Globally Consistent Depth Labeling of 4D Light Fields

Sven Wanner and Bastian Goldluecke
Heidelberg Collaboratory for Image Processing

Abstract

We present a novel paradigm to deal with depth reconstruction from 4D light fields in a variational framework. Taking into account the special structure of light field data, we reformulate the problem of stereo matching to a constrained labeling problem on epipolar plane images, which can be thought of as vertical and horizontal 2D cuts through the field. This alternative formulation allows to estimate accurate depth values even for specular surfaces, while simultaneously taking into account global visibility constraints in order to obtain consistent depth maps for all views. The resulting optimization problems are solved with state-of-the-art convex relaxation techniques. We test our algorithm on a number of synthetic and real-world examples captured with a light field gantry and a plenoptic camera, and compare to ground truth where available. All data sets as well as source code are provided online for additional evaluation.

1. Introduction

The 4D light field has been established as a promising paradigm to describe the visual appearance of a scene. Compared to a traditional 2D image, it contains information about not only the accumulated intensity at each image point, but separate intensity values for each ray direction. Thus, the light field implicitly captures 3D scene geometry and reflectance properties.

The additional information inherent in a light field allows a wide range of applications. Popular in computer graphics, for example, is light field rendering, where the scene is displayed from a virtual viewpoint [18, 14]. The light field data also allows to add effects like synthetic aperture, i.e. virtual refocusing of the camera, stereoscopic display, and automatic glare reduction as well as object insertion and removal [11, 10, 6].

In the past, light fields have been very hard to capture and required expensive custom-made hardware to be able to acquire several views of a scene. Straightforward but hardware-intensive are camera arrays [22]. Somewhat more practical and less expensive is a gantry construction consisting of a single moving camera [20], which is however restricted to static scenes, see figure 2. Recently, however, the first commercial plenoptic cameras have become available on the market. Using an array of microlenses, a single one of these cameras essentially captures an full array of views simultaneously. This makes such cameras very attractive for a number of industrial applications, in particular depth estimation and surface inspection, and they can also acquire video streams of dynamic scenes [8, 15, 16].

Figure 1. *Our novel paradigm for depth reconstruction in a light field allows to estimate highly accurate depth maps even for specular scenes.*

Naturally, this creates a high demand for efficient and robust algorithms which reconstruct information directly from light fields. However, while there has been a lot of work on for example stereo and optical flow algorithms for traditional image pairs, there is a lack of similar modern methods which are specifically tailored to the rich structure inherent in a light field. Furthermore, much of the existing analysis is local in nature, and does not enforce global consistency of results.

Contributions. In this paper, we introduce a framework for *variational light field analysis*, which is designed to enable the application of modern continuous optimization methods to 4D light field data. Here, we specifically address the problem of depth estimation, and make a number of important contributions.

- We introduce a novel local data term for depth estimation, which is tailored to the structure of light field data and much more robust than traditional stereo matching methods to non-Lambertian objects.

- We introduce a labeling scheme based on state-of-the-art convex relaxation methods [17, 21], which allows to estimate globally consistent depth maps which satisfy visibility constraints for all (possibly hundreds of) views simultaneously.

The robustness of our method is demonstrated on a large range of data sets, from synthetic light fields with ground truth available up to real-world examples from several sources, including a recent plenoptic camera. Source code for the method and all of our data sets are provided online on our web page.

2. Related work

The concept of light fields originated mainly in computer graphics. In this field, image based rendering [19] is a common technique to render new views from a set of images of a scene. Adelson and Bergen [1] as well as McMillan and Bishop [14] treated view interpolation as a reconstruction of the *plenoptic function*. This function is defined on a seven-dimensional space and describes the entire information about light emitted by a scene, storing an intensity value for every 3D point, direction, wavelength and time. A dimensionality reduction of the plenoptic function to 4D, the so called Lumigraph, was introduced by Gortler *et al*. [9] and Levoy and Hanrahan [12]. They introduced the two plane parametrization, which we also adopt in our work, where each ray is determined by its intersections with two planes.

A main benefit of light fields compared to traditional images or stereo pairs is the expansion of the disparity space to a *continuous space*. This becomes apparent when considering *epipolar plane images (EPIs)*, which can be viewed as 2D slices of constant angular and spatial direction through the Lumigraph. Due to a dense sampling in angular direction, corresponding pixels are projected onto lines in EPIs, which can be more robustly detected than point correspondences.

One of the first approaches using EPIs to analyze scene geometry was published by Bolles *et al*. [5]. They detect edges, peaks and troughs with a subsequent line fitting in the EPI to reconstruct 3D structure. Another approach is presented by Criminisi *et al*. [6], who use an iterative extraction procedure for collections of EPI-lines of the same depth, which they call an *EPI-tube*. Lines belonging to the same tube are detected via shearing the EPI and analyzing photo-consistency in the vertical direction. They also propose a procedure to remove specular highlights from already extracted EPI-tubes.

There are also two less heuristic methods which work in an energy minimization framework. In Matousek *et al*. [13], a cost function is formulated to minimize a weighted path length between points in the first and the last row of an EPI, preferring constant intensity in a small neighborhood of each EPI-line. However, their method only works in the absence of occlusions. Berent *et al*. [2] deal with the simultaneous segmentation of EPI-tubes by a region competition method using active contours, imposing geometric properties to enforce correct occlusion ordering.

(a) LEGO gantry [20] (b) Plenoptic camera [16]

Figure 2. *Acquisition devices which captured the real-world 4D light fields on which we test our algorithm.*

In contrast to the above works, we perform a labeling for all points in the EPI simultaneously by using a state-of-the-art continuous convex energy minimization framework. We enforce globally consistent visibility across views by restricting the spatial layout of the labeled regions. Compared to methods which extract EPI-tubes sequentially [5, 6], this is independent of the order of extraction and does not suffer from an associated propagation of errors. While a simultaneous extraction is also performed in [2], they perform local minimization only and require good initialization, as opposed to our convex relaxation approach. Furthermore, they use a level set approach, which makes it expensive and cumbersome to deal with a large number of regions. As a further novelty to previous work, we suggest to employ the *structure tensor* of an EPI to obtain robust local depth estimates.

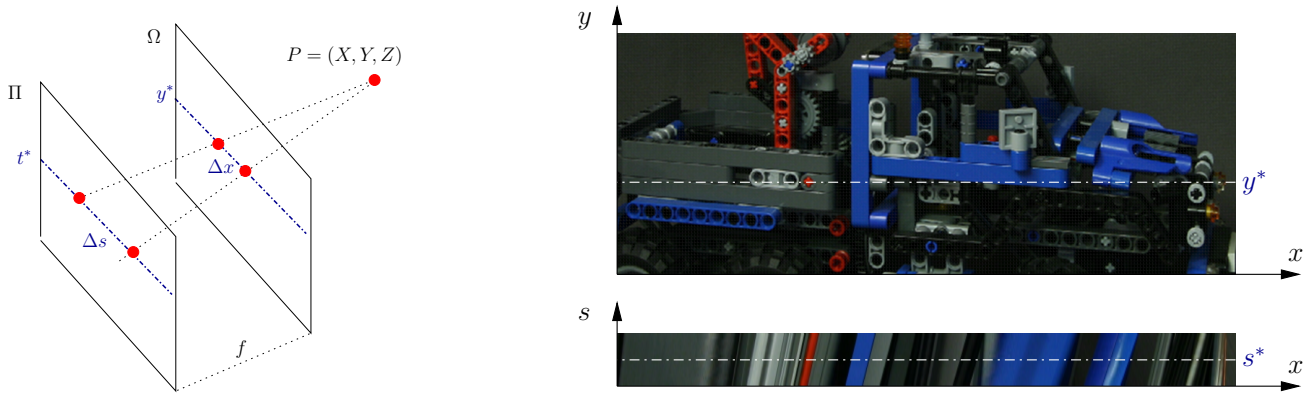
3. 4D light field structure

Several ways to represent light fields have been proposed. In this paper, we adopt the *two-plane parametrization*. One way to look at a 4D light field is to consider it as a collection of pinhole views from several view points parallel to a common image plane, figure 3. The 2D plane Π contains the focal points of the views, which we parametrize by the coordinates (s, t), and the image plane Ω is parametrized by the coordinates (x, y). A *4D light field* or *Lumigraph* is a map

$$L : \Omega \times \Pi \to \mathbb{R}, \qquad (x, y, s, t) \mapsto L(x, y, s, t). \tag{1}$$

It can be viewed as an assignment of an intensity value to the ray $R_{x,y,s,t}$ passing through $(x, y) \in \Omega$ and $(s, t) \in \Pi$.

For the problem of estimating 3D structure, we consider the structure of the light field, in particular on 2D slices



(a) Light field geometry

(b) Pinhole view at (s^*, t^*) and epipolar plane image S_{y^*,t^*}

Figure 3. *Each camera location (s^*, t^*) in the image plane Π yields a different pinhole view of the scene. By fixing a horizontal line of constant y^* in the image plane and a constant camera coordinate t^*, one obtains an epipolar plane image (EPI) in (x, s) coordinates. A scene point P is projected onto a line in the EPI due to a linear correspondence between its s- and projected x-coordinate, see figure (a) and equation (3).*

through the field. We fix a horizontal line of constant y^* in the image plane and a constant camera coordinate t^*, and restrict the light field to an (x, s) -slice Σ_{y^*,t^*}. The resulting map is called an *epipolar plane image (EPI)*,

$$\begin{aligned} S_{y^*,t^*} : \Sigma_{y^*,t^*} &\to \mathbb{R}, \\ (x, s) &\mapsto S_{y^*,t^*}(x, s) := L(x, y^*, s, t^*). \end{aligned} \tag{2}$$

Let us consider the geometry of this map, figure 3. A point $P = (X, Y, Z)$ within the epipolar plane corresponding to the slice projects to a point in Ω depending on the chosen camera center in Π. If we vary s, the coordinate x changes according to [5]

$$\Delta x = -\frac{f}{Z}\Delta s, \tag{3}$$

where f is the distance between the parallel planes[1].

Interestingly, a point in 3D space is thus projected onto a line in Σ_{y^*,t^*}, where the *slope of the line is related to its depth*. This means that the intensity of the light field should not change along such a line, provided that the objects in the scene are Lambertian. Thus, if we want to estimate the depth for a ray $R_{x,y,s,t}$, we can try to find the slope of level lines in the slices corresponding to the point. We now turn to formulating this problem as a variational labeling problem on the slice domains.

4. Local depth estimate on an EPI

We first consider how we can estimate the local direction of a line at a point (x, s) for an epipolar plane image S_{y^*,t^*}, where y^* and t^* are fixed. The case of vertical slices is analogous. The goal of this step is to compute a local *depth estimate* $l_{y^*,t^*}(x, s)$ for each point of the slice domain, as well as a *reliability estimate* $r_{y^*,t^*}(x, s) \in [0, 1]$, which gives a measure of how reliable the local depth estimate is. Both local estimates will used in subsequent sections to obtain a consistent depth map in a global optimization framework.

In order to obtain the local depth estimate, we need to estimate the direction of lines on the slice. This is done using the *structure tensor* J of the epipolar plane image $S = S_{y^*,t^*}$,

$$J = \begin{bmatrix} G_\sigma * (S_x S_x) & G_\sigma * (S_x S_y) \\ G_\sigma * (S_x S_y) & G_\sigma * (S_y S_y) \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix}. \tag{4}$$

Here, G_σ represents a Gaussian smoothing operator at an outer scale σ and S_x,S_y denote the gradient components of S calculated on an inner scale ρ.

The direction of the local level lines can then be computed via [3]

$$\boldsymbol{n}_{y^*,t^*} = \begin{bmatrix} J_{yy} - J_{xx} \\ 2J_{xy} \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix}, \tag{5}$$

from which we derive the local depth estimate via equation (3) as

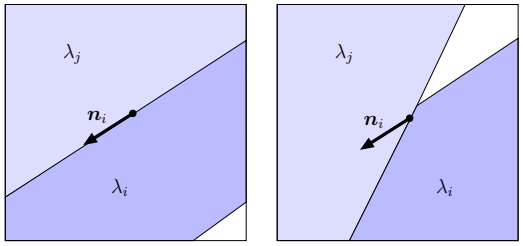
$$l_{y^*,t^*} = -f\frac{\Delta s}{\Delta x}. \tag{6}$$

As the reliability measure we use the *coherence* of the structure tensor [3],

$$r_{y^*,t^*} := \frac{(J_{yy} - J_{xx})^2 + 4J_{xy}^2}{(J_{xx} + J_{yy})^2}. \tag{7}$$

Using the local depth estimates d_{y^*,t^*}, d_{x^*,s^*} and reliability estimates r_{y^*,t^*}, r_{x^*,s^*} for all the EPIs in horizontal and vertical direction, respectively, one can now proceed to directly compute depth maps in a global optimization

[1] Note than to obtain this formula from figure 3(a), Δx has to be corrected by the translation Δs to account for the different local coordinate systems of the views.



(a) Allowed (b) Forbidden

Figure 4. *Global labeling constraints on an EPI:* if depth λ_i is less than λ_j and corresponds to direction $\boldsymbol{n}_i$, then the transition from λ_i to λ_j is only allowed in a direction orthogonal to $\boldsymbol{n}_i$ to not violate occlusing order.

framework, which is explained in section 6. However, it is possible to first enforce *global visibility constraints* separately on each of the EPIs. This step is computationally expensive, but leads to the best results. We explain it in the next section.

5. Consistent EPI depth labeling

The computation of the local depth estimates using the structure tensor only takes into account the immediate local structure of the light field. In truth, the depth values within a slice need to satisfy global visibility constraints across all cameras for the labeling to be consistent. In particular, a line with is labeled with a certain depth cannot be interrupted by a transition to a label corresponding to a greater depth, since this would violate occlusion ordering, figure 4.

In this section, we show how we can obtain globally consistent estimates for each slice, which take into account *all views simultaneously*. While this is a computationally very expensive procedure, it yields the optimal results. However, for less complex light fields or if computation time is important, it can be omitted and one can proceed with the local estimates to section 6.

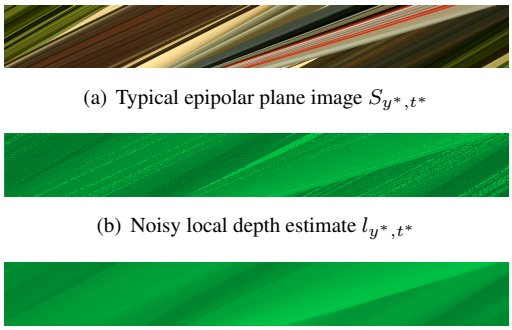
To satisfy the global visibility constraints on the slice level, we compute for each slice a new labeling d_{y^*,t^*} which is close to the local estimate l_{y^*,t^*}, but satisfies the constraints. The desired labeling is a map

$$d_{y^*,t^*} : \Sigma_{y^*,t^*} \to \{\lambda_1, \dots, \lambda_N\} \tag{8}$$

into a range of N discrete depth labels. We encode the minimization problem by using binary indicator functions $u_i : \Sigma_{y^*,t^*} \to \{0, 1\}$, one for each of the N labels, with the label uniqueness constraint that $\sum_{i=1}^{N} u_i = 1$. A region where $u_i = 1$ thus indicates a region of depth λ_i.

We define *local cost functions* $c_i(x, s)$, which denote how expensive it is to assign the depth label λ_i at a point (x, s), as

$$c_i(x, s) := r_{y^*,t^*}(x, s) \left| \lambda_i - l_{y^*,t^*}(x, s) \right|. \tag{9}$$



(a) Typical epipolar plane image S_{y^*,t^*}

(b) Noisy local depth estimate l_{y^*,t^*}

(c) Consistent depth estimate d_{y^*,t^*} after optimization

Figure 5. *With the consistent labeling scheme described in section 5, one can enforce global visibility constraints in order to improve the depth estimates for each epipolar plane image. Brighter shades of green denote larger disparities and thus lines corresponding to closer points.*

Note that the penalty is weighted with the local reliability estimate, so that points at which the depth estimate is likely inaccurate have less influence on the result.

The new slice labeling is then recovered by minimizing the global energy functional

$$E(\boldsymbol{u}) = R(\boldsymbol{u}) + \sum_{i=1}^{N} \int_{\Sigma_{y^*,t^*}} c_i u_i \, \mathrm{d}(x, s), \tag{10}$$

where $\boldsymbol{u} = (u_1, \dots, u_N)$ is the vector of indicator functions. The difficult part is now to define the regularizer $R(\boldsymbol{u})$ in a way that the labeling is globally consistent with occlusion ordering.

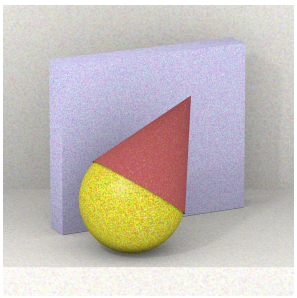
We observe that the visibility constraints restrict the allowed relative positions of labeled regions.

If $\boldsymbol{n}_i$ is the direction of the line corresponding to the depth λ_i, then the transition from λ_i to a larger depth label $\lambda_j, j > i$ is only allowed to happen in a direction orthogonal to $\boldsymbol{n}_i$, figure 4. Otherwise, a closer point would be occluded by a point which is further away, which is impossible.

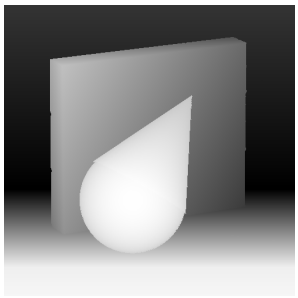
We enforce this occlusion ordering constraint by penalizing a transition from label λ_i to λ_j in direction $\boldsymbol{\nu}$ with

$$d(\lambda_i, \lambda_j, \boldsymbol{\nu}) := \begin{cases} 0 & \text{if } i = j, \\ \infty & \text{if } i < j \text{ and } \boldsymbol{\nu} \neq \pm \boldsymbol{n}_i^{\perp}, \\ 1 & \text{if } i < j \text{ and } \boldsymbol{\nu} = \pm \boldsymbol{n}_i^{\perp}, \\ d(\lambda_j, \lambda_i, \boldsymbol{\nu}) & \text{otherwise.} \end{cases} \tag{11}$$

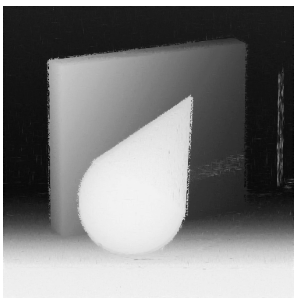
This problem fits into the minimization framework described in [21], where the authors describe the construction of a regularizer R to enforce the desired ordering constraints. We use an implementation of their method to minimize (10), and refer to their paper for more details. Note that the optimization scheme in [21] only imposes soft constraints, which means that there can still be some constraint



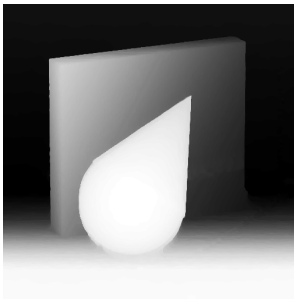
(a) Central view



(b) Ground truth



(c) Data term optimum



(d) After global optimization

Figure 6. *For a Lambertian light field with relatively simple geometry, our global approach already yields an almost perfect depth map. The expensive consistent estimate does not substantially improve the result anymore.*

violations left if the data term has a highly dominant preference for an inaccurate labeling.

Figure 5 demonstrates the result of enforcing global consistency for a single EPI of a light field. While the local depth estimate is noisy and of course does not satisfy any global constraints, the optimization yields a piecewise smooth estimate with sharp occlusion boundaries, which are aligned in the proper direction corresponding to the closer depth label of the transition. In particular, consistent labeling greatly improves robustness to non-Lambertian surfaces, since they typically lead only to a small subset of outliers along an EPI-line.

6. Global Integration

After obtaining EPI depth estimates d_{y^*,t^*} and d_{x^*,s^*} from the horizontal and vertical slices, respectively, we need to integrate those estimates into a consistent single depth map $u : \Omega \to \mathbb{R}$ for each view (s^*, t^*). This is the objective of the following section. We achieve our goal with a globally optimal labeling scheme in the domain Ω, where we minimize a functional of the form

$$E(u) = \int_\Omega g\,|Du| + \rho(u, x, y)\,\mathrm{d}(x, y). \qquad (12)$$

In figure 6, we can see that the per-slice estimates are still noisy. Furthermore, edges are not yet localized very

Source	*Data set*	*Views*	*Resolution*
Blender [4]	Conehead	21×21	500×500
	Buddha	21×21	768×768
	Mona	41×41	512×512
Gantry [20]	Truck	17×17	1280×960
	Crystal	17×17	960×1280
Plenoptic camera [16]	Elephant	9×9	980×628
	Watch	9×9	980×628

Figure 7. *4D Light field data sets used in our evaluation.*

Method	Δ	Dataset *Conehead*	*Mona*	*Buddha*
Stereo [7]	1	17.6	15.3	10.8
	5	64.3	43.3	45.6
Stereo [17]	1	34.5	9.5	19.1
	5	94.6	56.8	91.7
Local	1	21.5	8.1	26.4
(section 4)	5	77.1	43.0	79.6
Global	1	49.0	12.3	38.3
(section 4 + 6)	5	98.7	74.3	95.9
Consistent	1	51.1	15.5	39.6
(section 5 + 6)	5	98.9	80.1	97.1

Figure 8. *Comparison of estimated depth with ground truth for various approaches.* The table shows the percentage of pixels for which the error is less than Δ labels. Note that an error of one label corresponds to a relative depth error of only 0.2%, and the count is with respect to *all* pixels in the image without regard to occluded regions.

well, since computing the structure tensor entails an initial smoothing of the input data. For this reason, we encourage discontinuities of u to lie on edges of the original input image by weighting the local smoothness with a measure of the edge strength. We use

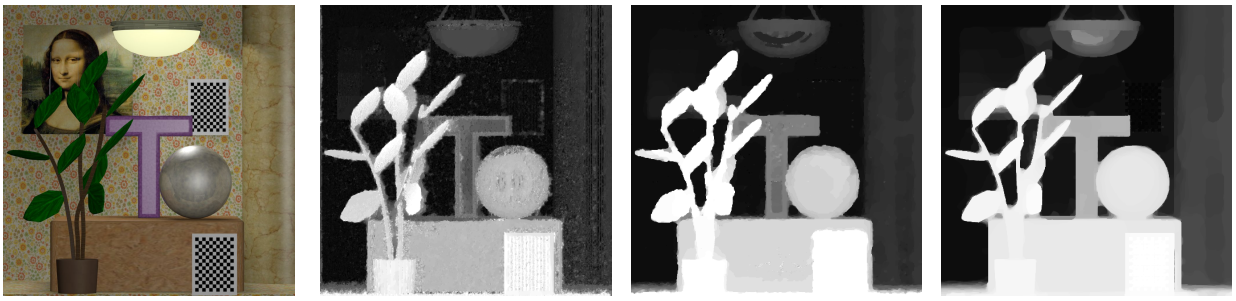
$$g(x, y) = 1 - r_{s^*,t^*}(x, y), \qquad (13)$$

where r_{s^*,t^*} is the coherence measure for the structure tensor of the view image, defined similarly as in (7).

In the data term, we want to encourage the solution to be close to either d_{x^*,s^*} or d_{y^*,t^*}, while suppressing impulse noise. Also, the two estimates d_{x^*,s^*} and d_{y^*,t^*} shall be weighted according to their reliability r_{x^*,s^*} and r_{y^*,t^*}. We achieve this by setting

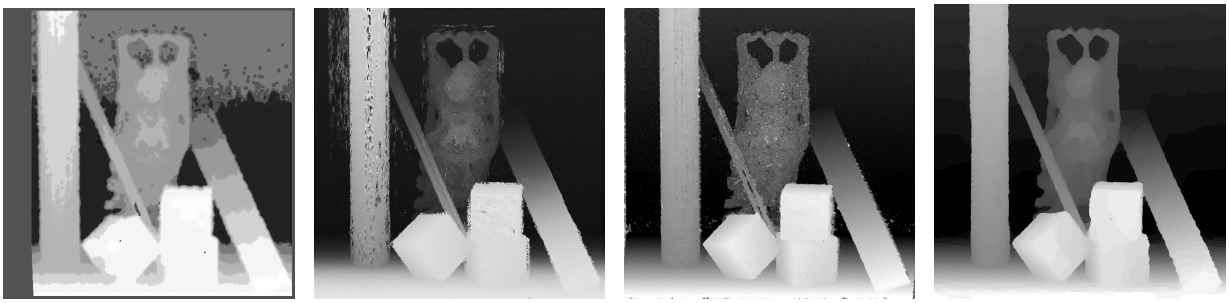
$$\begin{aligned}\rho(u, x, y) := \min(&r_{y^*,t^*}(x, s^*)\,|u - d_{y^*,t^*}(x, s^*)|\,,\\ &r_{x^*,s^*}(y, t^*)\,|u - d_{x^*,s^*}(y, t^*)|).\end{aligned} \qquad (14)$$

We compute globally optimal solutions to the functional (12) using the technique of functional lifting described in [17].



(a) Center view (b) Stereo [17] (c) Ours (global) (d) Ours (consistent)

Figure 9. *The synthetic light field "Mona" with some specularities and complex geometry is surprisingly difficult for depth reconstruction. Here, the consistent estimate yields a substantial improvement over global integration only, which already outperforms other methods.*



(a) Stereo [7], intensities rescaled (b) Ours (local) (c) Stereo [17] (d) Ours (global)

Figure 10. *With our methods, we can perform stable depth reconstruction in the synthetic light field "Buddha". The global stereo algorithm also achieves a satisfactory result, but has obvious difficulties on the specular statue and metal beam. See figure 1 for center view and result from our consistent estimation, which yields small, but visible and measurable improvements.*

7. Experiments

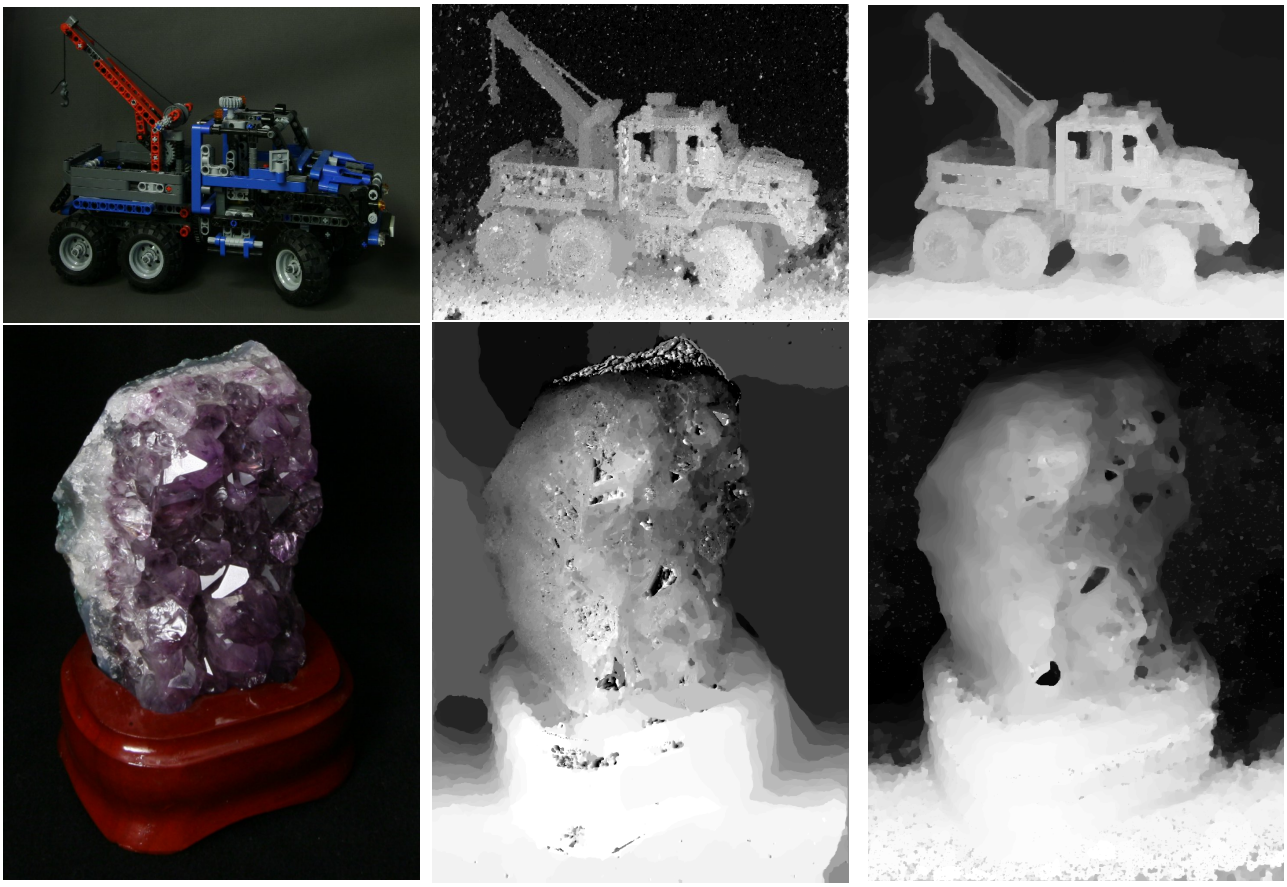
In this section, we verify the robustness and quality of depth estimates based on orientations in the epipolar plane image, and compare our results to stereo matching approaches. For this, we perform depth reconstruction on a number of data sets from various sources, see figure 7. The synthetic data sets and ground truth created with Blender will be provided on our webpage for future evaluations.

With our experiments we show that the light field paradigm provides better results if compared to stereo methods with comparable computational effort. In the first step, we compare our data term or local estimation using our structure tensor segmentation (section 4) to a local stereo approach [7]. In the second step, we compare the results of our global optimization (section 4 and 6) to a stereo matching approach within the same global optimization framework [17]. Finally, we demonstrate the potential of the very accurate but also time-consuming consistent approach (section 5 and 6).

Table 8 shows quantitative results of all the experiments with available ground truth data. For all methods, parameters were tuned to achieve optimum results. In our case, the dominant parameters are the scale parameters σ, τ for the structure tensor and a smoothing factor for global integration. We found that $\sigma = \tau = 0.8$ works well in most examples, and results are very robust to parameter variations.

Comparison of local estimates. Local estimation of depth for a single view performs at interactive frame rates. In table 8, one can see that our local data term on its own is often already more precise than a stereo approach [7]. This shows that the Lumigraph is a highly efficient data structure for depth estimation tasks, since with low amount of effort we obtain dense depth maps and for the most part bypass the occlusion problem by accumulating information from many surrounding views. Note that in this step, the local estimate performs depth estimation at floating point precision, which is the main cause of the high accuracy compared to the stereo matching method which is much more quantized, see figure 10.

Comparison of global estimates. Integrating the local depth estimates into a globally optimized depth map takes about 2-10 minutes per view depending on resolution and number of depth labels. Total computation times are comparable for both stereo matching as well as our novel technique based on structure tensor computation. Figures 9 to11 show visual comparisons of our global optimization to an approach which uses stereo matching [17], but which optimizes the same global functional (12) with a different data



(a) Center view (b) Stereo reconstruction [17] (c) Our method (global)

Figure 11. *Exploiting the structure of 4D light fields allows stable depth reconstruction even for very difficult scenes with reflective and specular surfaces, where methods based on stereo matching tend to fail. Note that the stereo result is already overly smoothed in some regions, while there is still a lot of noise left in the specular areas. Data sets "Truck" and "Crystal" from Stanford light field archive,* 17×17 *images at resolution* 1280×960*, runtime* 15 *minutes for both methods with* 128 *depth labels on an nVidia Fermi GPU.*

term. The numerical results in table 8 once again demonstrate the superiority of the light field structure for depth estimation.

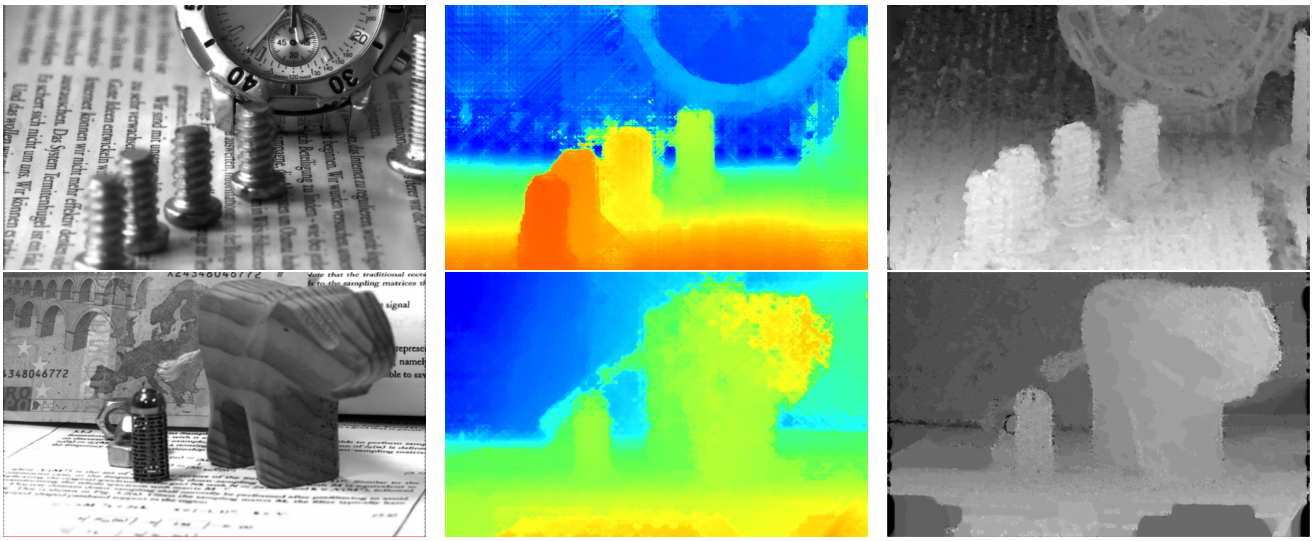
Consistent labeling. The full accuracy of the depth evaluation based on light fields is only achievable if we make use of the entire information provided in the EPI representation. For this, we need to impose the global occlusion ordering constraints described in section 5, which takes several hours to compute for all EPIs, but computes data for all views in the light field simultaneously. The visual and numerical improvements are shown in figures 9 and 10 as well as table 8.

Plenoptic camera images. We also experimentally demonstrate that our depth reconstruction technique works with images from a Raytrix plenoptic camera [16] and yields results which are superior to the reference algorithm provided by the manufacturer, figure 12. Note that in order to run our algorithm on this type of images, we first need to perform a conversion to a Lumigraph [23].

8. Conclusions

We have introduced a novel idea for robust depth estimation from 4D light field data in a variational framework. We locally estimate depth using dominant directions on epipolar plane images, which are obtained using the structure tensor. The local estimates are integrated into global depth maps with a globally optimal state-of-the-art convex optimization method. At greater computational cost, it is also possible to recover depth estimates which satisfy global visibility constraints. This can be done by labeling each epipolar plane image separately and imposing spatial layout constraints using a recent continuous optimization method based on convex relaxation.

We achieve state-of-the-art results, whose accuracy significantly surpasses that of traditional stereo-based methods. Furthermore, on data sets from various sources including plenoptic cameras, we demonstrated that by taking into account the special structure of 4D light fields one can robustly estimate depth for non-Lambertian surfaces.



(a) Center view (b) Reference algorithm by manufacturer (c) Our result (global)

Figure 12. *Depth estimates from lightfields acquired with the Raytrix plenoptic camera, which compare our method to a reference algorithm provided by the camera manufacturer on difficult real-world scenes with reflective metal objects. Our result clearly provides more details in the reconstruction.*

References

[1] E. Adelson and J. Bergen. The plenoptic function and the elements of early vision. *Computational models of visual processing*, 1, 1991. 2

[2] J. Berent and P. Dragotti. Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition. In *IEEE 8th Workshop on Multimedia Signal Processing*, pages 182–185, 2006. 2

[3] J. Bigün and G. Granlund. Optimal orientation detection of linear symmetry. In *Proc. ICCV*, pages 433–438, 1987. 3

[4] Blender Foundation. `www.blender.org`. 5

[5] R. Bolles, H. Baker, and D. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1(1):7–55, 1987. 2, 3

[6] A. Criminisi, S. Kang, R. Swaminathan, R. Szeliski, and P. Anandan. Extracting layers and analyzing their specular properties using epipolar-plane-image analysis. *Computer vision and image understanding*, 97(1):51–85, 2005. 1, 2

[7] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *Proc. ACCV*, 2010. 5, 6

[8] T. Georgiev and A. Lumsdaine. Focused plenoptic camera and rendering. *Journal of Electronic Imaging*, 19:021106, 2010. 1

[9] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The Lumigraph. In *Proc. ACM SIGGRAPH*, pages 43–54, 1996. 2

[10] A. Katayama, K. Tanaka, T. Oshino, and H. Tamura. Viewpoint-dependent stereoscopic display using interpolation of multiviewpoint images. In *Proceedings of SPIE*, volume 2409, page 11, 1995. 1

[11] M. Levoy. Light fields and computational imaging. *Computer*, 39(8):46–55, 2006. 1

[12] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. ACM SIGGRAPH*, pages 31–42, 1996. 2

[13] M. Matoušek, T. Werner, and V. Hlavác. Accurate correspondences from epipolar plane images. In *Proc. Computer Vision Winter Workshop*, pages 181–189, 2001. 2

[14] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. In *Proc. ACM SIGGRAPH*, pages 39–46, 1995. 1, 2

[15] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. Technical Report CSTR 2005-02, Stanford University, 2005. 1

[16] C. Perwass and L. Wietzke. The next generation of photography, 2010. `www.raytrix.de`. 1, 2, 5, 7

[17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*, 2010. 2, 5, 6, 7

[18] S. Seitz and C. Dyer. Physically-valid view synthesis by image interpolation. In *Proc. IEEE Workshop on Representation of visual scenes*, pages 18–25, 1995. 1

[19] H. Shum, S. Chan, and S. Kang. *Image-based rendering*. Springer-Verlag New York Inc, 2007. 2

[20] The (New) Stanford Light Field Archive. `http://lightfield.stanford.edu`. 1, 2, 5

[21] E. Strekalovskiy and D. Cremers. Generalized ordering constraints for multilabel optimization. In *Proc. International Conference on Computer Vision*, 2011. 2, 4

[22] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy. Using plane + parallax for calibrating dense camera arrays. In *Proc. CVPR*, 2004. 1

[23] S. Wanner, J. Fehr, and B. Jähne. Generating EPI representations of 4D light fields with a single lens focused plenoptic camera. *Advances in Visual Computing*, pages 90–101, 2011. 7