

Affine Epipolar Direction from Two Views of a Planar Contour^{*}

Maria Alberich-Carramiñana¹, Guillem Alenyà², Juan Andrade-Cetto³,
Elisa Martínez⁴, and Carme Torras²

¹ Departament de Matemàtica Aplicada I,
UPC Avda. Diagonal 647, 08028 Barcelona
`maria.alberich@upc.es`

² Institut de Robòtica i Informàtica Industrial,
CSIC-UPC Llorens i Artigas 4-6, 08028 Barcelona
`{galenya, torras}@iri.upc.edu`

³ Centre de Visió per Computador, UAB Edifici O,
Campus UAB, 08193 Bellaterra, Spain
`cetto@cvc.uab.es`

⁴ CiTS La Salle, Universitat Ramon Llull Pge. Bonanova 8, 08022 Barcelona
`elisa@salleurl.edu`

Abstract. Most approaches to camera motion estimation from image sequences require matching the projections of at least 4 non-coplanar points in the scene. The case of points lying on a plane has only recently been addressed, using mainly projective cameras. We here study what can be recovered from two uncalibrated views of a *planar contour* under *affine* viewing conditions. We prove that the affine epipolar direction can be recovered provided camera motion is free of cyclorotation. The proposed method consists of two steps: 1) computing the affinity between two views by tracking a planar contour, and 2) recovering the epipolar direction by solving a second-order equation on the affinity parameters. Two sets of experiments were performed to evaluate the accuracy of the method. First, synthetic image streams were used to assess the sensitivity of the method to controlled changes in viewing conditions and to image noise. Then, the method was tested under more realistic conditions by using a robot arm to obtain calibrated image streams, which permit comparing our results to ground truth.

1 Introduction

Recovering camera motion from image streams is an important task in a range of applications including robot navigation and manipulation. This requires a measure of the visual motion on the image plane and a model that relates this motion to the real 3D motion. Most of the existing work on motion recovery relies on a set of point matches to measure visual motion, and, depending on the acquisition conditions, different camera models have been used to emulate the imaging process [1,2]. 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.

This paper addresses the motion estimation problem in the context of an affine camera using active contours to measure visual motion. There are several previous motion estimation methods based on affine cameras [3,4]. A common feature of these algorithms is that they require the matching of at least four non-coplanar points and fail for planar structures [5]. The particular case of features lying on planes has not been analyzed in detail thus far. The formulation of this problem is the core of the present paper.

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] or of planar curves [10]. However, when perspective effects diminish, the relationship between two views of a planar structure becomes an affinity, which invalidates the methods based on collineations.

Following the stratified analysis of motion for affine viewing conditions introduced by Koenderink and van Doorn [3] and revisited by Shapiro et al. [4], we first explore what information of the affine epipolar geometry can be inferred from the affine deformation of the projection of a rigid and planar contour in two weak-perspective views. This sets the basis to derive the motion parameters in a second stage. We show that, under a 3D motion free of cyclorotation, the epipolar direction can be recovered by relating the two affine views of the contour. A series of experiments is performed to test the sensitivity of the method to the different conditions imposed.

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 and gives some prospects for future work.

2 Analytic Study of Two Weak-Perspective Views

2.1 The Camera Model

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. A new affine

coordinate frame associated with a second camera is given by the rows of \mathbf{R} and the new origin lies at $-\mathbf{R}^\top \mathbf{T}$ thus a point in this second camera is given by the expression

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{f'}{Z'_{\text{ave}}} \begin{bmatrix} X' \\ Y' \end{bmatrix}, \quad (1)$$

where $[X, Y, Z]^\top = \mathbf{R}[X', Y', Z']^\top + \mathbf{T}$, f' is the new focal length, and Z'_{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 views of the coplanar scene are related by the affinity given by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{M} \begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{t}, \quad (2)$$

with

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

$$\mathbf{t} = -\frac{f'}{Z'_{\text{ave}}} \begin{bmatrix} R_{1,1} & R_{1,2} & R_{1,3} \\ R_{2,1} & R_{2,2} & R_{2,3} \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} + c \begin{bmatrix} R_{1,3} \\ R_{2,3} \end{bmatrix}, \quad (4)$$

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 $\mathbf{M}\mathbf{v}$ of the second image \mathcal{R}' . Since the affine references chosen in 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 ρ has the form

$$\mathbf{R} = \begin{bmatrix} (1 - \cos \rho) \cos^2 \alpha + \cos \rho & \cos \alpha \sin \alpha (1 - \cos \rho) & \sin \alpha \sin \rho \\ \cos \alpha \sin \alpha (1 - \cos \rho) & (1 - \cos \rho) \sin^2 \alpha + \cos \rho & -\cos \alpha \sin \rho \\ -\sin \alpha \sin \rho & \cos \alpha \sin \rho & \cos \rho \end{bmatrix}. \quad (5)$$

Hence, the matrix \mathbf{M} is

$$\mathbf{M} = s \frac{f'}{f} \begin{bmatrix} (1 - \cos \rho) \cos^2 \alpha & \cos \alpha \sin \alpha (1 - \cos \rho) \\ + \cos \rho + a \sin \alpha \sin \rho & + b \sin \alpha \sin \rho \\ \cos \alpha \sin \alpha (1 - \cos \rho) & (1 - \cos \rho) \sin^2 \alpha \\ -a \cos \alpha \sin \rho & + \cos \rho - b \cos \alpha \sin \rho \end{bmatrix}, \quad (6)$$

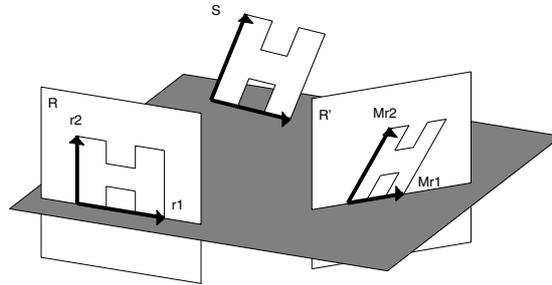


Fig. 1. Graphic illustration of Theorem 1. See text for details.

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

$$\mathbf{M}\mathbf{e} = s \frac{f'}{f} (\cos \rho + \sin \rho (a \sin \alpha - b \cos \alpha)) \mathbf{e}, \tag{7}$$

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 1 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 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 $\mathbf{M}\mathbf{r}_1$ 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 [11]. Within the same geometrical framework, this result is generalized to the affine camera model leading to Theorem 2. Let us sketch the main ideas of this generalized result; the reader is referred to [11] 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. 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 Π_∞) 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 \bar{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 Ω of the improper line r of \mathcal{R}), in the affine camera model the improper optical center \bar{O} may be any point in Π_∞ . 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 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. Let us state the announced generalization result:

Theorem 2. 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

$$\mathbf{x}' = \mathbf{M}\mathbf{x} + \mathbf{t} = \begin{bmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \quad (8)$$

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

$$M_{1,2}w^2 + (M_{1,1} - M_{2,2})w - M_{2,1} = 0. \quad (9)$$

If the motion is under the hypothesis of Theorem 1, then (9) 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 [12] 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 [13]. It has previously been shown [12] that the difference in control points $\mathbf{Q}' - \mathbf{Q}$ may be written as a linear combination of six vectors. Therefore, using matrix notation,

$$\mathbf{Q}' - \mathbf{Q} = \mathbf{W}\mathbf{S}, \quad (10)$$

where

$$\mathbf{W} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \mathbf{Q}^y \end{bmatrix}, \begin{bmatrix} 0 \\ \mathbf{Q}^x \end{bmatrix}, \begin{bmatrix} \mathbf{Q}^y \\ 0 \end{bmatrix} \right), \quad (11)$$

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

$$\mathbf{S} = [t_x, t_y, M_{1,1} - 1, M_{2,2} - 1, M_{2,1}, M_{1,2}]^\top, \quad (12)$$

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

Note that the dimension of the shape vector can be reduced if robot motion is constrained, for instance to lie on a plane [14].

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 [12], and the shape vector is updated at each frame.

In previous works [15,16], 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.

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 simul-

ating camera motion and computing projections under a full perspective camera model. Using this set, the sensitivity of the proposed algorithm to perspectivity effects is assessed by changing the distance of the target to the camera. A complete study involving the relaxation of all weak-perspective hypotheses can be found in [11].

The affine epipolar geometry is usually estimated using the Gold Standard algorithm [5]. 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.

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. In this setting, a comparison with the Gold Standard algorithm is also provided.

4.1 Simulations

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. Then, of course, perspectivity effects affect the recovery of the epipolar direction in the ways that will be analysed in the following.

In the first experiment we analyse how a decrement of the distance Z_{ave} from the camera to the target affects the computation of the epipolar direction. Decreasing the distance enlarges perspective effects, and consequently, should increase the error in epipolar direction recovery. For this experiment we consider distances of 500, 750, 1000, 1250, 1500, 1750 and 2000mm. The smallest of these, 500mm, corresponds to an extreme situation for the weak-perspective model, in which important unmodelled distortions in the projected control polygon are present. For larger depth values, the affine conditions are better satisfied, thus reducing the error, as shown in Figure 2. It is worth noting that even under these unfavourable conditions the recovery error stays below 0.6° .

The effects of relaxing other assumptions, such as lateral translations leading to uncentered targets, introducing depth relief, or having cyclorotation have also been explored and the results are given in [11], where the sensitivity to contour shape is also analysed.

Next we describe a comparison with a standard technique for computing the affine epipolar geometry, namely the Gold Standard (GS) algorithm [5]. 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 experimen-

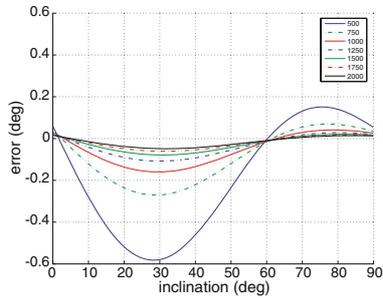


Fig. 2. Effects of relaxing one of the weak-perspective conditions by varying the distance from the camera to the target. The camera rotation is of 40° about an axis on the target with inclination of 45° .

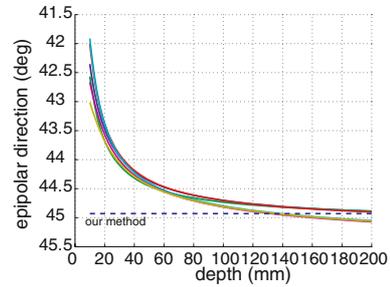


Fig. 3. Epipolar direction computed with the GS algorithm in the case of 2,4,...,12 out-of-plane points (a curve for each number) placed at increasing depths (in abscissae) above the H-shaped contour

tally the amount of depth information required by GS algorithm for it to provide equivalent epipolar direction recovery results to our procedure.

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 distance 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 3. It can be seen how as the depth of these points is increased, the error in the computation of the epipolar direction decreases. Moreover, 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 threshold indicating the error in the recovery of the epipolar direction using

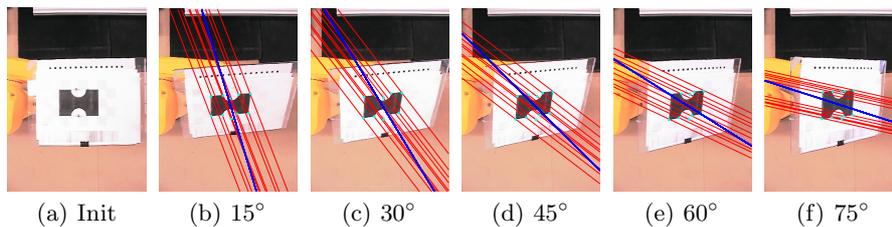


Fig. 4. The first experiment with real images entails pairs of views consisting of the initial one plus each of the other five, corresponding to camera rotations of 40° about an axis on the target with inclinations sampled at intervals of 15° . The epipolar direction computed by the proposed technique is displayed as a line passing through the target center, while the thin lines are the epipolar lines obtained with GS.

Table 1. Mean and standard deviation in degrees of the epipolar direction computed by the proposed technique and the GS algorithm from real images

epipolar direction	-15	-30	-45	-60	-75
$\bar{\theta}$	-16.63	-31.01	-45.00	-57.63	-72.04
σ	0.14	0.09	0.14	0.19	0.13
θ_{GS}	-18.53	-34.25	-49.46	-62.53	-76.36

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 those of the Gold Standard algorithm when the extra points are placed roughly at a distance equal to the target size (120 *mm* in our case).

Note the importance of parallax in the computation of the affine fundamental matrix with 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 also reduced, and the covariance of the 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.

4.2 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 from an active contour deformation, 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 active contour tracker. We are interested in using such setup in order to obtain a precise ground truth for the experiment. The initial distance from camera to target has had to be set to 500 *mm*. This corresponds to the extreme case discussed in Section 4.1, Fig. 2, and, therefore, we are testing the proposed approach under relaxed weak-perspective conditions. The acquired images have evident perspective effects, as shown in Figures 4 and 5, which make our algorithm work under extreme conditions. In order to provide depth information to the GS algorithm, the endpoints of two 20 *mm* screws placed at both sides of the contour are used as matching features in junction with the eight corners of the contour. Note that these are also extreme conditions for the GS algorithm to work, since

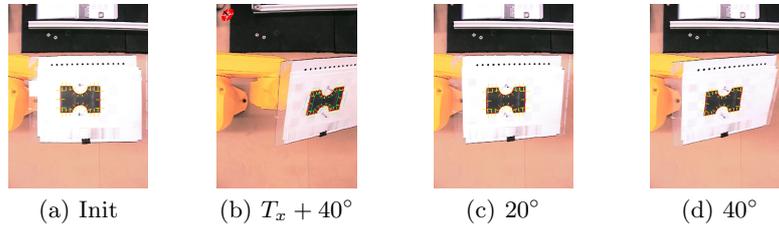


Fig. 5. Experiments with real images further relaxing weak-perspective conditions. The first sequence, entailing an uncentered target, starts at (a) and ends at (b). The next one departing from a non-frontoparallel target position starts at (c) and ends at (d).

Table 2. Mean and standard deviation of the epipolar direction computed over real images when weak-perspective conditions are further relaxed

<i>Frames</i>	$\bar{\theta}$	σ	θ_{GS}
Not Centered	-34.65	0.13	-56.29
Not Frontoparallel	-43.89	0.09	-49.78

very little depth information is provided: only two out-of-plane points. Thus, due to the setup we currently have, we are comparing both algorithms at the limit of their respective working conditions.

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. 2 with distance equal to 500 mm . Starting from the fronto-parallel position shown in Figure 4(a), the contour is tracked to each of the final views shown in the remaining frames of the figure. The epipolar direction computed by the proposed algorithm in each case is displayed as a line passing through the target center. Thin lines passing through the points correspond to the epipolar direction computed with the GS algorithm.

Table 1 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 ground truth as the angle departs from the 45° inclination. This should be interpreted in the light of Fig. 2 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. 2 when the distance Z_{ave} from the camera to the target increases. Results using the GS algorithm are slightly worse than those obtained with the proposed algorithm. This is due to perspective effects as well as to the poor depth information provided with the point matches used.

Two additional sequences were analyzed after further relaxing weak-perspective conditions. The first such sequence, labelled “Not centered”, starts at the

fronto-parallel initial position (Fig. 5(a)) and finishes at an uncentered position, after a translation of 100 *mm* along the x axis of the robot coordinate frame and a rotation of 40° about an axis at 45° inclination (Fig. 5(b)). Consistent with our simulated results [11], 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 2, first row, which is far from the motion assumption of Theorem 2. This pervasive effect appears also in the computation with the GS algorithm, yielding the largest error in the experiments.

The second experiment, labelled “Not Frontoparallel”, 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. 5(c) and, after a further rotation of 20° , finishes at 40° (Fig. 5(d)), all rotations about an axis at 45° inclination as before. Observe that the result is only a bit worse than that of the initial experiment, but with a similar standard deviation. The result with the GS algorithm here is similar as before.

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. 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.

Synthetic images were used to evaluate the results in a noise-controlled environment, 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 include an error analysis that involves positional errors on the contours due to the image acquisition process. Moreover, we 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. 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.

References

1. Beardsley, P.A., Zisserman, A., Murray, D.W.: Sequential updating of projective and affine structure from motion. *Intl. J. of Computer Vision* **23** (1997) 235–259
2. McLauchlan, P.F., Murray, D.W.: A unifying framework for structure and motion recovery from image sequences. In: *Proc. Intl. Conf. on Computer Vision*. (1995) 314–320
3. Koenderink, J., van Doorn, A.J.: Affine structure from motion. *J. Opt. Soc. Am. A* **8** (1991) 377–385
4. Shapiro, L., Zisserman, A., Brady, M.: 3d motion recovery via affine epipolar geometry. *Intl. J. of Computer Vision* **16** (1995) 147–182
5. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision* (Second Edition). Cambridge University Press (2004)
6. Sturm, P., Maybank, S.J.: On plane-based camera calibration: a general algorithm, singularities, applications. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. Volume 1. (1999) 432–437
7. Demirdjian, D., Zisserman, A., Horaud, R.: Stereo autocalibration from one plane. In: *Proc. 6th European Conf. on Computer Vision*. (2000) 625–639
8. Malis, E., Cipolla, R.: Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24** (2002) 1268–1272
9. Bartoli, A., Sturm, P., Horaud, R.: Structure and motion from two uncalibrated views using points on planes. In: *Proc. 3rd. Intl. Conf. on 3D Digital Imaging and Modeling, Canada* (2001) 83–90
10. Kaminski, J.Y., Shashua, A.: On calibration and reconstruction from planar curves. In: *Proc. European Conf. on Computer Vision*. (2000) 678–694
11. Alberich-Carramiñana, M., Alenyà, G., Andrade-Cetto, J., Martínez, E., Torras, C.: Affine epipolar direction from two views of a planar contour. Technical Report IRI-DT-2005/03, Institute of Robotics (IRI) (2005)
12. Blake, A., Isard, M.: *Active Contours*. Springer (1998)
13. Foley, J., van Dam, A., Feiner, S., Hughes, F.: *Computer Graphics. Principles and Practice*. Addison-Wesley Publishing Company (1996)
14. Alenyà, G., Martínez, E., Torras, C.: Fusing visual and inertial sensing to recover robot egomotion. *Journal of Robotics Systems* **21** (2004) 23–32
15. Martínez, E., Torras, C.: Qualitative vision for the guidance of legged robots in unstructured environments. *Pattern Recognition* **34** (2001) 1585–1599
16. Martínez, E., Torras, C.: Contour-based 3d motion recovery while zooming. *Robotics and Autonomous Systems* **44** (2003) 219–227