Plane-based Calibration Algorithm for Multi-camera Systems via Factorization of Homography Matrices

Toshio Ueshiba Fumiaki Tomita
National Institute of Advanced Industrial Science and Technology (AIST)
AIST Tsukuba Central 2, Tsukuba, 305-8568 Japan
{t.ueshiba, f.tomita}@aist.go.jp

Abstract

A new calibration algorithm for multi-camera systems using a planar reference pattern is proposed. The algorithm is an extension of Sturm-Maybank-Zhang style plane-based calibration technique for use with multiple cameras. Rigid displacements between the cameras are recovered as well as the intrinsic parameters only by capturing with the cameras a model plane with known reference points placed at three or more locations. Thus the algorithm yields a simple calibration means for stereo vision systems with an arbitrary number of cameras while maintaining the handiness and flexibility of the original method. The algorithm is based on factorization of homography matrices between the model and image planes into the camera and plane parameters. To compensate for the indetermination of scaling factors, each homography matrix is rescaled by a double eigenvalue of a planar homology defined by two views and two model planes. The obtained parameters are finally refined by a non-linear maximum likelihood estimation (MLE) process. The validity of the proposed technique was verified through simulation and experiments with real data.

1 Introduction

Recovering the three-dimensional (3D) structure of the scene from multiple images has been one of the most fundamental topics and attracting much attention in the vision community for a long time. Knowing both the intrinsic and extrinsic camera parameters, referred to as *camera calibration*, is an essential prerequisite for the reconstruction process. To meet the demands, many calibration techniques have been developed aiming at attaining high accuracy with minimum elaboration.

Classical camera calibration is performed by capturing a 3D reference object with a known Euclidean structure[14]. This type of technique yields the best results if the 3D geometry of the reference object is known with high accuracy. In addition, it is directly applicable to multi-camera systems by simply repeating the calibration process independently for each camera. However, setting up the 3D reference object with great accuracy is an elaborate task that requires special equipment and becomes more difficult as the dimensions of the view volume increase.

To avoid such difficulties, a simple and practical camera calibration technique using a *model plane* with a known 2D reference pattern was proposed by Sturm-Maybank[10] and Zhang[16] independently. The user freely places the model plane or the camera at two or more locations and captures images of the reference points. Camera parameters are recovered from homographies between the model plane and the image plane computed from correspondences between the reference points and their projections. Since the motion of the model plane or the camera need not be known, this technique is very handy in practice because the extent of the reference object is a 2D plane instead of a 3D volume.

Although this algorithm is very practical and yields good results when calibrating a *single* camera, a problem arises when applied to *multi-camera* systems by simply repeating the calibration procedure for each camera; calibrating each camera gives, as by-products, the positions and orientations of the model planes relative to the camera as well as the camera intrinsic parameters. Using this information, rigid transformations between any pairs of cameras could be determined through an arbitrarily chosen model plane. These transformations should be invariant, irrespective of the plane through which they are computed (see Fig. 1). However this constraint may not be necessarily satisfied in the presence of noise because the calibration is performed for each camera independently. This inconsistency can degrade the estimation accuracy of the relative displacements between the cameras, potentially causing a serious problem for multi-camera systems.

Sturm[8] proposed a pose estimation technique from multiple planar patterns with known metric structure incor-

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

IEEE COMPUTER SOCIETY

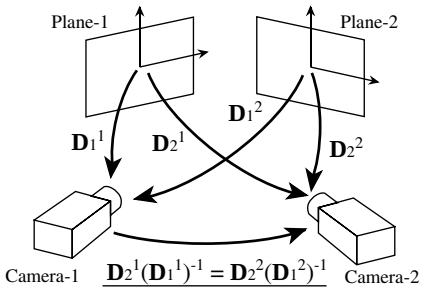


Figure 1. Consistency constraint in relative displacements between cameras

porating this consistency constraint. This method, however, assumes that all the cameras are calibrated beforehand. We show this restriction can be relaxed so that both camera calibration and pose estimation are solved simultaneously.

Proposed in this paper is integrating the Sturm-Maybank-Zhang style calibration technique[10, 16] with Sturm's pose estimation method[8] into a factorization-based multi-camera calibration algorithm implicitly incorporating the above-mentioned geometric constraint. Though the factorization technique was originally developed for 3D reconstruction of point features[11], recently it has also been applied to lines or planes[13, 7] since having been introduced in the extended version of a pioneer work[12] proposing self-calibration from planar scenes. In those works, a measurement matrix is constructed by stacking *inter-image* homographies induced by 3D planes and then factored into camera motions and object shapes. Though not categorized as a factorization method, Malis and Cipolla[4, 5] have proposed relevant self-calibration algorithms for planar scenes exploiting the constraints existing in the measurement matrix consisting of inter-image homographies. In contrast, our proposed algorithm factor the measurement matrix made from *plane-to-image* homographies into camera and plane parameters as in [8].

We first show that the homography between the model plane and the image plane is a composition of a camera projection matrix and a plane parameter matrix (section 2). This naturally leads us to constructing a measurement matrix by stacking homography matrices and then decomposing it into the camera projections and the plane parameters[11]. However, since each homography matrix can be determined only up to an unknown scale, we have to know the appropriate scaling factors that make the measurement matrix decomposable. To solve this problem, we use the recent findings on multi-view constraints on homography of Zelnik-Manor and Irani[15]. After factorization, the obtained projection matrices and plane parameter matrices are transformed from projective to Euclidean coordinate frames using the metric information of the reference planes. Finally all the estimation parameters are refined by non-linear optimization to achieve a statistically optimal and unbiased result (section 3). The validity of our proposed technique was verified by simulation and real experiments (section 4).

Notation Throughout this paper, vectors and matrices are denoted in bold-face. Especially, $\mathbf{0}$ and $\mathbf{I}$ represent the zero vector and identity matrix, respectively. Symbol $\top$ represents the transposition of vectors or matrices. Therefore $\mathbf{x}^\top\mathbf{y}$ is a scalar product of two vectors, $\mathbf{x}$ and $\mathbf{y}$. The equality of two vectors or matrices up to a non-zero scalar is denoted by the symbol $\simeq$.

2 Plane-to-image Homographies

Suppose that we have I cameras. The position and the orientation of the i-th camera are represented by a 3-vector $\mathbf{t}_i$ and a 3×3 orthogonal matrix $\mathbf{R}_i$, respectively. Without loss of generality, we assume that the world coordinate system is fixed to the first camera, i.e. $\mathbf{t}_1 = \mathbf{0}$ and $\mathbf{R}_1 = \mathbf{I}$. The image plane of the i-th camera is denoted by $\mathcal{I}_i$ $(i = 1, \ldots, I)$. Then a point in 3-space with inhomogeneous coordinates $\mathbf{X} \in \mathcal{R}^3$ is observed by the i-th camera as a 2D point on $\mathcal{I}_i$ at

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \simeq \underbrace{\mathbf{K}_i\mathbf{R}_i^\top \begin{bmatrix} \mathbf{I} & -\mathbf{t}_i \end{bmatrix}}_{\mathbf{P}_i} \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \tag{1}$$

where $\mathbf{K}_i$ is a 3×3 upper triangular matrix given by

$$\mathbf{K}_i = \begin{bmatrix} a_i k_i & s_i k_i & u_{0i} \\ 0 & k_i & v_{0i} \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

with the focal length k_i, the principal point $(u_{0i},\ v_{0i})$, the aspect ratio a_i and the skew parameter s_i[1]. The 3×4 matrix $\mathbf{P}_i$ in (1) is called the *camera matrix*.

Next we consider J planes π^j $(j = 1, \ldots, J)$ called the *model planes* with known 2D reference points. Let $\mathbf{p}^j$ and $\mathbf{q}^j$ be unit 3-vectors parallel to the horizontal and vertical axes of the 2D Euclidean coordinate frame fixed to π^j, and $\mathbf{d}^j$ be a 3-vector representing the position of its origin (see Fig. 2). From the orthogonality of the two axes and unity of $\mathbf{p}^j$ and $\mathbf{q}^j$, we have

$$\mathbf{p}^{j\top}\mathbf{p}^j = \mathbf{q}^{j\top}\mathbf{q}^j = 1, \quad \mathbf{p}^{j\top}\mathbf{q}^j = 0. \tag{3}$$

Then a point on π^j with coordinates (x, y) with respect to the 2D frame is located at

$$\begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{p}^j & \mathbf{q}^j & \mathbf{d}^j \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{Q}^j} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4}$$

IEEE COMPUTER SOCIETY

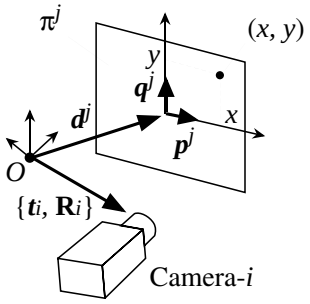


Figure 2. Geometrical configuration of cameras and model planes

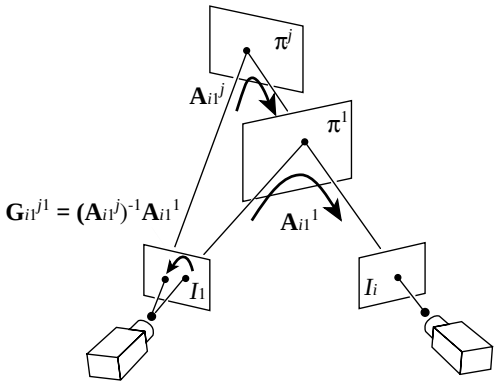


Figure 3. Inter-image and relative homographies

in 3-space where $\mathbf{Q}^j$ is a 4×3 matrix representing the model plane π^j and is referred to as the *plane matrix*.

Combining equations (1) and (4), we have

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \simeq \underbrace{\mathbf{P}_i\mathbf{Q}^j}_{\mathbf{H}_i^j} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{5}$$

Equation (5) means that a point (x, y) on π^j is mapped to a point (u, v) on $\mathcal{I}_i$ through a projective transformation (homography) represented by a non-singular 3×3 matrix $\mathbf{H}_i^j$ which is a composition of the camera matrix $\mathbf{P}_i$ and the plane matrix $\mathbf{Q}^j$. Each $\mathbf{H}_i^j$ can be recovered up to an unknown scale factor from the correspondences between the reference points on π^j and their projections on $\mathcal{I}_i$, i.e. $\{(x, y) \leftrightarrow (u, v)\}$.

3 Multi-camera Calibration Algorithm

3.1 Rescaling and factoring homography matrices

Stacking homography matrices $\mathbf{H}_i^j$ in (5) with all the cameras and model planes, we have

$$\underbrace{\begin{bmatrix} \mathbf{H}_1^1 & \cdots & \mathbf{H}_1^J \\ \vdots & & \vdots \\ \mathbf{H}_I^1 & \cdots & \mathbf{H}_I^J \end{bmatrix}}_{\mathbf{W}} = \underbrace{\begin{bmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_I \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \mathbf{Q}^1 & \cdots & \mathbf{Q}^J \end{bmatrix}}_{\mathbf{Q}}. \tag{6}$$

The $3I \times 3J$ matrix $\mathbf{W}$ is called the *measurement matrix*. Equation (6) says that this matrix is rank 4 and can be decomposed into a $3I \times 4$ matrix $\mathbf{P}$ representing the cameras and a $4 \times 3J$ matrix $\mathbf{Q}$ representing the model planes. Therefore, if all the plane-image homography matrices $\mathbf{H}_i^j$ are known including their absolute scales, we can recover the camera and the plane matrices (up to a common 4×4 non-singular matrix) using SVD in a manner similar to the factorization method proposed by Tomasi and Kanade[11].

Unfortunately, we can determine each homography matrix $\mathbf{H}_i^j$ from point correspondences only up to an unknown scale; all we have is $\tilde{\mathbf{H}}_i^j = \lambda_i^j\mathbf{H}_i^j$, where λ_i^j is an unknown scale factor. The measurement matrix $\tilde{\mathbf{W}}$ composed of $\tilde{\mathbf{H}}_i^j$ instead of $\mathbf{H}_i^j$ is not necessarily rank 4 and cannot be decomposed into the camera and plane matrices.

This situation is exactly the case where the factorization technique has been applied to perspective cameras in a structure-from-motion context[9, 3]. If cameras are calibrated, computing the scale factors λ_i^j is straightforward because the first two columns of the homography matrices are orthonormal[8]. In our case, however, recovering the scale factors is not simple because calibration is unknown. There are two solutions for this difficulty; one is to search for a set of scale factors $\{\lambda_i^j\}$ which makes $\tilde{\mathbf{W}}$ nearly rank 4 by iterative non-linear minimization[3]. Non-linear minimization, however, is computationally expensive and suffers from an initialization problem. We thus adopted another approach similar to [9] that exploits constraints on scale factors.

The model plane π^j induces an inter-image homography from $\mathcal{I}_1$ to $\mathcal{I}_i$ represented by $\mathbf{A}_{i1}^j = \mathbf{H}_i^j(\mathbf{H}_1^j)^{-1}$. Then a homography from $\mathcal{I}_1$ onto itself represented by $\mathbf{G}_{i1}^{j1} = (\mathbf{A}_{i1}^j)^{-1}\mathbf{A}_{i1}^1$ is a "relative homography" which maps a point from the first image to the i-th image through π_1 and then projects it back to the first image through π^j (see Fig. 3). Recently Zelnik-Manor and Irani[15] have shown that $\mathbf{G}_{i1}^{j1}$ is a planar homology[2] with a special form $\mathbf{G}_{i1}^{j1} = \mathbf{I} + \mathbf{e}\mathbf{f}^\top$ for some 3-vectors $\mathbf{e}$ and $\mathbf{f}$ (see appendix A), which means that $\mathbf{G}_{i1}^{j1}$ has a unit eigenvalue of multiplicity two[1]. Thus the relative homography matrix computed from the non-scaled plane-image homography matri-

[1] Associated subspace is of dimension two and perpendicular to $\mathbf{f}$.

IEEE COMPUTER SOCIETY

ces $\tilde{\mathbf{G}}_{i1}^{j1} = \tilde{\mathbf{H}}_1^j(\tilde{\mathbf{H}}_i^j)^{-1}\tilde{\mathbf{H}}_i^1(\tilde{\mathbf{H}}_1^1)^{-1} = (\lambda_i^1\lambda_1^j)/(\lambda_1^1\lambda_i^j)\mathbf{G}_{i1}^{j1}$ has an eigenvalue $\mu_i^j = (\lambda_i^1\lambda_1^j)/(\lambda_1^1\lambda_i^j)$ of multiplicity two which can be directly computed from $\tilde{\mathbf{G}}_{i1}^{j1}$. Using this value, we rescale $\tilde{\mathbf{H}}_i^j$ as

$$\tilde{\mathbf{H}}_i^j \leftarrow \mu_i^j\tilde{\mathbf{H}}_i^j \quad (i = 2, \ldots, I; j = 2, \ldots, J). \tag{7}$$

Since $\tilde{\mathbf{H}}_i^j = \mu_i^j\lambda_i^j\mathbf{H}_i^j = (\lambda_i^1\lambda_1^j/\lambda_1^1)\mathbf{P}_i\mathbf{Q}^j$ holds after this rescaling, the measurement matrix $\tilde{\mathbf{W}}$ composed of these rescaled homography matrices can be factored as

$$\begin{aligned}\tilde{\mathbf{W}} &= \begin{bmatrix} \tilde{\mathbf{H}}_1^1 & \cdots & \tilde{\mathbf{H}}_1^J \\ \vdots & & \vdots \\ \tilde{\mathbf{H}}_I^1 & \cdots & \tilde{\mathbf{H}}_I^J \end{bmatrix} \\ &\simeq \begin{bmatrix} \lambda_1^1\mathbf{P}_1 \\ \vdots \\ \lambda_I^1\mathbf{P}_I \end{bmatrix} \begin{bmatrix} \lambda_1^1\mathbf{Q}^1 & \cdots & \lambda_1^J\mathbf{Q}^J \end{bmatrix}. \end{aligned} \tag{8}$$

Note In reality the relative homography matrices $\tilde{\mathbf{G}}_i^j$ computed from the captured data are contaminated by noise. Thus computing the scaling factor μ_i^j as a double eigenvalue of $\tilde{\mathbf{G}}_i^j$ is not a good strategy because μ_i^j does not strictly become a double root. Appendix B shows that μ_i^j can be sub-optimally calculated in a linear fashion even in the presence of noise.

3.2 Imposing metric constraints

Suppose that the measurement matrix $\tilde{\mathbf{W}}$ composed of the rescaled plane-image homography matrices is factored as

$$\tilde{\mathbf{W}} = \begin{bmatrix} \mathbf{P}'_1 \\ \vdots \\ \mathbf{P}'_I \end{bmatrix} \begin{bmatrix} \mathbf{Q}^{1\prime} & \cdots & \mathbf{Q}^{J\prime} \end{bmatrix} \tag{9}$$

by using SVD[11]. Comparing eqn. (9) with (8), it can be seen that factored camera and plane matrices $\{\mathbf{P}'_i, \mathbf{Q}^{j\prime}\}$ are related with the "true" ones $\{\mathbf{P}_i, \mathbf{Q}^j\}$ by a set of projective transformations

$$\mathbf{P}'_i \simeq \mathbf{P}_i\mathbf{T}, \quad \mathbf{Q}^{j\prime} \simeq \mathbf{T}^{-1}\mathbf{Q}^j \tag{10}$$

where $\mathbf{T}$ is an unknown 4×4 non-singular matrix. To recover the camera matrices in an Euclidean world, we have to determine $\mathbf{T}$.

Without loss of generality, we can fix both projective and Euclidean coordinate frames to the first camera, which means $\mathbf{P}'_1 = [\mathbf{I} \;\; \mathbf{0}]$ and $\mathbf{P}_1 = [\mathbf{K}_1 \;\; \mathbf{0}]$. Then the transformation matrix $\mathbf{T}$ can be restricted to the following form

$$\mathbf{T} = \begin{bmatrix} \mathbf{K}_1^{-1} & \mathbf{0} \\ \mathbf{h}^\top & h \end{bmatrix} \tag{11}$$

where $[\mathbf{h}^\top \; h]$ composed of a 3-vector $\mathbf{h}$ and a scalar h represents the plane at infinity.

Introducing an unknown scale factor β^j, we rewrite the second equation of (10) as

$$\mathbf{T}\mathbf{Q}^{j\prime} = \beta^j\mathbf{Q}^j. \tag{12}$$

From equations (3) and (12), we have

$$\mathbf{p}^{j\prime\top}\boldsymbol{\omega}_1\mathbf{p}^{j\prime} = \mathbf{q}^{j\prime\top}\boldsymbol{\omega}_1\mathbf{q}^{j\prime} = (\beta^j)^2, \quad \mathbf{p}^{j\prime\top}\boldsymbol{\omega}_1\mathbf{q}^{j\prime} = 0 \tag{13}$$

where $[\mathbf{p}^{j\prime} \; \mathbf{q}^{j\prime}]$ is a 3×2 upper-left submatrix of $\mathbf{Q}^{j\prime}$ and $\boldsymbol{\omega}_1 = \mathbf{K}_1^{-\top}\mathbf{K}_1^{-1}$ is the image of the absolute conic observed by the first camera. We can determine $\boldsymbol{\omega}_1$ in an exactly same way as [10, 16]; ignoring β^j, we have two homogeneous constraints on $\boldsymbol{\omega}_1$ for each model plane π^j. Since $\boldsymbol{\omega}_1$ is symmetric and defined up to a scale, it has five degrees-of-freedom. Therefore, given three or more model planes, we have an over-determined linear system on the entries of $\boldsymbol{\omega}_1$ and can solve it in a least-squares sense. Then $\mathbf{K}_1$ is obtained by Cholesky factorization of $\boldsymbol{\omega}_1$ and normalizing its (3,3)-element to unity.

Once ω_1 is obtained, each β^j can be determined as

$$\beta^j = \sqrt{(\mathbf{p}^{j\prime\top}\boldsymbol{\omega}_1\mathbf{p}^{j\prime} + \mathbf{q}^{j\prime\top}\boldsymbol{\omega}_1\mathbf{q}^{j\prime})/2}. \tag{14}$$

Finally, we determine the plane at infinity $[\mathbf{h}^\top \; h]$ by using the fourth row of eqn. (12):

$$\begin{bmatrix} \mathbf{h}^\top & h \end{bmatrix}\mathbf{Q}^{j\prime} = \begin{bmatrix} 0 & 0 & \beta^j \end{bmatrix} \tag{15}$$

Stacking eqn. (15) for all the model planes π^j, we obtain $[\mathbf{h}^\top \; h]$ as a least-squares solution for this over-determined linear system, and this completes the computation of $\mathbf{T}$. All the camera matrices $\mathbf{P}_i$ $(i = 1, \ldots, I)$ are recovered by $\mathbf{P}_i \simeq \mathbf{T}^{-1}\mathbf{P}'_i$.

3.3 Nonlinear refinement

Up to now, we have solved for camera matrices $\mathbf{P}_i$ through a series of linear least-squares. However, the solution is not optimal and might be biased because the minimization criteria are based on algebraic errors. As recommended in [8, 16], the linear solution should be refined by a nonlinear minimization process based on statistically meaningful criteria for the best results.

Suppose that there are N^j reference points $\mathbf{x}_n^j = (x_n^j, y_n^j)$ $(n = 1, \ldots, N^j)$ on the model plane π^j. Let $\mathbf{u}_{in}^j = (u_{in}^j, v_{in}^j)$ be the projection of $\mathbf{x}_n^j$ on the i-th camera's image plane $\mathcal{I}_i$. Under the assumption that $\mathbf{u}_{in}^j$ is contaminated by independent and uniformly distributed Gaussian noise, the following criterion gives the maximum likelihood estimation (MLE) of the intrinsic and extrinsic camera

IEEE COMPUTER SOCIETY

parameters:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{n=1}^{N^j}\left\|\mathbf{u}_{in}^{j}-\hat{\mathbf{u}}(\mathbf{K}_i,\mathbf{R}_i,\mathbf{t}_i,\mathbf{p}^j,\mathbf{q}^j,\mathbf{d}^j;\mathbf{x}_n^j)\right\|^2 \rightarrow \min \quad (16)$$

where $\hat{\mathbf{u}}(\mathbf{K}_i,\mathbf{R}_i,\mathbf{t}_i,\mathbf{p}^j,\mathbf{q}^j,\mathbf{d}^j;\mathbf{x}_n^j)$ is the projection of $\mathbf{x}_n^j$ onto $\mathcal{I}_i$ computed according to eqn. (5). The minimization is performed using the Levenberg-Marquardt algorithm[6] initialized with the linear solution obtained so far.

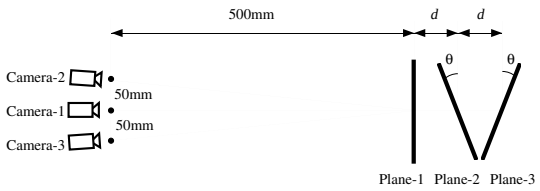


Figure 4. Spatial configuration of the simulation experiments

4 Experiments

The performance of our proposed algorithm has been tested through both simulation and experiments with real data.

4.1 Simulation results

Simulation experiments have been performed according to the spatial configuration displayed in Fig. 4. Three cameras are placed along a straight line at 50 mm intervals and oriented toward the fixation point at a 500 mm distance. The intrinsic parameters of the cameras are set to common values: $k = 900$, $u_0 = v_0 = 255$, $a = 1.3888$ and $s = 0.001212$. Each model plane is painted with $10 \times 14 = 140$ reference points at 18 mm intervals. The distance d and the orientation difference θ between the model planes are varied in the simulation. Independent Gaussian noise with 0 mean is added to the captured image points while varying its standard deviation at various levels. For each noise level, 100 trials are made and average results are displayed. Our proposed algorithm has been run in two modes; linear estimation part only (*L-proposed*) or with nonlinear refinement (*NL-proposed*). Sturm-Maybank-Zhang style calibration algorithm for a single camera[10, 16] with nonlinear refinement, denoted by *SMZ*, has also been implemented for comparison.

Figure 5 shows the errors in the estimated intrinsic parameters of the first camera under $d = 50$ mm and $\theta = 15^\circ$. It can be seen that not only *NL-proposed* but also *L-proposed* yielded better results than *SMZ* for all parameters. This is because our method was constrained by displacements between the cameras and therefore suffered less from over-fitting to the noise. Moreover, *L-proposed* performed almost as well as *NL-proposed*. Since there are less data in horizontal direction than vertical direction, the error in u_0 is larger than that in v_0.

Figure 6 shows the estimated position and orientation errors of the second and the third cameras relative to the first for both our proposed and SMZ algorithms. In the latter, calibration algorithm was run independently for each camera, and then rigid displacements between the cameras were computed by Sturm's plane-based pose estimation method[8] while fixing the intrinsic parameters to the values obtained in the calibration stage. The orientation errors are defined as the rotation angle of the error matrix $\Delta\mathbf{R}_i = \mathbf{R}_i\hat{\mathbf{R}}_i^\top$ where $\mathbf{R}_i$ and $\hat{\mathbf{R}}_i$ are the estimated and true rotation matrices of the i-th camera, respectively. The accuracy of *L-proposed* was very close to that of *NL-proposed* and much better than *SMZ*.

The final estimations by *SMZ* would be identical with *NL-proposed* if all the parameters were refined through the final nonlinear optimization step as recommended in [8]. We also tried this and encountered few convergence problems. However, we have also found that *SMZ* tended to fail more frequently than *NL-proposed* under near-degenerate configurations, i.e. with small values of θ (typically $\approx 5^\circ$), because the positive definiteness of the image of the absolute conic $\boldsymbol{\omega}$ was violated more easily in the presence of noise, which made Cholesky factorization of $\boldsymbol{\omega}$ impossible.

The algorithm *NL-proposed* was tested while changing the position and orientation of the model planes. Figure 7 shows the errors in focal length of the first camera while the distance parameter d was varied from 0 mm to 100 mm. It can be seen that the error is almost irrelevant to the value of d except in the vicinity of $d = 0$; we have found that the rescaling of homography matrices (see eqn. (7)) did not make the measurement matrix close enough to rank 4 even for relatively small noise level if $d \approx 0$, which produced a convergence problem. This is not due to degeneracy of the configuration because such sensitivity did not appear for zero or very small noise. Investigating the cause is still remained open.

The errors in focal length of the first camera under varying angle parameter θ are displayed in Fig.8. The improvement of accuracy can be achieved with large θ values, i.e. far from degenerate configuration.

4.2 Results with real data

Our proposed algorithm has been tested on real data as well. Three CCD cameras with lenses of 8 mm focal length

IEEE COMPUTER SOCIETY

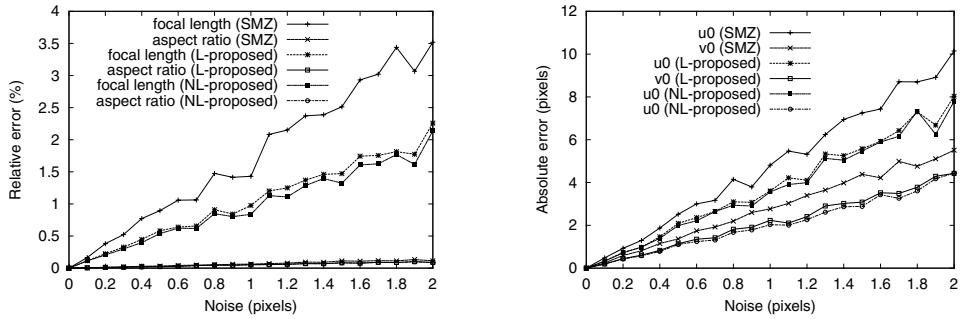


Figure 5. Errors in focal length, aspect ratio (left) and principal point (right) vs. the noise level ($d = 50$ mm, $\theta = 15^\circ$)

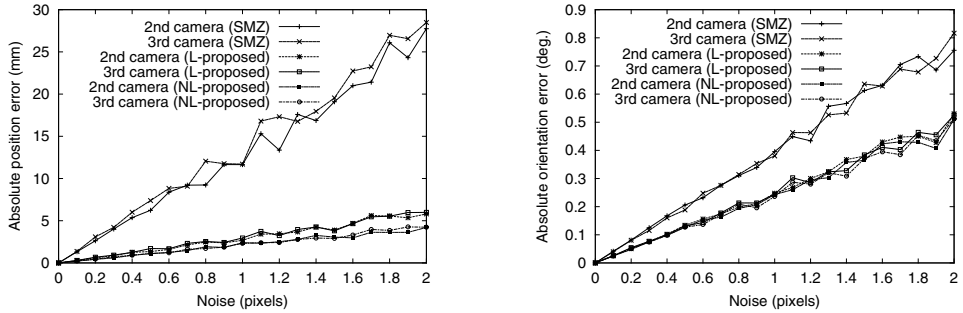


Figure 6. Errors in position (left) and orientation (right) vs. the noise level ($d = 50$ mm, $\theta = 15^\circ$)

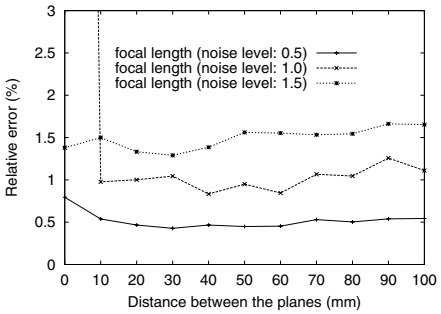


Figure 7. Errors in focal length vs. the distance between the planes ($\theta = 15^\circ$)

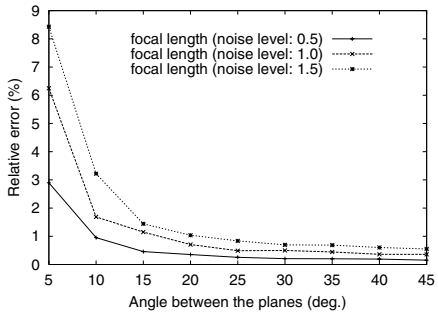


Figure 8. Errors in focal length vs. the angle between the planes ($d = 50$ mm)

were calibrated. A model plane with $13 \times 9 = 117$ lattice points was placed at six locations in different orientations.

Figure 9 shows the singular values of the measurement matrix $\tilde{\mathbf{W}}$ before and after rescaling (see eqn. (7)). It can be seen that the first four singular values are well separated from the rest after rescaling.

All of the estimated camera parameters are displayed in Table 1. The position and orientation of the i-th camera are represented by two 3-vectors $\mathbf{t}_i$ and $\boldsymbol{\theta}_i$ respectively, where $\boldsymbol{\theta}_i$ is of a magnitude equal to the rotation angle and parallel to the rotation axis. For each camera, the first column shows the initial guesses for the nonlinear refinement process and

IEEE COMPUTER SOCIETY

Table 1. Estimated camera parameters using 6 planes

parameters	first camera		second camera		third camera	
	initial	final	initial	final	initial	final
k	1092.42	1085.84±0.98	1099.75	1093.27±0.94	1114.19	1110.85±0.95
(u_0, v_0)	(321.06	(318.18±0.85	(352.36	(354.02±0.87	(376.67	(373.85±0.82
	218.16)	217.03±0.88)	236.03)	235.49±0.87)	227.68)	226.00±0.89)
a	1.00962	1.00829±0.00334	1.00803	1.00809±0.00321	1.00907	1.00600±0.00337
s	-0.00192	-0.00122±0.000592	0.003798	-0.001075±0.000582	0.001506	-0.001251±0.000603
$\mathbf{t}$ (mm)	(0	(0±0	(154.12	(154.30±0.32	(72.36	(72.42±0.20
	0	0±0	0.89	0.93±0.18	-74.33	-74.11±0.20
	0)	0±0)	22.66)	22.37±1.30)	19.88)	19.62±1.14)
$\boldsymbol{\theta}$ (deg.)	(0	(0±0	(-0.27	(-0.31±0.06	(-7.56	(-7.56±0.06
	0	0±0	-12.55	-12.40±0.06	-3.73	-3.75±0.06
	0)	0±0)	0.55)	0.57±0.01)	-1.18)	-1.12±0.01)

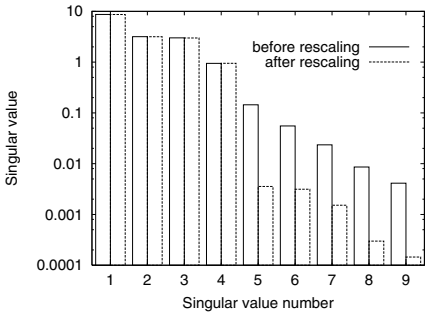
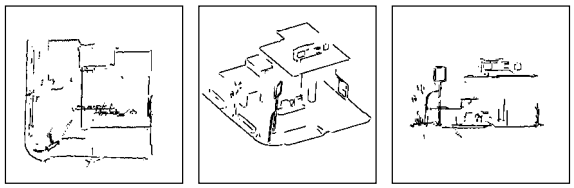


Figure 9. Singular values of measurement matrix before/after rescaling

(a) Input image triplet



(b) Reconstructed 3D structure: top view, upper-side view and front view

Figure 10. 3D reconstruction from three images captured by calibrated cameras

the second column shows the final optimal results with standard deviations. It can be seen that linearly computed estimations were very close to the optimal values. The refinement process was converged within three iterations.

We carried out 3D reconstruction from an image triplet captured by the calibrated cameras. A segment-based trinocular stereo vision algorithm was applied to the input images shown in fig. 10(a). Feature matching along corresponding epipolar lines and final triangulation in the 3D Euclidean space were performed using the estimated camera parameters. Several views of the reconstructed scene structure are displayed in fig. 10(b).

5 Conclusion

A new calibration algorithm for multi-camera systems using a model plane with known reference points has been described. Our proposed technique integrates the Sturm-Maybank-Zhang style calibration method with Sturm's plane-based pose estimation algorithm so that the rigid displacements between the cameras are consistently recovered as well as their intrinsic parameters. The algorithm inherits handiness and flexibility from the original method; the user can perform calibration only by moving the model plane freely and capturing its images at several locations. Simulation experiments showed that our proposed algorithm yielded better results than the preceding algorithm for both intrinsic and extrinsic parameters. Moreover, the linear solutions by the new algorithm were very close to the final nonlinear optimal solutions. Experiments with real data also have shown the validity of the proposed algorithm in practice.

A Relative Homography Matrix[15]

We rewrite $\mathbf{H}_i^j$ in (5) as

$$\begin{aligned}\mathbf{H}_i^j &= \mathbf{K}_i\mathbf{R}_i^\top \begin{bmatrix} \mathbf{I} & -\mathbf{t}_i \end{bmatrix} \begin{bmatrix} \mathbf{p}^j & \mathbf{q}^j & \mathbf{d}^j \\ 0 & 0 & 1 \end{bmatrix} \\ &= \mathbf{K}_i\mathbf{R}_i^\top \left(\begin{bmatrix} \mathbf{p}^j & \mathbf{q}^j & \mathbf{d}^j \end{bmatrix} - \mathbf{t}_i\mathbf{k}^\top \right)\end{aligned}$$

where $\mathbf{k}^\top = [0\ 0\ 1]$. Then, noting that $\mathbf{R}_1 = \mathbf{I}$ and $\mathbf{t}_1 = \mathbf{0}$, the inter-image homography between $\mathcal{I}_1$ and $\mathcal{I}_i$ induced by

IEEE COMPUTER SOCIETY

the model plane π^j is represented by

$$\begin{aligned}\mathbf{A}_{i1}^{j} &= \mathbf{H}_i^j(\mathbf{H}_1^j)^{-1} \\ &= \mathbf{K}_i\mathbf{R}_i^\top\left(\begin{bmatrix}\mathbf{p}^j & \mathbf{q}^j & \mathbf{d}^j\end{bmatrix} - \mathbf{t}_i\mathbf{k}^\top\right) \\ &\qquad\qquad \begin{bmatrix}\mathbf{p}^j & \mathbf{q}^j & \mathbf{d}^j\end{bmatrix}^{-1}\mathbf{K}_1^{-1} \\ &= \mathbf{K}_i\mathbf{R}_i^\top\left(\mathbf{I} - \mathbf{t}_i\frac{(\mathbf{p}^j\times\mathbf{q}^j)^\top}{(\mathbf{p}^j\times\mathbf{q}^j)^\top\mathbf{d}^j}\right)\mathbf{K}_1^{-1} \\ &= \mathbf{A}_{i1}^{\infty}\left(\mathbf{I} - \mathbf{e}_i\mathbf{n}^{j\top}\right)\end{aligned}$$

where $\mathbf{A}_{i1}^{\infty} = \mathbf{K}_i\mathbf{R}_i^\top\mathbf{K}_1^{-1}$ denotes a homography between $\mathcal{I}_1$ and $\mathcal{I}_i$ induced by the plane at infinity, $\mathbf{e}_i = \mathbf{K}_1\mathbf{t}_i$ is a projection of the i-th camera center to $\mathcal{I}_1$ (i.e. the epipole), and $\mathbf{n}^{j\top} = (\mathbf{p}^j\times\mathbf{q}^j)^\top\mathbf{K}_1^{-1}/(\mathbf{p}^j\times\mathbf{q}^j)^\top\mathbf{d}^j$ represents π^j. Hence the relative homography matrix $\mathbf{G}_i^j$ becomes

$$\begin{aligned}\mathbf{G}_{i1}^{j1} &= (\mathbf{A}_{i1}^{j})^{-1}\mathbf{A}_{i1}^{1} \\ &= \left(\mathbf{I} - \mathbf{e}_i\mathbf{n}^{j\top}\right)^{-1}\left(\mathbf{I} - \mathbf{e}_i\mathbf{n}^{1\top}\right) \\ &= \left(\mathbf{I} + \mathbf{e}_i\frac{\mathbf{n}^{j\top}}{1-\mathbf{n}^{j\top}\mathbf{e}_i}\right)\left(\mathbf{I} - \mathbf{e}_i\mathbf{n}^{1\top}\right) \\ &= \mathbf{I} + \mathbf{e}_i\frac{\mathbf{n}^{j\top} - \mathbf{n}^{1\top}}{1-\mathbf{n}^{j\top}\mathbf{e}_i}\end{aligned}$$

which has a form $\mathbf{G}_{i1}^{j1} = \mathbf{I} + \mathbf{e}\mathbf{f}^\top$.

B Finding μ s.t. $\mathbf{G} - \mu\mathbf{I}$ becomes rank 1

Let $\mathbf{G}$ be a 3×3 matrix having a form

$$\mathbf{G} = \mu\mathbf{I} + \mathbf{e}\mathbf{f}^\top \tag{17}$$

for some unknown scalar μ and unknown non-zero 3-vectors $\mathbf{e}$ and $\mathbf{f}$. We would like to find the value of μ from $\mathbf{G}$.

Equation (17) means that $\mathbf{G} - \mu\mathbf{I}$ is a rank 1 matrix, which in turn implies

$$\mathbf{a} - \mu\mathbf{i} \simeq \mathbf{b} - \mu\mathbf{j} \simeq \mathbf{c} - \mu\mathbf{k}$$

where $\mathbf{G} = [\mathbf{a}\ \mathbf{b}\ \mathbf{c}]$ and $\mathbf{I} = [\mathbf{i}\ \mathbf{j}\ \mathbf{k}]$. Therefore a cross product of any two vectors above vanishes. Taking a cross product of the first two vectors for instance, we have

$$(\mathbf{a}-\mu\mathbf{i})\times(\mathbf{b}-\mu\mathbf{j}) = \mathbf{a}\times\mathbf{b} + \mu(\mathbf{j}\times\mathbf{a} - \mathbf{i}\times\mathbf{b}) + \mu^2\mathbf{k} = \mathbf{0}. \tag{18}$$

Although this equation gives three constraints on μ, only two of them are independent. So, we pick up the ones that are linear in μ by taking scalar products of (18) with $\mathbf{i}$ and $\mathbf{j}$:

$$\mathbf{i}^\top(\mathbf{a}\times\mathbf{b}) + \mu\mathbf{k}^\top\mathbf{a} = 0, \quad \mathbf{j}^\top(\mathbf{a}\times\mathbf{b}) + \mu\mathbf{k}^\top\mathbf{b} = 0.$$

In a similar manner, four more linear constraints on μ can be derived from two other combinations of $\mathbf{a} - \mu\mathbf{i}$, $\mathbf{b} - \mu\mathbf{j}$ and $\mathbf{c} - \mu\mathbf{k}$. We can determine μ as a common root of these six linear constraints if there is no noise in $\mathbf{G}$, or as a least-squares solution in the presence of noise.

References

[1] O. D. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[3] A. Heyden. Projective Structure and Motion from Image Sequences Using Subspace Methods. In *Proc. 10th Scandinavian Conference on Image Analysis*, 1997.

[4] E. Malis and R. Cipolla. Multi-view constraionts between collineations: application to self-calibration from unknown planar structures. In *Proc. 6th European Conference on Computer Vision*, volume 2, pages 610–624, 2000.

[5] E. Malis and R. Cipolla. Self-calibration of zooming cameras observing an unknown planar structure. In *Proc. 15th International Conference on Pattern Recognition*, volume 1, pages 85–88, 2000.

[6] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1988.

[7] C. Rother, S. Carlsson, and D. Tell. Projective Factorization of Planes and Cameras in Multiple Views. In *Proc. 16th International Conference on Pattern Recognition*, volume 2, pages 737–740, 2002.

[8] P. Sturm. Algorithms for Plane-Based Pose Estimation. In *Proc. Computer Vision and Pattern Recognition*, volume 1, pages 706–711, 2000.

[9] P. Sturm and B. Triggs. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In *Proc. 4th European Conference on Computer Vision*, volume 2, pages 709–720, 1996.

[10] P. F. Sturm and S. J. Maybank. On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications. In *Proc. Computer Vision and Pattern Recognition*, volume 1, pages 432–437, 1999.

[11] C. Tomasi and T. Kanade. Shape and Motion from Image Streams under Orthography: a Factorization Method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[12] B. Triggs. Autocalibration from Planar Scenes. In *Proc. 5th European Conference on Computer Vision*, volume 1, pages 89–105, 1998.

[13] B. Triggs. Plane + Parallax, Tensors and Factorizaion. In *Proc. 6th European Conference on Computer Vision*, volume 1, pages 522–538, 2000.

[14] R. Y. Tsai. 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, Aug 1987.

[15] L. Zelnik-Manor and M. Irani. Multiview Constraints on Homographies. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(2):214–223, Feb 2002.

[16] Z. Zhang. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In *Proc. 7th International Conference on Computer Vision*, volume 1, pages 666–673, 1999.

IEEE COMPUTER SOCIETY