

Depth Range Accuracy for Plenoptic Cameras

Nuno Barroso Monteiro

*Institute for Systems and Robotics, University of Lisbon, Portugal
Institute for Systems and Robotics, University of Coimbra, Portugal*

Simão Marto

Institute for Systems and Robotics, University of Lisbon, Portugal

João Pedro Barreto

Institute for Systems and Robotics, University of Coimbra, Portugal

José Gaspar

Institute for Systems and Robotics, University of Lisbon, Portugal

1. Supplementary Material

1.1. Datasets for Reconstruction Accuracy

The experiments presented in this work to evaluate the reconstruction estimation accuracy were performed using the 1st generation Lytro camera (Figure 1). This camera was used to acquire seven datasets (Figure 4) under different zoom and focus settings¹. The zoom and focus step settings of each dataset were determined by placing a target object at a pre-determined depth of the encasing of the camera and autofocusing on this object. This allows to define a plane in focus by the main lens that is close the target object. Thus, the focus depth is assumed to be the depth of the target object.

The datasets acquired encompass images for calibration and depth range assessment. Each dataset is provided with a set of calibration plenoptic images since the camera parameters are different for each dataset. The calibration images are different from the depth images to ensure the results do not suffer from any type of overfitting effect. The calibration plenoptic images were captured using a 19×19 calibration grid of 3.18 mm cells placed at different poses and at different depths close to the target object depth bearing in mind that a minimum of 10 poses are required. On the other hand, the depth plenoptic

Email addresses: nmonteiro@isr.tecnico.ulisboa.pt (Nuno Barroso Monteiro),
smarto@isr.tecnico.ulisboa.pt (Simão Marto), jpbar@isr.uc.pt (João Pedro Barreto),
jag@isr.tecnico.ulisboa.pt (José Gaspar)

¹www.isr.tecnico.ulisboa.pt/~nmonteiro/datasets/plenoptic/cviu2017/



Figure 1: First generation Lytro camera.

images were captured using two different grid sizes: 19×19 grid of 6.10×6.08 mm cells and 5×7 grid of 26.50×26.38 mm cells. The grids for the depth plenoptic images were placed parallel to the encasing of the camera and at a regular spacing of 0.05 m from the camera for depth values ranging from 0.05 to 2.00 m. The two grid sizes are used for the depth plenoptic images since the depth range evaluated is wide and it is necessary to have a reasonable number of detections to assess the depth accuracy. The smaller grid size was placed up to a maximum depth of 1.0, 1.5 and 2.0 m according to the focus depth considered 0.05, 0.50 and 1.50 m. The bigger grid size was placed considering all depth range evaluated. Table 1 summarizes the properties of the datasets acquired.

Dataset	Zoom Step	Focus Step	Focus Depth (m)	Calibration Depth Range (m)	Calibration Poses	Ray Reprojection Error (mm)	Depth Poses
A	982	654	0.05	0.05 - 0.25	30	0.0993	60
B	754	941	0.05	0.05 - 0.35	30	0.1398	60
C	601	1212	0.05	0.10 - 0.40	14	0.2447	60
D	600	985	0.50	0.30 - 0.70	36	0.1357	70
E	335	1361	0.50	0.30 - 0.80	36	0.1267	70
F	337	1253	1.50	1.00 - 1.70	48	0.1806	80
G	100	1019	1.50	1.00 - 1.80	51	0.1381	80

Table 1: Information of the datasets acquired for calibration and depth range assessment. The ray reprojection error [1] corresponds to the error obtained using the full set of calibration plenoptic images.

The depth ranges used for the calibration procedure were defined relatively to the plane in focus by the main lens and considering the field of view of the camera. The depth range is defined relatively to the target object depth to have sharper viewpoint images which allow to detect more accurately the calibration grid points. The minimum depth value for the range was defined in order to have the full calibration grid in the viewpoint images. In Figure 2, one can see

35 the blurring that occurs for depths farther from the target object depth. On the other hand, the depth ranges used for the depth poses can be outside this range since the grids may fall out of the field of view. The number of calibration images is different among the several datasets to ensure a ray reprojection error [1] below 0.2 mm for each dataset (Table 1). The only dataset that has a ray
 40 reprojection error higher than 0.2 mm is Dataset C.

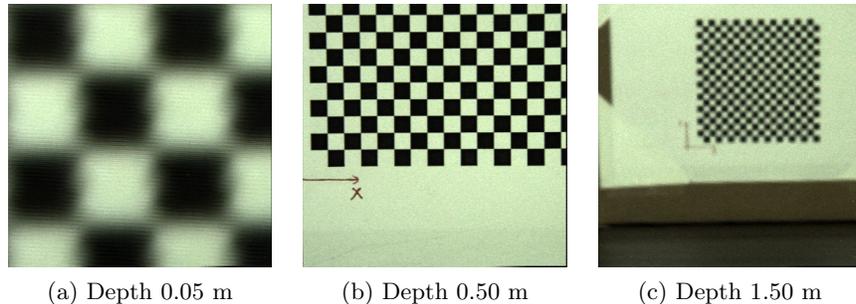


Figure 2: Viewpoint images obtained from the depth plenoptic images of the smaller grid of Dataset E. **(a)** Viewpoint image for smaller grid placed at 0.05 m. **(b)** Viewpoint image for smaller grid placed at 0.50 m. **(c)** Viewpoint image for smaller grid placed at 1.50 m. The Dataset E has the world focal plane at 0.50 m, which leads to sharper images near the world focal plane **(b)** and blurred images as we are farther from the world focal plane **(a)** and **(c)**.

The raw images acquired by the the standard plenoptic camera used have 3280×3280 pixels (Figure 3). After the decoding process described in [1] and removing a border of two pixels in i and j due to demosaicking and edge artifacts, the lightfield size for is $11 \times 11 \times 378 \times 379$ pixels. Thus, among all datasets, a
 45 wide range of viewpoint images can be obtained, more precisely 58 080 viewpoint images.

1.2. Projection Sets

In Section 4.1, two projection sets \mathcal{P}_{kl} and \mathcal{P}_{ij} are defined to obtain the multiple projections of a point on the lightfield captured by the camera (equations (8) and (9)). The projection sets allow to obtain the maximum number
 50 of projections on the lightfield. The conditions to apply each of the projection sets are defined using the slopes m_k and m_l in the ray-spaces (i, k) and (j, l)

$$m_k = -\frac{h_{si} + z h_{ui}}{h_{sk} + z h_{uk}} \quad , \quad m_l = -\frac{h_{tj} + z h_{vj}}{h_{tl} + z h_{vl}} \quad . \quad (1)$$

Namely, for $|m_{(\cdot)}| \leq 1$ we should use the projection set \mathcal{P}_{kl} while for $|m_{(\cdot)}| > 1$ we should use the projection set \mathcal{P}_{ij} . Nonetheless, these conditions do not
 55 directly relate with the depth of a point in the object space. One can redefine the conditions solving for the depth of the point, leading to projection rays defined by the projection set \mathcal{P}_{ij} whenever the depth of the point in the object

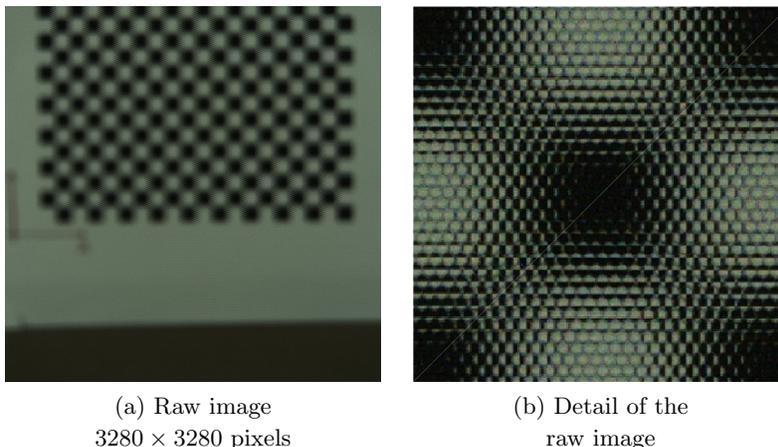


Figure 3: **(a)** Debayered raw image from a standard plenoptic camera. **(b)** Zoom of the debayered raw image to show the effect of the microlens array on the image captured by the sensor. These images are obtained from Dataset E.

space $z \in \chi$ where $\chi = \left] -\frac{h_{si} + h_{sk}}{h_{uk} + h_{ui}}, \frac{h_{si} - h_{sk}}{h_{uk} - h_{ui}} \right[$. The projection set \mathcal{P}_{kl} is used whenever $z \notin \chi$.

60 The depth limits of the set χ are not easily interpretable expanding the entries of the intrinsic matrix with the parameters defined by Dansereau et al. [1]. Thus, in Figure 5, we relate the projection sets with the depths of the points in the object space (in the world coordinate system) for the acquired Datasets D and F. The change of coordinate systems was performed considering the rigid
65 body transformations described in Section 6. The depth of the point in the world coordinate system is defined as the distance to the encasing of the camera. This figure depicts that the projection set defined by \mathcal{P}_{kl} is used for points farther from the camera while the projection set \mathcal{P}_{ij} is used for points near the camera.

1.3. Reconstruction Methodologies Comparison

70 In Section 5, two reconstruction methodologies are compared considering a projection error that is obtained by rounding the microlenses (k, l) and pixels (i, j) to the nearest integer. In this section, we assume that the projection error follows a Gaussian distribution with zero mean and different standard deviation values. Figure 6 depicts the results assuming that the projection error affects
75 all coordinates of the lightfield.

In this figure, one can see that the point reconstruction methodology (equation (11)) gives better results, relatively to the rounding error counterpart, when the noise added is lower than the error introduced by the rounding of microlenses and pixels indices. Furthermore, this figure shows that the point
80 reconstruction using line parameters (equation (14)) provide better results than the point reconstruction independently of the noise content and source (rounding or Gaussian).

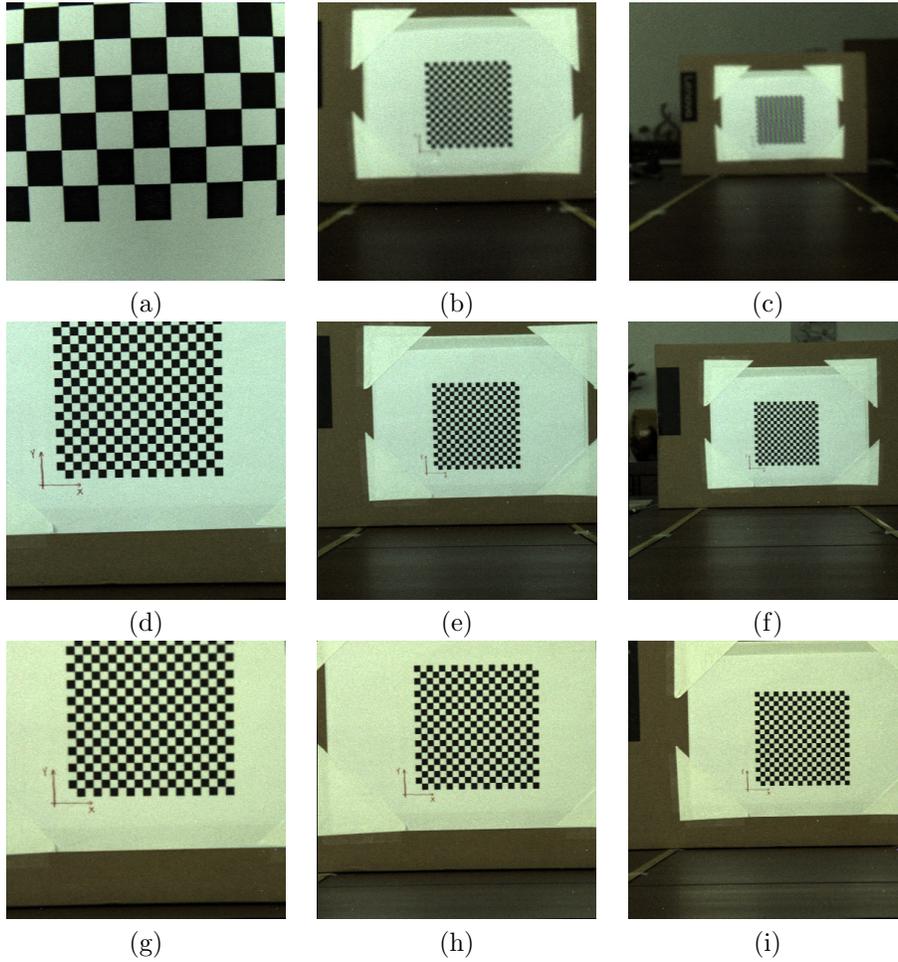


Figure 4: Viewpoint images with 378×379 pixels for grid poses at different depths for Datasets A, D and F. The viewpoint images for Dataset A correspond to grid poses at 0.05 m (a), 0.55 m (b) and at 1.00 m (c). The viewpoint images for Dataset D correspond to grid poses at 0.55 m (d), 1.10 m (e) and at 1.50 m (f). The viewpoint images for Dataset F correspond to grid poses at 1.10 m (g), 1.50 m (h) and at 2.00 m (i).

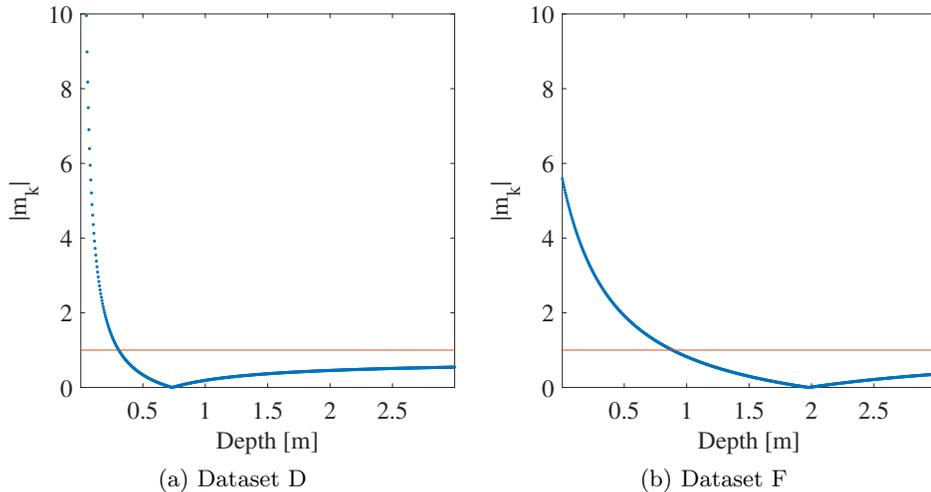


Figure 5: Evolution of slope $|m_k|$ (blue dots) with the depth of the points in the object space for Datasets D **(a)** and F **(b)**. The red line $|m_k| = 1$ defines the projection set that will be used to obtain the projection rays. For points below the red line, the set \mathcal{P}_{kl} is used. For points above the red line, the set \mathcal{P}_{ij} is used.

Additionally, since point reconstruction methodologies consider features detected with sub-pixel precision on images obtained from the lightfield, one modeled the error introduced by the feature detectors as a Gaussian distribution with zero mean and different standard deviation values. The lightfield allows to obtain viewpoint and microlens images by fixing either the (i, j) or (k, l) coordinates, respectively. Hence, in Figure 7, one considered that the feature detectors introduce error only in the (k, l) coordinates while the coordinates (i, j) are rounded to the nearest integer, i.e., features are detected on viewpoint images. In Figure 8, one considered that the feature detectors introduce error only in the (i, j) coordinates while the coordinates (k, l) are rounded to the nearest integer, i.e., features are detected on microlens images like in [2].

These figures continue to depict that the point reconstruction using line parameters (equation (14)) gives better results. Furthermore, one can see that the variance of the reconstructed depth is greater when adding noise to the coordinates (i, j) . This indicates that the reconstruction is more robust for noise added to the coordinates (k, l) .

1.4. Camera Parameters

In Section 6.1, we analyzed the camera parameters, zoom step and focus step, with the depth of a target plane. Due to lack of space, the results presented in Section 6.1 are not complete. In this section, we added additional 2D views of the volume in Figure 5.(a) (Figure 9).

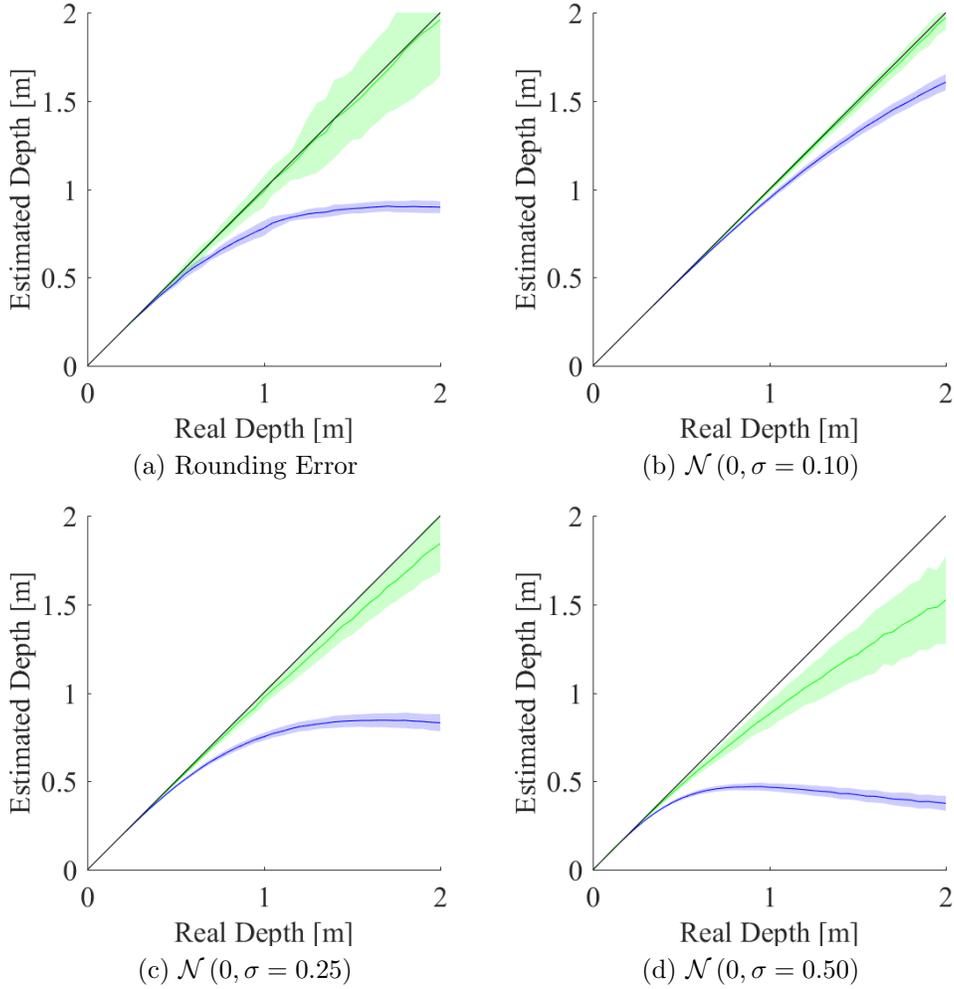


Figure 6: Reconstructed depth for randomly generated points at depths ranging from 0.01 to 2.00 meters. The projection error modeled as additive Gaussian noise affects all coordinates of the lightfield (b-d). The point reconstruction applied to the projection rays Φ_i (equation (11)) is presented in blue while the point reconstruction from line parameters (equation (14)) is presented in green. The mean for the estimated depth is presented as a darker line and the brighter shaded areas correspond to the standard deviation. The depth ground truth is represented with a black line.

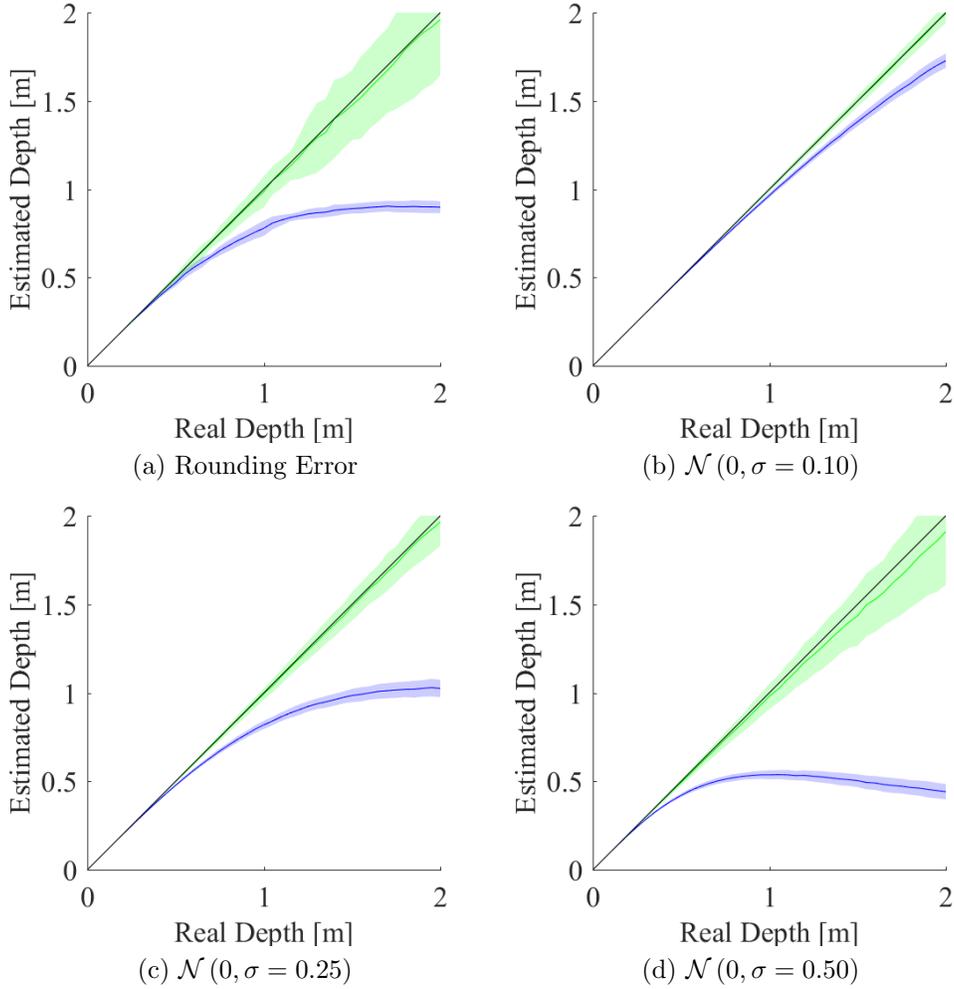


Figure 7: **Viewpoint Image Feature Detector Case.** Reconstructed depth for randomly generated points at depths ranging from 0.01 to 2.00 meters. The error modeled as additive Gaussian noise affects coordinates (k, l) of the light-field (b-d). The point reconstruction applied to the projection rays Φ_i (equation (11)) is presented in blue while the point reconstruction from line parameters (equation (14)) is presented in green. The mean for the estimated depth is presented as a darker line and the brighter shaded areas correspond to the standard deviation. The depth ground truth is represented with a black line.

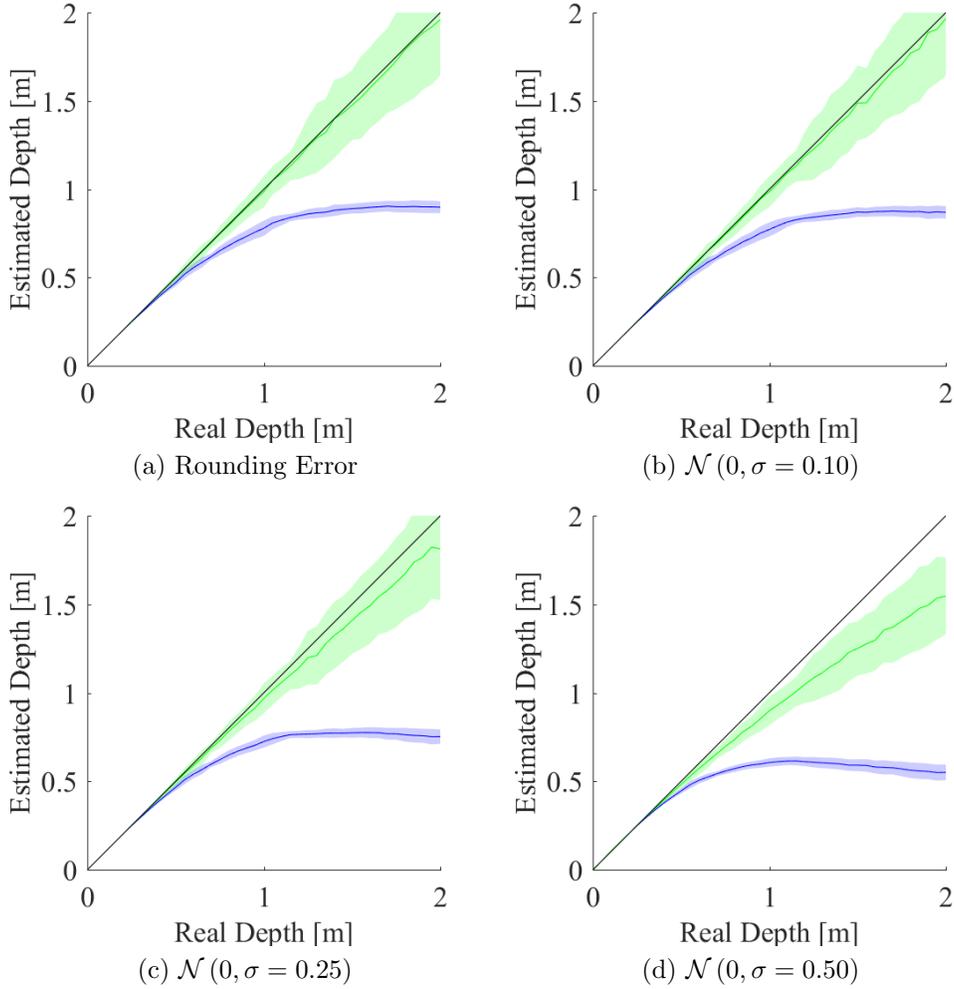


Figure 8: **Microlens Image Feature Detector Case.** Reconstructed depth for randomly generated points at depths ranging from 0.01 to 2.00 meters. The error modeled as additive Gaussian noise affects coordinates (i, j) of the light-field (b-d). The point reconstruction applied to the projection rays Φ_i (equation (11)) is presented in blue while the point reconstruction from line parameters (equation (14)) is presented in green. The mean for the estimated depth is presented as a darker line and the brighter shaded areas correspond to the standard deviation. The depth ground truth is represented with a black line.

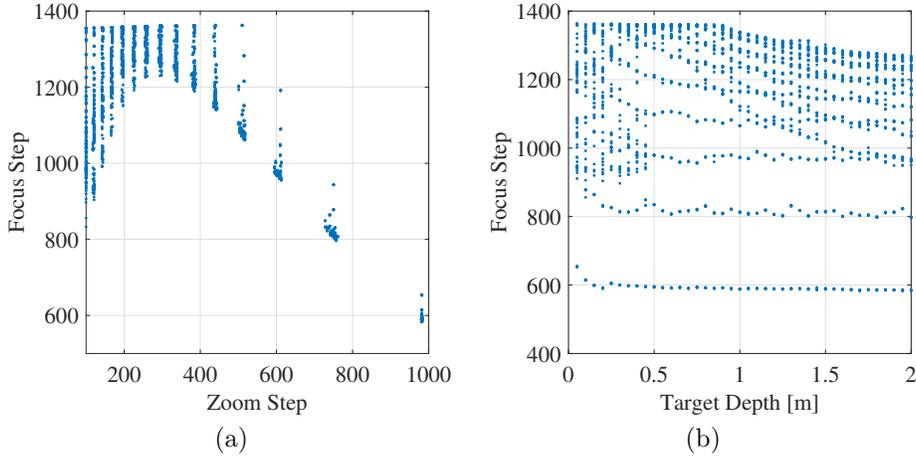


Figure 9: Camera autofocus given zoom step and target object depth. **(a)** represents the zoom step and focus step data points acquired. **(b)** represents the focus step and target object depth points acquired.

1.5. Experimental Results

105 In Section 6.2, we analyzed the influence of the world plane in focus on the reconstruction estimation accuracy. Due to lack of space, the results presented in Section 6.2 are not complete. In this section, we added figures with several dataset groupings for fixed focus depth (Figure 10) and for fixed zoom step (Figure 11).

110 Additionally, we added in Table 2, the results of the reconstruction estimation accuracy not removing the radial distortion component. Comparing with Table 3, one can see that these results are very similar leading to the conclusion that the radial distortion does not play an important role on the reconstruction estimation accuracy. Nonetheless, one has to notice that the depth range with
 115 normalized reconstruction error lower or equal to 10% is larger for Datasets E and F, and that Dataset C has a normalized reconstruction error that is always greater than the 10% for all depth range analyzed.

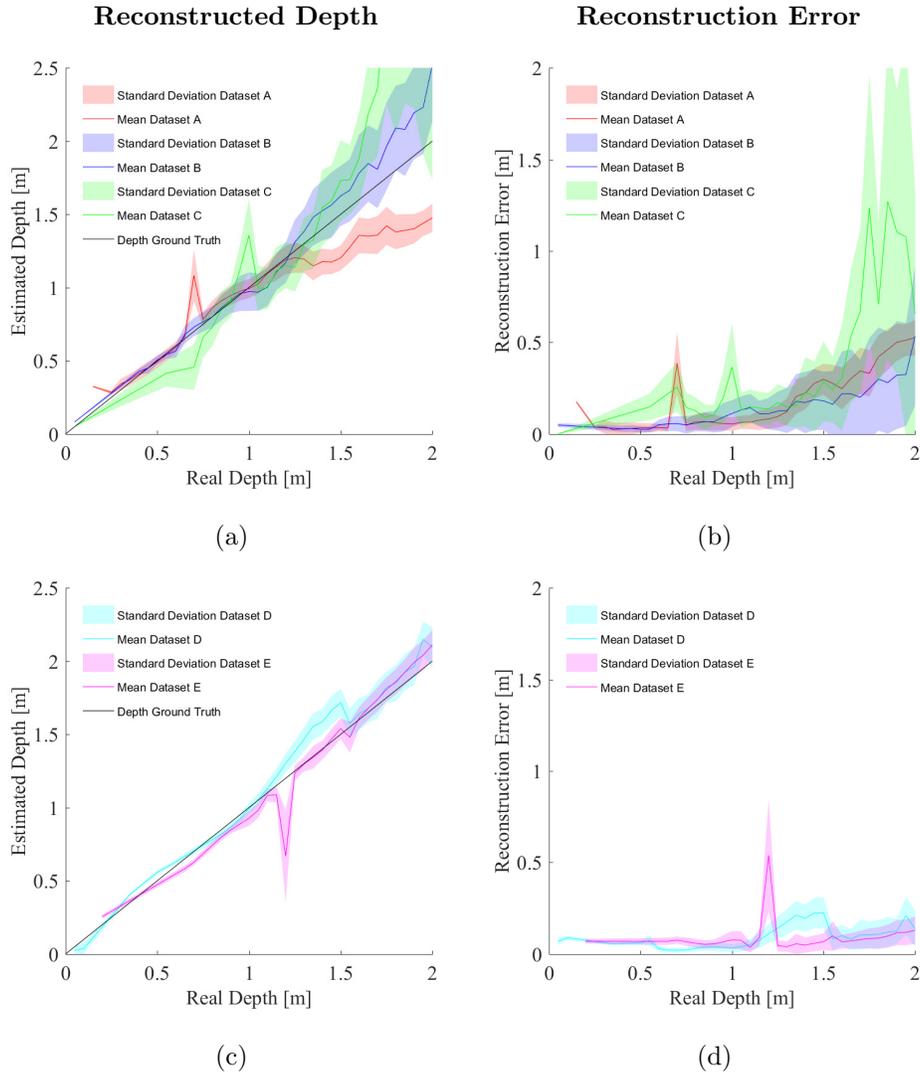


Figure 10: Reconstruction estimation accuracy with zoom step. The **first column** depicts the reconstructed depth while the **second column** depicts the reconstruction error for the estimated points obtained for datasets A through E. The **first row** groups the datasets with focus depth at 0.05 m (Datasets A, B and C) and the **second row** groups the datasets with focus depth at 0.50 m (Dataset D and E).

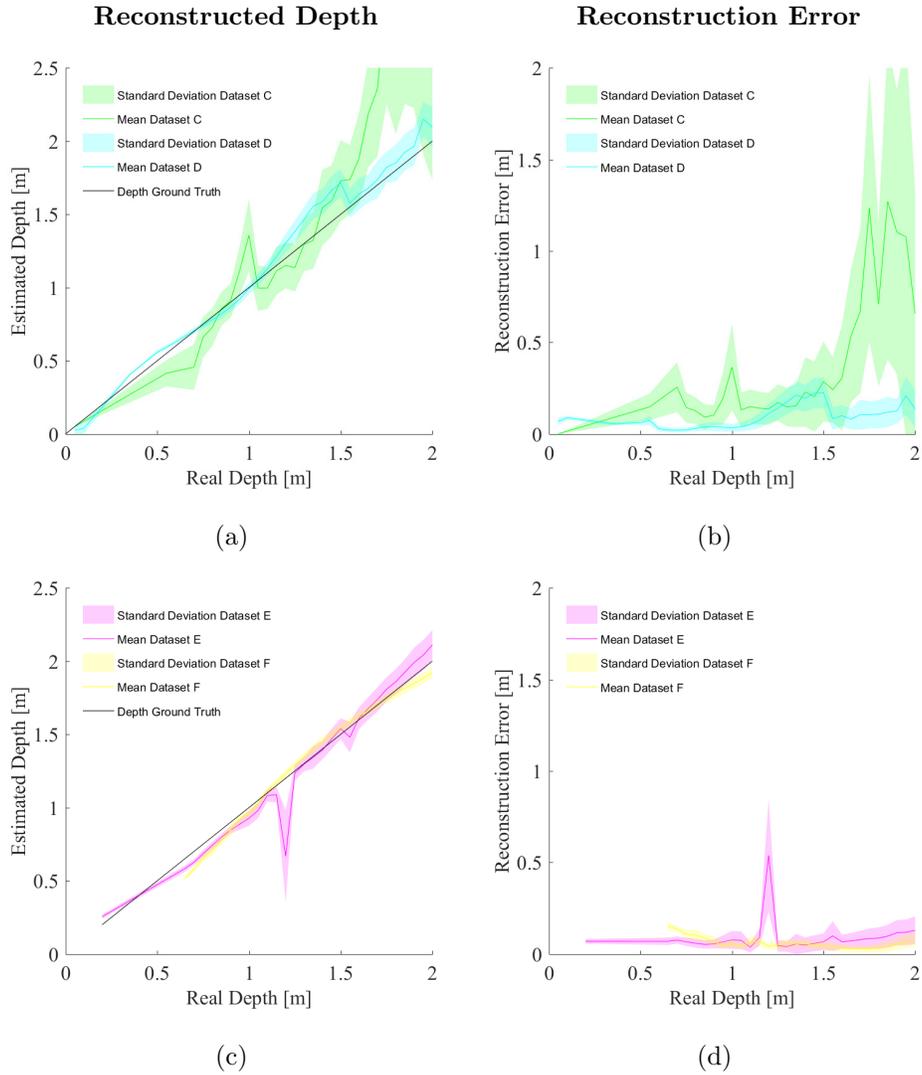


Figure 11: Reconstruction estimation accuracy with focus depth. The **first column** depicts the reconstructed depth while the **second column** depicts the reconstruction error for the estimated points obtained for datasets C through F. The **first row** groups the datasets with zoom step close to 600 (Datasets C and D) and the **second row** groups the datasets with zoom step close to 336 (Datasets E and F).

Dataset	Depth Range (m)	Mean \pm STD Error in Depth Range (%)	Mean \pm STD Error (%)
A	0.35 - 1.30	6.84 ± 5.11	16.67 ± 6.28
B	0.40 - 1.30	7.89 ± 5.96	13.72 ± 9.73
C	Not Defined	Not Defined	23.74 ± 17.72
D	0.60 - 2.00	5.18 ± 3.14	14.18 ± 4.87
E	0.65 - 2.00	5.48 ± 3.04	8.05 ± 4.01
F	0.85 - 2.00	3.97 ± 1.42	5.67 ± 1.57
G	1.50 - 1.85	1.94 ± 0.61	1.94 ± 0.61

Table 2: Depth ranges for the datasets acquired not removing the radial distortion component. The depth ranges are identified as the regions whose mean for the normalized reconstruction errors is lower or equal to 10%. The mean and standard deviation (STD) for the normalized reconstruction errors within the depth ranges defined and for all ground truth depths are also depicted.

References

- 120 [1] D. G. Dansereau, O. Pizarro, S. B. Williams, Decoding, calibration and rectification for lenselet-based plenoptic cameras, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 1027–1034.
- [2] Y. Bok, H.-G. Jeon, I. S. Kweon, Geometric calibration of micro-lens-based light field cameras using line features, *IEEE transactions on pattern analysis and machine intelligence* 39 (2) (2017) 287–300.