

Depth Estimation from Micro Images of a Plenoptic Camera

Jennifer Konz, Niclas Zeller, Franz Quint
 Karlsruhe University of Applied Sciences
 Moltkestr. 30, 76133 Karlsruhe, Germany
 {jennifer.konz, niclas.zeller, franz.quint}@hs-
 karlsruhe.de

Uwe Stilla
 Technische Universität München
 Arcisstr. 21, 80333 Munich, Germany
 stilla@tum.de

Abstract—This paper presents a method to calculate the depth values of 3D points by means of a plenoptic camera. Opposed to other approaches which use the totally focused image to detect points, we operate directly on the micro images taking thus advantage of the higher resolution of the plenoptic cameras raw image. Depth estimation takes place only for the points of interest, resulting in a semi-dense approach. The detected points can further be used in a subsequent simultaneous localization and mapping (SLAM) process.

Index Terms - depth estimation, focused plenoptic camera, micro images

I. INTRODUCTION

The concept of a plenoptic camera is known for over one hundred years (Ives, 1903 [1], Lippmann, 1908 [2]) but only due to the capabilities of nowadays graphic processor units (GPUs) an evaluation of video sequences with acceptable frame rates is possible.

The main advantage of a plenoptic camera is the depth information which can be estimated from only a single image. A traditional camera, which captures a 2D image of a scene, does not provide any depth estimation from one shot. In comparison, a plenoptic camera captures a complete 4D lightfield representation, which is suitable to calculate depth information [3][4].

There have been developed two concepts of plenoptic cameras: The unfocused plenoptic camera [5] and the focused plenoptic camera [6]. In this paper we work with focused plenoptic cameras. In these, a micro lens array (MLA) which is placed between the main lens and the sensor focuses the image of the former on the later. Thus, the raw image of this type of camera consists of many micro images, which each show a portion of the main lens image. These portions are pictured from a slightly different perspective in neighboring micro images. The disparities of corresponding points in the micro images enable the estimation of depth. The procedure for depth estimation with this type of camera is described e.g in [7]. This depth is in a first instance a virtual depth, i.e. related to internal parameters of the camera. However, calibrating the camera allows to compute metrical depth values. For the calibration process of the camera as well as for the procedures to synthesize a totally focused image please refer to [8].

A. Contribution of this work

This paper focuses on simplifying depth estimation by detecting points of interest (POI) directly in the raw image and not in the synthesized totally focused image. By matching POIs in the raw image that represent the same 3D point, its depth will be estimated. Because depth estimation mainly relies on the quality of matching the POIs from different micro images, different POI detectors are used and compared. It is of particular interest whether ordinary detectors which have been developed for traditional images can also be used for the raw image provided by a focused plenoptic camera.

Correspondence for points in different micro images is sought using epipolar geometry. Since the sensor is placed in front of the main lenses image, the image coordinates of points in the micro image differ from those in the (virtual) image of the main lens. This has to be accounted for in depth estimation, which is done by a linear regression. With a subsequent consideration of the depth information and calculation of the error between the actual projections and the matched features, the results of the depth estimation can further be improved.

The paper is structured as follows. In section II.A the concept of the focused plenoptic camera which is used in our approach is described. The depth estimation is explained based on this configuration. The particularities of POI detection in raw images are presented in section III. Section IV deals with matching the points highlighted by the detectors in different micro images. In section V we formulate the depth estimation for the matched image points. Section VI presents the results of the depth estimation by using different detectors and compares them. We conclude with section VII.

II. THE FOCUSED PLENOPTIC CAMERA

To highlight the differences between the set-up of a traditional camera and a plenoptic camera, the optical path of a thin lens is displayed in Figure 1. For a traditional camera the thin lens equation as given in eq. (1) can be used to describe the relation between the object distance a_L and the image distance b_L using the focal length f_L of the main lens.

$$\frac{1}{f_L} = \frac{1}{a_L} + \frac{1}{b_L} \quad (1)$$

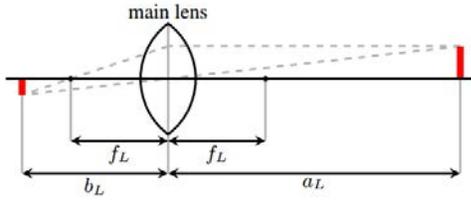


Fig. 1. Optical path of a thin lens [9]

With the configuration of traditional cameras displayed in Figure 1 the intensity of incident light is recorded on the image sensor. The 2D-image recorded by a traditional camera does not provide any information about the object distance. To gain information about the object distance a_L a plenoptic camera can be used.

A plenoptic camera consists of a micro lens array (MLA) which is placed between the main lens and the sensor. Regarding the position of the MLA and the sensor to the main lens image, two different configurations of a focused plenoptic camera are described by Lunsdaine and Georgiev [6][10].

In the Keplerian configuration the MLA and the sensor are placed behind the main lens images (s. Figure 2), whereas in the Galilean configuration MLA and sensor are in front of the main lens image (s. Figure 3). In the Galilean configuration the main lens image is only a virtual image. The plenoptic camera used in this work is with Galilean configuration.

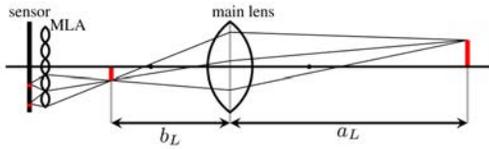


Fig. 2. Keplerian configuration [8]

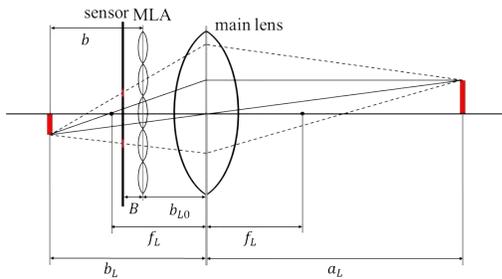


Fig. 3. Galilean configuration (based on [9])

The MLA of our camera has a hexagonal arrangement of the micro lenses (cf. Figure 4). There are three different types of micro lenses on the MLA, having different focal lengths. Thus different virtual image distances (resp. object distances) are displayed in focus on the sensor. Therefore, the effective depth of field (DOF) of the camera is increased compared to a focused plenoptic camera with a MLA consisting of micro lenses with the same focal length [8].

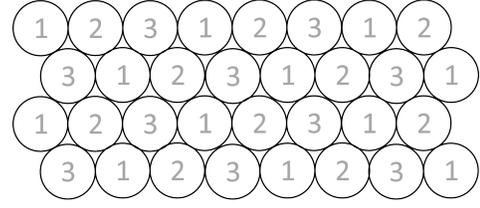


Fig. 4. Arrangement of the MLAs. Different micro lens types are marked by different numbers [11].

Each micro lens of the MLA produces a micro image on the sensor (s. Figure 5). However, depending on the focal length of the corresponding micro lens, a 3D point projected in different micro lenses will be focused in some of the micro images, while it is unfocused in others.

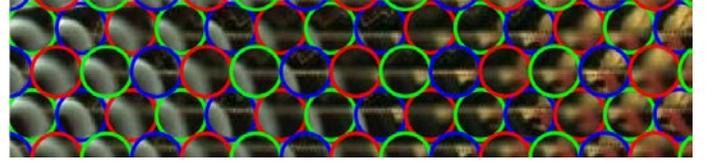


Fig. 5. Section of the micro lens images (raw image) of a Raytrix camera. Different micro lens types are marked by different colors. [8]

The thin lens equation of the main lens of a plenoptic camera in Galilean configuration can be written as in eq. (2) using the parameters of the camera. The parameter b_{L0} is the distance between the main lens and the MLA and b represents the distance between the MLA and the virtual image.

$$\frac{1}{f_L} = \frac{1}{a_L} + \frac{1}{b_L} = \frac{1}{a_L} + \frac{1}{b + b_{L0}} \quad (2)$$

The value of b_{L0} can be estimated with a previous calibration, using the method described by Zeller et al. [8]. In the same calibration process also the focal length f_L of the main lens is estimated. If one can estimate the virtual image distance b , then it is possible to calculate the depth of the object point using eq. (2). This is described in the following section.

A. Depth Estimation

To calculate the object distance of the 3D point, the virtual image distance b has to be determined. This is done using the coordinate system displayed in Figure 6, which is aligned to the MLA.

The geometrical relations of a virtual image point and the corresponding points in two micro images which result from the same 3D are displayed in Figure 7. The depth estimation is based on the method described by Zeller et al. [9] by using the disparity of a point in two micro images.

The principal points of the micro lenses (eq. (3)), the projection of a 3D point in the micro images (eq. (4)) and in the virtual image (eq. (5)) respectively are described with their three-dimensional position vectors in the coordinate system of Figure 6.

$$\vec{c}_i = [c_{x,i} \quad c_{y,i} \quad 0]^T \quad (3)$$

$$\vec{x}_{R,i} = [x_{R,i} \quad y_{R,i} \quad B]^T \quad (4)$$

$$\vec{x}_V = [x_V \quad y_V \quad v]^T \quad (5)$$

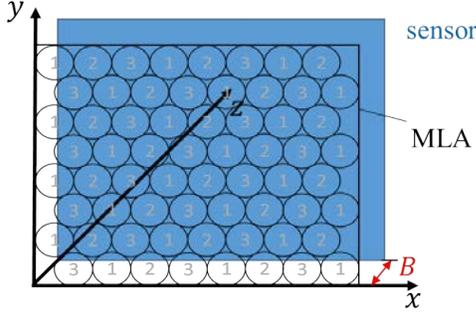


Fig. 6. Coordinate system for depth estimation

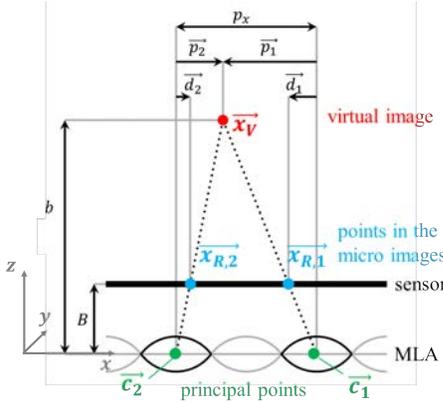


Fig. 7. Principle of the depth estimation (based on [9])

The intersection of the rays through the points in the micro images defines the virtual image in a distance b to the MLA. The triangles with a common vertex in the principal point of a micro lens and with their basis on the sensor (micro image) or in the virtual image respectively, are similar, leading to eq. (6).

$$\frac{\vec{d}_i}{B} = \frac{\vec{p}_i}{b} \quad (6)$$

The vectors \vec{d}_i (eq. (7)) define the distance between the principal points of a micro lens and the point in the corresponding micro image in x - and y -coordinates. The distance between a principal point and the virtual image in x - and y -coordinates is described with \vec{p}_i (s. eq. (8)). The x - and y -coordinates of both vectors are signed values and their sign is defined by the coordinate system in Figure 6.

$$\vec{d}_i = [d_{x,i} \quad d_{y,i}]^T \quad (7)$$

$$\vec{p}_i = [p_{x,i} \quad p_{y,i}]^T \quad (8)$$

With eq. (6) and the two-dimensional vectors, the distance b can be calculated using the x - or the y -coordinates. In the following calculations the x -coordinates are used.

The parallax p of the virtual image point, which defines the distance between the principal points of the two micro lenses, is described by eq. (9).

$$p = p_{x,2} - p_{x,1} \quad (9)$$

The disparity d is described as the difference between $d_{x,2}$ and $d_{x,1}$. With eq. (6) and (9) the definition for the disparity given in eq. (10) is received.

$$d = d_{x,2} - d_{x,1} = (p_{x,2} - p_{x,1}) \cdot \frac{B}{b} = p \cdot \frac{B}{b} \quad (10)$$

Equation (6) can be simplified using eq. (10) resulting in the distance b being defined as given in eq. (11).

$$b = \frac{p \cdot B}{d} \quad (11)$$

The disparity d and the parallax p are determined by the 3D point, respectively the distance between the micro images in which it is projected into. The fraction of the disparity with respect to the parallax is called the virtual depth v as given in eq. (12).

$$v = \frac{b}{B} \quad (12)$$

The virtual depth is proportional to the distance b between the MLA and the virtual image. This is needed to calculate the object distance a_L cf. eq. (2). The factor of proportionality is the inverse of b , i.e. the distance between MLA and sensor. This has to be estimated in a previous calibration step.

A point \vec{x}_V in the virtual image is defined as the intersection of the rays through the projected points in the micro images. Its position vector \vec{x}_V is given by eq. (13).

$$\vec{x}_V = (\vec{x}_{R,i} - \vec{c}_i) \cdot \frac{b}{B} + \vec{c}_i \quad (13)$$

It is straightforward to see that its x - and y -coordinates depend on the virtual depth v as indicated in the linear equations (14) and (15):

$$x_V = (x_{R,i} - c_{x,i}) \cdot v + c_{x,i} \quad (14)$$

$$y_V = (y_{R,i} - c_{y,i}) \cdot v + c_{y,i} \quad (15)$$

For two points which represent the same 3D point, the x - and y -coordinates calculated with eq. (14) and (15) have to be equal.

With this known model, the virtual image of a 3D point can be estimated if the 3D point is detected in more than two micro images. To calculate the object distance of the 3D point,

the relation between the virtual depth and the object distance is used (s. eq. (2) and (12)) and the following relation for the object distance $z_c = a_L$ holds:

$$z_c = \left(\frac{1}{f_L} - \frac{1}{v \cdot B + b_{L0}} \right)^{-1} \quad (16)$$

So to calculate the virtual depth v and the object distance z_c , points in the micro images have to be detected (s. section III) and matched (s. section IV).

III. POINT DETECTION IN MICRO IMAGES

To detect points in micro images, several detection methods like SURF, SIFT and Harris Corner Detector are applied directly to the raw image. They deliver a set of points of interest (POI), which will be used for estimation of the depth of the corresponding object points. As described in the previous section, the MLA is arranged hexagonally. By projecting the principal point of a micro lens orthogonally onto the sensor, the principal point in the corresponding micro image can be determined. With knowledge of the diameter of the micro lenses (measured in pixels on the sensor), the detected POI can be allocated to a certain micro image (s. Figure 8). To locate a point \vec{x}_R in the micro image, the vector as given in eq. (18) is calculated for each micro image with principal point \vec{c}_i . Because of the arrangement of the micro lenses in the MLA, a lens border (of 1.5 pixels) has to be defined which separates the micro images. The POI which do not fulfill eq. (19) for any micro image are not used in subsequent processing.

$$\vec{d}_{x,ci} = \vec{c}_i - \vec{x}_R \quad (17)$$

$$\left| \vec{d}_{x,ci} \right| \leq r_{lens} - l_{border} \quad (18)$$

With eq. (18) and (19) the POI can be allocated to the micro images. The allocated points are used in the following matching as described in section IV.

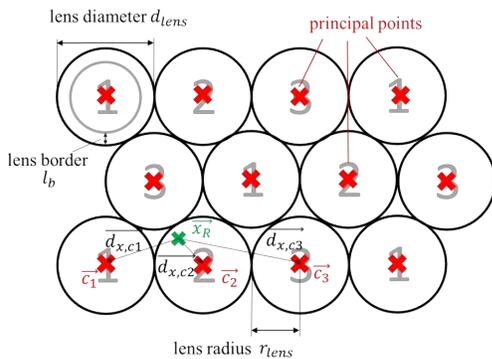


Fig. 8. Principle to determine the corresponding micro image for a detected POI

IV. POINT MATCHING

At first, only the adjacent micro images are used to find other points that match to the POI located in the reference micro image. The approach to match them in adjacent micro images is exemplified in Figure 9. The vectors of the detected points $\vec{x}_{R,i}$ and the principal points \vec{c}_i of the micro images in Figure 9 are two-dimensional. So only the x- and y-coordinates are used.

To match a POI to another POI in an adjacent micro image, the epipolar geometry can be used to restrict the search area. In Figure 10 the epipolar geometry between the virtual image and its representation in two micro images is displayed. Due to the plane-parallel arrangement of the microlenses, the epipolar lines will all be parallel to the line connecting the centers of the micro lenses (eq. (20)):

$$\vec{e}_{12} = \vec{c}_2 - \vec{c}_1 \quad (19)$$

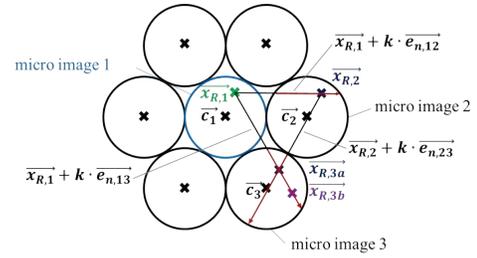


Fig. 9. Matching of the detected points using epipolar geometry

Starting from the detected POI $\vec{x}_{R,1}$, a corresponding POI would be located alongside the epipolar line. Due to manufacturing tolerances we allow a POI to be maximally one pixel away from the epipolar line to be matched to the reference POI. If several POIs are matched in the same micro image to a reference POI, those POI which do not belong to the same virtual image can be eliminated using already matched POIs from other micro images together with the epipolar geometry between those micro images (see Figure 9).

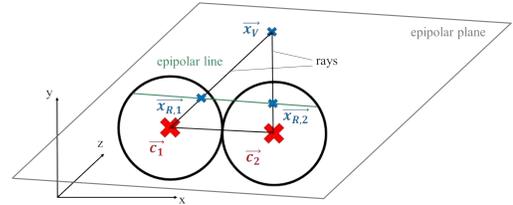


Fig. 10. Principle of the epipolar geometry with two micro images

After the matching is done in the adjacent micro images for every detected POI, the matched POIs are linked as indicated in Figure 11. This means that an uninterrupted chain of adjacent micro images is needed for POIs to be linked. However, since the micro lenses have different focal lengths and are arranged in the MLA as indicated in Figure 4, not only the directly adjacent micro lenses are considered for matching,

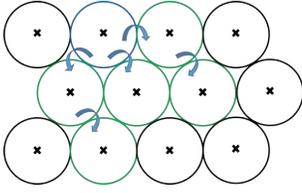


Fig. 11. Group of matched POIs with linkage

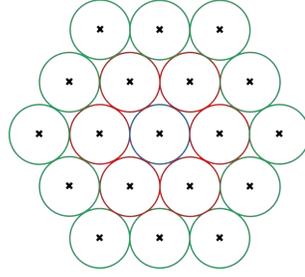


Fig. 12. Micro images used for direct POI matching

but also those with the same focal length which are just behind the directly adjacent ones (see Figure 12). This means that correspondences are searched in a total of 18 neighbouring micro images.

V. DEPTH ESTIMATION USING THE MATCHED POI

Each group of matched POI should represent the same virtual image point (resp. one 3D point) in the corresponding micro images. If the rays of the matched POI do not have a common intersection (s. Figure 13), the matched POI have to be optimized.

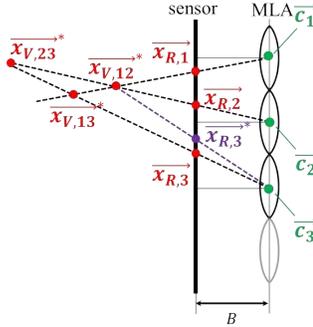


Fig. 13. Error estimation. Projecting a virtual image into the micro image

By choosing the intersection of the rays of $\vec{x}_{R,1}$ and $\vec{x}_{R,2}$, an error occurs for $\vec{x}_{R,3}$. The error for each POI can be described as given in eq. (20), where $\vec{x}_{R,i}^*$ is the virtual image point back-projected into the micro image.

$$\vec{\Delta}_i = [\Delta_{x,i} \quad \Delta_{y,i}]^T = \vec{x}_{R,i} - \vec{x}_{R,i}^* \quad (20)$$

The errors are minimized using the least square method. Because of the linear relation between the known values (POI coordinates, principal points) and the unknown values (coordinates of the virtual image point) (s. eq. (14) and (15)) the error minimization is performed using linear regression.

A. Linear Regression

For linear regression eq. (14) and (15) are transformed into the following functions:

$$f_{x,i} = x_{R,i} - c_{x,i} = \frac{x_V - c_{x,i}}{v} \quad (21)$$

$$f_{y,i} = y_{R,i} - c_{y,i} = \frac{y_V - c_{y,i}}{v} \quad (22)$$

According to this, the normalized coordinates of the virtual image point cf. eq. (23) are estimated using eq. (24), whereas the residual vector is defined with eq. (25) and the Jacobi matrix is given in eq. (26). For more details on linear regression we refer to [12]. Whereas N is the number of matched POI in one group, which can maximally be 18 (cf. Figure 12).

$$\vec{a} = \begin{bmatrix} x_V & y_V & 1 \\ v & v & v \end{bmatrix}^T \quad (23)$$

$$\vec{a}^* = \begin{bmatrix} x_V^* & y_V^* & 1 \\ v^* & v^* & v^* \end{bmatrix}^T = (J^T \cdot J)^{-1} \cdot J^T \cdot \vec{r} \quad (24)$$

$$\vec{r} = [x_{R,1} - c_{x,1} \quad y_{R,1} - c_{y,1} \quad \dots \quad x_{R,N} - c_{x,N} \quad y_{R,N} - c_{y,N}]^T \quad (25)$$

$$J = \begin{bmatrix} \frac{\partial f_{x,1}}{\partial \vec{a}} & \frac{\partial f_{y,1}}{\partial \vec{a}} & \dots & \frac{\partial f_{x,N}}{\partial \vec{a}} & \frac{\partial f_{y,N}}{\partial \vec{a}} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & -c_{x,1} \\ 0 & 1 & -c_{y,1} \\ \vdots & \vdots & \vdots \\ 1 & 0 & -c_{x,N} \\ 0 & 1 & -c_{y,N} \end{bmatrix} \quad (26)$$

B. Improvement of the calculated virtual images

To ensure a certain accuracy of the estimated virtual depth (resp. virtual image), the distances between the POIs in the micro images and the virtual image of the POI back-projected in these micro images are calculated as given in eq. (27) and (28). If any POI in a group of matched POIs has a distance error (s. eq. (39)) larger than one pixel, we repeat the linear regression but without the POI with the largest distance error. This is done until the remaining POIs do not have a distance error larger than one pixel or until only two POIs remain. In the last case the whole group of matched POIs is deleted.

$$\Delta x_{R,i} = x_{R,i}^* - x_{R,i} \quad (27)$$

$$\Delta y_{R,i} = y_{R,i}^* - y_{R,i} \quad (28)$$

$$\Delta_{R,i} = \sqrt{\Delta x_{R,i}^2 + \Delta y_{R,i}^2} \quad (29)$$

After this improvement step, the virtual depth v (resp. the object distance z_c) is calculated. Due to the configuration of our plenoptic camera, only a certain range for the virtual depth is reasonable. From this, a range for the object distance can be determined. In Figure 14 the function $z_c(v)$ is displayed for the camera parameters in Table 1. Only the estimated virtual images points (resp. 3D points) inside this range are classified as valid.

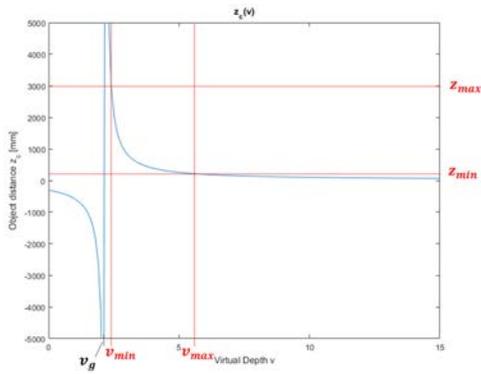

 Fig. 14. Function $z_c(v)$ with exemplary range for v and z_c

Table 1: Camera parameters

f_L	16.279748091856455 mm
b_{L0}	15.449618357330239 mm
B	0.38300659522738911 mm
d_{lens}	23.306472861260 pixels

VI. RESULTS

To evaluate the described methods, different POI detectors (SIFT, SURF and Harris Corner Detector (HCD)) are used from the openCV library. The estimated object distances for the image displayed in Figure 15 are compared to the values in the depth map generated with the method described in [13]. For the SIFT and SURF methods the default configuration of openCV is used and only the parameter for the amount of POIs is changed so that approximately 2000 POIs are detected. The Harris Corner Detector provides significantly less features for the captured scene (around 500 features).



Fig. 15. Raw image used for depth estimation

In the histograms displayed in Figure 16, Figure 17 and Figure 18 the absolute differences between the calculated object distance and the object distance from the depth map of [13] are displayed for the different detectors by their frequency. The HCD provides the best results regarding the absolute error of the object distances with maximum absolute errors around 50 cm.

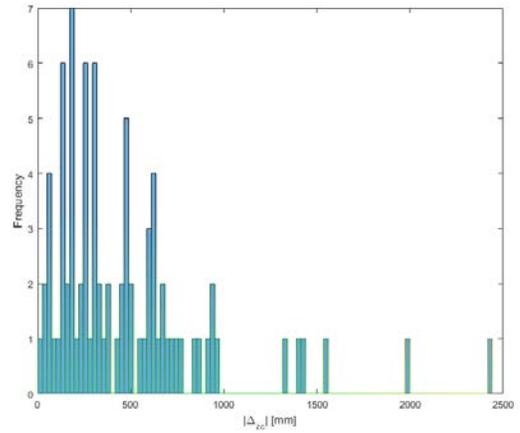


Fig. 16. Histogram of the absolute object distance errors (SIFT detector)

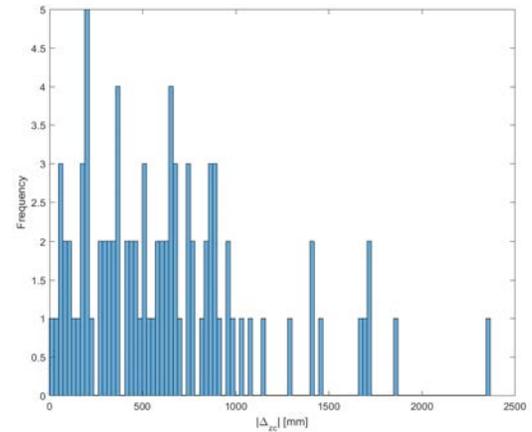


Fig. 17. Histogram of the absolute object distance errors (SURF detector)

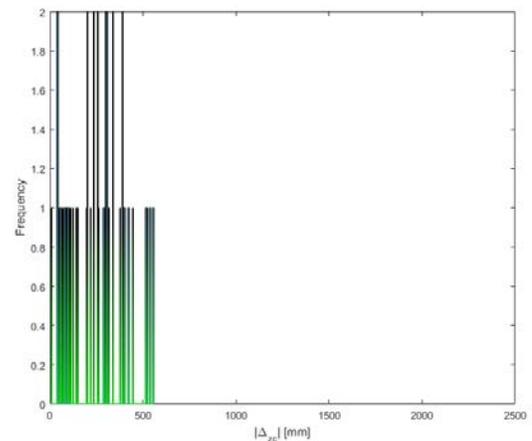


Fig. 18. Histogram of the absolute object distance errors (HCD)

VII. CONCLUSION

In general, the described approaches for feature matching and depth estimation provide to a certain extent acceptable results for common detectors. Although the HCD has the best results in comparison to SIFT and SURF, the number of corners which can be detected in micro images is too low to display a proper depth map.

By using only the detected POIs for matching and depth estimation, the accuracy of the estimation is limited. In further work, we want to improve the matching of POI by a correlation alongside the epipolar line to detect the corresponding points in the micro images. Furthermore, an edge detector specifically implemented for micro images could outperform SIFT and SURF, because in that case a set of edge points would be matched instead of single points.

A further application of the method described in this paper is to use it for finding matching POI in different raw images, recorded in a video. Then, this method can be used to implement simultaneous localization and mapping (SLAM) directly on raw images, without computing the totally focused image.

ACKNOWLEDGMENT

This research was done in the Bachelor Thesis of Jennifer Konz. It is part of the project Hypermod, which is funded by the Federal Ministry of Education and Research Germany in its program FHProfUnt. We gratefully acknowledge the support.

REFERENCES

- [1] F.E. Ives, "Parallax stereogram and process of making same", 1904.
- [2] G. Lippmann, "preuves rversibles donnant la sensation du relief", Journal de Physique Thorique et Applique, Vol. 7, No. 1, 1908.
- [3] E.H. Adelson, J.Y.A.Wang, "Single lens stereo with plenoptic camera", IEEE PAMI 14(2), pp. 99-106, 1992.
- [4] S.J. Gortler, R. Grzeszczuk, R. Szeliski, M.F. Cohen, "The lumigraph", Proc. 23rd SIGGRAPH, ACM, New York, pp. 43-54, 1996.
- [5] R. Ng, "Digital light field photography", PhD thesis, Stanford University, Stanford, USA, 2006.
- [6] A. Lunsdaine, T. Georgiev, "Full Resolution Lightfield Rendering", Adobe Technical Report, Adobe Systems Inc., 2008.
- [7] C. Perwass, L. Wietzke, "Single lens 3D-camera with extended depth-of-field", Human Vision and Electronic Imaging XVII, Burlingame, California, USA, 2012.
- [8] N. Zeller, C.A. Noury, F. Quint, C. Teulire, M. Dhome, "Metric calibration of a focused plenoptic camera based on a 3D calibration target", ISPRS Annals, Vol. III-3: 449-456, 2016.
- [9] N. Zeller, F. Quint, U. Stilla, "Calibration and accuracy analysis of a focused plenoptic camera", ISPRS Annals, Vol. II-3, pp. 205-212, 2014.
- [10] A. Lunsdaine, T. Georgiev, "The focused plenoptic camera", Proc. IEEE Int. Conf. on Computational Photography (ICCP), San Francisco, CA, USA, pp. 1-8, 2009.
- [11] Raytrix GmbH, "Raytrix Lightfield Camera Demo Images", 2013.
- [12] T. Strutz, *Data Fitting and Uncertainty: A practical introduction to weighted least squares and beyond*, 1st Edition, Vieweg+Teubner, 2011.
- [13] N. Zeller, F. Quint, U. Stilla, "Establishing a Probabilistic Depth Map from Focused Plenoptic Cameras", Proceedings, International Conference on 3D Vision (3DV), p. 91-99, Lyon, 2015.