Self-calibration of a Stereo Rig without Stereo Correspondence

F. Dornaika and R. Chung

Department of Mechanical & Automation Engineering
The Chinese University of Hong Kong, Shatin, Hong Kong
{dornaika, rchung}@mae.cuhk.edu.hk

Abstract – *In this paper, we introduce a new self-calibration method for a stereo rig using general rigid motions. Unlike the existing methods, our method does not involve any stereo correspondence. Rather the method requires the computation of camera motion between pairs of monocular images. We show that the computation of the stereo geometry, i.e. the rotation and the translation between the two cameras can be carried out using standard linear algebra tools. We provide a stability study for the method in the presence of noise. Synthetic and real experiments demonstrate the feasibility and robustness of the proposed method. Moreover, the developed method can be used to solve the stereo correspondence problem as well as the 3D shape of the observed scenes using a cooperative stereo-motion framework.*

Keywords *Self-calibration, stereo rig extrinsic parameters, motion estimation, linear algebra*

1 Introduction and approach

In the last decade a number of researchers developed self-calibration methods for vision sensors that require no known reference object. Such methods can be used to determine the intrinsic camera parameters, the stereo geometry as well as the 3D shape of the observed scene (see [3], [4], [8], [9], [12], for a single moving camera and [5], [18], [19], for a moving stereo rig). The usefulness of the self-calibration techniques can be tangible in some cases where the sensor parameters are subject to variations and no known reference objects are available (active vision, space robots). In [18], authors use motion and stereo correspondences across two stereo pairs (one motion of the stereo rig). They propose a method that recovers simultaneously the two internal parameters and the motion of each camera as well as the stereo geometry. In [5], authors use stereo correspondences across a sequence of stereo pairs. Using different projective reconstructions that are associated with each stereo pair, they propose an algorithm for the recovery of the internal parameters and the 3D Euclidean shape.

All previous self-calibration techniques for stereo entails the solving of the correspondence problem between left and right images which has been proved to be a difficult problem. Several factors make the stereo correspondence problem difficult: occlusions, large disparities, photometric and figural distortions. One can notice that in many cases it is much easier to find motion correspondences than solving for stereo correspondences.

In this paper we attempt to solve the following problem. Given a sequence of stereo pairs of unknown and arbitrary scenes (see Figure 1), we like to recover the stereo geometry without solving the stereo correspondence problem. In the sequel, we restrict our work to the estimation of the stereo geometry (the rotation and the translation between the two cameras) and suppose that the camera intrinsic parameters are known. This assumption is not unrealistic since the intrinsic parameters are usually known weakly or partially. On the other hand, there is no information about the geometry of the stereo rig.

Unlike the existing methods for self-calibration which consider stereo correspondences as given, our approach uses motion correspondences to recover the stereo geometry. The monocular correspondences between pairs of images are easier to obtain, as an arbitrary number of intermediate frames can be available. Once the stereo geometry is recovered, the stereo rig fundamental matrix will be known. Then, the stereo correspondences can be easily recovered. One way to do so is to use geometrical constraints derived from the motion correspondences and the computed fundamental matrix (cooperative stereo-motion) [1]. The 3D shape of the observed scenes will be straightforward. Thus, our proposed method is very useful in recovering the epipolar geometry of stereo rigs whose cameras have a large orientation w.r.t. to each other (in this case stereo correspondence will be difficult).

The proposed method is a two-step approach. In the first step, 3D motions of each camera are estimated using feature correspondences in two monocular images. In the second step, using the motion of the left and right cameras, the stereo geometry of the stereo rig is recovered by minimizing a cost function derived from the invariance of the stereo geometry. Since the proposed method does not require the same 3D features be visible in all monocular images, the stereo rig can perform any deliberate motion in its environment. This will improve the accuracy of the estimated parameters.

Central to our approach is the recovery of camera motion from monocular correspondences. Knowing the intrinsic parameters of the camera, the recovery of the motion parameters, i.e. the rotation and the translation (up to a scale factor) is a classical problem [13], [15], [16]. Closed-form and non-linear solutions can be used.

The organization of this paper is as follows. Section 2 presents in details the recovery of the stereo geometry, i.e. the rotation and the direction of the translation. Section 3 provides a stability study of the method in the presence of noise. Section 4 describes experiments with real images. Finally, section 5 provides a short discussion.

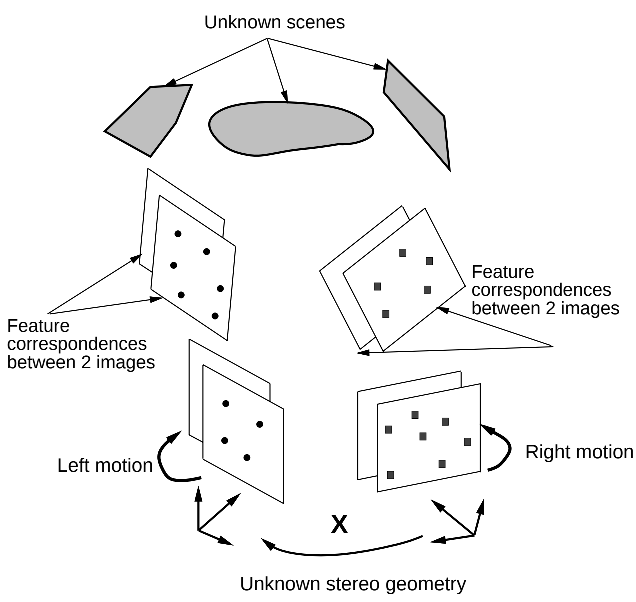


Figure 1: Self-calibration of a stereo rig using only monocular correspondences. It should be noted that feature correspondences between pairs of images can be different, that is, features need not be seen in all monocular images.

2 Stereo geometry estimation

In this section, we show that the geometry of the stereo rig can be recovered from camera motions estimated independently. The basic idea relies on solving a homogeneous system that is often used in hand-eye calibration problem [14], [11], [6]. However, the unknowns of our obtained system as well as the solution are different from those of the hand-eye calibration problem.

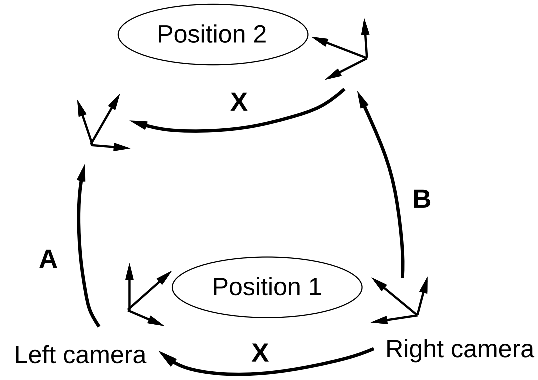


Figure 2: The stereo geometry $\mathbf{X}$ is related to the cameras motions by the homogeneous relationship $\mathbf{AX} = \mathbf{XB}$.

We consider one motion of the stereo rig or equivalently two positions. These positions are denoted by position 1 and position 2, respectively (see Figure 2). Let $\mathbf{A}$ be the transformation between the two positions of the left camera and $\mathbf{B}$ be the corresponding transformation of the right camera. Let $\mathbf{X}$ be the stereo geometry of the stereo rig that we are attempting to estimate. $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{X}$ are 4×4 matrices of the form:

$$\mathbf{A} = \begin{pmatrix} \mathbf{R}_A & \mathbf{t}_A \\ \mathbf{0}^T & 1 \end{pmatrix}$$

where $\mathbf{R}_A$ is a rotation matrix and $\mathbf{t}_A$ a translation vector.

Since the two cameras are rigidly linked, we obtain the following relationship (see Figure 2):

$$\mathbf{AX} = \mathbf{XB} \quad (1)$$

This equation can be decomposed into two equations: a matrix equation and a vector equation:

$$\mathbf{R}_A \mathbf{R}_X = \mathbf{R}_X \mathbf{R}_B \quad (2)$$

and:

$$(\mathbf{R}_A - \mathbf{I})\, \mathbf{t}_X = \mathbf{R}_X\, \mathbf{t}_B - \mathbf{t}_A$$

with $\mathbf{I}$ being the 3×3 identity matrix.

$\mathbf{A}$ and $\mathbf{B}$ are estimated from monocular correspondences. Therefore, $\mathbf{R}_A$ and $\mathbf{R}_B$ are known, and $\mathbf{t}_A$ and $\mathbf{t}_B$ are known up to a scale factor (each one has its own scale).

To solve Eq. (2) one may take advantage of the particular algebraic and geometric properties of rotation matrices. Indeed, this equation can be written as:

$$\mathbf{R}_A = \mathbf{R}_X \mathbf{R}_B \mathbf{R}_X^T$$

which is a similarity transformation, since $\mathbf{R}_X$ is an orthogonal matrix. Hence, matrices $\mathbf{R}_A$ and $\mathbf{R}_B$ have the same eigenvalues. A well-known property of a rotation matrix is that it has one of its eigenvalues equal to 1. Let $\mathbf{n}_B$ be the eigenvector of $\mathbf{R}_B$ associated with this eigenvalue. By postmultiplying Eq. (2) with $\mathbf{n}_B$, we obtain (we have $\mathbf{n}_B = \mathbf{R}_B\, \mathbf{n}_B$):

$$\begin{aligned} \mathbf{R}_A \mathbf{R}_X\, \mathbf{n}_B &= \mathbf{R}_X \mathbf{R}_B\, \mathbf{n}_B \\ &= \mathbf{R}_X\, \mathbf{n}_B \end{aligned}$$

and we conclude that $\mathbf{R}_X\, \mathbf{n}_B$ is nothing but the eigenvector of the matrix $\mathbf{R}_A$ associated with the eigenvalue 1:

$$\mathbf{n}_A = \mathbf{R}_X\, \mathbf{n}_B$$

One can notice that $\mathbf{n}_A$ and $\mathbf{n}_B$ are the rotation axes associated with the left and right rotations $\mathbf{R}_A$ and $\mathbf{R}_B$, respectively. Their derivation from $\mathbf{R}_A$ and $\mathbf{R}_B$ is straightforward .

To conclude the stereo geometry parameters $(\mathbf{R}_X, \mathbf{t}_X)$ satisfy the following constraints:

$$\begin{aligned} \mathbf{n}_A &= \mathbf{R}_X\, \mathbf{n}_B & (3) \\ (\mathbf{R}_A - \mathbf{I})\, \mathbf{t}_X &= \mathbf{R}_X\, \mathbf{t}_B - \mathbf{t}_A & (4) \end{aligned}$$

2.1 The rotation

Equation (3) is associated with one motion of the stereo rig. Thus, to estimate $\mathbf{R}_X$ at least two independent motions are necessary (two different directions $\mathbf{n}_A$). In the general case of n motions, one may estimate the rotation matrix $\mathbf{R}_X$ by minimizing this positive error function:

$$f(\mathbf{R}_X) = \sum_{i=1}^{n} \|\mathbf{n}_{Ai} - \mathbf{R}_X\, \mathbf{n}_{Bi}\|^2 \qquad (5)$$

Therefore, finding $\mathbf{R}_X$ is reduced to the problem of optimally estimating rotation from 3D to 3D vector correspondences (see [7]). To minimize f given by Eq. (5), we represent the rotation $\mathbf{R}_X$ by a unit quaternion. The solution for this unit quaternion can be found using the closed-form solution proposed by Faugeras and Hebert [2].

2.2 The direction of the translation

In this section, we will show how we can estimate the direction of the translation, i.e. $\mathbf{t}_X$ up to a scale factor. We stress the fact that the norm of the translation can not be recovered since there is no reference for 3D lengths.

Once the rotation of the stereo rig $\mathbf{R}_X$ has been determined, the direction of the translation can be determined using a set of Eq. (4). Thus, for n motions of the stereo rig we obtain the following set of vector equations:

$$\left\{ \begin{array}{c} \vdots \\ (\mathbf{R}_{Ai} - \mathbf{I})\, \mathbf{t}_X = \mathbf{R}_X\, \mathbf{t}_{Bi} - \mathbf{t}_{Ai} \\ \vdots \end{array} \right. \qquad (6)$$

Solving for $\mathbf{t}_X$ using the above system (6) is not as easy as in the off-line calibration case (see the hand-eye calibration in [14], [11]). In this case (off-line calibration), both $\mathbf{t}_{Ai}$ and $\mathbf{t}_{Bi}$ are known, therefore the translation $\mathbf{t}_X$ is obtained by solving a simple linear system. In our case, we only know the direction of $\mathbf{t}_{Ai}$ and $\mathbf{t}_{Bi}$; their norm can not be recovered since we have no knowledge about the observed scene.

In the remainder of this section, we present two linear methods as well as a non-linear one to estimate the direction of the translation $\mathbf{t}_X$.

2.2.1 First method

This method takes advantage of a well-known relationship between the left motion $\mathbf{A}_i$ and the right motion $\mathbf{B}_i$. This is given by:

$$\mathbf{n}_{Ai} \cdot \mathbf{t}_{Ai} = \mathbf{n}_{Bi} \cdot \mathbf{t}_{Bi} \qquad (7)$$

□ *Proof:* It is known that the vector $(\mathbf{R}_{Ai} - \mathbf{I})\, \mathbf{t}_X$ is perpendicular to the rotation axis of $\mathbf{R}_{Ai}$, i.e. $\mathbf{n}_{Ai}$. Therefore, we have:

$$\mathbf{n}_{Ai} \cdot \{(\mathbf{R}_{Ai} - \mathbf{I})\, \mathbf{t}_X\} = 0$$

Using Eq.(6) and the relations: (i) $\mathbf{n}_{Ai} = \mathbf{R}_X \mathbf{n}_{Bi}$ and (ii) $(\mathbf{R}_X \mathbf{n}_{Bi}) \cdot (\mathbf{R}_X\, \mathbf{t}_{Bi}) = \mathbf{n}_{Bi} \cdot \mathbf{t}_{Bi}$, one can successively write:

$$\begin{aligned} \mathbf{n}_{Ai} \cdot \{(\mathbf{R}_{Ai} - \mathbf{I})\, \mathbf{t}_X\} &= 0 \\ \mathbf{n}_{Ai} \cdot \{\mathbf{R}_X\, \mathbf{t}_{Bi} - \mathbf{t}_{Ai}\} &= 0 \\ (\mathbf{R}_X \mathbf{n}_{Bi}) \cdot (\mathbf{R}_X\, \mathbf{t}_{Bi}) - \mathbf{n}_{Ai} \cdot \mathbf{t}_{Ai} &= 0 \\ \mathbf{n}_{Bi} \cdot \mathbf{t}_{Bi} - \mathbf{n}_{Ai} \cdot \mathbf{t}_{Ai} &= 0 \; \square \end{aligned}$$

Let $\mathbf{u}_{Ai}$ and $\mathbf{u}_{Bi}$ be the unit vector of the translation vectors $\mathbf{t}_{Ai}$ and $\mathbf{t}_{Bi}$, respectively. Let λ_{Ai} and λ_{Bi} be the norms of these two vectors such that $\mathbf{t}_{Ai} = \lambda_{Ai}\,\mathbf{u}_{Ai}$ and $\mathbf{t}_{Bi} = \lambda_{Bi}\,\mathbf{u}_{Bi}$ (both λ_{Ai} and λ_{Bi} are unknown).

Therefore, (6) can be written as:

$$\left\{ \begin{array}{c} \vdots \\ (\mathbf{R}_{Ai} - \mathbf{I})\,\mathbf{t}_X = \lambda_{Bi}\,\mathbf{R}_X\,\mathbf{u}_{Bi} - \lambda_{Ai}\,\mathbf{u}_{Ai} \\ \vdots \end{array} \right. \quad (8)$$

where the unknowns are the components of $\mathbf{t}_X$ as well as the scalars λ_{Ai} and $\lambda_{Bi}, i = 1 \cdots n$.

Eq. (7) allows us to determine the ratio $\lambda_{Ai}/\lambda_{Bi}$, noted as k_i, $(\mathbf{n}_{Ai} \cdot \mathbf{t}_{Ai} = 0)$:

$$k_i = \lambda_{Ai}/\lambda_{Bi} = (\mathbf{n}_{Bi} \cdot \mathbf{u}_{Bi})/(\mathbf{n}_{Ai} \cdot \mathbf{u}_{Ai})$$

By substituting $\lambda_{Ai} = k_i\,\lambda_{Bi}$ in (8) and considering that $\mathbf{t}_X = \lambda_X\,\mathbf{u}_X$ ($\mathbf{u}_X$ is the unit vector associated with $\mathbf{t}_X$), (8) becomes:

$$\left\{ \begin{array}{c} \vdots \\ (\mathbf{R}_{Ai} - \mathbf{I})\,\mathbf{u}_X = (\lambda_{Bi}/\lambda_X)\,(\mathbf{R}_X\,\mathbf{u}_{Bi} - k_i\,\mathbf{u}_{Ai}) \\ \vdots \end{array} \right. \quad (9)$$

Each motion i provides 3 linear equations in $\mathbf{u}_X = (u_1, u_2, u_3)^T$ and λ_{Bi}/λ_X. Within these 3 equations, one can eliminate the scale factor (λ_{Bi}/λ_X) by dividing the first equation and the second one by the third one. Thus, each motion provides two homogeneous and linear equations in $\mathbf{u}_X$. For example, if we denote the vector $\mathbf{R}_X\,\mathbf{u}_{Bi} - k_i\,\mathbf{u}_{Ai}$ by $\mathbf{z}_i = (z_{1i}, z_{2i}, z_{3i})^T$ then these linear constraints will be given by (the subscript of the rotation $\mathbf{R}_{Ai}$ has been omitted):

$$\begin{array}{rcll} z_{3i}\,\{(r_{11} - 1)\,u_1 + r_{12}\,u_2 + r_{13}\,u_3\} & - & & \\ z_{1i}\,\{r_{31}\,u_1 + r_{32}\,u_2 + (r_{33} - 1)\,u_3\} & = & 0 & (10) \\ z_{3i}\,\{r_{21}\,u_1 + (r_{22} - 1)\,u_2 + r_{23}\,u_3\} & - & & \\ z_{2i}\,\{r_{31}\,u_1 + r_{32}\,u_2 + (r_{33} - 1)\,u_3\} & = & 0 & (11) \end{array}$$

with r_{ij} being the elements of the rotation matrix $\mathbf{R}_{Ai}$.

An algebraic and geometrical analysis of Eq. (9) shows that the resulting Eqs. (10) and (11) are equivalent (the rank of the matrix $(\mathbf{R}_{Ai} - \mathbf{I})$ is 2). Therefore, at least two independent displacements of the stereo rig are necessary to uniquely determine the unit vector $\mathbf{u}_X$. The resulting system of equations can be written in matrix form as:

$$\underbrace{\mathbf{G}}_{2n \times 3}\,\mathbf{u}_X = \mathbf{0}$$

For example the last row of $\mathbf{G}$ is given by the following 3-vector (see Eq. (11)): $(z_{3n}\,r_{21} - z_{2n}\,r_{31}, z_{3n}\,(r_{22} - 1) - z_{2n}\,r_{32}, z_{3n}\,r_{23} - z_{2n}\,(r_{33} - 1))^T$.

Finally, the solution for $\mathbf{u}_X$, in the sense of least squares, will be the unit eigenvector of $\mathbf{G}^T\mathbf{G}$ associated with its smallest eigenvalue.

2.2.2 Second method

This method consists of determining all the unknowns up to an overall scale factor by solving the homogeneous linear system (8). Therefore, if we represent the unknowns by a $(3+2n)$-vector $\mathbf{u}' = [\mathbf{t}_X^T, \cdots, \lambda_{Ai}, \lambda_{Bi}, \cdots]^T$ thus, the constraints provided by the system (8) can be written in the following matrix form:

$$\underbrace{\mathbf{G}'}_{3n \times (3+2n)}\quad \mathbf{u}' = \mathbf{0}$$

where $\mathbf{G}'$ is a $3n \times (3 + 2n)$ matrix.

Therefore, the solution for $\mathbf{u}'$, in the sense of least squares, will be the unit eigenvector of $\mathbf{G}'^T\mathbf{G}'$ associated with its smallest eigenvalue. Finally, the first three components of $\mathbf{u}'$ will give the direction of the translation vector, i.e. $\mathbf{t}_X$ up to a scale factor. The disadvantage of this method is that two additional unknowns will be obtained per supplementary displacement.

2.2.3 Non-linear optimization

Another alternative which is derived from the first method consists of solving the non-linear form of Eqs. (10) and (11). Thus, for each motion of the stereo rig we obtain the following two constraints:

$$\frac{(r_{11} - 1)\,u_1 + r_{12}\,u_2 + r_{13}\,u_3}{r_{31}\,u_1 + r_{32}\,u_2 + (r_{33} - 1)\,u_3} = \frac{z_{1i}}{z_{3i}}$$

$$\frac{(r_{21} - 1)\,u_1 + r_{22}\,u_2 + r_{23}\,u_3}{r_{31}\,u_1 + r_{32}\,u_2 + (r_{33} - 1)\,u_3} = \frac{z_{2i}}{z_{3i}}$$

Therefore, the direction of the translation $\mathbf{u}_X = (u_1, u_2, u_3)^T$ is obtained by minimizing a positive error function which has the following form:

$$\min_{\mathbf{u}_X}\; g(\mathbf{u}_X) = \sum_{i=1}^{n} (\frac{\hat{z}_{1i}}{\hat{z}_{3i}} - \frac{z_{1i}}{z_{3i}})^2 + (\frac{\hat{z}_{2i}}{\hat{z}_{3i}} - \frac{z_{2i}}{z_{3i}})^2 \quad (12)$$

where vectors $\hat{\mathbf{z}}_i$ and $\mathbf{z}_i$ represent the left side and the right side of the vector equation (9). Although the above minimization is non-linear, the optimal solution associated with the global minimum is easy to find since we only have 2 independent unknowns, i.e. the two spherical coordinates of $\mathbf{u}_X$. We have used the Levenberg-Marquardt technique in order to carry out the minimization [10].

This method is less sensitive to noise than the linear ones. The reason is the well-known fact that minimizing an error function $\sum(\frac{\alpha_i}{\beta_i} - \frac{\alpha'_i}{\beta'_i})^2$ is quite different from minimizing $\sum(\alpha_i\beta'_i - \alpha'_i\beta_i)^2$ because the latter is weighted by the variable quantity $\beta_i\beta'_i$. This will be confirmed in the next section.

3 Stability study

In this section we study the behavior of the proposed method as a function of motion noise. Since image data are noisy, uncertainties always affect the computation of camera motions regardless of the method being used. Since the stereo geometry is derived from camera motions, it follows that this computation has some uncertainties associated with it. To perform the stability study of the proposed self-calibration method, we developed the following framework:

1. Nominal value for the stereo geometry ($\mathbf{X}$) is provided. This nominal value can correspond to the geometry of a real stereo rig.

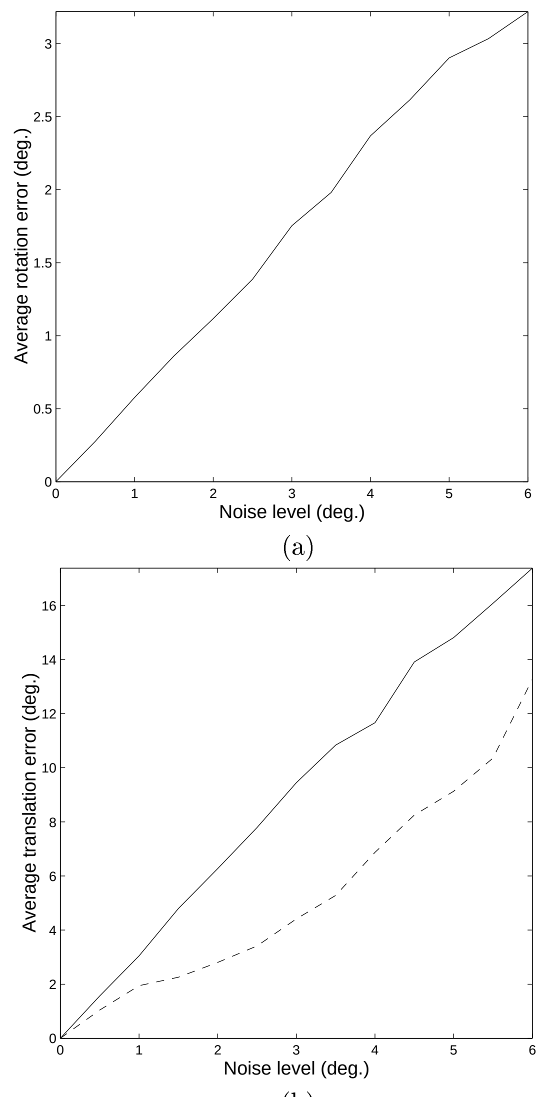
2. Also provided are n motions for the left camera from which the induced right camera motions can be computed with $\mathbf{B}_i = \mathbf{X}^{-1} \mathbf{A}_i \mathbf{X}$

3. Uniform noise is added to both the left and right camera motions, and the stereo geometry is estimated in the presence of this noise using the developed methods.

4. We study the variations of the estimated stereo geometry as a function of the noise.

Since both rotations and the translation directions may be represented as unit vectors, adding noise to motions consists of changing the direction of all the associated unit vectors. Therefore, perturbing camera motions is equivalent to adding noise to the spherical coordinates of all unit vectors. Simulated noise is obtained using a random number generator with a uniform distribution in the interval $[-\Theta_{max}/2, +\Theta_{max}/2]$. Thus, the level of noise is represented as the value of Θ_{max}. For each noise level we make run the self-calibration algorithm for a large number of trials. In the following we consider 3 motions of the stereo rig ($n = 3$).

The calibration errors are: rotation error and translation error. The rotation error is defined as the rotation angle of the 3D rotation required to align the two 3D rotations: the theoretical one (nominal) and the computed one. The translation error is defined as the angle between the two translation directions: the theoretical one and the computed one. Figure 3.a shows the average rotation error as a function of the uniform noise level over 500 trials. The noise level, i.e. Θ_{max} varies from 0 to 6 degrees. Figure 3.b shows the average translation error as a function of the noise level. The solid line corresponds to the linear method (first one) while the dashed one corresponds to the non-linear optimization. For example, when the level of noise is equal to 2 degrees, the error in rotation is 1.2 degree, and the error in the direction of the translation is 6.5 degrees for the linear method and 2.8 degrees for the non-linear method. We have found that the second linear method (which solves for the direction of the translation and the norms of translation vectors) has nearly the same accuracy as the first one. Though the first linear method slightly outperforms the second one.



(a)

(b)

Figure 3: Error in the stereo geometry estimation in the presence of uniform noise perturbing the left and right motions: (a) error in rotation and (b) error in translation, the solid curve corresponds to the linear method (first method), the dashed curve to the non-linear optimization.

4 Real image experiments

We have verified our method by experiments with real image data. To apply our proposed method of self-calibration, we must establish feature correspondences between two monocular images. For this purpose, we have used the *image-matching* algorithm that was developed at the INRIA laboratory [17]. This algorithm provides a set of feature correspondences between two arbitrary images. From these correspondences we have computed camera motions (factorization of the fundamental matrix). Finally, using left and right camera motions the stereo geometry is computed using the method described in Section 2.

All images are provided by two CCD cameras with resolution 512×512. Camera motions (both left and right) are computed between two consecutive positions of the stereo rig. In the first experiment we have used 7 stereo pairs of an indoor scene. Figure 4 shows one pair among them. Therefore, the motion data consist of 6 motions for each camera from which we have computed the stereo geometry. The top of this figure shows the matched features between two consecutive images that are used in computing the left first camera motion.

In the second experiment we have used 4 stereo pairs. Figure 5 shows one pair among them. Therefore, the motion data consist of 3 motions for each camera. The top of this figure shows the matched features between two consecutive images that are used in computing the left third camera motion.

In these two experiments, the number of feature correspondences between 2 consecutive monocular images varies from 70 to 190. One way to assess the quality of the self-calibration results is to compute the residual error of all used criteria. However, since the stereo rig has been calibrated off-line we can consider the result of this calibration as the ground truth and compare the self-calibration results with it.

Table 1 summarizes the residual errors using the self-calibration for the two experiments. The second column shows the residual error associated with the estimation of the rotation. The third column shows the residual error associated with the estimation of the translation. Table 2 shows, for the same experiments, a comparison between the self calibration and the off-line calibration which is considered as the ground truth. More precisely, the second column shows the error in rotation, the third columns the error in the direction of the translation. One can notice that these self-calibration results are quite accurate to obtain a precise 3D shape.

Figure 4: One among the seven stereo pairs used in the first experiment. The matched monocular features are shown with white crosses.

	$\sum \|\mathbf{n}_A - \mathbf{R}_X \mathbf{n}_B\|^2$	Eq. (12)
Exp. 1	0.13	4.5
Exp. 2	0.14	3.2

Table 1: Residual errors associated with the stereo geometry computation. The first row corresponds to the first experiment, the second row to the second experiment.

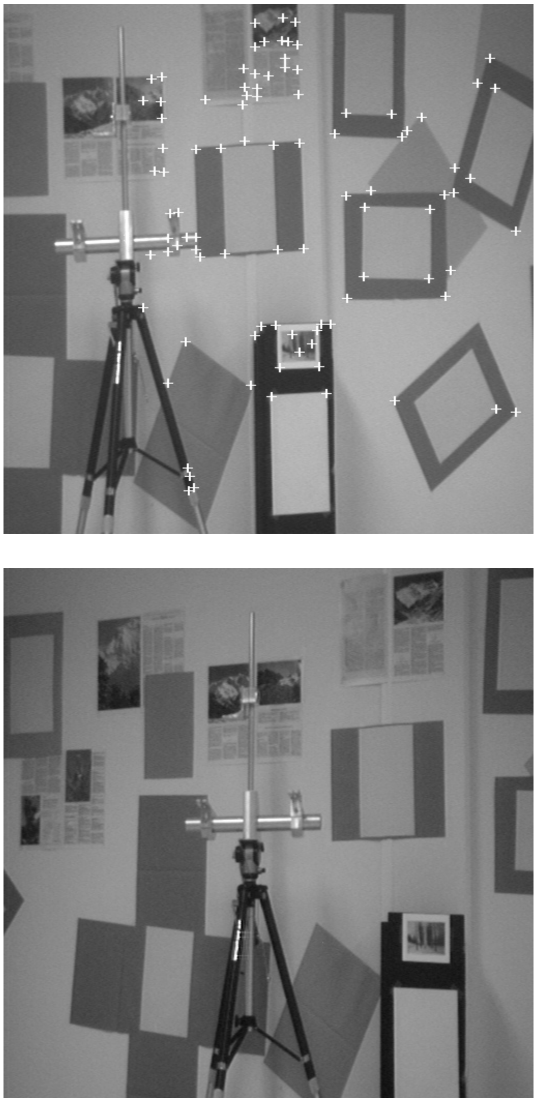


Figure 5: One among the four stereo pairs used in the second experiment. The matched monocular features are shown with white crosses.

	rotation err. (deg)	translation err. (deg)
Exp. 1	3.2	5.7
Exp. 2	4.9	4.5

Table 2: The deviation of the self-calibration results with respect to the off-line calibration. The first row corresponds to the first experiment, the second row to the second experiment.

5 Conclusion

In this paper, we have presented a new method for the recovery of the geometry of a stereo rig which can be used as an alternative for computing the epipolar geometry of the stereo rig. Neither known reference objects nor stereo correspondence is required; our system is therefore self-calibrating and stereo correspondence-less. The only requirement is that for each camera we are able to establish feature correspondences between pairs of monocular images which is more tractable than solving for stereo correspondence. Furthermore, the developed method can be integrated into a cooperative stereo-motion framework. Since the method decouples the stereo geometry estimation from the 3D reconstruction problem, tracking the same 3D scene is no longer required. Thus, deliberate motions can easily be performed in order to determine the stereo geometry in an accurate manner.

We have shown that the computation of the stereo geometry can be carried out using standard linear algebra tools. We also presented a non-linear optimization method for the recovery of the direction of the translation. A unique solution can be obtained when the stereo rig performs two motions with different rotation axes. Experimental results with real images as well as simulations demonstrate the feasibility and robustness of the proposed method. Interesting enough is the 3D accuracy obtained by the method. Although this accuracy is not the same as that of the off-line calibration it is still acceptable given that the stereo geometry is recovered by the use of unknown and natural scenes without any stereo correspondence.

In this work, we have assumed that the intrinsic parameters of the stereo rig are known. In case where they are not known, one can estimate them from motion correspondences using self-calibration techniques for a single camera [8]. Our future work would be on the extension of the current technique to simultaneously estimate the intrinsic and extrinsic parameters of the stereo rig. A more detailed version of this paper can be obtained from the authors on request.

Acknowledgment

The work described in this paper was substantially supported by a grant from the Research Grants Council of Hong Kong Special Administrative Region, China (RGC Ref. No. CUHK4114/97/E). It was also partially supported by the CUHK Postdoctoral Fellowship Scheme 1998 (CUHK Ref. No. 98/15/ERG).

References

[1] F. Dornaika and R. Chung. Cooperative stereo-motion: Matching and reconstruction. In *Proc. of the IEEE International Conference on Intelligent Vehicles*, October 1998.

[2] O. D. Faugeras and M. Hebert. The representation, recognition, and locating of 3D objects. *International Journal of Robotics Research*, 5(3):27–52, 1986.

[3] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd European Conference on Computer Vision*, pages 579–587, 1992.

[4] R. I. Hartley. An algorithm for self calibration from several views. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 908–912, 1994.

[5] R. Horaud and G. Csurka. Self-calibration and Euclidean reconstruction using motions of a stereo rig. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 96–103, January 1998.

[6] R. Horaud and F. Dornaika. Hand eye calibration. *International Journal of Robotics Research*, 14(3):195–210, 1995.

[7] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Amer. A.*, 4(4):629–642, 1987.

[8] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, March 1997.

[9] S. J. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–152, 1992.

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

[11] Y. C. Shiu and S. Ahmad. Calibration of wrist mounted robotic sensors by solving homogeneous transform equations of the form $AX = XB$. *IEEE Transactions on Robotics and Automation*, 5(1):16–29, 1989.

[12] B. Triggs. Autocalibration and the absolute quadric. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 609–614, June 1997.

[13] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of three-dimensional parameters of rigid objects with curved surfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6:13–27, 1984.

[14] R. Y. Tsai and R. K. Lenz. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE J. Robot. Automation*, 5(3):345–358, 1989.

[15] J. Weng, N. Ahuja, and T. S. Huang. Optimal motion and structure estimation. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 1989.

[16] J. Weng, T. S. Huang, and N. Ahuja. *Motion and Structure from Image Sequences.* Springer-Verlag, Berlin, 1993.

[17] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, October 1995.

[18] Z. Zhang, Q.-T. Luong, and O. Faugeras. Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction. *IEEE Trans. on Robotics and Automation*, 12(1):103–113, 1996.

[19] A. Zisserman, P. A. Beardsley, and I. D. Reid. Metric calibration of a stereo rig. In *Proc. of the IEEE Workshop on Representation of Visual Scenes*, pages 93–100, June 1995.