Variational Shape from Light Field

Stefan Heber, Rene Ranftl, and Thomas Pock*

Institute for Computer Graphics and Vision,
Graz University of Technology, Austria
{stefan.heber,ranftl,pock}@icg.tugraz.at
http://www.icg.tu-graz.ac.at/

Abstract. In this paper we propose an efficient method to calculate a high-quality depth map from a single raw image captured by a light field or plenoptic camera. The proposed model combines the main idea of Active Wavefront Sampling (AWS) with the light field technique, *i.e.* we extract so-called sub-aperture images out of the raw image of a plenoptic camera, in such a way that the virtual view points are arranged on circles around a fixed center view. By tracking an imaged scene point over a sequence of sub-aperture images corresponding to a common circle, one can observe a virtual rotation of the scene point on the image plane. Our model is able to measure a dense field of these rotations, which are inversely related to the scene depth.

Keywords: Light field, depth, continuous optimization.

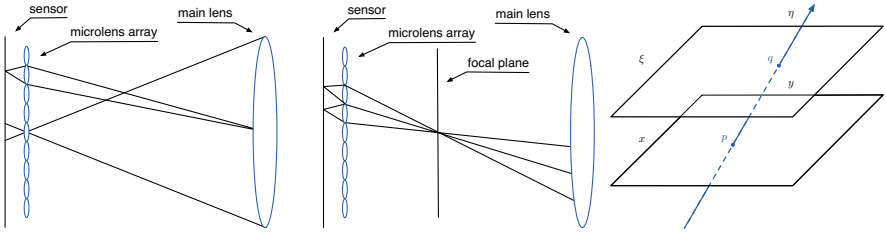
1 Introduction

In geometrical optics, rays are used to model the propagation of light. The amount of light propagated by a ray is denoted as radiance, and the radiance along all rays in a 3D space is called the plenoptic function [1]. The plenoptic function is a 5D function, due to the fact that a ray in 3D space can be parametrized via a position (x, y, z) and a direction (ξ, η). However, the 5D plenoptic function contains redundant information, because the radiance along a ray remains constant till it hits an object. Thus, the redundant information is one dimensional which reduces the 5D plenoptic function to the 4D light field [14] or Lumigraph [12].

There are different devices to capture light fields. The simplest ones are single moving cameras (gantry constructions), which allow a large baseline between captured viewpoints, but are limited to static scenes. Another way of capturing light fields is via multiple cameras. This approach allows to capture dynamic scenes, but is very hardware intensive. In order to capture a full light field, multiple cameras have to be arranged in a 2D array [22]. A further device which re-attracts attention in recent years is the light field [14] or plenoptic camera [2]. As long ago as in the year 1908, Lippmann [15] introduced the basic idea of such a camera. The technique has then been developed and improved by various authors [8–10], but it needed nearly a century till the first commercial plenoptic camera has become available [19, 18].

* This work was funded by the Austrian Science Fund (FWF).

A. Heyden et al. (Eds.): EMMCVPR 2013, LNCS 8081, pp. 66–79, 2013.



(a) plenoptic 1.0 (b) plenoptic 2.0 (c) parametrization

Fig. 1. (a) and (b) illustrates different types of light field or plenoptic cameras defined by Lumsdaine and Georgiev [16]. (a) Sketch of a traditional light field camera, also called *plenoptic 1.0 camera*. (b) Sketch of a focused light field camera, also denoted as *plenoptic 2.0 camera*. (c) Sketch of the used light field parametrization. The 4D light field $L(x, y, \xi, \eta)$ is rewritten as $L(p, q)$, where p is a point in the image plane and q is a point in the lens plane. Thus p represents the spatial component and q represents the directional component.

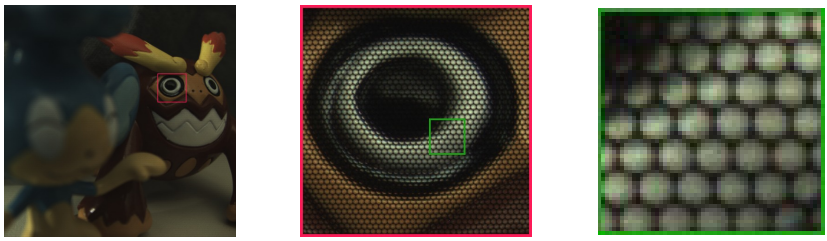
Compared to a conventional camera (2D photograph), which only captures the total amount of light striking each point on the image sensor, a light field camera records the amount of light of individual light rays, which contribute to the final image. Recording the additional directional information is achieved by inserting a micro-lens-array into the optical train of a conventional camera. This has the effect, that the micro-lenses separate the incoming light into several rays of different directions. The individual light rays are then captured at different locations on the sensor.

Depending on the focusing position of the main lens and the micro-lenses, Lumsdaine and Georgiev [16] distinguished between two types of plenoptic cameras, *plenoptic 1.0* and *plenoptic 2.0* (cf Figure 1(a) and 1(b)). In a traditional plenoptic camera (*plenoptic 1.0*) [2, 19, 18] the main lens is focused at the micro-lens plane and the micro-lenses are focused at the main lens (optical infinity). Thus, the position of each micro-lens captures spatial information, and the part of the sensor under each micro-lens captures angular information. Note, that each micro-lens spans the same angular range. Thus, the size of the individual micro-lenses sets the spatial sampling resolution, which leads to a fundamental trade-off: For a fixed sensor resolution, the increase of directional resolution, simultaneously decreases the spatial resolution of the final image. In a focused plenoptic camera (*plenoptic 2.0*) the micro-lenses are focused on the focal plane of the main lens, which has on the one hand the effect, that angular information is spread across different micro-lenses, but on the other hand the camera now records dense spatial information, rather than dense directional information, which results in a higher spatial resolution of the final image.

Using a two-plane parametrization (cf Figure 1(c)) the general structure of a light field can be considered as a 4D function

$$L : \Omega \times \Pi \to \mathbb{R}, \qquad (\boldsymbol{p}, \boldsymbol{q}) \mapsto L(\boldsymbol{p}, \boldsymbol{q}) \tag{1}$$

where $\boldsymbol{p} := (x, y)^{\mathrm{T}}$ and $\boldsymbol{q} := (\xi, \eta)^{\mathrm{T}}$ represent coordinate pairs in the image plane $\Omega \subset \mathbb{R}^2$ and in the lens plane $\Pi \subset \mathbb{R}^2$, respectively.



(a) (b) (c)

Fig. 2. Raw image data captured with a *plenoptic 1.0* camera. (a) shows the complete raw image and (b) and (c) are closeup views, which show the effect of the micro-lens array. Each micro-lens splits the incoming light into rays of different directions, where each ray hits the image sensor behind the micro-lens at a different location. Thus the use of such a micro-lens array makes it possible to capture the 4D light field.

There are several different visualizations of the light field data. The most obvious one, in the case of a plenoptic camera, is the raw image recorded by the sensor itself. This raw image is a composition of small discs, where each disc represents the image of a specific micro-lens. A typical example of a raw image obtained by a plenoptic camera is shown in Figure 2. Another representation can be obtained by extracting all values out of the raw image, which correspond to light-rays with the same direction. In the case of a *plenoptic 1.0* this means one has to consider all image values which are located at the same position in the disc-like image of each micro-lens. As in [18], images obtained in such a way will be referred to as sub-aperture images. In terms of the 4D light field $L(\boldsymbol{p}, \boldsymbol{q})$, a sub-aperture image is an image obtained by holding a direction $\boldsymbol{q}$ fixed and varying over all image positions $\boldsymbol{p}$. This sub-aperture representation provides an interesting interpretation of the light field data as a series of images with slightly different viewpoints, which are parallel to a common image plane. This shows that the light field provides information about the scene geometry. We also want to mention a more abstract visualization of the light field, which goes under the name epipolar image. An epipolar image is a slice of the light field, where one coordinate of the position $\boldsymbol{p}$ and one coordinate of the direction $\boldsymbol{q}$ is held constant.

It has been shown, that the light field can be used for different image processing tasks, like *e.g.* digital refocusing [13, 18], extending the depth of field [18], digital correction of lens aberrations [18], super-resolution [4, 24], and depth estimation [3, 23]. In this paper we will focus mainly on the latter task of depth calculation.

There has been done a lot of research in developing algorithm for traditional stereo estimation, but there is a lack of algorithms, which are capable of exploiting the structure within a light field. To the best of our knowledge, there are two recent works, which consider the task of depth estimation using a plenoptic camera. First, Bishop and Favaro [3] proposed an iterative multi-view method for depth estimation which considers all possible combinations of sub-aperture images on a rectangular grid aligned with the micro-lens center. Second, Wanner and Goldluecke [23] proposed a method for depth labeling of light fields, where they make use of the epipolar representation of

the light field. They also propose to enforce additional global visibility constraints, but even without this additional step, their method is computational expensive.

2 Methodology

The proposed method is motivated by the idea of Active Wavefront Sampling (AWS)[11], which is a 3D surface imaging technique, that uses a conventional camera and a so called AWS module. The simples type of an AWS module is an off-axis aperture, which is rotated around the optical axis. This circular movement of the aperture results in a rotation of a scene point's image on the image plane. This has the effect, that the scene point's depth is encoded by the radius of the according rotation on the image plane. A scene point located on the in-focus plane of the main lens will have a zero radius and thus its image will remain constant throughout all aperture positions, whereas scene points located at increasing distances from the in-focus plane will rotate on circles with increasing radii. Note, that a scene point which is located behind the in-focus plane will have an image rotation that is shifted by π on the rotation circle, compared to a scene point in-front of the in-focus plane. After calculating the circle rotation, the true depth can be obtained by simple geometric considerations.

Due to the fact that an image recorded with a traditional camera and an additional AWS module is similar to a specific sub-aperture image extracted from the light field, it is possible to apply the main idea of AWS to the light field setting. More precisely, in the case of a *plenoptic 1.0 camera*, *i.e.* that we have to extract sub-aperture images, where the directional positions lie on a circle centered at the origin of the lens plane Π.

Contrary to AWS, the light field provides much more information than just sub-aperture images corresponding to a single rotating off-axis aperture. The light field data allows to extract sub-aperture images corresponding to arbitrary circles and, what is even more important, it also allows to extract the center view. Moreover, it should also be mentioned, that in the light field setting the different sub-aperture images correspond to the same point in time. In the AWS setting images are captured at different times, where the time-difference depends on the time needed to mechanically move the rotating off-axis aperture from one position to the next.

We also want to note, that, in the AWS setting as well as in the light field setting, a circle movement is not the only path, which can be used. In general an arbitrary path can be used, and depth can be recovered as long as the path is known. However, in this work we will restrict ourselves to circular paths. On the one hand circular paths are used due to the fact that such patterns match the raw image data better than rectangular ones, and thus reduce unwanted effects like vignetting. On the other hand the use of circular patterns also simplifies the model.

3 Shape from Light Field Model

We now continue with formulating the proposed stereo model, which is based on variational principles. As an example we will briefly describe the general variational problem for stereo estimation, which is usually written as follows:

$$\min_d \lambda \Psi(d; I_L, I_R) + \Phi(d)\,, \tag{2}$$

where Ψ measures the data fidelity between two image I_L and I_R for a given disparity map d, Φ is a regularization function, and $\lambda \geqslant 0$ weights the influence of the data term. Common fidelity terms are *e.g.* the Euclidean norm, or the ℓ^1 norm.

The problem of stereo matching is in general ill-posed, therefore additional assumptions about the disparity map d are needed. This prior knowledge is added via the regularization function Φ. A popular regularization term is the Total Variation $\mathrm{TV}(x) = \int \mathrm{d}|\nabla x|$, which favors piecewise constant solutions.

In what follows we will present the proposed model for estimating a depth map for a given light field $L(\boldsymbol{p}, \boldsymbol{q})$. The model can be seen as a specialized multi-view stereo matching approach, where the rotation radius of a scene point's image is measured with respect to a given center position. We will first describe the data-fidelity term and discuss afterwards suitable regularization functions.

3.1 Data Fidelity Term

In our model we assume a *plenoptic 1.0 camera*, and we denote with $u : \Omega \to \mathbb{R}$ a function which defines for each scene point, captured at the position $\boldsymbol{p} \in \Omega$ in the center view $L(\cdot, \mathbf{0})$, the corresponding scene point's largest image rotation radius. This allows us to state the following energy in the continuous setting:

$$E_{data}(u) = \int_{\Omega} \int_{0}^{R} \int_{0}^{2\pi} \psi_{s,r}(\boldsymbol{p}, u(\boldsymbol{p}))\ \mathrm{d}(s,\, r,\, \boldsymbol{p})\,, \tag{3}$$

with

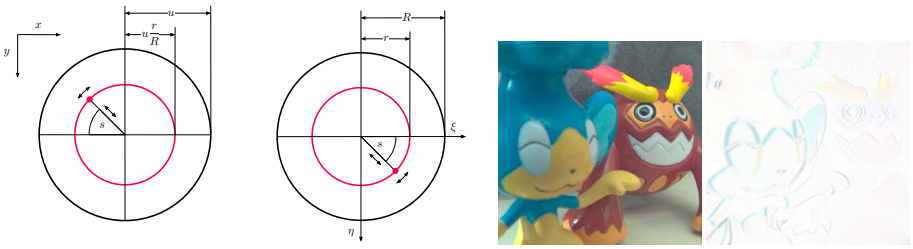
$$\psi_{s,r}(\boldsymbol{p}, u(\boldsymbol{p})) = \theta\left(L\left(\boldsymbol{p}, \mathbf{0}\right) - L\left(\boldsymbol{p} - u(\boldsymbol{p})\frac{\varphi_{s,r}}{R}\,, \varphi_{s,r}\right)\right)\,, \tag{4}$$

where $\varphi_{s,r} = r\left(\cos(s), \sin(s)\right)^{\mathrm{T}}$ is a circle parametrization with radius r and center at the origin, $\Omega \subset \mathbb{R}^2$ is the image domain, $\theta(\cdot)$ denotes an error estimator, and $R > 0$ is the predefined largest allowed circle radius in the lens plane, which corresponds to the largest possible aperture radius of the main lens. Thus, $\psi_{s,r}(\boldsymbol{p}, u(\boldsymbol{p}))$ measures the brightness difference between the center view $L(\cdot, \mathbf{0})$ at position $\boldsymbol{p}$ and the sub-aperture image $L(\cdot, \varphi_{s,r})$ at position $\boldsymbol{p} - u(\boldsymbol{p})\frac{\varphi_{s,r}}{R}$. Here, the latter position describes for varying s a circle in the image plane centered at $\boldsymbol{p}$ and with radius $u(\boldsymbol{p})\frac{r}{R}$ (cf Figure 3).

As already mentioned above, common fidelity terms are the Euclidean norm or the ℓ^1 norm. The corresponding error functions are the quadratic differences and the absolute differences, respectively. The Euclidean norm provides the advantage of being differentiable, which allows to apply standard optimization techniques. But it comes with the disadvantage of not being robust to outliers, which occur in areas of occlusions. The ℓ^1 norm on the other hand is non-smooth, but it is more robust to outliers, and hence we will make us of it, *i.e.* we choose $\theta(x) = |x|$.

In order to obtain a convex approximation of the data term (3) we use first-order Taylor approximations for the sub-aperture images, *i.e.*

$$\begin{aligned} &L\Big(\boldsymbol{p} - u(\boldsymbol{p})\frac{\varphi_{s,r}}{R}, \varphi_{s,r}\Big) \approx \\ &\quad L\left(\boldsymbol{p} - u_0(\boldsymbol{p})\frac{\varphi_{s,r}}{R}, \varphi_{s,r}\right) + \left(u(\boldsymbol{p}) - u_0(\boldsymbol{p})\right)\ \frac{r}{R}\,\nabla_{-\frac{\varphi_{s,r}}{r}} L\left(\boldsymbol{p} - u_0(\boldsymbol{p})\frac{\varphi_{s,r}}{R}, \varphi_{s,r}\right), \end{aligned} \tag{5}$$



(a) rotation in Ω (b) rotation in Π (c) center view (d) difference image

Fig. 3. (a) and (b) illustrate the parametrization used in (3). (a) sketches a scene point's image position (purple dot) and the corresponding rotation circle, and (b) shows the according directional sampling position in the lens plane for extracting the sub-aperture image. (c) and (d) provide a visualization of the scene point's image rotation. (c) is the center view and (d) shows the color inverted difference image between two sub-aperture images . The two fixed directional components of the light field, which are used to extract the sup-aperture images, are chosen to lie opposite to each other on a circle centered at zero.

where $\nabla_{-\frac{\varphi_{s,r}}{r}}$ denotes the directional derivative, with direction $[-\frac{\varphi_{s,r}}{r}, \mathbf{0}]^{\mathrm{T}}$.

Finally we want to note, that although the ℓ^1 norm is robust to outliers it is not robust to varying illumination. Thus, *e.g.* vignetting effects of the main lens and the microlenses, might lead to problems. In order to be more robust against illumination changes, we apply a structure-texture decomposition [25] on the sub-aperture images, *i.e.* we pre-process each image by removing its low frequency component.

3.2 Regularization Term

In this section we will briefly discuss different regularization terms, which can be added to the data-fidelity term proposed in Section 3.1. A regularization term is needed due to the fact, that the problem of minimizing (3) with respect to u is ill-posed, and therefore the fidelity term alone does not provide sufficient information to calculate a reliable solution. Thus, we additionally assume that u varies smoothly almost everywhere in the image domain. In order to add this assumption to our model, we will use an extension of Total Generalized Variation (TGV) [5]. As indicated by the name, TGV is a generalization of the famous Total Variation (TV). Whereas TV favors piecewise constant solutions, TGV^k favors piecewise polynomial solutions of order $k-1$, *e.g.* TGV^2 will favor piecewise linear solutions. Following the work by Ranftl *et al.* [21], we extend TGV^2 by using an anisotropic diffusion tensor $D^{\frac{1}{2}}$. This diffusion tensor connects the prior with the image content, which leads to solutions with a lower degree of smoothness around depth edges. This image-driven TGV regularization term can be written as

$$E_{reg}(u) = \min_{\boldsymbol{w}} \alpha_1 \|D^{\frac{1}{2}}(\nabla u - \boldsymbol{w})\|_{\mathcal{M}} + \alpha_0 \|\boldsymbol{\nabla} \boldsymbol{w}\|_{\mathcal{M}}\,, \tag{6}$$

where $\|\cdot\|_{\mathcal{M}}$ denotes a Radon norm for vector-valued and matrix-valued Radon measures, and $\alpha_0, \alpha_1 > 0$. Furthermore, as in [21]

$$D^{\frac{1}{2}} = \exp(-\gamma\,|\nabla I|^{\beta})\,\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}} + \boldsymbol{n}^{\perp}\boldsymbol{n}^{\perp^{\mathrm{T}}}, \tag{7}$$

where I denotes the center view $L(\cdot, \mathbf{0})$, $\boldsymbol{n}$ is the normalized gradient, $\boldsymbol{n}^{\perp}$ is a vector perpendicular to $\boldsymbol{n}$, and γ and β are predefined scalars. Combining the data term in (3) and the above regularization term (6) leads to a robust model, which is capable of reconstructing subpixel accurate depth maps.

3.3 Discretization

By discretizing the spatial domain Ω, and the involved circles in (3) we obtain

$$\hat{E}_{data}(u) = \sum_{\boldsymbol{p}\in\hat{\Omega}} \sum_{i=1}^{M} \sum_{j=1}^{N} \psi_{s_j,r_i}(\boldsymbol{p}, u(\boldsymbol{p})), \quad \text{with} \quad s_j = \frac{2\,\pi(j-1)}{N} \text{ and } r_i = \frac{R\,i}{M}, \tag{8}$$

where $\hat{\Omega} := \{(x, y)^{\mathrm{T}} \in \mathbb{N}_0^2 \,|\, x < n,\, y < m\}$ denotes the discrete image domain. Furthermore, M and N are the number of different sampling circles, and the number of uniform sampling positions on each circle, respectively.

3.4 Optimization

In order to optimize the complete discretized problem

$$\min_{u\in\mathbb{R}^{mn}} \lambda\,\hat{E}_{data}(u) + \hat{E}_{reg}(u) \tag{9}$$

we use a primal-dual algorithm, proposed by Chambolle *et al.* [7]. Therefore, we first have to rewrite (9) as a saddle point problem. Note, that $u \in \mathbb{R}^{mn}$ is now represented as a column vector, and $\hat{E}_{reg}(u)$ denotes the discrete version of (6). To simplify notation we first define $\tilde{A}_{ij}$ and $B_{ij} \in \mathbb{R}^{mn}$ as

$$\tilde{A}_{ij} := \left(\frac{r_i}{R}\,\nabla_{-\frac{\varphi_{s_j,r_i}}{r_i}} L\left(\boldsymbol{p} - u_0(\boldsymbol{p})\frac{\varphi_{s_j,r_i}}{R}, \varphi_{s_j,r_i}\right)\right)_{\boldsymbol{p}\in\hat{\Omega}}, \tag{10}$$

$$B_{ij} := \left(L\,(\boldsymbol{p}, \mathbf{0}) - L\left(\boldsymbol{p} - u_0(\boldsymbol{p})\frac{\varphi_{s_j,r_i}}{R}, \varphi_{s_j,r_i}\right)\right)_{\boldsymbol{p}\in\hat{\Omega}}, \tag{11}$$

and by setting $A_{ij} := \mathrm{diag}(\tilde{A}_{ij})$, it is possible to formulate (9) equivalently as the following saddle point problem

$$\min_{u,\boldsymbol{w}} \max_{\substack{\|\boldsymbol{p_u}\|_\infty \le 1 \\ \|\boldsymbol{p_w}\|_\infty \le 1 \\ \|\boldsymbol{p_{ij}}\|_\infty \le 1}} \Big\{ \lambda \sum_{i=1}^{M}\sum_{j=1}^{N} \langle B_{ij} - A_{ij}(u - u_0), \boldsymbol{p_{ij}}\rangle + \alpha_1 \Big\langle D^{\frac{1}{2}}(\nabla u - \boldsymbol{w}), \boldsymbol{p_u}\Big\rangle + \alpha_0\,\langle \boldsymbol{\nabla w}, \boldsymbol{p_w}\rangle \Big\}, \tag{12}$$

Algorithm 1. Primal-Dual Algorithm for Shape from Light Field

Require: Choose $\sigma > 0$ and $\tau > 0$, s.t. $\tau\sigma = 1$. Set Σ_{p_u}, Σ_{p_w}, $\Sigma_{p_{ij}}$, T_u, and T_w as in (13), $n = 0$, and the rest arbitrary.

while $n < iter$ **do**

// Dual step

$$\boldsymbol{p_u^{n+1}} \leftarrow \mathcal{P}_{\{\|\boldsymbol{p_u}\|_\infty \leqslant 1\}} \left(\boldsymbol{p_u^n} + \sigma \Sigma_{p_u} \alpha_1 \left(D^{\frac{1}{2}} \left(\nabla \bar{u}^n - \bar{\boldsymbol{w}}^{\boldsymbol{n}} \right) \right) \right)$$

$$\boldsymbol{p_w^{n+1}} \leftarrow \mathcal{P}_{\{\|\boldsymbol{p_w}\|_\infty \leqslant 1\}} \left(\boldsymbol{p_w^n} + \sigma \Sigma_{p_w} \alpha_0 \left(\boldsymbol{\nabla} \bar{\boldsymbol{w}}^{\boldsymbol{n}} \right) \right)$$

$$\boldsymbol{p_{ij}^{n+1}} \leftarrow \mathcal{P}_{\{\|\boldsymbol{p_{ij}}\|_\infty \leqslant 1\}} \left(\boldsymbol{p_{ij}^n} + \sigma\lambda \left(B_{ij} - A_{ij}(\bar{u}^n - u_0) \right) \right)$$

// Primal step

$$u^{n+1} \leftarrow u^n - \tau \, \mathrm{T}_u \left(\alpha_1 \nabla^{\mathrm{T}} \left(D^{\frac{1}{2}} \boldsymbol{p_u^{n+1}} \right) + \lambda \textstyle\sum_{i,j} A_{ij} \boldsymbol{p_{ij}^{n+1}} \right)$$

$$\boldsymbol{w^{n+1}} \leftarrow \boldsymbol{w^n} - \tau \, \mathrm{T}_w \left(\alpha_0 \boldsymbol{\nabla}^{\mathrm{T}} \boldsymbol{p_w^{n+1}} - \alpha_1 D^{\frac{1}{2}} \boldsymbol{p_u^{n+1}} \right)$$

$$\bar{u}^{n+1} \leftarrow 2\, u^{n+1} - u^n$$

$$\bar{\boldsymbol{w}}^{\boldsymbol{n+1}} \leftarrow 2\, \boldsymbol{w^{n+1}} - \boldsymbol{w^n}$$

// Iterate

$n \leftarrow n + 1$

end while

where the dual variables $\boldsymbol{p_{ij}}$ have the same dimension as B_{ij}. Now we can directly apply the algorithm proposed in [7] to solve (12).

An improvement with respect to convergence speed can be obtained by using adequate symmetric and positive definite preconditioning matrices [20], leading to the update scheme shown in Algorithm 1, which is iterated for a fixed number of iterations or till a suitable convergence criterion is fulfilled. Here $\mathcal{P}$ denotes a reprojection operator, and Σ_*, and T_* are the preconditioning matrices, given as follows

$$\begin{aligned} \Sigma_{p_u} = \frac{1}{3} I, \;\; \Sigma_{p_w} = \frac{1}{2} I, \;\; \Sigma_{p_{ij}} = I, \;\; \mathrm{T}_w = \frac{1}{\alpha_1^2 + 4\alpha_0^2} I \\ \mathrm{T}_u = \left(\alpha_1 \operatorname{diag}\left(D_x^{\mathrm{T}} D_x + D_y^{\mathrm{T}} D_y \right) \right)^{-1} + \sum_{i,j} \left(\lambda A_{ij} \right)^{-2}, \end{aligned} \tag{13}$$

where I denotes the identity matrix, $\left(\begin{smallmatrix} D_x \\ D_y \end{smallmatrix}\right) := D^{\frac{1}{2}} \nabla$, and $\operatorname{diag}(X)$ takes a matrix X and sets all elements, which are not on the main diagonal to zero.

Due to the fact, that the linear approximation (5) is only accurate in a small area around u_0 we embed the complete algorithm into a coarse-to-fine warping scheme [6].

Fig. 4. Examples of generated synthetic data. The figure shows pairs of images, where one is a closeup view of the synthetic raw image data and the other represents a corresponding sub-aperture image. From left to right the pairs show clean data, data with added vignetting effects, and data with added vignetting effect and additive Gaussian noise.

4 Experimental Results

In this section we first evaluate our method using the Light Field Benchmark Dataset (LFBD)[1], which is a dataset of synthetic light fields created with Blender[2]. All light field scenes within the dataset have a directional resolution of 9×9 pixels per micro-lens, and varying spatial resolutions, which are listed in Table 1. After the synthetic evaluation we will also present some qualitative results for real world data, where we use raw images captured with a Lytro[3] camera as input for the proposed algorithm.

4.1 Synthetic Experiments

We first evaluate our algorithm on synthetic images similar to the ones taken with a plenoptic camera (cf Figure 4), where the images are created using the LFBD. For the synthetic experiments we set $M = 1$ and $N = 8$, which means that we use a single circle and 8 sampling positions per circle. Other parameters are tuned for the different scenes. Figure 5 shows qualitative depth map results for the proposed approach as well as the ground truth data provided in the LFBD. It can be observed, that the proposed approach is capable of creating piecewise smooth depth maps with clear depth discontinuities. Further, we also visualize the relative depth errors in green (red), which are smaller (larger) than 0.2% and 1.0% in column four and five, respectively. It can be seen that the remaining errors are concentrated at occlusion boundaries or at positions of specular highlights.

Next we simulate vignetting effects of the micro-lenses and the main lens as well as additive image noise. For the vignetting effect of the main lens and micro-lenses we reduce the brightness of pixel based on their distance from the image center and the micro-lens center, respectively. As image noise we use additive Gaussian noise with zero mean and a variance with a σ equal to 2% of the image dynamic range. Example images are shown in Figure 4 and the quantitative results can be found in Table 1. To justify the high quality of our depth map results we also provide the results for the variational

[1] `http://lightfield-analysis.net`
[2] `http://www.blender.org/`
[3] `https://www.lytro.com/`

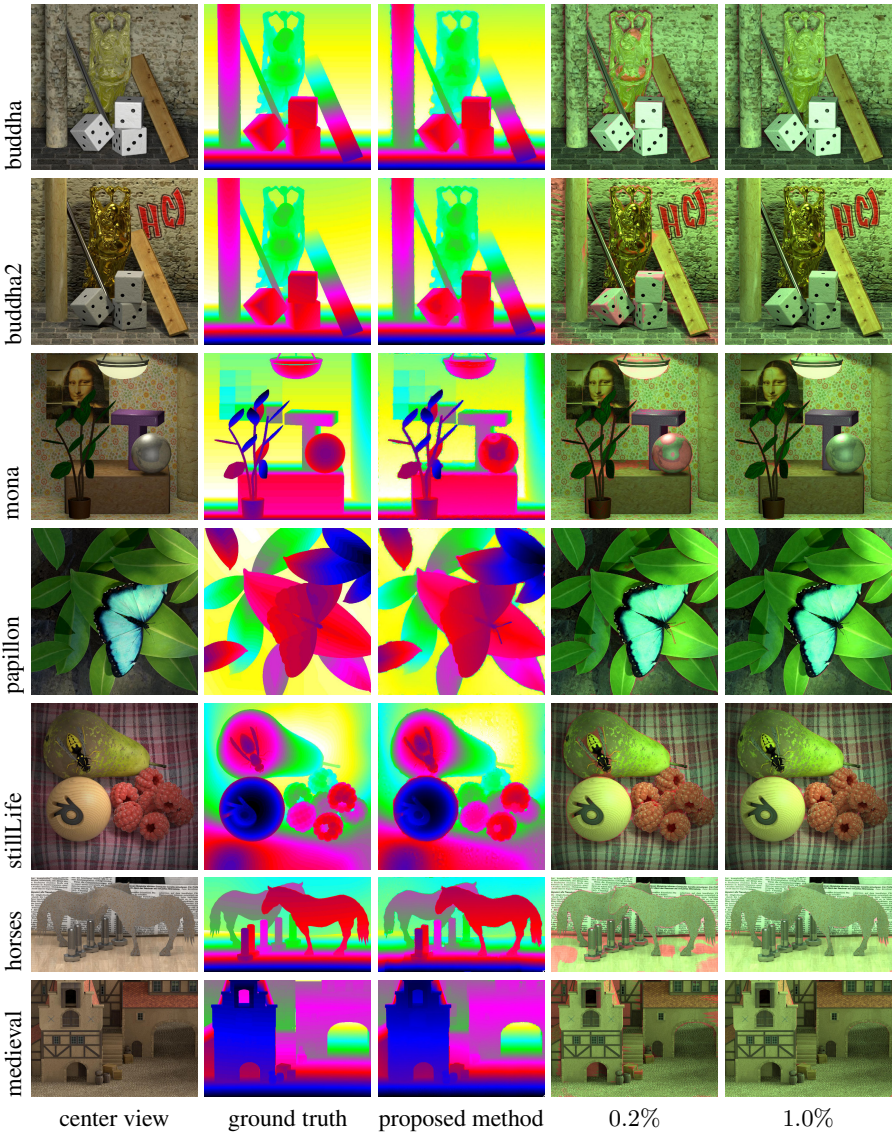


Fig. 5. Qualitative results for synthetic scenes. All scenes have a directional resolution of 9×9 pixels per micro-lens, and varying spatial resolutions, which are listed in Table 1. The figure shows from left to right, the center view, the color coded ground truth depth map, the color coded depth map results for the proposed method, and maps which indicate in green (red) the pixels with a relative depth error of less (more) than 0.2% and 1.0%, respectively.

Table 1. Quantitative results for the scenes shown in Figure 5. The table shows the percentage of pixels with a relative depth error of more than 0.2%, 0.5%, and 1.0%, for different synthetic scenes.

	clean			**vignetting**			**vignetting & noise**			spatial
	1.0%	0.5%	0.2%	1.0%	0.5%	0.2%	1.0%	0.5%	0.2%	resolution
buddha	2.12	3.91	8.37	2.23	4.41	10.38	2.42	4.97	15.28	768×768
buddha2	1.15	2.44	15.05	1.69	4.40	21.81	1.53	3.86	23.33	768×768
mona	2.08	4.66	12.90	2.44	6.47	19.54	2.82	9.46	22.11	768×768
papillon	2.42	3.82	8.79	2.37	4.17	12.97	3.08	6.26	19.12	768×768
stillLife	0.96	2.32	6.33	1.00	2.17	6.89	0.94	2.17	6.02	768×768
horses	1.98	4.47	16.83	2.56	6.98	21.37	3.71	11.05	27.86	1024×576
medieval	1.53	2.93	11.09	1.70	3.50	18.78	1.69	4.54	19.74	1024×720
average	1.75	3.51	11.34	2.00	4.59	15.96	2.31	6.04	19.07	

Table 2. Quantitative results of methods proposed by Wanner and Goldluecke [23] evaluated on an older version of the LFBD. The results are taken from [23], and provided here as a reference point to the results shown in Table 1. The scenes are quite similar to the ones shown in Figure 5 (especially the buddha* scene is nearly identical to the buddha2 scene of the current LFBD). The table shows the percentage of pixels with a relative depth error of more than 0.2% and 1.0%, for the different methods proposed in [23].

	Local [23]		**Global [23]**		**Consistent [23]**		directional	spatial
	1.0%	0.2%	1.0%	0.2%	1.0%	0.2%	resolution	resolution
conehead*	22.9	78.5	1.3	51.0	1.1	48.9	21×21	500×500
mona*	57.0	91.9	25.7	87.7	19.9	84.5	41×41	512×512
buddha*	20.4	73.6	4.1	61.7	2.9	60.4	21×21	768×768
average	33.4	81.3	10.4	66.8	8.0	64.6		

depth labeling approach presented in [23] (cf Table 2). These results were generated using an older version of the LFBD, which is unfortunately no longer available. However, certain scenes like *e.g.* the buddha* scene is nearly identical to the buddha2 scene in the current LFBD, which allows to draw a comparison. By doing so, we see that the proposed method outperforms the method in [23] in terms of accuracy by a large margin. Moreover, depending on the spatial resolution of the input images, the proposed approach takes about 5-10 seconds to compute, whereas the global approach and the consistent approach proposed in [23] take 2-10 minutes and several hours, respectively.

4.2 Real World Experiments

In this section we present some qualitative real world results obtained by using the proposed shape from light field method. For light field capturing we use a Lytro camera. Such a camera provides a spatial resolution of around 380×330 micro-lenses and a

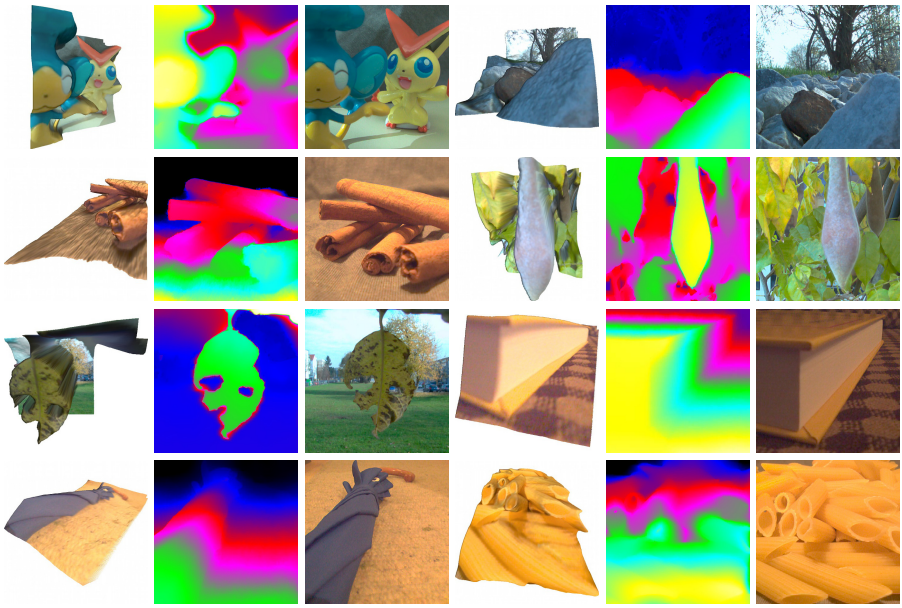


Fig. 6. Qualitative results, which demonstrate the effectiveness of the proposed model. The figure shows depth map results in terms of 3D views and color-coded disparity maps as well as all-in-focus results for various scenes.

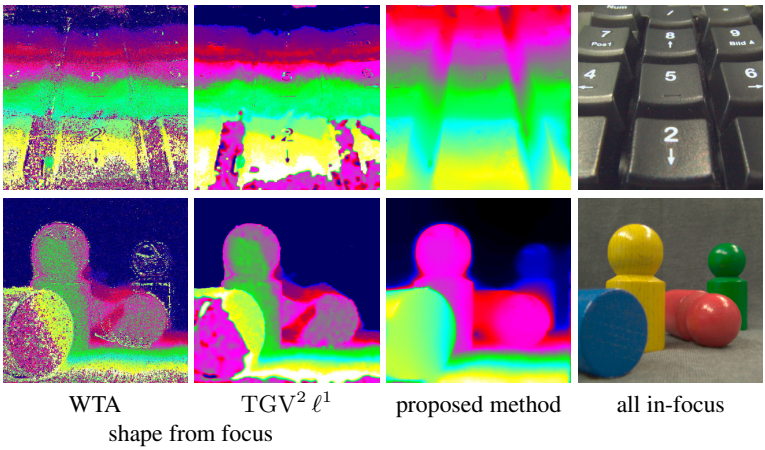


Fig. 7. Comparison to shape-from-focus. The figure shows shape from focus results in terms of the *winner-takes-all* (WTA) solution and a TGV^2 regularized version, where 12 digital refocused images (provided by the commercial Lytro software) were used as input. Furthermore, the figure shows the calculated depth maps of the proposed method, and corresponding all in-focus images.

directional resolution of about 10×10 pixels per micro-lens. For the real world experiments we set $M = 1$ and $N = 16$, and we again tune the other parameters. Figure 6 provides results of the proposed method for different scenes. Among others, Figure 6 shows 3D views created using the calculated depth maps, as well as color-coded disparity maps. Although the spatial resolution provided by the Lytro camera is quite low and the extracted sub-aperture images are quite noisy, the proposed method is again able to calculate piecewise smooth depth maps, with clearly visible depth discontinuities.

It should also be mentioned, that by using the calculated image rotation u, one can easily generate an all-in-focus image by summing up corresponding image locations in the sub-aperture images. All-in-focus results are shown in the third and sixth column of Figure 6.

In a final experiment we compare our results to the commercial Lytro software. Here we compare the proposed approach with a shape-from-focus approach [17]. For the shape-from-focus approach we use 12 digital refocused images, provided by the Lytro software, where we apply a high-pass filter on the images to measure the in-focus area in each image. The qualitative results in terms of the *winner-takes-all* (WTA) solution and a TGV^2 regularized versions are shown in Figure 7. It can be seen, that the results generated with the proposed approach are clearly superior.

5 Conclusion

In this work we proposed a method for calculating a high quality depth map out of a given light field, which was captured with a plenoptic camera. We first showed that it is possible to extract sub-aperture images out of the raw light field data, which are similar to those images captured with a rotating off-axis aperture in the AWS setting. Based on this observation we formulated a variational model which measures a dense field of scene point's image rotation radii over certain sub-aperture images, where these rotation radii encode depth information of the scene points in the image.

In the experiment section we showed on synthetic and real world examples that our model is capable of generating high-quality depth maps. We simulated vignetting effects and image noise on synthetic data. Moreover, we also showed that our model is robust enough to generate piecewise smooth depth maps with sharp depth discontinuities, out of the noisy and highly aliased sub-aperture images extracted form the raw light field data, which was captured with a Lytro camera.

References

1. Adelson, E.H., Bergen, J.R.: The plenoptic function and the elements of early vision. In: Computational Models of Visual Processing, pp. 3–20. MIT Press (1991)
2. Adelson, E.H., Wang, J.Y.A.: Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 99–106 (1992)
3. Bishop, T., Favaro, P.: Plenoptic depth estimation from multiple aliased views. In: 12th International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1622–1629. IEEE (2009)

4. Bishop, T.E., Favaro, P.: The light field camera: Extended depth of field, aliasing, and superresolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(5), 972–986 (2012)
5. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences 3(3), 492–526 (2010)
6. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
7. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision 40, 120–145 (2011)
8. Coffey, D.F.W.: Apparatus for making a composite stereograph (December 1936)
9. Dudnikov, Y.A.: Autostereoscopy and integral photography. Optical Technology 37(3), 422–426 (1970)
10. Fife, K., Gamal, A.E., Philip Wong, H.S.: A 3mpixel multi-aperture image sensor with 0.7μm pixels in 0.11μm cmos (February 2008)
11. Frigerio, F.: 3-dimensional Surface Imaging Using Active Wavefront Sampling. PhD thesis, Massachusetts Institute of Technology (2006)
12. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: SIGGRAPH, pp. 43–54 (1996)
13. Isaksen, A., McMillan, L., Gortler, S.J.: Dynamically reparameterized light fields. In: SIGGRAPH, pp. 297–306 (2000)
14. Levoy, M., Hanrahan, P.: Light field rendering. In: SIGGRAPH, pp. 31–42 (1996)
15. Lippmann, R.: La photographie intégrale. Comptes-Rendus, Académie des Sciences 146, 446–551 (1908)
16. Lumsdaine, A., Georgiev, T.: The focused plenoptic camera. In: Proc. IEEE ICCP, pp. 1–8 (2009)
17. Nayar, S., Nakagawa, Y.: Shape from Focus. IEEE Transactions on Pattern Analysis and Machine Intelligence 16(8), 824–831 (1994)
18. Ng, R.: Digital Light Field Photography. Phd thesis, Stanford University (2006)
19. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Technical report, Stanford University (2005)
20. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: International Conference on Computer Vision (ICCV), pp. 1762–1769. IEEE (2011)
21. Ranftl, R., Gehrig, S., Pock, T., Bischof, H.: Pushing the limits of stereo using variational stereo estimation. In: Intelligent Vehicles Symposium, pp. 401–407. IEEE (2012)
22. Vaish, V., Wilburn, B., Joshi, N., Levoy, M.: Using plane + parallax for calibrating dense camera arrays. In: Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2–9 (2004)
23. Wanner, S., Goldluecke, B.: Globally consistent depth labeling of 4D lightfields. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2012)
24. Wanner, S., Goldluecke, B.: Spatial and angular variational super-resolution of 4D light fields. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part V. LNCS, vol. 7576, pp. 608–621. Springer, Heidelberg (2012)
25. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An Improved Algorithm for TV-L^1 Optical Flow. In: Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.) Visual Motion Analysis. LNCS, vol. 5604, pp. 23–45. Springer, Heidelberg (2009)