Finding the Collineation between Two Projective Reconstructions

Gabriella Csurka, David Demirdjian, and Radu Horaud

GRAVIR-IMAG & INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 Montbonnot Saint Martin, France
E-mail: Radu.Horaud@inrialpes.fr

The problem of finding the collineation between two 3D projective reconstructions has been proved to be useful for a variety of tasks such as calibration of a stereo rig and 3D affine and/or Euclidean reconstruction. Moreover, such a collineation may well be viewed as a point transfer method between two image pairs with applications to visually guided robot control. Despite this potential, methods for properly estimating such a projective transformation have received little attention in the past. In this paper we describe linear, nonlinear, and robust methods for estimating this transformation. We test the numerical stability of these methods with respect to image noise, to the number of matched points, and as a function of the number of outliers. Finally, we devise a specialized technique for the case where 3D Euclidean coordinates are provided for a number of control points. © 1999 Academic Press

1. INTRODUCTION AND MOTIVATION

Until very recently it was believed that visual tasks require some form of off-line camera calibration. For example, a moving calibrated camera provides a sequence of images from which 3D Euclidean structure can in principle be recovered. More recently, it has been shown both theoretically and experimentally that an image sequence taken with an uncalibrated camera can provide 3D Euclidean structure as well [11]. The basic paradigm may well be viewed as a form of self- or on-line calibration and it consists of recovering projective structure first, then upgrading it to affine structure, and then to Euclidean structure.

Another possibility that has been recently investigated by a number of researchers is to consider a sequence of image pairs gathered with a moving uncalibrated stereo rig [1, 17, 2, 7]. With such an uncalibrated image pair, matched points are reconstructed in projective space [6]. If the stereo pair undergoes rigid motion, the two projective reconstructions (before and after the motion) are related by a 3D projective transformation which is conveniently represented by a 4×4 regular matrix—a collineation. Therefore, methods that attempt to recover projective, affine, or Euclidean structure from a moving stereo pair need to estimate this collineation. Interesting enough, with the exception of [1] which briefly outlines a method for estimating this projective transformation, there appears to be no published paper in the computer vision literature describing and evaluating methods for estimating this collineation from stereoscopic data.

In this paper we address the problem of estimating such a collineation from two projective reconstructions obtained precisely, we attempt to estimate the 3D projective transformation allowing us, in some optimal sense, to overlap one projective reconstruction onto the other. Therefore, the design of an optimal criterion, as well as the choice of a technique to minimize this criterion, are crucial. Since projective entities are defined only up to a scale factor, one cannot define a metric criterion in projective space.

The first method that we describe below "throws" out the scale factors and transforms the 4×4 matrix **H** representing the projective transformation into a 16-vector h. The solution is the null space of a measurement matrix **B**, i.e., $\mathbf{B}h = \mathbf{0}$.

The second method that we describe below attempts to estimate the scale factors rather than throwing them out. This formulation leads, as in the previous case, to a linear formulation. The solution vector s is found by solving the linear problem $\mathbf{C}s = r$.

When the 3D points are projected onto images, their 2D positions can be measured in pixel coordinates. Even if the internal camera parameters are not known, it is possible to define the Euclidean distance between two image points and build an error criterion based on the sum of the squares of these distances. The estimation of 3D projective transformations from such image measurements resembles bundle adjustment techniques used widely both in photogrammetry and computer vision for camera calibration. The advantage of these methods over the linear estimators mentioned above is better behavior in the presence of badly conditioned data. Unfortunately such an estimator leads to nonlinear minimization techniques and some form of initialization must be provided.

To summarize, neither of the above methods are entirely satisfactory. In order to overcome these drawbacks, a robust estimator in necessary. We describe such a robust estimator which makes use of either one of the two linear methods in its inner loop and which provides an initial guess to the nonlinear estimator.

Next, we study the behavior of the linear and nonlinear estimators in the presence of image noise and as a function of the number of matched points. We also study the behavior of the robust estimator in the presence of outliers. In the light of these

experiments it appears that the nonlinear method performs slightly better than the linear methods. However, for levels of noise below one pixel in magnitude, the linear method performs as well as the nonlinear ones and, hence, they should be preferred because they are less time consuming than the nonlinear methods.

Finally, we devise a specialized technique in the case where 3D Euclidean coordinates are provided for a number of control points.

2. LINEAR AND NONLINEAR ESTIMATION OF 3D COLLINEATIONS

Given a pair of images denoted by x and the fundamental matrix $\mathbf{F}_x$ that describes the epipolar geometry associated with the two cameras, there exists a projective basis of the projective space $\mathcal{P}^3$ such that the projections from this space onto the images are represented by[1] two 3×4 matrices [4, 12],

$$\mathbf{P}_x = (\mathbf{I}_3 \quad \mathbf{0}_3), \quad \mathbf{P}'_x = (\mathbf{G}'_x \quad \mathbf{e}'_x), \tag{1}$$

where $\mathbf{G}'_x$ is a 3×3 regular matrix and the 3-vector $\mathbf{e}'_x$ is called the epipole and is the projection onto the second image of the optical center associated with the first camera.

Let $\mathbf{x}$, $\mathbf{x}'$ be the images of a three-dimensional point M; then from the projection equations $\rho_x \mathbf{x} = \mathbf{P}_x \mathbf{X}$ and $\rho'_x \mathbf{x}' = \mathbf{P}'_x \mathbf{X}$, we obtain the projective coordinates $\mathbf{X}$ of point M in the projective basis $\mathcal{B}_x$ defined by $(\mathbf{P}_x, \mathbf{P}'_x)$ [12]. Notice that the above projective equalities are defined up to unknown scale factors ρ_x and ρ'_x.

Consider a second pair of images y of the same scene but taken from a different viewpoint and denote by $\mathbf{F}_y$ the corresponding fundamental matrix. The new projective matrices are

$$\mathbf{P}_y = (\mathbf{I}_3 \quad \mathbf{0}_3), \quad \mathbf{P}'_y = (\mathbf{G}'_y \quad \mathbf{e}'_y),$$

and if $\mathbf{y}$, $\mathbf{y}'$ are the new projections of the point M, its 3D projective coordinates in the basis $\mathcal{B}_y$ are denoted by $\mathbf{Y}$ which verifies $\rho_y \mathbf{y} = \mathbf{P}_y \mathbf{Y}$ and $\rho'_y \mathbf{y}' = \mathbf{P}'_y \mathbf{Y}$.

Assume now, that we have a set of m points, and let $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_m$ and $\mathbf{Y}_1$, $\mathbf{Y}_2$, ..., $\mathbf{Y}_m$ be their homogeneous representations in the projective basis $\mathcal{B}_x$, respectively $\mathcal{B}_y$. Then, there exists a 4×4 collineation matrix $\mathbf{H}$ that maps the points $\mathbf{X}_i$ to the points $\mathbf{Y}_i$,

$$\mu_i \mathbf{Y}_i = \mathbf{H} \mathbf{X}_i \tag{2}$$

with μ_i arbitrary nonnull scalars.

$\mathbf{H}$, defined up to a scale factor, has 15 degrees of freedom, so to determine it, we need at least five point correspondences since each pair $(\mathbf{X}_i, \mathbf{Y}_i)$ gives us three constraints after eliminating μ_i.

Indeed, five points in the projective space generally define a projective basis of $\mathcal{P}^3$ such that the collineation between the standard projective basis and the set of points $\{\mathbf{X}_i\}_{i=1..5}$ is given by the matrix

$$\mathbf{T}_1 = (\mu_1 \mathbf{X}_1 \quad \mu_2 \mathbf{X}_2 \quad \mu_3 \mathbf{X}_3 \quad \mu_4 \mathbf{X}_4),$$

where $(\mu_1, \mu_2, \mu_3, \mu_4)^\top = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)^{-1} \mathbf{X}_5$. The collineation $\mathbf{T}_2$ which maps the standard reference points to the second set of points $\{\mathbf{Y}_i\}_{i=1..5}$ can be obtained similarly. Finally, the collineation between $\{\mathbf{X}_i\}_{i=1..5}$ and $\{\mathbf{Y}_i\}_{i=1..5}$ is then:

$$\mathbf{H} = \mathbf{T}_2 \mathbf{T}_1^{-1}.$$

This method is straightforward, but unstable in the presence of noise since we only use the minimum number of points to compute the collineation matrix $\mathbf{H}$. If we have $m > 5$ pairs of points, it is better to use all of them and estimate $\mathbf{H}$ by some minimization technique.

2.1. Linear Method 1

The classical way to estimate the entries of $\mathbf{H}$ is to eliminate the scale factors μ_i. A homogeneous linear system in the entries of $\mathbf{H}$ is thus obtained [15],

$$
\begin{aligned}
Y_i^{(4)} V_i^{(1)} - Y_i^{(1)} V_i^{(4)} &= 0 \\
Y_i^{(4)} V_i^{(2)} - Y_i^{(2)} V_i^{(4)} &= 0 \\
Y_i^{(4)} V_i^{(3)} - Y_i^{(3)} V_i^{(4)} &= 0,
\end{aligned}
\tag{3}
$$

where we used the notation $\mathbf{V}_i = (V_i^{(1)}, V_i^{(2)}, V_i^{(3)}, V_i^{(4)})^\top$ for the vector $\mathbf{H} \mathbf{X}_i$:

$$V_i^{(j)} = H_{j1} X_i^{(1)} + H_{j2} X_i^{(2)} + H_{j3} X_i^{(3)} + H_{j4} X_i^{(4)}. \tag{4}$$

In the affine or Euclidean cases we have $Y_i^{(4)} = 1$ which is often at a different magnitude than $Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}$. This explains the choice of these three equations where $Y_i^{(4)}$ has the same role. In the projective case, we cannot say the same thing, so if we want all four coordinates to play the same role, we can add the three other possible equations (which are not independent from the previous ones):

$$
\begin{aligned}
Y_i^{(2)} V_i^{(1)} - Y_i^{(1)} V_i^{(2)} &= 0 \\
Y_i^{(3)} V_i^{(1)} - Y_i^{(1)} V_i^{(3)} &= 0 \\
Y_i^{(3)} V_i^{(2)} - Y_i^{(2)} V_i^{(3)} &= 0.
\end{aligned}
\tag{5}
$$

This system can be solved when $m \geq 5$ point correspondences are available and with an additional constraint, such as $\sum H_{ij}^2 = 1$.

---

[1] We denote by $\mathbf{I}_n$ the $n \times n$ matrix of identity and by $\mathbf{0}_n$ the n-vector containing n zeros.

If we denote $\boldsymbol{h} = (H_{11}, H_{12}, \ldots, H_{44})^\top$, Eqs. (3) and (5) can be written in the form,[2]

$$\mathbf{B}_i \boldsymbol{h} = \mathbf{0},$$

where $\mathbf{B}_i$ is a 6×16 matrix:

$$
\mathbf{B}_i =
\begin{pmatrix}
Y_4^i X_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & Y_i^{(1)} X_i^\top \\
\mathbf{0}_4^\top & Y_i^{(4)} X_i^\top & \mathbf{0}_4^\top & Y_i^{(2)} X_i^\top \\
\mathbf{0}_4^\top & \mathbf{0}_4^\top & Y_i^{(4)} X_i^\top & Y_i^{(3)} X_i^\top \\
Y_i^{(2)} X_i^\top & Y_i^{(1)} X_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\
Y_i^{(3)} X_i^\top & \mathbf{0}_4^\top & Y_i^{(1)} X_i^\top & \mathbf{0}_4^\top \\
\mathbf{0}_4^\top & Y_i^{(3)} X_i^\top & Y_i^{(2)} X_i^\top & \mathbf{0}_4^\top
\end{pmatrix}.
$$

Given m correspondences (X_i, Y_i), we are looking for $\boldsymbol{h}$ that minimizes the error function:

$$\sum_{i=1}^m \|\mathbf{B}_i \boldsymbol{h}\|^2 = \boldsymbol{h}^\top \left(\sum_{i=1}^m \mathbf{B}_i^\top \mathbf{B}_i \right) \boldsymbol{h}. \tag{6}$$

With the constraint $\sum H_{ij}^2 = 1$ the error function to be minimized becomes

$$\min_{\boldsymbol{h}} (\boldsymbol{h}^\top \mathbf{B} \boldsymbol{h} + \lambda(1 - \boldsymbol{h}^\top \boldsymbol{h})). \tag{7}$$

The 16×16 matrix $\mathbf{B} = \sum_{i=1}^m \mathbf{B}_i^\top \mathbf{B}_i$ being symmetric, semi-definite, and positive, the solution for $\boldsymbol{h}$ is the eigenvector of $\mathbf{B}$ corresponding to the smallest eigenvalue of $\mathbf{B}$:

$$\mathbf{B} \boldsymbol{h} = \lambda \boldsymbol{h}.$$

Indeed, for m point correspondences the rank of $\mathbf{B}$ is equal to 15 and, therefore, the null space of $\mathbf{B}$ is a nonnull 16-vector. In practice, due to noise, the rank of $\mathbf{B}$ is equal to 16 and the null-space becomes the eigenvector associated with the smallest eigenvalue. Let $\mathbf{B} = \mathbf{Q} \mathbf{D} \mathbf{Q}^\top$ be the eigen decomposition of $\mathbf{B}$ with $\mathbf{Q}$ being an orthogonal matrix whose columns correspond to the eigenvectors and $\mathbf{D}$ being a diagnoal matrix. The singular value decomposition of $\mathbf{B}$ is $\mathbf{B}^\top \mathbf{B} = \mathbf{Q} \mathbf{D}^2 \mathbf{Q}^\top$. Therefore, in practice, the solution to the minimization problem consists of computing the singular value decomposition of $\mathbf{B}$ and taking the last column of $\mathbf{Q}$.

2.2. Linear Method 2

An alternative solution is to estimate simultaneously the entires of $\mathbf{H}$ and the scale factors $\mu_1, \ldots \mu_m$. Equation (2) can be decomposed into four distinct linear constraints and, for exam-

ple, the first of these linear constraints can be written as:

$$H_{11} X_i^{(1)} + H_{12} X_i^{(2)} + H_{13} X_i^{(3)} + H_{14} X_i^{(4)} - \mu_i Y_i^{(1)} = 0.$$

Without loss of generality one scale factor is set to 1: $\mu_m = 1$. Therefore, we have 16 unknowns for the entries of $\mathbf{H}$ and $m - 1$ unknown scale factors. The four constraints available with each one of the m equations (2) can be written as a linear system, $\mathbf{C}\boldsymbol{s} = \boldsymbol{r}$, with $\boldsymbol{s} = (H_{11}, \ldots, H_{44}, \mu_1, \ldots, \mu_{m-1})^\top$, $\boldsymbol{r} = (\underbrace{0, \ldots, 0}_{4 \times (m-1)},$

$Y_m^{(1)}, Y_m^{(2)}, Y_m^{(3)}, Y_m^{(4)})^\top$, and

$$
\mathbf{C} =
\begin{pmatrix}
\mathbf{E}_1 & -Y_1 & \mathbf{0}_4 & \cdots & \mathbf{0}_4 \\
\mathbf{E}_2 & \mathbf{0}_4 & -Y_2 & \cdots & \mathbf{0}_4 \\
\vdots & & & & \vdots \\
\mathbf{E}_{m-1} & \mathbf{0}_4 & \mathbf{0}_4 & \cdots & -Y_{m-1} \\
\mathbf{E}_m & \mathbf{0}_4 & \mathbf{0}_4 & \cdots & \mathbf{0}_4
\end{pmatrix}.
$$

The 4×16 matrices $\mathbf{E}_i$ are defined by

$$
\mathbf{E}_i =
\begin{pmatrix}
X_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\
\mathbf{0}_4^\top & X_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\
\mathbf{0}_4^\top & \mathbf{0}_4^\top & X_i^\top & \mathbf{0}_4^\top \\
\mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & X_i^\top
\end{pmatrix}.
$$

This linear system consists of $4m$ equations. Since there are $16 + (m - 1) = 15 + m$ unknowns, we must have $m \geq 5$. The linear system can be solved using the standard pseudo-inverse technique,

$$\boldsymbol{s} = (\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{C}^\top \boldsymbol{r},$$

provided that the 3D points are not coplanar.

2.3. Nonlinear Method

One way to assess the quality of the estimated collineation $\hat{\mathbf{H}}$ is to compare the projections of $\hat{Y}_i = \hat{\mathbf{H}} X_i$ and of $\hat{X}_i = \hat{\mathbf{H}}^{-1} Y_i$ with the true image points. Let $\boldsymbol{x}_i$ and $\boldsymbol{x}_i'$ be the true image points (in the left and right images) from which the 3D point X_i was reconstructed, and let $\mathbf{P}_x$ and $\mathbf{P}_x'$ be the corresponding projection matrices (see (1)). The projections of $\hat{X}_i$ by $\mathbf{P}_x$ and $\mathbf{P}_x'$ are noted by $\hat{\boldsymbol{x}}_i$ and $\hat{\boldsymbol{x}}_i'$. Similarly we define $\boldsymbol{y}_i$, $\boldsymbol{y}_i'$, $\mathbf{P}_y$, $\mathbf{P}_y'$, $\hat{\boldsymbol{y}}_i$, and $\hat{\boldsymbol{y}}_i'$:

$$\hat{\boldsymbol{x}}_i = \mathbf{P}_x \hat{\mathbf{H}}^{-1} Y_i$$
$$\hat{\boldsymbol{x}}_i' = \mathbf{P}_x' \hat{\mathbf{H}}^{-1} Y_i$$
$$\hat{\boldsymbol{y}}_i = \mathbf{P}_y \hat{\mathbf{H}} X_i$$
$$\hat{\boldsymbol{y}}_i' = \mathbf{P}_y' \hat{\mathbf{H}} X_i.$$

With the notation $\boldsymbol{y}^\top = (\bar{\boldsymbol{y}}^\top \; 1)$, let $d(\bar{\boldsymbol{y}}, \hat{\boldsymbol{y}})$ denote the Euclidean distance between the image points $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$. The quality of the

---

[2] If we use only three equations one can remove the rows corresponding to the other equations and obtain $\mathbf{B}_i$ as a 3×16 matrix.

collineation is assessed by the quadratic error function:

$$f(\hat{\mathbf{H}}, \hat{\mathbf{H}}^{-1}) = \frac{1}{4m} \sum_{i=1}^{m} (d(\bar{\mathbf{x}}_i, \hat{\bar{\mathbf{x}}}_i)^2 + d(\bar{\mathbf{x}}'_i, \hat{\bar{\mathbf{x}}}'_i)^2 + d(\bar{\mathbf{y}}_i, \hat{\bar{\mathbf{y}}}_i)^2$$
$$+ d(\bar{\mathbf{y}}'_i, \hat{\bar{\mathbf{y}}}'_i)^2). \tag{8}$$

Direct minimization of this error function over the entries of $\mathbf{H}$ results in a highly nonlinear objective function because one has to express the entries of $\mathbf{H}^{-1}$ as a function of the entries of $\mathbf{H}$. Alternatively, one can minimize the objective function

$$\min_{\mathbf{H}, \mathbf{H}'} (f(\mathbf{H}, \mathbf{H}') + \eta \|\mathbf{H}\mathbf{H}' - \mathbf{I}_4\|^2), \tag{9}$$

where the term $\|\mathbf{H}\mathbf{H}' - \mathbf{I}_4\|^2$ is known as a penalty function. High numerical values for η enforce this constraint and, hence, guarantee that $\mathbf{H}' = \mathbf{H}^{-1}$.

Therefore, the minimization process considers $\mathbf{H}$ and $\mathbf{H}'$ as two independent sets of variables.

An alternative method is to use a less symmetric objective function and to minimize only the distances between $\hat{\bar{\mathbf{y}}}_i, \hat{\bar{\mathbf{y}}}'_i$ and the true image measurements $\bar{\mathbf{y}}_i, \bar{\mathbf{y}}'_i$ in the second pair of images. In this case we need no additional penalty term and the objective function becomes

$$\min_{\mathbf{H}} f_1(\mathbf{H}) = \min_{\mathbf{H}} \frac{1}{2m} \sum_{i=1}^{m} r_i \tag{10}$$

with the residual

$$r_i = d(\bar{\mathbf{y}}_i, \hat{\bar{\mathbf{y}}}_i)^2 + d(\bar{\mathbf{y}}'_i, \hat{\bar{\mathbf{y}}}'_i)^2. \tag{11}$$

3. EXPERIMENTAL COMPARISON USING SIMULATED DATA

We implemented the methods described above to estimate $\mathbf{H}$. Prior to the estimation of $\mathbf{H}$ projective reconstruction must be performed such that the coordinates of $\mathbf{X}_i$ are estimated from the image points $\mathbf{x}_i$ and $\mathbf{x}'_i$ and the coorrdinates of $\mathbf{Y}_i$ are estimated from the image points $\mathbf{y}_i$ and $\mathbf{y}'_i$. This is done using the triangulation method described by Hartley and Sturm [5]. The latter minimizes reconstruction errors by constraining the left-image and right-image points to lie on conjugated epipolar lines. Certainly, this method reduces the effect of image noise and biases the comparison described below. Nevertheless, the overall goal is to obtain the best available results and triangulation is beneficial with respect to such a goal.

We carried out a large number of experiments in order to compare the quality of the results. The algorithms have been tested with a set of 41 points.

The first experiment uses all the points and analyzes the sensitivity with respect to image noise. Gaussian noise with varying standard deviation (from 0 to 2 pixels) is added to the image point locations. The second experiment analyzes the sensitivity

as a function of the total number of points being used by the minimization processes, the minimum number of points being necessarily equal to 5.

With all the minimization methods and in both experiments, the same error function, i.e., Eq. (8), was used to assess the quality of the results.

Figure 1 shows the results of the first experiment and displays the median error over 100 trials of each algorithm and with various levels of noise. The following notations are used:

L1—linear method 1;

L2—linear method 2;

N1—nonlinear method, minimizing the back-projections in the two image pairs, and

N2—nonlinear method, minimizing the back-projections in the second image pair.

Note that the quality of the collineation linearly degrades with increasing image noise. The nonlinear methods (N1 and N2) sightly outperform the linear methods which, in turn, have an identical behavior.

Figure 2 shows the results of the second experiment, where three levels of noise were used in conjunction with the nonlinear method N2. This experiment reveals that in the presence of noise 15 points at least are necessary to properly estimate the transformation.

4. ROBUST ESTIMATION OF 3D COLLINEATIONS

A common source of errors in computer vision is image-to-image point matching. State-of-the-art algorithms consist of using robust estimators to determine the epipolar geometry [16]. Such an algorithm finds point matches that are geometrically correct; i.e., they lie on conjugated epipolar lines [9], but they may be physically incorrect: the point is correctly reconstructed from a mathematical point of view but it does not correspond to an actual feature.

If two image pairs are considered, as is the case in this paper, mismatches can result even from correctly matched points, based on the epipolar geometry associated with each one of the two image pairs. Such a mismatch will give a large error and the minimization methods described in the previous section cannot deal with such mismatches.

This problem is illustrated on Fig. 3. Let $\mathbf{x}, \mathbf{x}', \mathbf{y}$, and $\mathbf{y}'$ be a quadruple of matched points in between the two image pairs. Among these points, $\mathbf{x}'$ is a bad match because it results from the projection of 3D point N. Nevertheless, $\mathbf{x}$ and $\mathbf{x}'$ correctly satisfy the epipolar constraint because N lies in the same epipolar plane as M. The projective coordinates of M, when reconstructed from $\mathbf{y}$ and $\mathbf{y}'$ are denoted by the 4-vector $\mathbf{Y}$. This point projects onto the first image pair in $\mathbf{x}$ and $\mathbf{x}''$. The error between $\mathbf{x}'$ and $\mathbf{x}''$ is large and will have a dramatic effect on the nonlinear minimization process described in the previous section. Similarly, the linear methods will be affected as well because the 3D point M is wrongly matched with the 3D point Q.

Backprojected errors
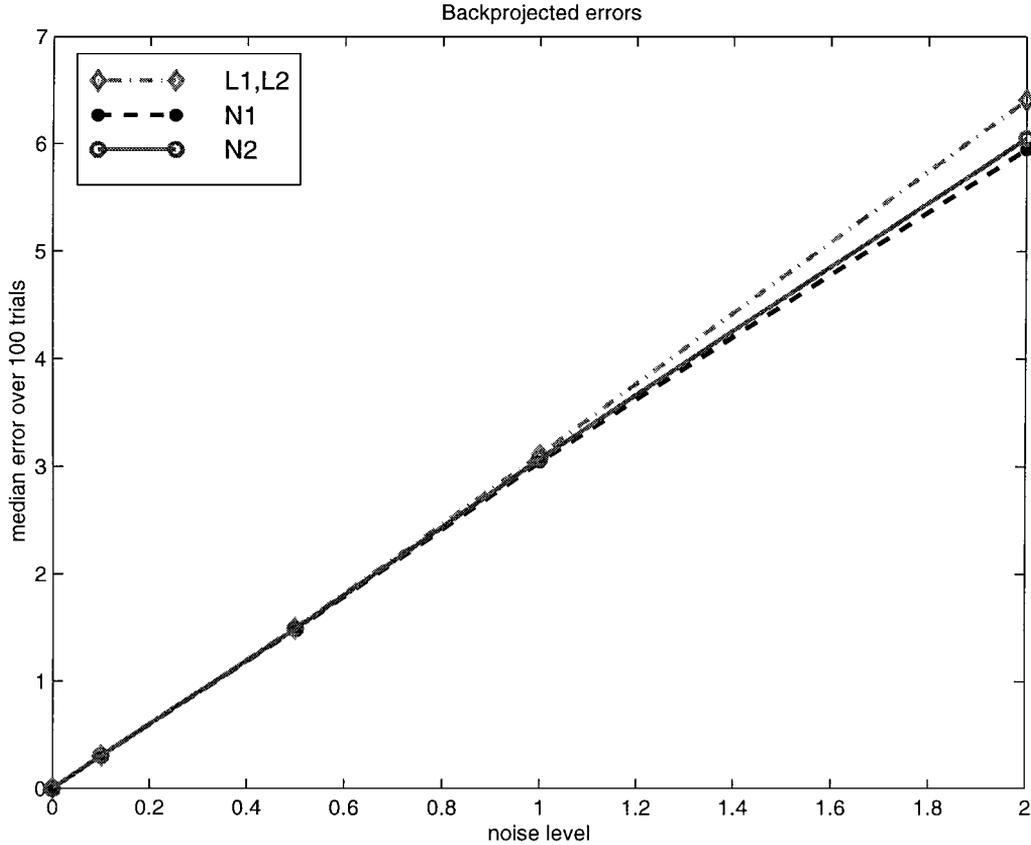
FIG. 1. Comparison of two linear methods and two nonlinear methods for computing the projective collineation in the presence of image Gaussian noise.

Commonly used robust methods include M-estimators, least-median-squares (LMedS), and random sample consensus [3]. We immediately ruled out M-estimators because in our case this method can tolerate only up to 7% outliers. The advantage of RANSAC over LMedS resides in the outlier rejection strategy. RANSAC requires a user-defined threshold while LMedS does not. Such a threshold is useful in practice for a number of reasons:

1. efficiency in the presence of nonhomogeneous noise;
2. Fifty percent outliers and above are allowed, and
3. increased time-efficiency because the inner loop of the algorithm can be quit as soon as a satisfactory solution is found.

When applied to the problem of estimating a 3D projective transformation, the robust method that we implemented can be summarized as:

1. Among the set of m matches consider N samples of five matches;
2. For each sample k, $1 \leq k \leq N$,
    2.1. Estimate a homography $\mathbf{H}_k$ from these five matches;
    2.2. Compute the total number of matches m_k consistent with $\mathbf{H}_k$, i.e., matches whose associated error is below a threshold, the error being defined by Eq. (11);
3. Select the homography $\mathbf{H}_k$ which is associated with the largest number of consistent matches;

4. Use all of these m_k to refine the estimation of the best available homography $\mathbf{H}_k$, and
5. Use the latter estimation to update the list of inliers and outliers.

In order to determine the number of samples N we must set the probability of selecting a good solution. The formula for this probability is

$$\gamma = 1 - [1 - (1 - \varepsilon_{\text{out}})^p]^N,$$

where p is the number of points in the sample ($p = 5$ for 3D collineations) and ε_{out} is the percentage of outliers that are to be found in the initial data.

For example, for $\gamma = 0.95$ and $\varepsilon_{\text{out}} = 0.5$ we obtain $N = 100$. Hence, the inner loop of the robust estimator algorithm has to be run at least 100 times. The inner loop main step (2.2) consists essentially in estimating a collineation $\mathbf{H}$ from five point matches. With five matches the solution is found in closed-form using either one of the two linear methods described above.

4.1. Experimenting with the Robust Estimator

In order to assess the performance of the robust estimator described above, we tested it in the presence of simulated data. For this purpose we used two sets of points: one set of 200 points
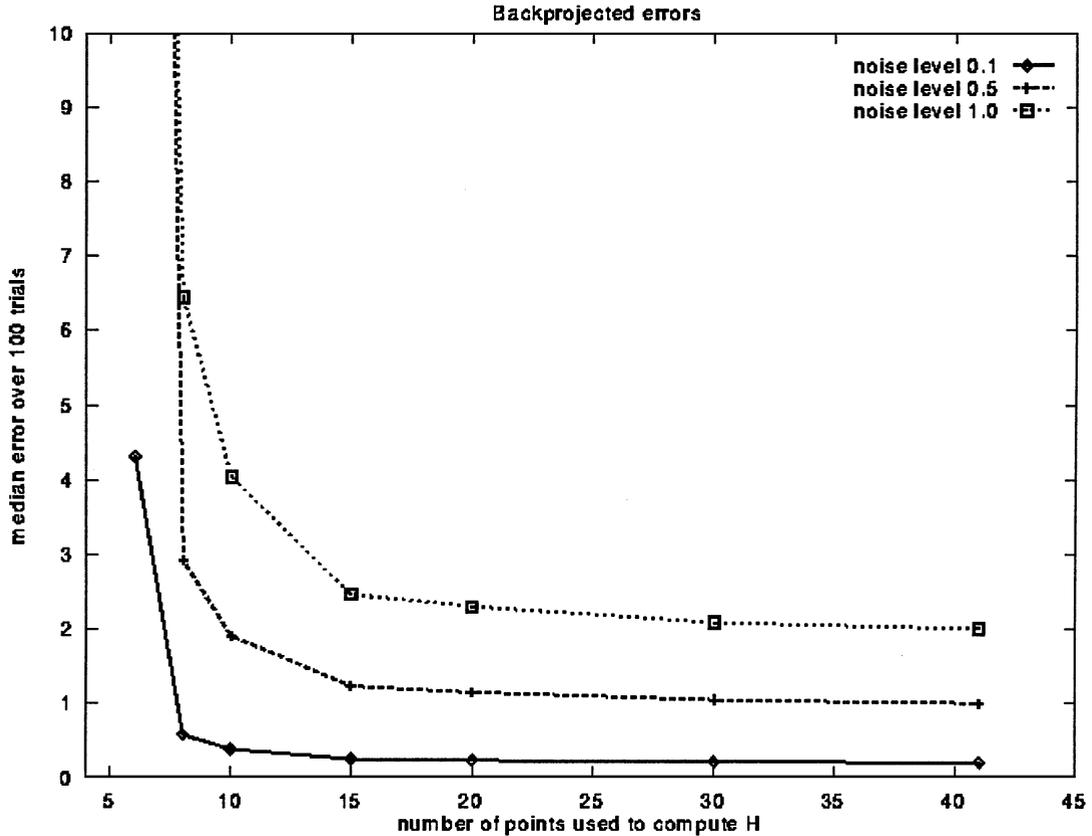
FIG. 2. The sensitivity of the nonlinear method N2 as a function of the number of points and for three levels of Gaussian noise with standard deviations 0.1, 0.5, and 1.

and another set of 200 points. The first set was perturbed with Gaussian noise with standard deviation equal to 0.5 pixels. The
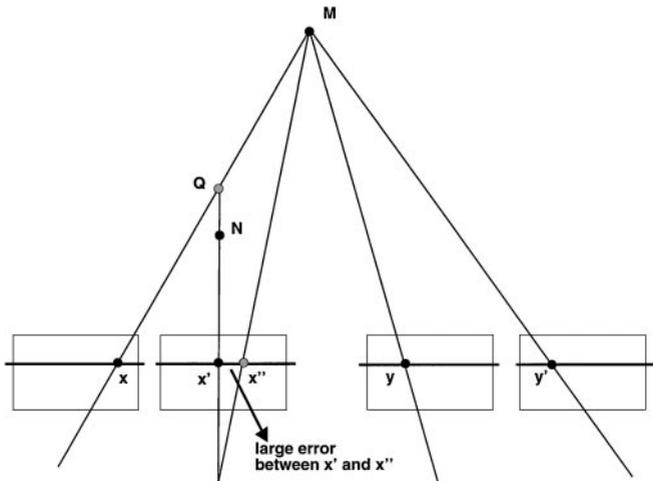

FIG. 3. Quadruples of mismatched image points can result even if the image points satisfy the epipolar geometry between the two image pairs. Points marked with a black dot are real points while points marked with a grey dot are virtual points predicted by geometrical computations.

second set was perturbed with white noise whose amplitude varies between 3 pixels and 20 pixels.

Furthermore, the algorithm's parameters were set for an expected number of outliers of 50% but the number of outliers varied between 0 and 75%. Figure 4 shows the result of running the algorithm. The average back-projected error per image point (the error measure is the one given by Eq. (8)) is shown as a function of the number of outliers present in the data.

5. DEALING WITH EUCLIDEAN CONTROL POINTS

Given a projective reconstruction of a set of 3D points, an important task is to upgrade it to Euclidean structure. In this case one may define the collineation $\mathbf{H}_u$ which maps or upgrades projective structure to metric structure. A number of authors studied the structure of $\mathbf{H}_u$ under various camera motions such as general motion of a stereo rig [7] and rotations of a single camera [4]. Here we take a different approach and assume that the Cartesian coordinates of a set of landmark points or control points are provided. A similar approach was taken by Mohr, Boufama, and Brand [10], but they were unable to express the solution in close form.

It is assumed that the Euclidean coordinates $\bar{Y}_i = (Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)})^\top$ of at least five points in general position are known in
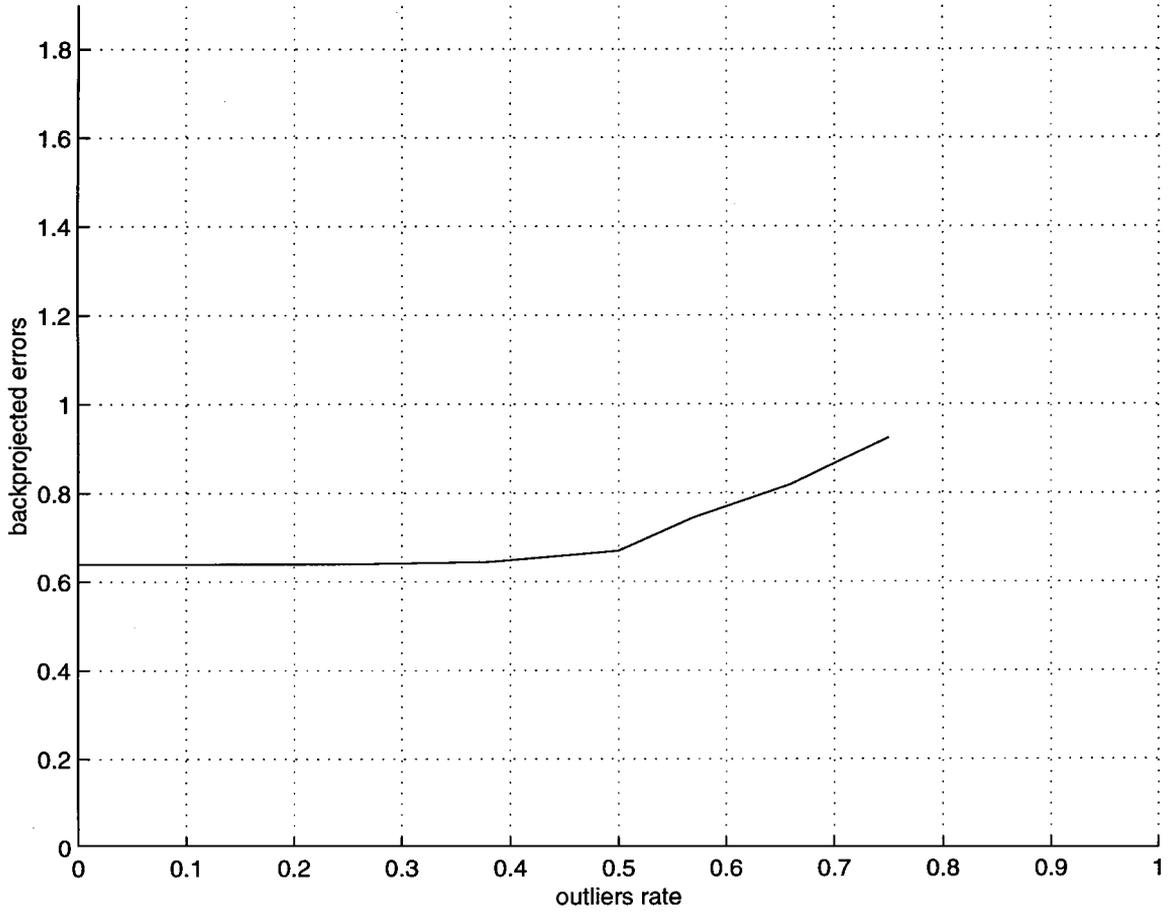
FIG. 4. This curve shows the back-projected image error as a function of the number of outliers. With 50% outliers and less, the back-projected error is almost constant and slightly higher than the initial image noise.

some Euclidean basis $\mathcal{B}_E$. Let, as before, X_i be the 3D projective coordinates of a point indexed by i. Since Euclidean space is a subspace of the projective space, one can compute the collineation $\mathbf{H}_u$ that maps X_i onto $Y_i = (\bar{Y}_i^\top \; 1)^\top$ (the homogeneous coordinates associated with $\bar{Y}_i$), using any of the linear or nonlinear techniques described in the previous sections. Here, however, we would like to take full advantage of the known Euclidean structure associated with the control points.

Consider the 4D vector space corresponding to the projective subspace generated by the Euclidean basis $\mathcal{B}_E$ [14]. In this subspace the hyper-plane $X^{(4)} = 1$ corresponds to the 3D Euclidean space. One way to estimate $\mathbf{H}_u$ is to minimize the sum of squares of the Euclidean distances between the 3-vectors $\bar{Y}_i$ and $\bar{V}_i$ with

$$\bar{V}_i = \left(\frac{(\mathbf{H}_u X_i)^{(1)}}{(\mathbf{H}_u X_i)^{(4)}}, \frac{(\mathbf{H}_u X_i)^{(2)}}{(\mathbf{H}_u X_i)^{(4)}}, \frac{(\mathbf{H}_u X_i)^{(3)}}{(\mathbf{H}_u X_i)^{(4)}} \right)^\top.$$

Hence, the error function to be minimized is

$$\min_{\mathbf{H}_u} \sum_{i=1}^{n} \varepsilon_i^2 \qquad (12)$$

with

$$\varepsilon_i^2 = \left(\frac{(\mathbf{H}_u X_i)^{(1)}}{(\mathbf{H}_u X_i)^{(4)}} - Y_i^{(1)} \right)^2 + \left(\frac{(\mathbf{H}_u X_i)^{(2)}}{(\mathbf{H}_u X_i)^{(4)}} - Y_i^{(2)} \right)^2 + \left(\frac{(\mathbf{H}_u X_i)^{(3)}}{(\mathbf{H}_u X_i)^{(4)}} - Y_i^{(3)} \right)^2$$

which is a nonlinear function in the entries of $\mathbf{H}_u$. To minimize this function, nonlinear optimization methods are necessary.

Alternatively, it is possible to define a metric in the 4D vector space associated with the Euclidean space when the latter is a subspace of the projective space. One way to define the Euclidean distance between Y_i (a control point) and $\mathbf{H}_u X_i$ (a transformed projective point) is to arbitrarily fix the scale factor associated with the homogeneous vector $\mathbf{H}_u X_i$ and to compute the distance d_i between this vector and its projection onto the direction of Y_i (see Fig. 5).

The projection of $\mathbf{H}_u X_i$ onto Y_i is given by their dot-product, normalized by the length of Y_i:

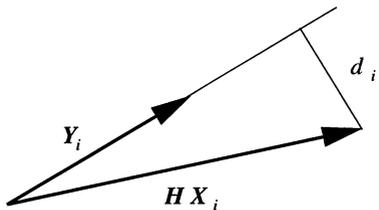$$p_i = \frac{\left(Y_i^\top \mathbf{H}_u X_i \right)^2}{Y_i^\top Y_i}.$$

FIG. 5. The Euclidean distance between vectors Y_i and $\mathbf{H}_u X_i$ in 4D vector space can be approximated by d_i.

The distance between the tip of $\mathbf{H}_u X_i$ and the tip of its projection is equal to

$$
\begin{aligned}
d_i^2 &= (\mathbf{H}_u X_i)^\top (\mathbf{H}_u X_i) - \frac{\left(Y_i^\top \mathbf{H}_u X_i\right)^2}{Y_i^\top Y_i} \\
&= X_i^\top \mathbf{H}_u^\top \mathbf{I}_4 \mathbf{H}_u X_i - \left(\frac{1}{Y_i^\top Y_i}\right) X_i^\top \mathbf{H}_u^\top Y_i Y_i^\top \mathbf{H}_u X_i \\
&= X_i^\top \mathbf{H}_u^\top \underbrace{\left(\mathbf{I}_4 - \frac{Y_i Y_i^\top}{Y_i^\top Y_i}\right)}_{\mathbf{A}_i} \mathbf{H}_u X_i \\
&= X_i^\top \mathbf{H}_u^\top \mathbf{A}_i \mathbf{H}_u X_i.
\end{aligned}
$$

Notice that matrix $\mathbf{A}_i$ just defined is a symmetric semi-definite positive matrix and its eigenvalues are equal to $(1, 1, 1, 0)$. With the notations already used in Sections 2.1 and 2.2 we have

$$
\mathbf{H}_u X_i = \begin{pmatrix} X_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & X_i^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \mathbf{0}_4^\top & X_i^\top & \mathbf{0}_4^\top \\ \mathbf{0}_4^\top & \mathbf{0}_4^\top & \mathbf{0}_4^\top & X_i^\top \end{pmatrix} \begin{pmatrix} H_{11} \\ \vdots \\ H_{44} \end{pmatrix} = \mathbf{E}_i h.
$$

The distance d_i becomes

$$
d_i^2 = X_i^\top \mathbf{H}_u^\top \mathbf{A}_i \mathbf{H}_u X_i = h^\top \mathbf{E}_i^\top \mathbf{A}_i \mathbf{E}_i h.
$$

By summation over all control points we obtain the error function

$$
\sum_i d_i^2 = h^\top \sum_i \left(\mathbf{E}_i^\top \mathbf{A}_i \mathbf{E}_i\right) h = h^\top \mathbf{A} h, \tag{13}
$$

where $\mathbf{A} = \sum_i (\mathbf{E}_i^\top \mathbf{A}_i \mathbf{E}_i)$ is symmetric, semi-definite, and positive. The optimization problem is therefore identical to the one expressed by Eq. (7) and the solution for h is the eigenvector associated with the smallest eigenvalue of $\mathbf{A}$.

In order to illustrate this method consider the pair of images of Fig. 6. Three-dimensional Euclidean coordinates of 26 data points are also provided with these images. Point correspondences were obtained interactively, and the fundamental matrix was estimated using the method described in [16]. Three-dimensional projective coordinates are obtained by triangulation [5]. The collineation between projective and Euclidean coordinates was estimated using the method described above and then used to upgrade the entire set of projective coordinates. This

FIG. 6. A pair of real images (left) and the Euclidean reconstruction of the scene (right).

resulted in a Euclidean map of some of the roofs which are displayed on Fig. 6.

6. CONCLUSION

In this paper we described various methods to estimate the collineation between two projective reconstructions. We described two linear methods which minimize an algebraic distance in 3D projective space, two nonlinear methods which minimize a metric distance in 2D image space, and a robust method which combines any of the linear methods with an outlier rejection strategy. Moreover, for the special case where Euclidean coordinates of a few control points are available, we devised a linear method based on a metric distance in the 4D vector space associated with homogeneous coordinates.

In order to evaluate these methods and assess their respective merits, we carried out a large number of experiments with both synthetic and real data. The synthetic data allowed us to add Gaussian noise to image data and to study the numerical stability of the various implemented methods. In the light of these experiments it appears that all the methods have an almost identical behavior and that the nonlinear methods outperform the linear ones. However, for levels of noise below one pixel in magnitude it is not possible to distinguish between the performances of the four methods. Another set of experiments allowed us to test the performance as a function of the number of points being considered. The conclusion of this set of experiments is, not surprisingly, that (i) the more point matches are available the better it is and (ii) below 15 to 20 point matches one should expect unreliable results.

For the special case where projective coordinates of a few control points are matched against Euclidean coordinates, we have been able to devise a linear method which is based on a metric distance in 4D space—the vector space associated with homogeneous coordinates of the 3D projective space and its Euclidean subspace. Such a linear error function could probably be extended to other estimation problems, such as camera calibration and bundle adjustment.

One interesting application associated with the computation of such a 3D projective transformation is the transfer of 2D points from one stereo image pair to another stereo image pair. This transfer technique is useful whenever a set of 3D points is observed with one stereo pair and one wants to display the same 3D set as would be observed by another stereo pair. This stereo transfer technique has been applied to a visual grasping method developed elsewhere [8]. Another application associated with the computation of a 3D collineation is the self-calibration of a stereo camera pair [7, 13].

Another promising outcome of this work that we plan to investigate is the use of projective motion estimation to perform motion segmentation. Indeed, in this paper we considered a moving stereo sensor observing a rigid scene. If the scene is not rigid, the scene points undergoing some motion will be detected as outliers. The outliers can be further analyzed in order to detect possible moving objects. It will therefore be possible to distinguish between ego-motion and moving scene objects.

REFERENCES

1. P. A. Beardsley, I. D. Reid, A. Zisserman, and D. W. Murray, Active visual navigation using non-metric structure, in *Proceedings, Fifth International Conference on Computer Vision, Cambridge, MA, June 1995*, pp. 58–64, IEEE Comput. Soc., Los Alamitos, CA.

2. F. Devernay and O. Faugeras, From projective to Euclidean reconstruction, in *Proceedings, Computer Vision and Pattern Recognition Conference, San Francisco, CA, June 1996*, pp. 264–269.

3. M. A. Fischler and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. ACM* **24**(6), 1981, 381–395.

4. R. I. Hartley, Self-calibration of stationary cameras, *Internat. J. Comput. Vision* **22**(1), 1997, 5–23.

5. R. I. Hartley and P. F. Sturm, Triangulation, *Comput. Vision Image Understanding* **68**(2), 1997, 146–157.

6. R. I. Hartley, R. Gupta, and T. Chang, Stereo from uncalibrated cameras, in *Proceedings, Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, 1992*, pp. 761–764.

7. R. Horaud and G. Csurka, Self-calibration and Euclidean reconstruction using motions of a stereo rig, in *Proceedings, Sixth International Conference on Computer Vision, Bombay, India, January 1998*, pp. 96–103, IEEE Comput. Soc., Los Alamitos, CA.

8. R. Horaud, F. Dornaika, and B. Espiau, Visually guided object grasping, *IEEE Trans. Rob. Automat.* **14**(4), 1998, 525–532.

9. Q-T. Luong and O. D. Faugeras, The fundamental matrix: Theory, algorithms, and stability analysis, *Internat. J. Comput. Vision* **17**(1), 1996, 43–75.

10. R. Mohr, B. Boufama, and P. Brand, Understanding positioning from multiple images, *Artificial Intelligence* **78**(1-2), 1995, 213–238.

11. M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, in *Proceedings of the 6th International Conference on Computer Vision, Bombay, India, January 1998*, pp. 90–95.

12. C. Rothwell, O. Faugeras, and G. Csurka, A comparison of projective reconstruction methods for pairs of views, *Comput. Vision Image Understanding* **68**(1), 1997, 36–58.

13. A. Ruf, G. Csurka, and R. Horaud, Projective translations and affine stereo calibration, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, June 1998*, pp. 475–481, IEEE Comput. Soc., Los Alamitos, CA.

14. J. G. Semple and G. T. Kneebone, *Algebraic Projective Geometry*, Oxford Science, New York, 1952.

15. Z. Zhang, A stereovision system for a planetary rover: Calibration, correlation, registration and fusion, in *Proceedings of the IEEE Workshop on Planetary Rover Technology and Systems, Minneapolis, Minnesota, April 1996*.

16. Z. Zhang, R. Deriche, O. D. Faugeras, and Q-T. Luong, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence* **78**(1-2), 1995, 87–119.

17. A. Zisserman, P. A. Beardsley, and I. D. Reid, Metric calibration of a stereo rig, in *Workshop on Representation of Visual Scenes, Cambridge, MA, June 1995*, pp. 93–100.