Structure from Motion from Three Affine Views

Long QUAN Maxime LHUILLIER
Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong.

Abstract

We describe a new method for Structure From Motion from three affine views. The central idea of the method is to explore the intrinsic three-view properties instead of previous two-view ones. The first key observation is that an affine camera is indeed essentially a one-dimensional projective camera operating on the plane at infinity : we prove that the essential motion—relative camera orientations—is entirely encoded by the infinity 1D trifocal tensor. From a practical point of view, this analysis allows the development of two new algorithms of SFM from three views. One based on entirely the minimal trifocal tensor and another on affine three-view constraints. Both algorithms are novel as all previous SFM from three views have been heavily based on only two-view constraint to extract Euclidean structure. These algorithms have been demonstrated on real image sequences.

1. Introduction

Motion/structure from orthographic or weak perspective views is a very old and popular topic. It is well known that at least 4 non-planar points over 3 orthographic or weak perspective views are sufficient to uniquely determine motion/structure up to a reflection about the image plane [21, 7, 9, 16]. Many algorithms have been published for this problem: the linear methods of Huang and Lee [7, 10], non-linear algebraic methods of Koenderink and Van Doorn [9, 2] and non-linear numerical method of Shapiro *et al.* [14]. A good review of the different methods can be found in [14]. The main drawback of existing methods is that they are essentially based only on two-view constraints. Very recently, multiple affine-view constraints have been intensively studied ([13, 8, 1, 19, 3]) the method proposed in [13] combined both 3-view and 2-view constraints. However no method exists for SFM from only three-view constraints probably due to the complicated relationship between euclidean motion parameters and 3-view constraints.

The central idea of this paper is to fully exploit the three-view constraints as they encode much richer motion information as the two-view constraints do. The first key observation is that an affine camera is indeed essentially a one-dimensional projective camera operating on the plane at infinity, as 1D projective cameras are encoded by the 1D trifocal tensor, we show that the essential motion is encoded by the infinity 1D trifocal tensor. Different algorithms are also proposed to determine motion parameters from the trifocal tensor. From a practical point of view, this analysis allows the development of two new algorithms of SFM from three views. One based on the minimal trifocal tensor and another on redundant affine three-view constraints. Both algorithms are novel as all previous SFM from three views have been based on only 2-view constraint to extract euclidean structure. These algorithms have been demonstrated on real image sequences.

The paper is organized as follows. In Section 2, we review the affine camera and the 1D projective camera models and discuss their relationship. Then, we describe how to determine motion parameters from the infinity trifocal tensor in Section 3. The computation of the infinity trifocal tensor is presented and discussed in Section 4. A short conclusion and future perspectives are given in Section 5.

2. 2D Affine and 1D Projective Cameras

Notation Throughout the paper, vectors are denoted in lower case boldface $\mathbf{x}$, $\mathbf{u}$ ..., matrices and tensors in upper case boldface $\mathbf{A}$, $\mathbf{T}$ ... (sometimes, matrix dimensions are made clearer with subscripts like $\mathbf{A}_{i\times j}$); Scalars are any plain letters or lower case Greek $a, u, A, \lambda \ldots$. The geometric objects are sometimes denoted by plain or Greek letters like l for a 2D line and L for a 3D line whenever it is necessary to distinguish the geometric object l from its coordinate representation by a vector $\mathbf{l}$. Covariant indices are written as subscripts and contravariant indices as superscripts. e.g. the coordinates of a point $\mathbf{x}$ in $\mathcal{P}^3$ are written with an upper index $\mathbf{x} = (x^1, x^2, x^3, x^4)^T$. The implicit summation convention is also adopted: Any index repeated as subscript and superscript in a term involving vectors, matrices and tensors implies a summation over the range of index values. e.g. the ith coordinate of the matrix product $\mathbf{Ax}$ is $A^i_j x^j$.
2D affine camera The affine camera first introduced by

Mundy and Zisserman [11] is the uncalibrated version of orthographic, weak-perspective and para-perspective projection models. It also describes a common degeneracy of the projective camera either when the viewing field is narrow or the scene is shallow compared to the average distance from the camera. Its broad usage not only lies in its algebraic simplicity, it is unavoidable for better numerical stability as it prevents the algorithms from their inherent ill-conditioning.

The key property is that parallelism is preserved by the affine camera $\mathbf{A}_{3\times4}$ so that the plane at infinity has been identified and the points at infinity are projected into points at infinity. The principal plane is sent to be confused with the plane at infinity, this is equivalent to having the third row of the projective camera matrix fixed as $(0,0,0,1)$ if the plane at infinity is identified as $x_4 = 0$:

$$\mathbf{A}_{3\times4} = \begin{bmatrix} \mathbf{M}_{2\times3} & \mathbf{t}_{2\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix}.$$

Finite points $\mathbf{x} = (\mathbf{x}_a, 1)^T$ are projected onto finite image points $\mathbf{u} = (\mathbf{u}_a, 1)^T$ as $\mathbf{u}_a = \mathbf{M}_{2\times3}\mathbf{x}_a + \mathbf{t}$. If we further use relative coordinates of the points with respect to a given reference point (for instance, the centroid of a set of points), the translation component $\mathbf{t}$ is canceled and the projection equation for finite points in relative coordinates are therefore $\Delta\mathbf{u} = \mathbf{M}_{2\times3}\Delta\mathbf{x}$.

This last equation is the basic projection equation for points in an affine camera when relative coordinates are used, and will be only denoted as $\mathbf{u} = \mathbf{M}_{2\times3}\mathbf{x}$ with always the implicit assumption that the centroid has been selected as the reference point throughout the paper.

1D projective camera One-dimensional projective camera has been first abstracted from the study of the geometry of lines under affine cameras [12, 5, 18]. It can also be defined by simple analogy to a 2D projective camera operating on lower dimensions.

A 1D projective camera projects a point $\mathbf{x} = (x^1, x^2, x^3)^T$ in $\mathcal{P}^2$ (projective plane) to a point $\mathbf{u} = (u^1, u^2)^T$ in $\mathcal{P}^1$ (projective line). This projection may be described by a 2×3 homogeneous matrix $\mathbf{M}$ as $\lambda\mathbf{u} = \mathbf{M}_{2\times3}\mathbf{x}$.

We now examine the geometric constraints available for points seen in multiple views similar to the 2D camera case [15, 17, 6, 20, 4]. There is a constraint only in the case of 3 views, as there is no any constraint for 2 views (two projective lines always intersect in a point in a projective plane).

Let the three views of the same point $\mathbf{x}$ be given as follows: $\lambda\mathbf{u} = \mathbf{M}\mathbf{x}$, $\lambda'\mathbf{u}' = \mathbf{M}'\mathbf{x}$, $\lambda''\mathbf{u}'' = \mathbf{M}''\mathbf{x}$. These can be rewritten in matrix form as

$$\begin{pmatrix} \mathbf{M} & \mathbf{u} & 0 & 0 \\ \mathbf{M}' & 0 & \mathbf{u}' & 0 \\ \mathbf{M}'' & 0 & 0 & \mathbf{u}'' \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ -\lambda \\ -\lambda' \\ -\lambda'' \end{pmatrix} = 0. \tag{1}$$

The vector $(\mathbf{x}, -\lambda, -\lambda', -\lambda'')^T$ cannot be zero, so

$$\begin{vmatrix} \mathbf{M} & \mathbf{u} & 0 & 0 \\ \mathbf{M}' & 0 & \mathbf{u}' & 0 \\ \mathbf{M}'' & 0 & 0 & \mathbf{u}'' \end{vmatrix} = 0. \tag{2}$$

The expansion of this determinant produces a trifocal constraint for the three views $T_{ijk}u^i u'^j u''^k = 0$, where T_{ijk} is a $2 \times 2 \times 2$ homogeneous tensor whose components T_{ijk} are 3×3 minors (involving all three views) of the 6×3 joint projection matrix by stacking $\mathbf{M}$, $\mathbf{M}'$ and $\mathbf{M}''$: $T_{ijk} = \det \begin{pmatrix} \epsilon_{ii'} \boldsymbol{M}^{i'} \\ \epsilon_{jj'} \boldsymbol{M}'^{j'} \\ \epsilon_{kk'} \boldsymbol{M}''^{k'} \end{pmatrix}$, where all indices vary from 1 to 2, $\boldsymbol{M}^i$ is the i-th line of $\boldsymbol{M}$ and $\epsilon_{11} = \epsilon_{22} = 0, \epsilon_{12} = -\epsilon_{21} = 1$.

This trifocal tensor encapsulates exactly the information needed for projective reconstruction in $\mathcal{P}^2$. Namely, it is the unique matching constraint, it minimally parameterizes the three views and it can be estimated linearly. Contrast this to the 2D image case in which the multi-linear constraints are algebraically redundant and the linear estimation is only an approximation based on over-parameterization.

2D affine camera is a 1D projective camera on the plane at infinity If we restrict ourself to the plane at infinity on which the affine camera direction (i.e. the affine camera center at infinity) lies, indeed a 1D projective camera occurs naturally. The 1D image line is the intersection of the image plane with the plane at infinity, i.e. the line at infinity of the image plane. The points at infinity are projected onto the points at infinity of the image plane which is now the 1D image line as affine camera preserves the points at infinity.

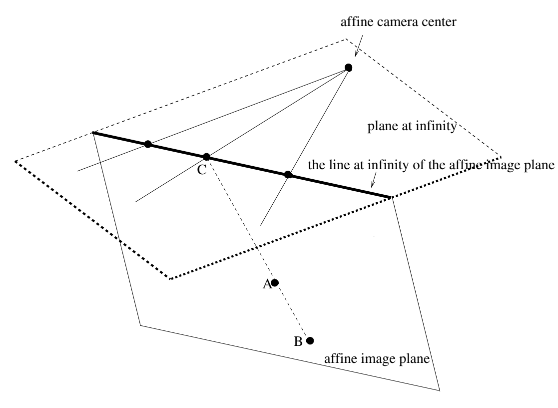

Figure 1. The relation between an affine camera and a 1D projective camera on the plane at infinity: for instance, any point pair A and B on the affine image plane gives also a 1D image point C on the line at infinity of the affine image plane.

3. SFM from infinity trifocal tensor

As 1D camera is governed by the trifocal tensor as just described above so does an affine camera, by its infinity trifocal tensor. The essential motion parameters should be contained in this trifocal tensor.

3.1. Extraction of epipoles from the trifocal tensor

The viewing directions/centers of the affine cameras are the centers of the 1D projective cameras on the plane at infinity. The 1D images of the camera centers are 1D epipoles. We first show that the epipoles are not algebraically independent, they are implicit in the infinity trifocal tensor.

From the fundamental trilinear constraint of the infinity trifocal tensor $T_{ijk}u^i u'^j u''^k = 0$, dualizing the tensor to move the first image coordinate to the right side, the trilinear constraint appears as a transfer equation from the second and third image coordinates: $\epsilon^{ii'} T_{ijk} u'^j u''^k \sim u^{i'}$, that is $T^i_{jk} u'^j u''^k \sim u^i$ ($\sim$ is the equality for all i up to a scale).

Now take the first camera center as a usual point, then its images should be respectively $\boldsymbol{0}$, $\boldsymbol{e}'_1$ and $\boldsymbol{e}''_1$ in the three images. As all corresponding points satisfy the trilinear constraint, so does the triplet of corresponding image points $\boldsymbol{0}$, $\boldsymbol{e}'_1$ and $\boldsymbol{e}''_1$. We have $T^i_{jk} {e'_1}^j {e''_1}^k = 0^i$.

This matrix equation says that the 2×2 matrix $T^i_{jk} {e'_1}^j$ has the vector $\boldsymbol{e}''_1$ as its kernel! The vector $\boldsymbol{e}''_1$ being an epipole in the third image could not be a zero-vector, so the matrix $T^i_{jk} {e'_1}^j$ must have rank 1. *I.e.*

$$\det(T^i_{jk} {e'_1}^j) = 0. \quad (3)$$

The expansion of this 2×2 determinant ends up as a quadratic equation in the unknown components of the epipole $\boldsymbol{e}'_1$.

Proceeding similarly by first dualizing the tensor for it being transfer equation from the first and second to the third image $T^k_{ij} u^i u'^j \sim u''^k$, then taking the corresponding triplet image points $\boldsymbol{e}_3$, $\boldsymbol{e}'_3$ and $\boldsymbol{0}$ of the third camera center, we obtain $T^k_{ij} e^i_3 {e'}^j_3 = 0^i$. We make the same observation that the matrix $T^k_{ij} {e'}^j_3$ has rank one, so that $\det(T^k_{ij} {e'}^j_3) = 0$. As $T^k_{ij} {e'}^j_3$ is a 2×2 matrix, the exchange of up and low indices makes no difference on the determinant constraint, so we have equally, $\det(T^i_{kj} {e'}^j_3) = 0$.

Thus, the determinant expansion leads exactly to the same quadratic equation (3) for both $\boldsymbol{e}'_1$ and $\boldsymbol{e}'_3$, so the unique quadratic equation gives two solutions, one is $\boldsymbol{e}'_1$ and the other is $\boldsymbol{e}'_3$, but the ordering remains undetermined.

Dualizing the tensor and contracting the tensor in different manners with the different camera centers, we obtain similar quadratic equations for all six epipoles. This gives a constructive proof that indeed all epipoles are contained in the trifocal tensor and are obtained by solving simple quadratic equations up to a two-way ambiguity of different ordering.

As such epipole is defined up to a scale, it counts only for 1 d.o.f., it is natural to see that the 7 d.o.f. tensor conveys more information than all epipoles, it still has 1 d.o.f. after removing the 6 d.o.f. for the 6 epipoles.

More practical estimation methods of trifocal tensor and epipoles in different configurations will be discussed in Section 4.

3.2. Determination of 1D camera matrices

Projection matrix gives a convenient and complete representation for cameras. The 1D camera matrices could be determined from the infinity trifocal tensor. This task is much simplified whence the epipoles have been determined from the trifocal tensor. Without loss of generality, we can always take the following normal forms for the 3 projection matrices $\mathbf{M} = \begin{pmatrix} \mathbf{I}_{2\times 2} & \mathbf{0} \end{pmatrix}, \mathbf{M}' = \begin{pmatrix} \mathbf{A}_{2\times 2} & \mathbf{e}'_1 \end{pmatrix}, \mathbf{M}'' = \begin{pmatrix} \mathbf{B}_{2\times 2} & \mathbf{e}''_1 \end{pmatrix}$. By noticing that each trifocal tensor component is linear in the entries of $\boldsymbol{A}$ and $\boldsymbol{B}$ for known $\boldsymbol{e}'_1$ and $\boldsymbol{e}''_1$, the estimate of the camera matrices is straightforward.

We can also notice that the camera matrices could not be determined only from epipoles, i.e. the trifocal tensor's extra d.o.f. allows the full determination of projection matrices.

3.3. Metrics of affine camera and the absolute conic:

The metric upgrade of affine structure for affine cameras is easier than projective camera as the plane at infinity has already been identified so that instead of the absolute quadric which is a space conic, we need only to determine a plane conic, i.e. the space conic restricted to the known plane at infinity. As the projection center of affine cameras lies on the plane at infinity, the projection of the absolute conic is now $\boldsymbol{M}\boldsymbol{\Omega}_\infty\boldsymbol{M}^T$ in dual coordinates, i.e. 1D dual absolute conic which is the pair of absolute points (or circular points) on the line at infinity of the affine image plane. As the infinity 1D projection center is the viewing direction of the affine camera, so the line at infinity of the affine image plane which is the intersection of the affine image plane with the plane at infinity is the polar line of the camera center w.r.t. the 2D absolute conic. The pair of absolute lines are tangent to the 2D absolute conic and touch it at the pair of circular points of the affine image plane as illustrated in Figure 2.

Determination of the absolute conic on the ∞-plane

- **Calibrated cameras** In 2D Euclidean space, the pair of 'circular points' characterizes the Euclidean structure of the plane. The known aspect ratio for the affine

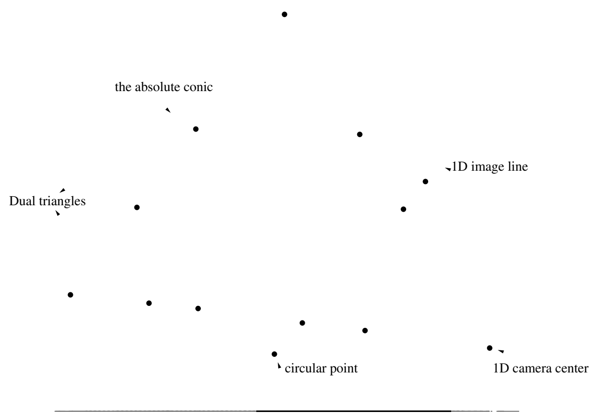

Figure 2. The metrics of three affine views on the plane at infinity: The triangle of 1D camera centers and the triangle of the 1D image lines are dual as the polar line of the 1D camera center is the image line w.r.t. the absolute conic.

camera is equivalent to the knowledge of the circular points on the affine image plane. The dual of the absolute conic on the plane at infinity could be determined by observing that the viewing rays of the circular points of each affine image plane are tangent to the absolute conic through the camera center.

If we take the dual form of the absolute conic ($\mathbf{\Omega}_\infty$), the viewing lines of the circular points $\mathbf{l} = \mathbf{c} \times I$ satisfy $\mathbf{l}^T \mathbf{\Omega} \mathbf{l} = 0$, just like a 'point'. Therefore $\mathbf{\Omega}_\infty$ can be linearly fitted from the 6 tangent lines.

- **Auto-calibration of 2D affine camera** If the cameras are uncalibrated, auto-calibration is also possible.

  Now take the internal calibration matrix $\boldsymbol{K}_{i,2\times 2} \sim \begin{pmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{pmatrix}$ for each affine image,

$$\boldsymbol{M}_{i,2\times 3} \mathbf{\Omega}_\infty \boldsymbol{M}_{i,2\times 3}^\top \sim \boldsymbol{\omega}_{2\times 2} = \boldsymbol{K}_i \boldsymbol{K}_i^T,$$

  where $\mathbf{\Omega}_\infty$ is the dual absolute conic on the plane at infinity and $\boldsymbol{\omega}_{2\times 2}$ is the line equation of the image of the absolute point-pair.

  The unknown scale factors can also be eliminated by treating the 2×2 symmetric matrices as 3-vectors, for the constant unknown aspect ratio, the auto-calibration constraints can be written as

$$(\boldsymbol{M}_i \mathbf{\Omega}_\infty \boldsymbol{M}_i^\top)_{ij} (\boldsymbol{\omega}_{2\times 2})_{mn} - (\boldsymbol{\omega}_{2\times 2})_{ij} (\boldsymbol{M}_i \mathbf{\Omega}_\infty \boldsymbol{M}_i^\top)_{mn} = 0$$
$$\text{where} \quad i \leq j, m \leq n = 1 \ldots 2.$$

  Five independent parameters are required to specify the Euclidean structure from the affine structure: the 5 parameters of the absolute conic on the plane at infinity. Since each image gives 2 independent constraints, generally 5 images are necessary for the five intrinsic calibration parameters. For constant unknown aspect ratio of the moving camera, three images are enough.

Determination of camera orientation parameters: Transforming the absolute conic to its canonical position converts all projective quantities into their true Euclidean counterparts. Euclidean 1D camera centers give the orientation of the affine cameras. Let the eigen-decomposition of $\mathbf{\Omega}_\infty$ be $\mathbf{QDQ}^T$, we can take $\mathbf{A} = \mathbf{QD}^{1/2}$. Then the projective transformation $\mathbf{A}$ which brings $\mathbf{\Omega}_\infty$ back to its true position $\mathbf{I}_{3\times 3}$ (i.e. $\mathbf{\Omega}_\infty = \mathbf{AIA}^T$), converts projective coordinates $\mathbf{x}$ into Euclidean $\mathbf{A}^{-1}\mathbf{x}$. All other remaining motion and structure parameters could be straightforwardly determined from this transformation.

4. From 3 affine views to the ∞-trifocal tensor

Now we describe how the infinity trifocal tensor could be estimated from multiple affine-view constraints. These constraints have been recently studied in [13, 1, 19, 14] following the same spirit for projective cameras [15, 17, 6, 20, 4]. We first briefly summarize the matching constraints for three affine views, then describe different strategies for obtaining the infinity trifocal tensor.

Matching constraints The most straightforward way is to use line correspondences as line directions naturally sit on the plane at infinity and satisfy the trilinear constraints on the plane at infinity [12]. The infinity trifocal tensor can therefore be linearly estimated from at least 7 line correspondences. All orientation components of the motion could be computed by extracting the epipoles from the infinity trifocal tensor as described in Section 3. The two-way ambiguity of orientations and the translation component could be uniquely fixed by the information provided by the position of the line features that has not yet been exploited [1, 12]. Compared with the method presented in [12] which was based on an explicit parameterization of the projection matrices by solving a quadratic equation in which the geometric interpretation was lost, the new method presented in Section 3 uses the decomposition of the trifocal tensor into epipoles as the intermediate step, this makes clear the geometric interpretation of the intrinsic relationship between trifocal tensor and epipoles and the inherent two-way ambiguity.

Less obvious, but it is still a trivial observation that not only line directions are governed by the infinity trifocal tensor, but also all point correspondences in relative coordinates as a relative point is *bona fide* a 'line segment' between the point and the reference point! The advantage of directly estimating the infinity trifocal tensor using trilinear constraints is that the parameter set is kept minimal so the numerical efficiency and consistency are guaranteed, but the main drawback is that the minimum number of points is five instead of 4 and there exists two-way ambiguity for extracting epipoles. Needless to say, points and lines mixed up nicely in this framework.

We can also use all available matching constraints of three affine views [13, 8]. Let the three views of the same point $\mathbf{x}$ be given as $\mathbf{u} = \mathbf{M}\mathbf{x}, \mathbf{u}' = \mathbf{M}'\mathbf{x}, \mathbf{u}'' = \mathbf{M}''\mathbf{x}$. These can be rewritten together in matrix form as $\begin{pmatrix} \mathbf{M} & \mathbf{u} \\ \mathbf{M}' & \mathbf{u}' \\ \mathbf{M}'' & \mathbf{u}'' \end{pmatrix} \binom{\mathbf{x}}{\lambda} = 0$, where $\lambda \neq 0$ encodes the (unrecoverable) global scale factor of the reconstruction. As the vector $(\mathbf{x}, \lambda)^T$ can not be zero, the rank of the coefficient matrix is at most 3, so all of its 4×4 minors vanish. Each vanishing 4×4 minor gives one constraint for the corresponding point. There are $C_6^4 = 15 = 3 + 4 + 4 + 4$ such minors, which can be divided into two types:

- Three 2-view constraints involving only two views with two rows from each view, $\boldsymbol{e}'^T \boldsymbol{u} + \boldsymbol{e}^T \boldsymbol{u}' = 0$ where $\boldsymbol{e}^i = \det \begin{pmatrix} \epsilon_{ii'} \boldsymbol{M}^{i'} \\ \boldsymbol{M}' \end{pmatrix}$ and $\boldsymbol{e}'^i = \det \begin{pmatrix} \boldsymbol{M} \\ \epsilon_{ii'} \boldsymbol{M}'^{i'} \end{pmatrix}$.
These are the affine epipolar geometry.
- Three sets of four 3-view constraints involving all three views with two rows from one view and one from each of the others, for instance, if we choose always two rows from the first view, we obtain
$T_{ijk}\boldsymbol{u}^i - \boldsymbol{e}''\boldsymbol{u}'^T + \boldsymbol{u}''\boldsymbol{e}'^T = \boldsymbol{0}_{2\times 2}, T_{ijk} = \det \begin{pmatrix} \epsilon_{ii'} \boldsymbol{M}^{i'} \\ \epsilon_{jj'} \boldsymbol{M}'^{j'} \\ \epsilon_{kk'} \boldsymbol{M}''^{k'} \end{pmatrix}$.
These are in fact uncalibrated version of the transfer equations over three views [22] that has been extensively used in object recognition.

12- and 20-parameter methods Using three two-view constraints for SFM has been the bases of almost all existing methods [7, 14, 16] due to the simple relation between the components of two-view constraints and the motion parameters. In [13], 2 two-view constraint and 1 three-view constraint are put together to form a 9-parameter linear system. Here we can use only the four three-view constraints as three-view constraints convey richer information than two-view constraints and therefore are preferable. The four three-view constraints give a **12-parameter** linear system including the 8 trifocal tensor components T_{ijk} and 4 for the two epipoles $\boldsymbol{e}'$ and $\boldsymbol{e}''$. The other four unknown epipoles could be extracted from T_{ijk} with the same procedure introduced in Section 3. The interesting thing here is that the two-way ambiguity of the computation of epipoles is removed thanks to that the two of the six epipoles have been uniquely determined.

By permuting three images, two other sets of 4 three-view constraints could be obtained. Obviously these constraints are algebraically redundant. Nevertheless, a **20-parameter** linear system could be obtained if we stack all three sets of 4 three-view constraints and three two-view constraints together as suggested in [8]. The main advantage is that all epipoles are simultaneously estimated, but it is a heavy over-parameterization and both efficiency and consistency of epipoles are difficult to handle.

Experiments The triplet of images on which the experiments have been performed is shown in the top of Figure 3. Face images give a typical good example of using affine camera model. About 50 points of interests are detected in each image and finally 29 of them are automatically matched across the three views through a multi-step robust matching procedure. The epipolar geometry of the three pairs of images computed using only four three-view constraints is also displayed on the original images in Figure 3. Both residual error of point-to-epipolar-line distances and visual inspection suggest excellent determination of the relative orientations between views.

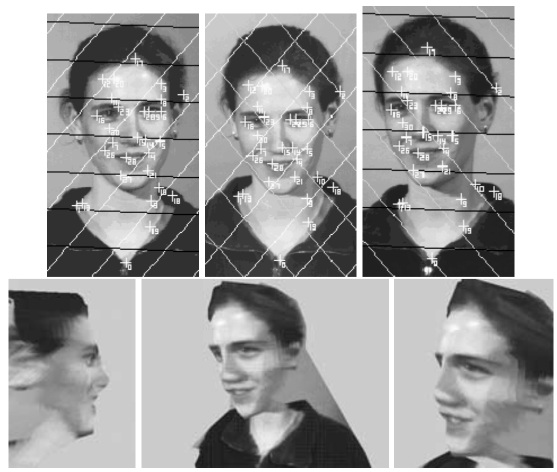

Figure 3. Top: a triplet of face images and the affine epipolar geometries for the three pairs of images computed from the three-view constraints. Bottom: some views of the reconstructed VRML model

A dense matching procedure is performed for each pair of views. The final disparity map for the triplet is obtained by combining those of the three pairs of images. Following the calibration method described in Section 3, the circular points derived from the known aspect ratio of the camera are used to fit the absolute conic. The eigen-decomposition of the matrix of the dual absolute conic fixes the metric properties of both structure and motion. The dense metric 3D reconstruction is obtained and a VRML has been created. Some sample views of the VRML file are illustrated in the bottom of Figure 3. We see that the depth for the major face parts is well determined from the side views of the face. The observed non-smooth transition in nose and mouth parts is due to imprecision of dense matching which has not been regularized. One of the key computational innovations described in this paper is to compute epipoles not from the direct affine epipolar geometry, but either from the minimal 8-parameter trifocal tensor or the 12-parameter three-view constraints. The epipoles computed from different methods are tabulated to compare the accuracy. Table 1 gives the comparative results between the 20-, 12- and 8-parameter

epipoles	e_2	e_3	e'_1	e'_3	e''_1	e''_2
20-param.	(0.648,-0.761)	(0.999,0.0432)	(0.650,0.760)	(0.647,-0.762)	(0.999,0.0364)	(0.642,0.767)
12-param.	(0.647,-0.762)	(0.999,0.04451)	(0.653,0.757)	(0.647,-0.762)	(0.999,0.0378)	(0.645,0.764)
8-param.	(0.705, -0.708)	(0.866,0.501)	(0.564,0.826)	(0.704,-0.710)	(0.928, 0.372)	(0.571, 0.821)
cosine 20-12	1	0.999999	0.999	1	0.999	0.999
cosine 20-8	0.997	0.886	0.994	0.997	0.941	0.996
cosine 8-12	0.997	0.887	0.994	0.997	0.942	0.996

Table 1. Table of the computed epipoles using the 20-, 12- and 8-parameter methods. The three last row show the cosine of the same epipolar direction obtained by two different methods.

methods. We note the extremely similar numerical behavior of 20- and 12- methods though the 12- method is clearly more efficient.

The computed epipoles from the 8-parameter method are still very stable, though the accuracy is gracefully degraded w.r.t. the 12- and 20-parameter methods. From all experiments we have conducted, we observe that the 8-parameter method gives very good results with its minimal parameterization. The 12-parameter and 20-parameter methods behave very similarly and both outperform the minimal 8-parameter method. As a trade-off between the efficiency/consistency and the ease of epipole extraction, the 12-parameter method is definitely the favorite method while the 8-parameter method is unavoidable for the case of pure line correspondences and some other minimal data cases.

5. Conclusion

We have developed a new method for structure from motion from three affine views. The central idea is to partly reduce the affine camera to a 1D projective camera on the plane at infinity. As 1D projective camera is governed by its trifocal tensor. We demonstrate that the essential motion parameters—the viewing directions of the affine cameras—are contained in the infinity trifocal tensor. We then described different methods of extracting orientation information from the trifocal tensor. This new approach contrasts greatly with the previous methods of structure from motion from affine views which have been heavily based on two-view constraints, and provides new insights on the intrinsic geometric structure of three affine views. From a computational point of view, it is worth investigating the benefits of the constraint nonlinear estimation methods using the tensor/epipole relations as the hard constraints.

References

[1] K. Astrom, A. Heyden, F. Kahl, and M. Oskarsson. Structure and motion from lines under affine projections. *ICCV'99*.

[2] B. Bennet, D. Hoffman, J. Nicola, and C. Prakash. Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America*, 6(7):1052–1069, 1989.

[3] L. Bretzner and T. Lindeberg. Feature tracking with automatic selection of spatial scales. *CVIU*, 71(3):385–392.

[4] O. Faugeras and B. Mourrain. About the correspondences of points between n images. *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, 1995.

[5] O. Faugeras, L. Quan, and P. Sturm. Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera. *ECCV'98*.

[6] R. Hartley. A linear method for reconstruction from lines and points. *ICCV'95*.

[7] T. Huang and C. Lee. Motion and structure from orthographic projections. *PAMI*, 11(5):536–540, 1989.

[8] F. Kahl and A. Heyden. Affine structure and motion from points, lines and conics. *IJCV*, 1999.

[9] J. Koenderink and A. van Doorn. Affine structure from motion. *Journal of the Opt. Soc. of America*, 8(2):377–385.

[10] C. Lee and T. Huang. Finding point correspondences and determining motion of a rigid object from two weak perspective views. *CVGIP*, 52:309–327, 1990.

[11] J. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. The MIT Press, Cambridge, MA, 1992.

[12] L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *PAMI*, 19(8):834–845, August 1997.

[13] L. Quan and Y. Ohta. A new linear method for euclidean motion/structure from three calibrated affine views. *CVPR'98*.

[14] L. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *IJCV*, 16(2):147–182.

[15] A. Shashua. Algebraic functions for recognition. *PAMI*, 17(8):779–789, August 1995.

[16] I. Shimshoni, R. Basri, and E. Rivlin. A geometric interpretation of weak-perspective motion. Technical report, Technion, 1999.

[17] M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. *Proceedings of DARPA IUW*, 1990.

[18] P. Sturm. *Vision 3D non calibrée : contributions à la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage*. PhD Thesis, INPG, France, 1997.

[19] T. Thorhallsson and D. Murray. The tensors of three affine views. *CVPR'99*.

[20] B. Triggs. Matching constraints and the joint image. *ICCV'95*.

[21] S. Ullman. *The Interpretation of Visual Motion*. The MIT Press, Cambridge, MA, USA, 1979.

[22] S. Ullman and R. Basri. Recognition by linear combinations of models. *PAMI*, 13(10):992–1006, 1991.