Attenuation-based Light Field Displays

Bachelorarbeit

der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Adrian Wälchli

2015

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

Abstract

Abstract goes here

Contents

Chapter 1

Introduction

1.1 Related Work

Chapter 2

Capturing a Light Field

2.1 The Plenoptic Function and the Light Field

The plenoptic function, as introduced by [AB91], 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

$$\begin{aligned} P\colon \mathbb{R}^3 \times [0, 2\pi) \times [0, \pi] \times \mathbb{R}^2 &\to \mathbb{R}^+ \\ (x, y, z, \theta, \phi, t, \lambda) &\mapsto P\,(x, y, z, \theta, \phi, t, \lambda)\,, \end{aligned}$$

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: $R\,(x, y, z, \theta, \phi, t) = \int_{\mathbb{R}} P\,(x, y, z, \theta, \phi, t, \lambda)\ \mathrm{d}\lambda$.

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 [LH96], 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. [LH96] 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 $\theta, \phi \in (0, \pi)$.

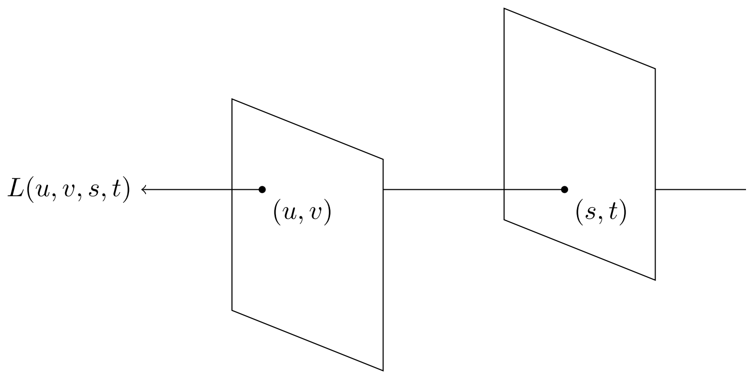


Figure 2.1: Parametrization of the light field with two planes.

2.2 Light Field Acquisition

For practical applications, the light field must be discretized and so an appropriate sampling method needs to be chosen. One way is to capture the light field with a grid of optical systems, e.g. cameras. Typically, the (u, v)-plane is sampled on a grid $G_{uv} = \{(u_i, v_j) \mid i = 1, \ldots, n, j = 1, \ldots, m\}$ on the (u, v)-plane with a resolution $n \times m$. The extent in horizontal (vertical) direction is called the horizontal (vertical) **baseline**. This means that only a slice of the actual light field can be captured and the two planes are clipped to form a rectangle.

Oblique Projection

Oblique projection, as shown in figure 2.2(a), 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. 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 2.2(b) by the transformation

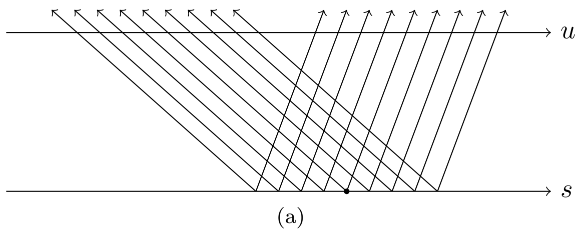
$$\theta = \arctan\left(\frac{u - s}{d}\right), \qquad \phi = \arctan\left(\frac{v - t}{d}\right).$$

However, this type of projection is only applicable for synthetic scenes that are rendered with a computer.

Perspective Projection

The angles of the rays in a light field captured by perspective projections are determined by the focal length 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,



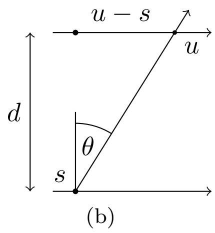
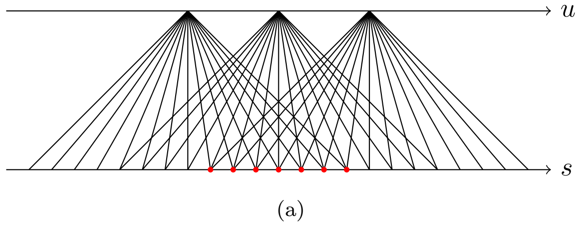


Figure 2.2: (a) Light field aquisition using oblique projection. (b) Reparameterization of the two-plane representation to angular coordinates.



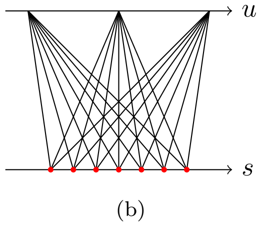


Figure 2.3: Perspective projections of a scene. (a) Projections with three pinhole cameras. (b) Discarding unused rays corresponds to cropping the camera images.

- 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 2.3(a) shows this scenario for three cameras in two dimensions. Each camera captures sample points on the (s, t)-plane, but not every point on the (s, t)-plane is captured by ever camera. As demonstrated in figure 2.3(b), the camera images need to be rectified such that all discrete coordinates (u, v, s, t) correspond to valid rays.

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 [BBM87] as follows. Consider a point P in 3D space and a pair of cameras with the optical axis pointing in the same direcion. 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 correspondance 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 2.4(b) and 2.4(c) are created by collecting the chosen pixel row (scanline) in every image and stacking it up.

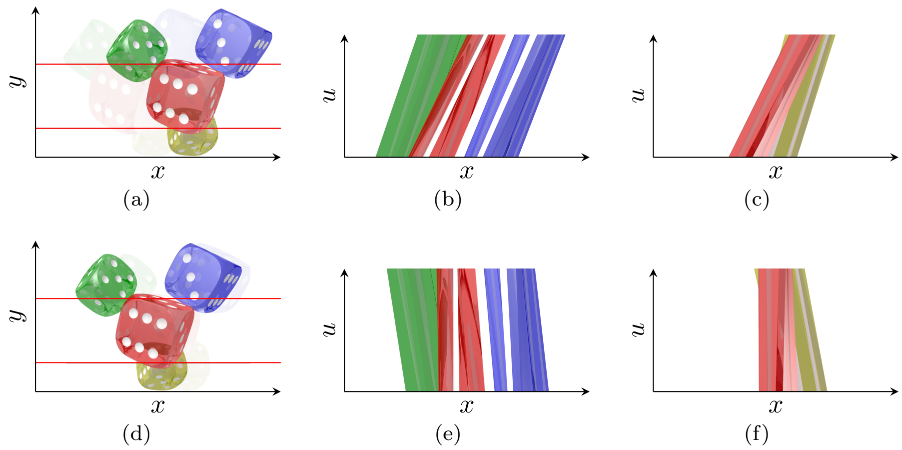


Figure 2.4: (a) Raw 3D light field rendered from 500 positions along a horizontal baseline. Two scanlines are extracted from every image. (b) The feature paths of the blue and green dice have a steeper slope than those of the red die. (c) Feature paths of the yellow die have an even steeper slope, indicating greater depth. (d) Images are cropped according to figure 2.3(b) 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.

The shift in pixels of projections of the same point P in consecutive images is called the (pixel) **disparity** D. Following the projection of the point P in every image corresponds to a line in the EPI with slope $1/D$. [BBM87] 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. The depth range in the light field can immediately be determined by identifying the maximum and minimum slope in the EPI.

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:

$$I = I_0 e^{-\int_{\mathcal{R}} \mu(r)\,\mathrm{d}r}. \tag{3.1}$$

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

$$\bar{I} := \log\left(\frac{I}{I_0}\right) = -\int_{\mathcal{R}} \mu(r)\,\mathrm{d}r. \tag{3.2}$$

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

$$\mathcal{R} = \left\{ \lambda a + b \;\middle|\; a = \begin{pmatrix} u-s \\ v-t \\ d \end{pmatrix}, b = \begin{pmatrix} s \\ t \\ 0 \end{pmatrix}, \lambda \in \mathbb{R} \right\}. \tag{3.3}$$

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

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

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 $\mathbb{R}^3$:

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

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).

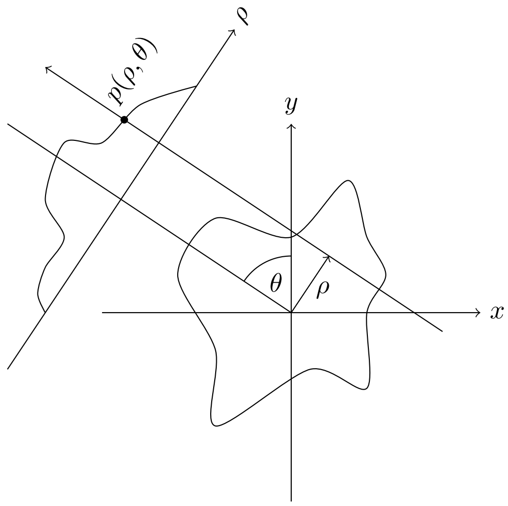


Figure 3.1: The 2D Radon transform of the ray (ρ, θ) passing a material with density $f(x, y)$.

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

$$p(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x \cos\theta - y \sin\theta - \rho) \,\mathrm{d}x \,\mathrm{d}y, \tag{3.6}$$

where $(\rho, \theta) \in \mathbb{R} \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 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

$$p(u, v, s, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y, z)\delta\left(a \times \left((x, y, z)^T - b\right)\right) \,\mathrm{d}x \,\mathrm{d}y \,\mathrm{d}z, \tag{3.7}$$

which is equivalent to equation 3.5. This shows that

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

or with the words of [WLHR11]: “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. [WLHR11] 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

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

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:

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

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, $P_m = \left(P_m^{(1)}, \ldots, P_m^{(N)} \right)$ 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 $\alpha = \left(\alpha^{(1)}, \ldots, \alpha^{(N)} \right)^T$. 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 overdetermined, which means that it has no solution in general. However, it is still possible to find values for α such that the error $\left\| P\alpha + \bar{L} \right\|$ is small. Thus, the objective becomes

$$\begin{aligned} \underset{\alpha}{\operatorname{argmin}} \quad & \left\| P\alpha + \bar{L} \right\|^2 \\ \text{subject to} \quad & 0 \leq \alpha \leq \infty. \end{aligned} \tag{3.11}$$

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 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 [AK84]. The update rule of SART for iteration $k = 0, 1, 2, \ldots$ is

$$\alpha^{(k+1)} = \alpha^{(k)} + \lambda C P^T R \left(\bar{L} - P\alpha^{(k)} \right), \tag{3.12}$$

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 convergence of SART has been studied by [JW01]. They have proven that it converges to a weighted least squares solution.

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 4.2.

4.1 Definitions

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

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

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

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

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,

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

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

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

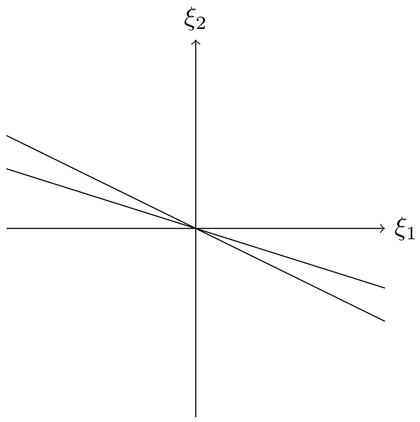


Figure 4.1:

4.2 Spectral Support of Display

Consider a scene with a bounded depth range $Z_{\min} \leq Z \leq Z_{\max}$. In section 2.3, it was shown that an object at a certain depth will be represented in the EPI as a line with a slope proportional to the depth.

Now, consider a single attenuation layer. The light field emitted by this layer has constant depth and thus, the lines in the EPI all have the same slope.

4.3 The Fourier Slice Theorem

Appendix A

ap1

A.1 apsec1

List of Tables

List of Figures

Bibliography

[AB91] Adelson, E. H. ; Bergen, J.: The Plenoptic Function and the Elements of Early Vision. In: *Computational Models of Visual Processing* (1991), S. 3–20

[AK84] Andersen, A. H. ; Kak, A. C.: Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of the ART Algorithm. In: *Ultrasonic Imaging* 6 (1984), Nr. 1, S. 81–94

[BBM87] Bolles, Robert C. ; Baker, H. H. ; Marimont, David H.: Epipolar-plane image analysis: An approach to determining structure from motion. In: *International Journal of Computer Vision* (1987), S. 7–55

[JW01] Jiang, Min ; Wang, Ge: Convergence of the Simultaneous Algebraic Reconstruction Technique (SART). In: *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers* 1 (2001), S. 360 – 364

[LH96] Levoy, M. ; Hanrahan, P.: Light Field Rendering. (1996), S. 1–12

[WLHR11] Wetzstein, G. ; Lanman, D. ; Heidrich, W. ; Raskar, R.: Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays. In: *ACM Trans. Graph.* 30 (2011), Nr. 4

[WLHR12] Wetzstein, G. ; Lanman, D. ; Hirsch, M. ; Raskar, R.: Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. In: *ACM Trans. Graph. (Proc. SIGGRAPH)* 31 (2012), Nr. 4, S. 1–11

[Yan10] Yan, Ming: Convergence Analysis of SART by Bregman Iteration and Dual Gradient Descent. (2010), S. 1–15

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