On the Optimization Criteria Used in Two-View Motion Analysis

Zhengyou Zhang, *Senior Member, IEEE*

Abstract—The three best-known criteria in two-view motion analysis are based, respectively, on the distances between points and their corresponding epipolar lines, on the gradient-weighted epipolar errors, and on the distances between points and the reprojections of their reconstructed points. The last one has a better statistical interpretation, but is, however, significantly slower than the first two. In this paper, I show that, given a reasonable initial guess of the epipolar geometry, the last two criteria are equivalent when the epipoles are at infinity, and differ from each other only a little even when the epipoles are in the image, as shown experimentally. The first two criteria are equivalent only when the epipoles are at infinity and when the observed object/scene has the same scale in the two images. This suggests that the second criterion is sufficient in practice because of its computational efficiency. Experiments with several thousand computer simulations and four sets of real data confirm the analysis. The result is valid for both calibrated and uncalibrated images.

Index Terms—Motion analysis, multiple-view geometry, 3D reconstruction, optimization criteria, algorithmic comparison, structure from motion, uncalibrated images.

1 Introduction

Two-view motion analysis has been an active area of research in computer vision since the late seventies, and a number of methods have been proposed. The focus was on the motion estimation between two images with known intrinsic parameters (we say they are calibrated). The reader is referred to [1], [2] for a review. Recently, analysis of uncalibrated images has attracted quite a number of researchers since the work of Faugeras [3] and Hartley [4]. There are several reasons: the calibration is fastidious and not very stable; it is impossible in many applications such as video sequence analysis; etc. The reader is referred to [5] for a review.

The two domains (calibrated or uncalibrated) share the same mathematical basis. If lens distortion can be ignored (see [6], [7] if distortion is considered), two images are related by a 3×3 matrix, which is called the *essential matrix* [8], [9] when image are calibrated and normalized image coordinates are used, and is called the *fundamental matrix* [10], [11], [12] when pixel image coordinates are used. The corresponding points between the two images must satisfy the epipolar constraint to be presented below. An important step in motion analysis, after the establishment of point correspondences, is to estimate the 3×3 matrix. The crucial part is the choice of an appropriate criterion for optimization. The three best-known criteria are based on the distances between points and their corresponding epipolar lines (denoted by J_1), the gradient-weighted epipolar errors (denoted by J_2), and the distance between the observation and the reprojection of the reconstructed structure (denoted by J_3). Table 1 lists a few references which use these criteria for determining *Euclidean* motion (when camera intrinsic parameters are known), *affine* epipolar geometry (when affine projection model is assumed), and *projective* epipolar geometry (i.e., fundamental matrix, when uncalibrated full perspective projection model is assumed). The optimization of these criteria is usually performed through an iterative numerical minimization procedure, which means that a reasonable initial guess of the epipolar geometry is required. Hartley's normalized eight-point algorithm [13] is recommended to obtain such an initial guess.

In this paper, we study the relationship of these three criteria. Data points are assumed to be corrupted by independent and identically distributed Gaussian noise. False matches are assumed to be already detected and discarded (see [5] on this topic). Uncalibrated images are considered, but the results are equally valid for calibrated images because only the general matrix form of the epipolar geometry is used in the analysis. Analytical analysis is carried out, which shows that, given a reasonable initial guess of the epipolar geometry, criteria J_2 and J_3 are equivalent when the epipoles are at infinity, and differ from each other only a little even when the epipoles are in the image, and that J_1 and J_2 are equivalent only when the epipoles are at infinity and when the observed object/scene has the same scale in the two images. The bias of J_1, although it is very small, is studied, which tends to make the object scales and the off-

TABLE 1
Sample References Which Use the Three Criteria Considered in This Paper (See Text)

criterion	Euclidean	Affine	Projective
J_1	[14]	[15]	[12]
J_2	[16]	[15]	[12]
J_3	[17]	[18]	[19]

• *The author is with Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA. E-mail: zhang@microsoft.com.*

Manuscript received 4 Aug. 1997; revised 27 Apr. 1998. Recommended for acceptance by D.J. Kriegman.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106854.

sets of the epipoles with respect to data points in both images similar. Experiments with several thousand computer simulations and four sets of real data confirm the analysis. In the appendix, the three criteria are reformulated to cases where the noise property of *each* point is known. The excellent approximation of J_3 by J_2 was also noted experimentally by Oliensis [20, Section 3.4.5].

2 Notation and Problem Statement

2.1 Notation

Let $\mathbf{x} = [x, y]^T$, we define $\tilde{\mathbf{x}} = [x, y, 1]^T$ and $\breve{\mathbf{x}} = [x, y, 0]^T$. Please note the difference: One is added as the last element in the former while zero is added in the latter. Furthermore, we define matrix

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{1}$$

We then have $\breve{\mathbf{x}} = \mathbf{Z}\mathbf{x}, \mathbf{x} = \mathbf{Z}^T\breve{\mathbf{x}}, \mathbf{Z}\mathbf{Z}^T = \mathrm{diag}(1, 1, 0)$, and $\mathbf{Z}^T\mathbf{Z} = \mathrm{diag}(1, 1)$.

A camera is described by a 3×4 projection matrix $\mathbb{P}$. The coordinates of a 3D point M = $[x, y, z]^T$ in a world coordinate system and its retinal image coordinates $\mathbf{m} = [u, v]^T$ are related by

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbb{P}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad or \quad s\tilde{\mathbf{m}} = \mathbb{P}\tilde{\mathrm{M}},$$

where s is an arbitrary scale factor. Note that the projection can be full perspective or affine (including orthographic and weak perspective).

The quantities related to the second camera are indicated by ′. For example, if $\mathbf{m}_i$ is a point in the first image, $\mathbf{m}'_i$ denotes its corresponding point in the second image.

A line $\mathbf{l}$ in the image passing through point $\mathbf{m} = [u, v]^T$ is described by equation $au + bv + c = 0$. Let $\mathbf{l} = [a, b, c]^T$, then the equation can be rewritten as $\mathbf{l}^T\tilde{\mathbf{m}} = 0$ or $\tilde{\mathbf{m}}^T\mathbf{l} = 0$. Thus, a 2D line is represented by a homogeneous 3D vector. Multiplying $\mathbf{l}$ by any nonzero scalar defines the same line. The distance from point $\mathbf{m}_0 = [u_0, v_0]^T$ to line $\mathbf{l} = [a, b, c]^T$ is given by

$$d(\mathbf{m}_0, \mathbf{l}) = \frac{au_0 + bv_0 + c}{\sqrt{a^2 + b^2}}.$$

Note that we here use the *signed* distance.

2.2 Epipolar Geometry

The epipolar geometry exists between any two camera systems. Consider the case of two cameras as shown in Fig. 1. Let C and C′ be the optical centers of the first and second cameras, respectively. Given a point $\mathbf{m}$ in the first image, its corresponding point in the second image is constrained to lie on a line called the *epipolar line* of $\mathbf{m}$, denoted by $\mathbf{l}'_{\mathbf{m}}$. The line $\mathbf{l}'_{\mathbf{m}}$ is the intersection of the plane Π, defined by $\mathbf{m}$, C and C' (known as the *epipolar plane*), with the second image plane $\mathcal{I}'$. This is because image point $\mathbf{m}$ may correspond to an arbitrary point on the semiline CM ($\mathbf{M}$ may be at infinity) and that the projection of CM on $\mathcal{I}'$ is the line $\mathbf{l}'_{\mathbf{m}}$. This is called the *epipolar constraint*. Algebraically, in order for $\mathbf{m}$ in the first image and $\mathbf{m}'$ in the second image to be matched, the following equation must be satisfied [21], [22]:

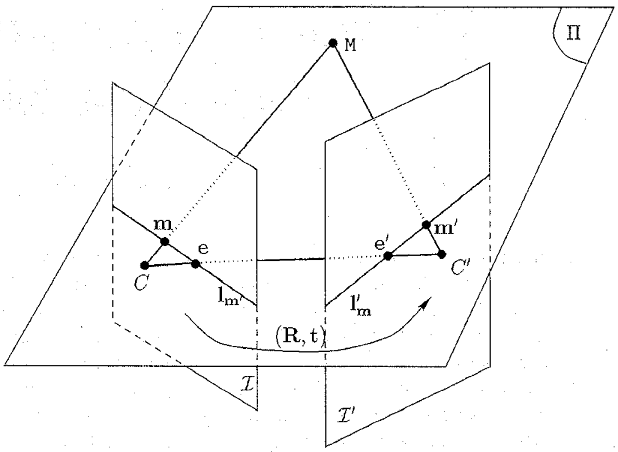


Fig. 1. The epipolar geometry.

$$\tilde{\mathbf{m}}'^T\mathbf{F}\tilde{\mathbf{m}} = 0, \tag{2}$$

where

$$\mathbf{F} = \left[\mathbb{P}'\mathbf{p}^{\perp}\right]_{\times}\mathbb{P}'\mathbb{P}^{+}. \tag{3}$$

Here, $\mathbb{P}^+$ is the pseudo-inverse of matrix $\mathbb{P}$: $\mathbb{P}^+ = \mathbb{P}^T(\mathbb{P}\mathbb{P}^T)^{-1}$, and $\mathbf{p}^{\perp}$ is any four-vector that is perpendicular to all the *row* vectors of $\mathbb{P}$, i.e., $\mathbb{P}\mathbf{p}^{\perp} = \mathbf{0}$. Thus, $\mathbf{p}^{\perp}$ is a null vector of $\mathbb{P}$. As a matter of fact, $\mathbf{p}^{\perp}$ indicates the position of the optical center (to which all optical rays converge). Geometrically, $\mathbf{F}\tilde{\mathbf{m}}$ defines the epipolar line of point $\mathbf{m}$ in the second image. Equation (2) says no more than that the correspondence in the right image of point $\mathbf{m}_1$ lies on the corresponding epipolar line. Transposing (2) yields the symmetric relation from the second image to the first image. The expression of the fundamental matrix given in (3) is a general one, which is valid for both perspective and affine cameras.

2.3 Problem Statement

The problem considered here is the estimation of $\mathbf{F}$ from a sufficiently large set of point correspondences:

$$\{(\mathbf{m}_i, \mathbf{m}'_i) | i = 1, \ldots, n\},$$

where $n \geq 7$. The point correspondences between two images can be established by a technique such as that described in [23]. We do not assume that the camera's intrinsic parameters are known, but the following analysis is equally valid for that case.

3 Three Criteria and Their Relationship

There are a number of techniques proposed in the literature to solve the above problem. The linear eight-point algorithm [8], [13] produces a closed-form solution, but the criterion used is not physically meaningful, and the result obtained is usually not satisfactory. Thus, this solution is usually refined through nonlinear optimization based on some appropriate criterion. Note, however, Hartley [13] shows that a simple normalization (translation and scaling) of image coordinates considerably improves the performance of the eight-point algorithm.

3.1 Three Criteria

For nonlinear estimation techniques, the three best-known optimization criteria are the following.

3.1.1 Distances Between Points and Epipolar Lines

The first is based on the distances between points and their corresponding epipolar lines:

$$\Sigma\left(d^2\left(\mathbf{m}_i', \mathbf{F}\tilde{\mathbf{m}}_i\right) + d^2\left(\mathbf{m}_i, \mathbf{F}^T\tilde{\mathbf{m}}_i'\right)\right).$$

Since the distance of a point $\mathbf{m}'$ to its corresponding epipolar line $\mathbf{l} = \mathbf{F}\tilde{\mathbf{m}}$ is given by

$$d(\mathbf{m}', \mathbf{l}) = d(\mathbf{m}', \mathbf{F}\tilde{\mathbf{m}}) = \frac{\tilde{\mathbf{m}}'^T\mathbf{F}\tilde{\mathbf{m}}}{\sqrt{\tilde{\mathbf{m}}^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}}},$$

we have the following criterion

$$J_1 = \sum_{i=1}^{n}\left(\frac{1}{\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i} + \frac{1}{\tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'}\right)\left(\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i\right)^2 \rightarrow \min \quad (4)$$

As can be observed, the two images play a symmetric role.

3.1.2 Gradient-Weighted Epipolar Errors

The second is based on the gradient-weighted epipolar errors. When data points are noisy, the epipolar constraint (2) is not exactly satisfied, i.e., $r_i = \tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i \neq 0$. If we assume that the points are perturbed by independent identically distributed Gaussian noise with mean zero and covariance matrix $\Lambda = \sigma^2 \operatorname{diag}(1, 1)$, then the variance of r_i is given, under the first order approximation, by

$$\begin{aligned}\sigma_{r_i}^2 &= \frac{\partial r_i}{\partial \mathbf{m}_i'}\Lambda\frac{\partial r_i^T}{\partial \mathbf{m}_i'} + \frac{\partial r_i}{\partial \mathbf{m}_i}\Lambda\frac{\partial r_i^T}{\partial \mathbf{m}_i} \\ &= \sigma^2\left(\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'\right)\end{aligned}$$

We can easily find that $\sigma_{r_i}^2$ varies from one point to another. Therefore, instead of minimizing $\Sigma_i r_i^2$, we should minimize the Mahalanobis distance, which is in our case the gradient-weighted quantity:

$$J_2 = \sum_{i=1}^{n}\frac{\left(\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i\right)^2}{\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'} \rightarrow \min. \quad (5)$$

Note that we have removed the constant σ^2 because dividing an objective function by a constant does not change the solution.

3.1.3 Distances Between Points and Reprojections

The third is based on the distances between points and the reprojections of their reconstructed points, that is

$$J_3 = \sum_{i=1}^{n}\left(\left\|\mathbf{m}_i - \hat{\mathbf{m}}_i\right\|^2 + \left\|\mathbf{m}_i' - \hat{\mathbf{m}}_i'\right\|^2\right) \rightarrow \min, \quad (6)$$

with

$$\hat{\mathbf{m}}_i = \frac{1}{\mathbf{p}_3^T\tilde{\mathrm{M}}_i}\begin{bmatrix}\mathbf{p}_1^T \\ \mathbf{p}_2^T\end{bmatrix}\tilde{\mathrm{M}}_i,\ \hat{\mathbf{m}}_i' = \frac{1}{\mathbf{p}_3'^T\tilde{\mathrm{M}}_i}\begin{bmatrix}\mathbf{p}_1'^T \\ \mathbf{p}_2'^T\end{bmatrix}\tilde{\mathrm{M}}_i, \quad (7)$$

where $\mathbf{p}_i^T$ and $\mathbf{p}_i'^T$ are the ith row of camera projection matrices $\mathbb{P}$ and $\mathbb{P}'$, which are related to $\mathbf{F}$, and $\tilde{\mathrm{M}}_i$ is the 3D projective point corresponding to $(\mathbf{m}_i, \mathbf{m}_i')$. This criterion can be derived from the maximum likelihood principle based on the same assumption on noise as that in the previous criterion. The minimization process is speeded up by separating the estimation of $\tilde{\mathrm{M}}_i$ from that of $\mathbf{F}$ (see [5] for details).

3.2 Relationship Between J_2 and J_3

Let us first consider J_3 (6). It is evident that $\hat{\mathbf{m}}$ and $\hat{\mathbf{m}}'$ satisfy the epipolar constraint (2). Therefore, minimizing J_3 is equivalent to minimize

$$J_3' = \sum_{i=1}^{n}\left(\left\|\mathbf{m}_i - \hat{\mathbf{m}}_i\right\|^2 + \left\|\mathbf{m}_i' - \hat{\mathbf{m}}_i'\right\|^2\right) \rightarrow \text{subject to } \tilde{\hat{\mathbf{m}}}_i'^T\mathbf{F}\tilde{\hat{\mathbf{m}}}_i = 0 \quad (8)$$

Define

$$\begin{cases}\Delta\mathbf{m} = \mathbf{m} - \hat{\mathbf{m}} \\ \Delta\mathbf{m}' = \mathbf{m}' - \hat{\mathbf{m}}'\end{cases} \Rightarrow \begin{cases}\tilde{\hat{\mathbf{m}}} = \tilde{\mathbf{m}} - \Delta\breve{\mathbf{m}} \\ \tilde{\hat{\mathbf{m}}}' = \tilde{\mathbf{m}}' - \Delta\breve{\mathbf{m}}'\end{cases} \quad (9)$$

The constraint (2) becomes

$$\tilde{\mathbf{m}}'^T\mathbf{F}\tilde{\mathbf{m}} - \Delta\breve{\mathbf{m}}'^T\mathbf{F}\tilde{\mathbf{m}} - \tilde{\mathbf{m}}'^T\mathbf{F}\Delta\breve{\mathbf{m}} + \Delta\breve{\mathbf{m}}'^T\mathbf{F}\Delta\breve{\mathbf{m}} = 0. \quad (10)$$

If we neglect the second order term $O_2 = \Delta\breve{\mathbf{m}}'^T\mathbf{F}\Delta\breve{\mathbf{m}}$ (see the end of this subsection for discussions) and use the Lagrange multiplier, we can transform the constrained minimization problem (8) into an unconstrained one:

$$\begin{aligned}J_3'' = \sum_{i=1}^{n}(&\Delta\mathbf{m}_i^T\Delta\mathbf{m}_i + \Delta\mathbf{m}_i'^T\Delta\mathbf{m}_i' + \\ &\lambda_i(\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i - \Delta\mathbf{m}_i'^T\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\Delta\mathbf{m}_i))\end{aligned} \quad (11)$$

Here we have used the equation $\breve{\mathbf{m}} = \mathbf{Z}\mathbf{m}$. If we let the first-order derivatives of J_3'' with respect to $\Delta\mathbf{m}_i$ and $\Delta\mathbf{m}_i'$ be zero, we have

$$\frac{\partial J_3''}{\partial \Delta\mathbf{m}_i} = 2\Delta\mathbf{m}_i - \lambda_i\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i' = 0 \quad \Rightarrow \Delta\mathbf{m}_i = \frac{\lambda_i}{2}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'$$

$$\frac{\partial J_3''}{\partial \Delta\mathbf{m}_i'} = 2\Delta\mathbf{m}_i' - \lambda_i\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i = 0 \quad \Rightarrow \Delta\mathbf{m}_i' = \frac{\lambda_i}{2}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i$$

Substituting them into the linearized constraint of (10) gives

$$\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i - \frac{\lambda_i}{2}\left[\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'\right] = 0,$$

which in turn yields

$$\lambda_i = 2\frac{\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i}{\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'}.$$

Substituting the obtained λ_i, $\Delta\mathbf{m}_i$, and $\Delta\mathbf{m}_i'$ back into (11) finally gives

$$\begin{aligned} J_3'' &= \sum_{i=1}^{n}(\Delta\mathbf{m}_i^T\Delta\mathbf{m}_i + \Delta\mathbf{m}_i'^T\Delta\mathbf{m}_i') \\ &= \sum_{i=1}^{n}\frac{\left(\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i\right)^2}{\left(\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'\right)^2} \\ &\quad \left(\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'\right) \\ &= \sum_{i=1}^{n}\frac{\left(\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i\right)^2}{\tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}'} \end{aligned} \tag{12}$$

This is *exactly* the gradient-based criterion J_2 (5).

Note that in the above derivation, the only term we neglect is the second-order term

$$O_2 = \Delta\tilde{\mathbf{m}}_i'^T\mathbf{F}\Delta\tilde{\mathbf{m}}_i.$$

This implies that *if* $\Delta\tilde{\mathbf{m}}_i'^T\mathbf{F}\Delta\tilde{\mathbf{m}}_i = 0$, *then the two criteria* J_2 *and* J_3 *are equivalent.* This is the case if the fundamental matrix is of the following form

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{bmatrix}.$$

For example, when the epipoles are at infinity or when the cameras are affine, i.e., when the epipolar lines are parallel, the two criteria are equivalent. The case for affine cameras has already been shown by Oxford group [15], [24]. If $\mathbf{F}$ is not of the above form, as long as either $\Delta\mathbf{m}_i$ is much smaller than $\mathbf{m}_i$ or $\Delta\mathbf{m}_i'$ is much smaller than $\mathbf{m}_i'$, then the second-order term is much smaller than one of the first-order terms: $\Delta\tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i$ or $\tilde{\mathbf{m}}_i'\mathbf{F}\Delta\tilde{\mathbf{m}}_i$. In practice, this condition is easily satisfied. For a particular point match, neglecting the second order term O_2 implies that the epipolar line does not change the orientation in its neighborhood. If a point is not close to the epipole, then the orientation of the epipolar line in the neighborhood does not change much indeed. For example, for a perturbation of one pixel perpendicular to the epipolar line, the angle between the new epipolar line and the original one is inversely related to the distance of the point to the epipole. If the distance is of 50 pixels, then the angle is only one degree. Experimental results provided in Section 5.3 show that J_2 is really a good approximation of J_3.

3.3 Relationship Between J_1 and J_2

Denote again the epipolar residual by r_i, i.e.,

$$r_i = \tilde{\mathbf{m}}_i'^T\mathbf{F}\tilde{\mathbf{m}}_i. \tag{13}$$

Let $w_i = \tilde{\mathbf{m}}_i'^T\mathbf{F}\mathbf{Z}\mathbf{Z}^T\mathbf{F}^T\tilde{\mathbf{m}}_i'$ and $w_i' = \tilde{\mathbf{m}}_i^T\mathbf{F}^T\mathbf{Z}\mathbf{Z}^T\mathbf{F}\tilde{\mathbf{m}}_i$, then J_1 (4) can be rewritten as

$$J_1 = \sum_{i=1}^{n}\left(\frac{1}{w_i} + \frac{1}{w_i'}\right)r_i^2 = \sum_{i=1}^{n}\frac{\left(w_i + w_i'\right)^2}{w_i w_i'}\frac{r_i^2}{w_i + w_i'}.$$

Therefore, J_1 is obtained by weighting each term of J_2 with

$$k_i = \frac{\left(w_i + w_i'\right)^2}{w_i w_i'} = \frac{\left(1 + s_i\right)^2}{s_i} \text{ with } s_i = \frac{w_i}{w_i'}.$$

When minimizing the criterion J_1, we will minimize both k_i and J_2. The minimum of k_i, which is equal to four, is achieved when $s_i = 1$, i.e., when $w_i = w_i'$. Thus, minimizing k_i is equivalent to estimating the fundamental matrix such that the contributions from both images are equal. This is exactly what we need when introducing J_1. Remember, however, that minimizing k_i should be accompanied by minimizing $r_i^2 / \left(w_i + w_i'\right)$, the latter is usually dominant, as was observed in our experiments. In consequence, the difference between the results obtained with J_1 and with J_2 is small, as will be shown in the experimental section.

4 Practical Equivalence

In practice, the rotation between two images is usually small. In order to gain a better understanding of the relationship between the three criteria, let us consider a special case in this section: a pure translation with $\mathbf{t} = [x, y, z]^T$. Note, however, the minimization is not constrained to pure translations.

Let the image scale factors be α_u and α_v, the coordinates of the principal point be u_0 and v_0, and the two image axes are orthogonal. Furthermore, we use $'$ to indicate the intrinsic parameters of the second camera. Then, the fundamental matrix is given by

$$F = \begin{bmatrix} 0 & -\frac{z}{\alpha_u'\alpha_v} & \frac{zv_0 + y\alpha_v}{\alpha_u'\alpha_v} \\ \frac{z}{\alpha_u\alpha_v'} & 0 & -\frac{zu_0 + x\alpha_u}{\alpha_u\alpha_v'} \\ -\frac{v_0'z + y\alpha_v'}{\alpha_u\alpha_v'} & \frac{u_0'z + x\alpha_u'}{\alpha_u'\alpha_v} & * \end{bmatrix}, \tag{14}$$

where the term "*" is irrelevant to our discussion here.

4.1 Practical Equivalence Between J_3 and J_2

The only term we neglect in deriving J_2 from J_3 is the second-order term $O_2 = \Delta\tilde{\mathbf{m}}_i'^T\mathbf{F}\Delta\tilde{\mathbf{m}}_i$. For our special $\mathbf{F}$ given in (14),

$$O_2 = z\left(\frac{\Delta u_i\Delta v_i'}{\alpha_u\alpha_v'} - \frac{\Delta u_i'\Delta v_i}{\alpha_u'\alpha_v}\right).$$

If $z = 0$ (i.e., the epipoles are at infinity), then this term is always zero, and J_2 and J_3 are, as we said before, equivalent. It is worth stressing, thanks to one of the reviewers, that the equality of J_2 and J_3 for translations with $z = 0$ (i.e., parallel to the image plane) is not strictly relevant to motion-recovery techniques minimizing these errors. The point is that the minimization is not restricted to such translations even if the exact translation is parallel to the image plane.

If $z \neq 0$, let's consider one of the first-order terms

$$\tilde{\mathbf{m}}_i'^T \mathbf{F} \Delta \breve{\mathbf{m}}_i = z\left(\frac{v_i'\Delta u_i}{\alpha_u \alpha_v'} - \frac{u_i'\Delta v_i}{\alpha_u' \alpha_v}\right) + o.t.$$

where $o.t.$ represents other terms. It is clear that if $\Delta u'_i \ll u_i'$ and $\Delta v_i' \ll v_i'$, then O_2 can be neglected compared to the first term of the right-hand side of the above equation. Similarly, if $\Delta u_i \ll u_i$ and $\Delta v_i \ll v_i$, then O_2 can be neglected compared to $\Delta \breve{\mathbf{m}}_i'^T \mathbf{F} \tilde{\mathbf{m}}_i$.

4.2 Practical Equivalence Between J_1 and J_2

We consider now the factor k_i, by which each term of J_2 is multiplied in order to obtain J_1. If $z \neq 0$, it is easy to compute, from **F**, the epipole in the first image as

$$[e_u, e_v]^T = [u_0 + \alpha_u x/z, v_0 + \alpha_v y/z]^T$$

and that in the second image as

$$[e_u', e_v]^T = [u_0' + \alpha_u' x / z, v_0' + \alpha_v' y / z]^T.$$

After some algebra, we finally obtain

$$s_i = \frac{w_i}{w_i'} = \frac{\alpha_u'^2 \alpha_v^2 (u_i - e_u)^2 + \alpha_u^2 \alpha_v'^2 (v_i - e_v)^2}{\alpha_u^2 \alpha_v'^2 (u_i' - e_u')^2 + \alpha_u'^2 \alpha_v^2 (v_i' - e_v')^2}. \tag{15}$$

The term k_i is thus minimized (i.e., $s_i \rightarrow 1$) if the epipoles in the two images have the same offset with respect to the matched points. If $z = 0$ (i.e., the translation is parallel to the image plane), then

$$w_i = F_{13}^2 + F_{23}^2 = x^2 / \alpha_v'^2 + y^2 / \alpha_u'^2$$

$$w_i' = F_{31}^2 + F_{32}^2 = x^2 / \alpha_v^2 + y^2 / \alpha_u^2$$

$$s_i = \frac{w_i}{w_i'} = \left(\frac{\alpha_u^2 \alpha_v^2}{\alpha_u'^2 \alpha_v'^2}\right)\left(\frac{\alpha_u'^2 x^2 + \alpha_v'^2 y^2}{\alpha_u^2 x^2 + \alpha_v^2 y^2}\right).$$

We can see that w_i, w_i', and s_i do not depend on a particular point match. They are the same for all point matches. Furthermore, we see that s_i is closely related to the ratio of the two image scales. If the two images have the same scales $(\alpha_u = \alpha_u'$ and $\alpha_v = \alpha_v')$, then $s_i = 1$. In that case, the criteria J_1 and J_2 are equivalent.

5 Experimental Results With Computer Simulations

In this section, we compare the three criteria with four different configurations of the epipolar geometry. In each configuration, a set of 104 noise-free 3D points are projected onto the images. The image resolution is 512 pixels × 512 pixels. The intrinsic parameters are the same for all images except for the second image of configuration 2. They are: $\alpha_u = 700$, $\alpha_v = 1{,}000$, $u_0 = 255$, $v_0 = 255$. The field of view is about 40 degrees. The three synthetic objects are at a distance of 2,600 mm, 3,200 mm, and 1,400 mm from the camera. Independent Gaussian noise with mean zero and standard deviation σ is added to the image points. The three criteria are then used to estimate the fundamental matrix from the *same set* of noisy image points. The three estimated fundamental matrices and the true one are compared to each other. The software Fdiff is used for the comparison. It computes the distance between two pencils of epipolar lines, defined by the two fundamental matrices, through sampling the whole visible 3D space. If the two fundamental matrices are equal to each other, then the two pencils of the epipolar lines coincide; otherwise, there is an difference. This distance is measured in image pixels (see [5] for more details). For each noise level σ, 200 trials have been conducted, and the average distances will be provided below. Furthermore, we have carried out the experiments with three noise levels: 0.25, 0.5, and one pixel.

TABLE 2

DISTANCES (IN PIXELS) BETWEEN THE ESTIMATED FUNDAMENTAL MATRICES AND THE TRUE ONE

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.00	0.08	1.67
$\mathbf{F}_2$		0.08	1.67
$\mathbf{F}_3$			1.70

(a)

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.00	0.08	3.43
$\mathbf{F}_2$		0.08	3.43
$\mathbf{F}_3$			3.46

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.01	0.25	7.33
$\mathbf{F}_2$		0.24	7.34
$\mathbf{F}_3$			7.37

(c)

Configuration 1. (a) $\sigma = 0.25$ pixel. (b) $\sigma = 0.5$ pixel. (c) $\sigma = 1$ pixel.

5.1 Comparison of the Estimated Fundamental Matrices

The estimated fundamental matrices based on the distances between points and their corresponding epipolar lines, the gradient-weighted epipolar errors and the distances between points and reprojections of the 3D reconstruction will be denoted by $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$, respectively. The true fundamental matrix is denoted by $\mathbf{F}_t$.

5.1.1 Configuration 1

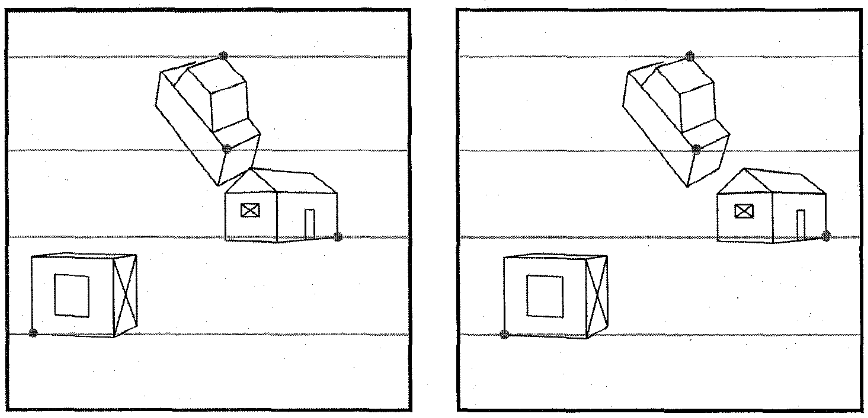
The second camera is displaced to the left by 100 mm, i.e., a pure translation with $\mathbf{t} = [100, 0, 0]^T$. Therefore, the epipoles in both images are at infinity. The test images are shown in Fig. 2, and the result is shown in Table 2.

5.1.2 Configuration 2

This configuration is the same as the first one, except that we simulate a zoom for the second image. The scale of the second image is 1.2 times that of the first image: $\alpha_u' = 840$, $\alpha_v' = 1{,}200$. The test images are shown in Fig. 3, and the result is shown in Table 3.

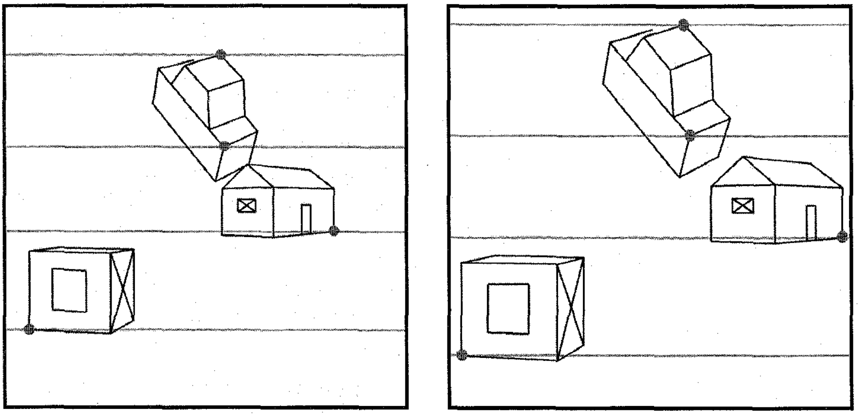
5.1.3 Configuration 3

The second camera first rotates to the left by 30 degrees and then translates to the right by 1,000 mm. Thus, the epipole in image 1 is $\mathbf{e} = [1{,}467.4, 255]^T$, while the epipole in image 2 is at infinity. The test images are shown in Fig. 4, and the result is shown in Table 4.



(a) (b)

Fig. 2. The test images used in Configuration 1. Four epipolar lines are shown.



(a) (b)

Fig. 3. The test images used in Configuration 2. Four epipolar lines are shown.

5.1.4 Configuration 4

The second camera advances by 250 mm, i.e., a pure translation with $\mathbf{t} = [0, 0, -250]^T$. The epipoles in both images are at the image center, i.e., $[255, 255]^T$. The test images are shown in Fig. 5, and the result is shown in Table 5.

In [17, Section VII.B], Weng et al. compared criteria J_2 and J_3 in one experiment, by measuring the rotation and translation errors separately, for different degrees of fields of view. The two criteria do behave approximately the same, but J_3 performs better for smaller fields of view (less than 30 degrees). The error of approximating J_3 by J_2 probably becomes larger in that case. Another possibility is probably due to the bas-relief ambiguity. In their experiment, the translations are smaller for small fields of view than for wide fields of view. It is well known that in case of small translations, it is not easy to distinguish the image effects of rotating and translating. Therefore, although J_2 has larger errors in both rotation and translation than J_3, the epipolar geometry estimated with J_2 may still not be distinguishable, in terms of image errors, from the one estimated with J_3.

TABLE 3
DISTANCES (IN PIXELS) BETWEEN THE ESTIMATED FUNDAMENTAL MATRICES AND THE TRUE ONE

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.02	0.02	1.59
$\mathbf{F}_2$		0.00	1.59
$\mathbf{F}_3$			1.59

(a)

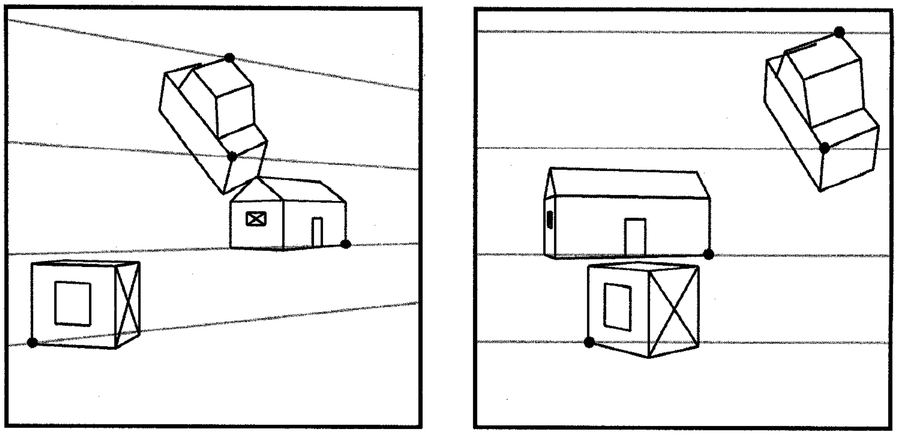
	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.07	0.07	3.78
$\mathbf{F}_2$		0.00	3.79
$\mathbf{F}_3$			3.79

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.29	0.29	7.81
$\mathbf{F}_2$		0.00	7.81
$\mathbf{F}_3$			7.81

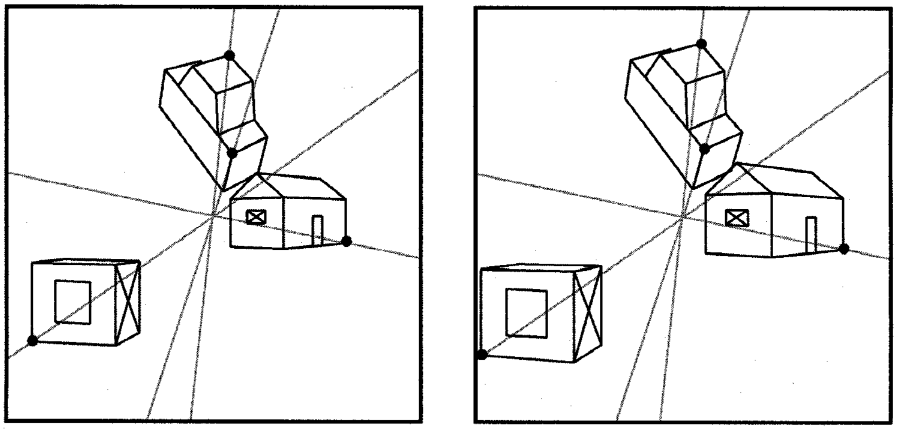
(c)

Configuration 2. (a) $\sigma = 0.25$ pixel. (b) $\sigma = 0.5$ pixel. (c) $\sigma = 1$ pixel.



(a) (b)

Fig. 4. The test images used in Configuration 3. Four epipolar lines are shown.



(a) (b)

Fig. 5. The test images used in Configuration 4. Four epipolar lines are shown.

TABLE 4
DISTANCES (IN PIXELS) BETWEEN THE ESTIMATED FUNDAMENTAL MATRICES AND THE TRUE ONE

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.00	0.00	0.18
$\mathbf{F}_2$		0.00	0.18
$\mathbf{F}_3$			0.18

(a)

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.01	0.01	0.38
$\mathbf{F}_2$		0.00	0.38
$\mathbf{F}_3$			0.38

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.03	0.03	0.72
$\mathbf{F}_2$		0.00	0.72
$\mathbf{F}_3$			0.72

(c)

Configuration 3. (a) $\sigma = 0.25$ pixel. (b) $\sigma = 0.5$ pixel. (c) $\sigma = 1$ pixel.

TABLE 5
DISTANCES (IN PIXELS) BETWEEN THE ESTIMATED FUNDAMENTAL MATRICES AND THE TRUE ONE

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.09	0.13	4.14
$\mathbf{F}_2$		0.06	4.13
$\mathbf{F}_3$			4.16

(a)

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	0.25	0.28	7.65
$\mathbf{F}_2$		0.06	7.65
$\mathbf{F}_3$			7.67

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_t$
$\mathbf{F}_1$	1.17	0.86	15.79
$\mathbf{F}_2$		0.55	15.92
$\mathbf{F}_3$			15.75

(c)

Configuration 4. (a) $\sigma = 0.25$ pixel. (b) $\sigma = 0.5$ pixel. (c) $\sigma = 1$ pixel.

TABLE 6
COMPARISON OF J1 AND J2 WITH RESPECT TO $s_i = w_i \,/\, w_i'$ FOR CONFIGURATION 2

	$\sigma = 0.25$			$\sigma = 0.5$			$\sigma = 1.0$		
	mean	dev.	diff.	mean	dev.	diff.	mean	dev.	diff.
$\mathcal{J}_1$	0.69486	6.40e-03	0.065	0.69545	1.21e-02	0.083	0.69412	2.49e-02	0.013
$\mathcal{J}_2$	0.69483	6.40e-03	0.060	0.69530	1.21e-02	0.070	0.69358	2.50e-02	0.035
ratio	1.00005	1.81e-05	2.763	1.00021	6.95e-05	3.022	1.00079	3.68e-04	2.148

5.2 Bias of J_1 With Respect to J_2

Now let us examine the relationship between J_1 and J_2 that we have analyzed. When the epipoles are at infinity, the quantity

$$s_i = w_i \,/\, w_i' = \left(F_{13}^2 + F_{23}^2\right) / \left(F_{31}^2 + F_{32}^2\right)$$

is a constant, and is closely related to the ratio of the image scales. For Configuration 1, the two images have the same scale, therefore $s_i = 1$. For Configuration 2, $s_i = 1/(1.2)^2$ because the scale of the second image is 1.2 times that of the first. We have conducted 200 trials on the data of Configuration 2. In each trial, s_i based on J_1, s_i based on J_2, and their ratio are estimated. Finally, their means and sample deviations are computed, as shown in Table 6. In the table, the columns corresponding to "**diff.**" are the absolute difference between the mean and the expected value *weighted* by the estimated sample deviation. If that value is small, then no bias is observed. This is clearly not the case for the ratio of s_is. Indeed, we have observed that the ratio is always larger than one in all trials. This implies that minimizing criterion J_1 indeed tends to bias s_i toward one, although the bias is very small (less than 0.1 percent). We have also examined the ratio with Configuration 1. The mean of the *ratio* is exactly one, which confirms that the two criteria are equivalent in that case.

For Configuration 4, the epipoles are not at infinity, and, thus, s_i usually changes from one point match to another. The values are however all around 0.77. From our analysis, we would expect that $s_i^{(1)}$s given by J_1 are larger than $s_i^{(2)}$s given by J_2 (please note the use of the superscript). If two images have the same aspect ratio (and this is the case for Configuration 4), (15) can be reduced to

$$s_i = \frac{\left(u_i - e_u\right)^2 + \left(v_i - e_v\right)^2}{\left(u_i' - e_u'\right)^2 + \left(v_i' - e_v'\right)^2}.$$

For each noise level, we conducted 200 trials. For each trial, we compute the *average ratio of* s_i as

$$\frac{1}{n}\sum_{i=1}^{n}\left(s_i^{(1)} \,/\, s_i^{(2)}\right),$$

which is expected to be larger than one because of the bias of J_1. When $\sigma = 0.25$, there are nine trials which give the average ratio less one, and the mean of the average ratios is 1.00002. When $\sigma = 0.5$, there are three trials which give the average ratio less one, and the mean of the average ratios is 1.00007. When $\sigma = 1$, there are five trials which give the average ratio less one, and the mean of the average ratios is 1.00017. There is clearly a bias when J_1 is used. It tends to favor the fundamental matrix whose epipoles have the same offset with respect to the image points in both images. Although small, the bias increases when the noise in the data points increases.

For general configurations involving rotation out of image plane between two images, such as Configuration 3, we still observe the bias with J_1. For Configuration 3, the values of s_i are also around 0.77. As for Configuration 4, we compute the average ratio $s_i^{(1)} \,/\, s_i^{(2)}$, which is expected to be larger than one if a bias does exist. When $\sigma = 0.25$, there are 46 among 200 trials which give the average ratio less one, and the mean of the average ratios is 1.00001. When $\sigma = 0.5$, there are 24 trials which give the average ratio less one, and the mean of the average ratios is 1.00003. When $\sigma = 1$, there are 29 trials which give the average ratio less one, and the mean of the average ratios is 1.00022. The bias is not as strong as in Configuration 4 because one of the epipole is at infinity.

5.3 Difference Between J_2 and J_3

As said in Section 4.1, J_2 approximates J_3 quite well if an epipolar line does not change its orientation too much in the neighborhood. They become equivalent when the epipoles are at infinity, in which case the epipolar lines all have the same orientation. In this section, we investigate the approximation error with respect to the distance of points to the epipoles.

We consider two images which differ from each other by a pure translation. The fundamental matrix is given by (14). We fix the translation magnitude to 200 mm and the y component to zero. We vary the angle between the translation vector and the z-axis. Therefore, the translation vector is given by $\mathbf{t} = 200[\sin\theta, 0, \cos\theta]^T$. We vary θ from zero degrees to 90 degrees. When $\theta = 0$ degrees, both epipoles are at $[255, 255]^T$; when $\theta = 5$ degrees, both epipoles are at $[316.242, 255]^T$; when $\theta = 90$ degrees, both epipoles are at $[+\infty, 255]^T$. We only consider one point match. The point in the first image is $\mathbf{m} = [245, 255]^T$; the corresponding point in the second image is $\mathbf{m}' = [250, 255]^T$. They satisfy the epipolar constraint for all θ. We consider this point match because they are close to the image center, and therefore, when $\theta = 0$ degrees, criterion J_2 is almost the worst approximation to J_3 that we can expect. Independent Gaussian noise with mean zero and standard deviation σ is then added to $\mathbf{m}$ and $\mathbf{m}'$. Four noise levels have been experimented: $\sigma = 0.5$, 1, 2, and 5 pixels. For each noise level, 200 trials have been conducted. For each noisy data, we compute the true reconstruction error based on J_3:

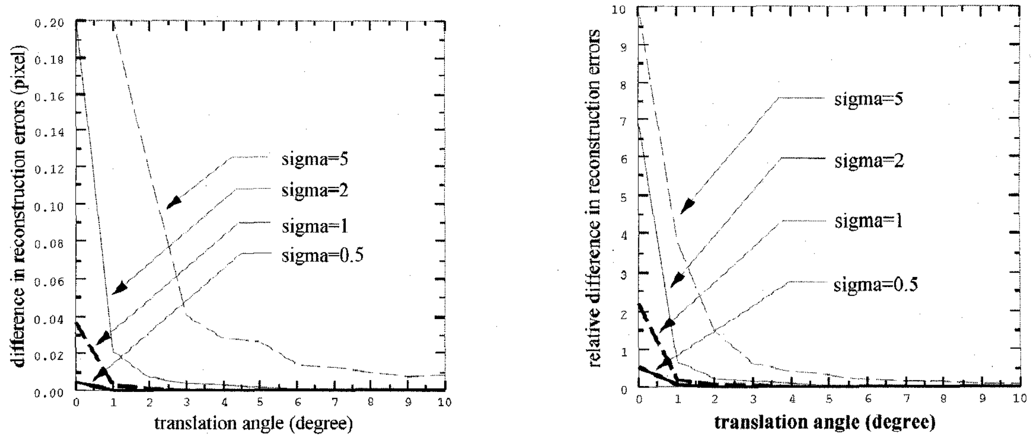
$$E_{J3} = \min_{\mathrm{M}} \sqrt{\|\mathbf{m} - \hat{\mathbf{m}}\|^2 + \|\mathbf{m}' - \hat{\mathbf{m}}'\|^2},$$

where $\hat{\mathbf{m}}$ and $\hat{\mathbf{m}}'$ are the projections of the reconstructed points, as given in (7), and the first-order approximation based on J_2:

TABLE 7
COMPARISON OF J2 AND J3 WITH RESPECT TO DIFFERENT TRANSLATION DIRECTION θ

θ	$d(\mathbf{m},\mathbf{e})$	$\sigma = 0.5$			$\sigma = 1$			$\sigma = 2$			$\sigma = 5$		
		$E_{\mathcal{J}3}$	E_{abs}	E_{rel}	$E_{\mathcal{J}3}$	E_{abs}	E_{rel}	$E_{\mathcal{J}3}$	E_{abs}	E_{rel}	$E_{\mathcal{J}3}$	E_{abs}	E_{rel}
0°	10.0	0.41	0.01	0.53	0.87	0.04	2.22	1.74	0.24	6.93	4.45	1.35	17.3
1°	22.2	0.42	0.00	0.04	0.77	0.00	0.16	1.55	0.02	0.66	4.00	0.31	3.79
2°	34.4	0.35	0.00	0.01	0.73	0.00	0.05	1.59	0.01	0.23	4.04	0.12	1.52
3°	46.7	0.40	0.00	0.01	0.78	0.00	0.03	1.64	0.00	0.12	3.79	0.04	0.60
4°	58.9	0.44	0.00	0.01	0.82	0.00	0.02	1.65	0.00	0.07	3.86	0.03	0.39
5°	71.2	0.38	0.00	0.00	0.80	0.00	0.01	1.58	0.00	0.05	4.05	0.03	0.30
6°	83.6	0.40	0.00	0.00	0.87	0.00	0.01	1.67	0.00	0.03	3.98	0.01	0.19
7°	95.9	0.43	0.00	0.00	0.79	0.00	0.01	1.60	0.00	0.02	4.00	0.01	0.15
8°	108.4	0.43	0.00	0.00	0.79	0.00	0.00	1.43	0.00	0.02	4.09	0.01	0.12
9°	120.9	0.40	0.00	0.00	0.86	0.00	0.00	1.64	0.00	0.02	3.85	0.01	0.09
10°	133.4	0.42	0.00	0.00	0.79	0.00	0.00	1.53	0.00	0.01	4.38	0.01	0.09
15°	197.6	0.41	0.00	0.00	0.83	0.00	0.00	1.57	0.00	0.01	3.97	0.00	0.03
20°	264.8	0.40	0.00	0.00	0.81	0.00	0.00	1.66	0.00	0.00	4.36	0.00	0.02
25°	336.4	0.41	0.00	0.00	0.92	0.00	0.00	1.49	0.00	0.00	4.45	0.00	0.01
30°	414.1	0.42	0.00	0.00	0.83	0.00	0.00	1.56	0.00	0.00	4.01	0.00	0.01
35°	500.1	0.41	0.00	0.00	0.84	0.00	0.00	1.63	0.00	0.00	4.18	0.00	0.01
40°	597.4	0.41	0.00	0.00	0.79	0.00	0.00	1.72	0.00	0.00	4.20	0.00	0.00
45°	710.0	0.42	0.00	0.00	0.80	0.00	0.00	1.56	0.00	0.00	3.99	0.00	0.00
50°	844.2	0.44	0.00	0.00	0.81	0.00	0.00	1.66	0.00	0.00	4.46	0.00	0.00
60°	1222.4	0.41	0.00	0.00	0.74	0.00	0.00	1.37	0.00	0.00	4.15	0.00	0.00
70°	1933.2	0.41	0.00	0.00	0.78	0.00	0.00	1.53	0.00	0.00	3.89	0.00	0.00
80°	3979.9	0.42	0.00	0.00	0.79	0.00	0.00	1.75	0.00	0.00	4.07	0.00	0.00
90°	∞	0.42	0.00	0.00	0.76	0.00	0.00	1.60	0.00	0.00	3.82	0.00	0.00

$d(\mathbf{m}, \mathbf{e})$ is the distance of the point $\mathbf{m}$ in the first image to the epipole. The distance of the point in the second image is $d(\mathbf{m}, \mathbf{e}) - 5$. E_{J3} is the reconstruction error in pixels, corresponding to the distance between points and the reprojection of the reconstructed point. E_{abs} is the absolute difference between E_{J3} and its approximation E_{J2}, in pixels. E_{rel} is the relative difference E_{abs}/E_{J3}, in percentage. All values, except $d(\mathbf{m}, \mathbf{e})$, are the average of 200 trials.



(a) (b)

Fig. 6. Comparison of J_2 and J_3 with respect to different translation direction θ. (a) Absolute difference. (b) Relative difference.

$$E_{J2} = \sqrt{\frac{\left(\tilde{\mathbf{m}}_i'^T \mathbf{F} \tilde{\mathbf{m}}_i\right)^2}{\tilde{\mathbf{m}}_i^T \mathbf{F}^T \mathbf{Z} \mathbf{Z}^T \tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T \mathbf{F} \mathbf{Z} \mathbf{Z}^T \mathbf{F}^T \tilde{\mathbf{m}}_i'}}\,.$$

Furthermore, we compute their absolute difference

$$E_{\text{abs}} = \left|E_{J3} - E_{J2}\right|$$

and the relative difference

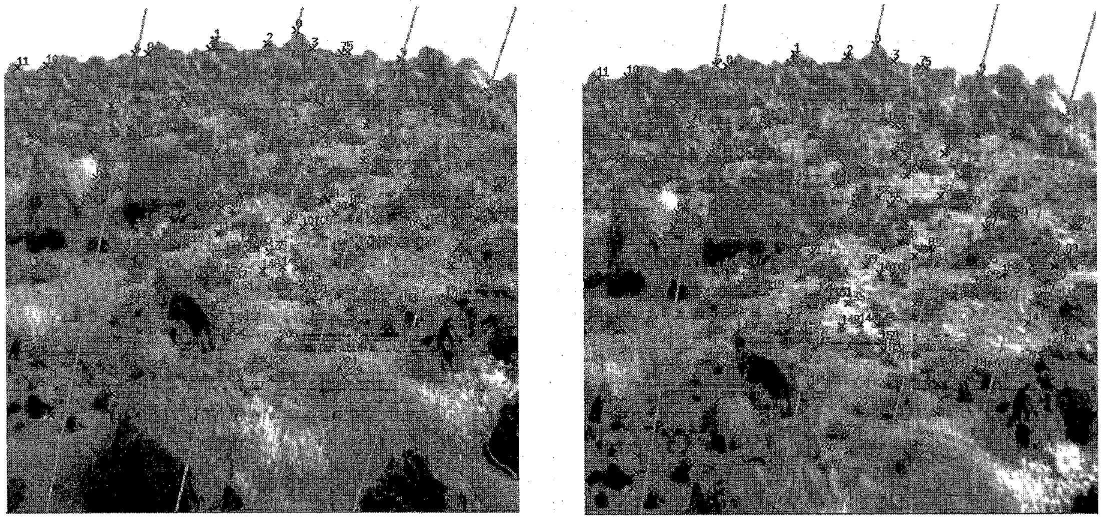
$$E_{\text{rel}} = E_{\text{abs}} / E_{J3}.$$

The average results of the 200 trials are shown in Table 7.

The absolute and relative differences in reconstruction errors between J_2 and J_3 are graphically displayed in Fig. 6. The approximation error decreases as the distance between the point and the epipole increases, and increases as the noise level increases. In all cases, when $\theta \geq 3$ degrees (i.e., $d(\mathbf{m}, \mathbf{e}) > 40$ pixels), the absolute difference is less than 0.1 pixel and the relative difference is less than 1 percent, which shows that the approximation of J_3 by J_2 is very good in practice.

6 Experimental Results With Real Data

In this section, we provide experimental results with four sets of real data, corresponding roughly to the four configurations described in the last section. The point matches have been established automatically by the software called image-matching [23], which is available from the author's home page. Fig. 7, Fig. 9, Fig. 11, and Fig. 13 show the four image pairs, the matched points, and several epipolar lines estimated with criterion J_3. Fig. 8, Fig. 10, Fig. 12, and Fig. 14 show the estimated fundamental matrices and their differ-



(a) (b)

Fig. 7. Real data 1: Rock scene, 242 point matches. Epipoles far away.

$$\begin{bmatrix} 1.56e-6 & 1.21e-4 & -3.65e-1 \\ -1.11e-4 & -1.67e-6 & -4.68e-2 \\ 3.65e-1 & 5.62e-2 & 0.8535 \end{bmatrix}$$

(a)

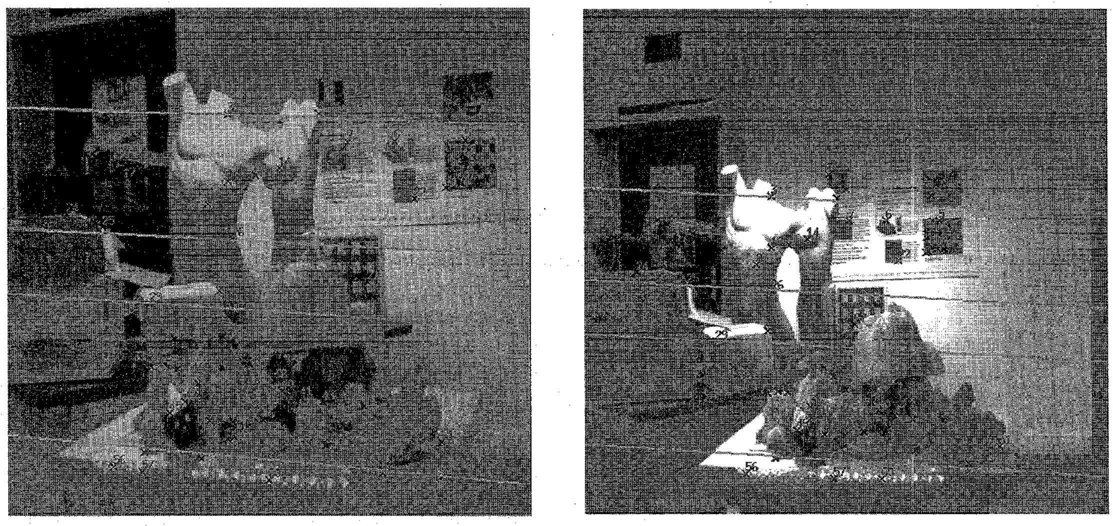
$$\begin{bmatrix} 1.56e-6 & 1.22e-4 & -3.65e-1 \\ -1.12e-4 & -1.67e-6 & -4.68e-2 \\ 3.65e-1 & 5.62e-2 & 0.8535 \end{bmatrix}$$

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$
$\mathbf{F}_1$	0.012	0.010
$\mathbf{F}_2$		0.002

(c)

Fig. 8. Real data 1: Comparison of different estimations of the fundamental matrix. (a) $\mathbf{F}_1$. (b) $\mathbf{F}_3$. (c) Difference (in pixels).



(a) (b)

Fig. 9. Real data 2: Indoor scene with two *different* image scales, 61 point matches. Epipoles far away.

ences for each image pair. The fundamental matrix estimated based on $\mathcal{J}_2$ is not shown for two reasons: The first is that it is very close to $\mathbf{F}_3$; the second is to save space. The same conclusion as with computer simulations can be drawn.

The three criteria have been implemented in a software called FMatrix, available from

http://www.inria.fr/robotvis/personnel/zzhang/

$$\begin{bmatrix} 8.92e-8 & 2.91e-7 & 5.07e-4 \\ -1.86e-6 & -2.43e-7 & -1.02e-2 \\ 7.54e-5 & 8.19e-3 & 0.9999 \end{bmatrix}$$

(a)

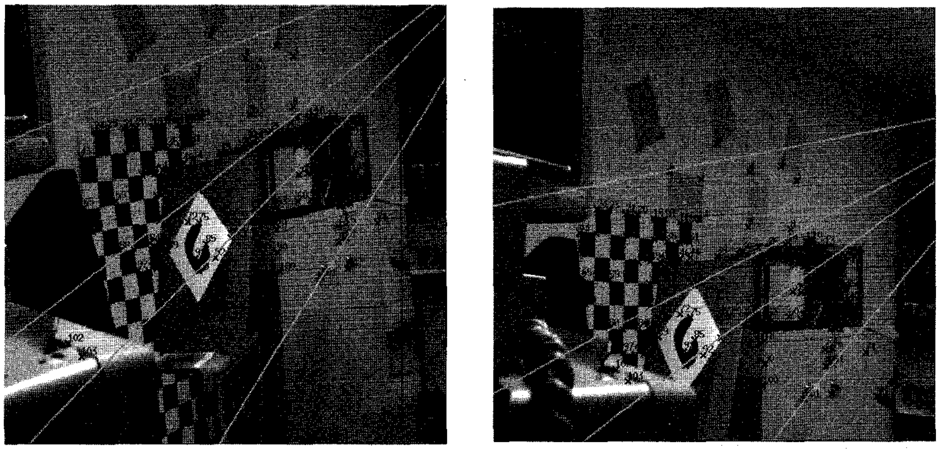
$$\begin{bmatrix} 9.07e-8 & 2.52e-7 & 5.26e-4 \\ -1.82e-6 & -2.36e-7 & -1.01e-2 \\ 5.58e-5 & 8.18e-3 & 0.9999 \end{bmatrix}$$

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$
$\mathbf{F}_1$	0.178	0.178
$\mathbf{F}_2$		0.002

(c)

Fig. 10. Real data 2: Comparison of different estimations of the fundamental matrix. (a) $\mathbf{F}_1$. (b) $\mathbf{F}_3$. (c) Difference (in pixels).



(a) (b)

Fig. 11. Real data 3: Indoor scene, 104 point matches. Epipoles close to the images.

$$\begin{bmatrix} -4.63e-6 & -2.10e-5 & 6.69e-4 \\ 2.13e-5 & -1.34e-6 & -1.35e-2 \\ 7.36e-4 & 1.37e-2 & 0.9998 \end{bmatrix}$$

(a)

$$\begin{bmatrix} -4.62e-6 & -2.10e-5 & 6.64e-4 \\ 2.13e-5 & -1.34e-6 & -1.35e-2 \\ 7.41e-4 & 1.37e-2 & 0.9998 \end{bmatrix}$$

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$
$\mathbf{F}_1$	0.038	0.037
$\mathbf{F}_2$		0.001

(c)

Fig. 12. Real data 3: Comparison of different estimations of the fundamental matrix. (a) $\mathbf{F}_1$. (b) $\mathbf{F}_3$. (c) Difference (in pixels).

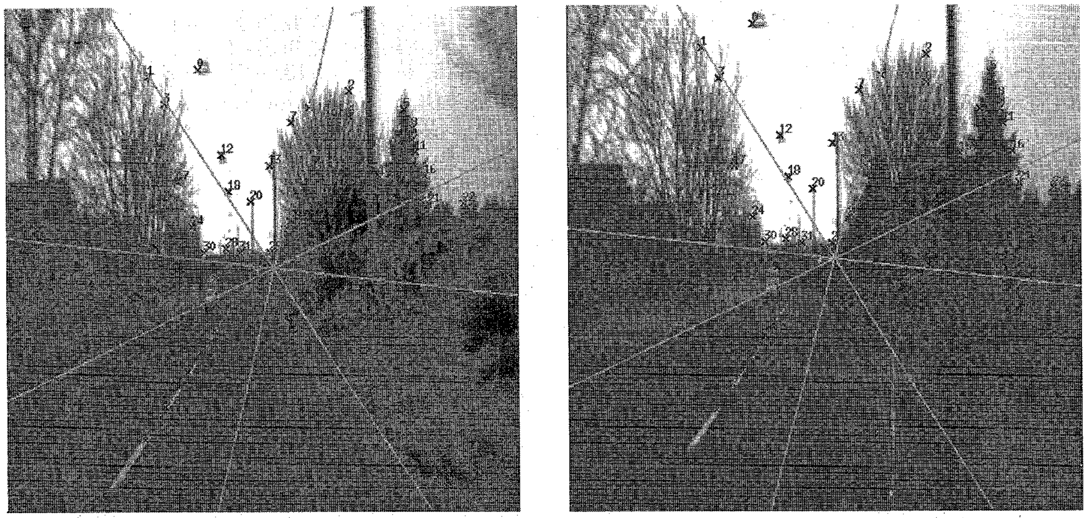
Table 8 compares the execution time of our particular implementation with the three criteria. They all start from the same initial guess which is provided by the eight-point algorithm. The execution time is measured on a Sun Sparc 10 workstation. In the last column, we have also shown the ratio of the CPU times spent on optimizing J_3 and J_2. This ratio increases when the number of points, denoted by N, increases. This is because the number of parameters to be estimated in J_3 (motion plus structure) increases with N, while that in J_2 remains constant.

7 Conclusion

In this paper, I have studied the relationship of the three best-known criteria, namely the one based on the distances between points and their corresponding epipolar lines (denoted by J_1), the one based on the gradient-weighted epipolar errors (denoted by J_2), and the one based the distances between the observed and reprojected points of the reconstruction (denoted by J_3).

Analytical analysis is carried out, which shows that, given a reasonable initial guess of the epipolar geometry, criteria J_2 and J_3 are equivalent when the epipoles are at infinity, and differ from each other only a little even when the epipoles are in the image, and that J_1 and J_2 are equivalent only when the epipoles are at infinity and when the observed object/scene has the same scale in the two images.

The bias of J_1 has been studied, which tends to make the object scales and the offsets of the epipoles with respect to data points in both images similar. The bias has been clearly observed, but is very small. Experiments



(a) (b)

Fig. 13. Real data 4: Outdoor scene, 51 point matches. Epipoles in the images.

$$\begin{bmatrix} 1.06e-6 & 2.93e-4 & -7.12e-2 \\ -2.93e-4 & 8.15e-6 & 7.65e-2 \\ 6.84e-2 & -7.98e-2 & 0.9889 \end{bmatrix}$$

(a)

$$\begin{bmatrix} 1.07e-6 & 2.93e-4 & -7.13e-2 \\ -2.93e-4 & 8.13e-6 & 7.65e-2 \\ 6.84e-2 & -7.98e-2 & 0.9890 \end{bmatrix}$$

(b)

	$\mathbf{F}_2$	$\mathbf{F}_3$
$\mathbf{F}_1$	0.111	0.112
$\mathbf{F}_2$		0.009

(c)

Fig. 14. Real data 4: Comparison of different estimations of the fundamental matrix. (a) $\mathbf{F}_1$. (b) $\mathbf{F}_3$. (c) Difference (in pixels).

with several thousand computer simulations and four sets of real data have confirmed our analysis. Since the optimization of J_3 is much more time consuming than the other two (about 50 times slower for 100 points), it is not recommended.

To summarize, I recommend the second criterion (gradient-weighted epipolar errors), which is actually a very good approximation to the third one. If a higher accuracy is required, the obtained estimation can be refined by using the third one, and this refining is much cheaper than if the third criterion would be directly used.

Note, however, that in the experiments with simulated data, the estimated fundamental matrix with J_1 is sometimes closer to the true fundamental matrix. The bias of J_1 may constrain the numerical minimization to behave better, which is a point to be studied in the future.

Appendix: General Formulation of the Three Criteria

The three criteria formulated in Section 3.1 assume that each point is corrupted by independent identically distributed Gaussian noise with mean zero and covariance matrix $\Lambda = \sigma^2 \operatorname{diag}(1, 1)$, where the knowledge of σ is not required. In this section, we reformulate the criteria if the covariance matrix for each point is different. Let us assume the covariance matrix for point $\mathbf{m}_i$ is $\Lambda_{\mathbf{m}_i}$. For the first criterion, the Euclidean distance of a point to the epipolar line in each image is replaced by the Mahalanobis distance. That is,

$$J_1 = \sum_{i=1}^{n} \left(\frac{1}{\tilde{\mathbf{m}}_i^T \mathbf{F}^T \mathbf{Z} \Lambda_{\mathbf{m}'_i} \mathbf{Z}^T \mathbf{F} \tilde{\mathbf{m}}_i} + \frac{1}{\tilde{\mathbf{m}}_i'^T \mathbf{F} \mathbf{Z} \Lambda_{\mathbf{m}_i} \mathbf{Z}^T \mathbf{F}^T \tilde{\mathbf{m}}'_i} \right) \left(\tilde{\mathbf{m}}_i'^T \mathbf{F} \tilde{\mathbf{m}}_i \right)^2$$

Similarly, the second criterion is reformulated as

$$J_2 = \sum_{i=1}^{n} \left(\frac{\left(\tilde{\mathbf{m}}_i'^T \mathbf{F} \tilde{\mathbf{m}}_i \right)^2}{\tilde{\mathbf{m}}_i^T \mathbf{F}^T \mathbf{Z} \Lambda_{\mathbf{m}'_i} \mathbf{Z}^T \mathbf{F} \tilde{\mathbf{m}}_i + \tilde{\mathbf{m}}_i'^T \mathbf{F} \mathbf{Z} \Lambda_{\mathbf{m}_i} \mathbf{Z}^T \mathbf{F}^T \tilde{\mathbf{m}}'_i} \right)$$

and the third one, as

TABLE 8
COMPARISON OF DIFFERENT CRITERIA IN TERMS OF CPU TIME IN SECONDS ON A SUN SPARC 10 WORKSTATION

Data set	J_1	J_2	J_3	J_3/J_2
1	0.55	0.55	53.12	96.6
2	0.18	0.21	5.86	27.9
3	0.24	0.27	13.74	50.9
4	0.21	0.26	7.66	29.5

$$J_3 = \sum_{i=1}^{n} \left(\Delta\mathbf{m}_i^T \Lambda_{\mathbf{m}_i}^{-1} \Delta\mathbf{m}_i + \Delta\mathbf{m}_i'^T \Lambda_{\mathbf{m}_i'}^{-1} \Delta\mathbf{m}_i' \right)$$

with $\Delta\mathbf{m}_i = \mathbf{m}_i - \hat{\mathbf{m}}_i$ and $\Delta\mathbf{m}_i' = \mathbf{m}_i' - \hat{\mathbf{m}}_i'$, where $\hat{\mathbf{m}}_i$ and $\hat{\mathbf{m}}_i'$ are the projections of the reconstructed points, as given in (7). Note that Kanatani [16] derives a more precise formula for the second criterion which involves the second order term in computing the variance.

Acknowledgments

This work benefits from many discussions with Andrew Zisserman during his stay at INRIA in 1996. The insightful comments from the reviewers have made several statements in this paper more precise. A short version of this paper, entitled "Understanding the Relationship Between the Optimization Criteria in Two-View Motion Analysis," appeared in *Proc. Sixth Int'l Conf. Computer Vision*, Bombay, India, 4–7 Jan. 1998. The author was with INRIA Sophia-Antipolis, France. The initial version of this paper was prepared when the author was on sabbatical leave in 1997 at ATR Human Information Processing Research Laboratories, Kroto, Japan. Thanks go to Shigeru Akamatsu for financial support.

References

[1] J. Aggarwal and N. Nandhakumar, "On the Computation of Motion From Sequences of Images—a Review," *Proc. IEEE*, vol. 76, pp. 917–935, Aug. 1988.

[2] T. Huang and A. Netravali, "Motion and Structure From Feature Correspondences: A Review," *Proc. IEEE*, vol. 82, pp. 252–268, Feb. 1994.

[3] O. Faugeras, "What Can Be Seen in Three Dimensions With an Uncalibrated Stereo Rig," *Proc. Second European Conf. Computer Vision*, pp. 563–578, Santa Margherita Ligure, Italy, May 1992.

[4] R. Hartley, R. Gupta, and T. Chang, "Stereo From Uncalibrated Cameras," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 761–764, Urbana-Champaign, Ill., June 1992.

[5] Z. Zhang, "Determining the Epipolar Geometry and Its Uncertainty: A Review," Tech. Rep. 2927, INRIA Sophia-Antipolis, France, July 1996. *Int'l J. Computer Vision*, vol. 27, no. 2, pp. 161-195, 1998.

[6] Z. Zhang, "On the Epipolar Geometry Between Two Images With Lens Distortion," *Int'l Conf. Pattern Recognition*, vol. 1, pp. 407–411, Vienna, Aug. 1996.

[7] G. Stein, "Lens Distortion Calibration Using Point Correspondences," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 602–608, Puerto Rico, June 1997.

[8] H. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene From Two Projections," *Nature*, vol. 293, pp. 133–135, 1981.

[9] R. Tsai and T. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects With Curved Surface," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 13–26, Jan. 1984.

[10] Q.-T. Luong, *Matrice Fondamentale et Calibration Visuelle sur l'Environnement-Vers une plus grande autonomie des systè*mes robotiques. PhD thesis, Université de Paris-Sud, Centre d'Orsay, Dec. 1992.

[11] O. Faugeras, "Stratification of 3-D Vision: Projective, Affine, and Metric Representations," *J. Optical Soc. Am. A*, vol. 12, pp. 465–484, Mar. 1995.

[12] Q.-T. Luong and O.D. Faugeras, "The Fundamental Matrix: Theory, Algorithms and Stability Analysis," *Int'l J. Computer Vision*, vol. 17, pp. 43–76, Jan. 1996.

[13] R. Hartley, "In Defense of the Eight-Point Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580–593, June 1997.

[14] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, Mass.: MIT Press, 1993.

[15] L. Shapiro, A. Zisserman, and M. Brady, "3D Motion Recovery via Affine Epipolar Geometry," *Int'l J. Computer Vision*, vol. 16, pp. 147-182, 1995.

[16] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Amsterdam: Elsevier, 1996.

[17] J. Weng, N. Ahuja, and T. Huang, "Optimal Motion and Structure Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 864–884, Sept. 1993.

[18] C. Tomasi and T. Kanade, "Shape and Motion From Image Streams Under Orthography: A Factorization Method," *Int'l J. Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[19] R. Hartley, "Euclidean Reconstruction From Uncalibrated Views," *Applications of Invariance in Computer Vision*, J. Mundy and A. Zisserman, eds., vol. 825, Lecture Notes in Computer Science. Berlin: Springer-Verlag, 1993, pp. 237–256.

[20] J. Oliensis, "A Multi-Frame Structure From Motion Algorithm Under Perspective Projection," tech. rep., NEC Research Institute, Apr. 1997. (Revised version, Mar. 1998.)

[21] G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition*. Kluwer Academic Publishers, 1996.

[22] Z. Zhang and G. Xu, "A General Expression of the Fundamental Matrix for Both Perspective and Affine Cameras," *Proc. 15th Int'l Joint Conf. Artificial Intelligence*, pp. 1,502–1,507, Nagoya, Japan, Aug. 1997.

[23] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry," *Artificial Intelligence J.*, vol. 78, pp. 87–119, Oct. 1995.

[24] I. Reid and D. Murray, "Active Tracking of Foveated Feature Clusters Using Affine Structure," *Int'l J. Computer Vision*, vol. 18, no. 1, pp. 41–60, 1996.

Zhengyou Zhang received the BS degree in electronic engineering from the University of Zhejiang, China, in 1985, the MS in computer science from the University of Nancy, France, in 1987, the PhD degree in computer science from the University of Paris XI, France, in 1990, and the Doctor of Science (*Habilitation a diriger des recherches*) diploma from the University of Paris XI, in 1994. He has been with INRIA (French National Institute for Research in Computer Science and Control) for 11 years: a research assistant from 1987 to 1990, a research scientist from 1990 to 1991, and a senior research scientist from 1991 until he joined Microsoft Research, Redmond, USA, in March 1998. In 1996-1997, he spent one year sabbatical as an Invited Researcher at the Advanced Telecommunications Research Institute International (ATR), Human Information Processing Research Laboratories, Kyoto, Japan. He is also a guest research professor in the Chinese Academy of Sciences since 1994, and a part-time professor in Northern Jiaotong University (Beijing, China) since 1996. His current research interests include 3D computer vision, dynamic scene analysis, vision and graphics, facial image analysis, and visual learning. He is an associate editor of the *International Journal of Pattern Recognition and Artificial Intelligence* and an action editor of *Videre: A Journal of Computer Vision Research*. He is a senior member of IEEE, and has coauthored the following books: *3D Dynamic Scene Analysis: A Stereo Based Approach* (Springer, Berlin, Heidelberg, 1992); *Epipolar Geometry in Stereo, Motion and Object Recognition* (Kluwer Academic Publishers, 1996); and *Computer Vision* (textbook in Chinese, Chinese Academy of Sciences, 1998).