Chapter in "Progress in Robotics and Intelligent Systems", George Zobrist and C.Y. Ho, editors, Ablex Publishing Corp., 1995.

Intelligent Control of a Vision-Based Spatial Reasoning System Integrated with a Robot Manipulator*

Michael Magee
Computer Science Department,
University of Wyoming,
Laramie, WY

William Hoff
Division of Engineering
Colorado School of Mines
Golden, CO

Lance Gatrell
Martin Marietta Astronautics Group,
Denver, CO

William J. Wolfe
Department of Computer Science and Electrical Engineering,
University of Colorado at Denver,
Denver, CO

1. INTRODUCTION

An intelligent robotic system must reason about a task, plan motion trajectories, avoid unnecessary contact with the environment, and manipulate target objects until a goal is achieved. The role of sensing in such a process can be very simple if the robot world is highly structured, but the evolution of intelligent robots hinges on their ability to understand dynamic environments, where sensing the right things at the right times is critical. Robots in simple environments can operate by sensing a critical feature at the start of a task, usually measuring the position of a target object, proceeding blindly with an action and then sensing a

* This research was partially supported by National Science Foundation Grants IRI-860255 and CDA-8820833.

feature for some terminal manipulation. Extending such systems to dynamic environments and complex tasks requires a much more sophisticated sensing strategy.

A common approach is to use a vision system to provide initial robot-to-target position measurement, tactile/force sensing to validate grasping, and collision avoidance sensors to monitor the entire process. In many tasks, the vision system can provide intermediate measurements that can correct for errors that may otherwise accumulate, provided that the camera is repositioned and recalibrated appropriately. In order to define principles of operation and clarify spatial analysis algorithms, the central test bed chosen for this research was a robotic task panel. The primary spatial tasks to be accomplished were the opening and closing of doors, drawers, and latches on the panel using a T_3-746 manipulator and a single movable camera.

This chapter describes a spatial reasoning system which incorporates intelligent sensor repositioning with robot planning. The fundamental operation of the system is divided into several phases. In the first phase the computer vision subsystem views the worksite (the task panel) and computes an initial registration for the primary structure. The spatial configurations of various substructures, such as doors and drawers, are determined using a generate-and-test paradigm in which knowledge of the topology, geometry, and kinematics of the task panel are used. Based upon the knowledge thus extracted, the spatial reasoning system then concentrates on an individual substructure to be manipulated and requests the robot path planner to reposition the camera so that a more advantageous viewpoint for the camera is achieved. Typically, the new position for the camera is such that the substructure occupies nearly the entire field of view. Successive refinement of the positional/orientational knowledge base is achieved by computing new registration parameters at the most recent sensor location. Once final substructure registration is accomplished, the robot path planner moves its end effector to grasp the substructure.

A description of the operation of the entire system as well as results which characterize the accuracy and reliability of the refinement algorithms are presented. It is shown that sensor repositioning can improve the quality of depth inferences up to two orders of magnitude. It is further demonstrated that knowledge about robotically controlled sensor motion along the optical axis of the camera can be used to recalibrate the camera as it approaches various substructures. This is important since the effective focal length changes between far and near views.

1.1. Related Work

Model-based spatial reasoning is a research area within the general context of computer vision which has as its basic premise the idea that the understanding of

visual (or other sensory) information is facilitated by knowing the important object features ahead of time. For example, a human observer at night can develop expectations for where the windshield and tires of an oncoming vehicle are, even though only two headlights are seen. In order to perform more specific functions (e.g., to identify the make and model) the observer may have to look for additional features as the vehicle approaches. These discriminating features are generally sought out using anticipated spatial relationships with other observed features based on models of known vehicles with which the observer is familiar.

For certain simple geometric patterns there have been numerous monoscopic approaches [1–5] to model-based computer vision which involve the comparative analysis of various combinations of the following steps:

1. Extracting primitives (for example, edges and corners) from the image
2. Assigning a correspondence between extracted primitives and object model features
3. Using a camera model to infer 3-D positions of the object that are consistent with the observed primitives
4. Introducing domain constraints (e.g., viewpoint assumptions) and iterating until a stable solution is derived.

The basic difficulties associated with each step are:

1. Extracting primitives is unreliable and inaccurate (noisy and incomplete).
2. The combinatorics of possible assignments of image-derived primitives to object features can be explosive.
3. The mathematical relationships between the geometry of solid objects and their image projections are sometimes difficult to solve, even under the assumption of perfect data.
4. Domain constraints can seriously limit the general applicability of a task-specific method.

The mathematics of image projections depends directly on the camera model used, for which the following are the simplest and most common assumptions found in the literature:

1. Orthographic camera model—perpendicular projection is assumed
2. Weak-perspective camera model—orthographic projection is followed by application of a scaling factor
3. Perspective camera model—a pinhole camera is assumed

In general, the orthographic assumption is the easiest to work with, but it is also the least generally applicable model. The weak-perspective assumption

approximates perspective projection, but degenerates with increasing depth of field. The perspective assumption is the most realistic, but is the most difficult to solve in closed form. Because of these trade-offs the computer vision literature is filled with a seemingly endless series of variations on camera models. (The analysis found in subsequent sections of this chapter uses the perspective camera model.)

Among the many image-derived primitives that might be used to suggest 3-D locations of a rigid object, the ones most commonly found in the literature are the following:

1. Three noncollinear points. The three points can be any three noncollinear points on the rigid object. They are usually geometrically significant features, such as the vertices of a polyhedron, but this need not be the case in general ([6–9]).
2. Vertex pair. This consists of a line segment (spine) between two corners detected in the image. It uses only the angle measurements at a single corner and although it is assumed that the detected corners correspond to object vertices, the spine does not have to correspond to an edge of the object ([10, 11]).
3. Four coplanar points. It is assumed that no three of the points are collinear and most typically we will assume that the points form a rectangle ([12–16]).

The use of three points is not very popular, with the notable exception of the work described in [9], primarily because of the four-fold ambiguity that can arise even when the image-to-object correspondence is known. The geometric nature of these ambiguities is the main focus of [17]. Fischler and Bolles [6] provided the foundational analysis of this problem (what they call the "perspective three-point problem"), setting up the constraining equations and demonstrating that there cannot be more than four solutions (they also give an example where there are exactly four solutions). Wolfe et al. [17] provide a systematic analysis of the geometric configurations that are associated with ambiguity sets of one, two, three, or four solutions and justifies the commonly held wisdom, as commented in [9], that the solution set usually consists of two configurations.

Many researchers have resorted to using four or more points and other features such as vertex pairs and edge/surface elements to remove the ambiguity inherent to three points. Fischler and Bolles [6] discuss the more general perspective n-point problem and [17] discusses the triangle problem in connection with vertex pairs and rectangles.

When considering higher levels of spatial reasoning, one of the most frequently cited works is that of the Acronym system ([18, 19]). This model-based system employed several fundamental control mechanisms including feature prediction and hypothesis verification. Known structures were represented as view-independent volumetric descriptors with simple objects being composed of generalized cylinders of specific dimensions. When presented with a digitized intensity image, Acronym would initially proceed in a bottom-up manner, finding edges and linking them to find higher-level structures (e.g. ellipses). These

higher-level structures were, in general, the result of projecting the three-dimensional edges of known structures. The system would then search for instances of object models within the context of the observed (projected) features. Constraint implications were propagated to lower levels during prediction and to higher levels during interpretation [20]. This control structure made Brooks' system particularly attractive since there is significant interaction between the low (pixel, edge) levels of processing and the higher levels of interpretation and understanding. Later searches for low-level features were constrained by what was known about models appearing in the scene.

Work by Martin and Aggarwal [21] used multiple-intensity images to build up volumetric descriptions of three-dimensional solids from a set of occluding contours (silhouettes). A subsequent study [22] demonstrated that the descriptors derived from observed data could be compared against descriptors of known models and recognized. Matching was achieved by comparing the principal moments and the three primary (orthogonal) silhouettes against similar parameterizations for the library of models. The system required specific knowledge regarding the relative location of the cameras, but was not constrained to represent objects based on a set of specific primitives.

The research of Boyer et al. [23] presents a system for robotic manipulation of objects which is based on structural stereopsis. One of the basic theses of their research is that establishing stereo correspondence is too computationally intensive and does not produce a scene representation that lends itself readily to the construction of a symbolic descriptor. The paradigm for primitive matching that they employ is to first extract low-level features from the left and right images and to represent these features symbolically. Parameterized structural descriptions (derived from each image) are then used as the basis for primitive matching. They are able to demonstrate a system which can reason about the orientations of several rods in a manner such that a robot manipulator is capable of moving them about. In addition to symbolic stereo from structural descriptions, [24] discusses several sensory subsystem alternatives for three-dimensional structural analysis, including the use of structured lighting, color-encoded light, and eye-in-hand configurations. These alternative approaches point out that there are many techniques that may be utilized to achieve such descriptors. Subsequent work by Chen ([25, 26]) demonstrates how representations which incorporate spherical octrees and which fuse sensory information can be applied to facilitating autonomous robot navigation.

Moving toward robotic systems which operate in well-structured domains, there is also a considerable body of research which has come from Oak Ridge National Laboratories. Specifically the works of Weisbin et al. [27] and Goldstein et al. [28] discuss the navigation, learning, and sensor fusion capabilities of an autonomous robotic system which is designed to reason in a complex but well-structured environment.

Other work by Magee et al. [29] demonstrated how complex but well-

structured objects with movable parts could be understood and reasoned about if the topology, geometry, and kinematics of various structures and substructures were known. In this study, the salient intensity-based features of a robotic task panel were modeled. These salient features facilitated the three-dimensional registration of the main body of the panel, location of parts of the panel which were immobile relative to the main structure, and inference of the (orientational and translational) states of movable substructures such as doors, drawers, and latches.

The remainder of this chapter is devoted to significant extensions of this work which greatly enhance the system's ability to handle problem cases that may arise when, for example, features are viewed from unfavorable positions that cause inaccurate reasoning or incorrect state inferences. The next section presents a brief overview of the spatial reasoning system as documented in [30]. This is followed by a discussion of some of the problems encountered and the solutions that can be achieved by closely integrating robot planning with vision-based spatial reasoning.

1.2. Research Background

In 1987, work was begun in order to develop a spatial reasoning system that was capable of reasoning about complex but well-structured objects with movable parts. The primary objectives of this early system were to observe a robotic task panel, reason about the states of its movable parts, and then to provide spatial information and knowledge to a robot planner which was capable of controlling a Cincinnati Milicron T_3 robot arm.

In its initial design, the system operated by establishing the registration parameters between the worksite and the sensor system. This phase was followed by determination of the state of each movable subpart and then robotically manipulating the relevant target substructure. The next subsections provide an overview of each of these processes.

1.2.1. Worksite Registration.

The first step in spatial reasoning is to register the main worksite (see Fig. 7.1) relative to the image sensor. There are three methods which have been successfully employed in order to accomplish spatial registration of the main panel. In the first of these methods, three noncollinear features, such as the circular targets near the panel corners, are located in left and right stereo image pairs and correspondence is established. The classical stereo equations are solved so that the three-dimensional coordinates of these features are determined and a homogeneous transformation is computed which defines how these features must have been transformed in order to produce the views (of the main panel body) which were observed. The stereo vision method proved to be robust, albeit computa-

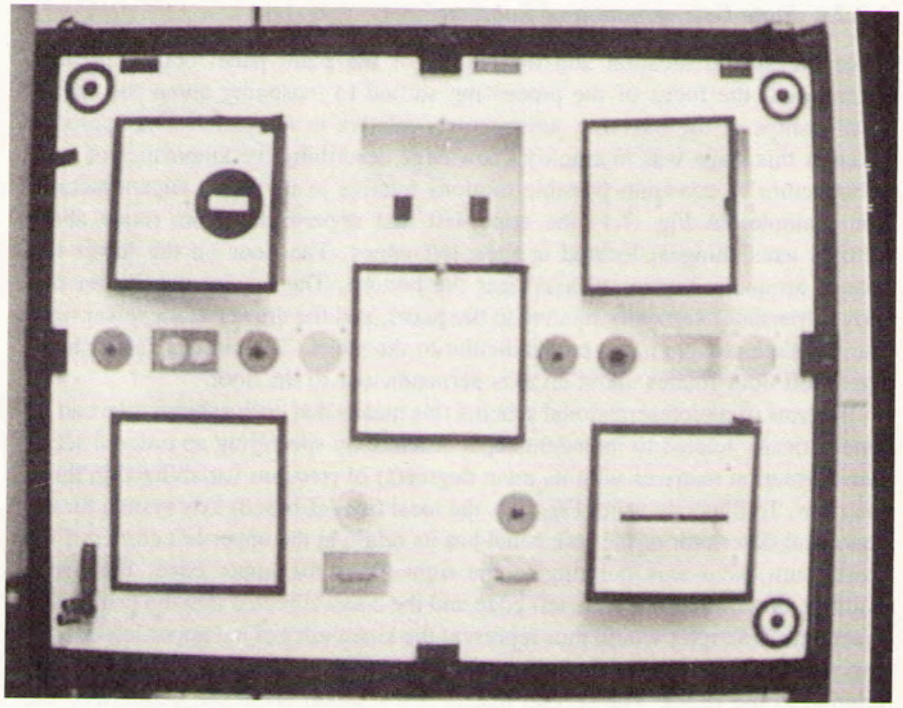

FIGURE 7.1. Raw image of the task panel at worksite registration time.

tionally expensive due to the necessity of using geometric characteristics in order to locate the primary registration features.

The second method relied on interactively locating the corners of the panel (with a mouse-controlled cursor) and applying the four-point three-dimensional worksite registration of Hung et al. [12]. This method reduced the amount of time required for worksite registration but injected a level of human interaction which would be undesirable for more fully autonomous environments. The quality of worksite registration was also directly dependent on the interactive operator's skill at locating the corners of the panel with pixel accuracy.

The third method for worksite registration relied on applying the Hung-Yeh-Harwood algorithm to active markings (for example, controllable light-emitting diodes (LEDs) located at the panel corners). From the standpoint of speed, accuracy, and repeatability of results, this method proved most desirable. The location of these active markings was fast since each marking could be easily found by subtracting an image in which all LEDs were off from an image in which each LED was active. An LED would thus appear as a bright spot in the difference image. Accuracy and repeatability were improved since subjective operator input was eliminated.

1.2.2. State Determination of Substructures.

Once the spatial location and orientation of the main panel body had been determined, the focus of the processing shifted to reasoning about the spatial relationships of the movable substructures relative to the panel. The approach taken at this stage was to employ knowledge describing the kinematics of each substructure to constrain possible motions relative to its parent superstructure. For example, in Fig. 7.1, the upper-left and upper-right doors rotate about vertical axes (hinges) located at their left edges. The door on the lower left rotates around a horizontal axis near the bottom. The upper-central door can move (translate) vertically relative to the panel, and the drawer at the lower right can translate along an axis perpendicular to the panel. The circular latch on the upper-left door rotates about an axis perpendicular to the door.

In terms of a representational schema this means that each substructure can be kinematically related to its parent superstructure by specifying an ordered set of transformation matrices with its main degree(s) of freedom variabilized in those matrices. To illustrate using Fig. 7.1, the local (model-based) axis system for the canonical descriptor of the task panel has its origin at the upper-left corner of the panel with the x-axis pointing to the right along the upper edge, the y-axis pointing downward along the left edge and the z-axis directed into the panel. The kinematic descriptor would thus represent the kinematics of the upper-left door as two matrices. The first of these matrices is a fixed translation of the door's hinge relative to the panel. The second matrix is a rotation about the y-axis with the magnitude of the rotation angle being variabilized, since it is not known a priori how far the door is open.

In order to determine the actual state of each substructural entity, variabilized transformations are generated within the kinematically possible limits of each entity. For example, each of the hinged doors can open approximately 90°. Hence, a discrete number of configurations between 0° and 90° are generated for each of the hinged doors. As each of these configurations is generated, a graphical display of the likely appearance of the entities is produced and projected into the image. This is possible since the projective characteristics of the camera are known, the homogeneous three-dimensional transformation of the panel has been calculated, and the kinematics of the substructures relative to the panel are part of the knowledge base. The thin dark lines near the edges of the movable substructures in Fig. 7.2 illustrate typical hypothesize-and-test processing for the original image shown in Fig. 7.1.

Now, as each kinematically possible configuration is hypothesized, it is tested in the actual observed data (image) by examining anticipated intensities. This matching procedure is one of iterative refinement. Initially, locations and orientations are generated and tested at a crude level of detail. For example, in the first pass, hinged-door configurations are generated at 15° intervals. When a good match is determined using crude increments, it is refined by reducing the incre-

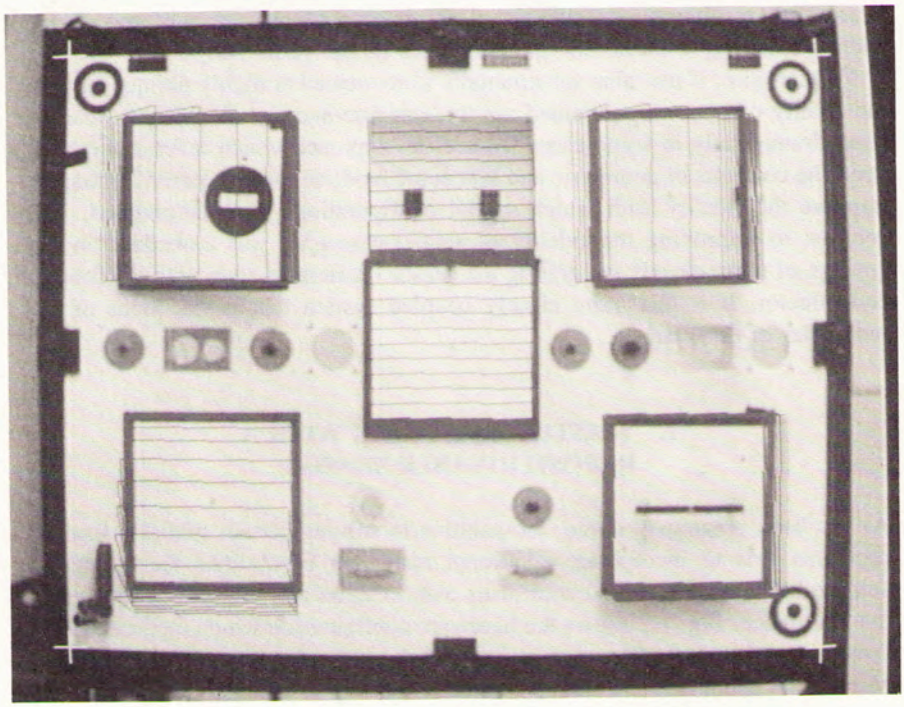

FIGURE 7.2. Generated and tested edges for primary substructures.

ment (to 2.5°) in the neighborhood of the movable feature. This is why the density of the graphically generated overlays in Fig. 7.2 increases as the true boundaries of the features are approached.

1.2.3. Robotic Manipulation.

Once the spatial configurations of the substructures relative to their superstructures have been determined, it is then possible to command a robot to manipulate them as long as the transformation of the camera coordinate system relative to the robot is known. This was in fact accomplished in the earlier study and is documented in detail in [30].

In spite of the fact that the earlier prototype system was able to provide useful spatial information to the robot arm and associated path planning system, there were, nevertheless, certain shortcomings which needed to be addressed. Specifically consider the integrated system's operation when attempting to reason about and grasp the handle of the drawer at the lower-right of the panel in Fig. 7.1. Both experimental and theoretical analyses demonstrated that image-processing errors of two pixels at each corner of the panel during the worksite registration

phase could result in three-dimensional translational (depth) errors of up to 8 cm. when attempting to locate and grasp features on the panel body.

Furthermore, if movable substructures were viewed at highly oblique angles, the quality of reasoning obtained via the generate-and-test mechanism deteriorated dramatically in some cases. Hence, an approach which refined and built upon the concepts of geometric and kinematic modeling was required in order to improve the fidelity with which spatial configurations were determined. The solution to improving the fidelity of spatial reasoning was embedded in the concept of more closely integrating the sensor (camera) system with the robotic end effector. It is this more closely coupled system that is the focus of the remainder of this chapter.

2. SPATIAL REASONING WITH A REPOSITIONABLE SENSOR

As has been previously stated, the solution to solving certain spatial reasoning problems was to incorporate additional reasoning capabilities that could be achieved if the sensor could control its own location and orientation relative to the workpiece. Fig. 7.3 shows the hardware configuration which facilitates such sensor repositioning. The camera is located on the faceplate of a Cincinnati Milicron T_3 robot arm in a manner similar to the way in which various other tools and/or end effectors would be. This gives the sensor the freedom to move about in the work environment so that new views can be suggested by the spatial reasoning system in order to avoid poor viewpoints or to improve the fidelity of the results obtained.

2.1. An Overview of Camera Repositioning

Prior to attempting to perform any actual robotic manipulation based on sensory input, the spatial reasoning system can request views from a maximum of four locations relative to a particular substructure. These will be termed the first, second, third, and fourth views. The first view of a substructure is always requested immediately following worksite registration and the substructure generate-and-test step. This view is achieved in order to reevaluate the registration of a substructure in closer proximity rather than on the initially inferred states which were determined at a considerably greater distance. Located near the corners of each substructure are four yellow circles which cannot even be detected in images taken at distances suitable for worksite registration. These markings can, however, be detected and used to reregister the substructure as the camera comes closer to the movable entity of interest. Hence, using one additional closer view can considerably enhance the accuracy of substructure state determination.

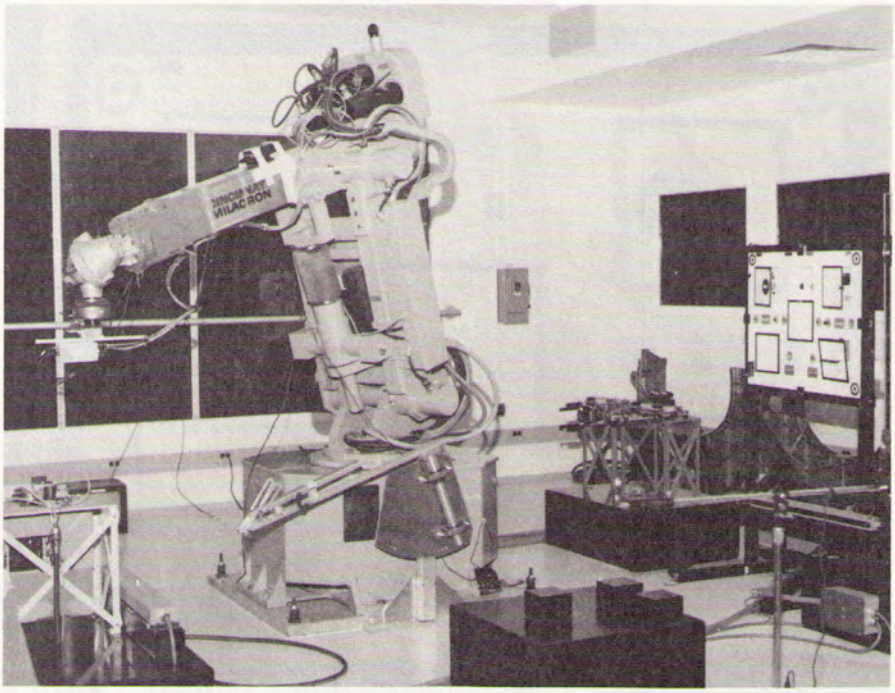

FIGURE 7.3. Laboratory configuration at worksite registration.

The second, third, and fourth views may (at the option of the operator) be taken for purposes of enhancing the accuracy of final robot end effector positioning. The primary purpose of these additional views is to recompute the important camera geometry characteristics (e.g. effective focal length) which may very well have changed due to lens refocusing between far (worksite registration) and near (first view) camera positions. The next two sections describe in detail both the objectives and methodologies employed in achieving these views.

2.2. Initial Camera Repositioning

The initial phases of processing with a repositionable camera are essentially the same as those previously discussed. That is to say, initial worksite registration and generate-and-test must be performed prior to repositioning the camera for purposes of fidelity improvement or obtaining better viewpoints of individual substructures. Once this initial processing has been accomplished, the spatial reasoning system suggests a prioritized set of advantageous camera locations which are then sought by the robot path planner. The graphical overlays in Fig. 7.4 show the prioritized set of desirable camera positions for the upper-right door

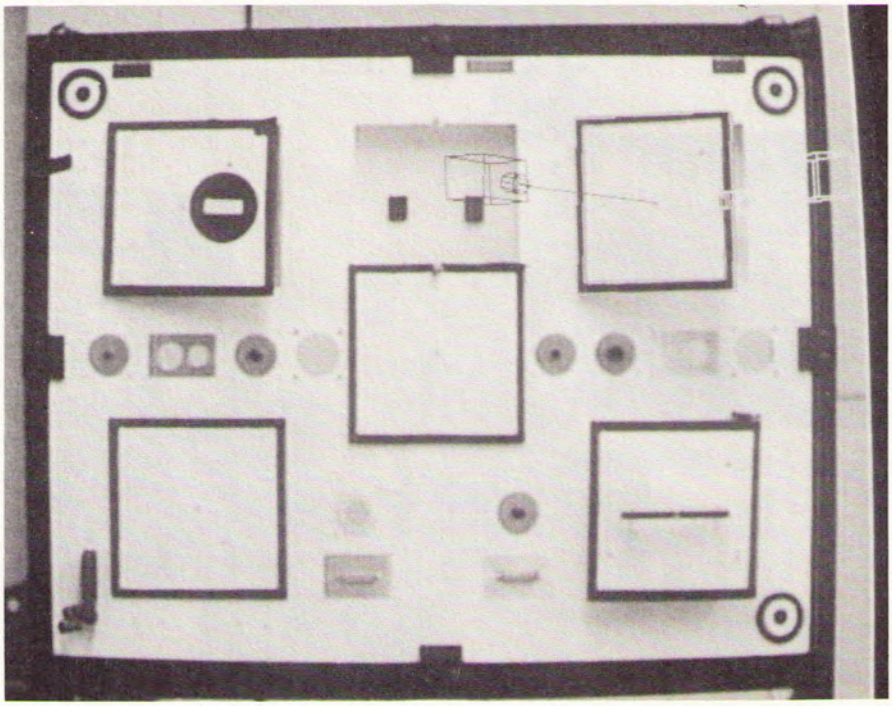

FIGURE 7.4. Initially proposed new viewpoints relative to upper-right door.

that are proposed as the result of viewing the panel from the locations shown in Fig. 7.3.

Potential new viewpoints are described as three-dimensional functions that are relative to the substructure being viewed. Currently, these functions seek as the preferred viewpoint a location that is 1 m away from the center of the substructure along a vector perpendicular to its main surface. In Fig. 7.4, the preferred viewpoint is shown as the left (dark) camera overlay. If the robotic path planner can move the camera to the proposed location the move is executed (see Fig. 7.5).

Now, upon completing the move to the new location, the spatial reasoning system views the substructure from the closer distance. This is termed the first view of the upper-right door after worksite registration. At this point it is, of course, entirely possible that errors in the initial worksite registration step as well as in the substructure generate-and-test could cause some discrepancy in where the actual substructure is in the new view relative to its anticipated location. The first view of the upper-right door shown in Fig. 7.6 dramatically illustrates the type of discrepancy which can arise. The (lightly shaded thin rectangular) overlays of the anticipated door edges are shifted to the right slightly. If the robot

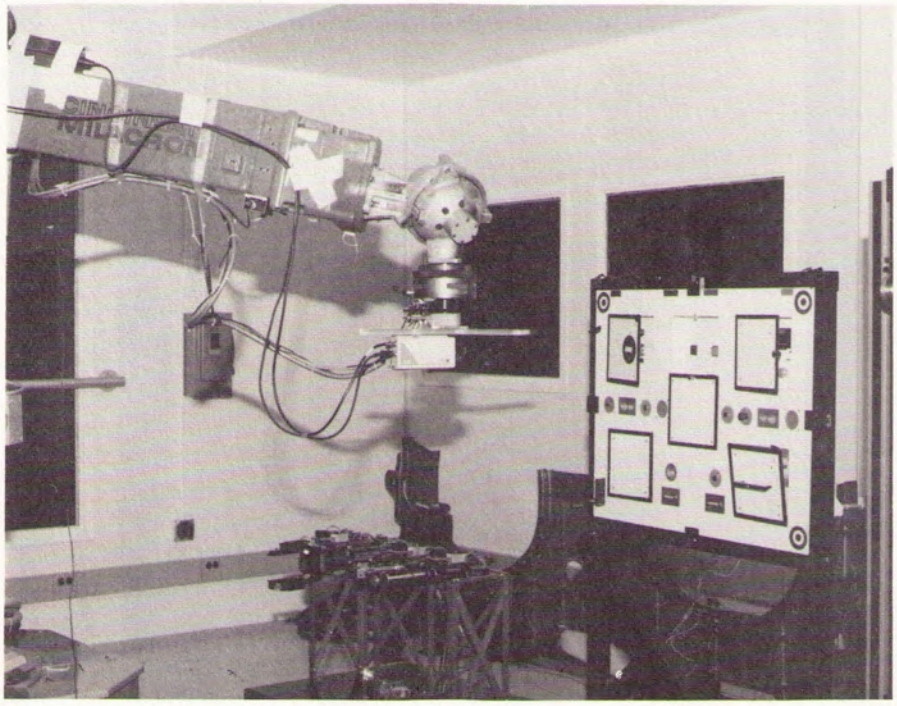

FIGURE 7.5. Configuration after initial repositioning of camera relative to upper-right door.

were to attempt to grasp the small rectangular latch on the right edge of the door at this point, it would be missed by a distance of approximately 1 cm (in the x-direction) and by (potentially) several centimeters in the z-direction. Clearly then, even the spatial information obtained at the new position must be refined.

2.3. Camera Recalibration

It should be noted at this point that because the camera has been moved between the initial (worksite registration) position and the position required to obtain the first view of the upper-right door, the effective geometry (focal length) of the camera itself has been altered. In an ideal situation, it could be assumed that one has a classical pinhole camera and that projections are always a function of a constant effective focal length f. Realistically, however, since camera lens systems are typically composed of multiple or compound lenses, the pinhole-projection assumption is generally invalid. It is therefore necessary to compensate for this lack of linearity in the projection system by recalibrating the camera

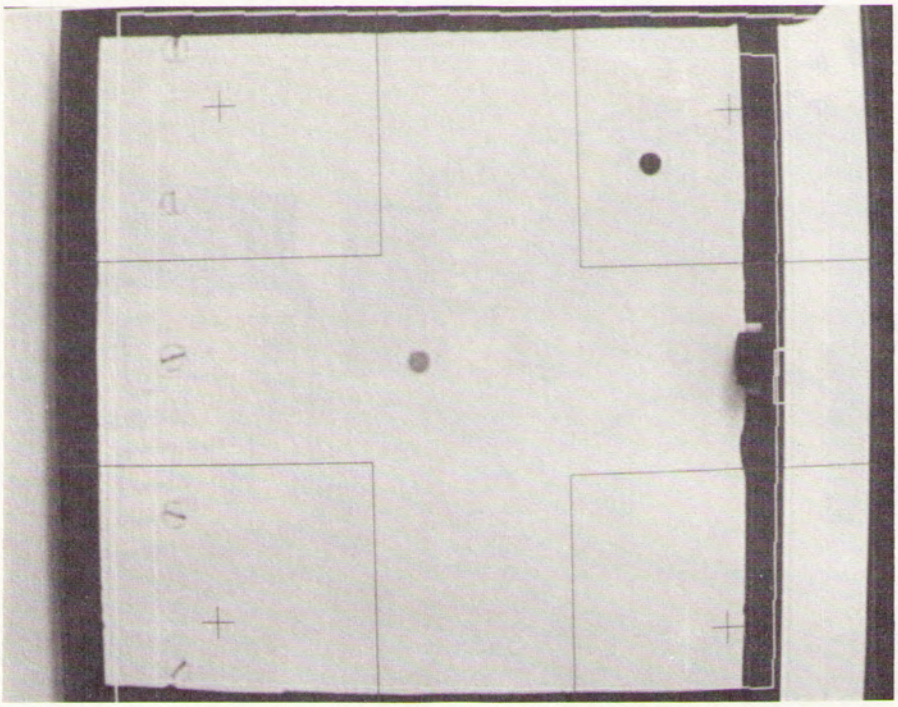

FIGURE 7.6. First view of upper-right door after worksite registration.

in an "on-the-fly" manner so that reliable projective characteristics can be obtained. This recalibration technique can be understood by examining Fig. 7.7.

Consider the case in which the effective (pinhole) focal length of the camera is not known. Suppose, however, that it is known that the camera moves in the direction of the optical (Z) axis by a distance of ΔZ away from a line of (perhaps unknown) length D, and that the optical axis is perpendicular to this line. This might reasonably occur if the camera has been repositioned on a vector orthogonal to a door and is viewing one of the edges of the door. From the similar triangle relationships in Fig. 7.7 it is clear that

$$\frac{d1}{f} = \frac{D}{Z1}$$

and

$$\frac{d2}{f} = \frac{D}{(Z1 + \Delta Z)}.$$

Hence $D*f = d1 *Z1 = d2 * (Z1 + \Delta Z)$.

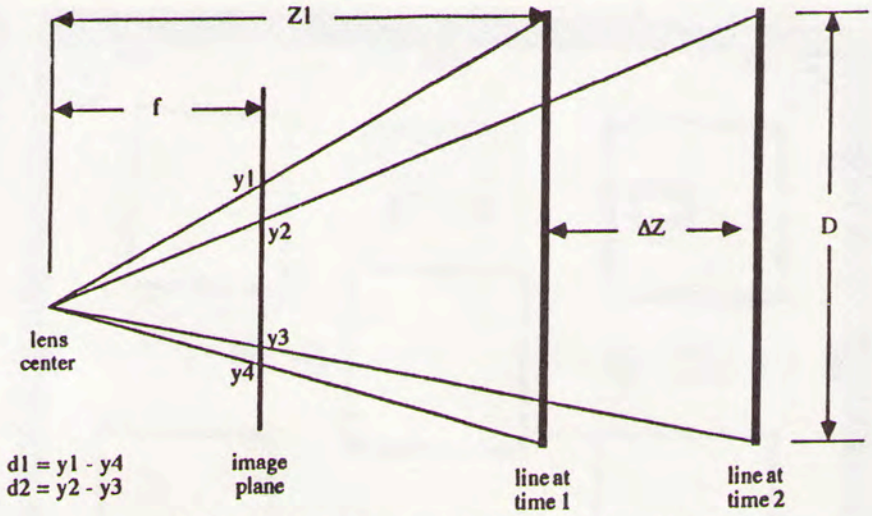

FIGURE 7.7. Geometry of camera recalibration.

Solving for Z1 it follows that $Z1 = d2 * \Delta Z / (d1 - d2)$. (Note: If the length of the line is known (e.g., it might be the length of the edge of a particular door) then the effective focal length can be computed as $f = d1 * Z1 / D$.)

The net effect is that if a linear feature which is perpendicular to the optical axis is observed at two different times such that movement of the camera is along the optical axis, the distance to the line can be computed. Furthermore, if the length of the linear feature is known, the current effective focal length can be calculated.

The actual mechanism used to achieve this recalibration is to reapply the Hung-Yeh-Harwood four-point algorithm to the yellow-colored circles identified by the lightly shaded large crosshair cursors in Fig. 7.6. These markers are found by searching the constrained darkly shaded rectangles in a neighborhood of the small black cursors. The black cursors denote the anticipated locations of each of the colored markers and are hence shifted relative to the actual locations by the same amount as the bounding rectangle for the anticipated door edges. Applying the four-point algorithm at this stage essentially reregisters the substructure (door) relative to the robotically mounted camera and provides the basis for a more accurate calculation of the vector which is orthogonal to the door. The camera is then repositioned along the refined normal vector (Fig. 7.8) and a new view is taken. This second view (Fig. 7.9) provides demonstrably better registration between the expected and actual viewed data. Note that the large cursors for the markers that were actually located and the small cursors for their anticipated locations nearly coincide. Furthermore, the search spaces shown by small black rectangles have significantly diminished in size relative to those shown in the first view (Fig. 7.6).

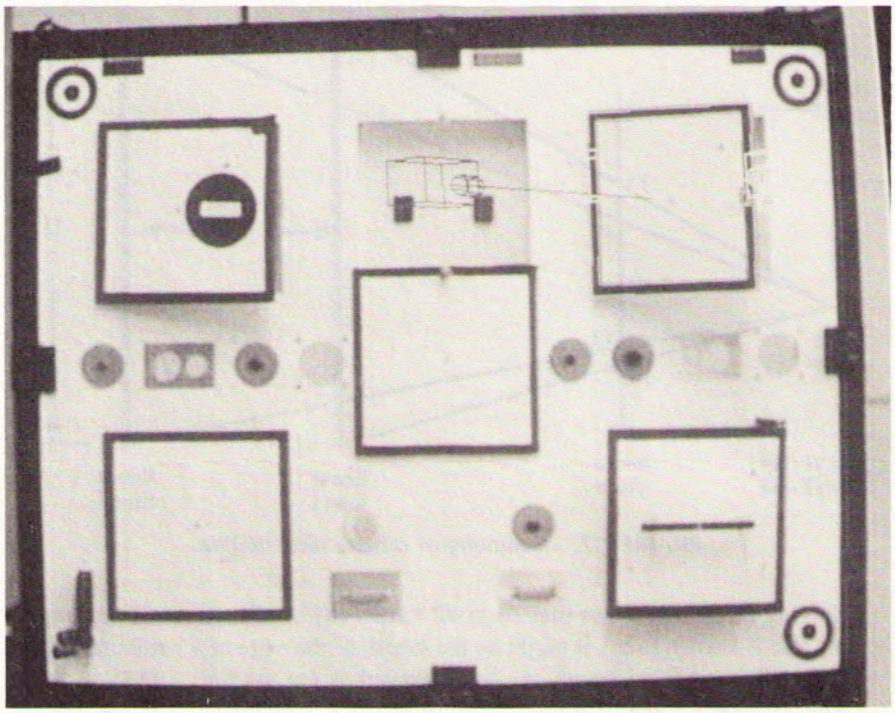

FIGURE 7.8. Proposed second viewpoint based on local reregistration of upper-right door.

Given this second view, camera recalibration can proceed under the assumption of at least an approximately orthogonal view. The camera is then moved under robot control a distance of 10 cm away from and then 20 cm toward the substructure. These third and fourth views provide the final images required to complete the accurate spatial reasoning process. Figs. 7.10 and 7.11 show these far and close calibration viewpoints, with Fig. 7.12 illustrating the locations of the four markings in all four views. Since it is known a priori how far the camera was moved along the optical axis by the robot, the effective focal length can now be computed based upon the known separation of the yellow marker points.

Applying this recalibration procedure significantly improves the results which would be obtained otherwise. For example, as target objects are approached from distances varying from 400 to 100 cm, changes in the effective focal length can cause repositioning errors on the order of 10 cm. when attempting to grasp a target object. However, incorporating the repositioning/recalibration mechanism reduces errors in the range of 1-2 orders of magnitude, such that millimeter accuracy in subsequent moves is achieved.

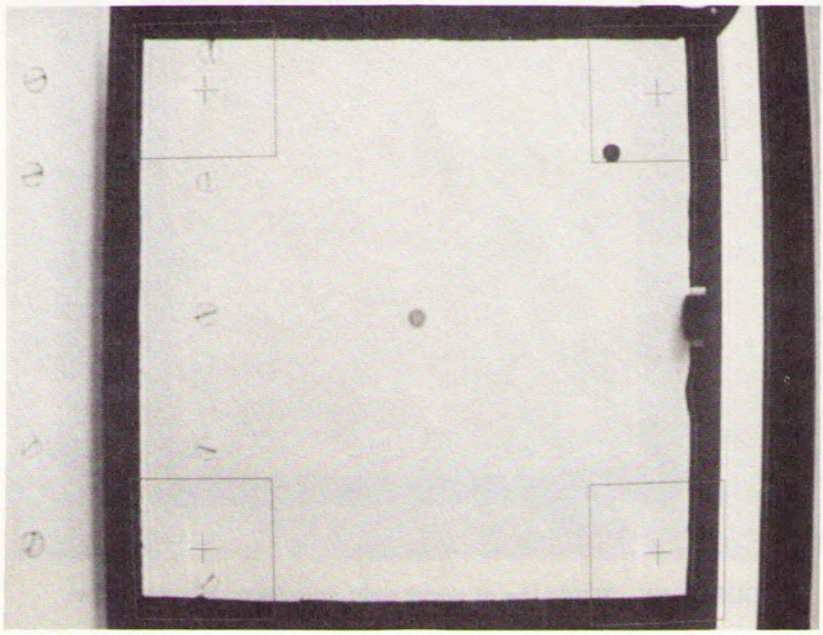

FIGURE 7.9. Second view of upper-right door.

2.4. Reasoning Prior to a Terminal Move

Once the effective focal length based on the third and fourth views is computed, it is used as the basis for determining the location of the robotically mounted camera relative to the substructure using the four-point algorithm. Since the effective focal length is changing between these two views, however, the four-point algorithm is applied to the markings found in the third view using the focal length computed based on the third and fourth views. Experimental results have shown that this procedure tends to minimize the errors in results that would be obtained if the four-point algorithm were applied to the registration points extracted from either the third or fourth views.

The terminal move is accomplished by moving a magnetic end effector to a position on the substructure having another magnet with opposite magnetic polarity (Fig. 7.13). The door is then opened by 30° (Fig. 7.14).

2.5. Repositioning for Exceptional Cases

In certain cases it is possible for the spatial reasoning system to encounter an exceptional case when attempting to determine the state of a particular substruc-

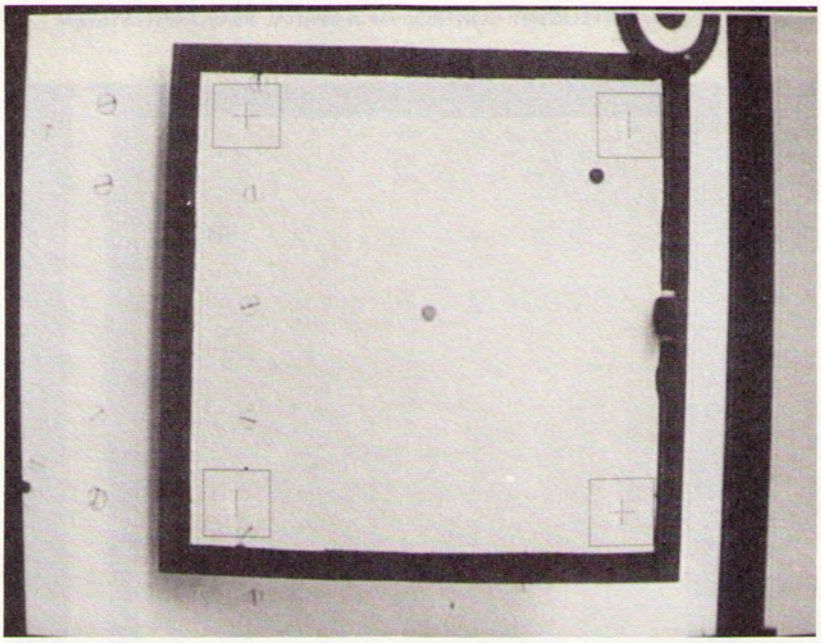

FIGURE 7.10. Third view of upper-right door (far calibration view).

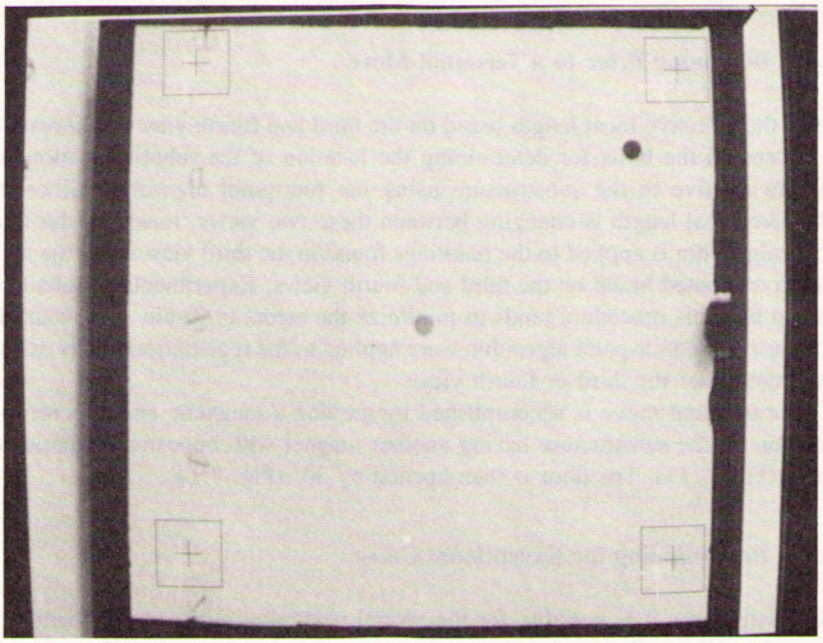

FIGURE 7.11. Fourth view of upper-right door (near calibration view).

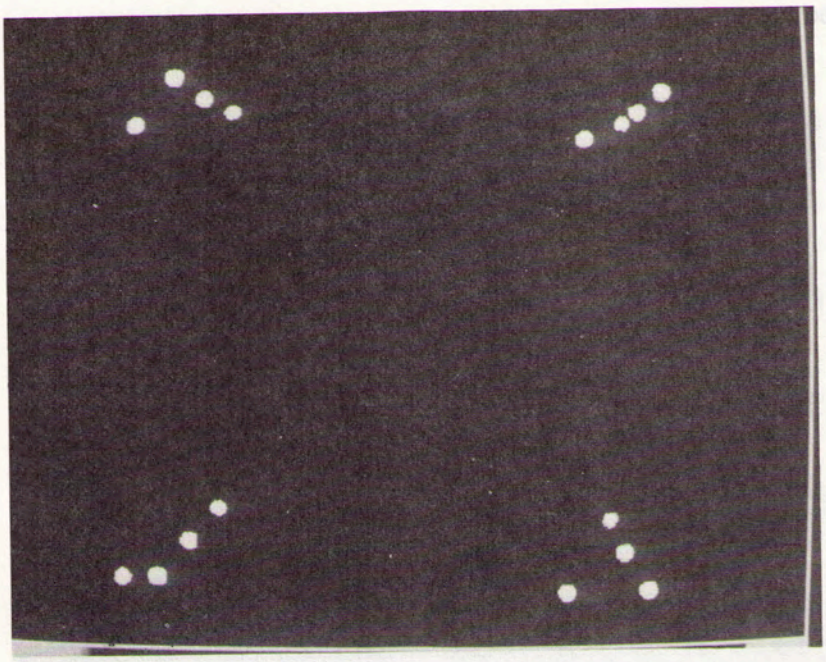

FIGURE 7.12. Composite of colored markers in all four images.

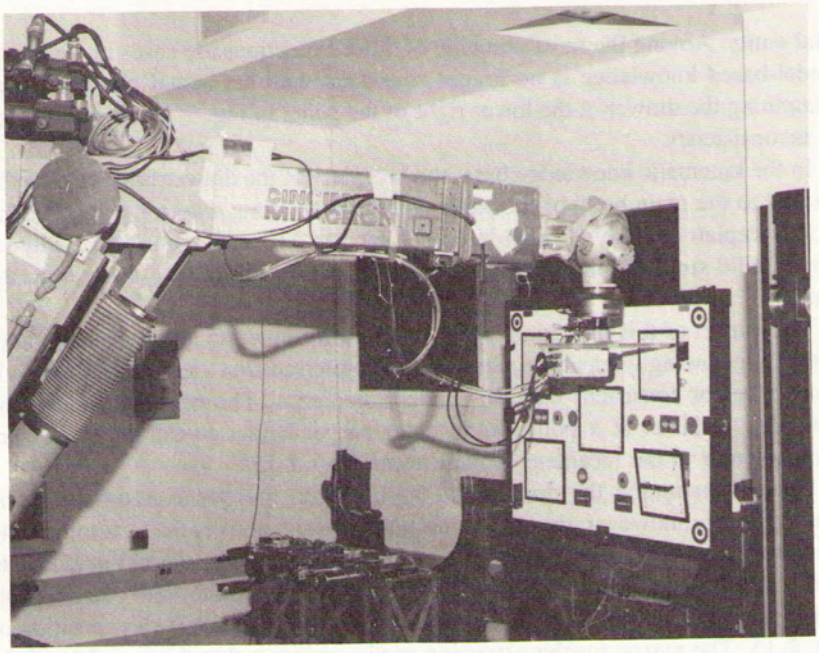

FIGURE 7.13. Grasping upper-right door.

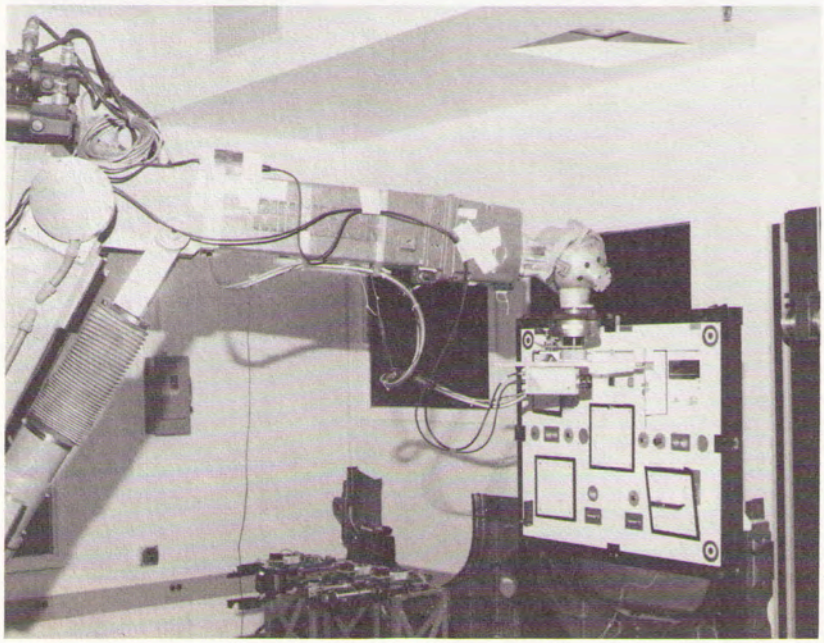

FIGURE 7.14. Opening the upper-right door by 30°.

tural entity. Among the most dramatic of these exceptions are cases in which the model-based knowledge is no longer consistent with the actual observed data. Examining the drawer at the lower right of the panel in Fig. 7.14 illustrates such an inconsistency.

In the kinematic knowledge base, the faceplate of the drawer is assumed to be parallel to the main body of the panel and to move along a vector perpendicular to its faceplate. It may be seen in Fig. 7.14 that this assumption is actually no longer valid since the faceplate has been bent outward at the top away from the panel.

The impact of this distortion is that when performing the initial generate-and-test step following worksite registration, the preferred first view of the camera is along a vector consistent with the model kinematics. The net effect is that the first view is taken at a point on a vector perpendicular to the panel and not perpendicular to the faceplate of the drawer (Fig. 7.15).

Upon reexamining the drawer and applying the four-point algorithm at the closer distance, however, the spatial reasoning system derives the location for the second view as being along a vector which is truly perpendicular to the faceplate of the drawer (Fig. 7.16). It should be noted that the second-view camera position in Fig. 7.16 is significantly below the position for the first-view position of Fig. 7.15. The reason for this alteration in viewpoints is that the faceplate of the

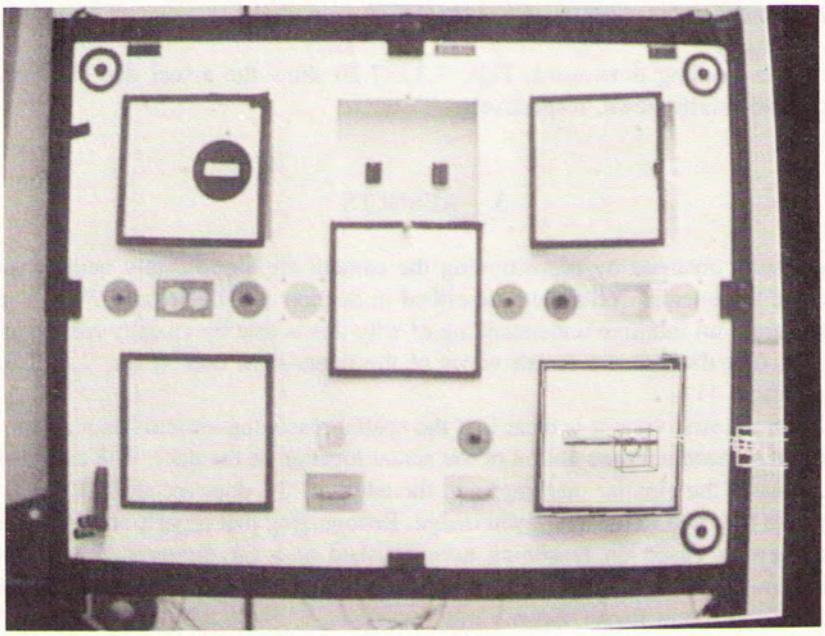

FIGURE 7.15. Proposed initial position for the camera to view the drawer after worksite registration.

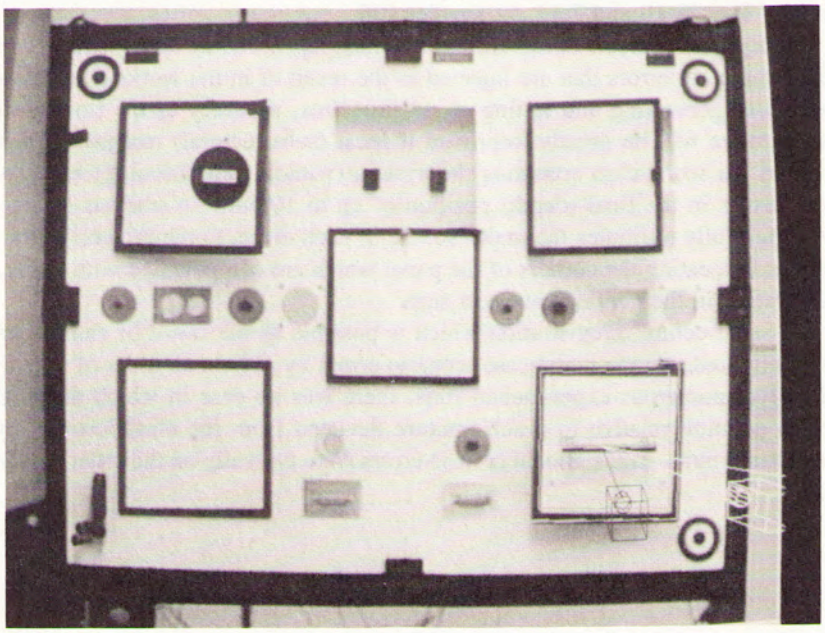

FIGURE 7.16. Revised position of camera selected to achieve view orthogonal to drawer surface.

drawer is bending downward. Figs. 7.17–7.20 show the actual first, second, third, and fourth views, respectively.

3. RESULTS

The results obtained by repositioning the camera are significantly better than those of the unenhanced system described in Section 1 of this chapter. One can easily grasp an intuitive understanding of why this is true by visually reexamining the first through the fourth views of the upper-right door (Figs. 7.6, 7.9, 7.10, and 7.11).

From the first view it is clear that the spatial reasoning system has a reasonable but somewhat crude notion of the actual location of the door. The expected locations of the circular markings and the edges of the door are shifted significantly to the right in the overlayed image. Recognizing that repositioning for the first view is based on reasoning accomplished at a far distance, the search windows for the circular markings are much larger than will be used for the second, third, and fourth views.

Upon repositioning for the second view, the search windows are considerably smaller and the crosshair cursors for the anticipated locations of the circular markings are almost directly on target. The search windows for the recalibration views of Figs. 7.10 and 7.11 are smaller still.

An empirical analysis of the utility of camera repositioning demonstrates that even if there are errors that are injected as the result of initial worksite registration and/or generating and testing of substructures, accuracy of the final positioning move will be greatly improved if local (substructural) reregistration is achieved. In worst-case scenarios the system (without repositioning) has produced errors in the final (depth) position of up to 10 cm. An analysis of such errors rightfully attributes the major source of such errors to minor (pixel) inaccuracies in locating the corners of the panel which are compounded with further inaccuracies in the generate-and-test step.

The substructure reregistration which is possible as the result of camera repositioning reduces the worst-case scenario errors by at least an order of magnitude. Over numerous experimental runs, there was no case in which the final (depth) position relative to a substructure deviated from the ideal location by more than 5 mms. Translational (x or y) errors were typically on the order of 1 to 2 mm.

4. CONCLUSIONS

It has been demonstrated that the ability of a model-based spatial reasoning system to request new viewpoints from a robotically controllable camera signifi-

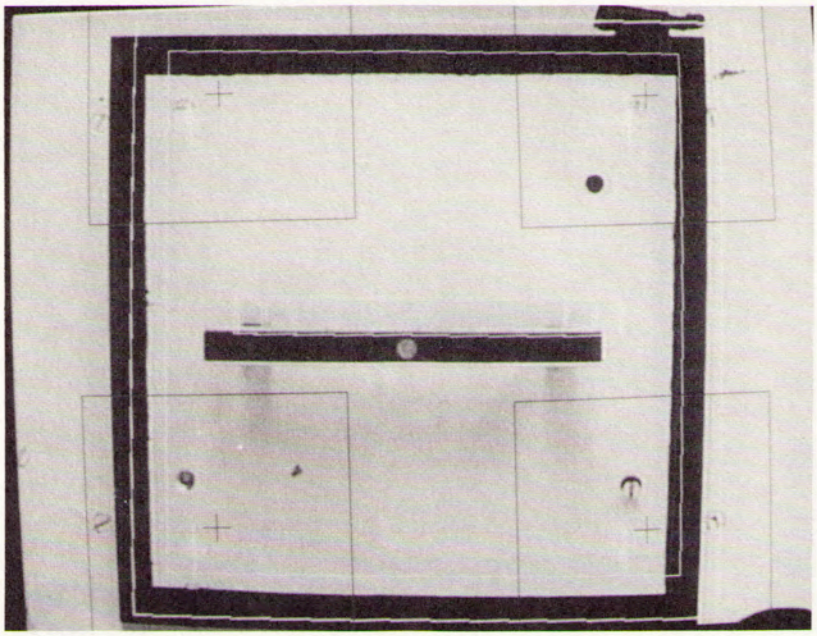

FIGURE 7.17. First view of the drawer after worksite registration.

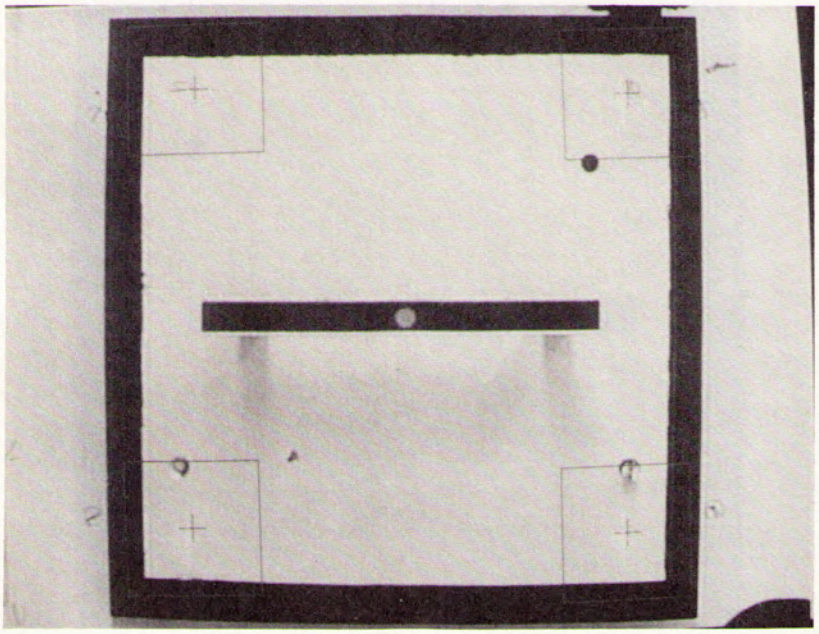

FIGURE 7.18. Second view of the drawer obtained after local reregistration.

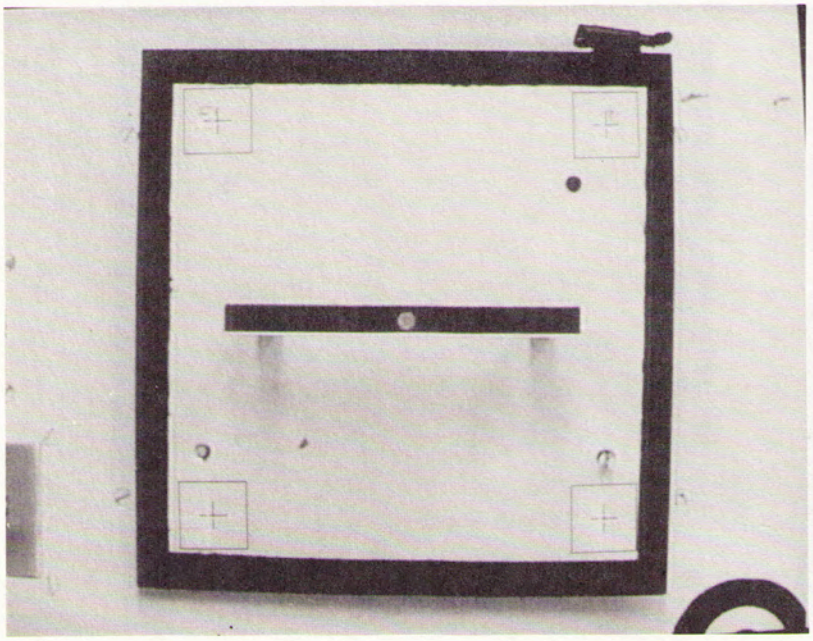

FIGURE 7.19. Third view of the drawer (far calibration view).

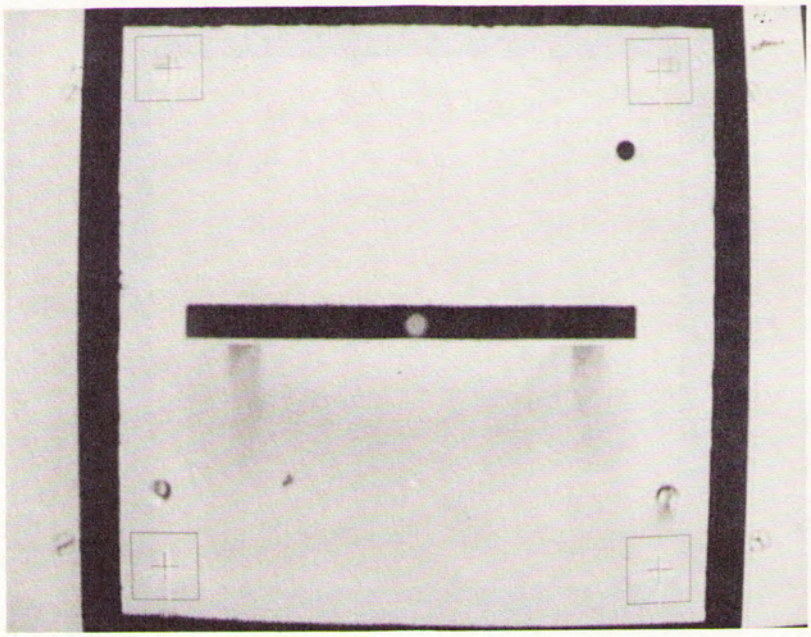

FIGURE 7.20. Fourth view of the drawer (near calibration view).

cantly enhances the quality and accuracy of spatial inferences. There are three primary reasons for the improvements that are realized.

The first of these reasons is that inferring substructural state information at large distances is a process that is highly sensitive to pixel-level errors that may arise when performing worksite registration and substructure generate-and-test. The second improvement that can be realized relates to the ability to achieve more accurate camera calibration as the result of knowing distances along the optical axis which the camera has traveled. Finally, since local reregistration of a substructure is accomplished in close proximity to the substructure based only on its local markings, exceptional conditions which run counter to the model descriptor can be handled.

REFERENCES

1. D. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment," in *DARPA Image Understanding Workshop,* April 1988 p. 1114.
2. D. Huttenlocher and S. Ullman, "Object recognition using alignment," in *Proc. First International Conference on Computer Vision,* 1987, pp. 102–111.
3. D. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence,* vol. 31, pp. 355–395, March 1987.
4. D. Lowe, *Perceptual Organization and Visual Recognition.* Boston, MA: Kluwer Academic Publishers, 1985.
5. T. Silberberg, D. Harwood, and L. Davis, "Object recognition using oriented model points," *Computer Vision, Graphics, and Image Processing,* vol. 35, pp. 47–71, July 1986.
6. M. Fischler and R. Bolles, "Random sampling consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. of the Association for Computing Machinery,* vol. 24, No. 6, pp. 381–395, June 1981.
7. W.J. Wolfe and K. Jones, "Camera calibration using the perspective view of a triangle," in *Proc. SPIE Conference on Automated Inspection and Measurement,* vol. 730, pp. 28–30, Oct. 1986.
8. J. Tietz and T. Richardson, "Development of an autonomous video rendezvous and docking system," Denver CO: Martin Marietta Denver Aerospace, Final Reports, Contract No. NAS8-34679, 1984.
9. S. Linnainmaa, D. Harwood, and L. Davis, "Pose determination of a three-dimensional object using triangle pairs," *Pattern Analysis and Machine Intelligence,* vol. 10, No. 5, pp. 634–647, Sept. 1988.
10. D. Thompson and J. Mundy, "Three-dimensional model matching from an unconstrained viewpoint," in *Proc. International Conference on Robotics and Automation,* 1987, pp. 208–220.
11. J. Mundy, A. Heller, and D. Thompson, "The concept of an effective viewpoint," in *DARPA Image Understanding Workshop,* p. 98, April 1988.
12. Y. Hung, P. Yeh, and D. Harwood, "Passive ranging to known planar point sets," in *Proc. IEEE International Conference on Robotics and Automation,* pp. 80–85, March 1985.
13. K. Cronils and P. Goode, "Location of planar targets in three space from monocular

image," in *Proc. Goddard Conference on Space Applications of Artificial Intelligence and Robotics,* May 1987.

14. R. Haralick, "Determining camera parameters from the perspective projection of a rectangle," Blacksburg, VA: Virginia Polytechnic Institute, Technical Note, June 1982.
15. W.J. Wolfe, D. Mathis, C. Weber, and M. Magee, "Integration of model-based computer vision and robotic planning," in *Proc. SPIE Cambridge Symposium on Advances in Intelligent Robotic Systems,* pp. 73–83, Nov. 1988.
16. W.J. Wolfe, G. White, and L. Pinson, "A multisensor robotic locating system and the camera calibration problem," in *Proc. SPIE Conference on Intelligent Robots and Computer Vision,* vol. 579, pp. 16–20, Sept. 1985.
17. W.J. Wolfe, D. Mathis, C. Weber, and M. Magee, "Locating known objects in 3-D from a single perspective view," in *Proc. SPIE Cambridge Symposium on Advances in Intelligent Robotic Systems,* pp. 550–556, Nov. 1988.
18. R.A. Brooks, "Symbolic reasoning among 3-D models and 2-D images," in *Artificial Intelligence 17,* Aug. 1981, pp. 285–348.
19. R.A. Brooks, "Model-based three-dimensional interpretations of two-dimensional images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 5, No. 2, pp. 140–149, March 1983.
20. P.L. Besl and R.C. Jain, "Three-dimensional object recognition," *Computing Surveys,* vol. 17, No. 1, pp. 75–145, March 1985.
21. W.N. Martin and J.K. Aggarwal, "Volumetric descriptions of objects from multiple views," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 5, No. 2, pp. 150–158, March 1983.
22. Y.F. Wang, M. Magee, and J.K. Aggarwal, "Matching three-dimensional objects using silhouettes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 6, No. 4, pp. 513–517, July 1984.
23. K.L. Boyer, A.J. Vayda, and A.C. Kak, "Robotic manipulation experiments using structural stereopsis for 3D vision," *IEEE Expert,* vol. 1, No. 3, pp. 73–94, Fall 1986.
24. A.C. Kak, K.L. Boyer, C.H. Chen, R.J. Safranek, and H.S. Yang, "A knowledge-based robotic assembly cell," *IEEE Expert,* vol. 1, No. 1, pp. 63–83, Spring 1986.
25. S.S. Chen, "An intelligent computer vision system," *International Journal of Intelligent Systems,* vol. 1, No. 1, pp. 15–28, 1986.
26. S.S. Chen, "Adaptive control of multisensor systems," in *Proc. SPIE-931 Conference on Sensor Fusion,* 1988, pp. 98–102.
27. C.R. Weisbin, G. de Saussure, J.R. Einstein, F.G. Pin, and E. Heer, "Autonomous mobile robot navigation and learning," *Computer,* vol. 22, No. 6, pp. 29–36, June 1989.
28. M. Goldstein, F.G. Pin, G. de Saussure, and C.R. Weisbin, "3-D world modeling based on combinatorial geometry for autonomous robot navigation," in *Proc. 1987 IEEE International Conference on Robotics and Automation,* 1987, pp. 727–733.
29. M. Magee, W.J. Wolfe, and B. Bloom, "Autonomous state determination using vision based spatial reasoning," *Cybernetics and Systems Research,* London: Kluwer Academic Press, 1988, pp. 909–916.
30. M. Magee, W.J. Wolfe, D. Mathis, and C. Weber-Sklair, "Model based spatial reasoning for hierarchically organized structured objects," *Advances in Spatial Reasoning,* (Volume 1) Ablex Publishing, 1990, pp. 181–218. S.S. Chen Ed.