HAL open science

Integrating grasp planning and visual servoing for automatic grasping

Radu Horaud, Fadi Dornaika, Christian Laugier, Christian Bard

To cite this version:

Radu Horaud, Fadi Dornaika, Christian Laugier, Christian Bard. Integrating grasp planning and visual servoing for automatic grasping. Oussama Khatib and J. Kenneth Salisbury. Experimental Robotics IV: The 4th International Symposium, 223, Springer Berlin / Heidelberg, pp.71–82, 1997, Lecture Notes in Control and Information Sciences, 10.1007/BFb0035198 . inria-00590074

HAL Id: inria-00590074

https://inria.hal.science/inria-00590074

Submitted on 31 Oct 2023

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.

Public Domain

Integrating Grasp Planning and Visual Servoing for Automatic Grasping*

Radu Horaud, Fadi Dornaika, Christian Bard, and Christian Laugier

GRAVIR–IMAG & INRIA Rhône-Alpes
655, avenue de l'Europe
38330 Monbonnot Saint Martin, FRANCE
Radu.Horaud@imag.fr

Abstract

In this paper we describe a method for aligning a robot gripper — or any other end effector — with an object. An example of such a gripper/object alignment is grasping. The task consists of, first computing an alignment condition, and second servoing the robot such that it moves and reaches the desired position. A single camera is used to provide the visual feedback necessary to estimate the location of the object to be grasped, to determine the gripper/object alignment condition, and to dynamically control the robot's motion. The original contributions of this paper are the following. Since the camera is not mounted onto the robot it is crucial to express the alignment condition such that it does not depend on the intrinsic and extrinsic camera parameters. Therefore we developp a method for expressing the alignment condition (the relative location of the gripper with respect to the object) such that it is *projective invariant*, i.e., it is view invariant and it does not require a calibrated camera. The central issue of any image-based servoing method is the estimation of the image Jacobian. This Jacobian relates the 3-D velocity field of a moving object to the image velocity field. In the past, the exact estimation of this Jacobian has been avoided because of the lack of a fast and robust method to estimate the pose of a 3-D object with respect to a camera. We discuss the advantage of using an exact image Jacobian with respect to the dynamic behaviour of the servoing process. From an experimental point of view, we describe a grasping experiment involving image-based object localization, grasp planning, and visual servoing.

1. Introduction and background

One of the most common operations in Robotics is grasping. Although the importance of grasping has been recognized for many years now, there are only a few grasping systems that can operate in a complex environment. This is mainly due to

*The work described herein has been supported by the European ESPRIT-III programme through the SECOND project (Esprit-BRA No. 6769).

the difficulty to execute precise robot hand motions in the presence of various perturbations: the robot's kinematic is known only partially, unpredictable obstacles may be located in the neighbourhood of the object to be grasped, and the location of the object to be grasped may not be known in advance. Moreover, the task of grasp planning itself is a complex one because the planner has to analyse a large number of situations.

One of our research goals is to demonstrate the advantage of integrating planning, sensing, and control in a number of robot tasks. Among these tasks, we have chosen grasping because of its importance in a number of applications: nuclear instrument manipulation, material handling, automated assembly, space robotics, etc.

Our approach to perform automatic grasping is as follows. A grasp plan is built off line [1]. This plan takes advantage of a priori knowledge such as the robot direct and inverse kinematics, the shape and the location of the object to be grasped, the shape of the gripper, etc. The output of this off line grasp planning process is the selection of a grasp, more precisely, the rigid transformation that maps the object's reference frame onto the gripper's reference frame.

Given such a grasp the problem to be solved on line is to control the robot's motion such that the final gripper/object alignment is as close as possible to the previously computed rigid transformation. In this paper we describe a *visual servoing* method for controlling the robot's motion (and hence, the gripper's motion) such that final goal position just described is properly reached. For that purpose, a camera is fixed in such a position and orientation that it sees both the gripper in some initial position and the object to be grasped.

The object-to-robot relationship is not known, the robot's kinematics is not perfectly known, and camera calibration is an almost impossible task; therefore, the robot's motion must be guided by some *relative rather than absolute* sensor measurements. To summarize, visually guided grasping proceeds in the following steps:

1. Locate the object to be grasped in some camera centered frame,
2. Predict the final position of the gripper in the image,
3. Detect the current gripper position in the image,
4. Attempt to align the gripper and the object by minimizing the discrepancy between the current gripper position in the image and its final image position (image-based servoing).

The problem of determining the position and orientation (location) of an object with respect to a camera is a classical and well studied one in computer vision. Recently we came up with an algorithm [10] which computes object location in 2 milliseconds. Because the gripper-to-object transformation to be reached is known in advance, one is able to combine these two relationships (gripper-to-object and object-to-camera) and to project gripper features onto the image. These image predictions amount to a goal position to be reached by the gripper in order to align with the object. The task of moving the robot such that it's gripper reaches this goal position is carried out by a visual servoing algorithm. Many such algorithms have been proposed in the literature [4], [9], [13], [8], [5], [3], [2], [15]. Among them we

found that the image-based visual servoing method developed by Espiau, Chaumette and Rives [4] is a very powerful one. With respect to the method cited above [4], we modified it such that (i) we were able to deal with a camera that was not rigidly attached to the robot being controlled and (ii) we showed that dynamic estimation of the hand-to-camera relationship improved the behaviour of the visual servoing algorithm [12].

The remainder of this paper is organized as follows. In section 2 we show how to represent an alignment between two objects in 3-D projective space. The alignment condition thus derived is projective-invariant in the sense that it can be used in conjunction with an un-calibrated camera (nor the intrinsic neither the extrinsic camera parameters are known) as a goal position for visual servoing. The method is inspired by recent work in computer vision and has some similarities with [7]. The major difference is that, with our approach, one camera is sufficient for carrying out the servoing while with the approach proposed in [7] two cameras are necessary. In section 3 we overview the visual servoing method which is an extension of the method described in [4]. Finally section 4 describes an experiment performed with a visually guided robot.

2. Projective-invariant alignment

Whenever one wants to align two objects (an object and a gripper, a peg and a hole, etc.) it is necessary to characterize somehow the relationship between these two objects. In Robotics this relationship is traditionally represented by some Euclidean transformation. In this section we show how to characterize alignments, like the ones just mentioned, such that the characterization is non Euclidean and hence is view-invariant. We will be considering alignment conditions associated with an un-calibrated camera and with the classical pin-hole model associated with the camera.

First we consider 6 points onto the 3-D object to be grasped, e.g., Figure 1. The approach can be easily generalized to any number of points provided that there are at least 6 points and that 5 among these points are in general position (i.e., no 4 among them are coplanar). This object has a Cartesian frame associated with it and the 3-D coordinates of the 6 (or more) object points are known in this frame – the object frame. We also consider a robot gripper and a number of points onto this gripper. The coordinates of these gripper points (or features) are known in some Cartesian gripper centered frame – the gripper frame.

The object-to-gripper Euclidean alignment condition states that the required 3-D rigid displacement D (a 4×4 homogeneous matrix) between the gripper frame and the object frame is known. We show now how to represent this alignment condition in the 3-D projective space rather than in the 3-D Euclidean space. We claim that this projective representation of the alignment condition is projective invariant because it is straightforward to map it onto any image *without calibrating the camera.*

Second we consider a 4×4 invertible matrix P that describes the mapping between the object Cartesian frame and the 3-D projective space. Matrix P is defined up to a scale factor and can be easily computed as follows. Let $\mathbf{A}_1$ through $\mathbf{A}_5$ be 4-vectors (i.e., homogeneous coordinates in the object frame) associated with 5 object points in general position and let $\mathbf{A}_6$ be the 4-vector associated with the sixth object point.

We denote by $\mathbf{A}_i^p$ the 4-vectors associated with the coordinates of the same

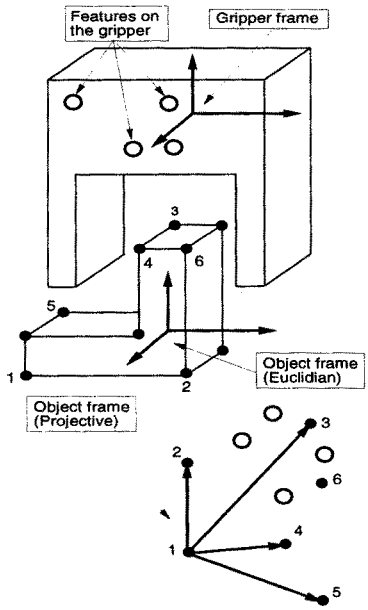

Figure 1. The rigid transformation that characterizes the grasping may be mapped onto a 3-D projective space.

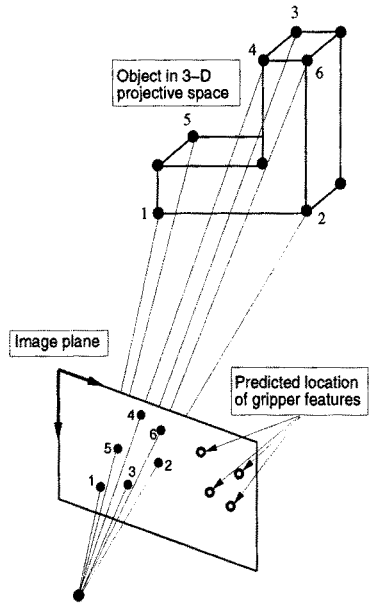

Figure 2. Object and gripper points may be mapped from the 3-D projective space onto an image, provided that at least 6 image-to-object point correspondences are known.

points in projective space. In this projective space let the first 5 points form a projective basis and hence have the following canonical coordinates associated with them:

$$(0\,0\,0\,1)\ \ (1\,0\,0\,1)\ \ (0\,1\,0\,1)\ \ (0\,0\,1\,1)\ \ (1\,1\,1\,1)$$

Every projective mapping:

$$\mathbf{A}_i^p = \lambda\, P\, \mathbf{A}_i \tag{1}$$

provides 3 linear constraints for the entries of P. Therefore, the 16 entries of P may be determined using 5 such point mappings (15 linear constraints) and an additional constraint. One possibility is to use the following constraint:

$$\sum_{i,j=1}^{4} p_{ij} = 1$$

Matrix P can therefore be determined using first 5 object points and their canonical projective coordinates. Once P is determined one can map the sixth object point (or any other point) onto the projective space spanned by the projective basis just defined.

Moreover, let $\mathbf{B}_j$ be a 4-vector associated with a gripper point. One can easily determine the 4-vector $\mathbf{B}_j^p$ which represents its projective coordinates in the canonical projective basis (D is the rigid transformation from gripper frame to object frame):

$$\mathbf{B}_j^p = \lambda\, P\, D\, \mathbf{B}_j$$

The sets of 4-vectors $\mathbf{A}_1^p$ through $\mathbf{A}_6^p$ and $\mathbf{B}_j^p$ ($j \geq 4$) constitute the projective invariant representation of the alignment condition. The reason for which there must be a minimum of 4 gripper points is because 4 object-to-image point matches constrain the position and orientation of the object with respect to the camera.

Finally, we consider a camera that "sees" the object to be grasped in an unknown position and orientation – neither the camera-to-object nor the camera-to-robot relationships are known, e.g., Figure 2. Let $\mathbf{a}_1$ through $\mathbf{a}_6$ be the images of the 6 object points. As already mentioned, the imaging process is a projective mapping from 3-D to 2-D and let M be a 3×4 matrix describing this mapping:

$$\mathbf{a}_i = M\ \mathbf{A}_i^p \tag{2}$$

with $\mathbf{a}_i = (su_i\ sv_i\ s)^T$ where u_i and v_i are the image coordinates of $\mathbf{a}_i$ and s is a scale factor. It is well known that at least 6 pairs of 3-D/2-D point correspondences are needed to completely define M. Therefore, the 3-D projective coordinates of the 6 object points together with their 2-D image coordinates completely define the matrix M.

Once the mapping M is determined it is straightforward to predict image coordinates for the gripper points by projecting their view invariant coordinates onto the image:

$$\mathbf{b}_j = M\ \mathbf{B}_j^p \tag{3}$$

It is important to stress the fact that the estimation of the projective mapping M outlined above does not amount for camera calibration. Indeed, matrix M maps points from a 3-D projective space onto the image and ***not*** from a 3-D Euclidean space onto the image. Therefore, the alignment condition as defined herein has no metrics associated with it.

In practice, the computations above are decomposed into two steps. The first step (estimation of $\mathbf{A}_1^p$ through $\mathbf{A}_6^p$ and $\mathbf{B}_j^p$) is performed off-line and the second step (estimation of M and of $\mathbf{b}_j$) is performed on-line. Notice that the numerical computations associated with the estimation of M are particularly simple because of the use of the canonical coordinates for 5 among the 6 object points.

The major difficulty associated with the approach described in this section is to find 6 matches between points onto the object to be grasped and image points. This matching problem amounts to object recognition from a single view and is a current research issue in its own right.

3. Visual servoing

In this section we consider a camera that observes a moving robot gripper. First we determine the image Jacobian associated with such a configuration. Second we define a visual servoing process that allows the camera to control the robot motion such that the gripper reaches a previously computed image position – one way to compute such an image position is, for example, to use the view-invariant alignment condition.

Let, as before, $\mathbf{B}_j$ be a 3-D point onto the robot gripper and let x_j, y_j, and z_j be its coordinates in the camera centered Cartesian frame. The projection of this point onto the image has as coordinates:

$$u_j \;=\; \alpha_u \frac{x_j}{z_j} + u_c \tag{4}$$

$$v_j = \alpha_v \frac{y_j}{z_j} + v_c \tag{5}$$

where α_u, α_v, u_c, and v_c are the well known intrinsic camera parameters associated with a pin-hole model.

Let's assume that the gripper moves in the space and that the translational velocity of the gripper frame origin is $\mathbf{V}_O$ and the angular velocity associated with the gripper frame is Ω_O. Let us denote by $\mathbf{T}_O$ this 3-D velocity screw (a 6-vector) expressed in the gripper frame and by $\mathbf{T}_c$ the 3-D velocity screw associated with the same gripper motion but expressed in the camera frame. The relationship between these velocity screws is:

$$\begin{pmatrix} \mathbf{V}_c \\ \Omega_c \end{pmatrix} = \begin{pmatrix} \mathbf{R} & -\mathbf{R}S(-\mathbf{R}^T\mathbf{t}) \\ 0 & \mathbf{R} \end{pmatrix} \begin{pmatrix} \mathbf{V}_O \\ \Omega_O \end{pmatrix} \tag{6}$$

In this equation $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and translation vector associated with the gripper-to-camera rigid transformation and $S(\mathbf{a})$ is the skew-symmetric matrix associated with a 3-vector $\mathbf{a}$. By computing the time derivatives of u and v in equations (4) and (5), we obtain the relationship between the velocity of a gripper point $\mathbf{B}_j$ and its image velocity:

$$\begin{pmatrix} \dot{u}_j \\ \dot{v}_j \end{pmatrix} = \begin{pmatrix} \alpha_u/z_j & 0 & -\alpha_u x_j/z_j^2 \\ 0 & \alpha_v/z_j & -\alpha_v y_j/z_j^2 \end{pmatrix} \begin{pmatrix} \dot{x}_j \\ \dot{y}_j \\ \dot{z}_j \end{pmatrix}$$

If one applies the well known law of rigid motion: $\mathbf{V}_{B_j} = \mathbf{V}_C + \Omega_C \times \overrightarrow{CB_j}$ relating the velocity of a point $\mathbf{V}_{B_j} = \begin{pmatrix} \dot{x}_j & \dot{y}_j & \dot{z}_j \end{pmatrix}^T$ and the translational and angular velocities $\mathbf{V}_C$ and Ω_C, and by combining with eq. (6), then it is straightforward to obtain:

$$\underbrace{\begin{pmatrix} \dot{u}_j \\ \dot{v}_j \end{pmatrix}}_{2\times 1} = J(\mathbf{B}_j) \underbrace{\begin{pmatrix} \mathbf{V}_O \\ \Omega_O \end{pmatrix}}_{6\times 1} \tag{7}$$

Since the goal of visual servoing is to execute robot motions, eq. (7) must be somehow inverted. This inversion supposes that (i) one is able to measure 2-D velocities of image points, (ii) at least three points are available such that one obtains at least 6 equations if no free motion is wished, (iii) the rank of the linear system of equations thus obtained is equal to 6, and (iv) than some reasonable model of the image Jacobian J is available. In what follows we will derive the robot control law and make clear the items mentioned above.

As it has been already explained, we consider 3-D points ($\mathbf{B}_j$) onto the robot gripper together with their projections onto the image ($\mathbf{b}_j$). Let $\mathbf{s}$ be the image vector formed with the coordinates of all the points $\mathbf{b}_j$. For n points, the vector $\mathbf{s}$ has $2n$ components:

$$\mathbf{s} = (u_1\ v_1 \ldots u_j\ v_j \ldots u_n\ v_n)^T$$

We denote by $\mathbf{s}^\star$ the vector of image point positions in the final (goal) position. This goal position may correspond, for example, to an alignment condition for grasping (section 2) or to any other goal position that one wants to reach.

Therefore, the task consists of moving the robot such that the Euclidean distance between the current position $\mathbf{s}$ and the goal position $\mathbf{s}^\star$ is minimized. Hence, one may constrain the image velocity of each point being considered to be proportional to the difference vector separating the current position from the goal position. This ideal desired behaviour writes as:

$$\dot{\mathbf{s}} = g\left(\mathbf{s}^\star - \mathbf{s}\right) \tag{8}$$

where g is a positive scalar that controls the convergence rate of the visual servoing.

It is now possible to combine eq. (7) with eq. (8) and we obtain:

$$J \begin{pmatrix} \mathbf{V}_O \\ \Omega_O \end{pmatrix} = g\left(\mathbf{s}^\star - \mathbf{s}\right) \tag{9}$$

With:

$$J = \begin{pmatrix} J(\mathbf{B}_1) \\ \vdots \\ J(\mathbf{B}_n) \end{pmatrix}$$

In these equations J is a $2n \times 6$ matrix that depends on the following parameters: α_u and α_v which are the horizontal and vertical scale factors associated with the camera, x_j, y_j, and z_j are the coordinates of the gripper points expressed in the camera frame, and $\mathbf{R}$ and $\mathbf{t}$ are the rotation and translation that map the gripper centered coordinate frame onto the camera centered coordinate frame. Notice that the latter parameters are time-varying because the camera is not rigidly attached to the robot.

It is therefore desirable to estimate the values of x_j, y_j, z_j and of $\mathbf{R}$ and $\mathbf{t}$ any time one needs to compute the matrix J. This can be done by a *pose* computation method. Pose computation is a classical problem in computer vision and photogrammetry and many closed-form and/or numerical solutions have been proposed in the past (see [14] for a review). Nevertheless, past solutions to the object pose computation problem are not entirely satisfactory. This is the main reason for which the current solution used in visual servoing consists of considering that the pose parameters do not vary too much over time and hence J has constant values for its coefficients [4]. Details of a real-time pose computation method that we recently developed can be found elsewhere [10].

Consequently, the control velocity screw may be computed as:

$$\begin{pmatrix} \mathbf{V}_O \\ \Omega_O \end{pmatrix} = g\left(J^T J\right)^{-1} J^T \left(\mathbf{s}^\star - \mathbf{s}\right) \tag{10}$$

The visual servoing algorithm can now be summarized as follows:

1. Grab an image of the robot gripper.

2. Detect image points that correspond to some pre-specified gripper points.

3. Match these current image point positions with the goal positions.
   If the current position is close enough to the goal position *then* stop.
   Else go to the next step.

4. Compute the pose of the gripper with respect to the camera.

5. Compute the matrix J as well as its pseudo-inverse.

6. Compute the velocity screw associated with the gripper (eq. (10) and update the current velocity screw with the newly computed values.

7. Go to step 1.

There are a number of advantages associated with steps 4. and 5. of this algorithm. First, an exact estimation of J guarantees an exponential decrease of the image error, i.e., $\|\mathbf{s}^\star - \mathbf{s}\|$. Second, the task of tracking the feature points in the image is more robust because one can predict more precisely the locations of these points. With an approximate Jacobian there is a risk that these feature points run out of the field of view of the camera because the transient image trajectory of the robot is unpredictable. Third, the time necessary for reaching the final position is optimized.

To conclude this section we compare the dynamic behaviour of the algorithm in the case of an approximated (constant) Jacobian with the case of a exact (continuously updated) Jacobian. Let $J^\dagger$ be the pseudo-inverse of J and let $\mathbf{T}$ be a 6-vector representing the velocity screw. Eq. (10) can be written as:

$$\mathbf{T}_O = gJ^\dagger(\mathbf{s}^\star - \mathbf{s})$$

In practice we use an estimation of J, $\hat{J}$, and the previous formula becomes:

$$\mathbf{T}_O = g\hat{J}^\dagger(\mathbf{s}^\star - \mathbf{s})$$

The time derivative of the image error $\mathbf{e} = (\mathbf{s}^\star - \mathbf{s})$ is:

$$\begin{aligned} \dot{\mathbf{e}} &= -\dot{\mathbf{s}} \\ &= -J\mathbf{T}_O \\ &= -gJ\hat{J}^\dagger\mathbf{e} \end{aligned}$$

The time derivative of the module of the image error vector $\mathbf{e}$ is:

$$\begin{aligned} \frac{d}{dt}\left(\frac{1}{2}\|\mathbf{e}\|^2\right) &= \mathbf{e}^T\dot{\mathbf{e}} \\ &= -g\mathbf{e}^T J\hat{J}^\dagger\mathbf{e} \end{aligned}$$

It is clear now that:

- If $\hat{J}$ is a good estimation such that $\hat{J} \approx J$ we have an exponential convergence of the algorithm because the matrix $J\hat{J}^\dagger$ is a positive matrix.

- If $\hat{J}$ is constant over time the positivity of $J\hat{J}^\dagger$ is not guaranteed for any robot position.

4. Experiments

As already mentioned, on-line grasping is performed in two steps: alignment and servoing. The task of alignment is to determine an image configuration associated with the final grasp configuration to be reached by the robot. The task of servoing is to actually move the robot from an arbitrary position to its final position.

In practice, alignment consists in the following steps:

- *Image processing*: Edges are extracted from the row image and these edges are further segmented into straight lines. The endpoints of these lines form junctions (Figure 3 top-right). A network of lines and junctions is thus built. This network is then treated as a graph which is further split into connected components. Finally, the connected component that is most likely to represent a polyhedral object is selected (Figure 3 middle-right).

- *Object model prediction*: A 3-D wire frame representation of the object is available. This wire frame is projected onto an image using approximate intrinsic and extrinsic camera parameters. The choice of these parameters may be done interactively such that the image of the model is as closed to the real image of the object as possible. The view of the model thus obtained is a network of lines and junctions.

- *Image-to-model matching*: In order to determine the 3-D to 2-D projective mapping M described in section 2 at least 6 point matches are necessary. This is done by directly matching the network (lines and junctions) associated with the image of the model against the network associated with the actual image of the object. Notice that, because of noise corruption and because the camera parameters are not exactly the same, this network-to-network matching is not a trivial task. Figure 3 (middle) shows a set of 9 junction matches that were obtained using the method described in [6].

- *Prediction of final gripper position*: The final alignment step is to be able to predict the position of the gripper in the image. The gripper has 4 white marks on it. The geometry of these marks — their x, y, and z coordinates in gripper frame — was determined off-line using a hand/eye calibration technique [11]. It is therefore possible to map the 3-D Euclidean coordinates of these marks onto the 3-D projective space and then to project them onto the image. The projected marks thus obtained are shown on Figure 3 bottom-left.

The image marks thus determined constitute the goal position to be reached by the gripper. Any other more "natural" marks could be used in theory. In practice, the white marks greatly simplify the low-level image processing associated with the servoing task. More formally, the image coordinates of the centroids of the marks are the components of the position vector $\mathbf{s}^\star$ — see section 3.

The object grasping with visual feedback method that we just outlined is illustrated on Figure 3. This figure shows the image of the object to be grasped together with the robot gripper (top-left). This image is segmented into edges and these edges are described in terms of lines and vertices (top-right). A matching algorithm establishes vertex-to-vertex assignments between the image and a wire-frame description of the object (middle-left and middle-right). A projective invariant characterization of the grasp allows to predict image locations for the gripper marks (bottom-left). Finally, the robot gripper is servoed such that the actual image positions of the gripper marks are aligned with their predictions (bottom-right).

5. Discussion

In this paper we described a method for aligning a robot gripper — or any other end effector — with an object. An example of such a gripper/object alignment is grasping. The task consists of, first computing an alignment condition, and second

servoing the robot such that it moves and reaches the desired position. A single camera is used to provide the visual feedback necessary to estimate the location of the object to be grasped, to determine the gripper/object alignment condition, and to dynamically control the robot's motion.

The original contributions of this paper are the following. Since the camera is not mounted onto the robot it is crucial to express the alignment condition such that it does not depend on the intrinsic and extrinsic camera parameters. Therefore we developed a method for expressing the alignment condition (the relative location of the gripper with respect to the object) such that it is *projective invariant*, i.e., it is view invariant and it does not require a calibrated camera. The central issue of any image-based servoing method is the estimation of the image Jacobian. This Jacobian relates the 3-D velocity field of a moving object to the image velocity. In the past, the exact estimation of this Jacobian has been avoided because of the lack of a fast and robust method to estimate the pose of a 3-D object with respect to a camera.

From an experimental point of view, we showed the interest of exact versus approximate Jacobian estimation. It is important to stress the fact that, in all the visual tasks that have been described above, the camera is either not calibrated or poorly calibrated. The only intrinsic parameter whose value is accurately needed, is the ratio between the horizontal and vertical scale factors. This ratio is known to be provided by the camera manufacturer with great accuracy.

The use of visual feedback for object grasping and for other alignment tasks is a promising method and an active research topic because its use allows for various disturbances and because it does not require robot-to-world calibration. An alternative to the use of vision is force feedback. However, vision and force are complimentary because force is effective only if the robot end effector touches something. Coordination of vision and force is a promising research topic.

References

[1] C. Bard, C. Bellier, J. Troccaz, C. Laugier, B. Triggs, and G. Vercelli. Achieving dextrous grasping by integrating planning and vision based sensing. *International Journal of Robotics Research*, 14, 1995. To appear.

[2] P. I. Corke. Video-rate robot visual servoing. In K. Hashimoto, editor, *Visual Servoing*, pages 257–283. World Scientific, 1993.

[3] P. I. Corke. Visual control of robot manipulators – a review. In K. Hashimoto, editor, *Visual Servoing*, pages 1–32. World Scientific, 1993.

[4] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, June 1992.

[5] J. T. Feddema, C. S. G. Lee, and O. R. Mitchell. Feature-based visual servoing of robotic systems. In K. Hashimoto, editor, *Visual Servoing*, pages 105–138. World Scientific, 1993.

[6] P. Gros. Matching and clustering: two steps towards automatic model generation in computer vision. In *Proceedings of the AAAI Fall Symposium Series: Machine Learning in Computer Vision: What, Why, and How?, Raleigh, North Carolina, USA*, pages 40–44, October 1993.

[7] G. D. Hager. Real-time feature tracking and projective invariance as a basis for hand-eye coordination. In *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 533–539, Seattle, Washington, June 1994.

[8] G. D. Hager, G. Grunwald, and G. Hirzinger. Feature-based visual servoing and its application to telerobotics. In *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, volume 1, pages 164–171, September 1994.

[9] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura. Manipulator control with image-based visual servo. In *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, volume 3, pages 2267–2272, Sacramento, California, April 1991.

[10] R. Horaud, S. Christy, F. Dornaika, and B. Lamiroy. Object pose: Links between paraperspective and perspective. In *Proceedings Fifth International Conference on Computer Vision*, pages 426–433, Cambridge, Mass., June 1995. IEEE Computer Society Press, Los Alamitos, Ca.

[11] R. Horaud and F. Dornaika. Hand-eye calibration. *International Journal of Robotics Research*, 14(3):195–210, June 1995.

[12] R. Horaud, F. Dornaika, C. Bard, and B. Espiau. Visually guided object grasping. Technical report, INRIA, March 1995. Submitted to *IEEE Trans. on Robotics & Automation.*

[13] N. Maru, H. Kase, S. Yamada, A. Nishikawa, and F. Miyazaki. Manipulator control by visual servoing with the stereo vision. In *Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 1866–1870, Yokohama, Japan, July 1993.

[14] T. Q. Phong, R. Horaud, A. Yassine, and D. T. Pham. Object pose from 2-D to 3-D point and line correspondences. *International Journal of Computer Vision*, 15(3):225–243, July 1995.

[15] R. Sharma and S. Hutchinson. On the observability of robot motion under active camera control. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, volume 1, pages 162–167, San Diego, California, May 1994.

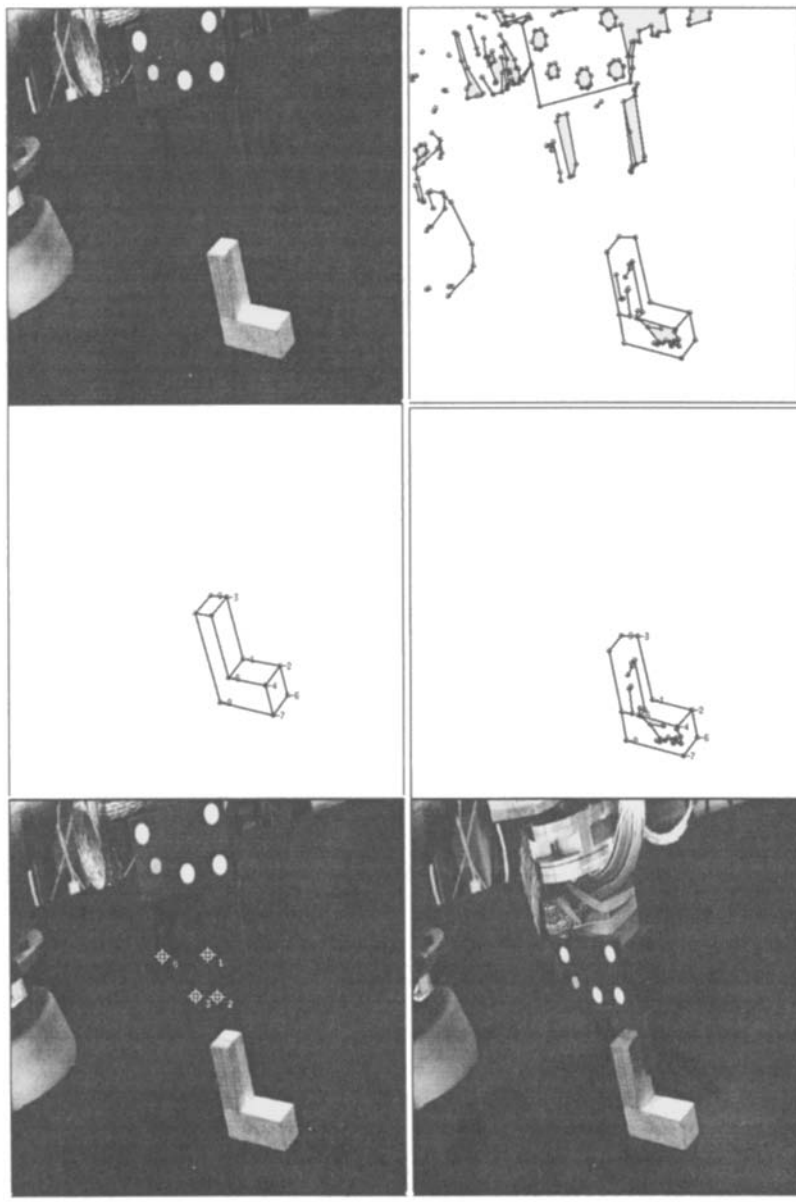

Figure 3. An example of applying the visually guided grasping method (see text). These images (top-left and bottom) are those grabbed by the camera performing both the alignment and servoing tasks.