data in post all input camera specifications like sensitivity, camera response curves and spectral responses need to be known. The needed color consistency between cameras also depends on the application mode. Synthetic aperture is relatively forgiving in terms of color variations due to the averaging of image information in the defocused areas.[120] Other applications like view rendering or special High-X applications[121] for superior image quality need higher color accuracy than it is available in consumer cameras or low cost sensors.

---

[113] Siragusano meeting records 2
[114] Wilburn (2004) 46
[115] Eberhardt meeting records 3
[116] Neuhaus meeting records 1
[117] Wilburn (2004) 11, 41
[118] Wilburn (2004) 30, Zilly et al. (2013) 3
[119] Siragusano meeting records 6
[120] Wilburn (2004) 18
[121] cf. Wilburn (2004) 17-19

3.2.1.4 Sensor Technology

To be able to compete with the current developments in the field of digital cameras, a light field camera system should offer 4K resolution, HFR (more than 25 or 30 fps) and HDR (high dynamic range) image output.[122] At the same time the light field benefits from cameras with a large depth of field, which means small sensors in the first place.[123]

3.2.2 Data Formats and Processing

3.2.2.1 Intelligent Data Processing and Storage

Ron Ganbar states that stereo 3D data is already perceived as a slowdown in the production process.[124] As the number of camera streams rises when shooting light field, it is a key requirement to process the data in an intelligent and as efficient as possible way.[125] Repetitive tasks should be automated and easy integrity and quality checks must be possible.[126] For efficient data storage and archiving as well as bandwidth reduction data compression should be explored.[127]

3.2.2.2 Solid Data Format and Metadata

This leads to the requirement of a solid data format for light field data that builds on today's standards and allows for easy interchange between applications and software packages. It would be best to keep the light field with all its information until the color grading and finishing stage.[128] This data format has to store multiple views, additional color channels for 3D data like depth and normals and extensive metadata. For VFX processes all set data is of critical importance. As much information as possible about timecode and camera settings has to be stored as well as light field related data.[129] The latter have to include the on-set decisions on focus depth or areas and viewpoint as a minimum.[130] Looking at future trends and global

[122] Kirchhoff, Neuhaus meeting records 5
[123] Kirchoff meeting records 6
[124] Ganbar meeting records 1,3
[125] cf. Siragusano, Ganbar meeting records
[126] Green et al. (2014) 215, Bugaj (2010) 789, 791
[127] Ganbar, Siragusano meeting records 2
[128] Siragusano meeting records 2
[129] Ganbar, Wieland, Siragusano, Kirchhoff, Neuhaus meeting records 3, 5
[130] Kirchhoff, Neuhaus meeting records 5

developments the data format should also be open-source to allow for a fast adoption and improved maintenance.[131]

3.2.3 Workflow Organization

3.2.3.1 Incorporating the DP

As already touched on briefly, there has to be a way to incorporate the DPs in the process of capturing and post processing a light field.[132] They are traditionally "responsible for achieving artistic and technical decisions related to the image"[133]. Together with the VFX supervisors in case of a VFX sequence they need to stay in control of all the image parameters, like framing and lighting and their vision should guide any post processing. To work efficiently on-set it also makes sense to decide on a "hero view" (preferred camera angle) and focus early on and store that information for postproduction.

3.2.3.2 Preview

This requires some sort of light field preview on-set.[134] At some point, a live preview is required, which might be of lower quality though.[135] In the end, ideally every department could access the whole light field and get a real-time preview of a basic camera simulation. Real-time in this context means a latency time that should not exceed 5 frames or roughly 200 ms, the ideal being one frame or 12 ms delay.[136]

3.2.3.3 Reliable Workflow

Wieland and Tukiendorf highlight the importance of a reliable workflow that makes all the needed information from set available until the finishing step and reduces manual or format specific work steps to a minimum.[137] As already mentioned above, a Quality Control or assistance system that helps achieving the best input data for the post processes is required in the long-term, too.[138]

---

[131] cf. Dabney (2013)
[132] Ganbar meeting records 1, Grandinetti meeting records 6
[133] "Cinematographer" (2014)
[134] Ganbar meeting record 3, Kirchoff 5
[135] Ganbar meeting records 4
[136] cf. Knopp (2014) 26
[137] Wieland, Tukiendorf meeting records 5
[138] Neuhaus meeting records 1

3.2.4 Postproduction and Software

3.2.4.1 High Quality Depth Maps

The depth information needs to be as accurate as possible and low on artifacts.[139] Lots of processes build on the depth data and therefore the overall image quality depends heavily on the depth maps.

3.2.4.2 Camera and Lens Simulation

A long-term requirement is the possibility to render light fields with parameters from existing or non-existing camera models and lenses.[140] For applications in VFX and film, simulation in real-world units is of great importance. In the use case of changing the aperture in postproduction, for example, a DP most likely wants to specify the new value as a number of stops. To start with, it must be possible to reproduce the characteristics of a conventional cinema camera, including the lens from the light field data captured by an array of smaller cameras.[141]

3.2.4.3 Image Quality

Image quality is critical for applications in media like film and TV.[142] This includes technical requirements of high dynamic range, sensitivity, a wide gamut of realistic colors, high resolution and sharpness, while keeping noise, aliasing and compression artifacts low in the image output of a light field pipeline.[143] These properties should be equal or superior to conventional camera systems in the corresponding domain. Additionally there also more subjective aspects of image quality that have been brought together by Yendrikhovskij in the following equation that predicts what quality means to an average viewer: Quality = Naturalness + Colorfulness + Discriminability. Discriminability means the ease of reading a scene and to understand an image in terms of composition.[144] Of course, this definition also applies to the image output of a light field system in the same way it does to conventional photography.

---

[139] Ganbar meeting reords 7

[140] Eberhardt, Siragusano, Ganbar meeting records 7

[141] Eberhardt meeting records 7

[142] Wieland, Ganbar, Siragusano meeting records 5

[143] Neuhaus, Kirchhoff meeting records 5

[144] Van Hurkman (2014) 403

3.2.4.4 Integrated Pipeline

Since the light field rendering is mostly a postproduction process at the moment, it needs a good connection from set to postproduction.[145] Tukiendorf and Wieland demand a reliable pipeline and good workflow integration.[146] There should be a set of, ideally open and customizable tools that can be integrated in existing pipelines and popular software packages.[147] A standard file format is one of the important and first things needed in this context.[148] The core functionality of the toolset has to be independent from different acquisition methods and camera types.[149]

3.2.4.5 Goal and Guideline

These quality requirements on light field technology shall serve as a guideline for the development of our workflow proposal and integration schemes. Even though the acquisition technology might be not there yet, we want to match the high-end work environment and standards of professional media productions. At this point it's also important to keep in mind that especially quality requirements affect each other, in some cases even contradict each other. These requirement need to be balanced in the development process later on.

Especially the named requirements related to workflow development and pipeline integration supported our idea of trying to integrate light field applications into modern film and VFX workflows. Although we used mainly one specific camera array in our tests that we are going to describe in detail in chapter 6.2.2, all applications should be seen as independent from the acquisition system.

3.3 Selecting Applications in Film and TV

3.3.1 Background – the digital evolution

One of the strengths of light field photography is the wide range of applications that profit from the additional scene data. In general it can help to gain a higher "degree of freedom", meaning less dependence on physical facts like sunlight, location or camera technology, which increases overall production

[145] Tukiendorf meeting records 5
[146] Wieland and Tukiendord meeting records 5
[147] Ganbar meeting records 6
[148] cf. Green et al. (2014) 103
[149] Wieland meeting records 5

efficiency.[150] Since we couldn't test all possible applications due to the limitation of this project and the current state of development, we had to choose a small number of feasible applications that hopefully meet the demands of the industry.

Creative processes are generally subject to constant change. Since the advent of digital film and video, the film industry finds itself in a phase of fundamental remodeling, though. Digital film and the file based workflow, or more accurately the continuous extinction of analog film, brought new distribution and financing models, faster turn around times, extensive postproduction as well as new roles and disciplines.[151] Video on demand, mini content marketing, crowdfunding, investment of advertisers and resellers and democratization of production processes are some popular keywords that bother producers at the moment.[152]

As Howard Lukk from Disney points out there is a trend to have smaller on-set crews and a tendency towards shifting decisions and processes to postproduction.[153] This is part of the everlasting challenge of getting more for less[154], but also a result of the continuously growing relevance of visual effects to film productions.[155] Consequently, a rising percentage of the production budget is related to postproduction and VFX in particular. 49 out of the past 50 academy awards in the category "best film" included VFX as a key storytelling device or as a digital character.[156] Also the documentary genre started to embrace the possibilities of digital postproduction.[157] Example use cases are motion graphics and sequences to visualize complex data as well as whole parts of VFX-driven reenactment[158]. This development addresses as well as causes changing viewing habits of the audience, a generation that spends enormous time looking at images on all sorts of screens.[159] Getting used to "Hollywood-like" effects, people expect things to look "real" to a today's and increasing standard.[160] After a time of fast development and technological inventions to create certain effects like hair, skin and fluids of emotive visual quality it is now going to be about more efficient ways to produce these effects.[161]

---

[150] cf. Knopp (2014) 8
[151] cf. ("14 Trends for Filmmakers to Look Out for in 2014 | Raindance")
[152] ("14 Trends for Filmmakers to Look Out for in 2014 | Raindance"), cf. Dabney (2013)
[153] Lukk (2014) cf. Neuhaus meeting records 1, Conelly, Court, and Cameron (2008)
[154] Romanek (2014)
[155] cf. Dabney (2013), Büttner (2004) 2-3
[156] Leberecht (2014)
[157] Collins (2009) 26, 30
[158] Collins (2009) 27
[159] cf. Collins 25, 42-54
[160] Romanek (2014), Büttner (2004) 3
[161] Visual Effects Society and Okun (2013) 11

Although VFX have been traditionally a (labor intensive and costly) post process, organizational structures are changing at the moment. In the context of virtual production VFX are part of the "world building" process.[162] Operations get more parallelized and borders between real and virtual, analog and digital as well as on-set and post get blurred (or fall entirely). VFX production nowadays usually starts in preproduction and goes on long after principal photography is done. According to producer Alex McDowell, the "cost that is killing productions is the transition between the analog process and the digital post". The seamless integration of virtual and real parts of a shot still poses a challenge and new and interactive ways are to be explored.[163] These are reasons why we focus on a workflow relying on visual effects.

As we talked to our industry partners we found an overall positive mindset towards new technologies and development.[164] "We use all the tools we can use"[165], says Tukiendorf, for example. Especially in VFX, people are willing to try new approaches in technology due to the competitive atmosphere between vendors and the fact that one of the main challenges usually is to create effects no one has seen before. But it's important to keep in mind that technology will be usually following the story to be successful in film productions as Jon Landau, producer of Avatar, emphasizes.[166] Therefore, to successfully apply light field to film and TV productions it should be treated as a storytelling device.

3.3.2 Use-Cases and selected applications

For each of the three major light field applications view rendering, digital focus and 3D data extraction we found some use-cases that address typical workflows and demands in film and TV productions. This chapter briefly describes the most promising.

3.3.3 Digital Focusing

Synthetic aperture or digital refocusing[167] can be used to help in situations with unpredictable movement or extremely shallow depth of field, where it is hard to manually pull focus and autofocus produces

---

162 McDowell (2014)

163 McDowell (2014), cf. Lukk (2014), Rogmans (2014)

164 e.g. Siragusano meeting records 2, Ganbar meeting records 1

165 Tukiendorf meeting records 6

166 Landau (2014)

167 in the following chapter and the main part of this thesis only the term „digital focus“ will be used to refer to the possibility of setting the depth of field at a different time from shooting an element. Additionally it emphasizes the common goal of positively setting the right focus instead of negatively correcting or undo the work of others.

unsatisfactory results.[168] At the moment you would have to stop the lens down or repeat a shot, maybe limiting creative possibilities in such circumstances. There is no working solution to change focus of a 2D image after it is taken.

The use of multiple or non-parallel focus planes could be realized fairly easy compared to traditional approaches, which rely on multiple exposures. This could be used as a storytelling device and add to a new and unique style of light field photography.

Less obvious, digital focusing can facilitate VFX processes, for example the integration of CG elements into live action plates.[169] Setting the focus at the right depth through all elements that comprise a shot is one of the important parts in VFX work. To be able to control the focus of live action plates at the same production stage as the focus of computer graphics elements makes shooting for VFX more efficient, since you don't have to worry about getting a previséd focus pull right, and compensates changes of a shot layout that occurs late in production. A good example for the latter would be an animated CG character whose animation is loved by the director and fits in the edit perfectly but is quite different from the one prevised. As Ganbar points out there are also other VFX processes that profit from in-focus plates. All masking operations like rotoscoping and color keying, paint tasks as well as motion tracking that uses computer vision and feature detection algorithms could be easier accomplished when scene elements have sharp edges.[170] This assumes that it will be possible to do these operations in an efficient way on a light field data set.

Although digital focusing is a quite popular application, setting focus is also one of the tasks that had to be mastered over the years by countless camera assistants and there is no urgent need for correction most of the time.[171]

3.3.4 View Rendering

View rendering or just light field rendering, as it is called by Levoy and Hanrahan in their 1997 paper of the same name, can be interpreted as computing a novel perspective view by extracting an appropriately positioned and oriented 2D slice from the 4D light field.[172] Theoretically this new viewpoint could be

[168] Ng (2006) 2-4, cf. Kirchhoff, Tukiendorf meeting records
[169] Wieland meeting records 5
[170] Ganbar meeting records 8
[171] Siragusano meeting records 2, Tukiendorf meeting records 2
[172] (Levoy) (2006) 3

anywhere outside the convex hull of the scene's objects.[173] In reality the degree of freedom as well as image quality depends on the camera array and camera specifications. For example, a two dimensional array of cameras would allow for a viewpoint anywhere on the 2D shape of the array, as well as in front and behind the virtual camera's sensor plane.

This technique could be used for reanimating camera movements, often used for stabilization and smoothing, or reframing in post. Both is already done today using 2D transformations and scale operations and usually takes place during the VFX or finishing stage.[174] Reasons to reframe are integration of VFX and establishing matching eye-lines between cuts, as well as guiding the audience view, compensate for angles not shot but needed in the edit or just improving the overall flow of the final edit at the finishing stage.[175] While resolution oversampling is not a big issue with nowadays' digital cameras for deliveries up to 2K that are mono 2D, it gets more of a challenge for Stereo 3D and 4K or higher resolved formats. These could be addressed by view rendering from a high-resolution light field. Especially in the context of stereoscopic images where viewing comfort can easily suffer from 2D cropping and transformations, light field could help.[176] Light field can be used to improve Stereo 3D post production since it allows to re-render the two views needed for a S3D delivery with all the camera parameters set and animated according to the context of the edit or the final shot layout if VFX are involved. It is also possible to generate different versions of a S3D program for different viewing conditions with appropriate inter-ocular distances.[177] Any number of additional views can be added after the fact, which ultimately could lead to filming a scene only once and generating the needed camera angles in post. Of course, a special camera setup or even stage setup similar to a camera dome would be needed. This could be interesting in the context of serial productions for TV, for example, that more than others demand for high efficiency.[178] In some cases view rendering could even replace motion control applications like in miniature or stop-motion shooting. In film and increasingly TV there is often the need to add a background to a foreground shot at a different location.[179] These scenarios that we refer to as backlot scenarios, could use light fields as live-action backgrounds to be able to adjust focus and viewpoint according to the foreground camera position while keeping correct parallax in the

[173] (Levoy) (2006) 3

[174] Tukiendorf meeting records 5, Kirchhoff meeting records 5, Ganbar, Siragusano meeting records 2

[175] Wieland meeting records 5, Kirchhoff meeting records 5, Neuhaus meeting records 6

[176] Transformations may change parallax and the stereographers assumptions on scene content, when cropping there is a high risk of introducing edge violations, compare for issues with cropping of S3D images: Zhang, Niu, and Liu (2013) 2-3, cf. Kirchoff meeting records 5, Mendiburu (2009) 108

[177] Kirchhoff meeting records 5

[178] Tukiendorf meetng records 5

[179] Tukienforf meeting records 5

background. The conventional approach would need another step of plate preparation that involves many hours of manual work to be able to show parallax in a 2D background. S3D deliveries would probably need even more preparation work or use a 3D environment altogether.

Figure 20: Virtual Backlot realized in a real-time color keying process by Stargate Studios Germany

View rendering could also be employed to simulate specific camera and lens models and is to a certain amount already part of the view rendering process as Ng states.[180] Simulating different camera models, rare and old lenses or even cameras that do not exist yet could be use cases not only in previs or VFX.

Not all cameras in an array have to be necessarily doing the same thing though. It would be possible to offset trigger times, exposure or spectral properties (including infrared) of different cameras for example and merge the different camera outputs to an image or sequence with superior attributes compared to a single camera's image. High-x applications, as Wilburn calls them in his dissertation "High Performance Imaging using arrays of inexpensive cameras", could make up for drawbacks of using small and cheap cameras in an light field array and create imagery for a variety of special use cases like high speed, high resolution, high signal-to-noise ratio, high dynamic range and wide gamut in a cost efficient way.[181] These applications need an array with the cameras "packed closely together" so the views provided by each camera can be made identical by a projective warp.[182]

---

[180] Ng (2006) 156
[181] Wilburn (2004) 15
[182] Wilburn (2004) 11

As we mentioned before there is an overall demand for higher quality digital images whereas view rendering always is a trade-off between the size of the acquisition system and the range of possible views in postproduction. Therefore light field systems are unlikely to replace dolly, steadycam or crane systems in the near future.[183]

3.3.5 3D Data

Since it is usually the task of the visual effects department to integrate effects and elements in a way that is perceived as photo-real, there is a natural demand for all data possible about the reality of shooting the effect plates. This survey data contains measurements of the set and the camera positions and parameters and reference photography from angles different than the shot camera likewise.[184] A light field consists of different views on a scene that can be used to calculate depth information or, in other words, distances of objects from the camera. A depth map together with the color values can then be used to generate 3D information like position data and scene normal vectors. The first task in VFX that would profit from this data is matchmoving. The solve algorithms could be constrained by the survey data and additional views.[185] Depth data gives the artists clues about relative position and distances in a scene in an easy accessible way compared to handwritten surveys that need to be linked to the right footage.

It can also be used to generate masks based on positions in 3D scene space for compositing or advanced secondary color correction.[186] Nowadays masking still involves a lot of manual work that could be reduced using light field data.[187] Maybe the traditional color keying workflows could also be improved using depth and position data if, for example, there are two objects of similar color at different depths. Moving point clouds of actors and objects can serve as guides for animators or modelers. A use case would be a digital double that needs to be animated according to a real actor, for example.[188] Today a labor-intensive technique called "rotomation" would be used to trace the 2D movement and apply it to a 3D model. The data from a light field could enable more of a 2.5D motion capture process. Depending on the camera configuration image based modeling algorithms could also be exploited to create 3D geometry from a light field. This can

---

183 Siragusano, Ruhrmann meeting records 2
184 Goulekas (2010) 127-139
185 Kurachi (2007) 136
186 Wieland meeting records 5
187 Ganbar, Tukiendorf meeting records 8
188 Wieland meeting records 5

be useful when a 3D model has to match an actual object to be able to interact with CG elements or particle effects like fluids and fire, for example.[189]

Light fields being basically computer graphics the 3D data could allow for editing live-action elements with the same tools as computer graphics. This can be seen as one step further to the goal of rendering all scene elements, computer generated and live-action in one physically correct environment, which should result in perfect integration.

One key aspect of integration is the lighting. "In a partly virtual scenery while shooting for example on a green screen stage the DP and the whole crew is challenged to create a scene, where all the elements are lit correctly."[190] The role of the cinematographer, traditionally in charge with the scene lighting, has been reconfigured by the possibilities of digital postproduction.[191] With the help of 3D data it is possible to relight a live-action element in postproduction, as part of a VFX process like it would be done with a green screen element, for instance, or in DI during grading. This can help integrate light emitting effects like fire as well as making small corrections to the set lighting like adding fill light to an actor's face or reducing light in the background to help guiding the viewer's focus.[192] The envisioned lighting by the DP is not always possible to achieve on-set due to spatial limitations or time pressure. Also, as mentioned above, the intention and layout of a shot might change during the editing phase. In a shot from Peter Jackson's King Kong relighting had to be used to blend the live-action plate of the actress Naomi Watts with all the other scene elements. Therefore, a complete digital double for Watts had to be created to seamlessly integrate the green screen plate of hers with the digital King Kong, a digital t-rex dinosaur as well as a background 2D matte painting and parts of a 3D CG environment.[193] This extreme example for relighting involved methods of image based modeling[194] and hours of manual animation. To date, colorists or compositors mostly modify the lighting by color correcting with hand drawn 2D shapes with soft edges that are positioned into place using manual keyframe animation or motion trackers. We will be covering this topic in more detail in 4.2.3.1. Generally, it can be said that it is still impossible to really change the lighting of an existing background plate in a drastical way.[195] For this to be possible you need the surface properties of the scene objects to correctly model reflections. Theoretically a light field could be used to extract surface properties, which would definitely be

189 Wieland meeting records 5
190 Knopp (2014) 69
191 Prince (2013) 65
192 Wieland, Siragusano, Ganbar meeting records 6, Van Hurkman (2014) 338-341
193 Knopp (2014) 69
194 "Light Stages"
195 Knopp (2014) 70

an interesting application for the future. But it is much easier to apply a generic shading model to the scene or object. Reshading can be used to add specular highlights to a diffuse surface or as digital makeup, for instance, as well as more distinct effects by merging a metal material with the skin tones of an actor.[196]

Artistic and new visual styles could be created using the 3D data also by visualizing the 3D information. Probably, this won't be the major use case but something like an animated point cloud has already been done in the Radiohead "House of Cards" music video.[197] Both real-time lidar equipment and structured light scanners were used.[198]

At the moment lidar scans and cyberscanning or several forms of image based modeling like the USC light stage[199] are the options to create partial or full 3D representations of a scene. They give satisfactory results in terms of quality and level of detail but do not work well for moving subjects like actors. Apart from that they need additional equipment, trained personnel and more importantly time on-set.[200] The advantage of light field systems is the fact that it captures 3D data at the moment of scene action and doing so anytime the camera is running.

Scene depth can also be captured using time of flight systems, but there are still limitations at the time of writing. As Knopp puts it in his survey on virtual production methods and guidelines:

> *For a professional end users standpoint current results indicate that the resolution and limitations of such technologies do not fulfill the requirements for the use in a film or TV production to date, because it would cause as much if not more workload to fix existing picture quality problems. Limitations of a time of flight system are for example scenes with background lighting, interferences with multiple cameras and reflections such as in glass or mirroring surfaces.*[201]

Even though light field technology still has quality issues with fine detail like hair and refractive or highly reflective objects, too, 3D data probably is the most important application for film and TV at the moment.[202] Until overall quality improves, it is also imaginable to use a hybrid approach for the extraction of 3D data.

---

[196] Siragusano meeting records 5, Van Hurkman (2014) 460-470
[197] cf. Failies (2014), ("Radiohead - House of Cards - YouTube")
[198] "House of Cards (Radiohead Song)", the source data is available under CC license on google code: https://code.google.com/p/radiohead/
[199] "Light Stages"
[200] Goulekas (2010) 137, Edwards (2015) 4, Lasky (2010) 141, 144-145, Clavadetscher (2010) 156-157
[201] Knopp (2014) 75
[202] experts meeting records 6

You could combine a small light field array for the 3D scene data with a digital cinema camera for the color values.[203]

3.4 Selected Applications

Based on the current stage of development and the results from the experts meeting that have been summed up in the previous chapters we chose some use cases to focus on. These use cases incorporate certain aspects of the following applications: 1. digital focusing for VFX and artistic purposes, 2. view rendering for small corrections and backlot scenarios, 3. 2.5D relighting with extracted 3D data and 4. compositing with depth and light field data.

Although digital focusing might not be the number one reason to use a light field system, we think it is a necessary step in a light field workflow. For the other applications to produce high quality results it is best to have an as large depth of field as possible in the input images. Plus, together with the other applications, changing focus at the final stage in the pipeline could be of great value at times. Finally, focusing with light field rendering is the only way to achieve a realistic behavior around object edges as parts of the background get revealed in semi-transparent out-of-focus areas. Depending on the camera configuration digital focusing delivers results that can be compared to the defocused look of real lenses. It demands a dense sampling and possibly the calculation of additional views through intelligent interpolation for final rendering.[204]

Likewise, view rendering is unique to light fields, but has a more distinguishable effect. The idea of repositioning or reframing an image meets the demands of the industry but is already done using 2D transforms and digital zooms at the expense of spatial resolution that is usually sufficiently available with modern cameras. Nevertheless the possibility to animate small camera movements independently from the camera's recording speed that contain correct parallax seems to be useful, especially in serial productions for TV or low budget productions.[205] "Very likely, a practically-sized array of 3×2 or 5×3 cameras would capture a sufficient range of viewpoints to fine-tune camera angles chosen by the di- rector and cinematographer"[206], Einarsson et al. mention in the context of the USC light stage project. The use of light field technology for typical second unit tasks like shooting backgrounds or aerial footage is probably more attractive and should

[203] Siragusano meeting records 5, Walter meeting records 7
[204] Siragusano meeting records 6, Ng (2006) 56-57
[205] cf. Tukiendorf meeting records 6
[206] Einarsson et al. (2006) 10

be our main application in the field of view rendering. The camera array's size isn't of primary importance for the crew in this case either.

Relighting can have an easy recognizable impact on an image and addresses interests of VFX producers and supervisors as well as colorists. As described above relighting with realistic results is only possible as a manual and labor-intensive process at the moment. Relighting by simulating real light sources in a 3D space could yield better results and cut costs. Small corrections like adding or removing fill and ambient light are done frequently during the grading and finishing process.[207] That's the reason why also the company Filmlight ltd. is working on an implementation of better and relighting techniques in their color correction tools.[208] The extracted 3D data needed for relighting doesn't have to meet the same quality requirements as the rgb-color output of a light field system. Siragusano states that, for a start, it is enough to be able to compete with a quickly drawn shape.[209] The 3D data only needs to approximate the real shape of objects and a lower level of detail can be considered enough for the time being. Therefore, relighting might become applicable in film and TV in the near future.

Besides relighting, there are lots of other tasks in a VFX pipeline that can profit from 3D scene data. We will focus on applications of depth and position data at the compositing stage that can help several integration tasks. The most popular topic in the context of 3D data is probably depth keying, extracting mattes for objects depending on their distance from the camera.[210] While depth keying is currently not able to compete with the quality of color keying, masks for color correction and atmospheric or volumetric effect, for example, seem very useful.[211] For these masking operations any depth map that is of higher quality than the one available from parallax of a stereo camera rig or a moving camera can be considered useful.[212] Finally we chose to give an outlook on compositing with one or more light fields. To work with light fields in an efficient way, one option is to rely on 3D data like depth and disparity between adjacent cameras.[213]

[207] Siragusano meeting records 2, 6; Van Hurkman (2014) 340, 347
[208] Siragusano meeting records 1
[209] Siragusano meeting records 6
[210] Knopp (2014) 75
[211] Ganbar meeting records 2,6
[212] Ganbar meeting records 7, see chapter 6.4.3.3 for a definition of quality of depth data
[213] Jarabo et al. (2014) 2

3.5 Summary

This chapter documented the preproduction phase of the project that lead to the concept for the test shoot. The experiences from the test shoot are then taken into consideration when describing models for a workflow integration of light field data.

After an introduction into the theory of the light field and technical developments in the field, starting from existing application presented in research projects and the first commercial products, ways to utilize light fields in the context of film and TV productions were explored. The results of the experts meeting, that was organized as part of the research project have been evaluated and categorized into overall requirements for the integration of light field technology in film and TV productions. Finally, light field applications further defined by use-cases were selected against the background of industry demands and practices.

4 Test Scenarios

Since there has been the intention to do our tests in the context of a narrative form, we chose the commercial spot. It has the advantage of a short form while still aiming for a cinematic visual style.

The first use-case we chose was a backlot scenario. Our goal has been to combine the applications view rendering and focus into one common real world scenario that incorporates visual effects[214]. The particular task is to insert a background in a foreground plate with green screen that shows a view outside a window. The foreground is shot in a studio and the background on location outdoors. A windowpane, reflections and a glass texture need to be added to the light field, too.

Secondly, we have a portrait that is a medium and close shot of a person. This can be seen as a typical scenario for relighting during color correction and grading. We shot the portrait with different light setups to be able to compare relighting results with reference footage.

As we shot our test in the style of a commercial spot, the most important shot would be the product shot or packshot. As the demands on a shot like this are of very high quality, it usually has to be repeated numerous times. Light field could take some of these iterations over to postproduction to make things more efficient and less expensive. This use-case incorporates relighting as it would be done in a VFX pipeline as well as compositing with depth data and digital focusing. Additionally, a 3D model rendered inside a conventional 3D software package is composited with a live-action background and relit for better integration as part of the process.

[214] cf. Tukiendorf meeting records 5-6

4.1 Virtual Backlot

By a Virtual Backlot we commonly understand the background of a scene, which is added in postproduction, often just window outlooks, the window itself and sometimes whole sets. Not to be confused with Stargate Studios' "*Virtual Backlot Live*". To replace a window or background the most commonly used technique is to shoot against a blank background, most of the times a green screen, which can then be replaced in postproduction with the help of colorkeying, often just called keying.

As Ganbar defines it in his book about professional compositing and visual effects in Chapter 7:

> *Keying is the process of creating a matte (an image that defines a foreground area) by asking the compositing system to look for a range of colours in the image. This is also sometimes called extracting a key. It's a procedural method, which makes keying a lot faster than rotoscoping, for example. However, it has its own problems as well.*
> *You have to shoot specifically for keying because you have to place the foreground object in front of a single-color screen. The color of the screen can't include any of the colors of the foreground object because the computer will be asked to remove this color. Usually this is done with either a blue or green backing – called bluescreen or green screen, respectively. The color you choose usually is related to the colors in the foreground object. If it's an actor wearing blue jeans, a greenscreen is used. If it's a green car, go with a bluescreen.*[215] –Ganbar (2011).

A major limitation with green screen backlots is the inflexible camera position, in order for a high realism the camera of the recorded on set footage has to exactly match the movement of the camera used for the background, for exact matches often *motion control camera rigs*[216] should be used. This technique has the advantage that you don't have to shoot on-location, saving a lot of time, money and permission-acquisition. Things that cost the production a lot of money, also there are cases where the environment does not exist in reality. The (artificial) environment is then added in postproduction.

[215] Ganbar (2011) 197
[216] Definition: a motion control camera rig is one that uses computer-controlled motors for accurate reproduction of camera movement with respect to time. (Okun and Zwerman (2012) 254)

Environments can either be built completely in 3D, as a still image (matte painting), layered images or a projected matte painting (2D images projected on 3D elements). Which technique to use depends on the wanted amount of realism and the different times and angles needed for the environment.[217]

4.1.1 Matte Painting

Even though Matte Paintings are among the oldest visual effects techniques in film, their purpose and necessity hasn't changed. Often the setting has to be fabricated artificially, either because it doesn't exist in reality or just because it's not possible to shoot at a desired location. Early on settings were painted on to canvas or later on layers of glass, when a high amount of realism was necessary. By layering the glass-paintings a parallax-effect could be achieved. Nowadays it's basically the same approach, but instead of paint, artists nowadays use pixels in digital imagery.[218]

Matte Paintings are one of the easiest ways to replace a green screen (or respectively). The common take is to simply replace the green screen with a still image or a video shot especially as background replacement (similar to a background projection in the analogue era). A big setback in terms of realism is hereby the lack of parallax movement. Missing parallax is often the reason a backlot is not perceived as real, because in everyday real life situations there is always parallax movement happening as soon as an individual is moving around. In a still image you don't have any movement in parallax. That's why people came up with the idea of separating the screen replacement in different layers, so the elements can shift in parallax to each other, just as matte paintings on layered glass (factor of parallax is depending on movement of the parts in correlation to each other). But with this 2.5D approach it is often difficult (or at least a lot of work) to fit in moving elements within a scene, so often still photographs are used and then animated for shift in parallax. Sometimes even the 2.5D approach isn't enough, especially for objects close to the camera, exposed to a big amount of movement or rotation. For up close objects, textures are often projected onto 3D geometry (with a low polygon count). The limitation in viewing angle is still given hereby, because the projected matte painting only covers a specific camera movement. The projection falls apart when the virtual camera deviates above a certain amount off its set path and will reveal smearing or distorted textures.

[217] cf. LightRay Visual Effects (2012)

[218] cf. Okun and Zwerman (2012) 575

4.1.2 3D Modelled Environments

While Projection Matte paintings have become the main approach for realistic Virtual Backlots in film production, in gaming fully modelled CG-Environments are necessary in order of freedom in terms of movement. But, even in movie productions it's not as uncommon as one might think. To gain flexibility and especially to freely move the camera in a synthetic landscape, 3D environments are becoming more and more popular. Due to the fact that you have a totally modelled and textured environment, it can be used not only for one sequence, but for all the sequences requiring a similar background setting, virtual set, set extension or others alike. Changes in terms of lighting, detail in modelling, retexturing or replacing objects can be done at any time without having to acquire new footage. 3D-modelling a whole environment means a lot of work, but sometimes it's worth it. Modelling a whole environment means paying attention to a lot of details. By texturing the geometry, every viewing angle is possible without any smearing, as it would happen when projecting onto geometry. In terms of realism, often movement of various items in the scene is simulated, i.e. trees, water, crowds.

Another big advantage of 3D environments is the fact that a director could freely try out different perspectives and movements with technologies like James Cameron's Virtual Camera, which was developed for the movie Avatar in order to see the actors' live performance on the motion capture stage in the virtual surroundings.[219]

As Anne Thompson stated:

> *This virtual camera allowed Cameron to shoot a scene simply by moving through the volume. Cameron could pick up the camera and shoot his actors photographically, as the performance occurred, or he could reshoot any scene by walking through the empty soundstage with the device after the actors were gone, capturing different camera angles as the scene replayed.*[220] – Thompson (2010)

4.1.3 State of the Art Technique

Nowadays, companies like Stargate Studios offer a photo real library of thousands of high quality live backlots shot at different locations around the world.

Stargate Studios' archive named "*virtual backlot*" is based on creation of 360-degree nodal point-panoramas and additional digital still photographs, which are then projected on to 3D geometry for the feeling of depth

[219] cf. Thompson (2010)

[220] Thompson (2010)

and making parallax movement possible. In terms of lighting the actors and partial sets still have to be lighted, like the captured backlot. As for CG Elements, a lighting map is extracted from the captured environment and transformed into a lighting map for any CG-Element in the environment.[221]

Stargate Studios' used projection-technique, giving great flexibility in camera movement and angles while shooting. If you use traditional techniques, you have to exactly matchmove the set-camera to the backlot-camera in order to realise a realistic movement of the combined scenes. Therefore it's always necessary to especially shoot the backlot for the wanted scene.

4.1.4 Possibilities of using a Light Field Array

By using a light field array, some restrictions are broken, but others come to mind.

One of the greatest benefactors of using a light field array, would be the possibility of capturing a background scene without the final on-set camera movement in mind. Depending on the construction and size of the camera array, the viewing angle and even the camera movement and parallax-properties could be simulated and matched to the scene shot in the studio. Of course you could use full-environments, like proposed above by Stargate Studios, or full CG environments. But these techniques require a lot of manual work in putting the scenes together. If you already know how you will approximately shoot your scene, you can use the light field array like a normal camera on the background-set and still keep flexibility in camera-movement, -angle and -parallax. The possible movement of the virtual camera depends on the configuration and setup of the camera-array. Even the simulation of different camera-sensors and optics is theoretically possible. Our vision is to automatically match the tracked camera movement of the camera to the backlot footage. Computation of new views and perspectives is called view rendering, as stated in 3.1, and provides the possibility of creating novel views between captured views. For a backlot-scenario, this means a huge advantage, as you can change the viewpoint afterwards depending on the size of your array, the wider the array the bigger the amount of horizontal movement possible. Not only can you move the virtual camera freely on horizontal-, vertical- and diagonal- paths on the 2D plane of the capturing array, but you can also animate a movement on these axes, as well as on the Z-Axis, including correct parallax-computation. So in summary, you get the benefits of a motion control system, without having to use a motion control system and without the need to exactly know the camera-path while shooting.[222] Not only the camera movement comes to mind, but also the camera and

---

[221] cf. Kaufman (2006)

[222] cf. Okun and Zwerman (2012) 173ff

camera-optics characteristics, like field of view, zoom effects and optical contortions could be simulated and matched to different camera types in postproduction.

With the use of a light field array, it is possible to realistically match the focus to a recorded scene with the help of depth information of a scene and the data collected while shooting on set. If both, the backlot and the on-set shooting were captured with a light field array, a global (de)focus could be set to both merged data-sets, being perfectly in line, with no manual matching necessary. The simulated depth of field either relies on computed depth data or synthetic aperture rendering, both possible with light fields. Also elements (window frames, reflections, smoke) added during compositing could be placed within correct depth and receive the matching focal values. Even without knowing the final outcome in terms of focus and depth of field of a future studio-shoot, you could already acquire a backlot flexible in terms of focus. Because light field based acquisition means basically capturing the whole shot in focus with the possibility to set the factors, leading to the depth of field and amount of defocus, at a later time in production. When shooting with a light field array, additional data besides the RGB Image can be generated, one of which is depth-data. In a backlot scenario, depth data could be especially helpful in order to correctly position objects in postproduction, as well as generate masks from the depth-maps by slicing certain parts of interest and converting the depth-map into an alpha-mask or matte, such masks could be used for grading depth layers differently or to remove/replace certain elements. Depth layers are also of use when the addition of volumetric components is planned, by dividing the sequence in different depth layers, fog for instance could move through the scene realistically, being (partly) occluded by certain objects. Depth data is not only of interest for keying purposes but also for matte extraction purposes where only certain areas of the image should be influenced, like grading, painting, sharpening etc.. With increasing quality of depth maps, the demand for depth keying increases. Traditional keying is based on color or luminance differences as mentioned earlier and often leads to problems with color spill or similar colors within the scene. Depth keying segments foreground objects from background objects using the objects relative distance to the camera. At the moment depth data is mostly available for CG renderings or Scenes filmed with additional ToF cameras relying on stereo vision algorithms. If the need for depth data arises during post production, depth maps are often manually created by hand drawing shapes.

Due to the possible extraction of normal-data, it is possible to match the lighting of a recorded scene to another in a certain range. It is possible to realistically add light-sources and relight objects. By constructing a partial 3D Model of a Scene, it is also thinkable to draw shadows from new light sources. The topic of lighting and relighting is further explained in the packshot scenario in chapter 4.2, which is why we won’t go into detail just here. In our case we used several advantages of the light field array.

4.2 Packshot

The product shot or *packshot* is the climax of any commercial spot. It shows the advertised product, usually in the most perfect way possible. Often the shoot of these parts takes a lot of time and consequently money to come up with results that the client and hopefully the audience will like. Since every aspect of the shot like, timing, shape, framing, composition, focus and lighting has to be perfect, it often has to be reshot several times in different versions.[223] In postproduction then, again, time is spent to further refine the shot or at least do some retouch. To gain absolute control over every aspect, while achieving ever new visual styles and effects it is common to use computer graphics in this context.[224] But it's not always possible to create full-CG environments[225] depending on the production budget and content of the shot. Additionally, the fast turnaround-time in commercial productions are challenging VFX productions.[226] Still visual effects can be often spotted in the context of commercials to help getting and keeping the viewer's attention. In our case we tested three applications in the context of the packshot, which consisted of a teapot and cup on a turntable in front of a studio wall with textured wallpaper. 1. We used the depth data to generate a matte for the product and the turntable to be able to replace the background. 2. The existing teapot and cup was relit with moving effect lights as well as an environment map to blend the foreground with the replaced background. 3. Additionally, we also used a CG rendered teapot and integrated it into a clean plate using relighting of both the CG and live-action element of the shot. Finally digital focus was applied.

Lighting is one very important aspect of filmmaking and particularly when presenting products. Since the basis for digital focus and view rendering has already been covered in 4.1, the following chapters will focus on strategies and workflows in lighting and relighting a scene in postproduction. After giving a short introduction on cinematic lighting and the term relighting we will present some approaches to relighting as they are available today. Based on that we are able to refine the definition of relighting with the proposition of three classes of relighting. From there we take a look at how the 3D data that can be extracted from light field data might improve the process keeping in mind the challenges of the packshot scenario.

[223] Siragusano meeting records 6

[224] Bunish (2012), already 1999 an article wrote about the increasing use of VFX in commercials: Ring of Fire Advanced Media Team (1999)

[225] cf. glossary "VFX", a shot that is solely put together of computer graphics elements, often with the intend to simulate real photography

[226] cf. McAuliffe (2007)

4.2.1 The Importance of Lighting

> *In motion pictures, lighting is not only used to help actors and sets look their best, but as an integral part of storytelling to set mood, direct attention, and underscore performance. This importance is reflected in the high proportion of time and expense spent on lighting: by many estimates, one half or more [Trumbull 2000] of the valuable time spent on a set is involved in setting up the lighting.*[227]

A key goal when setting the lights is to preserve the three-dimensionality of an object in the two-dimensional projection.[228] The shadows and reflections on surfaces give our biological visual system important cues on material properties and scene depth.[229] Light coming from the side often creates a dark atmosphere and feelings of loneliness or melancholy, whereas light from the top-front is usually perceived as neutral and positive, for example.[230] And these purposes of lighting are basically the same in animated computer-generated film and live-action picture as Sharon Calahan states.[231] The Director of Photography (DP) is in charge of designing the set lighting together with the director in terms of the emotional aspects. When combining elements shot at different locations or times as well as live-action and CG as part of a VFX process, the lighting between those elements has to match as closely as possible to achieve the goal of photorealism. Or in other words, differences in lighting easily give away the visual effect.[232] In contrary to other post processes like editing or sound mixing, the lighting of a live-action element traditionally can't be iteratively refined and improved after principal photography.[233] In postproduction the VFX supervisor or producer usually decides about the visual style.

4.2.2 Relighting

"Unfortunately, there's not always money or the time to accomplish detailed lighting setups for every scene."[234] That's especially the case with independent productions or documentaries. For full animation features the statement seems to be valid, too. Digital lighting is computationally expensive and, up to date, can't be fully evaluated in an interactive real-time environment, which makes it a costly process as well.[235]

---

[227] Wenger et al. 1
[228] Rausch (2013)20
[229] Rausch (2013) 21-25
[230] Rausch (2013) 44, 61
[231] Calahan (1999) 337-382
[232] cf. Whitehurst (2010) 656
[233] Wenger et al. (2005) 1
[234] Van Hurkman (2014) 347
[235] Saam (2010) 1

Visual Effects often face additional constraints. When shooting action in front of a green screen the lighting of the foreground often must be chosen before the virtual backgrounds are finalized, posing difficulties for achieving consistent illumination between foreground and background elements.[236] This could be the case in the virtual backlot scenario. It gets even more complex when new shot elements are introduced in postproduction that are light emitting objects and therefore need to interact lighting-wise with the already shot elements. In this particular example you would normally use a stand-in object or light on set to approximate the effect, which needs careful planning and possibly extra time on set, while limiting creative possibilities in postproduction.

In the context of the packshot scenario or portrait the tasks would be corrections for continuity, visual style and contrast as they would be addressed as part of the color grading or Digital Intermediate (DI) process most likely. Also relevant in the context of the packshot is the concept of seamless integration of CG elements, possibly the product that is only available as a CAD[237] model yet. There might also be demands for changes in product design and appearance after the spot is shot nowadays.

There have been a variety of different concepts presented in related publications on relighting. Based on the information on the website of the popular research project at the USC we define Relighting in general as techniques that allow to modify and design the lighting of an object or scene before or after the moment of initial recording or rendering.[238]

As relighting is a computer based process nowadays, it is done on digital images that could have been created by a camera or a 3D renderer. Most of the previous work is based on relighting computer generated imagery with attempts to save re-rendering cycles during the early lighting design stage. Since the ultimate goal for VFX productions would be to work in an integrated environment of computer generated graphics and live-action footage, we don't want to draw a line between CG and live-action footage in this context. Instead, in chapter 4.2.5 we will propose classes of relighting that are defined by the degree of freedom in parameters like shading and three-dimensionality.

---

[236] Wenger et al. 1

[237] cf. glossary "Computer Aided Design"

[238] "Light Stages"

4.2.3 Current State of the Art

Following, four approaches to changing the lighting of a scene are showcased, 2D methods based on masks for image segmentation, the light stages using time multiplexed illumination, 2.5D relighting using view dependent 3D data and prelighting as part of a lighting design workflow. This enables us to further refine the definition of relighting as a use-case of 3D scene data and to group the techniques in a system of classes.

4.2.3.1 Relighting in DI using 2D Shapes

Already at the time of analog film development there were simple techniques available to change the perception of light with masks and dodge or burn processes.[239] Nowadays, "Shape-based digital relighting" is a common procedure as part of the DI or color grading process.

> *The majority of films [...] now go through a creative grading process at the end of the postproduction pipeline. This stage results in the production of a digital intermediate and the process itself is now generally referred to as DI. The DI involves creatively changing color and contrast in filmed footage to create the finished look that the Director and DP designed.*[240]

A specialized operator, the colorist, usually uses just as specialized hard- and software for this task. Other than that there is no additional data needed apart from the 2D image. As Van Hurkman describes in his "Color Correction Handbook", one common correction is to cut down the ambient light in a shot to direct attention on important subjects.[241] This has traditionally been done with vignettes and is often realized using hand drawn shapes with soft edges nowadays. If the subject or camera is moving the shape's position needs to be animated using keyframe animation or the data from a motion tracker.[242] Accordingly the feeling of light can also be added to a scene by brightening a masked area of the image. Often human faces and especially the areas around the eyes are brightened to achieve an effect similar to using bounce cards on set.[243] Interestingly, oval shapes seem to work best in this context as they are "completely abstract and in fact end up looking a bit more similar to the effect of shining a practical light on the scene".[244] As the subject moves relative to a light source also the intensity of the correction needs to be animated, of course. Furthermore the corrections are limited by the ability to create shadows, which have to match inside and

[239] Ansel Adams used and wrote about dodging and burning extensively in his book „The Print" in the context of his zone system, cf. "Basic Darkroom Techniques"
[240] Whitehurst (2010) 659
[241] Van Hurkman (2014) 338-339
[242] Van Hurkman (2014) 355
[243] Van Hurkman (2014) 340, 356
[244] Van Hurkman (2014) 342

outside the corrected area for the correction to be invisible.[245] Another common goal during color grading is to add depth to an image. Colorists often use gradients to add an artificial falloff to the light in the scene.[246] Luminance shading and according to the neurobiologist Margaret Livingstone specifically gradients from dark to light trigger the brain's "where" system.[247] Other monocular depth cues that are simulated in grading are haze and airlight[248] as well as a reduction in saturation and texture, for example.[249] For complex scenes van Hurkman suggests to use custom shapes, again, to add "visual interest to an otherwise flat-looking scene".[250] The colorist would use soft polygonal shapes to segment the 2D image space based on depth layers in scene space. Subsequently the mid part of an image could be emphasized by bringing up the brightness, while adding shadow in the foreground and less contrast in the background to tell depth, for instance. Figure 21 shows a flat looking image from our test shoot. A hand-drawn mask and a gradient were used to make this still frame a bit more interesting and guide the view towards the face of the actor. The methods mentioned mostly affect the diffuse lighting component. To make a scene that has been filmed indoors feel like it's outdoors is a more challenging task, for example, as highlights need to be boosted or created. Some corrections can be made by selecting areas based on image luminance and then editing these accordingly.[251] Similar problems also occur when shooting day for night or vice versa. Here it might also be needed to *switch on* practical lights to tell the nighttime. In the context of the packshot it might be asked by the client to make the product look more shiny or glossy, which can't be done in DI. Sometimes results can be improved by multiplying the shape or luminance mask with an animated element like a procedural noise or smoke footage to create modulated values that have an organic feel to them.[252]

[245] Van Hurkman (2014) 341
[246] Van Hurkman (2014) 345
[247] Livingstone and Hubel (2002) 50-52, 109-112
[248] hue shift based on the scattering of light in the atmosphere that gets visible in the distance, cf. Aditi, Gupta, and Dua (2014) 200
[249] Van Hurkman (2014) 344-355
[250] Van Hurkman (2014) 350
[251] Squires (2014) 685
[252] Squires (2014) 685

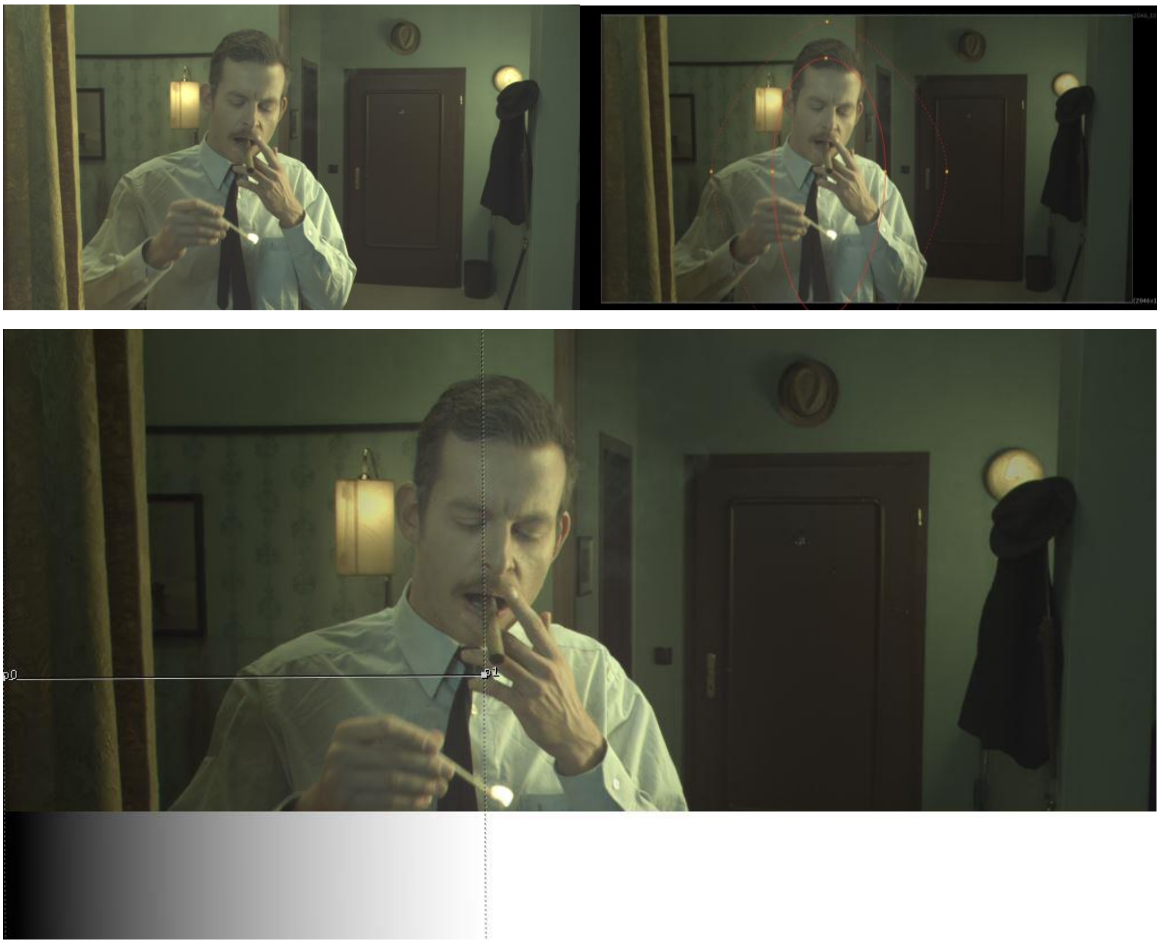


Figure 21: A rather flat looking image (top left) is color corrected, using a soft mask to darken the background (top right) as well as a gradient to emphasize the light direction (bottom).

4.2.3.2 Relighting with Time-Multiplexed Illumination

Another approach to relighting is the basic idea of capturing a scene under different basis lighting conditions. In postproduction the lighting then can be changed using linear combinations of the recorded lighting variants.[253] The research group at the University of Southern California (USC) around Paul Debevec is leading the research in this area and advanced the idea of relighting performances by capturing reflectance properties, too.[254] Performance relighting with time-multiplexed illumination, as it is called, uses specialized acquisition devices referred to as light stages.[255] For example, Light Stage 5 (Figure 22), sequentially

[253] e.g. (Horn and Chen) (2007) 5, (Debevec et al.) (2000)
[254] Malzbender et al. did similar research that is also referenced by Wenger et al.: Maizbender, Gelb, and Wolters (2001)
[255] Wenger et al. (2005) 3

numbered according to completion time, is a "2m diameter once-subdivided icosahedron, with the lowest five faces left open"[256] that is equipped with 156 equally spaced LED lights shining on the actor's face in the center of the apparatus. The performance is recorded by a high-speed camera with 2160fps that would capture 180 multiplexed lighting conditions in a time interval that is one 12th of a second, in this case. 10 diffusely lit Frames for motion estimation as well as 10 Frames with only a grey background illuminated by separate light sources for matting are added to the 156 lighting states. To increase the light intensity and thereby improving the SNR while reducing the strobe effect different lighting basis pattern were tested. A triangle pattern produces good results with a small cut in lighting resolution.[257] To compensate the motion between the 180 frames optical flow algorithms are used that interpolate the motion field between the diffuse lit frames. A 24fps output can also be achieved with optical flow retiming rather than doubling the frame rate to overcome limitations in light intensity and camera buffer size.[258] Finally motion blur is computed that fits to the final frame rate and shutter time. The captured data can then be used to estimate surface normals, albedo and later ambient occlusion[259] data per pixel and per frame, using a variant of photometric stereo and Lambertian reflectance model based on Woodhams publication from 1980.[260] To avoid errors caused by highlights or dark spots in the texture the pixels are filtered based on their luminance values before normal estimation. The normals can also be used to compute an ambient occlusion pass that describes the geometric self-shadowing. The generated data makes it possible to change reflectance parameters in a not totally physical accurate but straightforward way.[261] By dividing the matte frames by a clean-plate image of the background, a matte for compositing with a background image can be created.[262] A HDR spherical map of the environment could then also be used to relight the foreground accordingly.

---

[256] Wenger et al. (2005) 3

[257] Wenger et al. (2005) 5

[258] Wenger et al. (2005) 6

[259] When thought of a surface as a grid of small planes that are oriented in the direction of the surface tangents, normals are the vectors perpendicular to the surface planes; the albedo, a term introduced by Lambert in 1760 or reflection coefficient describes the diffuse reflectivity of a surface (cf. "Albedo"). This will be covered in more detail in chapter 4.2.3.3.

[260] Woodham (1980) 139–144

[261] Wenger et al. (2005) 7

[262] Wenger et al. (2005) 7

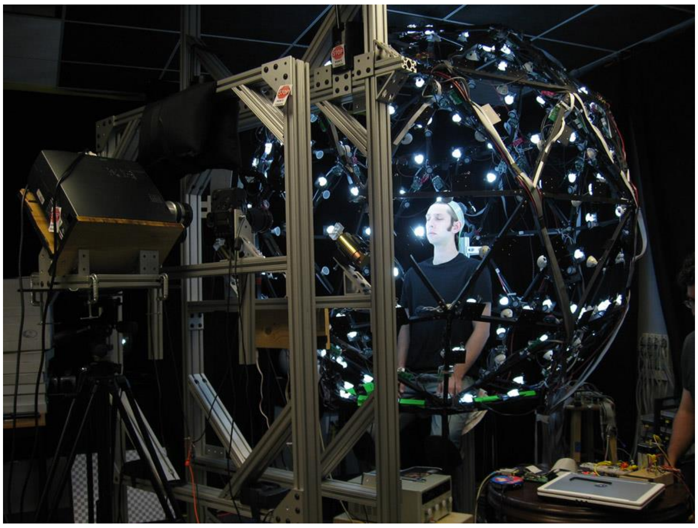


Figure 22: LightStage 5 used during the production of "Avatar" – actor Joel Moore is scanned for his character Norm Spellman

The disadvantage of this approach still is that it is highly data intensive and therefore recording time is limited by the buffer size of the high-speed camera, for example. As you need a controlled light environment and above all enough light in the first place, it needs special stages that offer only a limited space for performance. Also a spatially-varying specular map of good quality cannot be obtained with this setup due to the noise and the coarse lighting basis.[263] While giving surprisingly good results on the hair, the eyes definitely need to be relit with a more precise lighting model that takes spatially-varying specular reflections into consideration. Light Stages 6 to 10 address some of these issues as they got bigger and stage six could already capture a whole human body that performs on a rotating platform.[264] While being limited to only diffuse relighting, due to the limited number of captured lighting conditions and cyclic human motion like walking, Light Stage 6 allows for the change of the viewpoint. It captures a 7D dataset that can be seen as a light field and exists of 2D images over time (1D), illumination (2D) and viewpoint (2D). This enables a rendering process of image based relighting, image warping, light field interpolation, shadow rendering and compositing.[265] The view rendering is based on the plenoptic function mentioned in chapter one.[266]

Light Stage models 5 and 6 can and already are used as scanning devices for objects and human actors alike. It has been successfully used on productions like "Spider Man 2 and 3" (still light stage 2 and 3), "The Curios Case of Benjamin Button", Peter Jackson's "King Kong" and "Avatar" to create digital versions of

[263] Wenger et al. (2005) 7, Einarsson et al. (2006) 10
[264] "Light Stages", Einarsson et al. (2006) 3
[265] Einarsson et al. (2006) 7
[266] Einarsson et al. (2006) 8

actors. The Light Stage systems have been licensed to the company LightStage LLC, which offers commercial scanning services to the motion picture and interactive entertainment industries.[267]

4.2.3.3 2.5D Relighting

Since this approach to relighting has been used in the context of the test shoot, we will discuss it in more detail and also cover some basics of the shading process in computer graphics. How relighting was achieved in the context of the test scenarios will be presented as part of section 6.3.

A hybrid or 2.5D[268] approach to relighting is used to change or add lighting to 3D computer graphics after the 3D rendering step. It can be done inside a compositing package like Adobe's After Effects or The Foundry's Nuke, the latter bringing a relight tool as one of the standard nodes since version seven. These techniques are well described by Johannes Saam and Daniel Kubacki, for example.[269] To avoid the expensive step of iterative re-rendering computer graphics, the basic idea is to speed up the pipeline at some point by allowing for changes of the final image inside a compositing package without the need for any additional pre-computation. These techniques use only the 2D image and data passes that are outputted by the renderer at no or very little extra cost.[270] These additional passes are called Arbitrary Output Vectors (AOVs) and usually include a depth channel that codes the distance between the (virtual) camera and the scene object as a grey value for every pixel. As Saam shows, a depth channel is basically sufficient to calculate the data needed for relighting, which is a normal vector and a position in the 3D coordinate system for every pixel.[271] This information is also coded in rgb pixel values and can and usually will also be saved out during rendering, hence eliminating the need to recalculate them, since the data is needed for the rendering process anyway. The relighting itself is then done as a simplified shader operation or simulated 3D lighting.[272] The visibility check that normally is the first step in a rendering process is skipped as it is already burnt into the 2D image.

[267] LightStage is now part of the company OTOY that is best known for its GPU-based ray-tracer Octane; http://render.otoy.com/contact.php, OTOY also develops a light field viewing software at the time of writing; http://home.otoy.com/render/light-fields/, OTOY (2015) (press release)

[268] 2.5D meaning techniques "used to create the appearance of a complex 3D environment with a much simpler and faster 2D toolset", it often involves the use of simple, approximated geometry with a low number of polygons and a limited range of camera viewpoints, Spears (2010) 689

[269] Saam (2010), Kubacki (2009), Hassan Uriostegui showed an implementation with plotted point passes for Nuke in 2012, Double Negative used tools similar to the nuke relight node based on expressions already in 2009, cf. Roy Stelzer in Nuke Masterclass 2009, "Nuke 2009 Masterclass - 2.5D Re-Lighting (Video 2)"

[270] Saam (2010) 1

[271] Saam (2010) 1-2

[272] Saam (2010) 2

There is no explicit geometry involved. Shading means the process of giving each sample on a surface a value. A function or small program that does this is called a shader. A surface shader usually determines a color value for a certain pixel value based on its shading model, which describes on a basic level how an object responds to light.[273] To calculate a simple Lambertian diffuse shading the only thing missing now is a light vector which can be freely chosen. Then the new color value can be calculated as the result of the dot product between the incident light vector and normal vector. This relationship of the light and normal vector is also known as the cosine law in optics.[274] A light falloff can be added multiplying the length of the light vector, that is the distance between the shaded point and light in 3D space. The resulting grey value can then be tinted by any color, i.e. the light color.[275] The surface normal vector or normal for short refers to the term used in computer graphics and describes a vector that is perpendicular to the polygon plane of the 3D geometry.[276] In the case of 2.5D relighting every pixel can be seen as a small rectangular plane that has its normal. The vector parameters can be stored as red, green and blue values in a standard rgb file. The coordinate system is usually oriented relative to the camera plane, meaning the z-axis is parallel to the viewpoint vector. Depending on the origin point the 3D position data that is stored in red, green and blue values accordingly can also be in camera or projection space, in an object dependent or in an arbitrary world coordinate system.[277] Usually they are in a world coordinate system that is the same for all elements of a shot.

To get more realistic lighting for a greater variety of simulated surfaces the Lambertian shading model can be extended by a term for specular reflections based on Phong's work from 1975:[278]

$$f_{Phong}(\vec{l},\vec{v}) = \frac{c_d}{\pi} + c_s(\cos\alpha)^n$$

Figure 23: A simple shading equation based on Phong's reflection model

---

[273] Spears (2010) 672-673

[274] Phong (1975) (1975) 312

[275] Saam (2010) 2

[276] Geitz (2007) 15

[277] cf. Langer (2010) 1; to get things straight concerning coordinate systems a short definition is given: in CG every object has its model coordinate system, that defines the model's orientation and center of rotation, the object position in a scene can be defined in a world coordinate system. To model the image projection process inside a camera two more coordinate systems come into play. The camera coordinate system is a transformed world coordinate system with the origin as the center of projection inside the camera. This coordinate system can be tuned into the projection space by normalizing the coordinate in a zero to one space, which facilitates projection to image space.

[278] Phong (1975) 315

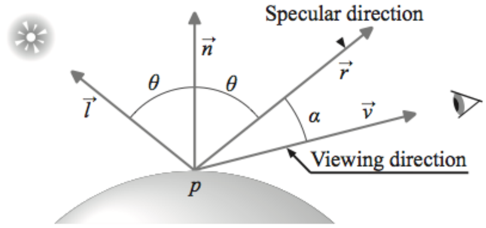


Figure 24: Geometry for the Phong model and general cosine lobe specular models.[279]

The specular reflection is controlled by the specular coefficient c_s and the specular exponent n that controls the falloff of the highlight. $\text{Cos}(\alpha)$ is the angle between the viewpoint or camera vector $\vec{v}$ and $\vec{r}$, which describes the specular direction that is dependent on the light direction $\vec{l}$ and the surface normal $\vec{n}$). Accordingly c_d is the diffuse coefficient. (figure 23 and 24) Having the freedom to change those shading coefficients you can essentially not only relight but also reshade the scene. This is also a restriction of this approach because you have to reshade your new lighting pass with a simplified shader unless you are dealing with computer graphics that were shaded with the same, known shaders as those available during relighting.

As we simulate real-world lighting the light intensities and effective pixel values are in linear relation to each other. Therefore, when finally adding the light to the picture this has to be done in scene-referred linear color space.[280] Generally, relighting should be done in this manner to remodel the real world process as accurate as possible and get predictable results. At the same time, relighting operations are extreme image editing operations and thus, benefit from high bit depths in the source material.

Since this kind of relighting is limited by the underlying vector math, only direct lighting with point, spot or directional lights can be used. Additionally, it can only be as good as the shading model used.[281] But in reality the more import drawback of this technique is the limited possibility to create shadows, which are needed for realistic lighting. This is discussed in more detail in 4.2.4. When relighting 3D an ambient occlusion pass is often available as AOV that holds information about self-shadowing of an object.

Nevertheless 2.5D relighting can be extremely useful to add small lighting effects or improved interaction in compositing. In the scenario of the packshot this could be the possibility to add some more fill-light to remove an unwanted shadow on some parts of the product or to add glow and light emitting particles to a

[279] Kurachi (2007) 218
[280] cf. Science and Technology Council (2013), Siragusano meeting records
[281] The shading models of Phong and Lambert are empirical approximations that work well under some circumstances and tend to give unrealistic results under others, cf Kurachi (2007) 220, Whitehurst (2010) 654

shot. In the context of VFX it is often used to add the light of a fire or explosion element to CG elements in a shot.[282]

4.2.3.4 3D Scene Prelighting for Look Development

In computer graphics the "process of balancing shaders, maps, and lighting until they match the reference is called look development"[283] Based on Gershbein and Hanrahan's work on "Interactive Cinematic Lighting Design" in 2000[284] the Lpics and later Lightspeed system at Pixar[285] is an example of a software tool that enables lighting Technical Directors (TDs) a more efficient look development workflow. This kind of relighting takes place before the final rendering stage and usually needs pre-computed data. To better understand the idea of relighting and the concept of integrating 3D computer graphics it makes sense to take a look at the lighting design workflow used in production first.

Usually the lighting process can be divided into two stages, the master lighting and the shot lighting. During master lighting the light for the background and the props of a sequence is set. The lighting artist selects static viewpoints on the scene that are characteristic for the sequence. The shot lighting refines the master lighting for a particular part of the sequence with a continuous camera position with the focus on the character and storytelling. The typical length of a shot is one to ten seconds and again the artist works only on a limited number of key frames.[286] At the moment the master lighting often involves Image Based Lighting (IBL) techniques, while, during shot lighting the artist often places extra lights to match the mood of other shots in the sequence. Usually it makes sense to only place one light at a time and check the result against a potential background plate and in context of the sequence regularly. Once the artist and supervisor is satisfied the shot gets rendered and delivered to the next production stage, which would usually be compositing.

The basic idea of Lpics is to interactively relight scenes based on cached data in image-space like displaced points and normal or calculated surface color on a per pixel basis and therefore allowing scenes of extremely

---

[282] cf. Duggal (2012)
[283] Whitehurst (2010) (2010) 656
[284] Gershbein and Hanrahan (2000)
[285] we chose Pixar as we have a good documentation of the different development stages of the relighting workflow at hand, but basically all big production houses at least had their own toolset for prelighting similar to Lpics
[286] Calahan (1999) 337-382

high geometric complexity. This helps to light, that is to say the placing and editing of the light shaders, the keyframes in master and shot lighting interactively but with a fixed viewpoint.[287] The light shaders, in the case of Lpics RenderMan shaders, are “simplified and translated into a hardware shading language” but keep their controls and attributes.[288] The surface shaders are executed by a conventional software renderer and the output is loaded as caches together with simplified geometry for shadow generation into Lpics. The rendering is then executed on the graphics processor (GPU) in realtime. This approach has been developed and further refined by many others, like Haan who implemented a full indirect illumination model using a wavelet basis.[289]

The final render usually still takes place on the central processing unit CPU to get the last five to ten percent of quality, although GPU based renderer are used from time to time as final image renderer nowadays.[290] If the scene hasn’t changed in terms of geometry and shaders the graphs could even be updated and the rendering could continue from the pre-computing stage without having to rerender the whole scene. Pixar recently moved to The Foundry’s Katana[291] for lighting and uses a GPU accelerated ray tracer for lighting preview that also allows to adjust the viewpoint and surface shaders freely at the lighting stage.[292] This trend can also be found at MPC who use a tool for fast lighting preview building on Nvidia’s Cuda based ray tracing engine OptiX that emerged from their Previs Engine called “Muggins”.[293] As Jedrzej Wojtowicz, head of the Shaders Department at Weta Digital points out, the idea of calculating the same shader models in preview and during final render is one of the key aspects of these technologies.[294] Since the movie Avatar Weta used to have its own tool “PantaRay” for relighting and lighting design that pre-calculated visibilities in a scene as spherical harmonics and allowed for fast diffuse lighting of very complex scenes by reducing the lighting to simple texture look-ups.[295] Weta’s new proprietary path tracer Manuka brings a prelighting and preview tool called Gazebo that is basically a GPU optimized and slightly less accurate version of the final frame renderer. Since all ray tracing algorithms produce a noisy image at first that gets progressively better, "it is not too difficult to implement on top of a path tracer some kind of re-lighting, interactive

---

[287] Fabio Pellacini et al. (2005) 467
[288] Fabio Pellacini et al. (2005) 464-465
[289] Haan (2006) 1
[290] cf. Nahmias (2014)
[291] Katana is an asset based software tool for lighting and look development that is highly customizable and can host a variety of renderers and shader libraries, http://www.thefoundry.co.uk/products/katana/, last visited 20.04.2015
[292] Nahmias (2014)
[293] Montgomery (2014)
[294] Seymour, “Manuka” (2014)
[295] cf. Seymour, “The Science of Spherical Harmonics at Weta Digital” (2013), Green (2003) 1-4

rendering kind of approach, and most renderers (which are path tracers) do that", says Luca Fascione, Rendering Research Supervisor at Weta Digital.[296] Additionally, the previously described techniques for 2.5D relighting can be used on that data, too. Oftentimes look development is also done in compositing software through blending of different light passes that have all been rendered with the same intensity. The mix values that can be evaluated in an interactive way are then transferred back to the shader setup in the 3D package.[297] In this area clearly the lines between relighting and lighting as well as preview and final image are more and more blurring in favor of more efficient workflows.

4.2.4 Shadows and Relighting

"Lighting is creating shadows".[298] Let alone, the word shading describes the importance of shadow to help telling a three dimensional shape in a 2D projection of reality. Besides, shadows play an important role in the perception of highlights on semi reflective materials like water, metal, glass or wet skin.[299] But as Steve Wright states, "Shadows are actually fairly complex creatures"[300] that are hard to model realistically.

As we can use normal lights with a positive intensity value, also lights with a negative intensity could be used and actually have been in use in computer graphics to correct the shot lighting in specific areas. When adding the negative intensity to the final image areas, which are affected by the light get darker.[301] This might be helpful for minor corrections when balancing light intensities throughout a shot but also poses some problems. Firstly, it is not very efficient to have to set up at least two lights to simulate one light source. Secondly, negative lights are risky because such setups don't obey the laws of physics like the law of conservation of energy, for example.[302]

Saam proposes the use of *bent normals* to fake shadows.[303] Bending normal vectors is a technique that is popular in game development and simulates self-shadowing by changing the normal vector's direction based on a calculated occlusion value. In computer graphics bending means averaging the direction of all unoccluded rays that are obtained during construction of the ambient occlusion map.[304] Still this doesn't

---

[296] Seymour, "Manuka" (2014)
[297] Xavier Bernasconi (2013), CG Supervisor Industrial Light & Magic
[298] Rausch (2013) 20
[299] Rausch (2013) 21
[300] Wright (2010) 225
[301] Whitehurst (2010) 661
[302] Whitehurst (2010) 661
[303] Saam (2010)
[304] Kurachi (2007) 212

address the shadows that are created on other objects in a scene like the ground, for example. Currently, ways to generate those are the use of approximated or a low-polygon version of the geometry in a fast 3D setup using projected shadows. These can often be accomplished inside a compositing application or going back to a full 3D renderer to only render the changed shadows.[305] The latter should still save some render time compared to a re-calculation of the color information but lacks interactivity. If there is no meshed geometry available it is possible to generate a point cloud from the position or depth pass in the context of 2.5D relighting. The point cloud can then also be used as light occluding object using projections or even a ray-tracing renderer.[306]

Again, the presented techniques don't cope with the shadows that are already in the picture you want to relight and which can cause serious problems selling the new light. Thus, it might make sense to consider the amount and kind of relighting you plan to do when setting up the lighting on set. In some situations a diffuse lit scene might make sense. Interestingly in the context of the packshot often a lighting situation with soft shadows is pursued as it is perceived as glamorous and noble. Apart from that, hard shadows are better avoided in any image analysis processes since they are no 3D geometry but might be interpreted as such by the 3D extraction algorithms causing false results.[307] Especially if the shadows are moving they should be masked out and left out of the computation or processed with specialized algorithms.

4.2.5 Classes of Relighting

So far, we presented some examples of relighting techniques that are employed at different stages of current production workflows. Based on this short review on industry practices we propose a classification of relighting approaches suitable for live-action footage to then highlight the possible improvements and possibilities that are gained using light field data. We define three classes that are characterized by the degree of freedom in postproduction. Similarly Choudhury et al. defined in their "survey of image-based relighting techniques" from 2006 three groups of image based relighting techniques based on their algorithmic structure.[308] Our approach tries to be more end-user oriented as it has the level of artistic control in mind.

[305] cf. Saam (2010) 2
[306] cf. Saam (2010) 2
[307] cf. Verma and Wu (2009) 1
[308] Choudhury and Chandran (2006) 176-177

It is inspired by the way the ideas regarding relighting were clustered during presentation at the experts meeting mentioned in chapter 3.2.[309]

The first class shall include 2D relighting and all processes based on masking and image segmentation techniques with 2D shapes using the standard color correction toolset as described above. This kind of relighting is only valid for one viewpoint and has to be adjusted manually for a moving camera or object. There is no specific representation of a light source or surface material. Masks or qualifiers could be created using procedural color or luminance keys as well as extracted 3D data like depth or position information.

As the next level or class two we see 2.5D relighting that enables the artist to work in a limited three-dimensional representation of the scene and uses synthetic shaders. Different kinds of light sources can be freely positioned and can have physical parameters like a light falloff. The surface material is simulated by a generic shader model like the one by Phong, which was described above. Like this reshading, i.e. changing the material properties like diffuse color or specular reflectivity, becomes possible, too. More complex effects like refraction or environment reflections can also be simulated in the context of this relighting class. Different shading models can be used and changed at any time without redoing the whole setup. Still there is only implicit and view dependent 3D geometry available which might limit the creation of shadows. Scenes with lots of objects or objects with materials of different reflectivity need to be simplified by one averaged shading model. Alternatively, a masking process could also be added to the workflow to split the scene into several relighting setups with different shading models. The possibility to change the surface parameters and light parameters independently and in a procedural way is probably the most important advantage compared to class one.

A special case of relighting is the use of an explicit 3D geometry that might be animated as well, to relight a live-action scene. This geometry can be created using traditional polygon modeling techniques or may involve 3D scanning and image-based modeling like in the case of the above mentioned example of Naomi Watts' digital double in King Kong.[310] This might mean that an element goes through the whole 3D pipeline just to create an additional light pass. These kinds of scenarios we also attribute to the second class of relighting as long as these use shaders that are based on empiric models.

The third class of relighting uses extracted surface reflection functions from real objects. This enables the exact remodeling of the surface properties in a scene and therefore promises to give the most realistic results

[309] Ganbar, Siragusano meeting records 3

[310] Prince (2011) 64-65

when changing the lighting drastically. The reflectivity of a surface can be described mathematically by the bidirectional reflectance distribution function (BRDF). It is "a function that relates the amount of light incident on a surface from one direction to the amount of light reflected off the surface in another direction".[311] Therefore a BRDF has to define two directions, which it does with two spherical variables (angles) each. Two more variables as parameters for surface position are added if the reflectance varies along the surface.[312] This makes the BRDF an equation of four or six variables. Since not all physical phenomena can be described with the standard BRDF, there are also enhanced versions like the bidirectional surface scattering distribution function (BSSRDF)[313] to model subsurface scattering in skin materials, for example. The BRDF can be fitted around empirical data, hence the equation above in the context of 2.5D relighting can also be written as a BRDF. But there are also methods to measure a BRDF of an existing object. For some materials like cloth, for example, it is hard to accurately craft an empirical mathematical model for.[314] In this case BRDF measurement can improve the visual quality of the shading. And it also makes sense in the context of relighting live-action footage where you don't have control over certain parts of the scene individually to apply different shading models. Most of the measurement techniques are based on images taken from different viewpoints and with different light positions along a sphere.[315] Often a static main camera is used in conjunction with a second camera that is movable.[316] It makes sense to capture HDR images to overcome the limited dynamic range of available digital cameras. Each pixel in the images serves a sample of the BRDF. Knowing the surface normal from 3D data extraction between views and the light position a light ray can be traced from the camera viewpoint to the light source giving the ratio of the outgoing radiance to the incoming radiance for a surface point.[317] Greg Ward describes a robotic device for the measurement process in "Measuring and Modeling Anisotropic Reflection" from 1992, called the gonioreflectometer.[318]

Although this class of relighting certainly has the advantage of being physically correct and photo-real it might not be that useful for current production methods. Often it is needed to adjust every part of the shading equation to realize a director's vision or to create materials that do not exist in the real world or are

311 Kurachi (2007) 213-214
312 Kurachi (2007) 214
313 cf. Jensen et al. (2001)
314 Whitehurst (2010) 654
315 cf. Kurachi (2007) 228-230, 234-240
316 Kurachi (2007) 234
317 Kurachi (2007) 235
318 actually Ward also invented a variant of the BRDF model, that enabled easy implementation of his technique and had an important impact on the research community, cf. Kurachi (2007) 229-230

not easily available for scanning.[319] We won't focus on this type of relighting due to the limited scope of this work and the unavailability for testing.

4.2.6 How Light Fields light the Way – possible Improvements

After showing the need for as well as different approaches to relighting and classifying different levels of relighting as they are relevant for postproduction of live-action footage and the scenario of the packshot, we will now take a look at how light field data might improve the process.

As technology advances the different approaches become interconnected. In the future software renderer like chaosgroup's vray RT[320] or Pixar's RenderMan that get available across 3D and compositing platforms might enable editing and rendering scenes in one unified environment using computer generated and live-action elements. This could result in more believable and more efficient integration of visual elements. Light field technology could be one way of capturing live-action input that enables for editing in a 3D environment. Apart from that, light field photography could enable extended editing possibilities of the parameters lighting and surface reflection for special shots like the product shot in a commercial. Color correction based on scene depth could help to emphasize the DPs vision in DI in a fast way and with improved quality compared to manual methods using shapes. Being able to position and move virtual light sources in a 2.5D or 3D environment and simulating falloff, color and surface shading would equip the colorist with new creative possibilities that could also improve the output quality for productions with limited time and budget on set. Actually, depending on the used camera configuration all classes of relighting can be realized with light field technology. For the near future 2.5D relighting with synthetic shaders will probably be the way to go. But it might also make sense to measure a BRDF of some product to enable accurate relighting of a packshot. A dense sampled light field could theoretically be used to estimate the BRDF of a surface at least for a limited viewing angle. Like we mentioned previously a BRDF estimation needs views on a scene in different lighting states. A light field usually only delivers a limited number of lighting conditions but from a number of viewpoints. In fact the plenoptic function is quite similar to the BRDF and Meneveaux and Fournier (2002) deal with the reflectance function extraction of light fields in their work "Reshading Light Fields".[321] They employ techniques more or less based on Sato and Ikeuchi.[322]

[319] Whitehurst (2010) 655

[320] http://www.chaosgroup.com/en/2/v-ray-nuke.html

[321] Meneveaux and Fournier (2002) 3

[322] cf. Sato, Wheeler, and Ikeuchi (1997), other works that cover BRDF extraction are amongst others Barrow and Tenenbaum (1978), Karsch et al. (2011)

Still the results include several assumptions as the equations are highly underdetermined. To accurately estimate a BRDF from an image parameterization of a real-world light field you would need to separate the global illumination component of the lighting first which is a whole research topic for itself.[323] An imprecise BRDF estimation from a light field recorded on set could also be linked to a database of separately measured BRDFs in postproduction.[324]

To replace the background in a packshot scenario like ours you could employ the depth data extracted from the light field to merge the main plate with another image based on each pixel's distance to the camera. This could save some time compared to conventional techniques of manually painting masks for the foreground elements and could overcome limitations of the color key process. This approach to merging images has been described in more detail in chapter 4.1.3

In this context digital focus as described in chapter 4.1.3 could also help to save time on set and during postproduction while gaining quality, that is to say correct focus on all frames, at the same time. The masking process for the foreground or background could be executed on objects with sharp edges, avoiding the problems connected to semi-transparent areas caused by defocus blur. Also the depth estimation itself, which is also needed to relight an image, benefits from sharp input images.

4.3 Portrait

The close and medium close shot are important parts of the cinematic language. They play a key role in transporting emotion to the viewer.[325] Humans are trained experts in reading and recognizing faces. In this context especially the eyes are of great importance as they carry emotional expressions and play a key role in the authenticity of the acting. We look at human faces since the first days of our lives. Already babies are most attracted by faces because they are of critical importance to the process of building relationships and attachment. [326] Therefore, we also tend to be unconsciously more aware of skin tone colors, reflectance properties and light and shadows in the area of the face compared to other scene objects. Thus a skin color that does not fit in the surrounding environment naturally, for example, can induce an uneasy feeling on the viewer's side. There is a tradition of lighting portraits that goes back to the first paintings. Together with the reflectivity that is traditionally controlled with make-up, already small nuances in lighting can have an

[323] Kurachi (2007) 240
[324] cf. Eberhardt meeting records 3
[325] cf. Moura (2014)
[326] cf. Stern (2010), Dornes (1993)

emotional impact. In a commercial spot the portrait shots often have the goal to connect with the viewer on an emotional level.[327] Examples for emotional effects of lighting directions can be found in chapter 4.2.1. In the following part the portrait shot, a classic scenario in filmmaking, is introduced and challenges as well as special VFX tasks connected to the scenario are described. Some editing techniques are presented that are currently available and address some of the potential problems. Finally possible advances in the area of post-processing portrait shots through the use of light field technology are evaluated.

4.3.1 Challenges of the Portrait Scenario

Figure 25: A final still frame from the commercial spot

An example of a portrait shot from the test shoot is shown in figure 25. In this particular example we darkened the background to direct the view to the actor and used a 2.5D relighting technique to reduce light that came from screen right while adding some fill light on the left side of the image. Additionally, some effect lights have been added to integrate the VFX elements of the shot better, the window reflecting a street in a city in the foreground. Finally, a shallow depth of field was applied using a digital focus approach. These edits show typical challenges of the scenario. Sometimes the lighting cannot be carried out in a perfect

[327] in fact everything in a commercial creates emotions on the the viewer's side. A description of the exact processes would be beyond the scope of this work. However, a common way to explain the idea of transfer of emotion between humands ist he mirror neuron system, a mechanism to imitate what we see in other humans or animals, including expressions of emotion. If watching commercials is interpreted as a learning process, a certain emotion is linked to the advertised product as a result; cf. e.g. Rizzolatti and Craighero (2004)

way due to limitations on set like time, space or weather. It might be the case that it doesn't match to a reverse shot or other shots in the sequence. When shooting with a very shallow depth of field there might be takes with good acting but slightly wrong focus that would have to be repeated or result in images of lesser quality. The postproduction of a portrait shot can also include VFX. As mentioned above not only lighting but also the makeup can control the look in a portrait shot. Methods to add digital make-up or retouch skin blemishes are increasingly used not only in zombie or vampire movies.[328] As it is still a difficult task to create human motion in a digital animation process, for visual complex characters or creature effects, CG enhancements or extensions are often added to a live-action plate instead of using a full CG character. This leads to the challenge of integration. When interaction with computer generated elements is aimed for, for example, when adding a CG creature or simulated fluid effects like fire or water, a 3D representation of the person is needed as colliding geometry, light occluding object or to generate hold-out masks. Hold-out masks are basically areas in a 2D image that get subtracted from an element to make room for an object that is in front of the element.[329] For complex objects it makes sense to create these masks inside a 3D package and to render them as part of the 3D render process which have to be re-rendered if the relative position of those two objects changes due to animation or layout revision.

4.3.2 From Digital Make-Up to Digital Double

Techniques used to change the lighting in postproduction have been presented in 4.2.3 and digital focus has been covered in detail as part of chapter 4.1. This chapter is going to focus on digital methods "to create effects that change the look of an actor that in previous years was solely the province of makeup"[330]. These techniques become more and more common having different goals like adding wounds, creature effects that change the appearance of an actor, metamorphosis effects and retouch for beauty to name a few. These edits should be seen as regular VFX processes of different complexity according to the task. In general, methods can be divided into 2D and 3D techniques, the latter being more expensive and usually needing more data from the set or even a separate 3D scan of body parts of an actor.[331]

Some of the 2D approaches can take place during color grading. Van Hurkman describes some ways to create digital makeup with the toolset of secondary color correction. Hue and saturation controls together

[328] Duggal (2012), Van Hurkman (2014) 460-467, Knoll (2010) 694-696

[329] cf. Okun and Zwerman (2010) 862

[330] Knoll (2010) 694

[331] Knoll (2010) 695-696

with manual or procedural generated masks can be used to add colorfulness or to accentuate blush. In a limited range the surface reflectivity can also be edited to make it glossier or to reduce shine. Color or luma keys as well as Hue-Saturation-Luminance qualifiers are used to select areas that can then be edited by adjusting midtones versus highlights. Blur and soften tools are employed to smooth complexions.[332] Van Hurkman also mentions techniques to create the look of a vampire or zombie skin using extreme color correction based on manual masking or color keys.[333] Other 2D techniques can be found in the domain of compositing that are similar to other painting techniques used for wire or rig removal.[334] Cosmetic cleanup would be done by cloning nearby texture to hide an unwanted artifact. Also a procedural fill tool could be used to generate interpolated pixel data for the "offending areas". For more complex edits that might go beyond basic cosmetic corrections, an area of the image could be painted as a still image and then be motion tracked and warped according to the movement of the area of interest.[335]

3D techniques usually need a virtual camera or at least a virtual representation of the actor that can be created through manual modeling or some 3D scanning technique like structured light, for example.[336] Either way this step needs the working time of several specialists. The model needs to be animated according to the actor's motion, too. A track of the subject's motion can be a good starting point. To accomplish this goal a secondary camera, a so-called witness camera, is often used on-set to provide a second view point that can be used for triangulation. It might also be necessary to add tracking markers on an actor's face. This animation can then be refined using a manual keyframe animation in a process often referred to as rotomation.[337] This is also a time consuming and therefore costly process. Once an animated 3D model is available, additional CG elements can be modeled and attached to the virtual representation. Naturally, in this case lighting and shading need to match to the surrounding image area.[338] If the goal of the process is mostly a change in shape like removing or transforming a body part, a re-projection of the 2D image onto the animated 3D model is often done. This method yields more convincing results than a pure 2D solution due to correct rendering of parallax and perspective.[339] The animated 3D model of a human being is also called digital double and can take the place of a real actor for the moment of a dangerous stunt or complex interaction with CG elements. Although it might not be as detailed as a digital double, a 3D representation

[332] Van Hurkman (2014) 460-464
[333] Van Hurkman (2014) 466-468
[334] cf. glossary "wire removal"
[335] Knoll (2010) 696
[336] Knoll (2010) 695-696
[337] cf. glossary "rotomation"
[338] Knoll (2010) 696
[339] Knoll (2010) 697

that is needed for the simulation of fluid effects or rigid body dynamics is mostly created in a similar way. Depending on the use-case a low-polygon approximation might be sufficient for a colliding object or only the silhouette of an actor is needed to create a holdout mask for a rendering of a volumetric fog or smoke element.[340] As computation power increases deep data is increasingly used in this context. The idea behind this concept goes back to the paper on "Deep Shadow Maps" by Lokovic and Veach from 2000 who presented a way to store "a representation of the fractional visibility through a pixel at all possible depths".[341] Applied to the process of general rendering and compositing a different way of working with visual elements evolved. Rather than rendering and re-rendering hold-out mattes for other objects and then layering a series of flat 2D images deep images provide a data structure in the rendered images that allow to create hold-out masks in compositing, i.e. after the 3D rendering step.[342] For each pixel position there is not only one color and transparency value available at a specific depth but an array of values at all depths that contribute to that pixel.[343] Consequently, image elements can be merged or cropped based on their distance to the camera in a realistic way without the need to render hold-out information. In the case of the effect of a volumetric fog, a live-action element like a person can be composited inside the 3D fog while keeping the information in front of and behind the person with respect to the person's exact position in space. Of course, this presumes that at least depth information about every pixel of the live-action footage is also available at the time of compositing. Most of the time, this data is generated by applying techniques similar to those involved in stereo conversion work. As described in chapter 4.1.1 manual masking and semi-automatic image segmentation techniques are used to isolate the scene elements at different depths, which are projected on cards in the 3D space of a compositing package, for example. The depth data outputted by the 3D renderer can then be converted in a deep representation using the DeepFromImage tool inside the Foundry's Nuke, for instance.[344] If the camera is moving and the scene shows enough parallax, depth data could also be extracted using algorithms borrowed from stereo vision and photogrammetry.[345] A drawback using deep data still is the high amount of storage capacity and bandwidth that deep files take as well as longer 3D rendering times. These limitations probably get less important as computer technology keeps developing.

[340] cf. Seymour, "The Art of Deep Compositing" (2014)
[341] Lokovic and Veach (2000) 1
[342] Seymour, "The Art of Deep Compositing" (2014)
[343] Heckenberg et al. (2010) 1
[344] Heinen (2013) 54-55
[345] photogrammetry means the general practice of determining geometry from photographs and is as old as photography itself. The process is called stereophotogrammetry and consists of three steps: camera calibration (creating a representation of the camera), stereo correspondence (defining the relationship between views) and stereo reconstruction (calculating coordinates of image points), cf. Kurachi (2007) 126-138

4.3.3 New Possibilities with Light Field Data

Digital focus can be used to achieve the desired look on any take. Also a very shallow depth of field would become possible without taking the risk of losing focus on one of the eyes. As described in 4.2.6 depth and normal data can be extracted from a light field recording that can be used for various relighting approaches. Using simulated light sources in a linear color editing environment delivers accurate results. By applying synthetic shaders to the surface, that is to say the actor's skin, effects similar to adding make-up could be achieved. For example, a metallic or glossy appearance could be easily added based on surface normals and a light source whose position is matched to the plate. When creating CG extensions, masks or even complete digital doubles, the 3D position data extracted from the light field can be used as a first starting point or as a guideline for animation. The advantage when recording a light field is that you have the 3D data for every shot. It is no longer an issue if a witness camera recording or survey data has not been collected on-location. Depending on the shot tracking markers could be reduced as well. A point cloud or simplified mesh[346] approximated from the point cloud could play the role of a collider object in a simulation saving some time modeling matching 3D geometry. Also the generation of holdout masks from live-action footage might get more efficient by using the position or depth data from the light field. This could result in easier and better integration of 3D effects. A deep data representation of a light field seems to be a reasonable option in this context as well. The multiple viewpoints of a light field even give multiple samples near the edges of an object. That could help to create better transitions between foreground and background in areas of fine detail and semi-transparency.[347] More on the deep data representation can be found in chapter 7.2.1.

4.4 Conclusion

In these chapters the three scenarios Virtual backlot, packshot and portrait shot have been introduced in detail. Potential challenges in the production process and current work practices and workflows have been presented. Also it has been discussed how light field data could help overcome some of the challenges or change the overall production process. Depending on the technological state and the goal of the production light field data might be able to improve efficiency and the quality of end results in some common scenarios of today's movie and TV productions. In the following chapters the requirements on workflow and pipeline

[346] cf. glossary "mesh"

[347] cf. Heinen (2013) 13-16

implementation for these possible applications will be analyzed starting with a look at the terms *workflow* and *pipeline* and how light field was integrated for the test shoot.

5 Workflow and Pipeline

Moving on, this chapter briefly introduces the term workflow and pipeline according to the concepts described under methodology earlier. The macro workflow and pipeline considerations of a typical film and TV production are examined in the form of a short overview. Finally on the macro level it is predicted which parts of the workflow get affected by the introduction of light field data. In chapter 8.3 examples for micro workflows in the context of tools for light field integration are investigated.

5.1 The Workflow

Bugaj defines a production workflow as "procedure or series of operations through which a task is performed and a deliverable produced".[348] Therefore, a workflow is characterized by its deliverable and the task, which is a goal-oriented description of the work. A workflow can be described at different levels of detail. The macro workflow consists of a number of micro workflows, each defined by an input deliverable, several work steps and the output deliverable.[349] On a higher level the production workflow in the case of a movie production can be defined by the deliverable of the movie as it is sent to a theatre and the task of producing it in an as efficient way as possible. The input deliverable might be an idea in this case. A micro workflow might be the editing step, defined by the input deliverable of the source footage and the steps and procedures that lead to the output deliverable, the edit decision list (EDL), for example. The steps leading to the output deliverable itself can also be characterized as workflows, for example the process of ingesting media, syncing audio, rough cut etc.. To understand the overall process and the interdependencies a limited level of detail should be sufficient. Still on a high level of abstraction, the rest of the chapter will focus on describing a simplified and representative model of a workflow for feature film or TV production at the time of writing that uses standard digital media as acquisition format.

A film production can be divided into three key phases, the preproduction, production and postproduction phase, which can be briefly defined as the initial preparation, the principal shooting of the movie and the

[348] Bugaj (2010) 784
[349] Bugaj (2010) 784-785

editing phase of the shot material until the release. Figure 26 shows a flow graph representation of the main production steps as it is commonly used to communicate the flow of information. In the beginning of the preproduction there is an idea that a producer or production company decides to turn into a movie. In a process that lasts from several weeks to over a year depending on the project the final script gets written, a storyboard created, crew casted, a production schedule and financing scheme defined, location scouted and set building started. This phase includes all the preparation work prior to the start of shooting on-location or set. Sometimes it already includes small production phases on its own in the context of pitchvis or previs.[350] The postproduction and VFX studios join the preproduction in the bidding process. As soon as a studio is contracted on a show it might be involved with producing first concepts and tests, previs and start with asset generation, i.e. often modeling of 3D objects and characters.[351] Already at the time of bidding in most cases also at least a small research and development team is working on developing new tools and techniques to achieve a certain effect in a more efficient way or an effect that did not exist before.[352] This process is very similar to a software development cycle. At the end of the preproduction phase all licenses and rights should be negotiated, too.

Next the production phase begins which consists of the actual shooting of the film – the principal photography. It lasts for anything between 30 days and a year. The exact shooting workflow can vary greatly from one production to another. The VFX crew collects survey data and reference images. If needed also lidar scans are done of the sets usually after the last good take of the action. The recorded footage of each shooting day is sent to the post or production house, sometimes also referred to as the digital lab. There, the raw material is checked and prepared for the dailies[353] that allow the director and some of the crew on set to evaluate the work. This might involve a first grading pass or effect layouts. Often editing of the show already starts during production and first drafts are shown during dailies, too. Editing is a powerful preview and quality control tool in this context.[354] The VFX studios continue with asset creation and research and development during production.[355]

---

[350] cf. glossary "previsualization"
[351] cf. Zwerman and Finance (2010) 97-107, Duggal (2012)
[352] Whitehurst
[353] in the UK and the rest of Europe often also called rushes, referring to the speed at which the preparation work is done
[354] cf. Beier (2013)
[355] Whitehurst

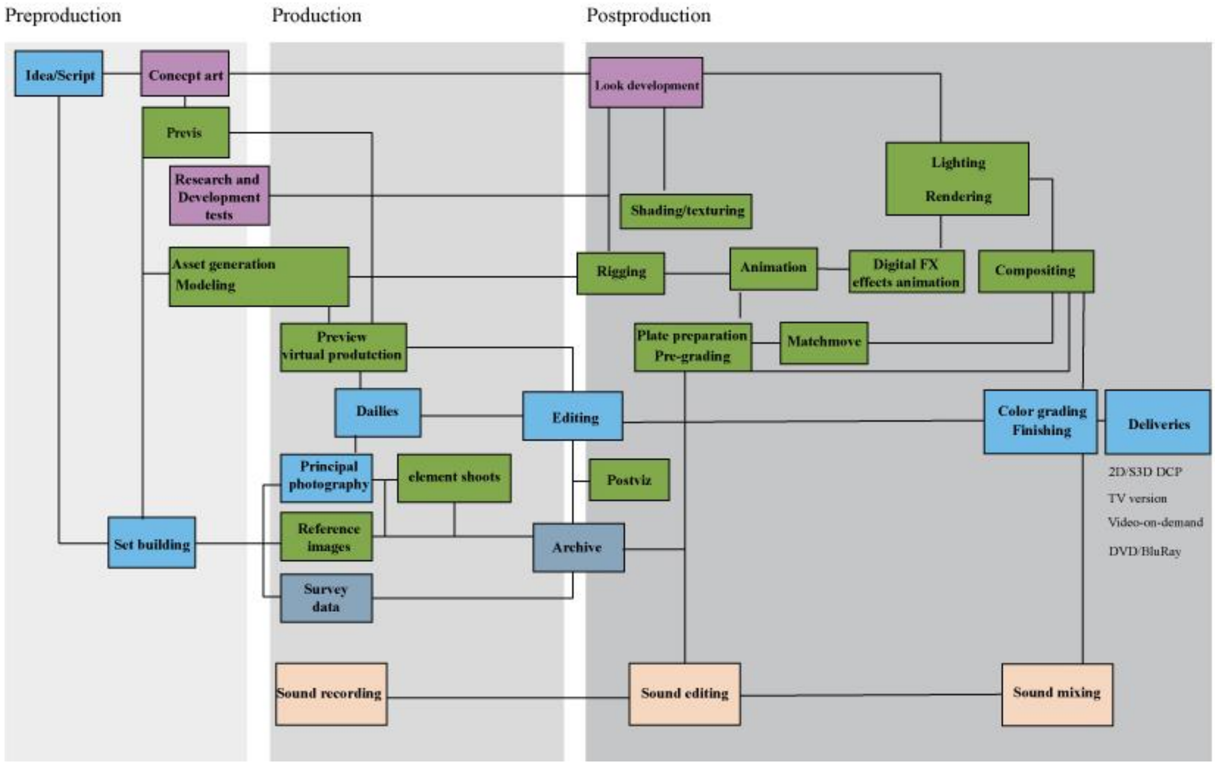


Figure 26: A simplified high-level flow-graph representation of a modern film production workflow (to be read from left to right), showing interdependencies and concurrences.

Originally postproduction has been defined as everything that happens after the creation of the final image, which would cover only editing and finishing. We will use a definition of postproduction that is more commonly used today and separates between photography on-set as production and all other image creating processes as postproduction.[356] Traditionally the first step of the postproduction phase is the editing. If not already done at the time of production the assistant editor begins with ingesting the footage and syncing audio using the camera and sound report from set as a guideline. Then, the editor creates a first rough cut. Based on the input of the director the edit ideally evolves to a locked version of the final edit, which is sent to add sound and music. At the same time also the visual effects editor takes over the edit to lay out the VFX sequences in the movie. As soon as a sequence is ready the shot data gets turned over to the VFX facility. Sometimes the visual effects edit also involves postvis[357] that can be done by just adding some layers in the editing software or by going through a low-quality VFX pipeline for complex shots. The time the editing process takes is highly dependent on the length of the film and the amount of footage going into the process but a typical time for editing a feature film would be about 3 months.

[356] Green et al. 35

[357] cf. glossary "previsualization"

Receiving the data from the editing department the VFX process during postproduction also starts with organizing data. Usually the filmed footage is converted into a special format suitable for VFX editing and saved on a centralized storage together with all the other data from set like survey photography, HDR images that hold lighting information from the set, as well as full 3D scans in the case of lidar data, for example. Some of this data needs manual cleanup and preparation work to be useful for the process. The footage often goes through a technical grading to match the look of the footage throughout a sequence before starting to work on individual shots.[358] Now the shot production begins with the matchmoving process that delivers a virtual representation of the real camera position and scene layout.[359] This process also relies on the measurements and reference images from the production stage. The virtual camera and scene data is given to the animation department and later also used for rendering and compositing. The modeling phase should be completed at that time and for objects and characters that need animation the rigging department created a system to control the movement, the rig. As it might be a very complex technical task, often the rig gets refined several times during the animation phase. Animation might also incorporate data from a performance capture session that took place at the production stage or is held during postproduction. Usually the animation is reviewed as grey-shaded playblasts.[360] The approved animation cache is then handed to the animation effects or CG effects department or directly to lighting and rendering if there are no further effects needed for a shot. Effects animation refers to animation that uses simulation and can be divided into the categories particles, rigid body dynamics and fluid simulations.[361] Effects like water, fire and destruction are created in this process. As soon as the 3D model of an object or character is considered ready it also enters the look development stage, at which textures and shaders are added to the model's surfaces.[362] The shading might take place in a dedicated department or as part of the lighting and rendering process where all animated CG objects are added to the scene created during matchmoving and lights are added. Then the scene gets calculated in the rendering process. Depending on the shot this may take from hours to days and is done on a multitude of processing units at a time. Finally, at the compositing stage all CG elements get combined and are integrated into the live-action footage. Also footage from a library or special element shoots can be used for effects like smoke or dust, for example. Shots that do not need computer graphics are often only worked on in compositing. Set extensions and digital backgrounds are typically created as digital

[358] Whitehurst

[359] cf. glossary "matchmoving"

[360] cf. glossary "playblast"

[361] Whitehurst

[362] see chapter 3.2.3.4 for more about the look develpment process

matte paintings[363] and added in compositing, too. Oftentimes the compositing also needs preparation work like keying, rotoscoping or wire removal, which might be done by a specialized department. The VFX production process takes anything from four to 12 months, sometimes even longer depending on the number and complexity of shots. It is also becoming more common to split the amount of work between studios to reduce the risk and speed up the process.[364]

All data, audio and video, comes together in the creative grading process again during the finishing process at the end of the postproduction phase. This stage results in the production of a digital intermediate and the process itself is now generally referred to as DI. The DI involves creatively changing color and contrast in filmed footage to create the finished look that the Director and DP designed.[365] The grading process can have slightly different results depending on the target media and usually takes longer for a movie than for a TV show. For a feature length film it takes five to 21 days. At the end of the DI the deliveries are created. These can be a DCP for digital distribution, an analog film print or a file format suitable for web distribution or BluRay- and DVD-authoring.

The individual workflow is shaped by the delivery or distribution format as well as the acquisition system. Some key considerations in the context of the acquisition are whether to shoot raw or processed data and whether the workflow is going to be an online or offline one, meaning using a copy of the original footage as much as possible or generating proxy data for most of the steps and going back to the original footage only at the end of the workflow. Raw data usually holds more image information but requires an additional processing step for the digital development that is computational expensive and therefore needs time. Although software now often supports to work with raw data, it usually is transcoded in some intermediate format at least for editing.[366]

As it is shown in figure 26, many processes are taking place simultaneously that are dependent on each other's data. One of the key challenges when designing a workflow implementation is to enable all of the departments to have the correct information and image data at any time of the production to make sure that no work needs to be redone. Chapter 5.2 will describe this in more detail in the context of the production pipeline.

---

[363] cf. glossary "matte painting"

[364] Ricklefs (2015)

[365] Whitehurst (2010) 659

[366] cf.. Van Hurkman (2014) 16-17

5.2 The Production Pipeline

According to Bugaj a pipeline is "a workflow specification plus a set of tools that are to be used to achieve the goal defined therein".[367] Therefore, "The goal of a production pipeline is to produce a product in an efficient and cost-effective way."[368] It has to "direct the flow of data from process to process"[369] and the "flow of work from task to task".[370] These two key tasks are linked and depend on each other. Often the pipeline is compared to the image of the assembly line, that is to say a chain of procedures where the second depends on the result of the first. This is only half true since the underlying creative process of a production pipeline is defined by circular feedback loops that need to be represented in the structure of the pipeline.[371] There is not much literature about the design process of a production pipeline that is a interdisciplinary effort that involves project management, IT and software development, because this knowledge can be the key to success or failure in today's production environment.

In the creation process of a pipeline the first thing to do is to define standards. This happens on several levels. It is defined which software is used for which task, including software that needs to be developed. File formats are defined and it has to be decided which attributes are stored in metadata. Metadata can be seen as "*data about the data*". "It is used to describe, control, structure or understand the content […]"[372] of a file and helps to produce confidence in a process this way. In the context of the file format aspects like compression and codecs, bit depth, color space and gamma are relevant and connected to the selection of the acquisition system and end format, of course. Then, workflow standards have to describe how assets are created and shared between departments and individuals. This also incorporates the use of a central database and common ways to communicate. And finally organizational standards have to define how tasks and reviews are scheduled or how the file system is structured and files are named, for example.[373] Although there are some off-the-shelf tools that help building a pipeline like the production management software shotgun[374] or ftrack[375], most pipeline implementations are unique and rely heavily on custom tools and scripts. Still the goals are quite similar. As mentioned above, the main idea is to gain efficiency by reducing

[367] Bugaj (2010) 786
[368] Green et al. (2014) 96
[369] Green et al. (2014) 96
[370] Green et al. (2014) 97
[371] cf. (Green, Hoesterey, Ricklefs, Streatfield, Theodore, Abecassis, Cole, et al.) (2014) 3
[372] McGraw (2014) 206
[373] Green et al. (2014) 103
[374] shotgun is a database driven toolset that is accessed via a web interface by artists, supervisors and producers; http://www.shotgunsoftware.com/
[375] ftrack takes a similar approach as shotgun, emphasizes on production tracking, https://www.ftrack.com/

manual repetitive work and delivering information to the right place at the right time since the biggest challenge of a film production is to organize the amount of information and data. A second issue in this context is accurate preview of the final image as soon as possible in the pipeline to reduce the number of revision cycles until it looks correct on the final media. This also points to color management that gets defined in the color pipeline, which describes how the colors of the input image relate to the colors of the output image and display device for each workflow step respectively.[376]

5.3 Light Field Data in a Production Pipeline

In this section of the work the term workflow has been defined in the context of a media production and the scheme of a film production workflow has been outlined. From there, it has been explored what pipeline means in this context and what needs to be taken into account when creating one. Next, in the form of a short overview potential changes in the production workflow are predicted that might happen when light field data is applied. The following chapter will discuss this in more detail as the process of the test production is presented.

At the current stage of development using light field data means a change in the acquisition system most likely and first of all. Therefore, the macro structure consisting of preproduction, production and postproduction will most likely stay the same. The preproduction will probably not change much either in the near future. The planning of the shoot might be a bit different, as it has to take into account a slightly longer setup time and the time saved due to less need for repetition of a take because of wrong focus or slightly wrong framing, for example. Probably the time needed on set will be similar to a stereo 3D shoot, which is not much more than a conventional 2D shoot. It also might make sense to do a tech vis[377] for some light field shots to get to know the new medium and be aware of the field of view of a certain array configuration, for example.

If light field as a data format can be successfully integrated into a production workflow experiences made in production and postproduction will probably feed back into the preproduction phase. Better preparation and planning for the shoot will be possible and new storytelling concepts might evolve.

Right before and in production the process of calibrating and characterizing the camera array has to be added to the workflow. While the working process between director and actors will stay the same, the other aspects of the production process will be affected when using a light field array instead of a 2D camera. Especially the cinematographer has to frame his shots taking into account all possible viewpoints in the light field and

---

[376] cf. Green et al. (2014) 28

[377] cf. glossary "previs"

light the scene differently, if he plans to relight parts of it, for example. The amount of data captured will be a lot more and all of that needs to be checked, archived and transferred to the post house properly. Once all the data is organized and parameterized the editing workflow will be similar to a conventional production and only incorporate some light field applications as some sort of preview most of the time. On smaller productions where there is no dedicated DI and grading, though, the editor will have to set the parameters for the final light field rendering, too. At this point it might make sense to introduce the possibility of an additional workflow step that has the goal of correcting and improving the visual quality and continuity between the light field cameras in terms of geometry and color. This might be similar to the sweetening process used in stereo 3D productions to improve the synchronicity between the two eyes in terms of timing, spatial relations and color.[378] Depending on the production scale it is probably a good idea to start with these further preparations that can incorporate rectification and semi-automatic color correction as well as disparity estimation of all captured views. The input deliverables of this process are the selected takes and frame ranges that are the result of the editing stage.

For the VFX part of the workflow there will be some changes since lots of VFX processes benefit from richness in scene information in the light field data. The compositing process will probably shift towards an approach similar to the procedure of compositing only CG like for animation features. Depth and other 3D data will be used and manual masking techniques might be reduced to a minimum in the long-term.

At the DI stage the trend to changing more parameters than just the color will be intensified with possibility to digitally set focus and relight the footage. Probably, extreme changes of the viewpoint will not be done after the editing stage in the near future.

It is important to keep in mind that this overview presumes the output format of a 2D or stereo 3D movie. If the output format is going to be some kind of light field format suitable for Virtual Reality (VR) systems, for example, depending on the amount of interactivity the workflow might change in the direction of a workflow that is similar to the one used in games and immersive entertainment. As mentioned before, this possibility is not going to be explored in detail as part of this work.

[378] Krause (2013) 45

6 Description of the Test Shoot

6.1 Preproduction

The preproduction of the shooting was essentially the same as in every common video/film production, with a traditional camera. Creating a production schedule, assembling a crew, screenwriting, storyboarding (visualizing the "look"), finding locations, casting, costume design, production design, prepare costumes, props, set dressing and VFX scheduling. Technical equipment, lights and all other necessities also had to be reserved for the filming.

Since a lot of visual effects shots had to be covered, preproduction also had to cover some planning of the coming postproduction work. Contrary to a big VFX production, we didn't have to keep an eye on budgeting and delegating shots to different postproduction companies, because everything was kept in-house and we were solely responsible for the visual effects and postproduction. More importantly, we had to think about designs and techniques used and select a VFX-Supervisor, making sure the shots were acquired in a way to get good results in postproduction.[379] The VFX-Supervisor is responsible for the visual effects shots and has to make sure visual effects shots are creatively and technically correct. In our case, the VFX-Supervisor was also the coordinator whilst shooting, making sure to take notes and logs for postproduction and handling the communication between different departments. Handling the reference items, such as chrome balls, acquiring HDRis of the sets, marker positioning etc. would also fall in the VFX-Supervisors range of duty. In like manner, the VFX-Supervisor had to fill the position of the Data-Collector, documenting the camera- and lens-information of all the cameras, taking measurements of set items and camera positions.[380]

The special aspects we had to keep in mind were that we had to show the functionality and the gain in production value of the light field camera, while writing the screenplay. Always keeping in mind, while drawing a storyboard, that we have the possibility to slightly change perspective, thanks to the camera array and that we don't have to determine the depth of filed just yet.

Maybe more than on traditional shoots, we were concerned about material properties because we knew about certain problems in depth-estimation and other algorithms concerning our light field camera array.

[379] cf. Okun and Zwerman (2012) 17/18

[380] cf. Okun and Zwerman (2012) 79

The algorithms of parallax calculation depend on diffuse surfaces. And for a successful depth estimation of every pixel in the image, a correct parallax calculation is necessary.

Specular reflections, highlights, (semi)transparent surfaces as well as areas out of the camera-sensors dynamic range (total black/white) can lead to problems in depth estimation. Reflections and highlights change position and shape in each viewpoint and reflect different amounts of light onto the sensor, making it difficult to find matching pixels for the algorithms. Transparent objects might lead to false depth estimation, because not the objects depth itself can be estimated, but only the depth of the objects lying behind the transparent object, the refraction, therefore giving false values in depth, because of distortions caused by the transparent object. Respectively, the amount of disparity is an indicator between the object and the camera, not being a problem when there is no transparent object between said elements, but if there is a transparent object in-between camera and object, the rays bend and change initial direction – comparing all viewpoints, giving non-linear ambiguous results in representation of disparity. Even though there are attempts to fix the false depth-estimation caused by transparent objects, i.e. the method of Maeno et al., which is taking the distortion of the background in account, we will not rely on the method because of the uncertainty of being able to embed Maeno et al.'s approach[381] in current algorithms provided to us. Also areas out of the dynamic range of a sensor can lead to false depth estimation, because the areas don't have any specific color values for a pixel which can be matched to another pixel with a similar value, instead whole areas are completely black or white, creating false values (holes) in depth maps.

So, in order to keep the quality of our depth maps as high as possible, we had to keep reflections, refractions, transparency, highlights and patterns in mind.

6.2 Production

Production is commonly defined by the actual filming of the live action on a set or location. Departments involved in filming were production, camera & lighting, art department, costume, hair and make-up. The positions were director, producer, production manager & -coordinator, location manager & -scout, script supervisor (continuity), director of photography, camera operator, first assistant camera, digital image technician, gaffer, best boy, lighting technician, key grip, dolly grip, production designer, set decorator, set dresser, props master, costume designer, make-up artist, production sound mixer, boom operator, post production supervisor, visual effects supervisor. Most of the departments and positions being the same as in any traditional shooting, we will only conduct the positions with differences to shooting with a regular

[381] cf. Maeno et al. (2013) 2788f

camera setup in detail. In our case the awareness of the term light field, the theory behind the term and the general understanding of the possibilities were given for all crew members. On a traditional set, the position of a light field supervisor might be necessary.

The production itself was on a tight schedule, a large amount of shots were shot in 3 days. In order to accomplish the filming of all shots, we had to regard the production schedule persistently. Due to the time limitations we had, we tried to use as little takes as necessary. Taking technical notes was an important part of the production, not only because of the lack of metadata on the light field array, but also because of information for tracking and match moving shots. Simultaneously we were taking records of all the shots and takes concerning special abnormalities like imperfect theatrical performance by the actor, shaky camera movement, differences in prop placement, sound recording problems, different variations of a shot, etc..

Those notes were necessary to select certain takes later on without reviewing all the recorded footage and to know what to watch out for in certain takes.

Most of the planned shots were shot according to plan, only a few optional shots were rejected due to time limitations. Prior to filming, there was not much time for camera calibration and characterization, neither testing with acquired data. Some minimalistic tests were conducted and the synchronicity between the cameras could be tested.

6.2.1 Positions with differing Tasks to Traditional Filming.

6.2.1.1 Director

The Directors on-set work was quite difficult to a traditional workflow, he usually has a certain image in mind and does everything to accomplish the look he desires. During our production he still had to follow the image he desires, in order to get the hero-cameras image the way he wants it to look, but at the same time keeping the flexibility and open mindedness of possible shift in camera-position and look of the light field multi camera array in mind. Also his creative and aesthetic ambition was limited to the chosen, fixed focal length (as described in 6.2.2.1) restricting him somehow in resourcefulness, as well as limiting distances while shooting

6.2.1.2 Director of Photography & Camera Operator

The camera operator, who was at the same time our director of photography, not only had to handle the considerable camera setup consisting of a stereo rig, a hero camera, the 9 industrial cameras for light field capturing and all the monitoring devices and cabling, but also had to check the framing of the hero camera and the multi camera array continuously to find a good average in terms of field of view. Our digital image technician also gave the possibility to preview both cameras on a camera-mounted monitoring device. The hero camera footage hat to be mirrored in order to be viewed correctly, due to the way it was mounted and due to the fact that it was filming off the reflective mirror. The multi camera array's preview had to be screen-captured from its recording device and then be injected back to the monitoring device on the camera-rig, giving it a bit of temporal offset, making camera panning difficult, so mostly the preview of the hero camera was used in order for smooth non-offset preview in camera movement and getting the right timing.

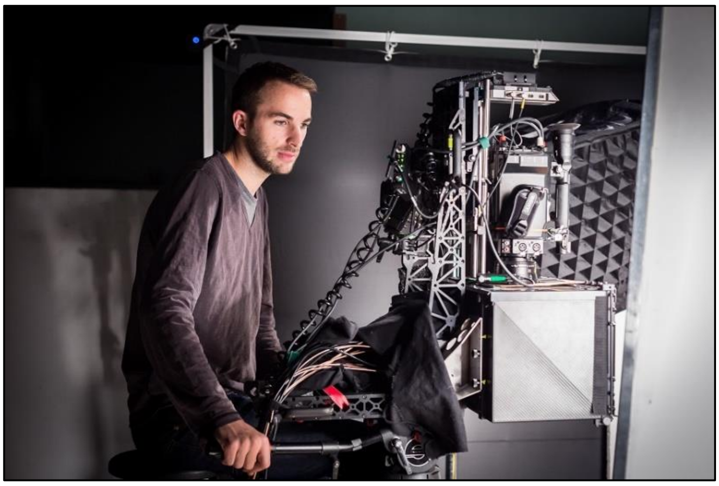


Figure 27: The Camera Operator handling the camera.

6.2.1.3 Assistant Camera / Focus Puller

The camera assistant was simultaneously responsible for focus pulling, even though we were working with a stereo-rig, the focus only had to be pulled on the hero-camera as reference. The multi camera array didn't need focus pulling, because the focus/defocus can be simulated in postproduction. Still the multi camera array needed to be setup the right way depending on scenes, which meant setting the aperture as low as possible, allowing enough light to fall on the sensor and keeping the gain and associated noise down, while still keeping everything, foreground and background, in focus at the same time. For scenes only covering a short distance in depth, it meant to be able to use a smaller aperture-number and for scenes with foreground and background relatively far apart, it meant a large aperture-setting. The focus had to be set accordingly, the sharpest point in the picture always being the actor, or the depth-plane on which the actor is moving. The aperture and focus had to be controlled manually for all of the 9 cameras, making it time-consuming, so we chose the order in which we filmed according to different scene-depths to keep the changes of focus/aperture to a minimum.

6.2.1.4 Digital Image Technician

During our shoot 2 digital image technicians were necessary in order to operate all the cameras and connected hardware. The first DIT was primarily responsible to check the camera outputs and provide the needed images for the camera operator and AC to the Astro monitor mounted on the mirror rig, which was connected to the Sony PWM-F3 (Input A) and the video village/recording station via a stereobrain image processor (Input B). The stereobrain was used to flip and flop the Sony PWM-F3 image, this was necessary because the F3 was the camera seeing the image reflected by the mirror, and to check alignment between the F3 and the center camera of the multi camera array with help of a difference image. The main control monitor on set was a 24” TVLogic screen. A preview of any of the Basler cameras mounted within the light field camera array was possible by mirroring the recording-station’s screen via a DVI to HDMI adapter, followed by an HDMI to SDI adapter and inputting the signal to the stereobrain and from there on having the possibility to send a preview image to the cameras control monitor as well as the TVLogic monitor. The preview of the Basler cameras suffered from a noticeable latency due to the preview latency of the capturing software.

A replay option of the light field array cameras recorded footage was possible by converting the packed raw files with the help of a batch conversion script, which was outputting 8 Bit Tiffs and a Quicktime file in sRGB colorspace for quality checks.

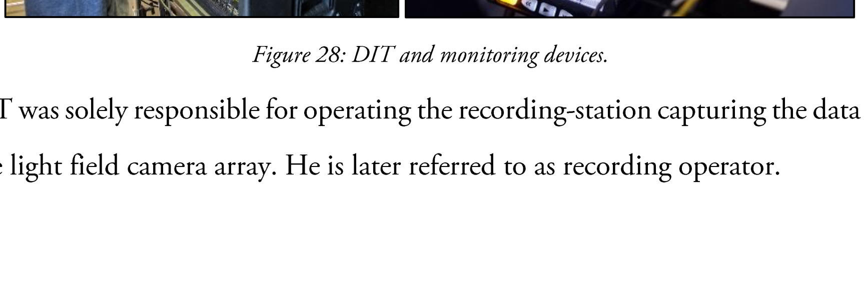


Figure 28: DIT and monitoring devices.

The second DIT was solely responsible for operating the recording-station capturing the data of all 9 cameras mounted in the light field camera array. He is later referred to as recording operator.

Figure 29: Recording operator.

6.2.2 Camera Array

As stated in chapter 3.2.1, the proposed Requirements for a light field camera array were:

- same size as standard production camera or an up to date stereo 3D camera rig
- max. 12kg and dimension of 50 x 20 x 35 cm (single camera, fully rigged)
- easy, intuitive UI for DP and AC
- fast alignment and calibration
- freedom in configuration (array layout)
- previous calibration and characterization
- 4K Resolution
- HFR (more than 25/30fps)
- HDR image output
- Large depth of field
- Recording metadata (i.e. timecode, lens type, camera information/settings
- On set live-preview, selectable hero-view
- Light field preview, preview of camera simulation and look-development

Those Requirements could only be partly matched for our light field multi camera array, due to the fact that our camera requirements were partly predetermined, due to the prototype provided by the Fraunhofer IIS. Consisting of a 3x3 camera array fitted with Basler ace series acA2000-50gc industrial cameras, capable only of full HD resolution and not the desired 4K resolution and the desired 25fps (up to 50fps possible) with a GigE interface. One of the main factors why the camera was chosen, was the 2/3" (11.26 mm x 5.98 mm) CMOS sensor and as a consequence thereof, the global shutter, avoiding motion blur as much as possible, easing up synchronisation between cameras and avoiding rolling shutter, which would strongly influence the look, if we desire to do an animated view rendering of a single frame. The 2/3" sensor also was providing the large amount of depth of field needed for the light field, in order to refocus in postproduction. With a pixel bit depth of 12 bits, the sensor was supposed to have a good dynamic range, as well as the possibility to support PoE while simultaneously gathering data on the same cable and therefor keeping the cabling to a minimum.

Concerning the cameras layout, we decided to go with a fixed array layout and spacing, which was proven to be working nicely in previous tests. As part of that decision, the requirement of freedom in array configuration was not overlooked, the freedom of different array layout is possible by every means, we just decided to go with one layout which covered our requirements, and not changing the layout also meant to maintain our tight production schedule.

The weight and size of the camera array shouldn't exceed the dimensions of a fully rigged traditional digital film camera (as stated in chapter 2.3.1 camera and hardware), in order to be easily handled by any DP. A traditional rigged camera can easily reach 12Kg and camera-only-dimensions of ~ 50x20x35 cm , the camera array by Fraunhofer IIS undercut those numbers by a huge amount, weighing approximately 3 Kg and measuring only 10x25x25 cm

The cameras also featured the ability to be programmed via the provided Basler API, making it possible to set parameters for one camera and then adopt those parameters for every camera in the array. The API also gave us the possibility of previewing all the cameras views, therefore, partly fulfilling the demand of a hero-preview and at least a step in the right direction concerning intuitive handling and user-interfaces.

As desired, most of the requirements to the cameras hardware and handling could be met, but not all of them. We had to refrain from HDR and HFR capturing, latter of which only due to the fact of an excessive amount of data and the limitation on our recording-systems bandwidth, exhausted by the data flow of 9 cameras.

The biggest drawback however was the inability to capture metadata. Meaning we had no timecode and no additional information whatsoever in our captured data, being a problem especially for postproduction and its workflows.

6.2.2.1 Optics

As for the optics, it was decided to use a normal focal length, as it provides a good mean value for all applications. Our 2/3" sensor with dimensions of 11.26 mm x 5.98 mm has a diagonal of approximately 12.74 mm ($\sqrt{(11.26^2+5.98^2)}$). The sensors diagonal measurements are at the same time the definition of the normal focal length. As a result of the calculated normal focal length, it was decided to take a Kowa 12 mm lens, being the nearest available focal factor available for our cameras C-Mount and also having the advantage of barely having lens-distortions at all.

The 12mm lens also provided the correct focal length in coherence with the chosen array spacing, for the highest flexibility in different light field applications.

6.2.2.2 Output Format

The video output format was supposed to be a 12 Bit packed Bayer GR 12 format, but unfortunately, we later found out we were only capturing in 8 Bit because the capturing software only supported writing 8 Bit files.

6.2.2.3 Camera Spacing

The camera spacing was chosen according to use cases and to provide enough matching features for algorithms (overlap in cameras). A lot of overlap was needed to have parallax data for almost every pixel in the image.

6.2.2.4 Hero Camera

The primary goal on behalf of Fraunhofer IIS was to show the possibility of capturing cinema quality footage with the multi camera array. In order to be able to compare the camera array to a traditional state of the art digital film camera, the decision to simultaneously shoot with an additional hero camera was made. The Sony PWM-F3 Super 35mm CMOS-Sensor Camera was the model of choice, offering high image quality and a relatively compact housing. Due to the different sensor size and therefore different normal focal length, a lens giving roughly the same field of view as the 12mm lenses on the multi camera array, had to be chosen. The cameras calculated normal focal length was approximately 27mm ($\sqrt{}(23.6mm^2+13.3mm^2)$), ultimately it was decided to use a focal length of 32mm, being a bit longer than calculated and giving a narrower field of view. It was decided to use rather 32mm and a narrower field of view, but having corresponding image-data from the multi camera array for every pixel of the hero-camera, then having a wider field of view and no matching image information in the edge regions. The main reasoning for doing so was the idea of trying to match the image information of the high quality hero camera to the data captured by the array in postproduction.

6.2.2.5 Combination of Multi Camera Array and Hero Camera

To match the sight of the cameras at any given moment during recording, we decided to use a beam splitter, traditionally used for filming stereo. In stereo acquisition, one of the cameras is usually mounted on a (motorized) horizontal plate, allowing rotation (for convergence) and translation (interocular) and often also tilting and is looking through the beam-splitters 45° mounted mirror. The second camera is mounted vertical, mostly fixed, pointing downwards and shooting the reflection created by the mirror. The main reason for mirror-rigs in stereo-filming is the possibility of a really small interocular distance between the cameras, without the physical limitation you have with a side-by-side rig.[382]

Since we didn't have the intention of filming stereo, but actually the intention of having a viewpoint as close as physically possible to the one of the other camera, a beam splitter was the only possibility to achieve such a solution. The 9 cameras of our light field array were just small enough to fit into a modified mirror rig, provided by Christian Meyer (Munich), mounted on the bottom, enabling them to see through the mirror. The beamsplitter had to be modified, the adjustable mounting plate, which is normally located on the horizontal part of the beam splitter, was moved to the vertical position in order to have enough space to fit the 3x3 multi camera array inside the beam-splitter. As a result of this, the multi camera array had to be mounted in place, not being able to adjusted, while the hero camera was mounted on top of the beam splitter with an adjustable mounting plate, in order to move the cameras optical centre to the optical centre of one of the multi camera array cameras.

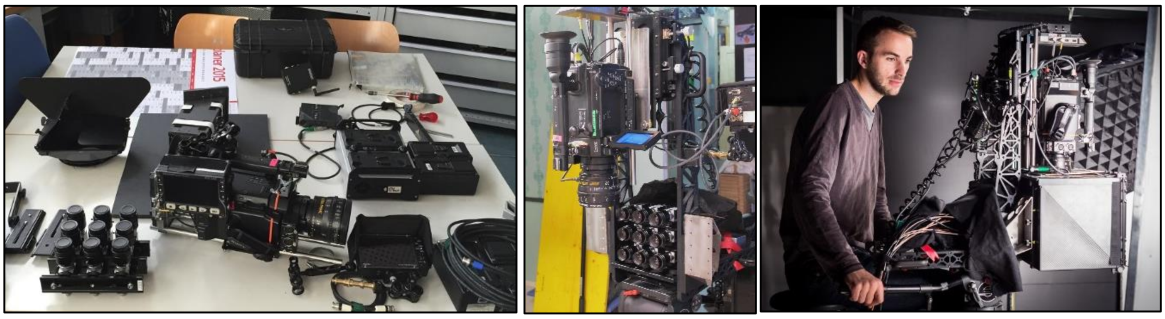


Figure 30: From left to right: Components for the rig setup, camera rig with mounted multi camera array and hero camera, camera operator with completely assembled camera rig.

[382] cf. Okun and Zwerman (2012) 430

6.2.2.6 Synchronisation

The synchronisation of the cameras was established by an Aja Gen 10 Sync Generator and Fraunhofer IIS' Ubuntu Linux PC (HHI proprietary recording software based on pylon API and openCV[383]). Synchronicity was checked, using a custom built light impulse generator from Fraunhofer IIS. To check the results a short recording with both cameras was made and was checked by comparing if the frames of the two cameras line-up at different points in the sequence.

The multi camera array was set up with a sync-delay of approximately 1,5ms. Synchronicity was checked, using a custom built light impulse generator. For that a short recording of a slate was made and then checked offline in a compositing package on a laptop to see whether the frames of the two cameras line-up at different points in time.

6.2.2.7 Calibration

The calibration of the F3 and the Basler cameras was established by previewing the center Basler camera and the Sony F3 on the same screen, with a difference overlay and manually adjusting the cameras with the help of a Siemens star, so rotation and translation of the images was aligned as good as possible.

Figure 31: Adjusting the cameras to one another.

6.2.2.8 Optics Calibration

The aperture and focus on the multi camera arrays lenses had to be chosen to keep as much of the scene as possible in focus, whilst at the same time providing enough light for the sensor to not produce too much

[383] *OpenCV* is an open source computer vision library originally developed by Intel. It is free for commercial and research use under a BSD license.

noise. All the aperture values had to be matched by comparing a brightness value of the preview screen of each camera and afterwards setting focus with the help of a Siemens star for each camera.
During the production it turned out to be difficult and time consuming to find the right aperture of all the cameras only according to a light value measured at the same spot in the cameras image and hard to manually focus the cameras while they were mounted on the camera array and only having a few millimetres to operate the focus and aperture.

6.2.2.9 GoPro Camera Array

Figure 32: GoPro Array during the backlot shooting.

In addition to our hero camera (F3) and the Basler multi camera array, a multi camera array consisting of 14 synchronized GoPro Hero 3+ cameras was used. The cameras were arranged in two rows, with 7 cameras mounted in each row. The camera array was mounted on top of a tripod and fitted with a preview screen. By providing batteries for all the cameras as well as the preview monitor, the system was completely self-sufficient and didn't need external power. The camera array was build this way to capture a wide viewing angle and allow a fairly large amount of horizontal movement in later view rendering applications. Since the cameras are not identical because the sensors are glued in place and can differ by a certain amount, a checkerboard was recorded by each camera from the exact same position to undistort each image individually during postproduction.
The camera was used for the outdoor backlot shooting, it was an experimental approach to see if the algorithms could handle the data, distortion, noise and deviating images (due the cameras implemented automatic settings). The first try to estimate disparity and generate good depth maps didn't turn out positive and due to lack of time for further research, we didn't use the acquired footage during our production.

6.2.3 On-Set Workflow

After hearing about the camera side of the workflow on set it is now described how the shoot was carried out in terms of the organization of operation. Most importantly it will be analyzed how preview and on-set video was involved.

At the start of the shoot, for each scene and whenever the camera rig had been moved from or to a dolly or tripod, the array had to be calibrated. As the interocular distances of the light field cameras would not change during the studio shoot, calibration meant to set and match focus and the iris of all nine cameras as well as aligning the middle camera of the array, in our case camera number four, with the position of the hero camera. Sometimes also synchronicity was checked using a custom built light impulse generator. To do that, a short recording of a slate was made and then checked offline in a compositing package on a laptop to see whether the frames of the two cameras line-up at different points in time.

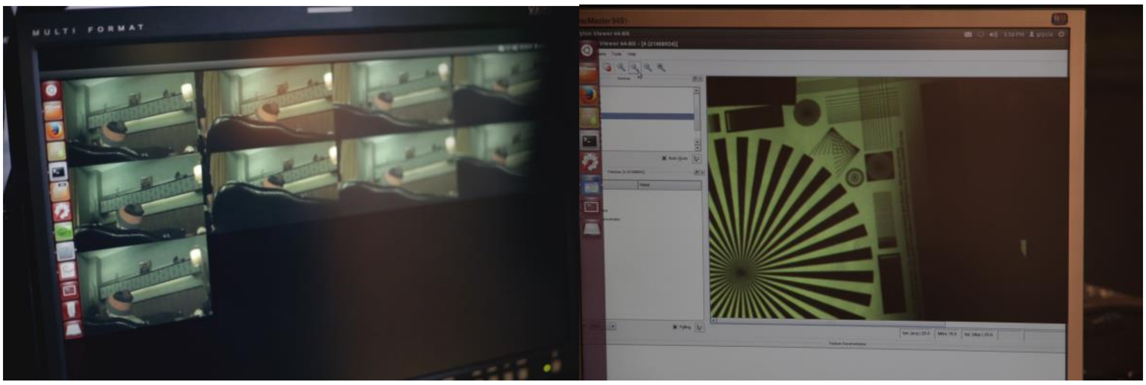
As we were constrained to fixed focus and focal length[384], the focal plane had to be adjusted to the scene layout. Shooting in the studio we actually only changed the focus distance once when going from wider to close shots. Generally there were two viewing modes available that could be outputted from the recording station and displayed on the tv logic monitor or the Astro monitors (cf. figure 34): A single view from either one of the nine Basler cameras or the hero camera or a multi-view image of all nine light field cameras. Figure 33 shows the viewing modes in operation during the test shoot. The viewing mode and the camera's number in the array had to set by the operator of the recording station or via a remote desktop from a computing device on set. In our case we had a separate operator at the recording PC who communicated with the set over radio. The image from the individual light field cameras was the result of a fast preview debayering with no color calibration applied. Additionally, the image of the Sony camera was also available at all times and the digital image technician (DIT) could switch between the two views on his monitors as well as route the desired signal to the camera operator's Astro monitor. First, focus was set for one of the light field cameras, typically the middle one. To do that focus charts containing some variations of Siemens stars were positioned on stands at the distances that needed to be more or less in focus in the following scene. By toggling between the single and multiview of the light field array it was assured that the charts could be seen from all viewpoints. While digitally zooming into the picture to be able to estimate sharpness better, the focus was iteratively adjusted between the middle, front and back of a scene while locking the position of the camera rig. Once the focus seemed to be the best match for a certain distance range, a mobile chart was used to check the distance of critical defocus to know how close or far some relevant content could be

[384] see also chapter 6.2.2.1

to the camera. This was a rather subjective decision of the DP that could only be checked against the hero camera set to equivalent focus distance. Altogether this experimental approach worked well and having also calculated the theoretical perfect focus distance in the beginning we found it to be the most effective and reliable approach since we had to check it visually anyway due to imperfections of the lenses and in the flange focal length of each camera. As soon as the focus was considered good for one camera, we transferred the setting to all other eight cameras by using the found distance value as a starting point and then checking again with the charts still standing.

Next, exposure was set. Since a high depth of field was needed to be able to estimate good depth data and set focus at a wide range of distances in post, the f-stop should be as high as possible and was only limited by the light we had available in the studio. Especially for wider shots the lighting was more difficult than we expected as we had to face the paradox of high contrast cinematic lighting to tell the mood of an early evening while avoiding shadow parts in the image. Having two camera systems with dramatically different dynamic range and sensitivity added to this issue. In some shots even a lower f-stop was used for the Sony PMW-F3 to shoot a reference of a different depth of field for comparison. We had to come up with a compromise between the right exposure for the Sony F3 and the light field array using the built-in neutral density (ND) filters of the Sony camera. Again, the decision for the f-stop was made for one of the cameras of the array first using the single view from the camera as preview that was routed to a waveform monitor to roughly evaluate exposure of the scene. The setting was then also applied to the other lenses according to the labeling. Using the utility software provided by Basler we were able to sample the pixel values in the image during live preview. Using the multi-view mode we selected an area of the chart that was seen by all cameras. Then the exact pixel values were matched between the views in single view mode by adjusting the aperture setting of the remaining eight cameras.

The third and final step of calibration was to align the camera positions of the Sony F3 hero camera and the middle camera of the light field array. This was done using the procedure known from Stereo 3D. The stereo image processor "Stereobrain" was used to output a difference image between the two cameras while mirroring one of the images (figure 28). Then the alignment was checked using charts at different depths and the mechanical controls for roll, tilt and level for the Sony Camera. This process required at least two people. As we did not use the exact same lenses and image formats a perfect match as it is needed for successful stereo vision was not achievable. Instead we aimed for a best match with an even amount of errors over different distances from the camera. The whole calibration process needed about 45 to 50 minutes for an initial calibration and about 20 to 30 minutes for readjustment.



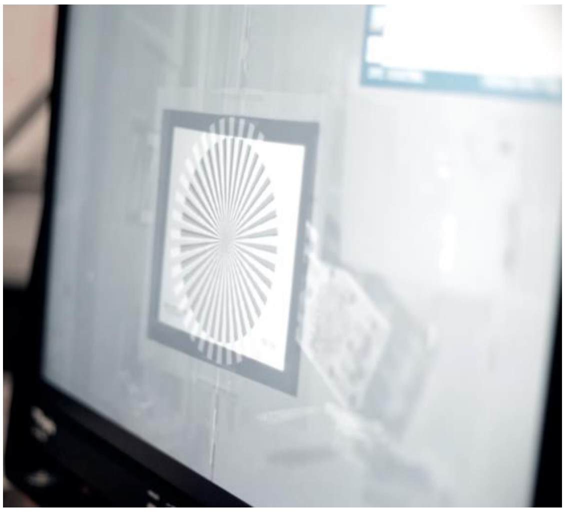


Figure 33: Multiview mode (top left, single view on recording station during focus calibration (top right), difference image for rig alignment (right)

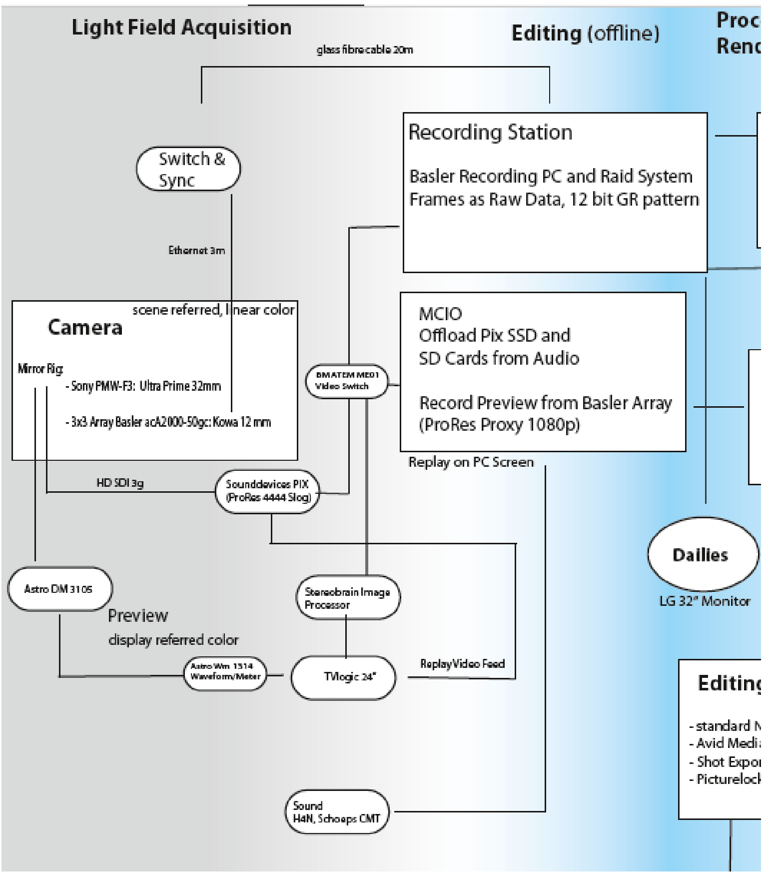


Figure 34: The on-set data flow of the test shoot

Setting up a shot was organized similar to a traditional shoot. After a first blocking the right camera angle was evaluated using the single middle view of the rig and then checked against the multi-view version. Then some adjustments to the camera position or set were made, e.g. some rigging removed or decoration moved in the outer regions of the region of view.[385] Also the lighting and exposure was fine-tuned by moving lights or adjusting brightness and checked with measurement tools like the waveform, histogram and vectorscope for both the hero camera and middle array view. A comfortable way to control the lights was achieved by using a dmx-based[386] dimmer system that could be controlled from a central control desk. While often having a stand-in for the actor at this stage, the exact framing was decided with the real actor in the scene, next. At this stage also the view of the hero camera was used for preview as it had a smaller field of view compared to the array cameras in our case. Because the light field recording needed about 10 times the disk space compared to the hero camera only, good rehearsal played an important role in the organization of the shoot. Once the director and DP were satisfied and the boom operator would know where to stand without

[385] region of view refers to the combined field of view of all cameras of the array

[386] DMX512 stands for digital multiplex with 512 individual pieces of information and is a standard that describes data transmission between controllers and lighting equipment and accessories; "DMX512 FAQ"

showing up in one of the views – again, the multi-view mode could be used to verify this in one rehearsal run – the first take was recorded. In any case a slate was filmed using an analog clapperboard as we did not have a common timecode between the cameras that could have helped the synchronization step in postproduction. In this context also taking measurements and notes for a detailed camera report was of great importance. This task was assigned to the DIT and the assistant director as well as the operator of the recording station. This included all the camera and rig data that is often saved as part of the metadata today as well as scene information like objects' dimensions and distances. For the shots that needed VFX additional measurements like camera position and angle were taken as usual.[387] Recording was started manually at the recording PC for all light field views at a time and on a sound devices stand-alone recorder for the Sony camera at the video village. The camera operator usually framed the image using the preview of the hero camera due to the longer latency of the array cameras' preview. But, one of the light field views was also captured by the DIT on a laptop attached to broadcasting hardware and could be previewed by the director in the video village. This lower-quality version was essentially the output from the recording PC's screen and could be used for fast replay and to solve continuity issues. This recording station basically consisted of an Intel Xeon based multicore PC system with 64 GB Ram running the recording software on an Ubuntu operating system. Two 10 gigabit fibre channel connections made connections to the camera array and a dedicated storage system. As mentioned before the recording PC only recorded an experimental, uncompressed and header-less 12 bit packed raw format that could not be previewed in real-time. The bandwidth and storage use by the nine cameras was between 700 and 800 MB/s. During breaks or at the end of a shooting day the middle camera view of the array was converted to 8 bit tiff image sequences and a low quality Quicktime movie for preview in a dailies session the next morning. The files were transferred via 10Gb Ethernet to a second NAS for backup at the end of each day, too. In total we ended up with recording 2,86 TB in the course of three shooting days. This data increased to 5,5 TB after the image processing and 3D data extraction process. Unfortunately, we were not able to simulate the light field applications described in chapter one and two for preview at any time during the shoot. We only tested the depth estimation process for a couple of frames from the light field array on the first production day on a laptop equipped with an early built of the proprietary software by Fraunhofer IIS. For the source footage we decided on a folder structure organized in shooting days, camera model and clip-number with a separate subfolder for each view. As we only had limited control over the naming of the source files the on-set reports

[387] cf. Goulekas (2010) 128-131

had to be used to link the files back to the shot names. The system time of all computer systems had been synchronized and could also be used to find correlating takes later.

6.3 Postproduction

The initial consideration for our postproduction was to edit light fields all the way through postproduction in order to maintain flexibility in output, especially the possibility to render different views after applying VFX and basic grading and the option to render out a 3D stereo version of our shots and maybe even shots for auto-stereoscopic or multiview/light field-displays. Since we had major limitations and also the factor of very limited time for editing, we had to decide against that initial approach. Editing light fields would have meant a great amount of work and data-transfer between the Stuttgart Media University and the Fraunhofer Institute Erlangen. The following steps would have been necessary in order to do light field editing:

- Image acquisition at Stuttgart Media University
- Image conversion by Fraunhofer Institute Erlangen
- Editing and Shot selection at Stuttgart Media University
- Rectification and depth estimation of the initial images at IIS
- Compositing and Visual Effects for all 9 rectified cameras
- Converting the selected takes and frame ranges to the proprietary file format supported by Fraunhofer's Avid Media Encoder Light Field Plugin in order to create a dense light field at IIS.
- View Rendering, Depth of Field editing in Avid Media Encoder (IIS Plugin) at Stuttgart Media University
- Final Grading in Stuttgart Media University

Not only the amount of work and the temporal limitation dissuaded us from the initial approach, but also the fact that the view-rendering results of the light field plugin were still generating a lot of artefacts (Figure 35), not meeting our quality demands and especially not the quality requirements of a cinema-like-quality final delivery.

Figure 35: Avid Viewrendering – Artefacts.

6.3.1 Postproduction Software

6.3.1.1 Editing

In postproduction, we were partly bound to Avid Media Composer, due to the availability of the Fraunhofer IIS light field plugin being only available for Avid. So in order to keep the amount of used postproduction software to a minimum, we decided to do the editing in Avid, a 2D non-linear editing (NLE) software.

The selected takes were chosen by reviewing the footage captured by our hero camera. The recorded data from our light field array was in a packed raw format, which needed to be converted into an image sequence capable of being handled by Avid Media Composer. The chosen proxy output format for editing was 8 bit TIFF (due to the fact that we recognized we only recorder in 8bit instead of 12 bits). Because we already chose what needed to be converted by reviewing the footage of our hero camera, the time for conversion could be kept to a minimum and could be limited to certain frame ranges. The sound was linked to the camera footage by slate-point.

The only difference to traditional editing was the fact that we had image sequences of nine cameras available, but since there wasn't a big difference in field of view, we decided to do the editing only with our center view and the matching shots from our hero-camera. It wasn't as easy to synchronize the shots in editing, since the cameras of the light field array didn't provide any metadata whatsoever, meaning there was no timecode available for editing.

Subsequently to selecting the shots we decided, what kind of data we would need from the IIS to establish the desired visual effects. For most of the basic VFX-shots, we just requested the center camera and corresponding depth data. For some Shots, all of the 9 cameras and normal- and position data.

6.3.1.2 Image Processing and Preparation

Before the shot production could start, another step of image processing needed to be done at the Fraunhofer IIS including the image rectification as well as depth, normal and position data generation. The editing decisions and a visual reference were sent to the Fraunhofer IIS. The framecounts and selected takes had been noted down in an Excel sheet since there was no way of transferring the frame ranges of both cameras securely from the NLE to an EDL.

The rectification and depth estimation used the original camera footage. A pre-grade before the processing step was planned but could not been tested due to time constraints. More time than we initially thought, was needed to cope with the raw camera format. It would have made sense to debayer the data in a scene linear or logarithmic image format at the beginning of the postproduction process. With the help of Daniele Siragusano we tried to create a basic characterization for the Basler cameras resulting in a matrix or look-up table (LUT) based on McBeth Charts we shot earlier. Unfortunately, these attempts gave unsatisfying results. Since we did not have enough time to test the cameras sufficiently we guessed that the poor quality of the color filters on the sensor are one reason for this. But it could also be related to problems with the recording format or undocumented in-camera data processing. Even though we came up with a LUT to transform the debayered camera data to Arri LogC in the end, which worked well color-wise, we were only able to use it on few shots. As it turned out the recording had been made with only 8bit of dynamic range that was not even equally distributed over the range of values due to a gamma that was applied in-camera. This limited the range of post processing possibilities severely and lead to artifacts in the midtones when using the LUT.

After debayering the data using the mathlab libraries the data set was rectified first.[388] Then, based on a second list of data requirements on a per shot basis, the 3D data extraction was done incorporating the Fraunhofer IIS algorithms.

The output containing the special file structure needed or the Avid plug-in, disparity or depth data, normal and position data as image sequences was sent back to the HDM. Most of the image data was stored as 8bit Tiff or 16bit Tiff files in the case of the data passes. For preview inside of the Avid NLE png-sequences were used. Over the course of the project we received several generations of 3D data of improving quality.

At HDM we also conducted some tests of depth generation of our own. We used the off-the-shelf nuke tools to generate depth from an image sequence and a moving camera. The moving camera was created using the Nuke camera tracker plug-in and reordering the array configuration as a sequence of images using retiming tools.[389] This helped to understand the concept but suffered from poor performance and could not keep up with the image quality of the results from Fraunhofer IIS, of course.

The edit was transferred to the Nuke Studio environment and conformed with data from Fraunhofer IIS..Again, the middle array camera was used to represent the light field. Also a denoise, sharpening and basic color correction setup was created to be part of a script template for all shots. The Sony F3 footage was also imported into the Nuke studio session and lens correction presents configured. But at the time of writing we only did a few tests regarding the possibility to merge the F3 footage with the Basler data.

As part of the compositing process itself we did some additional corrections and manipulation to the 3D data. This will be described in more detail in chapter 6.3.3.

6.3.1.3 Visual Effects & Compositing

Since there are two completely different workflows in digital compositing, it had to be decided which one to use and why. One can decide between a layer-based or a node based workflow. A *layer-based compositing* software relies on timeline editing, with each media object being a separate layer with its own effects, keyframes and time-boundaries. The timeline editing is an advantage, if you're used to working with NLE Software, because the basic workflow is essentially the same. The final composite is reached by layering on top of previously composited layers, in order to reach the final result. This workflow is well suited for limited 3D effects, such as in motion graphics and also for 2D effects, but for complex composites it quickly becomes

---

[388] see also chapter 8.3.6 for more on the rectification process
[389] This approach is described in more detail in 6.3.4.

confusing due to a large amount of layers. Also the Limitation in 3D Compositing is a factor, which should be thoroughly considered depending on the postproduction work necessary.

Node Based Compositing is handling complex composites with a different approach, by linking together several (simple) image operations. Each operation or effect is a so called *node*, by joining several nodes a composite is represented as a tree graph, often also called process tree, flow- or node graph, which appears similar to a flowchart and has the same diagrammatic representation. Said representation makes it easy to follow the approach to a solution, in our case the final composite, by following the node-path without getting lost in stacks of layers. This type of workflow allows great flexibility and the possibility to modify parameters of each node (in context) at any time during compositing. [390] [391] [392] Many of the node based compositing software packages are also including the possibility for limited 3D capabilities, giving the 2D artist the opportunity to work in a 3D environment and incorporating effects and tasks which used to be handled by the CG department.[393]

In postproduction for the test-shooting we decided to go with a node based compositing workflow, due to the complexity of our shots, the necessity for a 3D-Compositing-Environment and of course it being the state of the art workflow used for complex shots in a cinematic pipeline.

The Foundry's Nuke became the software of choice, because of previously mentioned node based workflow allowing iterative processes throughout the whole compositing and it being the current industry standard, offering extensive channel support, powerful image manipulation tools, a 3D compositing environment as well as a resolution-independent compositing system. For each of our shots a basic compositing setup was created with the help of Nuke Studio, including all the necessary read nodes for the RGB images, including correct setup for colorspace, as well as the 3D data. The needed 3D data was shuffled into the RGB stream of the associated camera and all the different cameras read nodes were joined into a single stream, by assigning dedicated view names. The setup also provided the needed frame ranges, as well as a gamma neutralization and denoising for the whole sequence. Since the cameras could not be exactly calibrated before shooting, a color pipeline was implemented in each nuke setup, to match the colors of the cameras to one another and reduce problems like flares and stray light, which occurred in some situations when light was falling inside the beam splitter from behind, due to our imperfect prototype setup. A basic grading was also

[390] cf.Byrne (2009) 13-15
[391] cf.Wright (2010) 10-11
[392] cf.Ganbar (2011) 1
[393] cf. Wright (2010) 3

done to provide a feeling of the final look and especially in order for the artist to work with a look that feels natural. In some cases, previously tracked 3D camera matchmoves where provided for the compositing artist. At the end of each initial basic compositing setup, reformatting to the correct output format took place, a timecode was added according to the shot and a customized write node was implemented to make sure the output format and path are correct. Many of our shots were edited in prototype-like compositing setups, trying out different approaches with the provided 3D data and multiple views.
The individual compositing process of different shots will be explained in detail in 6.3.2, 6.3.3 and 6.3.4, concerning the post production of three quintessential scenarios involving light field data.

6.3.1.4 Grading

Most of the grading was done during compositing for each shot. Nuke Studio turned out very helpful in order to compare the grading of each shot and match the grading with its implemented basic color correction functionalities. The final grading only consisted of minor corrections and was done in DaVinci Resolve as a last step before the finishing.

6.3.1.5 Finishing

The finishing consisted of two delivery formats for the final commercial spot as well as the breakdown. Delivery formats were, H264 Quicktimes in 1080p resolution, as well as a DCPs in 2K Flat. For the final commercial two versions were rendered in each delivery format, consisting of a version including voiceover sound and one version with sound effects and on-set audio only.

6.3.2 Postproduction of the Backlot Scenario

Prior to the compositing of each individual shot we decided on a conforming workflow with The Foundry's Nuke Studio in order to provide a homogenic pregrading, denoise and slight image enhancement for all the shots, which gave the artist working on the shot a rough idea of the final look and a nicer look to work with.

For our backlot scenario three shots were selected, two of them were from the same take, while another one was different in terms of camera movement and position. The Shots consisted of the sequences acquired during the studio shooting, a reflection plate, an HDRi also shot during the studio-shooting, the sequences shot during our outdoor backlot-shooting, and various elements which were photographed during the shootings. The Studio Set was built as a hotel room with a window outlook, the area behind the window completely covered with green screen and tracking markers in various depth, in order to later replace the green screen with our virtual backlot.

Fraunhofer IIS received our editing reports and the frame ranges we needed for each of the sequences, in order to rectify the images and compute the normal, depth and position passes which were necessary. The important part of matchmoving[394] was one of the first steps as soon as we received the rectified data and needed 3D passes from Fraunhofer IIS. The machmoving was done with The Foundry NukeX's internal 3D camera tracker (Figure 36). The tracking results were surprisingly good from the beginning on, due to the fact we had no out of focus areas thanks to the light field array. The matchmove data was needed in order to position and move other elements correctly according to camera movement in the 3D space.

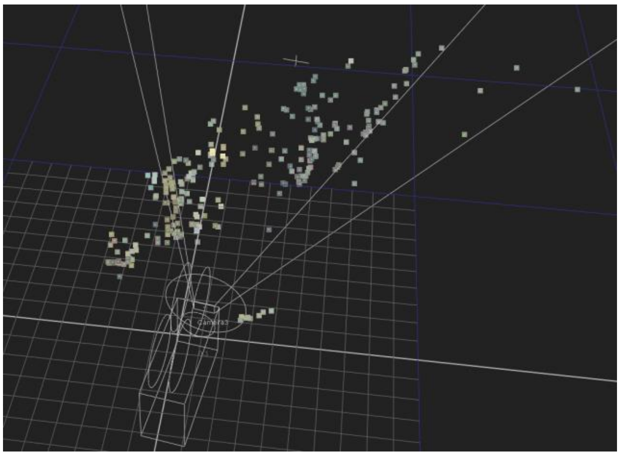


Figure 36: Matchmoved Camera and Point Cloud, generated by Nuke's Camera Tracker.

The next step was keying[395] the green screen inside the window outlook in order to replace it with the background we desired. The keying process itself required a combination of various keys and manually created garbage rotoshapes to get the desired alpha-matte. Since the green screen reflects unwanted bright green light onto the scene and the actor, the so called green spill, a spill reduction hat to be applied to remove unwanted green reflections.

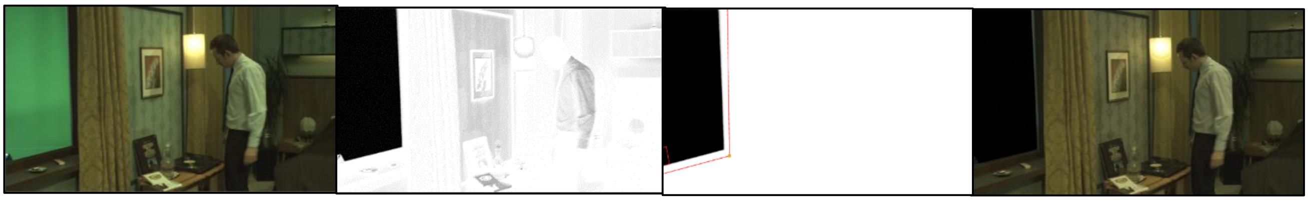


Figure 37: From left to right: Plate, keying, final matte, final key including despill.

[394] glossary
[395] glossary

Depending on the camera angle a background was chosen from 3 initial recording positions and was projected onto a card in 3D space placed behind the window with some distance in between. The intention of using a light field backlot filmed by a camera array of 2x7 GoPro cameras was discarded after receiving confirmation on the part of Fraunhofer IIS that at the moment it is not as easily possible to generate good disparity with the captured footage from the camera array. Luckily we decided to also shoot every angle with our previously used Sony PWM-F3 in case anything doesn't work out the way it should. The regular 2D footage was used as a background for our composites and mentioned previously projected onto a card in 3D space according to out 3D tracking data.

To add more realism, the backlot footage was graded according to the desired daytime we wanted to show. Since we removed the glass window during the shooting to eliminate the risk of reflection of staff members while shooting, every missing element of the window had to be added during postproduction. Elements like window framing and a knob had to be added, as well as reflections of the actor and the room, to mediate the feeling of glass. For the actors reflection we used a take we shot with the window in place, showing us the reflection. Excessive retiming, warping and rotoscoping was necessary, to match the reflection-plate to the movement in our scene, hide crew members and remove unwanted artefacts. To add reflections of the surroundings, in our case the room which was built in the studio, we used 360° HDRis we shot close to the cameras position, giving us the possibility to choose exactly which part of the room should be reflected in the window. Since the Reflection was supposed to be only slightly visible in the window, the lack of movement within the scene didn't matter and the static HDRi provided a satisfying result.

Figure 38: 360° HDRi of the Set.

Even with all the reflections, the window still didn't feel right, so we added another image layer consisting of dirt and smudges to the window plane, imparting a window glass at the right depth.

Due to arranging all the additional elements within the 3D space of the scene we had the correct depth data of all elements available for further steps. The combination of depth maps gave us the possibility to use a global ZDefocus in order to simulate the desired depth of field. Previously the initial depth maps of our studio shooting were normalized with the help of measurements taken from the set, after the normalization the values were in the format 1/distance in meters. For further refinement of our depth data we denoised the depth maps and filled holes by sampling neighbouring color values (see Figure 39).

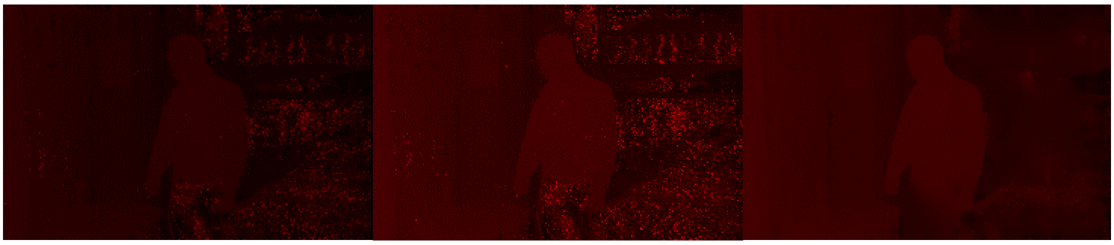


Figure 39: From left to right: depth map, normalized depth map (1/distance in m), refined depth map.

Figure 40: From left to right: Backlot depth map, Plate depth map, combined depth map.

Figure 41: Focal Plane Setup. From left to right: Focus on actor, focus on middleground, and focus on background.

Figure 42: DoF Simulation. From left to right: Focus on actor, focus on middleground, and focus on background.

With the initial plan, we also would have had the correct depth data of our backlot-plate, giving the possibility to animate the depth of field from the nearest to the furthest point in our compositing. Posterior to the production a rough disparity was calculated by using two of the 14 GoPro in order to show the feasibility and the possibility of generating depth maps and point clouds, as seen in Figure 43 & Figure 44.

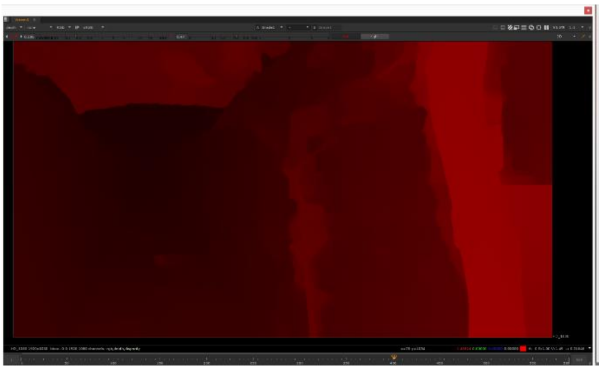


Figure 43: Backlot depth map generated by using 2 GoPro Cameras from our GoPro Array. Note: The depth estimation of the sky lead to wrong values.

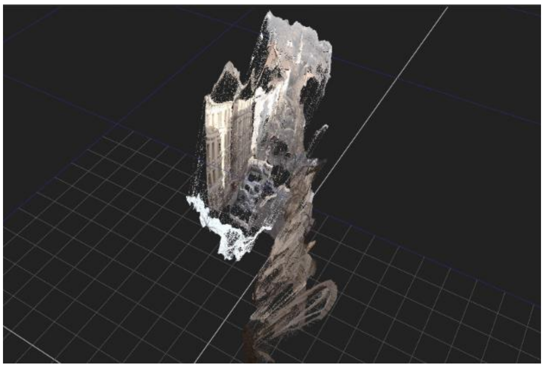


Figure 44: Pointcloud of the Backlot, generated with previously calculated depth map.

Instead, we used a very rough depth map, generated by using a grey image as depth layer and applying a gradient according to the depth of our scene. The nearest point in the depth map being the elements right in front of the window and the furthest being the end of the street seen in the shot.

Color grading and little retouches like replacement of a lamp in the scene were also applied at the end. As a last step, synthetic grain was added to give the final composite a more lively and consistent look.

6.3.3 Postproduction of the Packshot Scenario

The postproduction of the packshot scenario can be divided into two parts. First, we tested an approach for relighting CG and methods to integrate CG with a 2D view from the light field in a more theoretical way. In the second part we could transfer the techniques and experiences on the production of a second shot that only uses the data from the light field teapot.

As with the other shots we did a pre-grade and slight image enhancement in the form of a customized denoise and sharpening for the packshot, which gave a first idea of the final look. For our work with the packshot we had selected two takes, one with the real teapot and cup and one with the empty turntable only. As soon as we got the rectified plates from Fraunhofer IIS we started with a fast matchmove[396] and object track for the movement of the turntable. This got loaded into the 3D package Autodesk Maya to create a CG representation of the turntable. This could then be used to position and render a CG version of a teapot

[396] cf. glossary “matchmove“

including its shadow cast on the turntable. This was done in the conventional way of producing and rendering digital assets to today's standards using the high dynamic range imagery that was shot on set for lighting and realistic reflections on the teapot. Rendering the teapot with normal and position passes as additional AOVs enabled us to use the build-in tools inside of The Foundry's Nuke to relight the teapot at the compositing stage. The normals were in a camera oriented coordinate system and the position data as usually world oriented in this case. The node[397] called *relight* is based on the 2.5D relighting approach described in 4.2.3.3 and is frequently used in production for use-cases also described in 3.3.5 and 4.3. For the relight node to be able to do its calculations it needs four data streams: 1. An image data stream, that includes RGBA[398] color information as well as the normal and position data as two additional sets of RGB channels, 2. A light from Nuke's 3D toolset or a 3D scene containing several lights, 3. A camera node representing a virtual camera position and 4. A shader node providing a shading model. In Nuke's 3D system are three types of lights available, the point light, directional light and spot light that correspond to the basic light types in computer graphics. The relighting node itself takes the roll of a simplified rendering program in the relighting process. In the case of the teapot a variant of Phong's shading model[399] was used since the teapot is an object with specular reflections. The shader node allows for setting the constants of the underlying equation that is then solved by the relight node using the camera information as viewpoint vector, the light vector provided by the light node, the normal vector from the image input for each pixel and the position information to calculate the light falloff and scene scale.[400] The world position pass can also be used to plot a point cloud as a 3D representation of the scene object inside a 3D coordinate system (figure 45). Although this point cloud can be seen more like a relief than a 3D model it enables interactive position of lights in a 3D coordinate system and then feeding the position and direction into the relight node.

---

[397] being a node-based compositing package, Nuke consists of a variety of different small programs, calles nodes that can be a graphically assembled and linked in a flowgraph to formulate a script that is executed on a per-frame basis at render time, Wright (2010) 10

[398] channels of a color image formed by an additive color model: red, green and blue, A stands for the alpha or opacity channel that is needed to perform a compositing operation as described in Duff and Porter (1984) 254

[399] Kurachi (2007) 218

[400] cf. "Nuke Node Reference Guide" 690

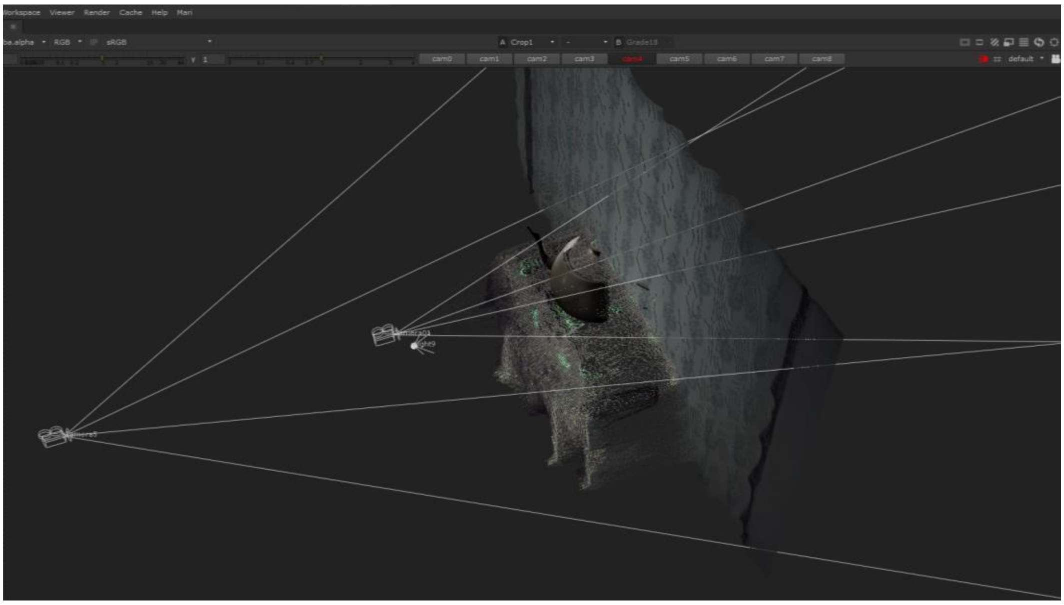


Figure 45: 3D coordinate system with point cloud of teapot and the live-action background

The resulting output of the relight node is an RGB image of the object with the shader uniformly applied to the whole object and the effect of the lights plugged into the node. It can be described as a light pass that shows the contribution of one light to the objects appearance. Usually this image is not the final effect that is needed in the relighting process. Usually the new light pass is merged with the original color image or beauty pass in the terminology of AOVs by added or removing, i.e. a plus or minus operation, the light pass. Since the relight node is unaware of phenomena like self-shadowing or light masking that are connected to the surface geometry the result will often look wrong in some areas depending on the texture and surface topology of the object. This issue is often addressed by blending some of the original texture and color information with the light pass before adding it to the original image, for example. The overlay and hard- or soft-light operations are usually used depending on the intensity of the desired effect. These blend modes originated in Adobes Photoshop and are now found in almost all compositing tools. The overlay mode, the simplest of the three, works as follows where B is the base layer, the original color image, and A is the blend layer or the light pass:[401]

[401] Wright (2010) 189

```
if (B < 0.5) then
    2*A*B,
else
    1-2*(1-A)*(1-B)
```

The logic of the overlay blending mode

If the pixel value in B is less than 0.5, a multiply operation is done. If it is more than 0.5, it switches to an operation similar to the screen blending mode.[402] Of course, this logic is applied to all pixels of all color channels respectively and presumes images in a linear colorspace. If pixel values are out of the zero-to-one range or only in one half of the spectrum some of the blending modes give undesirable results. It sometimes makes sense to scale the values before blending them and possibly reverse the scale operation afterwards if you are working in a floating point environment. If you have an ambient occlusion pass available that got rendered from the 3D package, for example, it can also be multiplied with the newly generated light pass to simulate the self-shadowing. As physically based rendering and path tracing is increasingly used in production it is not that common to render ambient occlusion passes anymore as the occlusion information is baked into the path traced image. In the case of the teapot we did not have an ambient occlusion image, for example.

Still, one problem concerning the output of the relight node remains. Since the data passes only contain discrete data at each pixel position the resulting light pass shows aliasing artifacts in edge regions, for example. The correct way of fixing this would be to render the data passes, i.e. the normal and position layers, in at least twice the resolution of the color pass according to the Nyquist–Shannon sampling theorem[403]. This might not always be an efficient option and that is why often a workaround is applied. The data passes are scaled up using an interpolation method that applies sharpening like the Mitchell algorithm before the relighting operation and resized afterwards using a soft scaling algorithm like the Gaussian or impulse method.[404] This reduces the artifacts to a sufficient minimum in most cases. Like this we were able to add or remove light from the object and to change the surface color and properties inside Nuke.

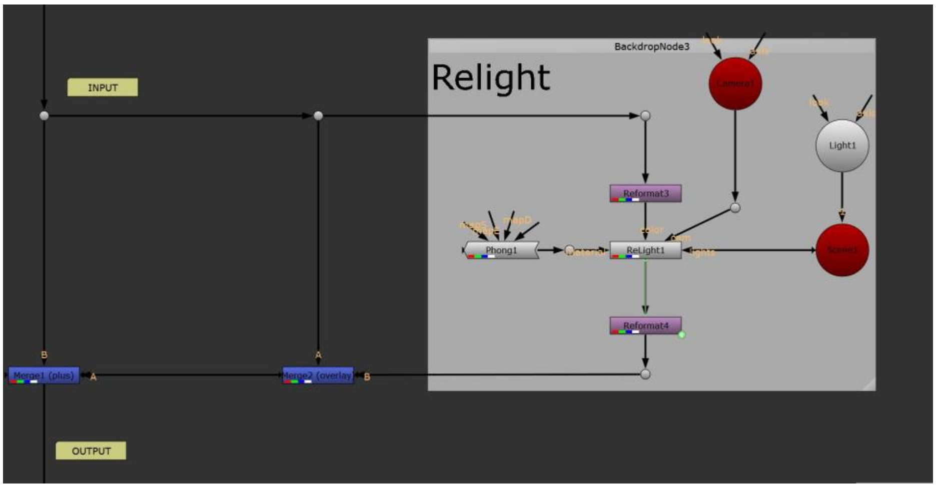
---

[402] Wright (2010) 189

[403] a mathematical description of the theorem can be found at http://ptolemy.eecs.berkeley.edu/eecs20/week13/nyquistShannon.html

[404] cf. Perez (2014)

Figure 46: A relighting setup as used regularly to relight CG at the compositing stage



Next, the idea was to also relight the clean plate that we used as the background for the CG teapot with the same lights that were used in the context of the teapot. We decided to transfer this existing 2.5D technique to the relighting of live-action footage in the case of the test shoot. Being production proven and part of a known and maintained software package this approach is relatively stable and an easy way of communicating the concept to media professionals.

Again, we got a depth map from Fraunhofer IIS but this time also position data and normal maps of the same resolution. We started off with analyzing the data passes and improving the overall quality. We will describe the term quality in the context of depth data in more detail in 6.4.3.3. Mainly two types of errors occurred: noise or spots of wrong information and holes, bigger areas of wrong information due to occlusion. In this case the depth data was filtered with different median filters first to reduce the number of small spots of wrong intensity. Being computational intensive we had to come up with a compromise between render time and quality. In the end we did not want a pre-render of the data passes to take more than a minute per Frame. Then, holes were filled in a variety of ways. Usually this was a two-step process of first generating a mask for the area that has no or wrong depth information and secondly resampling this area taking into account nearby pixels. Often the mask could be created in a semi-automatic process that consisted of a quickly drawn 2D shape as rough mask and a procedural generated mask for the detailed shape. The latter could sometimes be created from the depth data itself and be iteratively refined as the depth data got better. Sometimes known tools like a color or luma keyer were used, too. The generation of the new pixels incorporated plug-ins like the MultiSampler node from the Stereo Compositing suite Ocula[405] by The Foundry or a free gizmo collection like pixelfudger[406]. Or it was done using customized setups of stacked blurs or colored shapes. If some parts of the scene were static like the background in the case of the packshot,

---

[405] http://www.thefoundry.co.uk/products/ocula/

[406] http://www.pixelfudger.com/

we opted to take a freeze frame for that area and blend it with the moving parts of the image to get rid of noise over time. This type of noise was a huge problem with the first generations of normal maps. Sometimes brute force in the form of a denoise algorithm helped in this context.

Even though the data passes extracted from the light field contained normal and position data they differed from the passes generated by a 3D renderer. The position pass was oriented in a screen space or projection space coordinate system instead of a scene referred or world coordinate system as you would normally output from 3D. A reason to choose this approach is the independence of the camera model. This means the x- and y-coordinates are the pixel positions in the pixel grid of the image. The z-coordinate stored as values in the blue channel is a normalized depth map. This meant we could use our enhanced and corrected depth map also for an improved position pass. But at the same time the coordinate system of the live action plate, in our case the background, would not match to the coordinate system of the CG teapot in the foreground. The position data, that existed as numbers somewhere between zero and one, needed to be transformed to fit into the coordinate system of the teapot. Since the scale of the position data from the light field was very small and it was cumbersome to work in such a small coordinate system inside Nuke we chose this approach over scaling the teapot position data. Later a small tool was written to make this process faster and a bit more intuitive as the user would be able to control the position of the point pass in a 3D environment with a locator object (figure 47). A point cloud could be generated from the position data the same way as described before in the context of the teapot. Since it is not possible to rotate the position data in an easy way it makes sense to have the camera of the 3D scene in parallel orientation to one of the coordinate axis.

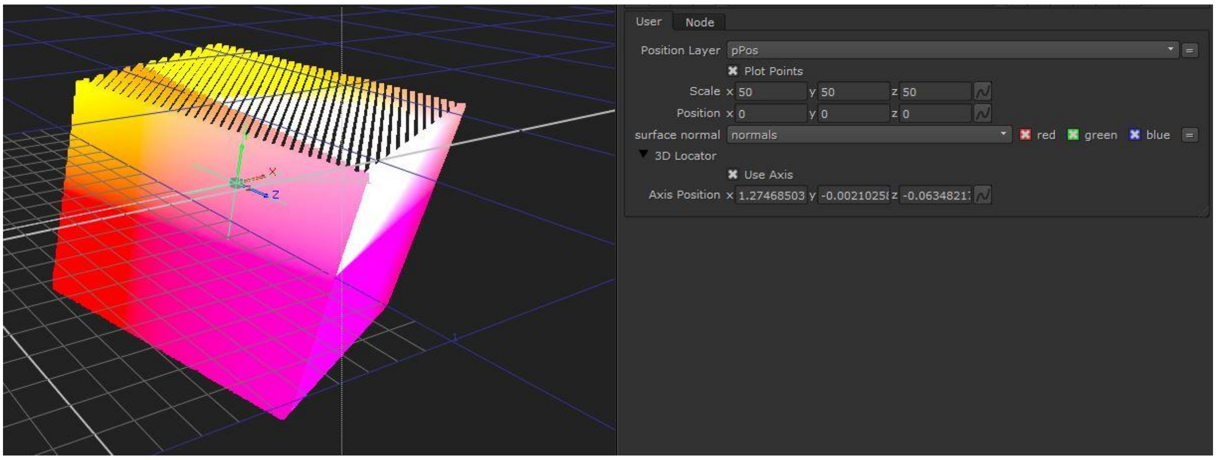


Figure 47: A small custom tool to transform the position data in a more intuitive way. The point cloud of a generic cube gets transformed.

Additionally, the teapot was rendered in a 3D scene that contained a virtual camera matched to the one filming the shot whereas the data pass from the light field had no connection to the camera whatsoever. This is not a problem as long as there is no camera movement in the scene. Actually a static planar camera could be used for the relighting process and give acceptable results. For more advanced setups the position data from the light field would have to be converted into a world coordinate system, though. To perform this conversion at least information about a camera is needed, i.e. the viewpoint and direction inside a higher level coordinate system. Ideally the exact parameters of the light field array as well as its position in the scene are taken into account when generating the position pass.

The normals were viewpoint dependent in both cases. Consequently, all necessary data was available to use a new instance of the relight node inside Nuke. We used the same shading model and light. As we did not have a way to exactly match the position data to the matchmoved camera during our test production, we either used a planar camera or a rough match. The shot camera from the 3D scene could actually been used in this case as for most of the shading operations a vector describing the view direction is sufficient. However, specular reflections might look wrong if the camera is not in the right position. Depending on the quality of the normals we usually decided to only use diffuse lighting anyway. Errors in the normal map kind of created surface structure that did not correspond to the color image which made specular reflections appear at places and in size that could not be justified in the context of the image. Yet a diffuse light pass could be created for the turntable and background area. As described before a blend operation was used to bring back some color and texture information. The aliasing problem could have also been fixed with the transform operation trick mentioned above but has not been necessary as we blurred the light pass anyway. To get rid of artifacts introduced by the remaining errors in the depth channel we had to introduce some kind of filter operation. Often a simple Gaussian blur was sufficient as we only used the diffuse component of the relighting operation most of the time. In the end we were able to render a sequence with a moving spot light that lights the teapot and the turntable in a realistic way.

We also tried generating a screen space ambient occlusion image based on the normal and position data to recreate the self-shadowing. This approach would have been based on methods used in game technology.[407] Due to the limited quality of the normal vectors generated from the light field data the results were not very promising while taking significantly longer to compute and crashing Nuke from time to time. Still this path may be an option worth pursuing for future developments as it may increase the visual quality noticeably. Furthermore, 2.5D relighting is not able to provide realistic reflections and shadows out of the box. Shadows could be approximated using one of the following techniques. Having roughly matched geometry by either manual modeling or from the point cloud it is possible to render shadows inside the Nuke 3D system. This

[407] an efficient technique to calculate ambient occlusion data from depth buffers has been first implemented by the Crysis Engine in 2007; cf. Bavoli and Sainz (2008); Carlsson (2010), see chapter 4.2.4 for more

can be done by either using geometry to cast the shadow on or by creating a camera projection at the position of the light. We tested both approaches but did not have enough time to come up with usable results yet. Issues are still the density of the point cloud, performance and again to find the correct scene scale without having the position data matched to a camera view first.

Figure 48: CG teapot and live-action background were relit with a spot light.

Figure 49: Live-action packshot source footage and a relighting pass (left to right).

Figure 50: Depth map (left) and generated mask for the foreground (right)

As shown in figure 49, we transferred this approach to the shot with the real teapot and cup and were able to relight the scene with several moving lights of different colors. A matte for the foreground area containing the turntable, teapot and cup was created using the improved depth data (figure 50). This enabled us to transition to a synthetic background similar to an animated matte painting in the course of the shot. The matte was refined using techniques known from the processes of color keying like combining a hard matte with high-contrast for the core area and a soft matte for the edge detail. As we blended to a different background we used relighting to also change the lighting condition. In this case the background showed a sky with clouds and sunlight coming from the right side of the screen. We matched the sunlight with a warm directional light. To get a feel of the blue skylight we tried an image based lighting (IBL)[408] approach. The idea is to use the color and luminance information in an image, often a HDRi as some kind of spherical map, to control the lighting of a scene. There are different ways to achieve this effect that is commonly used in path tracing renderers today. We tried a simplified technique that worked inside the compositing package Nuke and is freely available online as "Env Relight".[409] This tool computes reflections of an image map in lat-long[410] format only with normal and position images as input as well as a camera node. This tool is based around a very simple shading operation still the look can be adjusted by preprocessing the image map. Again, occlusions by the object itself or other objects are not taken into consideration using this method.

Finally 2D light effects for better integration were applied and the focus was set using the depth data, again.

6.3.4 Postproduction of the Portrait Scenario

One example of a portrait shot would be the shot of the man standing at the window and lighting his cigar. We shot the scene without the window itself and, therefore, it had to be added in a VFX process. Additionally, the shot aimed for an atmospheric and dim lighting, which could not be fully achieved on-set.

Figure 51: The original plate out of the middle array camera (left) and a first color correction using depth data to darken the background (right).

[408] cf. glossary "IBL"

[409] http://www.nukepedia.com/gizmos/other/envrelight

[410] a spherical image projection in lattidude and longitude coordinates, cf. glossary "spherical map"

This meant that color correction and creative color grading would be the main part of the postproduction process for this shot. Since we decided to do all our finishing work inside the Nuke Studio environment we could extend our grading tool arsenal by the relighting techniques described in 6.3.3 above. First we applied our image enhancement setup to the shot and created an improved version of the depth map for the desired view position, in this case the middle camera of the array. We always normalized the depth map according to the real distance of scene that we measured on-set. After setting white balance and black values as well as the overall hue, we then used the depth data to create a mask for the pixels in the background. This was easily possible by setting black- and whitepoint of the depth channel in a way that outputs a black and white image where all pixels at a certain distance become white while the foreground stays black. Only a little bit of edge treatment like blurring or eroding was needed to fit it nicely. This mask controlled a grade node that removed ambient light from the background by gaining it down in the selected area of the image (figure 51). Although the actor moved throughout the shot the correction did not affect any foreground pixels. At the same time we did a matchmove for the whole camera rig. This was done inside of Nuke with the *camera tracker* node. We experimented with the possibility to track several views at once. This option has been implemented for Stereo Shots and constraints the solving algorithms to the presumption of two cameras that are roughly at the same did not change inter-ocular distance over the course of the shot. Unfortunately these presumptions did not help in the case of a two-dimensional camera array as the presumption of all views being shot at the same height could not be turned off and gave wrong results. Also the performance was affected by the high number of views, the tracking process took up to 15 seconds per frame. The best results were achieved feeding the light field as a sequence of frames into the tracker moving along the rows and columns of the camera array for each point in time. (cf. figure 52) We could then extract the position data from the one camera moving around over time and recreate the geometry of the camera array. For static shots like the portrait shot presented here one frame would have been enough to extract the data, actually.

Figure 52: The nine rectified views of the shot.

To remove some of the light on his shoulder a relighting setup was used, that got the normal and position passes fed into. The quality of the normal and position data was good for the foreground but showed some errors and noise in the background. Since the background was static and not of key interest for the shot we used the already existing mask to merge the animated data passes of the foreground with a static and manually improved image of the background. A point light was positioned in 3D space using the point cloud representation of the position data. Looking at the output of the relight node at the same time as it updates in real-time made the process of setting the light's position and parameters like falloff, color and intensity an interactive process. This time the light pass was not overlaid with the original color image but merely subtracted from the plate instead. The strength of the effect could be controlled by adjusting the light intensity or the mix value of the minus operator. A second relighting operation was used to give the fire on the match more effect when lighting the cigar. Again, a point light source was positioned in the 3D space and the position value keyframe animated to follow the position of the burning match. (figure 53) Finally a static spot light simulated the light coming from a street lamp outside and another animated spotlight the blue light of a police car driving in the street.

Figure 53: The raw light pass (top left), the result of a blending operation with the original color image (top right) and the light added tot he composite (bottom left)

Next, the window was added as a 2D image projection onto simple low-polygon geometry. The lights could just be plugged into the 3D scene containing the window geometry and be transformed to the right scale to affect the window layer as well. Lastly the reflection on the glass was added as conventional 2D images on 3D cards, too. A rough approximation of a depth map for the reflection plate was created using image segmentation with manual drawn masks. Having depth information of all scene elements from the 3D setup, a manually created depth channel and the light field it is now possible to control the focus in the scene with a defocus effect inside the compositing package at the end of the process (figure 54). To achieve a more realistic behavior in defocused edge regions the defocus effect was cloned two times and applied to the reflection layer, the window and the original plate independently. Still it could be controlled in one user interface.

Figure 54: Digital focus set at different distances. The final composite (right).

Having the relation between scene and camera coded into the normals and position data the edit could be transferred to other viewpoints as long as there were depth, position and normal passes available for that view, too. As long as the camera models are identical the method by which the normal and position data has been created, generates data of equal scale in a common projection space coordinate system. Only the input data stream in the relighting setups of the script had to be replaced to generate a new view of the scene. The other elements of the scene could be shifted into the right perspective by replacing the camera information with the camera corresponding to the active viewpoint. This is another reason for using 3D setups for the window and the reflection in the glass. Even though this approach gave plausible results it is not a trivial process and might have some issues regarding the exact scene scale, in particular matching the proportions in the world space to the proportions in the projection space. In our case this was done by scaling the point data in view space proportionally to then match the scene scale of the objects in world space by eye. In the future more effort needs to be taken into the matchmove process to create an accurate scene representation that can be used to position all scene elements including the point clouds used for relighting if this approach is to be adopted.

6.4 Further Requirements

To sum up this chapter, we present a list of additional requirements based on the experience of the test shoot. Some of the requirements relate to the initially mentioned requirements in chapter 3.2, further refined by knowledge gained during production.

6.4.1 Preproduction

In Preproduction, no extraordinary problems occurred, the workflow and needs are the same as on a traditional production, since every production is different and has special requirements.

6.4.2 Production

6.4.2.1 Data Format

As mentioned in 5.2 one of the first considerations when designing a pipeline are file formats. To facilitate the on-set workflow as well as the transition to and from the successive postproduction steps a common data format for the light field is needed. It should be easily scalable in terms of the size of the light field, allow high bit depths and HDR content as well as different frame rates. The test shoot showed that if a light field containing way more than nine views is captured, compression schemes have to be part of that standard allowing for fast random access of any part of the light field and if possible also levels-of-detail (LOD). Apart from the light information coded in color values, it makes sense to also store the estimated depth or normal data in postproduction in the same data format. The format should be able to keep all data of a shot together and protect against corruption throughout the pipeline. A wide range of metadata should be part of the files or link to shot related information in a database. Finally the data format must be supported by a variety of production and more importantly postproduction tools. Even though this should be a long-term goal, the phase of experimentation and exploration that light field technology is currently should not be limited by a rigid data format.[411] At this point in time still it is important to decide for and stick to file formats, folder structures and naming convention throughout one production. Then the workflow can also be optimized using custom scripting solutions.

6.4.2.2 Metadata

The test production involved lots of manual work regarding the organization of files in cutlists, synchronization between cameras and keeping track of different versions and data passes for each shot. A

[411] cf. Root (2015)

key requirement for data management in today's media productions is metadata as described in 5.2. This can be part of the individual frames or movie files or stored in a dedicated database that links to unique shot identifiers. Based on the experience in the context of the test shoot it has been noticed that light field probably requires extended metadata. The conventional metadata like timecode and lens data including distortion values, focal length, f-stop needs to be stored for each view in the light field acquisition system. The data should contain information about the number and position of the camera or cameras in the array and the array configuration itself in terms of interocular distances and grid size as well as height, roll or tilt, for example. If there is calibration data available for each camera it also makes sense to store it for every shot take to have it linked to a certain take later in postproduction, too. As viewpoint and focus can be changed later, it is critical to keep the intention of the DP on-set, ideally as part of the metadata. This could include viewpoint, focus distance or focus points and areas in the image, if the director or DP is after a nonconventional type of focus effect. Ideally this data refers to a preview of the light field applications available on set.

6.4.2.3 Preview System

Throughout the production pipeline but most importantly on-location a preview system for the light field needs to be available. Requirement 3.2.3.2 preview can be extended by the option to view difference or disparity images between any of the views as well as the display of epipolar lines as overlay. This would be helpful for further technical quality checks performed by the DIT and calibration process, for example. Apart from the technical part this also needs clearly defined quality standards for light fields to make sure the final image is of predictable quality. For instance, this means avoid ghosting or edge artifacts when applying digital focus.

6.4.2.4 Calibration Workflow

In the case of the test shoot it was not a big deal to keep the scene available for calibration for quite some time. But this is not a good idea in a real production environment. A standardized testing environment and procedure is needed to allow for fast and efficient array calibration. Probably this should incorporate a clearly defined workflow and special test charts as well as a guideline to choosing the right array configuration in terms of inter-ocular distances and number of cameras. It might makes sense to use a computer based assistance system like the STAN (Stereoscopic Analyzer)[412] also developed by the Fraunhofer IIS.

6.4.2.5 The Light Field Report

On-set a detailed report on the camera parameters and quality of the light field should be created and delivered to the postproduction studio or digital lab. This can be compared to the “Stereo Quality Report” done by the stereographer in a Stereo 3D production.[413] Moreover, this links to the requirement 6.4.2.6 “competence”.

6.4.2.6 Competence

The situation that almost every person on-set knew something about the theory of the light field and the anticipated challenges in postproduction, affected the production in a positive way. There was less need for discussion on-set about certain technical requirements and special procedures that did usually not occur in traditional filmmaking. This leads to the requirement of special training of the set crew to shoot light fields. It is probably a good idea to have a specialized supervisor on set for all things light field related, much like the stereographer when shooting stereo 3D. Automated assistance systems as part of the camera system could also help in this context in the long-term.

6.4.3 Postproduction

6.4.3.1 Tools for Quality Enhancement

The test production showed that the 3D data generated from light fields needs improvement and editing at different steps of the postproduction process at the moment. This leads to the requirement of specialized tools that take user input to enable correction and editing of the data passes in an efficient way that might

[412] http://www.hhi.fraunhofer.de/departments/vision-imaging-technologies/products-technologies/capture/stan-stereoscopic-analyzer.html
[413] (Krause) (2013) 42

involve the whole light field at a time. This user input should be as intuitive as possible like sketching paint strokes or drawing rough 2D shapes with Bezier curves[414].

6.4.3.2 Transformation and Conversion of Data Passes

As mentioned in 6.3.3 the position data extracted from the light field usually had to be transformed in the process of relighting. Ways to transform the position data interactively and to convert it between coordinate systems based on provided camera information and origin points are to be explored.

6.4.3.3 Quality of Depth Maps

As described in requirement 3.2.4.1 high quality depth maps there is a need for depth data of a certain quality. We could verify this aspect in the context of the test shoot. Based on the experiences we propose a definition for the quality of depth data based on the following four parameters: 1. Temporal continuity, which means consistent depth values for objects in a scene over time or and different points in time respectively, 2. A high signal-to-noise ratio, 3. Minimal amount of artifacts caused by occlusions and 4. A high spatial resolution to reduce aliasing-artifacts. These requirements might not be able to satisfy with today's acquisition systems. Therefore, efficient solutions to correct or improve the data in postproduction are demanded as part of this requirement. Additionally, there is also a demand for a standard format of depth information regarding its relations to real-world scale and orientation.[415]

The quality requirement also transfers to other 3D data like the normal and position maps. Since they are based on the quality of the depth estimation in the tested approach we will not go into further detail.

6.4.3.4 Requirements to handle Light Field Data

During the shoot and especially the postproduction, the limitations of the new technology reached their limits. We knew we were bound to what the Fraunhofer Institute could offer us and knew we had to live with limitations.

The biggest limitation during postproduction, was the fact that we couldn't do actual light field compositing, due to the circumstances that we were only able to use view-rendering in Avid. In compositing

---

[414] glossary: bezier curves, rotoscope

[415] cf. Lukk (2014), Edwards (2015)

we were limited to the discrete nine initial views and the 3D data (disparity, depth, normal, position) gathered by generating disparity information between the views.

We developed helpful tools to load all the 9 cameras at once and also the appropriate 3D data for all the views in order to save time and to minimize user-mistakes. In order to improve the quality of the depth maps, we also worked with semi-automatic node base scripts to fill holes and reduce noise in the depth maps.

6.5 Conclusion / Chapter Summary

In summary, this chapter documented the test shooting, conducted to prove the feasibility of a light field camera array. Resulting from the shooting, it described the differences to a traditional production in terms of staff, hardware and workflows. Further, three test scenarios have been described in detail and resulting requirements for all the steps of a production were worded in order to understand what is necessary to integrate light field data.

7 Integration of Light Field Data

7.1 Macro Stages

Based on the experiences of the test shoot and the requirements collected in 3.2 and 6.4 this chapter aims at presenting some models and ideas regarding the successful integration of light field data in future production workflows. First, an approach taking into account the overall production workflow and the expected changes as described in chapter 4 and 5.3 is developed. As one example of the basic considerations the data format will be discussed in more detail as it is of major importance for efficient production and subsequent software development.[416] Chapter seven will then focus on micro workflows affecting the postproduction mostly and present an idea for a pipeline implementation of the macro workflow outlined here.

As mentioned in chapter 2.2 a light field as the term is used in the context of this work usually is parameterized as a 2D array of 2D images. This means the input for the workflow basically is a high number of images showing a scene from slightly different viewpoints. Consequently the overall workflow can be seen as a multi-view workflow. This is not to be confused with a multi-camera workflow where several discrete camera positions are used to capture a continuous event from different angles at the same time.[417] Even though adding more and more cameras to a scene would eventually lead to the acquisition of a light field, the term multi-camera would mean the use of several camera arrays to capture different light fields at the same time in this context. A good example of a multi-view workflow is a Stereo 3D (S3D) workflow where two views on a scene are captured simultaneously. Stereo photography has been done since the advent of the first cameras. The stereo 3D boom started by Cameron's movie Avatar in 2009 made possible by technological developments in the context of the digital cinema resulted in a range of tools and knowledge that can be used as a starting point for the integration of light field systems in a digital production workflow.[418] A simplified workflow scheme of a film production shooting S3D is shown in figure 55.

The second aspect that defined a workflow according to 5.1 is the deliverable, which is a movie or TV show in the scope of this work. These might be in a planar 2D format or Stereo 3D considering the formats delivered to cinemas today. But the multi-view workflow approach would also allow for output formats ranging from the full light field for interactive application as needed in the field of virtual or augmented reality to media for autostereoscopic displays containing a limited number of views across a horizontal axis.

[416] cf. requirement 2.3.4.3, "The State of Rendering – Part 1"
[417] cf. Jago (2015)
[418] cf. Näther (2009) 122, Engelhardt (2012) 3

Depending on the end format the unnecessary views can be dropped at any time in the proposed workflow model to gain efficiency without having to change the entire production pipeline.

The third assumption that is made is the existence of a VFX process as part of the workflow. Like with Stereo 3D the imperfection of todays recording hardware as well as the time-pressure on a film set probably leads to the need of corrections for almost all of the shots.[419] In the case of the test shoot this has been partly done as part of the rectification process at the Fraunhofer IIS and in the correction phase of the data passes at the compositing stage, again. Apart from that, VFX is a core part of todays storytelling techniques as mentioned in 3.3.1. This also points to the non-linearity of today's production workflows. As briefly touched on in the context of 5.1 today's production workflows are international and split between several facilities. It is worked simultaneously in different departments and iteratively across workflow steps. Additionally, the D.I. or color grading stage adopts more tools and techniques that belonged traditionally to VFX. That is why it is called finishing in the flowgraph of figure 56 and it is proposed to share technology between the steps VFX and finishing.

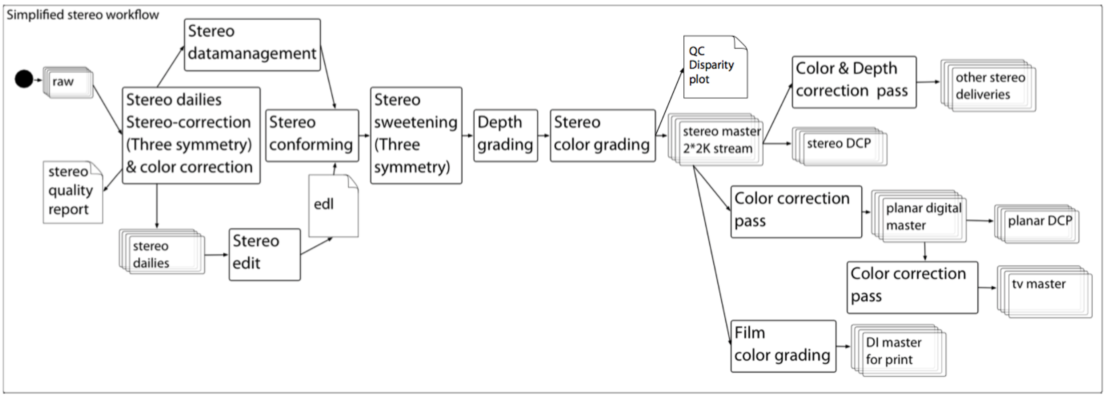


Figure 55: A simplified stereo 3D production workflow

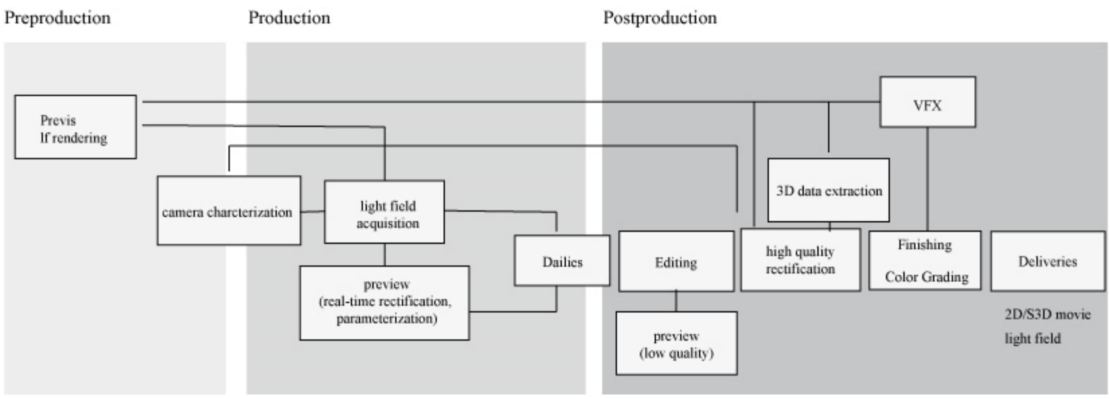


Figure 56: An even more simplified scheme of a future light field workflow

419 Townsend (2010) 443, Krause (2013) 42, Mendiburu (2009) 34

Like in the stereo 3D workflow an additional workflow step has been added for the image rectification, i.e. the correction of the input footage. The depth estimation or generation of normal and position data could also be done at this stage as this is a computational intensive step that would slow down the process later in the workflow. At the current state of technology this process stage would require a render farm most likely.[420] Therefore, it would probably be done at a specialized facility or incorporate cloud computing. Likewise, this workflow step as well as the camera characterization as required per 3.2.1.3 "camera characterization" for post and array calibration needs to be done by specialized tools and personnel. Apart from that, the team members keep their creative freedom and tasks compared to a conventional workflow as described in 5.1.

A workflow that integrates the described light field applications defined in 3.4 and used in one of the scenarios mentioned in chapter four, would then be described as follows.

During the production phase a special camera system is used to capture light fields of a scene. A fast, ideally real-time, preview enables the director and the other departments on set a preview of the view rendering and digital focus. Color Correction can be applied, too. It is possible to interactively apply changes to these parameters. Also relighting is possible as a low-resolution preview that can also be part of a virtual production process containing VFX elements. The settings made inside the light field viewer get stored in metadata together with the raw data from the light field array and converted into a special data format. The original data is archived. This special data format can be replayed with the settings from the set preview in slightly better quality during dailies. The editing takes place in a conventional non-linear editing system (NLE) that has the same viewer logic applied to render a preview of viewpoint, focus and possibly lighting. The changes made during editing can be stored in a new version of the metadata. The editing list is sent to process the high quality rectification and depth estimation. Depending on the acquisition system a draft version of depth data already existed. The processed files still contain the metadata and can now be edited as part of the VFX as well as at the D.I. stage. In VFX the 3D data is also used apart from the viewing and rendering toolset. The viewer technology is extended by a toolset that enables efficient editing of all views simultaneously or transferring the edits from one view to the other. Some research in this direction has been done by Jarabo, Masia, and Gutierrez, for example.[421].The light field and CG elements might be converted into a common data representation during the compositing process. Eventually the final image is rendered at the finishing stage or during the VFX process using the metadata from the editing, set or new settings and the viewer set to a higher quality mode. If the end format is going to be a light field the data is either

[420] cf. Lukk (2014)

[421] cf. Jarabo, Masia, and Gutierrez (2011), Jarabo, Masia, and Gutierrez (2014)

converted to a data format suitable for a light field display device discarding the metadata that control viewpoint and focus. Alternatively, the data format stays the same and just changes the compression scheme as the end-user's viewer application uses the same core technology as the viewer in the production process. If the main deliverable is a 2D theatrical release, the view rendering process has to add back any image distortion that has been removed as result of the rectification process to achieve a look that has the characteristics of a traditional lens.

Summing up, a multi-view workflow is favored that is built around a data format with strong metadata allowing for ideally non-destructive edits along the production stages incorporating methods to efficiently propagate edits from one view to the others. Against this background the final part of the work will focus on models to built tools for micro-workflows in postproduction.

7.2 Data Format

7.2.1 Data Representations

Moving on, a closer look is taken at some of the aspects that have to be taken into account when deciding for a data format. In the course of this interlude also some possible solutions will be presented.

Much like its analog counterpart the purpose of a file in the world of computers is to safely store information in an organized way. In the case of image data the information are a grid of color values, i.e. numbers that get extended by metadata, which is often some kind of text-based data.[422] In the context of light fields the data consists several views that need to be stored at one point in time according to the parameterization introduced in 2.1. Essentially, the idea of using light field technology in the first place is to have more information of a scene at hand in postproduction. Having already to cope with lots of data in a film production it makes sense to organize the data in a way that makes things simple and clearly laid out but easy to access at the same time. As required in 3.2.2.2 "Solid Data Format and Metadata" the data format's task is to keep the different views together once their symmetry is verified. The latter should be done on set as part of the quality check procedure done by the DIT and data wrangler. Symmetry means, like in the context of other multi-view systems, the geometric alignment, temporal synchronicity and matching colors

[422] cf. Kainz and Stanczyk (2010) 470-471

between views.[423] To avoid errors down the production line and avoid mismatches when working with proxy files there is a strong trend to working in an online workflow[424] as technology advances. If a raw format from the cameras is recorded, it usually needs to be debayered or demosaiced and transformed into a known color model before viewed on a screen. There is now increasing real-time support for a low-quality processing inside of editing, finishing and compositing software.[425] Even though it is seldom used in production at the moment it probably will be in the future. This leads to the requirement that the data format should support both raw data and processed images, which might be connected to the need for different bit depths. As per requirement 3.2.1.4 "Sensor Technology" HDR data has to be supported, too, as modern cinema cameras have more than the five stops of dynamic range possible in BT.709 and the tone mapping process will only be done as part of the DI.[426]

For light field data, it seems efficient to only have to keep track of one file instead of nine and more. That would mean storing the information of all views in one big file. But, the contrary is the case as long as all views have to be transferred from the storage to the application just to display one view. This comes down to the requirement of levels of detail, compression and random access as mentioned in 3.2.2.1. Yang et al. presented a concept for real-time applications that they call "finite viewpoint design".[427] It uses a server-based compositor to assemble only the pixel data from the views needed for a specific request. As part of the SCENE research project a more abstract scene representation format has been proposed. The Scene Representation Architecture (SRA) defines three layers that aim at a unified scene description across sample based scene data and object based CG data.[428] The base layer consists of basic spatio-temporal objects, so-called Atomic Scene Elements (ACELS). These can be arranged and combined in the scene layer (description of ACELS relations) and modified in the director's layer (rules of interaction). ACELS could be represented as superpixels, which are temporally consistent clusters of pixels grouped by similar color and shape. This might also be an interesting approach for long-term developments but won't be discussed in further detail here. Currently light fields still seem to be stored as a set of individual images or pixel clusters most of the time.

---

423 Krause (2013) 42-43

424 see chapter 5.1

425 cf. Van Hurkman (2014) 17

426 Van Hurkman (2014) 145-146

427 Yang et al. (2002) 3-4

428 Rogmans (2014)

In chapter 7.1 a workflow is proposed that separates the data workflow from the image manipulation workflow. The original image data stays untouched for most of the time and the edits only get applied at the end of the process. This approach can also be found in today's pipelines in the context of color grading, for example, that often only gets applied as viewer look-up-tables (LUTs) until the final grading process at the end of the production workflow. This concept has also been applied in TV productions by now.[429] It has the advantage that there is no reduction in image quality due to resampling until the very last step. And storage space is saved, as there is not a big need for intermediate file versions and pre-renders, which might be of significance considering the file size of light fields. A popular implementation of the idea of a non-destructive raw file workflow is the Adobe Raw file reader and the associated Adobe Digital Negative (DNG) format in the context of photography.[430] The edits done in Adobe's Lightroom, for example, are stored in an xml-format and can be reapplied when a file is opened in Adobe Photoshop.

An existing standard, that seems to fulfill most of the requirements is the image format OpenEXR that has been developed by Lucasfilm's Industrial Light & Magic in 2003 as free software under a modified BSD[431] license.[432] At the time of writing the set of C and C++ libraries is available in version 2.x. As it is free and open-source, OpenEXR support is wide spread across different software tools today. This also complies with requirement 2.2.2.2 that demands a free data format for light fields to make it a standard across platforms and software vendors easily. Recently all major software developers and studios have successfully adopted open data formats like Alembic, OpenVDB or PTex.[433] OpenEXR supports an unlimited number of views with an arbitrary number and combination of channels each.[434] It has extensive options for fully customizable metadata and other attribute data like information about camera position or color timing.[435] It supports three data types, 16bit (half) float and 32bit float as well as 32bit unsigned integer numbers. 16bit floating point numbers can store a dynamic range of 30 f-stops with full precision at 1024 steps per f-stop color resolution and another 10 f-stops with some loss of precision at the low end. That is sufficient for all currently existing HDR acquisition systems.[436] 32 bit of precision are only used in the context of CG

---

[429] cf. Tukiendorf meeting records, "Grundy Ufa Nutzt Filmlight Flip" (2013), "A Look at FLIP with GrundyUFA" (2015)

[430] cf. https://helpx.adobe.com/photoshop/digital-negative.html

[431] Berkeley Software Distribution (BSD), a free Unix-like operating system gave this family of permissive free software licenses its name; "BSD Licenses" (2015)

[432] Kainz et al. (2013) 3

[433] cf. glossary

[434] Kainz et al. (2013) 3,7, 11; Hillman and Welford (2007)

[435] Kainz et al. (2013) 4

[436] Kainz et al. (2013) 3

data passes like position or depth at the moment. Unfortunately there is no common color management standard for OpenEXR but the Academy Color Encoding System (ACES) by The Academy of Motion Picture Arts and Science is increasingly adopted in today's pipelines.[437] Color management is an important part in the light field acquisition process and it would make sense to use the ACES system, too. The exact implementation will depend on the design of the acquisition system, of course.

Relevant to the application for light fields, multi-part files can be created, that organize image parts into sets of channels that can be accessed individually.[438] In this context the ability to encode and decode files in a multi-threaded way is also a basic requirement that is supported by OpenEXR.[439] In tiled image mode different levels of detail per tile, so-called MIPMAP or RIPMAP-levels[440], can be stored to speed up the file access and processing as well.[441] A variety of both lossless and lossy compression models are supported, too.[442] But these only work on a per view basis and do not compress data across several views. This might be a drawback when working with light fields as it will be discussed in more detail in chapter 6.2.2. The topics compression and transfer speed get more important as the light fields get more dense with way more than just 10 or 20 views. Usually working in a server based infrastructure transferring these files will be a bottleneck. In the context of the test shoot the limit for real-time playback over a gigabit network connection has already been reached with one uncompressed view. Additionally, the file reader in a software will be overloaded with such a high number of views since the basic file handler process is still single threaded in most cases. To improve performance for files with lots of channels production pipelines usually tend to split the files in several smaller files to speed things up nowadays. Custom scripts or software then assembles those files in a user-friendly way.[443] To protect the files against corruption checksums need to be computed for every file transfer. These can also be stored as part of the metadata to save time the next time the file is copied.

Apart from the multi-view format there is another useful type of data representation that has been introduced in chapter 4.3.3: deep data. Deep images, opposed by flat images, are also supported by OpenEXR. Even a mixture of deep and flat images in one file is possible, although it might not be a very good idea organization-

---

[437] Kainz and Stanczyk (2014) 621-622; "The Art of Digital Color"
[438] Kainz et al. (2013) 4,5
[439] Kainz et al. (2013) 4
[440] cf. glossary "texture maps"
[441] Kainz et al. (2013) 11
[442] Kainz et al. (2013) 15
[443] reasons probably are inefficient reader plug-ins mostly, cf. Sciolette et al. (2009)

wise. [444] Instead of just one sample per pixel in conventional data formats, deep data stores an arbitrary number of samples per pixel that can also vary from pixel to pixel.[445] This data representation can be very useful at the compositing stage to combine light fields with CG elements as described in chapter 4.3.3. By comparing position data from different viewpoints color values at different depths could be figured out. These can then be mapped to pixel positions in a new view created during view rendering forming a deep image. A deep data file stores this information as several subsamples at different depths of a certain pixel position. Tracing a light ray backwards from the position of the camera, opacity and depth information is sampled whenever the ray comes across a surface (figure 57). If the surface has a volume, a set of deep-front and –back values is stored. Later additional sub-samples can be computed using the Beer-Lambert equation that describes the absorption of light travelling through a light absorbing material.[446]

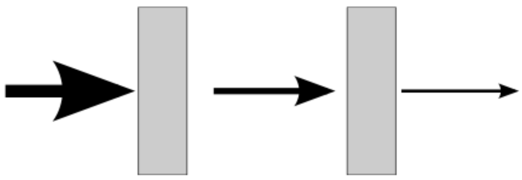


Figure 57: Light passing through objects from left to right, each volume attenuates a fraction of the light.

Of course, it should be explored if it is more efficient to render a CG element as a light field instead in the long-term. Chen and Horn present the possibility of merging light fields on a level of light rays in their work on the LightShop system.[447] Still, this approach needs opacity information about the objects in the scene and otherwise will assume fully opaque objects only.

To compensate for different requirements due to different host applications and potential light field acquisition systems an approach based on an independent framework could also be an option. The Moving Pictures Company (MPC) developed a proprietary framework called Muggins that includes tools to store data in a common, application independent data format that can be outputted to any format needed by a certain task.[448] The conversion to and from application specific data formats happens automatically in the background inside the framework. This framework has been further developed as the so-called SceneHub that now relies heavily on the framework of the fabric engine.[449] Fabric software describes their engine as "a

---

[444] cf. Kainz et al. (2013) 5
[445] Kainz et al. (2013) 5
[446] Heinen (2013) 11, cf. Hillman (2013)
[447] Horn and Chen (2007)
[448] Green et al. (2014) 104
[449] Ricklefs (2015)

framework for building high-performance content creation and content deployment tools".[450] It is a set of libraries that focus on high performance multi-threaded execution and provide the abstract programming language *KL* that can be compiled at runtime. As it can be integrated in almost every digital content creation package, it could not only be used to define a future data representation for light fields but also to develop a portable light field viewer application. A free fifty seat license for development makes fabric engine a competitive alternative to application dependent developments and workarounds.

7.2.2 Compression

When looking at the amount of data, produced by a single recorded frame and its calculated 3D Data the need for compression instantly seems reasonable (see Figure 58).

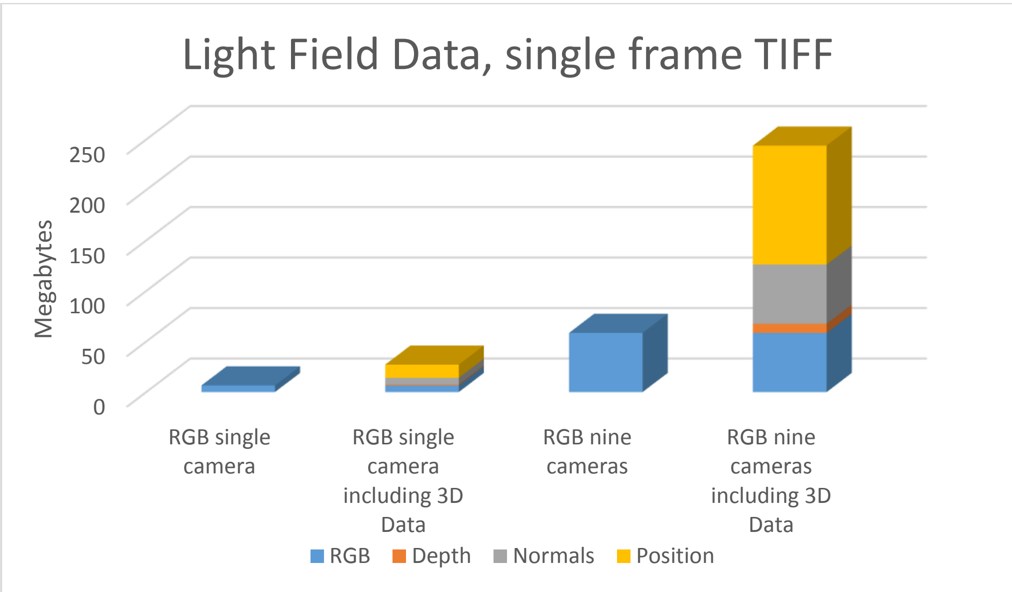


Figure 58: storage consumption of various data sets.

Naturally, a light field captured as a stack of 2D images, has a lot of redundancy due to similar contents. This chapter focuses on how images needed for light field postproduction are stored and handled. To minimize storage issues and simultaneously maximizing performance when using data of multiple cameras, compression is a topic that quickly comes to mind. Compression can be split into two main categories, lossless and lossy.

[450] Doyle, Taylor, and Zion (2014)

Lossy:

If the algorithm is allowed to alter an image, great image compression rates can be achieved. The file therefore becomes very small, but a lossy compression also means that the file is altered in such way that it can't be reconstructed to its original state. Lossy compression reduces image information, which is considered less relevant or not of meaning for the human visual system. But nevertheless, you lose data in a lossy compression and by decompressing you only get an approximation to the original.[451]

Lossless:

Lossless compression is achieved by removing redundancies, by eliminating redundant information and encoding it to a smaller file. No original data is lost, so the original image can always be completely reconstructed.[452] According to Shannon, the maximal compression possible is determined by the entropy of the source. The more redundancy there is, the more it can be compressed.[453].

Light field images introduce additional information into the general compositing workflow. This information has to be as accurate as possible, but at the same time, it needs to be stored efficiently and easily accessible in order to speed up the process of file-reading and file-writing as much as possible and not to slow down the postproduction workflow. File Transfer between facilities is also a key element when talking about compression.

Magnor & Girod introduced two ways of light field compression. One is a modified version of a video-compression codec and the other is entirely based on disparity-compensated image prediction. They were able to achieve compression ratios between 100:1 and 2000:1, depending on the quality and the image characteristics.[454]

The necessity of compression is described by Magnor/Girod in a very understandable way:

> "Compression is necessary to transmit the data as well as to fit all information into local memory during rendering, while fast access to arbitrary light field segments is crucial to enable interactive rendering rates." [455]

---

[451] cf. Okun and Zwerman (2010) 477
[452] cf. Okun and Zwerman (2010) 472-473
[453] cf. Shannon (1948) 14
[454] cf. Magnor/Girod (2000) 338-339
[455] Magnor/Girod (2000) 338

The first mentioned method of compression is based on video compression codecs: light field images are hereby divided into blocks, each block is coded with a fitting block coding mode. The image is then converted into YUV colorspace and the chrominance components are down sampled by a factor of 2 both in H- and V-directions. The coding process then starts by dividing images from the light field array in *intra (i)* and *predicted (p)* images. I-images are selected evenly throughout the clip range and are then compressed with a discrete cosine transformation (DCT) and coefficient quantization. I-images then serve as reference for p-images.[456]

By further comparing adjacent light field images, points/information often appear in both (multiple) images, but at different positions, due to parallax. This displacement is the so called *disparity*. Only the amount of disparity needs to be coded, because the disparity direction between the images can be deduced from known image recording positions.[457].The major drawback of this technique is the use of p-frames, therefore it is not suitable as an intermediate format for postproduction, due to the lack of being able to edit single frames, but of huge interest as final output format, where it is necessary to keep the light field functionality. Fields of application could be virtual-reality applications (Oculus, auto stereoscopic images, light field monitors), as well as immersive streaming services with flexible viewing positions.

The second mentioned compression method is entirely based on disparity-compensation. In order for this method to work, disparity information has to be available for every light field image. The amount of disparity is then found by comparing blocks of an image to its four neighbouring image-blocks and building an average, which is then compared to the original blocks disparity. The disparity value with the smallest prediction error is then chosen as the block's disparity magnitude. This technique also allows a fairly good estimation of intermediate images.[458]

Depending on the encoding and decoding speeds of this algorithm, it might be of interest for postproduction as well as final output formats.

In general, the introduced algorithms still take too long to compute the compression, making it for now not suitable in a real postproduction environment for now. Current workflows probably have to work with low-quality proxies and common (fast) compression-algorithms like zip, RLE or wavelet-compression of the light field data, in order to maintain accustomed habits and working-speeds. For the near future, the disparity-

---

[456] cf. Magnor/Girod (2000) 339
[457] cf. Magnor/Girod (2000) 339-340
[458] cf. Magnor/Girod (2000) 341-342

compensated approach is more suitable due to the possibility of random access of every frame, instead of only being able to access i-frames. So, a combination of disparity-compensated compression and traditional wavelet compression, modified for higher dimensional data, would probably deliver good results.

Another possibility for compression is down sampling a light field into a sparse representation, which later can be up sampled to achieve higher quality.

It's hard to define a general compression approach, fitting all light field use cases. The decision on which compression to use strongly depends on the desired output format, needed quality and the necessary decompression speeds. Web and mobile content might be a candidate for lossy compression with relatively low quality in order to keep the amount of data as low as possible and reduce loading times. Virtual Reality, Film and immersive gaming applications have high quality demands, and often the need for fast decompression algorithms, in order to provide real time content without delay. In general, it is necessary to divide compression into intermediate codecs, delivery codecs and archive codecs.

7.3 Summary

This chapter presented some ideas on integration schemes for light field technology on a higher workflow level. Taking into account the requirements collected as part of chapter 2 and 5 as well as the foundations on film production workflows and pipeline from chapter 4 a multi-view workflow was described as an example for light field integration. In this context some basic considerations could be presented and applied to the new medium light field. One of them is the data format when it comes to building a pipeline. The second part of the chapter therefore focused on requirements and some available solutions for the development of a data format that can serve as a standard for light field production and distribution. As light fields hold a lot of information methods for light field compression have been discussed as part of the considerations.

8 Light Field Toolset / Plugin

As with every emerging technology, the support by major software packages available for postproduction, is not given right from the beginning. For now, only proprietary software or plugins for specific file formats are available for handling light field data. None of the publicly available products is capable of handling moving images in a sufficient way. Most of them are only able to handle their specific data formats, because of no general agreement on how to handle light field data. In this chapter an overview over current commercial available Software solutions provided by RayTrix and Lytro is given. As well as an insight on Fraunhofer Institute IIS' dedicated light field viewer, and their Light Field Plugin for Avid Media Encoder. Propositions, on how to edit light fields in a more advanced way, are given and an advanced editing Interface created by Jarabo et al. during a study concerning user behaviour during light field editing is presented.

The Industry has to come to an understanding in aspects of meta-data-support, compression and data-formats, as mentioned in chapter 7.3. Despite all that, it is necessary to implement special functionalities for the major postproduction applications. As we cannot cover a suggestion of tools for each specific postproduction software, we will concentrate on The Foundry's Nuke, it being the industry standard compositing software at the moment, and give a proposition on how to handle light field data and what needs to be included in a basic toolset, as well as some insights what might be necessary for an advanced toolset, from there on later adaptions for other software suites are possible. The node based workflow allows image refinement throughout the compositing workflow, opposing to current software solutions. The proposition can afterwards be derived to other software, with slight modifications necessary, but keeping the basic idea and workflow in order to develop a compatibility between the applications. In order for light field postproduction to work properly, a global interface and interchange-capability between the major postproduction applications is necessary.

In this chapter we will investigate possibilities of editing light fields with the potential benefits of using depth-, normal- and position information. In addition, we will again consider imperfect depth reconstruction and its influences in the editing process and how high the imperfections can be to still be forgivable for certain solutions. Due to time limitations, we were not able to program the tools, neither was is possible for us to experiment with a user-group. This part has to be investigated posterior.

A light field is a four dimensional data structure, thus it is a challenging task to edit a light field with common displays and input devices. Also the much bigger amount of computational performance needed and the amount of storage, due to redundancies, are another factor and local edits have to be propagated accordingly to every possible view of the light field. We will specifically concentrate on an approach suitable for current state of the art compositing workstations with regular 2D displays and input devices.

8.1 Available Light Field Software

8.1.1 Raytrix

Raytrix is offering a User Interface called Raytrix LightfieldViewer capable of viewing their proprietary light field format *TRIX* (.trix). Since the functionality is limited to viewing captured light field data the User Interface is kept really simple only consisting of a 2D-View capable of refocussing an image by moving a slider (Figure 59) and showing the according depth map of an image. As well as a 3D-View capable of a limited geometric reconstruction of the captured object (Figure 60). The 3D View is also capable of viewing the depth of the assigned focal-plane (right image in Figure 60).

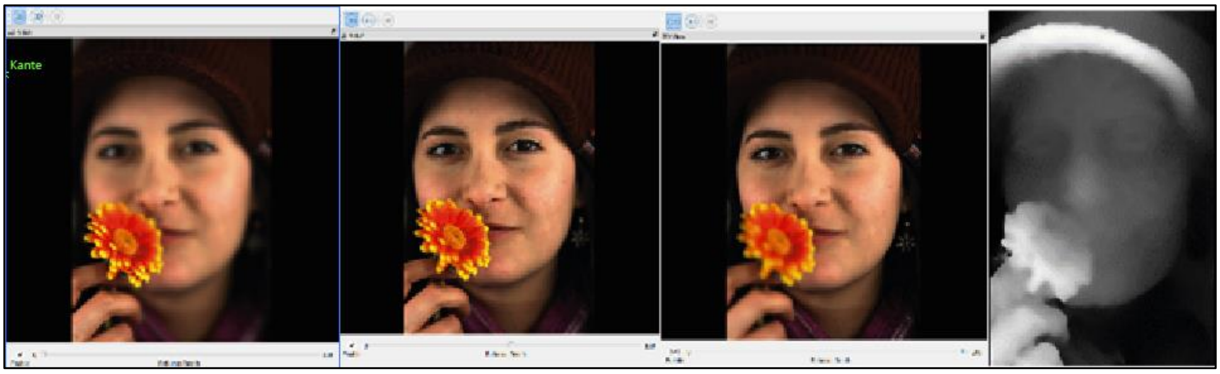


Figure 59: Raytrix User Interface. From left to right: Focus at 0, Focus at 50, Focus at 100, depth map.

Figure 60: Raytrix User Interface. From left to right: 3D View, different perspective in 3D View, Focal Plane visualized in the 3D View.

Advanced Options are also offered by Raytrix, but again very limited. Limitations being Depth Scaling for more accurate focus position, amount of depth of field and the strength of blur. Export is only possible as a 2D-Image (.png).

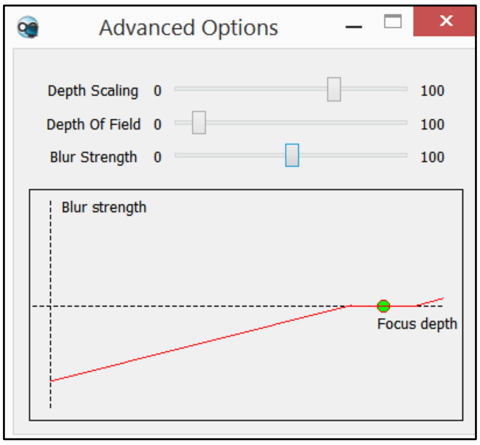


Figure 61: Raytrix's Advanced Options

8.1.2 Lytro

Lytro presents its current software Lytro Desktop 4.2.1 for its own Cameras LYTRO and Lytro Illum and their proprietary data-format with a user interface similar to most photo-editing software environments. Lytro Desktop is segmented into 5 main categories. Library is giving the user the possibility to organize and handle all their stored light field images inside the Software, capable of filtering images for certain flags, sorting them in folders and flagging, rejecting and rating the images. The library also is offering the possibility of adding captions to certain images, sharing them to different media platforms, playing a gallery slide show and playing previously animated images. The basic light field features such as digital focusing and view rendering can also be previewed in the library panel, but without the option of setting parameters for said actions.[459]

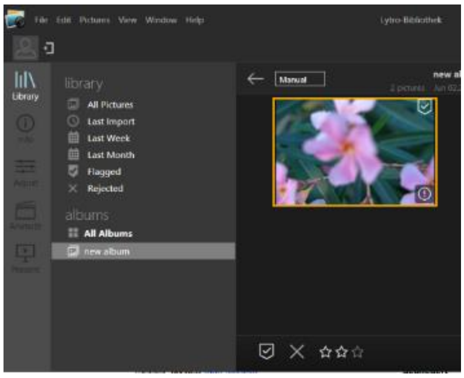


Figure 62: Lytro's Library.

The ability of view rendering is possible throughout all the different panels by clicking and dragging inside an image.

Lytro Desktop 4 is giving the user the choice of four configuration options in terms of image rendering quality. The default setting being *Balanced* can be set according to one's special needs and hardware configuration. (Figure 63)

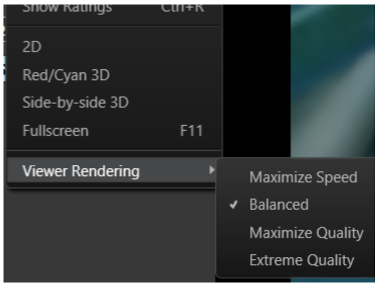


Figure 63: Lytro's View Rendering Performance Settings.

[459] cf. Lytro - Desktop 4 - Using Libraries to manage pictures (2014)

The Info panel is providing the user with information about the capture settings for a selected image. The following information is displayed: Taken (date), Model of the Camera, Shutter, ISO, F-Stop, EV, Focal Length, Mode, Flash, Captions, Tags and Rating.[460]

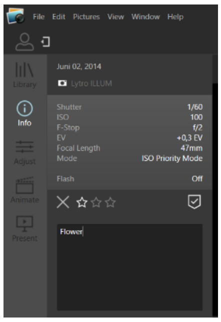


Figure 64: Lytro's Info Panel.

The most important panel when it comes to editing light field images is the one named Adjust, which contains image adjustment tools for enhancing and correcting the captured images. Only the adjustment tools concerning light field editing will be described since the other functionalities (seen in Figure 65) should be known from traditional image adjustment.

Figure 65: Traditional image adjustment options.

Lytro groups the special functionalities for light field images in the panel "Virtual Camera" (Figure 67), allowing the user to adjust aperture and perspective posterior to acquisition. Aperture, Depth Map, Focus Spread and Tilt are the four sections of which the Virtual Camera panel consists. The **Aperture** slider gives the user the control of simulating a virtual aperture in a range of f/1 up to f/16 (shown in Figure 66).[461]

Figure 66: Virtual Aperture settings from left to right: f/1.0, f/4.0, f/16.

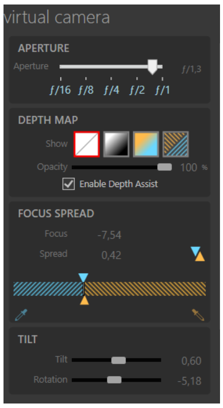
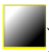
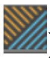


Figure 67: Virtual camera settings.

[460] cf. Lytro - Desktop 4 – The Info Panel (2014)
[461] cf. Lytro - Desktop 4 – Adjust (2014)

The **Depth Map** panel makes it possible to:

- reveal the depth map of the image, with darker being closer and lighter being further in terms of depth (Greyscale )
- exactly set focus according to depth displayed in gradiations of blue (foreground) to orange (background) (Two Tone )
- click on an area in your picture to see what's in front of your click (rendered blue)and behind your click(rendered orange)(Depth Assist )
- easily identify depthmap errors by overlaying one of the depth map options with lower opacity and comparing them to the original image. (Opacity 

)[462]

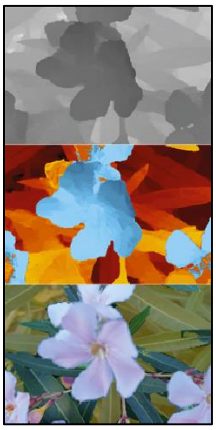


Figure 68: Depth Map viewing options.

Lytro Desktop also gives the possibility to export and reimport (Figure 69) an editable depth map in case you identified depthmap errors which can easily be fixed with an external image editor such as photoshop or similar.

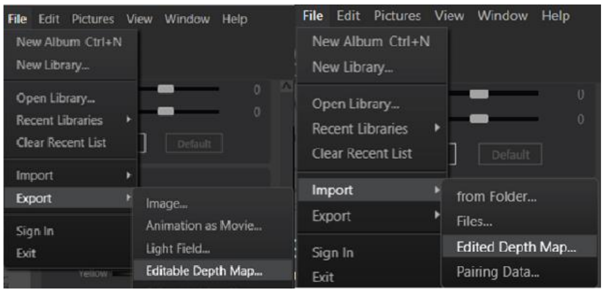


Figure 69: Editable Depth Map Export and Import.

Focus Spread enables the user to set the in- and out-of-focus area of the image, even going so far to behaving like having two different apertures in an image by having the possibility to set the defocused areas to a f/1.0 aperture (really blurry) while keeping the primary suspect or the area of interest in focus with an aperture of f/16.

Focus being the point desired to be compeltely sharp in the image either picked by clicking in the image on the desired area or by sliding the white bar (Figure 70) to the desired point. Spread being the area supposed to be in focus, the bigger the "spread", the bigger the amount of in-focus pixels in the image. The Spread is controlled by either picking certain depth areas with the droppers () or by adjusting the sliders ().[463]

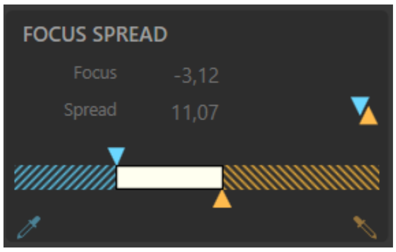


Figure 70: Focus Spread tool.

[462] cf. Lytro - Desktop 4 – Adjust (2014)
[463] cf. Lytro - Desktop **4** – Adjust (2014)

The **Tilt** tool allows a user to adjust the perspective of an image after capture, similar to the functionality of a tilit-shift-lens. Tilt tilts the image around the focal plane, decreasing depth of field and Rotation rotates the image around the focal plane, also decreasing depth of field. Making it possible to set multiple elements in focus which are originally not in the same focus-depth. [464]

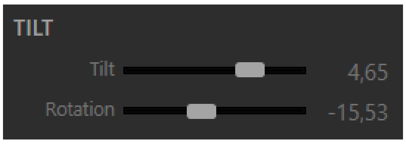


Figure 71: Tilt tool.

Lytros **Animate** panel in Lytro Desktop 4 makes it possible to animate an actual still image, but due to the light field possibilities a still image can be converted to a moving image by animating various combinations of focus, pan, zoom, perspective, aperture and tilt.[465]

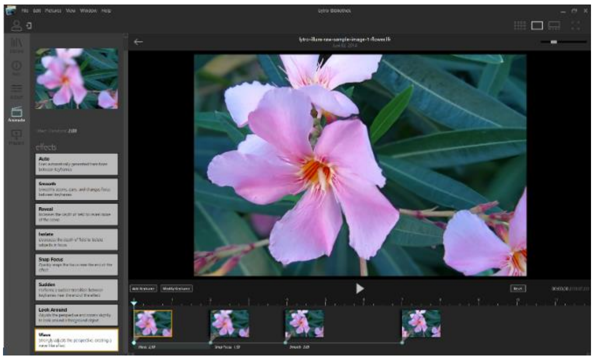


Figure 72: Lytro's Animate Panel.

The following description of the process of animating inside Lytro Desktop 4 is completely taken from the Lytro Desktop User-Manual as it is not primarily in scope of our work:

Animations are added from the left-hand panel, and modified using keyframes. The animation panel will open with the currently selected living picture. Living pictures have a starting and ending keyframe, by default, displayed in the timeline. When a living picture is opened in the animation panel, the starting and ending keyframes will be displayed by default.

To add an animation, first, select the starting point for your animation by clicking on the first keyframe in the timeline. Then, click on the animation that you want to add from the available options.

---

[464] cf. Lytro - Desktop **4** – Adjust (2014)
[465] cf. Lytro - Desktop **4** – Animate (2014)

- *Select a keyframe by clicking on it in the timeline, and then click on an animation option to add it to the selected living picture*
- *Click and drag keyframes to adjust the playing time of the selected animation*
- *Click the Play icon in the middle above the timeline to view the animation*
- *Manually add changes by adding and modifying keyframes*
    - *Click **Add Keyframe** to add a Keyframe*
    - *Click and drag to move your living picture within the frame, or*
    - *Click to refocus, or*
    - *Zoom using the **zoom slider** or mouse wheel*
    - *Then click **Modify Keyframe** to save the change into the animation*
    - *This will allow you to manually modify the animation effects, so that you can choose framing and focus as the animation plays*
    - *Add and modify as many keyframes as desired*
    - *To remove a Keyframe, click on it and click "Delete" on your keyboard*
- *Only one animation can be applied at a time*
- *When you add or modify an animation, changes will be automatically saved to that living picture's timeline, and available the next time you open it in the "Animate" panel*
- *To view an animation, in the Library mode click the "Play" icon*[466]

Lytro Desktop 4 – **Presentation Mode**

The presentation panel is mainly a presenter to show images and also animated images in fullscreen. We won't go into detail of traditional presentation methods, but lytro integrated features which are of intereset concerning the output of light field data. By integrating the possibility to output each of the edited images automatically either in traditional 2D view, anaglyph 3D and side-by-side 3D (Figure 74) Lytro is stepping in a direction of giving the user flexibility in viewing-formats.

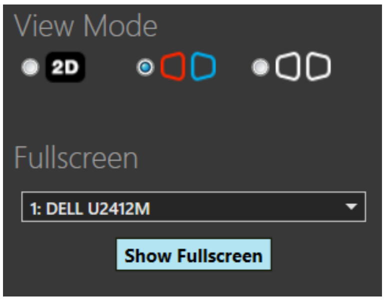


Figure 73: Lytro's View Mode output options and device selection.

Figure 74: Output modes from left to right: 2D View, anaglyph 3D, side-by-side 3D.

[466] Lytro - Desktop 4 – Animate (2014)

8.1.2.1 Additional Export Features

Lytro also implemented other helpful export and import features, next to editable-depth map export/import mentioned earlier. The formats can be summarized as 2D formats, 3D formats, Lightfield formats,Editable Depth Map, Animation Export and Online Uploads.

The export of 2D Formats includes exporting the refocused image as a still in different formats called *Refocused Jpeg* (.jpg) and *TIFF* (.tif). JPEG-Files will be exported in sRGB jpeg standard color space, while TIFF-Files are stored in 16bit ProPhoto RGB color space. The 2D output resolution is usually 2450 x 1634 pixels, depending on crop. *3D Red / Cyan Jpeg* (.jpg), *Stereo Jpeg* (.jpg) are categorised as 3D formats, inheriting the same export options as 2D JPEG-Files.

The Lightfield formats include 3 different types of export. First of which being the *Lightfield Camera Raw Image* (.lfr) which is appropriate as a backup format, with approx. 50MB each, since it is containing all light field data captured by the Lytro camera. The *XRAW* (.lfx) format which can only be captured b the Lytro Illum contains all light field data as well as all the camera calibration data, approx. file size of about 100MB each. The most flexible but also most space consuming, approx. 150MB, export format is the *Editable living picture* which exports a TIFF stack, containing 7 .tif images, and a stack lfp (stack.lfp) file. These TIFF Images can be edited, edits are necessary thorughout the TIFF Stack, in any 3rd part editing software and then be reimported into Lytro Desktop.

The *Editable Depth Map* (.png) export, as covered before, is a format to export the unique depth map of an image. The export itseld will output 3 files, the depth map as a PNG, the image with extended focus (.tiff) and a .txt file containing data information for the depth map.[467]

Animation Export offers the posibility to export an animated light field as a H.264 compressed movie or as a series of PNG images. Common frame ranges, Render Quality and 720p/1080p image sizes can be chosen.

[467] Lytro - Desktop 4 – Exporting (2014)

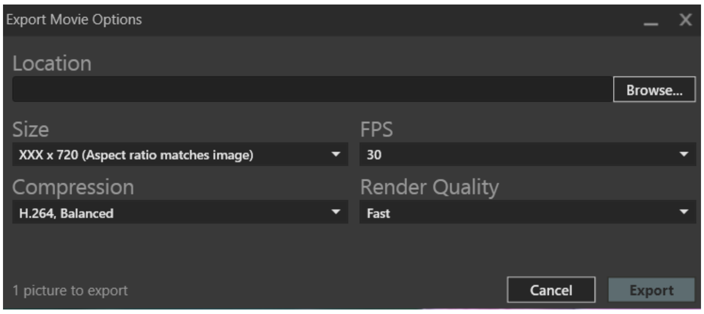


Figure 75: Animation Export options.

Lytro Desktop 4 also offers the possibility to upload its living pictures to its own hosting service as well as upload to 500px and Facebook, both also supporting the lytro uploads and user interaction.

8.1.3 Fraunhofer IIS Light Field ViewRenderer

In 2014 The **Fraunhofer Institute of Intergrated Circuits (IIS)** demonstrated their first own light field camera system at NAB Show in Las Vegas with a proprietary software called *LightField View Renderer* capable of handling the down sampled light field data from the multi camera array in real-time. The standalone software allows change in perspective, refocusing, synthetic aperture and 3D output.

Change in perspective or *View Navigation* is done by dragging the circle seen in Figure 76, representing the view position, in the designated area. The possibility of moving freely in X-/Y-Direction within the boundaries of the used camera array is hereby given. (Figure 77)[468]

Figure 76: View navigation interface.

Figure 77: Different viewing positions of the same camera.

In Addition to View Rendering it is possible to change the Focus for preview purposes by clicking at a desired point within the image or by moving the slider (Figure 76) to the desired point. If desired, the focus-effect can be further improved by changing the synthetic aperture and focus amount within the advanced settings to create a shallower/deeper depth of field (Figure 78).[469]

The System can also be used to output data for 3D stereo viewing with a flexible baseline as well as multiview output for autostereoscopic viewing devices (Figure 78 – stereo rendering).[470]

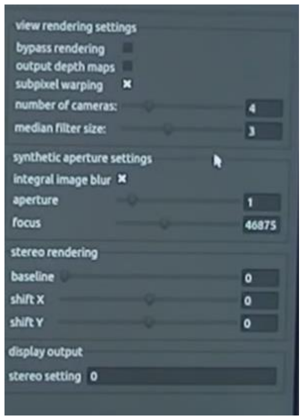


Figure 78: view rendering settings, synthetic aperture settings and stereo rendering settings.

[468] cf. InsightMediaTV (2014)
[469] cf. InsightMediaTV (2014)
[470] cf. InsightMediaTV (2014)

The Fraunhofer IIS continued working on their ViewRenderer and showed the next step in development later in 2014, a Light Field Plugin for the NLE Software Avid Media Composer. The basic functionality of their prior software was kept but modified in a way to animate it with help of keyframes throughout a sequence. The User Interface was changed dramatically in order to be aligned with avid media composer, now consisting of the view-renderer-settings, the initial camera views from the capturing device and a preview of the rendered view with applied effects.

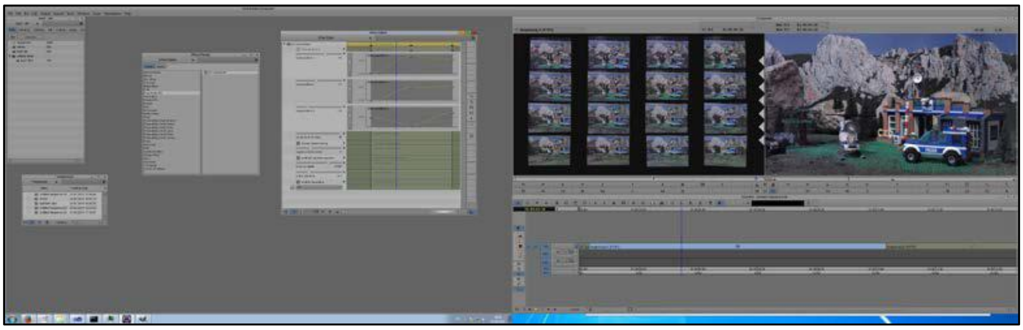


Figure 79: Screenshot of the Fraunhofer IIS ViewRenderer Plugin for Avid, showing the effect editor (left), multiview – displaying the initial cameras and a preview rendering with the adjusted settings.

In addition to the previous version, view rendering in Z-Position is now also possible, as well as Digital Zoom with realistic parallax. Export options include the output of additional disparity for the complying image.

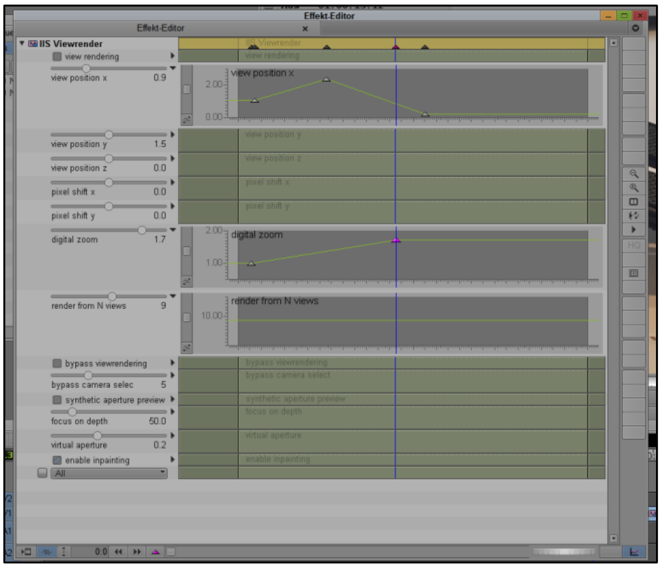


Figure 80: Screenshot of the Fraunhofer IIS ViewRenderer Plugin for Avid, showing the effect editor in detail with a few set keyframes (triangles).

8.1.4 Conclusion

With Raytrix Software only able to view but not edit light field data, Lytros Software only having the ability of still-image-refocussing and view-rendering, Fraunhofer IIS's plugin for avid supporting moving images, view rendering, refocussing, etc. but each only with a unique file format supported by them and therefore not being able to be used by anyone else or any third party data. In terms of refining data during the postproduction, Lytro offers the possibility to easily export, edit and reimport its depth maps while the Fraunhofer Plugin doesn't allow for easy refinement while working on a sequence. All in all current editing capabilities are limited to changing focus, perspective and applying filters. With the possibility of exporting of Lytro's *Editable living picture* which exports a TIFF stack it is possible to edit light fields in third party image editing software, but not in a very convenient way and not at all suited for moving image editing.

8.2 Editing Light Fields

In general the tools for two-dimensional devices can be divided into two categories, first of which is the option to edit light fields according to view-points of a virtual-camera called *multiview*, depending on parallax information between the views. Second of which is the use of depth reconstruction of a scene to create effects like Depth-of-Field and to isolate specific layers of an image. From said depth reconstruction it is also possible to derive other 3D-Reconstructions like Normal Data and Position data, which as well can be used to edit images in a realistic way.

The combination of both leads to a flexible way of editing light fields, using the additional data provided for editing a certain view while being able to propagate the changes to any possible view automatically.

To fulfil the multiview editing aspect propagation of the edit is necessary for all viewpoints. Different approaches for automatic or semi-automatic propagation of edits throughout all viewpoints of a light field are suggested by Jarabo et al., Hasinoff et al., Seitz and Kutulakos, Wang et al. and Zhang et al. among many others.

Semi-Automatic Systems requiring user intervention might be of use for still image light field editing, but for moving pictures the amount of work necessary to indicate correspondences between multiple coherent views is not justifiable. Therefor the Approach of manually positioning polygons in several views, proposed by Wang et al. and Zhang et al., in order to morph between (but also applicable for other edits) light fields is of little interest for our work.[471] Hasinoff et al.'s approach originally designed for editing among multiple Photos within a Personal Photo Collection seems irrelevant concerning the topic, but the proposed system inspired by search-and replace editing for text and computer vision for image matching, is built to propagate edits automatically among different images by matching the edited regions across images.[472]

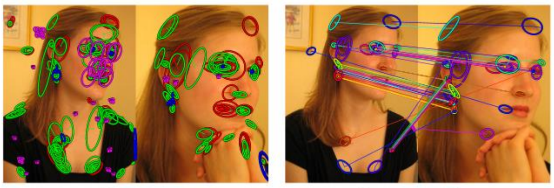


Figure 81: Feature detection and feature matching as proposed by Hasinoff et al..

The interest for light field editing becomes immediate when thinking about redundant features of images captured by a light field acquisition system. Feature matching algorithms known from computer vision and stitching software could be of use when the necessity of propagating edits among a whole image set of a light fields viewpoints is necessary.

Seitz and Kutulakos' present another attempt of image editing among multiple images of a 3D Scene by altering "[…] a scene's plenoptic function

Figure 82: Screenshots from an interactive plenoptic painting tool proposed by Seitz and Kutulakos (2002).

[471] cf. Zhang et al. (2002) & Wang et al. (2005)

[472] cf. Hasinoff et al. (2010)

in a physically-consistent way, thereby affecting scene appearance from all viewpoints simultaneously."[473] In order to do so a plenoptic reconstruction method, called *plenoptic decomposition,* decomposing the plenoptic function into radiance and shape components is applied, giving the possibility of transforming any 2D image edit within a plenoptic image into an operation in 3D and therefor giving the opportunity of efficiently propagating edits among different views.[474] Jarabo et al.'s approach also relies on editing the 3D representation of a light field by propagating sparse user edits in the full light field. They also propose a downsampling technique to handle the large amount of data before editing and finally upsampling back to the full resolution when an edit is final.[475]

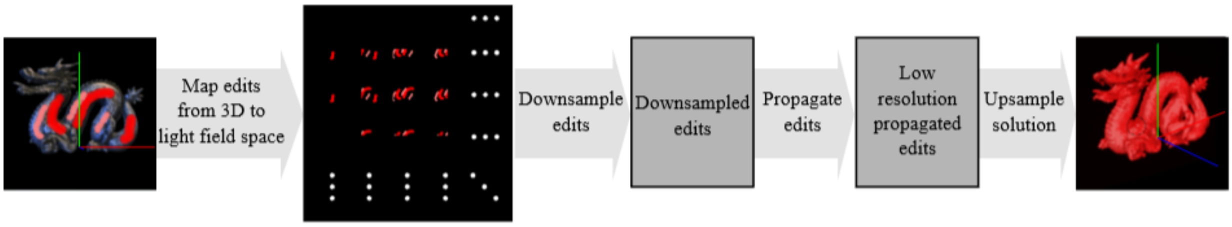


Figure 83: Interactive editing pipeline proposed by Jarabo et al.: Over the 3D representation of the light field, the user can draw sparse strokes. To propagate them, edits have to be first mapped from screen space to light field and then downsampled. Once downsampled, they are propagated in the downsampled domain. Finally the solution is upsampled.

[473] Seitz & Kutulakos (2002) 1
[474] cf. Seitz & Kutulakos (2002) 10
[475] cf. Jarabo, Masia and Gutierrez (2011)

8.2.1 Software for Advanced Light Field Editing

Jarabo et al. conducted a study to evaluate different light field editing interfaces, tools and workflows from a user perspective. During their evaluation they collected lots of data on how users behave when editing light fields and which methods they prefer. During their study they also created an interface in order to evaluate the user's behavior.[476] The study itself is indeed of interest but not topic of this chapter, this chapter will focus solely on the user interface provided by Jarabo et al. during their evaluation. The proposed interface incorporates the two main paradigms for light field editing, *multiview-* and *focus-based editing*.[477] Due to the fact that they are using synthetic light fields and light fields captured by plenoptic cameras a feature matching algorithm or other algorithms to propagate edits to multiple views are not necessary due to the rectified nature of the microlens light fields and the known positions of each corresponding pixel within the multiple views.

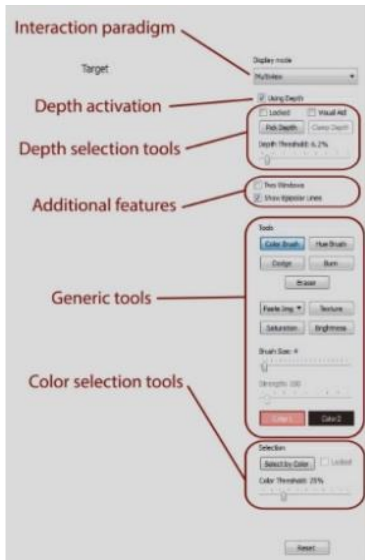


Figure 84: User Interface developed by Jarabo, Masiaand Gutierrez (2011).

The User interface is divided into three categories: *Display mode*, *Tools* and *Selection* (see Figure 84). The latter being of lowest interest. *Display mode* consists of the opportunity to choose between Multiview- and Focus-mode and additional depth-use "Using Depth" for both modes as well as settings for said addition. *Jarabo, Masiaand Gutierrez (2011)*

The *Tools* panel is offering different options for painting and image manipulation while the *Selection* panel offers to select certain image areas within a defined color threshold.

The two display modes can again be divided into 2 different modes each. Therefor possible editing modes are: Focus, Focus with depth, Multiview and Multiview with depth.

8.2.1.1 Focus Mode

In Focus mode the image plane which is supposed to be edited has to be put in focus by scrolling the mouse wheel to adjust the focal-plane. After setting the in focus objects the user can paint on the whole image in the previous adjusted depth layer. As soon as the user refocuses the image the painted strokes will also be out of focus, according to the depth layer they were painted on. Since the depth-values of the other views are the same the paint stroke is at the same XYZ-position in every view.

[476] cf. Jarabo et al. (2014) 1

[477] cf. Jarabo et al. (2014) 2

Figure 85: From left to right: illustrating the same depth for k views, paint stroke in front in focus/back defocussed, paint stroke in back in focus/front defocussed.

8.2.1.2 Focus with depth Mode

Focus with depth actually follows the exact same principle as *Focus* mentioned above with the difference that the in focus plane does not have to set manually. The "using depth" feature allows for the area below the mouse always to be in focus and automatically adjusted. Basically the user is always painting on the in focus region below the mouse.

Figure 86: Left: Area below the mouse pointer is in focus, painted stroke on this "active" depth is also in focus. Right: previous paint stroke out of focus, because mouse pointer is now on foreground depth layer.

8.2.1.3 Multiview Mode

In *Multiview* editing mode a stroke is painted in one view (view 1 in Figure 87) and the correct depth position has to be adjusted afterwards by adjusting the stroke in depth along epipolar lines in different views (view 2 or view k in Figure 87).

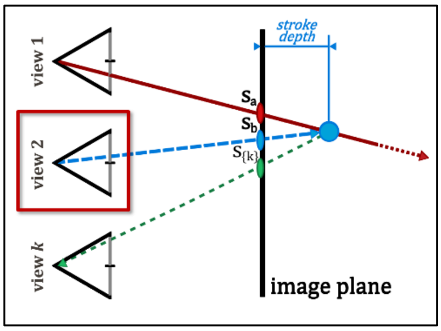


Figure 87: In Multiview the user first selects one view of the light field, and draws a stroke. (left image in Figure 88), this stroke can be anywhere along that red line drawn(in this illustration). To fix its actual position, the user needs to shift to another view (middle image in Figure 88), and adjust the depth of the stroke based on the observed parallax(right image in Figure 88). When satisified, the user fixes the position of the stroke, which can then be correctly observed on all other views.

Figure 88: Left: Paint stroke on initial view (view1 in Figure 82), Middle: Paint stroke on wrong depth in new view (view2 in Figure 82, blue dot would be at wrong position), Right: Corrected depth of paint stroke along epipolar lines (view2, blue dot at right position).

8.2.1.4 Multiview with Depth Mode

Same feature as Multiview but with the possibility of using depth. The edit hereby automatically snaps to the depth at the mouse cursors position for all views.

8.2.1.5 Other Features

Depending on the depth layer, painted strokes on the edits are realistically occluded by image-features in a nearer depth-plane (seen in Figure 89).

Figure 89: Paint stroke at the depth of the wall, depending on viewpoint the stroke gets occluded by foreground objects.

Visual Aid can be activated to view the active depth range when using depth information, the Treshhold of the depth can be easily adjusted with the depth treshhold slider. When using automatic depth selection the user can also lock the depth to a desired layer by checking *Locked* and picking the depth of the desired image feature.

8.3 Light Field Toolset Proposition

As we cannot cover the proposition of a complete toolset necessary for editing light fields we will provide an overview of necessary tools and a proposition of its key functionality. We will not cover algorithms necessary to be implemented in order to achieve said functionalities, only the models and ideas for implementation. The mentioned tools are all fictive and are meant to describe possible models for functionality, requirements in present tense and micro workflows in the context of the postproduction incorporating the three applications of light field data inside the three scenarios. Due to Nuke's node based structure iterative processes are possible and even welcome in order to modify and enhance data, such as depth maps, during the compositing and directly providing the modified data for other nodes. In a final proposition, the order of the tools should derive from the imaging process of a camera and traditional processing.

8.3.1 LF_Read

8.3.1.1 Basic Workflow and Requirements

Similar to The Foundry's *Read* node the purpose of this tool is to read images from a storage device, using their native resolution and sequence frame range. The difference to the standard read node is the necessity to load whole light field data sets of various kinds instead of just a traditional image sequences.

The regular Read node is able to handle light field data depending on the delivery format, but a lot of manual work has to be done by the user to set up all the read nodes necessary to load a whole data set of a light field sequence (Figure 90).

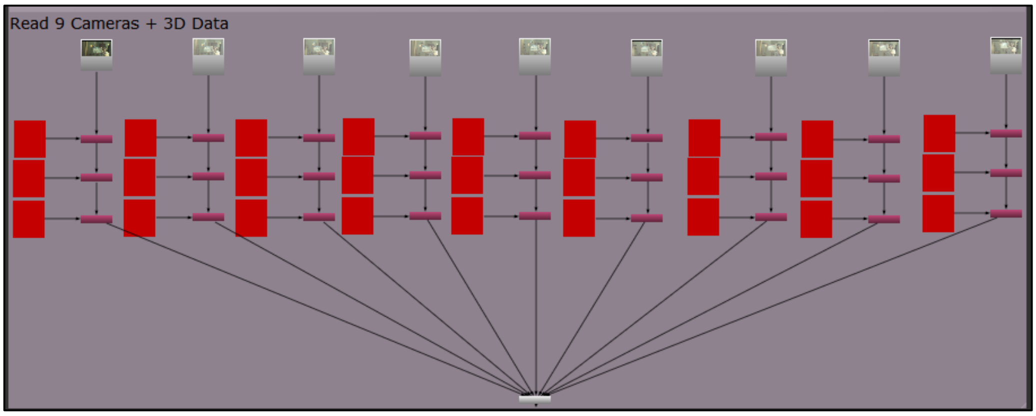


Figure 90: Example from our test shoot, 36 read nodes are necessary to load the data from the camera array consisting of 9 cameras, including normal,position and depth data.

Similar to the regular read node all imported sequences, regardless of format, need to be converted to Nuke's native 32-bit linear RGB colorspace.

The user is able to handle light field data intuitively and able to choose what kind of data format needs to be handled. The light field read node is also able to analyze metadata from the image files to determine the file format and necessary user interface-settings. The light field read node supports multiple types of light field data, as well as regular image file formats in order to maximize compatibility. Similar requirements exist in basic functions but need to be extended in multiple ways in order to be able to work with different types of light field data.

8.3.1.2 Possible Models for Implementation

The tools core functionality should be the correct setup of the necessary files for light field compositing according to camera type, camera arrangement, additional layers and a pass through of metadata for succeeding nodes. Factors such as the type of data i.e. multiple camera views in a single image sequence, multiple camera views in a single image sequence including additional 3D Data, multiple camera views in individual image sequences, multiple camera views in individual image sequences including additional 3D Data, microlens light fields etc..

A completely manual method for setting up the correct handling of data is also necessary if the metadata is corrupted or missing. The tools functionality should be given for different types of light field data, i.e. data from multi camera array recording systems, as well as microlens recording systems.

A provided functionality to convert the loaded pictures to a common intermediate light field format would also be of interest in order to ease up the process of loading light field data from there on.

8.3.1.3 User Interface

The user interface is kept as simple as possible. A first basic proposition was used during the postproduction of our test shoot, relying on Nuke's integrated Read-, Write- and Channel-Nodes. The main goal hereby was to decrease the user's amount of work necessary to import all the image sequences from a multiple camera array. The custom light field handler needed a single read node from one of the array cameras in order to check for related cameras and then creating all the read nodes for all other cameras within the specific folder-structure. The possibility of fetching all the cameras and joining them with a *JoinViews* node was given, as well as the option to sort the cameras in layers of a single view stream. The option of adding 3D data by shuffling it into the respective camera stream was also given. The user interface is shown in Figure 91.

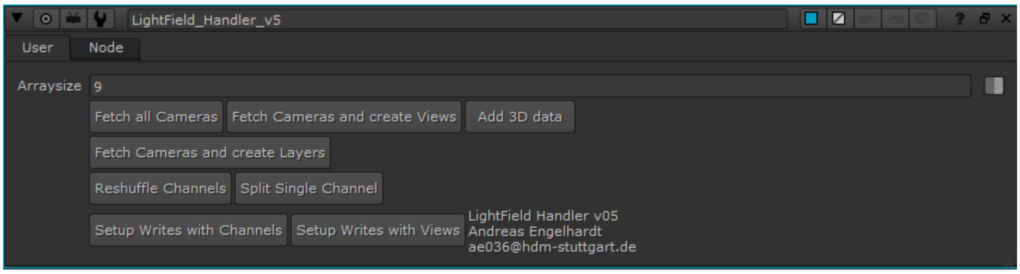


Figure 91: The LightField_Handler tool developed during our postproduction, able to semiautomatically load related cameras within a specific folder structure.

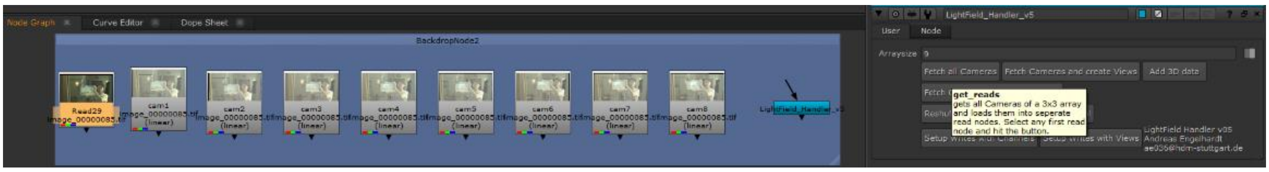


Figure 92: A single cameraview was imported manually with a Read node, the LightField_Handler created all related camera views.

Since the tool was very limited in functionality and developed for our very specific data-structure and camera system it is only mentioned in context.

The actual tool needed to import light field data in Nuke's Compositing Environment is more flexible and doesn't rely on the basic functionalities of the common Read Node. In fact it supports the workflow right from the beginning by automating as much of the import process as possible while still giving the user the option to interfere and change any setting manually.

The first step of importing data is to specify the type of data to be imported. By default the selection of the data format is *microlens array*, being the most common technique to capture light fields, followed by *multi camera array*. The following step is to define the location of the data. The tool then analyzes the provided meta data, proposes the correct settings and loads related data. If the meta data is missing the user gets the alternative to enter all the settings manually.

The settings being:

- Choose camera preset (automatically fills many of the setting-parameters)
- Filepath for each camera (only for multi camera arrays)
- Format
- Proxy Filepath & Format
- Specify paths for 3D Data (if not embedded in original image data)
- Filepath for calibration file (especially for microlens systems)
- Frame Range
- Rectified yes/no (only for multi camera arrays)
- Array setup (amount and arrangement of cameras) (only for multi camera arrays) – possibility to enter desired matrix size and order. H/V camera distance

Figure 93: Proposition for flexible array setups and variable camera-numbering.

- Principal View (on multi camera arrays)
- Array type: having custom subsettings for each type of camera identical cameras, multiple focus array, monolithic array, etc.
- Colorspace and option for raw data

8.3.2 View Rendering

8.3.2.1 Basic Workflow and Requirements

Since View Rendering is one of the main features of light field technology it has been used many times before and implemented in various ways. Software using View Rendering was described in chapter 8.1. The goal of this tool is a solution to quickly and easily preview various perspectives within the captured light field, while providing enough resolution to assess the image. Depending on the desired workflow the tool could be used in two ways. First being the use of View Rendering at the beginning of a compositing to determine a camera perspective or animated camera perspective for further post production with the option to only edit the chosen view in the further process and going back to traditional 2D compositing, after deciding on a view. Second being the option to work with a light field during the whole compositing process and being able to interact with the viewpoint at any step during the compositing process including the ability to view all previous edits (propagated to all views) from any perspective within the light field.

In terms of matching the view animation of a light field to a recorded scene, the option to exactly match camera movement somehow needs to be established. This could be of great interest for use cases like the described backlot scenario in chapter 4.1.

Several performance enhancing, respectively quality diminishing, options are also be provided in terms of the tools operating speed and quick preview capability. For multi camera array captured images, bypassing of view rendering is thinkable in order to speed up the process, making only the captured views viewable, but relinquish rendering new virtual views and interpolating intermediate images.

Depth of field simulation is provided as well, but only as a preview tool to decide on an arbitrary view with the final look in mind. Basic depth based DoF simulation with a fixed focal range and adjustable focal plane is sufficient for this use case, any further DoF simulation is handled within the suggested discrete DoF tool.

8.3.2.2 Possible Models for Implementation

The most essential part is, that the tool needs to incorporate algorithms to interpolate in-between the initially provided views, combine and resample the available images or redisplay previously rendered or digitized views in order to freely set the desired view within the light field. Basically a dense light field has to be generated from the recorded sparse light field, either up front or as close to real time as possible. Ng, Hanrahan, Levoy, Gortler, Zilly et al. provide solutions for virtual camera views, often called view rendering or light field rendering. Zilly et al.'s approach relying on Kauff et al.'s approach described in "Depth map creation and image-based rendering of advanced 3DTV services providing interoperability and scalability" (2007) uses Depth Image Based Rendering to interpolate additional intermediate views. Levoy and Hanrahan describe a solution for generating new views by combining and resampling the available images in their 1996 paper "Light Field Rendering", while Gortler et al. draw back on raytracing algorithms, known from computer vision, in their 1996 paper "The Lumigraph", to generate new views. Zilly et al.'s approach might be of the biggest interest concerning sparse light field acquisition with a multi camera array while the others are of interest for relatively dense light fields, for example captured by plenoptic cameras.

A basic DoF simulation algorithm relying on depth data, as mentioned in chapter 8.3.1 should also be integrated in the tool in order to quickly preview a DoF simulation for a certain view.

8.3.2.3 User Interface

The User Interface for the view rendering tool is as intuitive as possible, but still giving the possibility to define exact values for each action. In order to set the camera position on the X/Y plane an input device similar to a trackpad or the proposed *View navigation* interface by Fraunhofer IIS in subsection 8.1.3 Figure 76 are implemented. In addition a slider for each direction (X and Y) is provided with the option to manually enter the desired amount of movement from the center view. The Z-Position is also changeable with a slider or by entering a value. The virtual camera position can also be controlled by a previously tracked camera or any desired input camera in order to match the movement within different scenes (backlot scenario). Options for digitally zooming inside an image are provided, maybe even with the possibility of correct parallax change depending on the zoom factor, as well as a traditional digital zoom which enlarges an image. While enlarging the image, super resolution approaches, like Bishop et al.'s and Georgiev & Lumsdaine's, might be of great interest in order to retain as much quality as possible. All of the above mentioned functionalities contain the option of keyframe animation in order to simulate camera movement within a scene.

Several options to enhance the tool's performance are provided in a drop down menu, one of them giving the possibility to bypass the rendering of intermediate views completely, another by using lower quality proxies to preview a generated arbitrary view.

Previewing DoF simulation can optionally be implemented as an extra tab within the tool, with a simple depth based approach mentioned in subsection 8.3.4.

For downstream performance it is possible to select whether the user desires to continue working with the whole light field or limit the view-possibilities to only a certain amount of views, for example stereo editing, horizontal/vertical multiview applications or just a single 2D image. This constraint might be useful if the user knows his final delivery format and spares to carry the whole light field through the compositing pipeline. For stereo and multiview output, settings such as flexible baseline and shift are possible. A limitation of the light field's zone of view, for certain applications like Virtual Reality or autostereoscopic displays, might also be of interest to reduce the amount of data.

8.3.1 LF_DepthOfField

8.3.1.1 Basic Workflow and Requirements

Even though the possibility of a focus preview is given in the proposed view-rendering node, a special node with more preferences is necessary for a final result, giving the possibility to edit and refine the settings necessary for calculating a high quality depth of field and provide good view outputs for setting up the DoF simulation.

8.3.1.2 Possible Models for Implementation

In order of applying synthetic depth of field to an image, different approaches are of interest. The simplest, fastest and most used in compositing is the use of depth data to simulate depth of field effects, to be specific blurring the image according to depth values. Sophisticated solutions for such single view methods of depth based DoF simulations already exist. *ZDefocus*[478] is a tool already integrated in Nuke's Environment and *Bokeh*[479] is a Nuke Plugin developed by Peregrine Labs offering even more options to control and simulate the DoF. Therefor it would not be necessary to develop new User interfaces regarding depth based DoF simulation. The major drawback of these single view simulation methods are visual artefacts like depth incontinuity and edge bleeding artefacts. As a resolution sophisticated techniques to reduce such artefacts have been introduced by Kass et al. (2006), Zhou et al. (2007), Lee et al. (2009) among others.[480]

The more advanced model for implementing Synthetic DoF is a multiview approach originally developed in order to get realistic DoF in computer generated imagery. Hereby from a single rasterized view and its

[478] cf. The Foundry (ZDefocus)
[479] http://peregrinelabs.com/bokeh/
[480] for further reading: Kass et al. (2006), Zhou et al. (2007), Lee et al. (2009)

depthmap a light field is dynamically generated and used to synthesize DoF effects.[481] Due to the fact of having a real light field available for DoF simulation, the necessity to dynamically generate a light field is no longer required. DoF effects can be simulated by averaging multiple images and applying *accumulation buffering*[482] techniques known from computer vision.

This approach is similar to *synthetic aperture focussing* and *dynamically reparametrizing light fields*.

Vaish et al. describe the process of synthetic DoF mentioned above as follows:

> *"Synthetic aperture focusing consists of warping and adding together the images in a 4D light field so that objects lying on a specified surface are aligned and thus in focus, while objects lying off this surface are misaligned and hence blurred.[…]"*[483]

> *"Synthetic aperture focusing (also called dynamically reparametrized light fields) is a technique for simulating the defocus blur of a large aperture lens using multiple images of a scene, such as from a light field. The process consists of acquiring images of a scene from different viewpoints, projecting them on to a desired focal surface, and computing their average. In the resulting image, points that lie on the focal surface are aligned and appear sharp, whereas points off this surface are blurred out due to parallax. Researchers in computer vision and graphics have used synthetic aperture focusing [...]"*[484]

The amount of defocus hereby depends on the size of the synthetic aperture, succeeding on the deviation in viewing angle covered by the light field's images. The smaller the aperture, the less deviation in views, followed by a slighter amount of defocus. Depending on the density of a light field, sub image interpolation is needed in order to obtain a smooth blur effect, as seen in figure 94. In order for good sub image interpolation, needed to minimize missing image areas due to previous perspective occlusions, accurate disparity information is necessary to allow in-painting of areas without image information.

Figure 94: Left: sparsely sampled light field. Right: additional interpolated sub-images and smoother blur.

[481] cf. Yu et al. (2011)

[482] Accumulation Buffer: An extra image buffer to accumulate composite images. Useful for effects like antialiasing, DoF and motion blur.

[483] Vaish et al. (2005) 1

[484] Vaish et al. (2005) 1

In conclusion the latter mentioned model is more realistic in physical terms but also takes longer to calculate because a lot of data is necessary to compute the final result. In general the simulated depth of field should be applied not only to the RGB channel, but also to other channels, as for example alpha, depth and normal channels, in order to have matching layers to the RGB image in the further compositing process.

8.3.1.3 User Interface

The User Interface for DoF Simulation based on depth data does not have to be modified by a large amount, given the circumstances of the availability of sophisticated tools.

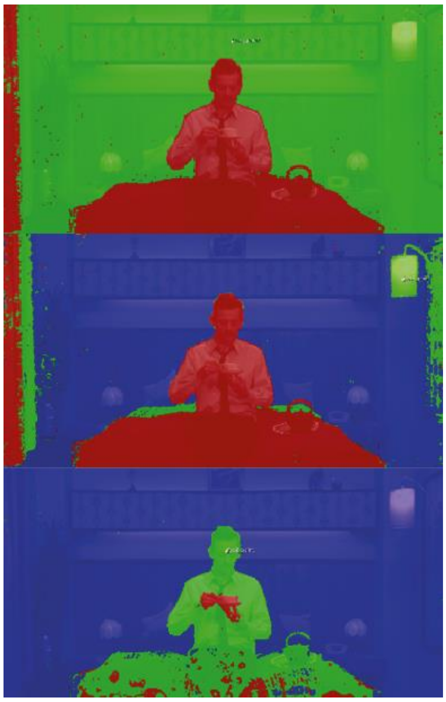


Figure 95: ZDefocus' focal plane setup. Top: Background in focus. Middle: Lamp and middleground in focus. Bottom: Actor in focus.

ZDefocus is a tool for DoF Simulation already integrated in The Foundry's Nuke offering a good solution to simulate depth of field blur with the use of depth data. The user interface is shown in Figure 97. To simulate DoF, ZDefocus splits the image into layers of different depth giving each of the layers an adequate blur size. The amount of layers and spacing between the layers can be specified, the less layers the faster is the process of DoF calculation, the more layer the smoother is the gradation. Automatic Layer spacing is also available, figuring out automatically how many depth layers to use based on the defined maximum blur size. After processing all the layers, they are blended together to preview the applied DoF result. The DoF simulation is by default applied to all layers, but can also be limited to specific layers, if desired. By default the standard output is the *result*, showing the applied changes of the ZDefocus node. In order to set up the desired in focus and out of focus areas of an image the *focal plane setup* and *layer setup* output can be of great help, offering the possibility to exactly identify which regions are in front of the focus (red), completely in focus (green), and behind the area of focus (blue), shown in Figure 95. To adjust a focal plane a user can either manually set the focal plane by changing the value of *focus plane (C)* or by moving the focal point widget inside the viewer window specifying the point of interest and picking the according depth data automatically from the depth channel. The *depth of field* slider defines the amount of in-focus area on the image while *size* defines the amount of blur at infinite depth value.[485] For further reading and more options see [The Foundry "Nuke Online Help – Zdefocus"].

Figure 96: Tool to pick depth according to area of interest, in this case: The actor's eye.

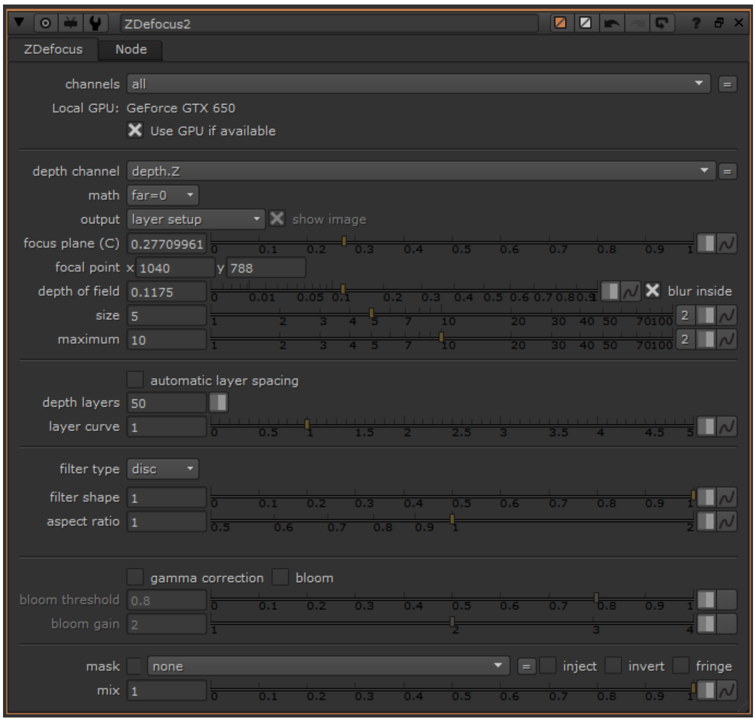


Figure 97: Nuke's ZDefocus user interface.

[485] cf. The Foundry (ZDefocus)

Peregrine Labs' Plugin *Bokeh* offer further possibilities compared to *ZDefocus*, as well as support for Deep Data and an optional 3D camera, overriding the focal plane, etc. of the Bokeh node. Setting up the node is quite similar to ZDefocus, also providing an output to preview the focal setup, in this case called *Focal Distance Visualization*. Focal Distance Visualization is similar to Nuke's *layer setup* output, but the color coding is different (red - in focus, blue - far, green – near). Bokeh also offers the option to output only the region in front or respectively in the back of the focused area, as well as adjust them differently. In terms of user interaction, focal plane and size setup are similar to ZDefocus. The real difference over ZDefocus comes with the *Lens* tab in Bokeh (Figure 100), offering a solution to simulate and match the DoF to physical lens characterstics. The *Real World Lens Simulation* takes Focal Length, Aperture, World Scale, World Scale Multiplier and the Film Format into account when simulating a physical lens. If the desired Film Format is not provided, a custom format can be configured as well. The options to add optical artefacts (Figure 101) and correct rendering artefacts (Corrective Slices tab) are given as well as the Kernel tab, providing a setup similar to ZDefocus' filter setup.

Figure 98: Bokeh's Focal Distance Visualization.

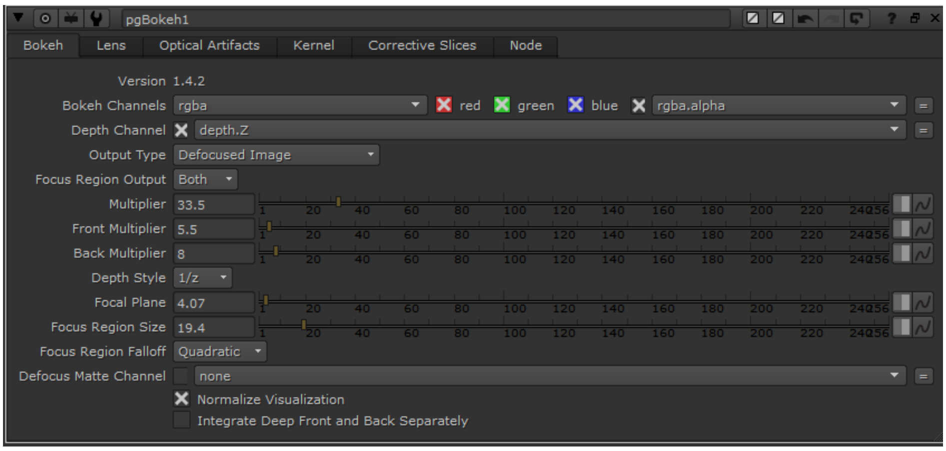


Figure 99: Basic DoF settings in Peregrine Labs Bokeh tool for Nuke.

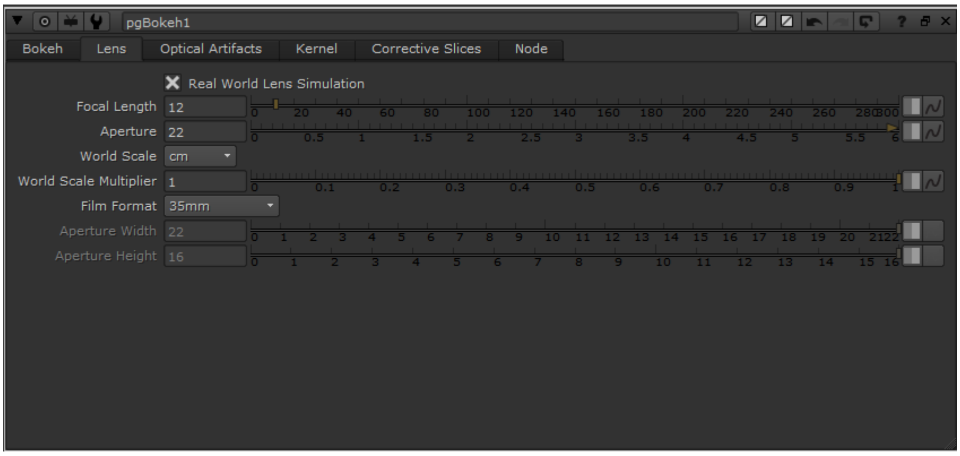


Figure 100: Lens settings in Peregrine Labs Bokeh tool for Nuke.

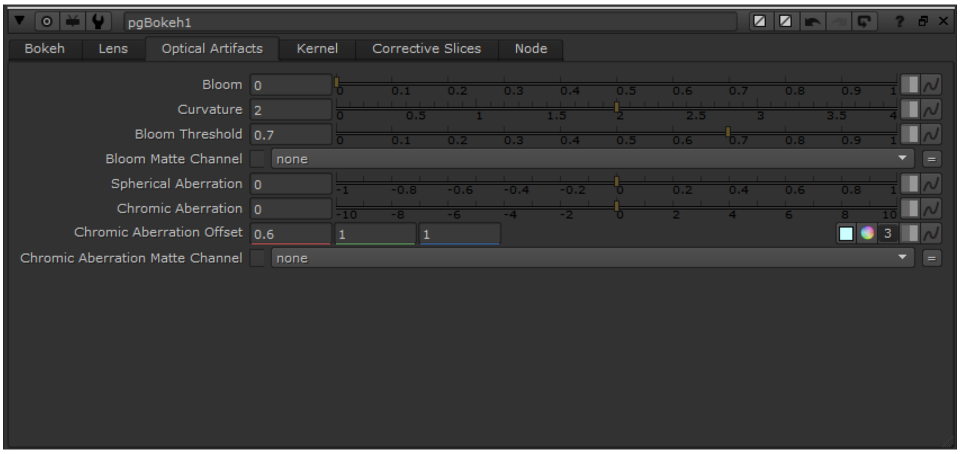


Figure 101: Optical Artifacts settings in Peregrine Labs Bokeh tool for Nuke.

For a physically more correct DoF implementation using a synthetic aperture approach described previously, the tool provides a list of camera and lens presets offering the option to simulate desired looks. The choice of a manual DoF setup is also given, with parameters to set the virtual aperture, focal point, film format/sensor size and the possibility to choose an arbitrary lens shape (i.e. round, bladed, anamorphotic, etc.). A focal plane setup similar to the ones proposed within depth based DoF tools is also provided.

8.3.2 LF_Roto(Paint)

8.3.2.1 Basic Workflow and Requirements

Manipulating images is one of the main tasks in VFX, often shapes and strokes are needed for simple, as well as complex tasks. Rig removal, rotoscoping, (garbage) matting, dustbusting, visual refinement and others are essential parts of the compositing process and require a lot of work. In light field compositing the requirements are the same as in traditional 2D compositing. To reduce the amount of work, rotoscoping and paint work needs to be automatically propagated to all views within a light field. The functionality of Nuke's Roto and RotoPaint Node is provided within the tool proposed here. No extra functionality, besides automatic propagation to associated views and limitation of an edit to certain depth areas is necessary.

8.3.2.2 Possible Models for Implementation

To implement the feature of automatic view propagation and depth editing, suggestions were mentioned in section 8.2 by Hasinoff et al. and subsection 8.2.1 by Jarabo et al.. Hasinoff et al.'s approach relies on feature matching between images while Jarabo et al.'s solution is to tie the edit to a specific depth layer or correctly propagate it to different views with the help of epipolar lines. The correct positioning of an edit among several views with the help of epipolar lines is only shown in a still image in subsubsection 8.2.1.3, but since the multi camera array's arrangement, respectively the microlens layer layout, won't change during a sequence the epipolar line setup has to be done once.

Traditional tracking and matchmoving techniques provide the correct positioning of an edit within a sequence, just as in traditional compositing. Feature detection algorithms as proposed by Hasinoff et al. could also position edits among a sequence of images, searching for similarities within images of a sequence. But since it was not necessary in the past, a 3D matchmove or point/planar tracking should suffice and offer a more performant option due to not making it necessary to search for matching features within multiple perspectives of a light field scene.

8.3.2.3 User Interface

The user interface is oriented on Nuke's existing Roto(Paint) node (Figure 102) with the extension of using depth and locking image edits to certain depth areas, similar to the functionalities proposed inhapter 8.2.

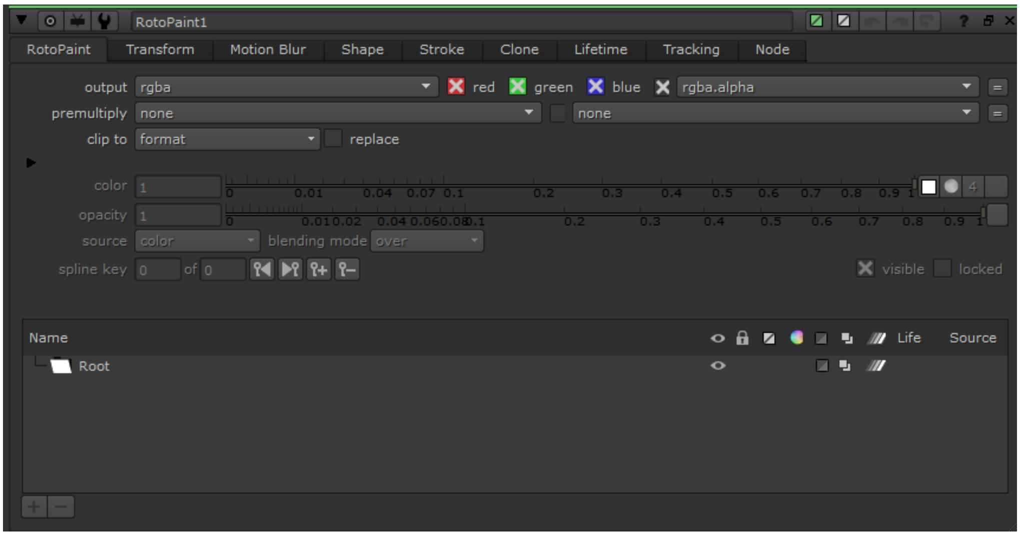


Figure 102: User Interface of Nuke's RotoPaint tool.

8.3.3 LF_DepthSampler

8.3.3.1 Tool Description and Requirements

As required in 6.4.3.3 "Quality of depth maps" this tool addresses the correction and improvement of 3D data that might has been generated earlier in the pipeline. The input is a light field data stream that includes depth and other 3D data for one or more views. Otherwise a single monochromatic image can also fed into the piece. The output is the light field data set with updated depth data for all views that are selected or a single monochromatic image with updated depth information for the selected view.

There are two basic functions available. First, the depth data of all views can be improved by taking into account the depth data from surrounding views if that has not already been done as part of the estimation process. And second, the depth data for a selected number of views can be edited by a customizable chain of filters. There is also user input possible to intuitively help the algorithms distinguish between foreground, background and different objects.

8.3.3.2 Implementation Models

The efficient generation and filtering of depth data is a big field of research. A method to refine normal and position data has been presented by Nehab et al., for example. It uses the data of both the normal and position passes to improve each other.[486] Smirnov et al. give a good overview about different filtering techniques presented in the past, also mentioning approaches that take into account color information.[487]

In the case of the depth sampler tool, one part of the functionality would be the improvement of the depth data of a selected view by taking into account the depth data of the surrounding views. This can be accomplished with algorithms similar to the ones used by the Fraunhofer IIS in the case of our test shoot. Based on occlusion information data from other views can be transferred to the depth map using disparity information.

The depth data of the individual views could be improved by a series of filter operations. For example, a stack of blurs can be used to fill holes. This can be combined with intelligent sampling and interpolation of surrounding pixels like it is done in the MultiSample node as part of the Ocula toolset or the Pixel Filler tool as described in subsection 6.3.3. Caselles, for instance, presents a good overview over current approaches to more sophisticated approaches to inpainting. There is a trend to combine texture based, taking into account texture repetition across the whole image, and geometry based inpainting, trying to interpolate local geometric structures like lines and total variation.[488] An option to take user input into consideration would be necessary in this context. This can be realized with function to select values by example from the image preview. Input samples can then be taken into account to constrain the filling and smoothing operations to a certain range of values in certain image areas.

Other filter operations include different kinds of bilateral filtering and other smoothing algorithms that might incorporate pattern recognition or median filters.

When filtering an image sequence it would make sense to take into account the frames before and after, i.e. using algorithms related to optical flow.[489] An option similar to a temporal denoiser like neat video[490] could be also useful.

User input could be based on graph cuts or other ways to create rough masks for objects and and to distinguish between foreground and background pixels.[491] This mask input can also come from the depth data itself. In an iterative loop the depth data can be refined based on the rough depth data already available,

[486] Nehab et al. (2005)
[487] Smirnov et al. (2012) 9-12
[488] Caselles (2011)
[489] cf. Kanumuri et al. (2008), Liwei Guo (2007)
[490] http://www.neatvideo.com/
[491] cf. Wang and Cohen (2008)

similar to the technique we used during the test shoot and described in subsection 6.3.3. Again, the basic idea is to constrain the depth estimation and in this case the depth correction to a range of plausible values.

Finally a function to scale or grade the depth map can be achieved with basic math. Based on input samples or survey data the depth map could be normalized to a standard format as required in 6.4.3.3 "Quality of Depth Maps".

Of course, multiple instances of the tool can be stacked to improve the data iteratively. The tool is aware of the instances around them via metadata streams.

8.3.3.3 User Interface

The user interface consists of drop-down menus to select views for preview or output and to set the operation mode to single image, multi-image or custom.

A graphical representation like a flow graph could be used to control the chain of filter operations. Each of them can be bypassed for preview or for final rendering independently. A control slider can adjust the number of iterations and the quality for each filter operation. A tooltip-type of advice can be used to estimate the setting needed for a certain output quality versus processing time.

Masks can be fed into the tool as Bezier curves or monochromatic images from the host application. A special mode allows the user to paint on one or several views to define regions in depth with different colors. A mode to sketch object contours of the depth data is also available. This can be taken into account for local repairs or to help the algorithm to distinguish between foreground and background.

Finally a mode of interaction is available, where the user can pick a depth value from a previously created depth map and use that as part of the normalization process or to constrain an area of pixels to this value. In the latter case a tolerance value can be adjusted.

8.3.4 LF_Edit3D

8.3.4.1 Basic Workflow and Requirements

The goal of this tool is to provide methods to intuitively transform 3D data like depth and position maps which has been generated earlier in the pipeline. Although normals are also subject of editing processes like in the context of bending them with occlusion data, transferring them from camera to world space or transferring them to a simulated camera model, this tool will only focus on position data. According to the outcome of the test shoot there is a higher demand for the editing of position data than normals at the moment. The user interface and core functionality can be transferred to another tool later.

Input is the data stream containing the light field and position data for one or more views and a matchmoved camera or other camera information in real-world units possibly. Additionally a scene description can be provided to define an origin point and scene scale. This could be used to match the position data as well as the camera to another scene or scale derived from a virtual environment, for example. If neither is provided the tool offers basic transformation, rotation and scale options that can also work non-proportional.

In the long-term it might make sense to have an exact camera model of the cameras used to shoot the light field and already take that into account when estimating depth and consequently normal and position data. Like this, the data could be in real-world units from the start. But that means, that either some kind of measurement or camera tracking has to done on-set or as an additional task for each shot of a production. With todays technology and workflows that won't be practical in the short-term. In the context of VFX a matchmove is frequently done, though.

8.3.4.1 Possible Models for Implementation

The core of the tool will definitely be a series of matrix transformations that can be applied to all views simultaneously. The existing tool, that only operates in the domain of channel expressions has to be extended to be able to import camera data. This data format could be the one of a host application that needs to be converted to an internal representation first or it is extracted from the metadata. Knowing the camera position in scene or world space coordinates as well as the basic camera parameters in terms of focal length, filmback size and image aspect-ratio the position data can be transformed from projection space to scene space. This can be seen as basically the reverse procedure of the image formation process inside a camera (figure 105). In this case the transformation matrix would consist of a projection to view space transformation that scales the position data from the pixel coordinate between zero and one to real-world units in camera or view space, first (figure 103). Knowing the sensor or filmback size of the camera model, the distance between pixels can be calculated to figure out the correct scale. Now the scene is in the desired scale with the camera, more precisely the center of projection, as origin point. To transform from view space to world space a transformation matrix is used that is located where the camera is and is oriented so that the Z-axis is perpendicular to the imaginary film plane. It is the same matrix that is used to put the camera into world-space and usually only consists of rotation and translation (figure 104). All desired operations could be concatenated in one matrix. Of course, it would also be possible to do the opposite - to transform position data from scene space to view space or even projection space to store it in a camera independent format. This transformation process has to be done for each frame of an image sequence. It might be a good idea to encourage the user to do the conversion for the whole sequence at a time and save out a new version of the light field metadata that can be read by the tools further down the pipeline if a workflow approach as described in 7.1 is taken. If the camera information is not provided, the editing could work as a composed transformation matrix combining translation, scale and rotation that is controlled by a locator object in 3D space or numeric input in an intuitive way. Such a locator object could also provide input to a conversion process from one coordinate system to another if it is interpreted as new origin point. If position data originally in projection space is transformed it might make sense to add an option to output an updated camera model as a result of the transformation process. This would be done by aligning the original camera space along the scene camera space, if no specific origin point is provided. User input specifying focal length and sensor size should be taken into account. (figure 105) In this context ways could also be explored to generate animated cameras based on the scene information inside the position and depth data of the light fields. Taking camera specifications like sensor size and focal length as well as coordinates of an origin point as input a moving camera could be calculated assuming a static scene. As with traditional matchmoving methods the moving parts in a scene must be ignored to get correct results.

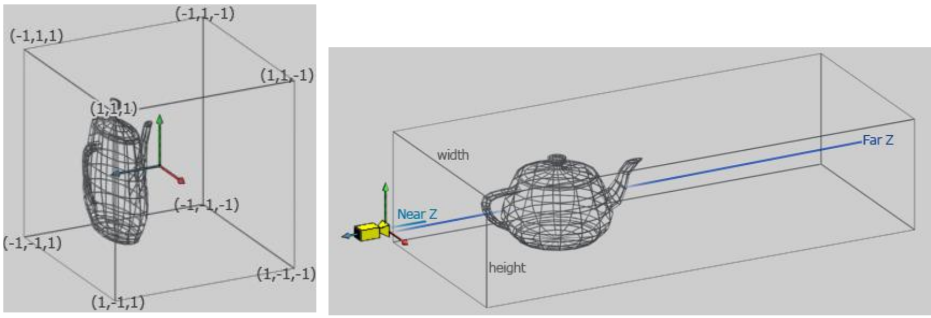


Figure 103: Teapot in the cuboid projection space (left), the scene needs to be scaled according to the camera specifications (right).For simplification, an orthographic camera projection is shown, it can also be accomplished for a perspective projection.

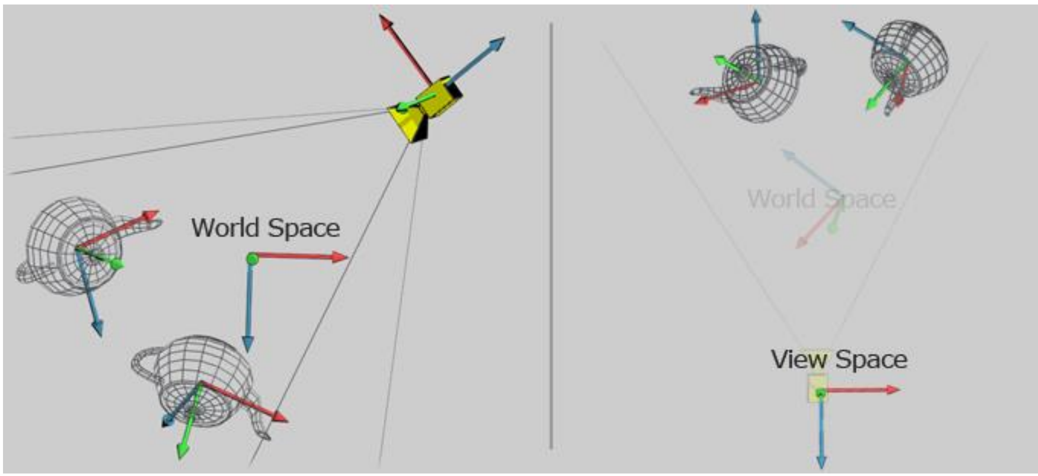


Figure 104: A scene consisting of two teapots seen in camera or view space (right) and in world space (left).

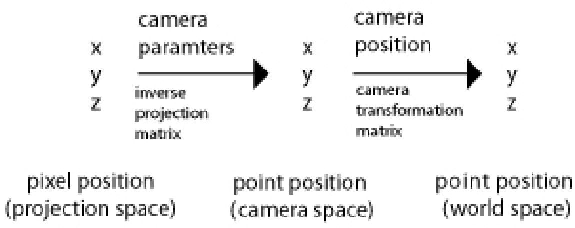


Figure 105: The processing logic of the automatic coordinate transformation.

Of course, all this can be formulated in one compound matrix.

8.3.4.2 User Interface

A starting point for the user interface can be the tool mentioned in 6.3.3 that has been prototyped in the context of the test shoot. It needs to be extended by options to select an existing camera description. There should be a switch to select the source of the camera information if more than one is provided. The transformation and scaling can be done with numerical input or interactively in 3D space using the point cloud plot. To avoid distraction most of the logic should remain invisible for this tool.

8.3.5 LF_Relight

8.3.5.1 Basic Workflow and Requirements

A relighting tool takes a light field or multi-view stream as an input and outputs one or several color images that represent the effect of one or more simulated light sources on the scene. It will be referred to the output as raw light pass in the following chapter. The raw light pass can be further edited before merged with the input data. The tool allows to individually bypass all workflow steps after the relighting operation itself to live up to the demands of a VFX workflow. The system is aware of a 3D coordinate system and ideally takes the light format of the host application as additional input. There are four light types available: point lights, spot lights, directional lights and image based lights. The position and characteristics of the light can be controlled interactively with exact values and graphical representations. A shading model can be selected from a variety of presents that include standard models to approximate diffuse and specular reflections. They can all be edited and customized by adding or removing terms from the equations and changing the constants. The camera data is either extracted from the metadata of the input stream or can be fed into the tool as an additional input in the format of the host application's camera model. If there is no camera model available at all, a planar projection model according to the input light field is fitted. If the light field has been relight at an earlier stage in the workflow the control data can be extracted from the metadata. According to the controls described in 7.4.3 the whole setup can also be transformed and moved into position of a new or existing coordinate system keeping the scale and relations of the scene objects and lights.

There is also an option to approximate shadows taking into account other geometry in the 3D scene as well as self-shadowing and self-masking (figure 106).

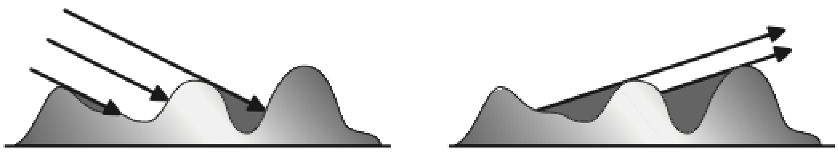


Figure 106: The difference between self-shadowing (left) and self masking (right)

The relighting tool is one of the bigger pieces in the toolset and might be developed as a set of different tools depending on the algorithmic implementation and performance optimization. The relighting tool can operate at different quality levels reaching near real-time performance in the suitable hard- and software environment. The user interface can be simplified for applications during the production stage or in an editing or color grading environment.

8.3.5.1 Possible Models for Implementation

Relighting has been a part of several projects researching ways to edit light fields. Meneveaux and Fournier present a system that determines the geometric shape of the scene by creating an octree of cells occupied by an object, for example. Surface normals are then estimated in object space using boundary points using a method like the one described by Gopi (2002).[492] Finally a surface reflection function is estimated.[493] As mentioned in 3.3.4 this approach profits from capturing a scene in as many different lighting conditions as possible. This approach to generate normals and a BRDF can be assigned to the third class of relighting techniques as defined in 4.2.5.

Horn and Chen used linear combinations of light rays from light fields captured in different lighting conditions in the context of light field editing system LightShop.[494] This editing system provides a language to define ray-shading programs that can be executed efficiently on graphics hardware. Also a light representation in the form of a light field can be used to cast shadows on light fields.[495] But it has to be noted that the LightShop system is primarily a rendering tool to create 2D projections from several 4D light fields that can be combined using the compositing operators introduced by Porter and Duff. The illumination of the input light fields is fixed and can only be changed by traditional compositing techniques, yet in an environment that does support all other light field applications like digital focus and view rendering.

Considering the implementation of relighting of live-action light fields in a production pipeline it makes sense to only change what is necessary at first. It is predicted that this tool is used mostly at the compositing and finishing stage probably. Therefore, an approach to relighting is proposed that is based on the concept of relighting CG at the compositing stage as mentioned in 4.2.3.3. This might increase acceptance and interest on the user side while addressing the requirement of an integrated editing environment for CG and live-action footage.[496] Having tools that can be used for live-action and CG elements alike also broadens the range of potential use-cases and consequently users. Finally it fulfills the requirement of staying in or defining standards as it transfers existing standard procedures to a new data format. Communication and information exchange can resort to existing notation and terminology.

2.5D relighting as in the context of the second class of relighting approaches and defined in 4.2.5 is selected as the basic model for the relighting tool presented in this subsection. This method has been successfully

[492] Meneveaux and Fournier (2002) 3
[493] Meneveaux and Fournier (2002) 1
[494] Horn and Chen (2007) 5
[495] Chen et al. (2006) 5
[496] cf. Ganbar meeting records 6

tested with multi-view data from the test shoot as part of the postproduction of the packshot and portrait scenarios as mentioned in 6.3.3 and 6.3.4.

It is a method based on implicit geometry. More precisely, it assumes a point cloud as geometric representation of the scene. It is likely that depth data will be estimated for one or all views as part of most pipelines as it has several applications in the postproduction process. From there normal and position data either has been generated at the rectification stage or can be calculated before entering the relighting stage. Different algorithmic approaches of different quality exist but won't be evaluated as part of this work.

The position data can be transformed into the scene space before the relighting operation using the tool presented in 8.3.4 or can be transformed as part of the relighting process using the same basic logic as mentioned in 8.3.4.2. It makes sense to provide both workflows as there might be existing metadata that only works correctly in the existing scene coordinates.

The core functionality of the tool is described in subsection 6.3.3 in detail in the context of the Nuke relight node. This tool will build on a simplified shading algorithm. As shading models the Phong and Lambert models can be used to model diffuse and specular reflections that can be extended with models for refractions and reflections depending on the renderer chosen. Anyway, it makes sense to limit the tool to simple shading models in the near future to provide reasonable performance. All parameters should be editable in an interactive way. Light sources are represented as vectors in the 3D space, which is part of the tool or the host application. Ideally both are possible and coordinate systems are converted internally to be able to work across applications.

To be able to relight light fields in an efficient there must be a way to relight multiple views or the whole light field at once. It might be an option to run several instances of the relight tool for all views simultaneously. As this leads to caching and memory issues probably, it would be better to run the relighting passes sequentially. The setup can be copied to the other views automatically using simple scripting. Compared to the performance of the nuke standard node utilizing multi-threading for all tasks could make it even possible to relight several views at once. Frameworks like Fabric Engine as mentioned in 7.2.1 could easily generate optimized code that can be executed on multicore CPUs as well as GPUs. Graphics processors can speed up vector calculations like the dot product needed for this relighting approach drastically.[497] Still this approach of re-rendering the relighting setup seems to waste processing time redoing similar calculations. This leads to the demand of intelligent propagation algorithms. Two examples are the voxel

[497] Georgiev and Salesin (2010)

carving method described by Seitz or disparity-based 2D warping techniques to transform edits from one view to another.[498] The voxel carving method can be seen as an editing-by-example approach that relies on an explicit geometry representation, a three-dimensional grid of voxels that is fitted to the rough shape of the scene automatically. The edits done on one or several image samples are transferred to all other images by projection them on the voxel geometry. Disparity based approaches are known from the domain of stereo 3D postproduction and approaches to depth based image rendering. Once disparity information is available for all views, the latter approach will be faster. But both approaches will result in image artifacts or holes in occluded areas. Similar problems have been solved in the context of Stereo 3D image synthesis by a combination of the translation or warping approach based on view disparity and a selective re-rendering of occluded parts.[499] There are some robust techniques to detect occlusions like the one presented by Zitnick and Kanade.[500] The algorithms developed by Fraunhofer IIS incorporate similar methods, too.

Image-based lighting could be realized by remapping the HDR input map with the help of the normal and position data. This approach is implemented in the tool mentioned in 6.3.3 and can be compared to a look-up and warping function. A glossy or brushed surface can be simulated like this. To speed the mapping process up spherical harmonics could be employed as a data representation in the context of diffuse relighting. For example, this approach is part of the Escher Tools by Artixels, a plug-in suite for Nuke.

One possibility to compute shadows it using the point cloud representation and project the shape of the point cloud onto other geometry in the scene that can be imported from the host system. Potentially a conversion of scale or data representation has to be part of that process. Additionally, a ground plane could be created by the relighting tool itself to cast a basic shadow onto. Other approaches to shadow creation in the context of relighting are also described in 4.2.4. Increasingly also path-tracing renderer get introduced in the context of compositing tools that might be an option to render high-quality shadows in an offline pass as interactive rates can not be reached at the moment for quality results. The tool could allow for export of a scene description containing light positions and parameters in a standard format to go into a specialized software package for shadow creation, too. In any case there should be filtering tools available to post-process shadows in an efficient way.

To simulate self-shadowing a technique could be borrowed from game development. As games face the challenge of interactive real-time performance an efficient method has been developed to calculate so-called

---

[498] Seitz and Kutulakos (1998) 2
[499] Adelson and Hodges (1992); cf. Masia, Jarabo, and Gutierrez (2014) 3
[500] Zitnick and Kanade (1999) 5-7

screen-space-ambient occlusion (SSAO) from depth data.[501] It is only an approximation of the real occlusion data but can help to add more realism to the relighting output. Maybe there are also ways to incorporate the data from different view positions to extend this approach for more realistic results.

A support of bent normal as introduced in 4.2.4 is also possible but is only of limited use if ambient occlusion data is available that can be used directly for image manipulation.

Finally, the requirement for quality levels can be addressed by reducing resolution and showing only the view that is currently worked on. Additionally the number of points in a point cloud representation can be reduced to gain processing speed. The renderer could only solve the shading equation for a limited number of pixel positions and then interpolate the result for preview purposes.

The output can be filtered with a variety of blur and sharpen algorithms before blended or merged with the input color image. Blend modes suitable are variants of the overlay mode described in 6.3.3. The final merge is a compositing operation based on the operators defined by Duff and Porter.[502] Basically, light is merged with an additive operation or subtractive operation in the case of a negative light.

All settings of the relighting tool can be outputted as metadata inside the merged image or set of images.

8.3.5.2 User Interface

The user interface should be organized supporting the workflow. Therefore, assuming a western cultural background of the user it should start with at the shading options at the top and move down to the options related to merging the light pass with the input.

Shading models can be edited by setting the constants of the basic shading equation. This can be done with input into numeric fields. As shaders get more complex often flowgraphs are used to visualize the data flow inside of 3D software packages (figure 107). The logic can be described by moving around nodes and setting their parameters then. This comes down to a visual programming approach much like the eXpresso language in Cinema4D or the interface in Houdini. For the ease of use there should be some presets that represent known real-world materials to be used at starting points or rough sketches. Basic materials could be brushed metal, glossy metal, plastic, wood and concrete as they can be found as part of most renderers and 3D software packages.

---

[501] cf. Carlsson (2010)

[502] Porter and Duff (1984) 255

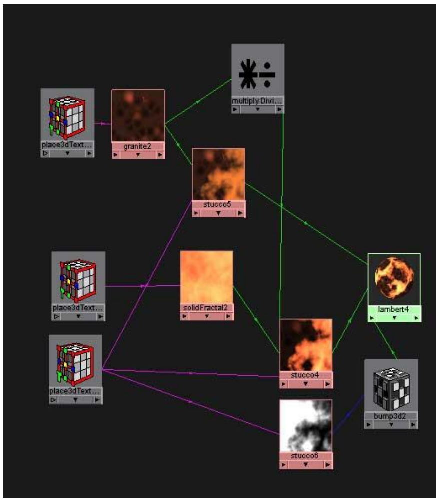


Figure 107: A shading graph inside Maya to define a lava material.

Two view modes should be available: a 3D coordinate system and a 2D image. The latter can be either the input file or the output light pass. Lights are created inside a hierarchical list that enables the parenting of lights for efficient positioning. When adding a light, there is also a graphical representation created that can be viewed in the 3D viewing mode or as overlay in the 2D view. The overlay shows a 2D projection of the light object with handles to interactively move it along or rotate it around the three axis of the light's coordinate system.

The light is represented as a sphere in the case of the point light and as a vector if it is a spot or directional light. An arrow that changes intensity, for example, can visualize the falloff parameter. But there are also slider and numerical inputs to set all parameters. To keep the user-interface clean the sliders can be hidden. The data input can be in real-world units or in an arbitrary scale depending on the setting in the coordinate system and dialog at the top of the tool. Workflows to position and edit lights can be found in the domain of the traditional 3D content creation packages like Autodesk Maya or 3dsMax and the interaction approach with the axis operator presented above can actually be found there.[503] However, as these tools grew in complexity over the years it might be a better idea to look at examples of a more simplified approach. Todays Compositing packages like Fusion or Nuke bring a basic 3D system that is optimized for simple tasks by limiting the number of the available options and operators.

[503] while 3dsmax used to be employed in the context of industry design and visualization and maya for animation, both packages are complete content creation packages nowadays, http://www.autodesk.com/products/maya/overview; http://www.autodesk.com/products/3ds-max/overview

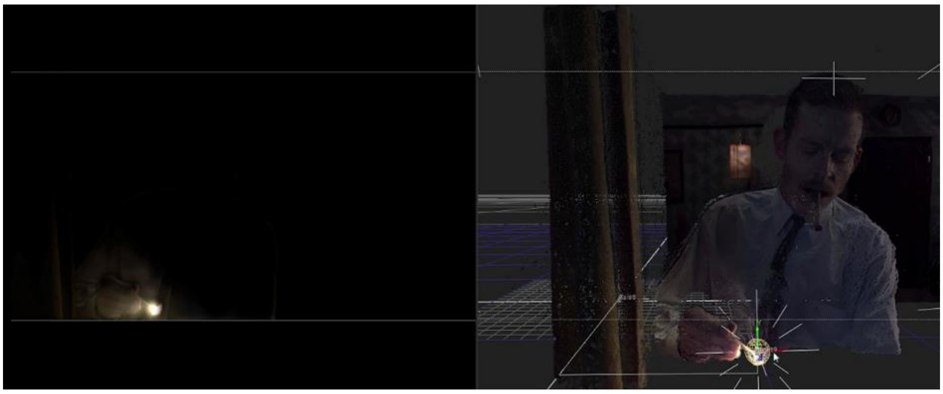


Figure 108: Interactive relighting with a point cloud representation of a scene and a simple geometry representing the light.

Having the demand for interactivity in mind it is probably a good idea to go after an interface design similar to compositing tools. Figure 108 shows a screen grab from the relighting process inside Nuke. Still worth mentioning is Katana by The Foundry that is a dedicated tool just for lighting that offers a streamlined workflow also for very complex scenes allowing for full customization and different viewport renderer.[504]

A drop-down menu offers various blend modes and post filters at the bottom part of the interface. A checkbox that is labeled as "bypass" disables a feature.

8.3.6 LF_Rectify

8.3.6.1 Basic Workflow and Requirements

When using synthetic light fields the rectification process is not necessary due to perfect alignment of the view points, but when using a real camera array the perfect alignment in terms of position and orientation of a camera within the array cannot be guaranteed. Often flexible array setups are wanted, leading to various distances between cameras. Therefore, alignment tools are necessary to provide an initial situation for further light field editing.

Before generating new views the acquired images of the multi camera array have to be aligned vertically and horizontally for example with an algorithm proposed by Zilly et al. in 2012 published in Multi-Camera Rectification using Linearized Trifocal Tensor at the International Conference on Pattern Recognition.

Currently only tools for horizontal alignment of cameras are available in Nuke and also limited to a stereo pair of cameras, for example the O_VerticalAligner included in The Foundry's Ocula Plugin. The node offers the possibility to warp views vertically to match corresponding features horizontally, by applying a

[504] https://www.thefoundry.co.uk/products/katana/; Seymour, "Katana in Production" (2012)

global transform. When maintaining convergence is of importance the tool also offers *Vertical Skew* and *Local Alignment* to warp views while maintaining the horizontal pixel position. [505] In order to align the cameras a previous solve, describing the cameras geometric relationship to another is necessary (further information see [The Foundry Ocula] "Solver" p. 64).

8.3.6.2 Possible Models for Implementation

In order to match features on camera setups flexible in arrangement and amount of cameras other rectification approaches are necessary. For example the rectification algorithm proposed by Zilly et al. using Linearized Trifocal Tensor. [506] The trifocal tensor hereby represents the geometry between a group three cameras at a time. Besides the more precise estimation of horizontal disparity, also vertical disparities are obtained when using a trifocal tensor. Previous approaches were often bound to calibrated cameras or specific camera layouts. Zilly et al.'s approach is suitable for uncalibrated cameras and not bound to specific calibration patterns.[507] The algorithm is then applied to each individual row of cameras and subsequently to all columns. The image coordinates of corresponding pixels in neighbouring views then have a simple geometric relationship, after a rectifying homography is applied to each of the camera images. Epipolar lines are now parallel in neighbouring camera images, vertical lines being in the same column and horizontal epipolar lines in the same row.[508]

Algorithms like the one proposed by Zilly et al. need to be implemented as a user-friendly tool in order to rectify images from various multi camera setups.

Figure 109: Left: Unrectified images. Right: Rectified Images.

505 The Foundry Ocula (2014) 82-84
506 cf. Zilly (2012)
507 cf. Zilly et al. (2012) 2727-2728
508 cf. Zilly et al. (2013) 3

8.3.6.3 User Interface

The user interface is kept very minimalistic, allowing to select various keyframes within a sequence and a function to analyze the images. Preview options for horizontal, as well as vertical, alignment are embedded in order to visualize the success of the rectification process.

8.3.7 Further Required Tools

The following tool suggestions are only mentioned because of their significance in light field postproduction. The description of each proposed tool from here on will be very limited and rudimentary, only describing the essential function.

LF_DisparityGenerator

A tool to generate disparity maps for light field images, describing the location of corresponding pixels within different views. The user interface is similar to Ocula's ODisparityGenerator node, but supporting any desired amount of views. A cloud computing approach is also of interest, especially for smaller companies and end users, because of the amount of necessary computing power in order to calculate accurate inter image disparity.

LF_DepthGenerator

The LF_DepthGenerator node produces a depth map for each view of light field image. The depth map is generated from the disparity of the views. Automated options for refinement of the depth maps, like hole-filling algorithms, denoising and filtering are supplied. Further refinement is possible with the LF_DepthSampler node described in 8.3.5.

LF_NormalGenerator

The LF_NormalGenerator generates normal maps for the desired images and is based on depth maps, and can be further refined by including the rgb-channels in its generation process

LF_PositionGenerator

The LF_PositionGenerator generates position maps from the disparity or depth map.

LF_CameraGenerator

LF_CameraGenerator makes it possibile to generate a virtual camera array from disparity, i.e. for export to Maya, Cinema4D, etc.. Sidenote: The Foundry recently announced the support for multi camera support, during NAB Show 2015, and plans to integrate a tool named *VR_CameraAligner* in their Software package. VR_CameraAligner has a function to solve cameras and generate a virtual camera rig.

LF_ColourMatcher

The LF_ColourMatcher is a tool to match cameras in terms of color, similar to Ocula's matchgrade node, but with the support for more than 2 views.

LF_Convert2D

Light field compositing means a lot of data in the compositing pipeline. Sometimes it's not necessary to have the light field possibilities all the way to the final composite. As soon as the benefits of the light field data aren't needed anymore, this tool provides the possibility to convert the data to a 2D dataset in order to provide a sophisticated and fast 2D compositing workflow.

LF_GenerateGeo / LF_Pointcloud

LG_GenerateGeo and LF_Pointcloud are providing the possibility to create a (partial) geometry of the scene, useful for CG integration, relighting setups, volumetric CG collision and more. Pointclouds and meshes are generated either just by using the depth and position data or in combination with traditional pointcloud generation algorithms known from Nuke's PointCloudGenerator, which uses the information generated by a CameraTracker node.

LF_DeepConvert

Due to the possibility of multiple samples per pixel, because of overlapping camera images, information about occluded surfaces and depth data for the images, a transformation to deep image data is thinkable and useful for some scenarios, as for example mentioned in section 3.4 and 6.2.1. After a conversion with the LF_DeepConvert tool, already established deep compositing workflows can be applied.

LF_PositionToCamera

A node able to calculate the camera position during a sequence, from the position pass. LF_PositionToCamera outputs a three-channel vector value representing the camera's position in world space. The export functionality to other software, i.e. Maya, Cinema4D, etc. is given. The LF_PositionToCamera node is related to the LF_Edit3D node mentioned in 8.3.6.

LF_ZSlice

LF_ZSlice offers the ability to select a slice of the image according to depth and/or position values, similar to Nuke's ZSlice node, but with a better interface, enabling the user to select screen coordinates. The tool can generate masks from position data.

LF_Merge

LF_Merge allows merging light fields or certain elements of light fields on ray basis, similar to the approach by Horn and Chen[509], mentioned earlier in this work. A mask input is supplied, as well as different merge-operations and selectable layers to merge.

LF_RayWarp

Ray warping is mostly used to simulate deformation of a light field or to simulate refractive effects, but also to change position and orientation of scene objects. LF_RayWarp takes the input rays of an image and provides the ability to output new modified rays. A similar approach was mentioned by Horn and Chen on

[509] cf. Horn and Chen

their work "LightShop: Interactive Light Field Manipulation and Rendering". Horn and Chen propose two ways of ray warping. The first is a procedural warp, consisting of linear transformation of 3D vectors. The other option is to use a 4D lookup table, which approximates the ray warping with a number of samples.[510] A ray look up table that can be precomputed by a ray-tracing algorithm. Functions like these cannot be established with traditional 2D tools.

LF_Write

Similar to Nuke's *Write* Node this tool is able to render the results of all preceding nodes to a selected storage device. Input is the data stream containing the light field, 3D data, other channels and layers, as well as metadata for multiple views. Support for multiple file formats is given in order to be able to choose various final output formats such as 2D, 3D Stereo, Multiview (H/V or both), and light fields (sparse/dense).

A selection for varying colorspaces, file types, datatypes, compression algorithms, required views, required layers/channels, metadata, desired type of output (as mentioned above) is provided.

The most important output type will be a light field intermediate format, allowing further light field possibilities like synthetic DoF and view rendering. Depending on the intended use of the images, the importance of containing all the data can be weighed and unnecessary information can be discarded. As for example the use of the light field data for a autostereoscopic display able to show horizontal parallax, vertical views can be limited to a single row of views in order to keep the amount of data as low as possible.

8.4 Conclusion

In conclusion, the proposed tools should be as automated as possible, but still enable the compositing artist to intervene at any time it is necessary. Due to the node based nature of Nuke, iteration should be allowed and is desired in order to achieve the desired result. Essentially the order of the processes should be followed, but the iterative process allows each component to be repeated until a problem is solved, respectively a result is satisfactory. Even though the order of operations should be followed, it's not necessary to do so. To preserve a creative process and experiment with different approaches a change of the tool order is possible at any time during compositing.

[510] cf. Horn and Chen 3

9 End

9.1 Conclusion/Review

> *Can we request that Photography renders the full variety offered by the direct observation of objects? Is it possible to create a photographic print in such a manner that it represents the exterior world framed, in appearance, between the boundaries of the print, as if those boundaries were that of a window opened on reality.* -Gabriel Lippmann, 1908.

This quote by Lippmann might have been the motivation when light fields were first employed in the context of computer graphics. They promised "photo-real" renderings without the need for complex geometric scene representations and processing hardware as well as algorithms that had yet to be invented. Today the computer graphics community discusses hyper-real renderings and the movie and TV industry is capable of showing visuals of everything a director can describe.

Although light field technologies, like the LightStages, have been and are used in the context of media productions, the major part of the research in the field focused on other application in the past years. As technology advanced the first commercially available light field cameras appeared. Light fields as one part of the wider field of computational photography might be able to bring unique and new aspects to digital photography, a discipline that has long been busy trying to copy the analog process from the past. Looking at an industry that is faced with constantly reinventing photography while holding up to the traditions of more than hundred years of storytelling achievement, this work tried to present promising applications of light field data in the context of todays film and TV production workflows. As part of a research project shared between the Fraunhofer IIS in Erlangen and the Stuttgart Media University (both in Germany), a meeting with industry experts and consequently a test shoot with an experimental camera system had been organized. In the first part of the work an introduction on the basic concept of light fields and some existing acquisition systems and research was given before reviewing the pre-production phase of the test shoot. Building on existing applications of light field technology and the outcome of the experts meeting some promising use-cases for the main applications of light field data, view rendering, digital focus and extraction of 3D data, could be found that address effective demands of film and TV productions. Light field technology seemed to be able to improve production efficiency and to introduce new artistic possibilities that could help generate superior results compared to conventional approaches in the long-term. Next, three scenarios were selected that represent common procedures of film and TV productions. These scenarios lead to the design of the test shoot and provided the scope for the workflow design as part of this work. Also, overall requirements for light field technology from the industries' perspective were defined that helped guiding considerations towards feasibility.

In the second part of the work the backlot, packshot and portrait scenarios were described in detail regarding possible challenges, current production procedures and workflow practices as well as possible applications of light field technology. It has been shown that light fields have the potential to improve different aspects of production workflows. After broadening the view to define a common basis for the terms workflow and pipeline for a moment, the practical test shoot has been described. Some approaches to light field integration into a production workflow could be tested in the course of the test production. These have been analyzed to be able to formulate further requirements at the end of chapter five. Based on the experiences from the test shoot, some theoretical considerations and following a list of requirements some ideas could be presented on how to integrate light field data in a production workflow for future applications. Finally, requirements and implementation models for the development of a production pipeline have been defined in the form of an open toolset for image processing that addresses some of the demands we presented earlier in the context of the three scenarios.

The initial question about requirements for successful integration of light field data has been answered for the three scenarios. Requirements were collected and presented in a comprehensive way supported by literature and industry practices. These might be used as a guideline for future developments and could also be of use for research and development in other areas. Also some applications for light field data that address the demands of today's media productions were proposed. Still, not all requirements can be fulfilled by current approaches and available light field technology. At the current stage a clear answer cannot be given whether incorporating light field technology will help productions to gain efficiency and quality. At the time of writing it seems like Technology and software still have to develop to efficiently integrate light fields in a production workflow for traditional film and TV output. The main aspects are the need for extra time and manual work and image quality issues resulting from the small cameras and imperfect algorithms at the moment. We were able to successfully apply light field data in some use-cases, showing that integration is possible while being constraint to the acquisition system of a sparse sampled light field or compressed light field. With algorithms still at an early stage of development this lead to a big amount of manual tweaks and cleanup work on the 3D data to match the requirements regarding image quality at the end of the production pipeline. Only one light field application, the extraction of 3D data actually provided results of production quality after all. But it should be kept in mind that only the sum of possible applications will make light field technology attractive for producers. The richness of the scene data can enable lots of different applications after the actual shoot, which is of great use if post work cannot be fully foreseen due to the complexity of the process. These special cases will probably be the occasions were light fields can be put into use first. The scenarios backlot and second unit photography in the context of e.g. aerial photography will be an interesting field of applications that has a strong need for new acquisition systems. Looking at the technological obstacles and amount of data produced it probably won't be used as a main production camera system soon.

Clearly, the proposed workflow has to be interpreted as a broad guideline for future developments only and might not be ideal for applications in the near future. Although we focused solely on the creation of conventional formats for film and TV the multi view workflow could also be relevant for applications like virtual reality and expanded vision in the context of physical installations. Nevertheless the approach using a toolset across workflow stages and departments definitely makes sense as it follows the trend of non-linear production workflows and interdependencies across departments, facilities and workflow steps. Like the multi-view approach in the context of the workflow outline the toolset might also be used in other areas of media production. Lots of other multi-camera applications and hybrid approaches that do not necessarily output a dense 4D light fields have also been discussed as part of this project, as mentioned in chapter 6. This might be a first step on the way to a light field film camera until technological limits have been overcome.

Looking back, the combination of the algorithmic knowledge and theoretical context of the Fraunhofer IIS with the production know-how at the HdM turned out to be a fruitful undertaking. Together we were able to successfully demonstrate the feasibility of using light field technology in the production of a fictive commercial short form. The experts meeting organized by the HdM brought valuable insights and new aspects into the project that helped moving it to the next development stage. The existing infrastructure at HdM could be used to comfortably produce the test shoot with a very short production time. As with most new technologies the first attempts will probably not result in perfect products. In our case the short preparation time caused some problems in the course of the production as the cameras could not been fully tested and characterized prior to shooting. This lead to some problems later in production as there has not been a known workflow available regarding debayering of the raw data and color management. In our tests we were limited only one acquisition system. However, the workflow considerations presented in this study were developed as camera independent as possible.

Unfortunately, due to limited time a scientific approach to fundamental testing and, most of all, evaluation of the results of tests was not possible. The team of four to nine people was busy with producing the shots for a presentation reel in the first place. This gave a good impression of the needs for integrating light field data into an efficient and running production workflow and the experiences helped to define the requirements as part of this work. But it would have been desirable to also make workflow decisions based on a scientific analysis and comparing different solutions in terms of performance and image quality, for example to be able to give an educated recommendation on workflow and pipeline design. We always planned and kept several workflow variants but never had the resources to test them under equal conditions. For example, it could not be tested what effect image preprocessing like denoising or transforming into a specific colorspace and grading could have on the result of the depth estimation.

For people thinking about employing light field technology in production this work offers a quick overview about the state of the arts, the basic theory and ideas for efficient application. It might be too early to define things like a data format and we would like people to encourage trying different approaches and to discuss the results. This explorative study only captured the requirements at a certain point in time. Hopefully, these can be useful when going further on the path of light field integration. Still, like in every fast moving field of research, theories and models can turn out useless in the end.

By now we also received news from the NAB 2015 in Las Vegas, USA. An edited version of the commercial spot has been presented together with the research project. The spot shows how the light field application 3D data extraction for a single view has been used to relight shots, generate masks based on depth data as well as preview the concept of digital focus with a depth based blur algorithm as part of a VFX process. The presentation got positive feedback from a big audience of film and TV industry professionals.

And the project already continued with developing plug-ins similar to the ones described in chapter seven. While still different acquisition systems are to be tested and developed we advocate the development of postproduction tools and standards. Apart from the aspect of the data format the rectification and depth estimation algorithms are a key part of the postproduction pipeline. Especially in the case of Fraunhofer's camera system, all applications depend on the quality of the depth data. In this context also a calibration workflow for the camera and the algorithms is a topic of higher priority. At the same time ways to improve performance have to be explored to achieve interactive editing for all light field applications.

As mentioned before, in this thesis we looked at things from the perspective of a film production based on live-action footage as primary source of the process. Using light fields as a data representation for CG renderings, as the term is used in this work mostly for synthetic renderings, could also be explored. The outcome of such experiments could change some of the assumptions made in the context of a pipeline. Film being a storytelling business content plays an important role for the success and acceptance of a new technology. Possible subjects could be related to the aspects time and space that can be experienced and manipulated in unconventional ways when using light field technology. Also, artifacts from the 3D extraction process can be used to create a unique visual style in commercials, video clips or stories about digital worlds.

Taking into account industry trends and technological development the following and final part presents some ideas for future development, also outside the boundaries of traditional film.

9.2 Outlook/Perspective

"We reached that penultimate level of winning an Oscar and now it's time to really show people what we can do in the more traditional filmmaking world"

-Ian Hunter, 2015.

During an interview with Will Mason, author at uploadVR, and Ian Hunter, co-founder of New Deal Studios, responsible for effects in Interstellar, X-Men: Days of Future Past, The Avengers, Inception among dozen others, said this concerning the Studios future shift in orientation from creating effects to content creation focusing on VR. This is just an example on how big the demand for new content is getting in Hollywood, not only concerning movie tie in projects, which were always connected to gaming and other interactive technologies, but also movies and TV-productions itself.

Since our goal was to show the possibilities of integrating light field technology in current production environments and giving new creative opportunities for postproduction and content creation, the above quoted statement from Hunter fits quite well. But in order to get there and show new approaches to traditional filmmaking, a few steps are still necessary.

In the following sections, we're going to introduce some near-future options concerning the presented research as well as technologies benefiting from light fields, apart from the ones already described.

9.2.1 Future of our Research Project

As mentioned in Section 9.1 this work summarizes the current state of the research project and the used multi camera array prototype. The Feedback we received for our work at NAB Show 2015 was very positive, but yet the cooperation between Stuttgart Media University and Fraunhofer IIS has just started and the project will continue in order to develop the industrial maturity needed to show the benefits of using light fields in modern production environments. Reaching industrial maternity would also mean reaching a TRL 6. To get to this stage, a further evaluation of our test results is necessary and additional scientific tests have to be conducted. Workflows have been proposed by us, but only in a minor magnitude, therefor workflow approaches have to be developed, tested and compared, as well as their resulting outcomes. Fraunhofer IIS is currently working on the implementation of their *Light Field ViewRenderer* for Nuke and is working with The Foundry as a partner in research. As soon as the basic functionalities are implemented, the goal is to further develop tools for light field editing and compositing. Hopefully our suggestions for a Nuke toolset are of help during the development-phase. The implementation of editing tools is currently the next step, followed by excessive testing and consistent further development. Various user studies, consisting of different types of users (f.i. VFX students, junior artist, senior artists and developers), have to be conducted, similar to Jarobo et al.'s studies, in order to improve the user interfaces and workflows, as well as to focus on improvements where they are most needed. User studies will also show what features are desirable for everyday work and which implementations and functionalities can be neglected or limited by certain amounts. The next step would be the test of the implemented tools in a real production environment, comparable to our fictional production, but further down the course of development, with more knowledge up front and a camera system adequate for high quality productions.

During the NAB Show 2015, Jon Wadleton presented upcoming features of The Foundry's Ocula for Nuke, concerning VR and AR content. Even though our research does not concern VR/AR at the moment, Ocula's future support for multiple cameras, as yet limited to two cameras, is of great interest. Color matching for multiple cameras will be integrated, as well as a camera alignment and solving, making it possible to generate a multi camera rig. The multi camera support will be of big help when implementing support for the Fraunhofer IIS camera rig.

Parallel to software integration, user interface development and excessive evaluation, the need for faster and more precise algorithms, higher quality images, storage solutions, and improvement of on-set work is existent and should not stay unattended. This means improvement on behalf of the used software, but likewise improvements in hardware and maybe even completely new approaches, in order to obtain the desired output quality. The handling of camera, calibration and recording devices, among others, needs to be addressed in order to support an effective production workflow. As mentioned by Siragusano during our experts meeting, the cameras don't have to necessarily be, neither do, the same. In order to gain more image information companies like Pelican Imaging already derived from many others by using monolithic sensors in their camera array in order to eliminate color crosstalk. Each camera in their array is sensitive to a single color (respectively color spectrum) and can be controlled independently to improve fidelity and reduce the need for aggressive color corrections. If one thinks even further it might be of interest, to not only use sensors that are sensitive to a single visible color spectrum, but sensors that are sensible to other wavelengths outside the visual spectrum. By staggering trigger times of multiple cameras in an array it would become possible to capture HFR, even if the cameras themselves might not be able to capture high frame rates. Algorithms for picture reconstruction would be challenged even further in order to reconstruct different views of a certain time, because the neighboring images aren't captured at the exact same time, but optical flow estimation and warping algorithms are also still improving. Besides HFR, HDR would also be possible with a multi camera array, by altering the exposure values among different cameras in order to obtain more dynamic range, as mentioned in section 3.2.1.4.

Depending on the intention of the production, respectively the postproduction, it might not be necessary to use a multi camera array. Plenoptic cameras only allow minor change in perspective (view rendering), but often there is no need for this application. The benefits of using a plenoptic camera over an array camera can be quite large, because only a single camera needs to be handled and correspondingly only one lens. State of the art cinema or broadcast cameras like the *Red WEAPON*, *Astrodesign* ***AH-4800***, *NHK SHK-810* and *Hitachi SK-UHD8060*, recently presented at the NAB Show 2015, offer 8K resolution with up to 75 frames capturing speed. By applying a microlens array to these systems the benefits of light field postproduction in terms of DoF, grading, relighting, etc. would be given while still maintaining a resolution adequate for cinematic projection. The resulting effective resolution is drastically reduced by using a microlens array, Lytro is currently using approximately 10% of the initial sensor resolution, while Raytrix seems to be able to use 25%, in the latter case this would mean an output resolution of 2K, when using an 8K camera and therefore comply with today's cinema standards.

In terms of software and algorithms, the user interface has already been mentioned above, but algorithms may have to be adapted or modified to handle new hardware approaches. Current algorithms have to be refined in a way to always deliver the best quality possible with the least amount of work necessary. Especially alignment, sub-pixel interpolation, intermediate view interpolation and disparity refinement have to be addressed. Superresolution is of great interest when acquiring images with multiple capturing devices.

Approaches, like the ones proposed by Georgiev and Lumsdaine (2009), Bishop et al. (2009), Wanner and Goldluecke (2013) should be taken into consideration in order to maximize the output quality.

9.2.2 Necessity for Light Fields in VR

The desire for immersive, emotionally engaging entertainment has been present for decades and was mostly linked to gaming. Since the big hype of virtual reality over the last few years and especially the last few months, due to an evolving market and competing developers such as Sony, Oculus, Valve and Samsung, new fields of application besides gaming, are recognized. Suddenly everybody is investing in VR, broadcasting, cinema, gaming, software companies, etc. and new companies spring up like mushrooms, trying to get in early. Still, little is known about possible business and distribution models, but according to Digi-Capital the virtual reality market is projected to generate $30 billion in revenue.[511]

While it's fairly easy to display computer generated content in VR and on HMD[512] devices, it's still difficult to capture and display real-world images in a way to be perceived as realistic. The problem of capturing live action VR is that you're bound to the perspective of the cameras and lenses used during a shoot. Distortion, image interpolation, stitching, etc. make adjustments possible, but the image was acquired in a certain perspective and all the manipulation is just trying to rebuild the reality and often leads to artefacts and other limitations like the problem of focus-convergence mismatch in classical stereo 3D.

Lytro, known for its light field cameras, is also joining the VR movement. They recently raised $50 Million in order to shift to new areas, including virtual reality and video. While light field photography is mainly known for the possible focus shifting, the real interest for VR lies in perspective. A simulation of real two-eye perspective and parallax in horizontal and vertical direction is necessary to perceive an image as "real". Most of the current real-world VR panoramas offer no parallax or only slight horizontal parallax from stereo acquisition. In addition to that, the user's perspective has to be calculated in real time, matching the movement of his head. At FMX 2015, in Stuttgart, OTOY's Chief Scientific Consultant, Paul Debevec, announced the "*first ever capture of a spherical light field of a real world environment used to produce a completely accurate, navigable scene in virtual reality*". They captured still image light field data of a room, which could

[511] $150 million, when combined with augmented reality revenues.

[512] See glossary

be rendered instantaneously in VR, providing 360° view, correct parallax movement and the possibility to move around in space (by a limited amount).

Companies like NextVR, are following similar approaches, by incorporating light field technology to provide the viewer with live action VR.[513] They already tested stereoscopic 360° live streaming for VR Headsets successfully in March 2015. Endgadget reports, they are currently working on positional tracking and adding light field technology to their 360° camera rigs, by applying plenoptic lenses to their cameras.

With our main interest in capturing light fields, they are not only of interest when it comes to real-world content creation for VR, but also for displaying the content in the HMDs. Displaying a light field would mean, seeing an image as close to real as currently possible, with the ability to focus on various depths of a scene. At SIGGRAPH 2013, Nvidia presented a HMD featuring light field displays, enabling the user to refocus at multiple depths of an image, giving a more immersive feeling. The prototypes displays only achieve a spatial resolution of 146 x 78 pixels and is therefore currently not matching the resolution needed for VR Headsets. But the approach of being able to realistically refocus with your eyes, already briefly mentioned in subsection 2.3.4 is an approach worth following.

One of the industries most talked about topics in the last months was Magic Leap, a Florida based Start-up Company, which, according to Forbes, received $542 million in funding during the third quarter of 2014. By now, no specifics concerning their technologies are known, only rumours and patent and trademark filings. Those filings describe a head mounted display device presenting realistic 3D imagery, able to create similar 3D patterns of light rays (light fields), like the ones our eyes receive from real objects. One of the patents also describes the functionality to produce virtual objects at different distances. Instead of microlenses, Magic Leap relies on many small curved mirrors, titled *Waveguide Reflector Array Projector.* Motion sensors, eye-tracking, cameras, infrared sensors and ultrasonic sensors are also part of the patent filings.[514]

It remains to be seen where virtual reality leads us and what technological breakthroughs will be accomplished.

---

[513] Note: statements concerning their technology vary, in some statements from NextVR they talk about light field technology, in others they mention 3D mapping with the help of LiDar scans.
[514] cf. Simonite, MIT Technology Review (2014)

9.2.3 Further Trends relevant for Light Fields

Besides the already mentioned need for further research and development and the necessities for light fields in VR, other trends and techniques are also applicable to, respectively of interest to, light field topics. Democratization of content, open source and cloud computing are major trends in the media industry and of great interest when thinking about development, solutions, delivery formats, algorithms and computation.

Several years ago, especially in the analog world, it was time consuming, expensive and difficult to create a film. The process of film making has been the domain of the film and TV industry and was hardly accessible for the average citizen. The digital era and the internet changed all that. Nowadays, in the DIY[515] digital world, almost everybody is able to create content, backed by technology and ingenuity. Our desire to create, edit and publish content is driving the industry to new measures to keep up with the consumer's demands. Wrestling the control of content distribution out of the hands of a few media production companies into the hands of anyone with a video camera and website is referred to as democratization of content. Democratization of content helps to generate new and fresh ideas and has the ability to change and empower storytelling and film making in many ways. Big production companies are far from obsolete, but further have the opportunity to react according to the consumers demands, formulated in the form of content. For light field technology this could mean great benefits and possibly never before thought of fields of application and new creative ways of storytelling.

Experimentation leads to breakthrough, it has always been this way and will stay this way. In many instances technological innovation is not driven by proprietary development, but rather by experimenting. Thousands of open source projects are leading innovation in the area of technology, showing that rather than on a single company, a development process can also rely on the distribution of R&D[516]. Open source is of vast interest concerning the development of tools for light field editing, such as those mentioned in section 8.3. It would be utterly exciting to see what the "community" comes up with. The development could be spread all over the world, with developers exchanging ideas and helping mutual colleagues in order to implement new approaches.

---

[515] DIY: do it yourself

[516] R&D: Reasearch and Development

On their blog, Lytro recently (05/2015) announced they are looking for alpha users. By opening their platform for users to experiment with, they want to see what users are able to do with their Lytro Power Tools (LPT) and gain valuable feedback. The power tools are a set of command line utilities, making light field exploration directly accessible.[517] Even though Lytro is not going open source, it's a step in the same direction of giving users the freedom to explore and develop.

Ultimately, open source could lead to a common standard by combining the requirements collected from a large amount of experiments and research.

By dynamically supplying resources like computational power, data storage, software and networking capacities over the Internet, users and developers alike are enabled to use resources which would otherwise be most likely inaccessible. Those resources could be a solution for the enormous computing power, at the time being, necessary to use some of the algorithms and practices presented in this work. Infrastructure as a Service, often referred to as IaaS, could provide virtualized computer hardware resources for computational needs and rendering. For developers, access to programming- and runtime environments could be established as "Platform as a Service", often referred to as PaaS. Alike, access to software, needed for special tasks, could be provided as SaaS or "Software as a Service". Those models are fundamental models of so called cloud computing.

[517] http://blog.lytro.com/post/118970358050/looking-for-alpha-users

Acknowledgements

We wish to thank everybody who supported this project and helped to accomplish this work. We would first like to thank Prof. Katja Schmid, for her guidance and especially appreciate Prof. Schmid's encouragement to pursue this research project. Further we are grateful for the great assistance of Daniele Siragusano and Ron Ganbar with our work. To the employees of the Stuttgart Media University, Peter Ruhrmann, Simon Walter, and Michael Kirschenlohr, thank you for your feedback, input and assistance when we needed it.

To the research team at Fraunhofer IIS consisting of Frederik Zilly, Matthias Ziegler, Joachim Keinert, Lukas Kolhagen, we are thankful for the opportunity to work with you on a project of this scale. Thank you for providing the technology and needed backgrounds to accomplish our research.

Gratitude is extended to the whole team of students working on this research project, Amelie Dillig, Clemens Helmchen, Johannes Hölz, Larissa Kurtz, Marlene Fritzsche, Moritz Wetzig, Sebastian Zeeden, Sergej Feininger and Tim Heumesser. Finally we'd like to thank the team of industry experts for their help and opinions concerning light fields and steering our research in the right direction.

Appendix A – Glossary

Alembic	Alembic is an open-source framework for storing and sharing scene data across applications that includes a C++ library, a file format, and client plugins and applications. It was initially developed in 2010 by teams from Sony Pictures Imageworks and Industrial Light & Magic, and development continues today. Being a data storage format it is intended to store a baked representation of scene data. Compared to other scene exchange formats like OBJ or FBX it has the advantage of a more efficient use of memory and disk space.[518]
Bezier Curve	A curved line or path defined by mathematical equations. The method was named after Pierre Bézier who developed this technique of computer drawing at Renault in the late 1960s.[519]
CAD (Computer Aided Design)	CAD is a group of applications in computer graphics. Computer aided design describes techniques to design and visualize products.
IBL image based lighting	Made popular by the research of Paul Debevec IBL is a standard tool for lighting computer graphics nowadays. Usually a set of HDR images is taken on set to capture the lighting conditions. This imagery, often as a spherical map, is then used to provide information for realistic diffuse or specular reflection. When projected on a sphere the surface shader can perform a look-up of values for the diffuse and specular directions. For diffuse lighting the map usually gets blurred with a convolution filter first. All major renderers currently support IBL.[520]
Matchmove	camera tracking or matchmoving is a two step process. Firstly, the movement of 2D features in an image is tracked using methods based on pattern recognition algorithms, then several of those movements are fed into photogrammetry like calculations to get a camera vector. Having two or more views on a scene can add useful constraints to the solving process.
Matte painting	Traces back to the analog process of painting background images. Nowadays, digital paintings are often created with software tools like Adobe's Photoshop. These can incorporate photos or film stills or can be based on a still frame from a movie sequence. Nevertheless matte paintings are stills that could be animated at a later stage.
Mesh	Similar to the real word, a mesh describes in computer graphics a collection of vertices, edges and faces that define the shape of an object

[518] https://code.google.com/p/alembic/

[519] Okun and Zwerman (2010) 844

[520] cf. Whitehurst (2010) 665

in 3D space.[521] These surfaces are usually made up by triangle or quadrilaterals but can also have other general polygons. There are different kinds of mesh representations and also a wide field of research regarding generation of meshes from a collection of 3D points as part of an automatic process.

MIPmap and RIPmap (texture maps)

Texture mapping is the process of overlaying an image onto a CG object.[522] UV coordinates traditionally describe the polygon surface of objects and are used to map an image file onto the surface. If an object is closer to camera these texture images need to be of higher resolution compared to situations where an object is only part of the background of a scene. Originated in game development, MIP maps are rectangular texture tiles that hold multiple resolution levels in one file. Depending on the distance between object and camera the corresponding image data is used for rendering. In the case of mipmaps the width and height of each level is a power of two smaller than the previous level, RIPmaps can store arbitrary levels of detail.[523]

OpenVDB

> *OpenVDB, now available in version 3, is an open source C++ library comprising a novel hierarchical data structure and a suite of tools for the efficient storage and manipulation of sparse volumetric data discretized on three-dimensional grids. It is developed and maintained by DreamWorks Animation for use in volumetric applications typically encountered in feature film production.*[524]

It is supported by all major rendering applications. While being a data representation basically, it enables more efficient computation for large scenes in simulation tasks like fluids, for example.

Optical flow

A technique to procedurally determine the movements of objects in a a sequence of images by examining the movement of smaller blocks of pixels within the image.[525] Usually, the output is a vector field defining directions of movement for each pixel or pixel block, the velocity field. In Visual Effects it is often used as part of retiming algorithms to interpolate in-between images or to apply realistic motion blur to a images. Apart from that, optical flow is part of compression algorithms like MPEG.[526]

Plate

A piece of original photography that is intended to be used as an element in a composite.[527]

---

521 “Polygon Mesh” (2015)
522 Okun and Zwerman (2010) 886
523 Windows CE .NET (2004)
524 http://www.openvdb.org/
525 Okun and Zwerman (2010) 872
526 Seymour, “Art of Optical Flow” (2006)
527 Okun and Zwerman (2010) 875

Playblast

Is a fast preview version of an animation. It is of lower quality and usually created using the hardware acceleration inside the viewport or preview screen of a 3D software package, therefore having the same look as the preview image used by animator, for example.

Previsualization, previs

The Joint Technology Subcommitee on Previsualization which is a collaboration between the Visual Effects Society (VES), the Art Direcotr's Guild (ADG) and the American Society of Cinematographers (ASC) proposes the following definition:

> *Previs is a collaborative process that generates preliminary versions of shots or sequences, predominantly using 3D animation tools and virtual environment. It enables filmmakers to visually explore creative ideas, plan technical solutions, and communicate a shared vision for efficient production [...]*[528]

We would like to emphasize that Previs is above all an internal tool that enables efficient and creative communication. Depending on the production stage that Previs is used in, different terms are used as follows:

- Pitchvis visualizes the potential of a project before it is entirely financed or "greenlit".
- Techvis or Technical Previs incorporates and generates accurate camera, lighting design and scene layout information to help define technical aspects and requirements of a shoot. Usually real world terms and measurements are used.
- On-set Previs is often related to Visual Effects as it creates a fast visualization of planned effects on location for the VFX, director and Director of Photography (DP).
- Postvis combines shot elements and digital assets to validate footage selection and preview for editorial.
- D-vis or design visualization takes place in preproduction, again. It's part of the look development process and can utilize virtual frameworks to generate virtual design spaces.

PTex

Ptex is a texture mapping system developed by Walt Disney Animation Studios for production-quality rendering that does not need uv-coordinates. A separate texture is applied internally to every face of polygon mesh.[529] Additionally, it promises better performance for models with a large number of textures due to better I/O and caching. Since 2010 it is free open source.

Rotoscope (Roto), rotomation

The rotoscope originaly is a device patented in 1917 by Max Fleischer that helped in cell animation. It basically projected a frame on a glass screen. The arist could then draw on that screen. The term is now used to describe any process of creating imagery or mattes on a frame-by-frame basis by hand. Rotomation refers to 3D animation

[528] Mat Beck (2010) 54-55

[529] http://ptex.us/

that is also done on a manual frame by frame basis to match the movement of an object or character in a 2D image.

Spherical map, latlong Spherical representation of a 360° Panorama as a latitude/longitude projection.

Visual Effects VFX refers to a number of individual tasks in film, TV and advertising postproduction. They constitute a sub-category of special effects (SFX) and are a post process, opposed to Practical or On Set SFX (sometime called only SFX, too). Nowadays VFX implicitly means the use of computers and aims at the integration of an effect to a final image that appears to be photo-real. Computer graphics and animation play a key role as elements of VFX at the moment. There can be found three major areas of VFX:
1. Manipulation of filmed footage by adding, removing or combination of one or more source images
2. Creation of new image elements like computer graphics (CG) or 2D methods like matte painting and the combination with shot live-action footage
3. Generation of moving image sequences digitally with very little or no shot elements.[530]

Wire removal, rotopaint Wire removal is a generic term to describe the process of using digital painting and compositing techniques to remove undesirable harnesses, rigs and wires that may have been needed to aid stunts or practical effects.[531] Nowadays this work incorporates all kinds of VFX techniques and often is done by specialized facilities.

[530] Assoc. Prof. Benjamin Seide 2013, VFX course I, NTU Singapore
[531] Okun and Zwerman (2010) 892

Appendix B – References

"A Look at FLIP with GrundyUFA." URL: http://www.filmlight.ltd.uk/customers/case-studies/grundyufa.php. *filmlight.ltd.uk*. 2015. Web. [Date of access: 3 Feb. 2015].

"Basic Darkroom Techniques." URL: http://www.kodak.com/global/en/consumer/education/lessonPlans/darkroom/fullCourse.shtml. *Kodak.* Web. [Date of access: 15 Apr. 2015].

"DMX512 FAQ." URL: http://www.usitt.org/content.asp?contentid=373. *USITT.* Web. [Date of access: 11 May 2015].

"Ftrack | Creative Project Management & Team Collaboration." URL: https://www.ftrack.com/. *ftrack | Production Tracking, Asset Management, Review and Team Collaboration*. N.p., n.d. Web. [Date of access: 7 May 2015].

"Grundy Ufa Nutzt Filmlight Flip." URL: http://www.film-tv-video.de/newsdetail+M5e38e9994a6.html. *Film-TV-Video.* Feb. 2013. Web. [Date of access: 12 May 2015]

"Light Stages." URL: http://ict.usc.edu/prototypes/light-stages/. *USC Institute for Creative Technologies.* ICT Graphics Lab, 2015. Web. [Date of access: 6 Mar. 2015].

"Light-Field." URL: http://www.iis.fraunhofer.de/en/ff/bsy/tech/lichtfeld.html. *Fraunhofer Institute for Integrated Circuits IIS.* N.p., n.d. Web. [Date of access: 16 Feb. 2015].

"Nuke Node Reference Guide." URL: http://thefoundry.s3.amazonaws.com/products/nuke/releases/9.0v5/NukeReferenceGuide.pdf. *The Foundry.* 2015. Web. [Date of access: 11 May 2015].

"NVIDIA Collaborates With Weta to Accelerate Visual Effects for Avatar." URL: http://www.nvidia.com/object/wetadigital_avatar.html. *Nvidia Filmmaking and Animation.* N.p., n.d. Web. [Date of access: 19 Apr. 2015].

"Pf Track." URL: http://www.thepixelfarm.co.uk/product.php?productId=PFTrack. *The Pixelfarm.* N.p., n.d. Web. [Date of access: 19 Apr. 2015]

"Shotgun Software | Production Tracking, Scheduling, Review and Asset Management." URL: http://www.shotgunsoftware.com/. *Shotgun Software.* N.p., n.d. Web. [Date of access: 7 May 2015].

"Technology Readiness Level." *NASA.* URL: http://www.nasa.gov/content/technology-readiness-level. 6 June 2013. Web. [Date of access: 6 Feb. 2015].

"The Nyquist-Shannon Sampling Theorem." URL: http://ptolemy.eecs.berkeley.edu/eecs20/week13/nyquistShannon.html. *University of California.* 2015. Web. [Date of access 23 Apr. 2015].

"What Is a Mipmap?" URL: https://msdn.microsoft.com/en-us/library/ms892793.aspx. *Windows CE .NET.* Apr. 2004. Web. [Date of access: 24 Apr. 2015].

Abbott, Joshua E. "Interactive Depth-Aware Effects for Stereo Image Editing." Master Thesis, Brigham Young University, 2013: n. pag. Print.

Adelson, Edward H., and John Y. A. Wang. "Single Lens Stereo with a Plenoptic Camera." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, 1992: 99–106. Print.

Adelson, Stephen Joel, and Larry F. Hodges. "Stereoscopic Ray-Tracing." URL: https://smartech.gatech.edu/handle/1853/3672. Technical Report. N.p., 1992. Web. [Date of access: 30 May 2015].

Aditi, Schailender Gupta, and Mansi Dua. "A Hybrid Approach for Dehazing Images." *International Journal of Hybrid Information Technology.* Vol. 3. N.p., 2014. 199–210. Print.

Alamia, Marco. "Article - World, View and Projection Transformation Matrices." URL: http://www.codinglabs.net/article_world_view_projection_matrix.aspx. *Codinglabs.* Web. [Date of access: 10 May 2015].

Anthony, Sebastian. “Google Joins the VR War, Invests in Light Field ‘cinematic Reality’ Company Magic Leap (updated).” URL: http://www.extremetech.com/extreme/191909-google-joins-the-vr-war-invests-in-light-field-cinematic-reality-company-magic-leap. *ExtremeTech.* N.p., 21.10.2014. Web. [Date of access 5 Feb. 2015].

Baroti, Tekle. “14 Trends for Filmmakers to Look Out for in 2014 | Raindance.” URL: http://www.raindance.org/14-trends-for-filmmakers-to-look-out-for-in-2014/. *Raindance* N.p., Dec. 2013. Web. [Date of access: 25 Feb. 2015].

Barrow, H.G., and J.M. Tenenbaum. “Recovering Intrinsic Scene Characteristics from Images.” Academic Press, 1978: n. pag. Print.

Bavoli, Louis, and Miguel Sainz. “Screen Space Ambient Occlusion.” Nvidia, Sept. 2008: n. pag. Print.

Beck, Mat. “What Is Previs?” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 53–55. Print.

Beier, Bela. “Hobbit in HFR.” *Digital Production* May 2013: 56–59. Print.

Bernasconi, Xavier. “Lighting and Rendering Pipeline”. ADM, Nanyang University, Singapore. 2013. Lecture.

Bonnington, Christina. "Ren Ng Shares His Photographic Vision: Shoot Now, Focus Later." URL: http://www.wired.com/2011/06/ren-ng-lytro/. 2011. Web. [Date of access: 19 Mar. 2015]

Bugaj, Stephan Vladimir. “Analysis of a Production Workflow.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 784–794. Print.

Bunish, Christine. “Reflections on VFX in Commercials.” URL: http://markeemagazine.com/wp/reflections-on-vfx-in-commercials/. *MarkeeMagazine.com.* N.p., Dec. 2012. Web. [Date of access: 27 May 2015].

Büttner, Manfred. “Special-Effects - Mehr als Zauberei.” Medientage München. 2004. Keynote.

Byrne, Bill. “The Visual Effects Arsenal – VFX Solutions for the Independent Filmmaker” Focal Press Elsevier, 2009. Print

Calahan, Sharon. “Storytelling through Lighting, a Computer Perspective.” *Advanced Renderman: Creating CGI for Motion Pictures*. Ed. Morgan Kaufmann. San Francisco: Morgan Kaufman Publishers, 1999. 337–382. Print.

Cardinal, David. "Pelican Imaging promises freedom from focusing." URL: http://www.extremetech.com/computing/152761-pelican-imaging-promises-freedom-from-focusing 2013. Web. [Date of access: 12 Mar. 2015]

Carlsson, Joakim. “Realistic Ambient Occlusion In Real-Time Survey of the AOV Algorithm and Propositions for Improvements.” University of Gothenburg, 2010: n. pag. Print.

Caselles, V. “Exemplar-Based Image Inpainting and Applications.” URL: https://www.siam.org/pdf/news/1941.pdf. *SIAM News* 44.10 (2011): n. pag. Web. [Date of access: 10 May 2015].

Chen, Billy. “Novel Methods to Manipulating and Combining Light Fields.” Phd dissertation, Stanford University, 2006. Print.

Chen, Shenchang Eric, and Lance Williams. “View Interpolation for Image Synthesis.” *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*. New York, NY, USA: ACM, 1993. 279–288. Web. SIGGRAPH ’93. [Date of access: 9 March 2015]

Chip.de mwa/mil. "Lytro Illum: Die Zukunft der Fotografie im Praxis-Test" URL: http://www.chip.de/news/Lytro-Illum-Die-Zukunft-der-Fotografie-im-Praxis-Test_69510670.html. 2014. Web. [Date of access: 12 Mar. 2015]

Choudhury, B., and S. Chandran. “A Survey of Image-Based Relighting Techniques.” *Proceedings of the First International Conference on Computer Graphics Theory and Applications*. Setubal, Portugal: INSTICC-INST SYST TECHNOLOGIES INFORMATION CONTROL & COMMUNICATION, 2006. 176–183. *dspace.library.iitb.ac.in*. Web. [Date of access: 26 Apr. 2015].

Cioni, Michael, and Ian Vertovec. *Dragon Tattoo 4K DI Workflow Overview.* URL: http://michaelcioni.tumblr.com/post/14267159477/dragon-tattoo-4k-di-workflow-overview. Interview by Debra Kaufman. 2011. Web. [Date of access: 21 Jan. 2015]

Citizen Kane. Dir. Orson Welles. RKO Radio Pictures, 1941. Film.

Clark, Barbara, and Susan J. Spohr. *Guide to Postproduction for TV and Film - Managing the Process*. 2nd ed. Focal Press, 2002. Print.

Clavadetscher, Charlie. “Lighting Data.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 150–158. Print.

Collins, Grant Vincent. “Entering the Era of Post Production – The Role Visual Effects Will Pay in Documentary.” Master Thesis, Queensland University. 2009: n. pag. Print.

Conelly, Robert, David Court, and Allan Cameron. “Embracing Innovation: A New Methodology for Feature Film Production in Australia.” AFTRS, 2008: n. pag. Print.

Dabney, B. “What Does the Future Hold for the Visual Effects Industry?” URL: http://www.gnomonschool.com/blog/career/what-does-the-future-hold-for-the-visual-effects-industry. *Gnomon School of Visual Effects*. N.p., Oct. 2013. Web. [Date of access: 16 Feb. 2015].

Debevec, Paul et al. “Acquiring the Reflectance Field of a Human Face.” *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., 2000. 145–156. Print.

DiFrancesco, David. "SUPER LIGHT-FIELD LENS WITH DOUBLET LENSLET ARRAY ELEMENT." URL: http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.html&r=1&f=G&l=50&s1=%2220110249341%22.PGNR.&OS=DN/20110249341&RS=DN/20110249341 US Patent & Trademark Office 2011. Web. [Date of access: 14 Mar. 2015]

Dornes, Martin. *Der Kompetente Säugling: Die Präverbale Entwicklung Des Menschen*. Originalausg. Frankfurt am Main: Fischer Taschenbuch Verlag, 1993. Print.

Doyle, Paul, Philip Taylor, and Peter Zion. “What Is Fabric Engine? Fabric Software Whitepaper.” URL: http://fabricengine.com/wp-content/uploads/2012/04/WhatIsFabric1.pdf. *Fabric Software.* 2014: n. pag. Web. [Date of access: 29 Apr. 2015]

Duggal, Sheena. Interview with John Montgomery. Fxphd course *vfx220 Feature Film Vfx Supervision*. 2012. Video.

Dunlop, Renee. *Production Pipeline Fundamentals for Film and Game*. Burlington, MA: Focal Press, 2014. Print.

Eberhardt, Bernhard. Interview during expert meeting. 5. Nov. 2014

Ebert, Christof. *Systematisches Requirements Engineering : Anforderungen Ermitteln, Spezifizieren, Analysieren Und Verwalten*. 4th ed. Heidelberg: dpunkt-Verlag, 2012. Print.

Edwards, Graham. “L Is for Lidar.” URL: http://www.cinefex.com/blog/. *Cinefex.com*. N.p., Mar. 2015. Web. [Date of access: 3 Apr. 2015].

Einarsson, Per et al. “Relighting Human Locomotion with Flowed Reflectance Fields.” *Proceedings of the 17th Eurographics Conference on Rendering Techniques*. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2006. 183–194. Web. EGSR ’06. [Date of access: 8 March 2015]

Engelhardt, Andreas. “Stereo 3D Ohne Kopfschmerzen.” 2012: n. pag. Print.

Fabio Pellacini et al. “Lpics: A Hybrid Hardware-Accelerated Relighting Engine for Computer Cinematography.” *ACM Trans. Graph.* 24 (2005): 464–470. Web. [Date of access: 19 Apr. 2015]

Failies, Ian. “Going Full Vamp: Dracula Untold.” URL: http://www.fxguide.com/featured/going-full-vamp-dracula-untold/. *fxguide*. N.p., 13 Oct. 2014. Web. [Date of access: 21 Jan. 2015].

Fraunhofer IIS. "What is Light-Field Technology." URL: http://www.iis.fraunhofer.de/en/ff/bsy/tech/lichtfeld.html. 2014. Web. [Date of access: 10 Apr. 2015]

Ganbar, Ron. Nuke 101 - Professional Compositing and Visual Effects. Berkeley, CA: Peachpit Press, 2011. Print

Geitz, Bob. “Vector Geometry for Computer Graphics.” URL: https://www.cs.oberlin.edu/~bob/cs357.08/VectorGeometry/VectorGeometry.pdf. *Oberlin College, Computer Science.* 2007. Web. [Date of access: 21 Jan. 2015].

Georgiev, Todor, and David Salesin. “Future of Adobe Photoshop.” NVIDIA GPU Conference. San Jose, California. 2010. Conference Presentation.

Gershbein, Reid, and Pat Hanrahan. “A Fast Relighting Engine for Interactive Cinematic Lighting Design.” Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000. 353–358. Web. SIGGRAPH ’00. [Date of access: 19 Apr. 2015]

Gortler, Steven J. et al. "The Lumigraph". Microsoft Research. 1996. Print.

Gortler, Steven J. et al. “The Lumigraph.” Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New York, NY, USA: ACM, 1996. 43–54. Web. SIGGRAPH ’96.

Goulekas, Karen. “On-Set Data Acquisition.” The VES Handbook of Visual Effects. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 127–139. Print.

Green, Robin. “Spherical Harmonic Lighting: The Gritty Details.” Sony Computer Entertainment America. 16 Jan. 2003: n. pag. Print.

Green, Tim, Matt Hoesterey, Hannes Ricklefs, Mark Streatfield, Steve Theodore Ben Carter, et al. “Asset Management.” *Production Pipeline Fundamentals For Film and Games*. Ed. Renee Dunlop. 1st ed. Burlington, MA: Focal Press, 2014. 211–233. Print.

Green, Tim, Matt Hoesterey, Hannes Ricklefs, Mark Streatfield, Steve Theodore, Laurent M. Abecassis, Ben Cole, et al. “Introduction.” *Production Pipeline Fundamentals For Film and Games*. Ed. Renee Dunlop.1st ed. Burlington, MA: Focal Press, 2014. 1–20. Print.

Green, Tim, Matt Hoesterey, Hannes Ricklefs, Mark Streatfield, Steve Theodore, Laurent M. Abecassis, Rob Blau, et al. “The Basic Functionality of a Pipeline.” *Production Pipeline Fundamentals For Film and Games*. Ed. Renee Dunlop. 1st ed. Burlington, MA: Focal Press, 2014. 96–124. Print.

Green, Tim, Matt Hoesterey, Hannes Ricklefs, Mark Streatfield, Steve Theodore, Laurent M. Abecassis, Holly Newman, et al. “The Stages of Production.” *Production Pipeline Fundamentals For Film and Games*. Ed. Renee Dunlop. 1st ed. Burlington, MA: Focal Press, 2014. 21–39. Print.

Greengard, Samuel. “Computational Photography Comes into Focus.” *Communications of the ACM* 57.2 (2014): 19–21. *EBSCOhost*. Web. [Date of access: 21 Jan. 2015].

Griggs, Mary Beth. "The Lytro Light-Field Camera: How It works." URL: http://www.popularmechanics.com/technology/gadgets/reviews/a6900/the-lytro-light-field-camera-how-it-works/. 2011. Web. [Date of access: 05 Mar. 2015]

Haan, M. “Direct-to-Indirect Transfer for Cinematic Relighting.” *ACM Trans. Graph.* 25 (2006): 1089–1097. Web. [Date of access: 19 Apr. 2015]

Hasinoff, Samuel W. et al. "Search-and-Replace Editing for Personal Photo Collections." Massachusetts Institute of Technology, 2010. Print.

Heckenberg, Daniel et al. “Deep Compositing.” 2010: n. pag. Print.

Heinen, Patrick. “Deep Image Compositing in a Modern Visual Effects Pipeline.” URL: http://opus.bsz-bw.de/hdms/volltexte/2014/1869. 2013. Web. [Date of access: 6 Oct. 2014]

Hillman, Peter, and Matt Welford. “Storing Multi-View Images in OpenEXR Files.” URL: http://openexr.com/MultiViewOpenEXR.pdf. *OpenEXR.com.* 29 Nov. 2007. Web. [Date of acess: 12 Feb. 2015].

Hillman, Peter. “The Theory of OpenEXR Deep Samples.” URL: http://openexr.com/TheoryDeepPixels.pdf. *OpenEXR*.com. Nov. 2013: n. pag. Web. [Date of acess: 12 Feb. 2015].

Hirsch, Matthew, Wetzstein, Gordon., Raskar, Ramesh, "A compressive light field projection system" ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2014, 2014. Print

Horn, Daniel Reiter, and Billy Chen. “LightShop: Interactive Light Field Manipulation and Rendering.” *Proceedings of the 2007 Symposium on Interactive 3D Graphics and Games*. New York, NY, USA: ACM, 2007. 121–128. Web. I3D ’07. [Date of access 5 Feb. 2015]

Huang, F. et al., "Eyeglasses-free Display: Towards Correcting Visual Aberrations with Computational Light Field Displays" ACM Transaction on Graphics, 2014. Print.

Hullfish, Steve. *The Art and Technique of Digital Color Correction*. Amsterdam ; Boston: Focal Press/Elsevier, 2008. Print.

Hunter, Ian. “Creating Cinematic Virtual Reality.” FMX 2015. Stuttgart. 2015. Conference Presentation.

InsightMediaTV. "NAB 2014 - Fraunhofer IIS Demonstrates 4x4 Light Field Camera System." URL: https://www.youtube.com/watch?t=88&v=pXPS4yB1xMM 2014. Web. [Date of access: 12 Mar. 2015]

Isaksen, Aaron, Leonard McMillan, and Steven J. Gortler. "Dynamically Reparameterized Light Fields." *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000. 297–306. Web. SIGGRAPH '00. [Date of access 27 Jan. 2015]

Jago, Maxim. "Adobe Premiere Pro Help - Multi-Camera Editing Workflow." URL: https://helpx.adobe.com/premiere-pro/using/create-multi-camera-source-sequence.html. Adobe Systems, 2015. Web. [Date of access: 23 Mar. 2015].

Jarabo, Adrian et al. "How Do People Edit Light Fields?" ACM Trans. Graph. 33.4 2014: 146:1–146:10. Web. [Date of access: 27 Dec. 2014]

Jarabo, Adrian et al. "Main Video - How Do People Edit Light Fields?" URL: http://giga.cps.unizar.es/~ajarabo/pubs/lfeiSIG14/videos/Jarabo2014_main_video.mp4 (SIGGRAPH 2014) (2014) Video. [Date of access: 12 Apr. 2015]

Jarabo, Adrian et al. "Presentation - How Do People Edit Light Fields?" URL: http://giga.cps.unizar.es/~ajarabo/pubs/lfeiSIG14/downloads/Jarabo_sig14_Presentation.zip (SIGGRAPH 2014) (2014) Presentation. [Date of access: 12 Apr. 2015]

Jarabo, Adrian, Belen Masia, and Diego Gutierrez. "Efficient Propagation of Light Field Edits." *Proceedings of SIACG*. Faro, Portugal: N.p., 2011. 75–80. Print.

Jarabo, Adrian. Masia, Belen. Gutierrez, Diego. "Efficient Propagation of Light Field Edits." Universidad de Zaragoz, 2011. Print

Jensen, Henrik Wann et al. "A Practical Model for Subsurface Light Transport." Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York, NY, USA: ACM, 2001. 511–518. Web. SIGGRAPH '01. [Date of access: 12 Apr. 2015]

Jones, A., et al. "Rendering for an Interactive 360º Light Field Display" ACM SIGGRAPH 2007 Papers Proceedings, 2007. Print.

Jones, Andrew et al. "Rendering for an Interactive 360° Light Field Display." URL: http://gl.ict.usc.edu/Research/3DDisplay/. *Rendering for an Interactive 360° Light Field Display*. N.p., 2007. Web. [Date of access 2 Apr. 2015].

Kainz, Florian et al. "Technical Introduction to OpenEXR." URL: http://openexr.com/TechnicalIntroduction.pdf. *OpenEXR.com*. 2013: n. pag. Web. [Date of access: 6 Jan. 2015]

Kainz, Florian, and Piotr Stanczyk. "Image Compression/File Formats for Post-Production." *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 469–482. Print.

Kainz, Florian, and Piotr Stanczyk. "Image Compression/File Formats for Post-Production." *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. URL: http://documents.routledge-interactive.s3.amazonaws.com/9780240825182/Chapter%206.pdf. Web Content. Oxford: Elsevier/Focal Press, 2014. 469–482. Web. [Date of access: 18 Jan. 2015]

Kamiguri, Takashi. "Toshiba putting focus on taking misfocusing out of photos." URL: http://ajw.asahi.com/article/business/AJ201212270054. 2012. Web. [Date of access: 19 Mar. 2015]

Kanumuri, S et al. "Temporal Flicker Reduction and Denoising in Video Using Sparse Directional Transforms." Vol. 7073. N.p., 2008. *CiteULike*. Web. [Date of access: 30 May 2015].

Karsch, Kevin et al. "Rendering Synthetic Objects into Legacy Photographs." Proceedings of the 2011 SIGGRAPH Asia Conference. New York, NY, USA: ACM, 2011. 157:1–157:12. Web. SA '11. [Date of access: 29 Apr. 2015]

Kass, M., Lefohn, A., Owens, J. "Interactive depth of field using simulated diffusion on a GPU." Tech. rep., Pixar. 2006. Print.

Kaufman, Debra. "Shoot It on the Virtual Backlot." URL: http://www.studiodaily.com/2006/05/shoot-it-on-the-virtual-backlot/. 2006. Web. [Date of access: 21 Apr. 2015]

Kim, Changil, et al., "Multi-Perspective Stereoscopy from Light Fields" Proceedings of ACM SIGGRAPH Asia 2011, 2011. Print.

Kirchhoff, Sönke. Interview during expert meeting. 5. Nov. 2014

Knoll, John. “Cg Prosthetics and Actor Enhancements.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2014. 694–697. Print.

Knopp, Thomas. “Virtual Production – Methods, Guidelines and Scenarios.” URL: http://www.dreamspaceproject.eu/Documents. Filmakademie Baden-Württemberg, Animationsinstitut, Apr. 2014. Web. [Date of access: 22 Feb. 2015]

Krause, Julian. “Ganz neue Dimensionen.” *Digital Production* June 2013: 42–49. Print.

Kubacki, Daniel. “Image Relighting Using Depth.” URL: http://www.ifp.illinois.edu/~kubacki1/Doc/ECE547FinalProjectReport.pdf. 2009. Web. [Date of access: 6 Jan. 2015].

Kurachi, Noriko. *The Magic of Computer Graphics*. Boca Raton: A K Peters/CRC Press, 2007. Print.

Kurke, Tom. “Light Field Cameras for 3D Imaging and Reconstruction | Thoughts from Tom Kurke.” URL: http://3dsolver.com/light-field-cameras-for-3d-imaging/. *3D Solver*. N.p., n.d. Web. [Date of access: 26 Jan. 2015].

Landau, Jon. Professional Spotlight: Jon Landau at FMX 2014. URL: http://www.awntv.com/videos/professional-spotlight-jon-landau-fmx-2014-part-1. *Animation World Network. May* 2014. Video. [Date of access: 20 Jan. 2015]

Langer, Michael. “Camera Model.” URL: http://www.cim.mcgill.ca/~langer/558/2009/558.html. COMP 558 lecture 4. CIM McGill, Sept. 2010: n. pag. Web. [Date of access: 15 May 2015]

Lanman, D., Luebke, D. "Near-Eye Light Field Displays" NVIDIA Research, 2013. Print.

Lanman, Douglas et al. “Polarization Fields: Dynamic Light Field Display Using Multi-Layer LCDs.” *ACM SIGGRAPH Asia 2011. Transactions on Graphics 30(6*. N.p., 2011. Print.

Lanman, Douglas, and David Luebke. “Near-Eye Light Field Displays.” ACM Trans. Graph. 32.6 (2013): 220:1–220:10. Web.

Lanman, Douglas, et al. "Polarization fields: dynamic light field display using multi-layer LCDs! Proceedings of the 2011 SIGGRAPH Asia Conference Article No. 186 , 2011. Print.

Lasky, Alan. “Lidar Scanning and Acquisition.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 140–145. Print.

Leberecht, Scott. *Life After Pie*. N.p., 2014. Film.

Lee S., Eisemann E., Seidel H.-P.: "Depth-of-field rendering with multiview synthesis." ACM Trans. 2009. Print.

Levoy, Marc et al. “Light Field Microscopy.” *ACM SIGGRAPH 2006 Papers*. New York, NY, USA: ACM, 2006. 924–934. Web. SIGGRAPH ’06.

Levoy, Marc et al. “Light Fields and Computational Imaging.” *IEEE Computer Society* 018-9162/06 (2006): 46–55. Print.

Levoy, Marc, and Pat Hanrahan. “Light Field Rendering.” *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. New York, NY, USA: ACM, 1996. 31–42. Web. SIGGRAPH ’96. [Date of access: 30 Nov. 2015]

Levoy, Marc, and Sean Anderson. “Unwrapping and Visualizing Cuneiform Tablets.” URL: http://graphics.stanford.edu/papers/cuneiform/. *IEEE Computer-Graphics and Applications* 22 (2002): 82–88. Web. [Date of access: 5 Feb. 2015].

Levoy, Marc. "Stanford Spherical Gantry." URL: http://graphics.stanford.edu/projects/gantry/. 2002. Web. (Date of access: 18 Feb. 2015).

Levoy, Marc. “Light Fields and Computational Imaging.” Stanford University, 2006. Print.

LightRay Visual Effects. "What VFX can do?-Part 2: Environments." URL: https://vfxforfilm.wordpress.com/2012/09/23/vfx_environments/. 2012. Web. [Date of access: 14 Apr. 2015]

Lippmann, G. "Epreuves réversibles. Photographies intégrales." Comptes Rendus de l'Académie des Sciences, 1908. 446-451. Print.

Livingstone, M, and D Hubel. Vision and Art: The Biology of Seeing. N.Y.: Harry N. Abrams, 2002. Print.

Liwei Guo, O. C. Au. "Temporal Video Denoising Based on Multihypothesis Motion Compensation." URL:http://vc.cs.nthu.edu.tw/home/paper/codfiles/cmhung/201110130958/Temporal%20Video%20Denoising%20Based%20on%20Multihypothesis%20Motion%20Compensation.pdf. *Circuits and Systems for Video Technology, IEEE Transactions on* 10 (2007): 1423 – 1429. Web. [Date of access: 23 May 2015]

Lokovic, Tom, and Eric Veach. "Deep Shadow Maps." URL: https://graphics.stanford.edu/papers/deepshadows/deepshad.pdf. *Conference: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* 2000: 385–392. Web. [Date of access: 3 May 2015]

Lukk, Howard. "The Integration of Computational Cinematography into the Workflow." Cinec Workshop Paradigm Shift in Media Production. München. 2014. Conference Presentation.

Lytro. "Desktop 4 - Animate." URL: https://support.lytro.com/hc/en-us/articles/202612244 Lytro Inc., 2014. Web. [Date of access: 10 May 2015]

Lytro. "Desktop 4 - Exporting." URL: https://support.lytro.com/hc/en-us/articles/200865460 Lytro Inc., 2014. Web. [Date of access: 10 May 2015]

Lytro. "Desktop 4 - Using Libraries to manage pictures." URL: https://support.lytro.com/hc/en-us/articles/202208894 Lytro Inc., 2014. Web. [Date of access: 10 May 2015]

Lytro. "Desktop 4 – Adjust" URL: https://support.lytro.com/hc/en-us/articles/202208994 Lytro Inc., 2014. Web. [Date of access: 10 May 2015]

Lytro. "Desktop 4 – The Info Panel" URL: https://support.lytro.com/hc/en-us/articles/202414820 Lytro Inc., 2014. [Date of access: 10 May 2015]

Lytro. "LYTRO ILLUM." URL: https://www.lytro.com. N.p., n.d. Web. [Date of access: 4 Feb. 2015].

Maeno, Kazuki. et al. "Light Field Distortion Feature for Transparent Object Recognition" CPVR, 2013. Print.

Magnor, Marcus and Girod, Bernd. "Progressive Compression and Rendering of Light Fields." University of Erlangen-Nuremberg, 2000. Print

Maizbender, T., D. Gelb, and H. Wolters. "Polynomial Texture Maps." Proceedings of SIGGRAPH 2001. N.p., 2001. 519–528. Print.

Mankins, John C. "Technology Readiness Levels." URL: http://www.hq.nasa.gov/office/codeq/trl/trl.pdf. Apr. 1995. Web. [Date of access 11 Dec. 2014].

Masia, Belen, Adrian Jarabo, and Diego Gutierrez. "Favored Workflows in Light Field Editing." Universidad de Zaragoza, 2014: n. pag. Print.

McAuliffe, Geoff. "The Digital Eye: VFX Supervision in Commercial Production." URL: http://www.awn.com/vfxworld/digital-eye-vfx-supervision-commercial-production. *Animation World Network*. N.p., Feb. 2007. Web. [Date o access: 27 May 2015].

McDowell, Alex, Mona Hamdy, and Uli Hanisch. "Collaboration in Creation: Making Worlds." FMX 2015. Stuttgart. 2015. Conference Presentation.

McDowell, Alex. Professional Spotlight: Alex McDowell at FMX 2014. URL: http://www.awntv.com/videos/professional-spotlight-alex-mcdowell-fmx-2014-part-1. *Animation World Network*. 2014. Video. [Date of access: 10 Jan. 2015]

McGraw, Dan. "Interlude: Metadata." *Production Pipeline Fundamentals For Film and Games*. Ed. Renee Dunlop. 1st ed. Burlington, MA: Focal Press, 2014. 206–210. Print.

Mendiburu, Bernard. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Oxford: Focal Press, 2009. Print.

Meneveaux, Daniel, and Alain Fournier. "Reshading Light Fields." (2002): n. pag. Print.

Montgomery, John. “NVIDIA GPU Tech Conference: Report 2.” URL: http://www.fxguide.com/quicktakes/nvidia-gpu-tech-conference-report-2/. *Fxguide,* fxguide, Mar. 2014. Web. [Date of access: 2 Apr. 2015].

Moura, Gabe. “SHOT SIZES: Telling What They See.” URL: http://www.elementsofcinema.com/cinematography/framing-shot-sizes/. *Elements of cinema.* N.p., 2014. Web. [Date of access: 28 May 2015].

Murray, Janet Horowitz. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace.* [Repr.]. Cambridge, Mass: MIT Press, 1998. Print.

Nahmias, Jean-Daniels. “Interactive GPU Ray Tracing for Physical Based Lighting Preview at Pixar.” FMX 2014. Stuttgart. 2014. Conference Presentation.

Näther, Markus. “Immer Mehr Kinos Setzen Auf Die Dritte Dimension.” *Digital Production* Aug. 2009: 122–125. Print.

Nehab, Diego et al. “Efficiently Combining Positions and Normals for Precise 3D Geometry.” *ACM SIGGRAPH 2005 Papers*. New York, NY, USA: ACM, 2005. 536–543. Web. SIGGRAPH ’05. [Date of access: 30 Apr. 2015]

Neuhaus, Günter. Interview during expert meeting. 5. Nov. 2014

Ng, Ren. “Digital Light Field Photography.” PhD dissertation, Stanford University, 2006. Print.

Nolf, Marcus. "Adobe LichtFeld Kamera Prototypen." URL: http://lightfield-forum.com/lichtfeld-kamera-prototypen/adobe-lichtfeld-kamera-prototypen/. Web. [Date of access: 12 Mar. 2015]

Nolf, Marcus. "CAFADIS LichtFeld Objektiv Prototyp." URL: http://lightfield-forum.com/light-field-camera-prototypes/cafadis-lightfield-lens-prototype/. Web. [Date of access: 12 Mar. 2015]

Nolf, Marcus. "LichtFeld Kamera Prototypen." URL: http://lightfield-forum.com/lichtfeld-kamera-prototypen/. Web. [Date of access: 12 Mar. 2015]

Nolf, Marcus. "Pelican Imaging Array-Kamera: Lichtfeld-Modul für Smartphones." URL: http://lightfield-forum.com/lichtfeld-kamera-prototypen/pelican-imaging-array-kamera-lichtfeld-modul-fur-smartphones/. Web. [Date of access: 12 Mar. 2015]

Nolf, Marcus. "Raytrix: Neue LichtFeld Kamera mit Bildauflösung von 25 % der Sensor-Auflösung." URL: http://lightfield-forum.com/2012/05/raytrix-neue-lichtfeld-kamera-mit-bildauflosung-von-25-der-sensor-auflosung/. 2012. Web. [Date of access: 12 Mar. 2015]

Nolf, Marcus. "Toshiba LightField Camera Module." URL: http://lightfield-forum.com/light-field-camera-prototypes/toshiba-lightfield-camera-module/. Web. [Date of access: 12 Mar. 2015]

Okun, Jeffrey A., and Susan Zwerman. *The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures*. Oxford: Elsevier/Focal Press, 2014. Print.

Okun, Jeffrey A., and Susan Zwerman. *The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures*. Oxford: Elsevier/Focal Press, 2010. Print.

OTOY. “OTOY Demonstrates First Ever Light Field Capture of a Real World Environment, Producing Photorealistic, Navigable Virtual Reality Experience.” URL: http://www.reuters.com/article/2015/05/07/otoy-fmx-lightfield-idUSnPn2wX1hz+88+PRN20150507. Reuters 7 May 2015. Reuters. Web. Press release. [Date of access: 28 May 2015].

Owano, Nancy. "Toshiba smartphone camera sensor has eye on future." URL: http://phys.org/news/2012-12-toshiba-smartphone-camera-sensor-eye.html. Phys.org. 2012. Web. [Date of access: 14 Mar. 2015]

Perez, Victor. “Nuke Advanced Relighting and Retexturing”. URL: https://cmivfx.com/videos/566-Nuke-Advanced-Relighting-and-Retexturing. Video lecture. cmiVFX, cmiVFX, 2014. Web.[Date of access: 10 Nov. 2014]

Perwaß, Christian and Wietzke, Lennart. "Single Lens 3D-Camera with Extended Depth-of-Field." Raytrix GmbH, 2012. Print.

Phong, Bui Tuong. “Illumination for Computer Generated Pictures.” Commun. ACM 18.6 (1975): 311–317. Web. [Date of access: 10 May 2015]

Porter, Thomas, and Tom Duff. “Compositing Digital Images.” Computer Graphics 18.3 1984: 253–259. Print.

Prince, Stephen. Digital Visual Effects in Cinema: The Seduction of Reality. New Brunswick, NJ: Rutgers University Press, 2011. Print.

Radiohead. “Radiohead - House of Cards - YouTube.” Online video clip. *Youtube.* Youtube, 07.2008. Web. [Date of access: 6 Mar. 2015].

Rausch, Oliver. *Gestalten mit Licht und Schatten: vom Porträt bis zur Landschaft.* München: Pearson, 2013. Print.

Raytrix. "R5 Light Field Camera Systems." Raytrix GmbH, 2014. Print.

Raytrix. “Raytrix Light Field Camera - Nvidia GTC 2012.” URL: http://www.raytrix.de/tl_files/downloads/Raytrix_Slides.pdf. Nvidia GTC. San Jose, CA. 2012: n. pag. Web. [Date of access: 4.Feb. 2015]

Ricklefs, Hannes. “Pipeline for Feature Film VFX.” FMX 2015. Stuttgart. 2015. Conference Presentation.

Ring of Fire Advanced Media Team. “Using Visual Effects in Commercials: A Guide for Success.” *Animation World Magazine* 4.3.07.1999, 1999: n. pag. Print.

Rizzolatti, Giacomo, and Laila Craighero. “The Mirror-Neuron System.” URL: http://dx.doi.org/10.1146/annurev.neuro.27.070203.144230. *Annual Review of Neuroscience* 27.1 (2004): 169–192. *Annual Reviews.* Web. [Date of access: 28 May 2015].

Robles, Patricio. “Lytro Releases Dev Kit for Light Field Camera Technology.” URL: http://www.programmableweb.com/news/lytro-releases-dev-kit-light-field-camera-technology/2014/11/11. *ProgrammableWeb.* N.p., 11.2014. Web. [Date of access: 23 May 2015].

Rodríguez-Ramos, Luis Fernando. "Restoring lost resolution of plenoptic images." URL: http://spie.org/x113095.xml 2015. Web. [Date of access: 14 Mar. 2015]

Rogmans, Sammy. “SCENE Tools for Next Generation Post-Production.” Cinec Workshop Paradigm Shift in Media Production. München. 2014.

Romanek, Neal. “VFX Workflows: Q&A With Simon Robinson Of The Foundry.” URL: http://www.c2meworld.com/creation/vfx-workflows-qa-with-simon-robinson-of-the-foundry/. Interview. *c2meworld.com.* N.p., 27 June 2014. Web. [Date of access: 21 Jan. 2015].

Root, John. “The Future of Immersive Entertainment.” FMX 2015. Stuttgart. 2015. Keynote.

Rupp, Chris. *Requirements-Engineering und -Management: Aus der Praxis von klassisch bis agil.* 6th ed. München: Hanser, 2014. Print.

Saam, Johannes. “Creative Relighting in Compositing Based on Z Depth.” National center of computer animation Bournemouth University, 2010: n. pag. Print.

Sato, Yoichi, Mark D. Wheeler, and Katsushi Ikeuchi. “Object Shape and Reflectance Modeling from Observa- Tion.” *Siggraph 1997.* Los Angeles: N.p., 1997. Print.

Schreer, Oliver. *Stereoanalyse und Bildsynthese: mit 6 Tabellen.* Berlin [u.a.]: Springer, 2005. Print.

Science and Technology Council. “The Academy Color Encoding System.” URL: http://www.oscars.org/sites/default/files/acesoverview.pdf. 2013. Web. [Date of access: 18 May 2015]

Sciolette, Michele et al. “Multi Layer Exr Files Performance.” URL: http://community.thefoundry.co.uk/discussion/topic.aspx?f=189&t=96403. *The Foundry Community* Apr. 2009. Web. [Date of acces: 23 May 2015]

Seitz, Steven M. and Kutulakos, Kiriakos N. "Plenoptic Image Editing." University of Washington and University of Toronto, 2002. Print

Seitz, Steven M., and Kiriakos N. Kutulakos. “Plenoptic Image Editing.” *Proc. ICCV 98.* N.p., 1998. Print.

Seymour, Mike “The Art of Deep Compositing.” URL: http://www.fxguide.com/featured/the-science-of-spherical-harmonics-at-weta-digital/. *fxguide.* Fxguide 27 Feb. 2014. Web. [Date of access: 3 May 2015].

Seymour, Mike. “Art of Optical Flow | Fxguide.” URL: http://www.fxguide.com/featured/art_of_optical_flow/. *Fxguide,* Fxguide, Feb. 2006. Web. [Date of access: 29 May 2015].

Seymour, Mike. “Katana in Production: Changing the Structure of the Workflow.” URL: http://www.fxguide.com/featured/katana-in-production-changing-the-structure-of-the-workflow/. *fxguide*. Fxguide, July 2012. Web. [Date of Access: 28 May 2015].

Seymour, Mike. “Manuka: Weta Digital’s New Renderer.” URL: http://www.fxguide.com/featured/manuka-weta-digitals-new-renderer/. *fxguide*. Fxguide, 8 June 2014. Web. [Date of access: 19 Apr. 2015].

Seymour, Mike. “The Art of Digital Color.” URL: http://www.fxguide.com/featured/the-art-of-digital-color/. fxguide. Fxguide, Aug. 2011. Web. [Date of access: 17 May 2015].

Seymour, Mike. “The Science of Spherical Harmonics at Weta Digital.” URL: http://www.fxguide.com/featured/the-science-of-spherical-harmonics-at-weta-digital/. fxguide. Fxguide, 25 Apr. 2013. Web. [Date of access: 19 Apr. 2015].

Seymour, Mike. “The State of Rendering – Part 1.” URL: http://www.fxguide.com/featured/the-state-of-rendering/. fxguide. Fxguide, July 2013. Web. [Date of access: 8 Mar. 2015].

Shannon, C.E. "A Mathematical Theory of Communication. Bell System." Technical Journal, 1948. Print.

Shih, Gerry. "Lytro unveils a more sophisticated 'light-field' camera." URL: http://in.reuters.com/article/2014/04/23/us-lytro-camera-idINKBN0D903A20140423. San Francisco, 2014. Web. [Date of access: 13 Apr. 2015]

Simonite, MIT Technology Review “How Magic Leap’s Augmented Reality Works“ URL: http://www.technologyreview.com/news/532001/how-magic-leaps-augmented-reality-works/ , MIT Technoloy Review, 2014. Web. [Date of access: 27 May 2015]

Sinko, Adam. "Lytro." URL: http://www.lichtfeld-kamera.net/p/lytro.html. Web. [Date of access: 13 Apr. 2015]

Sinko, Adam. "Raytrix." URL: http://www.lichtfeld-kamera.net/p/raytrix.html. Web. [Date of access: 13 Apr. 2015]

Siragusano, Daniele. Interview during expert meeting. 5. Nov. 2014

Smirnov, Sergey, Atanas Gotchev, and Karen Egiazarian. “Methods for Depth-Map Filtering in View-plus-Depth 3D Video Representation.” URL: http://dx.doi.org/10.1186/1687-6180-2012-25. *EURASIP Journal on Advances in Signal Processing* 2012.1 (2012): n. pag. Web. [Date of access: 26 Apr. 2015]

Spears, Derek. “3D Compositing.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 681–690. Print.

Spears, Derek. “Shader Basics.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 672–681. Print.

Squires, Scott. “What Makes a Good Visual Effects Artist?” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2014. 663–690. Print.

Stern, Daniel N. *Die Lebenserfahrung des Säuglings: mit einer neuen Einleitung des Autors*. 10. Aufl. Stuttgart: Klett-Cotta, 2010. Print.

Szeliski, Richard. “Open Platforms for Computational Photography.” *Communications of the ACM* 55.11 (2012): 89–89. *EBSCOhost*. Web. [Date of access: 21 Jan. 2015].

Tao, Michael W. et al. “Depth from Combining Defocus and Correspondence Using Light-Field Cameras.” URL: http://graphics.berkeley.edu/papers/Tao-DFC-2013-12/. *International Conference on Computer Vision (ICCV)*. N.p., 2013. Web. [Date of access 26 Jan. 2015]

The Foundry "Nuke Online Help – ZDefocus." URL: http://help.thefoundry.co.uk/nuke/9.0/Default.html#cshid=ZDefocus2 The Foundry Visionmongers Ltd. Internet. [Date of access: 16 May 2015] The Foundry Ocula "OCULA USER GUIDE VERSION 4.0V2." The Foundry Visionmongers Ltd. 2014. Print.

The Foundry. “Nuke 2009 Masterclass - 2.5D Re-Lighting (Video 2).” Online video clip. *Youtube*. Youtube, 12.2011. Web. [Date of access: 23 Apr. 2015]

Thompson, Anne. "How James Cameron's Innovative New 3D Tech Created Avatar." URL: http://www.popularmechanics.com/culture/movies/a5067/4339455/. 2010. Web. [Date of access: 16 Apr. 2015]

Townsend, Christopher. “3D Stereoscopic Visual Effects.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 427–448. Print.

Tukkiendorf, Jens Ernst. Interview during expert meeting. 5. Nov. 2014

Uriostegui, Hassan. “Point Pass Plotted to Geometry Cloudlet for 3D Surface Relighting in Nuke v1.2.” URL: http://vfxwarrior.com/cloudlight.pdf. *VFX Warrior*. 2012. Web. [Date of access: 22 Apr. 2015].

Vaish, Vaibhav et al. "Synthetic Aperture Focusing using a Shear-Warp Factorization of the Viewing Transform." Stanford University, 2005. Print

Van Hurkman, Alexis. *Color Correction Handbook: Professional Techniques for Video and Cinema*. 2nd edition. San Francisco, California: Peachpit Press, 2014. Print.

Venkataraman, Kartik et al. “PiCam: An Ultra-Thin High Performance Monolithic Camera Array.” 2013: n. pag. Print.

Venkataraman, Kartik. et al. "PiCam: An Ultra-Thin High Performance Monolithic Camera Array." Pelican Imaging Corporation, 2013. Print.

Verma, Chaman Singh, and Mon-Ju Wu. “Photometric Stereo.” URL: http://pages.cs.wisc.edu/~csverma/CS766_09/Stereo/stereo.html. 2009. Web. [Date of access: 1 May 2015]

Visual Effects Society, and Jeffrey A. Okun. “VES Whitepaper - ‘VES Issues ’The State of the Global VFX Industry 2013".” URL: http://www.visualeffectssociety.com/innews/ves-whitepaper-ves-issues-state-global-vfx-industry-2013. Visual Effects Society, 15 Apr. 2014. Web. [Date of access: 20 Jan. 2015].

Walter, Simon. Interview during expert meeting. 5. Nov. 2014

Wang, Jue, and Michael F. Cohen. Image and Video Matting: A Survey. Now Publishers Inc, 2008. Print.

Wang, Lifeng et al. "Light Field Morphing Using 2D Features." IEEE Trans Vis Comput Graph., 2005. Print

Ward, Gregory J. “Measuring and Modeling Anisotropic Reflection.” Proceedings Og Siggraph ’92, Computer Graphics. Vol. 26:2. N.p., 1992. 265–272. Print.

Wenger, Andreas et al. “Performance Relighting and Reflectance Transformation with Time-Multiplexed Illumination.” ACM SIGGRAPH 2005 Papers. New York, NY, USA: ACM, 2005. 756–764. Web. SIGGRAPH ’05. [Date of access: 8 March 2015]

Whitehurst, Andrew. “Digital Lighting.” *The VES Handbook of Visual Effects*. Eds. Jeffrey Okun and Susan Zwerman. Oxford: Elsevier/Focal Press, 2010. 649–672. Print.

Whitehurst, Andrew. “The Visual Effects Pipeline.” URL: http://www.andrew-whitehurst.net/pipeline.html. 2015. Web. [Date of access: 21 Jan. 2015].

Wieland, Georg. Interview during expert meeting. 5. Nov. 2014

Wikipedia contributors. "Lytro." URL: http://en.wikipedia.org/w/index.php?title=Lytro&oldid=655759229. Wikipedia, The Free Encyclopedia. 2015. Web. [Date of access: 16 May 2015]

Wikipedia contributors. “Albedo.” URL: http://en.wikipedia.org/w/index.php?title=Albedo&oldid=655376612. *Wikipedia, the free encyclopedia* 7 Apr. 2015. Web. [Date of access: 19 Apr. 2015].

Wikipedia contributors. “BSD Licenses.” URL: http://en.wikipedia.org/w/index.php?title=BSD_licenses&oldid=662514527. *Wikipedia, the free encyclopedia* 15 May 2015. Web. [Date of access: 17 May 2015].

Wikipedia contributors. “Cinematographer.” URL: http://en.wikipedia.org/w/index.php?title=Cinematographer&oldid=645953919. *Wikipedia, the free encyclopedia* 6 Feb. 2015. Web. [Date of access 21 Feb. 2015].

Wikipedia contributors. “House of Cards (Radiohead Song).” URL: http://en.wikipedia.org/w/index.php?title=House_of_Cards_(Radiohead_song)&oldid=639508522. *Wikipedia, the free encyclopedia,* 24 Dec. 2014. *Wikipedia*. [Date of access: Web. 7 Mar. 2015].

Wikipedia contributors. “Moore’s Law.” *Wikipedia, the free encyclopedia* 3 Feb. 2015. Web. [Date of access 5 Feb. 2015].

Wikipedia contributors. “Polygon Mesh.”
URL: http://en.wikipedia.org/w/index.php?title=Polygon_mesh&oldid=651774087. *Wikipedia, the free encyclopedia* 17 Mar. 2015. Web. [Date of access: 30 May 2015].

Wilburn, Bennett. “High Performance Imaging Using Arrays of Inexpensive Cameras.” URL: http://graphics.stanford.edu/~wilburn/wilburn_thesis.pdf. PhD dissertation, Stanford University, 2004. Web. [Date of access 21 Jan. 2015]

Wilburn, Bennett. “High Performance Imaging Using Large Camera Arrays.” ACM Transactions on Graphics Vol 24. N.p., 2005. 765–776. Print.

Wilburn, Bennett. et al. “The Light Field Video Camera.” Stanford University, 2012. Print.

Woodham, R.J. “Photometric Method for Determining Surface Orientation from Multiple Images.” Optical Engineering 19. Vol. 1. N.p., 1980. 139–144. Print.

Wright, Steve. *Compositing Visual Effects: Essentials for the Aspiring Artist*. Amsterdam; Boston: Elsevier/Focal Press, 2008. Print.

Wright, Steve. *Digital Compositing for Film and Video*. 3rd ed. Amsterdam; Boston: Elsevier/Focal Press, 2010. Print. Focal Press Visual Effects and Animation Series.

Yang, Jason C. et al. “A Real-Time Ditrbitued Light Field Camera.” *Eurographics Workshop on Rendering*. N.p., 2002. 10. Print.

Yang, Ruigang et al. “Toward the Light Field Display: Autostereoscopic Rendering via a Cluster of Projectors.” *IEEE transactions on visualization and computer graphics* 14.1 2008: 84–96. *NCBI PubMed*. Web. [Date of access: 15 Feb. 2015]

Yu, Zhan. et al. "Dynamic Depth of Field on Live Video Streams: A Stereo Solution." University of Delaware, 2011. Print.

Zhang, Fan, Yuzhen Niu, and Feng Liu. “Making Stereo Photo Cropping Easy.” 2013 IEEE International Conference on Multimedia and Expo (ICME). N.p., 2013. 1–6. IEEE Xplore. Web. [Date of access: 20 Apr. 2015]

Zhang, Michael. "The First Plenoptic Camera on the Market." URL: http://petapixel.com/2010/09/23/the-first-plenoptic-camera-on-the-market/. 2010. Web. [Date of access: 16 Mar. 2015]

Zhang, Zhunping et al. "Feature-Based Light Field Morphing." Microsoft Research Asia, 2002. Print.

Zhou, T., Chen, J.,AND Pullen, M. "Accurate depth of field simulation in real time." Computer Graphics Forum 2007. Print.

Zilly et al. “IIFOP 2.0 Projektskizze Technologiestudie”. Erlangen: Fraunhofer IIS, 2014. Print.

Zilly, F., Riechert, C., Müller, M. and Kauff, P. "Multi-Camera Rectification using Linearized Trifocal Tensor" Fraunhofer Institut for Telecommunications, Berlin and Technische Universität Berlin, Germany. International Conference on Pattern Recognition (ICPR 2012) 2012. Print.

Zilly, Frederik et al. “Light-Field Acquisition System Allowing Camera Viewpoint and Depth of Field Compositing in Post-Production.” Amsterdam: N.p., 2013. Print.

Zitnick, C. Lawrence, and Takeo Kanade. “A Cooperative Algorithm for Stereo Matching and Occlusion Detection.” URL: http://research.microsoft.com/en-us/um/people/larryz/zitnick_charles_1999_1.pdf. Carnegie Mello University, Pittsburgh. Oct. 1999: n. pag. Web. [Date of access: 18 May 2015]

Zwerman, Susan, and Charles Finance. *The Visual Effects Producer - Understanding the Art and Business of VFX*. Oxford: Focal Press Elsevier, 2010. Print.

Appendix C – List of Figures

Appendix D – List of Experts

The following list quickly introduces the external attendees of the expert meeting held at the HDM in November 2014.

Sönke Kirchhof, producer and stereographer from Berlin has a broad expertise in the field of immersion, VFX and stereoscopic filmmaking. He is the founder and CEO of realifefilm (http://www.realifefilm.com).

Daniele Siragusano is currently working as a workflow consultant at FilmLight Ltd. Munich and London (http://www.filmlight.ltd.uk/). As stereographer, postproduction supervisor and later head of technology at CinePostproduction GmbH, he played a key role in the development of a stereoscopic workflow for CinePostproduction.

Georg Wieland, Global Software Manager at MPC (http://www.moving-picture.com), London, is responsible for the development of MPCs global pipeline and proprietary VFX tools. He has a strong background in producing and managing international film productions.

Jens Ernst Tukiendorf is the DP supervisor and visual consultant at UFA Serial Drama GmbH in Potsdam (http://www.ufa-serialdrama.de) as well as member of the 'Wold Wide Competence Center' of the Fremantle Group. He has lots of experience with on-set work and the connected creative challenges.

Günter Neuhaus, branch manager of the digital film department of Ludwig Rental Berlin, is in close contact with his clients and knows the demands of present media productions.

Ron Ganbar worked in the VFX industry since 1995 and gained experience from commercial, animation, motion-graphics and feature film work. He is a VFX supervisor, digital compositor, consultant and trainer, now based at Tel Aviv.

Two more project partners could not attend the meeting in person. These were **Patrick Heinen**, a compositing artist at the Mill, Los Angeles who has a strong background in deep compositing, and **Jan Fröhlich**, former Technical Director at CinePostproduction GmbH, Germany and at the time of writing PhD student at HDM with a research interest in color management, high dynamic range imaging and gamut mapping.

Furthermore, Prof. Katja Schmid, Prof. Stefan Grandinetti, Prof. Dr. Bernhard Eberhardt, Simon Walter, Michael Kirschenlohr and Peter Ruhrmann from HDM Stuttgart attended the meeting. Frederik Zilly, Michael Ziegler, Joachim Keinert and Lukas Kolhagen represented Fraunhofer IIS. The group of students consisted of Clemens Helmchen, Johannes Hölz, Larissa Kurtz, Marlene Fritzsche, Moritz Wetzig, Sebastian Zeeden, Sergej Feininger, Stefan Müller and Andreas Engelhardt.

Appendix E – Dataset

Attached is a DVD containing footage (RGB & 3D Data), elements, HDRIs, tools, meeting records, a making of, a breakdown, the fictional commercial clip and a Lego stop motion short.

The folder *DATA* contains following subfolders:

- Footage
  - 4 sets of footage captured with our light field array and additional 3D data
- Set HDRis
  - 3 HDRis from our studio set
- Teapot
  - Projectdata of a teapot rendering used during our testing
- Tools
  - Camera specifications
  - Directory helper tool
  - Light Field Handler tool
  - PointTransform tool

The folder *meeting_records* contains:

- Questions we asked the experts during the meeting
- Protocols of each group of interviewees

The folder *Video* contains:

- Coming Home – Commercial
- Coming Home – Breakdown
- Coming Home – Making Of
- Lego Stopmotion Viewrendering