Journal of Mathematical Imaging and Vision, 9, 213–229 (1998)

A Unified Theory of Uncalibrated Stereo for Both Perspective and Affine Cameras

ZHENGYOU ZHANG
INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France
zzhang@sophia.inria.fr

GANG XU
Computer Science Department, CV Lab, Ritsumeikan University, Kusatsu, Shiga 525, Japan
xu@cs.ritsumei.ac.jp

;

Abstract. This paper addresses the recovery of structure and motion from uncalibrated images of a scene under full perspective or under affine projection. Particular emphasis is placed on the configuration of two views, while the extension to N views is given in an appendix. A unified expression of the fundamental matrix is derived which is valid for any projection model without lens distortion (including full perspective and affine camera). Affine reconstruction is considered as a special projective reconstruction. The theory is elaborated in a way such that everyone having knowledge of linear algebra can understand the discussion without difficulty. A new technique for affine reconstruction 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.

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

1. Introduction

Since the work of Koenderink and van Doorn [15] on affine structure from motion and that of Forsyth et al. [12] on invariant description, the development of non-metric vision has attracted quite a number of researchers [5, 13, 26, 17] (to cite a few). We can find a range of applications: object recognition [12], 3D reconstruction of scenes [15, 27, 9], image matching [35], visual navigation [3, 33], motion segmentation [20, 30], image synthesis [8], etc.

This paper mainly addresses the recovery of structure and motion from two uncalibrated images of a scene under full perspective or under affine projection. The extension to N views is given in Appendix D. There is already a large amount of work reported in the literature [5, 7, 13, 26, 38], 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.

Contributions of this paper are the following:

- A unified 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.
- Affine reconstruction is treated as a special projective reconstruction. 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. A comparison of our work with previous work is given in Sect. 5.3.
- The theory is elaborated in a way such that everyone having knowledge of linear algebra can understand the discussion without difficulty. This arrangement, of course, sometimes sacrifices the elegance of the formulations if Projective Geometry were used.

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.

Appendix A recapitulates the technique described in [25] for estimating the affine epipolar geometry from a set of point matches. Appendix B describes a robust technique based on least-median-squares principle which detects false matches and estimates the affine epipolar geometry at the same time. Appendix C presents a technique which computes the affine transformation between two sets of 3D affine points. All these algorithms together with affine reconstruction have been implemented in C and the software `AffineF` is available from the following Web page:
`http://www.inria.fr/robotvis/personnel/zzhang/`

A review on different techniques for estimating fundamental matrix under perspective projection is also available [34].

Appendix D extends the 2-view analysis to N views, also in a unified way for both perspective and affine cameras.

2. Perspective Projection and its Approximations

If the lens distortion can be ignored, the projection from a space point $\mathtt{M} = [X, Y, Z]^T$ to its image point $\mathbf{m} = [x, y]^T$ can be represented linearly by

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} , \tag{1}$$

where $x = U/S$, and $y = V/S$ if $S \neq 0$, and $\mathbf{P}$ is the 3×4 projection matrix which varies with projection model and with the coordinate system in which space points $\mathtt{M}$ are expressed. Given a vector $\mathbf{x} = [x, y, \cdots]^T$, we use $\widetilde{\mathbf{x}}$ 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

$$s\widetilde{\mathbf{m}} = \mathbf{P}\widetilde{\mathtt{M}} , \tag{2}$$

where $s = S$ is an arbitrary nonzero scalar.

Without loss of generality, we temporarily assume that the space points $\mathtt{M}$ are expressed in the camera coordinate system and that the cameras are normalized (see e.g., [6]). Under full perspective projection, the projection matrix (identified by the subscript p) is

$$\mathbf{P}_p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} . \tag{3}$$

Expanding it we have $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$. This is a nonlinear mapping, which makes many vision problems difficult to solve, and more importantly, they can become ill-conditioned when the perspective effects are small. Sometimes, if certain conditions are satisfied, for example, when the camera field of view is small and the object size is small enough compared to the distance from the camera, the projection can be approximated by a linear mapping [1]. For orthographic projection, the projection matrix (identified by the subscript o) is

$$\mathbf{P}_o = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{4}$$

Substituting it for (1), we can easily see that the image coordinates are the same as the X and Y coordinates, and the depth Z is lost. For weak perspective projection, the projection matrix (identified by the subscript wp) is

$$\mathbf{P}_{wp} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & Z_c \end{bmatrix} , \tag{5}$$

where Z_c is the average depth of the object, which is the depth of the object centroid. The difference between this and the orthographic projection is that while the right bottom component of P_{wp} is Z_c, that of P_o is 1. The further the object moves away from the camera, the smaller its image becomes. This is the reason why the weak perspective is also called *scaled orthographic projection*. For the paraperspective projection, the projection matrix (identified by the subscript pp) is

$$\mathbf{P}_{pp} = \begin{bmatrix} 1 & 0 & -X_c/Z_c & X_c \\ 0 & 1 & -Y_c/Z_c & Y_c \\ 0 & 0 & 0 & Z_c \end{bmatrix} , \tag{6}$$

where (X_c, Y_c, Z_c) is the position of the object centroid.

If the extrinsic parameters are considered, then the above projection matrices should be multiplied from the right by a 4 × 4 matrix

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} ,$$

where $(\mathbf{R}, \mathbf{t})$ is the rotation and translation relating the world coordinate system to the camera coordinate system. If the camera intrinsic parameters are considered, then the projection matrices should be multiplied from the left by a 3 × 3 matrix (see e.g., [6] for its general form). The projection matrix corresponding to the full perspective is then of the form:

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} , \tag{7}$$

which is defined up to a scalar factor. This implies that there are only 11 degrees of freedom in a full perspective projection matrix.

If we examine the camera projection matrices for orthographic, weak perspective, and paraperspective projections (see (4), (5) and (6)), we find that they all have the same form:

$$\mathbf{P}_A = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & P_{34} \end{bmatrix} . \tag{8}$$

Depending on different projection models, some constraints exist on the elements of matrix $\mathbf{P}_A$ except for P_{31}, P_{32}, and P_{33}, which are equal to 0. If we ignore the constraints on the matrix elements, $\mathbf{P}_A$ becomes the so-called *affine camera*, introduced by Mundy and Zisserman [19].

3. Fundamental Matrix for Any Projection Model

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

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

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

$$\widetilde{\mathtt{M}}' = s'\mathbf{P}'^{+}\widetilde{\mathbf{m}}' + \mathbf{p}'^{\perp} , \tag{9}$$

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

$$\mathbf{P}'^{+} = \mathbf{P}'^{T}(\mathbf{P}'\mathbf{P}'^{T})^{-1} , \tag{10}$$

and $\mathbf{p}'^{\perp}$ is any 4-vector that is perpendicular to all the *row* vectors of $\mathbf{P}'$, i.e.,

$$\mathbf{P}'\mathbf{p}'^{\perp} = \mathbf{0} .$$

Thus, $\mathbf{p}'^{\perp}$ is a null vector of $\mathbf{P}'$. 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 (9) corresponds to a point on the optical ray defined by $\mathbf{m}'$. Equation (9) is easily justified by projecting $\mathtt{M}'$ onto the second image, which indeed gives $\mathbf{m}'$.

Similarly, an image point $\mathbf{m}$ in the first image defines also an optical ray. Requiring the two rays to intersect in space implies that a point $\mathtt{M}'$ corresponding to a particular s' in (9) must project onto the first image at $\mathbf{m}$, that is

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

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

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

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

$$\widetilde{\mathbf{m}}^T\mathbf{F}\widetilde{\mathbf{m}}' = 0 \ , \tag{11}$$

where $\mathbf{F}$ is a 3×3 matrix, called *fundamental matrix*:

$$\mathbf{F} = [\mathbf{P}\mathbf{p}'^{\perp}]_{\times}\mathbf{P}\mathbf{P}'^{+} \ , \tag{12}$$

where we use the notation $[\mathbf{x}]_{\times}$ to denote the 3×3 antisymmetric matrix defined by a 3-vector $\mathbf{x}$ such that $\mathbf{x} \times \mathbf{y} = [\mathbf{x}]_{\times}\mathbf{y}$ for any 3-vector $\mathbf{y}$. More precisely, if $\mathbf{x} = [x_1, x_2, x_3]^T$, then

$$[\mathbf{x}]_{\times} = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \ .$$

Equation (11) is the well-known epipolar equation [13, 10, 16], but the form of the fundamental matrix (12) is general and, to our knowledge, is not yet reported in the literature. It does not assume any particular projection model. Indeed, equation (12) only makes use of the pseudo-inverse of the projection matrix (which is valid for full perspective as well as for affine cameras). In [16], for example, the fundamental matrix is formulated only for full perspective, because it involves the inverse of the first 3×3 submatrix of $\mathbf{P}$ which is not invertible for affine camera. In [38], a separate fundamental matrix is given for affine cameras. Our formula (12) works for both. We will specialize it for affine cameras in Sect. 5.1.

The fundamental matrix $\mathbf{F}$ recapitulates all geometric information between two images. The nine elements of $\mathbf{F}$ are not independent from each other. In fact, $\mathbf{F}$ has only 7 degrees of freedom. This can be seen as follows. First $\mathbf{F}$ is defined up to a scale factor because if $\mathbf{F}$ is multiplied by any nonzero scalar, the new $\mathbf{F}$ still satisfy (11). Second, the rank of $\mathbf{F}$ is at most 2, i.e., $\det(\mathbf{F}) = 0$. 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, (11) can also be rewritten as $\widetilde{\mathbf{m}}'^T\mathbf{F}^T\widetilde{\mathbf{m}} = 0$. 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

$$\mathbf{p}'^{\perp} = (\mathbf{I} - \mathbf{P}'^{+}\mathbf{P}')\boldsymbol{\omega} \ , \tag{13}$$

where $\boldsymbol{\omega}$ is an arbitrary 4-vector. To show that $\mathbf{p}'^{\perp}$ is perpendicular to each row of $\mathbf{P}'$, we multiply $\mathbf{p}'^{\perp}$ by $\mathbf{P}'$ 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 $\mathbf{P}'$. If $\mathbf{P}'$ is of rank 3 (which is usually the case), then $\mathbf{p}'^{\perp}$ is unique up to a scale factor.

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. In the calibrated case, the relative position (i.e., the rotation and translation) of the two cameras is known. The problem has already been extensively studied in stereo [2, 6]. In the uncalibrated case, like the one considered here, we assume that the fundamental matrix between the two images is known (e.g., computed with the methods described in [35]), 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 (12). Let $\mathbf{A}$ and $\mathbf{A}'$ be the 3×3 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

$$\mathbf{P} = \mathbf{A}\,[\mathbf{R}\ \mathbf{t}] \quad \text{and} \quad \mathbf{P}' = \mathbf{A}'\,[\mathbf{I}\ \mathbf{0}]\ ,$$

where $(\mathbf{R}, \mathbf{t})$ is the rotation and translation relating the two camera coordinate systems, and $\mathbf{I}$ is the 3×3 identity matrix and $\mathbf{0}$ is a zero 3-vector.

It is not difficult to see that

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

This yields:

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

Using the property $(\mathbf{A}\mathbf{x}) \times (\mathbf{A}\mathbf{y}) = \det(\mathbf{A})\mathbf{A}^{-T}(\mathbf{x} \times \mathbf{y}),\quad \forall \mathbf{x}, \mathbf{y}$, and the general expression of the fundamental matrix (12), we have

$$\begin{aligned} \mathbf{F} &= [\mathbf{P}\mathbf{p}'^{\perp}]_\times \mathbf{P}\mathbf{P}'^{+} = [\mathbf{A}\mathbf{t}]_\times \mathbf{A}\mathbf{R}\mathbf{A}'^{-1} \\ &\cong \mathbf{A}^{-T}[\mathbf{t}]_\times \mathbf{R}\mathbf{A}'^{-1}\ , \end{aligned} \tag{14}$$

where $\cong$ means "equal" up to a scale factor. Equation (14) is the usual form of the fundamental matrix (see e.g., [16]).

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 [5, 13, 18, 7].

At this point, we have to introduce some elementary notation of projective geometry (an introduction can be found in [7]). For a 3D point $\mathtt{M} = [X, Y, Z]^T$, its homogeneous coordinates are $\tilde{\mathbf{x}} = [U, V, W, S]^T = \lambda \widetilde{\mathtt{M}}$ where λ is any nonzero scalar and $\widetilde{\mathtt{M}} = [X, Y, Z, 1]^T$. This implies: $U/S = X$, $V/S = Y$, $W/S = Z$. If we include the possibility that $S = 0$, then $\tilde{\mathbf{x}} = [U, V, W, S]^T$ are called the *projective coordinates* of the 3D point $\mathtt{M}$, which are not all equal to zero and defined up to a scale factor. Therefore, $\tilde{\mathbf{x}}$ and $\lambda\tilde{\mathbf{x}}$ $(\lambda \neq 0)$ represent the same projective point. When $S \neq 0$, $\tilde{\mathbf{x}} = S\widetilde{\mathtt{M}}$. When $S = 0$, we say that the point is at infinity. A 4×4 nonsingular matrix $\mathbf{H}$ defines a linear transformation from one projective point to another, and is called the *projective transformation.* The matrix $\mathbf{H}$, of course, is also defined up to a nonzero scale factor, and we write

$$\rho \widetilde{\mathbf{y}} = \mathbf{H}\widetilde{\mathbf{x}}\ , \tag{15}$$

if $\widetilde{\mathbf{x}}$ is mapped to $\widetilde{\mathbf{y}}$ by $\mathbf{H}$. Here ρ 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 $\mathbf{P}$ and $\mathbf{P}'$. From (12), 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 $\mathbf{F}$ can be uniquely determined from two images. We are now interested in recovering $\mathbf{P}$ and $\mathbf{P}'$ from $\mathbf{F}$, 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.*

More precisely, this proposition says that if $\mathbf{P}$ and $\mathbf{P}'$ are two camera projection matrices consistent with the fundamental matrix $\mathbf{F}$, then $\widehat{\mathbf{P}} = \mathbf{P}\mathbf{H}$ and $\widehat{\mathbf{P}}' = \mathbf{P}'\mathbf{H}$ are also consistent with the same $\mathbf{F}$, where $\mathbf{H}$ is any projective transformation of the 3D space. Therefore, we only need to prove

$$[\widehat{\mathbf{P}}\widehat{\mathbf{p}}'^{\perp}]_\times \widehat{\mathbf{P}}\widehat{\mathbf{P}}'^{+} = \lambda\mathbf{F} \equiv \lambda[\mathbf{P}\mathbf{p}'^{\perp}]_\times \mathbf{P}\mathbf{P}'^{+}\ , \tag{16}$$

where $\widehat{\mathbf{p}}'^{\perp} = (\mathbf{I} - \widehat{\mathbf{P}}'^{+}\widehat{\mathbf{P}}')\widehat{\boldsymbol{\omega}}$ with $\widehat{\boldsymbol{\omega}}$ any 4-vector, and λ is a scalar since $\mathbf{F}$ is defined up to a scale factor.

Proof: After some simple algebra, we have

$$\widehat{\mathbf{P}}\widehat{\mathbf{p}}'^{\perp} = \mathbf{P}\mathbf{x} ,$$

where $\mathbf{x}$ is a 4-vector given by

$$\mathbf{x} = \left(\mathbf{I} - \mathbf{H}\mathbf{H}^T\mathbf{P}'^T(\mathbf{P}'\mathbf{H}\mathbf{H}^T\mathbf{P}'^T)^{-1}\mathbf{P}'\right)\mathbf{H}\widehat{\boldsymbol{\omega}} .$$

Multiplying $\mathbf{x}$ by $\mathbf{P}'$ from the left yields $\mathbf{P}'\mathbf{x} = \mathbf{0}$, which implies that $\mathbf{x}$ is a null vector of $\mathbf{P}'$. Since in general $\text{rank}(\mathbf{P}') = 3$ (i.e., the three row vectors are independent of each other, which is the case for both perspective projections and affine cameras), there is a *unique* null vector of $\mathbf{P}'$, defined up to a scale factor; we thus have $\mathbf{x} = \lambda\mathbf{p}'^{\perp}$ where $\mathbf{p}'^{\perp}$ is given by (13) and λ is a scale factor. Therefore, we have

$$\widehat{\mathbf{P}}\widehat{\mathbf{p}}'^{\perp} = \lambda\mathbf{P}\mathbf{p}'^{\perp} . \tag{17}$$

Next, let us examine $\widehat{\mathbf{P}}\widehat{\mathbf{P}}'^{+}$, which is equal to

$$\widehat{\mathbf{P}}\widehat{\mathbf{P}}'^{+} = \mathbf{P}\mathbf{H}\mathbf{H}^T\mathbf{P}'^T(\mathbf{P}'\mathbf{H}\mathbf{H}^T\mathbf{P}'^T)^{-1} \equiv \mathbf{P}\mathbf{N} ,$$

where $\mathbf{N} = \mathbf{H}\mathbf{H}^T\mathbf{P}'^T(\mathbf{P}'\mathbf{H}\mathbf{H}^T\mathbf{P}'^T)^{-1}$ is a 4×3 matrix. It is easy to verify that $\mathbf{P}'\mathbf{N} = \mathbf{I}$, so we must have

$$\mathbf{N} = \mathbf{P}'^{+} + \mathbf{Q} ,$$

where $\mathbf{Q}$ is a 4×3 matrix such that $\mathbf{P}'\mathbf{Q} = \mathbf{O}$, i.e., each column vector $\mathbf{q}_i$ ($i = 1, 2, 3$) of $\mathbf{Q}$ must be the null vector of $\mathbf{P}'$. That is, $\mathbf{P}'\mathbf{q}_i = \mathbf{0}$ for $i = 1, 2, 3$. Since the null vector is unique (up to a scale factor) and $\mathbf{P}'\mathbf{p}^{'\perp} = \mathbf{0}$, we have $\mathbf{q}_i = \alpha_i\mathbf{p}'^{\perp}$, where α_i is some scalar. Therefore, we have

$$\widehat{\mathbf{P}}\widehat{\mathbf{P}}'^{+} = \mathbf{P}\mathbf{P}'^{+} + \mathbf{P}\mathbf{Q} . \tag{18}$$

Combining (17) and (18) gives

$$[\widehat{\mathbf{P}}\widehat{\mathbf{p}}'^{\perp}]_\times\widehat{\mathbf{P}}\widehat{\mathbf{P}}'^{+} = \lambda[\mathbf{P}\mathbf{p}'^{\perp}]_\times\mathbf{P}\mathbf{P}'^{+} + \lambda[\mathbf{P}\mathbf{p}'^{\perp}]_\times\mathbf{P}\mathbf{Q} .$$

Because of the structure of matrix $\mathbf{Q}$ and the operator $[\cdot]_\times$, the second term of the right side of the above equation is a zero matrix, i.e., $[\mathbf{P}\mathbf{p}'^{\perp}]_\times\mathbf{P}\mathbf{Q} = \mathbf{O}$. The above equation is finally reduced to (16), which completes the proof. □

The consequence of this proposition is the following: if the true structure is $\mathtt{M}$, then the structure reconstructed from image points is $\mathbf{H}^{-1}\widetilde{\mathtt{M}}$, i.e., up to a projective transformation. This is because $\widehat{\mathbf{P}}\mathbf{H}^{-1}\widetilde{\mathtt{M}} = \mathbf{P}\widetilde{\mathtt{M}}$ 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 $\mathbf{F}$ with respect to a projective basis, which can be arbitrary because of Proposition 1. One way is to use a canonical representation [17, 3], as described below. The fundamental matrix $\mathbf{F}$ can be factored into a product of an antisymmetric matrix $[\mathbf{e}]_\times$ and a matrix $\mathbf{M}$, i.e.,

$$\mathbf{F} = [\mathbf{e}]_\times\mathbf{M} , \tag{19}$$

where $\mathbf{e}$ is the epipole in the first image because $\mathbf{F}^T\mathbf{e} = \mathbf{0}$, and $\mathbf{M}$ is a 3×3 matrix which is in general not unique because if $\mathbf{M}$ is a solution then $\mathbf{M} + \mathbf{e}\mathbf{v}^T$ is also a solution for any 3-vector $\mathbf{v}$ (indeed, we have always $[\mathbf{e}]_\times\mathbf{e}\mathbf{v}^T = \mathbf{O}$). Since $\mathbf{F}^T\mathbf{e} = \mathbf{0}$, the epipole in the first image is given by the eigenvector of matrix $\mathbf{F}\mathbf{F}^T$ associated to the smallest eigenvalue. Using the relation

$$\|\mathbf{v}\|^2\mathbf{I}_3 = \mathbf{v}\mathbf{v}^T - [\mathbf{v}]_\times^2 \qquad \forall\mathbf{v} ,$$

we have

$$\begin{aligned}\mathbf{F} &= \frac{1}{\|\mathbf{e}\|^2}(\mathbf{e}\mathbf{e}^T - [\mathbf{e}]_\times^2)\mathbf{F} \\ &= \frac{1}{\|\mathbf{e}\|^2}\underbrace{\mathbf{e}\mathbf{e}^T\mathbf{F}}_{\mathbf{O}} + [\mathbf{e}]_\times\underbrace{\left(-\frac{[\mathbf{e}]_\times}{\|\mathbf{e}\|^2}\mathbf{F}\right)}_{\mathbf{M}} .\end{aligned}$$

The first term on the right hand is a zero matrix because $\mathbf{F}^T\mathbf{e} = \mathbf{0}$. We can thus define matrix $\mathbf{M}$ as

$$\mathbf{M} = -\frac{1}{\|\mathbf{e}\|^2}[\mathbf{e}]_\times\mathbf{F} . \tag{20}$$

Once $\mathbf{F}$ is decomposed as above, the camera projection matrices can be chosen as

$$\mathbf{P} = [\mathbf{M} \quad \mathbf{e}] \quad \text{and} \quad \mathbf{P}' = [\mathbf{I} \quad \mathbf{0}] . \tag{21}$$

It is easy to verify that the above $\mathbf{P}$ and $\mathbf{P}'$ do yield the fundamental matrix $\mathbf{F}$. Another way is to choose five point matches, each of four points not being coplanar. The five points can be real as in [5] or be virtual as in [36].

Once $\mathbf{P}$ and $\mathbf{P}'$ are determined, the 3D reconstruction can be done in a much similar way as with calibrated cameras. Given two matched points $\mathbf{m}$ and $\mathbf{m}'$, we can estimate the corresponding 3D point $\mathtt{M}$ by minimizing the following criterion:

$$\mathcal{F}(\mathtt{M}) = \|\mathbf{m} - \widehat{\mathbf{m}}\|^2 + \|\mathbf{m}' - \widehat{\mathbf{m}}'\|^2 ,$$

where $\widehat{\mathbf{m}}$ and $\widehat{\mathbf{m}}'$ are projected points of $\mathtt{M}$ according to $\mathbf{P}$ and $\mathbf{P}'$, respectively. The reader is referred to [14, 22, 32] for more details.

5. Affine Reconstruction

This section deals with two images taken by an affine camera at two different instants or by two different affine cameras. We show that the structure can only be recovered up to an affine transformation in 3D space, and a new method for affine reconstruction is developed.

5.1. *Affine Fundamental Matrix*

In the case of a general affine camera [19, 25], the projection matrix (8) can be rewritten as

$$\mathbf{P}_A = \begin{bmatrix} \mathbf{p}_1^T & \\ \mathbf{p}_2^T & \mathbf{p}_4 \\ \mathbf{0}_3^T & \end{bmatrix} , \tag{22}$$

where $\mathbf{p}_4 = [P_{14}, P_{24}, P_{34}]^T$. We now derive the specific fundamental matrix for affine cameras from the general form of the fundamental matrix (12).

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 $\mathbf{p}_1'^T \mathbf{p}_3' = 0$ and $\mathbf{p}_2'^T \mathbf{p}_3' = 0$, we can verify that $\mathbf{p}'^{\perp}$ is indeed perpendicular to $\mathbf{P}'$:

$$\mathbf{P}'\mathbf{p}'^{\perp} = \frac{1}{\|\mathbf{p}_3'\|} \begin{bmatrix} \mathbf{p}_1'^{\,T} & \\ \mathbf{p}_2'^{\,T} & \mathbf{p}_4' \\ \mathbf{0}_3'^{\,T} & \end{bmatrix} \begin{bmatrix} \mathbf{p}_3' \\ 0 \end{bmatrix} = \mathbf{0}_3 .$$

Now, multiplying $\mathbf{P}$ with $\mathbf{p}'^{\perp}$ yields

$$\mathbf{P}\mathbf{p}'^{\perp} = \begin{bmatrix} \mathbf{p}_1^T \mathbf{p}_3' \\ \mathbf{p}_2^T \mathbf{p}_3' \\ 0 \end{bmatrix} , \tag{23}$$

or equivalently

$$\left[\mathbf{P}\mathbf{p}'^{\perp}\right]_{\times} = \begin{bmatrix} 0 & 0 & \mathbf{p}_2^T \mathbf{p}_3' \\ 0 & 0 & -\mathbf{p}_1^T \mathbf{p}_3' \\ -\mathbf{p}_2^T \mathbf{p}_3' & \mathbf{p}_1^T \mathbf{p}_3' & 0 \end{bmatrix} . \tag{24}$$

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

$$\mathbf{P}'\mathbf{P}'^{+} = \begin{bmatrix} \mathbf{p}_1'^T \mathbf{Q} \\ \mathbf{p}_2'^T \mathbf{Q} \\ \mathbf{0}_3^T \end{bmatrix} + \mathbf{p}_4' \mathbf{q}_4^T = \mathbf{I}_3 ,$$

$\mathbf{q}_4$ can be uniquely determined:

$$\mathbf{q}_4 = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{P_{34}'} \end{bmatrix} . \tag{25}$$

The constraint for matrix $\mathbf{Q}$ is then

$$\begin{bmatrix} \mathbf{p}_1'^T \\ \mathbf{p}_2'^T \end{bmatrix} \mathbf{Q} = \begin{bmatrix} 1 & 0 & -\frac{P_{14}'}{P_{34}'} \\ 0 & 1 & -\frac{P_{24}'}{P_{34}'} \end{bmatrix} . \tag{26}$$

It is evident that $\mathbf{Q}$ cannot be uniquely determined. In other words, any $\mathbf{Q}$ that satisfies the above equation suffices.

Now substituting these matrices for (12), we have

$$\mathbf{F}_A = \begin{bmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{27}$$

where

$$\begin{aligned} a_{13} &= \frac{P_{34}}{P_{34}'} \mathbf{p}_2^T \mathbf{p}_3', \\ a_{23} &= -\frac{P_{34}}{P_{34}'} \mathbf{p}_1^T \mathbf{p}_3', \\ a_{31} &= (-\mathbf{p}_2^T \mathbf{p}_3' \mathbf{p}_1^T + \mathbf{p}_1^T \mathbf{p}_3' \mathbf{p}_2^T) \mathbf{q}_1, \\ a_{32} &= (-\mathbf{p}_2^T \mathbf{p}_3' \mathbf{p}_1^T + \mathbf{p}_1^T \mathbf{p}_3' \mathbf{p}_2^T) \mathbf{q}_2, \\ a_{33} &= (-\mathbf{p}_2^T \mathbf{p}_3' \mathbf{p}_1^T + \mathbf{p}_1^T \mathbf{p}_3' \mathbf{p}_2^T) \mathbf{q}_3 \\ &\quad -\frac{P_{14}}{P_{34}'} \mathbf{p}_2^T \mathbf{p}_3' + \frac{P_{24}}{P_{34}'} \mathbf{p}_1^T \mathbf{p}_3' . \end{aligned}$$

The fact that the affine fundamental matrix has the form of (27) is mentioned in [38]. Defined up to a scale factor, $\mathbf{F}_A$ has only 4 degrees of freedom.

The corresponding points in the two images must satisfy the following relation, called the *affine epipolar equation*:

$$\tilde{\mathbf{m}}^T \mathbf{F}_A \tilde{\mathbf{m}}' = 0 \;. \qquad (28)$$

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

$$a_{13}x + a_{23}y + a_{31}x' + a_{32}y' + a_{33} = 0 \;. \qquad (29)$$

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 $\mathbf{F}_A$ can be estimated (see [25] and appendix 7 of this paper for details). We are now interested in recovering $\mathbf{P}_A$ and $\mathbf{P}'_A$ from $\mathbf{F}_A$, and once they are recovered, the structure can be redressed in 3D space.

Since $\mathbf{P}_A$ and $\mathbf{P}'_A$ 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

$$\tilde{\mathbf{m}} = \mathbf{P}_A \widetilde{\mathtt{M}} \quad \text{and} \quad \tilde{\mathbf{m}}' = \mathbf{P}'_A \widetilde{\mathtt{M}} \;.$$

Note that there is no more scale factor in the above equations. From the affine epipolar equation (28), it is easy to obtain

$$\widetilde{\mathtt{M}}^T \underbrace{\mathbf{P}_A^T \mathbf{F}_A \mathbf{P}'_A}_{\mathbf{S}} \widetilde{\mathtt{M}} = 0 \;, \qquad (30)$$

where

$$\mathbf{S} = \begin{bmatrix} P_{11} & P_{21} & 0 \\ P_{12} & P_{22} & 0 \\ P_{13} & P_{23} & 0 \\ P_{14} & P_{24} & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} P'_{11} & P'_{12} & P'_{13} & P'_{14} \\ P'_{21} & P'_{22} & P'_{23} & P'_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 & S_{14} \\ 0 & 0 & 0 & S_{24} \\ 0 & 0 & 0 & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} ,$$

with

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

Equation (30) becomes

$$\begin{aligned} (S_{14} + S_{41})X + (S_{24} + S_{42})Y & \\ + (S_{34} + S_{43})Z + S_{44} = 0 \;. & \end{aligned}$$

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:

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

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 $\mathtt{M}$. We need to show that the new structure $\widetilde{\widetilde{\mathtt{M}}} = \mathbf{H}_A^{-1}\widetilde{\mathtt{M}}$ 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 $\widetilde{\widehat{\mathtt{M}}}$ 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 (31), where P_{ij} and P'_{ij} should be replaced by $\widehat{P}_{ij}$ and $\widehat{P}'_{ij}$. Indeed, multiplying both sides of (31) 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 (31). In the following, we propose a similar method to what we used in (21). We will consider affine reconstruction as a special projective reconstruction. The affine fundamental matrix $\mathbf{F}_A$ can always be decomposed into $\mathbf{F}_A = [\mathbf{e}]_\times \mathbf{M}$ as in (19), where $\mathbf{e}$ and $\mathbf{M}$ can be simply computed as:

$$\mathbf{e} = \begin{bmatrix} -a_{23} \\ a_{13} \\ 0 \end{bmatrix}$$

(note that the last element is 0, implying that the epipole is at infinity.)

$$\mathbf{M} = -\frac{1}{\|\mathbf{e}\|^2}[\mathbf{e}]_\times \mathbf{F} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ 0 & 0 & 1 \end{bmatrix} ,$$

with $m_{ij} = -a_{i3}a_{3j}/(a_{13}^2 + a_{23}^2)$ for $i = 1, 2$ and $j = 1, 2, 3$.

If we conduct projective reconstruction, we can construct $\mathbf{P}$ and $\mathbf{P}'$ as in (21), which gives

$$\mathbf{P} = \begin{bmatrix} \mathbf{M} & \mathbf{e} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & -a_{23} \\ m_{21} & m_{22} & m_{23} & a_{13} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{P}' = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} .$$

However, they do not satisfy the definition of affine cameras (8), although they give the same fundamental matrix as $\mathbf{F}_A$. Let us define a special projective transformation $\mathbf{H}_A$ as

$$\mathbf{H}_A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} , \tag{32}$$

which simply swaps the last two columns of $\mathbf{P}$ and $\mathbf{P}'$, or equivalently swaps the third and fourth coordinates of a 3D projective point. If we apply $\mathbf{H}_A$, then we get

$$\mathbf{P}_A = \mathbf{P}\mathbf{H}_A = \begin{bmatrix} \mathbf{M} & \mathbf{e} \end{bmatrix} \mathbf{H}_A = \begin{bmatrix} m_{11} & m_{12} & -a_{23} & m_{13} \\ m_{21} & m_{22} & a_{13} & m_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{33}$$

$$\mathbf{P}'_A = \mathbf{P}'\mathbf{H}_A = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{H}_A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{34}$$

As we have shown in the previous section, multiplication of a projective transformation does not change the fundamental matrix. Furthermore, $\mathbf{P}_A$ and $\mathbf{P}'_A$ are now *affine* camera projection matrices.

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

$$\widehat{\mathbf{m}} = \begin{bmatrix} \widehat{u} \\ \widehat{v} \end{bmatrix} = \begin{bmatrix} m_{11}X + m_{12}Y - a_{23}Z + m_{13} \\ m_{21}X + m_{22}Y + a_{13}Z + m_{23} \end{bmatrix}$$

$$\widehat{\mathbf{m}}' = \begin{bmatrix} \widehat{u}' \\ \widehat{v}' \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} ,$$

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

$$\mathbf{B} = \begin{bmatrix} m_{11}^2 + m_{21}^2 + 1 & m_{11}m_{12} + m_{21}m_{22} & -m_{11}a_{23} + m_{21}a_{13} \\ m_{11}m_{12} + m_{21}m_{22} & m_{12}^2 + m_{22}^2 + 1 & -m_{12}a_{23} + m_{22}a_{13} \\ -m_{11}a_{23} + m_{21}a_{13} & -m_{12}a_{23} + m_{22}a_{13} & a_{23}^2 + a_{13}^2 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} u' + m_{11}u + m_{21}v \\ v' + m_{12}u + m_{22}v \\ -a_{23}u + a_{13}v \end{bmatrix} \ .$$

The 3D reconstructed point is then given by $\mathtt{M} = \mathbf{B}^{-1}\mathbf{b}$.

5.3. Relation to Previous Work

There already exist a number of algorithms for the recovery of affine structure from two affine images. They can be divided into two categories. The first relies on use of a local coordinate frame by choosing four non-coplanar points to form the affine basis [15, 4, 21, 31]. One drawback is that the error in the basis points directly affects the precision of the entire solution. The second category is characterized by the work of Shapiro [24]. Inspired by the work of Tomasi and Kanade [29] for a long image sequence under orthography, Shapiro uses the singular value decomposition technique (SVD) to determine the affine cameras and the scene structure simultaneously with the whole set of points. Our work uses also the whole set of points, but we first recover the affine epipolar geometry and then determine the scene structure. Instead of conducting a SVD of a $4 \times n$ matrix as in [24] where n is the number of point matches, we solve now two smaller problems:

- determination of the affine epipolar geometry, which involves the computation of the eigenvector of a 4×4 symmetric matrix associated with the smallest eigenvalue (see [25] appendix 7 of this paper for more details);
- 3D reconstruction, which involves an inverse of a 3×3 symmetric matrix, which is the same for all points, and a multiplication of a 3×3 matrix with a 3-vector for each point.

The new technique is thus more efficient.

6. Experimental Results with Affine Reconstruction

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

6.1. Synthetic Data

The parameters of the camera set-up are taken from a real stereovision 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 $400 \times 250 \times 300$ 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 512×512 pixels^2, and the projection of the object occupies a surface of about 130×120 pixels^2.

The method described in [25] 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, the affine transformation which brings the set of affinely reconstructed points to the original set of 3D points. The reader is referred to appendix 6 of this paper for details on how to estimate the affine transformation between two sets 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.

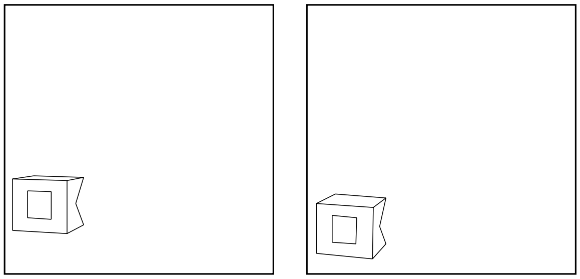

Fig. 1. Two perspective images of a synthetic object

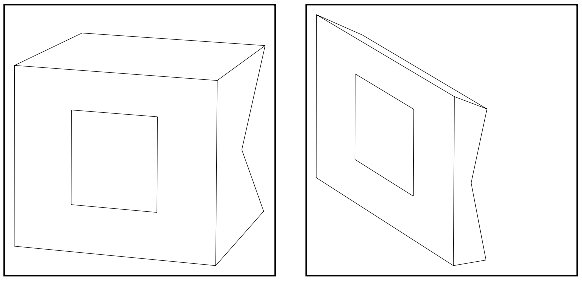

Fig. 2. Two orthographic views of the affine reconstruction

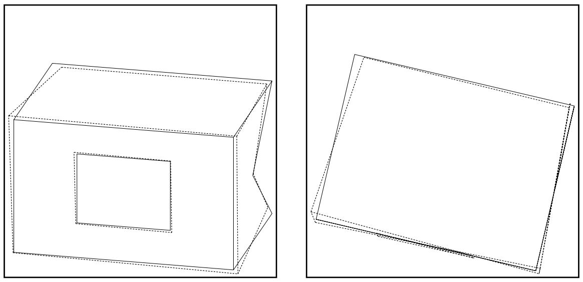

Fig. 3. Two orthographic views of the superposition of the original 3D data (in solid lines) and the transformed affine reconstruction (in dashed lines)

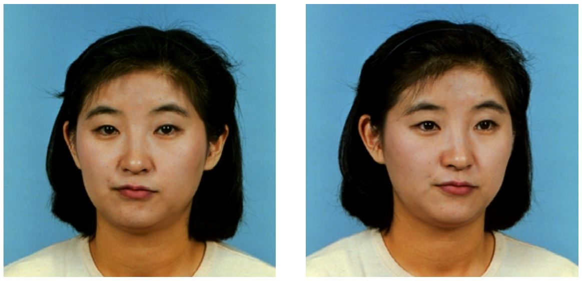

Fig. 4. Two original facial images

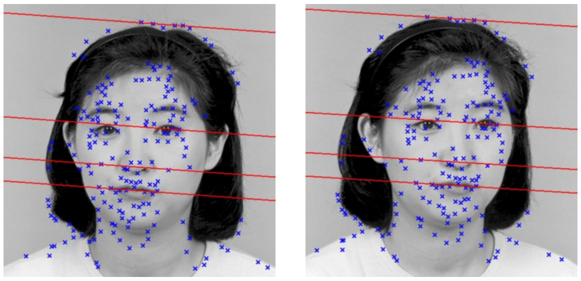

Fig. 5. Matched points (indicated by crosses) between two facial images with four corresponding epipolar lines overlaid

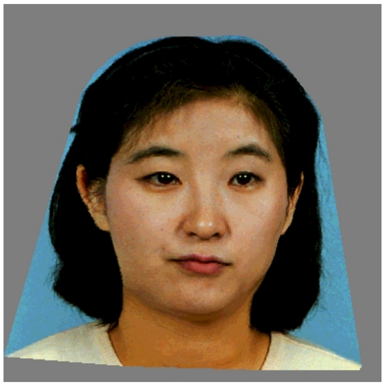

Fig. 6. A new image synthesized from the two real images shown in Fig. 4

6.2. *Application to Image Synthesis*

In this subsection, we apply the affine reconstruction technique to synthesize new images from real images (see [37] for more details). Figure 4 shows two original facial images differed by a rotation in depth of about 20 degrees. Although the projection is not affine, the images will be considered to be taken under affine projection.

We first extracted a set of characteristic points from each image, and then tried to establish an initial set of point matches using correlation and relaxation techniques as described in [35]. Finally, the robust technique described in Appendix B was applied to detect the false matches and to estimate the affine fundamental matrix between the two images. The good matches and the epipolar geometry thus obtained are shown in Fig. 5. (Note that our current version of image matching uses black-white images.)

Once the affine fundamental matrix is estimated, we can conduct affine reconstruction for each point match with respect to an implicit affine coordinate system as described in the last section. Once the desired position of a new image is specified, the reconstructed points are projected onto the new image. Using points as vertices, we can divide the new image into a set of triangular patches. Finally, textures (colors) from the original images are mapped to the triangular patches. An example is shown in Fig. 6, which roughly corresponds to the intermediate position of the two images shown in Fig. 4.

7. 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 unified expression of the fundamental matrix has been derived which is valid for any projection model without lens distortion (including full perspective and affine camera). Affine reconstruction is considered as a special projective reconstruction. 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.

Appendix A

Estimation of the Affine Epipolar Geometry

In this section, we present the technique described in [25] for estimating the affine epipolar geometry from a set of point matches. Consider a pair of matched points: $\mathbf{m}_i = [x_i, y_i]^T$ and $\mathbf{m}'_i = [x'_i, y'_i]^T$. The affine epipolar equation (29) can be regarded as a hyperplane in 4D and rewritten as $\mathbf{r}_i^T \mathbf{n} + a_{33} = 0$, where $\mathbf{r}_i = [x_i, y_i, x'_i, y'_i]^T$ contains the two image coordinates, and $\mathbf{n} = [a_{13}, a_{23}, a_{31}, a_{32}]^T$ is the 4D normal vector. The perpendicular distance from $\mathbf{r}_i$ to this hyperplane is given by $(\mathbf{r}_i^T \mathbf{n} + a_{33})/\|\mathbf{n}\|$. Given n point matches, we can estimate the affine epipolar geometry by minimizing the following cost function:

$$\mathcal{F}(\mathbf{n}, a_{33}) = \sum_{i=1}^{n} \frac{(\mathbf{r}_i^T \mathbf{n} + a_{33})^2}{\|\mathbf{n}\|^2} , \qquad (1)$$

which is the sum of the squared 4D perpendicular distances.

Since the five coefficients of the affine epipolar geometry are defined up to a scale factor, we can impose $\|\mathbf{n}\| = 1$. Using a Lagrange multiplier λ, we get

$$\mathcal{F}'(\mathbf{n}, a_{33}) = \sum_{i=1}^{n} (\mathbf{r}_i^T \mathbf{n} + a_{33})^2 + \lambda(1 - \mathbf{n}^T \mathbf{n}) .$$

We solve by setting the partial derivatives of $\mathcal{F}'(\mathbf{n}, a_{33})$ to zero. Differentiating with respect to a_{33} gives

$$\frac{\partial \mathcal{F}'(\mathbf{n}, a_{33})}{\partial a_{33}} = 2 \sum_{i=1}^{n} (\mathbf{r}_i^T \mathbf{n} + a_{33}) = 0 ,$$

which leads to

$$a_{33} = -\frac{1}{n} \sum_{i=1}^{n} (\mathbf{r}_i^T \mathbf{n}) = -\mathbf{n}^T \bar{\mathbf{r}} ,$$

where $\bar{\mathbf{r}}$ is the centroid of the 4D vectors $\mathbf{r}_i$. The optimal solution $\mathbf{n}$ thus passes through the data centroid $\bar{\mathbf{r}}$.

Substituting a_{33} into $\mathcal{F}'(\mathbf{n}, a_{33})$ and denoting the centered points by $\mathbf{v}_i = \mathbf{r}_i - \bar{\mathbf{r}}$, we obtain

$$\begin{aligned} \mathcal{F}''(\mathbf{n}) &= \sum_{i=1}^{n} (\mathbf{v}_i^T \mathbf{n})^2 + \lambda(1 - \mathbf{n}^T \mathbf{n}) \\ &= \mathbf{n}^T \mathbf{W} \mathbf{n} + \lambda(1 - \mathbf{n}^T \mathbf{n}) , \end{aligned}$$

where $\mathbf{W} = \sum_{i=1}^{n} \mathbf{v}_i \mathbf{v}_i^T$ is a symmetric matrix. Differentiating with respect to $\mathbf{n}$ gives

$$\frac{\partial \mathcal{F}''(\mathbf{n})}{\partial \mathbf{n}} = 2\mathbf{W}\mathbf{n} - 2\lambda\mathbf{n} = \mathbf{0} \ ,$$

or $\mathbf{W}\mathbf{n} = \lambda\mathbf{n}$. Thus, $\mathbf{n}$ is a unit eigenvector of $\mathbf{W}$ corresponding to the eigenvalue λ. To decide *which* eigenvalue, we substitute into $\mathcal{F}''(\mathbf{n})$,

$$\mathcal{F}''(\mathbf{n})_{\min} = \mathbf{n}^T\mathbf{W}\mathbf{n} = \lambda\|\mathbf{n}\|^2 = \lambda \ ,$$

showing that λ is the *smallest* eigenvalue of $\mathbf{W}$ (and $\mathbf{n}$ its associated eigenvector).

Although the above solution is optimal in terms of 4D perpendicular distances, it is shown in [24, 32] that $\mathcal{F}(\mathbf{n}, a_{33})$ is equivalent to the sum of squared distances between the observed image locations and the locations predicted by projecting the computed affine structure $\mathtt{M}_i$ onto an image using the computed affine cameras ($\mathbf{P}_A$ and $\mathbf{P}'_A$). This solution is thus also optimal in terms of an *image* distance measure.

Appendix B

False Match Detection and Robust Estimation

We have adapted a previously developed robust technique [35] to affine cameras. It is based on the least-median-squares method [23], and is able to detect false matches as many as 50% of the whole set of data and at the same time produce an accurate estimation of the affine epipolar geometry.

Given n point correspondences: $\{(\mathbf{m}_i, \mathbf{m}'_i) | i = 1, \ldots, n\}$, we proceed the following steps:

1. A Monte Carlo type technique is used to draw m random subsamples of $p = 4$ different point correspondences (recall that 4 is the minimum number to determine the affine epipolar geometry).
2. For each subsample, indexed by J, we use the technique described previously to compute the affine fundamental matrix $\mathbf{F}_J$.
3. For each $\mathbf{F}_J$, we can determine the median of the squared residuals, denoted by M_J, with respect to the whole set of point correspondences, i.e.,

   $$M_J = \operatorname*{median}_{i=1,\ldots,n} \frac{(\mathbf{r}_i^T\mathbf{n} + a_{33})^2}{\|\mathbf{n}\|^2} \ .$$

   Here, the squared 4D perpendicular distances are used, see (1).
4. Retain the estimate $\mathbf{F}_J$ for which M_J is minimal among all m M_J's.
5. Compute the *robust standard deviation* estimate

   $$\widehat{\sigma} = 1.4826[1 + 5/(n-p)]\sqrt{M_J} \ .$$

6. Declare a point match as a false match if its 4D perpendicular distance is larger than $(k\widehat{\sigma})^2$, where k is set to 2.5.
7. Discard the false matches and re-estimate the affine fundamental matrix using all remaining point matches.

More implementation details can be found in [35].

Appendix C

Estimation of the Affine Transformation

In this section, we present a technique which computes the affine transformation from a set of affinely reconstructed 3D points, denoted here by $\mathbf{x}_i = [x_i, y_i, z_i]^T$, to a set of 3D reference points, denoted here by $\mathbf{x}'_i = [x'_i, y'_i, z'_i]^T$. Let n be the number of points. Let $\mathbf{A}$ and $\mathbf{t}$ be the 3×3 matrix and 3-vector representing the affine transformation. For each pair of points, we then have

$$\mathbf{x}'_i = \mathbf{A}\mathbf{x}_i + \mathbf{t} \ .$$

The estimation of the affine transformation can be formulated as a least-squares by minimizing the following cost function:

$$\mathcal{F}(\mathbf{A}, \mathbf{t}) = \sum_{i=1}^{n} (\mathbf{A}\mathbf{x}_i + \mathbf{t} - \mathbf{x}'_i)^T (\mathbf{A}\mathbf{x}_i + \mathbf{t} - \mathbf{x}'_i) \ .$$

The solution of $\mathbf{t}$ is obtained by setting the first derivative of $\mathcal{F}(\mathbf{A}, \mathbf{t})$ with respect to zero:

$$\frac{\partial \mathcal{F}(\mathbf{A}, \mathbf{t})}{\partial \mathbf{t}} = 2\sum_{i=1}^{n} (\mathbf{A}\mathbf{x}_i + \mathbf{t} - \mathbf{x}'_i) = \mathbf{0} \ ,$$

which leads to

$$\mathbf{t} = \bar{\mathbf{x}}' - \mathbf{A}\bar{\mathbf{x}} \ ,$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i$ and $\bar{\mathbf{x}}' = \frac{1}{n}\sum_i \mathbf{x}'_i$ are the centroids of the two point sets. The optimal solution $\mathbf{A}$ thus passes through the data centroids $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}'$.

Substituting $\mathbf{t}$ into $\mathcal{F}(\mathbf{A}, \mathbf{t})$ and denoting the centered points by $\mathbf{y}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ and $\mathbf{y}'_i = \mathbf{x}'_i - \bar{\mathbf{x}}'$, we get

$$\mathcal{F}'(\mathbf{A}) = \sum_{i=1}^{n} (\mathbf{A}\mathbf{y}_i - \mathbf{y}'_i)^T (\mathbf{A}\mathbf{y}_i - \mathbf{y}'_i) \ .$$

Now let us define the derivative of a scalar λ with respect to a matrix $\mathbf{A}$:

$$\mathcal{D}(f, \mathbf{A}) = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \frac{\partial f}{\partial a_{13}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \frac{\partial f}{\partial a_{23}} \\ \frac{\partial f}{\partial a_{31}} & \frac{\partial f}{\partial a_{32}} & \frac{\partial f}{\partial a_{33}} \end{bmatrix} \ .$$

The solution for $\mathbf{A}$ is then given by setting $\mathcal{D}(\mathcal{F}'(\mathbf{A}), \mathbf{A}) = \mathbf{O}$, where $\mathbf{O}$ is the 3×3 zero matrix. If we consider only one term in $\mathcal{F}'(\mathbf{A})$, it can be easily verified that

$$\mathcal{D}((\mathbf{A}\mathbf{y}_i - \mathbf{y}'_i)^T(\mathbf{A}\mathbf{y}_i - \mathbf{y}'_i), \mathbf{A}) = 2\mathbf{A}\mathbf{y}_i\mathbf{y}_i^T - 2\mathbf{y}'_i\mathbf{y}_i^T \ .$$

We have thus

$$\mathcal{D}(\mathcal{F}'(\mathbf{A}), \mathbf{A}) = 2\mathbf{A}\mathbf{Y}\mathbf{Y}^T - 2\mathbf{Y}'\mathbf{Y}^T \ ,$$

where $\mathbf{Y}$ and $\mathbf{Y}'$ are $3 \times n$ matrices given by

$$\begin{aligned} \mathbf{Y} &= [\mathbf{y}_1, \ldots, \mathbf{y}_n] \ , \\ \mathbf{Y}' &= [\mathbf{y}'_1, \ldots, \mathbf{y}'_n] \ . \end{aligned}$$

The solution of $\mathbf{A}$ is then given by

$$\mathbf{A} = \mathbf{Y}'\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1} \ .$$

Appendix D

Epipolar Geometry of N Views

In this appendix, we show that the same algebraic manipulations can be easily extended to N-view analysis.

Let us consider first the case of three views where a set of point matches are available, denoted by $\{\mathbf{m}_i, \mathbf{m}'_i, \mathbf{m}''_i\}$ $(i = 1, \ldots, n)$. To derive the constraints on the images between three views, we follow the same idea as that presented in [28] for calibrated perspective images, but as for the fundamental matrix, we will not assume any particular projection model. Therefore, the result will be valid for both perspective and affine cameras.

Consider one point match $(\mathbf{m}, \mathbf{m}', \mathbf{m}'')$ (we omit here the subscript to simplify the notation). Let the corresponding structure in 3D space be $\mathtt{M}$. In general (full perspective or affine projection), we have

$$\begin{aligned} s\widetilde{\mathbf{m}} &= \mathbf{P}\widetilde{\mathtt{M}} \ , & (1) \\ s'\widetilde{\mathbf{m}}' &= \mathbf{P}'\widetilde{\mathtt{M}} \ , & (2) \\ s''\widetilde{\mathbf{m}}'' &= \mathbf{P}''\widetilde{\mathtt{M}} \ . & (3) \end{aligned}$$

Using pseudo-inverse matrices, we can get

$$\widetilde{\mathtt{M}} = s\mathbf{P}^+\widetilde{\mathbf{m}} + \mathbf{p}^\perp \ . \quad (4)$$

Substituting this for (2) and (3) yields

$$\begin{aligned} s'\widetilde{\mathbf{m}}' &= s\mathbf{P}'\mathbf{P}^+\widetilde{\mathbf{m}} + \mathbf{P}'\mathbf{p}^\perp \ , \\ s''\widetilde{\mathbf{m}}'' &= s\mathbf{P}''\mathbf{P}^+\widetilde{\mathbf{m}} + \mathbf{P}''\mathbf{p}^\perp \ . \end{aligned}$$

Define $\mathbf{B}' \equiv \mathbf{P}'\mathbf{P}^+$, $\mathbf{B}'' \equiv \mathbf{P}''\mathbf{P}^+$, $\mathbf{b}' \equiv \mathbf{P}'\mathbf{p}^\perp$ and $\mathbf{b}'' \equiv \mathbf{P}''\mathbf{p}^\perp$. Then we have

$$\begin{aligned} s'\widetilde{\mathbf{m}}' = s\mathbf{B}'\widetilde{\mathbf{m}} + \mathbf{b}' \ , \\ s''\widetilde{\mathbf{m}}'' = s\mathbf{B}''\widetilde{\mathbf{m}} + \mathbf{b}'' \ . \end{aligned}$$

To eliminate the unknown structure parameters s' and s'', we take the cross product of the above two equations with $\widetilde{\mathbf{m}}'$ and $\widetilde{\mathbf{m}}''$, respectively, which leads to

$$\begin{aligned} s\widetilde{\mathbf{m}}' \times (\mathbf{B}'\widetilde{\mathbf{m}}) + \widetilde{\mathbf{m}}' \times \mathbf{b}' = \mathbf{0} \ , \\ s\widetilde{\mathbf{m}}'' \times (\mathbf{B}''\widetilde{\mathbf{m}}) + \widetilde{\mathbf{m}}'' \times \mathbf{b}'' = \mathbf{0} \ . \end{aligned}$$

There is still one unknown s. We rearrange the terms of these equations as

$$\begin{aligned} s[\widetilde{\mathbf{m}}']_\times\mathbf{B}'\widetilde{\mathbf{m}} = -[\widetilde{\mathbf{m}}']_\times\mathbf{b}' \ , \\ -[\widetilde{\mathbf{m}}'']_\times\mathbf{b}'' = s[\widetilde{\mathbf{m}}'']_\times\mathbf{B}''\widetilde{\mathbf{m}} \ . \end{aligned}$$

Remember that $[\cdot]_\times$ denotes an antisymmetric matrix defined by a vector. To eliminate s, we take the outer product of both sides of the above two equations, which gives

$$[\widetilde{\mathbf{m}}']_\times\mathbf{B}'\widetilde{\mathbf{m}}\mathbf{b}''^T[\widetilde{\mathbf{m}}'']_\times = [\widetilde{\mathbf{m}}']_\times\mathbf{b}'\widetilde{\mathbf{m}}^T\mathbf{B}''^T[\widetilde{\mathbf{m}}'']_\times \ ,$$

or

$$[\widetilde{\mathbf{m}}']_\times\mathbf{G}(\widetilde{\mathbf{m}})[\widetilde{\mathbf{m}}'']_\times = \mathbf{O}_3 \ ,$$

where $\mathbf{O}_3$ is a 3×3 zero matrix, and

$$\mathbf{G}(\widetilde{\mathbf{m}}) \equiv \mathbf{b}'\widetilde{\mathbf{m}}^T\mathbf{B}''^T - \mathbf{B}'\widetilde{\mathbf{m}}\mathbf{b}''^T \ .$$

If $\widetilde{\mathbf{m}} = [u, v, t]^T$ (in general $t = 1$), then $\mathbf{G}(\widetilde{\mathbf{m}})$ can be expressed as the sum of three matrices:

$$\mathbf{G}(\widetilde{\mathbf{m}}) = u\mathbf{K} + v\mathbf{L} + t\mathbf{M} \tag{5}$$

with

$$\mathbf{K} = \mathbf{b}'\mathbf{b}_1''^T - \mathbf{b}_1'\mathbf{b}''^T \ , \tag{6}$$

$$\mathbf{L} = \mathbf{b}'\mathbf{b}_2''^T - \mathbf{b}_2'\mathbf{b}''^T \ , \tag{7}$$

$$\mathbf{M} = \mathbf{b}'\mathbf{b}_3''^T - \mathbf{b}_3'\mathbf{b}''^T \ . \tag{8}$$

Here, $\mathbf{b}_i'$ is the i^{th} column vector of $\mathbf{B}'$ (i.e., $[\mathbf{b}_1', \mathbf{b}_2', \mathbf{b}_3'] \equiv \mathbf{B}'$]); similarly for $\mathbf{b}_i''$ (i.e., $[\mathbf{b}_1'', \mathbf{b}_2'', \mathbf{b}_3''] \equiv \mathbf{B}''$]). It is easy to see that the three matrices $\mathbf{K}$, $\mathbf{L}$, and $\mathbf{M}$ are all singular. By defining the following operation:

$$(\mathbf{K}, \mathbf{L}, \mathbf{M}) * \widetilde{\mathbf{m}} = u\mathbf{K} + v\mathbf{L} + t\mathbf{M} \ , \tag{9}$$

we finally obtain the following matrix equation for points between three views

$$[\widetilde{\mathbf{m}}']_\times \left[(\mathbf{K}, \mathbf{L}, \mathbf{M}) * \widetilde{\mathbf{m}}\right] [\widetilde{\mathbf{m}}'']_\times = \mathbf{O}_3 \ . \tag{10}$$

We have therefore a set of nine equations, of which not all are independent. And it can be shown that there are only *four linearly independent equations* in the 27 elements of matrices $\mathbf{K}$, $\mathbf{L}$ and $\mathbf{M}$, and that there are exactly *three algebraic equations* in the camera parameters. The latter is easily understood: Equations (1) to (3) have 9 scalar equations but 6 unknowns ($\mathtt{M}$, s, s' and s').

For full perspective, equation (10) is trilinear in image coordinates, i.e., each term contains at most one coordinate of a point. For affine cameras, matrices $\mathbf{K}$, $\mathbf{L}$ and $\mathbf{M}$ are in the following form:

$$\begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 0 \end{bmatrix} , \quad \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 0 \end{bmatrix} , \quad \text{and} \quad \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & 0 \end{bmatrix} ,$$

respectively. Equation (10) is then linear in image coordinates.

If 4 or more views are considered, no more information will be available than if we consider any subset of 3 views among them [11]. A point in an additional view adds two equations exactly in the same way as in the third view.

Acknowledgment

We thank Olivier Faugeras, Stphane Laveau and Charlie Rothwell for valuable discussions, and anonymous reviewers for comments.

References

1. J. Aloimonos. Perspective approximations. *Image and Vision Computing*, 8(3):179–192, Aug. 1990.
2. N. Ayache. *Artificial Vision for Mobile Robots*. MIT Press, 1991.
3. P. Beardsley, A. Zisserman, and D. Murray. Navigation using affine structure from motion. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 85–96, Stockholm, Sweden, May 1994. Springer-Verlag.
4. S. Demey, A. Zisserman, and P. Beardsley. Affine and projective structure from motion. In *British Machine Vision Conference*, pages 49–58, Leeds, UK, Sept. 1992.
5. O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 563–578, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.
6. O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
7. O. Faugeras. Stratification of 3-D vision: projective, affine, and metric representations. *Journal of the Optical Society of America A*, 12(3):465–484, Mar. 1995.
8. O. Faugeras and S. Laveau. Representing three-dimensional data as a collection of images and fundamental matrices for image synthesis. In *Proceedings of the International Conference on Pattern Recognition*, pages 689–691, Jerusalem, Israel, Oct. 1994. Computer Society Press.
9. O. Faugeras, S. Laveau, L. Robert, C. Zeller, and G. Csurka. 3-d reconstruction of urban scenes from sequences of images. In A. Gruen, O. Kuebler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 145–168, Ascona, Switzerland, Apr. 1995. ETH, Birkhauser Verlag. also INRIA Technical Report 2572.
10. O. Faugeras, T. Luong, and S. Maybank. Camera self-calibration: theory and experiments. In G. Sandini, editor, *Proc 2nd ECCV*, volume 588 of *Lecture Notes in Computer Science*, pages 321–334, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.
11. O. Faugeras and B. Mourrain. About the correspondences of points between n images. In *Proceedings of the Workshop on the representation of visual scenes*, Cambridge, Massachusetts, USA, June 1995.
12. 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, Oct. 1991.

13. R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 761–764, Urbana Champaign, IL, June 1992. IEEE.
14. R. Hartley and P. Sturm. Triangulation. In *Proceedings of the ARPA Image Understanding Workshop*, pages 957–966. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc., 1994.
15. J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *Journal of the Optical Society of America*, A8:377–385, 1991.
16. Q.-T. Luong and O. D. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *The International Journal of Computer Vision*, 17(1):43–76, Jan. 1996.
17. Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume 1 of *Lecture Notes in Computer Science*, pages 589–599, Stockholm, Sweden, May 1994. Springer-Verlag.
18. S. Maybank. *Theory of reconstruction From Image Motion*. Springer-Verlag, 1992.
19. J. L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, 1992.
20. E. Nishimura, G. Xu, and S. Tsuji. Motion segmentation and correspondence using epipolar constraint. In *Proc. 1st Asian Conf. Computer Vision*, pages 199–204, Osaka, Japan, 1993.
21. L. Quan and R. Mohr. Towards structure from motion for linear features through reference points. In *IEEE Workshop on Visual Motion*, New Jersey, 1991.
22. C. Rothwell, G. Csurka, and O. Faugeras. A comparison of projective reconstruction methods for pairs of views. In *Proceedings of the 5th International Conference on Computer Vision*, pages 932–937, Boston, MA, June 1995. IEEE Computer Society Press.
23. P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.
24. L. Shapiro. *Affine Analysis of Image Sequences*. PhD thesis, University of Oxford, Department of Engineering Science, Oxford, UK, Nov. 1993.
25. L. Shapiro, A. Zisserman, and M. Brady. Motion from point matches using affine epipolar geometry. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, volume II of *Lecture Notes in Computer Science*, pages 73–84, Stockholm, Sweden, May 1994. Springer-Verlag.
26. 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.
27. A. Shashua and N. Navab. Relative affine structure: Theory and application to 3D reconstruction from perspective views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, June 1994. IEEE.
28. M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. Technical Report CAR-TR-482, Computer Vision Laboratory, University of Maryland, Dec. 1989.
29. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *The International Journal of Computer Vision*, 9(2):137–154, 1992.
30. P. Torr and D. Murray. Stochastic motion clustering. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision*, pages 328–337, Vol.II, Stockholm, Sweden, May 1994.
31. D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *Proceedings of the 4th International Conference on Computer Vision*, pages 675–682, Berlin, Germany, May 1993. IEEE Computer Society Press.
32. G. Xu and Z. Zhang. *Epipolar Geometry in Stereo, Motion and Object Recognition*. Kluwer Academic Publishers, 1996.
33. C. Zeller and O. Faugeras. Applications of non-metric vision to some visual guided tasks. In *Proceedings of the International Conference on Pattern Recognition*, pages 132–136, Jerusalem, Israel, Oct. 1994. Computer Society Press. A longer version in INRIA Tech Report RR2308.
34. Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. Technical Report 2927, INRIA Sophia-Antipolis, France, July 1996. To appear in the *International Journal of Computer Vision*.
35. 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, Oct. 1995.
36. Z. Zhang, O. Faugeras, and R. Deriche. An effective technique for calibrating a binocular stereo through projective reconstruction using both a calibration object and the environment. *Videre: A Journal of Computer Vision Research*, 1(1):58–68, 1997.
37. Z. Zhang, K. Isono, and S. Akamatsu. Euclidean structure from uncalibrated images using fuzzy domain knowledge: Application to facial images synthesis. In *Proceedings of the 6th International Conference on Computer Vision*, Bombay, India, Jan. 1998. IEEE Computer Society Press.
38. A. Zisserman. Notes on geometric invariants in vision. BMVC92 Tutorial, 1992.

Zhengyou Zhang is a Senior Research Scientist at INRIA (French National Institute for Research in Computer Science and Control), France. He was born in China, received the B.S. degree in electronic engineering from the University of Zhejiang, China, in 1985, the M.S. in computer science from the University of Nancy, France, in 1987, the Ph.D. degree in computer science from the University of Paris XI, France, in 1990, and the Doctor of Science (*Habilitation à diriger des recherches*) diploma from the University of Paris XI, in 1994.

He works at INRIA since 1987, and has been a Senior Research Scientist since 1991. In 1996–1997, he spent one year sabbatical as an Invited Researcher at Advanced Telecommunications Research Institute International (ATR), Human Information Processing Research Laboratories, Kyoto, Japan. He is also a Guest Research Professor in Chinese Academy of Sciences since 1994, and a Part-time Professor in Northern Jiaotong University (Beijing, China) since 1996. His current research interests include computer vision, dynamic scene analysis, vision and graphics, facial image analysis, and visual learning. He is a Senior Member of IEEE, an Associate Editor of the *International Journal of Pattern Recognition and Artificial Intelligence*, 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); *Computer Vision* (textbook in Chinese, Chinese Academy of Sciences, 1997).

Gang Xu received the B.E. degree in radio engineering from the Nanjing Institute of Technology (now Southeast University) in 1982, the M.S. and Ph.D. degrees in Control Engineering from Osaka University, in 1986 and 1989, respectively. From 1990 to 1996, he was an assistant professor at the Department of Control Engineering, Osaka University. In 1994, he was a visiting scientist in the Harvard Robotics Laboratory. In April 1996, He became an associate professor of the Department of Computer Science, Ritsumeikan University, and a co-director of the Rits Computer Vision Laboratory. His research interests include computer vision, computer graphics and other image-related fields. He has coauthored two books: *Epipolar Geometry in Stereo, Motion and Object Recognition* (Kluwer Academic Publishers, 1996) and *Three-Dimensional Vision* (textbook in Japanese, Kyoritsu, 1998).