

Attenuation-based Light Field Displays

Bachelorarbeit

der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Adrian Wälchli

2016

Leiter der Arbeit:
Prof. Dr. Matthias Zwicker
Institut für Informatik und angewandte Mathematik

Abstract

Abstract goes here

Acknowledgements

Special thanks goes to Prof. Dr. Matthias Zwicker for his supervision and support of this thesis and for giving me the opportunity to work together with the Computer Graphics Group at the University of Berne. A big thank you to secondary advisor Siavash Bigdeli for his time and effort to help me throughout the project. Lastly, I would like to thank my family, friends and colleagues for the great time I had with them during the time of my bachelor studies.

Contents

1	Introduction	1
1.1	Light Fields	1
1.2	Stereoscopic Displays	2
1.3	3D Displays	2
1.4	Related Work	3
2	Capturing a Light Field	4
2.1	The Plenoptic Function and the Light Field	4
2.2	Light Field Acquisition	5
2.3	Visualization	7
2.4	The Plenoptic Camera	8
3	Light Field Tomography	9
3.1	A Model for Light Attenuation	9
3.2	Discrete Attenuation Layers	11
3.3	Ray Casting	11
3.4	Iterative Reconstruction	13
4	Spectral Analysis	14
4.1	Definitions	14
4.2	Spectral Support of Light Fields	15
4.3	Spectral Support of Layered 3D Displays	16
4.4	The Fourier Slice Theorem	17
5	Implementation and Assessment	19
5.1	Oversampling	19
5.2	Baseline Scaling and Back Projection	19
5.3	Attenuator Tiling and Blending	19
5.4	Benefits and Limitations	20
A	Appendix	21
A.1	Backlight Fabrication	21
	List of Tables	23
	List of Figures	23
	Bibliography	24

Chapter 1

Introduction

Over the last few years, devices capable of displaying 3D content have shown to become increasingly popular. Most of the moviegoers have long accustomed to the variety of movies releasing in 3D every year, and with affordable 3D television screens on the market, movies with the extra dimension can be enjoyed in the living room. It is obvious that for the viewer, the most important part of the display is the content shown by it. Current graphics processors together with state-of-the-art rendering algorithms bring the 3D experience to the video game consumer, allowing for a higher immersion into the virtual world. But there is also the desire to view real-world photos or videos on such a display, elevating the need for 3D capturing devices.

1.1 Light Fields

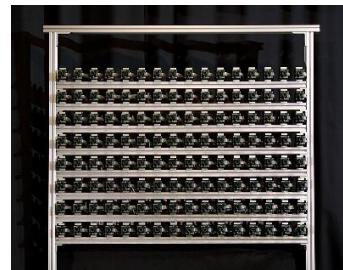
Light fields are the foundation for image based rendering, a technique to present different perspectives of a scene without the need to store geometry data, texture or lighting information. The light field is a simplified version of the more general plenoptic function, first characterized by Adelson and Bergen [1991]. It can be thought of a snapshot of the light in the entire scene, a database storing the radiance for every possible ray in the scene. This data can be captured by a grid of cameras such as the one depicted in figure 1.1b. In fact, a light field is formed inside every conventional camera, but the per-ray radiance information is lost when the light striking the sensor is accumulated over all angles. A camera that does not discard the additional radiance information is called a plenoptic camera or light field camera, shown in figure 1.1a. As described by Ng et al. [2005], an application for plenoptic camera is digital refocusing, the process of refocusing an image after it was taken. Chapter 2 gives an overview of the properties of light fields and how they are used in this work.

Mention
lumi-
graph?

The ideal 3D display should be able to display any light field, meaning it should emit light rays with radiance equal to value in the database. It turns out it is not so easy to build these displays. On the one hand are physical challenges, such as the direction of light in different angles or the correct depiction of color, contrast and brightness. On the other side comes a lot of data from the light field that needs to be processed, desirably in real-time and full resolution. Despite these challenges, many types of displays have been developed, although



(a) Image courtesy D-Kuru Wikimedia Commons



(b)

Figure 1.1: (a) Hand-held plenoptic camera from Lytro. (b) The Stanford multi-camera array, holding 128 video camera.

not necessarily based around light field technology, all having different trade-offs and limitations. There are two main categories, stereoscopic- and true 3D displays.

1.2 Stereoscopic Displays

Stereoscopic displays are based on the principles of binocular vision. The objective is to provide two distinct images to the human visual system, one for each eye, presenting the content from two slightly different perspectives. The disparities between the two images translate to depth cues in the human brain and allow for depth perception. The pair of images presented to the eyes remains constant when the viewer moves in front of the device. This effect distinguishes stereoscopic displays from 3D displays. Modern technologies include head-mounted displays, polarization systems, active shutter systems and autostereoscopy. Although not as comfortable to wear, head-mounted displays have separate high resolution screens for each eye allowing for a high degree of immersion. Polarization screens show the image pair superimposed with different polarization of the light, which is separated again by different polarization filters in the right and left side of the viewers eyeglasses. Active shutter systems use special eyeglasses that alternately block the light for one eye, letting the opposite eye see the corresponding image on the synchronized screen. Autostereoscopic displays present stereo content to the viewer without the need of special glasses. The technology is based on a lenticular lens or parallax barriers, which requires the viewer to be in a fixed and predefined position.

Ask for permission to use image showing the stanford camera array.

1.3 3D Displays

Real 3D displays ideally show the full 3D information to the observer. In contrast to stereoscopic displays, the person is able to move in front of the screen and view the content from a desired perspective. Present technologies include volumetric displays, holography, integral imaging and compressive light field displays. Volumetric displays reproduce a physical volume emitting the light of

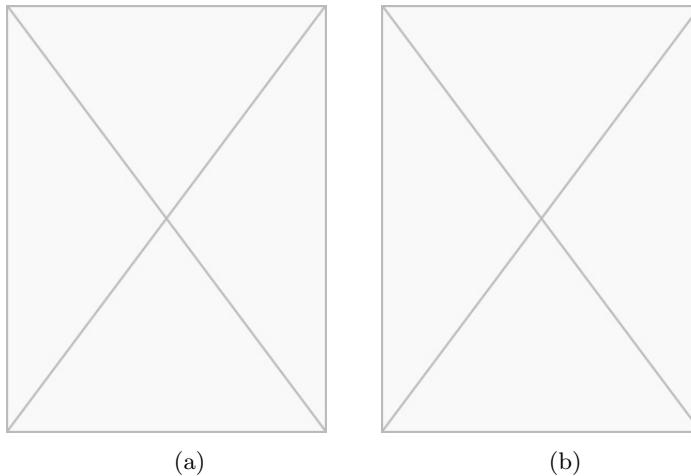


Figure 1.2: Attenuation layers between glass plates. (a) Front view of the display. (b) Side view: Ten pieces of 2 mm thick glass plates hold the five layers of transparencies, with a 4 mm separation between them.

virtual objects inside, allowing for a full 360 degree viewing angle. Holographic displays are based on conventional LCD panels equipped with a diffraction layer making it possible to project images in different directions in space. Integral imaging devices achieve the same result with a microlens array in front of the screen similar to lenticular lenses. Finally, compressive light field displays, also called tensor displays, consist of multiple LCD panels forming a stack of time multiplexed, light attenuating layers.

The work in this thesis is based on a much simpler version of these light field displays, called *Layered 3D*, which was first realized by Wetzstein et al. [2011]. The display is able to present a static, full 4D light field without the need of special glasses. It consists of masks, printed on transparencies, which attenuate light from a backlight in a multiplicative manner.

1.4 Related Work

This thesis was inspired by and builds upon the work of Wetzstein et al. [2011]. In their paper, they present a model for an inexpensive 3D display built from light attenuating, multiplicative masks. The simulated reconstructions they achieve are very convincing, although they only worked with synthetic light fields from oblique projections. With a conclusive spectral analysis, they show that multi-layer displays have increased depth of field and improved spatial resolution in comparison to other automultiscopic displays such as parallax barriers or integral imaging. The idea for this thesis is to extend the work of Wetzstein et al. to support camera light fields (e.g. captured with a plenoptic camera or camera array) in addition.

Chapter 2

Capturing a Light Field

Add small introduction

2.1 The Plenoptic Function and the Light Field

The plenoptic function, as introduced by Adelson and Bergen [1991], is a 7D function that describes the intensity of light for every frequency, along every light ray in space, at any time. It is defined as

$$P: \mathbb{R}^3 \times [0, 2\pi) \times [0, \pi] \times \mathbb{R}^2 \rightarrow \mathbb{R}^+$$
$$(x, y, z, \theta, \phi, t, \lambda) \mapsto P(x, y, z, \theta, \phi, t, \lambda),$$

where the parameters (x, y, z) are the coordinates of a point in 3D space and the angles (θ, ϕ) describe the direction of an incoming light ray at time t . The light's intensity is given for every wavelength λ and thus, the plenoptic function not only captures the visible frequency spectrum but all electromagnetic waves. A commonly used measure for light is the radiance, which is obtained from P by integrating over all wavelengths: $R(x, y, z, \theta, \phi, t) = \int_{\mathbb{R}} P(x, y, z, \theta, \phi, t, \lambda) d\lambda$.

In practice, it is impossible to acquire all the data needed to model the 7D plenoptic function and hence it is reasonable to consider only a subset of the parameters. Dropping the time parameter t in $R(x, y, z, \theta, \phi, t)$ yields a 5D function for the radiance in a static scene. As described by Levoy and Hanrahan [1996], this five dimensional representation can further be reduced to four dimensions in the following way. The radiance along a line is constant in free space and so, the 5D plenoptic function holds redundant information for the points on this line. Ignoring this redundancy leads to the equivalent 4D parameterization of the ray space. Levoy and Hanrahan [1996] propose a parameterization by two parallel planes, as seen in figure 2.1, where the coordinates of the lines (rays) are given by the intersections with the two planes. The **4D light field** $L(u, v, s, t)$ is therefore defined as the radiance along the line intersecting the two planes at coordinates (u, v) and (s, t) . This two-plane parameterization of the light field is the most common one seen in literature, but there are many ways to choose a parameterization. For instance, one can use a plane and two angles to define each ray passing this plane, which would result in a light field $L(u, v, \theta, \phi)$, where $\theta, \phi \in (0, \pi)$.

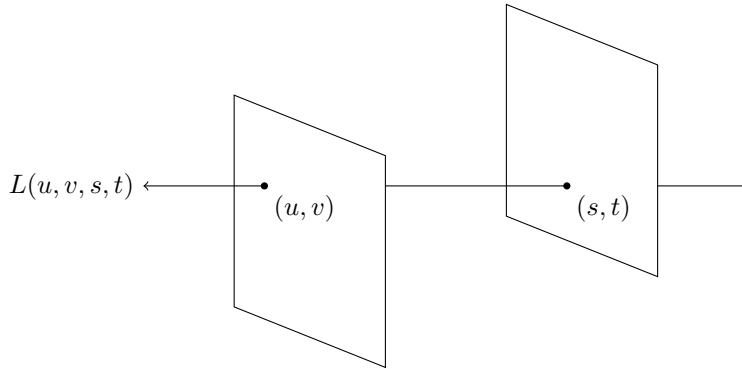


Figure 2.1: Parametrization of the light field with two planes.

2.2 Light Field Acquisition

For practical applications, the light field must be discretized and so, an appropriate sampling method needs to be chosen. This means that only a slice of the actual light field can be captured and the two planes are clipped to form rectangles. In this work, the term *light field* is used for both the infinite, continuous light field as well as the discrete collection of data samples.

Oblique Projection

Oblique projection, as shown in figure 2.2a, is a special case of orthographic projection: The parallel rays do not need to be perpendicular to the image plane of the camera. The advantage is that there is a one-to-one correspondence between camera position and ray angle, since all rays in one camera are parallel. This means that the angular resolution is simply the number of cameras, and the spatial resolution is the number of pixels in the image plane. The angular extent from θ_{\min} to θ_{\max} is called the **field of view** (FOV) of the light field and should not be confused with the field of view of a conventional camera. For a uniform angular sampling with resolution $N_\theta \times N_\phi$, the angles θ_i and ϕ_j are

$$\theta_i = \theta_{\min} + (i - 1) \frac{\text{FOV}_\theta}{N_\theta - 1}, \quad \phi_j = \phi_{\min} + (j - 1) \frac{\text{FOV}_\phi}{N_\phi - 1}, \quad (2.1)$$

where $i = 1, 2, \dots, N_\theta$ and $j = 1, 2, \dots, N_\phi$.

Given a light field $L(u, v, s, t)$ and the distance d between the two planes, a re-parameterization $L'(\theta, \phi, s, t)$ can be obtained according to figure 2.2b by the transformation

$$\theta = \arctan \left(\frac{u - s}{d} \right), \quad \phi = \arctan \left(\frac{v - t}{d} \right). \quad (2.2)$$

Note that uniform sampling in angular dimension does not yield a uniform grid in the (u, v) -plane. Despite the simplicity of this projection type, it is not feasible to build cameras of this type and so, oblique projection is left to be used exclusively by computers for rendering synthetic scenes.

Refer to figure with synthetic scenes (dice).

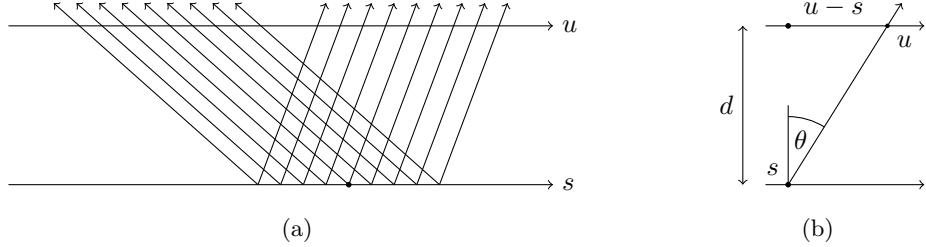


Figure 2.2: (a) Light field acquisition using oblique projection. (b) Re-parameterization of the two-plane representation to angular coordinates.

Perspective Projection

Another way to capture the light field is with a grid of optical systems, e.g. cameras. Typically, the (u, v) -plane is sampled on a grid $G_{uv} = \{(u_i, v_j) \mid i = 1, \dots, n, j = 1, \dots, m\}$ on the (u, v) -plane with a resolution $n \times m$. The extent in horizontal (vertical) direction is called the horizontal (vertical) **baseline**. Although it is strictly speaking not correct, the resolution of the (u, v) -plane is often referred to as the angular resolution. The angles of the rays in a light field captured by perspective projections are determined by the focal length, the sensor size and the sensor resolution of the camera. For a camera light field, typically it is expected that

- All cameras are placed at grid positions in G_{uv} on the same plane, called the (u, v) -plane,
- The optical axes of the cameras are orthogonal to the (u, v) -plane,
- All cameras have the same intrinsic parameters (e.g. focal length).

In this case, the focal planes of all cameras coincide with a common focal plane, the (s, t) -plane. Figure 2.3a shows this scenario for three cameras in two dimensions. Given images $I_{uv}(x, y)$ with respect to a coordinate system centered at the camera position (u, v) , the coordinates on the (s, t) -plane are $s = u + x$, and $t = v + y$. Thus, the light field in continuous coordinates is obtained by

$$L(u, v, s, t) = L(u, v, u + x, v + y) = I_{uv}(x, y). \quad (2.3)$$

In the discrete case, each camera captures sample points on the (s, t) -plane, but not everyone of these sample points on the (s, t) -plane is captured by every camera. So, as demonstrated in figure 2.3b, the camera images need to be rectified such that all discrete coordinates (u, v, s, t) correspond to valid rays. This rectification process is equivalent to a re-parameterization L' of the continuous light field L , given by the formula

$$L'(u, v, s', t') = L(u, v, \gamma(s' - u) + u, \gamma(t' - v) + v), \quad (2.4)$$

where $\gamma = \frac{d}{d'}$ and d' is the distance between the (u, v) -plane and the new (s', t') -plane. As derived by Isaksen et al. [2000], this re-parameterization is equivalent to a 4D shear.

A different way to understand this coordinate change is to imagine the (u, v) - and (s, t) -plane being the aperture and sensor planes respectively, resulting in

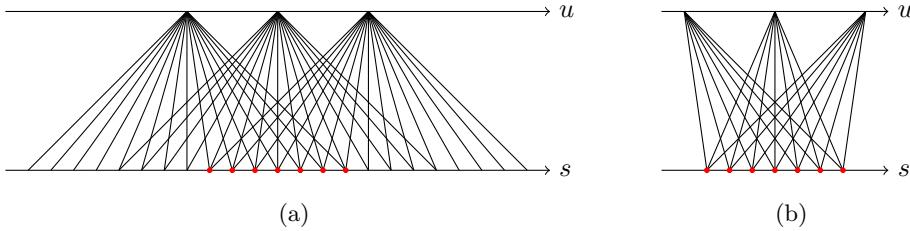


Figure 2.3: Perspective projections of a scene. (a) Projections with three pin-hole cameras. (b) Discarding unused rays corresponds to cropping the camera images.

one big camera in which a light field is formed. Changing the distance between the two planes is now equivalent to changing the focal length of this one camera. The effect on the light field inside is similar to refocusing, except that in a conventional camera, the image on the sensor is formed by a weighted integral over u and v such that the angular information vanishes. Objects at focal distance from the camera would appear sharp and objects away from the focal point would become blurred.

From stereo vision, it is known that the displacement of the projections in the image planes of two cameras is only dependent on the focal length f , the baseline Δu and the distance z , and the relation is given by $\Delta x = f\Delta u/z$. This knowledge can directly be applied to the two-plane parameterization. For the continuous light field, it amounts to

$$ds = \frac{z - Z_{st}}{z - Z_{uv}} du \quad \text{and} \quad dt = \frac{z - Z_{st}}{z - Z_{uv}} dv, \quad (2.5)$$

with Z_{uv} and Z_{st} denoting the placement of the (u, v) - and (s, t) -planes in Z -direction. Usually, the coordinate system is chosen such that $Z_{uv} = 0$. In the discrete case, the displacement Δs or Δt is also called the **disparity** and is often measured in pixel units.

Refer to
section
about the
plenoptic
camera

2.3 Visualization

The epipolar-plane image (EPI) allows for a very intuitive visualization of depth from a 4D light field. It was first defined by Bolles et al. [1987] as follows. Consider a point P in 3D space and a pair of cameras with the optical axis pointing in the same direction. The plane passing through P and the two centers of projection is called the **epipolar plane**. The epipolar plane projects to a line on each of the camera image planes, named the **epipolar line**. This line represents a constraint for the projection of P in each of the images and it is used to solve the correspondence problem in computer vision. The notion of epipolar lines can be directly applied to a multiple camera setup. In figure 2.4, a synthetic scene is rendered in 500 different positions along a horizontal baseline. Since the camera movement is in horizontal direction only, the epipolar lines correspond to a fixed pixel row in each image. The EPIs shown in figures 2.4b and 2.4c are created by collecting the chosen pixel row (scanline) in every image and stacking it up.

Mention
other
types of
visualiza-
tion? 2D
grid of im-
ages, etc.

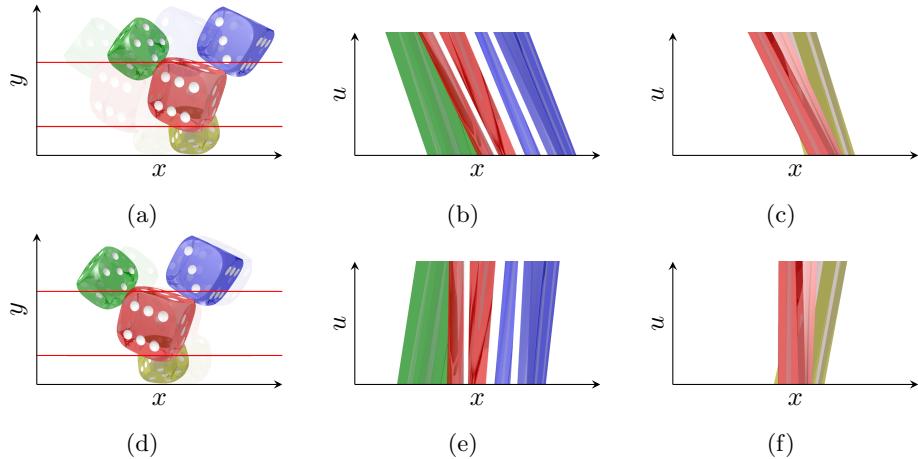


Figure 2.4: (a) Raw 3D light field rendered from 500 positions along a horizontal baseline. Two scanlines are extracted from every image. (b) The feature paths of the blue and green dice have a steeper slope than those of the red die. (c) Feature paths of the yellow die have an even steeper slope, indicating greater depth. (d) The light field is rectified according to figure 2.3b such that the disparities of the red die are approximately zero. (e) - (f) EPIs from the same scanlines. The slopes of the feature paths stay the same relative to each other.

As described in the previous section, the depth component of P occurs as a displacement of the projections in consecutive images. Under the assumption that the (u, v) -plane is sampled uniformly, the disparity D with respect to P stays constant from one image to the next. Thus, following the projection of the point P in every image corresponds to a line in the EPI with a slope proportional to $1/D$. Bolles et al. [1987] refer to this line as the **feature path**. This means that points farther away from the camera will appear as a feature path in the EPI with steeper slope than points close to the camera. Note that the depth range in the light field can immediately be determined by identifying the maximum and minimum slope in the EPI. Also, for a perfectly Lambertian scene, each line in the EPI has a uniform color.

2.4 The Plenoptic Camera

Introduce
Lytro
Camera,
mention
refocusing

Chapter 3

Light Field Tomography

3.1 A Model for Light Attenuation

The light field display is modeled by a volumetric attenuator $\mu(x, y, z)$ that attenuates the light traveling through its material. According to the Beer-Lambert law, the intensity of a light ray $\mathcal{R} \subset \mathbb{R}^3$ passing through the material decreases exponentially over distance:

$$I = I_0 e^{-\int_{\mathcal{R}} \mu(r) dr}. \quad (3.1)$$

The incident intensity I_0 is the intensity of the ray before it enters the attenuator. Equation 3.1 can be rewritten into

$$\bar{I} := \log \left(\frac{I}{I_0} \right) = - \int_{\mathcal{R}} \mu(r) dr. \quad (3.2)$$

Now, let the attenuator $\mu(x, y, z)$ be a cubic slab of height d in Z-direction and let $L(u, v, s, t)$ be the two-plane parameterization of the light field such that the (s, t) -plane coincides with the (x, y) -plane of the attenuator and the (u, v) -plane is at distance d . The set of points describing the ray defined by the coordinates (u, v, s, t) is

$$\mathcal{R} = \left\{ \lambda a + b \mid a = \begin{pmatrix} u-s \\ v-t \\ d \end{pmatrix}, b = \begin{pmatrix} s \\ t \\ 0 \end{pmatrix}, \lambda \in \mathbb{R} \right\}. \quad (3.3)$$

A point $p = (x, y, z)^T$ is part of the ray \mathcal{R} if and only if

$$\exists \lambda \in \mathbb{R} : p = \lambda a + b \iff a \times (p - b) = 0, \quad (3.4)$$

where \times denotes the cross product. Now, I can be replaced with the light field L and the right hand side of equation 3.2 can be written as an integral over \mathbb{R}^3 :

$$\bar{L}(u, v, s, t) = - \int_{\mathbb{R}^3} \mu(p) \delta(a \times (p - b)) dp. \quad (3.5)$$

- Explain linearity in log domain
- Mention other tomographic projection types
- Explain "intensity of ray" and unit

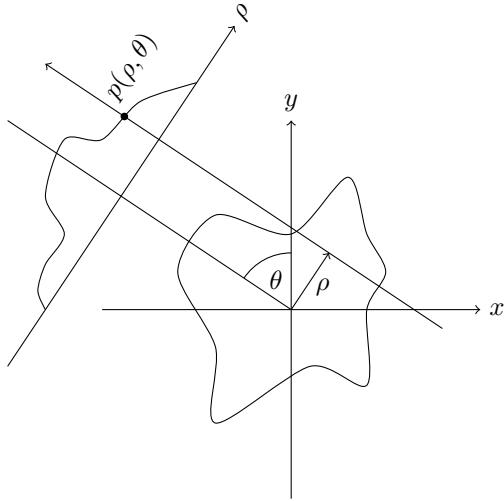


Figure 3.1: The 2D Radon transform of the ray (ρ, θ) passing a material with density $f(x, y)$.

Here, δ denotes the Dirac delta function on \mathbb{R}^3 and μ is zero outside the boundaries of the slab. This means that the integrand is only non-zero for points on the ray with coordinates (u, v, s, t) .

Combining equation 3.1 and 3.5 gives the light field emitted by the attenuator. The goal is to produce such an attenuation display that emits a given target light field.

In computed tomography, the **Radon transform** of a real valued and compactly supported, continuous function $f(x, y)$ on \mathbb{R}^2 is defined as

$$p(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta - y \sin \theta - \rho) dx dy, \quad (3.6)$$

where $(\rho, \theta) \in \mathbb{R} \times (-\frac{\pi}{2}, \frac{\pi}{2})$ defines a ray as shown in figure 3.1. Because the Radon transform is essentially a line integral, it can be generalized to three or more dimensions. Adapting the notation from the two-plane parameterization, the Radon transform of the attenuation map μ along ray \mathcal{R} becomes

$$p(u, v, s, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y, z) \delta(a \times ((x, y, z)^T - b)) dx dy dz, \quad (3.7)$$

which is equivalent to equation 3.5. This shows that

$$\bar{L}(u, v, s, t) = -p(u, v, s, t), \quad (3.8)$$

or with the words of Wetzstein et al. [2011]: “The logarithm of the emitted light field is equivalent to the negative Radon transform of the attenuation map.”

3.2 Discrete Attenuation Layers

The previous section introduced a continuously varying attenuation map to model the display. Wetzstein et al. [2011] propose to represent the attenuator with a set of N two-dimensional layers, also called masks.

Let $L_{ijkl} = L(u(i), v(j), s(k), t(l))$ be the matrix of samples from the light field and for simplicity, let $m := m(i, j, k, l)$ be a linear index of the 4D indices. Equation 3.1 suggests a per-ray constraint in the form

$$L_m = L_0 \prod_{n=1}^N t^{(n)}(h(m, n)), \quad (3.9)$$

where $h(m, n)$ is the (discrete) 2D coordinate of the intersection of the m -th ray with the n -th layer, and $t^{(n)}(\xi)$ is the **transmittance** of layer n at that coordinate. Having a constraint for each ray, the goal is to solve for the transmittance t . However, the system of equations in 3.9 is non-linear and cannot directly be solved. One can obtain a linear system of equations by taking the logarithm in 3.9:

$$\bar{L}_m = \sum_{n=1}^N \log(t^{(n)}(h(m, n))) = -\sum_{n=1}^N a^{(n)}(h(m, n)) = -P_m \alpha. \quad (3.10)$$

Here, $a^{(n)} := -\log t^{(n)}$ denotes the **absorbance** of layer n . This relation between transmittance and absorbance also directly follows from the Beer-Lambert law. Here, $P_m = (P_m^{(1)}, \dots, P_m^{(N)})$ is a binary row vector, encoding the intersection of the ray with the pixels on each layer. The unknown absorbance is represented by the column vector $\alpha = (\alpha^{(1)}, \dots, \alpha^{(N)})^T$. Each $\alpha^{(i)}$ is just a flattened representation of the absorbance matrix $a^{(i)}$. Note that equation 3.10 is the equivalent of the continuous version in 3.8, since P_m encodes the Radon transform. Finally, the above equations indexed by m can be combined into one large linear system $P\alpha = -\bar{L}$.

In most cases, P is not a square matrix and the system can become over-determined, which means that it has no solution in general. However, it is still possible to find values for α such that the error $\|P\alpha + \bar{L}\|$ is small. Thus, the objective becomes

$$\begin{aligned} \operatorname{argmin}_{\alpha} \quad & \|P\alpha + \bar{L}\|^2 \\ \text{subject to} \quad & 0 \leq \alpha \leq \infty. \end{aligned} \quad (3.11)$$

Finally, when optimal values α are found, the transmittance used to fabricate the layers is obtained by calculating $e^{-\alpha}$. Also note that the matrix P is very sparse because it is assumed that a ray passes through each layer at exactly one pixel no more than once and inter-reflections between the layers are not supported by the model. Thus, P can be efficiently stored using an appropriate data structure.

3.3 Ray Casting

To obtain the linear system P , the intersections between the rays and the attenuation layers have to be calculated. This calculation depends on the parameterization of the light field. For continuous light fields, it is always possible

to apply a re-parameterization to get the desired representation (e.g. the two-plane parameterization) and then compute the intersection in a standard way. For discrete light fields however, this would require a suitable interpolation in ray-space, which gives poor results when the distribution of samples in the target space becomes too sparse.

What follows is a description of two methods to compute the indices for the non-zero entries in P . For simplicity, only a two-dimensional attenuator of size w is assumed, consisting of N layers at various depths $Z_{\min} = z_1 < z_2 < \dots < z_N = Z_{\max}$. It is also assumed that the image plane s of the virtual cameras is bisecting the attenuator in the middle, at depth $Z_s = \frac{Z_{\max} - Z_{\min}}{2}$.

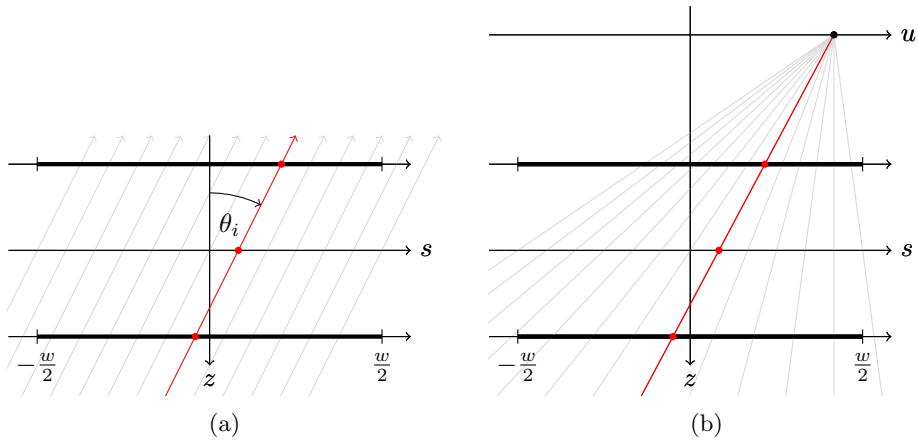


Figure 3.2: Computation of the ray-layer intersections from oblique (a) and perspective (b) projections. Two attenuation layers are drawn (top and bottom) with the virtual image plane in the center. Light rays intersect the layers at positions to be calculated.

Oblique Projection

The setup for the oblique projection type is illustrated in figure 3.2a. Let θ_i denote the angle of the i -th oblique view from the light field. Following the notation in previous sections, the linear index $m = m(i, k)$ identifies the ray $(\theta_i, s(k))$. The intersection of ray m with the n -th layer is simply

$$h(m, n) = s(k) + \Delta z \tan(\theta_i), \quad (3.12)$$

where $\Delta z = Z_s - z_n$ is the displacement of layer n from the image plane. The next step is to compute the pixel index at the point $h(m, n)$. The shift in pixel units can directly be derived from the shift in equation 3.12 given the pixel size Δs , yielding

$$\Delta k = \left[\frac{\Delta z \tan(\theta_i)}{\Delta s} \right], \quad (3.13)$$

and the new index is $k' = k + \Delta k$. The brackets in the above equation denote the rounding operation. Finally, this information is stored in the propagation matrix with the assignment $P_{mk'}^{(n)} = 1$.

Perspective Projection

For the perspective projection, it is assumed that the (u, v) -plane (or u -plane here) is at depth Z_u , with a distance $Z_s - Z_u$ from the image plane. Again, let $m = m(i, k)$ be the index that identifies the ray $(u(i), s(k))$. Finding the coordinates of the ray (u, s) on a layer is similar to the re-parameterization in equation 2.4. Setting $\gamma = \frac{Z_s - Z_u}{z_n - Z_u}$ for layer n , the formula for the intersection is

$$h(m, n) = \gamma(s(k) - u(i)) + u(i). \quad (3.14)$$

As before, the new pixel index $k' = k + \Delta k$ is computed from the shift in the number of pixels

$$\Delta k = \left\lceil \frac{h(m, n) - s(k)}{\Delta s} \right\rceil, \quad (3.15)$$

and the assignment $P_{mk'}^{(n)} = 1$ is made.

3.4 Iterative Reconstruction

The optimization problem in equation 3.11 is essentially a fitting problem. Theoretically, it can be solved in a least squares sense using the normal equation $P^T P \alpha = P^T \bar{L}$ and by inverting the matrix $P^T P$. For high resolution light fields, the matrix P becomes extremely large and it is unfeasible to compute the inverse of $P^T P$.

In general, the approach to solve these kind of problems is to use iterative methods. The choice of the method depends on the type of problem and the structure of the design matrix. In computed tomography, a variety of iterative solvers have been developed to solve the exact same problem. Among the different methods is the Simultaneous Algebraic Reconstruction Technique (SART) first proposed by Andersen and Kak [1984]. The update rule of SART for iteration $k = 0, 1, 2, \dots$ is

$$\alpha^{(k+1)} = \alpha^{(k)} + \lambda C P^T R \left(\bar{L} - P \alpha^{(k)} \right), \quad (3.16)$$

where λ is a relaxation factor. R and C denote the diagonal matrices with entries $R_{ii} = \frac{1}{r_i}$ and $C_{ii} = \frac{1}{c_i}$, where r_i and c_i are the sum of the elements in the i -th row and column of P respectively. The parts in 3.16 involving P and P^T are also referred to as the **forward-** and **back-projection** respectively.

The convergence of SART has been studied by Jiang and Wang [2001]. They have proven that it converges to a weighted least squares solution.

Show plot of residual norm in each iteration

Compare to linear least squares method from MATLAB

Mention choice of relaxation parameter lambda = 1

Chapter 4

Spectral Analysis

This chapter is intended to give an overview of the spectral properties and limitations specific to multiplicative light field displays. Spectral analysis is a crucial method for the quality assessment and it is the origin of a comprehensive understanding of 3D displays. A light field emitted by the display can be interpreted as a signal that is composed of sine waves with different amplitude, phase and frequency. Section 4.1 introduces the Fourier transform, an operation that decomposes such a signal into the frequencies that produce it. The spectral support, i.e. the range of frequencies the display is able to produce, is analyzed in section 4.3.

Introduction
to fourier
transform:
Why is
it need-
ed/impor-
tant

Decide if
Fourier
slice the-
orem is
needed

4.1 Definitions

The **Fourier transform** \hat{f} of an integrable function $f: \mathbb{R}^n \rightarrow \mathbb{C}$ is defined as

$$\hat{f}(\xi) = \mathcal{F}(f)(\xi) := \int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot \xi} dx \quad (4.1)$$

for any $\xi \in \mathbb{R}^n$. According to the Fourier integral theorem, if both f and \hat{f} are absolutely integrable and f is continuous, then the inverse transform

$$f(x) = \mathcal{F}^{-1}(\hat{f})(x) := \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x \cdot \xi} d\xi \quad (4.2)$$

is well-defined. The domain of f is called the **spatial domain** and the domain of \hat{f} is referred to as the **frequency domain**. An important property of the Fourier transform is that a convolution in the spatial domain becomes a multiplication in the frequency domain, or in other words,

$$\widehat{(f * g)}(\xi) = \hat{f}(\xi) \cdot \hat{g}(\xi) \quad (4.3)$$

for integrable functions $f, g: \mathbb{R}^n \rightarrow \mathbb{C}$. On the other hand, a multiplication in the spatial domain becomes a convolution in the frequency domain after applying the Fourier transform, that is

$$\widehat{(f \cdot g)}(\xi) = (\hat{f} * \hat{g})(\xi). \quad (4.4)$$

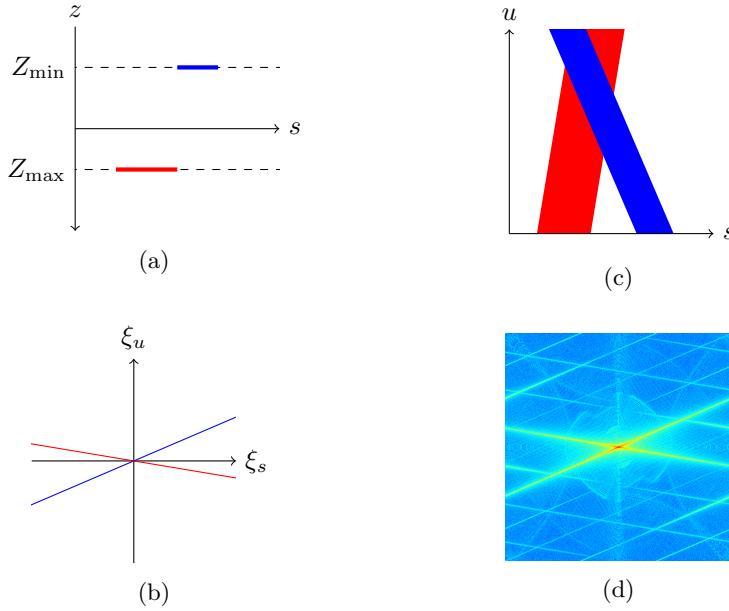


Figure 4.1: (a) Two objects (red and blue) placed at the bounds of the depth range. (b) The EPI representing the 2D light field of the scene. (c) Fourier transform of the EPI. The red and blue line mark the bounds for the spectral support. (d) Discrete Fourier transform of the EPI. Absolute values (magnitude response) are presented with colors on a logarithmic scale.

4.2 Spectral Support of Light Fields

Consider a scene with a bounded depth range between Z_{\min} and Z_{\max} . The two objects at the boundaries are shown in figure 4.1a, with the virtual image s plane between them. The consequent 2D light field $L(u, s)$ (or EPI) is depicted in figure 4.1c. From equation 2.5 it follows that objects appear in the EPI with a slope $\frac{du}{ds} = \frac{z - Z_u}{z - Z_s}$. Substituting z with Z_{\min} and Z_{\max} gives the slopes for the red and blue objects at the boundary, defining the range of slopes in the EPI for objects between the two.

Applying the Fourier transform to the continuous light field reveals that the frequency response is non-zero on lines $\frac{ds}{du}\xi_s + \xi_u = 0$. Again, for the scene with bounded depth range, this yields two lines representing the limits of the spectral support as shown in figure 4.1b. Objects between the red and blue ones will also have a frequency response within the fan spanned by the two lines. Therefore, the region of support for a continuous light field with bounded depth range can be defined in the following way.

$$\mathcal{S}(\xi_u, \xi_s) := \begin{cases} 1, & \text{if } Z_{\min} \leq \frac{Z_u \xi_u + Z_s \xi_s}{\xi_u + \xi_s} \leq Z_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (4.5)$$

A similar expression follows for the 4D light field, defining a 4D hyperfan for the region of support as derived by Dansereau et al. [2015]. Note that occlusions

as well as specular reflections are not incorporated in the above expression. These effects introduce additional discontinuities in the EPI that result in a high frequency response possibly outside the fan defined in equation 4.5.

In the case of sampled light fields, aliasing can occur due to a small sampling rate in either angular- or spatial direction. Chai et al. [2000] analytically derived the minimum sampling rate required for alias-free light field rendering and proposed a reconstruction filter from known depth boundaries. The region of support $\mathcal{S}(\xi_u, \xi_s)$ can also be thought of an ideal filter. As equation 4.3 shows, multiplying $\mathcal{S}(\xi_u, \xi_s)$ in the frequency domain is equivalent to a convolution in the spatial domain.

4.3 Spectral Support of Layered 3D Displays

With light field displays, it is of course desirable to achieve the same spectral coverage for the emitted light field as for the original. Again, the analysis starts with the assumption of a continuous light field and an attenuator with N continuously varying layers. Each layer by itself creates a light field, and since the layer is at constant depth, the frequency response is non-zero along a slanted line as demonstrated before. Let L_1, \dots, L_N denote the constant depth light fields per layer and let's assume all are parameterized with respect to the same (u, v) - and (s, t) -plane. The light field produced by all layers together is $L' = L_0 \cdot L_1 \cdots L_N$, where L_0 is the uniform illumination from the backlight. This directly follows from equation 3.9. With the multiplication theorem from equation 4.4, the Fourier transform of L' can be expressed as

$$\widehat{L}'(\xi) = (\widehat{L}_0 * \widehat{L}_1 * \cdots * \widehat{L}_N)(\xi), \quad (4.6)$$

where $\xi = (\xi_u, \xi_v, \xi_s, \xi_t)$, or $\xi = (\xi_u, \xi_s)$ for the two dimensional case. For the case of discretely sampled layers, the frequency support of the individual layer will be limited by its spatial cutoff frequency, that is the highest frequency it can produce with a given pixel size p . A signal with a period smaller than two pixels can not be reproduced by the layer's pixel grid and thus, the spatial cutoff frequency is defined as $\xi_0 = \frac{1}{2p}$ cycles/m. The sketch in figure 4.2a illustrates this for the case of three layers that are bandlimited by $\pm \xi_0$. The three lines convolved produce a diamond shaped region of support as shown in figure 4.2b, which is the effective spectral support of the display. This means that a light field with high frequencies outside the spectral support of the display will not be correctly displayed, or in other words, the display acts as a low-pass filter.

The **depth of field** of an automultiscopic display, as explained by Zwicker et al. [2006], is the depth range that can be reproduced by the display in full spatial resolution. Thus, the boundary of the spectral support describes an upper bound on the depth of field for any automultiscopic display, including layered displays. It turns out to be quite hard to analytically derive an exact expression for the upper bound. Wetzstein et al. [2011] present a statistical approach and give an approximation for the upper bound on the depth of field $|\xi_a|$ for a plane a placed at depth z_a from a N -layer display with a thickness $h = z_N - z_1$:

$$|\xi_a| \leq N \xi_0 \sqrt{\frac{(N+1)h^2}{(N+1)h^2 + 12(N-1)(z_a - Z_s)^2}} \quad (4.7)$$

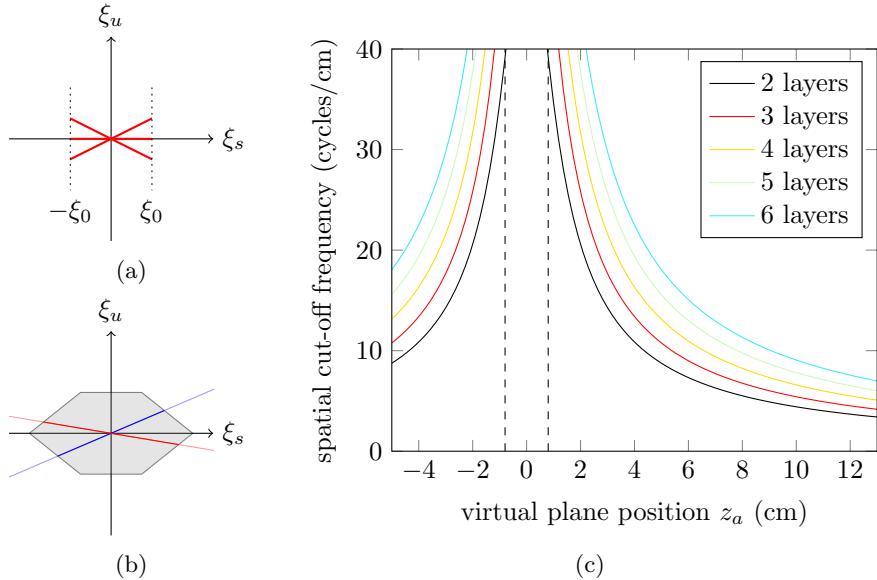


Figure 4.2: Spectral analysis for layered 3D displays. (a) Spectral support of individual layers (red) from a display with three layers, superimposed in the same frequency domain. The dashed lines mark the spatial cutoff frequency ξ_0 . (b) Combined spectral support of all three layers (gray), obtained by the convolution. The light field from figure 4.1 can be displayed with frequencies within the region of support. (c) Approximate upper bound on the depth of field for layered 3D displays with different a number of layers (Wetzstein et al. [2011]). The displays extent is depicted by the dashed lines.

The right side of the equation is plotted in figure 4.2c for different positions z_a of the virtual plane. It shows that the spatial cut-off frequency drops rapidly when moving the virtual plane away from the display. In fact, the drop-off is inversely proportional to z_a as equation 4.7 shows. As figure 4.2b suggests, the display is theoretically able to produce spatial frequency that exceeds the layers cut-off ξ_0 , even for content outside the display enclosure. The highest spatial frequency is achieved in the middle of the display, on the (s, t) -plane. This theoretical

4.4 The Fourier Slice Theorem

Explain usage of the theorem in this work

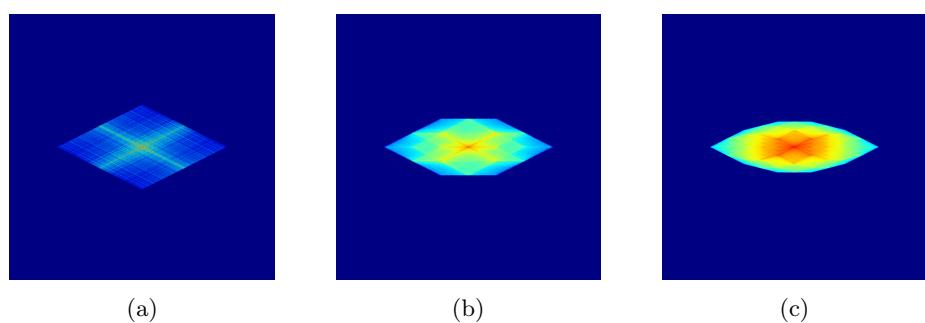


Figure 4.3: text

Chapter 5

Implementation and Assessment

5.1 Oversampling

5.2 Baseline Scaling and Back Projection

5.3 Attenuator Tiling and Blending

High resolution light fields can take up a significant amount of space in memory. For example, a light field taken with a Full HD camera from 17×17 angles would take up $1920 \cdot 1080 \cdot 17^2 \cdot 3 \cdot 8 / (1024^3) = 13.3947$ Gigabyte of memory. In addition, the propagation matrix stores information about every pixel in the light field and thus, can take up Gigabytes of space depending on the resolution of the attenuation layers. The proposed approach divides the attenuation layers into tiles. Figure 5.1 shows how the tiles are laid out. The inputs for the tiling algorithm are the resolution of the tiles $r = (r_x, r_y)$ and the overlap in horizontal and vertical direction, $o = (o_x, o_y)$. The tiles are then laid out in a grid beginning in the top left corner of the layer. The number of tiles needed to cover the plane can be calculated by

$$N_x = \left\lceil \frac{R_x - o_x}{r_x - o_x} \right\rceil \quad \text{and} \quad N_y = \left\lceil \frac{R_y - o_y}{r_y - o_y} \right\rceil. \quad (5.1)$$

The combination of the same tile from each layer forms a new attenuator of smaller size and lower resolution. The optimization is then performed for every tile separately, resulting in a smaller propagation matrix per tile. In the end, the optimized tiles are put together to form the complete attenuation layers.

In general, the borders of the attenuator contain less ray-propagation information and thus provide a higher degree of freedom for the optimization. This introduces artifacts that are clearly visible in the reassembled layers as shown in figure ???. To solve this issue, the tiles have to overlap. In this case, when reassembling the layers from the tiles, the overlaps need to be blended with a mask: After the optimization, each tile gets multiplied with a mask shown in figure ???. The finished layers are then obtained by summing the tiles and dividing by the sum of the blending masks shown in figure ???.

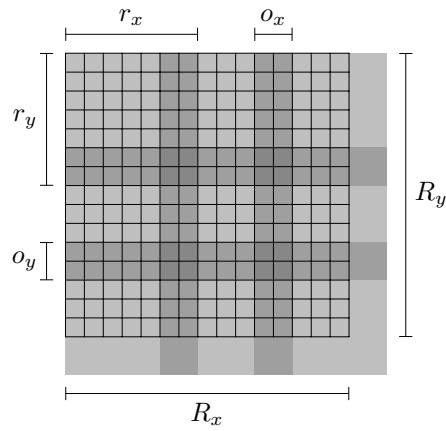


Figure 5.1: Layout of the tiles that cover the attenuation layers. The pixel grid of size $R_x \times R_y$ is covered by tiles of $r_x \times r_y$ pixels with an overlap of o_x in horizontal and o_y in vertical direction.

5.4 Benefits and Limitations

Appendix A

Appendix

A.1 Backlight Fabrication

For an optimal viewing experience, a uniform backlight is needed to place behind the glass plates which hold the printed transparencies. LEDs are the optimal choice for a simple white backlight: They are small sized, power saving, affordable and don't produce a lot of heat when turned on. The ones used for this build are average, consumer grade quality LEDs bought on Amazon. A detailed specification of the product is given in table A.1.

A five meter long and 10 mm wide LED strip is cut into pieces of 15 cm length, each having nine LEDs. These smaller strips are placed next to each other and glued onto a wooden plate forming the base for the backlight. Corresponding ends of the strips are reconnected by soldering small pieces of wire onto the contacts of the strips. For an easier soldering process and to reduce the risk of an electrical short, the strips are alternately offset by a small amount as shown in figure A.1c. A connector for the power supply is mounted on one end of the strip. Wooden stand-offs glued to the base hold the diffuser plate 17 mm above the LEDs. The diffuser is simply a white, milky acrylic plate from a hardware store, cut to the right size. Finally, the build is concluded with a wooden frame covering the LEDs and wires and holding the glass plates in place. The frame is painted with a color varnish to protect the wood from scratches and to beautify the product.

A total of three backlights were produced for this project. All three displays are powered by a 12V/6A DC power supply. The set also includes a remote control to turn the displays on and off or to adjust the brightness level.

Add dimensions of the display and frame

Figures:
All three displays
+ look inside, showing LED and soldering job.

LED Chip	SMD 5050
Electric current	12V DC
Color	Cold white
Color temperature	6000 Kelvin
Luminous flux	4500 Lumen
Emission angle	120°
Power consumption	60 Watt
Average lifespan	50000 hours

Table A.1: Specification of LEDs used for the backlight.

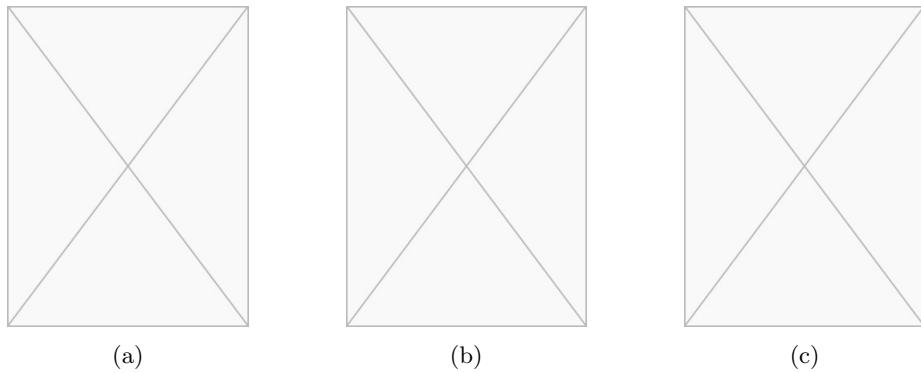


Figure A.1: (a) Three fully assembled displays with backlight. (b) Close up of the inside with the outer frame removed. Top to bottom: Glass plates holding attenuation layers, diffusion plate, LED grid. (c) Close up of the LED grid with the diffuser removed.

List of Tables

A.1 LED specification	22
---------------------------------	----

List of Figures

1.1	Light field aquisition devices	2
1.2	Attenuation layers between glass plates	3
2.1	Parametrization of the light field with two planes	5
2.2	Parameterization for light fields from oblique projections	6
2.3	Parameterization for light fields from perspective projections	7
2.4	Visualization of the light field with epipolar plane images	8
3.1	The Radon transform	10
3.2	Computation of the ray-layer intersections from oblique and perspective projections	12
4.1	Spectral analysis for light fields with bounded depth range	15
4.2	Spectral analysis for layered 3D displays	17
4.3	text	18
5.1	Tiling layout	20
A.1	Handcrafted backlights for the attenuation displays	22

Bibliography

- E. H. Adelson and J. Bergen. The plenoptic function and the elements of early vision. *Computational Models of Visual Processing*, pages 3–20, 1991.
- A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.
- R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, pages 7–55, 1987.
- J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum. Plenoptic sampling. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 307–318. ACM, 2000.
- D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenslet-based plenoptic cameras. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1027–1034. IEEE Computer Society, 2013.
- D. G. Dansereau, O. Pizarro, and S. B. Williams. Linear volumetric focus for light field cameras. *ACM Trans. Graph.*, 34(2):15:1–15:20, 2015.
- F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. X. Sillion. A frequency analysis of light transport. *ACM Trans. Graph.*, 24(3):1115–1126, 2005.
- S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 43–54, New York, NY, USA, 1996. ACM.
- F.-C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar. Eyeglasses-free display: Towards correcting visual aberrations with computational light field displays. *ACM Trans. Graph.*, 33(4):59:1–59:12, 2014.
- A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 297–306. ACM, 2000.
- M. Jiang and G. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 360–364, 2001.

- C. Kim, A. Hornung, S. Heinze, W. Matusik, and M. Gross. Multi-perspective stereoscopy from light fields. *ACM Trans. Graph.*, 30(6):190:1–190:10, 2011.
- M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pages 31–42, 1996.
- C.-K. Liang and R. Ramamoorthi. A light transport framework for lenslet light field cameras. *ACM Trans. Graph.*, 34(2):16:1–16:19, 2015.
- Y. Lu, W. Wang, S. Chen, Y. Xie, J. Qin, W.-M. Pang, and P.-A. Heng. Accelerating algebraic reconstruction using CUDA-enabled GPU. In *Sixth International Conference on Computer Graphics, Imaging and Visualization*, pages 480–485. IEEE, 2009.
- R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Trans. Graph.*, 34(4):59:1–59:12, 2015.
- R. Ng. Fourier slice photography. *ACM Trans. Graph.*, 24(3):735–744, 2005.
- R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11), 2005.
- V. Vaish, B. Wilburn, N. Joshi, and M. Levoy. Using plane + parallax for calibrating dense camera arrays. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1. IEEE, 2004.
- G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar. Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph.*, 30(4):95:1–95:12, 2011.
- G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph.*, 31(4):80:1–80:11, 2012.
- M. Yan. Convergence analysis of SART by bregman iteration and dual gradient descent. pages 1–15, 2010.
- M. Zwicker, W. Matusik, F. Durand, H. Pfister, and C. Forlines. Antialiasing for automultiscopic 3D displays. In *ACM SIGGRAPH 2006 Sketches*, SIGGRAPH '06, New York, NY, USA, 2006. ACM.

Erklärung

gemäss Art. 28 Abs. 2 RSL 05

Name/Vorname:

Matrikelnummer:

Studiengang:

Bachelor

Master

Dissertation

Titel der Arbeit:

.....
.....

LeiterIn der Arbeit:

.....

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetztes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.

.....
Ort/Datum

.....
Unterschrift