A General Expression of the Fundamental Matrix for Both Perspective and Affine Cameras

Zhengyou Zhang*
ATR Human Information Processing Res. Lab.
2-2 Hikari-dai, Seika-cho, Soraku-gun
Kyoto 619-02 Japan
zzhang©hip.atr.co.jp

Gang Xu
Department of Computer Science
Ritsumeikan University
Kusatsu, Shiga 525, Japan
xu©cs.ritsumei.ac.jp

Abstract

This paper addresses the recovery of structure and motion from two uncalibrated images of a scene under full perspective or under affine projection. Epipolar geometry, projective reconstruction, and affine reconstruction are elaborated in a way such that everyone having knowledge of linear algebra can understand the discussion without difficulty. A general expression of the fundamental matrix is derived which is valid for any projection model without lens distortion (including full perspective and affine camera). A new technique for affine reconstruction from two affine images is developed, which consists in first estimating the affine epipolar geometry and then performing a triangulation for each point match with respect to an implicit common affine basis. This technique is very efficient.

Keywords: Motion Analysis, Epipolar Geometry, Uncalibrated Images, Non-Metric Vision, 3D Reconstruction, Fundamental Matrix.

1 Introduction

Since the work of Koenderink and van Doom [Koenderink and van Doom, 1991] on affine structure from motion and that of Forsyth et al. [Forsyth *et al,* 1991] on invariant description, the development of non-metric vision has attracted quite a number of researchers [Faugeras, 1992; Hartley *et* a/., 1992] (to cite a few). We can find a range of applications: object recognition, 3D reconstruction of scenes, image matching, visual navigation, motion segmentation, image synthesis, etc.

This paper addresses the recovery of structure and motion from two uncalibrated images of a scene under full perspective or under affine projection. There is already a large amount of work reported in the literature [Faugeras, 1992; 1995; Hartley *et al,* 1992; Shashua, 1994; Zisserman, 1992], and it is known that the structure of the scene can only be recovered up to a projective transformation for two perspective images and up to an affine transformation for two affine images. We cannot obtain any metric information from a projective or affine structure: measurements of lengths and angles do not make sense. However, projective or affine structure still contains rich information, such as coplanarity, collinearity and ratios. The latter is sometimes sufficient for artificial systems, such as robots, to perform tasks such as navigation and object recognition. We believe that these results are important to the whole community of computer vision. However, they are usually stated using tools from Projective Geometry, which is not accessible to most researchers. One objective of this paper is to present these results in a way such that everyone having knowledge of linear algebra can understand without difficulty.

Other contributions of this paper are the following:

- A general expression of the fundamental matrix for any projection model is presented. Previously, the fundamental matrix is formulated separately for full perspective and affine projection. Our formula is valid for both.
- A new efficient technique for affine reconstruction from two affine images is developed. We decompose the problem into two subproblems: recovery of affine epipolar geometry and 3D reconstruction with respect to an implicit affine basis.

This paper is organized as follows. Section 2 presents different camera projection models. Section 3 derives an expression of fundamental matrix which is valid for any projection model (ignoring the lens distortion). Section 4 describes the projective reconstruction from two uncalibrated perspective images. In Section 5, we first specialize the general fundamental matrix to the case of affine cameras and then show that only affine structure can be recovered, and finally a new technique for affine reconstruction is proposed.

2 Perspective Projection and its Approximations

If the lens distortion can be ignored, the projection from a space point M = $[X, Y, Z]^T$ to its image point m =

*On leave from INRIA Sophia-Antipolis, France

$[x, y]^T$ can be represented linearly by

$$[U, V, S]^T = \mathbf{P}[X, Y, Z, 1]^T , \quad (1)$$

where $x = U/S$, and $y = V/S$ if $S \neq 0$, and P is the 3 x 4 projection matrix which varies with projection model and with the coordinate system in which space points M are expressed. Given a vector $\mathbf{x} = [x, y, \cdots]^T$, we use $\tilde{\mathbf{x}}$ to denote its augmented vector by adding 1 as the last-element, i.e., $\tilde{\mathbf{x}} = [x, y, \cdots, 1]^T$. Now we can rewrite the above formula concisely as

$$s\tilde{\mathbf{m}} = \mathbf{P}\tilde{\mathbf{M}} , \quad (2)$$

where s = 5 is an arbitrary nonzero scalar.

The projection matrix corresponding to the full perspective is of the form:

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} , \quad (3)$$

which is defined up to a scalar factor. This implies that there are only 11 degrees of freedom in a full perspective projection matrix. In terms of the intrinsic and extrinsic parameters of a camera, P can be decomposed as P = A[R t], where A is a 3 x 3 matrix defined by the intrinsic parameters (see e.g., [Faugeras, 1993]), and (R, t) is the rotation and translation (extrinsic parameters) relating the world coordinate system to the camera coordinate system.

The affine camera, introduced by Mundy and Zisserman [Mundy and Zisserman, 1992] as a generalization of orthographic, weak perspective and paraperspective projections, has the following form:

$$\mathbf{P}_A = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & P_{34} \end{bmatrix} . \quad (4)$$

The elements P_{31},P_{32} and *P33* are equal to 0. This is an approximation to the full perspective, and works quite well when object size and depth is small compared with the distance between the camera and object.

3 Fundamental Matrix for Any Projection Model

Consider now the case of two images whose projection matrices are P and P', respectively (the prime ' is used to indicate a quantity related to the second image). A point m in the first image is matched to a point m' in the second image. From the camera projection model (2), we have

$$s\tilde{\mathbf{m}} = \mathbf{P}\tilde{\mathbf{M}}' \quad \text{and} \quad s'\tilde{\mathbf{m}}' = \mathbf{P}'\tilde{\mathbf{M}}' .$$

An image point m' defines actually an optical ray, on which every space point $\tilde{\mathbf{M}}'$ projects on the second image at $\tilde{\mathbf{m}}'$. This optical ray can be written in parametric form as

$$\tilde{\mathbf{M}}' = s'\mathbf{P}'^{+}\tilde{\mathbf{m}}' + \mathbf{p}'^{\perp} , \quad (5)$$

where $\mathbf{P}'^{+}$ is the pseudo-inverse of matrix P':

$$\mathbf{P}'^{+} = \mathbf{P}'^{T}(\mathbf{P}'\mathbf{P}'^{T})^{-1} , \quad (6)$$

and $\mathbf{p}'^{\perp}$ is any 4-vector that is perpendicular to all the *row* vectors of P' i.e., $\mathbf{P}'\mathbf{p}'^{\perp} = \mathbf{0}$. Thus, $\mathbf{p}'^{\perp}$ is a null vector of P'. As a matter of fact, $\mathbf{p}'^{\perp}$ indicates the position of the optical center (to which all optical rays converge). We show later how to determine $\mathbf{p}'^{\perp}$. For a particular value s', equation (5) corresponds to a point on the optical ray defined by m'. Equation (5) is easily justified by projecting M' onto the second image, which indeed gives m'.

Similarly, an image point m in the first image defines also an optical ray. Requiring the two rays to intersect in space implies that a point M' corresponding to a particular *s'* in (5) must project onto the first image at m, that is

$$s\tilde{\mathbf{m}} = s'\mathbf{P}\mathbf{P}'^{+}\tilde{\mathbf{m}}' + \mathbf{P}\mathbf{p}'^{\perp} .$$

Performing a cross product with $\mathbf{P}\mathbf{p}'^{\perp}$ yields

$$s(\mathbf{P}\mathbf{p}'^{\perp}) \times \tilde{\mathbf{m}} = s'(\mathbf{P}\mathbf{p}'^{\perp}) \times (\mathbf{P}\mathbf{P}'^{+}\tilde{\mathbf{m}}') .$$

Eliminating *s* and *s'* by multiplying $\tilde{\mathbf{m}}^T$ from the left (equivalent to a dot product), we have

$$\tilde{\mathbf{m}}^T\mathbf{F}\tilde{\mathbf{m}}' = 0 , \quad (7)$$

where F is a 3 x 3 matrix, called *fundamental matrix:*

$$\mathbf{F} = [\mathbf{P}\mathbf{p}'^{\perp}]_\times \mathbf{P}\mathbf{P}'^{+} . \quad (8)$$

Equation (7) is the well-known epipolar equation [Hartley *et al.*, 1992; Faugeras *et al.*, 1992; Luong and Faugeras, 1996], but the form of the fundamental matrix (8) is general and, to our knowledge, is not yet reported in the literature. It does not assume any particular projection model. Indeed, equation (8) only makes use of the pseudo-inverse of the projection matrix (which is valid for full perspective as well as for affine cameras). In [Luong and Faugeras, 1996], for example, the fundamental matrix is formulated only for full perspective, because it involves the inverse of the first 3 x 3 submatrix of P which is not invertible for affine camera. In [Zisserman, 1992], a separate fundamental matrix is given for affine cameras. Our formula (8) works for both. We will specialize it for affine cameras in Sect. 5.1.

The fundamental matrix F recapitulates all geometric information between two images. The nine elements of F are not independent from each other. In fact, F has only 7 degrees of freedom. This can be seen as follows. First F is defined up to a scale factor because if F is multiplied by any nonzero scalar, the new F still satisfy (7). Second, the rank of F is at most 2, i.e., det(F) = 0. This is because the determinant of the antisymmetric matrix $[\mathbf{P}\mathbf{p}'^{\perp}]_\times$ is equal to zero. Another thing to mention is that the two images play a symmetric role. Indeed, (7) can also be rewritten as $\tilde{\mathbf{m}}'^{T}\mathbf{F}^{T}\tilde{\mathbf{m}} = \mathbf{0}$. It can be shown that $\mathbf{F}^{T} = [\mathbf{P}'\mathbf{p}^{\perp}]_\times \mathbf{P}'\mathbf{P}^{+}$.

The vector $\mathbf{p}'^{\perp}$ still needs to be determined. We first note that such a vector must exist because the difference between the row dimension and the column dimension is

one, and that the row vectors are generally independent from each other. Indeed, one way to obtain $\mathbf{p}'^{\perp}$ is

$$\mathbf{p}'^{\perp} = (\mathbf{I} - \mathbf{P}'^{+}\mathbf{P}')\boldsymbol{\omega}\,, \tag{9}$$

where $\boldsymbol{\omega}$ is an arbitrary 4-vector. To show that $\mathbf{p}'^{\perp}$ is perpendicular to each row of P', we multiply $\mathbf{p}'^{\perp}$ by P' from the left: $\mathbf{P}'\mathbf{p}'^{\perp} = (\mathbf{P}' - \mathbf{P}'\mathbf{P}'^{T}(\mathbf{P}'\mathbf{P}'^{T})^{-1}\mathbf{P}')\boldsymbol{\omega} = \mathbf{0}$, which is indeed a zero vector. The action of $\mathbf{I} - \mathbf{P}'^{+}\mathbf{P}'$ is to transform an arbitrary vector to a vector that is perpendicular to every row vector of P'. If P' is of rank 3 (which is usually the case), then $\mathbf{p}'^{\perp}$ is unique up to a scale factor. See [Xu and Zhang, 1996] for more details.

4 Projective Reconstruction

We show in this section how to estimate the position of a point in space, given its projections in two images whose epipolar geometry is known. The problem is known as *3D reconstruction* in general, and *triangulation* in particular. We assume that the fundamental matrix between the two images is known (e.g., computed with the methods described in [Zhang *et al.*, 1995]), and we say that they are *weakly calibrated.*

4.1 Fundamental Matrix for Full Perspective

We now derive a usual form of fundamental matrix for full perspective from the general expression (8). Let A and A' be the 3 x 3 matrices containing the intrinsic parameters of the first and second image. Without loss of generality, we choose the second camera coordinate system as the world coordinate system. Then, the camera projection matrices are P = A [R t] and P' = A' [I 0] , where (R, t) is the rotation and translation relating the two camera coordinate systems, and I is the 3 x 3 identity matrix and 0 is a zero 3-vector.

It is not difficult to see that

$$\mathbf{P}'^{+} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0}^T \end{bmatrix} \mathbf{A}'^{-1}\,, \quad \text{and} \quad \mathbf{p}'^{\perp} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}.$$

This yields:

$$\begin{aligned} \mathbf{P}\mathbf{p}'^{\perp} &= \mathbf{A}[\mathbf{R}\ \mathbf{t}]\begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \mathbf{A}\mathbf{t}\,, \\ \mathbf{P}\mathbf{P}'^{+} &= \mathbf{A}'[\mathbf{I}\ \mathbf{0}]\begin{bmatrix} \mathbf{I} \\ \mathbf{0}^T \end{bmatrix} \mathbf{A}'^{-1} = \mathbf{A}\mathbf{R}\mathbf{A}'^{-1}\,. \end{aligned}$$

Using the property $(\mathbf{A}\mathbf{x}) \times (\mathbf{A}\mathbf{y}) = \det(\mathbf{A})\mathbf{A}^{-T}(\mathbf{x} \times \mathbf{y}),\ \forall \mathbf{x}, \mathbf{y}$, and the general expression of the fundamental matrix (8), we have

$$\mathbf{F} = [\mathbf{P}\mathbf{p}'^{\perp}]_{\times}\mathbf{P}\mathbf{P}'^{+} = [\mathbf{A}\mathbf{t}]_{\times}\mathbf{A}\mathbf{R}\mathbf{A}'^{-1} \cong \mathbf{A}^{-T}[\mathbf{t}]_{\times}\mathbf{R}\mathbf{A}'^{-1}\,,$$

where $\cong$ means "equal" up to a scale factor. The above equation is the usual form of the fundamental matrix (see e.g., [Luong and Faugeras, 1996]).

4.2 Projective Reconstruction

In the calibrated case, a 3D structure can be recovered from two images only up to a rigid transformation and an unknown scale factor (this transformation is also known as a *similarity),* because we can choose an arbitrary coordinate system as a world coordinate system (although one usually chooses it to coincide with one of the camera coordinate systems). Similarly, in the uncalibrated case, a 3D structure can only be performed up to a projective transformation of the 3D space [Faugeras, 1992; Hartley *et al.*, 1992; Maybank, 1992; Faugeras, 1995]. A 4 x 4 nonsingular matrix H defines a linear transformation from one projective point to another, and is called the *projective transformation.* The matrix H, of course, is also defined up to a nonzero scale factor, and we write $\rho\tilde{\mathbf{y}} = \mathbf{H}\tilde{\mathbf{x}}$, if $\tilde{\mathbf{x}}$ is mapped to $\tilde{\mathbf{y}}$ by H. Here *p* is a nonzero scale factor.

Now we are given two perspective images of a scene. The intrinsic parameters of the images are unknown. Assume that the true camera projection matrices are P and P'. From (8), we have the following relation $\mathbf{F} = [\mathbf{P}\mathbf{p}'^{\perp}]_{\times}\mathbf{P}\mathbf{P}'^{+}$. Given 8 or more point matches in general position, the fundamental matrix F can be uniquely determined from two images. We are now interested in recovering P and P' from F, and once they are recovered, triangulation can be conducted to reconstruct the scene in 3D space.

Proposition 1 *Given two perspective images of a scene whose epipolar geometry (i.e., the fundamental matrix) is known, the camera projection matrices can only be determined up to an unknown projective transformation.*

The proof of this proposition is omitted due to space limitation. The consequence of this proposition is the following: if P and P' are two camera projection matrices consistent with the fundamental matrix F, then $\hat{\mathbf{P}} = \mathbf{P}\mathbf{H}$ and $\hat{\mathbf{P}}' = \mathbf{P}'\mathbf{H}$ are also consistent with the same F, where H is any projective transformation of the 3D space. If the true structure is M, then the structure reconstructed from image points is $\mathbf{H}^{-1}\tilde{\mathbf{M}}$, i.e., up to a projective transformation. This is because $\hat{\mathbf{P}}\mathbf{H}^{-1}\tilde{\mathbf{M}} = \mathbf{P}\tilde{\mathbf{M}}$ gives the exact projection for the first image; the same is true for the second image. Although the above result has been known for several years, we believe that it is easier to understand our discussion than what has been presented in the literature.

In order to reconstruct points in 3D space, we need to compute the camera projection matrices from the fundamental matrix F with respect to a projective basis, which can be arbitrary because of Proposition 1. One way is to use a canonical representation [Luong and VieVille, 1994; Beardsley *et al.*, 1994]:

$$\mathbf{P} = [\mathbf{M}\ \ \mathbf{e}] \quad \text{and} \quad \mathbf{P}' = [\mathbf{I}\ \ \mathbf{0}]\,,$$

where $\mathbf{e}$ is the epipole in the first image ($\mathbf{F}^T\mathbf{e} = \mathbf{0}$) and $\mathbf{M} = -\frac{1}{\|\mathbf{e}\|^2}[\mathbf{e}]_{\times}\mathbf{F}$.

5 Affine Reconstruction

This section deals with two images taken by an affine camera at two different instants or by two different affine cameras.

5.1 Affine Fundamental Matrix

In the case of a general affine camera, the projection matrix (4) can be rewritten as

$$\mathbf{P}_A = \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \mathbf{0}_3^T \end{bmatrix} \quad \mathbf{p}_4 \quad \text{where } \mathbf{p}_4 = [P_{14}, P_{24}, P_{34}]^T.$$

We now derive the specific tundamental matrix for affine cameras from the general form (8).

For any affine camera, we can construct $\mathbf{p}'^{\perp}$ as

$$\mathbf{p}'^{\perp} = \frac{1}{\|\mathbf{p}'_1 \times \mathbf{p}'_2\|} \begin{bmatrix} (\mathbf{p}'_1 \times \mathbf{p}'_2) \\ 0 \end{bmatrix} \equiv \frac{1}{\|\mathbf{p}'_3\|} \begin{bmatrix} \mathbf{p}'_3 \\ 0 \end{bmatrix}.$$

Here, we have defined $\mathbf{p}'_3 = \mathbf{p}'_1 \times \mathbf{p}'_2$. From $\mathbf{p}'^T_1 \mathbf{p}'_3 = 0$ and $\mathbf{p}'^T_2 \mathbf{p}'_3 = 0$, we can verify that $\mathbf{p}'^{\perp}$ is indeed perpendicular to $\mathbf{P}'$. Multiplying $\mathbf{P}$ with $\mathbf{p}'^{\perp}$ yields

$$\mathbf{P}\mathbf{p}'^{\perp} = \begin{bmatrix} \mathbf{p}_1^T \mathbf{p}'_3 \\ \mathbf{p}_2^T \mathbf{p}'_3 \\ 0 \end{bmatrix}. \tag{10}$$

Let us assume $\mathbf{P}'^{+} = [\mathbf{Q}^T \ \mathbf{q}_4]^T$, where $\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{q}_2 \ \mathbf{q}_3]$ is a 3×3 matrix and $\mathbf{q}_4$ is a 3-vector. Since

$$\mathbf{P}'\mathbf{P}'^{+} = \begin{bmatrix} \mathbf{p}'^T_1 \mathbf{Q} \\ \mathbf{p}'^T_2 \mathbf{Q} \\ \mathbf{0}_3^T \end{bmatrix} + \mathbf{p}'_4 \mathbf{q}_4^T = \mathbf{I}_3,$$

$\mathbf{q}_4$ can be uniquely determined: $\mathbf{q}_4 = [0, 0, 1/P'_{34}]^T$. The constraint for matrix $\mathbf{Q}$ is then

$$\begin{bmatrix} \mathbf{p}'^T_1 \\ \mathbf{p}'^T_2 \end{bmatrix} \mathbf{Q} = \begin{bmatrix} 1 & 0 & -P'_{14}/P'_{34} \\ 0 & 1 & -P'_{24}/P'_{34} \end{bmatrix}. \tag{11}$$

It is evident that Q cannot be uniquely determined. For us, any Q that satisfies the above equation suffices.

Now substituting these matrices for (8), we have

$$\mathbf{F}_A = \begin{bmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{12}$$

where a^ are related to the coefficients of the camera projection matrices. The fact that the affine fundamental matrix has the form of (12) is mentioned in [Zisserman, 1992]. Defined up to a scale factor, F>i has only 4 degrees of freedom. The corresponding points in the two images must satisfy the following relation, called the *affine epipolar equation:*

$$\tilde{\mathbf{m}}^T \mathbf{F}_A \tilde{\mathbf{m}}' = 0. \tag{13}$$

Expanding the epipolar equation, the left-hand side is a first-order polynomial of the image coordinates, and we have

$$a_{13}x + a_{23}y + a_{31}x' + a_{32}y' + a_{33} = 0. \tag{14}$$

It means that the epipolar lines are parallel everywhere in the image, and the orientations of the parallel epipolar lines are completely determined by the affine fundamental matrix.

5.2 Affine Reconstruction

Given a sufficient number of point matches (at least 4) between two images, the affine fundamental matrix *FA* can be estimated (see [Shapiro *et a*/., 1994]). We are now interested in recovering *PA* and P'A from FA, and once they are recovered, the structure can be redressed in 3D space.

Since P_A and P_A are defined up to a scale factor, without loss of generality, we assume $P_{34} = P'_{34} = 1$. Then the relation between a 3D point and its 2D image is given by $\tilde{\mathbf{m}} = \mathbf{P}_A \tilde{\mathbf{M}}$ and $\tilde{\mathbf{m}}' = \mathbf{P}'_A \tilde{\mathbf{M}}$. Note that there is no more scale factor in the above equations. From the affine epipolar equation (13), it is easy to obtain

$$\tilde{\mathbf{M}}^T \underbrace{\mathbf{P}_A^T \mathbf{F}_A \mathbf{P}'_A}_{\mathbf{S}} \tilde{\mathbf{M}} = 0 \text{ with } \mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & S_{14} \\ 0 & 0 & 0 & S_{24} \\ 0 & 0 & 0 & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix}, \tag{15}$$

and

$$\begin{aligned} S_{14} &= a_{13}P_{11} + a_{23}P_{21}, & S_{24} &= a_{13}P_{12} + a_{23}P_{22}, \\ S_{34} &= a_{13}P_{13} + a_{23}P_{23}, & S_{41} &= a_{31}P'_{11} + a_{32}P'_{21}, \\ S_{42} &= a_{31}P'_{12} + a_{32}P'_{22}, & S_{43} &= a_{31}P'_{13} + a_{32}P'_{23}, \\ S_{44} &= a_{13}P_{14} + a_{23}P_{24} + a_{31}P'_{14} + a_{32}P'_{24} + a_{33}. \end{aligned}$$

Equation (15) becomes

$$(S_{14} + S_{41})X + (S_{24} + S_{42})Y + (S_{34} + S_{43})Z + S_{44} = 0.$$

Since this equation should be true for all points, the four coefficients must be all zero, which leads to

$$\begin{aligned} a_{13}P_{11} + a_{23}P_{21} + a_{31}P'_{11} + a_{32}P'_{21} &= 0 \\ a_{13}P_{12} + a_{23}P_{22} + a_{31}P'_{12} + a_{32}P'_{22} &= 0 \\ a_{13}P_{13} + a_{23}P_{23} + a_{31}P'_{13} + a_{32}P'_{23} &= 0 \\ a_{13}P_{14} + a_{23}P_{24} + a_{31}P'_{14} + a_{32}P'_{24} &= -a_{33}. \end{aligned}$$

We thus have 4 simple constraints on the coefficients of the projection matrices, which is consistent with the number of the degrees of freedom in an affine fundamental matrix. Writing them in matrix form gives:

$$\begin{bmatrix} a_{13} \\ a_{23} \\ a_{31} \\ a_{32} \end{bmatrix}^T \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P'_{11} & P'_{12} & P'_{13} & P'_{14} \\ P'_{21} & P'_{22} & P'_{23} & P'_{24} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -a_{33} \end{bmatrix}^T. \tag{16}$$

We now show the following proposition.

Proposition 2 *Given two images of a scene taken by an affine camera, the 3D structure of the scene can be reconstructed up to an unknown affine transformation as soon as the epipolar geometry (i.e., the affine fundamental matrix) between the two images is known.*

Let the 3D structure corresponding to the true camera projection matrices $\mathbf{P}_A$ and $\mathbf{P}'_A$ be M. We need to show that the new structure $\tilde{\tilde{\mathbf{M}}} = \mathbf{H}_A^{-1}\tilde{\mathbf{M}}$ is still consistent with the same sets of image points (i.e., with the affine fundamental matrix), where

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

is an affine transformation of the 3D space, $\mathbf{A}$ is a 3×3 matrix, and $\mathbf{t}$ is a 3-vector. It follows that $\widehat{\mathtt{M}} = \mathbf{A}\mathtt{M} + \mathbf{t}$.

Proof: The camera projection matrices corresponding to the new structure $\widehat{\mathtt{M}}$ are:

$$\widehat{\mathbf{P}}_A = \mathbf{P}_A\mathbf{H}_A \quad \text{and} \quad \widehat{\mathbf{P}}'_A = \mathbf{P}'_A\mathbf{H}_A \,.$$

We only need to show that the new affine projection matrices $\widehat{\mathbf{P}}_A$ and $\widehat{\mathbf{P}}'_A$ satisfy the same relation as (16), where P_{ij} and P'_{ij} should be replaced by $\widehat{P}_{ij}$ and $\widehat{P}'_{ij}$. Indeed, multiplying both sides of (16) by $\mathbf{H}_A$ from the right, i.e.,

$$\begin{bmatrix} a_{13} \\ a_{23} \\ a_{31} \\ a_{32} \end{bmatrix}^T \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P'_{11} & P'_{12} & P'_{13} & P'_{14} \\ P'_{21} & P'_{22} & P'_{23} & P'_{24} \end{bmatrix} \mathbf{H}_A = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -a_{33} \end{bmatrix}^T \mathbf{H}_A$$

yields

$$\begin{bmatrix} a_{13} \\ a_{23} \\ a_{31} \\ a_{32} \end{bmatrix}^T \begin{bmatrix} \widehat{P}_{11} & \widehat{P}_{12} & \widehat{P}_{13} & \widehat{P}_{14} \\ \widehat{P}_{21} & \widehat{P}_{22} & \widehat{P}_{23} & \widehat{P}_{24} \\ \widehat{P}'_{11} & \widehat{P}'_{12} & \widehat{P}'_{13} & \widehat{P}'_{14} \\ \widehat{P}'_{21} & \widehat{P}'_{22} & \widehat{P}'_{23} & \widehat{P}'_{24} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -a_{33} \end{bmatrix}^T .$$

This completes the proof. ■

Because of the above result, there is no unique determination of $\mathbf{P}_A$ and $\mathbf{P}'_A$ from $\mathbf{F}_A$ based on (16). One simply way is the following:

$$\mathbf{P}_A = \begin{bmatrix} a_{31} & 0 & 0 & 0 \\ 0 & a_{32} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{P}'_A = \begin{bmatrix} -a_{13} & 0 & a_{32} & -a_{33}/a_{31} \\ 0 & -a_{23} & -a_{31} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Once $\mathbf{P}_A$ and $\mathbf{P}'_A$ are determined from $\mathbf{F}_A$, the 3D structure can be uniquely recovered. Let $\mathbf{m} = [u, v]^T$ and $\mathbf{m}' = [u', v']^T$ be the observed image points which have been matched between the two images. Let $\mathtt{M} = [X, Y, Z]^T$ be the corresponding space point to be estimated, which projects on to the two cameras $\mathbf{P}_A$ and $\mathbf{P}'_A$ as

$$\widehat{\mathbf{m}} = \begin{bmatrix} a_{31}X \\ a_{32}Y \end{bmatrix} \text{ and } \widehat{\mathbf{m}}' = \begin{bmatrix} -a_{13}X + a_{32}Z - a_{33}/a_{31} \\ -a_{23}Y - a_{31}Z \end{bmatrix} ,$$

Because the observations are made in image plane and the noise level can be reasonably assumed to be the same for each extracted image point, a physically meaningful criterion is to minimize, over the structure parameter $\mathtt{M}$, the point-to-point distances between the observed locations ($\mathbf{m}$ and $\mathbf{m}'$) and the image projections of the estimated scene structure ($\widehat{\mathbf{m}}$ and $\widehat{\mathbf{m}}'$): $\mathcal{F}(\mathtt{M}) = \|\mathbf{m} - \widehat{\mathbf{m}}\|^2 + \|\mathbf{m}' - \widehat{\mathbf{m}}'\|^2$. The solution is obtained by setting the derivative of $\mathcal{F}(\mathtt{M})$ with respect to $\mathtt{M}$ to zero, i.e., $\partial\mathcal{F}(\mathtt{M})/\partial\mathtt{M} = \mathbf{0}$. This yields a vector equation $\mathbf{B}\mathtt{M} = \mathbf{b}$, where $\mathbf{B}$ is a 3×3 matrix and $\mathbf{b}$ is a 3D vector. The 3D reconstructed point is then given by $\mathtt{M} = \mathbf{B}^{-1}\mathbf{b}$.

5.3 Experimental Results

We have tested the proposed technique with computer simulated data under affine projection, and very good results have been obtained. In this subsection, we show the result with data obtained *under full perspective projection* but treated as if it were obtained under affine projection.

The parameters of the camera set-up are taken from a real stereo vision system. The two cameras are separated by an almost pure translation (the rotation angle is only 6 degrees). The baseline is about 350 mm (millimeters). An object of size 400 x 250 x 300 mm^3 is placed in front of the cameras at a distance of about 2500 mm. Two images of this object under full perspective projection are generated as shown in Fig. 1. Line segments are drawn only for visual effect, and only the endpoints (12 points) are used in our experiment. The image resolution is 512x512 $pixels^2$, and the projection of the object occupies a surface of about 130 x 120 $pixels^2$.

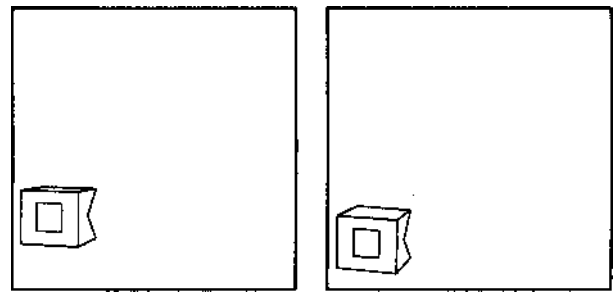

Figure 1: Two perspective images of a synthetic object

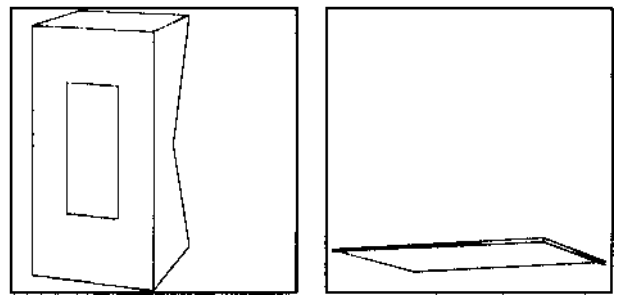

Figure 2: Two orthographic views of the affine reconstruction

The method described in [Shapiro *et* a/., 1994] is used to compute the affine epipolar geometry, and the root of the mean point-to-point distance is 0.065 pixels. This implies that even the images are perspective, their relation can be quite reasonably described by the affine epipolar geometry. The affine reconstruction result obtained with the technique described in this paper is shown in Fig. 2.

In order to have a quantitative measure of the reconstruction quality, we estimate, in a least-squares sense,

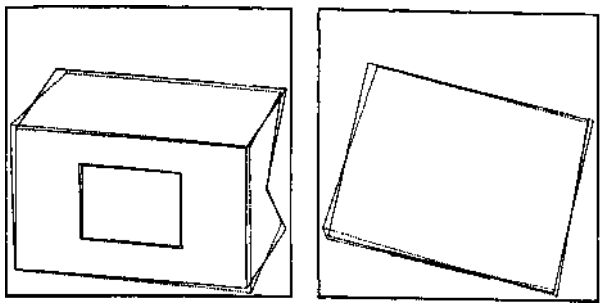

Figure 3: Two orthographic views of the superposition of the original 3D data (in solid lines) and the transformed affine reconstruction (in dashed lines)

the affine transformation which brings the set of affinely reconstructed points to the original set of 3D points. The root of the mean of the squared distances between the corresponding points is 10.4 mm, thus the error is less than 5%. The superposition of the two sets of data is shown in Fig. 3. It is interesting to observe that the reconstruction of the near part is larger than the real size while that of the distant part is smaller. This is because the assumption of an affine camera ignores the perspective distortion in the image.

6 Conclusion

We have addressed in this paper the problem of determining the structure and motion from two uncalibrated images of a scene under full perspective or under affine projection. Epipolar geometry, projective reconstruction and affine reconstruction have been elaborated in a way such that everyone having knowledge of linear algebra can understand without difficulty. A general expression of the fundamental matrix has been derived which is valid for any projection model without lens distortion (including full perspective and affine camera). A new and efficient technique for affine reconstruction from two affine images has been developed, which consists in first estimating the affine epipolar geometry and then performing a triangulation with respect to an implicit affine basis for each point match.

References

[Beardsley *et a*l., 1994] P. Beardsley, A. Zisserman, and D. Murray. Navigation using affine structure from motion. In J-O. Eklundh, editor, *Proc. 3rd European Conf. on Computer Vision,* volume 2, pages 85-96, Stockholm, Sweden, May 1994.

[Faugeras *et al,* 1992] O. Faugeras, T. Luong, and S. Maybank. Camera self-calibration: theory and experiments. In G. Sandini, editor, *Proc 2nd ECCV,* pages 321-334, Santa Margherita Ligure, Italy, May 1992.

[Faugeras, 1992] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision,* Santa Margherita Ligure, Italy, May 1992.

[Faugeras, 1993] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint* MIT Press, 1993.

[Faugeras, 1995] O. Faugeras. Stratification of 3-D vision: projective, affine, and metric representations. *Journal of the Optical Society of America A,* 12(3):465 484, March 1995.

[Forsyth *et a*l., 1991] D. Forsyth, J.L. Mundy, A. Zisserman, C. Coello, A. Heller, and C. Rothwell. Invariant Descriptors for 3D Object Recognition and Pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 13(10):971-991, October 1991.

[Hartley *et al,* 1992] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conf Computer Vision Pattern Recognition,* pages 761 764, Urbana Champaign, IL, June 1992.

[Koenderink and van Doom, 1991] J.J. Koenderink and A.J. van Doom. Affine structure from motion. *Journal of the Optical Society of America,* A8:377-385, 1991.

[Luong and Faugeras, 1996] Q.-T. Luong and O.D. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *The International Journal of Computer Vision,* I(17):43-76, January 1996.

[Luong and Vieville, 1994] Q.-T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. In J-O. Eklundh, editor, *Proc. 3rd European Conf on Computer Vision,* volume 1, pages 589 599, Stockholm, Sweden, May 1994.

[Maybank, 1992] S.J. Maybank. *Theory of reconstruction From Image Motion.* Springer-Verlag, 1992.

[Mundy and Zisserman, 1992] J.L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision.* MIT Press, 1992.

[Shapiro *et al.,* 1994] L.S. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. In J-O. Eklundh, editor, *Proc. 3rd European Conf. on Computer Vision,* volume II, pages 73-84, Stockholm, Sweden, May 1994.

[Shashua, 1994] A. Shashua. Projective structure from uncalibrated images: structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 16(8):778-790, 1994.

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

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

[Zisserman, 1992] A. Zisserman. Notes on geometric invariants in vision. BMVC92 Tutorial, 1992.

PANEL