# Recovering Epipolar Direction from Two Affine Views of a Planar Object[⋆]

Maria Alberich-Carramiñana

*Departament de Matemàtica Aplicada I, UPC*
*Avda. Diagonal 647, 08028 Barcelona*

Guillem Alenyà, Carme Torras

*Institut de Robòtica i Informàtica Industrial, CSIC-UPC*
*Llorens i Artigas 4-6, 08028 Barcelona*

Juan Andrade-Cetto

*Centre de Visió per Computador, UAB*
*Edifici O, Campus UAB, 08193 Bellaterra, Spain*

Elisa Martínez

*CiTS La Salle, Universitat Ramon Llull*
*Pge. Bonanova 8, 08022 Barcelona*

**Abstract**

The mainstream approach to estimate epipolar geometry from two views requires matching the projections of at least 4 non-coplanar points in the scene, assuming a full projective camera model. Our work deviates from this in three respects: affine camera, planar scene and active contour tracking instead of point matching. Using results in Projective Geometry, we prove that the affine epipolar direction can be recovered provided camera motion is free of cyclorotation. A setup consisting of a Staübli robot holding a planar object in front of a camera is used to obtain calibrated image streams, which are used as ground truth to evaluate the performance of the method, and to test its limiting conditions in practice. The fact that our method (applicable to planar, poorly textured scenes) and the Gold Standard algorithm (applicable to highly textured scenes with significant relief) produce comparable results shows the potential of our proposal.

*Key words:* Affine epipolar direction estimation, active contours, precision analysis

--------

[⋆] A preliminary version of this work was presented at ACIVS[1]

## 1 Introduction

Recovering camera motion and scene structure from image streams is an important task in a range of applications including robot navigation and manipulation. As a basic step underlying this task, extensive work has been devoted in the last decade to estimating epipolar geometry from two views. The methods proposed address different versions of the problem depending on: a) the camera model they assume, b) the types of scenes to which they apply, and c) how visual motion is measured on the image plane.

The present work addresses epipolar estimation for an *affine* camera viewing a *planar* scene, and using *active contours* to measure visual motion within images. Below we discuss each of these three options in turn.

The camera model relates visual motion on the image plane to 3D camera motion. Depending on the viewing conditions, different camera models have been used to emulate the imaging process [2,3]. The full perspective model (the pinhole camera), in either its calibrated (perspective camera) or uncalibrated (projective camera) versions, has proved to be too general when perspective effects diminish. Under weak-perspective viewing conditions (small field of view, or small depth variation in the scene along the line of sight compared to its average distance from the camera), simplified camera models, such as orthographic, scaled-orthographic or their generalization for the uncalibrated case, the affine camera model, provide an advantageous approximation to the pinhole camera, which avoids computing ill-conditioned parameters by explicitly incorporating the ambiguities due to weak perspective into the model.

Epipolar geometry estimation under affine assumption has received a lot of attention in the literature [4,5]. However, most of the known methods assume that:

(1) the scene contains **depth information**, and thus the algorithms fail when the scene configuration approaches a planar structure. For instance, our experimentation with the Gold Standard algorithm (see Section 6) shows that the relief of the target object should be of the same order of magnitude as the object length/width in order to obtain acceptable results; and

(2) the scene is **textured** enough to allow visual motion estimation from point correspondences.

Our work explores *an alternative approach* applicable when these two assumptions do not hold.

Previous attempts at eluding assumption 1), and thus estimating epipolar geometry from a dominant plane, have all used a full-perspective camera. It is

well known that two views of a plane are related by a collineation under full perspective projection. Several authors have used this fact to propose algorithms for camera calibration [6], self-calibration [7,8], or extraction of structure and motion from uncalibrated views of points on planes [9,10,11]. However, when perspective effects diminish, the relationship between two views of a planar structure becomes an affinity, which invalidates the methods based on collineations. To the best of our knowledge, there are no previous works eluding assumption 1) under affine viewing.

Concerning assumption 2), we estimate the visual motion between two views using active contours, thus being able to handle poorly textured scenes. The affinity relating two views, which we obtain from the tracking of a planar contour, is theoretically equivalent to that resulting from three point matches. However, as mentioned, our method eludes assumption 2), and benefits also from other advantages of active contours, such as their robustness to occlusions and noise, as well as the simplicity of the tracking procedure, avoiding point matching.

Using results in projective geometry, we show that, under a 3D motion free of cyclorotation, the epipolar direction can be recovered by relating the two affine views of a planar contour. To validate the theoretical proofs and to test their limiting conditions (relaxation of weak perspective, effects of cyclorotation and noise), a series of experiments has been performed. This testing is not aimed at a particular application at this stage, since we view the results of this paper as a necessary step towards the recovery of other motion and structure information (such as plane orientation) from three affine views of a planar object.

The paper is organized as follows. Section 2 contains the analytic study of two weak-perspective views and provides the basis for the recovery of the epipolar direction. Section 3 explains how the parameters of the affinity relating the two views are extracted in our implementation, based on a contour tracker. Section 4 is devoted to experimentation, using both synthetic and real image streams. Finally, Section 5 summarizes our contribution. Furthermore, Appendix A contains a geometric proof of our main contribution in Section 2, whereas Appendix B contains a generalization to two affine views.

## 2 Analytic study of two weak-perspective views

### 2.1 The camera model

A weak-perspective (or scaled-orthographic) camera $\psi$ projects a scene point first orthographically onto the average depth plane $\mathcal{R}^C$ (the plane parallel to the image plane $\mathcal{R}$ containing the centroid $C$ of the scene object) and then perspectively from this fronto-parallel plane $\mathcal{R}^C$ onto the image $\mathcal{R}$ (Fig. 1). Thus the second perspective projection (of proper optical center, say $P$) simply introduces a scale factor. Modulus this scale factor, the weak-perspective camera is an orthographic camera: all the projection rays are parallel with direction orthogonal to the image plane $\mathcal{R}$.

Let an affine world frame $\psi$ associated to the weak-perspective camera be chosen as follows: take any affine coordinate frame in the image plane $\mathcal{R}$, and add to it an orthogonal vector to $\mathcal{R}$ in order to obtain a global affine world frame. Under these frames, the $x$ and $y$ image coordinates coincide with the $X$ and $Y$ world coordinates, respectively, and the $Z$ coordinate coincides with the viewing direction, and thus the weak-perspective projection is given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z_{\mathrm{ave}}} \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{1}$$

where $f$ is the focal length, and $Z_{\mathrm{ave}}$ is the average distance of the object from the camera (that is, the plane $\mathcal{R}^C$ has equation $Z = Z_{\mathrm{ave}}$). When the depth variation of the object is small compared to $Z_{\mathrm{ave}}$, and the principal point is close to the centroid of the projected object, then the weak-perspective camera model gives an approximation of the perspective projection [12].

We assume that the scene object is stationary and that the camera translates by $\mathbf{T}$ and rotates by $\mathbf{R}$ around the object, and possibly zooms, thus giving a second weak-perspective camera $\psi'$. The new affine coordinate frames associated with the second camera are given by the rows of $\mathbf{R}$ and the new origin lies at $-\mathbf{R}^\top \mathbf{T}$, thus $\psi'$ has the expression

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{f'}{Z'_{\mathrm{ave}}} \begin{bmatrix} X' \\ Y' \end{bmatrix}, \tag{2}$$

where $[X', Y', Z']^\top = \mathbf{R}[X, Y, Z]^\top + \mathbf{T}$, $f'$ is the new focal length, and $Z'_{\mathrm{ave}}$ is the average distance to the object from the second camera.

Consider the equation $aX + bY + c = Z$ of a world plane $\mathcal{S}$. Then the two
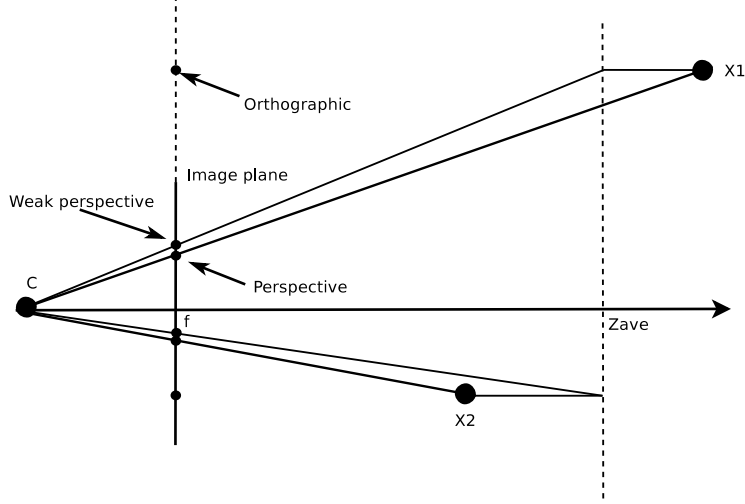
Figure 1. Affine projection compared to the (full) perspective projection and the orthographic camera models. Affine projections (and in particular, a weak-perspective projection) give a better approximation to the perspective projection than the other simplified model, the orthographic camera. In the affine camera model, a parallel projection is made onto the average depth plane, prior to an overall perspective projection (scaling). For example, in the paraperspective model, this projection is in the direction parallel to the perspective projection of the target centroid. However, in the weak-perspective model, the parallel projection onto the average depth plane is made in a direction perpendicular to the image plane. Under the condition that the target is seen with a small field of view, these two fixed directions of parallel projection give a good approximation for all the points of the target, compared to the orthographic camera model.

views of the coplanar scene are related by the affinity given by

$$
\begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{M} \begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{t}\,,
\tag{3}
$$

with

$$
\mathbf{M} = s\frac{f'}{f} \begin{bmatrix} R_{1,1} + aR_{1,3} & R_{1,2} + bR_{1,3} \\ R_{2,1} + aR_{2,3} & R_{2,2} + bR_{2,3} \end{bmatrix},
\tag{4}
$$

$$
\mathbf{t} = \frac{f'}{Z'_{\text{ave}}} \begin{bmatrix} T_x + cR_{1,3} \\ T_y + cR_{2,3} \end{bmatrix},
\tag{5}
$$

and where $s = Z_{\text{ave}}/Z'_{\text{ave}}$ is the scale factor that accounts for depth variation ($s > 1$ if the second camera approaches the scene object, and $s < 1$ if it departs from it), and $R_{i,j}$ are the elements of the rotation matrix $\mathbf{R}$.

A direction $\mathbf{v} = [x, y]^\top$ of the first image $\mathcal{R}$ is mapped by the above affinity to the direction $\mathbf{Mv}$ of the second image $\mathcal{R}'$. Since the affine references chosen in

5

the two cameras match by the displacement, we can superpose the two images and it has sense to consider directions invariant by $\mathbf{M}$.

## 2.2 Recovery of the epipolar direction

Consider an orthonormal coordinate frame associated to the first image (for instance, normalized pixel coordinates, when aspect ratio and skew are known). The rotation matrix about the unit axis $[\cos\alpha, \sin\alpha, 0]^\top$ and angle $\rho$ has the form

$$
\mathbf{R} = \begin{bmatrix}
(1-\cos\rho)\cos^2\alpha + \cos\rho & \cos\alpha\sin\alpha(1-\cos\rho) & \sin\alpha\sin\rho \\
\cos\alpha\sin\alpha(1-\cos\rho) & (1-\cos\rho)\sin^2\alpha + \cos\rho & -\cos\alpha\sin\rho \\
-\sin\alpha\sin\rho & \cos\alpha\sin\rho & \cos\rho
\end{bmatrix}.
\tag{6}
$$

Hence, the matrix $\mathbf{M}$ is

$$
\mathbf{M} = s\frac{f'}{f}\begin{bmatrix}
\begin{array}{c} (1-\cos\rho)\cos^2\alpha \\ +\cos\rho + a\sin\alpha\sin\rho \end{array} & \begin{array}{c} \cos\alpha\sin\alpha(1-\cos\rho) \\ +b\sin\alpha\sin\rho \end{array} \\
\begin{array}{c} \cos\alpha\sin\alpha(1-\cos\rho) \\ -a\cos\alpha\sin\rho \end{array} & \begin{array}{c} (1-\cos\rho)\sin^2\alpha \\ +\cos\rho - b\cos\alpha\sin\rho \end{array}
\end{bmatrix},
\tag{7}
$$

where $\mathbf{a} = [\cos\alpha, \sin\alpha]^\top$ is the direction of the rotation axis. The orthogonal vector $\mathbf{e} = [-\sin\alpha, \cos\alpha]^\top = \mathbf{a}^\perp$ is the epipolar direction. A straightforward computation shows that

$$
\mathbf{M}\mathbf{e} = s\frac{f'}{f}(\cos\rho + \sin\rho(a\sin\alpha - b\cos\alpha))\mathbf{e},
\tag{8}
$$

thus giving an analytic proof of the following result:

**Theorem 1** If the rigid motion between two weak-perspective cameras is assumed to be free of cyclorotation, then the epipolar direction $\mathbf{e}$ can be recovered as one of the two eigenvectors of the vectorial part $\mathbf{M}$ of the affinity that relates two views of a planar scene.

As a consequence, the direction $\mathbf{a} = \mathbf{e}^\perp$ of the axis of rotation can also be recovered.

Figure 2 illustrates the above result. Two views $\mathcal{R}$ and $\mathcal{R}'$ of a planar H-shaped object are shown, which are related by a rotation about an axis parallel to the
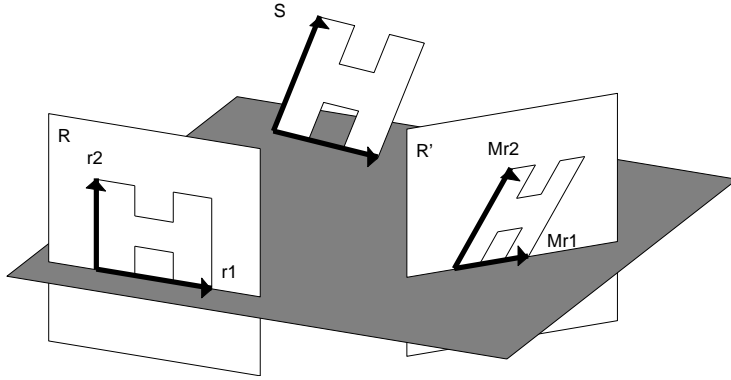
Figure 2. Graphic illustration of Theorem 1. See text for details.

image plane (i.e., free of cyclorotation). For simplicity of illustration, a basis $\{\mathbf{r_1}, \mathbf{r_2}\}$ is chosen aligned with the main axes of the H, and the axis of rotation is taken to be parallel to $\mathbf{r_2}$. Thus, the gray plane swept by $\mathbf{r_1}$ is left invariant by the rotation. Note, then, that the epipolar direction is that of $\mathbf{r_1}$ in $\mathcal{R}$ and that of $\mathbf{Mr_1}$ in $\mathcal{R}'$, and its perpendicular within each image is the direction of the rotation axis.

A geometric proof of Theorem 1 is included in Appendix A. Within the same geometrical framework, in Appendix B, this result is generalized to the affine camera model giving Theorem 4. Let us sketch the main ideas of this generalized result; the reader is referred to Appendix B for the details of the proof. The main advantage of this generalization is that, within the affine camera model, the projected target does not need to be centered in the image (assuming that the image center is a good approximation to the principal point). This enables us to handle a broader range of situations where the condition of small field of view is satisfied but the condition of being centered is relaxed, as, for instance, those arising in the experiments performed in Section 4.3.3. The affine camera model, which encloses the weak-perspective one, projects a scene point first under a fixed direction (which corresponds to a point $\overline{O}$ lying on the plane at infinity $\Pi_\infty$) onto the average depth plane $\mathcal{R}^C$ (the plane parallel to the image plane $\mathcal{R}$ containing the centroid $C$ of the scene object), and then perspectively from this fronto-parallel plane $\mathcal{R}^C$ onto the image $\mathcal{R}$ (Fig. 1). When $\overline{O}$ equals the direction $O$ orthogonal to the image plane, the affine camera becomes a weak-perspective camera. By this projection procedure it is inferred that the affine camera, as well as the weak-perspective camera, preserves parallelism.

While in the weak-perspective camera model the improper optical center $O$ is determined by the orientation of the image plane (i.e., $O$ is the pole with respect to the absolute conic $\Omega$ of the improper line $r$ of $\mathcal{R}$), in the affine camera model the improper optical center $\overline{O}$ may be any point in $\Pi_\infty$. In fact, the direction of parallel projection, i.e., the improper optical center, depends on the position of the projected target within the image plane. This implies, on

7

the one hand, that the same (pinhole) camera under affine viewing conditions can take two affine views with different improper optical centers (but keeping the same image plane). On the other hand, this also implies that, while the orientation of the image plane (and hence the improper optical center in case of a weak-perspective camera) is determined by the displacement performed by the camera, the improper optical center is not determined by the camera motion in the more general case of an affine camera. This is one of the reasons that makes the affine camera model more difficult to handle than the weak-perspective one.

Since the improper optical centers lie at infinity, the epipoles (of the first and second affine cameras) are also located at infinity in the image planes, i.e., the epipolar lines in both views are parallel. But, while in the weak-perspective cameras the epipoles coincide with the orthogonal direction (in the image plane) of the axis of rotation, in the general affine cameras the epipoles are no more related to this distinguished direction and, thus, a priori, they do not provide information about the rigid motion between the two affine cameras. This explains why most of the literature about the general affine camera model switches to the weak-perspective camera model when the question of inferring camera motion is addressed. By this same reason, our generalization to the affine camera is valid only for viewing directions that agree with these distinguished directions (which would be, in fact, the epipolar directions if the camera would be considered as a weak-perspective one). Let us state the announced result:

**Theorem 2** *(see Theorem 4 in Appendix B)* Assume that the rigid motion between two affine cameras is free of cyclorotation and that the target projections are shifted (from the center of the image) along the direction orthogonal to the axis of rotation. Then the epipolar direction can be recovered as one of the two eigenvectors of the vectorial part $\mathbf{M}$ of the affinity that relates the two affine views of a planar scene.

*2.3   Computing the epipolar direction from the affinity parameters*

Fix any coordinate frame in the image (for instance pixel coordinates, since orthonormality is not required) and assume that the affinity that relates the two views has the expression

$$\mathbf{x}' = \mathbf{M}\mathbf{x} + \mathbf{t} = \begin{bmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \tag{9}$$

In virtue of Theorem 1, the epipolar direction is one of the eigenvectors of $\mathbf{M}$.

An eigenvector $[1, w]^\top$ of $\mathbf{M}$ satisfies the equation

$$M_{1,2}w^2 + (M_{1,1} - M_{2,2})w - M_{2,1} = 0 \,. \tag{10}$$

If the motion is under the hypothesis of Theorem 1, then (10) must have two real solutions $w_1$, $w_2$, and the epipolar direction is $\mathbf{e} = [1, w_i]^\top$, for some $i \in \{1, 2\}$ (or $[0, 1]^\top$, in case $M_{1,2} = 0$).

## 3  Extracting the affinity parameters in our implementation

The affinity that relates two affine views is usually computed from a set of point matches. However, point matching is still one of the key bottlenecks in computer vision. In this work an active contour [13] is used instead. The active contour is fitted to a target object and the change of the active contour between different views is described by a shape vector deduced as follows. The contour is first represented as a parametric spline curve as it is common in Computer Graphics [14], $\mathbf{d}(s) = (d_x(s), d_y(s))^\top$, where both $d_x(s)$ and $d_y(s)$ are B-spline curves. They can be written as a function of their control points $\mathbf{Q}$,

$$\mathbf{d}(s) = \begin{bmatrix} d_x(s) \\ d_y(s) \end{bmatrix} = \begin{bmatrix} \mathbf{B}(s)\mathbf{Q^x} \\ \mathbf{B}(s)\mathbf{Q^y} \end{bmatrix} = \begin{bmatrix} \mathbf{B}(s) & \mathbf{0}^\top \\ \mathbf{0}^\top & \mathbf{B}(s) \end{bmatrix} \begin{bmatrix} \mathbf{Q^x} \\ \mathbf{Q^y} \end{bmatrix} = \mathbf{U}(s)\mathbf{Q} \,, \tag{11}$$

where $\mathbf{Q^x}$ and $\mathbf{Q^y}$ are column vectors of the $x$ and $y$ components of the control points, $\mathbf{0}$ is a column vector of zeros, and $\mathbf{U}(s) = \mathbf{I} \otimes \mathbf{B}(s)$ is the Kronecker product between the identity matrix $\mathbf{I}$ and the row vector $\mathbf{B}(s)$ of B-spline basis functions [13,14].

When the contour is observed in two different views, equation (3) can be generalized to describe their relation as:

$$\mathbf{d}'(s) = \mathbf{M}\mathbf{d}(s) + \mathbf{t} \,, \tag{12}$$

where $\mathbf{d}'(s)$, $\mathbf{d}(s)$ are the contour in the two views. Getting expression (11) into equation (12), we obtain

$$\mathbf{d}'(s) = \mathbf{M}\mathbf{U}(s)\mathbf{Q} + \mathbf{t} \,. \tag{13}$$

Now, defining $\mathbf{1}$ as a column vector of ones, one can observe that $\mathbf{B}(s)\mathbf{1} = 1$ from the convex hull property of B-spline curves, and using equation (11) again, the difference between $\mathbf{d}'(s)$ and $\mathbf{d}(s)$ can be written as

$$\mathbf{d}'(s) - \mathbf{d}(s) \;=\; t_x \mathbf{U}(s) \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} + t_y \mathbf{U}(s) \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} + (M_{11} - 1)\mathbf{U}(s) \begin{bmatrix} \mathbf{Q^x} \\ \mathbf{0} \end{bmatrix}$$

$$+ \; M_{12}\mathbf{U}(s) \begin{bmatrix} \mathbf{Q^y} \\ \mathbf{0} \end{bmatrix} + M_{21}\mathbf{U}(s) \begin{bmatrix} \mathbf{0} \\ \mathbf{Q^x} \end{bmatrix} + (M_{22} - 1)\mathbf{U}(s) \begin{bmatrix} \mathbf{0} \\ \mathbf{Q^y} \end{bmatrix} \quad (14)$$

On the other hand, the difference between $\mathbf{d}'(s)$ and $\mathbf{d}(s)$ is, from equation (11),

$$\mathbf{d}'(s) - \mathbf{d}(s) = \mathbf{U}(s)(\mathbf{Q}' - \mathbf{Q}) \,. \qquad (15)$$

Hence, comparing this result with expression (14), one can conclude that the difference in control points $\mathbf{Q}' - \mathbf{Q}$ may be written as a linear combination of six vectors. Therefore, using matrix notation,

$$\mathbf{Q}' - \mathbf{Q} = \mathbf{W}\mathbf{S} \,, \qquad (16)$$

where

$$\mathbf{W} = \left( \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} \mathbf{Q^x} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{Q^y} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{Q^x} \end{bmatrix}, \begin{bmatrix} \mathbf{Q^y} \\ \mathbf{0} \end{bmatrix} \right), \qquad (17)$$

and $\mathbf{S}$ is a vector with the six parameters of the linear combination, the shape vector

$$\mathbf{S} = [t_x, t_y, M_{1,1} - 1, M_{2,2} - 1, M_{2,1}, M_{1,2}]^\top \,, \qquad (18)$$

which encodes the relation between different affine views of the planar contour.

Figure 3 (reproduced from [15]) shows the graphical effect of each parameter: two translations and four scalings. Note that the dimension of the shape vector can be reduced if robot motion is constrained, for instance to lie on a plane [16,17].

Once the compact representation of the contour in terms of control points and knots is obtained, a Kalman filter is used to track the contour along the sequence [13], and the shape vector is updated at each frame.

In previous works [18,19], the continuously updated shape vector was used to estimate robot egomotion in practice, provided data from other sensors (such as an inclinometer) or scene information (such as depth) were supplied. Here we focus on the extraction of epipolar direction from the shape vectors of just two views, and the analysis of the attainable accuracy in the different possible working conditions.
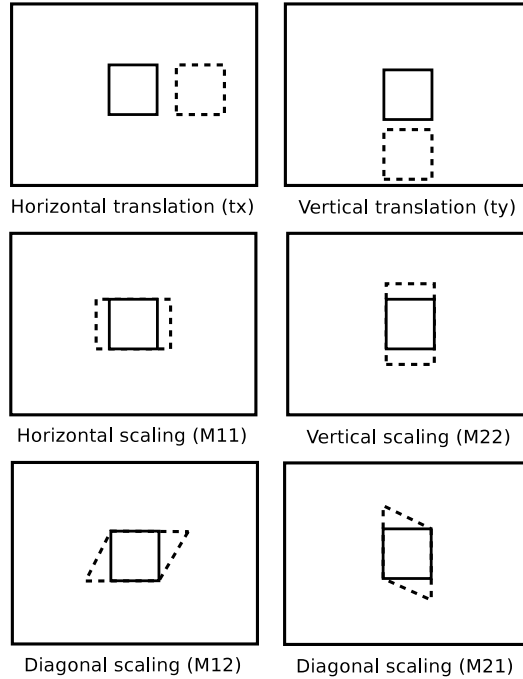
Figure 3. Affine deformations in the image plane induced by 3D camera motion.

## 4 Experimentation

Two sets of experiments were performed to evaluate the accuracy of the proposed method. The first set uses synthetic image sequences generated by simulating camera motion and computing projections under a full perspective camera model. Using this set, the sensitivity of the proposed technique is assessed: first, with respect to the characteristics of the contour shape; second, by relaxing the different weak-perspective conditions one at a time; third, when the motion has some amount of cyclorotation; and fourth, by adding variable amounts of pixel noise. Note that this would be hard to do using real images.

The affine epipolar geometry is usually estimated using the Gold Standard algorithm [20]. This technique requires image correspondences of at least 4 non-coplanar points. Using also our synthetic experimental testbed, we show the effects of approaching coplanarity for this configuration, and compare the results with those of our method, which computes the affine epipolar direction only from planar contour matches.

The second set of experiments uses real images taken by a robot arm moving along a calibrated path, showing the performance of the approach under realistic imaging conditions. Finally, when using a frontoparallel centered target, the matrix $\mathbf{M}$ in (9) becomes symmetric. We also show the consequences of exploiting this fact to reduce the dimension of the shape vector (18) from 6

11

(a) 7 control points

(b) 7 control points and 1 double knot
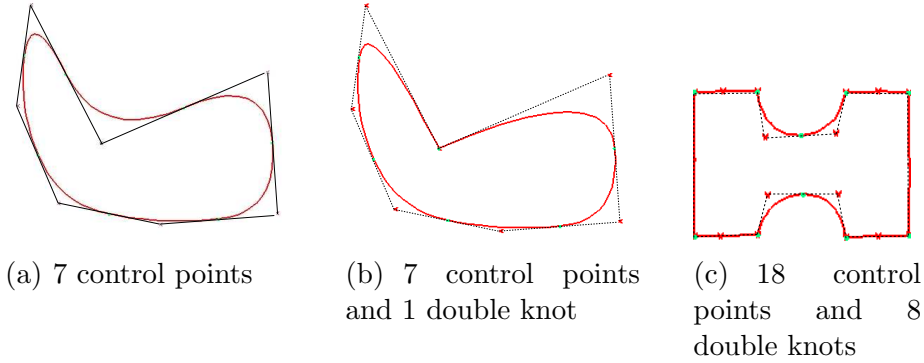
(c) 18 control points and 8 double knots

Figure 4. Representing contours using B-splines: (a) and (b) show that different contours can be modelled with the same control polygon but different knot multiplicities, while (c) displays the modelling of the H-shaped contour used in the experiments.

to 5 when parameterizing contour deformations.

### 4.1  Preliminary considerations

#### 4.1.1  Contour representation

As mentioned in Section 3, the contour $\mathbf{d}(s)$ in (11) is represented by using B-splines. Briefly, the sequence of control points, together with the multiplicity of the knots joining the different polynomial curve segments, are used to represent the contour [13]. The knot multiplicity determines the smoothness of the curve at that point. In our representation with quadratic B-splines, double multiplicity is used to encode corners. As can be seen in Figures 4(a) and (b), by just doubling the multiplicity of one knot, the modelled contour changes, even if the sequence of control points remains the same. Thus, both the set of control points and the knot multiplicities should be used to obtain the shape vector $\mathbf{S}$ from equation (16).

This is how it is done when we use the tracker in the experiments using real images. However, in the experiments using synthetic images, we compute the pseudoinverse of the shape matrix with only the control points, disregarding knot multiplicity. This simplification is valid because the H-shaped contour used in the experiments (Fig. 4(c)) contains many double knots and, thus, its control polygon fits well the actual contour. The same is true for the square shape also used in the synthetic experiments.
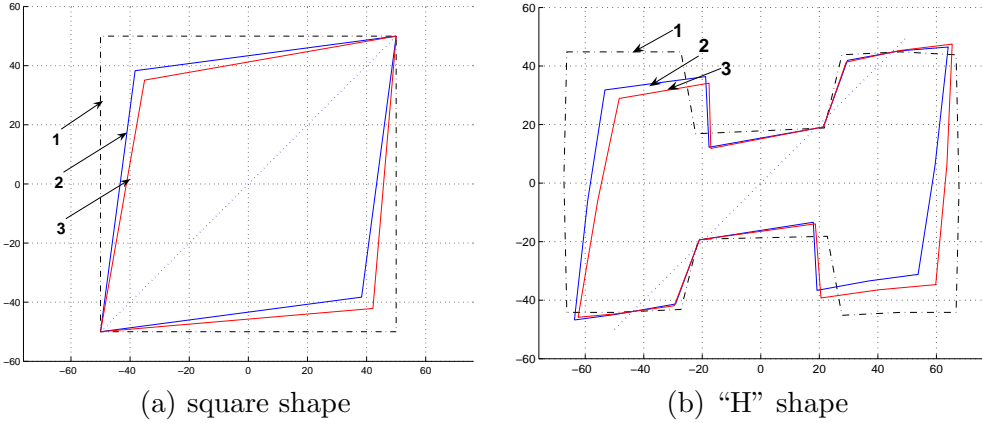
(a) square shape          (b) "H" shape

Figure 5. Difference in the projection of two control polygons for the two camera models discussed, after rotating the camera 40° about a 45° axis on the target, passing through the target center. 1) original contour, 2) weak-perspective projection, and 3) full-perspective projection.

### 4.1.2   Camera model used to generate synthetic images

When synthetic images are generated using an affine camera model (i.e., assuming perfect weak-perspective conditions), the epipolar direction is exactly recovered with the proposed method. However, we would like to assess the validity of the method under more general conditions. To this end, we generate the test set of synthetic images using a full perspective camera model (refer to Fig. 1). Then, of course, perspectivity effects affect the recovery of the epipolar direction in the ways that will be analysed in the following sections.

An illustration of the differences in projected contour control polygons for both camera models can be seen in Figure 5 for two different targets: a simple square (Fig. 5(a)) and an "H" form (Fig. 5(b)). In the plot, 1) represents the original template, 2) corresponds to the weak-perspective projection and 3) indicates the full perspective projection. These plots were obtained after a camera rotation of 40° about an axis centered on the target centroid and having an inclination of 45°, as drawn in the figure. The target is located at a distance of $500mm$ from the camera plane, and the camera is simulated with a focal distance value of $5mm$. Note how this configuration has purposely been chosen away from weak-perspective assumptions, in order to observe perspective effect differences in the projection of the two models.

### 4.2   Effects of the contour shape

To determine the influence of the contour shape on the recovery of the epipolar direction, we use the square and H-shaped contours in Fig. 5. Synthetic images of the two contours are generated for any positioning of the camera on the
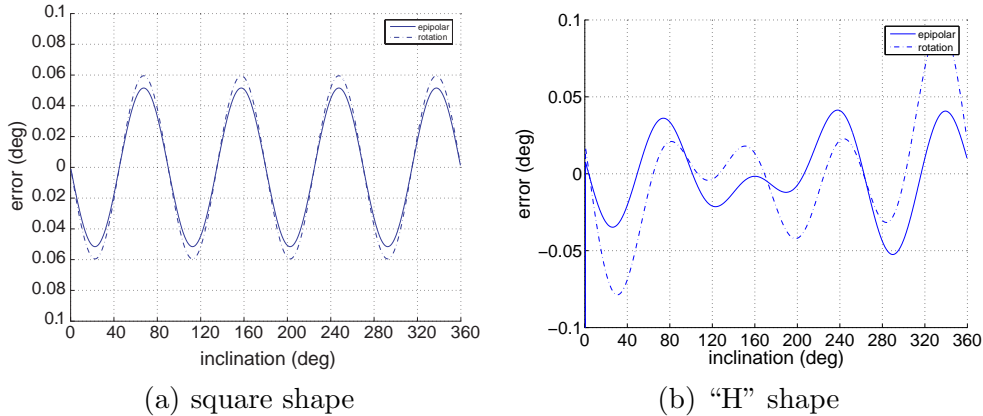
13

(a) square shape          (b) "H" shape

Figure 6. Error in the recovery after a camera rotation of 40° about an axis passing through the target centroid and whose orientation varies from 0° to 360°. The continuous line shows the error in the recovery of the epipolar direction, while the dotted one indicates the error in the recovery of the orientation of the rotation axis.

surface of a sphere centered on the target centroid and of radius $Z_{\text{ave}}$. This is equivalent to a pure rotation of the camera about an axis on the target, passing through the target centroid. As mentioned in the preceding section, projections are performed under a full-perspective camera model.

Figure 6(b) shows the evolution of the recovery error for a rotation of 40° about an axis whose orientation varies from 0° to 360°. $Z_{\text{ave}}$ is fixed to $1500mm$. Observe that the error remains very low throughout: less than 0.1°. For the square shape, the error shows a periodicity of 90°, and it is zero for the specific rotations at $0°, 45°, \ldots, 360°$. In upcoming sections we will see that this effect is preserved even after weak-perspective imaging assumptions are relaxed. However, for the more complex H shape, the epipolar recovery error is not equally distributed in the four quadrants, and the rotations for which the error is zero do not have the same periodicity as for the simple square model.

From this and similar experiments, we conclude that the shape of the contour being tracked has an influence on the distribution of the error for different camera motions, but not much on the average recovery error. This is presumably due to differences in shape symmetries and in the number of points conforming the control polygon.

### 4.3 Relaxation of weak-perspective conditions

#### 4.3.1 Camera translation along the optical center: distance to target

We analyse how a decrement of the distance from the camera to the target $Z_{\text{ave}}$ affects the computation of the epipolar direction. Decreasing the distance enlarges perspective effects, and consequently, should increase the error
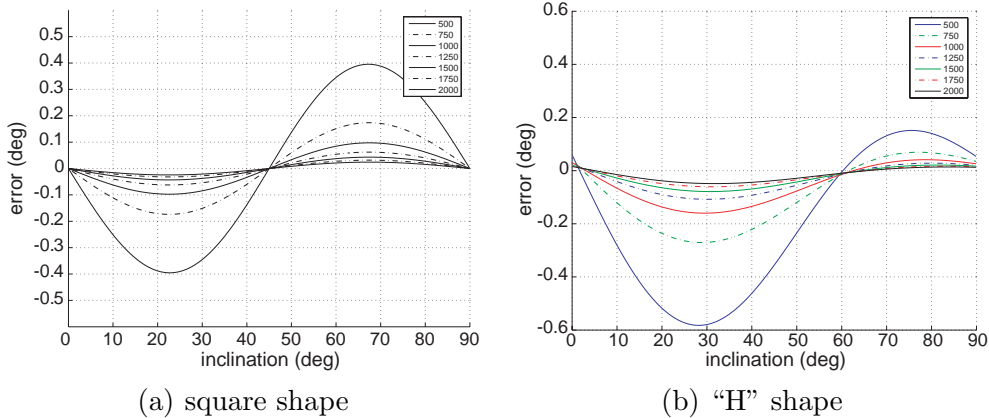
|   |   |
|:-:|:-:|
| (a) square shape | (b) "H" shape |

Figure 7. Error in the recovery of the epipolar direction when the distance from the camera to the target varies from 500 to 2000$mm$. The plot shows results for a camera rotation of 40° about an axis on the target with inclinations ranging from 0° to 90°.

in epipolar direction recovery. For this experiment we consider distances of 500, 750, 1000, 1250, 1500, 1750 and 2000$mm$. The smallest of these, 500$mm$, corresponds to an extreme situation for the weak-perspective model, in which important unmodelled distortions in the projected control polygon are present (refer to Fig. 5(b)). For larger depth values, the affine conditions are better satisfied, thus reducing the error, as shown in Figure 7. It is worth noting that even under these unfavourable conditions the recovery error stays below 0.6°.

### 4.3.2 Camera rotation: depth relief of the target

The weak-perspective model assumes that the depth relief of the target is very small compared to the distance $Z_{ave}$ from the camera to the target centroid. In our case in which the target is a planar contour, this means that the distance from the contour in 3D to its projection onto a plane centered on the target and parallel to the image plane is very small compared to $Z_{ave}$.

Starting from a fronto-parallel view of the contour, the amount of depth relief is directly related to the amount of camera rotation about an axis placed on the contour. Thus, we perform an experiment to analyse the effect of depth relief by fixing a rotation axis on the target at various inclinations (7.5°, 15°, ..., 82.5°) and then rotating the camera about each of these axes from 5° to 50°, with same focal length and $Z_{ave}$ values as before.

Figure 8 shows the results. In good accordance with Figure 6 for the square shape, the epipolar direction recovery error is zero for any positioning of the camera resulting from a rotation about an axis having an inclination of 45°. For the "H" shape, the zero error occurs at different inclination values for different amounts of rotation. Note that, for both target shapes, larger rota-

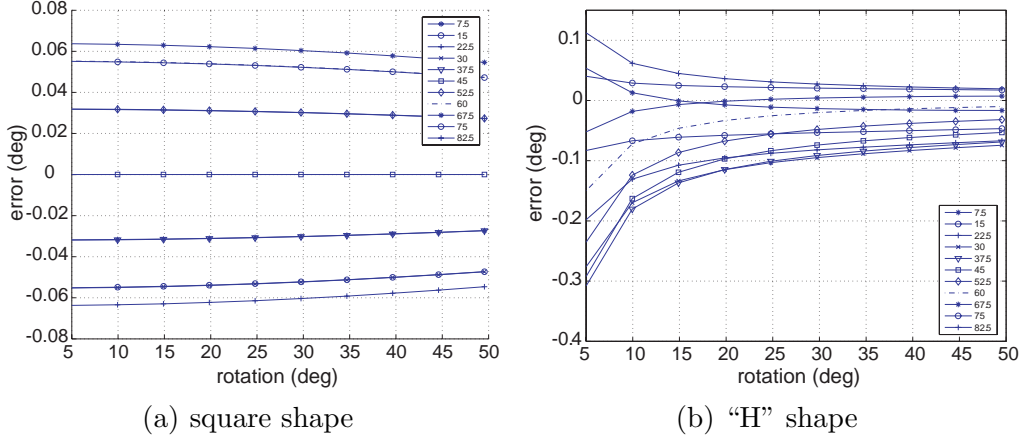|     |     |
| --- | --- |
| (a) square shape | (b) "H" shape |

Figure 8. Error in the recovery of the epipolar direction for different rotation axis inclinations when the rotations through these axis range from 5° to 50°.
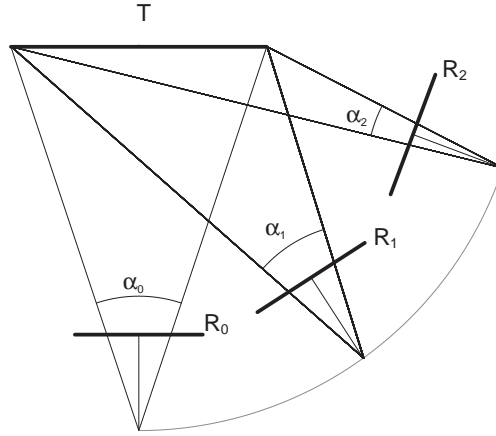


Figure 9. Starting with a frontoparallel target $T$, a first view $\mathcal{R}_0$ is taken and then two other views $\mathcal{R}_1$ and $\mathcal{R}_2$ are taken; $\mathcal{R}_2$ after a larger amount of rotation than $\mathcal{R}_1$. The maximum angle $\alpha_i$ between the rays of the target projection decreases (and, thus, the projection becomes closer to a parallel projection) as the amount of rotation increases.

tions lead to smaller recovery errors. A geometric explanation of this effect is that, starting from a fronto-parallel target position, the larger the amount of rotation, the closer the target projection is to the parallel rays characteristic of weak-perspective projection (Fig. 9). Therefore, it follows that unmodelled perspective effects produce larger errors for small rotation amounts.

In sum, the error in the recovery of the epipolar direction due to depth relief is negligible.

### 4.3.3 Lateral camera translation: Uncentered contour

The weak-perspective camera model assumes that the distance of the contour control points to the principal ray are small, again in relation to the distance
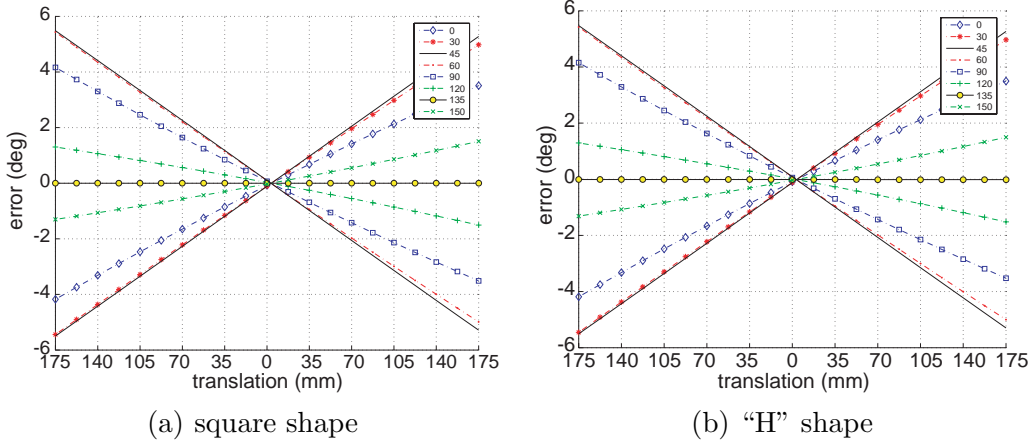
| (a) square shape | (b) "H" shape |

Figure 10. Error in the recovery of the epipolar direction for a fixed rotation of 40° about an axis inclined 45° on the target plane and an initial distance of 5*m*, when lateral translations of the camera along the several directions listed in the insert are performed.

$Z_{ave}$ from the camera to the target. This condition can be satisfied with a small field of view, and keeping the projected target centered in the image (supposing that the image center is very close to the principal point). In order to evaluate the effects of relaxing the condition that the target is centered, we simulate a series of camera translations parallel to the target plane. We show results for a series of translations at orientations $0°, 30°, \ldots, 150°$ on the image plane, and the special cases at $45°$ and $135°$. These translations range from $-175mm$ to $175mm$, spanning the entire simulated image area.

Compared to the previously evaluated conditions, camera translations that result in target shifts in the image lead to the largest errors in the computation of the epipolar direction. As can be seen in Figure 10, contour shape effects are negligible compared to the error induced by lateral camera translation. Moreover, epipolar recovery errors are more significant for translations along directions orthogonal to the epipolar direction $(135°)$, and negligible along the epipolar direction. See Appendix B for a geometrical explanation of this effect, which corresponds to a particular case of Theorem 4 proved therein.

## 4.4  Effects of cyclorotation

Our algorithm assumes that there is no cyclorotation in the camera motion. As can be easily seen in Fig. 2, performing a camera motion that contains cyclorotation wouldn't preserve any image direction. We like to quantify the error in the recovery of the epipolar direction as the amount of cyclorotation increases.

17

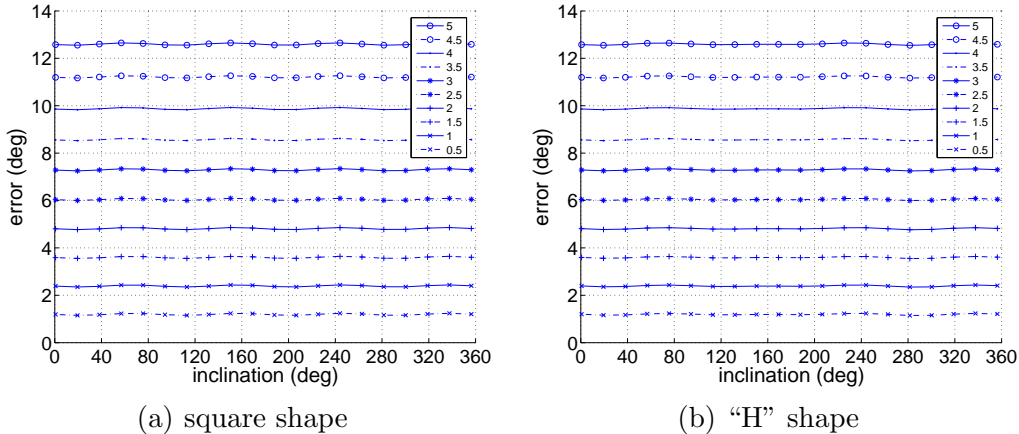|  |  |
|:---:|:---:|
| (a) square shape | (b) "H" shape |

Figure 11. Effects of introducing cyclorotation in the camera motion. Each line is the recovery error after a camera rotation about an axis passing through the target, and whose orientation varies from $0°$ to $360°$ in the plane, and from $0.5°$ to $5°$ out of the plane (increasing the amount of cyclorotation).

Figure 11 shows the large errors resulting from the same experiment of Section 4.2 but introducing different degrees of cyclorotation. For each experiment, the rotation is $40°$ about an axis passing through the target center whose orientation varies from $0°$ to $360°$, and whose cyclorotation component ranges from $0.5°$ to $5°$. As expected, the error in the recovered epipolar direction severely increases with respect to the amount of cyclorotation.

## 4.5 Sensitivity to pixel noise

The affinity $\mathbf{M}$ relating two views nears the identity for small motions, and thus the computation of the epipolar direction based on $\mathbf{M}$ is very sensitive to perturbations in the projection of the control points. To evaluate this sensitivity, we consider rotations ranging from $5°$ to $50°$ about an axis oriented at $-45°$ on the target plane, for which the recovery error neared zero under noise-free conditions (Fig. 8). Gaussian noises with standard deviations equal to multiples of 0.25 *pixels* are added to both the $x$ and $y$ components of the image projections of the control points. A Montecarlo simulation is performed for 10000 samples, and the results are plotted in Figure 12. For each shape, there are 10 plots corresponding to the rotation angles from $5°$ to $50°$ and, within each plot, the mean epipolar direction and its standard deviation are represented as a vertical segment for each noise level from 0.25 to 1.00 *pixels*.

Note that, in accordance with the effect illustrated in Figure 9, larger rotations lead to smaller recovery errors. Moreover, as could be expected, higher amounts of noise yield a poorer recovery, although it is worth mentioning that for rotations larger than say $20°$ the recovery is quite good even in the presence
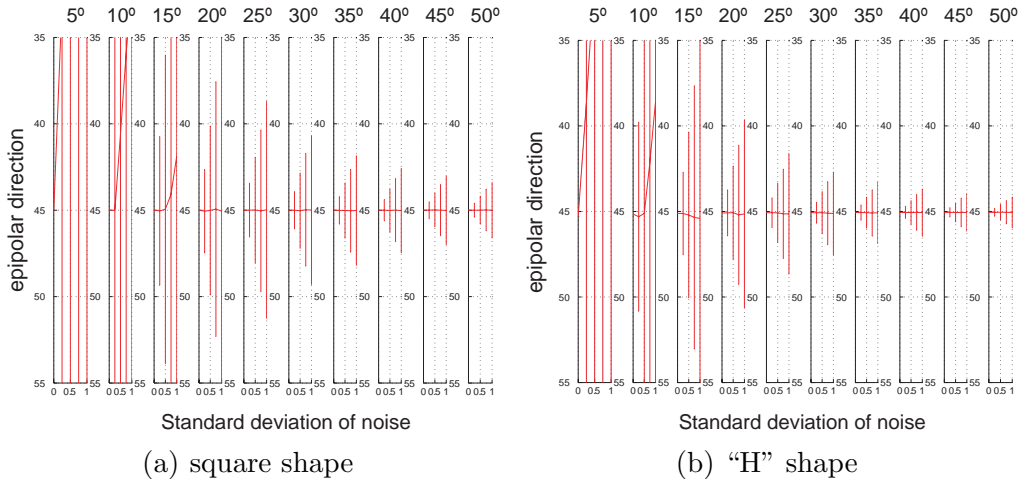
18

(a) square shape     (b) "H" shape

Figure 12. Mean (horizontally oriented line) and standard deviation (vertical line) of epipolar directions computed in the presence of Gaussian noise with standard deviations from 0.25 to 1.00 *pixels*. Ten graphs corresponding to rotations ranging from 5° to 50° about an axis oriented at −45° are displayed.

of considerable noise. Conversely, for rotations smaller than 10°, our technique shows not to be reliable under noisy conditions. Observe that the results for the "H" shape are much better, especially for small rotation angles, than those for the square, due to the fact that a larger number of control points results in a better estimation of the affinity.

## 4.6 Comparison with the Gold Standard algorithm

To this point, we have analysed the effects of several factors on the accuracy of our proposed procedure. Now, and still in a simulated setting, we would like to carry out a comparison with another technique for computing the affine epipolar geometry, namely the Gold Standard (GS) algorithm [20]. This algorithm, contrary to our procedure, needs non-coplanar point correspondences in order to compute the maximum likelihood estimate of the affine fundamental matrix. While in theory, only four non-coplanar points would suffice for computing the affine epipolar geometry using the GS algorithm; its performance is affected by the amount of non-coplanar information provided, both in terms of depth range and in the number of points used. The idea is to establish experimentally the amount of depth information required by the GS for this algorithm and our procedure to provide equivalent epipolar direction recovery results.

To this end, we set first an experiment in which we add a range from two to twelve extra points to the H-shaped contour, varying their distance with respect to the contour plane. Camera parameters are fixed at: 500*mm* dis-
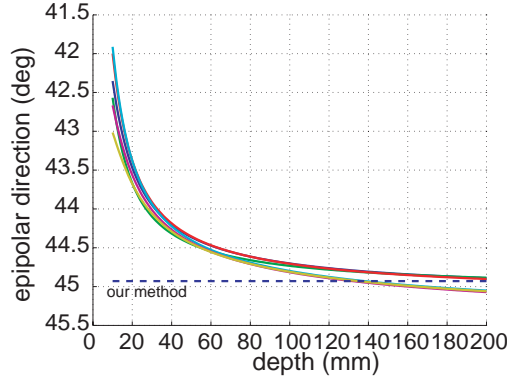
Figure 13. Epipolar direction computed with the GS algorithm in the case of 2,4,...,12 out-of-plane points (a curve for each number) placed at increasing depths (in abcissae) above the H-shaped contour. Notice that the relevant factor is not the number of points, but the depth at which they are placed.

tance to target and a focal distance of $767 pixels$. As before, camera motion is achieved via a rotation of $40°$ about an axis placed at an orientation of $45°$ on the target plane. The results are shown in Figure 13. It can be seen how as the depth of these points is increased, the error in the computation of the epipolar direction decreases. Furthermore, it turns out that the number and $xy$ location of these points have little effect in the computation of the epipolar direction. The figure contains plots of the resulting errors in the computation of the affine epipolar direction with the GS algorithm for different numbers of out-of-plane points, and a cut threshold indicating the error in the recovery of the epipolar direction using our proposed technique under the same experimental conditions (the additional points out of the contour plane are evidently not used in this case). As shown in the figure, for the given experimental conditions, the results of our technique are comparable to the results of the Gold Standard algorithm when the extra points are placed roughly at a distance equal to the target size ($120mm$ in our case).

Note the importance of parallax in the computation of the affine fundamental matrix in the Gold Standard algorithm. As the target points approach coplanarity, the parallax vector, which determines the epipolar direction, is monotonically reduced in length. Consequently, the accuracy of the line direction is correspondingly reduced, and the covariance of an estimated affine fundamental matrix increases. This situation does not occur in our procedure, as it has been devised precisely to compute the affine epipolar direction from two views of a plane.

Our second experiment evaluates the noise sensitivity of both techniques for computing the affine epipolar direction, and consists in repeating the same experiment multiple times adding Gaussian noise to the control polygon projections. The experiment is performed for our approach using the planar "H" shape, and for the GS algorithm using: a) the planar "H" contour and two

20

| case | | noise (pixels) | | | | |
|---|---|---|---|---|---|---|
| | | 0.00 | 0.25 | 0.50 | 0.75 | 1.00 |
| planar contour "H" | $\mu$ | -44.97 | -44.95 | -44.96 | -45.04 | -45.07 |
| with our technique | $\sigma$ | 0 | 0.193 | 0.492 | 0.552 | 0.876 |
| planar contour "H" and 2 | $\mu$ | -43.76 | -43.77 | -43.78 | -43.74 | -43.85 |
| points 120$mm$ off the plane (GS) | $\sigma$ | 0 | 0.143 | 0.245 | 0.436 | 0.486 |
| 2 layers of the contour | $\mu$ | -44.91 | -44.91 | -44.89 | -44.92 | -44.91 |
| "H" at 0$mm$ and 120$mm$ (GS) | $\sigma$ | 0 | 0.068 | 0.140 | 0.210 | 0.324 |
| Randomized depth values | $\mu$ | -44.81 | -45.75 | -45.03 | -44.51 | -44.84 |
| for the control polygon points (GS) | $\sigma$ | 0 | 0.294 | 0.486 | 0.712 | 0.625 |

Table 1

Epipolar direction recovery using our procedure and the Gold Standard algorithm, for increasing levels of Gaussian pixel noise. Ground truth is a camera rotation of 40° about an axis at 45° on the target.

off-the-plane control points at a distance of 120$mm$, b) 24 control points divided in two layers resembling two "H" shapes at different distances, and c) a randomized depth assortment of the 12 "H" control points. Table 1 shows a comparison of the results obtained for an increasing standard deviation of 0 to 1 *pixels* in image noise.

Observe that our procedure and GS using two layers of control points provide the most consistent results, while GS with only two out-of-plane points yields a somewhat downward-biased estimation. Overall, though, no big differences in accuracy are observed.

### 4.7   Experiments using real images

We present now results on image sequences in a controlled setting of our technique for computing the affine epipolar direction from pairs of views of a plane only. The goal of this work is not tracking, but computing the affinity and using it to estimate the epipolar direction induced by the two views. To this end, we facilitate the tracking phase by moving a simple target placed on a manipulator end-effector, and focus on evaluating the accuracy of the direction recovered in different situations, compared to robot motion ground truth.

The experimentation setup consists of a Stäubli RX60 manipulator holding the target pattern on its end-effector. This target is a planar artificial H-shaped figure with corners and curved edges, which can be easily tracked with our

(a) Initial            (b) 15°            (c) 30°

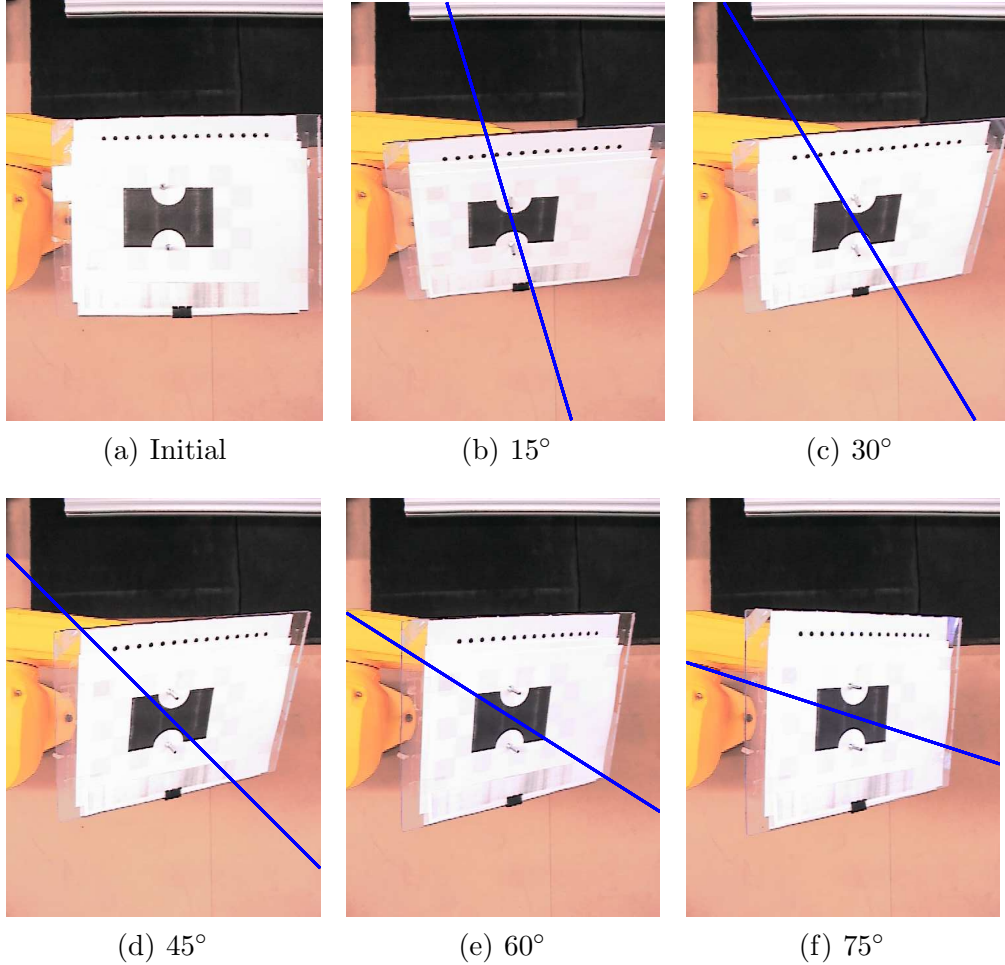(d) 45°            (e) 60°            (f) 75°

Figure 14. The first experiment with real images entails pairs of views consisting of the initial one plus each of the other five, corresponding to camera rotations of 40° about an axis on the target with inclinations sampled at intervals of 15°. The computed epipolar direction is displayed as a line passing through the target center.

active contour tracker. Images are acquired using a Sony DWL500 firewire camera with a calibrated focal length of $767 pixels$. The working zone of the robot arm is highly limited, and this restricts the repertoire of experiments that we can perform with it. So the initial distance from camera to target has had to be set to $500mm$. This corresponds to the extreme case discussed in Section 4.3.1, Fig. 7, and, therefore, we are testing the proposed approach under relaxed weak-perspective conditions, as studied in simulation in Section 4.3. The acquired images have evident perspective effects, as shown in Figures 14 and 15.

The first experiment entails camera motion induced by a rotation of 40° about an axis on the target at various inclination angles sampled at intervals of 15°. This, thus, relates to Fig. 7(b) with distance equal to $500mm$. Starting from the fronto-parallel position shown in Figure 14(a), the contour is tracked to each of the final views shown in the remaining frames of the figure. The epipo-

22

| epipolar direction | -15 | -30 | -45 | -60 | -75 |
|---|---|---|---|---|---|
| $\bar{\theta}$ | -16.63 | -31.01 | -45.00 | -57.63 | -72.04 |
| $\sigma$ | 0.14 | 0.09 | 0.14 | 0.19 | 0.13 |

Table 2
Mean and standard deviation in degrees of the epipolar direction computed by the proposed technique from real images.

lar direction computed by the proposed algorithm in each case is displayed as a line passing through the target center.

Table 2 presents the numerical values obtained in the computation of the epipolar direction. Standard deviation is computed by acquiring 300 images in the final position, estimating the shape vectors and then computing the corresponding epipolar directions. Note that the standard deviations are all very similar, and the mean values deviate more from the ground truth as the angle departs from the 45° inclination. This should be interpreted in the light of Fig. 7 as meaning that the tracker amplifies the recovery error due to perspectivity effects unmodelled by the weak-perspective camera. Consequently, under true weak-perspective conditions, the errors should be much lower as indicated by the shrinking of the error curves in Fig. 7 when the distance $Z_{ave}$ from the camera to the target increases.

Three additional sequences were analyzed after further relaxing weak-perspective conditions, in the same way as in the synthetic case. The first such sequence, labelled "Not centered", starts at the fronto-parallel initial position (Fig. 15(a)) and finishes at an uncentered position, after a translation of $100mm$ along the $x$ axis of the robot coordinate frame and a rotation of 40° about an axis at 45° inclination (Fig. 15(b)). Consistent with our simulated results (refer to Section 4.3.3), this lateral camera translation is by far the violation of weak-perspective conditions that has the most pervasive effect on the computation of the epipolar direction. See the numbers in Table 3, first row.

The second experiment, labelled "Not Frontoparallel A", corresponds to the same rotation described above, but the initial frame is not frontoparallel. The sequence starts with the target already rotated 20° as shown in Fig. 15(c) and, after a further rotation of 20°, finishes at 40° (Fig. 15(d)), all rotations about an axis at 45° inclination as before. Observe that the result is a bit worse than that of the initial experiment, but with a similar standard deviation.

Finally, the last experiment, labelled "Not Frontoparallel B", corresponds to a sequence from a non-frontoparallel initialization of the contour at 40° of rotation (Fig. 15(d)) going back to a frontoparallel target position (Fig. 15(a)). The result is very similar to that of the preceding experiment.

23

(a) Initial

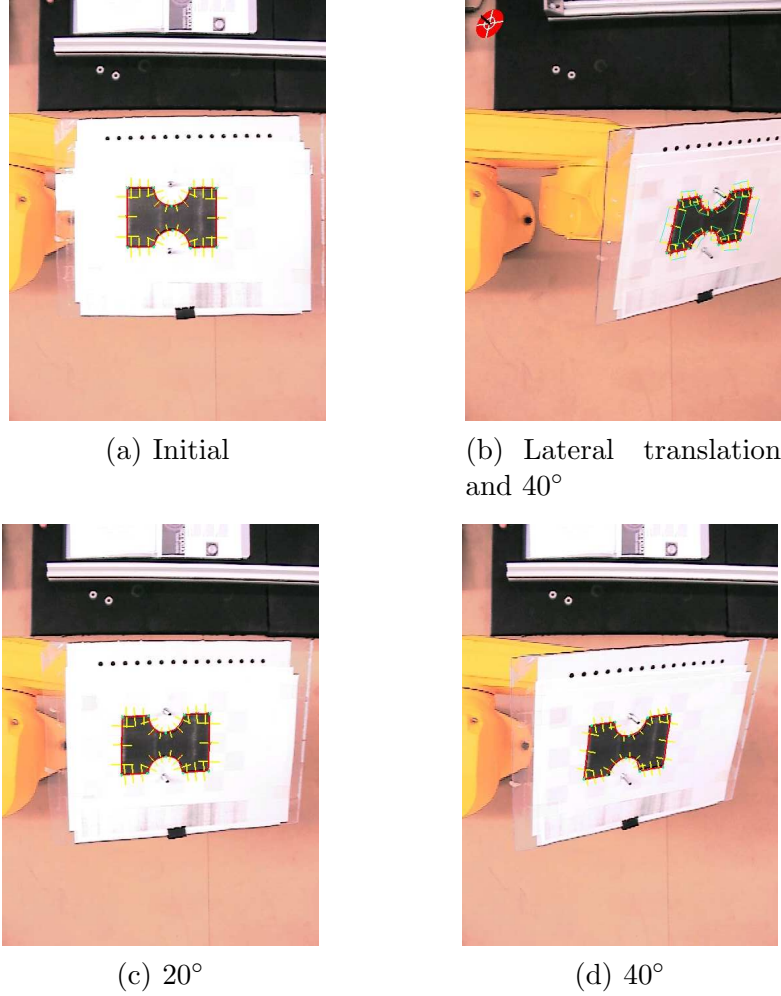(b) Lateral translation and 40°

(c) 20°

(d) 40°

Figure 15. Experiments with real images further relaxing weak-perspective conditions. The first sequence, entailing an uncentered target, starts at (a) and ends at (b). The next sequence departing from a non-frontoparallel target position starts at (c) and ends at (d). The last sequence, testing also non-frontoparallel initialization, starts at (d) and ends at (a).

| Frames | $\theta$ | $\sigma$ |
|---|---|---|
| Not Centered | -34.65 | 0.13 |
| Not Frontoparallel A | -43.89 | 0.09 |
| Not Frontoparallel B | 43.96 | 0.10 |

Table 3
Mean and standard deviation of the epipolar direction computed over real images when weak-perspective conditions are further relaxed.

*4.8 A remark: interest of reducing shape space*

Using a frontoparallel centered target, the matrix $\mathbf{M}$ in (12) turns out to be symmetrical. We can exploit this fact to reduce from 6 to 5 the dimension of
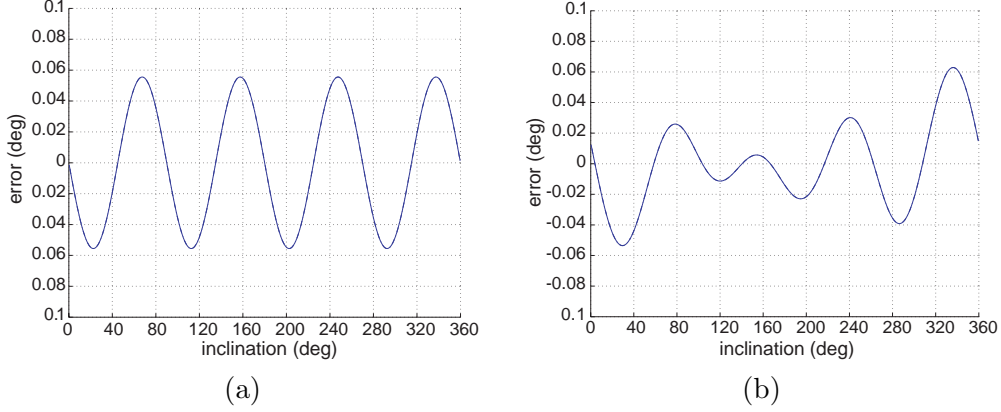
Figure 16. The same plots as in Fig. 6 but using a 5D shape vector instead of a 6D one.

the shape vector (18) used to parameterize contour deformations to

$$
\mathbf{W} = \left( \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^{\mathbf{x}} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}^{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^{\mathbf{y}} \\ \mathbf{Q}^{\mathbf{x}} \end{bmatrix} \right). \tag{19}
$$

Reproducing the experiments of Section 4.2 with this reduced shape matrix results in exactly perpendicular solutions for the epipolar direction and the rotation axis. This is because the vectorial part of the weak-perspective deformation is forced to be symmetric even with the perspective effects present. As a result, the computed errors are the same for both. Figures 16(a) and 16(b) show the values for both the epipolar direction and the rotation axis superimposed one on top of the other. Note that the angles at which the error becomes 0° are preserved, as the symmetry axes in the contour have not changed. This may not be of great benefit when weak-perspective image conditions are satisfied, but when perspective effects are strong, reducing the shape space greatly diminishes the distorsions that these effects cause in the recovery of the epipolar direction.

As mentioned before, we have exploited this advantage for improving egomotion recovery in some practical settings [16,17].

### 4.9  Applicability considerations

Throughout this experimental section we have made an effort to gradually relax the imaging conditions away from the weak-perspective model, in order to assess the validity of the proposed method even under unfavourable conditions. We have seen in Section 4.3.1 that the error introduced by small translations along the optical axis is negligible, and in Section 4.3.2 that the

depth relief assumption is not really required in normal conditions. Contrarily, in Section 4.3.3 we have shown that lateral camera translations that shift the contour away from the center of the image introduce critical errors. In a realistic application one needs to ensure that the target projects always to the center of the image. To do so, an active vision system that keeps the tracked contour always centered in the image, regardless of camera motion, should be used.

Moreover, we have seen in Section 4.5 that the epipolar direction cannot be accurately computed for small rotation values in the presence of pixel noise. In the implementation with our robotic arm and real image sequences, accurate estimation of the epipolar direction could be obtained once a rotation of 15° was exerted, or, equivalently, once the camera was translated laterally roughly one fifth of its distance to the target whilst keeping the target centered.

Finally, if we can assume that the target is frontoparallel in some frame of the motion sequence, one should consider the use of the reduced shape space equations defined in Section 4.8.

## 5 Conclusions

The recovery of camera motion and scene structure from uncalibrated image sequences has received a lot of attention lately due to its numerous applications, which range from robot localization and navigation, to virtual reality and archeology, to name just a few. Most works rely on detecting a set of non-coplanar points in the scene and matching their projections on the different views. Good results are achieved only when scenes are textured enough to allow visual motion estimation from point correspondences. In this paper we have departed from this main stream, by dealing with a less informative situation, namely features lying on a plane, and recurring to contour tracking instead of point matching.

Our main result is that, under weak-perspective conditions and assuming a camera motion free of cyclorotation, the epipolar direction can be recovered from the affinity relating two views of a planar scene. Two proofs of this result are provided, one within the framework of projective geometry, and the other through an analytic development.

Two sets of experiments have been undertaken to study the performance of the proposed method. In both, the affinity relating two views is derived by tracking a planar contour from one view to the other, and then extracting the affinity parameters from the observed contour deformation, coded as a 6D (or 5D) shape vector. First, synthetic images were used to evaluate the results in

a noise-controlled environment by relaxing one by one the conditions of the imaging model, and then to compare the accuracy of our method with that of the Gold Standard algorithm, which relying on matches of non-coplanar points falls in the main stream mentioned above.

The outcome of the comparison has been very encouraging, since with less scene information (only from a plane) and with a much simpler processing (solving a single second-order equation), we are able to obtain the epipolar direction with similar accuracy. It is worth reminding, however, that our method is less general in that it requires a camera motion free of cyclorotation.

The second experimental set consisted of image sequences that were used to validate the proposed approach under real imaging conditions. Note that the objective of the paper is to show what can be obtained from the affine deformation of two views of a contour, and not to validate the robustness of the contour tracker used. For this reason, simple and well-calibrated image sequences were used in order to have a good basis for ground truth comparison.

Future work will try to unravel under what circumstances additional information on camera motion and scene structure can be recovered from two (or more) uncalibrated views of a planar object under weak perspective viewing conditions. We have already mentioned that, if the aspect ratio and skew (i.e., the intrinsic parameters determining the metric structure in the image) are known, then the direction of the axis of rotation can also be recovered. Along the same line, we will tackle the recovery of the orientation of the scene plane, as well as what occurs in degenerate situations in which such orientation is the same as that of the image plane, or when both planes have a common direction.

## A    Geometric study of two weak-perspective views

### A.1    The camera model

Consider a weak perspective (or scaled orthographic) camera $\psi$, which projects a scene point first orthographically onto the average depth plane $\mathcal{R}^C$ (the plane parallel to the image plane $\mathcal{R}$ containing the centroid $C$ of the scene object) and then perspectively from this fronto-parallel plane $\mathcal{R}^C$ onto the image $\mathcal{R}$. Let $P$ be the proper optical center of the second perspective projection, which simply introduces a scale factor. Modulus this scale factor, the weak-perspective camera is an orthographic camera: all the projection rays are parallel with direction orthogonal to the image plane $\mathcal{R}$; consequently, if $O$ is the improper optical center (lying on the plane at infinity $\Pi_\infty$), then the
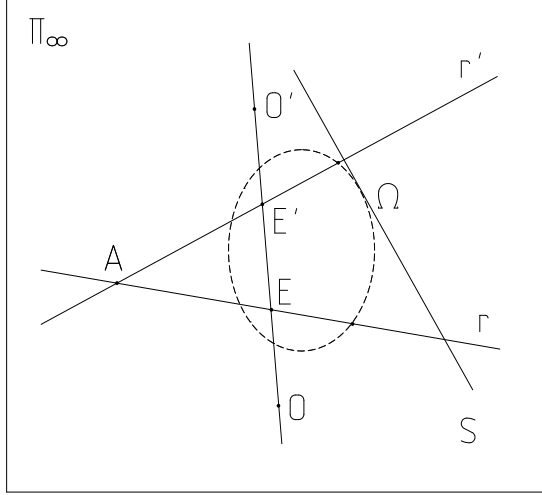
Figure A.1. Geometry on the plane at infinity of two weak-perspective cameras. The notations are explained in the text.

improper line $r$ of $\mathcal{R}$ or $\mathcal{R}^C$, that is, $r = \mathcal{R} \cap \Pi_\infty = \mathcal{R}^C \cap \Pi_\infty$, is the polar of $O$ with respect to the absolute conic $\Omega$, lying on $\Pi_\infty$.

Consider a world point $X$ projected by the weak-perspective camera $\psi$ to the image point $x = \psi(X)$, namely

$$\psi : \mathbb{P}^3 \xrightarrow{\vee O} O^* \xrightarrow{\cap \mathcal{R}^C} \mathcal{R}^C \xrightarrow{\vee P} P^* \xrightarrow{\cap \mathcal{R}} \mathcal{R} \,, \tag{A.1}$$

and $x = (((X \vee O) \cap \mathcal{R}^C) \vee P) \cap \mathcal{R}$, where $O^*$ denotes the rays through $O$, $\vee O$ denotes the projection with center $O$, and $\cap \mathcal{R}$ denotes the section with the plane $\mathcal{R}$. From (A.1) it is inferred that the weak-perspective camera $\psi$ (as well as the affine camera) has the property of mapping world directions to image directions: an improper world point $X \in \Pi_\infty$ is projected to an improper image point $x = \psi(X) \in \Pi_\infty$ given by $x = (X \vee O) \cap r$. Thus, a world plane $\mathcal{S}$ is mapped by the affinity $\psi_{|\mathcal{S}}$ to the image plane, i.e. the camera preserves parallelism.

Consider a second weak-perspective camera $\psi'$, whose image plane, improper and proper optical centers are $\mathcal{R}'$, $O'$ and $P'$, respectively. Let $\mathcal{S}$ be a world plane of improper line $s = \mathcal{S} \cap \Pi_\infty$. The restriction to a coplanar scene induces a projectivity $\psi'_{|\mathcal{S}} \circ (\psi_{|\mathcal{S}})^{-1}$ between the two weak-perspective images. Furthermore, it is an affinity, since image directions are mapped to image directions: $\psi'_{|\mathcal{S}} \circ (\psi_{|\mathcal{S}})^{-1}(r) = r'$, where $r' = \mathcal{R}' \cap \Pi_\infty$. Thus, we can consider the induced projectivity $\Psi = (\psi'_{|\mathcal{S}} \circ (\psi_{|\mathcal{S}})^{-1})_{|r}$ between the improper image lines (see Fig. A.1), which is given by

$$\Psi : r \xrightarrow{\vee O} O^* \xrightarrow{\cap s} s \xrightarrow{\vee O'} O'^* \xrightarrow{\cap r'} r' \,. \tag{A.2}$$

28

Once a reference world point $X_0 \in \mathcal{S}$ and its respective images $x_0 \in \mathcal{R}$ and $x_0' \in \mathcal{R}'$ are chosen, the affinity $\psi_{|\mathcal{S}}' \circ (\psi_{|\mathcal{S}})^{-1}$ is determined by $\{x_0, x_0', \Psi\}$, modulus a scale factor. In our application, the centroid of a world planar contour will be taken as the reference point, i.e., $X_0 = C$, and its projections $x_0$ and $x_0'$ will be easily recognized in the image planes as the centroids of the respective target contours, since the images are related to the scene plane by an affinity. Observe that, after the analysis of the camera model carried out, most of the relevant information of the affinity relating two weak-perspective views is enclosed by the projectivity $\Psi$ at infinity.

## A.2   The geometry at infinity

Let us analyze the situation in the plane at infinity. Since the optical centers of the two weak-perspective cameras lie at infinity, the epipoles $E = (O \vee O') \cap r$ and $E' = (O \vee O') \cap r'$ (of the first and second cameras, respectively) are also located at infinity in the image planes, i.e., the epipolar lines are parallel.

The rigid motion between the two cameras is decomposed into a rotation and a translation. The translation restricted to the plane at infinity is the identity. Hence only the change of pose modifies the location at infinity.

The rotation, say $\phi$, sends $r'$ to $r$ and it is decomposed into two parts (cf. Koenderink and van Doorn rotation representation [4]): a rotation about an axis of direction $A = r \cap r'$ and angle $\rho$, and a rotation about an axis of direction $O$ and angle $\theta$. The first rotation is a pure rotation *out of* the image plane, and the second is a rotation *inside* the image plane (also called *cyclorotation*). Let $\Phi : r' \to r$ be the restriction of the rotation to the improper line of the second image. The decomposition of the rotation restricted to $r'$ gives

$$\Phi : r' \xrightarrow{\Phi_\rho^A} r \xrightarrow{\Phi_\theta^O} r\,, \tag{A.3}$$

Observe that the polar of $A$ with respect to the absolute conic $\Omega$ is the line $E \vee E' = O \vee O'$. Hence $E \vee E'$ is the direction of the planes orthogonal to the axis of the first rotation, and then $\Phi_\rho^A(E') = E$ (see Fig. A.1). On the other hand, by (A.2) we obtain that $\Psi(E) = E'$. This gives a geometric proof of

**Theorem 3 (Geometric version of 1)** If the rigid motion between two weak-perspective cameras is assumed to be free of cyclorotation, then the epipolar direction $E$ can be recovered as one of the two fixed points of the improper image line homography $\Phi \circ \Psi$, which is induced by the affinity $\phi \circ \psi_{|\mathcal{S}}' \circ (\psi_{|\mathcal{S}})^{-1}$ that relates two views of a planar scene.

As a consequence, if the aspect ratio and skew (i.e. the intrinsic parameters determining the metric structure in the image) are known, the direction $A =$

$E^\perp$ of the axis of rotation can also be recovered.

Notice that, in image coordinates, the affinity $\phi \circ \psi'_{|\mathcal{S}} \circ (\psi_{|\mathcal{S}})^{-1}$ is given by Equation (3) and that the line homography $\Phi \circ \Psi$ is given by the matrix $\mathbf{M}$, appearing in the same Equation (3).

## B    Geometric study of two affine views

Keep the notations of Appendix A. The weak-perspective camera model assume that the distance of the target points to the principal ray are small. This can be achieved with a small field of view, when the projected target is centered in the image (assuming that the image center is a good approximation to the principal point). If the condition of being centered is relaxed, we need the more general affine camera model, which encloses the weak-perspective one: an *affine camera* $\overline{\psi}$ projects a scene point first under a fixed direction $\overline{O} \in \Pi_\infty$ onto the average depth plane $\mathcal{R}^C$ (the plane parallel to the image plane $\mathcal{R}$ containing the centroid $C$ of the scene object) and then perspectively from this fronto-parallel plane $\mathcal{R}^C$ onto the image $\mathcal{R}$. When $\overline{O}$ equals the direction $O$ orthogonal to the image plane, the affine camera becomes a weak-perspective camera.

If $P$ is the proper optical center of the second perspective projection, which simply introduces a scale factor, then the affine projection is

$$\overline{\psi} : \mathbb{P}^3 \xrightarrow{\vee \overline{O}} \overline{O}^* \xrightarrow{\cap \mathcal{R}^C} \mathcal{R}^C \xrightarrow{\vee P} P^* \xrightarrow{\cap \mathcal{R}} \mathcal{R} \,, \tag{B.1}$$

that is, analogous to the weak-perspective projection but for the improper optical center $\overline{O}$. Notice that, modulus a scale factor, all the projection rays of the affine camera are parallel with direction $\overline{O}$; and that, as well as the weak-perspective camera, a world plane $\mathcal{S}$ is mapped by the affinity $\overline{\psi}_{|\mathcal{S}}$ to the image plane, i.e. the camera preserves parallelism.

While in the weak-perspective camera model the improper optical center $O$ is determined by the orientation of the image plane (i.e., $O$ is the pole with respect to the absolute conic $\Omega$ of the improper line $r$ of $\mathcal{R}$), in the affine camera model the improper optical center $\overline{O}$ may be any point in $\Pi_\infty$. To see this, suppose that you have a camera under affine viewing conditions, that is, the target is seen under a small field of view. Then the direction of parallel projection, i.e., the improper optical center, depends on the position of the projected target within the image plane. Hence the same camera (with the same image plane) can take two affine views with different improper optical centers. This is one of the reasons that makes the affine camera model more difficult to handle than the weak-perspective one. Thus, we will speak of *a*
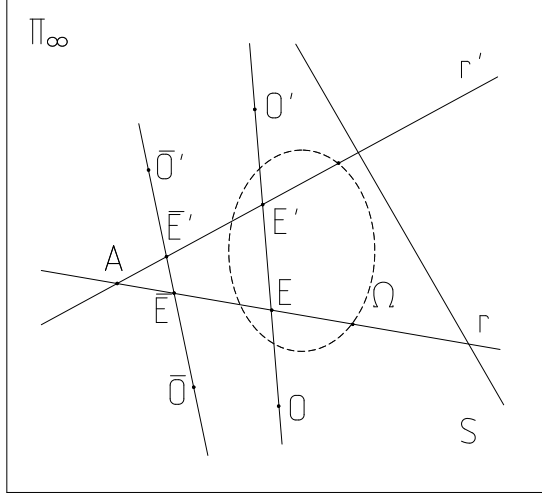
Figure B.1. Geometry on the plane at infinity of two general affine cameras. The notations are explained in the text.

*view of improper center $\overline{O}$ taken by an affine camera. If $\overline{O} = O$, we say that it is a weak-perspective view.*

Consider a second affine camera $\overline{\psi}'$, whose image plane, improper and proper optical centers are $\mathcal{R}'$, $\overline{O}'$ and $P'$, respectively; let $O'$ be the pole with respect to the absolute conic $\Omega$ of the improper line $r'$ of $\mathcal{R}'$. Let $\mathcal{S}$ be a world plane of improper line $s = \mathcal{S} \cap \Pi_\infty$. As in the case of two weak-perspective cameras, the projectivity $\overline{\psi}'_{|\mathcal{S}} \circ (\overline{\psi}_{|\mathcal{S}})^{-1}$ between the two images, induced by the restriction to a coplanar scene, is in fact an affinity, since image directions are mapped to image directions: $\overline{\psi}'_{|\mathcal{S}} \circ (\overline{\psi}_{|\mathcal{S}})^{-1}(r) = r'$, where $r' = \mathcal{R}' \cap \Pi_\infty$. We also consider the induced projectivity (see Fig. B.1) $\overline{\Psi} = (\overline{\psi}'_{|\mathcal{S}} \circ (\overline{\psi}_{|\mathcal{S}})^{-1})_{|r}$ between the improper image lines, which is given, in the general affine case, by

$$\overline{\Psi} : r \xrightarrow{\vee \overline{O}} \overline{O}^* \xrightarrow{\cap s} s \xrightarrow{\vee O'} \overline{O}'^* \xrightarrow{\cap r'} r'. \tag{B.2}$$

Consider the Koenderink and van Doorn decomposition [4] of the change of pose of the two affine cameras: a rotation about an axis of direction $A = r \cap r'$ and angle $\rho$, and a rotation about an axis of direction $O$ and angle $\theta$, and its restriction to the improper lines:

$$\Phi : r' \xrightarrow{\Phi^A_\rho} r \xrightarrow{\Phi^O_\theta} r. \tag{B.3}$$

The polar of $A$ with respect to the absolute conic $\Omega$ is the improper line $O \vee O'$, which will be called *virtual epipolar line*. The virtual epipolar line is the direction of the planes orthogonal to the axis of the first rotation, and then $\Phi^A_\rho(E') = E$, where $E = (O \vee O') \cap r$ and $E' = (O \vee O') \cap r'$ are called *virtual epipoles*.
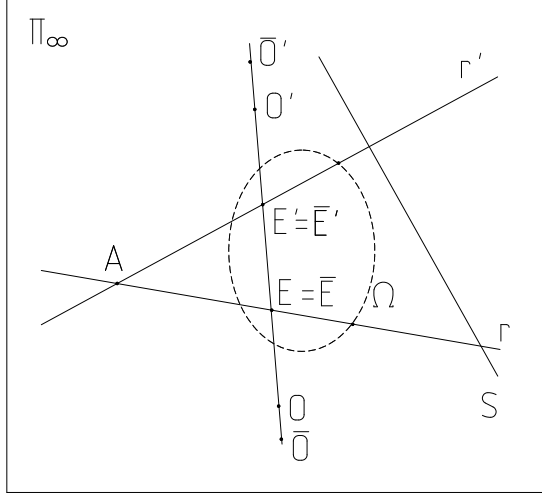
Figure B.2. Geometry on the plane at infinity of two affine cameras under the hypothesis of Theorem 4. The notations are clear from the text.

As in the case of two weak-perspective cameras, the epipoles $\overline{E} = (\overline{O} \vee \overline{O}') \cap r$ and $\overline{E}' = (\overline{O} \vee \overline{O}') \cap r'$ (of the first and second affine cameras, respectively) are also located at infinity in the image planes, i.e., the epipolar lines in both views are parallel. But, while in the weak-perspective cameras the epipoles coincide with the virtual epipoles, in the general affine cameras the epipoles are no more related with the virtual epipoles and, thus, a priori, they do not provide information about the rigid motion between the two affine cameras. This is one of the reasons why most of the literature about the general affine camera model switches to the weak-perspective camera model when the question of inferring camera motion is tackled.

In the more general setting of affine cameras we obtain the following generalization of Theorem 1, owing to the fact that $\overline{\Psi}(\overline{E}) = \overline{E}'$ by (B.2):

**Theorem 4** Assume that the rigid motion between two affine cameras is free of cyclorotation and that their improper centers lie on the virtual epipolar line. Then the epipolar direction $\overline{E} = E$ can be recovered as one of the two fixed points of the improper image line homography $\overline{\Phi} \circ \overline{\Psi}$, which is induced by the affinity $\overline{\phi} \circ \overline{\psi}'_{|\mathcal{S}} \circ (\overline{\psi}_{|\mathcal{S}})^{-1}$ that relates two views of a planar scene (see Fig. B.2).

As a consequence, in the situation where the affine camera takes first a weak-perspective view, and then moves and takes an affine view of improper center lying on the virtual epipolar line, the epipolar direction $\overline{E} = E$ can be recovered as one of the two fixed points of the improper image line homography, as well. This explains the results obtained in the experiments performed in Section 4.3.3.

# References

[1] M. Alberich-Carramiñana, G. Alenyà, J. Andrade-Cetto, E. Martínez, C. Torras, Affine epipolar direction from two views of a planar contour, in: Advanced Concepts for Intelligent Vision Systems, Vol. 4179 of Lect. Notes Comput. Sci., Antwerp, 2006, pp. 944–955.

[2] P. Beardsley, A. Zisserman, D. Murray, Sequential updating of projective and affine structure from motion, Int. J. Comput. Vision 23 (3) (1997) 235–259.

[3] P. F. McLauchlan, D. W. Murray, A unifying framework for structure and motion recovery from image sequences, in: Proc. IEEE Int. Conf. Comput. Vision, Massachusetts, 1995, pp. 314–320.

[4] J. Koenderink, A. J. van Doorn, Affine structure from motion, J. Opt. Soc. Am. A 8 (2) (1991) 377–385.

[5] L. S. Shapiro, A. Zisserman, M. Brady, 3D motion recovery via affine epipolar geometry, Int. J. Comput. Vision 16 (2) (1995) 147–182.

[6] P. Sturm, S. Maybank, On plane-based camera calibration: a general algorithm, singularities, applications, in: Proc. 13th IEEE Conf. Comput. Vision Pattern Recog., Vol. 1, Fort Collins, 1999, pp. 432–437.

[7] D. Demirdjian, A. Zisserman, R. Horaud, Stereo autocalibration from one plane, in: D. Vernon (Ed.), Proc. 6th European Conf. Comput. Vision, Vol. 1842, 1843 of Lect. Notes Comput. Sci., Springer-Verlag, Dublin, 2000, pp. 625–639.

[8] E. Malis, R. Cipolla, Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations, IEEE Trans. Pattern Anal. Machine Intell. 24 (9) (2002) 1268–1272.

[9] A. Bartoli, P. Sturm, R. Horaud, Structure and motion from two uncalibrated views using points on planes, in: Proc. 3rd. Intl. Conf. on 3D Digital Imaging and Modeling, Canada, 2001, pp. 83–90.

[10] O. Chum, T. Werner, J. Matas, Two-view geometry estimation unaffected by a dominant plane, in: Proc. 19th IEEE Conf. Comput. Vision Pattern Recog., San Diego, 2005, pp. 772–780.

[11] R. Tsai, T. Huang, W. Zhu, Estimating 3-d motion parameters of a rigid planar patch ii: Singular value decomposition, IEEE Trans. Acoust., Speech, Signal Processing 30 (8) (1982) 525–533.

[12] R. Horaud, F. Dornaika, B. Lamiroy, S. Christy, Object pose: the link between weak perspective, paraperspective, and full perspective, Int. J. Comput. Vision 22 (2).

[13] A. Blake, M. Isard, Active contours, Springer, 1998.

[14] J. Foley, A. van Dam, S. Feiner, F. Hughes, Computer Graphics. Principles and Practice, Addison-Wesley Publishing Company, 1996.

[15] E. Martínez, Recovery of 3d structure and motion from the deformation of an active contour in a sequence of monocular images, Ph.D. thesis, Universitat Politï¿$\frac{1}{2}$nica de Catalunya (2000).

[16] G. Alenyà, E. Martínez, C. Torras, Fusing visual and inertial sensing to recover robot egomotion, Journal of Robotic Systems 21 (2004) 23–32.

[17] G. Alenyà, J. Escoda, A.B.Martínez, C. Torras, Using laser and vision to locate a robot in an industrial environment: A practical experience, in: Proc. IEEE Int. Conf. Robot. Automat., Barcelona, 2005, pp. 3539–3544.

[18] E. Martínez, C. Torras, Qualitative vision for the guidance of legged robots in unstructured environments, Pattern Recognition 34 (2001) 1585–1599.

[19] E. Martínez, C. Torras, Contour-based 3d motion recovery while zooming, Robotics and Autonomous Systems 44 (2003) 219–227.

[20] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2nd Edition, Cambridge University Press, 2004.