Shape from Refocus

R. Huber-Mörk, S. Štolc, D. Soukup, and B. Holländer

Safety & Security Department, AIT Austrian Institute of Technology GmbH, Austria

Abstract. We present a method exploiting computational refocusing capabilities of a light-field camera in order to obtain 3D shape information. We consider a light-field constructed from the relative motion between a camera and observed objects, i.e. points on the object surface are imaged under different angles along the direction of the motion trajectory. Computationally refocused images are handled by a shape-from-focus algorithm. A linear sharpness measure is shown to be computational advantageous as computational refocusing to a specific depth and sharpness assessment of each refocused image can be reordered. We also present a view matching method which further stabilizes the suggested procedure when fused with sharpness assessment. Results for real-world objects from an inspection task are presented. Comparison to ground-truth data showed average depth errors on the order of magnitude of 1 mm for a depth range of 1 cm.

1 Introduction

The goal of *Shape from X* (SfX) methods is to extract 3D information from intensity or color images. In SfX methods it is not necessary to establish visual correspondence between images, when compared to other image based methods for 3D information extraction. This is of great computational advantage and overcomes the correspondence problem, although not entirely [1]. *Shape from Focus* (SfF) [2] and *Shape from Defocus* (SfD) [3] usually make use of at least two images taken by a single camera focused to different distances. SfF and SfD typically require active camera systems, i.e. variation of the focal lengths and/or aperture settings under computer control. In SfD, the depth is usually estimated from two observations with different focal or aperture settings. In SfF one varies the focal or aperture settings and depth estimation is obtained from comparison of sharpness measures.

One property of *light-field* data [4] is the possibility of focusing after a scene was acquired, a capability called *refocusing*. Computational refocusing naturally enables the use of SfD and SfF approaches for range sensing. Closely related work includes the following papers: Tao et.al. [5] demonstrated the use of focus cues as a complement to *epipolar plane image* (EPI) analysis of light-field data obtained by a plenoptic camera. Vaish et.al. [6] compare stereo and SfF, as well as robust measures for depth estimation based on light-field data obtained by a camera array spanning a so-called *synthetic aperture*.

This paper is organized as follows. We review related work in Sec. 2. The suggested *Shape from Refocus* (SfR) approach is introduced in Sec. 3 and more

G. Bebis et al. (Eds.): ISVC 2014, Part II, LNCS 8888, pp. 153–162, 2014.

details on this method are given in Sec. 4. Experimental results are provided in Sec. 5 and, finally, conclusions are drawn in Sec. 6.

2 Related Work

The 4-D radiance function of 2D position and 2D direction in regions of space free from occluders is commonly termed light-field [7]. Practical light-field acquisition can be performed in various ways, e.g. by a multi-camera array [8], a gantry system [4] or a plenoptic camera [9]. Computational refocusing and depth estimation are popular capabilities of light-field cameras. Depth estimation is related to the detection of linear structures in an EPI and is practically performed using methods like slope hypothesis testing [10] or structure tensor analysis [11]. The *Depth of Field* (DoF) for plenoptic cameras was investigated by Georgiev and Lumsdaine [12] for the so called *Plenoptic 2.0* camera design. Perwaß and Wietzke [13] discussed the DoF of their multi-focus plenoptic camera design in detail.

Defocusing properties of images were utilized by Pentland [3] and two approaches for SfD were suggested. The first one makes use of a sharpness measure applied to regions of steep discontinuities extracted from a single image. Depth estimation with ambiguity was obtained, i.e. points with similar blurring behind and in front of the exact focus could not be discriminated. The second approach uses two images of the same scene with different aperture setting. The so called *spectral ratio*, i.e. the ratio of local high-frequency content in the images, is then used to infer the distance to the exact focus. A bifocal sensor encorporated into a real-time system for SfD was described by Nayar et.al. [14], where, among other ideas, an external aperture was suggested in order to ensure equal magnification for the different focal planes.

SfF was discussed by Krotkov and Martin [2], where a theory of defocus was laid out and a number of focus criteria and a search strategy for automatic focusing to a point in an image were given. Nayar and Nakagawa [15] describe a focus measure operator and an automated SfF system including interpolation of depth estimates. Recently, the method of focus variation was presented as robust technology for high resolution surface metrology [16]. An extensive evaluation of focus measures used in SfF approaches was published [17].

3 Suggested Approach - Shape from Refocus

The suggested *Shape from Refocus* (SfR) is essentially a SfF technique exploiting light-field refocusing capabilities. We discuss the issues of DoF, depth resolution, refocusing and sharpness assessment.

3.1 Light Field Acquisition

In the approach by Štolc et.al. [18], a fast CMOS area-scan sensor is used to capture object positions (u, v) under varying angle s along the transport direction

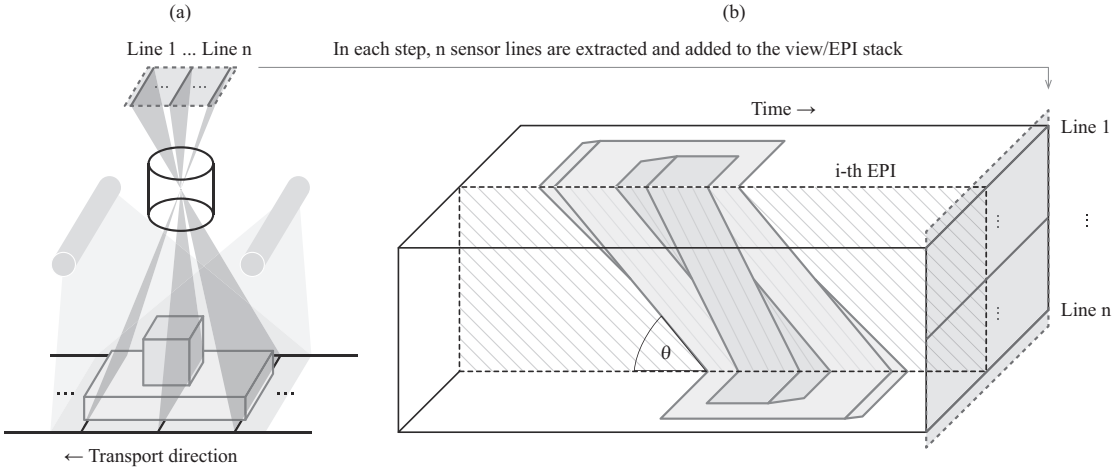

Fig. 1. Multi-line-scan acquisition of a simple 3-D object: (a) a blue cube standing on top of a green cuboid, (b) in each time step, the lines extracted from the sensor are inserted vertically into the EPI stack on the right

over time. As there is no variation of the angle t measured across the transport direction, a 3D slice of the complete 4D light field is recorded. Figure 1(a) shows an area-scan sensor observing multiple object lines at one time. Due to synchronized motion and image acquisition, which is a common technique in industrial machine vision, a specific object line is then observed under varying angles over time.

During the acquisition, the object is moved orthogonally to both the camera's optical axis as well as the orientation of the sensor lines. In each time step, a region of interest consisting of several lines is read out from the area-scan sensor, as shown in Fig. 1(b). By collecting all corresponding lines acquired over time, a 3D light-field data structure is produced. This data structure represents multiple views of the object observed from different viewing angles w.r.t. the system optical axis (i.e., all first lines constitute the first image, all second lines constitute the second image, etc.).

3.2 Refocusing Using Light Field

Refocusing in the context of light fields can be seen as summation along distinct slopes in the EPI domain. An EPI stack $E(u, v, s)$ is defined as a 3D data structure containing view images, i.e. the horizontal slices in Fig. 1, indexed by the line number $s = 1, \ldots, n$. The individual pixels in each view are indexed by (u, v). Summation along the slope given by θ in Fig. 1(b) focuses to the focal point, i.e. to the point P shown in Fig. 4. Summation along slopes tilted more to the right focuses computationally to points closer to the camera and summation along slope directions tilted to the left focuses to points further away from the camera, respectively.

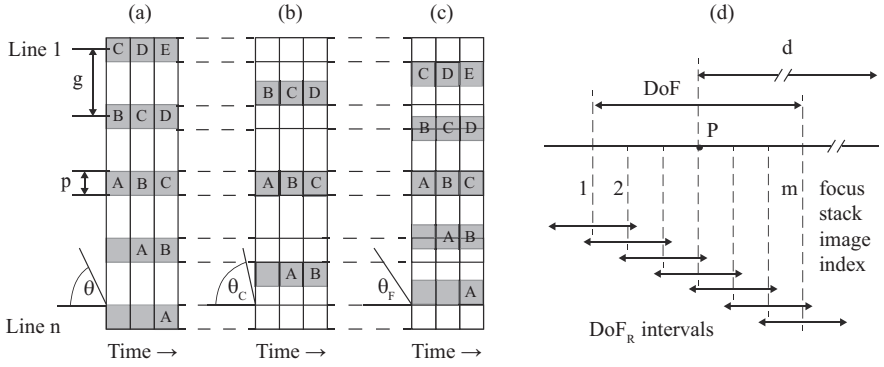

Fig. 2. Computational refocusing of object points acquired over time to (a) the focal point, (b) a point closer than the focal point and (c) a point further away than the focal point, (d) m individual focal planes with Depth of Field DoF_R covering the optical DoF

For refocusing at some image position along a slope θ we define a sheared EPI stack E_θ by

$$E_\theta(u, v, s) = E(u + (s - \hat{s})\theta, v, s), \quad s = 1, \ldots, n \tag{1}$$

where $\hat{s}$ is the index of the reference view. The number of views n extracted from the image sensor depends mainly on the applicable *Field of View* (FoV), which is limited by properties of optics as well as illumination. The refocused radiance values R are given by

$$R_\theta(u, v) = \sum_{s=1}^{n} E_\theta(u, v, s). \tag{2}$$

Fig. 2(a) shows three consecutively imaged sensor regions consisting of n lines each, i.e. those would be the three rightmost vertical planes inserted into the EPI shown in Fig. 1. Refocusing to the focal point is done by summation along the angle θ. The image shows single points "A" to "E" in the extracted sensor region. Assuming that the points "A" to "E" are all in the focal plane, all object regions are correctly summed up by integration along θ. Refocusing to a point closer than the focal point requires an angle $\theta_C > \theta$ in order to correctly sum up corresponding object regions, see Fig. 2(b). Analogously, Fig. 2(c) shows the situation for points further away than the focal point where we observe $\theta_F < \theta$. Figs. 2(a)-(c) also show that not all sensor lines are used for the construction of the EPI, a *stride* g defines the spacing between used lines. The senor pixel size is denoted by p.

Refocusing computationally increases the aperture diameter A by the size of the FoV. The FoV depends mainly on the focal length f and object distance d. In particular, we extract n lines from the CMOS sensor. The number of lines in the extracted region is practically limited by the extent of a sufficiently illuminated area on the object. Again, assuming the thin-lens-model with $1/f = 1/d + 1/b$

we obtain an increase of the aperture diameter by A_I, which is derived from congruent triangles shown in Fig. 3

$$A_I = 2 \cdot d \cdot \lfloor n/2 \rfloor \cdot g \cdot p \cdot \frac{d-f}{f}, \tag{3}$$

where p is the pixel pitch. The number n is usually chosen as a odd number and the central line contains the principal point, thus looking straight towards the object.

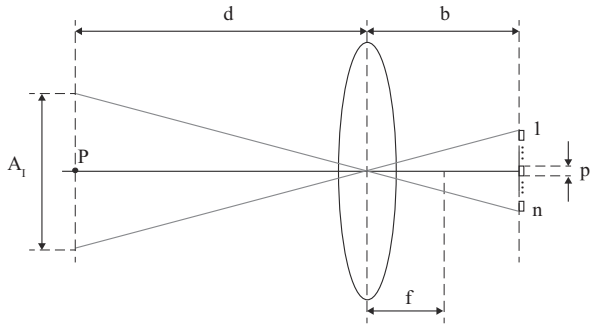

Fig. 3. Field of view A_I

The DoF relates image sharpness to object distance for an image acquisition system. We recall the basic DoF formulas for the thin-lens-approximated system shown in Fig. 4. We assume a system with a focal length f, an object distance d and a circle of confusion with diameter c, which is usually chosen to be equal to the sensor's pixel size p. The acceptable *depth of focus* (dof) w.r.t the sensor plane and the DoF in the object domain, where $dof = d_1 + d_2, d_1 = d_2$ and $DoF = D_1 + D_2, D_1 < D_2$, are related via the magnification M of the system, which is typically smaller than 1. The front DoF D_1 and back DoF D_2 give the total DoF using

$$D_{1,2} = \frac{Fcd^2}{f^2 \pm Fcd}, \quad DoF = \frac{2Fcd^2f^2}{f^4 - F^2c^2d^2}, \tag{4}$$

where $F = f/A$ is the *F-number*, and A is the aperture diameter. Note the different sizes C_1 and C_2 of the backprojected circle of confusion at back and front distance of the DoF, i.e. the height of vertical bars at C_1 and C_2 limiting the DoF in Fig. 4.

It can be shown, that for a reasonable range of working distances and magnifications the DoF is approximately proportional to the F-number F, see Eq. 4. Using $F = f/A$ the DoF for refocused images becomes $DoF_R \approx DoF \cdot A/(A+A_I)$. A shallow DoF_R, which turns out to be favorable in sharpness assessment of refocused images, is obtained by choosing n, g and d as large as possible, see Eq. 3.

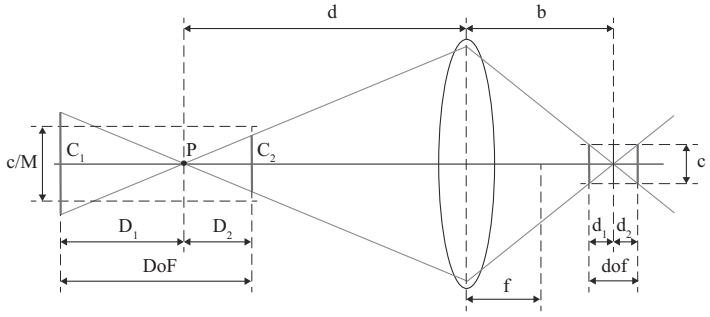

Fig. 4. Parameters describing the imaging process

A so called *focus stack* $R_\theta(u,v), \theta \in \Theta$ is constructed for a number m of different slopes θ corresponding to different *focal planes*. The number of focal planes m in the stack is chosen so as to cover the full DoF by overlapped intervals of refocused depth slices. Fig. 2(d) shows the organization of the focus stack along the axial depth direction. Overlapping intervals, with DoF_R each, span the single view DoF.

As shown in Fig. 1, the angle θ is related to the depth and becomes 45° (in this special case) for an object in the focal plane. An object closer to the sensor gets magnified and θ becomes larger, for objects behind the focal plane θ gets smaller. In general, the angle θ_R for refocusing to a plane at a distance d_R from the optical focal plane is given by (for simplicity, the stride g is not taken into account)

$$\theta_R = \arctan\Big(\frac{M'}{M}\Big), \qquad \frac{M}{M'} = \frac{d}{d-d_R}, \tag{5}$$

where M/M' is the relative magnification of the image.

Also note, that we achieve an improvement of the signal to noise ratio (SNR) for each refocused image by a factor of $\sqrt{n}$ when compared to a single view, which is the same as for digital CMOS time delay and integration (TDI) [19].

3.3 Sharpness Assessment

A number of measures for sharpness assessment used in autofocusing and SfF algorithms exist [17]. Linear operations, especially based on local first and second derivatives, are favorable due to their efficient implementability. We used a variant of the modified Laplacian suggested by Nayar [15]. In our case, as we are working with 3D light-field data, it is sufficient to estimate sharpness along one spatial direction only. Therefore, the sharpness measure $\Phi_\theta(u,v)$, i.e. the modified 1D Laplacian, for an image in the focus stack simply becomes

$$\Phi_\theta(u,v) = \big|\Delta\big(R_\theta(u,v)\big)\big| = \big|-R_\theta(u-1,v) + 2R_\theta(u,v) - R_\theta(u+1,v)\big|. \tag{6}$$

Spatial aggregation, e.g. local spatial average filtering, is recommended. The aggregated sharpness measure is denoted by $\bar{\Phi}_\theta(u,v)$.

The depth estimation is obtained from the focal plane index corresponding to the maximum of $\bar{\Phi}_\theta(u,v)$ over the set Θ of all integration angles θ

$$\theta_{\mathrm{MAX}}(u,v) = \arg\max_{\theta\in\Theta}\big(\bar{\Phi}_\theta(u,v)\big). \tag{7}$$

4 Implementation Details

In practical implementation, we make use of efficient computation in the case of dense focal stacking and optionally reuse the sharpness measures for view comparison.

4.1 Reversal of Refocusing and Sharpness Assessment

Using a focus measure Φ such as given in Eg. 6 with the expanded refocusing given in Eq. 2 we get

$$\Phi_\theta(u,v) = \Big|\Delta\Big(\sum_{s=1}^{n} E_\theta(u,v,s)\Big)\Big| = \Big|\sum_{s=1}^{n} \Delta\big(E_\theta(u,v,s)\big)\Big|, \tag{8}$$

provided that the operation Δ is linear. Reversing of the initial step of refocusing, i.e. summing over the views, with sharpness assessment is especially favorable if the number of focal planes is larger than the number of views, i.e. for $m > n$. Furthermore, the linear operation is well suited to be efficiently implemented in hardware right after image acquisition.

Integration along arbitrary slopes θ typically involves interpolation. Integration along slopes is equivalent to working with sheared focal stacks, e.g. see [5], where interpolation is moved into the shearing operation instead of the refocusing step. Nevertheless, from the computational point of view there is no advantage of maintaining a sheared focus stack.

4.2 Fusion with View Comparison

Spatial comparison of the reference view with refocused views is optionally used in order to increase robustness. In this case, Eq. 7 is extended to a combination of absolute sharpness assessment with the similarity between refocused and reference image sharpness

$$\theta_{\max}(u,v) = \underset{\theta\in\Theta}{\arg\max} \left(\bar{\Phi}_\theta(u,v) - \overline{\left| n\cdot\Delta\left(E_\theta(u,v,\hat{s})\right) - \sum_{s=1}^{n} \Delta\left(E_\theta(u,v,s)\right)\right|} \right). \tag{9}$$

5 Results

We present results for two objects, a printed circuit board (PCB) and a banknote mapped onto a 3D-printed wave.

A PCB shown in Fig. 5(a) was acquired from a distance $d = 370$mm. A lens with $f = 50$mm was used and a total number of $n = 9$ lines with a stride of $g = 32$ pixels were extracted from the sensor having a pixel size of $p = 7\mu$m. The achieved spatial resolution was 48μm per pixel. The PCB was transported on a conveyor belt with image acquisition synchronized with transport.

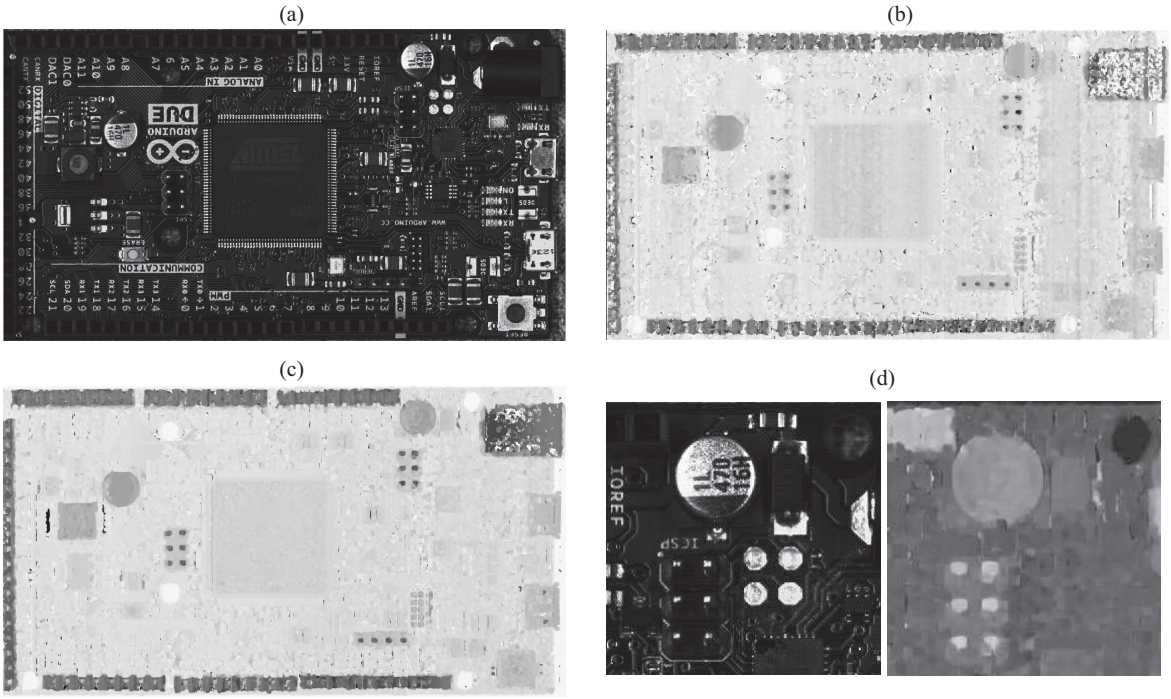

Fig. 5. Image of a PCB: (a) reference view, result of (b) SfR with sharpness assessment only, (c) fused with view comparison and (d) magnified regions

A total number of $m = 11$ planes for computational refocusing were selected. The initial step of focus assessment, i.e. the application of Δ was done on the original images by the modified Laplacian (see Eq. 6). The aggregation domain to obtain $\bar{\Phi}$ was 11×11 pixels. Results without (see Eq. 7) and with view comparison (see Eq. 9) are shown in Fig. 5(b) and (c), respectively. Fine details, e.g. the pin connectors in the upper right region, become clearly visible, see Fig. 5 (d). Furthermore, the fusion suggested in Sec. 4.2 helps to reduce wrong depth estimates.

In the next experiment a banknote was mapped onto a 3D-printed waveform, see Fig. 6(a), in order to have access to ground truth data. The ground truth, shown in Fig. 6(b), is a sinus undulation with a peak-to-valley range of 10 mm. Parameters were similar to the PCB experiment, with the exception of $f = 20$mm, a stride of $g = 16$ pixels and a spatial resolution of 120μm.

The estimated depth for using sharpness assessment only is shown in Fig. 6(c) and fails in the unstructured region on the left. In fact, this region contains a fine printed structure which is not resolved at 120μm pixel size. Fusion with view comparison largely solves this problem, see Fig. 6(d). Numerically, the mean absolute difference for the result from SfR compared to the ground truth were 1.77 mm with the median of 0.85mm. Fusion of SfR with view comparison achieved a mean absolute difference of 1mm with the median of 0.63mm. Absolute depth deviations from the ground truth scaled to the range $[0, 1]$, where white corresponds to a mean absolute error of 1cm and black means no error, are shown for different algorithmic alternatives in Fig. 6(e) and (f), respectively.

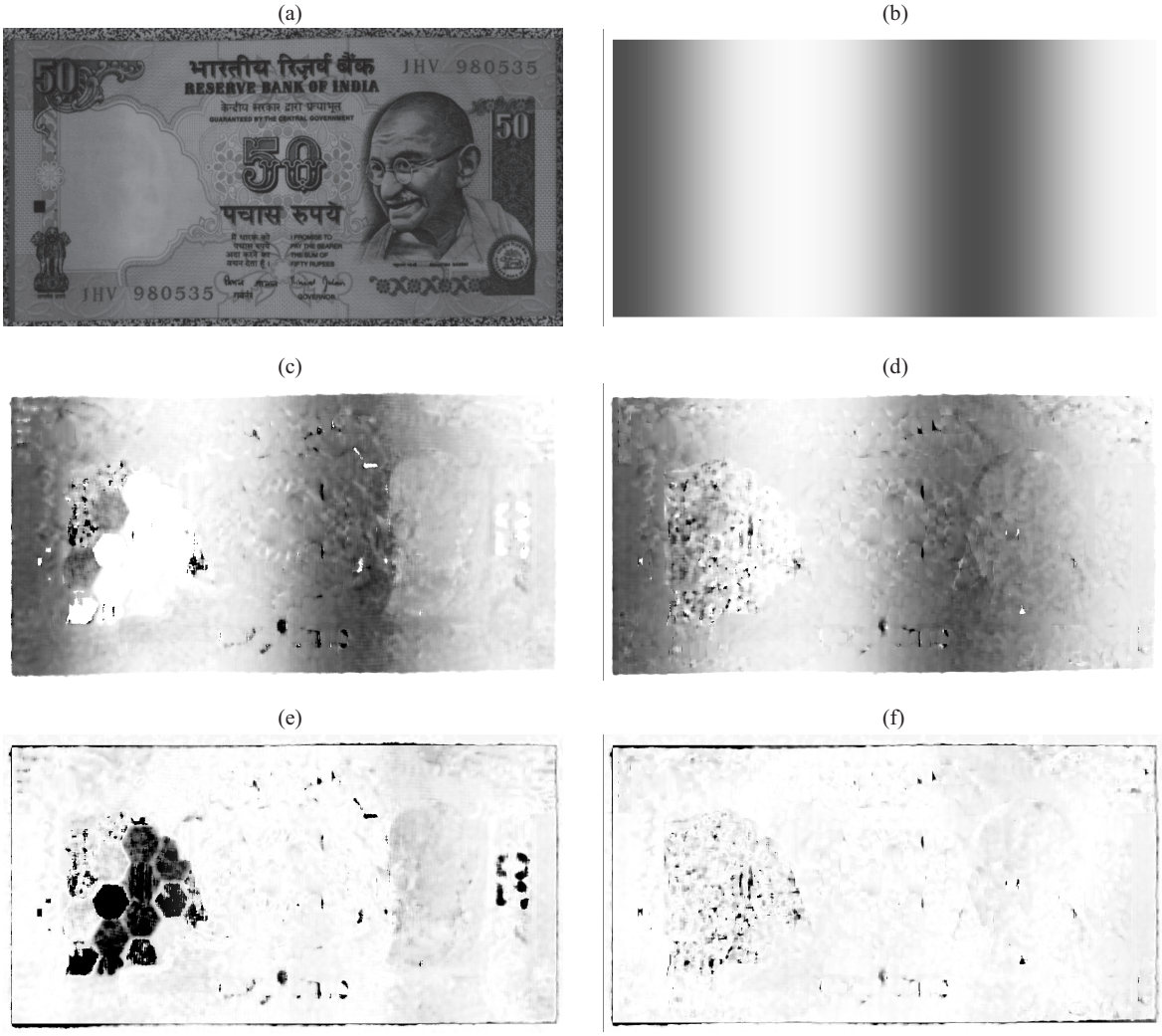

Fig. 6. Image of a 50 Rupee banknote mapped onto a 3D-printed wave: (a) reference view, (b) ground truth, result of (c) SfR with sharpness assessment only, (d) fused with view comparison, absolute difference images for (e) SfR with sharpness assessment only and (f) fused with view comparison

6 Conclusions

We have presented a method for depth estimation based on computational refocusing. The intended area of application is industrial inspection of moving objects. Synchronization of object motion and image acquisition is a prerequisite, which is state-of-the-art in machine vision systems. Operating in a 3D light-field and arranging operations in appropriate manner makes the method suited for real-time operation. Undergoing work includes integration into an industrial machine vision setup. Further research include modifications on the optical system, i.e. added apertures, in order collect the different views towards each object point over a shorter period of time. Issues related to inhomogeneity of illumination and loosely synchronized acquisition and transport will be reduced by this measure.

References

1. Schechner, Y.Y., Kiryati, N.: Depth from defocus vs. stereo: How different really are they? Intl. J. of Comp. Vis. 39(2), 141–162 (2000)
2. Krotkov, E., Martin, J.P.: Range from focus. In: Proc. of Intl. Conf. on Robotics and Automation, pp. 1093–1098 (1986)
3. Pentland, A.P.: A new sense for depth of field. IEEE Trans. on Pat. Anal. and Mach. Intel. 8(4), 523–531 (1987)
4. Levoy, M., Hanrahan, P.: Light field rendering. In: Proc. of Conf. on Comp. Graph. and Interactive Tech., pp. 31–42 (1996)
5. Tao, M.W., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from combining defocus and correspondence using light-field cameras. In: Proc. of Intl. Conf. on Comp. Vis (ICCV) (December 2013)
6. Vaish, V., Szeliski, R., Zitnick, C.L., Kang, S.B., Levoy, M.: Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In: Proc. of Conf. on Comp. Vis. and Pat. Rec. (CVPR) (2006)
7. Bolles, R.C., Baker, H.H., David, Marimont, H.: Epipolarplane image analysis: an approach to determining structure from motion. Intl. J. of Comp. Vis. 1(1), 7–55 (1987)
8. Wilburn, B., Joshi, N., Vaish, V., Talvala, E.-V., Antunez, E., Barth, A., Adams, A., Horowitz, M., Levoy, M.: High performance imaging using large camera arrays. ACM Trans. on Graph. 24(3), 765–776 (2005)
9. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: "Light field photography with a hand-held plenoptic camera," Tech. Rep. CSTR 2005-02, Stanford University (April 2005)
10. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.: Scene reconstruction from high spatio-angular resolution light fields. ACM Trans. on Graph. 32(4), 73:1–73:12 (2013)
11. Wanner, S., Goldlücke, B.: Globally consistent depth labeling of 4D light fields. In: Proc. of Conf. on Comp. Vis. and Pat. Rec (CVPR), pp. 41–48 (2012)
12. Georgiev, T., Lumsdaine, A.: Depth of field in plenoptic cameras. In: Proc. of Eurographics (April 2009)
13. Perwaß, C., Wietzke, L.: Single lens 3D-camera with extended depth-of-field. In: Proc. of SPIE-IS&T Elec. Imag., vol. 8291, pp. 8–15 (2012)
14. Nayar, S.K., Watanabe, M., Noguchi, M.: Real-time focus range sensor. IEEE Trans. on Pat. Anal. and Mach. Intel. 18(12), 1186–1198 (1996)
15. Nayar, S.K., Nakagawa, Y.: Shape from focus. IEEE Trans. on Pat. Anal. and Mach. Intel. 16(8), 824–831 (1994)
16. Danzl, R., Helmli, F., Scherer, S.: Focus variation a robust technology for high resolution optical 3D surface metrology. J. of Mech. Eng. 57(3), 245–256 (2011)
17. Pertuz, S., Puig, D., Garcia, M.A.: Analysis of focus measure operators for shape-from-focus. Pat. Rec. 46(5), 1415–1432 (2013)
18. Štolc, S., Huber-Mörk, R., Holländer, B., Soukup, D.: Depth and all-in-focus images obtained by multi-line-scan light-field approach. In: Proc. of SPIE-IS&T Elec. Imag., vol. 9024, pp. 7–16 (2014)
19. He, X.-F., Nixon, O.: Time delay integration speeds up imaging. Photonics Spectra 46(5), 50–55 (2012)