USE OF 3D VISION FOR FINE ROBOT MOTION

Anatole Lokshin and Todd Litwin

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove, Pasadena, Ca. 91109

I. Introduction

A wide use of robotic manipulators is handicapped by their inability to operate in an unstructured environment. The existing industrial systems operate by moving along pretaught trajectories, or in the best case rely on *a priori* CAD world models. Since these motions are "blind" any distortion in robot trajectory due to calibration errors can be catastrophic. Besides the obvious inflexibility of such approach, significant problems of robot calibration have to be solved in order to exercise these motions [Stone].

An integration of 3D vision systems with robot manipulators will allow robots to operate in a poorly structured environment by visually locating targets and obstacles. However, by using computer vision for objects acquisition makes the problem of overall system calibration even more difficult. Indeed, in a CAD based manipulation a control architecture has to find an accurate mapping between the 3D Euclidean work space and a robot configuration space (joint angles). If a stereo vision is involved, then one needs to map a pair of 2D video images directly into the robot configuration space. Neural Network approach [Kup88] aside, a common solution to this problem is to calibrate Vision and Manipulator independently, and then tie them via common mapping into the task space. In other words, both Vision and Robot refer to some common *Absolute* Euclidean Coordinate Frame via their individual mappings.

This approach has two major difficulties. First a Vision System has to be calibrated over the total work space. And second, the Absolute Frame, which is usually quite arbitrary, has to be the same with a high degree of precision for both Robot and Vision subsystem calibrations.

This paper describes a work aimed on the use of computer vision to allow robust fine motion manipulation in a poorly structured world which is currently in progress at the JPL along with the preliminary results, encountered problems, and directions for the future work.

II. System Setup Description

The JPL Telerobot Testbed, where the described work has been done, includes several subsystems working cooperatively. There are low level mechanization subsystem - Manipulation and Control Mechanization (MCM), Run Time Control (RTC) subsystem, Operator Control Station (OCS), AI Task Planner, and Sensing and Perception (S&P) subsystems. More detailed description of the Telerobot Testbed and individual subsystems can be found elsewhere [Matij],[Kan 1-2], [Stone-2]. Robot Vision is provided by the S&P subsystem which determines object 3-D positions and orientations by matching a pair of stereo images with prestored polyhedral models. A detailed description of the algorithms and special purpose hardware used by the S&P is given in [Gennery].

The Telerobot Testbed has five video cameras which are available for the Sensing and Perception subsystem (S&P). Three cameras are stationary and provide wide angle - F = 12.5 mm - view of the total workspace. Two cameras are mounted on a wrist of a Puma-560 manipulator, called Camera Arm, and can be used to provide a close view stereo (F = 25.0 mm). The stationary cameras are used mainly to acquire and track large objects such as satellites, and do not provide enough resolution for fine motion manipulation. Only the movable close view cameras can be used to provide positional information for the objects in a task space. Here we will be concerned only with the Puma arm mounted stereo cameras. Therefore the terms "cameras" or "vision" will assume only a pair of potentially movable cameras and not the three stationary wing cameras.

The movable cameras can be brought close enough to the Task Board to provide a good view of a particular object of interest, however the total work volume where an object can be focused in both cameras is only about 1'x1'x1'. As a result, any realistic Payload Servicing Task requires an ability to move the cameras and to perform machine-vision operations from arbitrary locations.

The Sensing and Perception (S&P) subsystem was designed around stationary cameras which have to be calibrated only once. Due to a very time-consuming nature of a camera calibration process the same approach cannot be used for the movable cameras. It is absolutely required, that a 2D → 3D mapping of the cameras be found only once at a particular configuration of the Camera Arm, and then could be recomputed for any arbitrary Camera Arm configuration. The following method allows such an alteration.

III. Extension of the Vision Work Space

The camera model used by the Sensing and Perception Subsystem is composed of four 3-vectors: c, a, h, and v, which are known as the center, axis, horizontal, and vertical vectors, respectively (Fig. 1). These vectors describe the relationship between the 3D coordinates of the Telerobot Reference Frame (used as Absolute Frame) and the 2D image coordinates of a stationary camera. They are derived from a rather extensive and time-consuming calibration process. As long as the camera is not moved from its calibrated position then these vectors are sufficient as they are. If, however, the camera needs to be moved, then the vectors of the camera model must be altered to reflect the movement.

For each camera, there is defined an arbitrarily placed camera reference point which is rigidly attached to the camera. While this additional coordinate frame can be set anywhere, to simplify computations this frame was coincided with the wrist origin of the Camera Arm. If this point is located in the Telerobot Reference Frame at position p (a 3-vector) and orientation q (a quaternion) before the camera is moved, then the camera model (c, a, h, v) can be transformed to a new camera model (c', a', h', v') at location (p', q') after the move by the following:

$$c' = p' + R(c - p)$$
$$a' = Ra$$
$$h' = Rh \hspace{4cm} (*)$$
$$v' = Rv$$

where

$$R = r(q')r^T(q)$$

and $r(q)$ is a function which transforms a quaternion q into a rotation matrix (Fig. 1).

However, a practical application of this algorithm leads to a significant mismatch between the real and S&P perceived absolute object positions.. The errors are about 2 − 4 cm and result from the combination of errors in the Puma Arm nominal kinematics, Vision calibration based on a fictitious absolute frame, and measuring errors in the abs positions of the task objects. These errors could be significantly reduced by using more detailed model for arm kinematics [Hayati] and by much more elaborate measurements for the Task Data Base. However, it is our opinion, that an application of the computer vision in robotics should lead to the inherently robust algorithms which provide an ability to work with nominally known systems

operated in an unstructured world. To achieve these results, we substituted the whole concept of Vision Based Absolute Motion, by a Vision Based Relative Motion.

The next section describes an approach which allows employment of moving cameras by using independent calibration of Vision and Manipulation Systems and without any significant improvements in the determination of p and p' absolute coordinates.

III. Absolute vs. Relative Robot Motions

Let us compare the accuracy in the positioning of a Robot End Effector (EE) in two cases. In one case the Robot is asked to reach a position P given in some Absolute Frame, say Telerobot Reference Frame (Fig.2). Alternatively, for a second case we will command the Robot to move its EE to a position P_1 which is known relative to the EE start location P_0 (Fig. 3).

We will use standard matrix notations to describe object positions and orientations in the standard $R^3 \times SO^3$ space [Paul]. Let Z be a transformation from the origin of the absolute frame to the shoulder of Puma manipulator, and T^{0P} gives transformation from the shoulder to some point P, say EE or camera reference point. For any true value Y we will denote its computed or obtained estimate as $\hat{Y}$.

A. Absolute Motion

Let us put the EE at absolute location P which is described in the abs frame of reference by a transformation $T_{des}^{EE}(P)$. The desired joint angles can be found (and are actually computed in the existing Telerobot Testbed) via inverse kinematics of $\hat{T}^{0E}(P)$, where

$$T_{des}^{EE}(P) = \hat{Z}\hat{T}^{0E}(P) \tag{1}$$

therefore

$$\hat{T}^{0E}(P) = \hat{Z}^{-1}T_{des}^{EE}(P) \tag{2}$$

A Nominal Inverse Kinematics (NIK) algorithm transfers $\hat{T}^{0E}(P)$ into a set of desired joint angles J_{des}, which are sent to the joint controllers. As a result, the arm moves to some set of joint angles J which is slightly different from J_{des}. Due to the imperfection of the NIK model and joint controllers the achieved location of the EE P_1 is different than the desired P. The resulting shoulder-EE transform $T^{0E}(P_1)$ is different from the commanded $\hat{T}^{0E}(P)$ by some error $D_{kin}(P)$ (if P and P_1 are close then $D_{kin}(P) \sim D_{kin}(P_1)$)

$$T^{0E}(P) = \hat{T}^{0E}(P)D_{kin}(P) \tag{3}$$

The final position of the EE becomes

$$T_{abs}^{EE}(P_1) = ZT^{0E}(P_1) \tag{4}$$

and by substituting (3) into (4) we have

$$T_{abs}^{EE}(P_1) = Z\hat{T}^{0E}(P)D_{kin}(P') \tag{5}$$

257

By substituting (2) and (1) in (5) we can find an absolute error in the EE positioning

$$T_{abs}^{EE}(P_1) = Z\hat{Z}^{-1}T_{des}^{EE}(P)D_{kin}(P) \qquad (6)$$

If an estimate of the shoulder transform differs from its true value as

$$Z\hat{Z}^{-1} = D_Z \qquad (7)$$

then the final position of the EE becomes

$$T_{abs}^{EE}(P_1) = D_Z T_{des}^{EE}(P)D_{kin}(P) \qquad (8)$$

One can see from (8) that if $T_{des}^{EE}(P)$ involves large translational motions, which is usually the case, even a small rotational error in Z may lead to a significant absolute difference between the desired and actual position of the End Effector.

If T_{des}^{EE} is computed from vision or CAD data, it has its own error

$$T_{des}^{EE}(P) = \hat{T}_{abs}^{EE}(P) = T_{abs}^{EE}(P)D_{vis}(P) \qquad (9)$$

and the final difference between the desired position P and achieved P_1 becomes

$$T_{abs}^{EE}(P_1) = D_Z T_{abs}^{EE}(P)D_{vis}(P)D_{kin}(P) \qquad (10)$$

The experimental results with the PUMA 560 setup at the Telerobot Testbed showed, that a Puma arm with a carefully calibrated NIK can achieve about 5 mm accuracy. After the CAD Data Base was updated by touching a number of points on the Task Board with a special tool, an accuracy was improved to about 2 mm.

The use of the Vision System in a single, carefully calibrated configuration leads to about the same result, however, if a "moving camera model" described in sec. II is used, then the vision perceived object positions are about 20 – 40 mm off. Fig. 4 shows a Vision Overlay obtained with the stationary calibrated cameras. Fig. 5 shows performance degradation when the described above "Moving Cameras" method (eq. (*)) had been used.

B. Relative Motion

In the previous paragraph we have shown that the major potential source of the robot positioning errors is an uncertainty in the robot location relative to the "Absolute" coordinate frame.

Suppose now that we want to make a relative motion of EE from its current position P_0 to P_1 on some well known transformation A. Then the desired end position is

$$T_{des}^{EE}(P_1) = T_{abs}^{EE}(P_1) = T_{abs}^{EE}(P_0)A$$

An estimate of the desired position has to be based on the perceived current location of the EE

$$\hat{T}_{des}^{EE}(P_1) = \hat{T}_{abs}^{EE}(P_0)A \tag{11}$$

or from (1), (2), (3) and (11)

$$\hat{T}_{des}^{EE}(P_1) = \hat{Z} D_{kin}^{-1}(P_0)T^{0E}(P_0)A = \hat{Z} D_{kin}^{-1}(P_0)T^{0E}(P_1) \tag{12}$$

By substituting (12) into (6) we have

$$\hat{T}_{abs}^{EE}(P_1) = Z \hat{Z}^{-1} \hat{Z} D_{kin}^{-1}(P_0)T^{0E}(P_1)D_{kin}(P_1) = Z D_{kin}^{-1}(P_0)Z^{-1}T_{abs}^{EE}(P_1)D_{kin}(P_1) \tag{13}$$

We see from the (13), that if an arm has a good kinematics model then its relative motion can be done very accurately. But the most important feature of the derived relation (13) from the Vision – Motion coordination point of view is its independence on $\hat{Z}$. It shows that Cameras and Manipulator can be calibrated independently relative to their local frames. Namely, one needs to know only the manipulator's inverse kinematics and to be able to determine a true 3D relative position between two objects.

This property of (13) may be used for fine motion of Robot EE. Say a particular object has to be grasped by the EE. Suppose also, that the grasping should start from an approach position which is located A relative to the object. In other words the desired absolute position of the EE is

$$T_{abs}^{EE}(P_{apr}) = T_{abs}^{obj}A \tag{14}$$

If the current position of the EE is P then a relative motion B can move it to the P_{apr}, where

$$T_{abs}^{EE}(P)B = T_{abs}^{obj}A \tag{15}$$

Suppose now, that instead of the true *abs* positions in (15) one is using the vision derived values. Then the desired adjustment B is

$$\hat{B} = [T_{cam}^{EE}(P)]^{-1}T_{cam}^{ob}(P_{obj})A \tag{16}$$

Let $T_{abs}(P) = T_{cam}(P)D_{vis}(P)$ then from (16)

$$\hat{B} = [D_{vis}^{-1}(P)T_{abs}^{EE}(P)]^{-1}D_{vis}^{-1}(P_{obj})T_{abs}(P_{obj})A \tag{17}$$

If P is close to P_{obj}, which is usually the case, then $D_{vis}(P) \sim D_{vis}(P_{obj})$ (since obviously $D_{vis}(X) = D_{vis}(Y)$ when $X = Y$ and we expect the error to be a smooth function of $X - Y$) and A is small, eq. (17) gives a very good approximation to the relative distance B (vision error is uniformed across the field of view) then

$$\hat{B} = B \tag{18}$$

Eq. (18) together with (12) shows, that using Vision Based Relative Motion vision and manipulator calibration can be separated and performed in the relative sense only. Namely, instead of accuracy of absolute

259

motions only manipulators Inverse and Forward Kinematics – motions in the shoulder frame should be calibrated. Similarly, only relative camera calibration (relative to their field of view) is important. Such calibration can be done only once with a high precision calibration fixture and then eq. (*) can be used.

IV. Discussion

The Vision Based Relative Motion calls for the following possible scenario. Operator moves Cameras to position an object of interest in Cameras field of view. Then the object is manually designated (say by using an overlay wireframe) and its 3D position and orientation T_{vis}^{obj} is computed by vision using (*) calibration correction. The EE then is positioned (autonomously) at $T_{vis}^{obj}A$ which is probably 2 – 3 cm away from the desired position $T_{abs}^{obj}A$. Then the position of the EE is verified by the Vision as T_{vis}^{EE} and eq. (16) is used. After the EE is accurately positioned, grappling of parts mating operation can be done by using compline motion, etc.

The preliminary experiments on the JPL Telerobotics Testbed have shown that the described method allows reduction of the EE positioning error from about 2 cm to 1 mm with no noticeable orientation errors.

The most serious problem now is to design a "Vision – friendly" EE to facilitate fast and accurate EE verification by the Vision System.

In the future research, if vision bandwidth allows, a Vision based motion can be implemented, by substituting a single approach point A by a sequence $[A_1, A_2, ..., A_n]$ and by vision assist hopping between A_i.

V. Summary

The Vision Based Relative Motion together with teleoperation may work as a connecting link between autonomous free motions and specific task oriented macros based on the complaint motion, proximity sensors, and objects CAD models. A right combination of these technologies will allow us to perform sophisticated assembly and servicing operations in an unstructured world by using independently calibrated systems and a limited number of robotics primitives.

A work in this direction is continuing at the JPL Telerobot Testbed.

Acknowledgment.

Literature

[Kup88] Kuperstein M., "Implementation of an Adaptive Visually-Guided Neural Controller for Single Postures". *Proc. ACC*, June 15-17, 1988, Atlanta, Ga., p. 2282.

[Stone] Stone H.W., *Kinematical Modeling, Identification, and Control of Robot Manipulators*, Klewer Academic, Boston, 1987.

[Paul] Paul R., *Robot Manipulators: Mathematics, Programming, and Control*, MIT Press, Cambridge, Mass., 1981.

[Hayati] Hayati S. and Mirmirani M., "Improving the Absolute Positioning Accuracy of Robot Manipulators", *Journal of Robotic Systems*, Vol. II, No. IV, 1985.

[Matij] J. Matijevic, W. Zimmerman, S. Dolinsky, "The NASA–OAST Telerobot Testbed Architecture",

Proc. NASA Conf. on Space Telerobotics, Pasadena, Jan. 31 – Feb. 2 , 1989.

[Kan 1] E. Kan, et al.,"The JPL Telerobot Operator Control Station : Part I – Hardware". *Proc. NASA Conf. on Space Telerobotics*. Pasadena, Jan. 31 – Feb. 2 , 1989.

[Kan 2] E.Kan, et al.,"The JPL Telerobot Operator Control Station : Part II – Software". *Proc. NASA Conf. on Space Telerobotics*. Pasadena, Jan. 31 – Feb. 2 , 1989.

[Stone–2] H.Stone, et al.,"Experiences with the JPL Telerobot Testbed : Issues and Insights." *Proc. NASA Conf. on Space Telerobotics*. Pasadena, Jan. 31 – Feb. 2 , 1989.

[Gennery] Gennery et al., "Sensing and Perception Research for Space Telerobotics at JPL." *IEEE Proc. on ICRA*, March 1987, Releigh, NC.
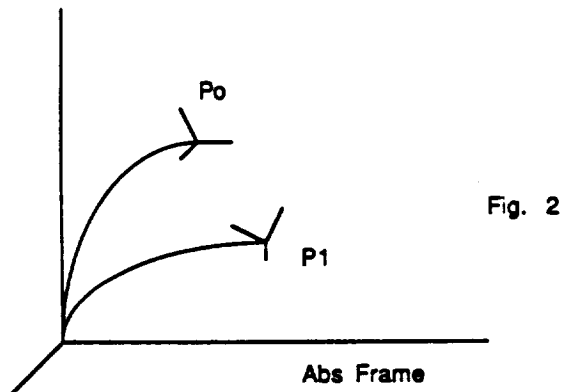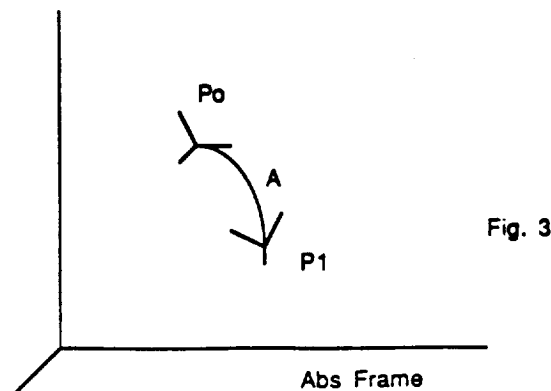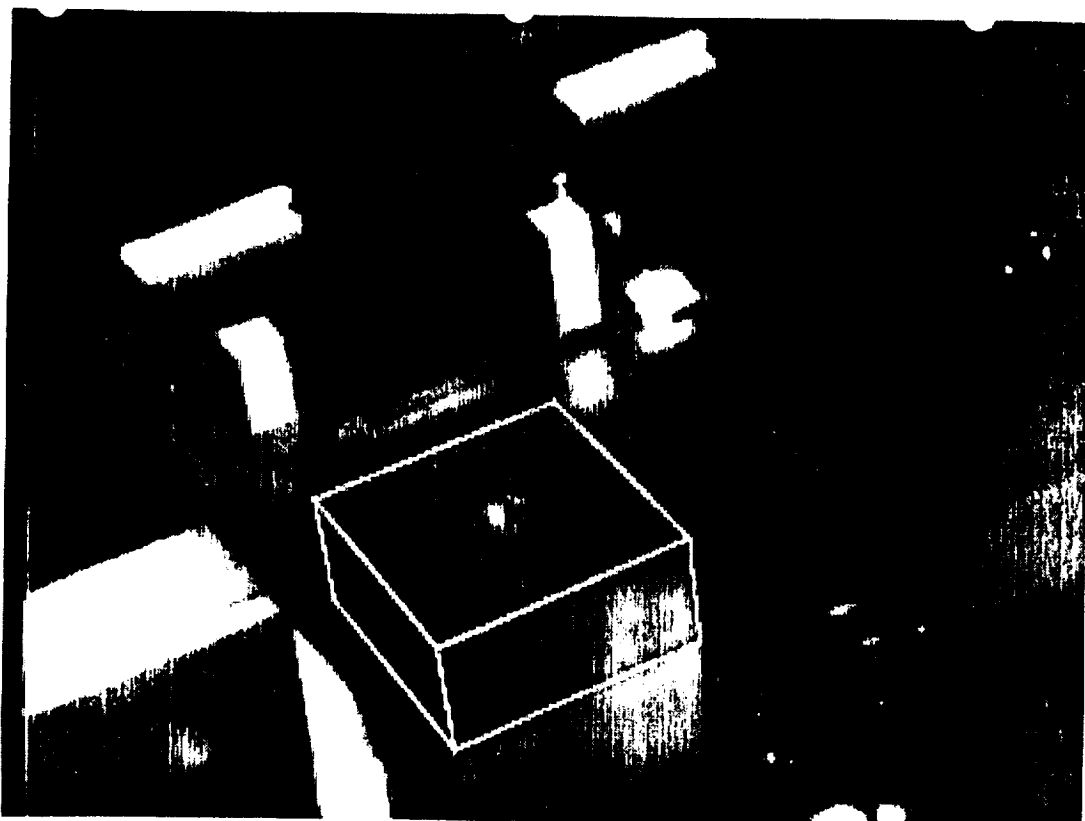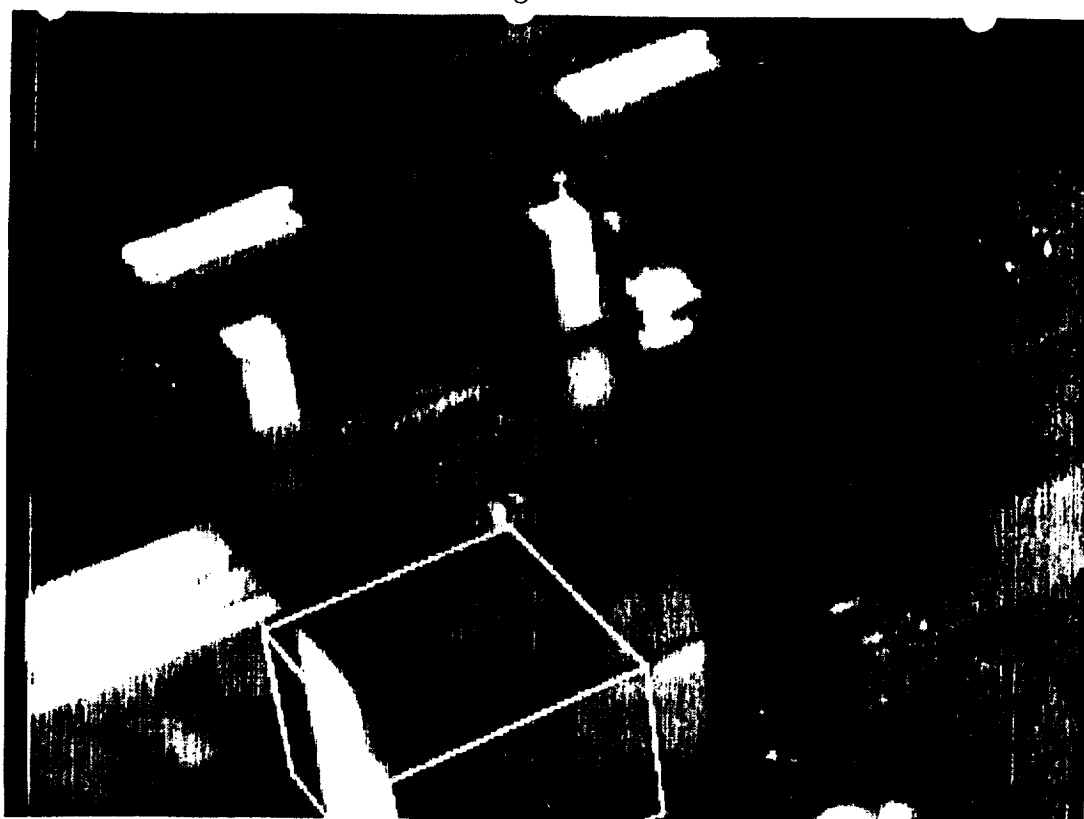
Fig. 1

Abs Frame

Fig. 2

Abs Frame

Fig. 3

Abs Frame
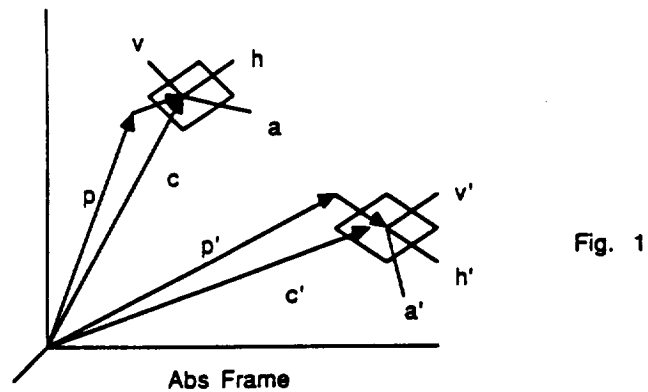
Figure 4

Figure 5

TELEROBOTS 2