Depth Estimation within a Multi-Line-Scan Light-Field Framework

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

AIT Austrian Institute of Technology GmbH, Intelligent Vision Systems,
Safety & Security Department, Donau-City-Straße 1, 1220 Vienna, Austria

Abstract. We present algorithms for depth estimation from light-field data acquired by a multi-line-scan image acquisition system. During image acquisition a 3-D light field is generated over time, which consists of multiple views of the object observed from different viewing angles. This allows for the construction of so-called epipolar plane images (EPIs) and subsequent EPI-based depth estimation. We compare several approaches based on testing various slope hypotheses in the EPI domain, which can directly be related to depth. The considered methods used in hypothesis assessment, which belong to a broader class of block-matching algorithms, are modified sum of absolute differences (MSAD), normalized cross correlation (NCC), census transform (CT) and modified census transform (MCT). The methods are compared w.r.t. their qualitative results for depth estimation and are presented for artificial and real-world data.

1 Introduction

We address the processing of light-field data for a specific light-field acquisition setup in industrial machine vision. A number of algorithms, with potential real-time in-line operation, are compared w.r.t. the achievable depth estimation quality on simulated and real-world data. Light field-based processing is performed in the EPI domain [1], originally introduced for structure from motion estimation. All considered algorithms are based on slope hypothesis testing in the EPI domain. Image features appear as linear structures in EPIs and the slopes of those structures can be related to image depth information.

An alternative to hypothesis testing is direct estimation of the principal orientation of linear structures in EPIs, e.g. using the *structure tensor* [2]. As structure tensor estimation suffers from sensitivity to noise, a global optimization strategy was suggested, which significantly increases accuracy of depth maps, but also introduces high computational demands. Besides the robustness w.r.t. noise, an EPI analysis algorithm has to consider the object reflectance properties. The assumption of static-Lambertian scene behavior is typically violated for real scenes, especially when dealing with dynamically constructed light fields. The static-Lambertian assumption presumes that all object points preserve their radiance values regardless of the viewing angle and the illumination is expected to stay constant over time.

G. Bebis et al. (Eds.): ISVC 2014, Part II, LNCS 8888, pp. 471–481, 2014.

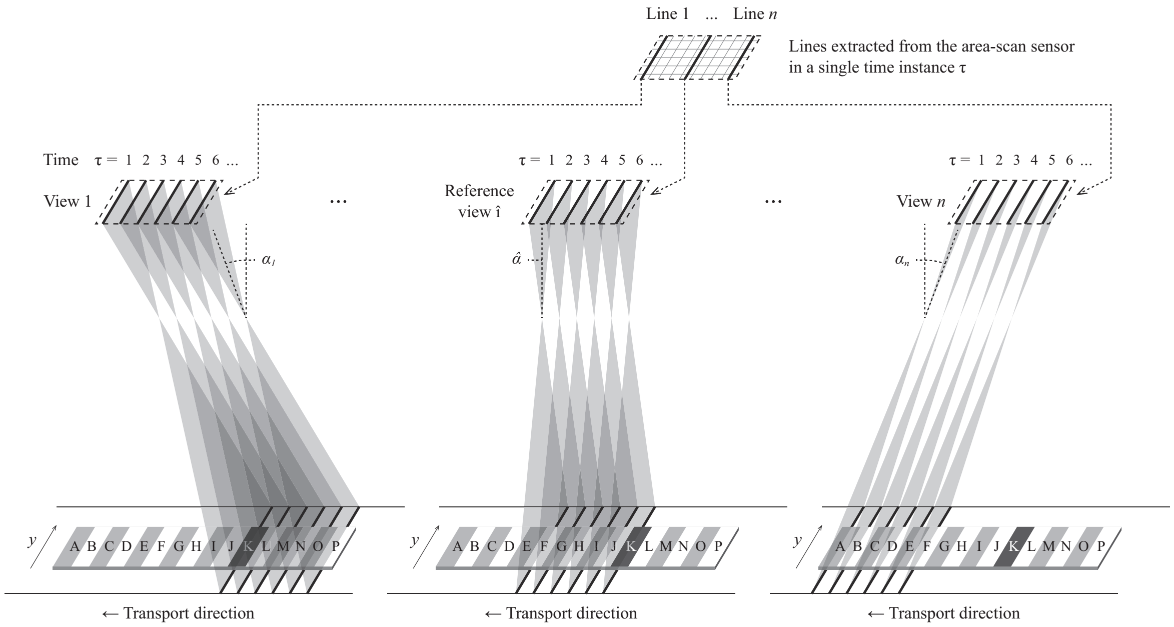

Fig. 1. Formation of multiple object views by the light-field multi-line-scan camera over time $\tau \in \{1, 2, 3, \cdots\}$. The obtained 3-D light-field data structure is shown in Fig. 2.

Although all algorithms presented here are generally applicable to light fields acquired by a broader range of recording devices, we point out some unique acquisition characteristics. During our light-field acquisition, the object is moved orthogonally to the camera's optical axis and the orientation of the sensor lines. By collecting all corresponding lines acquired over time (i.e. all 1st lines form one image, all 2nd lines form another image, etc.), a 3-D light field is produced. Multiple views of the object observed from different viewing angles are obtained.

The approaches for EPI analysis which are most similar to ours, i.e., also based on hypothesis testing, are those by Kim et al. [3] and Venkataraman et al. [4]. Kim et al. rely on the static-Lambertian assumption and identify the best hypotheses as those which minimize the overall deviation of radiance values along hypothetical slopes w.r.t. the value in the reference view. Venkataraman et al. evaluate the sum of absolute distances (SAD) of radiances between different views for a discrete number of hypothesized depths.

The paper is organized as follows: Section 2 discusses light-field imaging in general and our 3-D light-field acquisition in detail. In Section 3, we introduce the algorithms used for hypothesis testing. Experimental results are given in Section 4 and conclusions are drawn in Section 5.

2 Multi-Line-Scan Light Fields

In general, a light field is defined as a 4-D radiance function of 2-D position and direction in regions of space free from occluders [5]. Light-field acquisition can be realized in various ways, e.g., by a multi-camera array capturing scene from different viewing angles [6]. A gantry system uses a single camera which is mechanically displaced during acquisition of a static scene [5]. PiCam realizes a monolithic camera array using a 4x4 lens array [4]. Recently, an unstructured acquisition using a moving hand-held device was also suggested [7]. Another way of light-field acquisition is splitting the optical path using filters, masks, code patterns, etc., which is termed *coded aperture imaging* [8]. Plenoptic cameras use microlens arrays placed in front of the sensor plane to obtain 4-D light fields [9–11].

In our approach, an area-scan sensor is used to capture object positions (x, y) under varying angle α along the transport direction over time. In this setup, there is no variation of the angle measured across the transport direction. Therefore, only a 3-D slice of the complete 4-D light field is acquired. In Fig. 1, three sections depict exemplary settings for three sensor lines repeatedly extracted from the area-scan sensor. Each of these lines (indexed $i = 1, \ldots, \hat{i}, \ldots, n$) observes some region of a conveyor belt under a different but fixed viewing angle α_i. Since the objects are transported in front of the camera, each sensor line observes every object region at distinct time instances τ. In this manner, 1st sensor line sees the object region "K" at time instant $\tau = 1$, $\hat{i}$-th sensor line sees "K" at time instances $\tau = 6$, and n-th sensor line doesn't see "K" at any of the time instants $\tau = 1$ through 6, only few steps later. In the following, we refer to an image formed by all object lines collected over time characterized by a constant viewing angle α_i as a *view*, which is denoted as $I_i(x, y)$. The dimension x expresses the spatial extent of the object in the transport direction, which is generated over time τ. The dimension y corresponds with the spatial extent along sensor lines.

The EPI data corresponding to varying observation angles α_i at each line of an area-scan sensor is shown in Fig. 2. E.g., the object region "K" in the focal plane is seen under different angles α_i at different time instances τ, which corresponds to dimension x. Integration along a line with the slope θ provides the irradiance estimate for the in-focus object region "K".

Parts of the object that do not reside in the camera focal plane are mapped differently to the sensor plane. Also here corresponding radiance values map to linear structures, but with different slopes depending on a distance between camera and the corresponding object point.

Our system can be seen as a narrow-baseline multi-view stereo system, which means that the field of view is typically very narrow (approx. 2°). This significantly reduces geometrical distortions and causes that the system directly fulfills the epipolar constraint, as well as reduces the correspondence problem of stereo vision. Given the non-canonical verged stereo geometry (i.e. the zero disparity is obtained in a finite distance from the camera), it can be shown that there is a simple linear relationship between the depth and detected disparity [12].

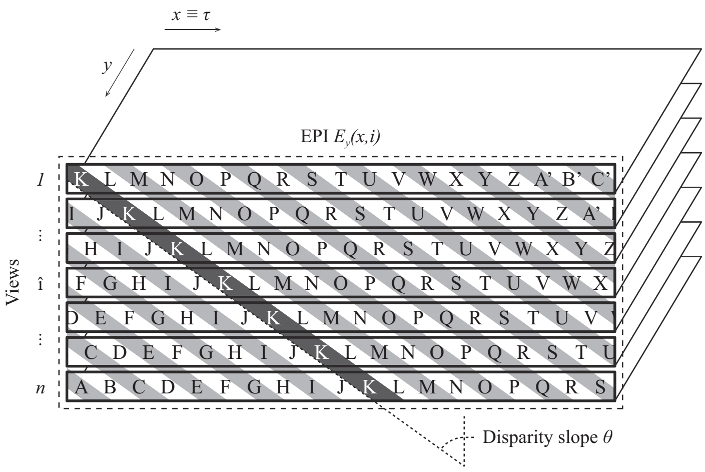
x ≡ τ
y
EPI E_y(x,i)
Views
1
⋮
î
⋮
n
K L M N O P Q R S T U V W X Y Z A' B' C'
I J K L M N O P Q R S T U V W X Y Z A'
H I J K L M N O P Q R S T U V W X Y Z
F G H I J K L M N O P Q R S T U V W X
D E F G H I J K L M N O P Q R S T U V
C D E F G H I J K L M N O P Q R S T U
A B C D E F G H I J K L M N O P Q R S
Disparity slope θ

Fig. 2. 3-D light-field obtained by the multi-line-scan camera. Letters "A","B","C",... represent visual structures residing in the object plane, as shown in Fig. 1. These structures map into diagonal lines in the EPI domain, slopes of which θ are proportional to the distance between the camera and the corresponding object point.

3 Disparity Estimation

We adopt the method of slope hypothesis testing for analyzing orientations of the EPI structures. For each spatial location (x, y) in the reference view $I_{\hat{i}}$ (typically the central view), a number of slope hypotheses are generated in the corresponding EPI $E_y(x, i)$ and the best one is taken as the orientation estimate. In the following, we provide descriptions of block-matching approaches, which are capable of handling setups that violate the static-Lambertian scene condition and are rather robust to noise.

Let us denote an $m \times m$ image patch at position (x, y) in a view I_i as $\mathcal{P}(x, y, i)$. Further, we define a set of image patches $\Omega(x, y, \theta)$ collected along an imaginary line within the EPI $E_y(x, i)$ with a slope θ intersecting the reference view $I_{\hat{i}}$ in the position (x, y) (compare Fig. 2):

$$\Omega(x, y, \theta) = \left\{ \mathcal{P}(x + (i - \hat{i})\,\theta,\ y,\ i) \middle| i = 1, \ldots, n \right\},$$

where n stands for the number of views. In other words, for a position (x, y), from each view one $m \times m$ patch is included in $\Omega(x, y, \theta)$, whose center resides on that imaginary line passing through (x, y) with slope θ in the EPI domain. In general, the construction of the sets $\Omega(x, y, \theta)$ requires interpolation. In our experiments, we used cubic interpolation in order to get high-quality results.

A simple way to measure patch similarities would be to employ the SAD measure similar to Venkataraman et al. [4]. Note that SAD cannot handle very well non-static-Lambertian scenes. In [13], the authors showed that an adaptation of SAD, the so-called modified SAD (MSAD), is more appropriate in such situations, because illumination and contrast differences between views are factored out in a pre-processing step. This pre-normalization is accomplished by subtracting the mean $\mu(P)$ of each patch from all its elements and by dividing them by their standard deviation $\sigma(P)$. We refer to the resulting normalized patch as $\tilde{P} = (P - \mu(P))/\sigma(P)$. Correspondingly, a set $\Omega(x, y, \theta)$ consisting of pre-normalized patches is denoted as $\tilde{\Omega}(x, y, \theta)$.

Based on the set $\tilde{\Omega}(x, y, \theta)$ we define an auxiliary set $\Delta_{MSAD}(x, y, \theta)$ consisting of distances of all normalized image patches from the corresponding normalized patch in the reference view:

$$\Delta_{MSAD}(x, y, \theta) = \Big\{ \sum_{\substack{\text{patch}\\ \text{pixels}}} |P - \mathcal{P}(x, y, \hat{\imath})| \,\Big|\, P, \mathcal{P} \in \tilde{\Omega}(x, y, \theta) \Big\},$$

where the element-wise absolute differences are obtained by a simple subtraction of gray-scale values or the Euclidean distance in the case of color images.

Consequently, the overall MSAD cost function $C_{MSAD}(x, y, \theta)$ for each object point (x, y) and any given hypothesized slope θ is defined as follows:

$$C_{MSAD}(x, y, \theta) = \sum_{\delta \in \Delta_{MSAD}(x,y,\theta)} \delta.$$

Another well known measure for comparing images which is insensitive to illumination changes is NCC. In the second approach, we employ NCC to assess the similarity of image patches. Analogously to the MSAD approach, we compute a set of NCC coefficients between image patches in all views with respect to the reference view as follows:

$$\Delta_{NCC}(x, y, \theta) = \Big\{ \Big\langle P, \mathcal{P}(x, y, \hat{\imath}) \Big\rangle \,\Big|\, P, \mathcal{P} \in \tilde{\Omega}(x, y, \theta) \Big\},$$

where $\langle \cdot, \cdot \rangle$ denotes a vector scalar product. Consequently, the NCC cost function can be defined as:

$$C_{NCC}(x, y, \theta) = \sum_{\delta \in \Delta_{NCC}(x,y,\theta)} -\delta.$$

The 'minus' before δ transforms the similarity measure into a cost function. In the case of color images, the patches in the individual color channels are treated separately giving three times as many NCC coefficients in $\Delta_{NCC}(x, y, \theta)$, which are equally included in the final cost assessment.

Although, MSAD and NCC are invariant to illumination variations, they are both computationally quite expensive, which makes them less suitable for time-critical applications. Regarding real-time applications, one would need a similarity measure that is robust against illumination variations, but at the same time is computationally efficient. The CT [14] as well as MCT [15] approaches have this property, which has been exploited in various applications, e.g., Ambrosch et al. [16] reported on the use of (M)CT for efficient stereo vision applications.

In the CT approach, the images are first census-transformed in a pre-processing step. The actual patch comparisons take place later on based on CT-transformed image patches. Both CT as well as MCT yield a bit vector, which serves as a descriptor of the image patch centered at the corresponding pixel. Each entry in that bit vector reflects whether the corresponding pixel value is lower or equal (0) or greater (1) than the central pixel of the patch. In the case of MCT, the patch values are compared to the mean of the patch rather than to the central pixel, which makes the transform in general more robust against noise. Dissimilarity between obtained (M)CT bit vectors is measured by means of the Hamming distance. Analogously to $\tilde{\Omega}$ in the case of normalized image patches, we introduce a set $\bar{\Omega}(x, y, \theta)$ of bit vectors obtained by the (M)CT transform from image patches in $\Omega(x, y, \theta)$. A corresponding hypothesis cost function for disparity estimation is given by the sum of Hamming distances of bit vectors along a hypothesized slope in the EPI domain w.r.t. the bit vector in the reference view $I_{\hat{\imath}}$:

$$C_{(M)CT}(x, y, \theta) = \sum_{\delta \in \Delta_{CT}(x,y,\theta)} \delta,$$

where

$$\Delta_{(M)CT}(x, y, \theta) = \left\{ H(P, \mathcal{P}(x, y, \hat{\imath})) \middle| \, P, \mathcal{P} \in \bar{\Omega}(x, y, \theta) \right\},$$

$H(\cdot, \cdot)$ is the Hamming distance. Color images are handled analogously to the NCC approach.

Whichever approach is chosen, for every hypothesized slope θ and object point (x, y) a cost value is obtained. Thus, each hypothesis gives a cost map covering the entire spatial image domain. In order to suppress spurious disparity estimates, all cost maps are box-filtered with a filter mask of the same size as the image patches used for cost computations. In order to handle noisy data more effectively, one may need to use standard or bilateral median filters instead of the proposed box filter, however, at higher computational cost. Finally, in each object point (x, y), the slope associated with the minimal cost is chosen as the final disparity estimate.

In order to reduce the computational load associated with testing a lot of hypotheses, we use quite large steps between hypothesized slopes. Nevertheless, we compensate for this loss of granularity at the end by applying so-called *sub-hypothesis refinement* [12]. In each object point (x, y), we consider the corresponding cost values $C_*(x, y, \theta)$ as a function of θ and locally fit a quadratic function to the cost values around the determined minimal cost hypothesis θ_*. Only the minimum of this quadratic fit refers to the actual final disparity estimate. Sub-hypothesis refinement is computationally very efficient compared to an approach where dense hypotheses sets have to be evaluated and our results show smooth variation of disparities for gentle depth changes.

4 Experimental Results

In this section, we compare the described disparity estimation algorithms based on synthetic data, where the ground-truth disparity values are explicitly known. Although a lot of work has been published on the topic of comparing block-matching algorithms applied to different tasks, to our knowledge, no prior work

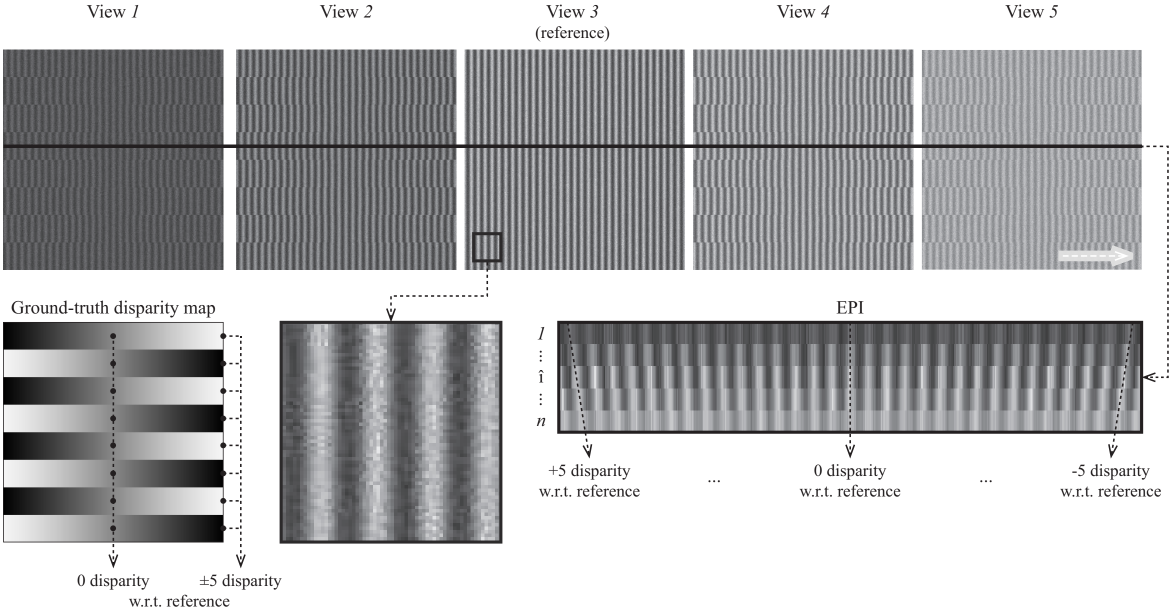

Fig. 3. Example of the synthetic ground-truth light field comprising 5 views of the signal of 8 times the Nyquist wavelength (i.e. 16 pixels per period). In View 5, the dashed arrow marks the emulated transport direction.

investigated the impact of different block-matching methods on depth estimations from light-field data.

Fig. 3 shows an example of synthetic ground-truth data. The performance of all methods was compared on the disparity interval of $[-5, 5]$, which corresponds with the narrow baseline multi-view stereo system described above. The light fields consisted of $\{3, 5, 7, 9, 11\}$ views and the simulated signal wavelengths were $\{2, 4, 8, 16\}$ times the Nyquist wavelength. Block-matching and spatial filter sizes were equal and chosen from two options 3×3 and 7×7. The camera noise was simulated by additive Gaussian noise with an amplitude of 10 dB. Properties of a scene violating the static-Lambertian condition were mimicked by making the signal contrast and bias considerably different in each light-field view.

Tab. 5 shows root mean-squared errors (RMSE) of disparity values delivered by all methods, when compared with corresponding ground-truth disparity maps. The smaller patch size (3×3) provided significantly worse results due to aperture problems at longer wavelength image structures. Certainly, the 7×7 configuration would also run into the same problems if not enough high-frequency structures were available. Hence, in such cases one may need to employ additional algorithmic efforts such as, e.g., pyramid processing [3].

Tab. 5 reveals that numbers of views less than 7 generate significantly less accurate disparity estimates than setups with 7 or more views. For the cases with 3 and 5 views, the obtained disparity errors are high due to the correspondence problem caused by an insufficient sampling density in the viewing angle domain. Therefore, it is advisable to use light fields with $\geq$7 views.

Although RMSE for NCC are apparently smaller, it is remarkable that all four methods show quite similar results. MSAD is only slightly worse than NCC.

Table 1. RMSE results of the ground-truth experiment with the MSAD, NCC, CT, and MCT methods applied to light fields with different numbers of views and signals of different wavelengths. The best performing configurations with RMSE within the $[0, 1)$ are marked in a bold font. Medium-quality configurations with RMSE ranging in $[1, 2)$ are marked in light gray. Low-quality configurations providing RMSE ≥ 2 are marked in dark gray.

Method	Domain	Small patch size (3 × 3)					Large patch size (7 × 7)				
	λ Views	2	4	8	16	← Avg.	2	4	8	16	← Avg.
MSAD	3	4.25	4.67	1.54	3.50	3.49	3.82	4.25	**0.35**	1.44	2.46
	5	4.23	**0.22**	1.27	3.41	2.28	3.51	**0.10**	**0.40**	1.32	1.33
	7	**0.07**	**0.19**	1.19	3.36	1.20	**0.07**	**0.10**	**0.40**	1.25	**0.46**
	9	**0.07**	**0.18**	1.16	3.35	1.19	**0.07**	**0.10**	**0.41**	1.25	**0.46**
	11	**0.07**	**0.16**	1.07	3.28	1.14	**0.07**	**0.09**	**0.36**	1.12	**0.41**
NCC	3	4.14	4.65	1.47	3.46	3.43	3.52	4.18	**0.36**	1.34	2.35
	5	4.07	**0.19**	1.19	3.35	2.20	3.06	**0.13**	**0.40**	1.24	1.21
	7	**0.05**	**0.17**	1.11	3.29	1.15	**0.06**	**0.13**	**0.41**	1.17	**0.44**
	9	**0.05**	**0.16**	1.08	3.27	1.14	**0.06**	**0.14**	**0.42**	1.18	**0.45**
	11	**0.05**	**0.14**	**0.99**	3.21	1.10	**0.06**	**0.14**	**0.39**	1.06	**0.41**
CT	3	4.48	4.75	2.23	4.08	3.88	3.87	4.37	**0.49**	1.84	2.64
	5	4.63	**0.39**	1.85	3.99	2.72	3.77	**0.13**	**0.51**	1.72	1.53
	7	**0.16**	**0.35**	1.75	3.95	1.55	**0.06**	**0.12**	**0.50**	1.65	**0.58**
	9	**0.16**	**0.33**	1.70	3.94	1.53	**0.06**	**0.12**	**0.50**	1.63	**0.58**
	11	**0.16**	**0.31**	1.59	3.87	1.48	**0.06**	**0.12**	**0.44**	1.51	**0.53**
MCT	3	4.24	4.69	1.61	3.59	3.53	3.71	4.36	**0.38**	1.46	2.48
	5	4.18	**0.23**	1.29	3.48	2.30	3.19	**0.11**	**0.41**	1.34	1.26
	7	**0.07**	**0.20**	1.21	3.41	1.22	**0.07**	**0.11**	**0.41**	1.27	**0.47**
	9	**0.07**	**0.19**	1.18	3.38	1.20	**0.07**	**0.11**	**0.42**	1.27	**0.47**
	11	**0.06**	**0.17**	1.09	3.32	1.16	**0.07**	**0.10**	**0.38**	1.15	**0.42**

Especially for relevant numbers of views (≥7), all methods perform practically equally well. CT shows the highest RMSE, but on the other hand, can run significantly faster than NCC or MSAD. Thus, it is also an interesting outcome that MCT outperforms CT that much (in average by about 16%, see Tab. 5, last column), as its computational effort would be only slightly larger than of CT.

Finally, we present also disparity maps obtained for a real-world object – a printed circuit board (PCB) – captured with our multi-line-scan light-field setup (see Fig. 4). Each disparity map was computed with one of the four compared disparity estimation algorithms. It is perceivable that CT and MCT results are noisier than those obtained by MSAD and NCC. Nevertheless, all four results show quite accurate rendering of the object's 3-D structure. While these results certainly coincide with a common knowledge about block-matching methods in stereo vision, our results confirm similar behavior in the context of depth estimation from light fields supported by numerical data.

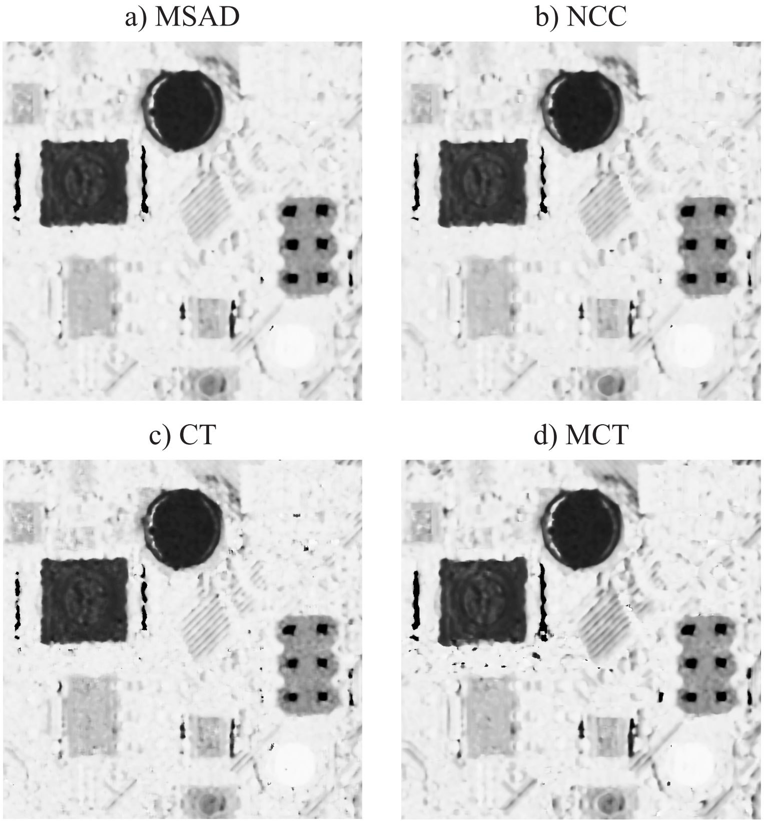

Fig. 4. Disparity maps of PCB obtained by the MSAD, NCC, CT, and MCT methods from the light field comprising 9 views

5 Conclusions and Discussion

In this paper, we presented a study on depth estimation in the EPI domain, whereas we considered the special case of a 3-D light field acquired under non-static-Lambertian scene conditions by a narrow-baseline stereo system. The quality of the obtained depth estimates of four investigated block-matching methods NCC, MSAD, CT, and MCT was compared on the basis of synthetic ground-truth data comprising different frequencies of image structures, different numbers of views, and disparities within a certain range. We presented reconstruction errors w.r.t. ground-truth data as well as actual depth maps for a real-world object. The experiments revealed that for a number of views ≥ 7, all four algorithms are capable of generating accurate dense depth maps. Compared to usual depth or optical flow estimation methods, where usually only pairs of views are taken into consideration, this result demonstrates how far the EPI-based depth estimation methodology, which naturally involves many views, makes depth estimation more robust. Moreover, the narrow-baseline in our system further facilitates the estimations, since there are no severe geometrical distortions to be expected in different views. While we saw that all four block-matching metrics yielded fairly good depth estimations, NCC provided the best results, followed by MSAD.

However, both methods require a relatively large computational effort. On the other hand, CT and MCT proved to be noisier, but are much less computationally demanding. Finally, our results indicate that high-quality depth estimates can be reliably computed from 3-D light fields, suitable for time-critical applications.

References

1. Bolles, R.C., Baker, H.H., Marimont, D.H.: Epipolarplane image analysis: an approach to determining structure from motion. Int. J. Comput. Vis. 1(1), 7–55 (1987)
2. Wanner, S., Goldlücke, B.: Globally consistent depth labeling of 4D light fields. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Providence, RI, pp. 41–48. IEEE (2012)
3. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.H.: Scene reconstruction from high spatio-angular resolution light fields. ACM Trans. Graph. 32(4), 73:1–73:12 (2013)
4. Venkataraman, K., Lelescu, D., Duparré, J., McMahon, A., Molina, G., Chatterjee, P., Mullis, R., Nayar, S.: PiCam: an ultra-thin high performance monolithic camera array. ACM Trans. Graph. 32(6), 166:1–166:13 (2013)
5. Levoy, M., Hanrahan, P.: Light field rendering. In: Proc. Conf. on Computer Graphics and Interactive Techniques, SIGGRAPH, pp. 31–42. ACM, New York (1996)
6. 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. Graph. 24(3), 765–776 (2005)
7. Davis, A., Levoy, M., Durand, F.: Unstructured light fields. Comput. Graph. Forum 31(2pt.1), 305–314 (2012)
8. Liang, C.-K., Lin, T.-H., Wong, B.-Y., Liu, C., Chen, H.H.: Programmable aperture photography: multiplexed light field acquisition. ACM Trans. Graph. 27(3), 55:1–55:10 (2008)
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. Lumsdaine, A., Georgiev, T.: The focused plenoptic camera. In: Proc. IEEE Int. Conf. on Computational Photography (ICCP), San Francisco, CA, pp. 1–8. IEEE (2009)
11. Perwaß, C., Wietzke, L.: Single lens 3D-camera with extended depth-of-field. In: Proc. of SPIE – Human Vision and Electronic Imaging XVII, vol. 8291, pp. 829108-1–829108-15 (2012)
12. Štolc, S., Soukup, D., Holländer, B., Huber-Mörk, R.: Depth and all-in-focus imaging by a multi-line-scan light-field camera. Journal of Electronic Imaging 23(5), 053020 (2014)
13. Štolc, S., Huber-Mörk, R., Holländer, B., Soukup, D.: Depth and all-in-focus images obtained by multi-line-scan light-field approach. In: Niel, K.S., Bingham, P.R. (eds.) Proc. of SPIE-IS&T Electronic Imaging – Image Processing: Machine Vision Applications VII, San Francisco, CA (February 2014)

14. Zabih, R., Woodfill, J.: Non-parametric local transforms for computing visual correspondence. In: Eklundh, J.-O. (ed.) ECCV 1994. LNCS, vol. 801, pp. 151–158. Springer, Heidelberg (1994)
15. Cyganek, B.: Comparison of nonparametric transformations and bit vector matching for stereo correlation. In: Klette, R., Žunić, J. (eds.) IWCIA 2004. LNCS, vol. 3322, pp. 534–547. Springer, Heidelberg (2004)
16. Ambrosch, K., Zinner, C., Kubinger, W.: Algorithmic considerations for real-time stereo vision applications. In: Proc. Machine Vision and Applications (MVA), Keio University, Yokohama, JP, pp. 231–234 (May 2009)