N95-23731

Visually Guided Grasping to study teleprogrammation within the BAROCO testbed

M. Devy * V. Garric * M.Delpech ** C. Proy **

* L.A.A.S. - C.N.R.S.
7 Avenue du Colonel Roche
31077 Toulouse Cedex
FRANCE

Email: michel, garric@laas.fr

** CNES
18, avenue Edouard Belin
31055 Toulouse Cedex
FRANCE

Email: delpech, proy@philae.cnes.fr

KEY WORDS

Model-Based Vision, Localization , Grasping

INTRODUCTION

This paper describes vision functionalities required in future orbital laboratories; in such systems, robots will be needed in order to execute the on-board scientific experiments or servicing and maintenance tasks under the remote control of ground operators. For this sake, ESA has proposed a robotic configuration called EMATS; a testbed has been developped by ESTEC in order to evaluate the potentialities of EMATS-like robot to execute scientific tasks in automatic mode.
For the same context, CNES develops the BAROCO testbed [1] to investigate remote control and teleprogrammation, in which high level primitives like "Pick Object A" are provided as basic primitives.
In nominal situations, the system has an a priori knowledge about the position of all objects. These positions are not very accurate, but this knowledge is sufficient in order to predict the position of the object which must be grasped, with respect to the manipulator frame. Vision is required in order to insure a correct grasping and to guarantee a good accuracy for the following operations.
In this paper, we describe our results about a visually guided grasping of static objects.
It seems to be a very classical problem, and a lot of results are available [3]. But, in many cases, it lacks a realistic evaluation of the accuracy, because such an evaluation requires tedious experiments. We propose in this paper several results about calibration of the experimental testbed, recognition algorithms required to locate a 3D polyhedral object, and the grasping itself.

SYSTEM CALIBRATION

The figure 1 shows the LAAS experimental testbed: a 6 d.o.f. classical manipulator, with a camera mounted near the gripper. Before any experiment, a lot of knowledge must be learnt: we do not focus on these steps, but, the final results, and especially, the accuracy of the grasping, depends heavily on the calibration quality. In this

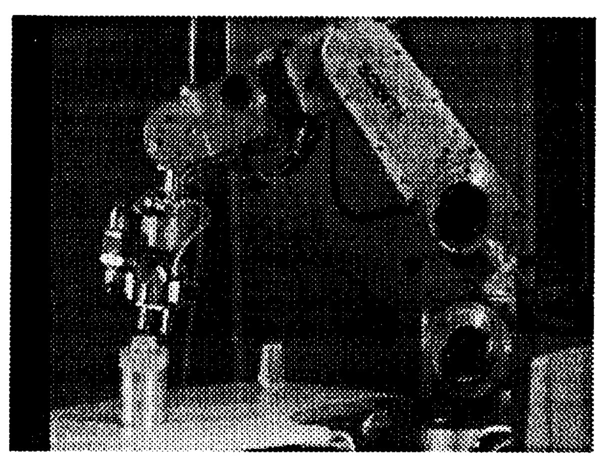

Figure 1: *The LAAS experimental testbed*

work, we only use a classical "Look and

Move" strategy in order to guide the manipulator towards the object.
On figure 2, the five different frames used during the Pick and Place task, are represented: the more important is R_{rob}, static frame linked to the robot, in which the position of the effector frame R_{eff} is known by the transform T_{re}. Two transforms must be estimated off line: T_{eg} and T_{ec}. The transform T_{co} must be estimated by the object localization from the image, corrected from distortions. In nominal situation, we have a rough estimate for the transform T_{ro}, from the a priori knowledge of the environment model.

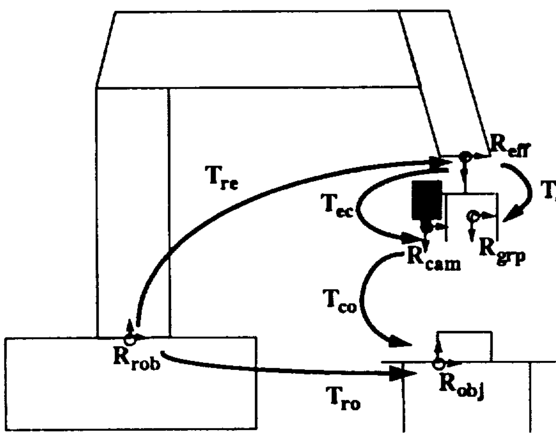

Figure 2: *Reference frames*

These gripper and hand-eye calibrations have been performed by the Tsai method [5], using a specific object (a dihedral part, fitted with visual patterns). We have evaluated the stability and the accuracy of the hand-eye calibration, for several positions of the camera around the object; we compare the estimations of the object position with respect to the robot frame R_{rob}; this position is computed by the transform product:
$T_{re} * T_{ec} * T_{co}$.
Then, the stability of this product means good estimations for T_{re} measured by internal sensors, T_{ec} estimated by the hand-eye calibration and T_{co}. We can use localization functions, which take as inputs, point matchings [4] : mean deviations of less than 1 mm for the translation, 0.06 degrees for the orientation.
Once the manipulator is calibrated, we must initialize an approximative environment model, such that the initial positions of the work areas and of the objects around the robot, are known with a maximum deviation of 5 cm in translation, and 15 degrees in orientation. At last, the object models are described by a R.E.V. graph. For each direction around the object, we index the visible 2D primitives, and we point to the discriminant clues which could provide good hypothesis, without time consuming: especially discriminant perceptual groupings, like a polygonal chain or a set of parallel segments.

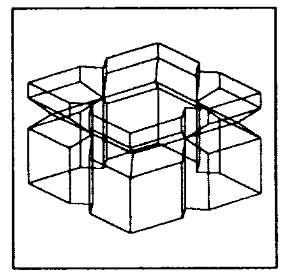

Figure 3: *Grasp interface*

The figure 3 presents the wireframe model of the grasp interface (3*3*2 cm cubic part) which will fit all equipments that the manipulator will have to pick.

OBJECT RECOGNITION

A general model-based method performs identification and localization of a 3D polyhedral object only from one image. The recognition algorithm is based on the R.E.V. models and the aspect graphs of the objects; it relies first on a generation of hypotheses, then on a verification of each pertinent hypothesis. Experiments have shown that this method required very good results for the segmentation , and that complexity could be very important (cluttered environments, occlusions, noisy images, ...). Nevertheless, 3D object recognition from a single image can provide fair results if it exists on the object model, some discriminant clues, from which a rigth hypothesis can be generated without any complexity.
Generally, for the generation, hypotheses are searched in a compatibility graph, in which each node corresponds to a so-called elementary hypothesis i.e., a matching between a scene feature and a model feature

(segments, regions, elliptic contours) , and each arc stands for the compatibility between two matchings; for each consistent hypothesis, the object position is computed. For the verification and refinement, we look for new matchings between scene features and predicted positions of model features. The generation of the elementary hypotheses relies on length criterion for single segments, or from different parameters for perceptual groupings (parallel or convergent segments). In order to determine if two elementary hypotheses are compatible, we use two kinds of constraint: *topological constraints* (connexity using the REV graph, and visibility, using the aspect graph), and *numerical constraints* (invariant measures according to affine tranformation).
Once the compatibility graph is built, the search for recognition hypotheses is performed by the maximal cliques algorithm. This method can be very expensive in computing time, due to their significant combinatorial complexity, especially if the compatibility graph is very large (too many elementary matchings, too weak compatibility criteria).
For each pertinent hypothesis, a first localization based on the segment matchings, is computed by [2]. Then, we can predict the object position in the image and infer (scene segments, model edges) matchings. If such matchings are not found, the confidence rate on this hypothesis must be reduced; otherwise, it can be increased, and a more accurate localization can be computed using Kalman filtering algorithm.

VISUALLY GUIDED GRASPING

Effectively, in the nominal case, when the system must execute a high level primitive "Pick object CYLINDER", the approximative position of CYLINDER can be found in the environment model. If this position was perfectly known, and with a perfect robot, we could directly command a movement towards the final grasp position from which the gripper could be closed.
In order to reach the actual grasp position, a vision procedure is required to correct the T_{ro} estimate during the approach, and to dynamically correct the error due to the geometrical model of the manipulator. The last movement towards the grasp position will be undertaken, only when the T_{ro} estimate will be refined and when the length of this last movement will be weak enough to insure that the grasp position will be reached with an error lower than the required tolerance (at this time, half a millimiter). So, through the first estimate of the object, T_{ro_0}, through the aspect graph which says what is the better view point to deal with the recognition of the grasp interface on CYLINDER, a planification module can off line select an optimal effector position T_{re_1}, from where an image is acquired and segmented (figures 4 and 5). From this

Figure 4: *First image*

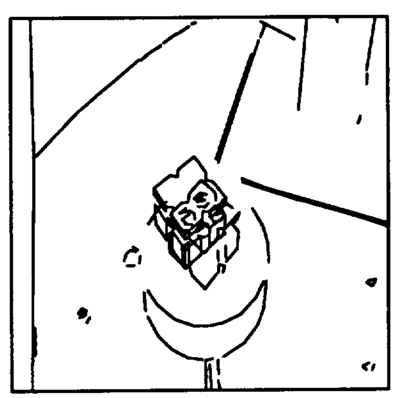

Figure 5: *Scene features*

image 1, the recognition of the cubic grasp interface, could be very simple, since the environment model gives directly the hypothesis on the object position according to the robot frame; using the different

transforms shown on figure 6 (the dashed box represents the estimated object position, according to the a priori knowledge), we can directly predict the object position T_{pred_0} with respect to the camera:
$T_{pred_0} = T_{ec}^{-1} * T_{re_1}^{-1} * T_{ro_0}$.
This prediction can replace the one given by the hypothesis generation procedure of a recognition system; it could be validated in the verification step. We show on figure 5 a possible predicted position of the object model.

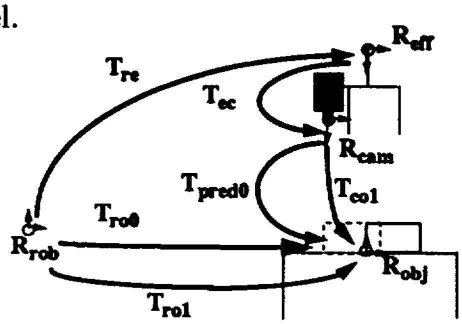

Figure 6: *Model prediction*

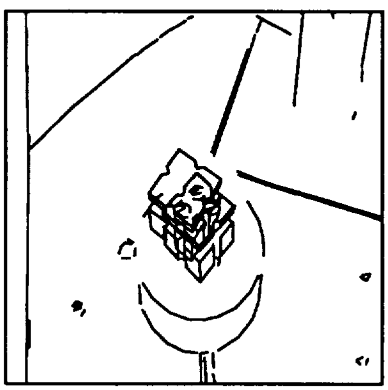
Figure 7: *First localization*

The final localization T_{co_1} is presented on figure 7. From this localization with respect to the camera frame, we can compute a better estimate T_{ro_1} of the object position with respect to the robot frame:
$T_{ro_1} = T_{ro_0} * T_{pred_0}^{-1} * T_{co_1}$
For the last iteration, the figure 8 shows the projection of the visible model edges for the prediction and for the final localization; the final localization seems perfect (model edges confounded with the scene segments). We have at this time some difficulties to estimate the error on the final grasp operation. The only result is visual; it seems we have about 1 mm error, when the effector reaches the grasp position.

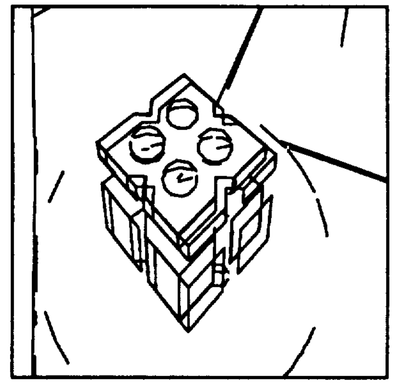
Figure 8: *Last iteration*

CONCLUSION

We have described in this paper, a perception application related to visually guided Pick and Place task which will be required in teleprogrammation mode to undertake scientific experiments in future in-orbit laboratories. Other research works will be done in order to improve the perceptual algorithms, especially to take in account more complex objects.

References

[1] M. Delpech, J.Y. Prado, P. Giron, and J. Tailhades. Baroco, a testbed to study teleoperation within the columbus environment. In *Third ISAIRAS, Toulouse (France)*, September 1992.

[2] M. Dhome, M. Richetin, J.T. Lapresté, and G. Rives. Determination of the Attitude of 3D Objects from a Single Perspective View. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, December 1989.

[3] G. Hirzinger, B. Brunner, J. Dietrich, and J. Heindl. Sensor-based space robotics-rotex and its telerobotic features. *IEEE Transactions on Robotics and Automation*, October 1993.

[4] P. Puget and T. Skordas. An Optimal Solution for Mobile Camera Calibration. In *1st ECCV (Nice)*, April 1990.

[5] R. Tsai and R. Lenz. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, June 1989.