Proceedings of the 2004 IEEE
International Conference on Robotics & Automation
New Orleans, LA • April 2004

Homography-Based Grasp Tracking for Planar Objects

Raffaella Carloni*, Gabriel Recatalá†, Claudio Melchiorri*, Pedro J. Sanz† and Enric Cervera†
* Dept. of Electronics, Computer Science and Systems (DEIS) - Lab. of Automation and Robotics (LAR)
University of Bologna, 40136 Bologna, Italy
Email:{rcarloni, cmelchiorri}@deis.unibo.it
† Dept. of Computer Science and Engineering - Robotic Intelligence Lab.
Jaume I University, 12071 Castellón, Spain
Email:{grecata, sanzp, ecervera}@icc.uji.es

Abstract—**The visual tracking of grasp points is an essential operation for the execution of an approaching movement of a robot arm to an object: the grasp points are used as features for the definition of the control law. This paper describes a strategy for tracking grasps on planar objects based on the use of a homography. In particular, the homography is used for transferring (translating) a grasp from a view of an object to a second one, providing in this way a correspondence in the second view. The grasp tracking procedure combines the search of a new grasp with the translation and the evaluation of that grasp between object views. Results of the proposed grasp tracking strategy are shown in case of grasps computed for a two-finger and a three-finger gripper.**

I. INTRODUCTION

One of the basic requirements for a robotic system acting in an unknown dynamic environment is the capability of detecting features in the surrounding environment and to trace them during the relative motion. This capability can be used e.g. when grasping unknown object, case in which proper grasping strategies must be adopted and the desired grasp configuration must be reached also in dynamic situations. In this paper, in particular, a method for tracking selected grasp points of unknown planar object is presented and illustrated.

The tracking problem consists on the search of matching points between sequencial views of the object. This problem can be solved with homography, a projective transformation in image space of two views of the same physical plane [1]. In order to compute the homography matrix, the availability of some initial correspondences between two views of the image plane is required. With this respect, in this paper, a methodology based on the analysis of the curvature of the object to be grasped is explained.

Besides the theoretical presentation of the proposed method, experimental results, obtained with two different laboratory setups, are illustrated (see Fig. 1). The first is present at the Robotic Intelligence Laboratory (Jaume I University of Castellón, Spain) and consists of a seven degrees of freedom (dof) robot arm, a gripper with two flat parallel fingers and a pair stereo. The second setup is available at the Laboratory of Automation and Robotics (University of Bologna, Italy) and consists of a six dof robot arm, a gripper with three one-dof fingers and a single camera.

(a) Setup in Castellón (Spain) (b) Setup in Bologna (Italy)

Fig. 1. The experimental setups.

II. GRASP 'TRANSLATION'

Several solutions to the tracking problem have been presented in the literature.

The control of the positioning movement between the robot and the object is usually based on some features other than the grasp points. For example, in [2], some marks explicitly set on the object and previously known are considered, as in the classical visual servoing techniques; in [3], some geometric parameters related to the projection of the object, such as the measure of the image velocity at each pixel, are used. In [4], a method based on homography has been presented in which the whole shape of the object is taken into account for the computation of a homography, assuming as known the matrix of the camera internal parameters. In some cases, as in [5], the grasp search and the computation of the target position in the control loop have been performed off-line. A grasp tracking strategy is proposed in [6], where for each image the previously computed grasp is applied and the robot movement is restricted to four dof (translation along x, y, z and rotation about z).

In the literature, a general solution to the problem of finding correspondences between images has not been presented yet: there are many works that use the fundamental matrix and the epipolar geometry, but with the hypothesis of not having parallel views of the object [4], [7], [8] and others that are based on the description of the grasp with B-splines that approximate the contour of the object [9], [10].

A. Overall description of the tracking problem

In the approaching movement of a robot arm towards the object to be grasped, the main problem is to realize a visual servoing control loop which could track significant points, such as the grasp points. Since the robot has to move in the free space, that is with six dof, it is necessary to find an invariant rappresentation of the coordinates of that object in the image space.

This allows, once the desired grasp configuration (i.e. the contact points) has been defined in the first image, to track the same configuration among a sequence of different (sequencial and/or pairs) images of the same object. In this way, the desired grasp configuration can be used in real time as final target for the robotic system, independently on its motion. Here, the tracking of the desired grasp along different images of the object is referred to as "grasp translation". Note that the desired grasp configuration consists of a finite number of points (two or three in this paper) depending on the geometry of the gripper.

The tracking algorithm is structured as follows:

Algorithm 1 Homography-based grasp translation

extract the external contour of the object in each image;
extract four correspondences from the contours;
compute the homography matrix by using those correspondences;
perform the grasp translation using the homography.

B. Computation of the homography

The search of correspondences between two image planes that contain two views of the same scene is a typical problem of projective geometry. In fact, in the projective plane model, transformations are represented by mappings of the plane into itself. In the projective space a transformation between two planes can be represented by a general linear transformation H, called *homography* [1].

Since the projective plane has three homogeneous coordinates, the transformation is represented by a 3x3 matrix with eight independent parameters. The scale of the matrix is not important because all the projective points are equivalent up to a scalar λ, so one element of the matrix can be set to a fixed value. Therefore, the set of distinct projective transformations becomes an eight-dimensional subspace of the nine-dimensional space defined by the matrix elements. The projective matrix requires eight independent parameters to define an unique mapping: so, it is necessary to find at least four initial correspondences (i.e. four no collinear point) between two projectively transformed planes to define the transformation matrix uniquely. With more than four correspondences, it is possible to compute the homography matrix using numerical, iterative methods, which are less sensitive to measurement errors in the initial correspondences.

Let the four corresponding points be (u_1, v_1), (u_2, v_2), (u_3, v_3), (u_4, v_4), in the plane I and (u_1', v_1'), (u_2', v_2'), (u_3', v_3'), (u_4', v_4'), in the plane I'. The relationship between these correspondences can be written as:

$$\begin{bmatrix} \lambda_i u_i' \\ \lambda_i v_i' \\ \lambda_i \end{bmatrix} = H \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}, \quad i \in [1, 4] \tag{1}$$

These equations define a linear system that can be used to compute the components of the homography matrix H.

This transformation is normally used in computer vision and lets to find correspondences between all the points of two different planes, i.e. of the two different image planes, that represent the same physical plane: given a homography matrix, for every generic point in one image, it is possible to find its corresponding point on the other image. With this regard it is not important if the two images come from a stereo ring images or are a sequence from a single camera.

In order to reduce the computational burden of obtaining H, a direct, non-iterative method has been used to solve the linear system (1). This direct method has the drawback that is more sensitive to errors in the position of the correspondences. However, the observed errors in the application of the homography to translate have resulted small and have been compensated with the correction method explained below.

In this paper, only four corresponding points have been considered for the on-line computation of H, and a strategy for the selection of these correspondences based on the analysis of the curvature of the contour in each view of the object is presented. Nevertheless, the visual tracking control loop is independent of the actual procedure considered for finding the initial correspondences.

C. Extraction of correspondences

In the literature, the initial correspondences required for the computation of the homography are often assumed or selected manually, [7]. In [8], Zhang et al. select corners in one image and try to match image sub-windows centered at them in the other image. In other works, the selection of correspondences is based on the search of epipolar tangences between two images [9].

In this paper, a matching procedure has been developed for the automatic selection of four correspondences between the projected silhouettes of the object. For simplicity, only the external contour of the object in each view has been considered. As the object is assumed to be planar, the homography computed using corresponding points from the external contour could be used to transfer points belonging also to any of the internal contours; anyway, an extension of this procedure to include both the external and a set of internal contours would imply the definition of an initial step to determine the correspondences.

The algorithm describing the extraction of correspondences is organized as follows:

Algorithm 2 Extraction of correspondences

compute curvature of each contour;
select 4 points with highest curvature in each contour;
find a matching between the selected points.

The proposed matching procedure uses the curvature to try to select points that are representative of the shape of the contour. In particular, it selects the corners that are identified as peaks in the vector of curvatures computed for that contour. A discrete curvature function was used to compute the curvature at each point of the contour. In particular, the k-torsion was considered for this purpose [11]. However, the matching is independent of the method for computing the curvature, so functions other than the k-torsion could also be used. This use of the curvature for the selection of relevant points has also been considered in other works [10].

As four point matches are required to compute the homography, the four points within the contour with the highest curvature are selected. First of all, by considering the curvature of the first object and its derivative function, it is possible to find the four points with the maximal curvature. Next, the procedure tries to match those points on the other corresponding contour.

Let I' and I be the images that contain the corresponding contours. As it can be noticed in Fig. 2, for each point p' in the first image the procedure considers a neighborhood of the point which is at the same position in the vector of the points of the contour of the second image, p_{corr}. Starting from that point, p_{corr}, the displacement from the two corresponding curvature is calculated. In particular for each point p' the algorithm calculate:

$$d_i = \sum_{i=-n}^{n} |v'(i) - v(i)|, \tag{2}$$

where v' is the value of the curvature of the point p' in the first image and v of p_{corr}. The value of v is calculated in a neighborhood of $2n$ of the point p_{corr}, where n depends on the size of the object. Then other displacements are calculated:

$$d'_i = \sum_{i=-n}^{n} |v'(i) - v(i \pm \xi)|, \tag{3}$$

where ξ is the generic dimension of the window in which the correspondence is searched. Here, big changes of scale between the two images are taken into account.

Finally, the point p that has the minimum displacement (in terms of curvature) from the point p', is selected as the corrersponding point of p' [12], that is:

$$min(d_i, d'_i). \tag{4}$$

The procedure of contour analysing is illustrated in Tab. I: three objects are shown with two views of each on and their relative curvature functions: the objects taken into account are a floppy disk, an Allen wrench and the handle of a tool. For each view, the four points with maximum curvature are selected on the curvature function (shown as -). These four maxima are used to select four points on the object view on top (shown as $*$). The correspondences of these four points in the other view (shown as $\bullet$) are found by determining the matching between the four maxima in each curvature function).

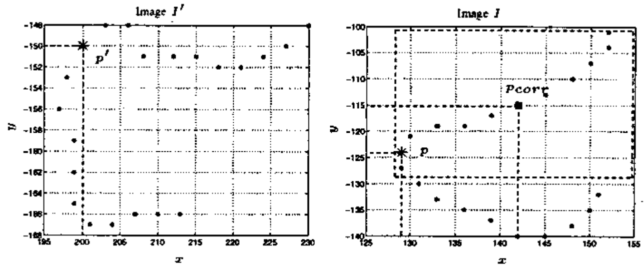
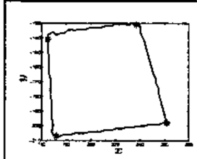
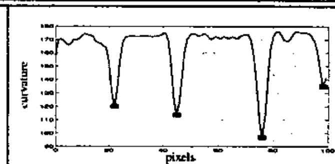
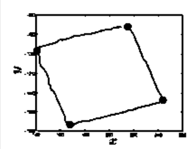
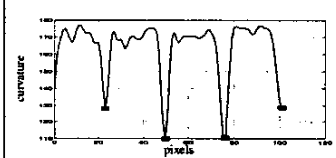
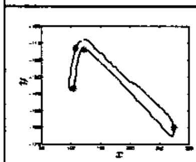
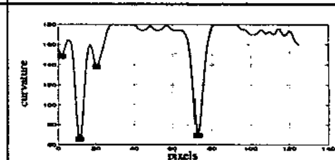
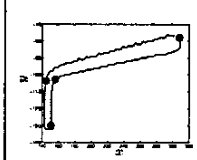
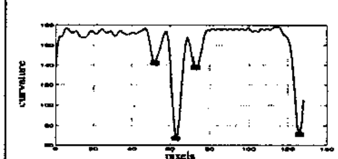
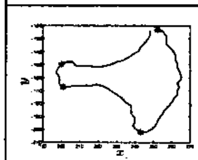
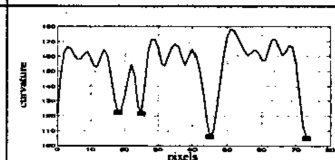
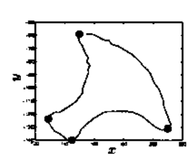
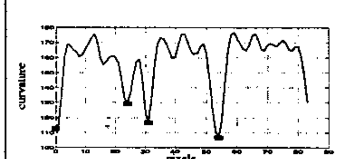


Fig. 2. Find correspondences.

TABLE I

SELECTED POINTS ($*$) AND CONTOUR CURVATURE ON THE FIRST CONTOUR AND THEIR CORRESPONDENCES ($\bullet$).

First/second contour:	Curvature of first/second contour:
	
	
	
	
	
	

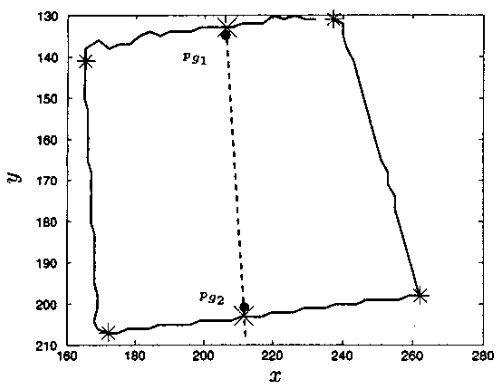


Fig. 3. Correction of the position of the grasp points.

D. Grasp translation procedure

Ideally, the points p_{g_i} produced by the homography should lie on the corresponding contour extracted from I. Nevertheless, due to errors in the contour extraction in I' and I, in the selection of the corresponding points for the computation of the homography matrix, and/or in the computation itself of this matrix, the produced points may happen not to be in that contour, but close to it.

In general, the correction in the position of points p_{g_i}, so that they lie on the contour, can be made considering the closest points to them that belong to the contour as the translated grasp points. A more precise correction can be achieved by taking into account known restrictions in the relationship between the grasp points.

For instance, in the case of a two fingers gripper, the two points produced by the homography define a line: this line is considered to be the grasp line of the translated grasp. The translated grasp points can be considered as the points of the contour that are the closest to points p_{g_i} and also belong to this line, as shown in Fig. 3.

For grasps with more than two points, the relative location of the grasp points is usually restricted by the mechanical design of the gripper. In the case of the three-finger gripper gripper considered in this paper, the normals to the contour at the contact points should define an angle of $120°$ between them. Therefore, the normals to the contour going through points p_{g_i} can be used to find the corrected position of the translated grasp.

III. Grasp Tracking

A translation mechanism such as the one described above is one of the fundamental steps in the tracking of a grasp.

A grasp tracking procedure for the case of a sequence of images provided by a single camera is outlined in algorithm 3. This algorithm is applied to each of the images in the sequence. According to this algorithm, as no grasp is available at the beginning of the sequence, a new one is searched for in the first images of the object until a stable grasp is found. For each new incoming image that grasp is translated from the previous image using the grasp translation algorithm described in section II-D and the stability of the new grasp is evaluated using criteria related to the considered grasp search strategy [6], [13], [14]. In case that the grasp is found to be unstable, it has to be discarded and a new one has to be searched for. A tolerance threshold is considered in order to allow a temporary negative evaluation in some consecutive images of the sequence, which may be due inherent errors in the procedures for the contour extraction and/or the translation of the grasp.

Algorithm 3 Sequence grasp tracking

```
if no previously computed grasp available then
  search for a new grasp;
else
  translate grasp from previous to current object view;
  evaluate translated grasp;
  if negative grasp evaluation then
    search for a new grasp;
  end if
end if
```

Algorithm 4 outlines the grasp tracking procedure in the case of a sequence of images provided by a stereo pair. In this case, a grasp tracking is performed, using algorithm 3, on the sequence of images provided by one of the cameras of the stereo pair. For each of these images, if a grasp is found that grasp is translated to the image provided by the other camera of the stereo pair.

Algorithm 4 Stereo grasp tracking

```
track grasp on one of the images of the stereo pair (sequence grasp
tracking, see alg. 3)
if grasp available then
  translate grasp on the other image of the stereo pair;
end if
```

A. Grasp search

The selection of the grasp on the object depends on the configuration of the gripper used and, in general, on the context of the manipulation task. Strategies can be found in some previous works: for two-finger grippers in [6], [15], while for three-finger grippers in [13], [14].

Once the homography matrix H has been computed as indicated in section II-B, all the points available in image I' are translated into points in image I, and in particular the grasp points. Let $\{P_{g_i} = (u_{g_i}, v_{g_i}),\ i \in [1, n_g]\}$ be the image coordinates of the set of the n_g grasp points in image I'. The corresponding coordinates $\{p_{g_i} = (u_{g_i}, v_{g_i}),\ i \in [1, n_g]\}$ of these points in image I can be computed using (1).

B. Visual tracking control loop

The grasp tracking procedure described above can be used as a mechanism to be able to robustly use a set of grasp points as the set of control features within a visual servoing control system. This allows the execution of tasks such as the positioning of the gripper with respect to the grasp points selected on an object. The dof controllable by such a system are conditioned by the degrees of freedom under which the grasp tracking and, in turn, the grasp translation among subsequent views are possible.

Since the homography is invariant with respect to the movements of the robot arm, that is any combination of translations and rotations with respect to the x, y and z axes of the cameras, the grasp tracking method allows the manipulator to move without particular constraints in the free space, giving six dof to the visual servoing control loop.

IV. Results

The procedure descibed in the previous section has been experimentally verified by means of two different laboratory setups, available in Castellón and in Bologna and briefly described in Sect. I.

Since the tracking is actually performed between pairs of images, the vision system can consist either of a pair stereo or a single camera, forming in both cases an eye-in-hand configuration. Even though the strategy can be applied to either translation of grasps within a pair stereo of images or along a sequence of images, the results shown in this paper correspond to the case of a single camera. The vision system is considered uncalibrated so the internal parameters are not used in the computation of the homography. The grasp translation procedure has been used to translate grasps generated both for two-fingered and three-fingered grippers.

The object is not previously known and no model of it is available. It is assumend to be rigid and planar; ideally, it can be considered as a shape that lies on a given plane, so it is considered relatively flat.

The tracking of the grasp points during the movement of the robot avoids the grasp search at each time instant, reducing the computational cost of each iteration of the control loop, and it prevents the grasp search engine, specially under noisy sensor data,.to select different grasp points.

The results of the translation of the grasp points within the pairs of figures shown in Tab. I can be observed in Tab. II and in Tab. III. The corresponding points used to compute the homography, the lines that join the grasp points according to the geometry of the grippers, and the grasp points, belonging to the contour of the object are pointed out.

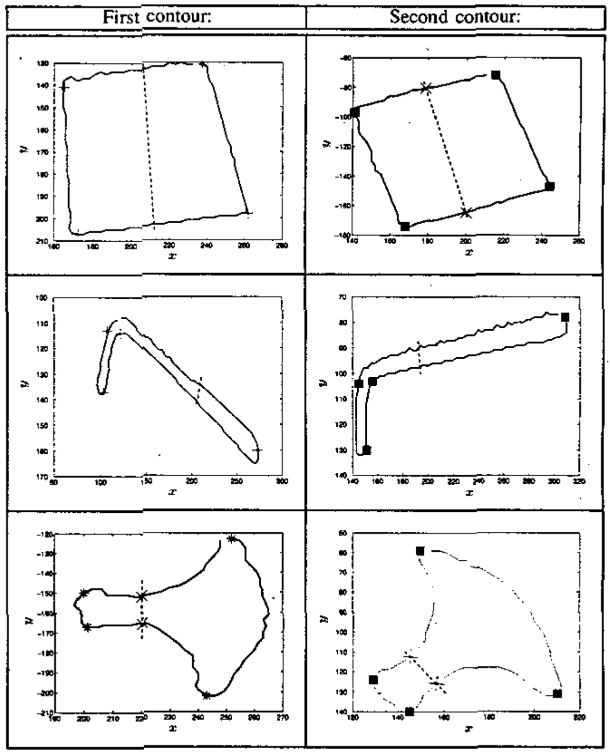
Tab. IV shows the application of the grasp tracking procedure based on the computation of a homography applied to a sequence of views of an object: the pairs of images that are taken into account for computing the homography and between which the grasp is translated are consecutive along the sequence. As it can be observed, the use of the homography provides an accurate tracking of the grasp.

V. Conclusions

In this paper, a method for the tracking of grasp points based on the homography concepts has been described. This strategy allows six dof motion of the robot and can be applied to either the translation of grasps within a stereo pair of images or along a sequence of images.

TABLE II

Grasp translation based on a homography (two fingered gripper).



The results have shown that it is possible to track grasps composed of a variable number of points independently of the particular gripper used. The strategy proposed could be combined with a mechanism to evaluate the translated grasp. However, it does not rely on any specific method for the selection or the evaluation of the grasp, as long as the grasp can be expressed as a set of points belonging to the contour of the object.

Future work will be devoted to a more extensive testing of the procedure, including a more structured integration of the grasp tracking mechanism within the loop of a visual servoing control system. Additionally, since only the external contour of the object has been considered here for the location and tracking of the grasp points, the use of the homography for tracking grasps on objects with internal holes will be considered, as well as the extension of the procedure to the 3D case.

Acknowledgment

This work is partially supported by projects DPI2001-3801, inf01-27, GV01-244, CTIDIA/2002/195, and P1-1B2001-28. The authors also wish to thank Youcef Mezouar and Ángel P. del Pobil for their helpful comments during the development of this work. R. Carloni thanks the University of Bologna and

TABLE III

GRASP TRANSLATION BASED ON A HOMOGRAPHY (THREE FINGERED GRIPPER).

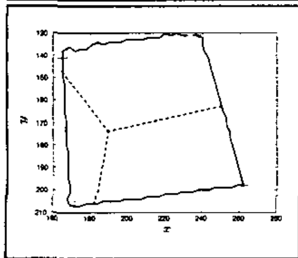
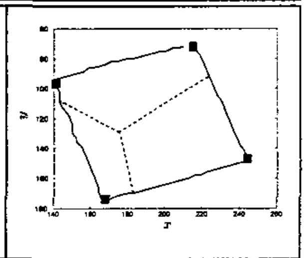
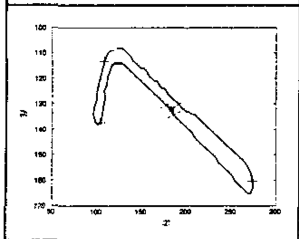
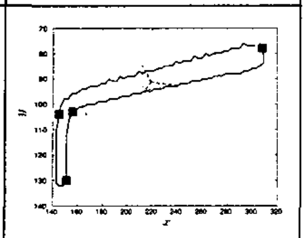
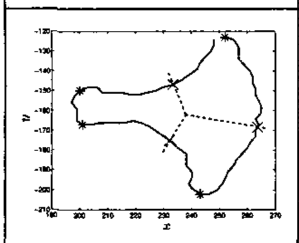
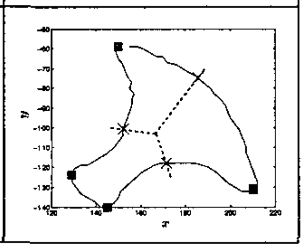
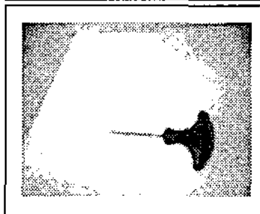
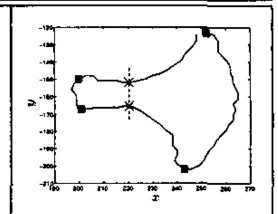
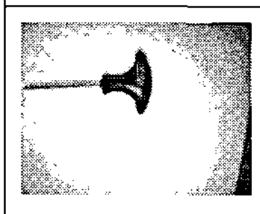
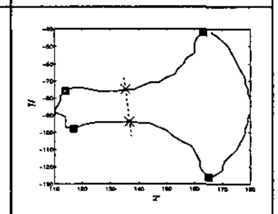
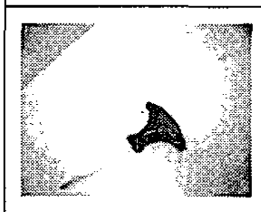
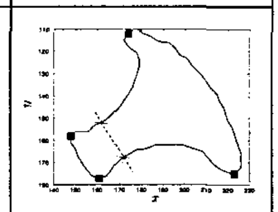
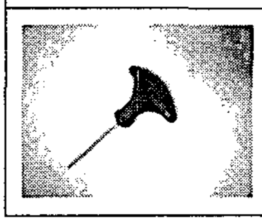
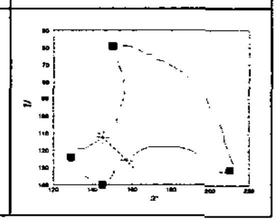
First contour:	Second contour:
	
	
	

TABLE IV

GRASP TRACKING BASED ON THE COMPUTATION OF A HOMOGRAPHY BETWEEN CONSECUTIVE CONTOURS.

Sequence of images:	Tracking of grasp points:
	
	
	
	

the University Jaume I for the financial support of the research period spent in Castellón.

REFERENCES

[1] J. Mundy and A. Zisserman, Eds., *Geometric Invariance in Computer Vision*. Cambridge (Massachussets), USA: The MIT Press, 1992.

[2] E. Malis, F. Chaumette, and S. Boudet, "Multi-cameras visual servoing," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, vol. 4, San Francisco (California), USA, Apr. 2000, pp. 3183–3188.

[3] P. K. Allen, A. Timcenko, B. H. Yoshimi, and P. Michelman, "Automated tracking and grasping of a moving object with robotic hand-eye system," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 2, pp. 152–165, Apr. 1993.

[4] G. Chesi, E. Malis, and R. Cipolla, "Automatic segmentation and matching of planar contours for visual servoing," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, vol. 3, San Francisco (California), USA, Apr. 2000, pp. 2753–2758.

[5] R. Horaud, F. Dornaika, and B. Espiau, "Visually guided object grasping," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 4, pp. 525–532, Aug. 1998.

[6] G. Recatalá, M. Sorg, J. Leupold, P. J. Sanz, and A. P. del Pobil, "Visual grasp determination and tracking in 2D dynamic scenarios," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, Lausanne, Switzerland, Oct. 2002, pp. 25–30.

[7] Q.-T. Luong and O. Faugeras, "The fundamental matrix: Theory, algorithms and stability analysis," *International Journal of Computer Vision*, vol. 17, pp. 43–75, 1996.

[8] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," INRIA, Sophia-Antipolis, France, Tech. Rep. 2273, May 1994.

[9] R. Cipolla and P. Giblin, *Visual Motion of Curves and Surfaces*. Cambridge, UK: Cambridge University Press, 1999.

[10] S. Glas, G. Recatalá, and M. Sorg, "Automatic reconstruction of silhouettes using B-splines," in *Image Analysis. 13th Scandinavian Conference, SCIA 2003. Proceedings*, ser. Lecture Notes in Computer Science 2749, J. Bigun and T. Gustavsson, Eds. Göteborg, Sweden: Springer-Verlag, June/July 2003, pp. 239–246.

[11] P. J. Sanz, A. P. del Pobil, J. M. Iñesta, and G. Recatalá, "Vision-guided grasping of unknown objects for service robots," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, Leuven, Belgium, May 1998, pp. 3018–3025.

[12] G. Recatalá, "Visual determination, tracking and execution of 2D grasps using a behavior-inspired approach," Ph.D. dissertation, University Jaume I, Castellón, Spain, 2003.

[13] A. Macchelli, C. Melchiorri, R. Carloni, and M. Guidetti, "Space robotics: an experimental set-up based on RTAI-Linux," in *4th Real-Time Linux Workshop*, Boston, USA, Dec. 2002.

[14] A. Morales, G. Recatalá, P. J. Sanz, and A. P. del Pobil, "Heuristic vision-based computation of planar antipodal grasps on unknown objects," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, Seoul, Korea, May 2001, pp. 583–588.

[15] D. Perrin, C. E. Smith, O. Masoud, and N. P. Papanikolopoulos, "Unknown object grasping using statistical pressure models," in *Proc. IEEE Intl. Conf. on Robotics and Automation*, San Francisco (California), USA, Apr. 2000, pp. 1054–1059.