Self-calibration of an uncalibrated stereo rig from one unknown motion

Z. Zhang Q.-T. Luong O. Faugeras
INRIA Sophia-Antipolis, 2004 route des Lucioles
BP 93, F-06902 SOPHIA-ANTIPOLIS Cedex, France
e-mail: zzhang@sophia.inria.fr

Abstract

We address in this paper the problem of self-calibration and metric reconstruction (up to a scale) from one unknown motion of an uncalibrated stereo rig. The epipolar constraint is first formulated for two uncalibrated images. The problem then becomes one of estimating unknowns such that the discrepancy from the epipolar constraint, in terms of sum of squared distances between points and their corresponding epipolar lines, is minimized. Redundancy of the information contained in a sequence of stereo images makes this method more robust than using a sequence of monocular images. Real data have been used to test the proposed method, and the results obtained are quite good. We also show experimentally that it is very difficult to estimate precisely the coordinates of the principal points of cameras. A variation of as high as several dozen pixels in the principal point coordinates does not affect significantly the 3-D reconstruction. A theoretical analysis is provided in this paper to explain this phenomenon.

1 Introduction

A wealth of work on camera calibration has been carried out by researchers either in Photogrammetry or in Computer Vision and Robotics (see [13] for a review). Recently, a number of researchers in Computer Vision and Robotics are trying to develop online camera calibration techniques, known as self-calibration. The idea is to calibrate a camera by just moving it in the surrounding environment. The motion rigidity provides several constraints on the camera intrinsic parameters. They are more commonly known as the *epipolar constraint*, and can be expressed as a 3×3, so-called *fundamental matrix*. Hartley [4] proposed a singular-value-decomposition method to compute the focal lengths from a pair of images if all other cameras parameters are known. Trivedi [11] tried to determine only the coordinates of the principal point of a camera. Maybank and Faugeras [8] proposed a theory of self-calibration. They showed that a camera can be in general completely calibrated from three different displacements. At the same time, they proposed an algorithm using tools from algebraic geometry. However, the algorithm is very sensitive to noise, and is of no practical use. Luong, in cooperation with them, has developed a real practical system as long as the points of interest can be located with sub-pixel precision, say 0.2 pixels, in image planes [7, 2].

In this paper, we describe a self-calibration method for a binocular stereo rig from one displacement using a simplified camera model (i.e., the principal points

BMVC 1994 doi:10.5244/C.8.49

are known). We have made this simplification because the position of the principal point is very difficult to be precisely estimated by calibration, and is in practice very close to the image center. We have shown this experimentally together with a theoretical analysis in this paper. The formulation, however, can be easily extended to include all parameters. Because of the exploitation of information redundancy in the stereo system, our approach yields more robust calibration result than only considering a single camera, as to be shown by experiments with real images.

2 Problem Statement and Notations

2.1 Camera Model

A camera is described by the widely used pinhole model. The coordinates of a 3-D point $M = [x, y, z]^T$ and its retinal image coordinates $\mathbf{m} = [u, v]^T$ are related by $s[u, v, 1]^T = \mathbb{P}[x, y, z, 1]^T$, where s is an arbitrary scale, and $\mathbb{P}$ is a 3×4 matrix, called the perspective projection matrix. Denoting the homogeneous coordinates of a vector $\mathbf{x} = [x, y, \cdots]^T$ by $\tilde{\mathbf{x}}$, i.e., $\tilde{\mathbf{x}} = [x, y, \cdots, 1]^T$, we have $s\tilde{\mathbf{m}} = \mathbb{P}\tilde{M}$.

The basic assumption behind this model is that *the relationship between the world coordinates and the pixel coordinates is linear projective*. This allows us to use the powerful tools of projective geometry, which is emerging as an attractive framework for computer vision [9]. With the state of the art of the technology, camera distortion is reasonably small, and the pinhole model is thus a good approximation.

The matrix $\mathbb{P}$ can be decomposed as

$$\mathbb{P} = \mathbf{A}[\mathbf{R}\ \mathbf{t}] ,$$

where $\mathbf{A}$ is a 3×3 matrix, mapping the normalized image coordinates to the retinal image coordinates, $(\mathbf{R}, \mathbf{t})$ is the displacement (rotation and translation) from the world coordinate system to the camera coordinate system. See [14] for more details.

2.2 Problem Statement

The problem is illustrated in Fig. 1. The left and right images at time t_1 are respectively denoted by I_1 and I_2, and those at time t_2 are denoted by I_3 and I_4. A point $\mathbf{m}$ in the image plane I_i is noted as $\mathbf{m}_i$, and a point M in 3-space expressed in the coordinate system attached to the i-th camera is noted as M_i. The second subscript, if any, will indicate the index of the point in consideration. Thus $\mathbf{m}_{ij}$ is the image point in I_i of the j-th 3-D point, and M_{ij} is the j-th 3-D point expressed in the coordinate system attached to the i-th camera.

Without loss of generality, we choose as the world coordinate system the coordinate system attached to the left camera at t_1. Let $(\mathbf{R}_s, \mathbf{t}_s)$ be the displacement between the left and right cameras of the stereo rig. Let $(\mathbf{R}_l, \mathbf{t}_l)$ be the displacement of the stereo rig between t_1 and t_2 with respect to the left camera. Let $(\mathbf{R}_r, \mathbf{t}_r)$ be the displacement of the stereo rig between t_1 and t_2 with respect to the right camera. Let $\mathbf{A}_l$ and $\mathbf{A}_r$ be the intrinsic matrices of the left and right cameras, respectively. The problem can now be stated as:

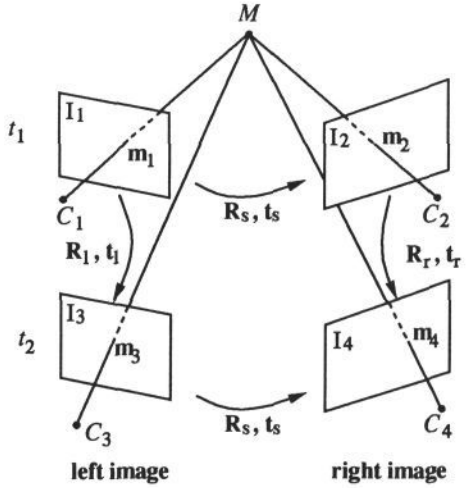


Fig. 1. Illustration of the problem to be studied

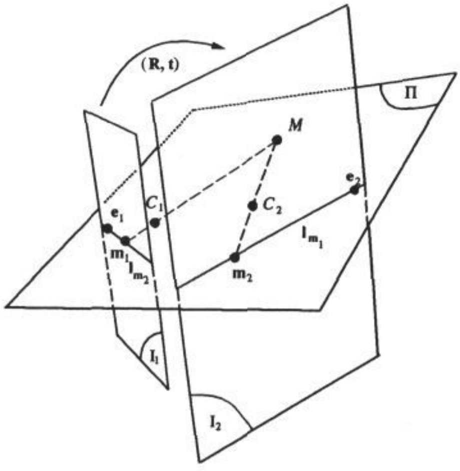


Fig. 2. The epipolar geometry

given • m point correspondences between I_1 and I_2, noted by $\{\mathbf{m}_{1i}^{12}, \mathbf{m}_{2i}^{12}\}$ $(i = 1, \ldots, m)$
• n point correspondences between I_3 and I_4, noted by $\{\mathbf{m}_{3j}^{34}, \mathbf{m}_{4j}^{34}\}$ $(j = 1, \ldots, n)$
• p point correspondences between I_1 and I_3, noted by $\{\mathbf{m}_{1k}^{13}, \mathbf{m}_{3k}^{13}\}$ $(k = 1, \ldots, p)$
• q point correspondences between I_2 and I_4, noted by $\{\mathbf{m}_{2l}^{24}, \mathbf{m}_{4l}^{24}\}$ $(l = 1, \ldots, q)$
estimate the intrinsic matrices $\mathbf{A}_\mathrm{l}$ and $\mathbf{A}_\mathrm{r}$, and the displacements $(\mathbf{R}_\mathrm{s}, \mathbf{t}_\mathrm{s})$, $(\mathbf{R}_\mathrm{l}, \mathbf{t}_\mathrm{l})$ and $(\mathbf{R}_\mathrm{r}, \mathbf{t}_\mathrm{r})$, and the 3-D structure of these points if necessary.

Refer to Fig. 1. As the relative geometry of the two cameras, i.e., $(\mathbf{R}_\mathrm{s}, \mathbf{t}_\mathrm{s})$, does not change between t_1 and t_2, the displacement of the left camera $(\mathbf{R}_\mathrm{l}, \mathbf{t}_\mathrm{l})$ and that of the right $(\mathbf{R}_\mathrm{r}, \mathbf{t}_\mathrm{r})$ are not independent from each other. Indeed, after some simple algebra, we have the following constraints:

$$\mathbf{R}_\mathrm{r} = \mathbf{R}_\mathrm{s}\mathbf{R}_\mathrm{l}{\mathbf{R}_\mathrm{s}}^T \tag{1}$$

$$\mathbf{t}_\mathrm{r} = \mathbf{t}_\mathrm{s} + \mathbf{R}_\mathrm{s}\mathbf{t}_\mathrm{l} - \mathbf{R}_\mathrm{r}\mathbf{t}_\mathrm{s} \ . \tag{2}$$

Furthermore, the overall scale can never be recovered by this system, we can set one of the translations to have unit length, say, $\|\mathbf{t}_\mathrm{s}\| = 1$.

Thus, there are in total 21 unknowns in this system: 5 parameters for each intrinsic matrix, 5 parameters for $(\mathbf{R}_\mathrm{s}, \mathbf{t}_\mathrm{s})$, and 6 parameters for $(\mathbf{R}_\mathrm{l}, \mathbf{t}_\mathrm{l})$.

3 Epipolar Constraint

Considering the case of two cameras as shown in Fig. 2. Let the displacement from the first camera to the second be $(\mathbf{R}, \mathbf{t})$. Let $\mathbf{m}_1$ and $\mathbf{m}_2$ be the images of a 3-D point M on the cameras. Under the pinhole model, we have the following two equations:

$$s_1\tilde{\mathbf{m}}_1 = \mathbf{A}_1\,[\mathbf{I}\ \mathbf{0}]\,\tilde{M}_1 \qquad s_2\tilde{\mathbf{m}}_2 = \mathbf{A}_2\,[\mathbf{R}\ \mathbf{t}]\,\tilde{M}_1\ ,$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are the intrinsic matrices of the first and second cameras, respectively. Eliminating M, s_1 and s_2 from the above equations, we obtain

$$\tilde{\mathbf{m}}_2^T \mathbf{A}_2^{-T} \mathbf{T} \mathbf{R} \mathbf{A}_1^{-1} \tilde{\mathbf{m}}_1 = 0 , \tag{3}$$

where $\mathbf{T}$ is an antisymmetric matrix defined by $\mathbf{t}$ such that $\mathbf{Tx} = \mathbf{t} \wedge \mathbf{x}$ for all 3-D vector $\mathbf{x}$ ($\wedge$ denotes the cross product).

Equation (3) is a fundamental constraint underlying any two images if they are perspective projections of one and the same scene. Let $\mathbf{F} = \mathbf{A}_2^{-T}\mathbf{T}\mathbf{R}\mathbf{A}_1^{-1}$, $\mathbf{F}$ is known as the fundamental matrix of the two images [7, 2]. Without considering 3-D metric entities, we can think of the fundamental matrix as providing the two epipoles (i.e., $\mathbf{e}_1$ and $\mathbf{e}_2$, the vertexes of the two pencils of epipolar lines) and the 3 parameters of the homography between these two pencils, and this is the only geometric information available from two uncalibrated images [8, 7].

4 Problem Solving

4.1 Formulation

From the preceding section, we see that each point correspondence provides one equation of the form Eq. (3). As we have in total $M = m + n + p + q$ point correspondences (see Sect. 2), we can estimate the intrinsic matrices $\mathbf{A}_\mathrm{l}$ and $\mathbf{A}_\mathrm{r}$, and the displacements $(\mathbf{R}_\mathrm{s}, \mathbf{t}_\mathrm{s})$, $(\mathbf{R}_\mathrm{l}, \mathbf{t}_\mathrm{l})$ and $(\mathbf{R}_\mathrm{r}, \mathbf{t}_\mathrm{r})$ by solving a least-squares problem, which minimizes the discrepancy from the epipolar constraint (3), i.e.,

$$\begin{aligned}\min \Bigg[&\sum_{i=1}^{m} \left((\tilde{\mathbf{m}}_{2i}^{12})^T \mathbf{A}_\mathrm{r}^{-T} \mathbf{T}_\mathrm{s} \mathbf{R}_\mathrm{s} \mathbf{A}_\mathrm{l}^{-1} \tilde{\mathbf{m}}_{1i}^{12}\right)^2 + \sum_{j=1}^{n} \left((\tilde{\mathbf{m}}_{4j}^{34})^T \mathbf{A}_\mathrm{r}^{-T} \mathbf{T}_\mathrm{s} \mathbf{R}_\mathrm{s} \mathbf{A}_\mathrm{l}^{-1} \tilde{\mathbf{m}}_{3j}^{34}\right)^2 + \\ &\sum_{k=1}^{p} \left((\tilde{\mathbf{m}}_{3k}^{13})^T \mathbf{A}_\mathrm{l}^{-T} \mathbf{T}_\mathrm{l} \mathbf{R}_\mathrm{l} \mathbf{A}_\mathrm{l}^{-1} \tilde{\mathbf{m}}_{1k}^{13}\right)^2 + \sum_{l=1}^{q} \left((\tilde{\mathbf{m}}_{4l}^{24})^T \mathbf{A}_\mathrm{r}^{-T} \mathbf{T}_\mathrm{r} \mathbf{R}_\mathrm{r} \mathbf{A}_\mathrm{r}^{-1} \tilde{\mathbf{m}}_{2l}^{24}\right)^2 \Bigg] .\end{aligned} \tag{4}$$

The total number of unknowns is 21 (see Sect. 2). On the other hand, we have three independent fundamental matrices, each providing seven constraints on the intrinsic and extrinsic parameters. We thus have in total 21 constraints. This implies that we can in principle solve all the unknowns. However, a set of extremely nonlinear equations are involved, making the parameter initialization impossible[1].

In this paper, we assume a simplified camera model: the angle between the retinal axes θ is $\pi/2$, and the location of the principal point (u_0, v_0) is known, and is assumed to be at the image center in this paper. What we need to estimate is then α_u, α_v for each camera. We have made such simplification for the following reasons:

- With the current technology, the angle θ can be made very close to $\pi/2$. Indeed, we have carried out a large number of experiments on our CCD cameras using a classical calibration method [3], and the differences between the estimated θ and $\pi/2$ are found to be in the order of 10^{-6} radians.
- The position of the principal point is in practice very close to the image center. On the other hand, it is very difficult to be precisely estimated.

[1] More correctly, we have not been able to work out such an algorithm.

Note that this simplification has also been adopted by many researchers [12, 6]. They claim that a deviation of the location of the principal point by a dozen of pixels from the real location does not produce any severe distortion in 3-D reconstruction. This has been confirmed by our experimentation (see Sect. 5), and a theoretical analysis will be given in Sect. 6 to explain this phenomenon.

Under this simplification, we are able to compute an initial estimate of all remaining unknowns from the three fundamental matrices, as can be found in [14].

4.2 Implementation details

Now we describe some implementation details. The minimization problem formulated by Eq. (4) is based on the epipolar constraint Eq. (3). However, it is clear from Eq. (3) that a translation can only be determined up to a scale. We thus normalize each translation such that its norm is 1. More precisely, each translation is represented by its spherical coordinates. A rotation is described by the Rodrigues matrix [10]:

$$\mathbf{R} = \frac{1}{1+\dfrac{a^2+b^2+c^2}{4}} \begin{bmatrix} 1+(a^2-b^2-c^2)/4 & -c+ab/2 & b+ac/2 \\ c+ab/2 & 1+(-a^2+b^2-c^2)/4 & -a+bc/2 \\ -b+ac/2 & a+bc/2 & 1+(-a^2-b^2+c^2)/4 \end{bmatrix} .$$

A rotation is thus represented by three parameters $\mathbf{g} = [a, b, c]^T$. Such a representation is valid for all rotations except for the one whose rotation angle is equal to π (which will not happen in the problem addressed in this paper). We have chosen this parameterization because of its relatively simple expression of its derivative and because there is no constraint on the parameters.

Regarding the constraints on the extrinsic parameters, it is rather easy to incorporate the constraint Eq. (1) in Eq. (4). We do it by simply replacing $\mathbf{R}_r$ by $\mathbf{R}_s\mathbf{R}_l\mathbf{R}_s{}^T$. For the constraint Eq. (2), however, it is much more difficult. As the scale of one of the translations can never be recovered by this system, we can set, say, $\|\mathbf{t}_s\| = 1$. The scales of the other two translations, $\mathbf{t}_l$ and $\mathbf{t}_r$, cannot be recovered by our algorithm. This is because we try to estimate the unknowns by minimizing the discrepancy from the epipolar constraint (quantified by the distance of a point to its epipolar line, see below), and the scales in the translations does not influence the functional to be minimized. This is also clear from the fact that the fundamental matrix is only defined up to a scale factor. Two unknown scales are thus involved in Eq. (2), which implies that Eq. (2) provides only one scalar equation. To be more precise, the scalar equation is given by

$$\left|\mathbf{R}_s\hat{\mathbf{t}}_l \quad (\mathrm{I}-\mathbf{R}_r)\hat{\mathbf{t}}_s \quad \hat{\mathbf{t}}_r\right| = 0 , \tag{5}$$

where $|\cdot|$ denotes the determinant of a 3×3 matrix, and $\hat{\mathbf{t}}$ denotes the unit translation direction vector, i.e., $\hat{\mathbf{t}} = \mathbf{t}/\|\mathbf{t}\|$. The constraint (5) says nothing more than that the three vectors $\mathbf{R}_s\hat{\mathbf{t}}_l$, $(\mathrm{I}-\mathbf{R}_r)\hat{\mathbf{t}}_s$ and $\hat{\mathbf{t}}_r$ are coplanar. This implies that the crossproduct of two vectors, say, $[(\mathrm{I}-\mathbf{R}_r)\hat{\mathbf{t}}_s] \wedge \hat{\mathbf{t}}_r$, should be orthogonal to the other vector, $\mathbf{R}_s\hat{\mathbf{t}}_l$. In our implementation, this constraint multiplied by a coefficient (Lagrange multiplier) is used as an additional measurement in the objective function. The constraint is not satisfied exactly, but has a small value. This value, noted as c, depends on the value of the Lagrange multiplier, noted as λ. The larger the value of λ is, the smaller the value of c is. We set $\lambda = 10^5$, which gives a value of c in the order of 10^{-8}. The unknown scales in $\mathbf{t}_l$ and $\mathbf{t}_r$ can be easily recovered in the 3-D reconstruction phase if, for example, one can identify one point in each image corresponding to a single point in 3-D space.

The minimization is performed by the `Nag` routine `E04GDF`, which is a modified Gauss-Newton algorithm for finding an unconstrained minimum of a sum of squares of M nonlinear functions in N variables ($M \geq N$). It demands the first derivative of the objective function and an initial estimate of the intrinsic and extrinsic parameters to be supplied. Due to space limitation, the reader is referred to [14] for a description of how to compute the initial estimation from Kruppa equations. If we have appropriate knowledge of the cameras (e.g., from constructors' specifications), we can directly use it as the initial estimation.

5 Experimental Results

We show in Fig. 3 the two pairs of stereo images used in this experiment. Two CCD cameras with resolution 512 × 512 are used. The points of interest used for self-calibration are also shown as indicated by the white crosses. These points are extracted with sub-pixel accuracy by an interactive program of Deriche and Blaszka [1]. However, a few points, especially those on the background are not well localized.

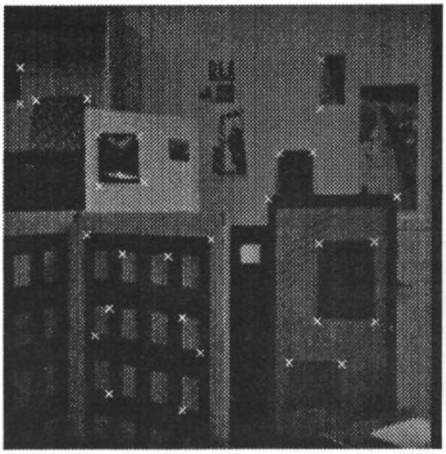
We show in Table 1 a subset of results of the self-calibration on this sequence of images, where the *distance* means the root of mean squares of distances between points and their corresponding epipolar lines (i.e., the root of the objective function divided by the number of correspondences). We see that the average distance decreases from 11.3 pixels to 0.5 pixels. This shows the advantage of our approach which takes into account the stereo correspondences over the previous approach based on calibrating separately each camera. The two cameras are almost the same as they have almost the same intrinsic parameters.

Table 1. Results of the self-calibration: Intrinsic parameters

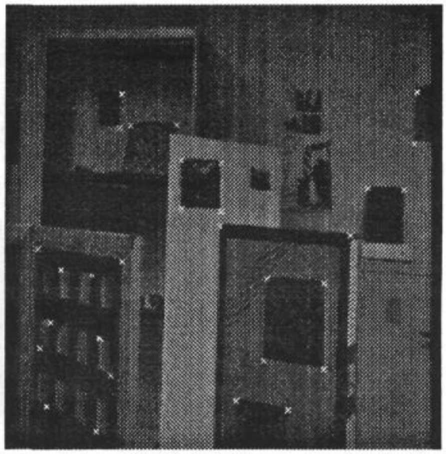
	Left camera		Right camera		distance
	α_u	α_v	α_u	α_v	(pixels)
initialization	597.15	786.99	598.31	902.26	11.3
final estimate	610.99	910.13	617.13	916.58	0.5

After having estimated the intrinsic and extrinsic parameters, we can perform metric reconstruction (up to a scale). In order to have a better visualization, the 3-D reconstructed points are artificially linked, and shown as line segments. Figure 4a shows the back projection of the reconstructed 3-D points on the left image at t_1, which are linked by black line segments, together with the original 2-D points as indicated by white crosses. The projected and original points coincide very well. Figure 4b shows the projection of the 3-D reconstruction on a plane perpendicular to the image plane, which is expected to be parallel to the ground plane. A closeup of the foreground is also given. We see clearly that the reconstructed points of the foreground lie in two planes. The reconstruction of the background is however noisier due to the poor location of the 2-D points.

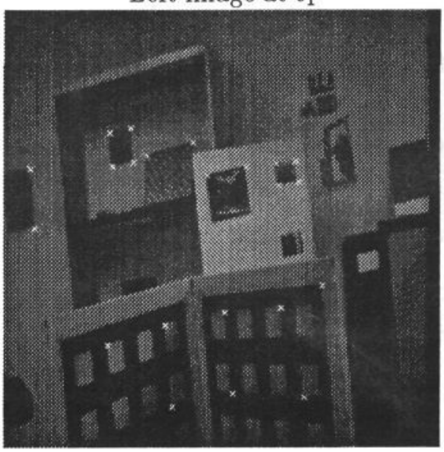
To give an idea of the quantitative performance, we consider the points on the grid pattern because we have manually measured their positions. Using an algorithm similar to that described in [5], we are able to compute the scalar factor,



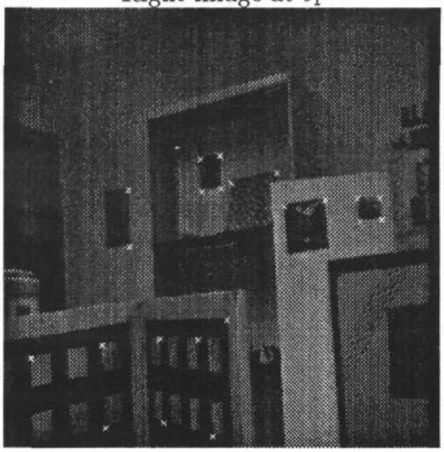
Left image at t_1



Right image at t_1



Left image at t_2



Right image at t_2

Fig. 3. Images with overlay of the points of interest used for self-calibration

the rotation matrix and the translation between the points reconstructed and those measured manually. The scale computed is 301.69. We then apply the estimated transformation to the points reconstructed, and eventually compute the distances between the transformed points and the manually measured ones. The error (the root of the mean squares of the distances) is found to be 0.86 millimeters (the grid size is about 300 millimeters), which is remarkably small remembering that *no knowledge has been used* except that the principal point is assumed to be located at the image center.

Now let us consider the effect of the position of the principal point on the reconstruction. The process is as follows. We shift the coordinates of the principal point from the image center by $(\delta_{u_0}, \delta_{v_0})$, and then carry out the same calibration procedure. Finally we compare the reconstruction result with the manually measured one, as described just above. The results are shown in Table 2, where the errors are still quantified as the root of the mean squares of the distances between the transformed points and the manually measured ones (in millimeters). The image center is at (255, 255) in our case. One example is the number at the third row and the eleventh column, 3.99 millimeters, which corresponds to the error of

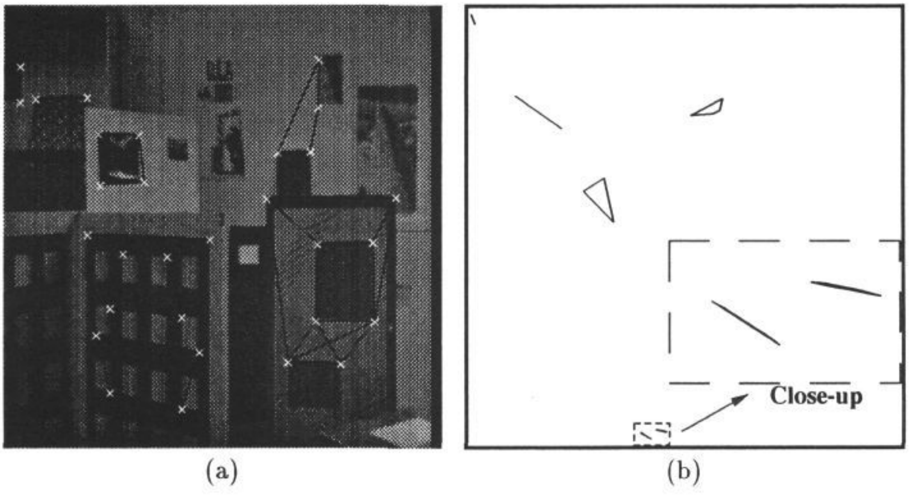


Fig. 4. Final 3-D reconstruction result. (a): Back projection on the left image at t_1; (b): Projection on a plane perpendicular to the image plane (top view)

the reconstruction with $(\delta_{u_0}, \delta_{v_0}) = 25 \times (1, 1)$, i.e., the principal point is assumed to be at (280, 280). From this table, we confirm that the 3-D reconstruction is not very sensitive to the location of the principal points of the cameras. Even with as high as 35 pixels of deviation from the image center, the 3-D reconstruction is still reasonable. Sect. 6 will explain this phenomenon through a formal analysis.

Table 2. Errors in reconstruction (in millimeters) versus positions of the principal points

$(\delta_{u_0}, \delta_{v_0})$	−25	−20	−15	−10	−5	0	5	10	15	20	25
(1, 0)	1.57	1.73	1.70	1.47	1.11	0.86	0.97	1.29	1.76	2.29	2.81
(0, 1)	1.62	1.37	1.15	0.97	0.87	0.86	0.95	1.11	1.30	1.52	1.76
(1, 1)	2.73	2.75	2.51	2.00	1.33	0.86	1.13	1.74	2.49	3.26	3.99
(1,−1)	1.09	1.02	1.06	1.04	0.94	0.86	0.88	0.95	1.14	1.40	1.69

6 Influence of the position of the principal point in 3-D reconstruction

In this section, we examine the reason why it is difficult to localize precisely the principal point. Without loss of generality, we assume that the world coordinate system coincides with the coordinate system of camera. Using the pinhole model in Sect. 2.1, we then have the perspective projection matrix

$$\mathbb{P} = \begin{bmatrix} a & c & u_0 \\ 0 & b & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} ,$$

and the relation between the 3D points M and the image points $\mathbf{m}$ is given by

$$s\tilde{\mathbf{m}} = \mathbb{P}\tilde{M} \tag{6}$$

where a, b, and c are related to the intrinsic parameters α_u, α_v and θ, and s is an arbitrary scale.

Now if the principal point (u_0, v_0) is not very well localized, say, if it is shifted from the true position by $(\Delta u, \Delta v)$, we want to know its influence in 3-D reconstruction. The perspective projection matrix is now

$$\mathbb{P}' = \begin{bmatrix} a & c & u_0 + \Delta u \\ 0 & b & v_0 + \Delta v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} .$$

For the same set of image points $\mathbf{m}$, the 3-D reconstruction M' will satisfy the following equation:

$$s\tilde{\mathbf{m}} = \mathbb{P}'\tilde{M}' \tag{7}$$

From Equations 6 and 7, we have

$$\begin{bmatrix} a & c & u_0 \\ 0 & b & v_0 \\ 0 & 0 & 1 \end{bmatrix} M = \begin{bmatrix} a & c & u_0 + \Delta u \\ 0 & b & v_0 + \Delta v \\ 0 & 0 & 1 \end{bmatrix} M' ,$$

or,

$$M' = \mathbf{D}M ,$$

where

$$\mathbf{D} = \begin{bmatrix} a & c & u_0 + \Delta u \\ 0 & b & v_0 + \Delta v \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} a & c & u_0 \\ 0 & b & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\frac{\Delta U}{a} + \frac{c\Delta v}{ab} \\ 0 & 1 & -\frac{\Delta v}{b} \\ 0 & 0 & 1 \end{bmatrix} .$$

Thus, due to the bad localization of the principal point, the reconstructed 3-D points are distorted from the true ones according to $\mathbf{D}$.

In practice, c is in the order of 10^{-3}; a and b are in the order of 10^3. Thus, it is clear that the matrix $\mathbf{D}$ is very close to identity, i.e., the distortion is very small, even when the position of the principal point is shifted by several dozen pixels from the true position. This is one main reason why it is difficult to localize precisely the principal point.

7 Conclusion

In this paper, we have described a new method for calibrating a stereo rig by moving it in an environment without using any reference points (self-calibration). The only geometric constraint between a pair of uncalibrated images is the *epipolar constraint*, which has been formulated in this paper from a point of view in Euclidean space. The problem of self-calibration has then been formulated as one of estimating unknowns such that the discrepancy from the epipolar constraint, in terms of sum of squared distances between points and their corresponding epipolar lines, is minimized. As the minimization problem is nonlinear, an initial estimate of the unknowns must be supplied. The initialization is done based on the work of Maybank, Luong and Faugeras on self-calibration of a single moving camera, which requires to solve a set of so-called Kruppa equations. One point which differs our work from the previous ones in that our formulation is directly built in the measurement space and is thus physically more meaningful. Furthermore, stereo setup is used in our work. Redundancy of the information contained in a sequence of stereo images makes this method more robust than using a sequence of monocular images. This has been demonstrated with real data. The results obtained are very good. We have also shown experimentally that it is very difficult to estimate precisely the principal point. A variation of as high as several dozens of pixels in the principal point coordinates does not affect significantly the 3-D reconstruction. This phenomenon has been explained through a theoretical analysis.

Our future work will be on the extension of the current technique to take into account all camera parameters and to include more image views.

References

[1] R. Deriche and T. Blaszka, "Recovering and characterizing image features using an efficient model based approach," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, (New York, NY), June 1993.

[2] O. Faugeras, Q.-T. Luong, and S. Maybank, "Camera self-calibration: theory and experiments," in *Proc. Second European Conf. Comput. Vision* (G. Sandini, ed.), pp. 563–578, Springer-Verlag, Lecture Notes in Computer Science 588, May 1992.

[3] O. Faugeras and G. Toscani, "The calibration problem for stereo," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, (Miami, FL), pp. 15–20, IEEE, June 1986.

[4] R. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *Proc. Second European Conf. Comput. Vision*, pp. 579–587, 1992.

[5] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 7, pp. 629–642, Apr. 1987.

[6] K. Kanatani and Y. Onodera, "Anatomy of camera calibration using vanishing points," *IEICE transactions on informations and systems*, vol. E74, pp. 3369–3378, Oct. 1991.

[7] Q.-T. Luong, *Matrice fondamentale et calibration visuelle sur l'environnement: Vers une plus grande autonomie des systèmes robotiques.* Dissertation, University of Paris XI, Orsay, Paris, France, Dec. 1992.

[8] S. Maybank and O. Faugeras, "A theory of self-calibration of a moving camera," *Int'l J. Comput. Vision*, vol. 8, pp. 123–152, Aug. 1992.

[9] J. L. Mundy and A. Zisserman, eds., *Geometric invariance in computer vision.* MIT Press, 1992.

[10] C. C. Slama, ed., *Manual of Photogrammetry.* American Society of Photogrammetry, fourth ed., 1980.

[11] H. P. Trivedi, "Can multiple views make up for lack of camera registration," *Image and Vision Computing*, vol. 6, pp. 29–32, Feb. 1988.

[12] R. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, (Miami, FL), pp. 364–374, IEEE, June 1986.

[13] R. Tsai, "Synopsis of recent progress on camera calibration for 3D machine vision," in *The Robotics Review* (O. Khatib, J. Craig, and T. Lozano-Pirez, eds.), pp. 147–159, MIT Press, 1989.

[14] Z. Zhang, Q.-T. Luong, and O. Faugeras, "Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction," Research Report 2079, INRIA Sophia-Antipolis, Valbonne, France, 1993.