Repeated Structures: Image Correspondence Constraints and 3D Structure Recovery.

Joseph L. Mundy[1] and Andrew Zisserman[2]

[1] The General Electric Corporate Research and Development Laboratory, Schenectady, NY, USA.
[2] Robotics Research Group, Department of Engineering Science, Oxford University, England

Abstract. Recently, a number of classes of 3D structures have been identified which permit structure recovery and 3D invariants to be measured from a single image of the structure. A large class with this property is the case of *repeated structures* where a structure (such as a pointset, curve or surface), and a transformed copy of the structure are both observed in a single perspective image. In general the 3D reconstruction is only possible up to a 3D projectivity of space, but smaller ambiguities are possible, depending on the nature of the 3D transformation between the repeated structures. An additional theme of the paper is the development of feature correspondence relations based on the epipolar geometry induced in the image by the repeated structure. In some cases, correspondence is based on projective homologies rather than a true epipolar geometry.

1 Introduction

The motivation for this paper arises from the case of repeated structures imaged by a single camera. An image of a repeated structure is equivalent to multiple views of the single structure. For example, an image of an object and a copy of the object translated to a new position is identical in projection properties to two images of the single object obtained by translating the camera.

Recent work [1, 3] has demonstrated that 3D structure can be recovered from a pair of uncalibrated cameras, up to a 3D projective ambiguity in general. Thus, stereo analysis can be applied to a single, uncalibrated camera view of repeated structure to recover 3D structure with the same projective ambiguity.

Repeated geometric configurations often occur in man-made structures and even some natural forms. Some examples are:

- An object repeated in a single scene by an Euclidean transformation, e.g. a row of development houses along a street.
- An object repeated with a more general transformation such as affine or projective, e.g., two ends of a wrench are related by an affine transformation.
- Various 3D spatial symmetries such as,
  - bi-lateral symmetry (e.g. a spoon)

- discrete axial symmetry (e.g. a hex bolt)
- rotational symmetry (e.g. a vase)

These structures impose an epipolar geometry on a single image which determines image correspondences between related points. That is, a point in a structure determines a line through the epipole, on which the corresponding point must lie. An example is given in Figure 1. This epipolar relationship reduces the complexity of determining feature correspondences. An epipolar structure is not always possible. For example, in the case of repetition by reflection, when the plane of symmetry passes through the camera center. However, a correspondence structure still exists, based on a specialized projective transformation, a planar homology which maps corresponding image features of the repeated structures.

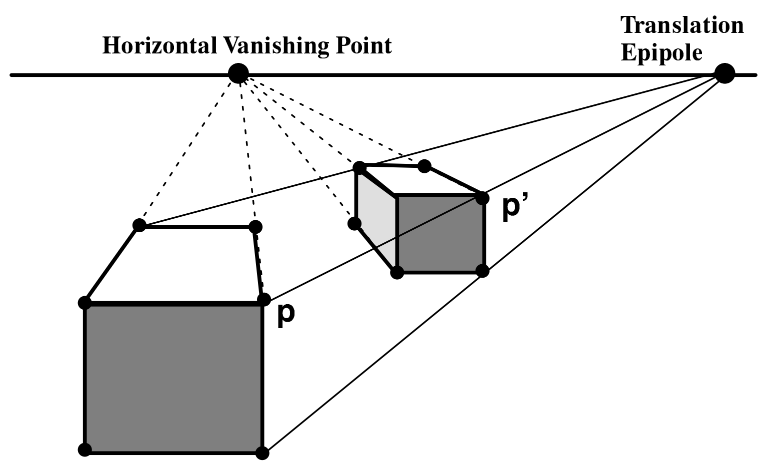

Fig. 1. Two identical structures separated by a horizontal 3D translation. For repetition by translation, corresponding points on the repeated structure generate lines which all converge to a single epipole. The case where corresponding points lie on the same epipolar line is called auto-epipolar.

The full 3D projective transformation is the most general ambiguity which arises from repeated structures viewed by an uncalibrated camera. Projective invariants can be constructed from a single image of repeated structures as for invariants of multiple views[9]. For example, six 3D points determine three independent projective invariants: five points determine an invariant projective coordinate frame so that the remaining point has three invariant coordinates in this five point basis. The ambiguity can be smaller if the camera calibration is partially known or if some restriction applies to the transformation which relates the repeated structures. When the coordinate frames of a stereo pair of identical uncalibrated cameras are related by translation alone, the resulting stereo reconstruction of space is ambiguous only up to a 3D affine transformation. This result was reported by Moons *et al* [7] and is elaborated elsewhere in this volume. The advantage of an affine reconstruction is that more invariants can be constructed from a given set of features[3].

In this paper we make the following contributions: First, a general process for determining 3D reconstruction ambiguity of various repeated structures is developed (Section 3). Second, a given repeated structure induces an epipolar or other correspondence relationship in the image which can be used to establish corresponding points from each copy of the repeated structure. The basic concepts of epipolar correspondence are introduced in Section 4. The overall approach is illustrated by a number of examples of repeated structure in Section 6. Finally, in some cases the repeated structure does not support a true epipolar structure, however a correspondence structure based on planar homology is still available. These results are discussed in Section 7.

2 Camera Models and Epipolar Geometry

The following analysis is based on techniques developed by Hartley for the study of the essential and fundamental matrix [3, 4].

2.1 Projective Cameras and Standard Forms

The perspective projection from 3D to the image plane, is modeled by a 3×4 projection matrix, $\mathtt{P}$, so that

$$\mathbf{x} = \mathtt{P}\mathbf{X} \tag{1}$$

where homogeneous coordinates are used, $\mathbf{X} = (X, Y, Z, 1)^T$, $\mathbf{x} = (x, y, 1)^T$ and equality is up to a non-zero scale factor in the case of homogeneous vectors.

The general perspective camera, $\mathtt{P}$, can be partitioned as

$$\mathtt{P} = [\mathtt{M} \;\; | - \mathtt{M}\mathbf{t}_0]$$

where $\mathtt{M}$ is an arbitrary 3×3 matrix and $\mathbf{t}_0$ is a 3-vector from the world coordinate origin to the center of projection.

We can put the camera in a *standard reference frame*, i.e., the principal ray along $\mathbf{Z}$ and $(\mathbf{u}, \mathbf{v})$ aligned along $(\mathbf{X}, \mathbf{Y})$, by a suitable 3D transformation of coordinates. That is,

$$\mathtt{P}_c = [\mathtt{I}|\mathbf{0}] = \mathtt{PB} \tag{2}$$

where the 4×4 matrix $\mathtt{B}$ is

$$\mathtt{B} = \begin{bmatrix} \mathtt{M}^{-1} & \mathbf{t}_0 \\ \mathbf{0}^T & 1 \end{bmatrix}$$

[3] A projective transformation of 3D space has 15 degrees of freedom while an affine transformation has only 12. Therefore, an invariant coordinate frame can be constructed from four 3D points in the affine case, while five points are required for the projective case.

To be more specific about B, the camera matrix, M, differs from the standard frame by a rotation of the center of projection as well as an arbitrary change in internal calibration. That is,

$$\mathtt{M} = \mathtt{KR}$$

where the matrix, K is a 3×3 upper triangular matrix and R is a rotation matrix. If $\mathtt{K} = \mathtt{I}$, which is the case for a calibrated camera, then B is simply a 3D Euclidean transformation. For an uncalibrated camera an affine transformation of space is required to bring P to standard form[4]

For example, when the camera has square pixels, the matrix, K, is given by,

$$\mathtt{K} = \begin{bmatrix} 1 & 0 & \frac{u_0}{f} \\ 0 & 1 & \frac{v_0}{f} \\ 0 & 0 & \frac{1}{f} \end{bmatrix}$$

where (u_0, v_0) is the principal point and f is the focal length. In this case P can be brought to the standard form with a 3D affine transformation,

$$\mathtt{B} = \begin{bmatrix} \mathtt{L} & \mathbf{t}_0 \\ \mathbf{0}^T & 1 \end{bmatrix}$$

where $\mathtt{L} = \mathtt{M}^{-1}$, or

$$\mathtt{L} = \mathtt{R}^T \begin{bmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & f \end{bmatrix}$$

2.2 Epipolar Geometry

For two cameras, $\mathtt{P}_a$ and $\mathtt{P}_b$,

$$\mathbf{x}_a = \mathtt{P}_a \mathbf{X} \qquad \qquad \mathbf{x}_b = \mathtt{P}_b \mathbf{X}$$

where $\mathbf{x}_a$ and $\mathbf{x}_b$ are corresponding points in the images formed by each camera. Corresponding points in the two images satisfy the epipolar constraint:

$$\mathbf{x}_b^T \mathtt{F} \mathbf{x}_a = 0 \tag{3}$$

where F is a 3×3 matrix of maximum rank 2, called the fundamental matrix. The epipolar line in image b corresponding to $\mathbf{x}_a$ is $\mathbf{l}_b = \mathtt{F}\mathbf{x}_a$, and in image a corresponding to $\mathbf{x}_b$ is $\mathbf{l}_a = \mathtt{F}^t \mathbf{x}_b$, where $\mathbf{l}$ is the vector of homogeneous line coefficients.

[4] In Section 3.5 we discuss the possibility of *self-calibration* where the internal parameters of a camera can be determined from a set of initially uncalibrated views. If such calibration is possible, then the ambiguity can be less than affine.

For two cameras, $[\mathtt{I}|\mathbf{0}]$ and $[\mathtt{M}|-\mathtt{M}\mathbf{t}]$, the fundamental matrix is given by[3],

$$\mathtt{F} = \mathtt{M}^{-T}\left[\mathbf{t}\right]_{\times} = \left[\mathtt{M}\mathbf{t}\right]_{\times}\mathtt{M} \tag{4}$$

where the notation $[\mathbf{v}]_{\times}$, with $\mathbf{v} = (x, y, z)^T$, is the matrix

$$[\mathbf{v}]_{\times} = \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}$$

Given some other three-element vector, $\mathbf{w}$, the cross-product, $\mathbf{v} \times \mathbf{w}$, is

$$[\mathbf{v}]_{\times}\mathbf{w}$$

In the case of calibrated cameras, where $\mathtt{K} = \mathtt{I}$, the fundamental matrix is known as the essential matrix, and equation (4) becomes

$$\mathtt{E} = \mathtt{R}\left[\mathbf{t}\right]_{\times} = \left[\mathtt{R}\mathbf{t}\right]_{\times}\mathtt{R}$$

The epipole is defined as the point common to all epipolar lines, i.e., ϵ_b is defined by,

$$\epsilon_b^T \mathbf{l}_b = \epsilon_b^T \mathtt{F}\mathbf{x}_a = 0$$

for all $\mathbf{x}_a$. Thus, $\epsilon_b^T\mathtt{F} = \mathbf{0}^T$ or $\mathtt{F}^T\epsilon_b = \mathbf{0}$ and ϵ_b is the null space of $\mathtt{F}^T$. Similarly, $\mathtt{F}\epsilon_a = \mathbf{0}$.

2.3 3D Reconstruction

For two cameras, $[\mathtt{I}|\mathbf{0}]$ and $[\mathtt{M}|-\mathtt{M}\mathbf{t}]$, the the rays in 3D space defined by corresponding points, $\mathbf{x}_a$, and $\mathbf{x}_b$ are,

$$\mathbf{x}_a = \lambda_a \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\mathbf{x}_b = \lambda_b \begin{bmatrix} u_b \\ v_b \\ 1 \end{bmatrix} = \mathtt{M}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathtt{M}\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Thus,

$$\lambda_a \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = \lambda_b \mathtt{M}^{-1}\begin{bmatrix} u_b \\ v_b \\ 1 \end{bmatrix} \tag{5}$$

Solving for λ_a and λ_b determines $(X, Y, Z)^T$.

3 Ambiguity of 3D Reconstruction

We will first proceed as if there are two cameras observing a single copy of the repeated structure and then show how standard stereo reconstruction is related to repeated structures in a single view. The ambiguity is determined in four stages:

1. The constraints that apply to the two cameras (arising from the transformation on the repeated structure) are found.
2. The cameras are transformed such that $\mathtt{P}_a$ takes the standard form. This is not strictly necessary but considerably simplifies the subsequent analysis.
3. The most general transformation of 3D is determined which preserves the constraints on the cameras.
4. The resulting ambiguity in the reconstructed 3D structure is then computed.

The approach to ambiguity analysis is illustrated by considering a specific example.

3.1 Two Cameras Related by Translation

Suppose we have two cameras, $\mathtt{P}_a$ and $\mathtt{P}_b$ whose external 3D coordinate frames differ only by a translation, $\mathbf{t}$. The 3×3 matrices, $\mathtt{M}_a$, $\mathtt{M}_b$ are related by,

$$\mathtt{M}_a = \mathtt{K}_a\mathtt{R}$$
$$\mathtt{M}_b = \mathtt{K}_b\mathtt{R}$$

where $\mathtt{K}_a, \mathtt{K}_b$ are upper triangular[5] and represent the internal calibration of each camera.

Applying the standard transformation to $\mathtt{P}_b$,

$$\mathtt{P}'_b = \mathtt{P}_b\mathtt{B} = \left[\mathtt{M}_b\mathtt{M}_a^{-1} | -\mathtt{M}_b\mathbf{t}\right]$$

But,

$$\mathtt{M}_b\mathtt{M}_a^{-1} = \mathtt{K}_b\mathtt{R}\mathtt{R}^{-1}\mathtt{K}_a^{-1} = \mathtt{K}_0$$

where $\mathtt{K}_0$ represents the difference in the internal calibration of $\mathtt{P}_a, \mathtt{P}_b$. The final form of the cameras is then:

$$\mathtt{P}'_a = [\mathtt{I}|\mathbf{0}] \qquad (6)$$
$$\mathtt{P}'_b = [\mathtt{K}_0| -\mathtt{M}_b\mathbf{t}]$$

For identical cameras, $\mathtt{K}_0 = \mathtt{I}$ (since $\mathtt{K}_a = \mathtt{K}_b$). For the case of repeated structures, the cameras are always identical.

[5] Note that upper triangular matrices form a group under multiplication, e.g. $\mathtt{K}_1\mathtt{K}_2 = \mathtt{K}_3$ and $\mathtt{K}_1^{-1} = \mathtt{K}_4$, where $\mathtt{K}_i$ are upper triangular.

3.2 Ambiguity of the Camera Relation

The next step in analyzing the ambiguity of reconstruction is to determine what transformations can be applied to the cameras without affecting their relationship. Consider a general projective transformation, $\mathtt{D}$, where

$$\mathtt{D} = \begin{bmatrix} \mathtt{E} & \mathbf{s} \\ \mathbf{a}^T & 1 \end{bmatrix}$$

In order to keep the standard form for the first camera,

$$[\mathtt{I}|\mathbf{0}]\mathtt{D} = [\mathtt{I}|\mathbf{0}]$$

so $\mathbf{s} = 0$, $\mathtt{E} = I$. Applying the resulting matrix $\mathtt{D}$ to the second camera, we have the following constraint on the vector, $\mathbf{a}$.

$$\mathtt{K}_0 - \mathbf{t}'\mathbf{a}^T = \mathtt{K'}_0$$

where $\mathbf{t}' = \mathtt{M}_b\mathbf{t}$ and $\mathtt{K'}_0$ is of the same form as $\mathtt{K}_0$, i.e. the same constraints on the matrix elements. For example, a zero element of $\mathtt{K}_0$ is also zero in $\mathtt{K'}_0$. Such constraints hold since the cameras in the standard frame differ only in internal calibration which must be the same form as $\mathtt{K}_0$. Considering identical cameras where $\mathtt{K}_0 = \mathtt{I}$ and expanding the matrices,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} t'_1a_1 & t'_1a_2 & t'_1a_3 \\ t'_2a_1 & t'_2a_2 & t'_2a_3 \\ t'_3a_1 & t'_3a_2 & t'_3a_3 \end{bmatrix} = k \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

So it follows that $k = 1$ and

$$a_1 = a_2 = a_3 = 0$$

since $\mathbf{t}'$ is a general vector. Thus the transformation, $\mathtt{D}$, is forced to be the identity.

In general, not all components of the vector $\mathbf{a}$ will be forced to zero. Examples where $\mathbf{a} \neq \mathbf{0}$ will arise in analyzing other repeated structure classes in Section 6.

3.3 Reconstructing 3D Geometry

From equation (5) we obtain in this case:

$$\lambda_a \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} + \mathbf{t}' = \lambda_b \mathtt{K}_0^{-1} \begin{bmatrix} u_b \\ v_b \\ 1 \end{bmatrix} \qquad (7)$$

This vector equation is homogeneous in $\lambda_a, \lambda_b, \mathbf{t}'$, so we can scale all the 3D coordinates of space by a constant k without affecting the solution. In general, uniform scaling is the only new source of ambiguity arising from the reconstruction of 3D space. Uniform scaling can be represented as,

$$\mathtt{S} = \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \mathbf{0}^T & s \end{bmatrix}$$

3.4 The Overall Ambiguity

The total ambiguity of reconstruction is found by multiplying the 3D transformation matrices encountered in producing the final reconstruction. For our example, the total transformation is $\mathtt{T} = \mathtt{BDS}$

$$\mathtt{T} = \begin{bmatrix} \mathtt{M}_a^{-1} & \mathbf{t}_0 \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathtt{I} & \mathbf{0} \\ \mathbf{0}^T & s \end{bmatrix}$$

Thus, for pure translation the overall reconstruction is 3D affine. To summarise, the overall reconstruction ambiguity is obtained by analyzing the following three stages of transformation.

1. Transform one of the cameras to the standard frame (using $\mathtt{B}$). For uncalibrated cameras, this transformation introduces an affine transformation of space. A general affine transformation exhibits 12 degrees of freedom, but the transformation to the standard camera frame is often restricted to 10 parameters since the internal calibration involves only unknown focal length, aspect ratio, and principal point. This ten parameter transformation is the least ambiguity that can occur for cameras with unknown internal calibration.
2. Determine what projectivities of 3D space leave the standard form for the cameras unchanged. At this stage, a full projective transformation might be allowable. For example, when the two cameras are related by a general 3D rotation and translation, all elements of $\mathbf{a}$ are non-zero.
3. Determine the ambiguity of stereo reconstruction itself. The only ambiguity which can occur is an isotropic scaling of space (which is covered by an affine transformation).

Even though our main interest is in repeated structures which corresponds to identical cameras, other cases have been considered according to these steps to compare with the results of Moons *et al* (in this volume).

The overall ambiguity for two cameras under translation is summarized in Table 1.

Case	Relation Between Cameras	Resulting Ambiguity
1	Identical Calibrated Cameras	Isotropic Scaled Euclidean
2	Identical Uncalibrated Cameras	Affine
3	Different Focal Lengths	Affine
4	Same Principal Points, $\neq \mathbf{0}$, Same f's	Affine
5	Same Principal Points, $\neq \mathbf{0}$, Different f's	Projective
6	Different Principal Points, Same f's	Affine
7	Different Principal Points, Diff. f's	Projective

Table 1. Results for the ambiguity of 3D reconstruction for cameras related by translation only. Only cases 1, 2 and 4 apply to repeated structures under translation.

3.5 Camera Self-Calibration

Recent work on camera calibration from images taken of arbitrary scenes and arbitrary viewpoints bears some relation to the investigation of ambiguity of reconstruction for repeated structures[6]. Suppose a 3D structure, $\mathbf{X}$, is known up to an affine ambiguity. We define an image pair $\mathbf{x}_a = \mathtt{P}_a\mathbf{X}$ and $\mathbf{x}_b = \mathtt{P}_b\mathbf{X}$ and the plane at infinity $X_4 = 0$. For points on the plane at infinity, $\mathbf{X}^\infty = (X, Y, Z, 0)^\top$, $\mathbf{x}_a = \mathtt{M}_a\mathbf{X}^\infty$, $\mathbf{x}_b = \mathtt{M}_b\mathbf{X}^\infty$, and

$$\mathbf{x}_b = \mathtt{M}_b\mathtt{M}_a^{-1}\mathbf{x}_a = \mathtt{H}_\infty\mathbf{x}_a \tag{8}$$

where $\mathtt{H}_\infty$ is the *infinite homography* [6] which maps image points from the first image to the second image for 3D points on π_∞, i.e. vanishing points are mapped to vanishing points. It can be shown [6] that under $\mathtt{H}_\infty$

$$\mathtt{C}_b = \mathtt{H}_\infty\mathtt{C}_a\mathtt{H}_\infty^\top \tag{9}$$

where $\mathtt{C}_i = \mathtt{K}_i\mathtt{K}_i^\top$. $\mathtt{C}_i$ is the image in view i of the dual (i.e. the inverse) of the absolute conic. The image of the absolute conic, $(\mathtt{K}_i\mathtt{K}_i^\top)^{-1}$ is independent of the camera's position and orientation, and only depends on the camera's intrinsic parameters [2]. Equation (9) is the transformation of a conic under the linear transformation $\mathtt{H}_\infty$.

If camera intrinsic parameters are fixed between views, $\mathtt{C}_a = \mathtt{C}_b = \mathtt{C}$ then

$$\mathtt{C} = \mathtt{H}_\infty\mathtt{C}\mathtt{H}_\infty^\top. \tag{10}$$

This is a linear equation for $\mathtt{C}$. In general there is a one parameter family of solutions (as well as an overal scale), but this is reduced to a two fold ambiguity by assuming there is no skew between the image axes (a quadratic constraint on the elements of $\mathtt{C}$). Once $\mathtt{C}$ is determined (up to scale), $\mathtt{K}$ can be obtained simply by a Choleski decomposition of $\mathtt{C} = \mathtt{K}\mathtt{K}^\top$. Subsequently, the affine structure ambiguity can be reduced to only a scaled Euclidean ambiguity by the transformation

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^S = \mathtt{K}^{-1}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^A \tag{11}$$

It can be shown [6] that

$$\mathtt{H}_\infty = \mathtt{K}\mathtt{R}\mathtt{K}^{-1} \tag{12}$$

where $\mathtt{R}$ is the rotation between images. If there is no rotation, equation (10) reduces to $\mathtt{C} = \mathtt{C}$ and there is no constraint on $\mathtt{C}$. It appears that in all self-calibration methods (e.g. [2, 5] and Hartley (in this volume)) if there is no rotation between views then there is no self-calibration constraint on the intrinsic parameters. Thus the ambiguity of reconstruction remains at least affine.

[6] This material was recently added in response to the paper by Luong and Vieville [6] since it is appropriate to characterize their results in the ambiguity framework just presented. The notation used here differs from that in [6], where $\mathtt{K}$ is the image of the dual of the absolute conic.

4 Epipolar Correspondence Structure

An important aspect of 3D reconstruction from repeated structures is the determination of epipolar geometry. Epipolar geometry defines the image relationships between corresponding features on the repeated structure. These relationships are encapsulated in the fundamental matrix, $\mathtt{F}$.

This is illustrated using again two identical cameras related by a translation. When the two camera frames are related by the translation vector, $\mathbf{t}$, the fundamental matrix is,

$$\mathtt{F}_t = [\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

and $\mathtt{F}_t^T = -\mathtt{F}_t$, $\epsilon_a = \epsilon_b = \epsilon$. where

$$\epsilon = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

so in image plane coordinates,

$$\epsilon_u = \frac{t_x}{t_z} \qquad \epsilon_v = \frac{t_y}{t_z}$$

Consequently, from the epipoles alone, $\mathbf{t}$ can only be recovered up to scale - this leads to an unknown scale in the stereo 3D reconstruction.

This case also illustrates a useful constraint which applies more generally: corresponding image points lie on the same epipolar line. That is,

$$\mathbf{l} = \mathtt{F}_t \mathbf{x}_a = [\mathbf{t}]_\times \mathbf{x}_a$$

But, $[\mathbf{t}]_\times \mathbf{x}_a = \mathbf{t} \times \mathbf{x}_a$ and

$$\mathbf{x}_a \cdot (\mathbf{t} \times \mathbf{x}_a) = \mathbf{x}_a \cdot \mathbf{l} = 0$$

so $\mathbf{x}_a$ also lies on $\mathbf{l}$. This convenient epipolar correspondence geometry is called *auto-epipolar*. In subsequent discussion, we will focus on repeated structure classes which result in auto-epipolar feature correspondence.

5 Repeated Structures

Now we relate the procedures developed for two cameras observing a single structure, the usual stereo configuration, to the case of a single image of a repeated structure.

> Suppose we have a structure, $\mathcal{S}$, and a transformation which generates a *copy* of $\mathcal{S}$, i.e. $\mathcal{S}' = \mathcal{T}_g(\mathcal{S})$, where $g \in G$, for some group G. $\mathcal{S}$ and $\mathcal{S}'$ are viewed in a single perspective image. This is equivalent to a stereo pair (with the *camera's* related by g^{-1}). As in the stereo case, there are three goals:

1. Determine the ambiguity of the reconstruction (and consequently the appropriate invariants).
2. Determine the correspondence geometry within the single image.
3. Develop an algorithm for carrying out the reconstruction and computing invariants based on the correspondence geometry.

In this general setting, it becomes clear that it is not necessary that we restrict ourselves to simple translations or rotations (Euclidean transformations) of $\mathcal{S}$. Many 3D objects can be represented as repeated structures where G is an equiform or affine transformation. For example, a rod with two different diameter spheres at each end, can be defined as a translation followed by an equiform scaling. The case of extruded surfaces can be viewed as translation and scaling of the cross-sectional boundary curve.

Repeated structures can also be defined by other transformational symmetries between $\mathcal{S}$ and $\mathcal{S}'$. A case we consider in detail below is bilateral symmetry where $\mathcal{T}_g$ is a mirror reflection about the symmetry plane. A similar transformation arises in the case of rotationally symmetric objects where the outline curve on the object surface can be considered as a bilateral symmetry structure where the plane of symmetry passes through the camera center of projection.

5.1 Ambiguity of Reconstruction

If the transformation which generates the copy is $\mathtt{T}_g$, then an equivalent camera configuration is given by $\mathtt{P}_a$ and $\mathtt{P}_a\mathtt{T}_g^{-1}$. Without loss of generality, we take

$$\mathtt{P}_a = \mathtt{K}[\mathtt{I}|\mathtt{0}]$$

Suppose $\mathtt{T}_g^{-1}$ is an affine transformation. This can be represented as:

$$\mathtt{T}_A = \begin{bmatrix} \mathtt{A} & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{bmatrix}$$

Then $\mathtt{P}_b$ is given by

$$\begin{aligned} \mathtt{P}_b &= \mathtt{K}[\mathtt{I}|\mathtt{0}] \begin{bmatrix} \mathtt{A} & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{bmatrix} \\ &= \mathtt{K}[\mathtt{A}|\mathbf{t}'] \end{aligned}$$

To determine the ambiguity we follow the procedure in Section 3, and first put $\mathtt{P}_a$ into standard form giving:

$$\begin{aligned} \mathtt{P}_a &= [\mathtt{I}|\mathtt{0}] \\ \mathtt{P}_b &= \mathtt{K}[\mathtt{AK}^{-1}|\mathbf{t}'] \\ &= \mathtt{KAK}^{-1}[\mathtt{I}| - (-\mathtt{KA}^{-1}\mathbf{t}')] \end{aligned} \quad (13)$$

where $\mathtt{P}_b$ has the canonical form $\mathtt{M}[\mathtt{I}| - \mathbf{t}]$. We describe a number of special cases of affine transformations in Section 6.

5.2 Epipolar Geometry

Using the second form from equation (4) the fundamental matrix for the equivalent camera pair given in equation (13) is

$$\mathtt{F} = [\mathtt{K}\mathbf{t}']_\times \mathtt{KAK}^{-1}$$

Using the identity,

$$[\mathtt{M}\mathbf{t}]_\times = \mathtt{M}^{-T} [\mathbf{t}]_\times \mathtt{M}^{-1} \tag{14}$$

gives

$$\mathtt{F} = \mathtt{K}^{-T} [\mathbf{t}']_\times \mathtt{AK}^{-1}$$

We now return to the question of when is the epipolar geometry auto-epipolar? From Section 4 this is when the quadratic form $\mathbf{x}^T \mathtt{F} \mathbf{x} = 0$. In this case,

$$\begin{aligned} \mathbf{x}^T \mathtt{F} \mathbf{x} &= (\mathtt{K}^{-1}\mathbf{x})^T [\mathbf{t}']_\times \mathtt{A}(\mathtt{K}^{-1}\mathbf{x}) \\ &= \mathbf{x}'.\, [\mathbf{t}']_\times \mathtt{A}\mathbf{x}' \end{aligned}$$

where $\mathbf{x}' = \mathtt{K}^{-1}\mathbf{x}$, which is again a quadratic form. The quadratic form is zero if the matrix

$$[\mathbf{t}']_\times \mathtt{A} \tag{15}$$

is skew, which provides a simple test for auto-epipolar geometries.

6 Examples of Repeated Structure

6.1 Translation

The transformation between a point $\mathbf{X}$ in $\mathcal{S}$ and the corresponding point, $\mathbf{X}'$ in $\mathcal{S}'$ is given by,

$$\mathbf{X}' = \mathbf{X} + \mathbf{t}$$

so that $\mathtt{A} = \mathtt{I}$ and $\mathbf{t}' = \mathbf{t}$. An example is shown in Figure 1.

The equivalent cameras in standard form are from equation (13)

$$\begin{aligned} \mathtt{P}_a &= [\mathtt{I}|\mathtt{0}] \\ \mathtt{P}_b &= [\mathtt{I}|\mathtt{K}\mathbf{t}] \end{aligned}$$

This is the case examined in Sections 3 and now considered as a single image of a repeated structure. The overall ambiguity is affine.

The correspondence geometry is clearly auto-epipolar, since $[\mathbf{t}]_\times$ is skew.

6.2 Translation and Rotation

Consider a structure and a copy of the structure which has been translated and then rotated and observed in a single view. In this example, the transformation between a point $\mathbf{X}$ in $\mathcal{S}$ and the corresponding point, $\mathbf{X}'$ in $\mathcal{S}'$ is given by,

$$\mathbf{X}' = \mathtt{R}\mathbf{X} + \mathbf{t}$$

where $\mathtt{R}$ is a rotation matrix. The equivalent cameras for this repeated structure are given by $\mathtt{A} = \mathtt{R}$ and $\mathbf{t}' = \mathbf{t}$:

$$\begin{aligned} \mathtt{P}_a &= [\mathtt{I}|\mathbf{0}] \\ \mathtt{P}_b &= \left[\mathtt{KRK}^{-1}|\mathtt{K}\mathbf{t}\right] \end{aligned} \tag{16}$$

The reconstructed 3D geometry is ambiguous up to a projectivity of space which is shown as follows. The form of camera $\mathtt{P}_b$ should be preserved by $\mathtt{D}$ so,

$$\mathtt{KRK}^{-1} + \mathbf{a}^T(\mathtt{K}\mathbf{t}) = k\mathtt{K}'\mathtt{R}'\mathtt{K}'^{-1}$$

In this case, all components of $\mathbf{a}$ are, in general non-zero, since the projective transformation is indistinguishable from a difference in camera calibration. That is $\mathbf{a}$ can be accounted for by the difference between $\mathtt{K}$ and $\mathtt{K}'$.

The epipolar correspondence structure with both translation and rotation is not as convenient as the case of pure translation as it is not auto-epipolar in general. This is clear from the skew test: equation (15) in this case is $[\mathbf{t}]_\times\mathtt{R}$ which is not skew in general.

6.3 Affine Repeated Structures

In this example, the transformation between a point $\mathbf{P}$ in $\mathcal{S}$ and the corresponding point, $\mathbf{P}'$ in $\mathcal{S}'$ is given by,

$$\mathbf{X}' = \mathtt{A}\mathbf{X} + \mathbf{t}$$

where $\mathtt{A}$ is an affine matrix. The equivalent cameras for this repeated structure are,

$$\mathtt{P}_a = [\mathtt{I}|\mathbf{0}] \tag{17}$$

$$\mathtt{P}_b = \left[\mathtt{KAK}^{-1}|\mathtt{K}\mathbf{t}\right] \tag{18}$$

Examples of the general affine copy transformation are shown in Figure 2. The ambiguity of reconstruction is in general projective, since the affine transformation of the copy is indistinguishable from the effects of $\mathbf{a}$. For example, consider the special case of scaling only in Z. In this case,

$$\mathtt{A}_s = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s \end{bmatrix}$$

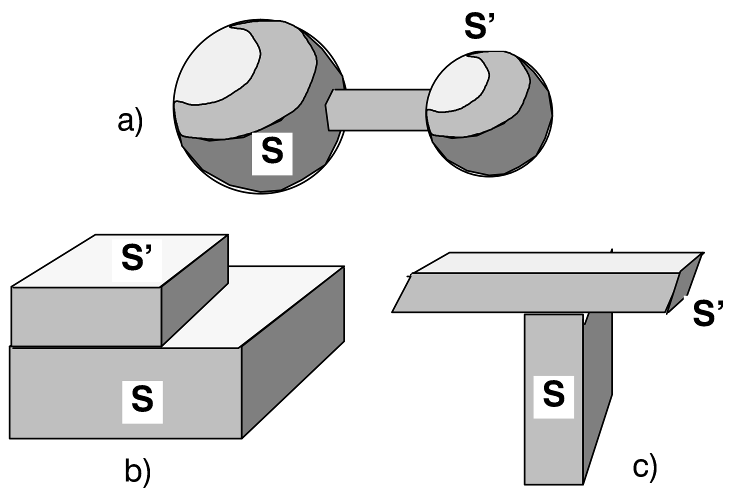

Fig. 2. Examples of repeated structures where the copy, $\mathcal{S}'$, is an affine transformation of $\mathcal{S}$.

Expanding the matrices,

$$\mathtt{KA}_s\mathtt{K}^{-1} = \begin{bmatrix} 1 & 0 & k_{13}(s-1)/k_{33} \\ 0 & 1 & k_{23}(s-1)/k_{33} \\ 0 & 0 & s \end{bmatrix}$$

where

$$\mathtt{K} = \begin{bmatrix} 1 & 0 & k_{13} \\ 0 & 1 & k_{23} \\ 0 & 0 & k_{33} \end{bmatrix}$$

In this case the requirement that $\mathtt{P}_b$ remain the same form under the transformation, $\mathtt{D}$ yields,

$$\begin{bmatrix} 1 & 0 & k_{13}(s-1)/k_{33} \\ 0 & 1 & k_{23}(s-1)/k_{33} \\ 0 & 0 & s \end{bmatrix} + \begin{bmatrix} t_1a_1 & t_1a_2 & t_1a_3 \\ t_2a_1 & t_2a_2 & t_2a_3 \\ t_3a_1 & t_3a_2 & t_3a_3 \end{bmatrix} = k \begin{bmatrix} 1 & 0 & k'_{13}(s'-1)/k'_{33} \\ 0 & 1 & k'_{23}(s'-1)/k'_{33} \\ 0 & 0 & s' \end{bmatrix}$$

It follows that a_3 is not constrained to be zero by this relation, and thus, for anisotropic scaling between $\mathcal{S}$ and $\mathcal{S}'$, the reconstruction ambiguity is projective.

The epipolar geometry is not auto-epipolar in general because $[\mathbf{t}]_\times \mathtt{A}$ is not skew in general. However, if the affine transformation is restricted to an isotropic scaling ($\mathtt{A}_i = s\mathtt{I}$) then the epipolar structure has the auto-epipolar form, since $[\mathbf{t}]_\times \mathtt{A}_i$ is skew. The case of simple scaling is shown by examples a and b in Figure 2.

6.4 Bilateral Symmetry

Rothwell *et al* [10] have studied the case of objects with a plane of symmetry, i.e., bilateral symmetry. In this case the relationship between a 3D point, $\mathbf{P}$ and its symmetric corresponding point, $\mathbf{P}'$ is given by,

$$\mathbf{P}' = \mathtt{T} \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathtt{T}^{-1} \mathbf{P}$$

where

$$\Sigma = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and $\mathtt{T}$ is an Euclidean transformation. $\mathtt{T}$ is given by,

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

where $\mathtt{R}$ is a 3D rotation matrix and $\mathbf{t}$ is a 3D translation. The coordinate system for bi-lateral symmetry is shown in Figure 3.

The composite reflection transformation is

$$\begin{aligned} \mathtt{A} &= \mathtt{R}\Sigma\mathtt{R}^{\mathtt{T}} \\ \mathbf{t}' &= -\mathtt{R}\Sigma\mathtt{R}^{\mathtt{T}}\mathbf{t} + \mathbf{t} = \mathtt{R}(\mathtt{I} - \Sigma)\mathtt{R}^T\mathbf{t} = \mathtt{R}\Gamma\mathtt{R}^T\mathbf{t} \end{aligned}$$

where

$$\Gamma = (\mathtt{I} - \Sigma) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

It can be shown that the ambiguity of reconstruction is projective. Note that when the plane of symmetry passes through the center of projection, $\mathbf{t} = \mathbf{0}$, and 3D structure cannot be established. This case is analogous to images related by rotation about the center of projection. We will return to this condition in Section 7.

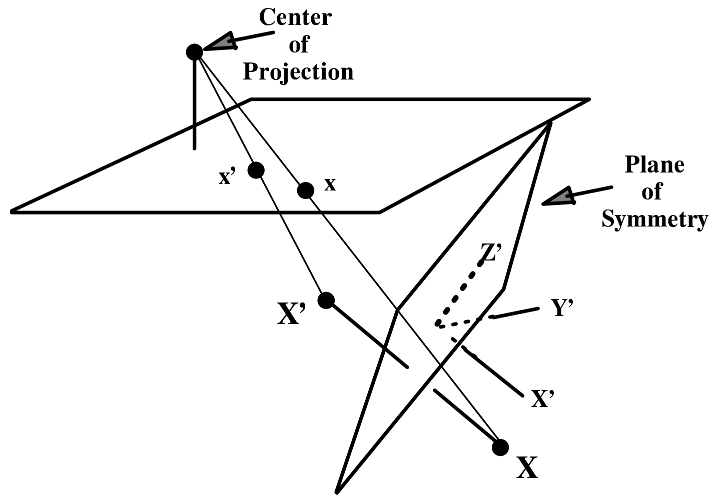

Fig. 3. The coordinate system for bi-lateral symmetry.

Epipolar Geometry The matrix for the skew test in this case is, from equation (15),

$$[\mathtt{R}\Gamma\mathtt{R}^T\mathbf{t}]_\times \mathtt{R}\Sigma\mathtt{R}^{\mathtt{T}}$$

Applying the identity (14) simplifies this to

$$\mathtt{R}[\Gamma\mathtt{R}^T\mathbf{t}]_\times \Sigma\mathtt{R}^{\mathtt{T}}$$

This is a skew matrix since, $[\Gamma\mathtt{R}^T\mathbf{t}]_\times$ is a matrix of the form

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -x \\ 0 & x & 0 \end{bmatrix}$$

and post multiplying by Σ maintains this skew form (the rotations simply rotate the quadratic form vectors). Consequently, the correspondences are auto-epipolar.

7 Non-Epipolar Correspondence Structures

We reserve the terminology "epipolar geometry" to those cases of repeated structure where the equivalent cameras, $\mathtt{P}_a$ and $\mathtt{P}_b$, actually have distinct centers of projection. There are, however, important cases of repeated structures where there is no translation between the centers.

7.1 Example - Bilateral Symmetry

Consider the case of bilateral symmetry when the plane of symmetry passes through the center of projection of $\mathtt{P}_a$. The cameras in standard form become,

$$\mathtt{P}_a = [\mathtt{I}|\mathbf{0}]$$
$$\mathtt{P}_b = \left[\mathtt{KR}\Sigma\mathtt{R}^T\mathtt{K}^{-1}|\mathbf{0}\right]$$

There is no epipolar geometry defined, however there is still a correspondence structure. The correspondence is defined by the relationship between the image projections of $\mathcal{S}$ and $\mathcal{S}'$. That is, if we define the image projection of a point in $\mathcal{S}$ as $\mathbf{x}$ and the corresponding point as $\mathbf{x}'$ then,

$$\mathbf{x} = [\mathtt{I}|\mathbf{0}]\,\mathbf{X}$$
$$\mathbf{x}' = \left[\mathtt{KR}\Sigma\mathtt{R}^T\mathtt{K}^{-1}|\mathbf{0}\right]\mathbf{X}$$

Thus,

$$\mathbf{p}' = \mathtt{T}\mathbf{p}$$

where $\mathtt{T} = \left[\mathtt{KR}\Sigma\mathtt{R}^T\mathtt{K}^{-1}\right]$. So $\mathbf{x}$ and $\mathbf{x}'$ are related by a planar projective transformation. However, this transformation is not an arbitrary 3×3 matrix and there results a convenient correspondence structure [8].

First, note that $\mathtt{T}^2 = \mathtt{I}$, so $\mathtt{T}$ has eigen-values ± 1, this is a two-cyclic homography. Further, it is clear from the special case when $\mathtt{RK} = \mathtt{I}$, that $\mathtt{T}$ has eigen-values $\{-1, 1, 1\}$ (i.e. there is a degenerate eigenvalue) because in this case $\mathtt{T} = \Sigma$. Since eigen-values are preserved by a similarity transformation this is true of $\mathtt{T}$ in general, and $\mathtt{T}$ is known as a planar harmonic homology [11].

7.2 Planar Homology

Eigen-vectors of a projective transformation determine the fixed points and fixed lines of the transformation. If two eigen-vectors have the same eigen-value then they define a line of fixed points. The fixed lines and points for a planar-homology are shown in Figure 4.

As shown by Springer, these eigenvalues define a pencil of fixed lines all intersecting at a fixed point called the *center* of the homology. The center is the fixed point corresponding to the eigen-value $\lambda = -1$. There is one line of the pencil which passes through any given point. This pencil defines the correspondences between image points which are corresponding projections of $\mathcal{S}$ and $\mathcal{S}'$. That is, each line of the pencil is fixed under the transformation which carries $\mathbf{x}$ to $\mathbf{x}'$. Therefore, given that $\mathbf{x}$ is on a line $\mathbf{l}$ of the pencil, then $\mathbf{x}'$ is on the same line.

A final interesting example of planar homology is provided by the case of a rotationally symmetric object. In this case we can consider that the outline curve on the object surface is a bilaterally symmetric pair with the plane of symmetry passing always through the center of projection. Even if the plane of symmetry is arbitrarily rotated about the center, the relationship between the two halves of the occluding boundary in the image is just a planar homology and the resulting image structure can be used to find corresponding image points on each side of the symmetry plane.

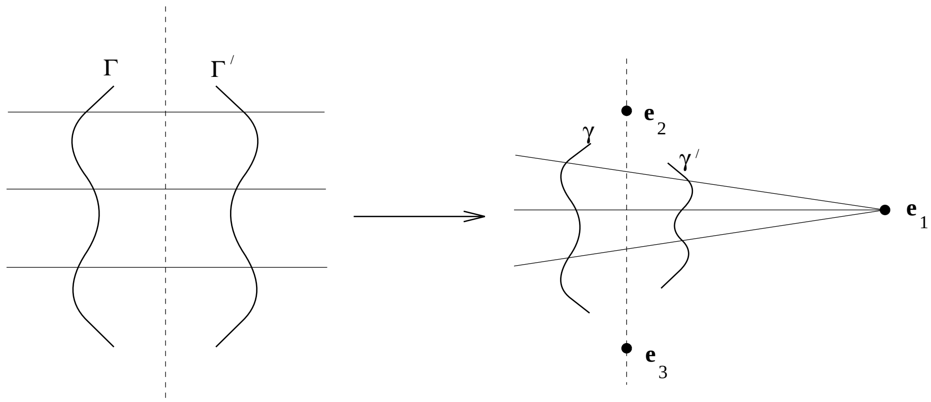

Fig. 4. Under a projective transformations parallel object correspondences converge to a vanishing point. Corresponding points are related in this case by a particular projective transformation, $\mathtt{T}$, called a planar-homology. Two of the eigenvalues of $\mathtt{T}$, corresponding to $\mathbf{e}_2$ and $\mathbf{e}_3$, are equal. The third, corresponding to $\mathbf{e}_1$, is distinct and non-zero. The line $\mathbf{e}_2 \times \mathbf{e}_3$ is a line of fixed points. Corresponding points, $\mathbf{x}'$ and $\mathbf{x}$, are collinear with $\mathbf{e}_1$. $\mathbf{e}_1$ defines a pencil of fixed lines.

References

1. Faugeras, O., "What can be Seen in Three Dimensions with an Uncalibrated Stereo Rig?" *Proc. of ECCV*, p.563-578, 1992.
2. Faugeras, O., Luong, Q.T., and Maybank, S.J., "Camera Self-Calibration: Theory and Experiments," *Proc. of ECCV*, 321-334, 1992.
3. Hartley, R.I., Gupta, R. and Chang, T., "Stereo from Uncalibrated Cameras," *Proc. of CVPR*, p.761-764, 1992.
4. Hartley, R.I., "Cheirality invariants", *Proc. DARPA IU Workshop*, 745–753, 1993.
5. Hartley, R.I., "Self-calibration of stationary cameras," *Proc. of ECCV*, 1994.
6. Luong, Q.T. and Vieville, T., "Canonic representations for the geometries of multiple projective views," *Tech. Rep.* ucb/csd-93-772, University of California, Berkeley, USA, 1993.
7. Moons, T., Van Gool, L., Van Diest, M. and Oosterlinck, A., "Affine structure from perspective image pairs under relative translations between object and camera," University of Leuven Report No. KUL/ESTAT/MI2/9306, 1993.
8. Mukherjee, D.P., Zisserman, A. and Brady, J.M., "Shape from Symmetry - Detecting and Exploiting Symmetry in Affine Images", Univ. of Oxford, Dept. Eng. Sci. Report No. OUEL 1988/93. To appear, *Proc. Royal Soc.*, 1994.
9. Mundy, J.L. and Zisserman, A. (editors), *"Geometric Invariance in Computer Vision,"*, MIT Press, Cambridge Ma, 1992.
10. Rothwell, C.A., Forsyth, D.A., Zisserman, A. and Mundy, J.L. "Extracting Projective Structure from Single Perspective Views of 3D Point Sets", *ICCV*, 1993.
11. Springer, C., "Geometry and Analysis of Projective Spaces," W.H. Freeman and Company, 1964.

This article was processed using the LaTeX macro package with LLNCS style