Filtering Probabilistic Depth Maps Received from a Focused Plenoptic Camera

Niclas Zeller, Franz Quint
Faculty of Electrical Engineering and Information Technology
Karlsruhe University of Applied Sciences
76133 Karlsruhe, Germany
Email: niclas.zeller@hs-karlsruhe.de,
franz.quint@hs-karlsruhe.de

Uwe Stilla
Department of Photogrammetry and Remote Sensing
Technische Universität München
80290 Munich, Germany
Email: stilla@tum.de

Abstract—This paper presents a filtering approach for semi-dense probabilistic depth map received from a focused plenoptic camera. In the probabilistic depth map each valid depth pixel contains, beside the depth value itself, a variance which gives a measure for the certainty of the estimated depth. This variance is used in a weighted filtering approach. Here, beside removing outliers and filling holes in the semi-dense depth map, pixel neighborhoods are modeled in a Markov-Random-Field (MRF). The presented approach is working in two steps, firstly a rough regularization is performed in each micro image separately and secondly, after projection to the virtual image space, another more precise regularization is performed. The presented filtering approach considers properties of the plenoptic imaging, like varying spatial resolution over object distance. Besides, it preserves discontinuities in the depth map. The algorithm aims for low complexity and due to its nature it can easily be implemented in parallel.

I. Introduction

While a monocular camera captures light intensities only on a 2D sensor, a plenoptic camera gathers intensities in a sampled 4D light-field representation.

Even though this concept to gather 4D light-field information was invented already more than a century ago [1], [2], it took until the last decade to put the first plenoptic cameras on the market. One reason therefor is the high computational cost which has to be spend to process the recorded light-field information. Nevertheless, today there are several algorithms available to process the recorded 4D light-field information in, or at least close to, real-time.

Processing tasks last from subsequent refocusing, over super resolution imaging, up to depth estimation. Especially plenoptic camera based depth estimation and the fact, that a plenoptic camera records much more information about scene structures than a monocular camera arouse attention in the computer vision community. Here, possible applications are for instance Visual Odometry (VO) or Simultaneous Localization and Mapping (SLAM) but also gesture and pattern recognition.

This paper continues the work of our prior publication [3] where a probabilistic depth estimation approach for a focused plenoptic camera was developed. The method in [3] establishes a semi-dense depth map, which means that depth is only estimated for regions of high contrast. Here we present a post
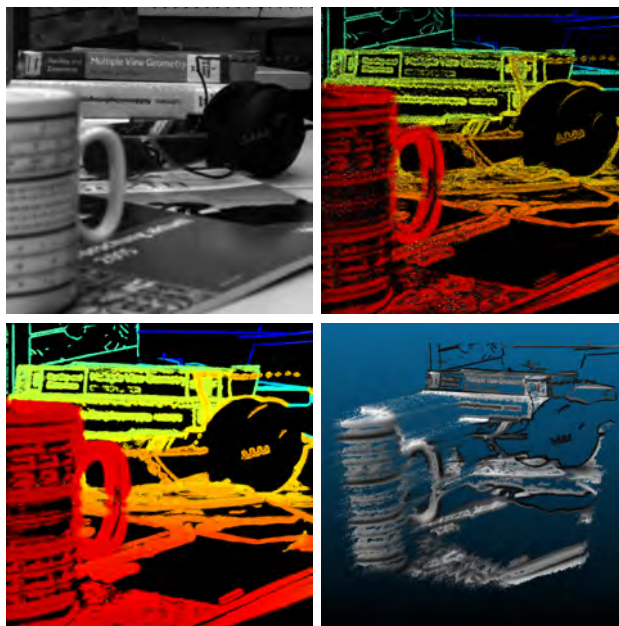

Fig. 1. Sample scene recorded by a focused plenoptic camera. Top left: Synthesized totally focused intensity image. Top right: Color coded unfiltered depth map. Bottom left: Color coded filtered depth map. Bottom right: Point cloud representation of the filtered depth map.

processing step to filter out outliers and stochastic noise to improve the quality of the depth map.

Due to the fact that we want to use the resulting depth information in Visual Odometry which is supposed to operate close to real-time, we were focusing on methods which have low complexity and can be implemented in an efficient manner.

In this article, we first briefly describe the concept of a focused plenoptic camera (Section II) which is requisite to understand our depth estimation approach. This approach is presented succinctly in Section III. Our filtering method based on the probabilistic depth map is introduced in Section IV. Section V shows how based on the estimated depth an intensity image with a very high depth of field (DOF) (totally focused image) can be synthesized. Section VI presents the evaluation of the filtering approach and Section VII concludes our work.
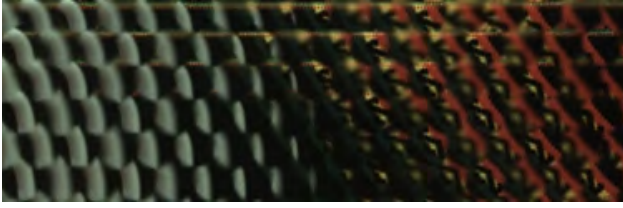
Fig. 2. Subsection of the raw image recorded by a focused plenoptic camera. The raw image consists of thousands of circular micro image arranged on a hexagonal grid.

II. CONCEPT OF THE FOCUSED PLENOPTIC CAMERA

A focused plenoptic camera (plenoptic camera 2.0) [4], [5], which the presented method relies on, slightly differs from a traditional, unfocused plenoptic camera [6], [7]. Nevertheless, both concepts are based on a micro lens array (MLA) which is placed in front of the sensor.

The main advantage of a focused plenoptic camera is that it produces a focused micro image of a subsection of the scene under each micro lens, as can be seen in the recorded sensor image in Figure 2. Thereby a higher spatial image resolution is achieve than with an unfocused plenoptic camera, where each micro lens just represents one spatial sample.

A focused plenoptic camera can be realized in two different setups [8], [4], the Keplerian and the Galilean mode. In the Keplerian mode the light-field sensor (image sensor and MLA) is placed in a distance further than the image distance b_L to the main lens of the camera, while in the Galilean mode the light-field sensor is placed closer than b_L to the main lens.

Figure 3 shows the interior of a focused plenoptic camera based on the Galilean setup. Here D_L defines the aperture of the main lens and D_M the one of a micro lens. b_L represents the image distance of the projected main lens image (virtual image). The relation between the image distance b_L and the object distance a_L is defined, dependent on the focal length f_L, by the thin lens equation:

$$\frac{1}{f_L} = \frac{1}{a_L} + \frac{1}{b_L} \tag{1}$$

In the following we will discuss only the Galilean mode. Nevertheless, similar assumptions can be made for the Keplerian mode.

In our research we are working with a Raytrix camera [5], which is a focused plenoptic camera operating in the Galilean mode. Besides, a Raytrix camera has as distinct feature an MLA that consist of three different types of micro lenses which are arranged in a hexagonal grid. Each type of micro lens has a different focal length and thus focuses a different image distance b_L and respectively object distance a_L on the sensor. The three focal lengths are chosen such, that the corresponding DOFs are just adjacent. Thus, the DOF of the camera is extended. Due to overlap of the micro images it can be assured, that each virtual image point occurs focused in at least one micro image.

From Figure 3 one can see, that a certain object point is projected to multiple micro images on the sensor. Thus, by
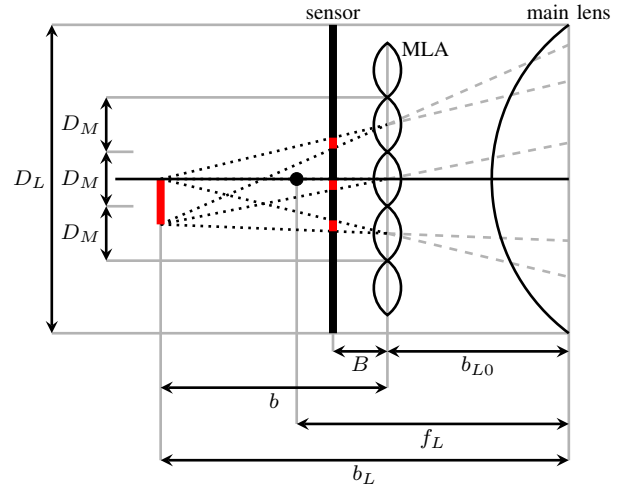

Fig. 3. Optical path inside a focused plenoptic camera based on the Galilean configuration. The MLA and the image sensor lie in front of the virtual image created by the main lens. A virtual image point in distance b behind the MLA results in multiple focused micro images on the sensor.

finding corresponding points in the micro images, the distance b between MLA and the respective virtual image point can be calculated by triangulation:

$$b = \frac{d \cdot B}{p_x} \tag{2}$$

Here, d represents the baseline distance between two micro lenses, B the distance between MLA and sensor and p_x the measured disparity between the corresponding points in the micro images. Since points with a large distance b are seen from micro lenses which are further apart, the lager baseline distance d can be used to improve the depth estimate. A detailed derivation of this equation can be found in [9]. Prior to a metric calibration, the distance B is not known precisely and thus the depth estimate is defined relative to this distance:

$$v = \frac{b}{B} = \frac{d}{p_x} \tag{3}$$

The distance v is called the virtual depth subsequently [5]. Besides, we defined the inverse virtual depth $z = v^{-1}$, which will be used in the following sections:

$$z = \frac{1}{v} = \frac{p_x}{d} \tag{4}$$

III. PROBABILISTIC DEPTH ESTIMATION

This section presents our probabilistic virtual depth estimation algorithm, originally published in [3]. The prior section showed that depth estimation in a focused plenoptic camera can be solved by finding pixel correspondences between micro images. Due to the high overlap of the micro images, this task can be considered as multi-view stereo problem. Since the focus of our approach lies on real-time applicability, the presented algorithm searches for corresponding pixel pairs based on local criteria. To combine all these correspondences

in one depth map, for each depth observation an uncertainty measure is defined which is used to establish a probabilistic depth map, as introduced in [10].

A. Probabilistic Virtual Depth

From eq. (4) one can see, that the inverse virtual depth z is proportional to the estimated disparity p_x. The disparity p_x of a point in two micro images is estimated based on intensity error minimization along the epipolar line. Thus, the sensor noise is considered to be the main error source effecting this estimate. Since the sensor noise is usually modeled as additive white Gaussian noise (AWGN), we also consider the estimated disparity p_x and hence also the inverse virtual depth to be a Gaussian distributed random variable Z, defined by the mean z and the variance σ_z^2.

$$f_Z(x) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{(x-z)^2}{2\sigma_z^2}} \qquad (5)$$

In [3] it is derived in detail how the inverse virtual depth variance σ_z^2 can be obtained.

B. Virtual Depth Observation

For depth estimation we want to establish stereo correspondences between any possible pairs of micro lenses. Thus, a graph of baselines is build which defines the micro image pairs used for stereo matching. For a certain micro lens this graph contains the baselines to all micro lenses right to the micro lens of interest. In that way it is assured that each pixel correspondence is established only once. Figure 4 shows five sample baselines, one for each of the five shortest baseline distances.

Each baseline in the graph is defined by its length d as well as its 2D orientation on the MLA plane $\boldsymbol{e}_p = (e_{px}, e_{py})^T$. Since the micro images are considered to be rectified, the orientation vector of the baseline is equivalent to that of the epipolar line of any pixel under the micro lens of interest.

The inverse virtual depth estimation is performed for each pixel in the raw image $p_R := \boldsymbol{x}_R = (x_R, y_R)^T$ which lies under a micro lens and has a certain intensity gradient along the epipolar line $\boldsymbol{e}_p$. For each pixels multiple stereo observations are estimated starting from the shortest baseline up to the longest possible baseline. This is done, since for a short baseline it is more likely to find unambiguous correspondences, while longer baselines improve the precision. Thus, for each observation the prior estimates are used to limit the search range.

C. Merging Inverse Virtual Depth Observations

The prior Section described already that for a pixel p_R in the raw image multiple inverse virtual depth observations can be obtained. Each new observation is incorporated into a inverse virtual depth hypothesis similar to the update step in a Kalman filter. Here, the new inverse virtual depth hypothesis $\mathcal{N}(z, \sigma_z^2)$
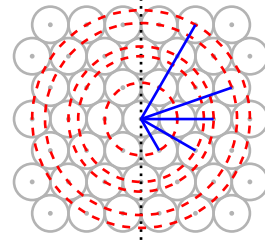

Fig. 4. Five shortest baseline distances in a hexagonal micro lens grid. For a micro lens stereo matching is only performed with neighbors for which the baseline angle ϕ is in the range $-90° \leq \phi < 90°$.

results from the hypothesis prior the observation $\mathcal{N}(z_p, \sigma_p^2)$ and the new observation $\mathcal{N}(z_o, \sigma_o^2)$ as follows:

$$\mathcal{N}(z, \sigma_z^2) = \mathcal{N}\left(\frac{\sigma_p^2 \cdot z_o + \sigma_o^2 \cdot z_p}{\sigma_p^2 + \sigma_o^2}, \frac{\sigma_p^2 \cdot \sigma_o^2}{\sigma_p^2 + \sigma_o^2} \right) \qquad (6)$$

IV. POST PROCESSING OF PROBABILISTIC DEPTH MAP

In the probabilistic depth estimation algorithm pixels correspondences are found only based on local criteria. Besides, each pixel is processed separately without considering a larger neighborhood. Thus, the estimated inverse virtual depth values z in the micro images can be considered to be more or less uncorrelated. This is actually not the case for depth maps of real scenes, where neighboring depth pixels usually are highly correlated. Thus, in this post processing step we model the connection between pixels in a certain neighborhood by a Markov-Random-Field (MRF).

The regularization is done in several steps. First outlier removal and hole filling is performed in each micro image separately (Sec. IV-A). Afterwards the pixels from the micro images are projected back into the virtual image space which is created by the main lens projection (Sec. IV-B). Here again outlier removal and hole filling is performed. Finally the depth map is updated based on an MRF structure, using all inverse virtual depth pixel hypothesizes within a certain neighborhood (Sec. IV-C).

A. Removing Outliers and Filling Holes in Micro Images

1) Removing Outliers: Due to ambiguous structures in the micro images it happens that wrong correspondences are established between pixels in the micro images. Thus, in a first step pixels which are outliers with respect to their neighborhood are removed. For each valid depth pixel in the raw image $p_R^{(i)}$, with the depth hypothesis $\mathcal{N}(z_i, \sigma_{zi}^2)$, an average inverse virtual depth $\overline{z_i}$ and a corresponding variance $\overline{\sigma_{zi}^2}$ of the valid depth pixels N_{Rvalid} within the neighborhood $N_R^{(i)}$ is defined:

$$\overline{z_i} = \frac{\sum_{k \in N_x^{(i)}, k \neq i} z_k \cdot \left(\sigma_{zk}^2\right)^{-1}}{\sum_{k \in N_x^{(i)}, k \neq i} \left(\sigma_{zk}^2\right)^{-1}} \qquad (7)$$

$$\overline{\sigma_{zi}^2} = \frac{|N_x^{(i)}| - 1}{\sum_{k \in N_x^{(i)}, k \neq i} \left(\sigma_{zk}^2\right)^{-1}} \qquad (8)$$

In eq. (8) $N_x^{(i)}$ defines the intersection between the set of neighborhood pixels $N_R^{(i)}$ of $p_R^{(i)}$ and the set of valid depth pixels N_{Rvalid} ($N_x^{(i)} = N_R^{(i)} \cap N_{Rvalid}$). Besides $|N_x^{(i)}|$ is the cardinality of the set $N_x^{(i)}$ and defines the number of elements in the set. Eq. (7) and (8) are actually quite similar to the definition of z and σ_z^2 in eq. (6). The only difference is, that the sum of the inverse variances is multiplied by the number of valid neighbors ($|N_x^{(i)}| - 1$).

Each pixel $p_R^{(i)}$ that has a inverse virtual depth estimate z_i which does not satisfy the following condition is classified as outlier:

$$(z_i - \overline{z_i})^2 \leq 4 \cdot \overline{\sigma_{zi}^2} \tag{9}$$

For the experiments in Section VI we defined a squared neighborhood of 5 pixel × 5 pixel.

2) Filling Holes: After removing the outlier in the micro images, pixels which have enough gradient but no valid depth estimate are filled based on the neighboring pixels. Therefore, again the average inverse virtual depth $\overline{z_i}$ within a neighborhood region is calculated based on eq. (7). The inverse virtual depth $\overline{z_i}$ gives the new depth value for the invalid pixel $p_R^{(i)}$, while the corresponding variance σ_{zi}^2 is initialized to some predefined high value. By setting the initial variance to some high value, these interpolated depth values can not negatively effect the later regularization.

B. Projecting Micro Images into Virtual Image Space

After removing outliers and filling holes in the micro images, all micro image pixels which have a valid depth hypothesis are projected into the virtual image space. We define the virtual image space similar to [3] by the pixel coordinates x_V and y_V and the virtual depth $v = z^{-1}$ ($\boldsymbol{x}_V = (x_V, y_V, v)^T$).

The virtual image coordinates x_V and y_V are calculated based on a central perspective projecting thru the corresponding micro lens as follows:

$$x_V = (x_R - c_x)v + c_x \tag{10}$$
$$y_V = (y_R - c_y)v + c_y \tag{11}$$

Here c_x and c_y is the center of the micro lens under which the pixel p_R with the coordinates $\boldsymbol{x}_R$ lies.

As described in Section II, a virtual image point is projected to multiple micro images. Vise verse, when performing the back projection to the virtual image space, multiple raw image pixels can fall on the same virtual image pixel and thus have to be merged. This is done similar to the merging step in Section III-C (eq. (6)).

C. Regularization of Virtual Image

In the virtual image space firstly outliers which were not detected in the micro images are removed. Besides, the inverse virtual depth values are smoothed, while considering the imaging concept of the camera.

For the virtual image space regularization again for each pixel $p_V^{(i)} := \boldsymbol{x}_V^{(i)}$ a neighborhood region $N_V^{(i)}$ is defined.

Each micro lens performs a central perspective projection and thus, objects with a high virtual depth occur smaller on the sensor than objects with a small virtual depth. Besides, virtual image points with a high virtual depth are observed by more micro lenses. Thus, vise verse back projected virtual image regions with a high virtual depth consist of more points which are spread over a larger region. Hence, for the virtual image regularization, the size of the neighborhood is defined dependent on the virtual depth v_i of the pixel of interest $p_V^{(i)}$. For each pixel $p_V^{(i)}$ a radius $r(v_i)$ is defined as follows:

$$r(v) = \lceil n \cdot v_i \rceil \tag{12}$$

Here n defines some constant parameter. For lower complexity in calculation, the radius $r(v_i)$ defines the maximum allowed Chebyshev distance L_∞ to the pixel $p_V^{(i)}$ instead of the Euclidean distance. A pixel $p_V^{(k)}$ for instance has the Chebyshev distance $L_\infty^{(k)}$ to the pixel $p_V^{(i)}$ defined as follows:

$$L_\infty^{(k)} = \max\left(|x_V^{(i)} - x_V^{(k)}|, |y_V^{(i)} - y_V^{(k)}|\right) \tag{13}$$

Thus, a squared neighborhood $N_V^{(i)}$ around $p_V^{(i)}$ is defined.

1) Removing Outliers: For removing outliers in the virtual image, similar to the micro images, the mean inverse virtual depth $\overline{z_i}$ (eq. (7)) and the mean inverse virtual depth variance $\overline{\sigma_{zi}^2}$ (eq. (8)) of valid depth pixels within the neighborhood region $N_V^{(i)}$ is defined:

$$N_x^{(i)} = N_V^{(i)} \cap N_{Vvalid} \tag{14}$$

Here, N_{Vvalid} is the set of all pixels p_V which have a valid depth estimate. Nevertheless, besides fulfilling the condition defined in eq. (9) a pixel $p_V^{(i)}$ has to have a density of valid depth pixels in its neighborhood above a certain threshold T_D.

$$T_D \leq \frac{|N_x^{(i)}|}{|N_V^{(i)}|} \tag{15}$$

In the following experiments the minimum density was set to $T_D = 0.25$.

2) Filling Holes: At that point there are no intensities available for the virtual image. Thus, pixel validity can not be defined based on the pixel's intensity gradient, as it is done in the raw image. Instead, each pixel in the virtual image p_V which has at least one direct neighbor with a valid depth hypothesis is filled by a weighted average of neighboring pixels, defined similar to eq. (7).

3) MRF based Noise Reduction: In a final step the actual correspondence between neighboring pixels in the virtual image is established by modeling the neighborhoods in an MRF. Here the same neighborhood $N_V^{(i)}$ as for outlier removal is defined.

In the MRF for a neighborhood $N_V^{(i)}$ two different states are defined to handle discontinuities in the depth map. Thus, either a pixel $p_V^{(k)}$ in the neighborhood $N_V^{(i)}$ belongs to the same object as $p_V^{(i)}$ or to a different object. These two objects can be considered as foreground and background.

Besides, a weighting function $w(d)$ is defined which models the lower correlation between further apart pixels. Here d defines the euclidean distance between two pixels. In our approach we use a Gaussian curve as weighting function, where the standard deviation σ_w is proportional to v.

$$w(d) = e^{-\frac{d^2}{2\sigma_w^2}} \qquad (16)$$

To model the two different states (foreground and background), the neighborhood is divided in two different sets of valid depth pixels, N_{sim} and N_{diff}. Here N_{sim} are all valid pixels which have a depth value similar to $p_V^{(i)}$ (including $p_V^{(i)}$) and N_{diff} contains all valid pixels with a depth value different to $p_V^{(i)}$. The updated depth value of the pixel $p_V^{(i)}$ is then calculated based on the larger one of the two sets as follows:

$$z_i = \frac{\sum_{k \in N_x} z_k \cdot \frac{w_k}{\sigma_{zk}^2}}{\sum_{k \in N_x} \frac{w_k}{\sigma_{zk}^2}} \qquad (17)$$

$$\sigma_{zi}^2 = \frac{\sum_{k \in N_x} w_k}{\sum_{k \in N_x} \frac{w_k}{\sigma_{zk}^2}} \qquad (18)$$

Here, N_x is the one of N_{sim} and N_{diff} with the higher cardinality.

$$N_x = \begin{cases} N_{sim} & \text{if } |N_{diff}| < |N_{sim}|, \\ N_{diff} & \text{else.} \end{cases} \qquad (19)$$

w_k is defined as follows:

$$w_k = w(d_k) = w\left(\|(x_V^{(i)}, y_V^{(i)})^T - (x_V^{(k)}, y_V^{(k)})^T\|\right) \qquad (20)$$

V. Intensity Image Synthesis

Additionally to the virtual depth map an intensity image is synthesized. This is done as described in [5]. For each point $p_V^{(i)}$, based on its inverse virtual depth estimate z_i a radius R_i is defined.

$$R_i = \frac{D_M}{2 \cdot z_i} \qquad (21)$$

Here, D_M is the diameter of a micro lens as shown in Figure 3. The radius R_i defines the region around the pixel $p_V^{(i)}$ in which micro lenses still see this virtual image point. Hence, the point $p_V^{(i)}$ is back projected to all micro images for which the center lies within this radius around the pixel. Here only the micro lens types are considered in which the virtual image point is in focus. Based on the corresponding interpolated intensity values a weighted mean is calculated.

$$I(p_i) = \frac{\sum_k I_k \cdot h_k}{\sum_k h_k} \qquad (22)$$

Here h_k are the respective intensity values received from a recorded white image. By calculating a weighted mean using the information of the white image for each pixel the optimum signal-to-noise ratio (SNR) is obtained.

For pixels p_V in the virtual image which have no depth estimate it is satisfactory to assume an approximate depth value to synthesize its intensity. This can be done since such
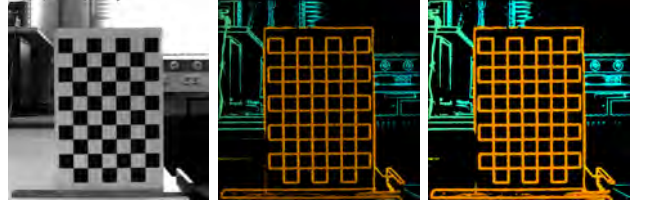

Fig. 5. Planar chessboard target used for quantitative evaluation. Left: Totally focused intensity image. Center: Color coded depth map before filtering. Right: Color coded depth map after filtering.
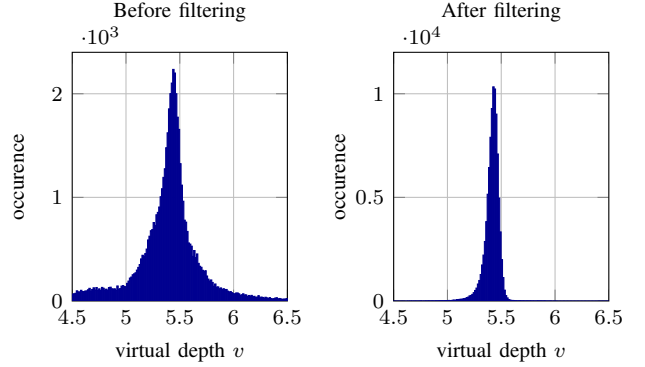

Fig. 6. Histogram of virtual depth v across chessboard target.

pixels anyways lie in a neighborhood of little texture and thus all pixels in the neighborhood have similar intensities.

VI. Results

In this section we evaluate the presented filtering method based on artificial as well as real scenes. In a first evaluation we try to measure the filtering results in numbers, as presented in Section VI-A, while Section VI-B presents a qualitative evaluation.

A. Quantitative Evaluation

For a quantitative evaluation we used a planar target as ground truth which was placed in front of the plenoptic camera, as shown in Figure 5. Here the virtual depth values across the planar target were evaluated and compared before and after filtering. Figure 6 shows the histograms of the virtual depth v before and after filtering.

Besides, Table I shows some calculated statistics for the virtual depth v across the chessboard target (mean $\overline{v}$, median $\tilde{v}$, standard deviation σ_v).

As one can see, the standard deviation σ_v is highly reduce by the filtering. Furthermore, the arithmetic mean $\overline{v}$ is effected quite a lot by the filtering. This is because the virtual depth is

	mean $\overline{v}$	median $\tilde{v}$	std. dev. σ_v
Before filtering	5.667	5.430	1.304
After filtering	5.413	5.423	0.071

TABLE I
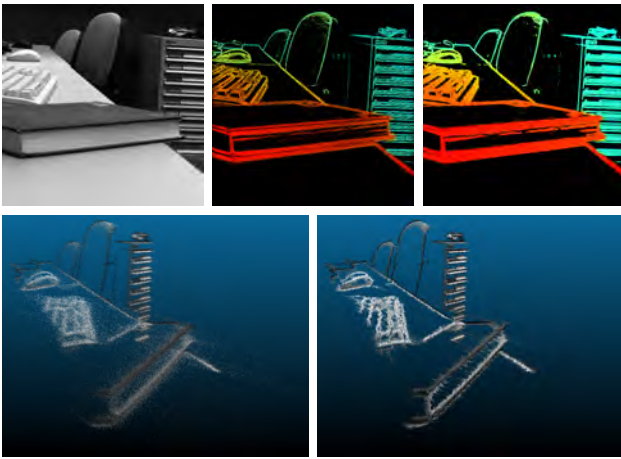STATISTICS CALCULATED FOR VIRTUAL DEPTH ACROSS CHESSBOARD PLANE.

11

Fig. 7. Real scene recorded by a focused plenoptic camera. Top left: Synthesized totally focused intensity image. Top center: Color coded unfiltered depth map. Top right: Color coded filtered depth map. Bottom left: Point cloud representation of the unfiltered depth map. Bottom right: Point cloud representation of the filtered depth map.
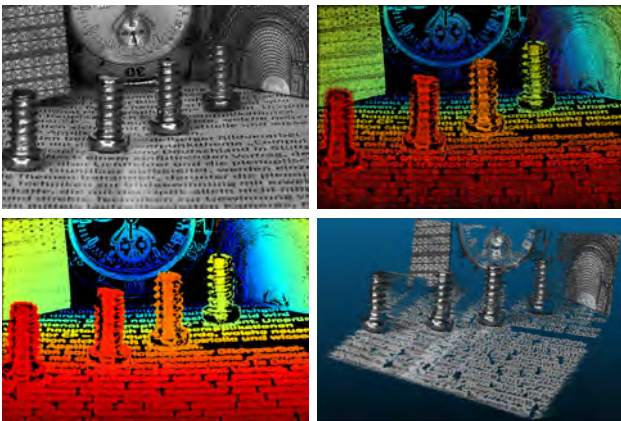

Fig. 8. Raytrix sample scene [11]. Top left: Synthesized totally focused intensity image. Top right: Color coded unfiltered depth map. Bottom left: Color coded filtered depth map. Bottom right: Point cloud representation of the filtered depth map.

not symmetrically distributed but the inverse virtual depth. This mapping from z to v has the effect, that the mean is shifted away from the distributions maximum towards higher virtual depth values. Nevertheless, the median $\tilde{v}$ stays more or less constant.

B. Qualitative Evaluation

In this part of the evaluation some real scenes are presented for which the depth was estimated and filtering was applied. Figure 1 and Figure 7 show two real scenes which were recorded with a Raytrix R5 camera. In Figure 8 a sample scene from Raytrix [11] is shown.

The color coded depth maps show quite well how holes in the depth map are filled and how pixel in a neighborhood of low density are removed. From Figure 7 one can see that this sometimes also removes some valid structures form far away regions (e.g. at the tool cabinet in the back). This could be

avoided by adjusting the parameters. Nevertheless, this also might cause some unwanted outliers to be kept.

The bottom row in Figure 7 shows a comparison of the point clouds before and after filtering. These point clouds nicely indicate how the noise is reduced but discontinuities are kept.

VII. CONCLUSION

In this paper we presented a method to filter semi-dense probabilistic virtual depth maps which were estimated based on a focused plenoptic camera. Besides removing outliers and filling holes, the approach uses an MRF to smooth the depth map, while preserving discontinuities.

The presented results show high improvement in comparison to the unfiltered depth information. Noise is highly reduced while most of the details are kept.

The presented method was developed in such manner, that it later can be implemented as real-time post processing for instance on a graphic processor unit (GPU). The properties of the resulting depth map are such that it supplies reliable depth information which later can be used in a plenoptic camera based Visual Odometry.

In future one could think of including intensity information for filtering the virtual image space, similar to the micro images. Nevertheless, since the correct intensity of a pixel relies on a correct virtual depth estimate, the problem becomes more complex.

REFERENCES

[1] F. E. Ives, "Parallax stereogram and process of making same," USA Patent US725 567, 04 14, 1903.

[2] G. Lippmann, "Epreuves reversibles. photographies integrales," *Comptes Rendus De l'Academie Des Sciences De Paris*, vol. 146, pp. 446–451, 1908.

[3] N. Zeller, F. Quint, and U. Stilla, "Establishing a probabilistic depth map from focused plenoptic cameras," in *Proc. International Conference on 3D Vision (3DV)*, 2015, pp. 91–99.

[4] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proc. IEEE International Conference on Computational Photography (ICCP)*, San Francisco, CA, April 2009, pp. 1–8.

[5] C. Perwaß and L. Wietzke, "Single lens 3d-camera with extended depth-of-field," in *Proc. SPIE 8291, Human Vision and Electronic Imaging XVII*, Burlingame, California, USA, January 2012.

[6] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, February 1992.

[7] R. Ng, M. Levoy, M. Brédif, G. Guval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford University, Computer Sciences, CSTR, Tech. Rep., 05 2005.

[8] A. Lumsdaine and T. Georgiev, "Full resolution lightfield rendering," Adobe Systems, Inc., Tech. Rep., 2008.

[9] N. Zeller, F. Quint, and U. Stilla, "Calibration and accuracy analysis of a focused plenoptic camera," *ISPRS Annals of Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. II-3, pp. 205–212, 09 2014.

[10] J. Engel, J. Sturm, and D. Cremers, "Semi-dense visual odometry for a monocular camera," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, Dec 2013, pp. 1449–1456.

[11] Raytrix GmbH, last accessed: July 30, 2015. [Online]. Available: http://www.raytrix.de/