A Geometric Interpretation of Weak-Perspective Motion

Ilan Shimshoni, *Member, IEEE*, Ronen Basri, *Member, IEEE*,
and Ehud Rivlin, *Member, IEEE*

Abstract—We present a geometric interpretation of the problem of motion recovery from three weak-perspective images. Our interpretation is based on reducing the problem of estimating the motion to a problem of finding triangles on a sphere whose angles are known. Using this geometric interpretation, a simple method to completely recover the motion parameters using three images is developed. The results of running the algorithm on real images are presented. In addition, we describe which of the various motion parameters can be recovered already from two images.

Index Terms—Weak-perspective projection, structure from motion.

——————————— ✦ ———————————

1 INTRODUCTION

DETERMINING the relative position and orientation of two or more cameras from images of an unknown scene taken by those cameras is known as the motion estimation problem in computer vision. In this paper we address the problem of determining the motion parameters between two and three weak-perspective (or scaled-orthographic) views. The weak-perspective model is appropriate when the observed object is far from the camera relative to its size. In this situation the perspective distortions are relatively small. Therefore, a full perspective model is not required, and this simpler and more robust model can be used. Below we present a geometric interpretation for weak-perspective motion from two and three images. This interpretation leads to a simple algorithm for recovering all the motion parameters from three images.

The problem of motion estimation from orthographic views has been studied first by Ullman [15], who showed that two orthographic images do not suffice to fully recover the motion parameters. Ullman then proposed a nonlinear solution to the problem of motion recovery from three images. He showed further that only two solutions exist in this case, and that these two solutions are identical up to a reflection with respect to the image plane. Later, Huang and Lee [7], [10] introduced a linear solution to this problem. Kontsevich [9] extended this algorithm to a camera model which allows for a uniform scaling factor (i.e., the weak-perspective camera model), and Shapiro, Zisserman, and Brady [13] developed a noise resistant motion recovery algorithm which also estimates the statistical confidence in the results. Tomasi and Kanade [14] developed an algorithm to recover structure and motion from a sequence of orthographic images which was later extended for scaled orthographic projection and the paraperspective models by Poelman and Kanade [12]. In addition, algorithms for segmenting multiple moving objects were proposed in [4], [6], [3].

- *I. Shimshoni is with the Department of Industrial Engineering and Management, Technion—Israel Institute of Technology, Technion City 32000, Israel. E-mail: ilans@ie.technion.ac.il.*
- *R. Basri is with the Department of Applied Math, The Weizmann Institute of Science, Rehovot 76100 Israel. E-mail: ronen@wisdom.weizmann.ac.il.*
- *E. Rivlin is with the Department of Computer Science, Technion—Israel Institute of Technology, Technion City 32000, Israel. E-mail: ehudr@cs.technion.ac.il.*

The major drawback of the linear algorithm presented in [7] is that it does not offer a geometric interpretation. In this paper, we present a geometric interpretation for weak-perspective motion from two and three images. We begin by analyzing what parameters can and cannot be recovered from two images. We show that given two images the two viewing directions corresponding to these images lie on a great circle on the viewing sphere whose orientation is parallel to the epipolar lines in the two images. Consequently, the epipolar lines determine the orientation of this great circle, but the amount of rotation along the great circle cannot be determined from only two images. We then analyze the case of recovering motion parameters from three images. We point out that in this case the three pairs of epipolar constraints determine a triangle on the viewing sphere. The angles of this spherical triangle are the angles between the epipolar lines in every image. The sides of this triangle, which correspond to the unknown angles of rotation, can be recovered from its angles simply by using the sine and cosine laws for spherical triangles yielding the rotation components of the transformation. Finally, we analyze which of the translation parameters of the camera can be recovered from two and three images.

The rest of the paper is organized as follows. The weak-perspective projection model is described in Section 2. Section 3 discusses the case of recovering weak-perspective motion from two images. A simple, geometric method for motion recovery from three images is presented and demonstrated in Section 4.

2 WEAK-PERSPECTIVE PROJECTION

The weak-perspective projection is defined as follows. Given n points

$$\mathbf{P}_1, ..., \mathbf{P}_n \in \mathbb{R}^3 \ (\mathbf{P}_i = (X_i, Y_i, Z_i)^T, 1 \le i \le n),$$

let $\overline{Z}$ denote the mean value of $Z_1, ..., Z_n$. A weak-perspective image of the points is constructed by first projecting the points parallel to the Z axis to the plane $Z = \overline{Z}$ followed by a perspective projection to the plane $Z = f$, where f is the focal length of the camera (see Fig. 1). The first projection has the effect of omitting the depth coordinates, Z_i, and the second projection has the effect of scaling the points by $f / \overline{Z}$. The weak-perspective projection, therefore, resembles the paraperspective projection, except that the first projection is orthogonal to the image plane (see [11], [1], [2]). Formally, denote by

$$\mathbf{p}_i = \left(x_i, y_i, z_i\right)^T = \frac{f}{\overline{Z}} \mathbf{P}_i \tag{1}$$

then the projected location of $\mathbf{P}_i$ is given by (x_i, y_i).

Suppose now that we rotate the object by a rotation $R \in SO(3)$ and translate it by $\mathbf{T} \in \mathbb{R}^3$ with respect to the camera location, that is,
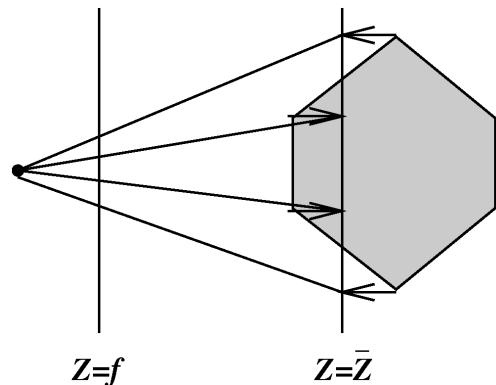

Fig. 1. The weak-perspective projection (a side view). An object is first projected onto the $Z = \overline{Z}$ in a direction orthogonal to the image plane, and then to the image plane ($Z = f$) by a central projection.

$$\mathbf{P}'_i = R\mathbf{P} + \mathbf{T}. \tag{2}$$

Consequently, the image position of $\mathbf{P}'_i$ is given by (x'_i, y'_i), where

$$\mathbf{p}'_i = \left(x'_i, y'_i, z'_i\right)^T = \frac{1}{\overline{Z}'}\mathbf{P}'_i \tag{3}$$

and $\overline{Z}'$ represents the new mean of $Z'_1, \dots, Z'_n$.

Combining (2) and (3), we obtain

$$\mathbf{p}'_i = s\left(R\mathbf{p}_i + \mathbf{t}\right), \tag{4}$$

where $s = \overline{Z}' / \overline{Z}$ and $\mathbf{t} = \left(t_x, t_y, t_z\right)^T = f\mathbf{T} / \overline{Z}$.

Alternatively, the object can be regarded as stationary, in which case the camera would translate and rotate around the object. The new coordinate frame associated with the camera is given by the rows of R, and the new origin lies at $-R^T\mathbf{T}$. The third row of R, denoted by $\mathbf{v}_z$ represents the new viewing direction, the direction along which the object is projected. The viewing direction is a unit vector and is therefore a point on the unit sphere which is known as the viewing sphere. The other two rows, denoted by $\mathbf{v}_x$ and $\mathbf{v}_y$, represent the main axes of the image. Given $\mathbf{v}_z$ the possible choices of $\mathbf{v}_x$ and $\mathbf{v}_y$ are related by a rotation in the image. It is important to note that there is no canonical way (due to the "Hairy Ball" Theorem, see, e.g., [8, p. 163]) to choose $\mathbf{v}_x$ and $\mathbf{v}_y$ for a given $\mathbf{v}_z$ which will guarantee that $\mathbf{v}_x(\mathbf{v}_z)$ and $\mathbf{v}_y(\mathbf{v}_z)$ will be continuous everywhere on the viewing sphere. In the discussion below we will decompose the rotation R between two images into two parts, a change in the viewing direction followed by a rotation in the image.

3 WEAK-PERSPECTIVE MOTION FROM TWO IMAGES

In this section, we show how to extract all the possible motion parameters from two weak-perspective images. Most of the results presented below were derived also in [10], [13]. We are given two images taken by a weak-perspective camera. As there is no known global coordinate frame with which to associate the objects in the scene, we will use a coordinate frame which is related to the first image. Under that frame the x and y image coordinates coincide with the X and Y world coordinates and the Z coordinate coincides with the viewing direction.

Given two images, the relationships between corresponding points in the two images are given by (4), namely,

$$\mathbf{p}'_i = s\left(R\mathbf{p}_i + \mathbf{t}\right), \tag{5}$$

where $\mathbf{p}_i = (x_i, y_i, z_i)^T$, $\mathbf{p}'_i = \left(x'_i, y'_i, z'_i\right)^T (1 \le i \le n)$, z_i and z'_i represent the unknown depth coordinates, and R, s, and $\mathbf{t}$ represent the unknown motion parameters. Our task is to recover as much information as possible about R, s, and $\mathbf{t}$ from the set of corresponding pairs $(\mathbf{p}_i, \mathbf{p}'_i)$. Consider the topmost two equations:

$$x'_i = sr_{11}x_i + sr_{12}y_i + sr_{13}z_i + st_x$$
$$y'_i = sr_{21}x_i + sr_{22}y_i + sr_{23}z_i + st_y \tag{6}$$

where r_{ij} are the components of R and $\mathbf{t} = (t_x, t_y, t_z)$.

Eliminating z_i from (6), we obtain

$$r_{23}x'_i - r_{13}y'_i + s\left(r_{21}r_{13} - r_{11}r_{23}\right)x_i +$$
$$s\left(r_{22}r_{13} - r_{12}r_{23}\right)y_i + s\left(t_yr_{13} - t_xr_{23}\right) = 0 \tag{7}$$

which relates corresponding points in the two images. Since R is orthonormal, this equation simplifies to

$$r_{23}x'_i - r_{13}y'_i + sr_{32}x_i - sr_{31}y_i + s\left(t_yr_{13} - t_xr_{23}\right) = 0, \tag{8}$$

which is a linear relationship of the form

$$Ax'_i + By'_i + Cx_i + Dy_i + E = 0. \tag{9}$$

This equation defines the epipolar relationship between corresponding points in the two images. A point $\mathbf{p}_i = (x_i, y_i)$ in the first image, defines a line in the other image, where its corresponding point must lie. The equation for this line is given by $Ax'_i + By'_i + K = 0$, where $K = Cx_i + Dy_i + E$. For a different point, the equation for the epipolar line will differ only by the constant K. Thus, the epipolar lines in weak-perspective projection are always parallel to each other. An example of the epipolar lines found for a pair of images can be seen in Fig. 2.

Given two images the unknown coefficients (A, B, C, D, and E) in (9) can be found using four pairs of corresponding points. When more than four correspondences are used, a more accurate solution is obtained using a linear least-squares algorithm. In addition, robust estimation techniques may be used to discard false correspondences [5]. Notice that (9) is homogeneous in these unknowns, and so a solution can be obtained only up to an arbitrary multiplicative factor. For convenience we scale (A, B, C, D, E) by a constant such that $A^2 + B^2 = 1$. Under this condition $r_{23} = kA$, $r_{13} = -kB$ for some $k \in \mathbb{R}$. Notice that since $r_{13}^2 + r_{23}^2 \le 1 = A^2 + B^2$, then $|k| \le 1$.

Given that

$$k(A, B, C, D, E) = (r_{23}, -r_{13}, sr_{32}, -sr_{31}, s(t_yr_{13} - t_xr_{23})), \tag{10}$$

we will now determine what motion parameters can and cannot be extracted from the two images.

We will first deal with the viewing direction. When (10) is substituted into the rotation matrix R, we get

$$R = \begin{pmatrix} * & * & -kB \\ * & * & kA \\ -kD/s & kC/s & \pm\sqrt{1-k^2} \end{pmatrix}. \tag{11}$$

Fig. 2. Two images of a scene. Ten pairs of corresponding points have been extracted and the points (denoted by the circles) and their respective epipolar lines have been overlaid over the original images.
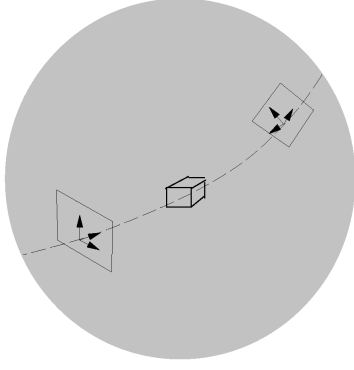
Fig. 3. Recovering the viewing direction component of motion. All that can be recovered is the great circle on the viewing sphere on which the viewing direction of the second image lies.

The entries denoted by * can also be given as functions of the linear coefficients, but as we will not be using them in the derivation, their expressions are omitted. Since $|k| \leq 1$, we can define an angle ϕ such that $\sin\phi \equiv k$, transforming R into

$$R = \begin{pmatrix} * & * & -B\sin\phi \\ * & * & A\sin\phi \\ -(D/s)\sin\phi & (C/s)\sin\phi & \cos\phi \end{pmatrix}. \quad (12)$$

The third row of R is the viewing direction of the second image. Thus the viewing directions for the two images are:

$$(0, 0, 1) \text{ and } (-\sin(\phi)D/s, \sin(\phi)C/s, \cos(\phi)).$$

Both viewing directions, regardless of the value of ϕ, are orthogonal to the vector $(C, D, 0)$, and therefore so is the plane defined by these two vectors (and the origin). Intersecting this plane with the viewing sphere yields a great circle on the viewing sphere (Fig. 3). A great circle on a sphere is the shortest path between two points on the sphere. Therefore, even though we do not know the viewing direction of the second image we know the shortest path on the viewing sphere connecting the two viewing directions. Note that the unrecoverable parameter ϕ represents the arclength extent between the two viewing directions along the viewing sphere. Alternatively, ϕ is the angle of rotation of the camera about the axis $(C, D, 0)$.

The method we will use to describe how to recover the other components of the motion is as follows. We will first eliminate parameters so as to cause the corresponding epipolar lines of $\mathbf{p}_i$ and $\mathbf{p}'_i$ to coincide and then try to make the actual points within the lines coincide.

We will deal first with image rotation. As mentioned above there is no canonic way to define $\mathbf{v}_x$ and $\mathbf{v}_y$ for a given $\mathbf{v}_z$. Therefore, we will define $\mathbf{v}_x$ and $\mathbf{v}_y$ for the second image related to the first image. While the viewing direction moves on the great circle toward the viewing direction of the second image $\mathbf{v}_x$ and $\mathbf{v}_y$ are defined such that the angles between the epipolar lines and the x and y coordinates in the image do not change. Thus, if the epipolar lines in the two images are not parallel to each other (Fig. 4 top), that is due to image rotation (Fig. 4 middle). To simplify the description we may instead rotate both images such that the epipolar lines in both images are parallel to the x axis (Fig. 4 bottom). Formally, the image rotation between the two images can be recovered by computing the angle between (A, B) and (C, D).

The scale is recovered by computing the ratio of the distances between the epipolar lines of $\mathbf{p}_i$ and $\mathbf{p}_j$ and those of $\mathbf{p}'_i$ and $\mathbf{p}'_j$ (Fig. 5 top and middle). This ratio is given by:
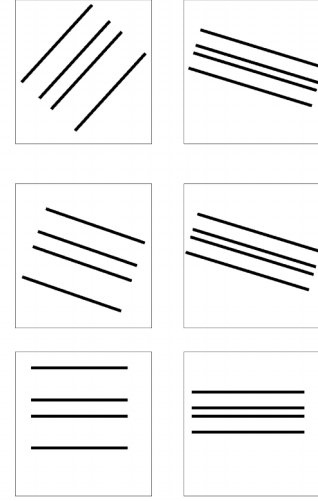

Fig. 4. Recovering image rotation. (Top) Epipolar lines of two images. (Middle) The first image is rotated making the epipolar lines of both images parallel to each other. (Bottom) Both images are rotated making the epipolar lines of both images parallel to the x axis.
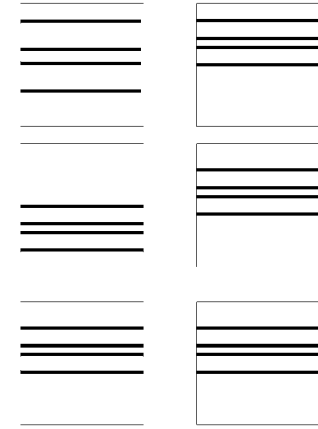

Fig. 5. Recovering image scale and translation. (Top) Epipolar lines of two images. (Middle) The first image is scaled making the distance between pairs of corresponding epipolar lines equal in the two images. (Bottom) The image is translated in the direction orthogonal to the epipolar lines such that the epipolar lines coincide.

$$\frac{C^2 + D^2}{A^2 + B^2} = \frac{s^2 k^2 \left(r_{32}^2 + r_{31}^2\right)}{k^2 \left(r_{23}^2 + r_{13}^2\right)} = \frac{s^2 k^2 \left(1 - r_{33}^2\right)}{k^2 \left(1 - r_{33}^2\right)} = s^2. \quad (13)$$

As $A^2 + B^2 = 1$, the above expression simplifies to $s = \sqrt{C^2 + D^2}$.

Finally, we turn to recovering the translation components. Under a scaled-orthographic projection the depth component of the translation, t_z, cannot be recovered. Of the two other components, t_x and t_y, only the component of translation orthogonal to the epipolar lines is recoverable. This component can be recovered by translating the image such that the epipolar lines of corresponding points coincide (Fig. 5 bottom). Since the linear coefficient E equals $E = -s(t_x, t_y) \cdot (A, B)^T$, $-E/s$ is the value of this component of the translation.

The component of translation along the epipolar lines cannot be recovered because the change in the position of points in this direction is due not only to translation but also due to changes in the viewing direction between the two images. The magnitude of motion of points along the epipolar lines is affected by their depth values, and since these values are unknown, it is impossible to decouple this motion from the amount of translation along the epipolar lines.

In [15], [7], [14], both translation components are eliminated by setting the origin at either one of the feature points or at the centroid of the points. This is achieved by translating the selected feature point (or the centroid) in each of the images to the same position. Notice first that the translation recovered in this method would be different according to the choice of an origin. If indeed the centroid of the points corresponds to the center of mass of the object then the translation recovered in this method would correspond to the actual fronto-parallel translation of the object assuming a static camera. However, if the scene is in fact static and the camera is moving the actual translation of the camera cannot be fully recovered in this method because such a recovery would require knowledge of the 3D position of the selected feature point (or the centroid, that is, knowledge of $\overline{Z}$) with respect to the camera. To see this, consider the example of two images which differ only by a change in viewing direction. If we pick an origin other than the true origin of the images then we will attribute the difference in the positions of the origin in the images to translation (although in fact it is due to the change in viewing direction). Furthermore, each different selection of an origin will imply a different translation.

In conclusion, from two weak-perspective images we are able to recover the great circle on the viewing sphere on which the viewing directions of the two images lie, the image rotation and scale, and the component of translation orthogonal to the epipolar lines. The angle of change in viewing direction and the component of camera translation along the epipolar lines cannot be recovered from two images.

4 WEAK-PERSPECTIVE MOTION FROM THREE IMAGES

As we have seen in the previous section, not all the components of the motion can be recovered from two images. When three images are given, Ullman [15] has shown that motion and structure (up to a reflection) can be recovered, and Huang and Lee [7] have presented a linear algorithm to recover the motion and structure.

Huang and Lee's method is based on the following principles. For each pair of images, the coefficients (A, B, C, D, E) are recovered. As mentioned above, these coefficients form constraints on the entries of the corresponding rotation matrix R. The three rotation matrices $R_{1 \to 2}$, $R_{2 \to 3}$, and $R_{1 \to 3}$ satisfy the following relation:

$$R_{1 \to 2} R_{2 \to 3} = R_{1 \to 3}. \tag{14}$$

Using this relation and the orthonormality of the rotation matrices, Huang and Lee describe a series of algebraic steps which eventually lead to the complete recovery of the three matrices.

The drawback of this method is that it does not offer a geometric interpretation of the problem. We will now use the results derived in the previous section to construct a simple geometric solution to the three image motion problem.

Given three weak-perspective images of a scene, our main problem is to find the viewing directions of the second and third images with respect to the first image. The geometric setting of the three images on the viewing sphere is shown in Fig. 6. The three viewing directions are unit vectors, and so they can be regarded as points on the viewing sphere. WLOG we can assume that the viewing direction of the first image, denoted by $\mathbf{v}_{z1}$, is (0, 0, 1). The viewing directions of the other two images, denoted by $\mathbf{v}_{z2}$ and $\mathbf{v}_{z3}$, are not yet known. The epipolar lines in the first image determine two great circles on the viewing sphere through $\mathbf{v}_{z1}$ along which $\mathbf{v}_{z2}$ and $\mathbf{v}_{z3}$ must lie (Fig. 6). These great circles are known, since they are contained in planes that are parallel to the epipolar lines of the first image. However, the position of $\mathbf{v}_{z2}$ and $\mathbf{v}_{z3}$ along these great circles is not yet known. A third great circle passing through $\mathbf{v}_{z2}$ and $\mathbf{v}_{z3}$ completes a spherical triangle whose vertices are the tips of the three viewing directions, $\mathbf{v}_{z1}$, $\mathbf{v}_{z2}$, and $\mathbf{v}_{z3}$. The sides of this
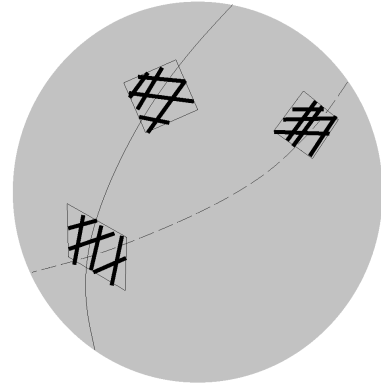

Fig. 6. The geometric setting of the three viewing directions. The unknown viewing directions lie on great circles which pass through the viewing direction of the first image. The epipolar lines of each pair of images are parallel to the great circle which runs between them.
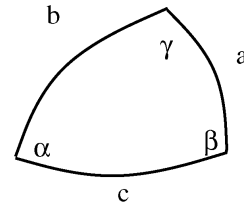

Fig. 7. A spherical triangle. The angles of the triangle are α, β, and γ. The angles between the vertices of the opposite edges and the center of the sphere are a, b, and c, respectively.

spherical triangle are the sections of the three great circles that connect the three viewing direction on the viewing sphere. The lengths of these sides correspond to the only missing unknowns in the three rotation matrices, the unknown angles $\phi_{i \to j}$ between the viewing directions $\mathbf{v}_{zi}$ and $\mathbf{v}_{zj}$ ($1 \leq i < j \leq 3$), which were mentioned in the previous section.

We thus need to recover the lengths of the three sides of the spherical triangle formed by the viewing directions in the three images. In each of the three images we recover the epipolar lines that relate it with the other two images. Since the epipolar lines are parallel to the planes that contain the sides of the spherical triangle, the angle between the two sets of epipolar lines in each image is identical to one of the angles of the spherical triangle. Thus, every image provides one angle of the triangle, and together we obtain all three angles, α, β, and γ, of the triangle.

Spherical triangles have the following well-known properties. The area of the spherical triangle with angles α, β, and γ on a sphere of radius r is given by: $S_r = r^2 \epsilon$, where $\epsilon = \alpha + \beta + \gamma - \pi$. Therefore, since we know the three angles, we can also recover the area of the triangle. For spherical triangles, there exist variants of the law of sines,

$$\frac{\sin \alpha}{\sin a} = \frac{\sin \beta}{\sin b} = \frac{\sin \gamma}{\sin c}, \tag{15}$$

and the law of cosines,

$$\cos a = \frac{\cos \alpha + \cos \beta \cos \gamma}{\sin \beta \sin \gamma}, \tag{16}$$

where a, b, c denote the lengths of the sides opposite to α, β, and γ respectively. (a, b, and c are the arc-lengths of the geodesic curves separating the three vertices of the triangle, or the angles formed by connecting the center of the sphere with any two vertices of the triangle, see Fig. 7.)

Using the law of sines and the law of cosines, we can recover a, b, and c and use them to recover also the unknown viewing direc-
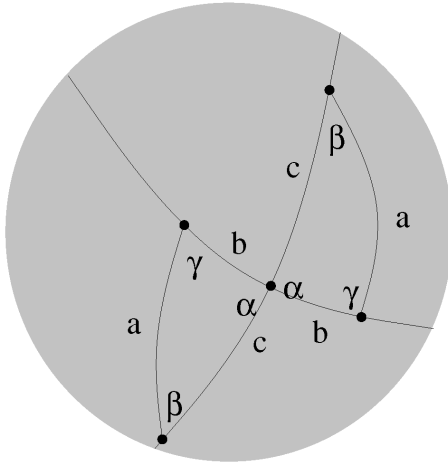
Fig. 8. The two solutions for the viewing directions of three given images as points on the viewing sphere yielding two triangles whose vertices lie on the corresponding great circles and whose angles are α, β, and γ.

tions and rotation matrices. Recovering the rotation matrix from these parameters is immediate. For instance, the rotation between the first and the second image, $R_{1 \to 2}$, can be recovered as follows. First rotate the first image so that its epipolar lines with respect to the second image will coincide with the x-axis. This is a rotation about the z-axis by an angle $\tan^{-1}(B/A)$ (where A and B are the coefficients of the epipolar lines from (9)). Next, rotate the points by the recovered angle $a = \phi_{1 \to 2}$ about the y-axis. Finally, undo the rotation that brings the epipolar lines in the second image with respect to the first image to the x-axis, that is, rotate about the z-axis by the angle $-\tan^{-1}(D/C)$.

Ullman [15] showed that there exist two solutions to the three image problem where in the second solution the 3D structure is the reflection with respect to the image plane of the 3D structure obtained in the first solution. We therefore expect to find a second triangle on the viewing sphere whose second and third vertices lie on the same two great circles, and whose angles are also α, β, and γ. The vertices of this triangle are obtained by moving away from $\mathbf{v}_{z1}$ along the same great circles the same distance, but in the opposite direction (Fig. 8).

It is straightforward to show that in this second solution, the rotation matrices $R'_{1 \to j}$, $j = 2, 3$ are:

$$R'_{1 \to j} = \begin{pmatrix} r_{11} & r_{21} & -r_{31} \\ r_{12} & r_{22} & -r_{32} \\ -r_{13} & -r_{23} & r_{33} \end{pmatrix}, \qquad (17)$$

where r_{lm} are the entries of $R_{1 \to j}$. It is also straightforward to show that if $\mathbf{P} = (X, Y, Z)$ is a 3D point recovered by the first solution, then $\mathbf{P}' = (X, Y, -Z)$, the reflected point, is the corresponding point for the other solution.

The algorithm above, like previous algorithms (e.g., [15], [7]), requires that all three angles, α, β, and γ will not vanish. In fact, as can be seen in (16), its results may be unstable if either of these angles approaches zero. This situation corresponds to the special case where the three viewing directions lie on the same great circle. Alternative methods can be used in this case, one of which is discussed in [16].

While rotation can be fully recovered given three weak-perspective images, translation cannot be fully recovered. We have seen in Section 3 that with two images only the component of translation orthogonal to the epipolar lines can be recovered, while the component of translation in the direction of the epipolar lines cannot be recovered. Given three images denote by $\mathbf{e}_{i \to j}$ the direction of the epipolar lines between image i and j ($1 \le i < j \le 3$), and by $\mathbf{e}_{i \to j}^{\perp}$ the orthogonal vector. Let $u_{i \to j}$ denote the unknown translation component along the epipolar lines, and let $v_{i \to j}$ denote the known translation component orthogonal to the epipolar line, then

$$\left(\mathbf{e}_{1 \to 2}, \mathbf{e}_{1 \to 2}^{\perp}\right)\begin{pmatrix} u_{1 \to 2} \\ v_{1 \to 2} \end{pmatrix} + \left(\mathbf{e}_{2 \to 3}, \mathbf{e}_{2 \to 3}^{\perp}\right)\begin{pmatrix} u_{2 \to 3} \\ v_{2 \to 3} \end{pmatrix} =$$
$$\left(\mathbf{e}_{1 \to 3}, \mathbf{e}_{1 \to 3}^{\perp}\right)\begin{pmatrix} u_{1 \to 3} \\ v_{1 \to 3} \end{pmatrix} \qquad (18)$$

These are two linear equation in three unknowns, the three $u_{i \to j}$s. The three unknowns, therefore, can be determined only up to one missing parameter, the translation along the epipolar lines between one image pair. For every given value of translation along the epipolar lines in any pair of images, there is a setting of translation between the two other pairs that will be compatible with the three images. It can be readily shown that the different translations can be obtained by setting the unknown value of $\overline{Z}$, the mean depth of the points.

To demonstrate the algorithm, we took three images of a scene. The camera was mounted on a robot arm, and its motion was con-

Fig. 9. (Top) Three images of a scene. (Bottom) Fifteen triplets of corresponding points have been extracted and the points (denoted by the circles) and their respective epipolar lines have been overlaid over the original images.
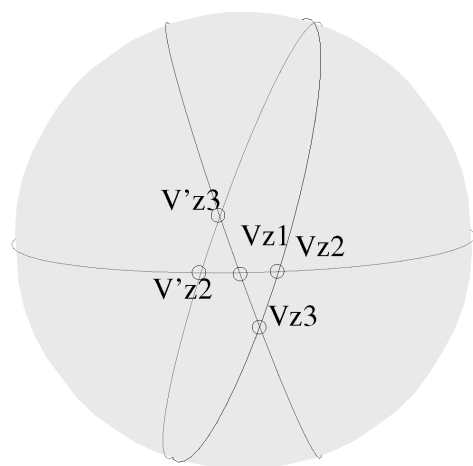
Fig. 10. The two solutions for the viewing directions of the three images shown in Fig. 9. The solutions are shown as points on the viewing sphere yielding two triangles whose vertices lie on the corresponding two great circles through $\mathbf{v}_{z1}$.

trolled. The true viewing directions were separated by $14.85°$, $9.66°$, and $14.85°$. We manually extracted and matched fifteen points on each image and used them to recover the epipolar lines. Fig. 9 shows an overlay of these epipolar lines on the three images. We then computed the angles between the epipolar lines in each image and recovered the viewing directions of the images. The two solutions for the viewing directions are shown in Fig. 10. The recovered viewing directions were separated by $13.97°$, $8.11°$, and $16.31°$, respectively. The error therefore did not exceed $1.55°$.

5 SUMMARY

We have presented in this paper a geometric interpretation of motion recovery from two and three images taken by a weak-perspective camera. In the case of two images we have shown which motion parameters can and cannot be recovered and how to derive those parameters from the epipolar constraints. We then used these results to develop a simple intuitive geometric algorithm to fully recover the motion parameters when three images are given. Our algorithm is based on reducing the problem of estimating motion from three images to a problem of finding on a sphere triangles whose angles are known. These angles are the angles between the epipolar lines in each of the three images. This new interpretation offers a better understanding of the motion recovery problem under weak-perspective projection and provides an algorithm which is straightforward to program.

ACKNOWLEDGMENTS

REFERENCES

[1] Y. Aloimonos, "Perspective Approximations," *Image and Vision Computing*, vol. 8, no. 3, pp. 179-192, Aug. 1990.

[2] R. Basri, "Paraperspective = Affine," *Int'l J. of Computer Vision*, vol. 19, no. 2, pp. 169-179, Aug. 1996.

[3] J. Costeira and T. Kanade, "A Multi-Body Factorization Method for Motion Analysis," *ICCV95*, 1995, pp. 1,071-1,076.

[4] C. Debrunner and N. Ahuja, "Motion and Structure Factorization and Segmentation of Long Multiple Motion Image Sequences," *Proc. European Conf. Computer Vision*, 1992, pp. 217-221.

[5] M.A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381-395, June 1981.

[6] W.C. Gear, "Feature Grouping in Moving Objects," *IEEE Workshop on Non-Rigid and Articulate Motion*, 1994.

[7] T.S. Huang and C.H. Lee, "Motion and Structure From Orthographic Projections," *Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 536-540, May 1989.

[8] J.J. Koenderink, *Solid Shape*. Cambridge, Mass.: MIT Press, 1990.

[9] L.L. Kontsevich, "Pairwise Comparison Technique: A Simple Solution for Depth Reconstruction," *J. Optical Soc. of Am.*, vol. 10, no. 6, pp. 1,129-1,135, June 1993.

[10] C.H. Lee and T.S. Huang, "Finding Point Correspondences and Determining Motion of a Rigid Object From Two Weak Perspective Views," *Computer Vision and Graphic Image Processing*, vol. 52, no. 3, pp. 309-327, Dec. 1990.

[11] Y. Ohta, and K. Maenobu, and T. Sakai, "Obtaining Surface Orientation From Texels Under Perspective Projection," *Proc. Int'l Conf. Artificial Intelligence*, 1981, pp. 746-751.

[12] C.J. Poelman and T. Kanade, "A Paraperspective Factorization Method for Shape and Motion Recovery," *Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 206-218, Mar. 1997.

[13] L.S. Shapiro, A. Zisserman, and M. Brady, "3D Motion Recovery via Affine Epipolar Geometry," *Int'l J. Computer Vision*, vol. 16, no. 2, pp. 147-182, Oct. 1995.

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

[15] S. Ullman, *The Interpretation of Visual Motion*. Cambridge, Mass.: M.I.T. Press, 1979.

[16] S. Ulmann and R. Basri, "Recognition by Linear Combination of Models," *Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 992-1,006, Oct. 1991.