Smart Camera Networks: An Analytical Framework for Auto Calibration without Ambiguity

Kiran Kumar Vupparaboina
Dept. Of Electrical Engineering
IIT Hyderabad
Andhra Pradesh, India-502205
Email: ee11p012@iith.ac.in

Kamala Raghavan
Dept. Of Electrical Engineering
IIT Madras
Tamil Nadu, India-600036
Email: ee10b126@smail.iitm.ac.in

Soumya Jana
Dept. Of Electrical Engineering
IIT Hyderabad
Andhra Pradesh, India-502205
Email: jana@iith.ac.in

***Abstract*—With the proliferation of smart environment, smart multi-camera networks assume growing significance. Specifically, non-intrusive calibration of such camera networks becomes imperative in smart applications such as telepresence systems, where multi-view imaging/recording needs to be performed in a dynamic setting with continuously changing intrinsic and extrinsic camera parameters. Unfortunately, popular auto calibration methods are known to introduce ambiguity or require manual intervention. In this backdrop, we propose a three-camera configuration (which can be generalized) with a stereo pair having known baseline distance and an additional (mono) camera positioned arbitrarily, and analytically establish the uniqueness of auto calibration in the proposed configuration.**

I. INTRODUCTION

In a variety of everyday applications, including security and surveillance, human-computer interfaces, smart homes and offices, smart cameras have begun to play a significant role. At the same time, certain specialized systems such as cooperative robotic networks and 3D telepresence equipment are built around smart multi-camera networks. Such smart systems often require the generation of 3D geometry/model of the scene from multiple 2D views captured by various cameras. As an illustration, consider a 3D telepresence system, schematically depicted in Fig. 1. A multitude of cameras take images of a target 3D object from various perspectives, and the geometry of the 3D object is reconstructed from those views by algorithmically solving the aforementioned inverse problem [1], [2]. In general, such 3D reconstruction algorithms require the knowledge of intrinsic camera parameters such as focal length as well as extrinsic camera parameters such as camera position and orientation for each camera in the network. Unfortunately, such parameters often vary with time. For example, during an extended telepresence session, the object of interest (possibly a human subject) potentially moves around, and hence each of the various camera potentially requires to adjust position as well as pose, and possibly focal length, in order to keep its focus on that object. In this backdrop, calibration, i.e., the task of obtaining various camera parameters, assumes significance.

Calibration strategies generally fall into two broad categories: (i) Calibration using an external calibration object, and (ii) Auto calibration where no extraneous object is necessary. In the first category, the external object is often chosen as a chess board due to its regular geometry. Tsai pioneered such calibration technique [3], which Zhang simplified later [4]. Other calibration objects have since been used including

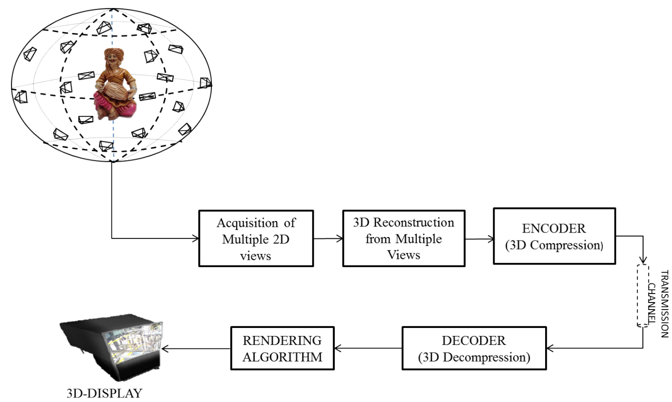

Fig. 1: Schematic diagram of a 3D telepresence system.

blinking lights on two ends of a rod [5], and multicolored cubes [6]. Although the calibration accuracy per se could be satisfactory, the main limitation of such techniques lies in their inherent intrusive nature. For example, it may not be practical to introduce external calibration objects during a real-time telepresence session or a medical procedure. Instead, in such applications one would prefer a non-intrusive method where the features from typical scenes/objects provide sufficient information to estimate the required camera parameters, both intrinsic and extrinsic.

Auto calibration techniques attempt to achieve the above goal, although various popular ones are known to allow ambiguities. The literature takes two distinct approaches to auto calibration. The first approach attempts to solve Kruppa equations, derived from the epipolar geometry of the absolute conic, which brings out the fact that the image of the absolute conic depends only on intrinsic parameters of the camera, and is invariant to geometric transformations [7]. Maybank and Faugeras reported early results in this direction [8]. Unfortunately, Kruppa equations lead to ambiguity because the solution for the fundamental matrix itself is nonunique. In addition, scaling of the aforementioned technique to large camera networks involves high complexity.

Consequently, a different approach involving factorization is often preferred in addressing auto calibration of large-scale multi-camera networks. An early attempt in this direction was made by Trigg [9], which has since been improved by

978-1-4799-2178-2/13/$31.00 ©2013 IEEE

Han and Kanade [10]. Specifically, the image matrix formed from multiple views is factorized with the help of singular value decomposition (SVD). Even this advancement provides a solution to the calibration problem only upto a projective ambiguity. Later Svoboda *et al.* removed the aforementioned ambiguity by introducing a virtual calibration object by projecting a laser pointer into the scene [11]. While Svoboda's method managed to put the constraint on the scene necessary to ensure unambiguous calibration, it also mandated external intervention.

In this backdrop, the exploration narrows down to finding an unambiguous auto calibration method that does not require manual intervention. Towards this goal, Lerma *et al.* consider a three-camera configuration, and imposes a baseline distance constraint between each of the three camera pairs [12]. Those authors empirically observe that such constraints improve the photogrammetric solution. However, a rigorous analytical justification is not provided. Further, the scalability of their method to large camera networks remains uncertain. In comparison, in this paper we propose a three-camera network with less restrictions. In particular, two of those cameras form a stereo pair with parallel optical axes with known baseline separation, and the third camera is arbitrarily chosen and positioned. In the proposed setup, we provide an analytical framework for auto calibration without ambiguity. Specifically, we advance the state of the art in two ways: (i) The three baseline distance constraints of the aforementioned work are now relaxed to only one, and (ii) unique estimation of parameters, both intrinsic and extrinsic, for all three cameras is rigorously established. Finally, although we demonstrate our technique for a three-camera network, our analysis extends to large multi-camera networks as long as the said network includes a stereo pair with known baseline distance as above, and certain technical assumptions are fulfilled.

II. Background

A. Linear Image Formation Model

Projection of a 3D point $[X\ Y\ Z\ 1]^T$ in homogeneous world coordinate onto the image plane by a camera is mathematically modeled by the linear transformation:

$$s\bar{x} = K[R|t]\bar{X}, \tag{1}$$

where $\bar{x} = [x\ \ y\ \ 1]^T$ indicates the homogeneous image coordinate,

$$K = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

indicates the intrinsic camera parameter matrix which include effective focal lengths (f_x, f_y) along both x and y directions (note: $f_x = f_y$ in case of square pixels), γ is the skew factor between x- and $y-$axis, principal point (u_0, v_0),

$$R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$$

denotes an unitary rotation matrix indicating the pose of the camera,

$$t = [\ t_X\ t_Y\ t_Z\ ]^T$$

denotes the vector of translation parameters along $X-$, $Y-$, and $Z-$axis, while s is a scaling/magnification factor. The linear model (1) of image formation provides an approximation where nonlinearity due to lens distortion is ignored.

B. Calibration using Extraneous Object

Many multicamera networks use traditional photogrammetric methods such as those due to Tsai [3] and Zhang [4] to calibrate one camera at a time. In Zhang's method, known metric information of chessboard corners is related to corresponding image coordinates to obtain the calibration parameters. In particular, the world coordinate is chosen to align the chessboard with the $Z = 0$ plane, and the $X-$ and $Y-$ axes with the chessboard sides, so that the coordinates of the corner points are known given the length of each square. Hence projective homography H between corner points in the image and corresponding 3D coordinates is given by

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ y_{11} & y_{12} & \dots & y_{1N} \\ 1 & 1 & \dots & 1 \end{bmatrix} = H \begin{bmatrix} X_1 & X_2 & \dots & X_N \\ Y_1 & Y_2 & \dots & Y_N \\ 1 & 1 & \dots & 1 \end{bmatrix}, \tag{2}$$

where

$$H = \frac{1}{s} \begin{bmatrix} r_1 & r_2 & t_X \\ r_4 & r_5 & t_Y \\ r_7 & r_8 & t_Z \end{bmatrix}.$$

Here the 3-rd column of the rotation matrix R is ignored because all 3D points of interest lies on the $Z = 0$ plane. As both 3D points and there images are known, the matrix H can be obtained from (2). Subsequently, various camera parameters can be found from H in a straightforward manner. In real-time applications, introducing a chessboard at $Z = 0$ each time camera parameters vary may not be realistic. For the same reason, improved techniques making use of other extraneous objects could still pose similar difficulties [5], [6].

C. Auto Calibration

1) Auto calibration using Kruppa equations: Kruppa equations relate image of the absolute conic with intrinsic parameters, and has been used for auto calibration [7], [8]. However, such equation allows ambiguity. To see this, consider a point $\bar{X} = [X\ Y\ Z\ 0]^T$ in homogeneous coordinates on the absolute conic (i.e., the conic section on the plane at infinity), and denote by $\bar{x}$ its projection onto the image plane. From (1), we obtain

$$\begin{aligned} \bar{x} &= K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} \\ &= KR\bar{X}', \end{aligned}$$

where $\bar{X}' = [X\ Y\ Z]^T$. Rearranging, one can write

$$\bar{X}' = RK^{-1}\bar{x}. \tag{3}$$

Since $\bar{X} = [X\ Y\ Z\ 0]^T$ lies on the absolute conic, it satisfies

$$\bar{X}'^T\bar{X}' = 0. \tag{4}$$

Using (3) in (4), and simplifying, one obtains

$$\bar{x}^T C\bar{x} = 0, \tag{5}$$

where $C = K^{-T}K^{-1}$ is the image of the absolute conic, which depends only on the intrinsic parameters and invariant to geometric transformations.

Further, let F indicate the fundamental matrix of two cameras with identical intrinsic parameters, and $F = U\Sigma V^T$ indicates its singular value decomposition (SVD), where U,V are orthogonal matrices and

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is the diagonal matrix containing singular values. Denoting the dual of image of the absolute conic by $D = KK^T$, (5) leads to the Kruppa equation

$$\frac{v_2^T D v_2}{\sigma_1^2 u_1^T D u_1} = \frac{-v_2^T D v_1}{\sigma_1 \sigma_2 u_1^T D u_2} = \frac{v_1^T D v_1}{\sigma_2^2 u_2^T D u_2}, \tag{6}$$

where u_i, v_i denotes i-th ($i = 1, 2$) column of the matrices U and V. Solving (6), one obtains intrinsic parameter matrix K. However, as various intrinsic matrices K may give rise to the same fundamental matrix F, the above solution is not unique. Also, methods based on Kruppa equations are hard to generalize beyond stereo pairs with identical intrinsic parameters.

2) Auto calibration via factorization of image matrix: In contrast with Kruppa-equations-based methods, a factorization method due to Han and Kanade applies to large camera networks [10]. The image formation of N 3D points via M cameras can be written as

$$W = PX, \tag{7}$$

where

$$W = \begin{bmatrix} s_{11}\begin{bmatrix} x_{11} \\ y_{11} \\ 1 \end{bmatrix} & \dots & s_{1N}\begin{bmatrix} x_{1N} \\ y_{1N} \\ 1 \end{bmatrix} \\ \vdots & \ddots & \vdots \\ s_{M1}\begin{bmatrix} x_{M1} \\ y_{M1} \\ 1 \end{bmatrix} & \dots & s_{MN}\begin{bmatrix} x_{MN} \\ y_{MN} \\ 1 \end{bmatrix} \end{bmatrix}$$

indicates the image matrix. Each row of W corresponds to images due to one particular camera, while

$$P = \begin{bmatrix} K_1[R_1|t_1] \\ K_2[R_2|t_2] \\ \vdots \\ K_M[R_M|t_M] \end{bmatrix}$$

is the overall projective transformation matrix, in which K_i and $[R_i|t_i]$ denotes intrinsic and extrinsic parameters of i-th camera ($i = 1, 2, ..., M$), and X is the collection of N 3D points. Noting that the right hand side of (7) has rank four, the image matrix W is approximated by a rank-4 matrix $\hat{W}$ based on SVD. An estimate $(\hat{P}, \hat{X})$ of (P, X) can be obtained by factorizing $\hat{W} = \hat{P}\hat{X}$. However, such factors are clearly nonunique, as $\hat{P}H$ and $H^{-1}\hat{X}$ are clearly factors as well, for any 4×4 invertible matrix H.

Fig. 2: Proposed three-camera configuration with a stereo pair with known baseline distance and an unconstrained mono camera.

3) Auto calibration without ambiguity: While the aforementioned factorization technique is quite general in scope, additional constraints are needed to remove the associated ambiguity. To this end, Svoboda *et al.* introduced a virtual calibration object by projecting a pointer into the scene, thereby putting constraints required to ensure uniqueness [11]. While attractive, this scheme still requires external intervention. It has since been observed by Lerma *et al.* that constraining baseline distances between cameras make calibration more accurate [12]. In particular, pairwise baseline distance constraints in a three-camera network help refine parameters obtained using epipolar geometry and fundamental matrix.

III. New Auto Calibration Framework

Inspired by reported effectiveness of baseline constraints, and attempting to avoid constraints requiring manual intervention, we explored whether auto calibration could be achieved without ambiguity. In this process, we aimed to restrict our system as little as possible.

A. Proposed Three-Camera Configuration

With a view to exploiting baseline distance constraint, while not overly restricting the system, we consider a three-camera configuration, where two of the cameras form a stereo pair with parallel optical axes with known baseline separation, and the third camera is arbitrarily chosen and positioned. See Fig. 2 for a schematic depiction. In practice, one may separately deploy a stereo camera pair and a mono camera, which are readily available, to implement our three-camera configuration. In the following, we analytically demonstrate that the proposed camera network admits a unique calibration under mild conditions. Our analysis assumes significance as similar demonstration of uniqueness has so far not been attempted for baseline-distance-constrained systems to the best of our knowledge.

B. Formalism

At this point, we turn to formally pose the calibration problem at hand. Consider N feature points on the 3D object in the 3D world coordinate system, each of which is visible through every one of the three cameras. Denote the j-th 3D point by $\bar{X}_j = [X_j \; Y_j \; Z_j \; 1]^T$ $(j = 1, 2, ..., N)$, and its projection on the image plane (in local image coordinates) of the i-th camera $(i = 1, 2, 3)$ is given by $\bar{x}_{ij} = [x_{1j} \; y_{1j} \; 1]^T$. Our task is to uniquely obtain the intrinsic and extrinsic parameters of all three cameras at hand from the given baseline distance l in the stereo pair, and the image points $\{\bar{x}_{1j}, \bar{x}_{2j}, \bar{x}_{3j}\}_{j=1}^N$. Of course, such unique calibration would also lead to the recovery of original feature points $\{\bar{X}_j\}_{j=1}^N$. Towards our goal, we find the desired solution in a closed form from the linear model (1) of image formation. For the sake of simplicity, we ignore nonlinear distortions.

1) Image formation by stereo camera pair: Without loss of generality, we shall index by '1' and '2' the two cameras in the stereo pair, and by '3' the mono camera. Further, fix the origin of the world coordinate system at the center of Camera 1. Next assume that the optical axes of Cameras 1 and 2 are parallel, and the baseline is aligned with X-axis, so that Camera 2 has its center at $[l \; 0 \; 0]^T$, where l is the given baseline distance of the stereo pair as mentioned earlier. Moreover, denoting by f_1 and f_2 the respective focal lengths of Cameras 1 and 2, and assuming zero skew factors and perfectly centered image planes, we obtain from (1) the following image formation equations for the stereo pair:

$$s_{1j}\bar{x}_{1j} = \begin{bmatrix} f_1 & 0 & 0 \\ 0 & f_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \bar{X}_j \tag{8}$$

$$s_{2j}\bar{x}_{2j} = \begin{bmatrix} f_2 & 0 & 0 \\ 0 & f_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -l \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \bar{X}_j, \tag{9}$$

where s_{1j} and s_{2j} are scaling/magnification factors for Cameras 1 and 2, respectively, the feature point $\bar{X}_j$ $(j = 1, 2, ..., N)$.

Rewriting (8) and (9), we obtain

$$\begin{aligned} s_{1j}x_{1j} &= f_1 X_j \quad (10) & s_{2j}x_{2j} &= f_2(X_j - l) \quad (13) \\ s_{1j}y_{1j} &= f_1 Y_j \quad (11) & s_{2j}y_{2j} &= f_2 Y_j \quad (14) \\ s_{1j} &= Z_j \quad (12) & s_{2j} &= Z_j. \quad (15) \end{aligned}$$

Now, from (10)–(15), we obtain unique expressions for X_j and Y_j $(j = 1, 2, ..., N)$ as follows. Dividing (10) by (11), we obtain

$$Y_j = \frac{y_{1j}}{x_{1j}} X_j. \tag{16}$$

Similarly, dividing (13) by (14), we obtain

$$Y_j = \frac{y_{2j}}{x_{2j}} (X_j - l). \tag{17}$$

Now, equating (16) and (17), we get the solution

$$X_j = \frac{x_{1j}y_{2j}l}{x_{1j}y_{2j} - x_{2j}y_{1j}}. \tag{18}$$

Finally, substituting (18) in (16), we also get

$$Y_j = \frac{y_{1j}y_{2j}l}{x_{1j}y_{2j} - x_{2j}y_{1j}}. \tag{19}$$

Thus (18) and (19) provide unique solutions for $\{(X_j, Y_j)\}_{j=1}^N$.

At the same time, dividing (11) by (14), and rearranging, we obtain

$$f_2 = \frac{y_{2j}}{y_{1j}} f_1. \tag{20}$$

Similarly, dividing (10) by (12), and rearranging, we get

$$Z_j = \frac{f_1 X_j}{x_{1j}}. \tag{21}$$

Clearly, (20) and (21) fail to provide unique solution to f_1, f_2 and $\{Z_j\}_{j=1}^N$, as f_1 could be arbitrarily chosen without violating any governing equation.

2) Image formation by mono camera: Next we show the following interesting result. The images of feature points formed by the mono camera, although arbitrarily configured, when considered in conjunction with those formed by the stereo pair, are sufficient to ensure the desired uniqueness. Image formation by the mono camera, i.e., Camera 3, is given by

$$s_{3j}\bar{x}_{3j} = \begin{bmatrix} f_3 & 0 & 0 \\ 0 & f_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} r_1 & r_2 & r_3 & t_X \\ r_4 & r_5 & r_6 & t_Y \\ r_7 & r_8 & r_9 & t_Z \end{array}\right] \bar{X}_j \tag{22}$$

Parameters in (22) have same meaning as that of (1) described in section-II. Now in addition to f_1, f_2 and Z_j we have 5 intrinsic and 6 extrinsic parameters of mono camera to find. Analyzing (22) will help in bring a closed form expression to find all the aforementioned parameters. Expanding equation (22), we get

$$\begin{aligned} s_{3j}x_{3j} &= f_3(r_1X_j + r_2Y_j + r_3Z_j + t_X) \quad (23) \\ s_{3j}y_{3j} &= f_3(r_4X_j + r_5Y_j + r_6Z_j + t_Y) \quad (24) \\ s_{3j} &= r_7X_j + r_8Y_j + r_9Z_j + t_Z \quad (25) \end{aligned}$$

Now, dividing (23) by (25), we get

$$x_{3j} = \frac{f_3(r_1X_j + r_2Y_j + r_3Z_j + t_X)}{r_7X_j + r_8Y_j + r_9Z_j + t_Z}, \tag{26}$$

and dividing (24) by (25), we get

$$y_{3j} = \frac{f_3(r_4X_j + r_5Y_j + r_6Z_j + t_Y)}{r_7X_j + r_8Y_j + r_9Z_j + t_Z}. \tag{27}$$

Cross multiplying and rearranging terms in (26), we obtain

$$\begin{aligned} x_{3j} = \quad & X_j\frac{f_3r_1}{t_Z} + Y_j\frac{f_3r_2}{t_Z} + \frac{f_3r_3Z_j}{t_Z} + \frac{f_3t_X}{t_Z} \\ & -x_{3j}X_j\frac{r_7}{t_Z} - x_{3j}Y_j\frac{r_8}{t_Z} - x_{3j}\frac{r_9Z_j}{t_Z}. \end{aligned} \tag{28}$$

In the above equation we have Z_j which is unknown Z-coordinate of the j-th point in world coordinate system. By using (21) we can express the Z_j in terms of f_1 which remains same for every point.

Next substituting (21) in (28), we get

$$\begin{aligned} x_{3j} = X_j\frac{f_3r_1}{t_Z} + Y_j\frac{f_3r_2}{t_Z} + \frac{X_j}{x_{1j}}\frac{f_1f_3r_3}{t_Z} + \frac{f_3t_X}{t_Z} \\ - x_{3j}X_j\frac{r_7}{t_Z} - x_{3j}Y_j\frac{r_8}{t_Z} - \frac{x_{3j}X_j}{x_{1j}}\frac{f_1r_9}{t_Z}. \end{aligned} \tag{29}$$

Similarly, from (27), we get

$$y_{3j} = X_j\frac{f_3 r_4}{t_Z} + Y_j\frac{f_3 r_5}{t_Z} + \frac{X_j}{x_{1j}}\frac{f_1 f_3 r_6}{t_Z} + \frac{f_3 t_Y}{t_Z} - y_{3j}X_j\frac{r_7}{t_Z} - y_{3j}Y_j\frac{r_8}{t_Z} - \frac{y_{3j}X_j}{x_{1j}}\frac{f_1 r_9}{t_Z}. \tag{30}$$

At this point, We make use of (29) and (30) to express the problem as a multi-objective constrained optimization.

C. Multi-objective constrained optimization

Right hand sides of (29) and (30) represent the x and y coordinates of reprojected j-th 3D point back to the image plane of the mono camera, while left hand side represents measured coordinates. We try obtain the objective function that minimizes reprojection error between both x and y coordinates combinedly, hence called multi-objective. First we bring (29) and (30) to standard mathematical structure and then define the objective function.

Rewrite (29) as

$$x_{3j} = \begin{bmatrix} X_j \\ Y_j \\ \frac{X_j}{x_{1j}} \\ 1 \\ -x_{3j}X_j \\ -x_{3j}Y_j \\ -\frac{x_{3j}X_j}{x_{1j}} \end{bmatrix}^T \begin{bmatrix} \frac{f_3 r_1}{t_Z} \\ \frac{f_3 r_2}{t_Z} \\ \frac{f_3 f_1 r_3}{t_Z} \\ \frac{f_3 t_X}{t_Z} \\ \frac{r_7}{t_Z} \\ \frac{r_8}{t_Z} \\ \frac{f_1 r_9}{t_Z} \end{bmatrix}. \tag{31}$$

Note that (31) is overdetermined (for sufficiently large N), and takes the form

$$A_1\bar{x}_1 = \bar{b}_1, \tag{32}$$

where

$$A_1 = \begin{bmatrix} X_1 & Y_1 & \frac{X_1}{x_{11}} & 1 & -x_{31}X_1 & -x_{31}Y_1 & -\frac{x_{31}X_1}{x_{11}} \\ \vdots & & & \ddots & & & \vdots \\ X_N & Y_N & \frac{X_N}{x_{1N}} & 1 & -x_{3N}X_N & -x_{3N}Y_N & -\frac{x_{3N}X_N}{x_{1N}} \end{bmatrix}$$

$$\bar{x}_1 = \begin{bmatrix} \frac{f_3 r_1}{t_Z} & \frac{f_3 r_2}{t_Z} & \frac{f_3 f_1 r_3}{t_Z} & \frac{f_3 t_X}{t_Z} & \frac{r_7}{t_Z} & \frac{r_8}{t_Z} & \frac{f_1 r_9}{t_Z} \end{bmatrix}^T$$

$$\bar{b}_1 = \begin{bmatrix} x_{31} & \dots & x_{3N} \end{bmatrix}^T.$$

Similarly, (30) can be written

$$y_{3j} = \begin{bmatrix} X_j \\ Y_j \\ \frac{X_j}{x_{1j}} \\ 1 \\ -y_{3j}X_j \\ -y_{3j}Y_j \\ -\frac{y_{3j}X_j}{x_{1j}} \end{bmatrix}^T \begin{bmatrix} \frac{f_3 r_4}{t_Z} \\ \frac{f_3 r_5}{t_Z} \\ \frac{f_3 f_1 r_6}{t_Z} \\ \frac{f_3 t_Y}{t_Z} \\ \frac{r_7}{t_Z} \\ \frac{r_8}{t_Z} \\ \frac{f_1 r_9}{t_Z} \end{bmatrix}, \tag{33}$$

which is also a overdetermined system in the form

$$A_2\bar{x}_2 = \bar{b}_2, \tag{34}$$

where

$$A_2 = \begin{bmatrix} X_1 & Y_1 & \frac{X_1}{x_{11}} & 1 & -y_{31}X_1 & -y_{31}Y_1 & -\frac{y_{31}X_1}{x_{11}} \\ \vdots & & & \ddots & & & \vdots \\ X_N & Y_N & \frac{X_N}{x_{1N}} & 1 & -y_{3N}X_N & -y_{3N}Y_N & -\frac{y_{3N}X_N}{x_{1N}} \end{bmatrix}$$

$$\bar{x}_2 = \begin{bmatrix} \frac{f_3 r_4}{t_Z} & \frac{f_3 r_5}{t_Z} & \frac{f_3 f_1 r_6}{t_Z} & \frac{f_3 t_Y}{t_Z} & \frac{r_7}{t_Z} & \frac{r_8}{t_Z} & \frac{f_1 r_9}{t_Z} \end{bmatrix}^T$$

$$\bar{b}_2 = \begin{bmatrix} y_{31} & \dots & y_{3N} \end{bmatrix}^T.$$

In view of (32) and (34), one needs to solve the multi-objective optimization problem minimizing reprojection error

$$\Phi^* = \arg\min_{\Phi} \lambda\|A_1\bar{x}_1 - \bar{b}_1\| + (1-\lambda)\|A_2\bar{x}_2 - \bar{b}_2\| \tag{35}$$

in order to estimate camera parameter vector

$$\Phi = \{f_1, f_3, r_1, r_2, r_4, t_X, t_Y, t_Z\}.$$

Here λ determines the relative weights on the respective objectives. Note that the uniqueness is ensured from the overdetermined nature of the problem.

Also note that all entries of the rotation matrix of mono camera are not considered as unknown quantities. This is because only three parameters, namely r_1, r_2 and r_4, determines the rotation matrix within eight alternatives as shown in the Appendix. In order to avoid ambiguity from this perspective, one can solve eight variants of the optimization problem (35), and the best of these solutions in terms of reprojection error.

IV. Discussion

In this paper, we proposed a three-camera configuration, where two of the cameras form a stereo pair with parallel optical axes with known baseline separation, and the third camera is arbitrarily chosen and positioned. We further presented a mathematical framework which demonstrates unambiguous auto-calibration of multi-camera networks in the proposed configuration. An expression has been obtained as a unique solution to a multi-objective constrained optimization function that minimizes the reprojection error. Care has been taken so that such expression preserves the structure and properties of the projective transformation matrix. Here we hasten to add that our solution is based on linear idealization of the image formation process. In order to apply our method for practical applications such as the 3D telepresence system, additional steps including lens distortion estimation will be necessary. Although we demonstrated our system for a three camera network, our analysis extends to general multi-camera networks as long as the said network includes a stereo pair with known baseline distance, and certain technical assumptions are fulfilled.

Appendix

One can easily see that a unitary rotation matrix $R = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix}$ is completely determined by r_1, r_2 and r_4 upto a finite number of possibilities. To this end, note the following. First r_3 can be obtained by equating norm of the first row to one, thus we get

$$r_3 = \pm\sqrt{1 - r_1^2 - r_2^2}. \tag{36}$$

Similarly r_7 and r_8 are given by

$$\begin{aligned} r_7 &= \pm\sqrt{1 - r_1^2 - r_4^2} \\ r_8 &= \pm\sqrt{1 - r_2^2 - r_5^2}. \end{aligned}$$

Now r_5 in (37) is still an unknown, which can be obtained by equating inner product of first two columns to zero and substituting (37) and (37). Thus r_5 is given by

$$r_5 = \frac{-r_1 r_2 r_4}{1 - r_1^2} \pm \frac{\sqrt{1 - 2r_1^2 - r_2^2 - r_4^2 + r_1^2 r_2^2 + r_1^2 r_4^2 + r_2^2 r_4^2 + r_1^4}}{1 - r_1^2}. \quad (37)$$

Next, using orthonormality, r_6 and r_9 can be expressed as

$$r_6 = \pm\sqrt{1 - r_4^2 - r_5^2} \quad (38)$$

$$r_9 = \pm\sqrt{1 - r_7^2 - r_8^2}. \quad (39)$$

As a result, for every choice of r_1, r_2 and r_4, there can be at most eight possible choices of unitary matrix R.

REFERENCES

[1] H. Aghajan, A. Cavallaro, "Multi-Camera Networks: Principles and Applications," *Academy Press*, May 2009.

[2] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press*, Second Edition, March 2004.

[3] R.Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses,"*IEEE J. Robotics and Automation* vol. 3, no. 4, pp. 323-344, Aug.1987.

[4] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, No. 11, November 2000.

[5] G. Kurillo, L. Zeyu, R. Bajcsy, "Wide-area external multi-camera calibration using vision graphs and virtual calibration object," *Second ACM/IEEE International Conference on Distributed Smart Cameras, 2008*, pp.1-9, 7-11 Sept. 2008.

[6] H. Baker, D. Tanguay, I. Sobel, D. Gelb, M. Gross, W. Culbertson, T.Malzenbender, " The coliseum immersive teleconferencing system," *In Proceedings of International Workshop on Immersive Telepresence* France 2002.

[7] R.I. Hartley, "Kruppas equations derived from the fundamental matrix," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, no. 2, 133–135, February 1997.

[8] S. Maybank, O.D. Faugeras, "A theory of self-calibration of a moving camera," *International Journal on Computer Vision*, vol. 8, no. 2, 123–151, 1992.

[9] B. Triggs, "Autocalibration from Planar Scenes,"*Proc. Fifth European Conf. Computer Vision*, pp. 89-105, June 1998.

[10] M. Han, T. kanade,"Creating 3D Models with Uncalibrated Cameras,"*Applications of Computer Vision, 2000, Fifth IEEE Workshop*, pp.178–185, 2000.

[11] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *PRESENCE: Teleoperators and Virtual Environments, MIT Press*, pp 407-422, vol. 14, no. 4, August 2005.

[12] J.L. Lerma, S. Navarro, M. Cabrelles, A.E. Segu, "Camera calibration with baseline distance constraints," *Photogrammetric Record*, vol. 25, no. 130, pp. 140–158, 2010.