



**HAL**  
open science

## Active rough shape estimation of unknown objects

C. Dune, E. Marchand, Christophe Collewet, Cédric Leroux

► **To cite this version:**

C. Dune, E. Marchand, Christophe Collewet, Cédric Leroux. Active rough shape estimation of unknown objects. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'08, 2008, Nice, France, France. pp.3622-3627. inria-00351864

**HAL Id: inria-00351864**

**<https://inria.hal.science/inria-00351864v1>**

Submitted on 12 Jan 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Active rough shape estimation of unknown objects

Claire Dune<sup>1,2</sup>, Eric Marchand<sup>1</sup>, Christophe Collwet<sup>1</sup>, Christophe Leroux<sup>2</sup>

**Abstract**—This paper presents a method to determine the rough shape of an object. This is a step in the development of a "One Click Grasping Tool", a grasping tool of everyday-life objects for an assistant robot dedicated to elderly or disabled. The goal is to determine the quadric that approximates at best the shape of an unknown object using multi-view measurements. Non-linear optimization techniques are considered to achieve this goal. Since multiple views are necessary, an active vision process is considered in order to minimize the uncertainty on the estimated parameters and determine the next best view. Finally, results that show the validity of the approach are presented.

## I. MOTIVATIONS

Much technology remains to be developed to assist people with disabilities. Robotic assistance may help these people to recover some manipulation capabilities by achieving actions such as carrying, grabbing, picking up and moving objects [11]. As the severity of the patient's handicap increases, the complexity of the devices to use increases. This requires significant efforts to simplify the interactions between the patient and these devices.

This paper presents our work and results produced on a "One Click Grasping Tool", a solution for grasping everyday life objects which only requires one click from the user. The aim of this work is to provide an intuitive tool able to seize objects presented in an unstructured manner and with no previously known shape [4]. The robotic system is a manipulator that can be mounted either on a wheelchair or on a mobile robot. It is equipped with a gripper and two cameras. One camera is mounted on the effector. The other one is mounted on the top of the wheelchair or on the mobile base and observes the scene. In a previous work [4] we have proposed a method that allows the user, through a single click, to select the object of interest in the remote camera view. After a search step, the effector camera is then automatically focused on the object.

However the system needs more information to achieve a relevant grasp of an unknown object in an unknown pose. This is the goal of this paper. It is assumed that a convex rigid object can be grasped by aligning the gripper with its minimal dimension while being perpendicular to its major axis. These are the features that have to be recovered from a set of images of the object. Let us advocate that our goal is not to build an accurate 3D object reconstruction but to

gather enough information to allow a manipulator to grasp it using visual servoing techniques.

Our approach is based on contour analysis and implicit 3D reconstruction methods. In this paper, the considered objects are modeled by quadratic functions. They have the property to project as conics that may be tracked over an image sequence. Then, the parameters of the quadric that fits the image data are estimated using non-linear minimization techniques. Depending of the taken views, some ambiguity may remain on the estimated parameters of the quadric. Hence, an active vision step is considered in order to minimize the uncertainty on the parameters estimation and determine the next best view.

## II. RELATED WORK

In the past decades grasping of unknown objects has been investigated. This kind of application raises several issues. The first one is the selection of the information that is used to perform the grasping. The second one deals with an "exploration" task that allows sensors to get relevant information.

Regarding image based methods for grasping an unknown object, [7] proposes to use stereo vision information to determine the grasping points. The object to grasp lies in a known scene. The object segmentation is performed by a simple image difference between the current view and the known background. Then the 3D skeleton of the object is computed and used to determine the grasping points. In [12] a method to seize unknown plane dark objects on a white table is proposed. The grasping points are determined using the 2D moments. In [6] the ellipsoid that best fits an object set on a textured background is found. A parametric motion estimation algorithm [10] is used to detect the object that is not in the major plane of the scene. Then, using a probabilistic scheme, the ellipsoid that is inside the intersection of all cones generated by the image contours is computed.

In [13] the grasping points of an object of unknown shape are defined by using a learned model data base. Five categories of object shapes and their grasping points are learnt. When a new object has to be grasped, it is associated to a category which determines its grasping points. Given these five models, a huge variety of objects are grasped in unstructured environments.

In [15] the object is represented by a set of super-ellipsoids. A laser sensor is servo in such a way to reduce the uncertainty on the estimated 3D functions. Even if it deals with the different field of object recognition, this approach can be related to ours. While this technique is based on the 3D points provided by a laser range finder, our technique

<sup>1</sup> INRIA Rennes-Bretagne Atlantique, Lagadic, Rennes, France

<sup>2</sup> CEA List, F-92265 Fontenay Aux Roses, France

This work has been done at IRISA-INRIA Rennes in collaboration with CEA List. It was also supported by CEA and by Brittany County Council under contribution to student grant.

used 2D information extracted from a set of views of the object.

Our method uses no a priori knowledge about the object except that it is roughly convex. We assume that the necessary features to perform a proper grasp are the following: the object major axis, its centroid position and its rough size. These features are contained by the quadric that best fits the object. To solve the first problem the parameters of this quadric are estimated in a determinist way using the 2D apparent contours extracted from the views of the object. The exploration problem is solved in a similar way to the solution proposed by [15]. The camera is controlled to reduce the uncertainty of the estimated parameters.

### III. MODELING ISSUE

This section proposes a method to describe an object using a compact representation based on the computation of the quadratic function that best fits 3D shape of the object. The aim is to find the major object axis orientation, the position of its centroid and its approximate scale. Such information is critical for a grasping task.

#### A. The quadric representation

Let  $\mathbf{h}({}^w\mathbf{X}, {}^w\mathbf{Q})$  be the quadric where  ${}^w\mathbf{X}$  is a 3D point expressed in the world frame  $\mathcal{R}_w$  and  ${}^w\mathbf{Q} = ({}^wq_0, \dots, {}^wq_9)^\top$  its parameters. Let us note that in the remainder of the paper  $\mathbf{Q}$  is expressed, if not specified, in  $\mathcal{R}_w$ . The equation of a quadric  $\mathbf{h}$  has the following expression:

$$\mathbf{h} = \mathbf{X}^\top \Gamma \mathbf{X}, \quad (1)$$

where  $\Gamma$  is the symmetric positive matrix associated to the quadric:

$$\Gamma = \begin{pmatrix} {}^wq_0 & {}^wq_3 & {}^wq_4 & {}^wq_6 \\ {}^wq_3 & {}^wq_1 & {}^wq_5 & {}^wq_7 \\ {}^wq_4 & {}^wq_5 & {}^wq_2 & {}^wq_8 \\ {}^wq_6 & {}^wq_7 & {}^wq_8 & {}^wq_9 \end{pmatrix} \quad (2)$$

#### B. Computation of the object features

Given the parameters  $\mathbf{Q}$  of the quadric in  $\mathcal{R}_w$ , the features needed for the grasping can be computed (see Fig. 1). The centroid of a quadric is the point  $\mathbf{X}_c$  where all the partial derivative  $\frac{\partial \mathbf{h}}{\partial X}$ ,  $\frac{\partial \mathbf{h}}{\partial Y}$ ,  $\frac{\partial \mathbf{h}}{\partial Z}$  are null. The inertia axes of the object can be found by considering the eigen space associated with  ${}^w\Gamma_q$ . The major axis direction is the eigen vector related to the largest singular value and the small axis is related to the smallest singular value.

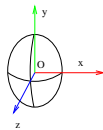


Fig. 1. Quadric Frame

#### C. Projection in the image plane

Using the perspective projection equations ( $x = X/Z$ ,  $y = Y/Z$ ) and given that the contour generator is a set of points of the quadric surface at which rays are tangent to the surface, it can be shown that the projection of a quadric in the image plane is a conic:

$$\mathbf{g}(\mathbf{x}, \mathbf{c}) = \mathbf{x}^\top \gamma \mathbf{x} \quad (3)$$

where the parameters  $\mathbf{c} = (k_0, k_1, k_2, k_3, k_4, k_5)^\top$  are related to the parameters of the quadric expressed in the camera frame as follows:

$$\begin{aligned} k_0 &= {}^c q_6^2 - {}^c q_9 {}^c q_0 & k_1 &= {}^c q_7^2 - {}^c q_9 {}^c q_1 \\ k_2 &= {}^c q_6 {}^c q_7 - {}^c q_9 {}^c q_3 & k_3 &= {}^c q_6 {}^c q_8 - {}^c q_9 {}^c q_4 \\ k_4 &= {}^c q_7 {}^c q_8 - {}^c q_9 {}^c q_5 & k_5 &= {}^c q_8^2 - {}^c q_9 {}^c q_2 \end{aligned} \quad (4)$$

Hence, we can compute the  $\mathbf{c}$  parameters of the projection of a quadric in the image plane knowing the position of the camera and the  $\mathbf{Q}$  quadric parameters. At the opposite, if we know a set of positions and the set of  $\mathbf{c}$  parameters of some conics measured for these positions of the camera, we can infer the parameters  $\mathbf{Q}$  of the corresponding quadric in the world frame.

### IV. ROUGH SHAPE ESTIMATION FROM MULTIPLE VIEWS

Given some views of an object and the positions from which they have been taken, we aim to find which quadric best fits the shape of the object. Figure 2 presents the general fitting scheme. The yellow ellipsoid is the quadric that best fits the real object, whereas the red ellipsoid is the currently estimated quadric. In each view, an image processing algorithm gives us a set of 2D points  $\mathbf{x}$  (in green) that are on the apparent contour of the object. Thus the fitting process consists in fitting the projection of the reel quadric with the green measured conic.

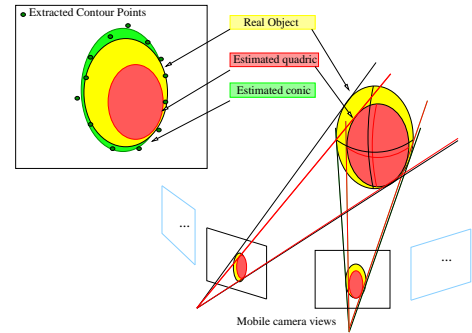


Fig. 2. Quadric fitting scheme

#### A. Conic parameters from 2D measurements

From a general point of view, given a set of points  $\mathbf{x}(x, y, 1)$  detected on the object apparent contour (see section VI), it is possible to estimate the parameters of the conic that best fits this set of points by solving a least square problem [16].

In every object view, we measure the best fitting conic of parameters  $\mathbf{c}$  and we store it along with the position from which the view has been taken as data set.

## B. Cost function

Given a set of initial estimated parameters  $\hat{\mathbf{Q}}$  and a set of positions, we compute the estimated projection  $\hat{\mathbf{c}}$ . We try to fit the real measurement  $\mathbf{c}$  with the estimated projection  $\hat{\mathbf{c}}$ . The cost function to minimize is thus:

$$f(\hat{\mathbf{Q}}) = \chi^2(\hat{\mathbf{Q}}, \mathbf{x}) = \sum_{ij} \frac{(\hat{\mathbf{c}}_{ij} - \mathbf{c}_{ij})^2}{\sigma_i} \quad (5)$$

where  $i \in [0, 5]$  is the index of the  $i^{\text{th}}$  parameter and  $j \in [0, N]$  the index of the  $j^{\text{th}}$  view. The cost function can also be written in a matrix form as the matrix  $\mathbf{F}$  of dimension  $\dim(\mathbf{F}) = (6N, 1)$  which rows  $i$  are  $\hat{c}_i - c_i$  and where  $N$  is the number of views we took. The corresponding Jacobian matrix is  $\mathbf{J} = \frac{\partial \mathbf{c}}{\partial \hat{\mathbf{Q}}}$  of dimension  $\dim(\mathbf{J}) = (6N, 10)$ .

Coefficients  $\hat{\mathbf{Q}}$  are estimated solving a non-linear system. Considering non noisy data, the cost function is quadratic in the neighborhood of  $\mathbf{Q}^*$  and its minimum value can be obtained using least square estimation techniques. To ensure the uniqueness of the solution we set the  $\mathbf{x}$  axis related coefficient to  $q_0 = 1$ , *i.e.*, we normalize all the coefficients by  $q_0 \neq 0^1$ .

There are several methods to update the estimated parameters during the optimization, among them, let us focus on the Gauss-Newton and the Levenberg-Marquardt approaches:

- Gauss-Newton:

$$\hat{\mathbf{Q}}_{k+1} = \hat{\mathbf{Q}}_k - \alpha_k \mathbf{H}_{\mathbf{k}}(\hat{\mathbf{Q}}_k)^{-1} \nabla f(\hat{\mathbf{Q}}_k) \quad (6)$$

- Levenberg-Marquardt:

$$\hat{\mathbf{Q}}_{k+1} = \hat{\mathbf{Q}}_k - (\mathbf{H}_{\mathbf{k}}(\hat{\mathbf{Q}}_k) + v_k \mathbf{D}_{\mathbf{k}})^{-1} \nabla f(\hat{\mathbf{Q}}_k) \quad (7)$$

In both cases  $\mathbf{H}_{\mathbf{k}}(\hat{\mathbf{Q}}_k) = \mathbf{J}(\hat{\mathbf{Q}}_k)^T \mathbf{J}(\hat{\mathbf{Q}}_k)$  is the Hessian of the cost function and  $\nabla f(\hat{\mathbf{Q}}_k)$  is defined by  $\nabla f(\hat{\mathbf{Q}}_k) = \mathbf{J}(\hat{\mathbf{Q}}_k)^T F(\hat{\mathbf{Q}}_k)$ .

On one hand, the Gauss-Newton step needs a linear estimation, in the downhill direction, of a criterion  $\alpha$  that can be time consuming while, on the other hand, the Levenberg-Marquardt method requires the computation of the parameter  $v_k$ .

When dealing with noisy data, computing the downhill direction given by  $\mathbf{H}^{-1} \mathbf{J}^T \mathbf{F}$  of the Gauss-Newton method becomes a real issue as soon as the Hessian matrix becomes ill conditioned and lead to a divergence of the minimization. For its part, Levenberg-Marquardt algorithm may require many iterations leading to a slower convergence. Thus we propose to use a mixed method between Levenberg-Marquardt and Gauss-Newton. It uses a step of Levenberg-Marquardt when the matrix is ill conditioned otherwise a Gauss-Newton step is considered.

An estimation error may also arise from the 2D conics measured in the views. To cope with the views for which the conics are ill measured, due to the noise, we propose either to use a robust estimation technique [9] to give a higher weight for the "good views" and a lower weight for the "outliers

<sup>1</sup>This can be done without loss of generality because the cases where  $q_0 = 0$  are degenerate quadric that cannot represent a real 3D object.

views" or to remove the views that are the more ambiguous at the end of the estimation process.

[15] shows that at a given camera position, and for some estimated parameters  $\hat{\mathbf{Q}}$ , the uncertainty  $U({}^w \mathbf{M}_{\mathbf{c}}, \hat{\mathbf{Q}})$  on the parameters  $\mathbf{Q}$  is:

$$U({}^w \mathbf{M}_{\mathbf{c}}, \hat{\mathbf{Q}}) \propto \sqrt{\mathbf{J}^T \mathbf{C} \mathbf{J}} \propto \sqrt{\mathbf{J}^T \mathbf{H}^{-1} \mathbf{J}} \quad (8)$$

Hence, we can use the  $U$  value either to remove the views that are the more ambiguous or use a robust estimation over all the views.

## V. CLOSING THE LOOP: COMPUTING OPTIMAL CAMERA MOTIONS

We have shown that the  $\hat{\mathbf{Q}}$  parameters of the currently estimated quadric can be inferred from several views of an object. However it can happen that the current views do not provide enough information to accurately compute the parameters. For example, when the successive positions of the camera are too close to each other, the information on the object depth is not reliable and the parameters related to it may be ill estimated although the measured and estimated conics fit perfectly in each image plane.

Thus, in order to refine the estimation of the parameters, more views need to be taken. The issue is then to find a criterion that can help us to compute the best direction for the camera motion. We now look into active reconstruction methods.

As shown in [15], it is possible to use the uncertainty on the estimated parameters as a criterion to control the camera displacement. The Shannon entropy can be related to covariance matrix  $\mathbf{C}$  of the estimated parameters. In [15] the authors show that the best next direction can be obtained by minimizing  $\det(\mathbf{C})$ , *i.e.* by finding the view where the uncertainty decrease the more or the parameters vary the more.

The determinant of the covariance matrix for the iteration  $k$  in the optimisation process can be defined as:

$$\det(\mathbf{C}_{\mathbf{k}}) = \det(\mathbf{H}_{\mathbf{k}})^{-1} \quad (9)$$

Besides, [15] shows that the Hessian matrix can be computed as the sum of the Hessian for each view:

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \mathbf{J}_{k+1}^T \mathbf{J}_{k+1} \quad (10)$$

where  $\mathbf{J}_{k+1}$  is the Jacobian matrix associated to the next tried view. It is directly related to the position  ${}^w \mathbf{M}_{ck+1}$  and the currently estimated parameters  $\hat{\mathbf{Q}}$ .

Then the next best location is

$${}^w \mathbf{M}_{ck+1} = \arg \max_{{}^w \mathbf{M}_{\mathbf{c}}} (\det(\mathbf{H}_{\mathbf{k}} + \mathbf{J}_{k+1}^T \mathbf{J}_{k+1})) \quad (11)$$

The relation between  $\det(\mathbf{C})$  and the uncertainty is true in the neighborhood of the current sensor position, where the linear approximation of the function is true [15]. The mobile camera velocity is computed so as to move in the direction of the next best view. Yet it can be noticed that as soon as one parameter is well estimated its related eigen value is zero. Consequently the determinant  $\det(\mathbf{C})$  is also zero even if the

other parameters are badly estimated. Hence we propose to use an other criterion that is the Froebenius norm of  $\mathbf{C}$ , that is the sum of the squared eigen values of the positive symmetric  $\mathbf{C}$  matrix. Besides, computing this criterion is less time consuming than computing the determinant. Some results are shown in the next section. The determination of the next camera location is a non linear minimization problem without analytic Jacobian computation. To tackle this issue, we have chosen to use the Simplex method of Nelder and Mead. This method only requires some evaluations of the cost function at some points that are the vertices of the current Simplex. Using this approach, only the translational velocity is computed. The rotational velocity is computed so as to keep the projection of the centroid of the estimated quadric in the center of the image plane by using a visual servoing scheme [5].

## VI. DETECTION OF THE OBJECT CONTOUR IN THE IMAGE PLANE

We proposed here to use active contours to extract the points required to compute the 2D conics as described in section IV-A. Indeed, since their first introduction [8], active contours have proven to be very powerful in various contexts, including edge detection and target tracking.

They are deformable curves which are defined in an image domain and move under the influence of internal forces coming from the curve itself and external forces generated from the image (typically gradient information for edge detection). We consider here a curve parameterized by a vector  $\mathbf{a}$ :  $\mathcal{C}_{\mathbf{a}}(\mathbf{x}(u))$ . This curve moves according to the evolution of the components of  $\mathbf{a}$  [14], [1], [2] while, with traditional methods (*i.e.* points-based), the curve directly moves from points, then these latter approaches are local. Conversely, using a parametric formulation  $\mathcal{C}_{\mathbf{a}}$  implicitly introduces global constraints. Consequently, they are usually more robust, especially when  $\dim(\mathbf{a})$  is low. They are also more suitable for control issue. Indeed, the computation duration is greatly reduced.

We define  $\mathcal{C}_{\mathbf{a}}(\mathbf{x}(u))$  by a polar description:

$$\mathcal{C}_{\mathbf{a}}(u) : \mathbf{x}(u) = \mathbf{x}_c + \rho_{\mathbf{a}}(u) \begin{pmatrix} \cos u \\ \sin u \end{pmatrix} \quad (12)$$

with  $u \in [0, 2\pi]$ ,  $\mathbf{x}_c$  being a point inside the contour. Besides, we perform a Fourier expansion of the radius  $\rho_{\mathbf{a}}(u)$ :

$$\rho_{\mathbf{a}}(u) = a_0 + \sum_{k=1}^{k=h} a_k \cos ku + b_k \sin ku \quad (13)$$

where  $h$  is the number of harmonics. Thus, we have here  $\mathbf{a} = (a_0, \dots, a_h, b_1, \dots, b_h)$ .

To derive the evolution equations of  $\mathbf{a}$ , the problem is set as the minimization of the curve energy  $E(\mathbf{a})$  with respect to the parameters  $\mathbf{a}$ , leading to the following simplified associated Euler-Lagrange equation:

$$\gamma_i \frac{da_i}{dt} + \frac{\partial E}{\partial a_i} = 0 \quad (14)$$

where the curve is supposed to be massless and embedded in a viscous medium of viscosity  $\gamma_i$ .

In our case, we only consider an external energy term  $E(\mathbf{a}) = E_{ext}(\mathbf{a})$ , that we write:

$$E_{ext} = - \int_{\mathcal{C}_{\mathbf{a}}} \nabla I(\mathbf{x}(u))^\top \mathbf{n}(u) du. \quad (15)$$

where  $\mathbf{n}(u)$  is the outward unit normal of  $\mathcal{C}_{\mathbf{a}}$  in  $u$ . Note that  $\nabla I(\mathbf{x}(u))$  must also be computed as an outward vector. To be more robust against the initialization of  $\mathcal{C}_{\mathbf{a}}$ , we also consider a similar pressure force as described in [3]:  $\mathbf{f}_p = k_1 \mathbf{n}(u)$  (with  $k_1$  a scalar).

## VII. EXPERIMENTAL RESULTS

This section presents a typical execution of the application presented above. The experimental set up is shown in Figure 3. The mobile camera is mounted on the end effector of a 6-DoF robot arm. Although we do not use any a priori knowledge about the object or about the scene, the method uses some heuristics. Firstly, the object is assumed to be roughly convex. Secondly, the object can be segmented from the background using an active contour.

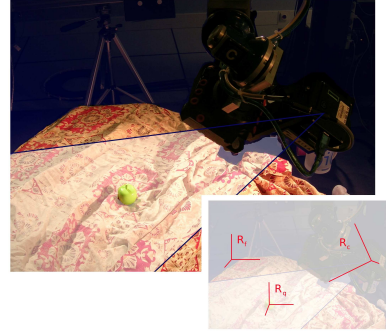


Fig. 3. Experimental set up: a mobile manipulator equipped with an eye in hand camera. An object which shape is a priori unknown lies in a scene with a textured background.

### A. Object segmentation and conic feature in 2D image.

The object is segmented from the background using the method presented in Section VI. One of the main issue when using active contours to detect an object is the initialisation step. For the purpose of our application, the aim would be to initialise the snake using only the user click on the image of the remote view. In a previous work [4], we have shown how to estimate the position of an object in the world frame from one click of the user. Using this estimation and the current position of the camera, the estimated position of the object can be projected in the mobile view and the neighborhood of this projection can be used as the initialisation for the snake.

Figure 4 shows that the segmentation is robust even in the case of a cluttered background. The two images of the top show the snake itself and the two other images show the conic fit. Even if the object is not an exact quadric, we can notice that the long axis of the conic is approximately the object long axis and that the size can be roughly estimated.

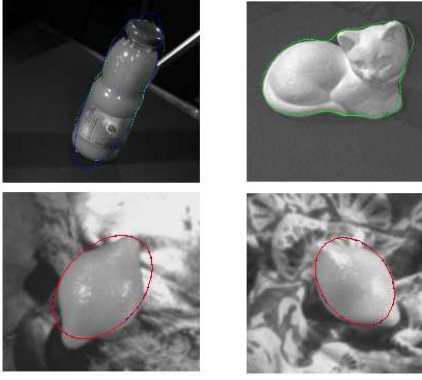


Fig. 4. Some results of contour extraction and conic fitting using active contours

### B. Quadric estimation

Assuming an initial rough estimation of the position of the object using the approach described in [4], the initial set of  $\hat{\mathbf{Q}}$  parameters are those of a sphere of radius the width of the gripper (8cm) and centered on the estimated position of the object. The projection of the estimated location can also be used to create the seed for the active contour in the effector camera view.

Considering a motion of the camera on a view sphere around the object, images are acquired, conics are extracted and finally a set of  $\hat{\mathbf{Q}}$  parameters is computed.

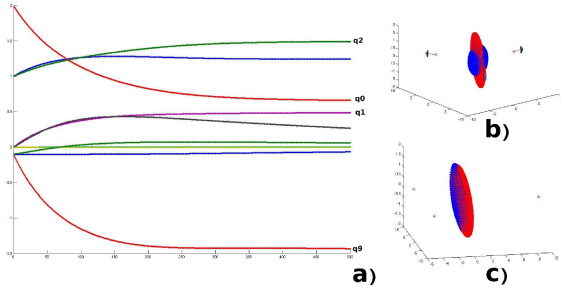


Fig. 5. a) Evolution of the quadric parameter estimation over the time for simulation data. b) Initial state: The camera positions are represented as small pyramids, the initial guess is represented in blue and real object in red. c) Final state. The final estimation (in blue) fits the real object.

Figure 5 shows the evolution of the  $\hat{\mathbf{Q}}$  parameters over the time for simulation data. It shows the initial guessed state and the final state computed using the projection of the real object in the camera views surrounding the object. The curves are smooth and the parameters of the quadric converges towards the real values. The left part of Figure 8 wrt the dotted line shows the estimation of the parameters of the quadric for real data. Because of the measurement noise, there is more overshoot before the optimisation process converges to the best values. Table I shows some results for different optimisation methods. The Levenberg-Marquardt method and our method give better results than the Gaussian estimation process.

Method	final cost	# iteration	norm of error
Levenberg-Marquardt	7.35e-13	7	0.089
Gauss-Newton	2.87e-10	101	0.126
GN robust	8.64e-06	101	0.034
LM robust	5.20e-06	101	0.191
Considered approach	5.60e-13	27	0.081

TABLE I

RESULTS OBTAINED ON QUADRIC PARAMETER ESTIMATION FROM 6 VIEWS USING VARIOUS MINIMIZATION TECHNIQUES

Although correctly estimated from the available data, some ambiguities may remain on the quadric parameters especially when the views are close to each other, the object depth may be ill defined.

### C. Selection of the next best direction

In order to reduce the remaining ambiguities in the quadric parameters estimation, we need to take more measurements, and to move the camera as described in section V. The next camera motion is computed in order to reduce the entropy on the parameters set. To do so, the camera is controlled toward the region where the parameters are the most likely to change *i.e.* region where the ambiguity is high.

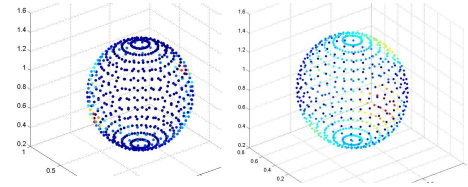


Fig. 6. The object has been seen from several views and a quadric that best fits its 3D shape has been estimated. However some ambiguities remain on the parameters and another view need to be taken. Those two vision spheres show the location where the ambiguities are the highest in hot colour. the left vision sphere is obtained by using the determinant criterion whereas the right one is obtained using the sum criterion.

Figure 6 highlights the regions on a view sphere where the ambiguity of the parameter set is high (hot colors) and those where the ambiguity is low (cold colors). The left sphere depicts the determinant of the estimated Hessian if the camera were moved on this position at the next step, *i.e.* the product of the eigen values of  $\mathbf{H}$ . The right sphere represents the Frobenius norm of the eigen values of  $\mathbf{H}$  if the camera were moved to these locations. We can see that many regions are in dark blue in the left sphere. This is due to the fact that a view is considered to have a low uncertainty as soon as one eigen value comes to zero whereas the other can still be high. Using the sum of the eigen values instead of the products tackles this issue.

Figure 7 illustrates the motion of the camera using both criteria. The use of the determinant results in an erratic path whereas the use of the sum results in a smooth path on the view sphere. The parameters of the quadric are found within approximately 10 views.

Figure 8 shows the evolution of the cost function and parameters over the time when adding new views. Adding

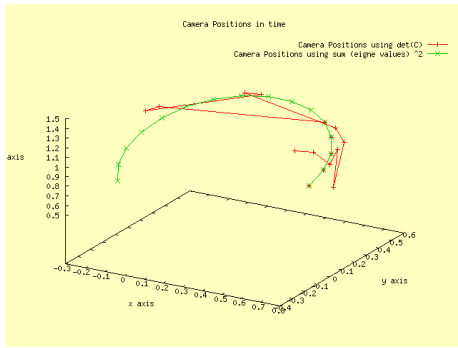


Fig. 7. Camera motion along the reconstruction process. The red path is the one found using the determinant criterion and the green path is the one found while using the sum of the eigen values criterion

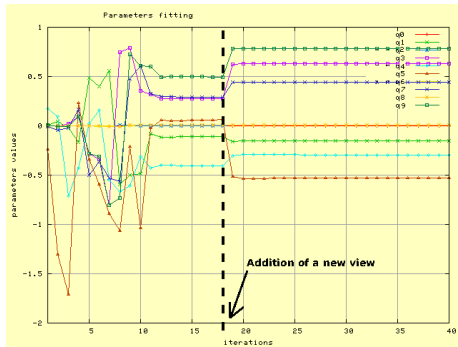


Fig. 8. Evolution of the quadric parameters estimation over the time for real data. The dotted line represents the time where an additional view has been taken. Every parameter changes until a minimum is reached.

new views leads to find a better estimation of the parameters  $\hat{Q}$ . If some ambiguity remains new views may be taken until the  $\hat{Q}$  parameters do not vary any more.

Along with the estimation of the quadric parameters, the object frame can be computed. Figure 9 shows some results obtained for different objects. The blue array is the major object axis, while the red array is the small object axis. The estimated quadric is projected in the view as the yellow conic. The red ellipse is the currently measured conic. Knowing the object frame we can perform a 3D servoing to align the gripper with the object frame and grasp it.

## VIII. CONCLUSION

In this paper, we have proposed an efficient solution to estimate the shape of an unknown object in order to grasp it with a robotic system. The object shape is approximated by a quadric surface whose parameters are estimated using multiple views of the scene. During the acquisition phase, the robot motions are continuously optimized in order to minimize the uncertainty on the estimated parameters by selecting the next best view. Wrt many classical solutions, our reconstruction process is directly based on the contours of the object in the image. The solution is thus very fast, allowing to compute the object shape in real time and to use it in a closed-loop grasping task with the robot. The proposed solution is fully generic and works for any roughly convex

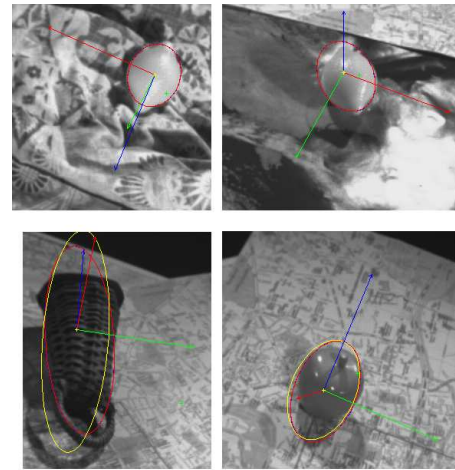


Fig. 9. Results of the optimisation process. The object frames are projected in the last view along with the currently measured conic (in red) and the estimated conic in yellow. The blue array is the major axis and the red array is the minor axis of the object.

object.

## REFERENCES

- [1] B. Bascle, P. Bouthemy, N. Deriche, F. Meyer. Tracking complex primitives in an image sequence. In *ICPR'94*, pp. 426–431, Jerusalem, Oct. 1994.
- [2] P. Brigger, J. Hoeg, M. Unser. B-spline snakes: a flexible tool for parametric contour detection. *IEEE T. on Image Processing*, 9(9):1484–1496, Sept. 2000.
- [3] L.D. Cohen, I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE PAMI*, 15(11):1131–1147, Nov. 1993.
- [4] C. Dune, E. Marchand, C. Leroux. One click focus with eye-in-hand/eye-to-hand cooperation. In *IEEE ICRA'07*, pp. 2471–2476, Rome, , Apr. 2007.
- [5] B. Espiau, F. Chaumette, P. Rives. A new approach to visual servoing in robotics. *IEEE T. on Rob. and Aut.*, 8(3):313–326, June 1992.
- [6] G. Flandin, F. Chaumette. Visual data fusion for objects localization by active vision. In *ECCV'02*, pp. 312–326, Copenhagen., May 2002.
- [7] A. Hauck, M. Rttinger, J. Sorg, G. Frber. Visual determination of 3d grasping points on unknown objects with a binocular camera system. In *IEEE ICRA'99*, pp. 272 – 278, 1999.
- [8] M. Kass, A. Witkin, D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1987.
- [9] E. Malis, E. Marchand. Experiments with robust estimation techniques in real-time robot vision. In *IEEE/RSJ , IROS'06*, pp. 223–228, Beijing, Oct. 2006.
- [10] J.-M. Odobez, P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, Dec. 1995.
- [11] World Health Organization. World health organization assessment, classification and epidemiology group, international classification of functioning and disability. World Health Organization, July 1999.
- [12] P.J. Sanz, A. Requena, J.M. Iesta, A.P. del Pobil. Grasping the not-so-obvious: vision-based object handling for industrial applications. *IEEE Robotics and Automation Mag.*, 12(3):44–52, Sep. 2005.
- [13] A. Saxena, J. Driemeyer, N. Andrew. Robotic grasping of novel objects using vision. *IJRR*, 27:157–173, 2008.
- [14] D. Terzopoulos, D. Metaxas. Dynamic 3d models with local and global deformations: deformable superquadrics. *IEEE PAMI*, 13:703–713, May 1991.
- [15] P. Whaite, Ferrie F.P. Autonomous exploration driven by uncertainty. In *IEEE PAMI*, pp. 193–205, 1997.
- [16] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing*, 15(1):59–76, Jan. 1997.