

Adopting an Unconstrained Ray Model in Light-field Cameras for 3D Shape Reconstruction

Filippo Bergamasco, Andrea Albarelli, Luca Cosmo, Andrea Torsello

Dipartimento di Scienze Ambientali, Informatica e Statistica
Università Ca' Foscari Venezia, Venice Italy

Emanuele Rodolà, Daniel Cremers

Department of Computer Science
Technische Universität München, Garching, Germany

Abstract

Given the raising interest in light-field technology and the increasing availability of professional devices, a feasible and accurate calibration method is paramount to unleash practical applications. In this paper we propose to embrace a fully non-parametric model for the imaging and we show that it can be properly calibrated with little effort using a dense active target. This process produces a dense set of independent rays that cannot be directly used to produce a conventional image. However, they are an ideal tool for 3D reconstruction tasks, since they are highly redundant, very accurate and they cover a wide range of different baselines. The feasibility and convenience of the process and the accuracy of the obtained calibration are comprehensively evaluated through several experiments.

1. Introduction

Due to their recent availability as off-the-shelf commercial devices, light-field cameras have attracted increasing attention from both scientific and industrial operators.

Traditional cameras are designed to capture the amount of light radiation directed toward an image plane. The captured rays can converge to a common projection point (as for the pinhole model), could go through a common axis (as for the models including radial distortion), or can follow any other distribution, even ditching any parametric model. However, regardless of the camera model, the mechanics remains basically projective, and the result of the imaging process is a 2D image. Light-field cameras pursue a different goal: to capture the full plenoptic function generated by each observed material point [1], which includes the intensity of the light radiating from each point along all the directions over the sphere. Of course, this goal is not practi-

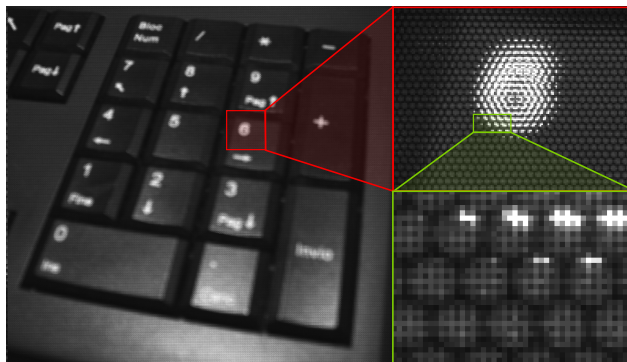


Figure 1. A real-world light-field camera is no more than a tightly packed array of very small (distorted) pinhole cameras sharing the same imaging sensor.

cally achievable by any physical sensor, due to the technical and theoretical problem involved. In practice, most, if not all, the light-field devices ever built are made up of an array (explicit or implicit) of traditional cameras, each one contributing to capture a portion of the plenoptic function. An example can be seen in Figure 1, where we show a detail of the composite image captured by a Lytro light-field camera [22]. The number, type and arrangement of such cameras, as well as their calibration, has been a very active topic in recent research. One of the main hurdles in plenoptic photography derives from the composite imaging formation process which limits the ability to exploit the well consolidated stack of calibration methods that are available for traditional cameras. While several efforts have been done to propose practical approaches, most of them still rely on the quasi-pinhole behavior of the single microlens involved in the capturing process. This results in several drawbacks, ranging from the difficulties in feature detection, due to the reduced size of each microlens, to the need to adopt a model with a relatively small number of parameters.

1.1. Related Work

Light-field cameras have been recently popularized thanks to the recent introduction of commercial devices such as Lytro and Raytrix models. The idea of capturing the plenoptic function, however, has been investigated by the scientific community for at least a couple of decades, starting with the seminal works by Adelson and Wang [1] and exploring several alternative designs, including the adoption of aligned pairs of attenuation masks through the whole light path [33] or just before the camera sensor [30], the introduction of hemispherical mirrors within the imaging path [25], a time multiplexed multiple exposure schema which allows to exploit the full sensor resolution [20] and the use of a simple but quite effective arrays of traditional cameras [32]. Such interest in light-field imaging devices is well justified by its numerous applications, ranging from depth map estimation [9, 13, 4] to super resolution [5, 31], refocusing after shooting [22] and creation of virtual points of view [19, 18].

Since most light-field devices are explained by a multipinhole model, most calibration methods proposed in recent literature exploit this fact and the related projective geometry properties. Such extension, however, is hardly straightforward. In fact, the narrow field of view and distribution of micro-lens make it difficult to get a full view of the calibration target and to obtain a resolution high enough for feature localization. Each method proposes a different solution to solve these limitations.

One of the most recent approaches, proposed by Bok et al. [6], where the difficulties in finding chessboard corners have been overcome by resorting to adopt the linear features resulting from the interface of white and black squares. Such features are indeed more abundant, since the probability of a linear object to be fully captured by a single micro-lens is much higher than the chance to properly observe a corner center and its surroundings. Differently, Cho et al. [10] introduce a method to calibrate micro-lens centers by searching for local maxima in the frequency domain over white images, that is images of uniform white targets. Further, they also suggest an interpolation method to reconstruct subaperture images, *i.e.*, images obtained by pixels capturing rays that converge to a single point, that is pinhole subimages of the captured light-field. Dansereau et al. [11] introduced a 15-parameters overall camera model (including a 5 parameter distortion), leading to a lower dimensionality with respect to modeling individual pinhole cameras. They also propose a rectification method that can be adopted for multiview stereo. Other methods include one proposed by Johannsen et al. [17], also using a 15-parameters model, calibrated using a small pattern of regular dots, and the work by Vaish et al. [29], which can be used to calibrate large arrays of discrete camera devices whose projection centers lie on the same plane.

It should be noted that virtually all the calibration methods to be found in the recent literature are designed to be used mainly to enhance the production of sub-aperture images. For this reason they are basically only concerned with the reconstruction of the imaging process. This paper takes a radically different path, focusing on the recovery of the geometry of generalized sensor rays in order to exploit them for accurate 3D surface triangulation.

1.2. Contribution

This paper makes two main contributions, that we feel to be important to the light-field community.

First, we analyze the use of a calibration method that escapes the need to adopt a parametric model by exploiting dense correspondences generated using phase coding technique [3]. While dense calibration has been already used in literature, this is the first time that it is attempted with light-field cameras and its correct behavior is not guaranteed. In fact both the initialization hurdles and the sparsity of the micro-lenses pixel could hinder the process, leading to unsatisfactory results. To this end, we perform an in-depth study of the different aspect of calibration under various conditions.

The second contribution is related to the recovery of 3D shapes. This is a common application of light-field cameras, especially using multiway-stereo algorithms. Unfortunately, most triangulation methods are based on epipolar geometry constraints, that can only be exploited if pinhole cameras are available. To this end, we propose a triangulation step suitable for any camera model and we use it to compare our method with the recent state-of-the-art.

2. Light-Field Calibration

All the calibration methods for a light-field camera must deal with a common hurdle: Each micro-lens, being pinhole, could be calibrated independently using standard target-based methods, ranging from the basic approach proposed by Tsai [28], to more advanced model that could account for the high distortion that micro-lenses usually exhibit [16, 12]. However, these approaches would incur in the disadvantage that the recovery of the target pose could be very ill-posed when performed on the basis of a single

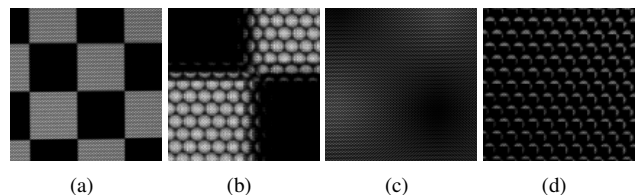


Figure 2. A chessboard pattern captured by a Lytro lightfield camera respectively at minimum and maximum zoom level.

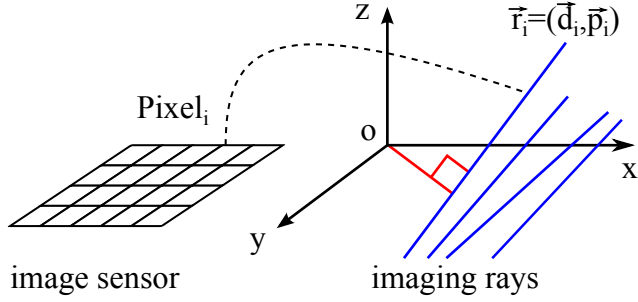


Figure 3. The non-parametric imaging model adopted in this paper (see text for details).

micro-lens image. In fact, each micro-lens only counts few pixels spanning a large view angle, resulting in poor angular resolution of each micro-lens in isolation. Furthermore, most approaches adopt image space techniques to localize target features, thus the limited span of each micro-lens image would severely reduce the number of recoverable correspondences and their accuracy. This problem can be better understood by looking at Figure 2, where the images of a chessboard acquired by a commercial Lytro lightfield camera are shown. In Subfigure 2(a) and 2(b) we show an overall frame and a detail of the image created on the CCD when the zoom of the main lens is set to the minimum value. The overall frame (about 3000 pixels wide) appears similar to what would be obtained using a standard projective camera. However, by looking at the detail of each microlens (about 10 pixels wide), it can be observed that most of them capture a fully black or fully white field and just a few see a chessboard corner. Under these conditions, classical calibration methods that need to relate target points to observed features are useless, hence the need for specialized technique. The behavior is even more extreme within Subfigure 2(d) and 2(c), where we display images of the same chessboard obtained with the maximum zoom level of the main lens. Here, the overall frame lost any connection to a projective transform and the images produced by the microlens are so wide that they extend beyond the span of a single check. The solution we are proposing is to refrain to use any global parametric model and to independently assess the characterization of every single imaging ray that insists on the camera sensor. To this end, we apply a dense target localization method which works in the time rather than the space domain, thus escaping the aforementioned hindrances. Such dense correspondences, in turn, enable the adoption of a parameter-free optimization for non-central cameras.

2.1. The Parameter-Free Camera Model

Following [3], we adopt a non-parametric camera model where each ray is modeled as an independent line within a common reference frame. Such reference frame is not di-

rectly related to the physical sensor. In fact, according to this model, image coordinates can be considered just labels for the imaging rays, which are not related to them by means of any analytic function (see Figure 3). More formally, index $i \in \{1..n\}$ ranges over all the n pixels of the camera sensor (in no particular order). The ray associated to pixel i can be written as $\vec{r}_i = (\vec{d}_i, \vec{p}_i)$, where $\vec{d}_i, \vec{p}_i \in \mathbb{R}$ represent direction and position of the ray respectively. These vectors satisfy $\|\vec{d}_i\| = 1$ (normalized direction) and $\vec{d}_i^T \vec{p}_i = 0$ (orthogonal position vector). Any point \vec{x} in the ray \vec{r}_i satisfies the parametric equation $\vec{x} = \vec{d}_i t + \vec{p}_i$ for some $t \in \mathbb{R}$. Lacking any explicit structural relation between rays, this model counts 4 degrees of freedom for each pixel, resulting in several millions unknowns.

A solution space this large needs an exceptional number of observations, and this can only be obtained using a dense coding strategy, which assigns to each image pixel (i.e. to each ray) a pair of coordinates on the calibration target. There are several ways to do this, in this paper we follow [3] adopting a flat monitor as the target [14, 24] and using a multi period phase shift coding [21] in order to obtain dense target coordinates on a Lytro camera sensor. The coding has been performed both horizontally and vertically. In Figure 4 an example of such dense coding can be seen. We used a scale of red and green values to show the recovered horizontal and vertical coordinates of the monitor overlaid to a white light imaging of the scene (resulting in a color blending). The code appears to be smooth, but of course this is an effect due to detail level of the figure, in practice the image presents the same repetition effects that can be seen in Figure 2 and that will be discussed in detail in the next section. The dense correspondences acquired over several poses of the target can be used to feed the iterative optimization method presented in [3] obtaining the characterization of each ray that has been correctly codified within a large enough number of different poses. Such method, however, has been designed to work on quasi-pinhole cameras and there is no guarantee that it works with a plenoptic camera. Neither it is obvious that the dense coding would work well with the considered imaging process, especially for the higher camera zoom levels shown in Subfigures 2(d) and 2(c).

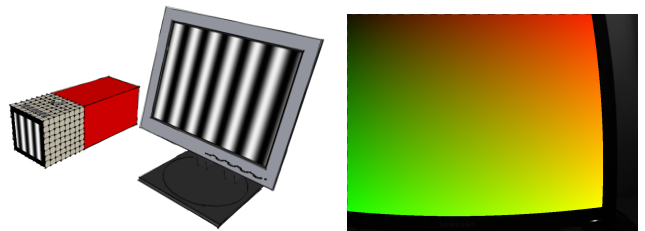


Figure 4. Dense calibration requires the user to acquire a sequence of horizontal and vertical patterns from a standard monitor.

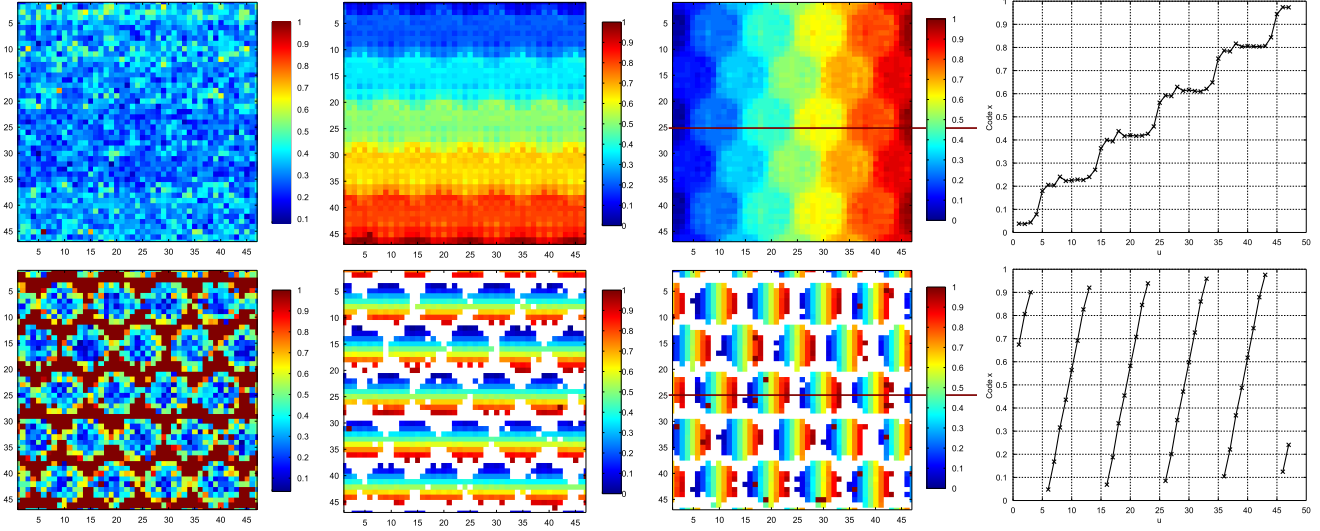


Figure 5. Behavior of the dense coding method applied to a Lytro lightfield camera at different zoom levels (see text for details).

In the next two sections we will study in detail the performance of the dense coding and of independent rays calibration when applied to a lightfield camera.

2.2. Dense Target Acquisition

All our experiments have been performed using a first generation Lytro plenoptic camera, equipped with a 3280x3280 pixels sensor. Throughout all the tests we used two zoom levels for the main lens, which we will refer to as *minzoom*, corresponding to the 1.0 level on the camera, and *maxzoom* corresponding to the 5.5 level (the reader can refer to Figure 2 to get an idea about what results can be obtained at such zoom levels).

Horizontal and vertical coding of the target has been performed using multi-period phase shift on a 21 inches monitor place at about 1 meter from the camera, with period lengths respectively 11, 13 and 17 monitor pixels and capturing 15 samples for each period (see [21] for details).

In Figure 5 we show the acquired codes within a small portion of the imaging sensor measuring about 50x50 pixels respectively for minzoom (upper row) and maxzoom (lower row) (note that the code has been normalized from 0 to 1 for better visualization). The first column shows the coding error according to [21], a lower value means a more accurate target position recovery, a value of 1 means that no recovery has been possible for that particular pixel. At minzoom level the coding error is low and isotropic, this is due to the low distortion of the microlens images and to the overall quasi-pinhole behavior of the camera. Note that, although it may be counter-intuitive, at minzoom even the usually disregarded space *between* microlens is correctly coded, thus recovering the intersection between the associated ray and the target.

This allows to effectively calibrate the full sensor, including inter-lens pixels. In the next section we will show that the resulting calibration exhibits a good accuracy. This is in strong contrast with standard parametric light-field camera calibration methods, which, even when capable of capturing features between microlens, would still be unable to describe the behavior of those rays with their models.

Conversely, at maxzoom the coding error is lower around the centers of the microlens and increases a bit when moving toward the edges. Note that outside the microlenses, the code is not recovered. This is not due to measuring errors, but rather to the low signals that reaches the sensor as a result of the strong vignetting (which can be observed also on Subimage 2(c)). In the two central columns of Figure 5 we show respectively the recovered horizontal and vertical codes. Both are dense and quite smooth, however the

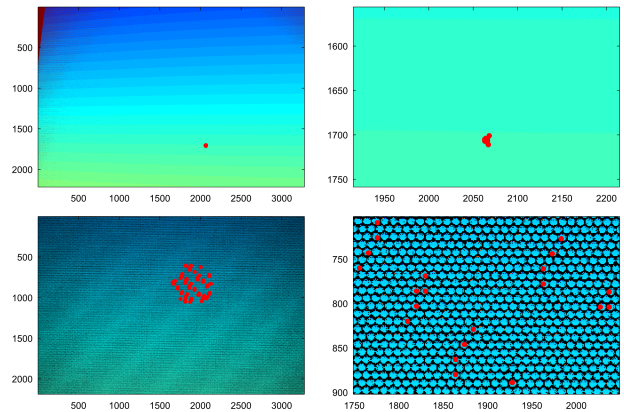


Figure 6. Distribution of points with similar code over the imaging plane.

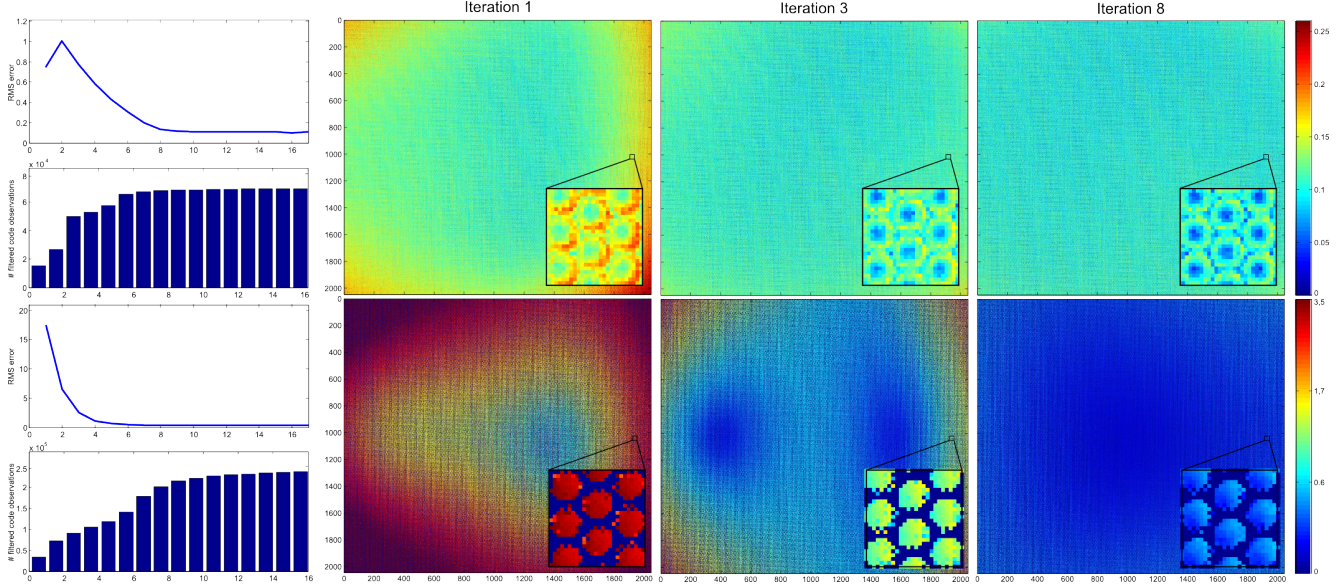


Figure 7. Evolution of the RMS error of rays during the optimization and number of culled outliers (see text for details).

maxzoom setting clearly result in micro images that cover a wider area of the target and include several repetitions of the same code. This property, obviously, is very useful if the data is to be used for surface reconstruction, since multiple and well separated observations of the same point will result (in principle) in a more accurate triangulation. The code distribution is even more apparent in the last column of the figure, where a slice of the coding has been plotted (corresponding to the black horizontal line on the third column of the figure).

The different distribution of codes can be understood very well also by looking at Figure 6. In the first row we overlay to a coded image obtained at minzoom the pixels with a code less than one code unit far from a given pair of coordinates, showing two level of details. In the second row we plot the same information over a coded image obtained at maxzoom setting. It can be seen that a minimum zoom level a point can be observed at most by one or two microlenses, while at the maximum level the same code is repeated over and over throughout a large number of different microlenses spanning a (relatively) large portion of the imaging sensor.

To this end we can affirm that, while lower zoom levels can be used for tasks such as refocusing of subaperture images creation, higher zoom levels are best suited for surface reconstruction tasks, where the larger disparity of the captured light rays results in a more accurate and robust triangulation. Specifically, in such cases accuracy would be granted by the wider angle between rays used for triangulation and robustness can be achieved out of their large number. In Sections 3.1 and 3.2 we will substantiate these statements.

2.3. Rays Optimization

In order to apply the ray optimization procedure presented in [3] we first need an initial estimation for the poses of the target and for the rays. We solved this problem by creating a quite raw sub-aperture image obtained by grouping all the microlens centers. With a very coarse approximation, such image can be considered as if it were produced by a pinhole camera, to this end we can get an initial estimate using the standard calibration procedure made available by OpenCV [7]. The optimization procedure has been performed using a total of 10 target poses. Since the non-parametric model does not provide an image plane, the ray optimization does not proceed by minimizing some reprojection error, as its common with standard calibration methods. Instead, [3] minimizes, with respect to ray and poses parameters, the sum of the squared distances between the 3D coordinates of a target point and the ray that observed it. Since our code advances of one unit for each monitor pixel, the RMS error is a metric measure in the Euclidean spaces expressed in units of pixels, that in our case, corresponds to about 0.25mm.

In Figure 7 we show the trend of this RMS error with respect to subsequent steps of the optimization process. As for the previous section, the first row represents the minzoom and the second the maxzoom setup (note that both for the plots and for the color-coded images the scales are different). In both cases the RMS error converges after just a few iterations to a final value that is well below one monitor pixel, representing in practice just a few hundredths of millimeter. Such accuracy should be put in context, considering that during the calibration procedure the target is placed about one meter far away from the camera. The col-

ored plots express the RMS error associated to each pixel of the camera sensors. While with our model pixels are just indexes, it is still interesting to see how the error is distributed on the sensor, both at a global and microlens scale. Specifically, we can observe that both zoom levels start with an anisotropic error distribution (probably due to the incorrect handling of the global distortion) and rapidly reach a lower and better distributed RMS error. This is a well-known feature of parameter-free models, which are very well capable at accommodating both local and global distortions.

Finally, the histograms of Figure 7 report the total number of ray outliers that have been culled after each iteration. Such culling happens when a ray received a valid code, but it cannot be correctly justified using the estimated poses, that is when it cannot pass close enough to the target codes it has observed. This can happen for a variety of reasons, including occlusions (especially on the border of the monitor) or wrong code recovery due to measurement errors. Note that, even at convergence, the removed rays are in a magnitude order between 10^4 and 10^5 , since the overall number of pixel is in the order of 10^7 we can conclude that more than 99% of sensor elements are correctly calibrated.

3. 3D Shape reconstruction

Generally speaking, the 3D position of an observed point can be recovered by triangulation [15] if it is observed by different points of view by means of an imaging process of known geometry. To this end, three sub-problems must be addressed:

1. the point must be identified for each point of view;
2. all the viewing direction must be recovered;
3. an intersection between them must be computed;

Sub-problem 1 can be solved in many different ways, ranging from image-based correspondences to structured light coding. Since the goal of this paper is not to introduce a matching method, and we are interested in factoring out most error sources that are not related to calibration. To this end, we solve the point identification problem using the same phase shift coding described in the previous section, which we have shown to be feasible and robust. On the other hand, with respect to subproblems 2 and 3 we introduce two task specific solutions.

3.1. Rays interpolation

One reason that makes constrained camera models such as [6] effective in practice is that exists a continuous mapping between any point (u, v) in the image plane and the corresponding ray exiting the camera. Consequently, 3D point triangulation can be solved by searching multiple occurrences of the same feature among the micro-lenses and intersecting the corresponding rays originating from the feature coordinates. In the case of phase-shift structured

light coding, the set of projected codes is known but is extremely unlikely that the camera probing rays would sample exactly such codes. However, under the assumption of locally planar 3D surface, each feature location (u, v) can be recovered by interpolating the observed codes in the image plane.

Conversely, if we model our camera as a generic sparse bundle of unconstrained probing rays, there is no trivial way to recover the ray \vec{r}_ℓ exiting the imaging device at any (possibly sub-pixel) image point p . Further, there is not even a concept of image plane but just some existing calibrated rays in space each one sampling an intensity value or, if we use a structured-light system, a two-dimensional phase code. In other words, the interpolation cannot be performed on the image plane but on a set of already known rays whose contribution in the estimation of \vec{r}_ℓ depends on what those rays are observing.

In the following Section we give a solution to the rays interpolation problem. Then, in Section 3.2 we describe in detail our proposed triangulation process for light-field cameras.

3.1.1 Rays manifold interpolation function

Let $R_d = \{\vec{r}_i\}$ a set of n known camera rays, and $\vec{w} = (w_1, \dots, w_n) \in \mathbb{R}^n$, $\sum_{i=1}^n w_i = 1$ a convex combination of weights.

We pose the ray interpolation problem in terms of rigid motions blending. Let $K \in SE(3)$, $K(\vec{r}_a) = \vec{r}_b$ be a the rigid motion that transforms a ray \vec{r}_a into \vec{r}_b . A famous result by Chasles [8] states that any rigid transformation is in fact a screw motion, i.e., a rotation around an axis placed anywhere in the 3D space, and a translation along the direction of the axis. when applied to rays \vec{r}_a and \vec{r}_b , the screw motion of all the points under a pure translation is limited by the length of the translation, while the motion induced by the rotation is unbounded. For this reason, we chose the rigid motion aligning \vec{r}_a to \vec{r}_b of minimal rotation as interpolant K_{ab} of the two rays. It is straightforward to see that the best possible rotation angle is the one between the two vectors \vec{d}_a and \vec{d}_b (i.e. $\text{acos}(\vec{d}_a^T \vec{d}_b)$) that rotates the first ray around the axis given by $\vec{d}_a \times \vec{d}_b$. When the rotation angle and axis is chosen, the optimal translation is the one moving \vec{r}_a according to a vector T orthogonal to $\vec{d}_a = \vec{d}_b$ whose length is equal of the distance between the two rays. In other terms, the best translation is the vector that connects the two nearest points \vec{s}_1 and \vec{s}_2 lying on \vec{r}_1 and \vec{r}_2 respectively. To summarize, given two rays \vec{r}_a and \vec{r}_b , we choose the interpolant K_{ab} as:

1. The rotation R_K around the axis $\vec{d}_a \times \vec{d}_b$ with angle $\text{acos}(\vec{d}_a^T \vec{d}_b)$
2. The translation $T_K = \vec{s}_2 - \vec{s}_1$

Given a set of interpolants mapping rays to rays, the problem of ray interpolation can be cast as one of averaging in the manifold of rigid transformations $SE(3)$. This is the path taken by *Dual-quaternion Iterative Blending* (DIB) that interpolates roto-translations in terms of a screw motion represented in terms of dual quaternion[26] and can be interpreted as computing a manifold averaging in $SE(3)$ endowed with the screw motion metric. More formally, DIB takes a set of rigid motions K_i with $i = 1, \dots, n$, and a set of weights w_i and finds the unique motion K^* that satisfies

$$\sum_{i=1}^n w_i \log(K_i K^{*-1}) = 0, \quad (1)$$

where \log is the logarithm map of the group $SE(3)$. This interpolation approach exhibits many useful properties such being *constant speed*, *shortest path* and *coordinate system independent*. Adopting this approach, given a set of rays $\{\vec{r}_1, \dots, \vec{r}_n\}$, we initialize the interpolated ray $\vec{r}_\ell = (\vec{d}_\ell, \vec{p}_\ell)$ as their weighted linear combination followed by a reprojection on the rays manifold:

$$\vec{d}_\ell = \frac{\sum_{i=1}^n w_i \vec{d}_i}{\|\sum_{i=1}^n w_i \vec{d}_i\|} \quad (2)$$

$$\vec{p}_\ell = \frac{\sum_{i=1}^n w_i \vec{p}_i}{\|\sum_{i=1}^n w_i \vec{d}_i\|} - \vec{d}_\ell \left(\frac{\vec{d}_\ell^T \sum_{i=1}^n w_i \vec{p}_i}{\|\sum_{i=1}^n w_i \vec{d}_i\|} \right) \quad (3)$$

Then, we compute the rigid transformations $K_{\ell,i}$ as the screw motion between \vec{r}_ℓ and each \vec{r}_i according to the procedure stated before. Once computed, all the $K_{\ell,i}$ are averaged via DIB with the weights \vec{w} to obtain K_{avg} . Finally, K_{avg} is applied to \vec{r}_ℓ to obtain a better estimate, and the procedure is repeated until convergence.

3.1.2 Interpolation weights estimation

In the structured light case, we base the weights estimation for the set of rays R_d in terms of the codes $c_1 \dots c_n \in \mathbb{R}^2$ observed by each $\vec{r}_1 \dots \vec{r}_n$.

In this work, we cast the weight estimation as a regularized barycentric interpolation. Following [23] we adopted inverse squared distance weight, but add a regularization factor λ . Specifically, let \mathbf{D} be the $n \times n$ diagonal matrix whose diagonal entries $d_{ii}, i = 1 \dots n$ are the squared distances between each observed code c_i and co . Then, the weight vector $\mathbf{w} = (w_1 \dots w_n)^T$ can be estimated as:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}^T (\mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{w} \\ \text{subject to} \quad & \mathbf{C} \mathbf{w} = co, \\ & \mathbf{1}^T \mathbf{w} = 1 \end{aligned} \quad (4)$$

$$\text{where } \mathbf{C} = \begin{pmatrix} | & & | \\ c_1 & \dots & c_n \\ | & & | \end{pmatrix}.$$

Problem 4 can be solved as a generalized least squared problem, yielding:

$$\mathbf{w} = \mathbf{A} \mathbf{C}^* (\mathbf{C}^* \mathbf{A} \mathbf{C}^{*T})^{-1} co^* \quad (5)$$

where $\mathbf{A} = (\mathbf{D} + \lambda \mathbf{I})^{-1}$, $\mathbf{C}^* = \begin{pmatrix} \mathbf{C} \\ \mathbf{1}^T \end{pmatrix}$ and $co^* = \begin{pmatrix} co \\ 1 \end{pmatrix}$.

3.2. Ray Selection and Triangulation

Since we are dealing with light-field cameras, we expect each known projected code co being visible in many different micro-lenses (see Fig. 6). Therefore, we start by searching in the acquired coded image all the pixel locations $\mathbf{u}_1 \dots \mathbf{u}_m$ whose codes $c(\mathbf{u}_1) \dots c(\mathbf{u}_m)$ are closer than a threshold dst_c to the projector code co . Due to their pinhole nature, we expect each \mathbf{u}_i to lie in a different micro-lens.

At this point, for each \mathbf{u}_i our aim is to create a new ray \vec{r}_{ui} that would have observed the code co from the same micro-lens where \mathbf{u}_i lies. Therefore, we look at the adjacent 8 neighbor pixels for valid codes and rays to be used as interpolation data. Then the ray is interpolated as explained above.

Once all the virtual rays $R_t = \{\vec{r}_{u1} \dots \vec{r}_{un}\}$ are collected, a robust least-square estimation is used to triangulate the 3D point P_{co} associated to the projected code co . Specifically, we adopt a RANSAC scheme to select a subset $R_{ti} \subseteq R_t$ producing a point P_{co} whose squared distance between each ray in R_{ti} is less than a threshold dst_s .

3.3. Quantitative Analysis and Comparisons

To assess the accuracy of the 3D surface reconstruction (and hence of the calibration) we need to define some proper error measure with respect to a ground truth. Here we opt to use the same monitor that has been used for the calibration by creating a set of additional shots and by triangulating its surface. This approach has the advantage of producing a planar surface that can be easily fitted allowing to compute the displacement error between the plane model and each triangulated point.

We tested a total of four setups at the maxzoom level (which is best suited for reconstruction for the reasons described in previous sections). The first one adopted the full method described in this Section. Subsequently, we disabled respectively the RANSAC selection and the ray interpolation, to study their role in the overall accuracy.

Finally, we adopted the calibration method proposed in [6], which has been introduced very recently and can be considered among the state-of-the art for parametric light-field calibration. This latter setup has been obtained using the software made available by the authors to calibrate the camera and, subsequently, by producing single rays r_i according to the imaging model presented in the original pa-

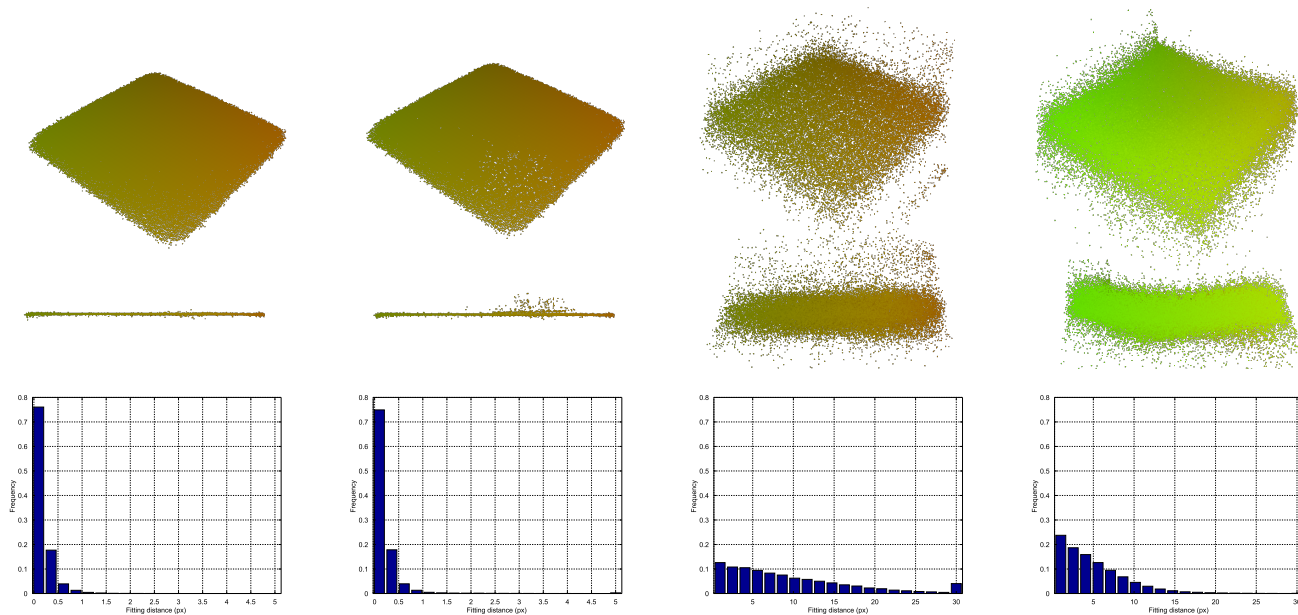


Figure 8. 3D reconstructions of the target plane using respectively our full method, disabling the RANSAC selection, disabling the interpolation and, finally, using the parametric calibration obtained with [6].

per. This way the obtained rays can be used directly within our pipeline. Note also that we applied both RANSAC and interpolation to these rays, in order to make the results comparable with the best results obtained with our method.

In Figure 8 we show the surfaces obtained from different points of view. We also show the distribution of the fitting errors for each method (note that the horizontal scales are different). The first two columns show the results obtained with our method with and without the RANSAC selection. The error distribution is quite similar, in fact just a few misplaced points appear when the consensus between ray is not enforced. This is somewhat expected since outlier culling has already been performed during calibration and we think that these inaccurate points result from coding errors that happened during the reconstruction shot. Conversely, ray interpolation is key to an accurate reconstruction. In fact, since a microlens covers a wide angle, we expect to observe quite large jumps in the codes between pixels (and rays), thus making the observation of exactly the same code (or a code near enough) from the different camera pixels very unlikely.

Finally, the triangulation obtained using the rays calibrated with [6] has a performance that is very similar to the one obtained without interpolation (even if, in this case, we are indeed interpolating the rays before triangulation). Moreover, a bit of global bending can be seen, probably due to a non perfect compensation of the distortion of the main lens. It should be noted that, while [6] is not really meant for 3D reconstruction, but only for image synthesis, it is

still one of the most relevant and recent calibration methods against which to compare.

4. Conclusions and Future Work

With this paper we studied the feasibility of the non-parametric calibration of light-field cameras using phase shift-based dense target coding and independent rays optimization. Despite the non conventional imaging process that characterizes plenoptic cameras, both target coding and rays calibration work well, to the extent that the complete camera sensor can be calibrated for moderated zoom levels. Additionally, we proposed specially crafted techniques for interpolating, selecting and triangulating the several views of the same material point that occurs in different microlens. The combination of such techniques with the high accuracy of the calibrated ray model enables the effective adoption of the light-field camera for 3D applications [2, 27] which could be unfeasible using parametric models. Such claims have been validated by comparing the surface reconstructed using the same correspondences respectively with our calibration and the one obtained using a state-of-the-art parametric method. Finally, this paper is focused on exploiting light-field cameras for 3D reconstruction, thus no effort is made to define a suitable (virtual) imaging plane to produce synthetic images for refocusing or subaperture synthesis purposes. This, however, is not an inherent limitation resulting from the adopted model and, indeed, the definition of an optimal imaging plane could be an interesting venue for further research.

References

- [1] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.
- [2] A. Albarelli, E. Rodolà, F. Bergamasco, and A. Torsello. A non-cooperative game for 3d object recognition in cluttered scenes. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2011*, pages 252–259, 2011.
- [3] F. Bergamasco, A. Albarelli, E. Rodolà, and A. Torsello. Can a fully unconstrained imaging model be applied effectively to central cameras? In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1391–1398, June 2013.
- [4] T. Bishop and P. Favaro. Plenoptic depth estimation from multiple aliased views. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 1622–1629, Sept 2009.
- [5] T. Bishop and P. Favaro. The light field camera: Extended depth of field, aliasing, and superresolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(5):972–986, May 2012.
- [6] Y. Bok, H.-G. Jeon, and I. Kweon. Geometric calibration of micro-lens-based light-field cameras using line features. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision ECCV 2014*, volume 8694 of *Lecture Notes in Computer Science*, pages 47–61. Springer International Publishing, 2014.
- [7] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, Inc., 1st edition, October 2008.
- [8] M. Chasles. Note sur les propriétés générales du système de deux corps semblables entr'eux et placés d'une manière quelconque dans l'espace; et sur le déplacement fini ou infiniement petit d'un corps solide libre. *Bulletin des Sciences Mathématiques, Astronomiques, Physiques et Chimiques*, 14:321–326, 1830.
- [9] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu. Light field stereo matching using bilateral statistics of surface cameras. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1518–1525, June 2014.
- [10] D. Cho, M. Lee, S. Kim, and Y.-W. Tai. Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3280–3287, Dec 2013.
- [11] D. Dansereau, O. Pizarro, and S. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1027–1034, June 2013.
- [12] F. Devernay and O. Faugeras. Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured environments. *Mach. Vision Appl.*, 13(1):14–24, Aug. 2001.
- [13] B. Goldluecke. Globally consistent depth labeling of 4d light fields. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CVPR '12*, pages 41–48, Washington, DC, USA, 2012. IEEE Computer Society.
- [14] M. Grossberg and S. Nayar. A general imaging model and a method for finding its parameters. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 108–115 vol.2, 2001.
- [15] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.
- [16] C. Hughes, P. Denny, E. Jones, and M. Glavin. Accuracy of fish-eye lens models. *Appl. Opt.*, 49(17):3338–3347, Jun 2010.
- [17] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwa. On the calibration of focused plenoptic cameras. In M. Grzegorzec, C. Theobalt, R. Koch, and A. Kolb, editors, *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, volume 8200 of *Lecture Notes in Computer Science*, pages 302–317. Springer Berlin Heidelberg, 2013.
- [18] A. Levin and F. Durand. Linear view synthesis using a dimensionality gap light field prior. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1831–1838, June 2010.
- [19] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '96*, pages 31–42, New York, NY, USA, 1996. ACM.
- [20] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen. Programmable aperture photography: Multiplexed light field acquisition. *ACM Trans. Graph.*, 27(3):55:1–55:10, Aug. 2008.
- [21] E. Lilienblum and B. Michaelis. Optical 3d surface reconstruction by a multiperiod phase shift method. *Journal of Computers (JCP)*, 2:73–83, 2007.
- [22] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11), 2005.
- [23] D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference*, ACM '68, pages 517–524, New York, NY, USA, 1968. ACM.
- [24] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In *Computer Vision - ECCV 2004*, volume 3022 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin Heidelberg, 2004.
- [25] Y. Taguchi, A. Agrawal, A. Veeraraghavan, S. Ramalingam, and R. Raskar. Axial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering. *ACM Trans. Graph.*, 29(6):172:1–172:8, Dec. 2010.
- [26] A. Torsello, E. Rodolà, and A. Albarelli. Multiview registration via graph diffusion of dual quaternions. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2441–2448, June 2011.
- [27] A. Torsello, E. Rodolà, and A. Albarelli. Sampling relevant points for surface registration. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2011*, pages 290–295, 2011.

- [28] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.
- [29] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy. Using plane + parallax for calibrating dense camera arrays. In *In Proc. CVPR*, pages 2–9, 2004.
- [30] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3), July 2007.
- [31] S. Wanner and B. Goldluecke. Spatial and angular variational super-resolution of 4d light fields. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *Computer Vision ECCV 2012*, volume 7576 of *Lecture Notes in Computer Science*, pages 608–621. Springer Berlin Heidelberg, 2012.
- [32] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph.*, 24(3):765–776, July 2005.
- [33] Z. Xu and E. Y. Lam. A high-resolution lightfield camera with dual-mask design, 2012.