 DOI: 10.3217/978-3-85125-603-1-08

Multi-camera Array Calibration for Light Field Depth Estimation

Bernhard Blaschitz[1], Doris Antensteiner[1] and Svorad Štolc[1]

***Abstract*— At the core of stereo methods for depth estimation and 3D reconstruction lies geometric calibration, i.e. the determination of intrinsic and extrinsic camera parameters and consecutive image rectification, such that the epipolar constraints are met in all views. In this spotlight paper, we present a multi-camera array calibration that fulfills the requirements for 3D reconstruction. The method is based on an optimization procedure that minimizes the reprojection error. We used it to calibrate the Xapt Eye-sect XA camera array with 4x4 camera modules equipped with identical wide-angle lenses. For this particular setup, we analyzed the algorithm's precision step by step, from initial pairwise multi-view stereo calibration to final bundle adjustment, to assess influence of each individual step. The conducted quantitative analysis based on the reprojection error revealed superiority of the bundle adjustment over all other considered intermediate steps yielding accuracy as much as 33x higher than the initial pairwise method. In order to demonstrate real-world performance of the calibrated camera array, we present a number of acquisitions of different physical objects along with estimated disparity maps and corresponding texture images generated by a light field multi-view stereo algorithm.**

I. Introduction

In recent years, there has been a boom of commercially available stereo and multi-camera systems for both consumer as well as industrial applications. Geometries of existing multi-camera systems are very diverse: from matrix cameras such as Xapt Eye-sect XA used in this paper, through plenoptic cameras such as Lytro or Raytrix, to unstructured multi-camera systems that make use of multiple free camera modules. Capturing multiple views of a scene by multi-camera systems is often interpreted as *light field*, which is the 4D radiance function of 2D position and 2D direction of each light ray propagating thorough space in regions free from occluders [1].

The steady improvement of stereo matching algorithms creates a high need for automated tools for calibrating such light field systems, consecutively allowing highly accurate depth estimation and 3D reconstruction.

The basis of multi view calibration is the geometric calibration, i.e. the determination of intrinsic and extrinsic camera parameters [2], as well as multi-view stereo and bundle adjustment [3] and image rectification.

In this case study, we present a multi-camera array calibration pipeline that fulfills the requirements for 3D reconstruction, such as epipolar constraints. The method is based on an optimization procedure which minimizes the reprojection error.

[1]all three are with AIT Austrian Institute of Technology, Giefingasse 4, 1210 Vienna, Austria firstname.lastname@ait.ac.at

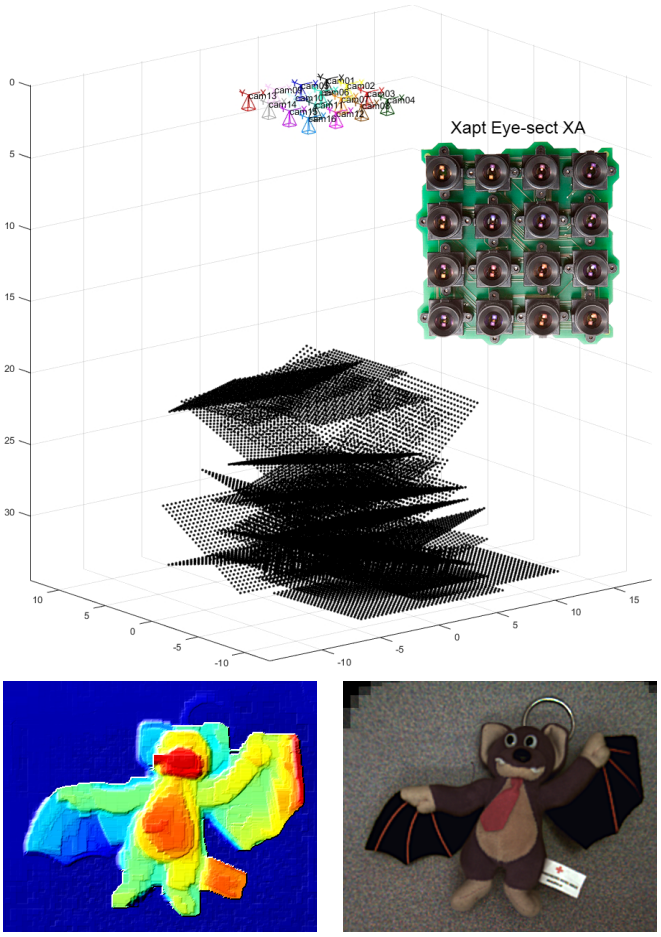

Fig. 1. *Top:* Estimated positions of the 16 camera modules of Xapt Eye-sect XA as well as estimated poses of the presented calibration patterns as a result of the proposed optimization procedure minimizing reprojection errors. Note that each camera module has a resolution of just 480×480 pixels. *Bottom:* Example of a depth model (left: disparity map; right: texture image) obtained by a light field multi-view stereo algorithm making use of the calibrated system. See also Figs. 4 and 5 for further examples of reconstruction from the same setup.

For a multi-view system, which is positioned in an unstructured manner and thus results in a irregular light field, it is favorable to implement a generic multi-view matching scheme. In this paper we considered an algorithm inspired by [4] and [5], extended by a real-time discrete-continuous optimizer [6] for a globally consistent depth map under the generalized first-order *total variation* (TV) prior.

In Sec. II we describe our multi-view calibration pipeline, which improves existing methods. In this case study, the calibration of the Xapt Eye-sect XA camera array, its accuracy as well as a number of depth reconstructions generated by a

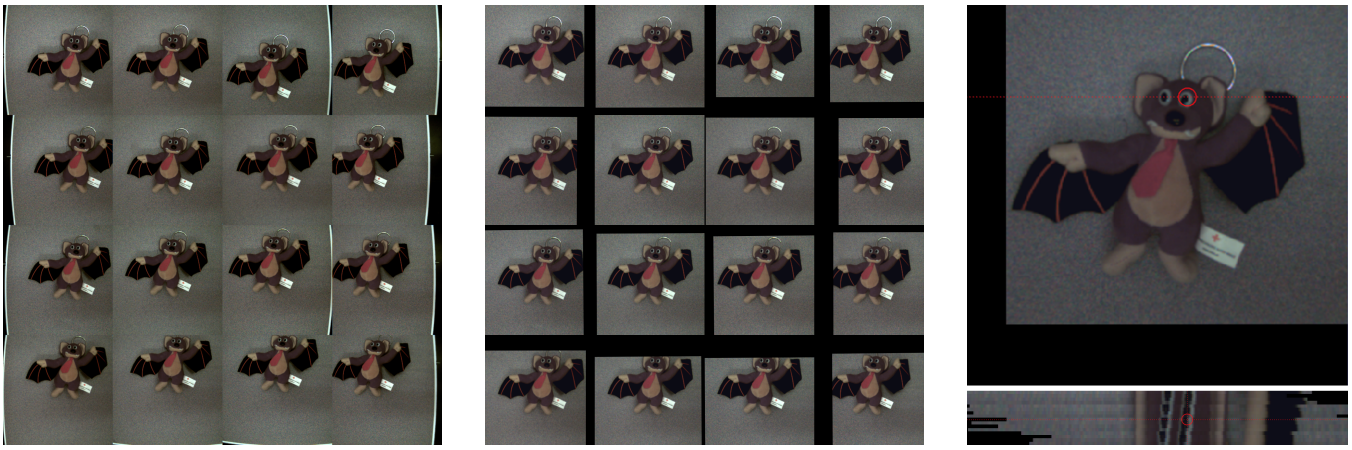

Fig. 2. *Left:* The input image taken by the Xapt Eye-sect XA camera. *Middle:* The undistorted rectified image. Note that all edges of the chessboard are bent in the left image (with distortion) and straight unbent (without distortion) in the middle image. *Right:* The rectified view of Camera 9 and the associated horizontal EPI image; the corresponding 3d reconstruction is in Fig. 1.

light field multi-view stereo algorithm are shown in Sec. III. Finally, in Sec. IV we conclude this study.

II. Calibration Method

We present a methodology, which was implemented in Matlab and relies on the *Complete Camera Calibration Toolbox for Matlab* [8], mainly for the intrinsic calibration. Our contribution improves over the prior art in the following points:

- It makes use of a high precision calibration target and accompanying algorithms [7], which improve the accuracy of automated pattern detection (see Fig. 3) and is stable to defocusing and sensitive to mirroring.
- It computes a true multi-view calibration instead of a pairwise stereo calibration. For this we use *bundle adjustment* [3], which optimizes intrinsic and extrinsic camera parameters by minimizing the overall reprojection error (see Sec. II-A). It can also cope with patterns that are not visible in all cameras.
- It allows image rectification suitable for light field processing, such as depth measurement and 3D reconstruction (see Sec. II-B).

A. Optimizing camera parameters

The notation complies with the camera model introduced in *OpenCV Toolbox* [9] and builds on the toolbox from [8].

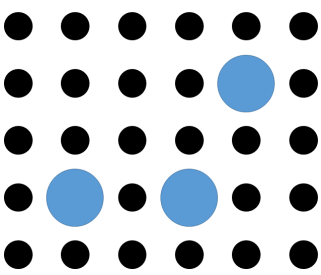

Fig. 3. We use an improved central element [7] for the calibration target, which has the advantage that only three dots in the center have to be visible in order to recognize the pattern, with a high robustness w.r.t. defocusing.

The intrinsic camera model has 10 degrees of freedom: two focal lengths f_x, f_y, two principal point coordinates c_x, c_y, camera skew α, three radial distortion parameters k_1, k_2, k_3 and two tangential distortion coefficients p_1, p_2, which comprise the distortion parameters $d = (k_1, k_2, p_1, p_2, k_3)$. Furthermore, there are 6 degrees of freedom for extrinsic camera parameters T, which comprise the position and rotation of the camera in a global coordinate system.

The bundle adjustment [3] is a non-linear method for refining extrinsic and intrinsic camera parameters, as well as the structure of the scene. It is characterized by minimizing the reprojection error by a standard least-squares approach

$$E(\mathbf{C},\mathbf{X}) = \sum_{i=1}^{n}\sum_{j=1}^{m} dist(x_{ij}, C_i(X_j)))^2, \qquad (1)$$

where $C_i(X_j) = C(H_i, T_i, d_i, X_j)$ is the *reprojected point*, i.e. the image of a point $X_j \in \mathbb{R}^3$ as observed by the i-th camera. Furthermore, x_{ij} is the corresponding detected point of the calibration pattern and $dist(x_{ij}, C_i)$ is the point's reprojection error.

We initialize the minimization with a single view calibration, choose one camera as the central view and initialize the other cameras' extrinsic parameters T by factoring out the average difference in pose of the detected calibration pattern. This is inspired by the initialization of the stereo calibration in [8], hence the assigned designation *pairwise* in Tab. I.

The quadratic objective function of Eq. 1 is minimized with a standard least squares solver. To avoid getting stuck in a local minimum, due to many degrees of freedom, an outer iteration for different optimization phases keeps certain parameters fixed.

For the phase *patterns* in Tab. I, only the positions and rotations of calibration patterns are optimized and all camera specific parameters remain unchanged. For the next phase, *extrinsics*, poses of calibration patterns as well as intrinsics are fixed and only the positions and rotations of all cameras are optimized. Finally, in the phase *bundle adjustment* all parameters are allowed to change.

B. Image rectification for light field processing

In order to facilitate light field / multi-view correspondence analysis methods, images captured by the system need to be rectified making use of the obtained calibration model. Since all cameras usually point to different directions and their locations are rarely coplanar, the standard stereo image rectification [10], which is defined for two cameras, cannot easily be generalized.

As described in [5], all camera views need to be reprojected to a common regression plane ε, which turns the costly warping necessary for cross-comparison between multiple images to simpler translation and scaling operations. If all camera centers are coplanar and ε is chosen parallel to the camera plane, the entire image manipulation needed for the correspondence analysis between multiple cameras reduces just to a translation, which poses a significant computational and algorithmic advantage over the standard stereo approach.

The rectified images have been computed as follows: the regression plane ε has been chosen parallel to the plane fitted through the camera centers and minimizing the squared distance to all calibration patterns. Then, all camera images have been projected onto ε and resampled with the same regular pixel grid. The obtained images form the rectified light field which is required to perform depth estimation.

III. Calibrating Xapt Eye-sect XA

With regard to demonstrating real-world performance of the proposed calibration model, we have taken an example of the Xapt Eye-sect XA camera array with 4x4 camera modules equipped with identical wide-angle lenses. Initially a number of images of AIT's calibration target [7] were acquired for estimating the camera's calibration model, see Fig. 1 (top) for a visualization.

A. Comparison of the reprojection errors

In order to compare the results of our method comprising bundle adjustment with the original method of Bouguet [8], we have conducted a quantitative accuracy analysis based on the reprojection error. The results of this analysis are shown in Tab. I.

For a typical set of calibration images, the reprojection errors resulting from Eq. (1), which are computed per camera after a pairwise optimization with respect to Camera 6, are shown in row *pairwise*. The reprojection errors after further optimization of the calibration pattern poses are given in row *patterns*. The row *extrinsics* shows errors after additional optimization of extrinsic parameters for all cameras. Finally, the reprojection errors after the full bundle adjustment, which also includes optimizing the intrinsic parameters of all cameras, are provided in row *bundle adjustment*.

This exemplary application shows that the biggest drop in the reprojection error occurs in the first step after the initialization, which is when the optimization of the calibration pattern poses took place. Nevertheless, the bundle adjustment showed superior results over all other considered intermediate steps yielding an accuracy which is 33x higher than the initial pairwise method.

B. Depth estimation using light fields

Light field data is captured with the Xapt Eye-sect XA camera by measuring the irradiance values from different viewpoints on the object. For each observation we thereby obtain 4×4 images with different viewpoints, each of which has a resolution of 480×480 pixels. For the multi-view correspondence analysis we considered 32 random camera pairs out of all 120 (i.e. $\binom{16}{2}$) possible pairs.

Since the Xapt Eye-sect XA shows irregularities in the system geometry, namely camera positions were off the grid by as much as 3% of the baseline, the implementation of a robust matching algorithm for 3D reconstruction is essential. Therefore we implemented a robust multi-view matching algorithm for a qualitative evaluation of the camera calibration as described below.

We generate *normalized gradient* features for comparing local image structures, which we compare using the *sum of absolute differences* (SAD). This approach proved more performant compared to the traditional *Census transform / Hamming distance* [11] tandem, especially when coupled with a subsequent regularizer.

Using the resulting features of the rectified light field images (which we obtained with the calibration model as described in Sec. II-B), we perform a correspondence analysis inspired by [4] in each spatial location $(x, y) \in X \times Y$ of a chosen reference view of the camera matrix. The analysis is conducted separately for pairs of cameras. Each hypothesis for a defined camera pair contributes to one global cost function. The resulting cost volume describes the matching quality of visual structures for defined disparity hypotheses within the light field views.

A globally consistent depth solution was obtained under the *total variation* (TV) prior, using a real-time discrete-continuous optimizer proposed in [6]. This algorithm shows exceptional performance, both concerning the speed as well as the solution quality. Further depth refinement methods can be implemented as described in [12].

Figs. 4 and 5 show qualitative reconstruction examples.

IV. Conclusions

We presented a pipeline for geometric multi-view calibration, which includes bundle adjustment. With our routines we have calibrated a multi-view camera and used it for capturing depth information.

The conducted quantitative analysis based on the reprojection error revealed superiority of the bundle adjustment over all other considered intermediate steps yielding an accuracy as much as 33x higher than the initial pairwise method. The largest refinement of the reprojection error was observed in the first step after the initialization during the optimization of the poses of the calibration pattern.

For a qualitative assessment of the calibration model we implemented a multi-view stereo matching algorithm which includes a real-time discrete-continuous optimizer which allows a globally consistent depth map under the generalized first-order *total variation* (TV) prior.

TABLE I

Comparison of the mean reprojection errors (in pixel) per camera of Xapt Eye-sect XA array for different phases of the proposed algorithm. Note that the algorithm was initialized with the *pairwise* method and reference camera 6, subsequently different parameters were optimized: in *patterns* poses of the calibration patterns, in *extrinsics* only the 16 cameras' poses and in *bundle adj.* all parameters, including camera intrinsics.

Phase/Camera	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}	Avg.
pairwise	2.28	8.75	9.32	10.29	2.27	0.09	8.49	16.64	6.21	1.91	13.47	12.73	4.06	8.18	9.01	8.44	7.63
patterns	0.37	0.44	0.57	0.41	0.44	0.54	0.62	0.42	0.38	0.41	0.51	0.77	0.46	0.34	0.49	0.93	0.51
extrinsics	0.16	0.24	0.33	0.22	0.24	0.54	0.58	0.25	0.22	0.15	0.29	0.28	0.29	0.24	0.37	0.29	0.29
bundle adj.	0.12	0.19	0.30	0.18	0.23	0.15	0.55	0.19	0.19	0.13	0.26	0.21	0.27	0.21	0.33	0.22	0.23

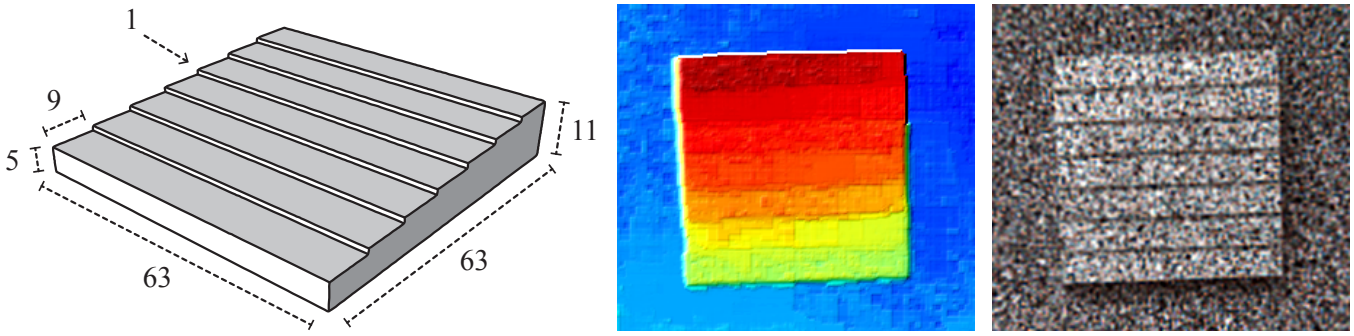

Fig. 4. The performance of the presented multi-camera array calibration was tested by means of a 3D printed staircase object with seven 1 mm steps. The object was acquired with the Xapt Eye-sect XA camera at the working distance of approx. 340 mm. The estimated system's baseline and the average focal length was approx. 90 mm and 710 mm, respectively. The corresponding depth resolution was $\Delta z \approx 1$ mm. Despite a low camera resolution and limited baseline, the obtained disparity map generated by the calibrated camera array accurately reproduced each individual step of the staircase, hence we consistently operate at or above the predicted depth resolution of this system. That is additional evidence of the calibration model's high accuracy additionally to low reprojection errors.

REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. of Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH, New York, NY, USA, 1996, pp. 31–42.

[2] F. Ciurea, D. Lelescu, P. Chatterjee, and K. Venkataraman, "Adaptive geometric calibration correction for camera array," *Electronic Imaging*, vol. 2016, no. 13, pp. 1–6, 2016.

[3] Y. Furukawa and J. Ponce, "Accurate camera calibration from multi-view stereo and bundle adjustment," *International Journal of Computer Vision*, vol. 84, no. 3, pp. 257–268, 2009.

[4] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 4, pp. 353–363, 1993.

[5] R. T. Collins, "A space-sweep approach to true multi-image matching," in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*. IEEE, 1996, pp. 358–363.

[6] A. Shekhovtsov, C. Reinbacher, G. Graber, and T. Pock, "Solving dense image matching in real-time using discrete-continuous optimization," in *21st Computer Vision Winter Workshop*, 2016.

[7] B. Blaschitz, S. Štolc, and D. Antensteiner, "Geometric calibration and image rectification of a multi-line scan camera for accurate 3d reconstruction," in *IS&T Electronic Imaging*, 2018.

[8] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2004.

[9] G. Bradski, "Opencv toolbox," *Dr. Dobb's Journal of Software Tools*, 2000.

[10] R. Klette, *Concise computer vision*. Springer, 2014.

[11] C. Zinner, M. Humenberger, K. Ambrosch, and W. Kubinger, "An optimized software-based implementation of a census-based stereo matching algorithm," in *International Symposium on Visual Computing*. Springer, 2008, pp. 216–227.

[12] D. Antensteiner, S. Štolc, and T. Pock, "A review of depth and normal fusion algorithms," *Sensors*, vol. 18, no. 2, 2018.

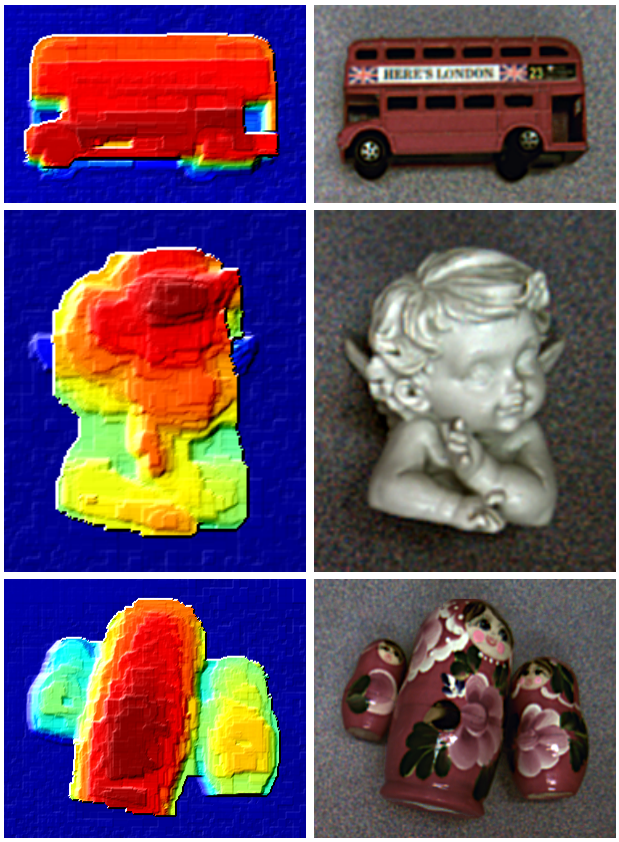

Fig. 5. Examples of the estimated disparity maps (left) and corresponding texture images (right) for several real-world objects. The disparity maps are displayed in pseudo colors, where blue stands for areas further away and red for areas closer to the camera. In order to increase readability of surface details, slight shading was applied to disparity maps.