Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction

Zhengyou Zhang, Quang-Tuan Luong, and Olivier Faugeras, *Senior Member, IEEE*

Abstract— **We address in this paper the problem of self-calibration and metric reconstruction (up to a scale factor) 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. Although the full self-calibration is theoretically possible, we assume in this paper that the coordinates of the principal point of each camera are known. Then, the initialization of the unknowns can be done based on our previous work on self-calibration of a single moving camera, which requires to solve a set of so-called Kruppa equations. 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 has 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.**

I. Introduction

IT IS well recognized [1]–[3] that stereoscopic cues are important in understanding the 3-D environment surrounding us and allow robust algorithms for 3-D reconstruction to be applied. In order for the 3-D reconstruction to be possible, one needs to know the relationship between the 3-D world coordinates and their corresponding 2-D image coordinates for each camera, and the relative geometry between the two cameras. This is the purpose of camera calibration. A wealth of work on camera calibration has been carried out by researchers either in Photogrammetry [4], [5] or in Computer Vision and Robotics [6]–[12] (see [13] for a review). The usual method of calibration is to compute cameras parameters from one or more images of an object of *known size and shape*, for example, a flat plate with a regular pattern marked on it. One problem is that it is impossible to calibrate online, while the cameras are involved in a visual task [14]. Any change of camera calibration occurring during the performance of the task cannot be corrected without interrupting the task. The change may be deliberate, for example the focal length of a camera may be adjusted, or it may be accidental, for example the camera may undergo small mechanical or thermal changes. In many situations such as vision-based planetary exploration, it is not very practical to calibrate cameras with a calibration apparatus. We can either send a calibration apparatus together with the planetary rover and observe it each time we need to calibrate the cameras, which is not very realistic, or we can precalibrate the stereo system on the ground, which is not reliable.

Recently, a number of researchers in Computer Vision and Robotics have been 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 [15] proposed a singular-value-decomposition method to compute the focal lengths from a pair of images if all other cameras parameters are known. Trivedi [16] tried to determine only the coordinates of the principal point of a camera. Maybank and Faugeras [14] 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 subpixel precision, say 0.2 pixels, in image planes [17], [18].

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 are known). We have made this simplification because it is very difficult to estimate precisely the position of the principal point by calibration, and is in practice very close to the image center. We have shown this experimentally in Section VI. 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 results than those which consider a single camera, as to be shown by experiments with real images. Section II describes the calibration problem to be addressed in this paper. Section III summarizes the epipolar constraint, which is the fundamental constraint underlying all self-calibration techniques. Section IV deals with the details of the problem solving, including additional constraints present in a stereo rig. As the problem is solved by a nonlinear optimization, an initial estimation of the camera parameters must be supplied, which is described in Section V. Experimental results with real data are provided in Section VI. In Section VII, we discuss the possibility to

Manuscript received November 15, 1993; revised July 25, 1994.

Z. Zhang and O. Faugeras are with INRIA Sophia-Antipolis, F-06902 Sophia-Antipolis, Cedex, France (e-mail: zzhang@sophia.inria.fr).

Q.-T. Luong is with SRI International, Menlo Park, CA 94025 USA.

Publisher Item Identifier S 1042-296X(96)01061-5.

include cross correspondences and the possibility to estimate all camera parameters.

This work is an extension of our previous work described in [19], where the intrinsic parameters of the cameras were supposed to be known while the extrinsic ones were unknown.

II. Problem Statement and Notations

A. 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

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbb{P}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

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 [20]. With the state of the art 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. The most general matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \begin{bmatrix} -fk_u & fk_u \cot\theta & u_0 \\ 0 & -\frac{fk_v}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where

- f is the focal length of the camera;
- k_u and k_v are the horizontal and vertical scale factors, whose inverses characterize the size of the pixel in the world coordinate unit;
- u_0 and v_0 are the coordinates of the principal point of the camera, i.e., the intersection between the optical axis and the image plane; and
- θ is the angle between the retinal axes. This parameter is introduced to account for the fact that the pixel grid may not be exactly orthogonal. In practice it is very close to $\pi/2$.

It is clear that we cannot separate f from k_u and k_v. In the following, we use the following notations: $\alpha_u = -fk_u$ and $\alpha_v = -fk_v$. We thus have five intrinsic parameters for each camera: α_u, α_v, u_0, v_0, and θ.

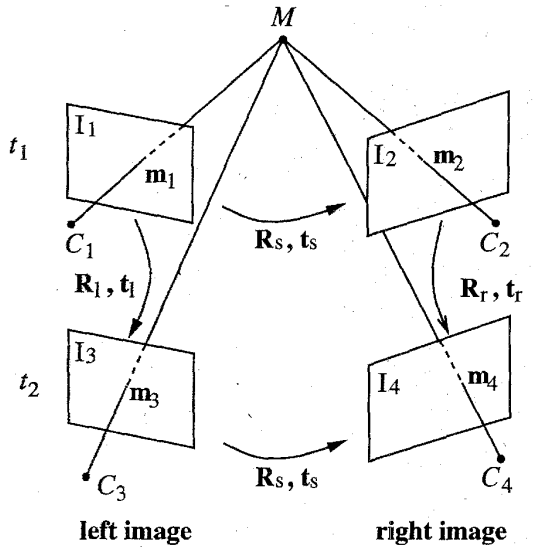


Fig. 1. Illustration of the problem to be studied.

B. 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 follows.

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 $\mathbf{A}_l$ and $\mathbf{A}_r$, and the displacements $(\mathbf{R}_s, \mathbf{t}_s)$, $(\mathbf{R}_l, \mathbf{t}_l)$ and $(\mathbf{R}_r, \mathbf{t}_r)$, and the 3-D structure of these points if necessary.

Refer to Fig. 1. As the relative geometry of the two cameras, i.e., $(\mathbf{R}_s, \mathbf{t}_s)$, does not change between t_1 and t_2,

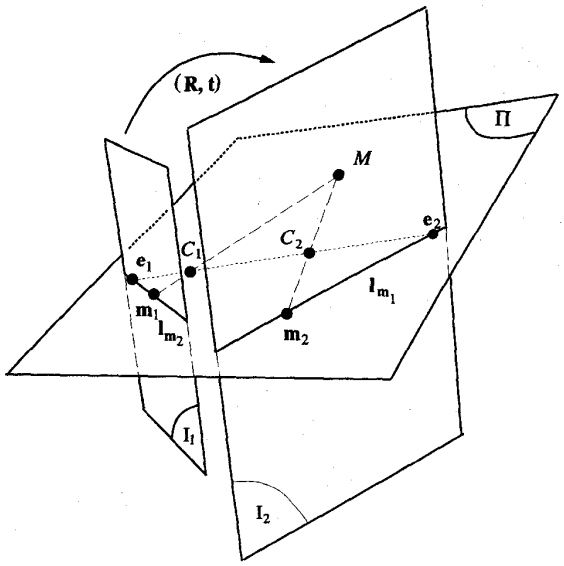


Fig. 2. The epipolar geometry.

the displacement of the left camera $(\mathbf{R}_l, \mathbf{t}_l)$ and that of the right $(\mathbf{R}_r, \mathbf{t}_r)$ are not independent from each other. Indeed, after some simple algebra, we have the following constraints:

$$\mathbf{R}_r = \mathbf{R}_s \mathbf{R}_l \mathbf{R}_s{}^T \tag{2}$$

$$\mathbf{t}_r = \mathbf{t}_s + \mathbf{R}_s \mathbf{t}_l - \mathbf{R}_r \mathbf{t}_s. \tag{3}$$

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

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

III. Epipolar Constraint

Let us consider the case of two cameras as shown in Fig. 2.

Let C_1 and C_2 be the optical centers of the first and second cameras, respectively. Given a point $\mathbf{m}_1$ in the first image, its corresponding point in the second image is constrained to lie on a line called the *epipolar line* of $\mathbf{m}_1$, denoted by $\mathbf{l}_{\mathbf{m}_1}$. The line $\mathbf{l}_{\mathbf{m}_1}$ is the intersection of the plane Π, defined by $\mathbf{m}_1$, C_1 and C_2 (known as the *epipolar plane*), with the second image plane I_2. This is because image point $\mathbf{m}_1$ may correspond to an arbitrary point on the semiline C_1M (M may be at infinity) and that the projection of C_1M on I_2 is the line $\mathbf{l}_{\mathbf{m}_1}$. Furthermore, one observes that all epipolar lines of the points in the first image pass through a common point $\mathbf{e}_2$, which is called the *epipole*. $\mathbf{e}_2$ is the intersection of the line C_1C_2 with the image plane I_2. This can be easily understood as follows. For each point $\mathbf{m}_{1k}$ in the first image I_1, its epipolar line $\mathbf{l}_{\mathbf{m}_{1k}}$ in I_2 is the intersection of the plane Π^k (defined by $\mathbf{m}_{1k}$, C_1 and C_2) with image plane I_2. All epipolar planes Π^k thus form a pencil of planes containing the line C_1C_2. They must intersect I_2 at a common point, which is $\mathbf{e}_2$. Finally, one can easily see the symmetry of the epipolar geometry. The corresponding point in the first image of each point $\mathbf{m}_{2k}$ lying on $\mathbf{l}_{\mathbf{m}_{1k}}$ must lie on the epipolar line $\mathbf{l}_{\mathbf{m}_{2k}}$, which is the intersection of the same plane Π^k with the first image plane I_1. All epipolar lines form a pencil containing the epipole $\mathbf{e}_1$, which is the intersection of the line C_1C_2 with the image plane I_1. If $\mathbf{m}_1$ (a point in I_1) and $\mathbf{m}_2$ (a point in I_2) correspond to a single physical point M in space, then $\mathbf{m}_1$, $\mathbf{m}_2$, C_1 and C_2 must lie in a single plane. This is the well-known *co-planarity constraint* or *epipolar equation* in solving motion and structure from motion problems when the intrinsic parameters of the cameras are known [21].

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:

$$s_1 \tilde{\mathbf{m}}_1 = \mathbf{A}_1 [\mathbf{I}\, \mathbf{0}] \begin{bmatrix} m \\ 1 \end{bmatrix}$$

$$s_2 \tilde{\mathbf{m}}_2 = \mathbf{A}_2 [\mathbf{R}\, \mathbf{t}] \begin{bmatrix} m \\ 1 \end{bmatrix}$$

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

$$\tilde{\mathbf{m}}_2^T \mathbf{A}_2^{-T} [\mathbf{t}]_\times \mathbf{R} \mathbf{A}_1^{-1} \tilde{\mathbf{m}}_1 = 0 \tag{4}$$

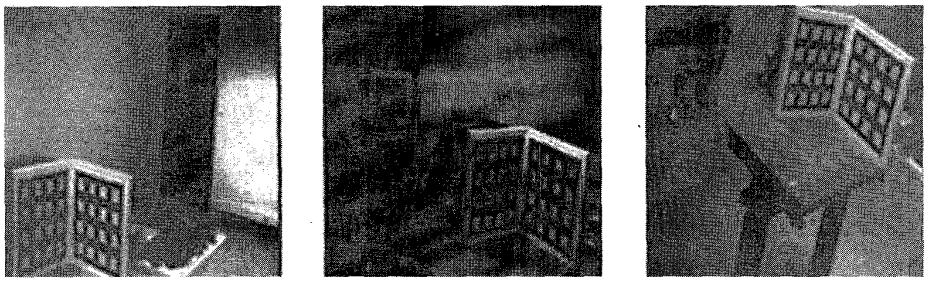
where $[\mathbf{t}]_\times$ is an antisymmetric matrix defined by $\mathbf{t}$ such that $[\mathbf{t}]_\times \mathbf{x} = \mathbf{t} \times \mathbf{x}$ for all 3-D vector $\mathbf{x}$ ($\times$ denotes the cross product).

Equation (4) is a fundamental constraint underlying any two images if they are perspective projections of the same static scene. Let $\mathbf{F} = \mathbf{A}_2^{-T} [\mathbf{t}]_\times \mathbf{R} \mathbf{A}_1^{-1}$, $\mathbf{F}$ is known as the fundamental matrix of the two images [17], [18]. 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 [14], [17]. This implies that the fundamental matrix has only seven degrees of freedom. Indeed, it is only defined up to a scale factor and its determinant is zero. Geometrically, $\mathbf{F}\tilde{\mathbf{m}}_1$ defines the epipolar line of point $\mathbf{m}_1$ in the second image. Equation (4) says no more than that the correspondence in the right image of point $\mathbf{m}_1$ lies on the corresponding epipolar line. Transposing (4) yields the symmetric relation from the second image to the first image.

IV. Problem Solving

A. Formulation

From the preceding section, we see that each point correspondence provides one equation of the form (4). As we have in total $M = m+n+p+q$ point correspondences (see Section II), we can estimate the intrinsic matrices $\mathbf{A}_l$ and $\mathbf{A}_r$, and the displacements $(\mathbf{R}_s, \mathbf{t}_s)$, $(\mathbf{R}_l, \mathbf{t}_l)$ and $(\mathbf{R}_r, \mathbf{t}_r)$ by solving a least-squares problem, which minimizes the discrepancy from the epipolar constraint (4), i.e.,



(a)

Image	u_0	v_0	α_u	α_v	$\lvert\theta - \pi/2\rvert$
1	211.77	305.16	684.27	1012.25	3.25×10^{-6}
2	297.11	294.23	697.05	1030.90	3.01×10^{-6}
3	299.29	186.17	685.10	1014.49	1.40×10^{-6}

(b)

Fig. 3. An example to show the difficulty of localizing the principal point of a camera. (a) Images. (b) Intrinsic parameters estimated with a classical calibration method.

$$\min\Bigg[\sum_{i=1}^{m}\left((\tilde{\mathbf{m}}_{2i}^{12})^T\mathbf{A}_r^{-T}[\mathbf{t}_s]_\times\mathbf{R}_s\mathbf{A}_l^{-1}\tilde{\mathbf{m}}_{1i}^{12}\right)^2 + \sum_{j=1}^{n}\left((\tilde{\mathbf{m}}_{4j}^{34})^T\mathbf{A}_r^{-T}[\mathbf{t}_s]_\times\mathbf{R}_s\mathbf{A}_l^{-1}\tilde{\mathbf{m}}_{3j}^{34}\right)^2 + \sum_{k=1}^{p}\left((\tilde{\mathbf{m}}_{3k}^{13})^T\mathbf{A}_l^{-T}[\mathbf{t}_l]_\times\mathbf{R}_l\mathbf{A}_l^{-1}\tilde{\mathbf{m}}_{1k}^{13}\right)^2 + \sum_{l=1}^{q}\left((\tilde{\mathbf{m}}_{4l}^{24})^T\mathbf{A}_r^{-T}[\mathbf{t}_r]_\times\mathbf{R}_r\mathbf{A}_r^{-1}\tilde{\mathbf{m}}_{2l}^{24}\right)^2\Bigg]. \quad (5)$$

The total number of unknowns is 21 (see Section II). 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 [7], 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 estimate it precisely. To show this, Fig. 3(a) shows three images taken from three different positions by the same camera. The corners on the grids have been used to estimate the intrinsic parameters using a classical calibration method [7]. The results are given in Fig. 3(b). We see that the variation of u_0 and v_0 is quite large from position to position.

Note that this simplification has also been adopted by many researchers [6], [22]. 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 Section VI).

Under this simplification, we are able to compute an initial estimate of all remaining unknowns from the three fundamental matrices, as to be described in Section V.

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

B. Implementation Details

Now we describe some implementation details. The minimization problem formulated by (5) is based on the epipolar constraint (see (4)). However, it is clear from (4) 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 [23], shown at the bottom of the page. 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

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

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 (2) in (5). We do it by simply replacing $\mathbf{R}_r$ by $\mathbf{R}_s\mathbf{R}_l{\mathbf{R}_s}^T$. For the constraint (3), however, it is much more difficult. As the scale of one of the translations can never be recovered by this system, we can set, for example, $\|\mathbf{t}_s\| = 1$. 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 (3), which implies that (3) provides only one scalar equation. To be more precise, the scalar equation is given by

$$\left|\mathbf{R}_s\hat{\mathbf{t}}_l \quad (I-\mathbf{R}_r)\hat{\mathbf{t}}_s \quad \hat{\mathbf{t}}_r\right| = 0 \tag{6}$$

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 (6) says nothing more than that the three vectors $\mathbf{R}_s\hat{\mathbf{t}}_l$, $(I-\mathbf{R}_r)\hat{\mathbf{t}}_s$ and $\hat{\mathbf{t}}_r$ are coplanar. This implies that the crossproduct of two vectors, say, $[(I-\mathbf{R}_r)\hat{\mathbf{t}}_s] \times \hat{\mathbf{t}}_r$, 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 problem described by (5) is to minimize a sum of terms of the form

$$\left(\tilde{\mathbf{m}}_k^T\mathbf{F}_{ij}\tilde{\mathbf{m}}_l\right)^2 \tag{7}$$

where $\mathbf{F}_{ij} = \mathbf{A}_i^{-T}[\mathbf{t}]_\times\mathbf{R}\mathbf{A}_j^{-1}$ $(i,j = l,r)$ is a fundamental matrix. This criterion, however, does not have direct interpretation in the measurement space, i.e., in the image planes. As $\mathbf{l}_l = \mathbf{F}_{ij}\tilde{\mathbf{m}}_l$ actually defines the epipolar line of $\tilde{\mathbf{m}}_l$, we can replace (7) by the squared distance from $\tilde{\mathbf{m}}_k$ to $\mathbf{l}_l$, i.e.,

$$d^2(\tilde{\mathbf{m}}_k, \mathbf{l}_l) = (\tilde{\mathbf{m}}_k^T\mathbf{F}_{ij}\tilde{\mathbf{m}}_l)^2/(l_1^2 + l_2^2)$$

where l_1 and l_2 are the first two elements of the vector $\mathbf{l}_l$. $d(\tilde{\mathbf{m}}_k, \mathbf{l}_l)$ is the Euclidean distance of point $\mathbf{m}_k$ to line $\mathbf{l}_l$ in the image plane. In order to maintain the symmetry between the two cameras to avoid any discrepancy in the epipolar geometry, the term (7) is in fact replaced by the following two terms:

$$d^2(\tilde{\mathbf{m}}_k, \mathbf{F}_{ij}\tilde{\mathbf{m}}_l) + d^2(\tilde{\mathbf{m}}_l, \mathbf{F}_{ij}^T\tilde{\mathbf{m}}_k).$$

The new criterion is more meaningful because Euclidean distances in the measurement space are used.

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 $(M \geq N)$ [24]. It requires the first derivative of the objective function and an initial estimate of the intrinsic and extrinsic parameters. In the next section, we shall describe how to compute the initial estimation. If we have appropriate knowledge of the cameras (e.g., from constructors' specifications), we can directly use it as the initial estimation.

In summary, there are $N = 16$ unknowns to be estimated (2 intrinsic parameters for each camera, 5 for $(\mathbf{R}_s, \mathbf{t}_s)$, 5 for $(\mathbf{R}_l, \mathbf{t}_l)$ and 2 for $\mathbf{t}_r$) under the constraint (6). We need at least $m = 15$ point correspondences. However, in order to use the initialization technique described in the next section, we must have $m + n \geq 8$, $p \geq 8$, and $q \geq 8$. In that case, we have $M \geq 24$ and the problem is always overdetermined.

V. Initialization of the Parameters to Be Estimated

In this section, we briefly outline the process of the initialization of the parameters to be estimated, and provide appropriate references if necessary.

First, based on the epipolar constraint (4), i.e.,

$$\tilde{\mathbf{m}}_k^T\mathbf{F}_{ij}\tilde{\mathbf{m}}_l = 0$$

we can estimate the fundamental matrices $\mathbf{F}_{ll}$, $\mathbf{F}_{rr}$, and $\mathbf{F}_{rl}$ between the left images, the right images, and the left and right images, respectively, from the given point correspondences. Several methods have been proposed in [25]. We have used a linear least-squares method followed by a nonquadratic minimization to improve the results.

Consider now the case of the left camera, i.e., $\mathbf{F}_{ll}$. For the reason of simplicity, the subscript will be omitted. From the definition of $\mathbf{F}$, we have $\mathbf{E} = \mathbf{A}^T\mathbf{F}\mathbf{A}$, where $\mathbf{E} = [\mathbf{t}]_\times\mathbf{R}$ is called the essential matrix. As is well known [26], $\mathbf{E}$ is subject to two independent polynomial constraints. This implies that the entries of $\mathbf{A}$ are subject to two independent polynomial constraints inherited from $\mathbf{E}$. As there are only two unknowns, α_u and α_v, they can then be solved. Let us now follow the theory of self-calibration developed by Maybank and Faugeras [14]. Consider the absolute conic: $x_1^2 + x_2^2 + x_3^2 = 0$, $x_4 = 0$. If we define $x = x_1/x_3$ and $y = x_2/x_3$, then the equation can be rewritten as $x^2 + y^2 = -1$, which represents an imaginary circle of radius $i = \sqrt{-1}$ on the plane at infinity. The important property of the absolute conic is that its image ω on the camera does not change from the position of the camera as long as the same intrinsic parameters as maintained. In fact, the image ω is a conic described by the matrix $\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1}$. This implies that the two epipolar lines *tangent* to ω in the first image must correspond to the two epipolar lines *tangent* to ω in the second image under the epipolar transformation defined by the fundamental matrix $\mathbf{F}$. This provided two independent, the so-called *Kruppa equations*. They are of degree 2 in the

entries of matrix $\mathbf{K}$,

$$\mathbf{K} \equiv \begin{bmatrix} -\delta_{23} & \delta_3 & \delta_2 \\ \delta_3 & -\delta_{13} & \delta_1 \\ \delta_2 & \delta_1 & -\delta_{12} \end{bmatrix}$$

where $\mathbf{K}$ is the matrix defining the dual conic of ω, i.e., $\mathbf{K} = \mathbf{B}^* = \mathbf{A}\mathbf{A}^T$, defined up to a scale factor. We thus have the following relations:

$$\begin{aligned} \delta_1 &= v_0 \\ \delta_2 &= u_0 \\ \delta_3 &= u_0 v_0 - \alpha_u \alpha_v \frac{\cot\theta}{\sin\theta} \\ \delta_{12} &= -1 \\ \delta_{23} &= -u_0^2 - \frac{\alpha_u^2}{\sin^2\theta} \\ \delta_{13} &= -v_0^2 - \frac{\alpha_v^2}{\sin^2\theta}. \end{aligned}$$

To be more precise, the Kruppa equations can be derived as follows. Consider the epipolar line $\mathbf{l}$ tangent to ω in the first image. It is defined by the epipole $\mathbf{e}_1$ and a point, say, $\mathbf{y}$, and therefore, $\mathbf{l} = \mathbf{e}_1 \times \mathbf{y}$. The epipolar line $\mathbf{l}$ is tangent to ω if and only if

$$(\mathbf{e}_1 \times \mathbf{y})^T \mathbf{K} (\mathbf{e}_1 \times \mathbf{y}) = 0 \quad or \quad \mathbf{y}^T [\mathbf{e}_1]_\times \mathbf{K} [\mathbf{e}_1]_\times \mathbf{y} = 0. \quad (8)$$

The epipolar line corresponding to the point $\mathbf{y}$ is $\mathbf{F}\mathbf{y}$ and is tangent to ω in the second image if and only if

$$\mathbf{y}^T \mathbf{F}^T \mathbf{K} \mathbf{F} \mathbf{y} = 0. \quad (9)$$

Writing that (8) and (9) are equivalent yield the so-called Kruppa equations. One possible way to do so is to take $\mathbf{y} = (1, \tau, 0)^T$ in the case where the epipoles are at finite distance. The relations (8) (respectively, (9)) take the form $P_1(\tau) = k_0 + k_1\tau + k_2\tau^2 = 0$ (respectively, $P_2(\tau) = k_0' + k_1'\tau + k_2'\tau^2 = 0$), and the Kruppa equations can be written as three proportionality conditions between these two polynomials, of which only two are independent:

$$\begin{aligned} k_2 k_1' - k_2' k_1 &= 0 \\ k_0 k_1' - k_0' k_1 &= 0 \\ k_0 k_2' - k_0' k_2 &= 0. \end{aligned}$$

The reader is referred to [14] and [17] for more details about the epipolar transformation and the Kruppa equations.

In our particular case (u_0, v_0 and θ are known), the Kruppa equations reduce to two simple equations of the form

$$a_1 x_1^2 + b_1 x_1 x_2 + c_1 x_1 + d_1 x_2 + e_1 = 0 \quad (10)$$

$$a_2 x_2^2 + b_2 x_1 x_2 + c_2 x_2 + d_2 x_1 + e_2 = 0 \quad (11)$$

where $x_1 = \alpha_u^2$, $x_2 = \alpha_v^2$, and $a_1, b_1, \cdots, e_2$ are coefficients which can be computed from $\mathbf{F}$, u_0, v_0 and θ. We must point out that we always have $a_1 a_2 = b_1 b_2$. We can thus solve x_2 in terms of x_1 from the first equation, and substitute it into the second equation to obtain an equation of degree 3 only in x_1. So, we can have at most three real solutions, and usually only one physical solution such that $x_1 \geq 0$ and $x_2 \geq 0$.

TABLE I
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

Once we have computed α_u and α_v, we can compute the essential matrix by $\mathbf{E} = \mathbf{A}^T \mathbf{F} \mathbf{A}$. The rotation and translation are then computed from $\mathbf{E}$ by any standard methods such as those described in [27]–[29]. We thus compute α_{ul}, α_{vl}, $\mathbf{R}_l$, and $\mathbf{t}_l$ for the left camera from the fundamental matrix $\mathbf{F}_{ll}$. In the same way, we can compute α_{ur}, α_{vr}, $\mathbf{R}_r$, and $\mathbf{t}_r$ for the right camera from the fundamental matrix $\mathbf{F}_{rr}$.

It remains to compute the rotation $\mathbf{R}_s$ and the translation $\mathbf{t}_s$ between the left and right cameras. Since the intrinsic matrices $\mathbf{A}_l$ and $\mathbf{A}_r$ are now available, we can easily compute the essential matrix $\mathbf{E}_s = \mathbf{A}_r^T \mathbf{F}_{rl} \mathbf{A}_l$. $\mathbf{R}_s$ and $\mathbf{t}_s$ can then be solved from $\mathbf{E}_s$ using the same method as for $(\mathbf{R}_l, \mathbf{t}_l)$ and $(\mathbf{R}_r, \mathbf{t}_r)$.

In summary, the initial estimates are obtained as follows.

- First, compute the fundamental matrices $\mathbf{F}_{rl}$, $\mathbf{F}_{ll}$, and $\mathbf{F}_{rr}$ from the stereo correspondences, the temporal correspondences of the left camera, and the temporal correspondences of the right camera, respectively.
- Second, compute $\mathbf{A}_l$, $\mathbf{R}_l$, and $\mathbf{t}_l$ from $\mathbf{F}_{ll}$; compute $\mathbf{A}_r$, $\mathbf{R}_r$, and $\mathbf{t}_r$ from $\mathbf{F}_{rr}$.
- Finally, compute $\mathbf{R}_s$ and $\mathbf{t}_s$ from $\mathbf{F}_{rl}$.

VI. EXPERIMENTAL RESULTS

We show in Fig. 4 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 subpixel accuracy by an interactive program of Blaszka and Deriche [30]. However, a few points, especially those on the background are not well localized.

We show in Table I 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.

After having estimated the intrinsic and extrinsic parameters, we can perform metric reconstruction (up to a scale factor). For comparison, we show both the 3-D reconstruction result with the initial estimate (Fig. 5) and that with the final estimate (Fig. 6). In order to have a better visualization, the 3-D reconstructed points are artificially linked, and shown as line segments. Figs. 5(a) and 6(a) show 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

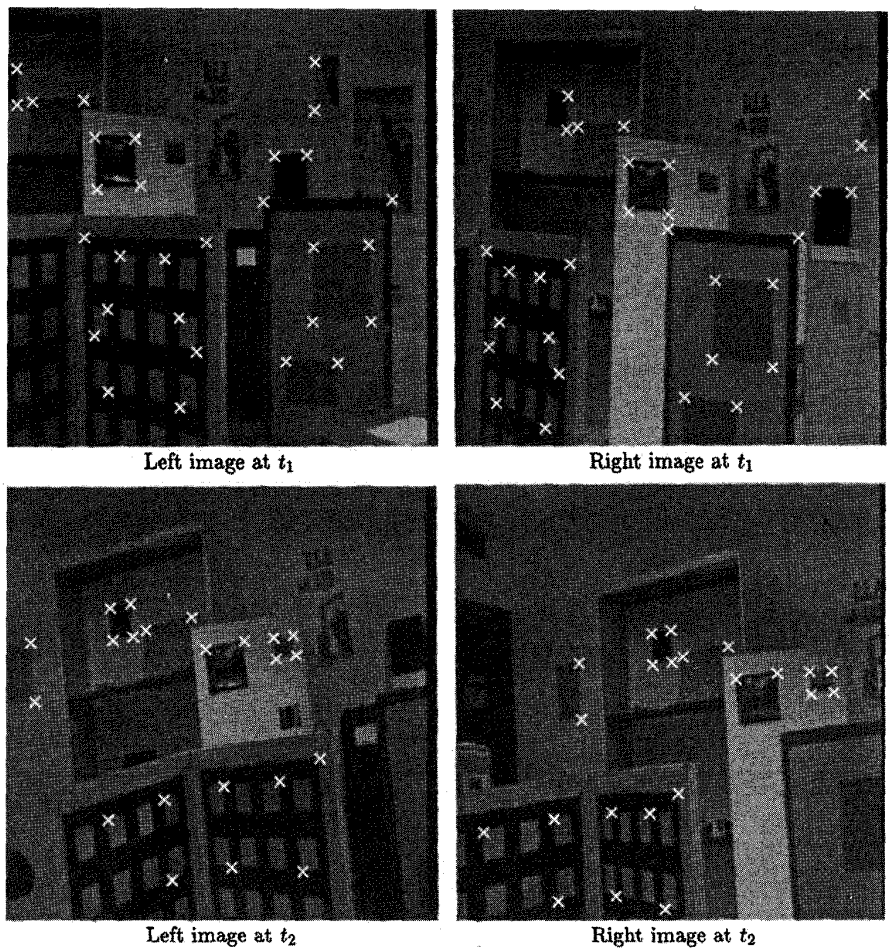
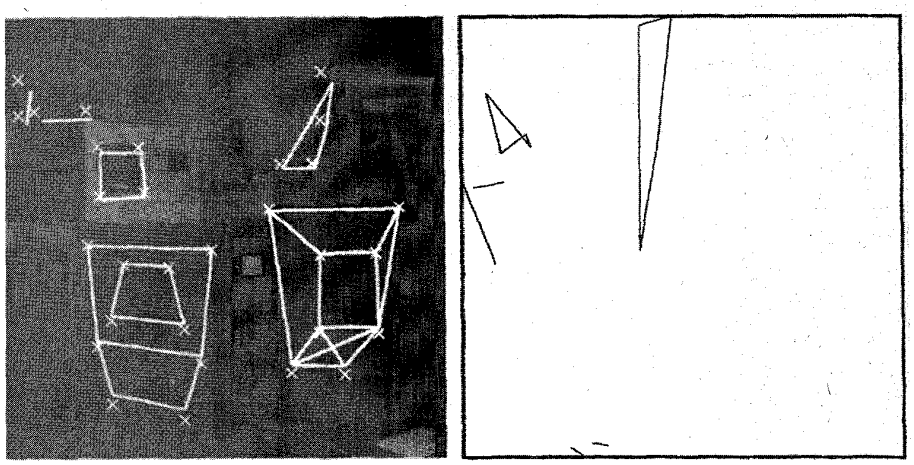


Fig. 4. Images with overlay of the points of interest used for self-calibration.

points as indicated by white crosses. The projected and original points with the technique described in this paper coincide very well as shown in Fig. 6(a), while this is not the case with the previous technique used for parameter initialization as shown in Fig. 5(a). Figs. 5(b) and 6(b) show 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 in Fig. 6(b). 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, especially with the previous technique as shown in Fig. 5(b). In Fig. 7, a stereogram of the final 3-D reconstruction is displayed. The reader can easily perceive the 3-D information by cross-eye fusion. In particular, he should see two planes in the foreground.

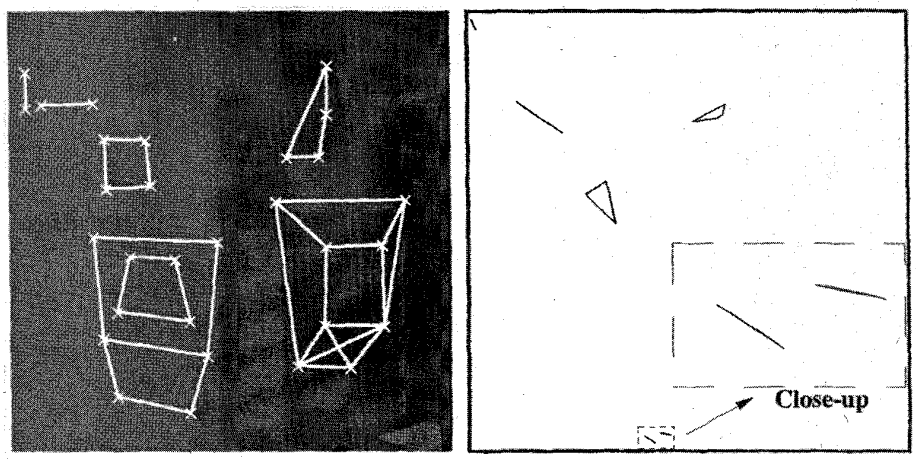
To give an idea of the quantitative performance, we consider the points on the grid pattern because we have manually measured their positions. Using an algorithm similar to that described in [31], we are able to compute the scalar factor, the rotation matrix and the translation between the points reconstructed 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 mm (the grid size is about 300 mm), 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 II, 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. For example, the number given at the first row and the third column, 1.70 mm, corresponds to the error of the reconstruction with $(\delta_{u_0}, \delta_{v_0}) = -15 \times (1, 0)$, i.e., the principal point is assumed to be at (240, 255). 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 a deviation as large



(a) (b)

Fig. 5. Initial 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).



(a) (b)

Fig. 6. 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).

as 35 pixels from the image center, the 3-D reconstruction is still reasonable.

VII. Discussions

In this section, we examine the exact number of geometric constraints which exist in the problem addressed in this paper.

A. Gain from the Correspondences Across Cameras

In this paper, we have ignored the correspondences between image 1 and image 4 and those between image 2 and image 3. In this section, we describe the benefit of using such cross correspondences[2].

Let $\mathbf{F}_{ij}$ be the fundamental matrix between image i and image j. The epipole in image i, $\mathbf{e}_{ij}$, satisfies $\mathbf{F}_{ij}\mathbf{e}_{ij} = 0$; the epipolar line in image j of a point in image i, $\mathbf{m}_i$, is given by $\mathbf{l}_{\mathbf{m}_i} = \mathbf{F}_{ij}\mathbf{m}_i$. Symmetrically, the epipole in image j, $\mathbf{e}_{ji}$, satisfies $\mathbf{F}_{ij}^T\mathbf{e}_{ji} = 0$; the epipolar line in Image i of a point in image j, $\mathbf{m}_j$, is given by $\mathbf{l}_{\mathbf{m}_j} = \mathbf{F}_{ij}^T\mathbf{m}_j$. Geometrically, the epipole $\mathbf{e}_{ij}$ is the projection in image i of the optical center of image j.

Assume we have obtained the fundamental matrices $\mathbf{F}_s$, $\mathbf{F}_l$, and $\mathbf{F}_r$. As said in Section III, each fundamental matrix

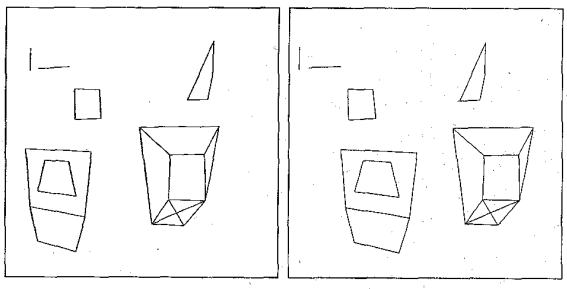


Fig. 7. Stereogram of the metric reconstruction for cross-eye fusion.

[2]This problem was raised by one of the anonymous reviewers of this article.

TABLE II
ERRORS IN RECONSTRUCTION (DISTANCES 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

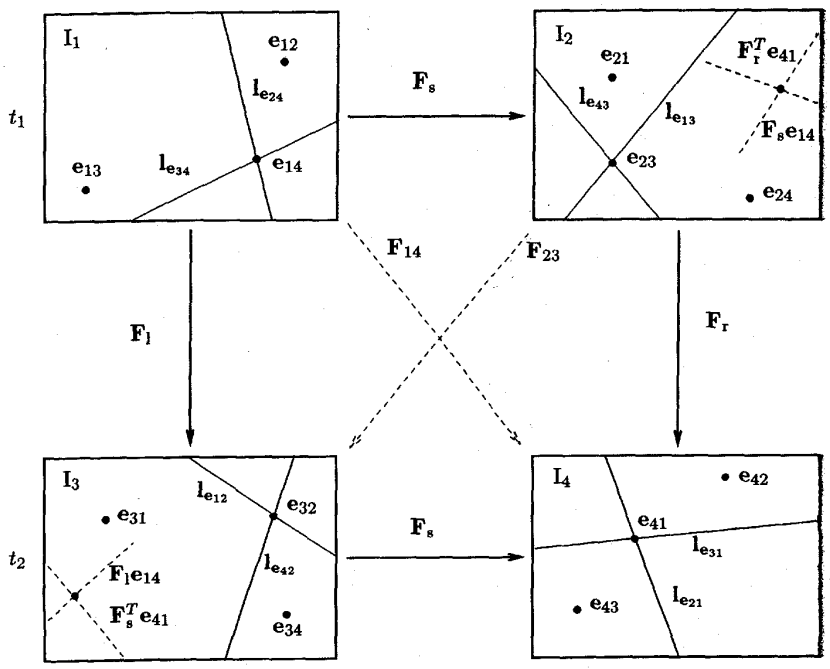


Fig. 8. Illustration of the epipolar geometry of a stereo rig in motion.

depends on seven parameters which can be interpreted geometrically as the coordinates of the two epipoles (four parameters) and the three parameters determining the homography between the two pencils of epipolar lines. Thus, the three fundamental matrices provide in total 21 constraints on the intrinsic and extrinsic parameters.

Consider now the epipolar geometry $\mathbf{F}_{14}$between images 1 and 4 (see Fig. 8). In image I_2, we know two epipoles e_{21} and e_{24}. Therefore we can build the epipolar line in I_1 of e_{24}, $l_{e_{24}} = \mathbf{F}_s^T e_{24}$, and the epipolar line in I_4 of e_{21}, $l_{e_{21}} = \mathbf{F}_r^T e_{21}$. These two epipolar lines are the intersections of the plane going through the three optical centers (the so-called *trifocal plane* [32]) with I_1 and I_4, respectively. Thus, we observe [32]:

- the two epipolar lines correspond to each other between I_1 and I_4; and
- the two epipoles e_{14} and e_{41} must lie on these two lines, respectively.

Exactly the same reasoning can be made with the triplet of cameras (I_3, I_1, I_4): We build two epipolar lines, $l_{e_{34}} = \mathbf{F}_l^T e_{34}$, and $l_{e_{31}} = \mathbf{F}_s^T e_{31}$, which correspond to each other between I_1 and I_4. Thus, we have two pairs of corresponding epipolar lines. The epipoles e_{14} and e_{41} are the intersection of $l_{e_{24}}$ and $l_{e_{34}}$, and that of $l_{e_{21}}$ and $l_{e_{31}}$, respectively. Since in addition to e_{14} and e_{41}, three correspondences of epipolar lines are required to determine completely $\mathbf{F}_{14}$. Therefore, the three fundamental matrices $\mathbf{F}_s$, $\mathbf{F}_l$ and $\mathbf{F}_r$ leave only one unknown parameter in $\mathbf{F}_{14}$.

Consider now the epipolar geometry $\mathbf{F}_{23}$between I_2 and I_3. Making exactly the same reasoning, we build two pairs of corresponding epipolar lines between I_2 and I_3 (as shown as solid lines in Fig. 8), which leaves one unknown parameter in $\mathbf{F}_{23}$. This unknown can, however, be determined as follows. In fact, we have not yet used the epipoles e_{14} and e_{41} in I_1 and I_4; they are *corresponding* points for these two images (they can be thought of as images of the point at infinity of the line going through the two optical centers of I_1 and I_4). Considering the triplet of cameras (I_1, I_2, I_4), we can construct in I_2 the point corresponding to e_{14} and e_{41} as the intersection of two epipolar lines (shown as dashed lines in Fig. 8). Similarly, considering the triplet of cameras (I_1, I_2, I_4), we can construct in I_3 the point corresponding to e_{14} and e_{41} as the intersection of two epipolar lines. These two points *correspond to each other*, and provide us with a third pair of corresponding epipolar lines (by linking them to the epipoles), which completes the determination of $\mathbf{F}_{23}$.

From the above discussion, it is clear that we can obtain, theoretically, *only one additional constraint* on the intrinsic and extrinsic parameters by using the correspondences between I_1 and I_4 and those between I_1 and I_4. In practice, we would expect to achieve a better self-calibration result because of data redundancy.

B. Estimation of All Camera Parameters

As described in Section IV, by counting the number of geometric constraints available and the number of intrinsic and extrinsic parameters to estimate, it is possible to achieve a full calibration of the stereo rig undergoing one motion. The formulation in Section IV is still valid for the full calibration, only the minimization should be carried out over a larger set of unknowns by including the location of the principal points. However, we have not been able until now to work out an algorithm to initialize the parameters from the fundamental matrices.

One possible solution could be the following: Assume the principal point is at the image center, compute the other parameters exactly as described in Section V, and eventually conduct a minimization of sum of squared distances between points and their epipolar lines, as described in Section IV, over all parameters. However, the self-calibration may be less stable because a larger set of unknowns is involved in the minimization leaving less constraint on the solution, as noticed by Luong [17] in calibrating a single camera from three views. We have not yet implemented this method.

VIII. 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.

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.

In order to apply the technique described in this paper, we must establish correspondences between two images (stereo or temporal) with unknown epipolar geometry. This is a difficult problem. We have recently developed an algorithm to solve it, and very good results have been obtained [33], [34]. The idea is to establish matches through the recovery of the unknown epipolar geometry. We first use classical techniques (correlation and relaxation methods in our particular implementation) to find an initial set of matches, and then use a robust technique—the Least Median of Squares (LMedS)—to discard false matches in this set. The epipolar geometry is then accurately estimated using a meaningful image criterion. This algorithm produces the fundamental matrix as well as the correspondences between two images. The executable code is available through anonymous ftp on `krakatoa.inria.fr` in file `pub/robotvis/BINAIRES/sun4OS4/image-matching.gz`.

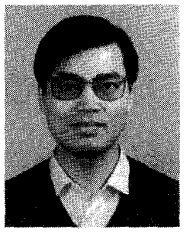
Acknowledgment

The authors would like to thank the anonymous reviewers for their insightful comments which improved this paper.

References

[1] D. Marr and T. Poggio, "A computational theory of human stereo vision," in *Proc. Royal Society London*, 1979, B. 204, pp. 301–328.
[2] J. E. W. Mayhew and J. P. Frisby, "Psychophysical and computational studies toward a theory of human stereopsis," *Artificial Intell.*, vol. 17, pp. 349–385, 1981.
[3] ——, Eds., *3D Model Recognition from Stereoscopic Cues*. Cambridge, MA: MIT Press, 1991.
[4] D. C. Brown, "Close-range camera calibration," *Photogrammetric Eng.*, vol. 37, no. 8, pp. 855–866, 1971.
[5] W. Faig, "Calibration of close range photogrammetry systems: mathematical formulation," *Photogrammetric Eng. Remote Sensing*, vol. 41, no. 12, pp. 1479–1486, 1975.
[6] R. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," in *Proc. IEEE Conf. Computer Vision and Pattern Recogition*, Miami, FL, June 1986, pp. 364–374.
[7] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, June 1986, pp. 15–20.
[8] R. Lenz and R. Tsai, "Techniques for calibrating of the scale factor and image center for high accuracy 3D machine vision metrology," in *Proc. Int. Conf. Robotics and Automation*, Raleigh, NC, 1987, pp. 68–75.
[9] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand-eye calibration," in *Proc. Int. Symp. Robotics Research*, Santa Cruz, CA, Aug. 1987, pp. 287–297.
[10] G. Toscani, *Système de calibration optique et perception du mouvement en vision artificielle*, Ph.D. dissertation, Univ. Paris XI, Orsay, Paris, France, 1987.
[11] G. Q. Wei and S. D. Ma, "Two plane camera calibration: A unified model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 1991, pp. 133–138.
[12] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distorton models and accuracy evaluation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 10, pp. 965–980, 1992.
[13] R. Y. Tsai, "Synopsis of recent progress on camera calibration for 3D machine vision," in *The Robotics Review*, O. Khatib, J. J. Craig, and T. Lozano-Perez, Eds. Cambridge, MA: MIT Press, 1989, pp. 147–159.
[14] S. J. Maybank and O. D. Faugeras, "A theory of self-calibration of a moving camera," *Int. J. Comput. Vision*, vol. 8, no. 2, pp. 123–152, Aug. 1992.
[15] R. I. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *Proc. 2nd European Conf. Computer Vision*, 1992, pp. 579–587.

[16] H. P. Trivedi, "Can multiple views make up for lack of camera registration," *Image Vision Comput.*, vol. 6, no. 1, pp. 29–32, Feb. 1988.
[17] Q.-T. Luong, *Matrice fondamentale et calibration visuelle sur l'environnement: Vers une plus grande autonomie des systèmes robotiques*, Ph.D. dissertation, Univ. Paris XI, Orsay, Paris, France, Dec. 1992.
[18] Q.-T. Luong and O. D. Faugeras, "Camera calibration, scene motion and structure recovery from point correspondences and fundamental matrices," *Int. J. Comput. Vision*, vol. 17, no. 12, Dec. 1995.
[19] Z. Zhang, "Motion and structure of four points from one motion of a stereo rig with unknown extrinsic parameters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 12, Dec. 1995. Also, in *Proc. IEEE Conf. CVPR*, June 1993, pp. 556–561.
[20] J. L. Mundy and A. Zisserman, Eds., *Geometric Invariance in Computer Vision*. Cambridge, MA: MIT Press, 1992.
[21] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, 1981.
[22] K. Kanatani and Y. Onodera, "Anatomy of camera calibration using vanishing points," *IEICE Transactions on Informations and Systems*, vol. E74, no. 10, pp. 3369–3378, Oct. 1991.
[23] C. C. Slama, Ed., *Manual of Photogrammetry*, 4th ed. Amer. Soc. Photogrammetry, 1980.
[24] Numerical Algorithms Group, *NAg, FORTRAN Library Manual*, mark 14 ed., Oxford, U.K., 1990.
[25] Q.-T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo, "On determining the fundamental matrix: Analysis of different methods and experimental results," INRIA Sophia-Antipolis, France, Rapport de Recherche 1894, 1993.
[26] T. S. Huang and O. D. Faugeras, "Some properties of the E matrix in two-view motion estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 12, pp. 1310–1312, Dec. 1989.
[27] O. D. Faugeras, F. Lustman, and G. Toscani, "Motion and structure from motion from point and line matches," in *Proc. 1st Int. Conf. Computer Vision*, London, UK, 1987, pp. 25–34.
[28] J. Weng, T. S. Huang, and N. Ahuja, "Motion and structure from two perspective views: Algorithms, error analysis and error estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 5, pp. 451–476, May 1989.
[29] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.
[30] R. Deriche and T. Blaszka, "Recovering and characterizing image features using an efficient model based approach," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, New York, NY, June 1993, pp. 530–535.
[31] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer. A*, vol. 7, pp. 629–642, Apr. 1987.
[32] O. D. Faugeras and L. Robert, "What can two images tell us about a third one?," *Int. J. Comput. Vision*, 1994, accepted for publication. Also INRIA Res. Rep. 2018, 1993.
[33] Z. Zhang, R. Deriche, Q.-T. Luong, and O. Faugeras, "A robust approach to image matching: Recovery of the epipolar geometry," in *Proc. Int. Symp. Young Investigators on Information\Computer\Control*, Beijing, China, Feb. 1994, pp. 7–28.
[34] 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 Intell. J.*, 1995, to appear. Also INRIA Res. Rep. 2273, May 1994.



Zhengyou Zhang was born in Zhejiang, China, in 1965. He received the B.S. degree in electronic engineering from the University of Zhejiang, China, in 1985, the D.E.A. diploma in computer science from the University of Nancy, France, in 1987, and the Ph.D. degree in computer science and the diploma *Habilitation à diriger des recherches* from the University of Paris XI, Orsay, France, in 1990 and 1994, respectively.

From 1987 to 1990 he was a Research Assistant in the Computer Vision and Robotics Group of the French National Institute for Research in Computer Science and Control (INRIA), France. Currently, he is a Research Scientist at INRIA. His current research interests include computer vision, mobile robotics, dynamic scene analysis and multisensor fusion. He is the author of the book (with O. Faugeras) *3D Dynamic Scene Analysis: A Stereo Based Approach* (Berlin, Heidelberg: Springer, 1992), and the forthcoming book (with S. Ma) *Computer Vision* (in Chinese) (Chinese Academy of Sciences, 1996).

Quang-Tuan Luong was born in Paris in 1964. He graduated from Ecole Polytechnique in 1987. He received the M.S. ("Magistere") degree from Ecole Normale Superieure in 1988, and the Ph.D. degree from the University of Paris XI, Orsay, France, in 1992.

Since then, he has been a Visiting Postdoctoral Researcher with the University of California at Berkeley before joining SRI International as a Computer Scientist in 1995. His research interests are in computer vision, which include color, calibration, 3-D reconstruction, projective geometry, autonomous vehicles, and virtual reality.

Olivier Faugeras (SM'95) is currently Research Director at INRIA where he leads the Computer Vision and Robotics group. His research interests include the application of Mathematics to computer vision, robotics, shape representation, computational geometry, and the architectures for artificial vision systems as well as the links between artificial and biological vision. He is an Associate Professor of Applied Mathematics at the Ecole Polytechnique in Palaiseau, France, where he teaches computer science, computer vision, and computational geometry.

Dr. Faugeras is an Associate Editor of several international scientific journals including the *International Journal of Robotics Research*, *Pattern Recognition Letters*, *Signal Processing*, *Robotics and Autonomous Systems*, and *Vision Research*. He is Editor-in-Chief of the *International Journal of Computer Vision*. He has served as Associate Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE from 1987 to 1990. In April 1989 he received the "Institut de France—Fondation Fiat" prize from the French Science Academy for his work in vision and robotics.