Multi-Focus Plenoptic Simulator and Lens Pattern Mixing for Dense Depth Map Estimation

R. Ferreira[1], J. Cunha [1] and N. Goncalves[1]†

[1]Institute of Systems and Robotics - University of Coimbra, Portugal

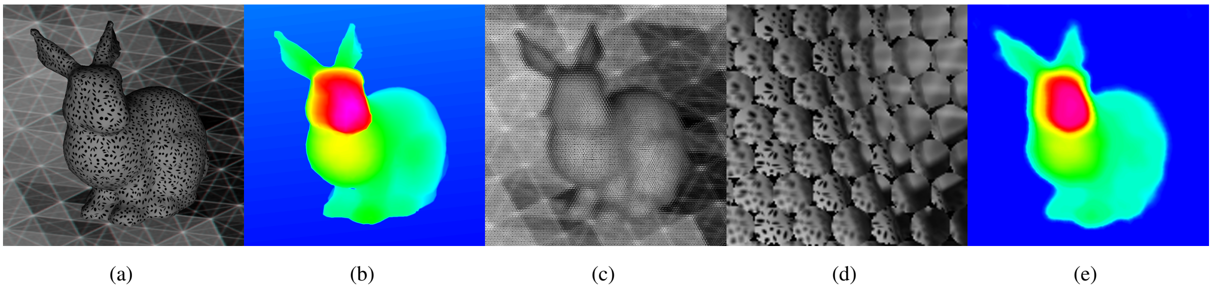


Figure 1: (a) Computer generated scene. (b) Depth ground truth of the generated scene. (c) Generated plenoptic image. (d) Closeup of the generated plenoptic image. (e) Depth estimation of our algorithm.

Abstract

Light field cameras capture a scene's multi-directional light field with one image, allowing the estimation of depth. In this paper, we introduce a fully automatic method for depth estimation from a single plenoptic image running a RANSAC-like algorithm for feature matching. The novelty about our method is the use of different focal-length lenses for multiple depth map refining, generating a dense depth map for future all-in-focus renders. We also present a plenoptic simulator which produces a plenoptic dataset from a 3D computer rendered scene. This simulator, which is unique, as far as we known, allows testing of plenoptic oriented algorithms since it can reproduce datasets with desired scene characteristics, providing the depth ground truth for error measurement. This work is a on-going project with promising results.

Categories and Subject Descriptors (according to ACM CCS): I.4.1 [Image Processing and Computer Vision]: Digitization and image capture—; I.4.8 [Image Processing and Computer Vision]: Scene analysis—

1. Introduction

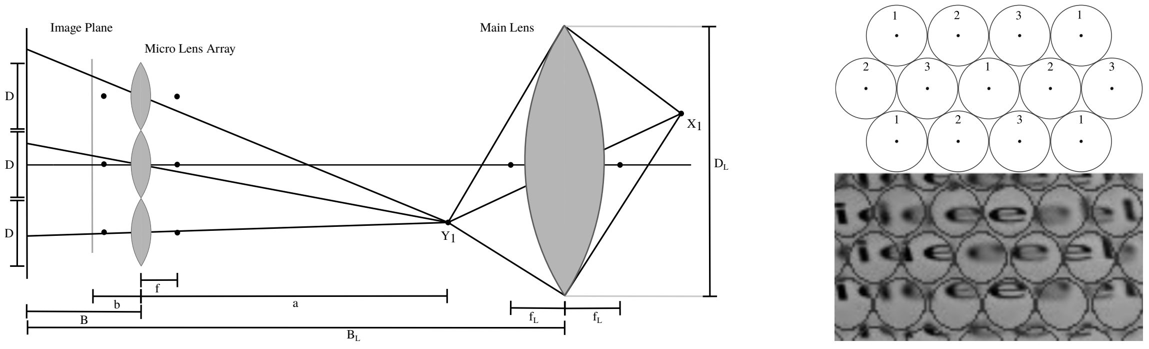
Plenoptic or light field cameras (PLF) are cameras that acquire the plenoptic function, that is to say that they know, for each pixel, the amount of light traveling in all directions. These cameras have received a lot of interest in the few last years since they inherently allow for multiple view geometry. Although formalized earlier (about 100 years ago), PLF cameras were commercially built only in the last one or two decades. These cameras are built by placing a microlens array behind the major optical lens of the system. This construction allows for the formation of an array of smaller images that compose the 4D light field and by easily sampling it. It is then straightforward to estimate the depth of the scene due to the redundancy created by the same point being imaged several times. The resolution of the images sampled is limited to the resolution of the CCD and this is the reason why only with nowadays CCDs resolutions we have medium to high quality PLF cameras (HD).

In 1908 Lippmann [Lip08] addressed the concept of plenoptic camera, suggesting the placement of an array of lenses between the camera's main lens and the film. This allows the capture of the light field of a scene. Now with digital image sensors we are able to extract information from the plenoptic images. This technology has several applications which can be divided into two major areas: depth estimation and image rendering.

With a raw plenoptic image we can achieve the scene depth, which is essential for image render. Dansearau and Bruton [DB04]

† e-mail: nunogon@deec.uc.pt

DOI: 10.2312/egsh.20161009



(a) (b)

Figure 2: (a) Plenoptic camera projection model where f is the micro-lens focal distance, D the micro-lens aperture, a is the distance from the micro-lens plane (MLP) to the main virtual image (VI), b distance from the MLP to main VI projected through the micro-lenses, B distance from the MLP to image plane (IP), f_L main lens focal distance, D_L main lens aperture and B_L the distance from the main lens plane to image plane. The point X_1 is projected through the main lens, obtaining the VI Y_1 (b) top image shows the hexagonal layout of different type lenses. The lens type is identified by number, while the bottom image is a sample from Raytrix dataset with different blurs in different lens types.

propose a depth estimation using 2D gradient operations. They were able to define the light field direction using a two plane parametrization (s,u) and (t,v), thus achieving the depth of the scene. Bishop and Favaro [BZF09] tried to compensate present aliasing since plenoptic cameras are not immune to spatial aliasing. Wanner and Goldluecke [WG12] used dominant directions on epipolar plane images to estimate the scene depth, claiming to have obtained results that surpassed the ones from Raytrix.

Most recently Fleischmann and Koch [FK14] approached the depth estimation paradigm with disparity between neighbor lenses. Using several types of regularization they achieve a per-lens dense map well suited for volumetric surface reconstruction techniques.

On the other hand, image rendering consists in converting the plenoptic image into a focused image, the same way a conventional camera would see the world. RenNg [NLB*05] proposed that each micro-lens contributes with only one pixel for the final rendered image. This is a fast processing method but produces low resolution images. Lumsdaine [LG08] suggested that each micro-lens contributes with a small patch of pixels for the final rendered image. This increases the final image resolution but generates artifacts. Perwass [PW12] achieves a final image with good resolution and low artifacts by back tracing each pixel to the image plane.

2. Multi-focus plenoptic cameras

The plenoptic 2.0 or multi-focus plenoptic camera has an array of multi-focal length micro-lens where each micro-lens have a different focal length from its neighbor lenses. There are at least three different focal lengths and each one is called a type. This construction allows to obtain a large depth of field and a rendered image with higher resolution. The most common lens type arrangement is hexagonal, as illustrated by the top image of figure 2b. Lenses with different focal lengths will present different blurs for the same plane and will be in focus for different depth ranges. The image on the bottom of figure 2b shows a sample of a scene at a constant depth. An object in front of the main lens will be projected through the main lens and then projected through the micro-lens array into the image plane, as show in figure 2a.

3. Feature detection and depth estimation

Our algorithm to estimate a dense depth map is based on photometric similarities between pairs of micro-lens images. We use SIFT descriptor to search for salient points. This method allows us to obtain the most significant points like corners, edges and contrast points only by adjusting threshold parameters. Salient points are then searched for in neighboring lenses to obtain correspondences, by relying on stereo epipolar geometry. Since we are provided a big number of salient points and their correspondences, we apply a RANSAC-method to obtain the best 3D point cloud. Our method is based on the back projection model presented by [PW12]. We estimate the depth for different micro-lens configurations instead of selecting micro-lens based on the effective resolution ratio (ERR) at the given virtual depth. We summarize our method as follows:

- Step 1 - **Selection of a subset of three lines** For each correspondence, a subset of three lines is considered. The central line is defined by the salient point and the test correspondence. For each central line two adjacent parallel lines are incorporated in the model, representing a one pixel error tolerance.
- Step 2 - **Estimation of the 3D virtual points** The previous defined lines are grouped two by two and for each pair it is computed the 3D point that minimizes the distance between them. The final 3D point has the median of their coordinates.
- Step 3 - **Testing the model** Having an hypothetical 3D point obtained in the previous step, we now need to test the hypothesis for this virtual point. The chosen error measurement is the distance of the virtual candidate point to all the correspondence lines obtained in the previous step.
- Step 4 - **Assessment of the model** A threshold is defined so we can distinguish the good from the bad estimations. This allows to assume which lines are suited to add to the model (labeled as

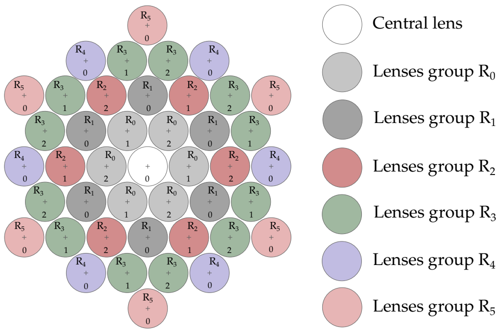


Figure 3: Illustration of the lens neighborhood, with every group labeled from R_0 to R_5, and lens type from 0 to 2. The lower value in the micro-lens illustration is the lens type.

inliers). If there is more than one outlier, the model is discarded and we go back to the first step. If not, we advance to Step 5.

- Step 5 - **Re-estimations of the 3D virtual point** This step is similar to Step 2. We re-estimate the 3D virtual point using only the inliers. These lines are again grouped two by two and the 3D point for every combination is the point that minimizes the distance between them. The final 3D point is the median coordinates of all points generated by every line combination.
- Step 6 - **Error metrics** In this step we evaluate the model in terms of error. It is a mean error from the inliers's distances obtained in Step 3.
- Step 7 - **Repeat steps 1-6 for every correspondence**

The output of the previous algorithm is a 3D point cloud of virtual points as projected by the main lens of the camera to their virtual image. At a final stage, a coarse regularization method will reproject the 3D points of the cloud to the micro-lens images and, thus, attribute an average depth value for every micro-lens. The final dense depth map is build by weighting the 3D point cloud depth and the coarse depth map. The main contribution of this algorithm is a smart mixture of neighboring micro-lenses of different type that, although with different blurs, are able to improve the depth estimation of the sparse point cloud and of the dense map.

As for the lens pattern used in Step 1 (where neighbor lenses are searched for replications of a given salient point) we use distinct combinations of lenses, generating different depth maps for this step. Knowing that for a multi-focus plenoptic camera there are lenses with different types, we define lens groups based on the lens type and the distance to the central lens. Figure 3 shows these configurations. We do a smart mixture of lens groups that, even mixing different blurs, is able to optimize the depth estimated, considering different depth ranges for the differently generated depth maps. Notice that the depth accuracy depends on the stereo baseline, which is smaller for higher scene depths. Our smart adaptive mixture of micro-lens is able to adjust baseline and range.

4. Plenoptic simulator

To test our algorithm to estimate the depth of a scene, we built a simulator of plenoptic images for a multi-focus plenoptic camera.

Each produced dataset must have a depth scene converted into gray-scale image, a micro-lens RGB image (plenoptic image) and a calibration file. The calibration file includes the simulated camera parameters used to generate the dataset.

4.1. 3D scene

Using OpenGL we are able to create a 3D world and access the depth buffer. For each scene we generate an RGB image of the scene and a gray-scale image with the depth values.

We created several scenarios to assess different characteristics of our algorithm. One of them, the "bolt" dataset, replicates the Raytrix "watch" dataset. It consists of a back plane with a watch image, two side planes, a floor plane and several cylinders representing the bolts. To replicate a silhouette with high detail, we created another dataset using the Stanford Bunny (figure 1a). We use the Bunny because it is a landmark within the world of computer vision and computer graphics.

For every dataset, textures are needed so that the depth estimation algorithm would find features. Textures are based on normalized images and are indexed to surface corners over the final scene rendering so that the surface will be filled with texture image.

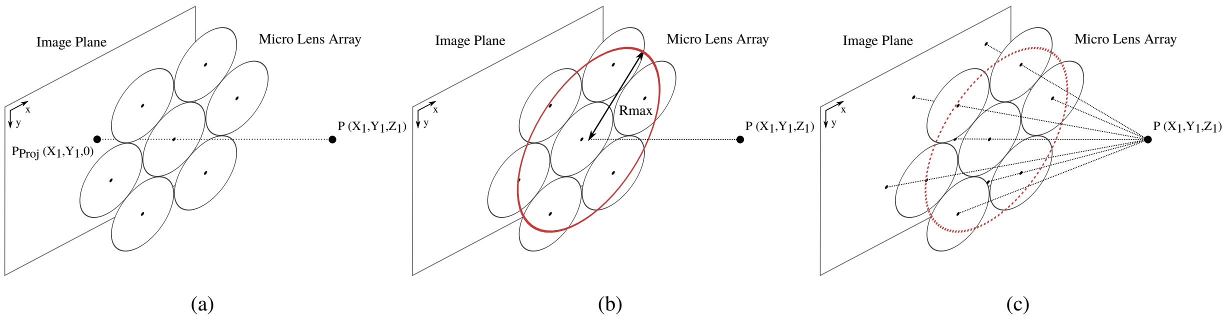
We introduced lighting to the object for the shadows to be noticed and thus the surface too, considering the presence of two different illumination types: ambient and diffuse. We opted to omit specular illumination since it removes detail in the object's texture, thus making feature recognition more difficult.

4.2. Micro-lens array projection

Having the 3D world, the next step is to project it into a plenoptic image. A plenoptic camera is constituted of a main lens, a micro lens array and a sensor. The main lens setting used mimics the 100*mm Zeiss* Planar because it is the main lens used by Raytrix cameras in their datasets. However the structure can be configured for different lenses. The micro lens array can also be fully configured by setting all focal distances, the distance from the sensor (image plane) to micro-lens array, lens diameters and pixel size.

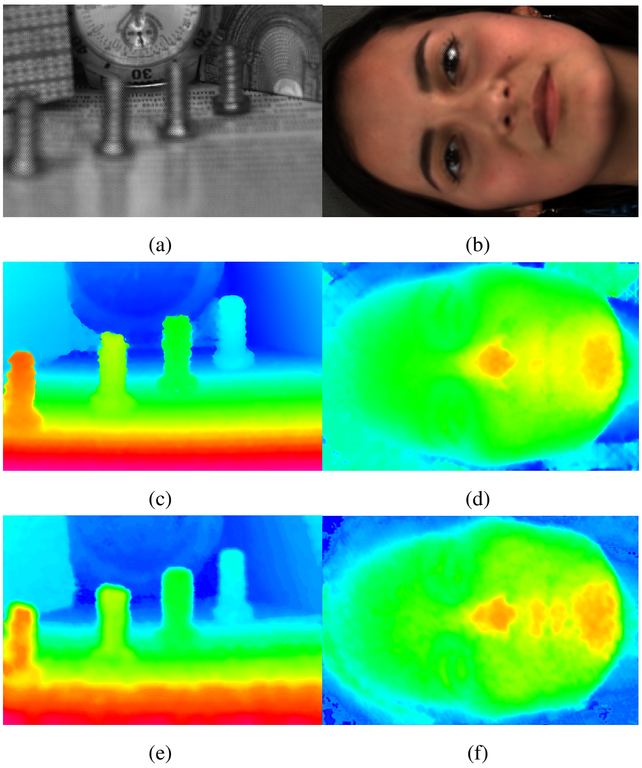
First, the world scene is projected through the main lens creating a virtual image. This virtual image is then projected through the micro lenses. The following steps describe this process.

- Step 1 - **Determine the central lens to which a point P belongs.** Project the point P to the image plane (figure 4a). If the projected point falls into the gab between micro-lenses, the lens selected is the one with the nearest center.
- Step 2 - **Determine the radius R_{max}.** Knowing this radius (equation (1)), we discard all lenses for which the distance from the center to point P exceeds R_{max}. This step is illustrated in figure 4b where R_{max} is drawn as a red boarder.
- Step 3 - **Lenses selection.** The calibration data provides the depth range for each lens and, from the previously selected lens group, we remove every lens with a different depth range (lens type) from the lens containing the projection of point P.
- Step 4 - **Compute the blur diameter for the virtual point projection.** A given virtual point projected through the micro lenses



(a) (b) (c)

Figure 4: Illustration of a few steps of the synthesization algorithm. The first step illustrated by (a) determines the central micro-lens to which point P belongs. Second step, illustrated by (b), is the R_{max} projection to determine which lenses project the virtual point P. Finally for the forth step, figure (c) illustrates the backtrace of P onto the image plane through the lenses within the R_{max} radius.



(a) (b) (c) (d) (e) (f)

Figure 5: Comparison of our and Raytrix's results for depth estimation on Raytrix's datasets. (a) Watch plenoptic image. (b) Andrea plenoptic image. (c) Raytrix's depth estimation for Watch. (d) Raytrix's depth estimation for Andrea. (e) Our depth estimation for Watch. (f) Our depth estimation for Andrea.

will have an associated blur. The blur diameter (circle of confusion) is given by $s = fD/z$, where s is the blur diameter, f is the focal length, D the aperture and z the virtual point's depth.

- Step 5 - **Apply the pixel value and the blur circle to the image.**

$$R_{max} = \frac{|\, z \,| \times D}{2 \times B} \tag{1}$$

The full dataset is composed of a plenoptic image, a depth ground truth image and a ".xml" file with its configurations.

5. Results and conclusions

Figures 1a, 1b and 1c show the generated dataset for the Stanford Bunny where a closeup on the rendered plenoptic image is shown in figure 1d. As for the depth estimation, figure 1e shows our depth estimation for the Stanford Bunny dataset. We also tested our algorithm on Raytrix datasets as shown in figure 5.

The results presented show that our algorithm is able to accurately estimate the depth of a scene and to improve the results in specially difficult areas such as the background and the silhouette of the foreground objects. This improvement is due to our mixture of lenses of different focal-lengths. Additionally, our algorithm is fully automatic with no human intervention.

This work is part of an on-going project with promising results and we intend to improve the algorithm and to extensively compare it against other methods.

References

[BZF09] Bishop T. E., Zanetti S., Favaro P.: Light field superresolution. *ICCP, IEEE International Conference on* (2009), 1–9.

[DB04] Dansearau D., Bruton L.: Gradient-based depth estimation from 4d light field. *Circuits and Systems. ISCAS 3* (2004), III – 549–52.

[FK14] Fleischmann O., Koch R.: Lens-based depth estimation for multi-focus plenoptic cameras. *Pattern Recognition* (2014), 410–420.

[LG08] Lumsdaine A., Georgiev T.: Full resolution lightfield rendering. *Indiana University and Adobe Systems, Tech. Rep* (2008).

[Lip08] Lippmann G.: Épreuves réversibles. photographies intégrals. *Comptes-Rendus Academie des Sciences 146* (1908), 446–451.

[NLB*05] Ng R., Levoy M., Brédif M., Duval G., Horowitz M., Hanrahan P.: Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR 2*, 11 (2005).

[PW12] Perwass C., Wietzke L.: Single lens 3d-camera with extended depht-of-field. *SPIE Human Vision and Electronic Imaging* (2012).

[WG12] Wanner S., Goldlueck B.: Globally consistent depth labeling of 4d light fields. *CVPR, 2012 IEEE Conference on* (2012), 41–48.