 DOI: 10.3217/978-3-85125-528-7-11

Depth estimation using light fields and photometric stereo with a multi-line-scan framework

Doris Antensteiner[1], Svorad Štolc[1], and Reinhold Huber-Mörk[1]

Austrian Institute of Technology
Digital Safety and Security Department, Austria
{*doris.antensteiner.fl,svorad.stolc,reinhold.huber-moerk*}*@ait.ac.at*

Abstract
In this paper we deal with a combination of two state-of-the-art computational imaging approaches - (i) light fields and (ii) photometric stereo - in order to improve the quality of 3D reconstructions within a multi-line-scan framework. Computational imaging uses a redundant description of an image scene to reveal information which would not have been available via conventional imaging techniques. In the case of light fields the redundancy is achieved by observing the scene from many different angles, which allows capturing 3D shapes in areas with a prominent surface structure using stereo vision techniques. Contrarily, photometry makes use of multiple illuminations in order to capture local surface deviations without the necessity of any surface structure. As photometric surface reconstruction is very sensitive to fine surface details and light fields excel in capturing global shapes, naturally a more complete description can be achieved through a combination of both techniques. We present a compact hybrid photometric light field setup with relatively low costs and improved accuracy, which is therefore well suited for industrial inspection. A multi-line-scan camera is statically coupled with an illumination source to obtain light field data which is also comprised of photometric information. Novel algorithms have been developed to use this data for an improved 3D reconstruction, which exhibits large-scale accuracy as well as sensitivity to fine surface details.

1. Introduction

Traditional film cameras as well as digital cameras capture light rays and project images of the environment onto a 2D plane. Non-traditional approaches such as multi-camera arrays, plenoptic cameras or coded apertures capture a portion of the so called 4D light field and further subdivide this ray space with respect to position and orientation. Photometric stereo makes explicit use of directional variation of illumination. The presence of different lighting conditions, in cases where the light field is dynamically constructed over time, introduces a photometric variation into the data structure. The combination of light fields, describing the variation of image content over observational directions, and photometric approaches, describing the variation of image content depending on lighting directions, is a promising research direction.

In general, the combination of methods which are (i) locally precise but globally inaccurate with (ii) globally accurate methods, which are lacking local structure, was approached from different perspectives, e.g. combining shading with RGB-D [12], improving a depth map from time-of-flight with polarization cues [3], or using stereo vision with photometric stereo [5]. Methods used for the combination are either based on the fusion of normal vectors provided by different approaches, the fusion of depth measurements, or employing variational methods.

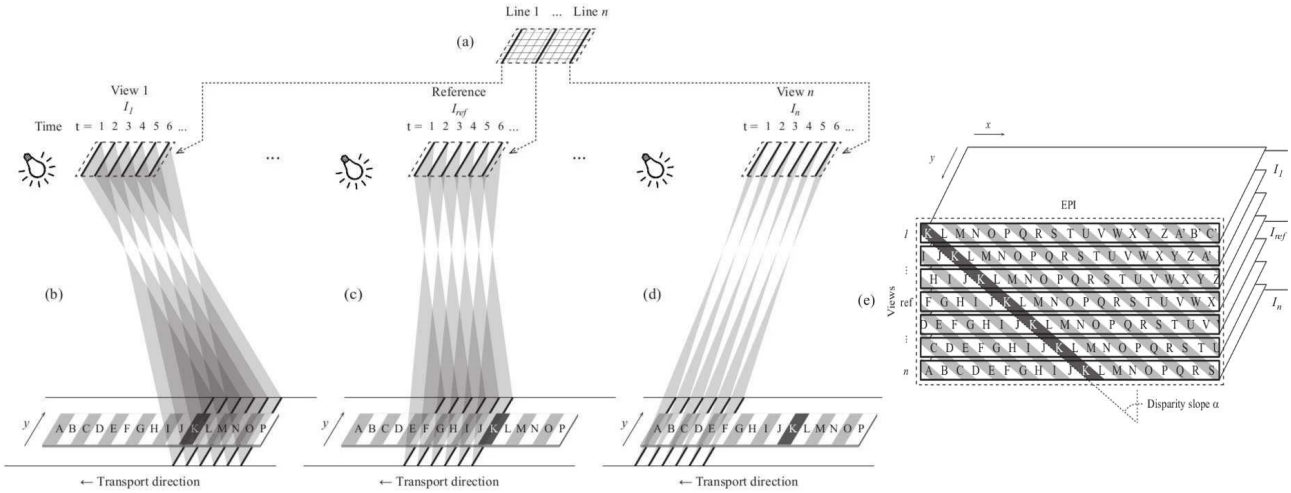


Figure 1: Multi-line-scan setup with directional lighting: a) multi-line sensor, views constructed over time, imaging the object area divided in uppercase letters: b) view 1, c) reference and top-down view, d) view n, e) EPI stack holding views denoting object lines by uppercase letters the disparity slope α.

The depth reconstruction from light field data is usually estimated through the epipolar plane image (EPI) data structure. EPIs were originally introduced for the estimation of structure from motion [1], but they also became a popular tool in light field processing [10],[4]. Kim et al. [4] use an easy criterion for ranking depth hypotheses, namely the best hypothesis is the one, for which as many radiance values as possible along the hypothesized slope in an EPI are similar enough to the radiance in the reference view. Venkataraman et al. [8] use pattern matching between different views, i.e. for a discrete number of hypothesized depths the sum of absolute differences (SAD) of radiances between different views is calculated. Wanner and Goldlücke [10] suggest a statistical approach to estimate the principal orientation of linear structures in EPIs via analysis of the structure tensor constructed locally in small EPI neighborhoods.

This paper is organized as follows. We describe the proposed setup in Sec. 2. In Sec. 3. we describe the fusion framework for light fields and photometric stereo. First results describing the work in progress is given in Sec. 4. In Sec. 5. we draw first conclusions and discuss further work.

2. Multi-Line-Scan Setup

Light fields provide 4-D information, consisting of two spatial and two directional dimensions. They can be captured e.g. by a multiple camera array [11], where each camera has a different viewing perspective of the scene, or by plenoptic cameras [6], which usually make use of a microlens array placed in front of the sensor plane to acquire angular reflectance information.

Our multi-line-scan framework [9] is a light field acquisition setup, where we use an area-scan sensor to observe the object under varying angles, while the object is transported in a defined direction over time. This setup works in real-time and in-line for industrial inspection setups. Fig. 1 illustrates how the light field data is obtained through multiple viewing angles on the moving object over time. Each sensor line observes the conveyor belt in a different viewing angle and captures a certain region. As the object moves under the observed sensor lines, see Fig. 1a, each sensor line captures every object region at distinct time instances, see Figs. 1b,c, and d. We represent the thereby captured light fields as light field image stacks, see Fig. 1e, in which each image is acquired from a slightly different

viewpoint along one direction.

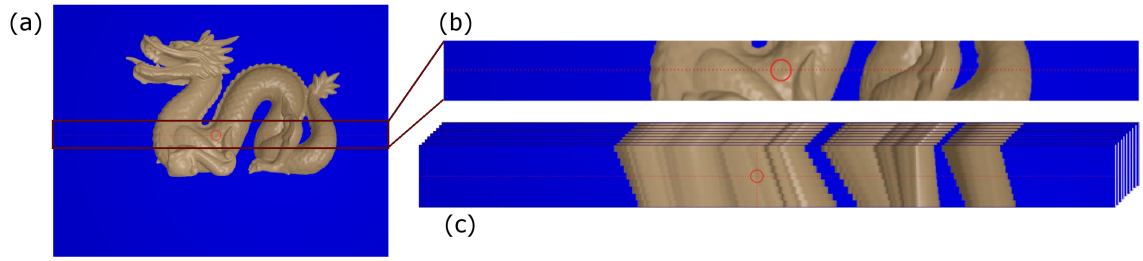


Figure 2: Dragon Stanford [13] scene image (a), zoomed in at the red dotted line (b), where EPI stacks with 9 different viewpoints are shown, each stack formed with a certain light direction (c).

We analyse the captured light fields in the EPI domain [1]. A cut through the light field stack shows linear slope structures, where the angle of the slope corresponds to the disparity and thereby the depth of the scene, as shown in Fig. 2. Each angle of a slope in the EPI stack corresponds to a defined distance between the camera and the object point. Photometric information is obtained by a static light source w.r.t. the sensor while the object is moving. As an object moves on the conveyor belt, the relationship between the illumination and the observation angle changes in a systematic way, so that the surface inclination in the transport direction can be estimated. This photometric information is used to estimate the surface normals of the object.

3. Combination of Light Fields and Photometric Stereo

Photometric stereo describes the surface variation w.r.t. the lighting direction. Reflections of the light on the objects' surface under different lighting orientations provide information about the surface normals at each object point. We combine the depth from light fields with fine surface structures as observed by photometric stereo, to gain an improved depth map of the scene. Figure 3 shows the depth estimation achieved using both light field and photometric stereo independently in a virtual test setup, where we simulated 81 camera viewpoints and 25 illumination angles.

3.1. Light Field Depth Estimation

Depth information of a scene can be retrieved by analyzing the slope angles in EPI stacks. An EPI slice of this stack is shown in Fig. 2, where each angle in the slice refers to a defined depth of a corresponding point in the scene. Using this data we gain a rough absolute depth estimation of the scene.

Analyzing the depth from EPI stacks can be seen as finding such an angle α^* for each point (x, y), where the difference between values at the sheared coordinates $(x(\alpha), y(\beta))$ in the light field L and the reference view I_0 is minimal [7].

$$\alpha^*(x, y) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{i=n} |I_i(x(\alpha), y(\beta)) - I_0(x, y)|$$

The use of a block-matching approach creates a higher robustness to both noise and non-Lambertian lighting conditions. As shown in Fig. 3b, light field data provides quite robust absolute depth estimation, but lacks precision in fine surface details.

(a) dragon (b) LF (c) PS

Figure 3: Light field (LF) and photometric stereo (PS) depth information.

3.2. Photometric Stereo Surface Normals

The surface appearance depends on the shape, reflectance, and illumination of the object. In our virtual test setup we arranged 25 light sources on a sphere around the object in order to capture several images from the same viewpoint under different illumination angles. Contrarily, in our real world setup different illumination angles are achieved by the movement of the object under the light source. We assume a Lambertian reflection model to describe the radiance. The pixel intensity vector $Iv = [i_1, ..., i_{25}]$ for each light source and pixel depends on the illumination vector $L = [L_1, ..., L_{25}]$, as well as on the estimated surface normal unit vector N to the surface and the surface albedo ρ.

$$Iv = \rho \cdot L \cdot N$$

Thereby, we solve the albedo and normal vectors with $\rho \cdot N = L^{-1} \cdot Iv$. The depth map is then integrated using the algorithm of Frankot and Chelappa [2].

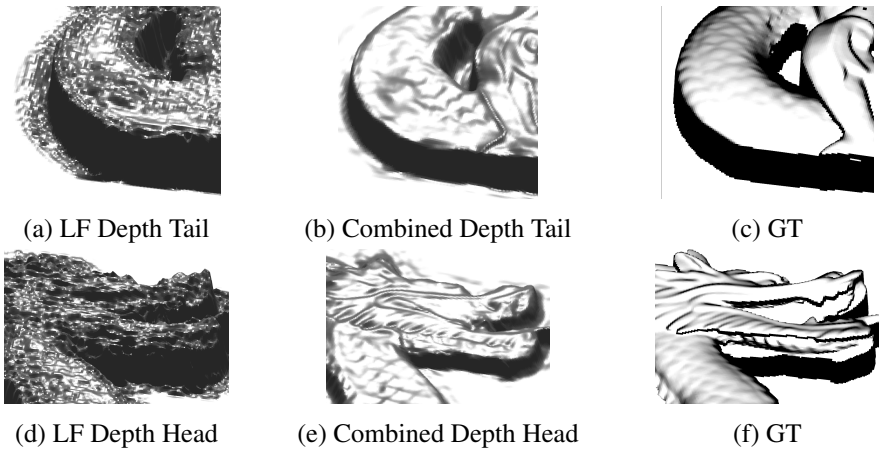
As shown in Fig. 3c, this photometric stereo approach results in fine depth measurements and a strong relative depth accuracy, while the absolute depth accuracy suffers from an accumulative offset. We use the benefits of both the photometric stereo and light field depth estimation to achieve an improved depth estimation result.

3.3. Combination

We refine the light field depth map, as shown in Fig. 3b, using high frequency photometric stereo depth information, as shown in Fig. 3c. Depth from light field yields reliable absolute depth measures, but suffers both from inaccurately estimated details in the structure and high frequency noise. Low frequencies in the light field depth map D_l are extracted using a bilateral smoothing filter f_l. High frequency components are taken from the photometric stereo depth map D_p, using a high-pass image filter f_h. Depth refinement is obtained by replacing high frequency information from the light field depth map by the according high frequencies in the photometric stereo depth map. Our final depth map D is thereby constructed as the linear combination of the low frequency components from the light field depth map and the high frequency components from the photometric stereo depth map, weighted by the factors λ_l and λ_p respectively.

$$D = \lambda_l \cdot D_l * f_l(u, v) + \lambda_p \cdot D_p * f_h(u, v)$$

Results are shown in Fig. 4, where 4a and 4d hold the depth data from light field images from both the head and the tail of the dragon object. The second column, see Figs. 4b and 4e, shows the photometrically refined depth map, using our combinational approach.



(a) LF Depth Tail (b) Combined Depth Tail (c) GT

(d) LF Depth Head (e) Combined Depth Head (f) GT

Figure 4: Combination of light field and photometric stereo in order to improve fine structures in the light field depth estimation. Both rows show a part of the tested object: (a),(d) original light field depth, (b),(e) improved depth by photometric stereo, (c),(f) ground truth depth map.

4. Experimental Results

We performed experiments on several coins, using the multi-line-scan setup, an example is shown in Fig. 5. Transporting the coin, shown in Fig. 5a, along the conveyor belt, we acquired light field and photometric stereo data. The depth information gained from light field data is shown in Fig. 5b. Surface normals are estimated from the same photometric light field data through the detection of a specular lobe in each image location. The specular lobe position corresponds with the local orientation of the surface. The image depth is calculated using the surface normals as described in Subsection 3.2. In Fig. 5c we obtained a refined solution by the combination of photometric stereo depth and light field depth, as described in Sec. 3.

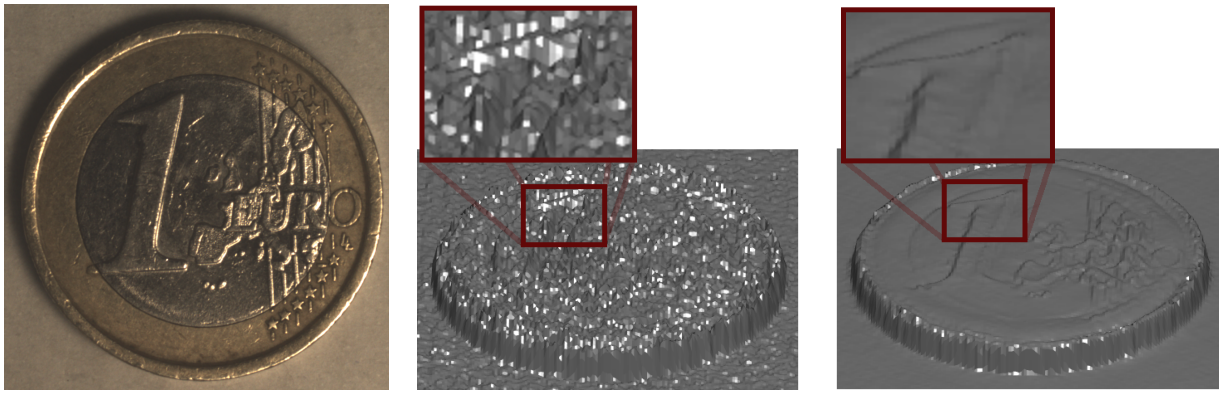
5. Conclusion and Discussion

We discussed the pros and cons of passive (light field) and active (photometric) stereo approaches for depth estimation on virtually generated data. A way to combine both approaches in order to achieve a more precise depth estimation was presented. We also showed initial results on our multi-line-scan setup, which is acquiring light field data and photometrically varying data at the same time, i.e. while observing the object under relative motion. Relative motion between the object and the acquisition device is a typical configuration in industrial vision systems, thus this setup fits well for such applications. Initial results were given and the suggested combination scheme was demonstrated. Further work will cover a more complete evaluation, as well as an improvement of the combination scheme.

References

[1] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolarplane image analysis: an approach to determining structure from motion. *Int. J. Comp. Vis.*, 1(1):7–55, 1987.

[2] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. Pat. Anal. and Mach. Intell.*, 10:439–451, 1988.



(a) Coin image (b) LF depth (c) Combined LF and PS

Figure 5: Coin image (a), acquired with our Multi-Line-Scan Setup, with an estimated depth both from light field and our improved result through combination with photometric stereo data.

[3] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar. Polarized 3D: High-quality depth sensing with polarization cues. In *Proc. Intl. Conf. Comp. Vis. (ICCV)*, 2015.

[4] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graphics*, 32(4):73:1–73:12, 2013.

[5] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM Trans. Graph.*, 24(3), August 2005.

[6] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. Technical Report CSTR 2005-02, Stanford University, April 2005.

[7] M. Tao, P. Srinivasa, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi. Depth from shading, defocus, and correspondence using light-field angular coherence. In *Proc. Comp. Vis. and Pat. Rec. (CVPR)*, June 2015.

[8] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar. PiCam: an ultra-thin high performance monolithic camera array. *ACM Trans. Graph.*, 32(5), 2013.

[9] S. Štolc, D. Soukup, B. Holländer, and R. Huber-Mörk. Depth and all-in-focus imaging by a multi-line-scan light-field camera. *J. of Electronic Imaging*, 23(5):053020, 2014.

[10] S. Wanner and B. Goldlücke. Globally consistent depth labeling of 4D light fields. In *Proc. of Comp. Vis. Pat. Rec. (CVPR)*, pages 41–48, 2012.

[11] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph.*, 24(3):765–776, July 2005.

[12] L. F. Yu, S. K. Yeung, Y. W. Tai, and S. Lin. Shading-based shape refinement of RGB-D images. In *Proc. of Comp. Vis. and Pat. Rec. (CVPR)*, pages 1415–1422, 2013.

[13] The Stanford 3D Scanning Repository, http://graphics.stanford.edu/data/3Dscanrep/.