https://doi.org/10.2352/ISSN.2470-1173.2017.9.IRIACV-268

High-Precision 3D Sensing with Hybrid Light Field & Photometric Stereo Approach in Multi-Line Scan Framework

Doris Antensteiner**, Svorad Štolc**, Kristián Valentín**, Bernhard Blaschitz**, Reinhold Huber-Mörk**, Thomas Pock**
* ***AIT Austrian Institute of Technology GmbH, Vision, Automation & Control, Vienna, Austria***
** ***Graz University of Technology, Institute for Computer Graphics and Vision, Graz, Austria***

Abstract

We present a hybrid multi-line scan approach which enables simultaneous acquisition of light field & photometric stereo data. While light fields capture mostly large-scale surface deviations and rely on visible surface structures, photometric stereo is primarily sensitive to fine surface deviations and does not rely on visible structures. The combination of both approaches yields a solid performance for a large variety of depths, ranging from macro- to microscopic scales. Contrary to traditional photometric stereo, that relies on a strobed illumination, our approach uses two constant light sources which, however, generate multiple illumination geometries in different portions of the camera's field of view. Our object is moving on a conveyor belt during the acquisition process. Due to our multi-line scan sensor the object is observed from several viewing angles. The object's movement is causing each object point to be illuminated under several illumination directions. Hence, during our acquisition process the object points are captured under all feasible viewing angles and lighting conditions. In our system, surface normals are derived making use of the Lambert's cosine law. However, due to the lack of illuminations spanning orthogonally to the transport direction, the surface normals can be inferred only in the transport direction. We present a variational approach for 3D depth reconstruction designed specifically for our hybrid setup that jointly takes into account the light field as well as photometric stereo depth cues and provides one globally consistent solution. Depth maps obtained by the proposed algorithm show both the large-scale accuracy as well as sensitivity to fine surface details.

Introduction

In recent years there has been a boom of 3D sensing techniques in the field of machine vision. Techniques based on the time of flight principle or stereo vision typically exhibit a solid performance on large scales but suffer from a lack of fine surface details. On the other hand, techniques such as photometric stereo or focus stacking provide vast local detail but lack global consistency. In an industrial environment, there are additional requirements of high speed and inline applicability, which renders many existing methods unsuited for those applications.

In our paper, we deal with a hybrid light field & photometric stereo machine vision approach specifically designed for industrial inline inspection applications that yields solid performance for a large variety of depths, ranging from macro- to microscopic scales. While light fields capture mostly large-scale surface deviations and ranging relies on visible surface structures, photometric stereo is mostly sensitive to fine surface deviations and does not require structured surfaces.

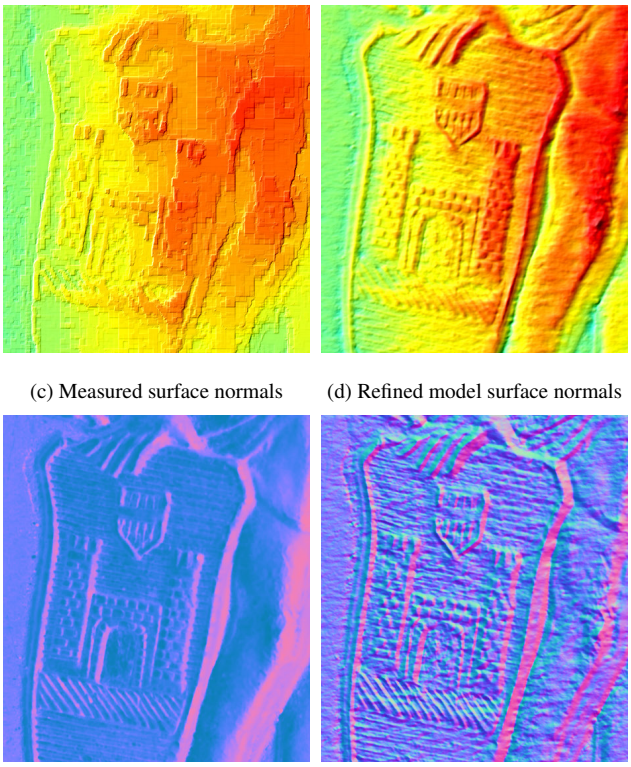

Figure 1: Close-up of a historical coin: (a) Initial depth map obtained from light field only by the CRF solver. (b) Final refined depth map obtained with the proposed joint light field & photometric stereo depth reconstruction algorithm. (c) Surface normals measured in the transport direction estimated from the acquired data making use of the refined depth model. (d) Surface normals derived from the refined depth model. Slight shading was applied to depth maps (first row) in order to increase readability of details.

Depth reconstruction from light field data is usually done making use of the epipolar plane image (EPI) structure, which was introduced in [1] for structure from motion analysis. The sheared EPI stack was used in [2] to calculate the depth by the radiance differences to a central image. A fine-to-coarse depth estimation approach using EPI stacks was introduced by [3], which allows sharper edges at depth discontinuities than the more traditional coarse-to-fine estimation while preserving homogeneous

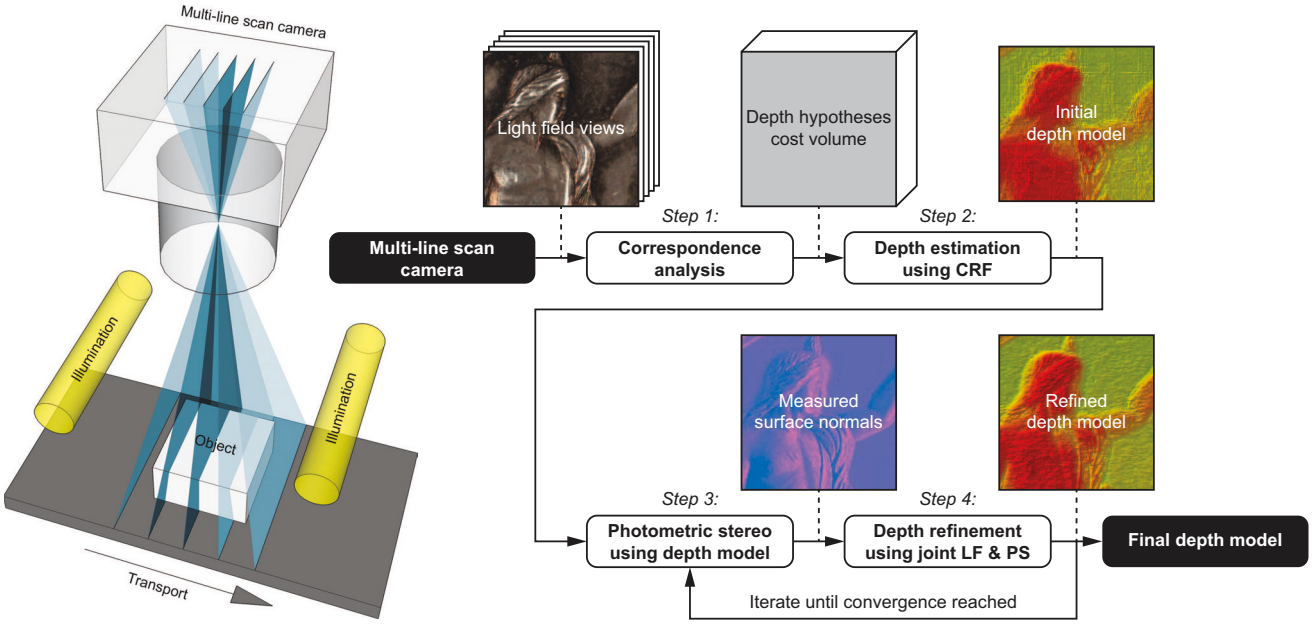

Figure 2: Left: Schematic of the AIT multi-line scan setup with a constant illumination. At each time step a set of lines is extracted from the sensor, then the object is moved (w.r.t. the camera) by exactly one line increment and another set of lines is extracted, and so on. Right: Data processing pipeline of the proposed joint light field & photometric stereo depth estimation algorithm.

depth regions. A structure tensor analysis of the EPI stack was employed by [4] for a fast and robust local disparity estimation.

Depth and surface normal information was previously combined in various ways. Range scanning data using stripe pattern projection was combined with photometric stereo from five fixed light sources in [5], taking low and high frequencies into account. A depth refinement method using photometric stereo was introduced by [6] which used RGB-D camera data to estimate the depth, lighting and albedo of the scene and minimized an energy function, optimizing the depth, smoothness, shading and temporal aliasing of a scene. Surface normals from polarization cues were used in [7] for depth refinement, where the depth surface normals are iteratively corrected using the polarization cues and a depth fidelity constraint preserves the global coordinate system at the normal-to-depth integration step. In [8] we introduced a fast high-pass / low-pass filter approach to combine light field data with photometric stereo in a multi-line scan setup. A gradient descent energy minimization approach in a multi-view light dome setup was investigated in [9].

The paper is organized as follows. We start with an overview of the state-of-the-art techniques combining light fields with photometric stereo. Then we describe our hybrid light field & photometric stereo setup implemented within the multi-line scan framework. Afterwards we introduce a joint depth estimation algorithm, which is specifically tailored to the multi-line scan system. The method's performance as well as optimal parametrization is analyzed making use of synthetic ground truth data. We provide also a number of real-world examples acquired with our multi-line scan camera prototype. Finally, we end with conclusions and future work.

Multi-line scan camera for joint light field & photometric stereo

For the acquisition of data comprising both the light field as well as the photometric stereo information, we propose to use a multi-line scan camera system as we described in [10]. Such a system consists of a multi-line scan camera, line illumination and a linear transport stage (see Fig. 2, left). In the prototype setup used in this paper, we employed a camera *Allied Vision BONITO CL-400C* equipped with a lens *Schneider-Kreuznach APO-COMPONON 4/45*, two light sources *Chromasens CORONA II* and a translation stage *Thorlabs LTS300/M*.

The way the light field and the photometric stereo data are obtained using the multi-line scan camera is explained in the following two sections.

Light field capture

During the acquisition process the object is moving under the camera on the linear stage. At each acquisition time t_i multiple equidistant lines are read out from the area scan sensor where each line contributes to a different view of the acquired object. In the time between two consecutive acquisitions t_i and t_{i+1} the object has to travel by a distance equivalent to one pixel in a defined working distance (typically the focal distance). That guarantees equivalent resolutions in both image dimensions (i.e. along the sensor lines as well as along the transport).

Unlike the usual 4D light field structure with two spatial and two directional dimensions, the described system produces a 3D light field structure with two spatial and one directional dimension. Such a light field can be represented as an image stack consisting of multiple object views (see Fig. 3) which allows EPI-based processing.

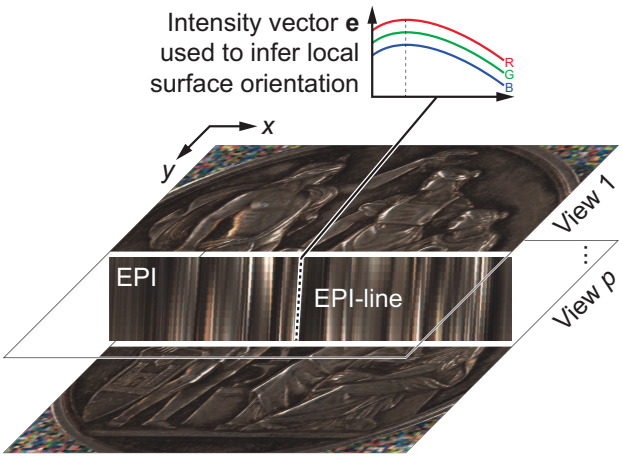

Figure 3: Visualization of a light field obtained by the multi-line scan camera with a constant illumination. In the middle, the epipolar plane image (EPI, i.e. a slice through the light field) is shown, containing multiple linear structures – EPI-lines, the slopes of which reflect local depths w.r.t. the camera. Above, the plot shows an intensity vector **e** extracted along one specific EPI-line (dashed) used to infer the local surface orientation.

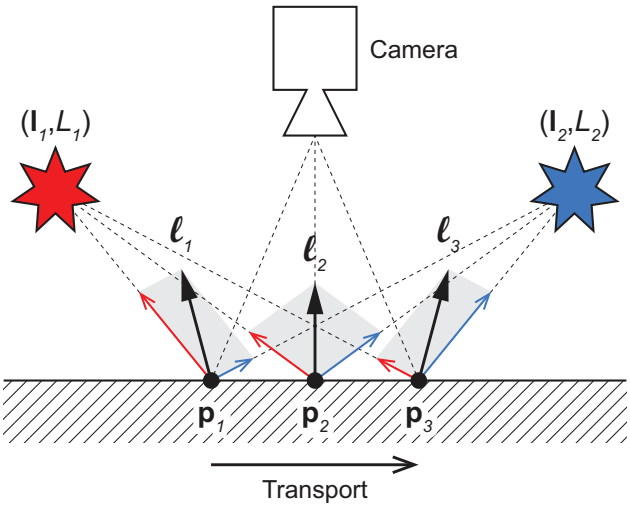

Figure 4: Assumed illumination model comprised of two constant line light sources at the positions $\mathbf{l}_1$ and $\mathbf{l}_2$, with the scalar intensities L_1 and L_2, respectively. Due to the inverse-square law, the integral of the two illuminations at the point $\mathbf{p}_i$ results in an effective illumination vector ℓ_i, that is different in each observed point (i.e. sensor line).

Photometric stereo capture

It is known that stereo vision techniques with a small baseline, such as most light field setups, are rather limited in depth accuracy on fine scales. Typical failure cases are areas with strong specular reflections or areas with little or no texture [4]. Nevertheless, given enough surface structure, they exhibit exceptional robustness and high accuracy on large scales. In order to obtain a competitive performance on both large as well as fine scales, we extended our multi-line scan light field system [10] by taking into account an additional photometric stereo cue, that is inherently comprised in this setup.

In a traditional photometric stereo, the still object is observed multiple times from one viewing perspective under different illumination conditions. This allows to derive local surface orientations (i.e. surface normals) from observed intensities making use of known illumination angles [11]. When assuming Lambertian reflectance, which is entirely valid only for matte materials, one can derive a simple cosine law for the determination of surface normals. Because the cosine law does not break down for processed materials that slightly violate this assumption, we employ this model for quite accurate surface normal estimates.

In contrast to traditional photometric stereo with multiple switched or strobed light sources, our approach uses two constant line light sources (see Fig. 2, left). In our setup, the light sources are located symmetrically around the optical axis in the transport direction in order to illuminate the observed area from two flat angles. As illustrated in Fig. 4, such an arrangement gives rise to different illumination configurations in every observed line. Since the employed line lights have very homogeneous emission along the sensor lines, all pixels in the same sensor line are illuminated almost equally and, therefore, share the same illumination parameters (i.e. the light direction and intensity). The downside of this illumination geometry is the lack of illuminations spanning orthogonally to the transport direction, resulting in a collinear set of illumination vectors. Consequently, the surface normals can be inferred only in the transport direction (i.e. x-dimension).

In this paper, we assume a simple illumination model based on the Lambertian assumption. Under this constraint, an effective illumination vector ℓ_i in an observed sensor line $\mathbf{p}_i$ is given as the sum of all elementary illumination vectors that contribute to that point (shown as thin red and blue arrows in Fig. 4):

$$\ell_i = \sum_{j=1}^{q} \frac{\mathbf{l}_j - \mathbf{p}_i}{|\mathbf{l}_j - \mathbf{p}_i|} \cdot \frac{L_j}{|\mathbf{l}_j - \mathbf{p}_i|^2}, \tag{1}$$

where $\mathbf{l}_j$ are the illumination positions and q stands for the number of light sources (in our case $q = 2$).

Due to the inverse-square law, the elementary light vectors are different in each observed line, which results in different effective illumination vectors as well. Due to the object's movement during the capture, each object point is eventually observed under every available illumination condition. Let $\mathbf{e} = [e_1 \dots e_p]^\top$ be the vector of intensities observed at the same object location under different illumination conditions and let $\mathbf{L} = [\ell_1 \dots \ell_p]^\top$ be the matrix of the respective effective illumination vectors, where p is the number of light field views. Then the corresponding surface normal vector $\mathbf{n}$ can be derived as follows:

$$\mathbf{n} = \mathbf{L}^+ \cdot \mathbf{e}, \tag{2}$$

where $\mathbf{L}^+$ stands for the pseudoinverse of the over-determined non-square illumination matrix $\mathbf{L}$.

This way it is possible to perform photometric stereo within the multi-line scan framework without the necessity of switching or strobing the illumination during the acquisition process. Hereinafter, the surface normals estimated by the described method will be referred to as *measured surface normals*, since they are assessed directly from the recorded data. On the other hand, normals derived from the reconstructed depth models will denoted as *model surface normals*.

As a result of parallax comprised in the light field, intensities associated with the same object location occur along an EPI-line, the slope of which depends on the absolute distance of that location from the camera (see Fig. 3). Therefore, with our approach the surface normal estimates are inherently linked with respective depth estimates. Hence, it is necessary to utilize a preliminary depth model to calculate surface normals, which can afterwards be used to improve the depth model, etc.

Hybrid light field & photometric stereo depth estimation algorithm

In this paper, we present a variational approach for depth reconstruction which jointly takes into account the light field as well as the photometric stereo depth cues and provides one global solution. Certain aspects of the method are specifically designed to work well with the multi-line scan setup. The proposed multi-step iterative algorithm is outlined in Fig. 2 (right). Individual steps of the algorithm are explained in detail in the following sections.

Step 1: Multi-view correspondence analysis

Starting from the compound light field & photometric stereo data obtained using the multi-line scan camera, we first perform a multi-view correspondence analysis in the EPI domain using the *census transform* (CT) image features. This method shows great robustness against brightness and contrast variations in different views, which happens quite often due to the presence of photometric effects in our data.

Let $\mathcal{X} = \{1 \dots m\}$ and $\mathcal{Y} = \{1 \dots n\}$ be sets of pixel indices in the x and y-dimension, respectively. Moreover, let $\mathcal{D} = \{d_z \in \mathbb{R} \mid z \in \mathcal{Z}\}$ be a linear set of disparity values, where $\mathcal{Z} = \{1 \dots k\}$ is a discrete set of disparity labels. Making use of a geometric camera calibration model, there is an isomorphic relationship between physical depths, disparity values and disparity labels. Hence, hereinafter these terms will be referred to as equivalent.

During the correspondence analysis a number of disparity hypotheses from $\mathcal{D}$ are tested in each pixel location from $(x,y) \in \mathcal{X} \times \mathcal{Y}$ which results in a cost volume $C \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}}$, where each value reflects the similarity of visual structures at the corresponding locations in the light field views. For the convenience, values in the cost volume are normalized to fit in the interval $[0,1]$.

In Fig. 2 (right), the depth hypotheses cost volume obtained in this step is depicted as a gray box between Step 1 and Step 2. For more details on the CT-based multi-view stereo matching within the multi-line scan framework, see [12].

Step 2: Initial depth estimation using CRF

In order to assess surface normals in Step 3, an initial approximate depth model must first be obtained. Given the precalculated hypothesis costs, we employ the discrete-continuous optimization algorithm based on conditional random fields (hereinafter referred to as the CRF algorithm, see [13]) to determine a quick yet accurate approximation of the global solution (i.e. globally consistent depth map), under the generalized first-order total variation (TV) prior.

Let $\mathcal{V} = \mathcal{X} \times \mathcal{Y}$ be a set of nodes, where each node $i \in \mathcal{V}$ corresponds to a pixel location. Moreover, let $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ be an edge set, where each $ij \in \mathcal{E}$ corresponds to an edge connecting the two pixel locations i and j. For efficiency reasons, we restrict the edge set to a 4-neighborhood, but the extension to a higher degree of connectivity is straight forward. To each pixel $i \in \mathcal{V}$ we associate a discrete disparity label $z_i \in \mathcal{Z}$.

Our goal is now to find an optimal discrete labeling $Z \in \mathcal{Z}^{\mathcal{V}}$ that minimizes the following conditional random field (CRF) energy:

$$\min_{Z} \sum_{i \in \mathcal{V}} f_i(z_i) + \sum_{ij \in \mathcal{E}} f_{ij}(z_i, z_j), \tag{3}$$

where $f_i(z_i)$ are the unary terms that are given by the CT matching costs for each depth hypothesis z_i and f_{ij} are the binary terms that apply a smoothness constraint to the optimal labeling. We use the generalized TV function from [13]. In other words, by minimizing the CRF energy, we try to find an optimal labeling Z that provides a trade off between minimizing the matching costs and minimizing the smoothness constraint.

The CRF minimization problem is combinatorial and hence NP-hard. However, as shown in [13] we can compute very good approximate solutions by means of the dual minorize maximize (DMM) algorithm. The idea of the DMM algorithm is to decompose the CRF energy into distinct chain problems for which the CRF energy can be minimized efficiently using dynamic programming. Then we consider the Lagrangian function obtained from introducing a vector of Lagrange multipliers that force the solutions of the distinct chain problems to agree in the optimum. The dual problem associated with the Lagrangian function is continuous, piecewise linear and concave but it should be noted that it only provides a lower bound to the original problem. The idea of the DMM algorithm is now to iteratively construct a sequence of minorants to the dual problem which can be efficiently maximized using dynamic programming. Once the dual problem is solved, the primal solution (and hence the depth map) is computed from the dual solution using again dynamic programming.

The complete algorithm is implemented on the GPU using the CUDA programming language. In Fig. 2 (right), the solution computed by the CRF solver is depicted as an initial depth model between Step 2 and Step 3. Examples of the CRF solutions can be seen in Fig. 1 (a) and Fig. 9 (a).

Step 3: Photometric stereo using previously assessed depth model

As soon as a previous discrete or continuous disparity labeling $Z \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$ is available either from Step 2 or Step 4, it can be used to extract the intensity vectors $\mathbf{e}$ in each pixel location along the corresponding EPI-lines by the provided disparity model. Subsequently these vectors are used to generate the surface normal field $N \in [\mathbb{R}_x, \mathbb{R}_y, \mathbb{R}_z]^{\mathcal{X} \times \mathcal{Y}}$ by applying Eq. (2) in all pixel locations.

For convenience in Step 4, the surface normals are expressed as disparity gradient field $G \in [\mathbb{R}_x, \mathbb{R}_y]^{\mathcal{X} \times \mathcal{Y}}$ that can be calculated from N as follows:

$$G = \left[-N_x / N_z \cdot g_x,\ -N_y / N_z \cdot g_y \right], \tag{4}$$

where g_x and g_y are scalar constants that are used to account for conversion between real-world units of surface normals and internal disparity units considered in Step 4. Their values are given by the system geometry and can be assessed through calibration.

Due to the lack of the photometric stereo evidence orthogonally to the transport direction, values in G_y (i.e. gradients in y-dimension) are always estimated to be zero. Hence, the first-order TV smoothness prior is implicitly applied in y-dimension.

In Fig. 2 (right), the measured surface normal map obtained in this step is depicted as a pinkish image between Step 3 and Step 4. Examples of the measured surface normal maps can be seen in Fig. 1 (c) and Fig. 9 (c).

Step 4: Depth refinement using joint light field & photometric stereo

The depth refinement algorithm performed in this step is formulated as a solution for the optimization problem, which is a natural extension of the generalized TV regularization, that takes advantage of the provided surface normals.

Let C be the hypothesis cost volume calculated in Step 1 and G is the disparity gradients field from previous Step 3. The goal of this step is to find a refined continuous disparity labeling $Z \in \mathbb{R}^{\mathcal{X}\times\mathcal{Y}}$ by solving the following continuous energy minimization problem:

$$\begin{aligned}\min_{Z} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} & C(x,y,Z(x,y))+ \\ & \lambda_x/2 \cdot r\left[\nabla_x Z(x,y) - G_x(x,y)\right]+ \\ & \lambda_y/2 \cdot r\left[\nabla_y Z(x,y) - G_y(x,y)\right].\end{aligned} \tag{5}$$

Regularization parameters λ_x and λ_y separately control the influence of the respective x and y gradients over the data term C. The function $r[\cdot]$ stands for a nonlinear penalty function, in our case it is the *truncated quadratic* function of the following form:

$$r[x] = \min\left[\left(x/a\right)^2,\ 1\right], \tag{6}$$

where a is the parameter governing the extent of truncation.

Operating on a discrete image domain, the energy term from Eq. (5) can be reformulated making use of four intermediate differences:

$$\begin{aligned}\min_{Z} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} & C(x,y,Z(x,y))+ \\ & \lambda_x/4 \cdot r\left[Z(x+1,y) - G_x(x+1,y) - Z(x,y)\right]+ \\ & \lambda_x/4 \cdot r\left[Z(x-1,y) + G_x(x-1,y) - Z(x,y)\right]+ \\ & \lambda_y/4 \cdot r\left[Z(x,y+1) - G_y(x,y+1) - Z(x,y)\right]+ \\ & \lambda_y/4 \cdot r\left[Z(x,y-1) + G_y(x,y-1) - Z(x,y)\right].\end{aligned} \tag{7}$$

To solve this optimization problem, we propose to use a primal block-coordinate descent algorithm [14] initializing with the CRF solution from Step 2. The algorithm operates on a discrete set of disparity labels $z \in \mathcal{Z}$ with a subsequent sub-label refinement to approximate the continuous solution. The initial solution provided by the CRF is already quite accurate, so it is unlikely to get stuck in local minima and therefore unnecessary to employ an elaborate optimization scheme such as the primal-dual approach. In order to ensure convergence, the algorithm uses an alternating "checkerboard" update pattern Ω_i, which defines the set of pixel coordinates that are updated simultaneously in i-th refinement iteration:

$$\Omega_i = \{(x,y) \in \mathcal{X}\times\mathcal{Y} \mid \text{if } x+y+i \text{ is even}\}. \tag{8}$$

Let Z_0 be either the approximate CRF solution obtained in Step 2 or an intermediate refined solution from previous Step 4. Then the solution is iteratively refined further according to the following update rule:

$$Z_{i+1}(x,y) = \begin{cases} \overline{\underset{z\in\mathcal{Z}}{\arg\min}} & C(x,y,z)+ \\ & \lambda_x/4 \cdot r\left[Z_i(x+1,y) - G_x(x+1,y) - z\right]+ \\ & \lambda_x/4 \cdot r\left[Z_i(x-1,y) + G_x(x-1,y) - z\right]+ \\ & \lambda_y/4 \cdot r\left[Z_i(x,y+1) - G_y(x,y+1) - z\right]+ \\ & \lambda_y/4 \cdot r\left[Z_i(x,y-1) + G_y(x,y-1) - z\right] \\ Z_i(x,y) & \text{otherwise.} \end{cases} \quad \forall (x,y)\in\Omega_{i+1}, \tag{9}$$

The term $\overline{\arg\min}$ refers to $\arg\min$ extended by the sub-label refinement. The refined version accepts the argument of the minimum of a parabola fitted into three values around the discrete energy minimum.

To facilitate the sub-label refinement with the right parametrization of the penalty function, we chose $a = 1.5$ in Eq. (6) which is the smallest value that produces at least three values in the quadratic non-truncated section around the minimum of the penalty function. That prevents formation of unwelcome discretization effects in the refined solution.

Regarding the number of refinement iterations performed in Step 4, we found it efficient to run at least 10 refinement steps using the same gradient field G assessed in previous Step 3. Afterwards, the algorithm either returns back to Step 3 and starts a new epoch or terminates if a sufficient number of epochs have been performed (in our case 50 epochs).

In Fig. 2 (right), the refined depth model obtained in this step is depicted as an image after Step 4. Examples of the depth models can be seen in Fig. 1 (b) and Fig. 9 (b).

Geometric camera calibration

All described algorithmic steps assume a calibrated light field system, requiring a line-scan calibration such as the one addressed in [15]. However, as our multi-line scan sensor is a repurposed area scan sensor, we can use a standard area scan calibration approach [16] to determine distortion, intrinsic and extrinsic parameters for the camera setup. As it turns out, perspective and lens distortions are negligible in our current setup, mostly because we use a high-end optical system.

Results

We tested our algorithm on a synthetically rendered scene from a multi-line scan setup simulated in Blender [17] as well as on real-world data acquired with our industrial multi-line scan prototype. Synthetic data allows a ground truth comparison with floating point accuracy and an optimal choice of regularization parameters. With the real-world setup we show the industrial applicability and performance.

Performance evaluation using synthetic data

Using our Blender multi-line scan simulation (see Fig. 5) we simulated the multi-line scan acquisitions process and generated ground truth depth data and surface normals. Two arrays of point lights simulate two line illumination sources and our surface object is moving in the transport direction under a multi-line scan

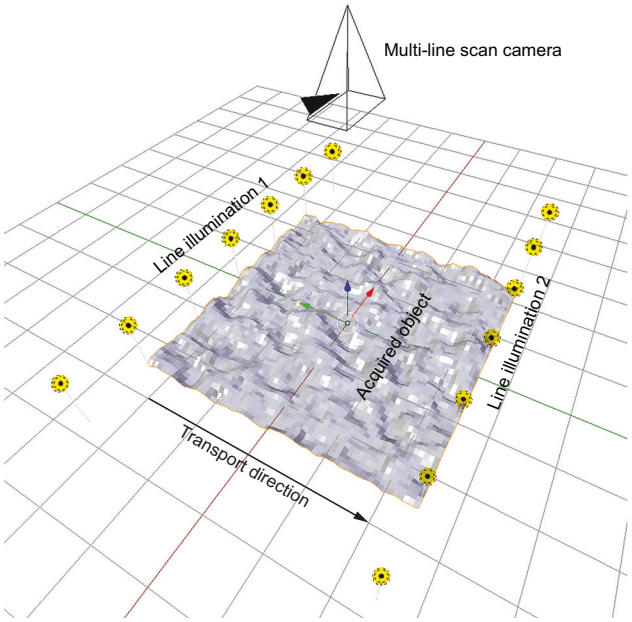

Figure 5: Blender [17] model of the multi-line scan system used for generating ground truth data.

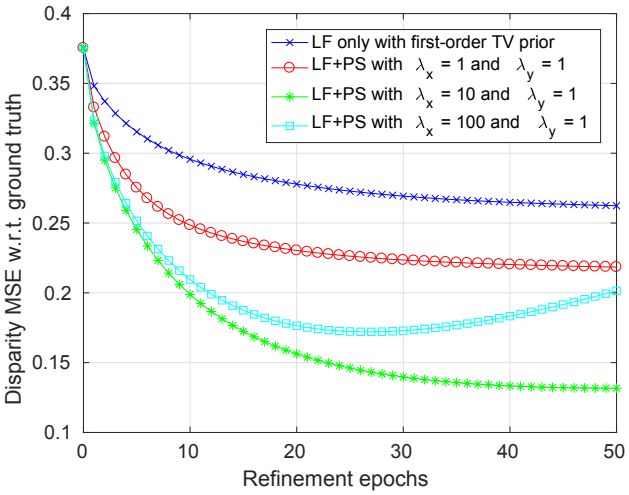

Figure 6: Mean squared error (MSE) of the disparity w.r.t. the ground truth during 50 refinement epochs of the proposed iterative algorithm. Each epoch consisted of 10 iterations of the energy minimization. Four different regularization configurations are shown: LF only with the first-order TV prior (crosses), and LF+PS with $\lambda_y = 1$ and $\lambda_x = 1$ (circles), $\lambda_x = 10$ (stars), and $\lambda_x = 100$ (squares), respectively.

camera. This setup was used for exhaustive tests with four different regularization configurations. The results are presented in Fig. 6 and the corresponding depth reconstructions at epoch 50 are shown in Fig. 7. We achieve a significant improvement in estimation in all configurations of joint light field and photometric stereo (LF+PS), compared with light field only (LF only). The best performance is achieved with the combined LF+PS method with $\lambda_x = 10$ and $\lambda_y = 1$, which exhibits 33% improvement of MSE already after 10 refinement epochs and 44% improvement after 20 epochs. A performance drop is observed for $\lambda_x = 100$ after 26 epochs, where the strong photometric influence drags the solution further away from the ground truth. This divergent effects is caused by the global bias of photometric stereo.

The improvement in accuracy is demonstrated also in Fig. 8 (a) by correlation in ∇X and ∇Y between the initial depth estimate by provided the CRF and the ground truth. In this case, the Pearson correlation coefficient reaches a value of 0.4861 for ∇X and 0.5431 for ∇Y. On the other hand, the ∇X and ∇Y values estimated from the refined depth model after running 50 epochs of our refinement algorithm show a significant improvement in accuracy with correlation coefficients of 0.9386 and 0.8105, respectively. Note that ∇Y improves even though there is no direct measurement of this property. It is thanks to joint use of light field, partial photometric stereo in x-dimension and smoothenss assumption in y-dimension. Limitations of this graceful behavior are observable in Fig. 9 (b, bottom row), where some long horizontal PCB tracks are not entirely recognizable.

Real-world examples

Using our industrial multi-line scan setup we demonstrate the real-world performance of our refinement algorithm. In Fig. 1, qualitative results are shown for close-up of a historical coin. The refined depth model obtained by our LF+PS method is clearly superior to the initial CRF model w.r.t. the amount of detail. The separate bricks in the castle and individual fingers on the hand are well visible in the refined depth contrary to the CRF solution.

In contrast to the refined model surface normals, the measured surface normals show more artifacts (e.g. ghosting around the coins edges) and lack horizontal edges.

Further examples of 2 Euro coin, a printed circuit board (PCB) with components and a bare PCB are shown in Fig. 9. In all presented cases, the refined depth (2nd column) yields a vast amount of detail (e.g. more readable 3D text, better defined ICs, clearly visible PCB tracks) compared with the initial model, despite a large diversity of material types present in these objects.

Conclusions

In this paper, we discussed a approach to acquisition and computational combination of two state-of-the-art methods — light fields and photometric stereo – in a machine vision setup suitable for industrial applications. We have shown how the multi-line scan camera can be used for the acquisition of data comprising both the light field as well as the photometric stereo information. Moreover we propose an algorithmic framework for processing the obtained data in order to provide competitive 3D reconstruction results.

Based on synthetic data we were able to show a significant decrease in MSE of the estimated disparity w.r.t. the ground truth on the order of magnitude of 50%. Remarkably, the reconstruction is also improved orthogonal to the transport direction despite of the lack of gradient information in this direction. Experiments with real-world objects such as coins and PCBs showed that faithful reproductions of fine details and consistent global reconstructions are possible with the proposed method.

In this paper, we have partly disregarded calibration issues for the camera (i.e. stereo calibration) as well as the illumination (i.e. photometric calibration). Instead we worked with assumed models derived from a good knowledge of the system. Therefore,

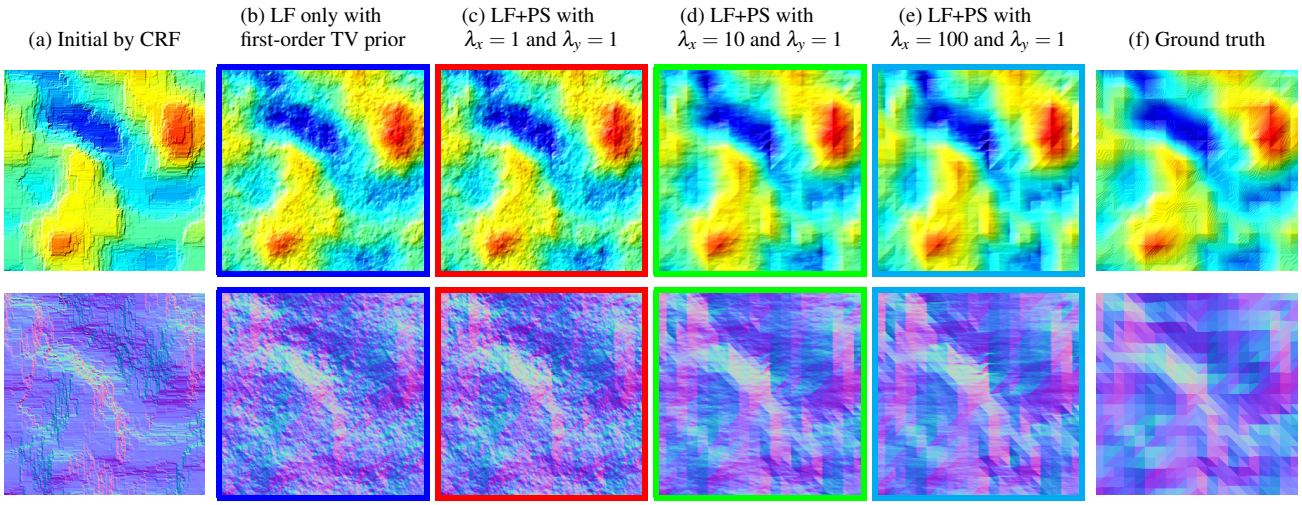

Figure 7: Depth models (first row) and corresponding surface normal maps derived from the models (second row) in the ground truth experiment: (a) initial by CRF, (b-e) after 50 epochs of the LF+PS refinement algorithm with different regularization configurations, and (f) the ground truth. Four refinement results (b-e) correspond with end points of the four curves in Fig. 6. Slight shading was applied to depth maps (first row) in order to increase readability of details.

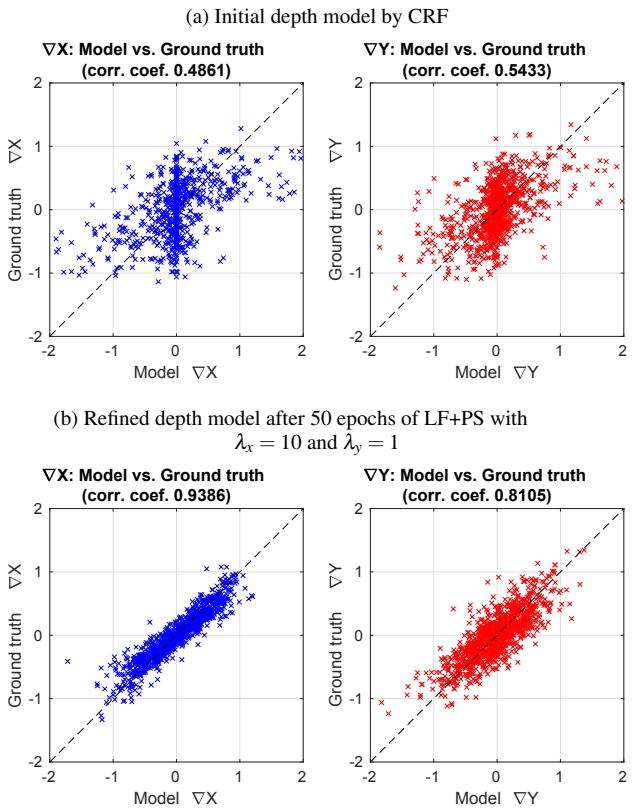

Figure 8: Correlation between ∇X (left) and ∇Y (right) derived from the reconstructed model vs. ground truth (a) before and (b) after running the proposed depth refinement algorithm. Note that despite the photometric stereo evidence in in x-dimension the correlation of ∇Y improves throughout the refinement process.

future work will include an implementation of the multi-line scan geometric calibration tailored to our system. To allow for more complex illumination geometries as well as handling of a broader range of material types, we intend to look into machine learning approaches for learning photometric models rather than relying on any model assumptions.

Acknowledgments

This work is supported by the research initiative "Mobile Vision" with funding from the Austrian Federal Ministry of Science, Research and Economy and the AIT Austrian Institute of Technology GmbH.

References

[1] Robert C Bolles, H Harlyn Baker, and David H Marimont. Epipolar-plane image analysis: an approach to determining structure from motion. *International Journal of Computer Vision*, 1(1):7–55, 1987.

[2] Michael W Tao and Pratul P Srinivasan. Depth from shading, defocus, and correspondence using light-field angular coherence. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[3] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Transactions on Graphics*, 32(4):1, 2013.

[4] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 36(3):606–619, 2014.

[5] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics*, 24(3):536, 2005.

[6] Chenglei Wu, Michael Zollhöfer, Matthias Nießner, Marc Stamminger, Shahram Izadi, and Christian Theobalt. Real-time shading-based refinement for consumer depth cameras. *Proceedings of ACM SIGGRAPH Asia*, 33:3, 2014.

[7] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3D: high-quality depth sensing with polarization cues. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 3370–3378, 2015.

[8] Doris Antensteiner, Svorad Štolc, and Reinhold Huber-Mörk. Depth estimation using light fields and photometric stereo with a multi-line-scan framework. In *Proceedings of Austrian Association for Pattern Recognition Workshop (OAGM)*, 2016.

[9] Doris Antensteiner, Svorad Štolc, and Reinhold Huber-Mörk. Depth estimation with light field and photometric stereo data using energy minimization. In *Proceedings of 21st IberoAmerican Congress on Pattern Recognition (CIAPR)*. Springer Lecture Notes in Computer Science, 2016.

[10] Svorad Štolc, Daniel Soukup, Branislav Holländer, and Reinhold Huber-Mörk. Depth and all-in-focus imaging by a multi-line-scan light-field camera. *Journal of Electronic Imaging*, 23(5):053020, 2014.

[11] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):191139–191139, 1980.

[12] Kristián Valentín, Svorad Štolc, and Reinhold Huber-Mörk. Improved cost computation with local binary features in a multi-view block matching framework. In *Proceedings of 10th International Conference on Measurement*, pages 21–24, 2015.

[13] Alexander Shekhovtsov, Christian Reinbacher, Gottfried Graber, and Thomas Pock. Solving dense image matching in real-time using discrete-continuous optimization. In *Proceedings of 21st Computer Vision Winter Workshop (CVWW)*, page 13, 2016.

[14] Stephen J Wright. Coordinate descent algorithms. *Mathematical Programming*, 151(1):3–34, 2015.

[15] Rajiv Gupta and Richard I Hartley. Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(9):963–975, 1997.

[16] David Ferstl, Christian Reinbacher, Gernot Riegler, Matthias Rüther, and Horst Bischof. Learning depth calibration of time-of-flight cameras. In *Proceedings of British Machine Vision Conference (BMVC)*, 2015.

[17] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Blender Institute, Amsterdam, 2016.

Author Biography

Doris Antensteiner is a PhD student at AIT, Vienna in the field of computational imaging and computer vision. She received her master's degree with distinction at the Technical University of Vienna in 2011 in the field of Computer Science. After that she worked at Kapsch, Vienna at the R&D unit "Video and Sensor" in the field of computer vision until she joined the AIT in 2015.

Svorad Štolc is a scientist AIT, Vienna. In 2002, he earned his master's degree in Computer Science from Comenius University, Bratislava and, in 2009, PhD degree in Bionics and Biomechanics from Technical University, Košice and Slovak Academy of Sciences, Bratislava. He is a (co)author of more than 50 peer-reviewed scientific papers and holds a number of patents in machine vision. His main research areas are computational imaging and machine vision with a focus on industrial inspection and document security.

Kristián Valentín received his PhD in Computer Science from Comenius University, Bratislava in 2015. Since 2014, he worked at AIT, Vienna in the field of computational imaging and computer vision. In 2017, he joined Photoneo, Bratislava to work on 3D scanners.

Bernhard Blaschitz earned his master's degree in Mathematics from the University of Vienna in 2008 and a PhD degree in Applied Geometry from Technical University of Vienna in 2014. He joined AIT in 2015 where he works as a scientist.

Reinhold Huber-Mörk received his PhD in computer science from the University of Salzburg in 1999. Since then he worked at the Aerosensing GmbH, Oberpfaffenhofen in remote sensing image analysis, at the Advanced Computer Vision GmbH, Vienna in computer vision and in 2006 he joined AIT, Vienna, where he is currently a senior scientist in the field of machine vision.

Thomas Pock received his MSc (1998-2004) and PhD (2005-2008) in Computer Engineering (Telematik) from Graz University of Technology. After a Post-doc position at the University of Bonn, he moved back to Graz University of Technology where he has been an Assistant Professor at the Institute for Computer Graphics and Vision. In 2013 he received the START price of the Austrian Science Fund (FWF) and the German Pattern recognition award of the German association for pattern recognition (DAGM) and in 2014, he received an starting grant from the European Research Council (ERC). Since June 2014, he is a Professor of Computer Science at Graz University of Technology (AIT Stiftungsprofessur "Mobile Computer Vision") and a principal scientist at the Center for Vision, Automation and Control at the Austrian Institute of Technology (AIT). The focus of his research is the development of mathematical models for computer vision and image processing in mobile scenarios as well as the development of efficient algorithms to compute these models.

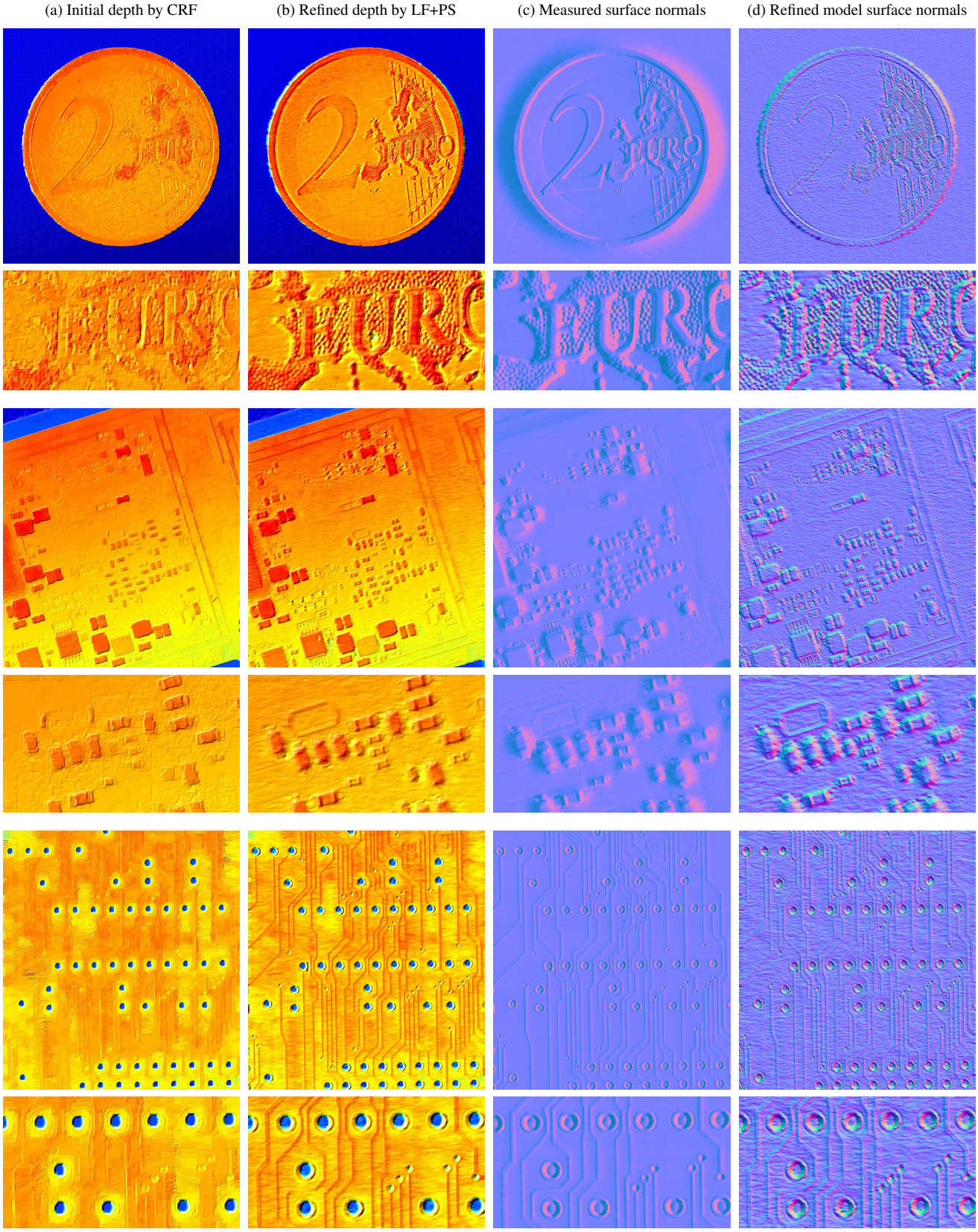

Figure 9: Additional real-world examples of refined depth models obtained with the proposed method: 2 Euro coin (top), printed circuit board (PCB) with components (middle), and bare PCB (bottom). (a) Initial depth map obtained from light field only by the CRF solver. (b) Final refined depth map obtained with the proposed joint light field & photometric stereo depth reconstruction algorithm. (c) Surface normals measured in the transport direction estimated from the acquired data making use of the refined depth model. (d) Surface normals derived from the refined depth model. Slight shading was applied to depth maps (first two columns) in order to increase readability of details.