Light Field Reconstruction using a Generic Imaging Model

David Uhlig and Michael Heizmann

Karlsruhe Institute of Technology,
Institute of Industrial Information Technology,
Hertzstraße 16, 76187 Karlsruhe, Germany

Abstract Light field cameras play an increasingly important role in computer vision and optical metrology. However, due to their complex design, their calibration is very difficult and usually precisely adapted to the respective light field camera type. We present a method that extracts a light field from an arbitrary light field imaging system without knowing and without modelling the internal optical elements. We calibrate the camera using a generic calibration procedure, transform the obtained set of rays into an equivalent light field representation and finally, reconstruct a rectified light field from the irregularly sampled data. Experimental results validate the method and demonstrate that the geometrical structure of the light field is preserved by an adequate rectification.

Keywords Light field, decoding, rectification, generic camera

1 Introduction

The light propagating in space contains a variety of different information. However, when an image is taken with a classic camera, a large proportion of the information contained in the light is lost due to the projection. Computational cameras can encode information that is not available using conventional cameras. The additional modification of the camera can be used to extract useful information from the raw data apart from only the intensity-based colored image of the scene. In recent years, research on light field cameras (plenoptical cameras) has become more and more important. In contrast to

DOI: 10.58895/ksp/1000124383-8 erschienen in:

Forum Bildverarbeitung 2020

DOI: 10.5445/KSP/1000124383 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000124383/

traditional cameras, light field cameras are able to capture both the angular and spatial information of the light rays that are propagated through space. They are thus able to obtain multiple views of the same scene in a single photographic image exposure, to estimate the depth of the scene or to shift the focus of the image after capturing the image [1]. These advantages have led to light field cameras becoming an important tool in image processing and optical metrology. As a result, a precise calibration of these cameras becomes increasingly important.

The first commercially available light field camera was presented by Ng [1]. He proposed a hand-held camera that consisted of an additional micro-lens array in front of the sensor. This array additionally allows to detect the directional dependencies of the rays, and thus a light field can be extracted. Since the design of microlens based cameras is not trivial, the light field has to be decoded from the raw sensor image using sophisticated algorithms. Furthermore, each lens (micro and main lens) is affected by the usual lens aberrations, *i. e.* a subsequent rectification of the light field is necessary to obtain correct geometric information relevant for image processing and metrology applications. Dansereau *et al.* [2] presented a method that first extracts a light field from the raw sensor data and then rectifies it by estimating the values of a 12-parameter camera model. Bok *et al.* [3], in contrast, presented a method that could extract the rectified light field directly from the raw sensor data by also using a low-dimensional camera model. In order to be able to extract any information about the light field, both methods must initially detect the microlenses very precisely. But, since the camera rays at the boundary of the microlenses are very difficult to model in both methods, these pixels are mostly discarded.

Another disadvantage of these methods is the model based calibration in general. It can't describe highly local errors such as the strong distortions at the boundaries of the microlenses using a low-dimensional model. As a consequence, in the recent years, new camera models were proposed that describe the camera as a generic imaging system. They are able to model the ray of each pixel individually and thus allow high-precision calibration [4, 5]. However, the biggest disadvantage of the common light field reconstruction methods is that they are only applicable for a single type of camera, *e. g.* microlens based light field cameras whose microlenses are exactly focused on the sensor. To our knowledge, there is no single method yet that can reconstruct a light field from any type of light field camera.

In this work we present a method to reconstruct a light field, that was captured by an arbitrary light field imaging system, without knowing the actually used configuration of optical elements inside the camera. We propose to use a generic camera calibration procedure to optimally calibrate each individual pixel of the camera, where all distortions of the optical elements are contained in the unconstrained bundle of sight rays, and thus are modeled very accurately. Further, we propose to use this bundle of rays to obtain an irregularly sampled presentation of the light field, and finally, we present a simple reconstruction method to interpolate a rectified light field from the irregularly spaced camera rays. We use the presented method to calibrate and reconstruct light fields from a commercially available Lytro Illum light field camera.

The paper is organized as follows: Section 2 provides the background about light fields and light field cameras as well as an introduction to the concept of generic camera calibration. Section 3.1 and 3.2 derive the 4D light field parameters from the unconstrained ray bundle obtained in the generic calibration. Section 3.3 describes the algorithm for the reconstruction of the light field from the rays' intensity values and finally, section 4 experimentally validates the proposed method by analyzing real light field images. At last, section 5 draws conclusions and presents directions for future work.

2 Background

2.1 Light Fields and Light Field Cameras

In the field of geometrical optics the light of a scene can be described by the plenoptical function with six variables: three spatial coordinates, two angular coordinates, one spectral value. In a conventional camera usually only a subspace of this function can be captured: two spatial coordinates with a color/intensity value. A light field camera allows to capture two additional angular dimensions. For this, the most common type are microlens based light field cameras. The design of these is similar to that of conventional cameras, with the difference that an array of microlenses is positioned in front of the sensor [1], see fig. 2.1. By adding the microlens array it is possible

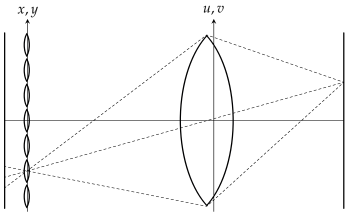

Figure 2.1: Schematic structure of the light field camera.

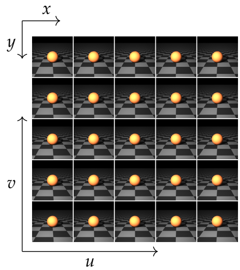

Figure 2.2: Interpretation of the light field as a camera array.

to capture a section of the light field $L(u, v, x, y)$ of a scene. Here x, y describe the coordinates of the microlenses in front of the sensor and thus, the spatial dimension of the light field. u, v describe the coordinates within the microlens relative to its center and implicitly provide information on where a light ray has passed through the main lens. They represent the angular information of the light field. Each u, v coordinate therefore represents a virtual subcamera, which observes only a part of the main lens, meaning that a light field camera can also be interpreted as a multi-camera array, whereby each subcamera has a slightly different view onto the scene, see fig. 2.2. The additional information compared to the standard camera allows to change the perspective on the scene after the exposure, which allows to extract depth information, or to shift the focus after the image capture.

In particular, there are different configurations, *e. g.*, the distance of the array to the sensor can be varied or microlenses with multiple focus lengths can be used [6]. Furthermore, there are coded aperture based light field cameras, kaleidoscope-like configurations and of course camera arrays [7, 8]. All have in common that decoding the light field from the sensor data and calibrating the camera is generally difficult. For example, to reconstruct the light field of microlens-based cameras, the centers of the microlenses, which are often arranged in a hexagonal grid, must be detected very accurately [9]. The 4D light field can then be extracted by shifting the pixels onto a rectangular grid and reshaping the 2D-microlens-images into a 4D array. This light field, however, generally still contains all the distortions of the main lens and the microlenses, which is why an additional rectification is necessary [2,3].

2.2 Calibration

The basis of the calibration is a precise modeling of the camera, which is of course strongly influenced by the camera type. Conventionally, low-dimensional models are used to model the entire camera. However, their disadvantage is that they have insufficient descriptive power. Consequently, with modern cameras or optical systems not all pixels can be described perfectly by these few model parameters. The more complex an optical system becomes, the more difficult it is to model it using a low-dimensional representation. Hence, the lack of flexibility and precision has led to the development of new camera models. Cameras are described as generic imaging systems, which are independent of the specific camera type and allow high-precision calibration [4,5]. An imaging system is modeled as a set of photosensitive pixels, where all other optical elements are represented by a black box. Each pixel collects light from a bundle of rays entering the imaging system, which is called *raxel*. The set of all *raxels* with the associated geometric parameters forms the complete generic imaging model.

The geometric parameters can be described for each pixel i by a single camera ray running through the center of the *raxel* along the direction of light propagation, $\vec{r}_i = (\vec{d}_i^{\mathrm{T}}, \vec{m}_i^{\mathrm{T}})^{\mathrm{T}}$, with a direction vector $\vec{d}_i$ and a start vector $\vec{m}_i$. Its calibration is usually performed by minimizing the Euclidean distance of the rays $\vec{r}_i$ to known reference points $\vec{p}_{ik}$ in space, also called ray re-projection error: $\epsilon_i = \sum_k d_{\mathrm{euclid}}(\vec{r}_i, \vec{p}_{ik})$. A minimization of the commonly used ray projection error is often not possible, because most generic models do not support a direct projection onto the pixel plane. See [5] for more details.

The advantage of this type of modeling is that there is no longer one global model that has to describe the camera over the entire pixel plane. Instead, with the generic model even high-frequency distortions in the optical imaging system can be modeled equally accurate both locally and globally, resulting in a highly accurately calibrated camera. This is specifically important for light field cameras, where it becomes very difficult to model distortions of the microlenses with a global model. In the end, however, one does not obtain an "image", but rather a set of rays with corresponding intensities. This does not interfere with many applications in optical metrology, *e. g.*, profilometry or deflectometry, where only the geometric ray properties are relevant [10]. But it can make other tasks more difficult, due to the loss of spatial correlations between pixels and their corresponding rays. The classic image processing algorithms cannot be applied without further effort. In the special case of the light field camera, algorithms such as the subsequent re-focusing of the image or a simple depth estimation can no longer be carried out using standard methods. Therefore, we propose to use the generic camera model to reconstruct the light field from the set of rays. And thus, we obtain a generic algorithm to extract the light field from an arbitrary optical imaging system, neglecting the actual design of the used light field camera.

3 Light Field Reconstruction

3.1 From Generic Camera Rays to Light Field Coordinates

In order to reconstruct the light field from the camera raw data, the camera must first be calibrated using a generic calibration method [5]. Since the camera is considered a black box, it is generally not possible to define a consistent camera coordinate system for every camera. Hence, the result of the generic calibration is not unique, *i. e.* the calibrated rays are represented in an arbitrary coordinate system, which depends on the starting configuration of the generic calibration procedure. To transform this arbitrary coordinate system into one that is fixed to the individual camera, a few steps are necessary. First, we need to define the origin as the point which has the smallest distance to all rays, *i. e.* it minimizes the mean Euclidean distance to all rays. For a light field camera this corresponds approximately to the center of the exit pupil. Further, we define the z-axis of the camera coordinate system as the average ray direction. The last remaining degree of freedom is the rotation around this z-axis. To determine it, we calculate the intersections of all rays with a distant plane orthogonal to the z-axis. Since light field cameras project the light perspectively onto a rectangular sensor, the pattern of the intersections will be its projection into space. Applying a principal component analysis (PCA) to this 2D point cloud results in a rotation which aligns the rectangle with the x- and y-axes. As final step, we transform the rays into light field coordinates. For this, we calculate the intersections of the rays with the 2-plane-representation of the light field. The u, v-plane is placed orthogonal to the z-axis into the origin of the coordinate system. The x, y-axis is placed parallel to this at an arbitrary distance f. Thus, each ray $\vec{r}_i$ can be described by four light field coordinates (u_i, v_i, x_i, y_i) with color value $L(u_i, v_i, x_i, y_i)$, see fig. 3.1.

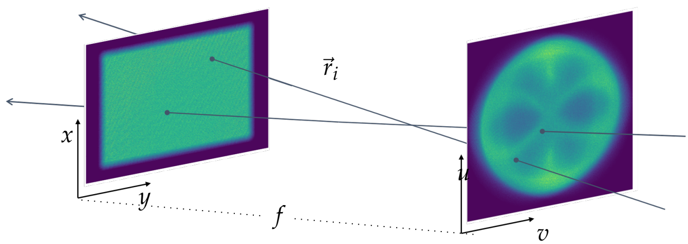

Figure 3.1: 2-plane-parameterization of the light field. The ray $\vec{r}_i$ intersects the u, v- and the x, y-plane in (u_i, v_i, x_i, y_i). The intensities in the planes visualize the spatial distribution of the intersection points as a 2D histogram.

3.2 Discrete Light Field

In order to reconstruct a light field from the bundle of rays belonging to the camera, the observed ray intensities must be interpolated to a discretized light field. We parameterize it to be interpolated into the same 2-plane-representation as before. The complete set of real camera rays described as a set of 4D-points is arranged in an irregular 4D-grid. Still, the classical light field algorithms require a regular grid with uniform spacing. Therefore, this irregular grid of continuous rays has to be interpolated to a discrete light field described by a regular grid. The number of 4D cubes in each direction and the length of their edges could in principle be defined arbitrarily, but it is advisable to incorporate knowledge about the physical camera. For example, our microlens-based light field camera (Lytro Illum) has about 14×14 pixels under each microlens. Thus, this sampling can be used directly as a basis for the discretization of the u, v-plane. The sampling of the x, y-plane can be determined in the same way by, *e. g.*, the number of microlenses in front of the sensor. This procedure leads to a regular grid with grid points $(u, v, x, y) \in U \times V \times X \times Y$ with the resolutions of the respective dimensions $U = V = [0, \ldots, 14]$, $X = [0, \ldots, 551]$, $Y = [0, \ldots, 383]$. After the discrete target light field has been defined, we need to transform the set of real camera rays. First, by means of a histogram analysis of the spatial density of the ray-plane intersection points, the domains of the real light field dimensions are determined, see fig. 3.1. In order to place the regular grid structure into the irregular data, we define the grid extension by using a threshold value on the histogram data. A threshold of, *e. g.*, 10% ensures that most of the camera rays are within the range defined by the regular grid. Since the real light field parameters are specified in physical units, *e. g. mm*, they have to be transformed to the previously defined discrete 4D-pixel grid by a shifting and scaling operation, *e. g.* $u_i \leftarrow \frac{u_i - \min(u_i)}{\max(u_i) - \min(u_i)} U_{\max}$. This still results in irregular spaced data, which however can now be interpolated more easily to the desired regularly sampled light field.

3.3 Reconstruction

After the parameters of the light field have been defined, each corresponding light field pixel can be determined for every ray, by finding the discrete grid point that is closest to the rays' light field representation. Since the rays and the grid are normalized to the same scale, these correspondences $\mathcal{N}_{u,v,x,y}$ can easily be found by a simple rounding operation to the closest integer $[\cdot]$. As a result, each light field pixel is only influenced by the rays that lie in the corresponding 4D-cube:

$$\mathcal{N}_{u,v,x,y} := \left\{ i \;\middle|\; 1 \geq \left\| (u, v, x, y)^{\mathrm{T}} - \left([u_i], [v_i], [x_i], [y_i] \right)^{\mathrm{T}} \right\|_0 \right\} . \quad (3.1)$$

The intensity of a discrete pixel can then be calculated from the intensity values of the corresponding rays as a weighted average:

$$L(u,v,x,y) = \frac{1}{\sum_{i \in \mathcal{N}_{u,v,x,y}} w_i} \sum_{i \in \mathcal{N}_{u,v,x,y}} w_i \, L(u_i, v_i, x_i, y_i) , \quad (3.2)$$

$$w_i = \frac{1}{\epsilon_i} \exp(- \left\| (u,v,x,y)^{\mathrm{T}} - (u_i, v_i, x_i, y_i)^{\mathrm{T}} \right\|_1^2) . \quad (3.3)$$

For the weighting factor we calculate the distance between the ray's light field parameters and its correspondence in the grid. In order to consider larger deviations less, the error is squared and exponentially weighted. An additional weighting of the different light field coordinates is not required, since these have already been brought to a unified basis by the normalization of section 3.2. To additionally benefit from the results of the generic calibration, the error ϵ_i of the calibration procedure is taken into account, *e. g.* the pixelwise ray-projection error [5]. This suppresses badly calibrated camera rays, which often do not have good optical properties, *e. g.* dead pixels or pixels at the edges of micro lenses, which can be strongly distorted.

4 Results

For the evaluation of the proposed method, the sight rays of a Lytro Illum light field camera were estimated using a generic camera calibration. Subsequently, these were used to reconstruct the light field of a scene, using the proposed method. The reconstruction of the central view of an example image is shown in fig. 4.1. Here, only rays from the center of the u, v-plane where used in the reconstruction. For a comparison to the state-of-the-art, the methods of Dansereau *et al.* [2] and Bok *et al.* [3] were evaluated too. It can be seen that the proposed method can reconstruct the scene correctly, although there were absolutely no presumptions about the internal optical structure

Figure 4.1: Center views of the light field. Dansereau *et al.* (top left), Bok *et al.* (top right), proposed method (bottom left). Detailed views: Dansereau *et al.* (top), Bok *et al.* (middle), proposed method (bottom).

of the camera and no information of the spatially correlated pixels was used. The reconstruction results of Dansereau *et al.* and Bok *et al.* are relatively similar, but show a sharper result compared to the proposed method. In detail it can be seen that the proposed method can reconstruct the light field even near object edges very well. The visibly larger blur is due to the relatively freely chosen sampling of the light field. A better optimized choice of the light field dimensions should result in less rays being summed up, thus reducing the blur. In addition, the arbitrary offset of the reconstruction grid produces interpolation-related blur. This should also be reduced by a further optimization of this offset.

Nevertheless, the advantage of the proposed method can be found in another area. Apart from the central view, the light field contains much more information. If one fixes an angular and a spatial coordinate in the 4D light field pointing in the same direction, *e. g.* u and x, one gets a 2D-slice of the light field, a so-called epipolar plane image (EPI) [1]. Lines of different slopes can be seen, whose orientation represents the depth of the observed object point [7]. The depth estimation is thus reduced to a simple local orientation estimation in the EPIs, whereby the quality of the estimation is significantly influenced by the calibration. The better the quality of the lines, the better the result of the depth estimation. Fig. 4.2 shows examples of a horizontal and a vertical EPI generated by fixing u and v to its center coordinates and by selecting pixel lines for the x and y coordinate, respectively. The EPI of Dansereau *et al.* shows strong deviations from the epipolar geomery, visible through the curvy epipolar lines. This is caused by the poor generalizability of the method, which was developed for the old Lytro camera and works only moderately well for the newer Lytro Illum. The EPI of Bok *et al.* on the other hand is much straighter. However, there are errors at the top and the bottom. These areas correspond to pixels which are located at the boundary of the microlenses, where the imaging is more strongly distorted. For the proposed method, it can be seen that the epipolar geometry is maintained much better, visualized by the straight lines in the EPIs. Also, the distortions of the lenses are compensated, resulting in a rectified light field. However, as before, due to generic nature of the method, the sampling is not yet ideal. This is visible by the overall lower resolution and the slightly more blurry appearance.

5 Conclusions

In this paper we presented a method that allows us to calibrate any light field camera (*e.g.* microlens-based, mirror-based, camera arrays) without having to model the exact optical properties. Using a generic calibration, we can precisely calibrate the individual camera rays. We normalized the result to transform it into an equivalent light field representation. Since classical algorithms require a regular sampling, we fit a regular 4D grid into the irregular camera rays. Summation of the rays' weighted intensity values finally resulted in the interpolation and reconstruction of the rectified light field. Experiments showed that the method can provide good reconstructions and that it returns rectified light fields. The epipolar geometry between the subcameras is preserved and shows even better results than the conventional methods. However, in detail it can be seen that the reconstructed light fields are more blurred in comparison

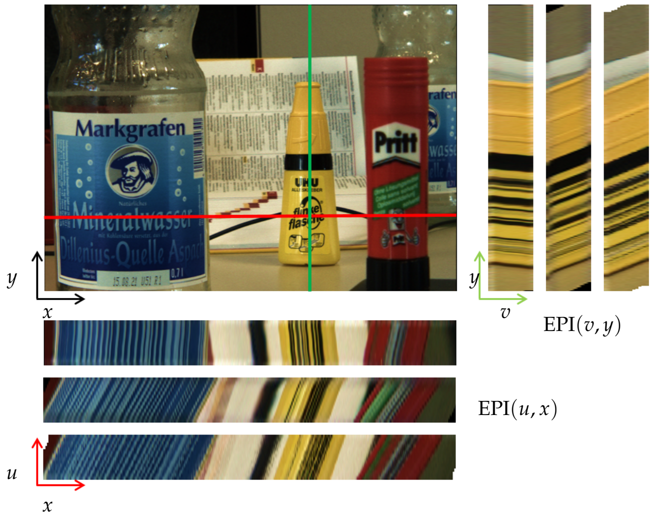

Figure 4.2: Horizontal (red) and vertical (green) EPIs in comparison: Dansereau *et al.* (top & left), Bok *et al.* (middle), proposed method (bottom & right).

to the standard methods. This can be explained by the sub optimal sampling of the light field coordinates. Therefore, further work is devoted to the improvement of the light field sampling, whereby both the desired resolution and the position of the grid points will be optimized and adapted to the used camera. Also, more experimental evaluation using different light field acquisition systems is in progress.

References

1. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.
2. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1027–1034.

3. Y. Bok, H.-G. Jeon, and I. S. Kweon, “Geometric calibration of micro-lens-based light field cameras using line features,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 287–300, 2017.
4. M. D. Grossberg and S. K. Nayar, “The raxel imaging model and ray-based calibration,” *International Journal of Computer Vision*, vol. 61, no. 2, pp. 119–137, 2005.
5. D. Uhlig and M. Heizmann, “A calibration method for the generalized imaging model with uncertain calibration target coordinates,” in *2020 IEEE Asian Converence on Coputer Vison*, H. Ishikawa, C.-L. Liu, T. Pajdla, and J. Shi, Eds. Springer International Publishing, 2020.
6. T. Georgiev and A. Lumsdaine, Eds., *The multifocus plenoptic camera*. International Society for Optics and Photonics, 2012.
7. S. Wanner and B. Goldluecke, “Variational light field analysis for disparity estimation and super-resolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014.
8. I. Ihrke, J. Restrepo, and L. Mignard-Debise, “Principles of light field imaging: Briefly revisiting 25 years of research,” *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 59–69, 2016.
9. M. Schambach and F. Puente León, “Microlens array grid estimation, light field decoding, and calibration,” *IEEE transactions on computational imaging*, vol. 6, p. 591–603, 2020.
10. S. Werling, M. Mai, M. Heizmann, and J. Beyerer, “Inspection of specular and partially specular surfaces,” *Metrology and Measurement Systems*, vol. 16, no. 3, pp. 415–431, 2009.