Dima, E.; Sjöström, M.; Olsson, R., "Assessment of Multi-Camera Calibration Algorithms for Two-Dimensional Camera Arrays Relative to Ground Truth Position and Direction", in 3DTV-Conference, 4-6 July 2016.

ASSESSMENT OF MULTI-CAMERA CALIBRATION ALGORITHMS FOR TWO-DIMENSIONAL CAMERA ARRAYS RELATIVE TO GROUND TRUTH POSITION AND DIRECTION

Elijs Dima, Mårten Sjöström, Roger Olsson

Dept. of Information and Communication Systems, Mid Sweden University
SE-851 70 Sundsvall Sweden

ABSTRACT

Camera calibration methods are commonly evaluated on cumulative reprojection error metrics, on disparate one-dimensional datasets. To evaluate calibration of cameras in two-dimensional arrays, assessments need to be made on two-dimensional datasets with constraints on camera parameters. In this study, accuracy of several multi-camera calibration methods has been evaluated on camera parameters that are affecting view projection the most. As input data, we used a 15-viewpoint two-dimensional dataset with intrinsic and extrinsic parameter constraints and extrinsic ground truth. The assessment showed that self-calibration methods using structure-from-motion reach equal intrinsic and extrinsic parameter estimation accuracy with standard checkerboard calibration algorithm, and surpass a well-known self-calibration toolbox, BlueCCal. These results show that self-calibration is a viable approach to calibrating two-dimensional camera arrays, but improvements to state-of-art multi-camera feature matching are necessary to make BlueCCal as accurate as other self-calibration methods for two-dimensional camera arrays.

Index Terms — Camera calibration, multi-view image dataset, 2D camera array, self-calibration, calibration assessment

1. INTRODUCTION

For accurate sampling of a scene's light field, systems composed of multiple digital cameras must undertake a camera calibration process. Calibration provides information on each camera's internal (intrinsic) parameters and their relative positions (extrinsic parameters), forming pinhole camera matrices [1] that are used in rendering new virtual views. Although various calibration techniques exist in the light field and computer vision community, it has not been reported how calibration techniques perform for two-dimensional camera arrays, in particular relative to ground truth camera intrinsic and extrinsic parameters.

Existing calibration techniques were evaluated on disparate datasets in [2][3][4][5] without an available ground truth for camera placement and properties, instead relying on reprojection errors. Some techniques have publicly available implementations [2][3][6], and some are theoretically described [5] in academic literature. Therefore, when constructing light field capture systems with two-dimensional multi-camera layouts, existing methods need to be evaluated for suitability on common grounds.

In this paper, freely available calibration implementations were assessed with focus on determining their suitability for use in our upcoming Light Field Evaluation System (LIFE). LIFE's capture component will consist of a 2-dimensional array of synchronized, coplanar color cameras, and is intended for use in indoor teleconferencing scenarios. Implementations of multi-camera calibration methods were assessed on a common dataset with 3 vertical by 5 horizontal viewpoint positions and known ground truth constraints on camera intrinsic and extrinsic parameters. The calibration methods' estimates were compared against each other and against the dataset's ground truth.

The novelties of this paper are following: (1) we evaluated several multi-camera calibration methods on a common, two-dimensional dataset representing a typical use-case scenario, (2) we conducted our evaluation based on known ground truth values and parameter equality constraints, and (3) we introduced a dataset for calibration evaluations of two-dimensional multi-camera arrays, with ground truth knowledge. The rest of the article is organized as follows: we describe existing calibration methods and motivate our selections in Chapter 2. Chapter 3 describes our experimental setup and dataset, and Chapter 4 describes the evaluation methodology. We present our results and analysis in Chapter 5, and conclude our work in Chapter 6.

2. CAMERA CALIBRATION

2.1 Overview of camera calibration methods

Current approaches used for camera calibration are generally classifiable as *object-calibration* methods, which make use of special calibration objects [2][6] with known dimensions, and *self-calibration* methods that rely on scene/image properties without a calibration object [3][5][7] and can be used in structure-from-motion reconstruction tools.

A seminal work in object-based camera calibration is Z. Zhang's proposition of the checkerboard calibration process [2][8]. The process involves capturing multiple images of a planar black-and-white checkerboard calibration object in different poses, taking up most of the camera's view. Points-of-interest are extracted from images via locating straight-line intersections. A closed form homography is established between detected checkerboard points and their relation to the absolute image conic in projective geometry. A Levenberg-Marquardt algorithm is employed to improve performance in noisy conditions and deal with nonlinear lens distortion. The general technique presented in [2] has been altered and reworked many times [6][9], with modifications ranging from changes to the calibration object/pattern, to adaptations of the homography estimation or solution optimization.

Self-calibration methods make use of alternate sources of feature correspondences for homography establishment. These correspondences can be obtained from image feature descriptors such as SIFT [10], or from forcing easy-to-detect dimensionless points into the scene, e.g. by using a light stick or a laser pointer, as suggested by T. Svoboda et al. [3]. Their method, implemented as "BlueCCal toolbox", uses synchronized camera capture with a non-deterministically moved point-light source, creating easily identifiable feature-point locations in cameras. The locations are validated via pairwise RANSAC analysis, and missing point projections are filled via projective depth estimation and ranked

978-1-5090-3313-3/16/$31.00 ©2016 IEEE

Figure 1. Left: A scene state captured in our dataset with 15 camera positions. Right: cameras c_1, c_2 and c_3 on a moving dolly in position t = 1.

matrix fitting to an incomplete noisy measurement matrix. Euclidean stratification is used to obtain projection matrices that can be decomposed into intrinsic and extrinsic camera matrices.

2.2 Selection of calibration methods

Both object-calibration and self-calibration approaches are valid for our capture system's use-cases. Ability to autonomously calibrate multiple (n>2) cameras in a system is a requirement for our application. We focused on calibration methods with freely available implementations to make our results more publicly useful, as motivated by Bakken et al. [9]. We avoided evaluating calibration methods with complex or unique calibration objects, or hundreds of synchronized captures, for the same reasons.

We chose to include Z. Zhang's checkerboard calibration algorithm [2] in our evaluation because it serves the purpose of our research and is a standard method for this calibration class [9]. The AMCC toolbox [11] (an automation wrapper for Bouguet's Matlab toolbox [6] of Zhang's algorithm [2]) implementation was selected for evaluation because it fully automates the checkerboard corner identification.

For the self-calibration class, we selected VisualSFM [4][12] and Bundler [7] Structure-from-Motion programs, which inherently incorporate camera calibration, rely on SIFT, and are readily usable. Because of prominence of BlueCCal [3] in self-calibration literature, it was also included in our evaluation. We added a SIFT-based (using VLFeat's [13] version of SIFT) feature multi-matching and filtering algorithm, as described by Goorts et al. in [14] and Dwarakanath et al. in [15], to transform BlueCCal into a calibration method that works without a point-light source.

3. EXPERIMENTAL SET-UP

We created a dataset[1] reflecting the intended scenarios for our upcoming light field capture system in order to evaluate the performance of the calibration methods. The properties of the dataset ensured that our evaluations were based on a 2D-array of high-resolution consumer cameras with constraints on intrinsic and extrinsic camera parameters, in an in-doors scene with and without a dedicated calibration object in n>10 positions and a non-uniform background environment.

The capture unit consisted of a rigid vertical stack of 3 Canon EOS M cameras c_1, c_2, c_3 (shown in Figure 1) mounted on a dolly with 5 equidistant horizontal translation positions (t = 1,..5). Because the same physical camera took images in each elevation level, there exists a constraint on intrinsic camera properties being identical in each camera 'row' in the dataset. The rigid vertical

Distance (d_1), top to middle cameras (c_1 to c_2)	0.352m ± 0.001m, fixed, identical for all horiz. positions (t = 1,..5)
Distance (d_2), middle to bottom cameras (c_2 to c_3)	0.345m ± 0.002m, fixed, identical for all horiz. positions (t = 1,..5)
Horiz. camera-to-camera distance	0.249m ± 0.001m
Camera rotation	Identical for each camera row, static between cameras for (t = 1,..5)
Camera intrinsic parameters	Identical for each camera row

Table 1: Known camera rig constraints.

system and calibrated dolly provided constraints on cameras' relative positioning, which was verified via a laser rangefinder before and after the capture session. For dataset details, see Table 1.

The dataset consisted of 18 captured states of the same scene, each with the 15 predefined camera positions ($c_{i,t}$, i = 1,..3, t = 1,..5). 17 scene states contained a checkerboard placed in different positions and orientations throughout the scene. The remaining state was without a checkerboard as a self-calibration scenario.

4. EVALUATION METHOD

We evaluated the selected calibration methods based on their estimated camera parameter outputs relative to known ground truths, instead of their reported point reprojection errors. The point reprojection error in multi-sensor systems is a cumulative metric with multiple, non-equally contributing factors, as demonstrated by Schwarz et al. [16]. Our assessment focus was placed on camera lens distortion, principal point, and extrinsic parameter estimates, because these are the main contributors of position and depth rendering error in multi-sensor systems [16].

Object-based calibration methods were evaluated on all scene images with checkerboard present. Self-calibration methods were evaluated with no checkerboard present. Evaluations of self-calibration methods were also conducted on scene captures with a single checkerboard present, to determine whether the presence of a checkerboard would affect the results of the calibration methods. Table 2 shows the full experimental setup variations. Calibration methods treated each translation t of cameras c_1, c_2, c_3 as separate cameras.

The lens distortion estimation was assessed based on first coefficient (k_1) for each of c_1, c_2 and c_3 (Figure 1). Each method estimated a different total number of distortion coefficients, reducing the significance of k_1 relative to other parameters. Each calibration method estimated k_1 five times, once for each horizontal translation of cameras in dataset. The distortion was expected to be identical for each lens at each t as per intrinsic constraint in Table 1. We measured the standard deviation (*std*) of k_1 at each position of the cameras. The principal point (x_0, y_0) estimation was likewise assessed, relying on intrinsic parameter equality per camera and evaluating *std* of x_0, y_0 at each position of each camera.

The extrinsic parameter estimation was assessed based on Euclidean distances between cameras, described by the functions d_1, d_2. The function $d_n(c_n,c_{n+1})$ equals the distance between cameras c_n and c_{n+1}. We assessed the Mean Square Error (MSE) of d_1, d_2 respective to ground truth, and the *std* of d_1 and d_2 estimated for each translation of each camera. AMCC took world-scale into account from known checkerboard corner distances. The other calibration methods estimated d_1, d_2 up to an arbitrary global scale factor, which we then matched to the known world-scale to enable result comparison. Rotation estimations of cameras were assessed

[1] Available at www.miun.se/stc/Realistic3D/Dima-2016-1

Name	Applied algorithms	Calibration input
AMCC	Zhang's calibration AMCC automation	17 checkerboard positions
Bundler 1	Snavely's calibration	No checkerboard
Bundler 2	Snavely's calibration	1 checkerboard position
BlueCCal 1	Svoboda's calibration SIFT feat. matching Goorts' filtering	No checkerboard
BlueCCal 2	Svoboda's calibration SIFT feat. matching	No checkerboard
VisualSFM 1	Wu's calibration	No checkerboard
VisualSFM 2	Wu's calibration	1 checkerboard position

Table 2: Experimental calibration tool test setups.

based on the *std* of camera-to-camera relative rotations. We compared angles a_1, a_2 between cameras, expecting a_1, a_2 to remain constant regardless of translation. The function $a_n(c_n,c_{n+1})$ equals the angular offset between cameras c_n and c_{n+1}.

5. RESULTS AND ANALYSIS

The calibration methods estimated k_1 to be fairly constant for c_1, c_2 and c_3, with a maximum *std* of 0.0113 (AMCC, c_2). The presence or absence of a checkerboard changed the k_1 estimate by a maximum of 0.0183 for VisualSFM. BlueCCal did not include distortion coefficients in its explicit output data, and thus is not part of Figure 2. The figure further shows that AMCC and VisualSFM calibration estimated similar k_1 values, whereas Bundler estimated larger k_1 for all three cameras. Bundler and VisualSFM exhibited a more consistent behavior for the estimates of k_1, when considering k_1 of c_1 relative to c_2 and c_3.

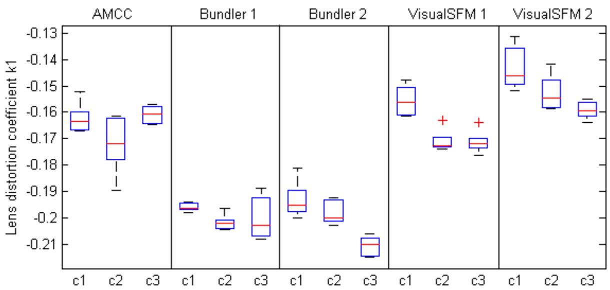

Figure 2. Estimates of lens distortion coefficient k_1 for top (c_1), middle (c_2), and bottom (c_3) cameras. Box plots show median, 25[th] & 75[th] percentile, whiskers show min and max of k_1 estimates, + are outliers.

Bundler and VisualSFM bypass principal point x_0, y_0 estimation by halving the image resolution. This implies that uncertainties in measurements are translated to two parameters (focal length and distortion) and thus implies lower variances than if all three parameters had been estimated. However, Figure 2 shows that k_1 variation was similar between the assessed methods.

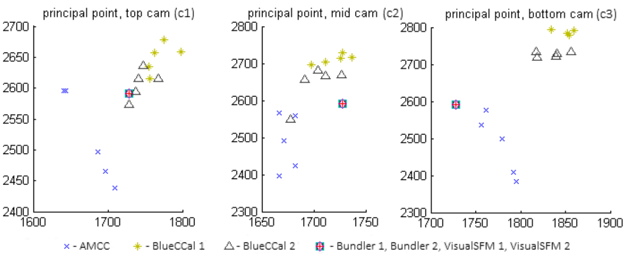

Figure 3. Estimated principal point pixel offset values for top (c_1), middle (c_2) and bottom (c_3) physical cameras.

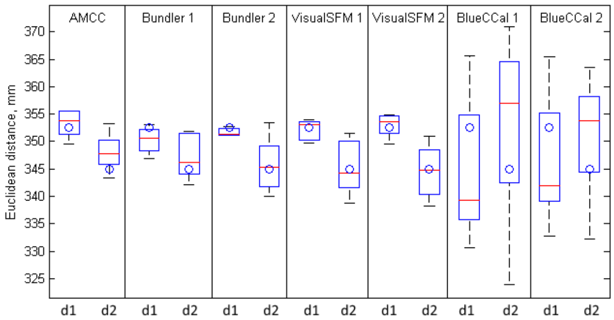

Figure 4. Inter-camera Euclidean distance estimates $d_1(c_1,c_2)$, $d_2(c_2,c_3)$. Circle shows ground truth, box plots show median, 25[th] & 75[th] percentile, whiskers show minimum and maximum of d_1, d_2.

BlueCCal and AMCC estimate x_0, y_0 based on internal estimates of the lens distortion and point reprojections. As shown in Figure 3, BlueCCal and AMCC estimated different principal point values for the same cameras at different translations, with a maximum $std(x_0) = 32.3$ and $std(y_0) = 82.0$ by AMCC.

For estimated camera-to-camera Euclidean distances d_1, d_2, all calibration methods exhibited inaccuracies ranging from 3mm to 25mm, with BlueCCal providing the least accurate position estimates in terms of variation. Figure 4 shows that presence or absence of checkerboard in the scene did not affect position estimation accuracy for Bundler and VisualSFM. Likewise, there was no notable difference in accuracy between the checkerboard-calibration method and the better-performing self-calibration methods, with maximum inaccuracy of 8mm by VisualSFM.

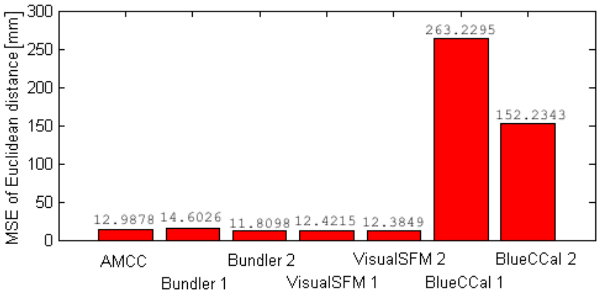

Figure 5. Mean Square Errors of Euclidean distances d_1, d_2 of estimated camera positions with respect to measured ground truth.

The MSEs of camera-to-camera distances in Figure 5 show that Bundler and VisualSFM were as accurate as AMCC with respect to the ground truth. BlueCCal's MSEs were larger by a factor of 11 to 20, indicating a lower position estimation precision.

Estimated camera-to-camera angles a_1, a_2 show that all calibration methods, except BlueCCal, were fairly constant in estimating relative camera orientation. Maximum deviation for AMCC was 0.0049 rad.

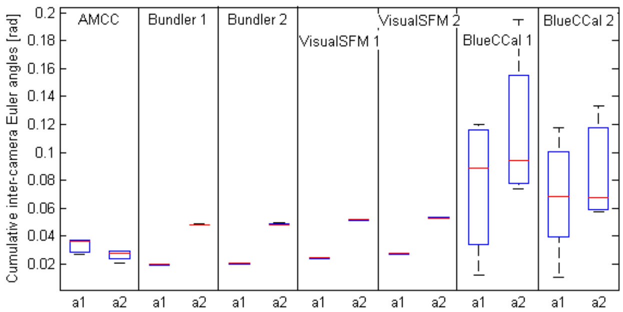

Figure 6. Estimates of inter-camera rotation difference a_1 (between cameras c_1,c_2), a_2 (between cameras c_2, c_3). Box plots show median, 25[th] & 75[th] percentile, whiskers show minimum and maximum of a_1, a_2.

Figure 6 shows that BlueCCal was less accurate by a factor of 9 to 10, exhibiting a maximum deviation of 0.0472 rad. Presence or absence of checkerboard did not affect the rotation estimation of Bundler and VisualSFM calibration, and both estimated identical a_1, a_2 with no notable deviations.

The presence or absence of checkerboard in inputs to self-calibration methods made no notable difference to parameter estimation accuracy. While checkerboard pattern corners are easier to detect than general image features, the self-calibration methods did not have the necessary detector optimizations to capitalize on this. Moreover, for all significant camera parameters as identified by Schwarz et al. [16], the assessed checkerboard-calibration method performed no better than the self-calibration methods in Bundler and VisualSFM. Self-calibration methods estimated more precise extrinsic parameters, as evidenced by distributions of d_1, d_2, a_1, a_2, whereas AMCC estimated additional intrinsic parameters x_0 and y_0, which likely caused greater estimation variations. AMCC's execution time was several orders of magnitude greater than VisualSFM's/Bundler's, with the largest time spent on checkerboard corner detection. However, this may have been caused by differences in implementation or optimization, which we did not focus on.

BlueCCal was consistently the least accurate of the assessed calibration methods. In particular, the estimate deviations in extrinsic parameters indicated that BlueCCal would produce more erroneous virtual views in the assessed configuration. The other self-calibration methods also relied on SIFT feature detection, implying that BlueCCal's inaccuracy may be caused by differences in match filtering. Estimation differences in Figure 5 and Figure 6 between BlueCCal 1 and BlueCCal 2 proved that pre-filtering of cross-camera feature matches can negatively affect estimation accuracy. We additionally tested BlueCCal with a Hessian-Laplace feature detector from VLFeat, which made BlueCCal gradually discard all but 20 detected feature matches as 'outliers' and subsequently fail to converge on any acceptable camera parameter sets.

6. CONCLUSIONS

We selected and evaluated 4 freely available tools for the purposes of multi-camera calibration. To measure estimated camera parameter values from calibration directly against known constraints, we captured a dataset with 15 camera positions and 18 scene states, using 3 cameras in a controlled-motion rig. Ground truth and equality constraints from physical cameras were used to verify calibration method accuracy based on estimation errors for camera parameters that are most significant in view reprojection.

Assessment results showed that SIFT-based self-calibration methods embedded in VisualSFM and Bundler structure-from-motion tools are more accurate than traditional autonomous checkerboard calibration for two-dimensional camera arrays. The choice of checkerboard calibration vs. self-calibration can therefore be determined by practical aspects such as expected scene properties and ability and time to manipulate checkerboards in a scene prior to data capture. Our results also showed that the most widely available Matlab self-calibration toolbox, BlueCCal, requires better than the existing, tested alterations in feature detection and matching in order to achieve acceptable accuracy in two-dimensional multi-camera systems without resorting to a point-light source in a dark room.

Our future work involves designing an integrated variation of the calibration methods used in Bundler/VisualSFM, adapted for our planned multi-camera capture system. An extension to enable principal point estimation is also being considered. Alternately, Zhang's traditional checkerboard calibration method may be adapted, but significant improvements to autonomous execution speed are necessary for practical use.

7. ACKNOWLEDGEMENT

This work has been supported by grant 20140200 of the Knowledge Foundation, Sweden.

8. REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Press, Cambridge, 2004.

[2] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, pp. 1330–1334, Nov. 2000.

[3] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multicamera self-calibration for virtual environments," *Presence*, vol 14, pp. 407–422, Aug. 2005.

[4] C. Wu, "Towards Linear-Time Incremental Structure from Motion", in *3D Vision – 3DV 2013, 2013 International Conference on*, Seattle, USA, June/July 2013, pp. 127–134.

[5] Z. Zhang, Y. Matsushita, and Y. Ma, "Camera calibration with lens distortion from low-rank textures", *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, Colorado Springs, USA, June 2011, pp. 2321–2328.

[6] J.-Y. Bouguet, Camera calibration toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib doc/, accessed on 21/02/2016.

[7] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo Tourism: Exploring image collections in 3D," in ACM Transactions on Graphics (TOG), vol. 25, pp. 835–846, July 2006.

[8] W. Sun and J. R. Cooperstock, "Requirements for Camera Calibration: Must Accuracy Come with a High Price?" in *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on*, Breckenridge, USA, Jan 2005, pp. 356–361.

[9] R. H. Bakken, B. G. Eilertsen, G. U. Matus, and J. H. Nilsen, "Semi-automatic camera calibration using coplanar control points," in *Proceedings of NIK Conference*, Trondheim, Norway, Nov. 2009, pp. 37–48.

[10] D. G. Lowe, "Object recognition from local scale-invariant features," *Computer Vision 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, pp. 1150–1157, Sept. 1999.

[11] M. Warren, D. McKinnon, and B. Upcroft, "Online Calibration of Stereo Rigs for Long-Term Autonomy," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, Karlsruhe, Germany, May 2013, pp. 3692–3698.

[12] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz, "Multicore Bundle Adjustment," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, Colorado Springs, USA, June 2011, pp. 3057–3064.

[13] A. Vedaldi and B. Fulkerson, "VLFeat: An Open and Portable Library of Computer Vision Algorithms," in *Proceedings of the 18th ACM international conference on Multimedia*, Firenze, Italy, Oct. 2010, pp. 1469–1472.

[14] P. Goorts, S. Maesen, Y. Liu, M. Dumont, P. Bekaert, and G. Lafruit, "Self-calibration of Large Scale Camera Networks," in *Proc. of the 11th Intl. Conf. on Signal Processing and Multimedia Applications (SIGMAP 2014)*, Vienna, Austria, Aug. 2014.

[15] D. Dwarakanath, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Faster and More Accurate Feature-Based Calibration for Widely-Spaced Camera Pairs," in *Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference on,* Bangkok, Thailand, May 2012, pp. 87 – 92.

[16] S. Schwarz, M. Sjöström, and R. Olsson, "Multivariate Sensitivity Analysis of Time-of-Flight Sensor Fusion," in *3D Research*, vol. 5:3, pp. 1 – 16, Sept. 2014