Towards Euclidean auto-calibration of stereo camera arrays

Roopak R. Tamboli[a], Kiran K. Vupparaboina[b], M. Shanmukh Reddy[c], Peter A. Kara[d],
Aron Cserkaszky[a], Maria G. Martini[d], Ashutosh Richhariya[b], and Soumya Jana[c]

[a]Holografika, Budapest, Hungary
[b]LV Prasad Eye Institute, Hyderabad, India
[c]Institute of Technology Hyderabad, Kandi, Sangareddy, India
[d]WMN Research Group, Kingston University, Kingston upon Thames, UK

ABSTRACT

Multi-camera networks are becoming ubiquitous in a variety of applications related to medical imaging, education, entertainment, autonomous vehicles, civil security, defense etc. The foremost task in deploying a multi-camera network is camera calibration, which usually involves introducing an object with known geometry into the scene. However, most of the aforementioned applications necessitate non-intrusive automatic camera calibration. To this end, a class of camera auto-calibration methods imposes constraints on the camera network rather than on the scene. In particular, the inclusion of stereo cameras in a multi-camera network is known to improve calibration accuracy and preserve scale. Yet most of the methods relying on stereo cameras use custom-made stereo pairs, and such stereo pairs can definitely be considered imperfect; while the baseline distance can be fixed, one cannot guarantee the optical axes of two cameras to be parallel in such cases. In this paper, we propose a characterization of the imperfections in those stereo pairs with the assumption that such imperfections are within a considerably small, reasonable deviation range from the ideal values. Once the imperfections are quantified, we use an auto-calibration method to calibrate a set of stereo cameras. We provide a comparison of these results with those obtained under parallel optical axes assumption. The paper also reports results obtained from the utilization of synthetic visual data.

Keywords: multi-camera network, camera calibration, stereo camera, Euclidean calibration

1. INTRODUCTION

3D visualization is instrumental in several applications such as telepresence, next-generation remote healthcare,[1] education and entertainment. The goal of re-creating the 3D-likeness of subjects, objects and scenes in these applications can be achieved by methods such as range scanning, structure-from-motion, stereo cameras, array of cameras etc. Since most of the aforementioned applications require non-intrusive and instantaneous acquisition, capturing multiple 2D views using an array of cameras is a method of wide interest.[2,3] In such methods, camera calibration is the foremost task where the intrinsic parameters such as focal length, as well as the extrinsic parameters such as camera position and orientation, are estimated. Traditionally, extraneous objects — e.g., chessboard patterns, checkered cubes and laser pointers — have been used for calibrating both single camera,[4,5,6] and network of cameras,[7,8].[9] The aforementioned applications, where an adaptive camera network dynamically calculates the calibration parameters, require scene-based auto-calibration. In such applications, pre-calibrated cameras could be restrictive, and extraneous objects may not be introduced.[10] Furthermore, applications such as multi-party telepresence, requiring the fusion of multiple scenes, require the preservation of scale as well.

Further author information: (Send correspondence to Roopak R. Tamboli)
Roopak R. Tamboli: E-mail: r.tamboli@holografika.com, ee13p0008@iith.ac.in
Kiran K. Vupparaboina: E-mail: vkiran@lvpei.org
M.Shanmukh Reddy: E-mail: ee14resch11006@iith.ac.in
Peter A. Kara: E-mail: p.kara@kingston.ac.uk
Aron Cserkaszky: E-mail: a.cserkaszky@holografika.com
Maria G. Martini: E-mail: m.martini@kingston.ac.uk
Ashutosh Richhariya: E-mail: ashutosh@lvpei.org
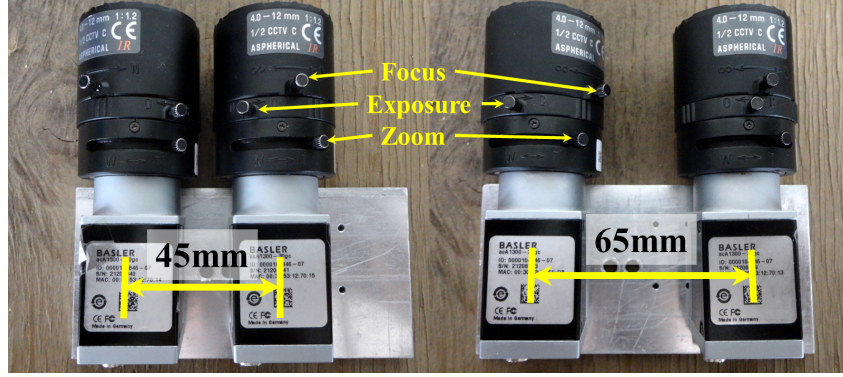Soumya Jana: E-mail: jana@iith.ac.in

Figure 1: Custom-made stereo pairs (Vupparaboina *et al.*[24]).

Conventional multi-camera networks contain several monocular cameras. For such networks, various methods offer metric auto-calibration, allowing faithful recovery of shape, but not scale.[2] Such techniques typically establish point correspondences across multiple 2D views, remove false correspondences, and then make use of either a factorization method,[11] or Kruppa equations.[12] Metric calibration has also been obtained by imposing other related constraints, such as constraints on camera extrinsic using special motions,[13],[14],[15] and constraints on the scene.[16,17] Furthermore, attempts have been made at combining factorization and Kruppa equations towards distributed auto calibration.[18] Subsequently, making use of efficient sparse bundle adjustment (SBA),[19] Furukawa and Ponce proposed a method for auto-calibration and dense 3D reconstruction.[20]

In comparison, significantly less effort has been directed towards Euclidean auto-calibration, required for the recovery of both shape and scale. Lerma *et al.* considered a three-camera configuration, imposed baseline constraints on each of the three camera pairs (resulting in an inflexible structure), and reported photogrammetrically improved auto-calibration.[21] Vuppparaboina *et al.* demonstrated that Euclidean auto-calibration of a camera network is achievable as long as the camera network contains a stereo camera with known baseline separation[22] and provided an extended analysis using a four-camera network which can be generalized.[23] The authors also reported the quantification of true-to-scale facial features using a multi-camera network.[24] Their proposed method recovers scale from the mild baseline constraint, while still achieving re-projection errors comparable to that achieved by the well-known Zhang and Pollefeys methods.[5,25]

As explained above, the assisted methods of camera calibration preserve scale by introducing an extraneous object with know geometry in the scene. On the other hand, non-intrusive methods impose constraints on the camera network, such as the inclusion of a stereo camera. The performance of both these kinds of calibration methods is heavily affected by the degree of closeness of the explicit or implicit constraints that these methods introduce, to their ideal values. For example, an assisted method that uses planar chessboards as calibration objects could be adversely affected if the chessboards have been bent by frequent use or the dimensions of the squares have an error in their measurement. In this regard, Huang *et al.* presented a flexible camera calibration using an unmeasured imperfect target.[26] Their method uses the geometry of the calibration target in an optimization process with constraints based on the pattern design. As a result, the camera can be calibrated in the absolute scale without taking any measurement of the imperfect calibration target.

In this paper, we take into account a few sources of errors that affect the non-intrusive methods of camera calibration. In particular, we adopt the Euclidean camera calibration framework proposed by Vupparaboina *et al.* As discussed earlier, the said method requires at least one stereo camera with known baseline separation to be included in the camera network. In an ideal stereo camera, the two optical axes are parallel. However, several custom-made stereo pairs, such as the ones shown in Figure 1, may not achieve this in practice. Furthermore, the distance between two camera centers can also contain error in its measurement due to imperfections in the scale used to measure it, wearing of the platform on which cameras are mounted, manufacturing defects or human errors in measurements. In this work, we attempt to examine how an imperfect stereo camera affects the performance of the aforementioned calibration method, in a simulated environment.

The remainder of this paper is organized as follows: In Section 2, we explain theoretical underpinnings of this work, originally proposed by Vupparaboina *et al.*[22] Here, we introduce some practical issues associated with camera calibration in general and our approach in particular. We attempt to address some of the issues in a simulated environment in Section 3. Section 4 presents results of this study. The paper is concluded in Section 5.

2. THEORETICAL BACKGROUND

We briefly present the theoretical principle that enables true-to-scale 3D reconstruction. A detailed discussion is provided by Vupparaboina *et al.*[23] In this research, we simulate a four-camera network consisting of two stereo camera pairs with known baseline separations. Each of the constituent cameras forms an image of the target object — here, a set of 3D points. An example of such a network is shown in Figure 2. In this network, cameras 'C1' and 'C2' form a stereo pair, and cameras 'C3' and 'C4' form the second stereo pair. Therefore, these cameras are also referred as SL-L, SL-R, SR-L, and SR-R, respectively. Although there were two stereo cameras pairs, the second pair is considered as two monocular cameras while deriving the expressions in the following. Without the loss of generality, this enabled using the previously proposed mathematical formulation. Nevertheless, the second stereo pair offers a higher number of point correspondences compared to two monocular cameras.

Let us consider N feature points on the object, with homogeneous coordinates $\bar{X}_j = [X_j \ Y_j \ Z_j \ 1]^T$ ($j = 1, 2, ..., N$), each visible to all the four cameras. Denote by $\bar{x}_{ij} = [x_{1j} \ y_{1j} \ 1]^T$ the projection of $\bar{X}_j$ on the i-th image plane (in local coordinates), where $i = 1, 2, 3, 4$ indicate the respective indices of the left and the right cameras in the stereo pair, and the left and the right monocular cameras (respective cameras). Now, ignoring lens distortion, image $\bar{x}_{ij}$ of $\bar{X}_j$ ($j = 1, 2, \ldots, N$) formed by the i-th ($i = 1, 2, 3, 4$) camera is given by[27]

$$s_{ij} \begin{bmatrix} x_{ij} \\ y_{ij} \\ 1 \end{bmatrix} = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1^i & r_2^i & r_3^i & | & t_X^i \\ r_4^i & r_5^i & r_6^i & | & t_Y^i \\ r_7^i & r_8^i & r_9^i & | & t_Z^i \end{bmatrix} \begin{bmatrix} X_j \\ Y_j \\ Z_j \\ 1 \end{bmatrix}, \tag{1}$$

where s_{ij} denotes a scale factor. In the right hand side of (1), the first matrix is diagonal, and it is specified only by focal length f_i under mild assumptions.[23] The second matrix consists of a 3×3 unitary rotation parameter matrix concatenated with a 3×1 translation parameter vector. Note that the focal lengths of the two cameras in a stereo pair need not be equal.

Now let us fix the origin of the world coordinate system at the center of Camera 1 (SL-L), and take its principal axis as the Z-axis. Thus the center of Camera 2 (SL-R) has the location $[l \ 0 \ 0]^T$, where l denotes the known baseline distance of the stereo pair. Furthermore, for the sake of simplicity, let us assume that the sensor arrays in Cameras 1 and 2 are shifted versions of each other such that both rotation matrices are identical.[28] Accordingly, solving (1) for $i = 1, 2$ yields

$$X_j = \frac{x_{1j} y_{2j} l}{x_{1j} y_{2j} - x_{2j} y_{1j}} \tag{2}$$

$$f_2 = \frac{y_{2j}}{y_{1j}} f_1 \tag{4}$$

$$Y_j = \frac{y_{1j} y_{2j} l}{x_{1j} y_{2j} - x_{2j} y_{1j}} \tag{3}$$

$$Z_j = \frac{f_1 X_j}{x_{1j}}. \tag{5}$$

In (2)–(5), we can observe that X_j and Y_j are unambiguously determined. However, f_2 and Z_j are obtained only in terms of f_1, which at present remains the only unresolved quantity related to the stereo pair ($i = 1, 2$). Clearly, if Camera 1 was calibrated, f_1 and hence all unknown quantities would be known, which in theory should have been enough for 3D reconstruction.[29] Yet in practice, such a method would not be robust, as a small inaccuracy in the value of f_1 would produce a large error in Z_j, the relative depth of the j-th feature point.

In the present setting (where Camera 1 is uncalibrated), f_1 can still be resolved using images formed by monocular Cameras 3 and 4, i.e., SR-L and SR-R. To achieve such a resolved state, one must note the following:

Using (5) in (1) for the monocular cameras ($i = 3, 4$), we obtain

$$
\begin{aligned}
x_{ij} = {}& X_j \frac{f_i r_1^i}{t_Z^i} + Y_j \frac{f_i r_2^i}{t_Z^i} + \frac{X_j}{x_{1j}} \frac{f_1 f_i r_3^i}{t_Z^i} + \frac{f_i t_X^i}{t_Z^i} \\
& - x_{ij} X_j \frac{r_7^i}{t_Z^i} - x_{ij} Y_j \frac{r_8^i}{t_Z^i} - \frac{x_{ij} X_j}{x_{1j}} \frac{f_1 r_9^i}{t_Z^i}
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
y_{ij} = {}& X_j \frac{f_i r_4^i}{t_Z^i} + Y_j \frac{f_i r_5^i}{t_Z^i} + \frac{X_j}{x_{1j}} \frac{f_1 f_i r_6^i}{t_Z^i} + \frac{f_i t_Y^i}{t_Z^i} \\
& - y_{ij} X_j \frac{r_7^i}{t_Z^i} - y_{ij} Y_j \frac{r_8^i}{t_Z^i} - \frac{y_{ij} X_j}{x_{1j}} \frac{f_1 r_9^i}{t_Z^i},
\end{aligned}
\tag{7}
$$

where respective left hand sides (LHS) of (6) and (7) denote observed x and y image coordinates of j-th object point in i-th monocular cameras, and the corresponding right hand sides (RHS) denote associated estimates. The equality in each of (6) and (7) holds only ideally without any model and measurement errors. In practice, respective differences between LHS and RHS in (6) and (7) indicate the coordinates of the re-projection error. The unknown quantities can now be obtained by minimizing the sum of squared re-projection errors, the sum running over all N feature points ($j = 1, 2, ..., N$) and both monocular cameras ($i = 3, 4$). At this point, one difficulty still remains: the unknown variables are clearly dependent. Fortunately, f_1 and $\{f_i, r_1^i, r_2^i, r_4^i, t_X^i, t_Y^i, t_Z^i\}_{i=3,4}$ constitute a set of 15 independent variables, from which the rest of the unknown quantities can all be derived, and it is enough to perform the aforementioned minimization with respect to only these 15 independent variables.

In this manner, one obtains all unknown quantities, including the absolute values of focal lengths, as well as rotations and translations relative to Camera 1, of the proposed four-camera network, and the process is referred as auto-calibration. Simultaneously, one also estimates (reconstructs) 3D locations $[X_j\ Y_j\ Z_j]$ ($j = 1, 2, ..., N$) without any scale ambiguity, as the baseline separation l acts as an internal reference of length. Clearly, auto-calibration and 3D reconstruction are intertwined operations.

In this paper, since we assume imperfect stereo cameras, we cannot arrive at (2)-(7) in the manner described above. This is because the rotation matrix for 'C2' (or 'C4') is no longer an identity matrix and similarly, the translation vector is different than $[-l, 0, 0]^T$. We deal with this issue by using the 'Fit3D Toolbox' for triangulation and calculation of re-projection error.[30]

3. METHODOLOGY

In this section, we describe our experimental setup and methodology. We resort to simulated environment in order to focus on a few imperfections in stereo cameras and avoid other sources of errors such as lens distortion, incorrect point correspondences, etc. The 3D point cloud data used in this work is a head model available publicly in the GrabCAD library[*]. We use only 50 3D points from this model in this study.

The intrinsic and extrinsic parameters of the four cameras, expressed as the projection matrix $P = K[R|T]$ are given below. We assume that the origin of the world co-ordinate system is aligned with the co-ordinate system of 'C1'.

$$
P_1 = \begin{bmatrix} 1500 & 0 & 0 \\ 0 & 1500 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 1 & | & 0 \end{bmatrix}
\tag{8}
$$

$$
P_2 = \begin{bmatrix} 1510 & 0 & 0 \\ 0 & 1510 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.9997 & 0.0174 & -0.0175 & | & -45 \\ -0.0171 & 0.9997 & 0.174 & | & 0 \\ 0.0178 & -0.0171 & 0.9997 & | & 0 \end{bmatrix}
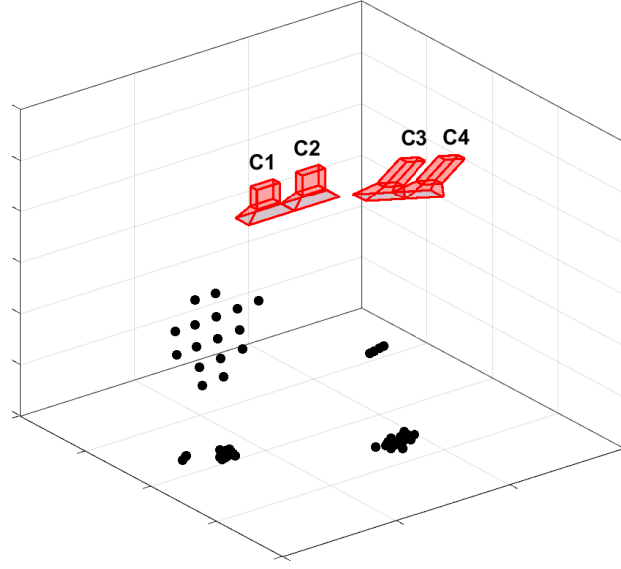\tag{9}
$$

Figure 2: Arrangement of synthetic cameras and 3D points (figure not to scale).

$$P_3 = \begin{bmatrix} 1505 & 0 & 0 \\ 0 & 1505 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.8660 & 0 & -0.5000 & | & -200 \\ 0 & 1.0000 & 0 & | & 0 \\ 0.5000 & 0 & 0.8660 & | & -50 \end{bmatrix} \tag{10}$$

$$P_4 = \begin{bmatrix} 1518 & 0 & 0 \\ 0 & 1518 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.8570 & 0.0174 & -0.5150 & | & -245 \\ -0.0061 & 0.9997 & 0.0237 & | & 0 \\ 0.5152 & -0.0171 & 0.8569 & | & -50 \end{bmatrix} \tag{11}$$

These values are chosen from the real-life setup used by Vupparaboina *et al.* in their work.[24] The baseline length of both the cameras was chosen to be 45mm, same as that of the stereo pair shown in Figure 1. Even in a real-life camera network, one can perform per-camera initial calibration using assisted methods to obtain an estimate of camera parameters and then calibrate the network of cameras. This justifies initialization of camera matrices with close-to-true values.

As briefed in Section 2, camera calibration and 3D reconstruction requires solving a multi-objective optimization problem with several variables. The number of variables depends on the size of the camera network.[23] In all our simulations described below, we used the `gamultiobj` solver for MATLAB. The cost function supplied to `gamultiobj` was re-projection errors for all the points seen by the cameras. Re-projection error is calculated as ℓ_2 norm of difference of original 2D point location and the 2D point obtained by projecting the 3D estimate at each iteration on the camera's image plane. An additional term to account for error in baseline length was included in the cost function.

3.1 Characterization of a stereo camera

We begin with a single stereo camera. As explained in Section 2, a standalone stereo camera cannot be calibrated unambiguously unless the focal lengths are known. Since our aim is to characterize the imperfections in a stereo camera, we assume that the focal lengths can be estimated using an assisted method of calibration. with these assumptions, we simulate two situations described in the following.

3.1.1 Characterization with known focal lengths

In this experiment, we sought for true values of camera parameters using 'gamultiobj', except the focal lengths f_1 and f_2, which were assumed to be known perfectly. Thus, the decision variables in this case were rotation parameters, translation parameters and baseline length. The cost function was evaluated for several pairs of $\{f_1, f_2\}$ and other parameters were allowed to vary. This simulation allowed us to roughly observe the behavior of the cost function and thereby design further simulations, as described next.

3.1.2 Characterization with unknown focal lengths

In the second experiment, it was assumed that the focal lengths are known with uncertainty. Therefore, focal lengths were included in the set of decision variables. Here, we simulated two situations. In the first one, the optical axes of two cameras in the stereo pair were assumed to be parallel, while the 2D projections were obtained using an imperfect stereo camera. Therefore, rotation parameters were constant and were not included in the decision variables. In the second situation, the stereo pair was assumed to be imperfect and the decision variables included the rotation parameters. By minimizing the re-projection error, we expect the decision variables to attain their true values. As long as the stereo camera — albeit imperfect — is unchanged, the camera parameters remain constant. This knowledge is believed to improve the accuracy of camera network calibration in terms of re-projection errors.

While we use synthetic camera matrices for the purpose of simulations, these settings can be used for a real stereo camera, because the baseline length is roughly known (by construction, supplied by manufacturer or measured) and as discussed earlier, other parameters can be estimated approximately using an assisted method of calibration.

## 3.2 Towards calibration of a camera network

Next, we simulate the setups presented by Vupparaboina *et al.*, where (i) one stereo and one monocular are used and (ii) two stereoscopic cameras are used.[22, 23] In these simulations, we evaluate two cases, namely (i) the imperfect stereo camera(s) are assumed to be perfect and (ii) the imperfection in stereo camera(s) are taken into account while solving the optimization problem. In the second case of two stereo cameras, multiple simulations can be carried out depending on how the second stereo pair is treated. It can be used as two monocular cameras or a stereo camera. We used 'C3' and 'C4' as a stereo pair and simulated only two situations, one in which both the stereo pairs are assumed to be ideal, and the second case where imperfections in both the pairs are characterized separately.

3.3 Bounds on decision variables

In all our simulations, the search space of gamultiobj was kept small by providing tight search bounds. In real-life cameras, initial estimates of decision variables can be obtained by Zhang's calibration.[31] The gamultiobj attains a minima for which values of the decision variables were observed to be close to the actual values. We then gradually relaxed the bounds in subsequent runs. While doing so, the initial population supplied the gamultiobj was the population that was at the end of the previous run. One may argue that prior knowledge of camera parameters helps gamultiobj attain the minimum. While this is true, it is still justified in a situation where a camera network is deployed to acquire a certain scene that does not change drastically with time, which, in turn, results in a small change in camera parameters. The heavy initial calibration step becomes necessary only in the beginning. Further runs can benefit from the previous run of the calibration.

4. RESULTS

We now provide results of our experiments described in Section 3. Figure 3 shows results for the experiment described in Section 3.1.1. Re-projection errors in X and Y as well as error in baseline length are denoted by e_x, e_Y and e_l, respectively. As seen from this figure, it is possible for gamultiobj to get trapped in one of the minima. This observation supports the choice of very tight bounds chosen for characterization of a single stereo camera. In particular, the following observations were made pertaining to characterization experiment of a single stereo camera: (i) No bound on decision variables and random initial population results in very high
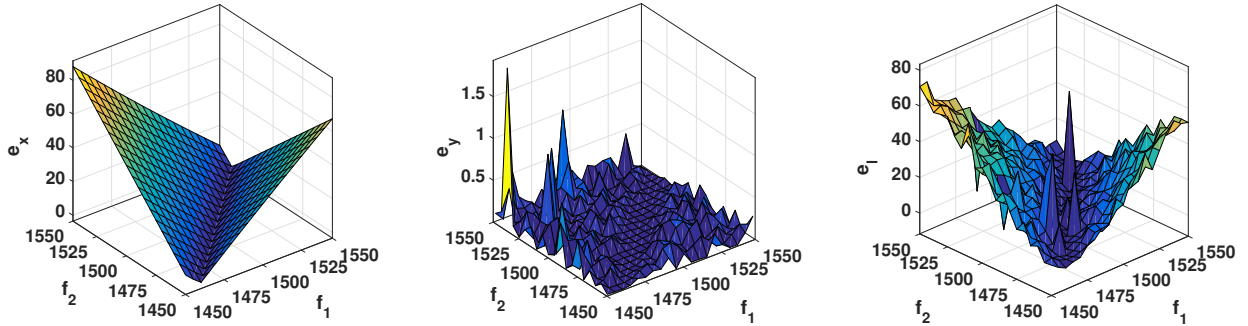
(a) Re-projection error in X-coordinate (b) Re-projection error in Y-coordinate   (c) Error in baseline length

Figure 3: Search over focal length space.

Table 1: Stereo camera characterization: Average re-projection in pixels.

Type of experiment	e_X		e_Y		e_l
	C1	C2	C1	C2	
Imperfections neglected	0.02	0.01	12.04	12.12	0
Imperfections considered	3.65×10^{-14}	4.08×10^{-12}	1.89×10^{-13}	1.83×10^{-13}	0

re-projection error. (ii) Tight bounds on decision variables and random initial population results in very low re-projection error. (iii) No bound on decision variables and one of the populations on the pareto front from (ii) as the initial population results in further reduction in re-projection error. After repeating this step for multiple times, the re-projection error eventually goes to zero. However, the angles and thereby the scale are not always preserved.

Next, we report the results of characterization of stereo camera with unknown focal lengths, as described in Section 3.1.2. In this experiment, all camera parameters were included as decision variables. In the first case, the imperfections in the stereo pair were neglected, and in the second case, they were taken into account. Since the re-projection errors were very small, we provide only the average errors in Table 1. When the stereo cameras were falsely assumed to be perfect, the value of e_Y was found to be little over 12 pixels. When the imperfections were accounted for, e_Y dropped to negligible values. The values of e_X and e_l were negligible and zero, respectively, in both the simulations.

With this observation, we now examine how an imperfect stereo camera affects the calibration of a camera network. We begin with calibration of a system of cameras comprising of an imperfect stereo camera (C1, C2) and a monocular camera (C3). To calibrate this system, `gamultiobj` was used with various combinations of values of 'population size' and 'number of generations'. Figure 4 shows results where the 'population size' was 5000 and the 'number of generations' was 100. The maximum values of e_X and e_Y were found to be 94 pixels and 17 pixels, respectively, if the stereo camera was assumed to be ideal. In the other case, the re-projection errors, though not negligible, were reduced. Note that at each iteration of `gamultiobj`, the 3D points are estimated as seen by the stereo camera (C1-C2), that is the world co-ordinate system and the re-projected on the monocular camera. Thus, the re-projection errors shown in Figure 4 are for the monocular camera.

Finally, we present the results for the calibration of 2 stereo cameras. As described earlier, the second stereo pair was treated as a stereo camera instead of two monocular cameras. Because of this, we only need to find the transformation from camera C3 to camera C1, which in turn helps localizing camera C3. We simulated only two cases where either imperfections in both the stereo cameras were considered or both the cameras were assumed to be ideal. Compared to the system of one stereo and one monocular cameras, the re-projection errors were found to be higher for the two-stereo camera system. Re-projection errors in Y-coordinates did not seem to differ significantly in either case. On the other hand, accounting for the stereo imperfections reduced the re-projection error for X-coordinate.
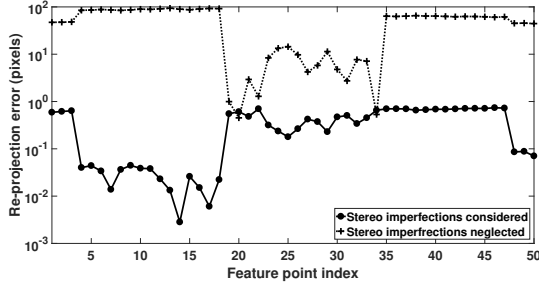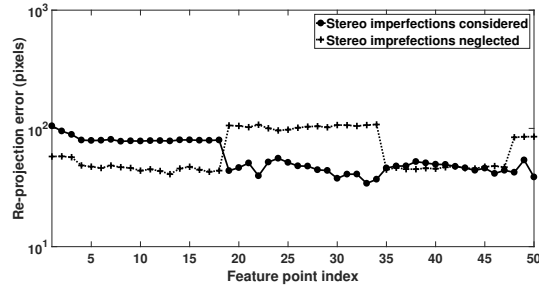
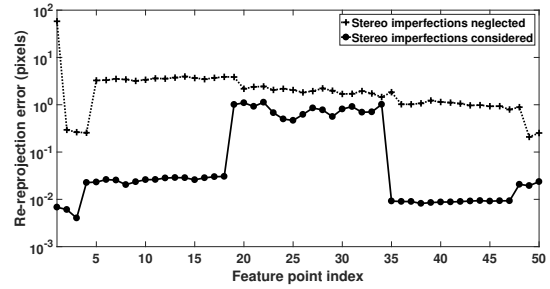(a) Re-projection error in X-coordinate
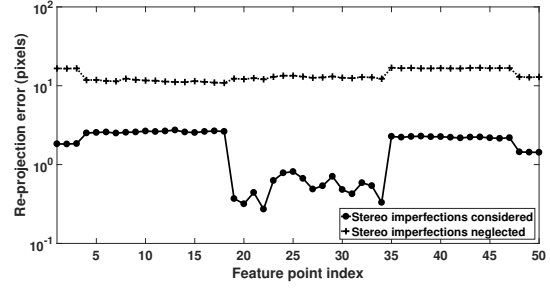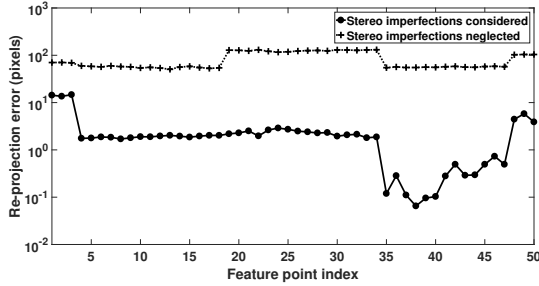

(b) Re-projection error in Y-coordinate

Figure 4: Calibration of a system of one stereo and one monocular camera. The Y-scale is logarithmic. The re-projection errors are for the monocular camera C3.
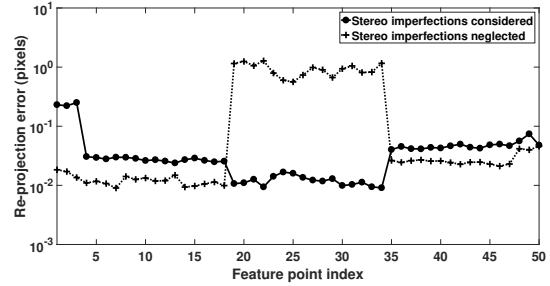

(a) Re-projection error in X-coordinate

(b) Re-projection error in Y-coordinate

(c) Re-projection error in X-coordinate

(d) Re-projection error in Y-coordinate

Figure 5: Calibration of a system of two stereo cameras. The Y-scale is logarithmic. (a) and (b): re-projection errors for 'C3'. (c) and (d): re-projection errors for 'C4'.

5. CONLCUSIONS

In this paper, we studied the effect of imperfections in a stereo camera on 3D reconstruction from multiple views. We used an Euclidean auto-calibration method proposed by Vupparaboina *et al.*,[23] which requires at least one stereo camera with the known baseline separation. First, we characterized the imperfections in a standalone stereo camera. Then, we calibrated a system of one stereo and one monocular camera as well as a system of two stereo cameras. In both cases, we observed that accounting for the imperfections in a stereo pair improves performance of the calibration method. We believe that this approach, albeit benefited from hindsight, would improve calibration of larger camera networks. Relaxation of parallel axis assumption allows one to utilize near-perfect stereo cameras instead of the legacy stereo cameras, yet assures true-to-scale 3D reconstruction. In the future, we shall make an extended analysis of the approach used in this paper, and use real-life cameras and models to investigate the imperfections arising in such setups. We also plan to take into account other sources of error, such as uncertainties in point correspondences and lens distortions.

ACKNOWLEDGMENTS

REFERENCES

[1] Vupparaboina, K. K., *Perspectives on Imaging and Image Analytics in Ophthalmology*, PhD thesis, Indian Institute of Technology Hyderabad (2017).

[2] Hartley, R. and Zisserman, A., [*Multiple view geometry in computer vision*], Cambridge university press (2003).

[3] Aghajan, H. and Cavallaro, A., [*Multi-camera networks: principles and applications*], Academic press (2009).

[4] Tsai, R. Y., "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal of Robotics and Automation* **3**(4), 323–344 (1987).

[5] Zhang, Z., "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(11), 1330–1334 (2000).

[6] Heikkila, J. and Silvén, O., "A four-step camera calibration procedure with implicit image correction," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], 1106–1112, IEEE (1997).

[7] Baker, H., Tanguay, D., Sobel, I., Gelb, D., Goss, M. E., Culbertson, W. B., and Malzbender, T., "The coliseum immersive teleconferencing system," in [*International Workshop on Immersive Telepresence*], **6** (2002).

[8] Kurillo, G., Li, Z., and Bajcsy, R., "Wide-area external multi-camera calibration using vision graphs and virtual calibration object," in [*Second ACM/IEEE International Conference on Distributed Smart Cameras*], (2008).

[9] Svoboda, T., Martinec, D., and Pajdla, T., "A convenient multicamera self-calibration for virtual environments," *PRESENCE: teleoperators and virtual environments* **14**(4), 407–422 (2005).

[10] Hemayed, E. E., "A survey of camera self-calibration," in [*IEEE Conference on Advanced Video and Signal Based Surveillance*], 351–357, IEEE (2003).

[11] Sturm, P. and Triggs, B., "A factorization based algorithm for multi-image projective structure and motion," in [*European conference on computer vision*], 709–720, Springer (1996).

[12] Maybank, S. J. and Faugeras, O. D., "A theory of self-calibration of a moving camera," *International Journal of Computer Vision* **8**(2), 123–151 (1992).

[13] Knight, J., Zisserman, A., and Reid, I., "Linear auto-calibration for ground plane motion," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], **1**, IEEE (2003).

[14] Zhong, H. and Hung, Y., "Self-calibration from one circular motion sequence and two images," *Pattern recognition* **39**(9), 1672–1678 (2006).

[15] Zhao, Z., Liu, Y., and Zhang, Z., "Camera calibration with three noncollinear points under special motions," *IEEE Transactions on Image Processing* **17**(12), 2393–2402 (2008).

[16] Malis, E. and Cipolla, R., "Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(9), 1268–1272 (2002).

[17] Wang, G., Jonathan Wu, Q., and Zhang, W., "Kruppa equation based camera calibration from homography induced by remote plane," *Pattern Recognition Letters* **29**(16), 2137–2144 (2008).

[18] Devarajan, D., Radke, R. J., and Chung, H., "Distributed metric calibration of ad hoc camera networks," *ACM Transactions on Sensor Networks (TOSN)* **2**(3), 380–403 (2006).

[19] Lourakis, M. I. and Argyros, A. A., "SBA: A software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software (TOMS)* **36**(1) (2009).

[20] Furukawa, Y. and Ponce, J., "Accurate camera calibration from multi-view stereo and bundle adjustment," *International Journal of Computer Vision* **84**(3), 257–268 (2009).

[21] Lerma, J. L., Navarro, S., Cabrelles, M., and Seguí, A. E., "Camera calibration with baseline distance constraints," *The Photogrammetric Record* **25**(130), 140–158 (2010).

[22] Vupparaboina, K. K., Raghavan, K., and Jana, S., "Smart camera networks: An analytical framework for auto calibration without ambiguity," in [*IEEE Recent Advances in Intelligent Computational Systems (RAICS)*], 310–315 (2013).

[23] Vupparaboina, K. K., Raghavan, K., and Jana, S., "Euclidean auto calibration of camera networks: baseline constraint removes scale ambiguity," in [*2016 Twenty Second National Conference on Communication (NCC)*], IEEE (2016).

[24] Vupparaboina, K. K., Tamboli, R. R., Reddy, M. S., Richhariya, A., Naik, M., and Jana, S., "Oculofacial surgical planning: True-to-scale 3D feature quantification using multicamera network," in [*2016 IEEE Annual India Conference (INDICON)*], (2016).

[25] Pollefeys, M., Koch, R., and Van Gool, L., "Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters," *International Journal of Computer Vision* **32**(1), 7–25 (1999).

[26] Huang, L., Zhang, Q., and Asundi, A., "Flexible camera calibration using not-measured imperfect target," *Applied Optics* **52**(25), 6278–6286 (2013).

[27] Hartley, R. and Zisserman, A., [*Multiple view geometry in computer vision*], Cambridge university press (2003).

[28] Chen, J., Mustafa, W., Siddig, A., and Kulesza, W., "Applying dithering to improve depth measurement using a sensor-shifted stereo camera," *Metrology and Measurement Systems* **17**(3), 335–347 (2010).

[29] Tzou, C.-H. J., Artner, N. M., Pona, I., Hold, A., Placheta, E., Kropatsch, W. G., and Frey, M., "Comparison of three-dimensional surface-imaging systems," *Journal of Plastic, Reconstructive & Aesthetic Surgery* **67**(4), 489–497 (2014).

[30] Esteban, I., Dijk, J., and Groen, F., "Fit3D toolbox: multiple view geometry and 3D reconstruction for matlab," in [*Electro-Optical Remote Sensing, Photonic Technologies, and Applications IV*], **7835**, International Society for Optics and Photonics (2010).

[31] Zhang, Z., "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence* **22**(11), 1330–1334 (2000).