2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Depth from the visual motion of a planar target induced by zooming

Guillem Alenyà, Maria Alberich and Carme Torras

Abstract— **Robot egomotion can be estimated from an acquired video stream up to the scale of the scene. To remove this uncertainty (and obtain true egomotion), a distance within the scene needs to be known. If no a priori knowledge on the scene is assumed, the usual solution is to derive "in some way" the initial distance from the camera to a target object. This paper proposes a new, very simple way to obtain such a distance, when a zooming camera is available and there is a planar target in the scene. Similarly to "two-grid calibration" algorithms, no estimation of the camera parameters is required, and no assumption on the optical axis stability between the different focal lengths is needed. Quite the reverse, the non stability of the optical axis between the different focal lengths is the key ingredient that enables to derive our depth estimate, by applying a result in projective geometry. Experiments carried out on a mobile robot platform show the promise of the approach.**

I. INTRODUCTION

This paper presents a new method for inferring depth information using a zooming camera. In previous works [1], [2] we have shown how to recover robot egomotion from the deformation of an active contour. We have proposed to express the deformation of the contour in the image with a 6-dimensional affine *shape vector*. Then, with a non-linear non-derivable algorithmic function the performed 3D motion can be recovered up to a scale factor (as it is common in monocular vision). Scaled 3D motion can be recovered also in the context of a zooming camera [3]. Studying further the characteristics of the proposed affine shape space, we will show how the initial distance can be computed from the affine shape deformation caused by a zoom-lens camera.

Being based on active contour tracking, our egomotion recovery algorithm requires that the whole object projection keeps into the image all along the robot trajectory. This is sometimes too restrictive with a fixed camera, as the allowed robot motion is highly limited. One of the more promising solutions we have considered is to provide motion to the camera by means of a pan-and-tilt unit, and to implement a control algorithm to keep the target centered in the image (or at least within the image) in the whole sequence. One of the main problems of the control algorithm is that different gains should be applied depending on the distance from camera to target. Observe that, as usual in monocular imaging, it is not possible to disambiguate a priori the motion of a closer and small object from that of a far and big one.

Metric egomotion may be obtained if some additional information can be gathered. The scale factor depends on the camera focal distance and also on the initial distance from the camera to the viewed target. The camera focal distance can be obtained easily by a camera calibration or autocalibration method, even for zooming cameras [4]. The initial distance from camera to target is harder to obtain. In [2] we used a laser and other authors have proposed, for example, to use the range scanner of an autofocus camera [5], stereo correspondence, trifocal tensors [6], depth from defocused images [7] and depth from zooming.

In *depth from zooming* both camera and scene should be stationary and image deformation be caused only by zooming. Ma and Olsen [8] proposed a method to recover depth information from the variation in the focal distance and the optical flow. They noticed that the equation that describes the displacement obtained by zooming is similar to the one describing the translation of a camera along the optical axis. They assumed a thin-lens camera model (that nowadays is known not to be the most suitable model for zoom lenses [6]). In their mathematical formulation, they assumed that the apparent object translation is due exclusively to focal length variation. Lavest et al. [9] showed that this is not correct. In their work they use the thick-lens camera model, which is more accurate in modelling the focal change process. The correspondence that they establish between a thick-lens model and the corresponding pinhole configuration is interesting. To obtain good reconstruction data, a very accurate calibration process should be performed, including intrinsic (with radial distorsion) and extrinsic parameters. They were forced to use high-quality lenses, as they assumed that the optical axis was stable during the zooming sequence.

Rodin and Ayache [10] introduced a calibration method that does not require a physical axial camera. They used a geometric rectification method, but distorsions were not taken into account and the triangulation base they used was very small (only 50 mm).

Later, Lavest et al. [11] proposed an implicit reconstruction method that uses a two-plane geometric calibration procedure. The method was originally developed by Martins et al. [12] to solve the back-projection problem, and extended by Gremban et al. [13] to include also a solution to the projection problem, formulated with systems of linear equations. The idea is to find, without any explicit camera model, the ray in space that defines the line of sight of a given pixel. To calibrate, Lavest et al. used a micrometric table to translate the calibration pattern, as the reconstruction method that they

This work is partially funded by the EU PACO-PLUS project FP6-2004-IST-4-27657.

The authors are with the Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Llorens i Artigas 4-6, 08028 Barcelona {galenya,torras}@iri.upc.edu

Maria Alberich is also with the Departament de Matemàtica Aplicada I, UPC, Avda. Diagonal 647, 08028 Barcelona maria.alberich@upc.edu

proposed requires a high-precision calibration process. A new point in the image (located manually in [11] and by means of an iterative algorithm in [14]) can be triangulated with the calibration data to find the 3D point location. This method has the advantages of taking into account all distorsions, the optical center displacement produced when zooming, and not requiring the estimation of the camera parameters. A common comment [15], [16] is that it doesn't take into account the blurring effects that in some situations are produced when zooming.

The article is structured as follows. Section II presents the shape space that parameterises the general 6 d.o.f motion, and the reduced space corresponding to a zooming camera used to extract the required scale. In Section III we present the calibration algorithm and the proposed method to infer depth. Experiments with real images taken from a mobile robot are explained in Section IV. Finally, in Section V some conclusions and ideas about the applicability of the method in current approaches that require an initial depth estimate are stated.

II. AFFINITY RECOVERY FROM THE DEFORMATION OF AN ACTIVE CONTOUR

Under weak-perspective conditions (i.e., when the depth variation of the viewed object is small compared to its distance to the camera), every 3D motion of a planar object projects as an affine deformation in the image plane.

The affinity relating two views is usually computed from a set of point matches [17], [18]. In this work an active contour [19] fitted to a target object is used instead. The contour, coded as a B-Spline [20], deforms between views leading to changes in the location of the control points. A relation can be established between some extracted point features and a contour, considering the list of points as the set of the B-Spline control points. As a consequence, the method presented next, that obtains a motion parameterisation through pseudoinverse multiplication, can be applied also with point correspondences (as will be proved in Sec. IV).

It has been formerly demonstrated [19], [1], [3] that the difference in terms of control points $\mathbf{Q}' - \mathbf{Q}$ that quantifies the deformation of the contour can be written as a linear combination of six vectors. Using matrix notation

$$\mathbf{Q}' - \mathbf{Q} = \mathbf{W}\mathbf{S} \tag{1}$$

where

$$\mathbf{W} = \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} \mathbf{Q^x} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{Q^y} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{Q^x} \end{bmatrix}, \begin{bmatrix} \mathbf{Q^y} \\ \mathbf{0} \end{bmatrix} \right) \tag{2}$$

and $\mathbf{S}$ is a vector with the six coefficients of the linear combination. This so-called shape vector

$$\mathbf{S} = [t_x, t_y, M_{1,1} - 1, M_{2,2} - 1, M_{2,1}, M_{1,2}] \tag{3}$$

encodes the affinity between two views $\mathbf{d}'(u)$ and $\mathbf{d}(u)$ of the planar contour:

$$\mathbf{d}'(u) = \mathbf{M}\mathbf{d}(u) + \mathbf{t}, \tag{4}$$

where $\mathbf{M} = [M_{i,j}]$ and $\mathbf{t} = (t_x, t_y)$ are, respectively, the matrix and vector defining the affinity in the plane.

The deformation of the contour parameterized as a planar affinity permits deriving the camera motion in 3D space [1] even in the presence of zooming [3]. It has shown before that different deformation spaces can be defined corresponding to several constrained robot motions [21]. I.e. in the case of a planar robot, with 3 degrees of freedom, the motion space is parameterised with two translations (T_x, T_z) and one rotation (θ_y) yielding a three-dimensional shape space, which should be enlarged with one additional degree of freedom to cope with misalignments of the camera and robot coordinate systems [2].

Here the proposed solution is similar to the one in [2]. We need to define a reduced shape space able to deal with all the possible image deformations caused by zooming. First, the effect of zooming by a factor ρ is to translate the image point x along a line going from the principal point v_0 to the point $x' = \rho x + (1 - \rho)v_0$. At practical effects, this can be implemented by multiplying the calibration matrix corresponding to the first frame by the factor ρ, and it can be introduced directly as one of the degrees of freedom in the reduced shape space that we want to build. Second, the optical axis in a zooming camera is not constant [9], since the principal point position changes when zooming. To be able to model the translation effects present when zooming, we use the horizontal and vertical translation degrees of freedom[1]. The resulting shape matrix is of the form

$$\mathbf{W_{zoom}} = \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \begin{bmatrix} \mathbf{Q^x} \\ \mathbf{Q^y} \end{bmatrix} \right) \tag{5}$$

and the shape vector is

$$\mathbf{S} = [t_x, t_y, \rho] \ . \tag{6}$$

III. DEPTH FROM THE AFFINITY

As we will show, the algorithm presented here shares the main advantages of the "two-grid calibration" algorithm [12], [11]: no estimation of the camera parameters is required, and no assumption on the optical axis stability between the different focal lengths is needed. Quite the reverse, the non stability of the optical axis between the different focal lengths is the key ingredient that enables to derive our depth estimate. Note that if we try to model a zooming camera with the pinhole model we can assume neither that the optical axis is constant nor that the projection center is at the same place [4]. We only assume that the optical axis varies always in the same way between some two given focal lengths. We also suppose that the relation between two views of the same scene taken by a static zooming camera is accurately approximated by a planar homothetic transformation (a change in scale and a translation). As explained before, the scale factor (equivalently, the ratio of the homothetic transformation) accounts for the change in focal length, and the translation accounts for the displacement of the principal point, due to the non stability of the optical axis.

[1]This can be derived in a similar manner as was done in [21].

Furthermore, the proposed algorithm overcomes one of the major difficulties of the existing algorithms: it works well under affine viewing conditions. Moreover, from a computational point of view, it is a straightforward calibration algorithm: it avoids time-consuming minimization calculations, since the input data are ratios of three planar homothetic transformations. The estimation of these ratios relies on the restriction of a planar affine shape-space, which parameterizes the deformation of the projected target in the image (see Sec. II), combined with a quick an robust feature location method, such as an active contour tracking [19] or an affine-transfer based method [22].

A. Calibration algorithm

A planar target is located at a distance z_1 of the camera. The target is viewed by the camera at zoom A. Then the camera switches to zoom B and the homothetic transformation h_1 (whose ratio will be denoted ρ_1) that relates these two views (from zoom A to zoom B) is computed. This process is repeated at a distance z_2 of the camera: a planar target (it may be different from the preceding one) is viewed by the zooming camera, from zoom A to zoom B, and the homothetic transformation h_2 (whose ratio will be denoted ρ_2) that relates the initial and final views is computed.

If a new planar target (at an unknown distance z) is acquired with the zooming camera, again from zoom A to zoom B, then the homothetic transformation h (whose ratio will be denoted by ρ) that relates the initial and final views is computed. We claim that the ratio of depths $\frac{z_2-z_1}{z-z_1}$ may be computed from the ratios of the preceding homothetic transformations and is given by $\frac{\rho(\rho_2-\rho_1)}{\rho_2(\rho-\rho_1)}$. Thus, we obtain a straightforward estimation of the unknown depth z, without knowing any camera parameter. Moreover, the tedious use of metric instruments, such as a micrometric table, is avoided in the calibration process, since the relative orientation between the planes containing the two calibration targets is not relevant; besides, there is no need to use grids, hence the two calibration targets may be familiar objects in the scene (such as a door, window, board ...). The problem of computing accurately the ratio of the homothetic transformation relating the initial and final views of a zooming camera is overcome by reducing the dimension of the shape vector, which encodes the affine relation between the two views (see Section II).

B. Inferring the depth

We will show, as announced, how the non stability of the optical axis between the different focal lengths is used to infer our depth estimate.

We suppose that the direction of the optical axis in focal length A differs slightly from the direction of the optical axis in focal length B. Hence there exists an optical ray l in zoom A, which goes through an image point x, whose direction equals the direction of the optical axis a_B in zoom B (see Fig. 1).

This ray l is close to the optical axis in zoom A, and it cuts the calibration planes in the points X_1 and X_2, and the target

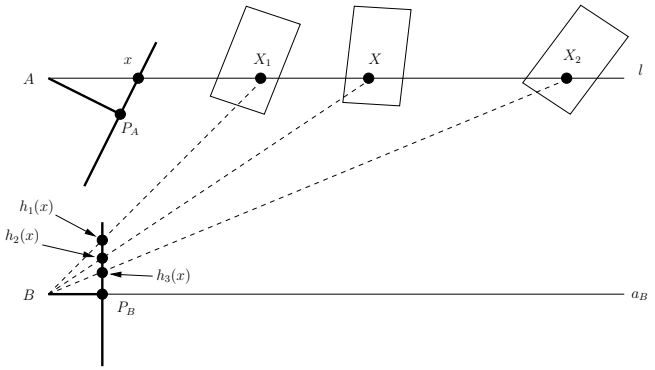

Fig. 1. A static zooming camera views the same scene with zoom A and zoom B. The variation of the optical axis between the two focal lengths has been magnified in order to exhibit the relevant features (see III-B) to infer the depth in the algorithm of Section III-A.

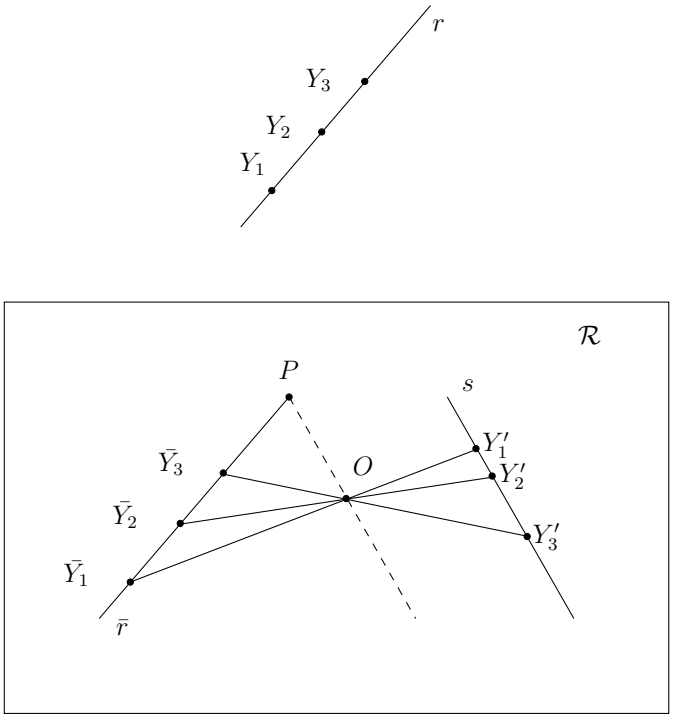

Fig. 2. Scene line r with three reference points Y_1, Y_2, Y_3 projected in the image $\mathcal{R}$ to $\bar{x}$ and $\bar{Y}_1$, $\bar{Y}_2$, $\bar{Y}_3$ respectively. P is the vanishing point of r. Auxiliary points are drawn on $\mathcal{R}$ and lines to derive the equality of simple ratios $(Y_1, Y_2, Y_3) = (Y'_1, Y'_2, Y'_3)$ claimed in Theorem 1.

plane in the point X. Thus the simple ratio of these points $(X_1, X_2, X) = \frac{d(X_1,X_2)}{d(X_1,X)}$ (where $d(Y_1, Y_2)$ is the distance between two points Y_1 and Y_2) is a sharp estimate of the ratio of depths $\frac{z_2-z_1}{z-z_1}$.

The scene points X_1, X_2 and X are projected in zoom B to the image points $h_1(x)$, $h_2(x)$ and $h(x)$, respectively (see Fig. 1). Our goal is to determine the simple ratio of the scene points (X_1, X_2, X) from the image points $h_1(x)$, $h_2(x)$ and $h(x)$. This is done by applying the following result of projective geometry:

Theorem 1: Given the vanishing point P of a scene line r, with three reference points Y_1, Y_2, Y_3, then the simple

Fig. 3. Pioneer 3AT mobile platform used in the experiments

ratio (Y_1, Y_2, Y_3) can be computed from their imaged points $\overline{Y}_1$, $\overline{Y}_2$, $\overline{Y}_3$ as follows: choose an image point O (not on the imaged line $\overline{r}$) and an image line s (not going through O) parallel to the line joining O and P; for $i = 1, 2, 3$, determine the point Y_i' lying on s and on the line joining O and $\overline{Y}_i$; then $(Y_1, Y_2, Y_3) = (Y_1', Y_2', Y_3')$ (see Fig. 2).

The case that concerns us is when $r = l$ and $Y_1 = X_1$, $Y_2 = X_2$, $Y_3 = X$. The vanishing point of l (the image point of the point at infinity of l) is the principal point $P_B = P$ in zoom B. The assumption that the optical axis varies always in the same way between zoom A to zoom B is equivalent to $P_B = h_1(P_A) = h_2(P_A) = h(P_A)$, where P_A is the principal point in zoom A. Therefore, if we fix an image reference system centered at $P = P_B$, with first vector in the direction of $\overline{r}$ and unit length $d(x, P_A)$, then $h_1(x)$, $h_2(x)$ and $h(x)$ have coordinates $(\rho_1, 0)$, $(\rho_2, 0)$ and $(\rho, 0)$, respectively. By choosing, for instance, $O = (0, -1)$ and the line $x = 1$, and by applying Theorem 1, we obtain the desired result

$$(X_1, X_2, X) = \frac{\rho(\rho_2 - \rho_1)}{\rho_2(\rho - \rho_1)}. \tag{7}$$

IV. EXPERIMENTS

The performance of the proposed algorithm has been tested on real images acquired with a Sony DFW-VL500 digital camera. The camera brochure states that the zoom of the camera can be moved to predefined positions ranging from 40 to 1432 corresponding to focal lengths from 5.5 to 64 mm. The camera is mounted on a Pioneer mobile platform (see Fig. 3). The translations performed with the robot are roughly estimated with marks on the floor. The drawers of a table and a stool serve as *natural* landmarks from which calibration information is extracted. Although the focus of the camera is kept constant, no defocus problems have been observed in the range of zoom positions and distances that we have used.

TABLE I

RESULTS OF ESTIMATED DEPTHS USING DIFFERENT CALIBRATION DISTANCES AND DIFFERENT TARGET OBJECTS.

Exp. ID	Cal1	Cal2	Estimated	Measurements
1	240	360	277.6	280
2			321.4	320
3			401.7	400
4			269.8	280
5	240	320	288.2	280
6			357.8	360
7			281.6	280
8	320	360	367.7	400

The robot takes an image pair with zoom in positions 40 and 708, at distances 240, 280, 320, 360 and 400 cm with respect of the table drawers. From Figure 4(a) to Figure 4(d) the image pairs corresponding to 240 and 360 cm are plotted. For the distance 280 we use also a wood stool (see Fig. 4(e) and 4(f)) to validate that the proposed method is only dependent on the zooming camera, and not on the calibration object. The idea is to perform the calibration off-line with a natural landmark, and use this calibration in real-time operations with any given new landmark, as usual with other calibration methods. The steps to compute the unknown depth are detailed in Alg 1.

1 **for** *i=1* **to** *2* **do**
2 Place camera at distance d_i from the calibration object
3 Compute the shape vector S_i produced by the deformation between the image taken at $zoom_1$ and the one at $zoom_2$
4 **end**
5 Place the camera at unknown distance from the target object
6 Compute the shape vector S produced by the deformation between the image taken at $zoom_1$ and the one at $zoom_2$
7 With S_1, S_2 and S find the unknown distance by applying (7)

Algorithm 1: Steps of the depth estimation algorithm

Four points are manually extracted for each drawer image in order to construct the corresponding shape vector. For the stool images, six points are extracted instead, in order to assess the robustness of the shape vector obtained. As the method to obtain the shape vector through pseudoinverse multiplication can be seen as a minimization [19], the more point location measures are available, the more precision can be obtained.

Some results are summarized in Table I. The columns labelled Cal1 and Cal2 indicate the two distances used to perform the geometric calibration, and the other two columns show the estimated distance by the presented algorithm and

(a) d=240cm, Zoom=40. (c) d=360cm, Zoom=40. (e) d=280cm, Zoom=40.

(b) d=240cm, Zoom=708. (d) d=360cm, Zoom=708. (f) d=280cm, Zoom=708.

Fig. 4. For each camera position two images are needed to estimate the scale factor. The shown images correspond to the experiment labelled 4 in Table I. (a) (b) First calibration pair. (c)(d) Second calibration pair. (e)(f) Testing pair. Note that the calibration object and the one used for testing are not the same, and also different numbers of location measures are used to estimate the shape vector, 4 for the drawer images and 6 for the stool ones.

the measured one. For the experiments labelled 1 and 2, the camera is placed at 280 and 320 cm from the drawer, respectively. These depths are between the two calibration distances (240 and 360 cm), and the estimated depth is correctly computed by the algorithm in each experiment. In the experiment labelled 3 the camera is placed farther than the second calibration distance (out of the calibration range), and the depth is also recovered with small error, compared to the measured one. With these calibration parameters we perform a fourth experiment (numbered 4) using the 6 points extracted from the stool images. In this case depth is also reasonably recovered, although worse than in the previous cases.

In experiments 5, 6 and 7 the calibration range is shortened, using the calibration distances 240 and 320 cm. When the distance is between the calibration ones, as in experiment 5, the error is of the same order as in the previous experiment. When the camera is located farther than the second calibration distance, the depth is correctly recovered but with more error, compared to experiment 3. As typical in geometric calibration, the depth is correctly recovered within the range defined by the first and the second calibration distances as the algorithm is interpolating. Out of this zone the depth can be also inferred extrapolating, but the error grows as the distance increases. We find also that the larger the distance between calibration positions the more precision is obtained.

Finally, with experiment 8 we test the effect of moving both calibration camera positions farther away. Calibration was done with images taken at 320 and 400cm. A test is performed placing the camera in the middle obtaining a correct recovered depth.

V. CONCLUSIONS AND FUTURE WORKS

We have presented a simple method to determine the depth of a robot placement with respect to a landmark. The image deformation caused by zooming is modelled by a 3 degrees of freedom shape vector in a presented shape space, where the third element is the scale of the associated homotecy. This simple scale value is recorded at each calibration step. When a new scale is computed from the zooming of a new object, it can be compared to the calibration scales and, knowing the depth of the calibration objects, deduce the depth of the current target with a simple operation.

A minimum set of 3 point correspondences are needed to construct the affinity, but more correspondences will result in a better shape vector estimation, as a minimisation process is used. Here we have presented experiments using 4 and 6 correspondences between zooming images.

With the experiments we have demonstrated the validity of the method. The distance between calibration positions determines a calibrated zone where the algorithm is more precise. Out of this zone the algorithm also infers the depth but is less precise as the distance increases. We have demonstrated that the required shape vector can be calculated from different objects and using different numbers of point correspondences.

We have observed that the zooming sometimes drops the target out of the image. For practical purposes it is convenient to calibrate with some different zoom positions to be able to find one zoom range that contains the target in both images and for which we have calibration information.

Our objective has been mainly to remove from the ego-motion algorithm the scaling uncertainty, common in all monocular systems. But this method can be used also for other purposes, for instance, the initialisation of the pan and tilt controllers of our active vision system. Experiments with the PTZ control show that the obtained precision is enough to initialize the controllers in a good response zone.

In [2] we estimate the initial distance with a laser, and in [23] with a calibration pattern. Several other algorithms could benefit from the estimation of the initial distance of a given landmark. Let us just enumerate a few. Davison [24] estimate the depth of a landmark in monocular vision using a particle filter. In order to acquire the scale of the scene in the first frame a known object is used. Our method can be used thus changing the known object by any object in the scene. Sola [25] proposed to solve the depth initialisation problem with an approximation of the Gaussian Sum Filter, and Jensfelt et. al. [26] proposed to exclude from the SLAM process those features for which the depth had not been determined. When little disparity between matched features is present, for example in approaching robot motions and distant targets, all these methods could not extract significant information.

Recently Caballero et. al. [27] presented a monocular visual odometer for aerial vehicles. They proposed to measure the distance between the camera and the various targets used in the experiments with a sonar or a laser range sensor, but finally they did it manually.

Obviously, for traditional point-based maps it is not practical to perform the zoom positioning for each landmark initialisation. However, the presented algorithm is useful for those situations where an average depth is needed, as those mentioned before.

REFERENCES

[1] E. Martínez and C. Torras, "Qualitative vision for the guidance of legged robots in unstructured environments," *Pattern Recognition*, vol. 34, pp. 1585–1599, 2001.

[2] G. Alenyà, J. Escoda, A.B.Martínez, and C. Torras, "Using laser and vision to locate a robot in an industrial environment: A practical experience," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA'05)*, Barcelona, Apr. 2005, pp. 3539–3544.

[3] E. Martínez and C. Torras, "Contour-based 3d motion recovery while zooming," *Robotics and Autonomous Systems*, vol. 44, pp. 219–227, 2003.

[4] M. Li and J.-M. Lavest, "Some aspects of zoom-lens camera calibration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 11, pp. 1105–1110, November 1996.

[5] J. A. Fayman, O. Sudarsky, E. Rivlin, and M. Rudzsky, "Zoom tracking and its applications," *Machine Vision and Applications*, vol. 13, no. 1, pp. 25 – 37, 2001.

[6] B. Tordoff, "Active control of zoom for computer vision," Ph.D. dissertation, University of Oxford, 2002.

[7] Y. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: How different really are they?" *Int. J. Comput. Vision*, vol. 39, no. 2, pp. 141–162, 2000.

[8] J. Ma and S. I. Olsen, "Depth from zooming," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 1883–1890, oct 1990.

[9] J.-M. Lavest, G. Rives, and M. Dhome, "Three-dimensional reconstruction by zooming," *IEEE Trans. Robot. Automat.*, vol. 9, pp. 196–206, 1993.

[10] V. Rodin and A. Ayache, "Axial stereovision: Modelization and comparison between two calibration methods." in *Proc. 1stIEEE Int. Conf. Image Process.*, Austin, Texas, Nov. 1994, pp. 725–729.

[11] J. Lavest, C. Delherm, B. Peuchot, and N. Daucher, "Implicit reconstruction by zooming," *Comput. Vis. Image Und.*, vol. 66, no. 3, pp. 301–315, June 1997.

[12] H. A. Martins, J. R. Birk, and R. B. Kelley, "Camera models based on data from two calibration planes," *Comp. Graph. Image Processing*, vol. 17, no. 2, pp. 173–180, 1981.

[13] K. Gremban, C. Thorpe, and T. Kanade, "Geometric camera calibration using systems of linear equations," in *Proc. Image Understanding Workshop*, Cambridge, 1988, pp. 820–825.

[14] C. Delherm, J.-M. Lavest, M. Dhome, and J.-T. Laprest, "Dense reconstruction by zooming," in *Proc. 4th European Conf. Comput. Vision*, ser. Lect. Notes Comput. Sci., B. Buxton and R. Cipolla, Eds., vol. 1065. London, UK: Springer-Verlag, Apr. 1996, pp. 427–438.

[15] M. Baba, N. Asada, A. Oda, and T. Migita, "A thin lens based camera model for depth estimation from blur and translation by zooming," in *Proc. 15th Int. Conf. Vision Interface*, Calgary, May 2002, pp. 274–281.

[16] Z. Myles and N. da Vitoria Lobo, "Recovering affine motion and defocus blur simultaneously," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 6, pp. 652–658, June 1998.

[17] J. Koenderink and A. J. van Doorn, "Affine structure from motion," *J. Opt. Soc. Am. A*, vol. 8, no. 2, pp. 377–385, 1991.

[18] L. S. Shapiro, A. Zisserman, and M. Brady, "3D motion recovery via affine epipolar geometry," *Int. J. Comput. Vision*, vol. 16, no. 2, pp. 147–182, 1995.

[19] A. Blake and M. Isard, *Active contours*. Springer, 1998.

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

[21] E. Martínez, "Recovery of 3d structure and motion from the deformation of an active contour in a sequence of monocular images," Ph.D. dissertation, Universitat Politï¿½nica de Catalunya, 2000.

[22] B. Tordoff and D. Murray, "Reactive control of zoom while fixating using perspective and affine cameras," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 1, pp. 98–112, January 2004.

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

[24] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proc. IEEE Int. Conf. Comput. Vision*, Nice, Oct. 2003, pp. 1403–1410.

[25] J. Sola, A. Monin, M. Devy, and T. Lemaire, "Undelayed initialization in bearing only SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Edmonton, Aug. 2005.

[26] P. Jensfelt, D. Kragic, J. Folkesson, and M. Bjorkman, "A framework for vision based bearing only 3d slam," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA'06)*, Orlando, May 2006, pp. 1944–1950.

[27] F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "A visual odometer without 3d reconstruction for aerial vehicles. applications to building inspection," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA'05)*, Barcelona, Apr. 2005, pp. 4673–4678.