Massachusetts Institute of Technology

Artificial Intelligence Laboratory

A.I. Memo 772 August 1984

Determining Grasp Points using Photometric Stereo and the PRISM Binocular Stereo System

Katsushi Ikeuchi
H. Keith Nishihara
Berthold K. P. Horn
Patrick Sobalvarro
Shigemi Nagata

Abstract

This paper describes a system which locates and grasps doughnut shaped parts from a pile. The system uses photometric stereo and binocular stereo as vision input tools. Photometric stereo is used to make surface orientation measurements. With this information the camera field is segmented into isolated regions of continuous smooth surface. One of these regions is then selected as the target region. The attitude of the physical object associated with the target region is determined by histograming surface orientations over that region and comparing with stored histograms obtained from prototypical objects. Range information, not available from photometric stereo is obtained by the PRISM binocular stereo system. A collision-free grasp configuration and approach trajectory is computed and executed using the attitude, and range data.

Keywords: Photometric stereo, Binocular stereo, Extended Gaussian image, Legal grasp configuration, Bin-picking, Robot vision.

Acknowledgment: This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Office of Naval Research under Office of Naval Research contract N00014-77-C-0389, in part by the System Development Foundation, and in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505.

1. Overview

Image understanding research has produced various techniques for extracting information about visible surfaces from a scene. Two lines of research that have been investigated extensively are shape from shading [Horn 75] and binocular stereo [Marr&Poggio 79]. One of the next problems that should be attacked, is how to use these methods to solve practical problems in robot manipulation. This paper explores the complementary use of photometric stereo and binocular stereo to solve problems in locating good grasp points on a doughnut shaped part in a bin of parts. The task requires the following steps:

(1) identify the *location* of the part in a complex scene,

(2) measure the *attitude* of the part,

(3) measure the *elevation* of the part above some reference plane, and

(4) compute a collision free *grasp point*.

An earlier paper [Ikeuchi&etal 83] presented techniques for using photometric stereo to accomplish the first two tasks, in addition to determining the class to which an object belongs from a set of known shape classes. In this paper we combine that system with a binocular stereo system, PRISM, designed for use in robotics [Nishihara 84], to assist with the last two tasks.

Photometric stereo determines the surface orientation at a point on an object's surface from the image brightnesses obtained at the corresponding point in the image under three different illumination conditions. Distortions in brightness values due to mutual illumination or shadowing between neighboring objects are detected by the method as "impossible" brightness triples. The locations of these triples was used to segment the visual scene into isolated regions corresponding to different objects. The distribution of surface orientations—an orientation histogram—measured over one of these isolated regions was used to identify the shape from a catalogue of known shapes. The object's attitude in space was also obtained as a by-product of the matching process.

The part's elevation, however, was not known and had to be measured by moving the manipulator hand down the camera line of sight towards the part until a light beam between the fingers was broken. Finally, with the elevation known, the manipulator was retracted and a second approach made along a trajectory appropriate to the part's attitude.

There were two problems with this approach:

(1) The pickup motion required two separate arm motions: The first, to measure elevation, and the second, to grasp the object.

(2) Collisions of the gripper with neighboring objects could not be predicted since their distances relative to that of the target were not available to the system.

In the hybrid approach presented here, a binocular stereo system is used to produce a coarse elevation map for determining a collision-free configuration for the gripper, and to measure the absolute height at the selected pickup point.

2. Basic Modules

There are four basic modules in our system: photometric stereo, binocular stereo using the PRISM algorithm, extended Gaussian image matching, and collision-free configuration planning for the gripper.

2.1. Reflectance Map and Photometric stereo

The reflectance map [Horn 77] represents the relationship between surface orientation and image brightness. Since the direction of a surface normal has two degrees of freedom, we can represent surface orientation by points on a sphere or in a two dimensional plane. The brightness value associated with each surface orientation—assuming a fixed light source and viewing configuration—can be obtained either empirically [Woodham 79] or analytically from models of the surface micro-structure and the surrounding light source arrangement [Horn&Sjoberg 79].

The photometric stereo method takes multiple images of the same scene from the same camera position with various illumination directions in order to determining surface orientation [Horn&etal 78, Woodham 78, Silver 80, Woodham 80, Ikeuchi 81b, Coleman&Jain 81]. This setup gives multiple brightness values at each picture cell. Since different images are taken from the same point, there is no disparity between the images, as there is with binocular stereo; thus, no correspondence problem has to be solved.

Each illumination configuration has a unique reflectance map associated with it, and so each of the three brightness measurements is consistent with a different set of surface orientations. Each of these sets corresponds to an iso-brightness contour on the reflectance map associated with that lighting configuration. The intersection of the three contours obtained will typically yield a unique surface orientation.

This method is implemented using a lookup table. If we assume both the viewer and the light source are far from the object, then both the light source directions and the viewer direction are essentially constant over the image. Thus, for a particular light source, the same reflectance map applies everywhere in the image. In practice, a calibration object of known shape is used to determine the relationship between brightness and surface orientation. The points where iso-brightness lines cross can be pre-calculated and stored as a table of surface orientations indexed by triples of brightness values. Thus the main operation of the algorithm is table lookup! This makes it possible to determine surface orientations very rapidly.

The result of the application of the photometric stereo method is called a needle diagram, since it can be shown as a picture of the surface covered with short needles, each parallel to the local normal. The length of a line, which is the image of one of the needles, depends on how steeply inclined the surface is, and the orientation of the line indicates the direction of steepest descent.

2.2. The PRISM system

The PRISM stereo-matching algorithm was designed to produce range measurements rapidly, in the presence of noise. The algorithm is built on the zero-crossing stereo theory of Marr and Poggio [Marr&Poggio 79]. Their approach uses scale specific image structure in a coarse-guides-fine matching strategy.

Their matching primitive was defined in terms of local extrema in the image brightness gradient, after approximate lowpass filtering with a two dimensional Gaussian convolution operator. The low pass filtering serves to attenuate high spatial frequency information in the image so that local maxima in the gradient would correspond to coarse scale properties of the image. These locations are approximated by zero-crossings in the Laplacian of the Gaussian filtered image, or equivalently, zeros in the image convolved with a Laplacian of a Gaussian, $\nabla^2 G$, [Marr&Hildreth 80]. The PRISM algorithm, however, does not explicitly match zero-crossing contours.

The zero-crossing contours are, for the most part, stably tied to fixed surface locations, *but* their geometric structure carries more information, some components of which are closely coupled to system noise. As a consequence, algorithms which explicitly match zero-crossing contours tend to be more noise sensitive than is necessary [Nishihara 84]. Matching the dual representation—regions of constant sign in the $\nabla^2 G$ convolution— produces useful results over a broader range of noise levels and more rapidly than algorithms that explicitly match the shape of the contours bounding regions of constant sign.

An additional consideration that has influenced the design of this system, is the specific nature of most sensory tasks in robotics [Nishihara&Poggio 83]. Our view in this design has been, that by avoiding the computation of details not necessary for accomplishing the task at hand, a simpler, faster, and possibly more robust performance can be obtained. The PRISM system [Nishihara 84] was designed to test this notion.

The initial design task of the implementation was to rapidly detect obstacles in a robotics work space and determine their rough extents and heights. In this case speed and reliability are important while spatial precision is less critical.

Four components make up the system, first an *unstructured* light source is used to illuminate the workspace. A simple slide projector covers the viewed surfaces with a random texture pattern to provide a high density of surface markings to drive the binocular matching. The specific geometry of the markings is not important to the matching, thus markings already present in the physical surface do not interfere with, and in fact assist, the matching process. This is not the case with single camera *structured* light systems which depend on the measurement of the fine geometric structure of a known projected pattern.

The second component is a high speed convolution device [Nishihara&Larson 81] which applies a 32×32 approximation of the $\nabla^2 G$ operator to the left and right camera images.

The third component uses a binary correlation technique to determine the relative alignments between patches of the left and right filtered images which produce the best agreement between the convolution signs. This operation is accomplished at three scales of resolution using a coarse-guides-fine control strategy. The result is a disparity measurement indicating the best alignment, along with a measure of the quality of the match between left and right images, at that alignment.

The final component handles the conversion of image position disparity to physical height. Two conversion tables are used. One gives absolute elevation as a function of horizontal disparity. The other table gives vertical disparity as a function of horizontal disparity. Together they allow cameras with large—but stable—geometric distortion to be used. Both mappings depend on position in the image.

The test system uses a pair of inexpensive vidicon cameras. Vidicons were selected over solid state cameras for the first implementation to allow an assessment of the approach with particularly bad geometric distortion and limited brightness resolution. The cameras are mounted above the workspace of a commercial manipulator, the Unimation PUMA. The digitized video signals are fed to the high speed digital convolver which applies a 32×32 approximation of the $\nabla^2 G$ operator to the images at a 10^6 picture cell per second rate.

Matching is accomplished in software on a Lisp machine. The basic module of the program performs a test on a single patch in the image at a single disparity and determines whether or not a correlation peak occurs nearby. If one does, the approximate distance and direction in disparity to that peak is estimated. The "detection range" of this module is determined by the size of the convolution operator used. With the largest operator, a single application of the module covers a range of about 12 picture cells in disparity. Repeated applications of this module are used to produce a 36 × 26 matrix of absolute height measurements—accurate to approximately 10 mm with a repeatability about 5 times better. The matching covers a 100 picture cell disparity range and takes 30 seconds from image acquisition to final output.

2.3. Extended Gaussian Image Matching

The extended Gaussian image (EGI) of an object can be approximated by the histogram of its surface orientations. Let us assume that there is a fixed number of patches per unit surface area and that a unit normal is erected on each patch. These vectors can be moved, without changing the direction they point in, so that their "tails" are at a common point and their "heads" lie on the surface of a unit sphere. Each point on the sphere corresponds to a particular surface orientation. This mapping of points from the surface of the object onto the surface of a unit sphere is called the Gaussian image and the unit sphere used for this purpose is called the Gaussian sphere [Do Carmo 76].

Imagine now attaching a mass to each end-point, equal to the area of the patch it corresponds to. The resulting distribution of masses is called the *extended* Gaussian image (EGI) of the object [Smith 79, Bajcsy 80, Ballard&Sabbah 81, Ikeuchi 81a, Horn 83], in the limit as the density of surface patches becomes infinite. It has several interesting properties: the total mass is equal to the surface area of the object, the center of mass is at the center of the sphere, and there is only one convex object corresponding to any (valid) EGI.

The EGI is invariant with respect to translation of the object. If it is normalized, by dividing by the total mass, then it is also invariant with respect to scaling. When the object rotates, the EGI is changed in a particularly simply way: it rotates in the same fashion as the object. These properties make it attractive for determining the attitude of an object.

A surface patch is not visible from a particular viewing direction if the normal to the surface makes an angle of more than 90° with respect to the direction towards the viewer. The orientations which correspond to those patches that *are* visible, lie on a hemisphere, obtained by cutting the Gaussian sphere with a plane perpendicular to the direction towards the viewer. This hemisphere will be referred to as the *visible hemisphere* [Ikeuchi 83]. It should be clear that we can estimate only one half of the EGI from data obtained using photometric stereo or depth ranging.

We will call the point where the direction towards the viewer intersects the surface of the visible hemisphere the *visible navel.* Surface patches that are visible, have orientations which correspond to points on the Gaussian sphere whose distance from the navel, measured on the surface of the sphere, is no more than $(\pi/2)$.

There are two problems in matching the EGI estimated from experimental data with those obtained from object models and stored in the computer: the number of degrees of freedom of the attitude of an object, and the effects of self-obscuration on the observed EGI's for objects that are not convex.

The attitude in space of an object has three degrees of freedom. Correspondingly, there are three degrees of freedom in matching the observed EGI and a prototypical EGI. Two degrees of freedom correspond to the position on the prototypical Gaussian sphere of the visible navel of the observed EGI (That is, the direction towards the viewer). The remaining degree of freedom comes from rotation of the observed EGI, relative to the prototypical EGI, about its visible navel (That is, the rotation of the object about the direction towards the viewer). One approach, is to evenly sample the space of rotations and perform a match for every trial attitude. This brute force method can be somewhat expensive, if reasonable precision in determining the attitude is required, since the space of rotations is three dimensional.

We use two notions to constrain orientation. First of all, note that the apparent (cross-sectional) area of an object depends on where it is viewed from. It can be shown that the height of the center of mass of the visible hemisphere of the EGI, above the plane through the edge of the hemisphere, is equal to the ratio of the apparent to the actual area. So the location of the center of mass of the observed EGI constrains the possible positions of the visible navel on the prototypical EGI

(Note that the center of mass of the *whole* EGI is at the center of the sphere and so of no use). Secondly, the direction of the axis of least inertia of the observed EGI can be used to determine the relative rotation between the two EGI's for a particular position of the navel on the prototypical EGI [Ikeuchi 83].

In the case of a convex object, the EGI obtained from a needle diagram, taken from a particular direction, is equal to the full EGI of the object, restricted to the corresponding visible hemisphere. This is not the case, in general, when dealing with a non-convex object: Some surface patch may be obscured by another part of the object, and thus not visible, even if the normal makes an angle of less than 90° with the direction towards the viewer. So the contributions of surface patches to the EGI will vary with viewing direction. One can deal with this by defining a viewer-direction dependent EGI, which takes into account the effects of obscuration. The disadvantage of this approach is, of course, that we now have to store many EGIs to represent one object instead of a single one. We can store these EGIs in a table whose rows correspond to rotations about the line of sight, and whose columns correspond to different positions of the navel on the Gaussian sphere.

2.4. Grasp Configuration

The grasp configuration should satisfy the following two conditions (assuming friction):

(1) It should produce a mechanically stable grasp, given the gripper's shape and the object's shape. Such configurations will be called *legal grasp* configurations.

(2) The configuration must be achievable without collisions with other objects. Such configurations are limited by the relationship between the gripper's shape and the shapes of neighboring obstacles. Configurations satisfying this condition will be called *collision-free* configurations.

These configurations depend on the type of gripper. We assume that the gripper has a pair of parallel rectangular jaws, as is commonly the case in current industrial robots.

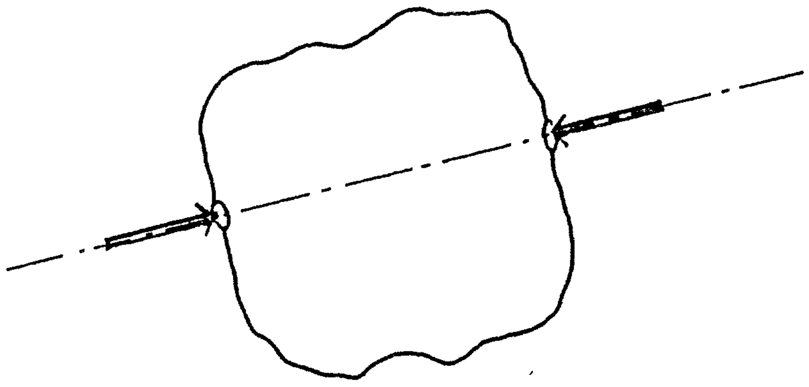

Figure 1. The two applied forces: The applied forces should be the same in magnitude, of opposite direction, and be along the line between the two contact points.

2.4.1. Legal Grasp Configuration

There are several definitions of optimal grasping [Hanafusa&Asada 77, Brady 82]. We define the optimal grasping configuration as the one in which the object satisfies the following two conditions:

(1) The object is not free to translate while the gripper is grasping the object.

(2) The object is not free to rotate while the gripper is grasping the object.

A parallel jaw gripper applies forces at two points. In order to guarantee conditions (1) and (2), the two applied forces should be the same in magnitude, opposite in direction, and lie along the line connecting the contact points, as indicated in Figure 1.

Consider the force at one of these points of contact. Let the friction angle be the arc-tangent of the coefficient of friction. If the angle between the surface normal direction and the line connecting the two grasping points is less than the friction

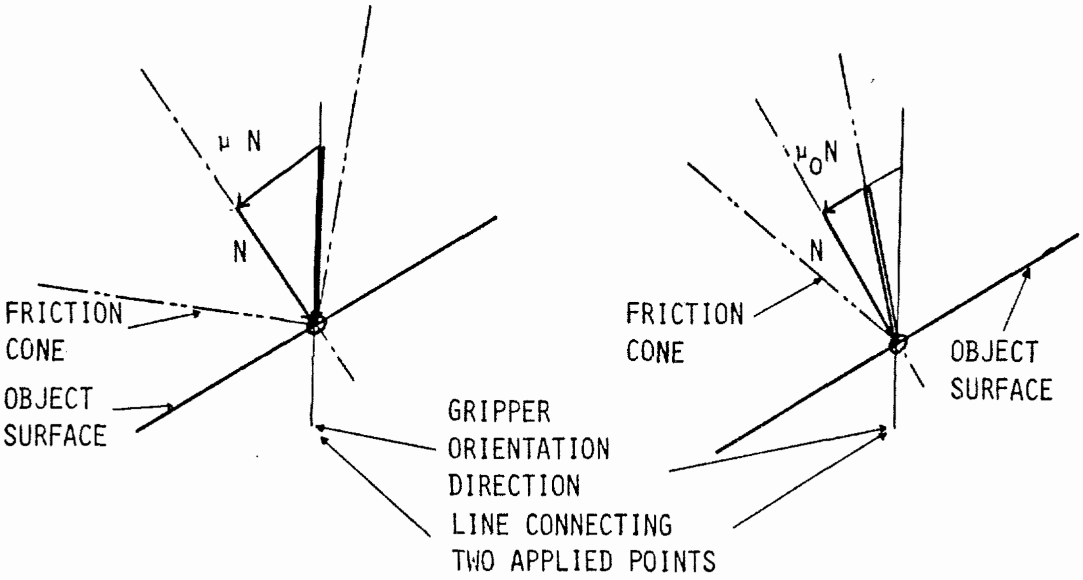

Figure 2. Friction cone and applied force. If the angle between the surface normal direction and the line connecting the two grasp points is less than half of the zenith angle of the friction cone, the direction of the force applied by the gripper coincides with the line connecting the two grasp points. Otherwise, the forces do not lie along the line, because the friction can only contribute $N\mu_0$ in the direction parallel to the surface, where N is the applied force perpendicular to the surface and μ_0 is the coefficient of friction.

angle, then the direction of the force applied by the gripper can agree with the line connecting the two points of contact (See Figure 2(a)). If the angle is larger, the force does not lie along that line (See Figure 2(b)), because friction can only contribute $N\mu_0$ in the direction parallel to the surface, where N is the normal force and μ_0 is the coefficient of friction. In cases where we cannot predict the magnitude of the friction angle, the most conservative solution is one in which the surface normals at the two contact points lie on the same line. This is a necessary and sufficient condition for satisfying conditions (1) and (2) in the absence of friction information.

2.4.2. Determining Legal Grasp Configurations from Object Shape

The next task is to extract legal grasp points by using the previous rule. This can be done by exploring the surface of the object. Let us assume that the surface normal direction at some point P can be determined. We will construct a line, in a direction opposite to that of the surface normal, and extend the line until it reaches the other side of the object. We will call the point reached Q. If the surface normal at the point Q agrees with the direction of the line, then the pair of points (P, Q) is added to the list of possible legal positions. (It is possible that no such pairs are found. In that case this simple algorithm decides that the object is ungraspable. Usually, however, there is an infinite number of point pairs satisfying this condition.)

For a smoothly curved object, the silhouette is of particular interest, since it can be determined from the image. There the surface normal is parallel to the image plane, and perpendicular to the silhouette in the image.

At some points—for example, at a crease in an object—the surface orientation may vary discontinuously with position on the surface. We cannot use such a point as the first point, P, in the above algorithm, because we cannot determine the surface orientation there. Such a point may, however, be used for grasping, *if* it happens to be found as the second point, Q', in the above algorithm, when starting from some other initial point P'.

Figure 3 shows examples of legal grasping points on various objects. At this stage, the gripper's shape is treated simply as a pair of points. The attitude in space of the gripper is not fully defined at this point; only the direction of the line between the two grasping points is known.

The gripper has another degree of freedom, in that it can rotate about the line connecting the two grasping points. The range of rotation about this axis is

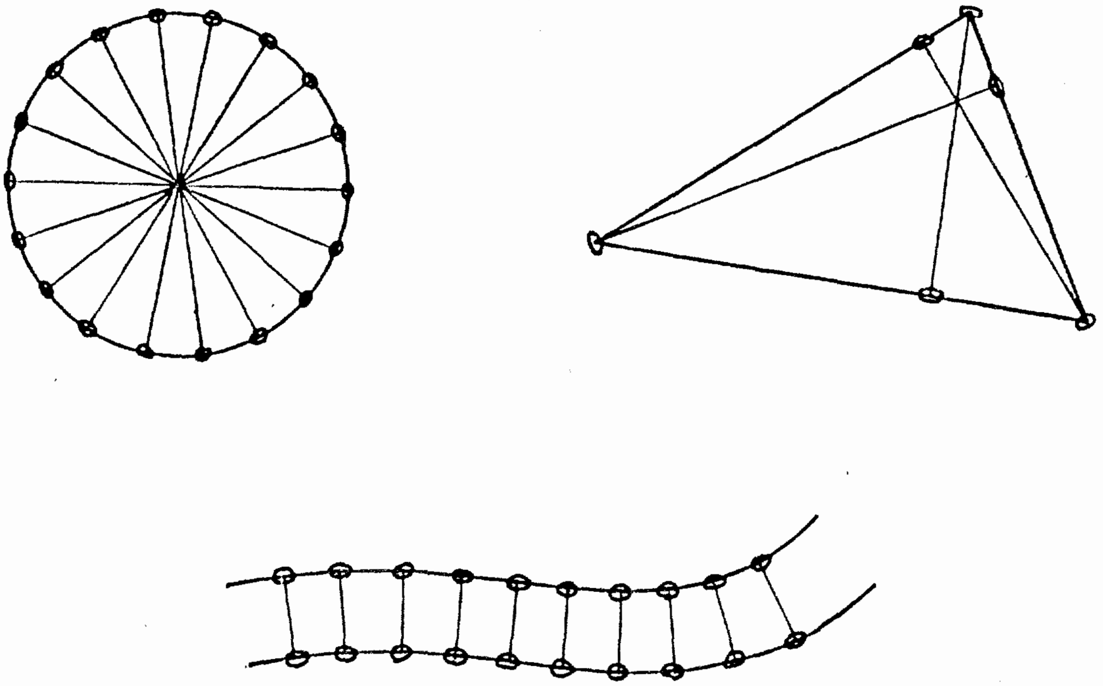

Figure 3. Examples of legal grasping points

constrained by the shape of the gripper and the shape of the object. We will call this degree of freedom the legal *rotation* of the gripper. The legal grasp configuration is a general name for the legal grasp points *and* the legal grasp rotation. If we use the point half way between the grasping points to represent the position of the gripper, then our legal grasp configuration becomes identical to Lozano-Perez's Legal Grasp Position (GSETS) [Lozano-Perez 76, 81].

2.5. Collision Free Configurations

Legal grasp configurations only describe the relationship between the gripper

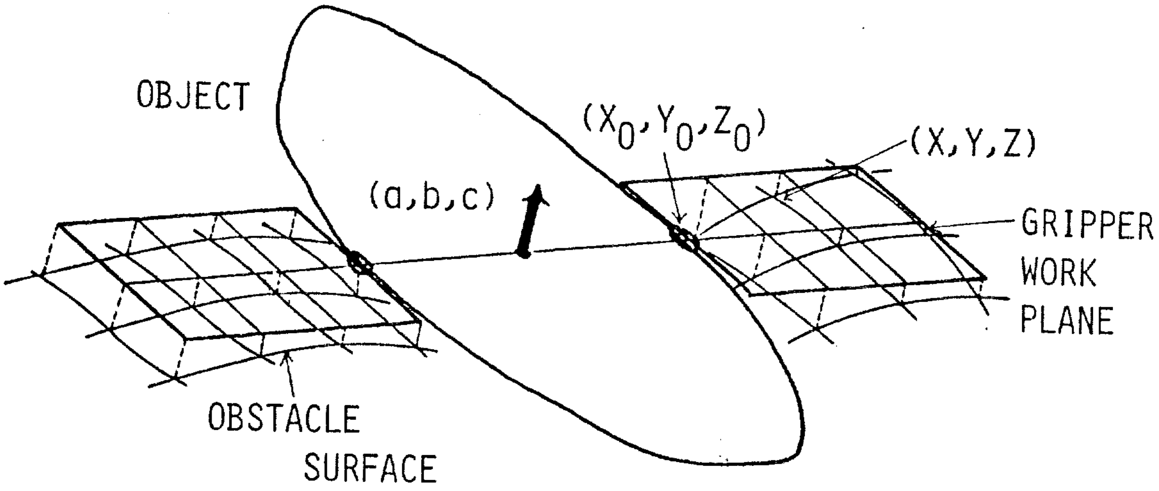

Figure 4. Gripper work space and obstacle surface.

and the object grasped. Among these legal grasp positions, we have to choose a grasp position which can be achieved without hitting other objects.

One approach to doing this is to use the method of configuration space obstacles (CSO) [Lozano-Perez 81, 83] which uses an equivalent representation in which the obstacles are enlarged and the gripper is reduced to a point. We do not follow this approach, however, since the number of neighboring obstacles in bin-picking tasks can be quite large and the computation of the CSOs correspondingly expensive. Also, the obstacles typically overlap and so individual CSOs must be combined to make composite CSOs.

Instead, we use a direct method. The central idea is to check every candidate grasp configuration among the legal grasp configurations, one after another, to see

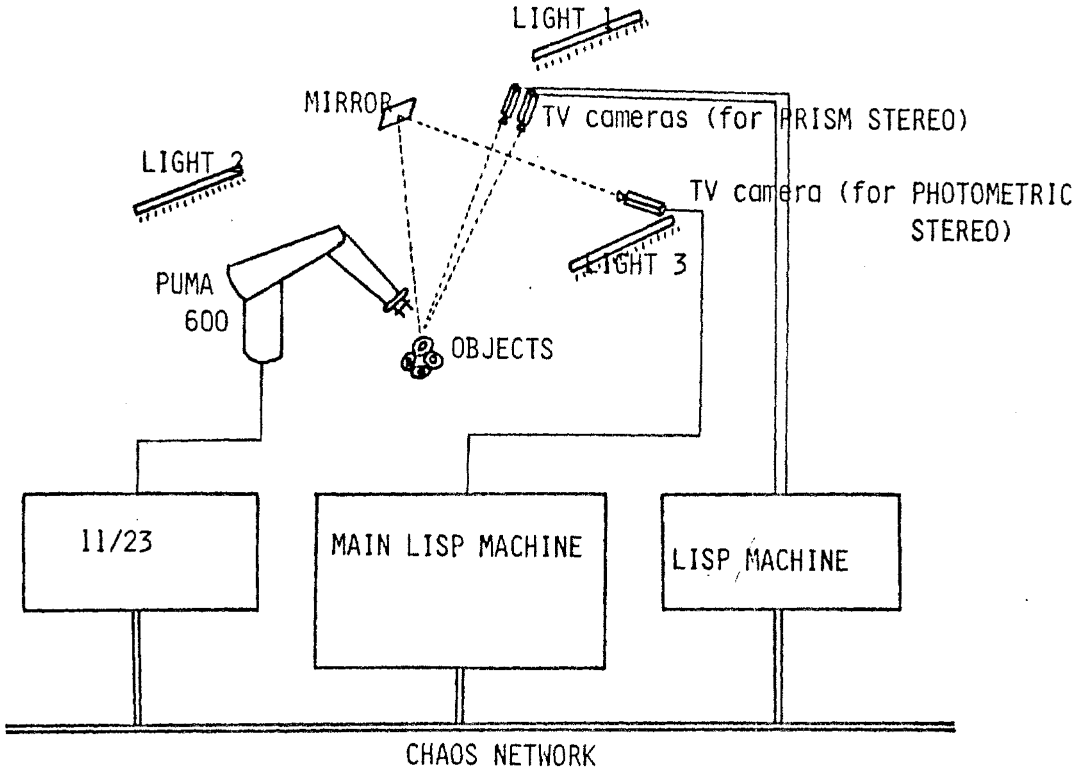

Figure 5. Hardware Configuration.

whether or not the gripper would hit an obstacle in that configuration.

The grasp motion sweeps out a pair of rectangular volumes which will be occupied by the fingers. The inner faces of these volumes pass through the legal grasp points; their orientation is determined by the legal grasp rotation; and their width and thickness correspond to the dimensions of the fingers. We will check whether these rectangular areas intersect the other objects or not.

Each of the rectangular areas lies in a plane:

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

where (x_0, y_0, z_0) is one of the legal grasp points, and $(-a, -b, -c)$ is the gripper approach direction. We check z values (elevation supplied by binocular stereo) within the two rectangular footprints to see that they are below this plane. If any point is not below the plane, the gripper will collide in that configuration. Conversely, if the left hand side of the above equation is less than zero for all points in the footprint, then the configuration is a collision free configuration (See Figure 4). One may even chose the best grasping configuration in the sense of the one where the highest point of the obstacles has the lowest height relative to the rectangular areas representing the gripper jaws.

3. System Details

The photometric stereo method and the matching of orientation histograms is implemented on a Lisp machine. This Lisp machine also controls the flow of execution. The PRISM stereo system is implemented on another Lisp machine running in parallel. Both Lisp machines, and the PUMA arm controller are connected via a local area network, the Chaos net, as shown in Figure 5.

The system has evolved through three generations, incorporating pickup point selection strategies of increasing sophistication. In the first system, the pickup point was selected without concern for possible collisions with neighboring objects and the range information was used only to set the height of the approach trajectory. The second and third versions use the PRISM elevation map to identify a collision free grasp point.

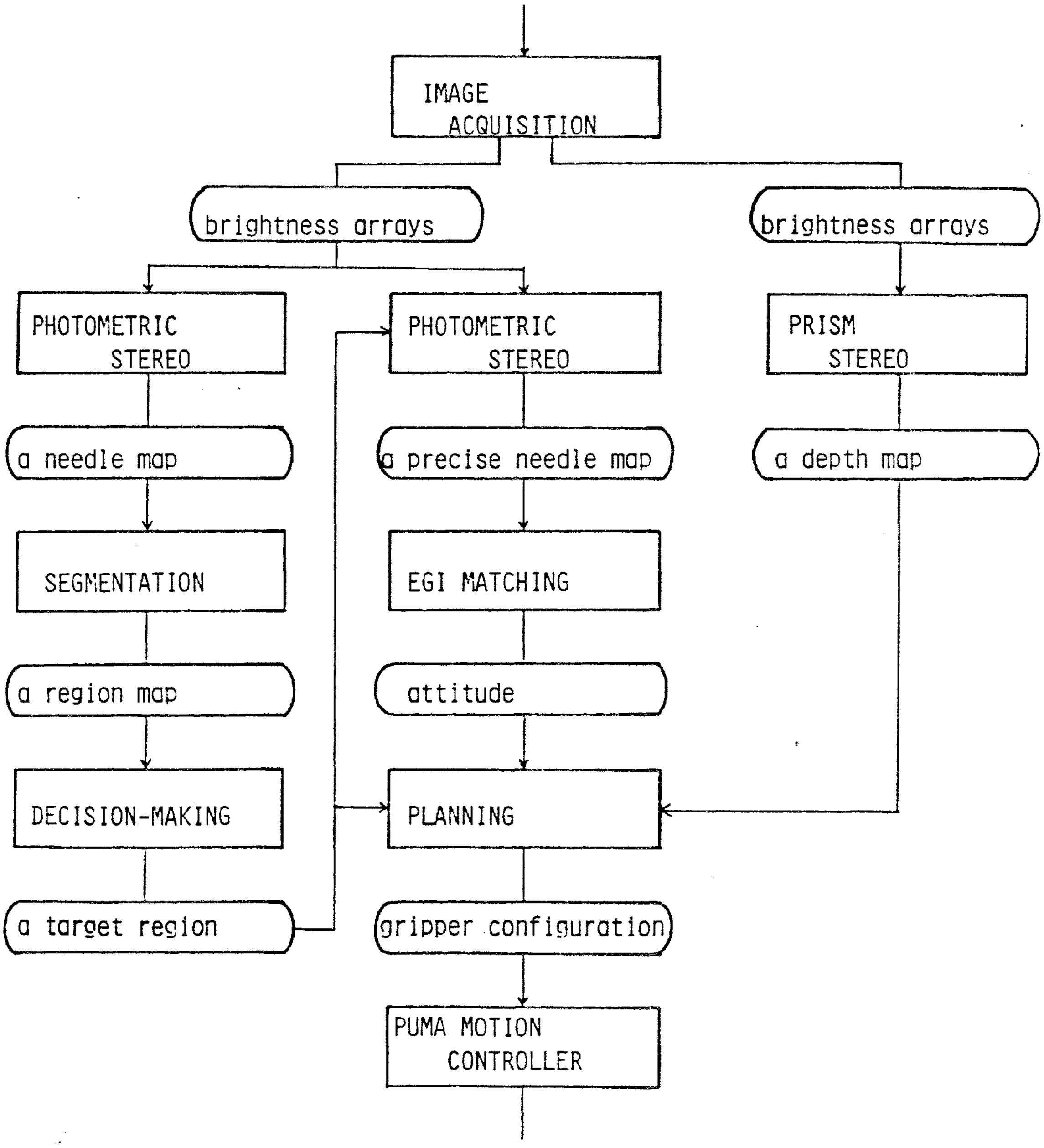

Figure 6. Information flow in the vision part.

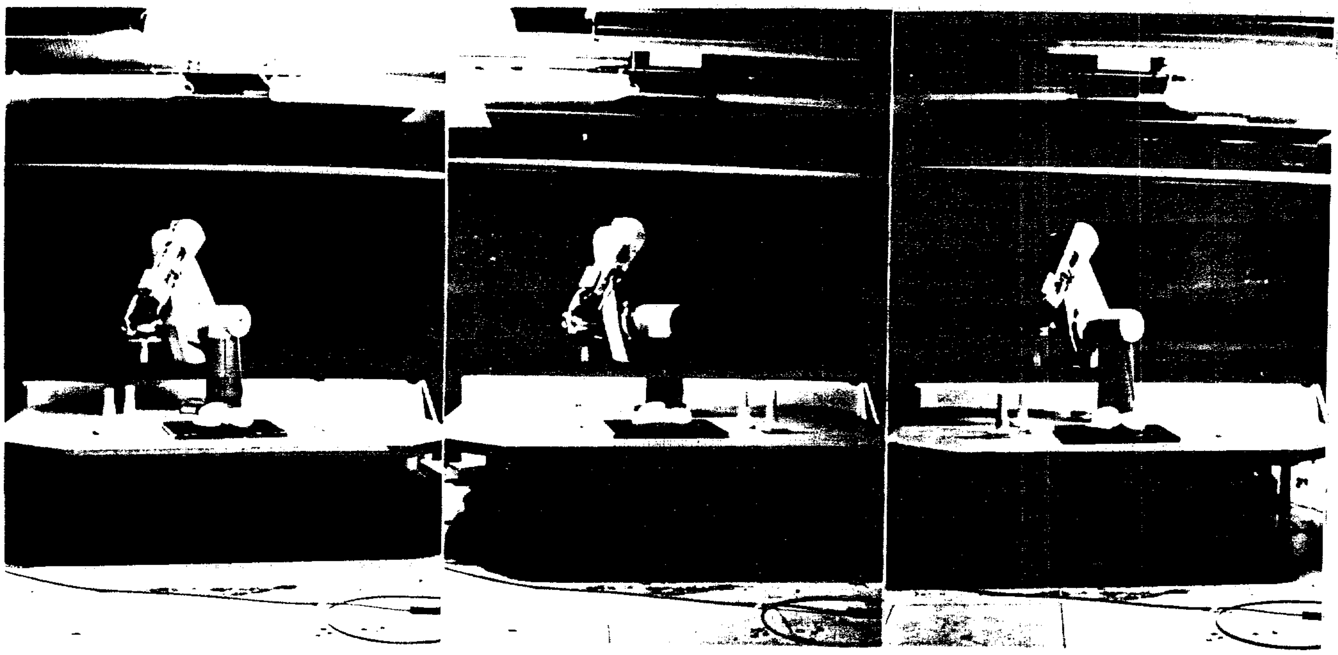

Figure 7. Light source for photometric stereo.

Figure 8. Light source for binocular stereo.

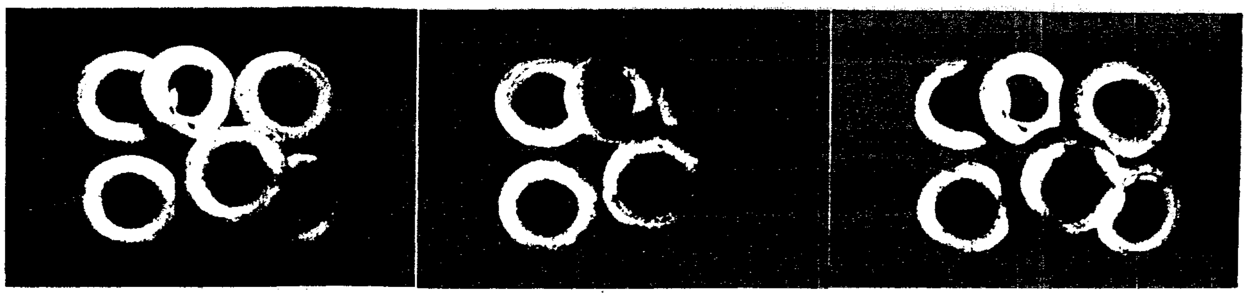

Figure 9. Three brightness arrays.

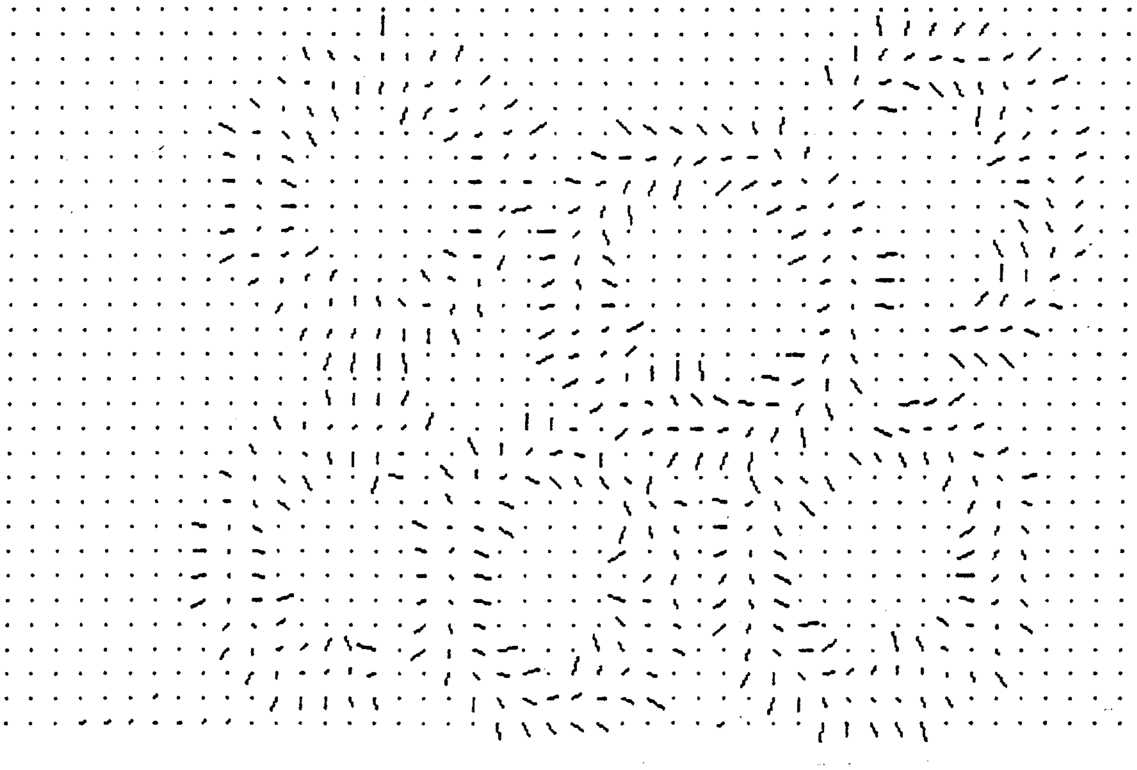

Figure 10. A needle diagram generated using photometric stereo.

3.1. Strategy 1

Choosing the highest point of the target object as the grasp point minimizes the likelihood of collision with neighboring objects. The position of the highest point is determined analytically, from the attitude of the object obtained using photometric stereo and matching of orientation histograms. Figures 6-18 illustrate the basic steps of this approach. Information flow in the vision part is shown in Figure 6.

(1) Three images are obtained of the scene with three different light sources (banks of ordinary fluorescent lamps) using a single CCD TV camera for photometric stereo (See Figure 7). A pair of images are also obtained under the random texture illumination, using a pair of vidicon TV cameras for the PRISM stereo system. (See Figure 8.)

(2) The photometric stereo module generates a needle diagram of the scene by means of the lookup table developed using a calibration object. Figure 9 shows three images obtained under different illumination conditions. Figure 10 shows the resulting needle diagram.

(3) The segmentation process divides the input scene into isolated regions based on the needle diagram. Segmentation is based on:

(a) areas where the surface normal varies discontinuously with position,

(b) areas where the system cannot determine surface orientation due either to shadowing or mutual illumination.

The isolated regions are shown in Figure 11.

(4) One target region is selected among the isolated regions based on the Euler number and the area of the region, as shown in Figure 12.

(5) The photometric stereo module is run again on the original image data, using a different lookup table, to obtain more detail in the regions near the edge of the target object. (One could actually use new images here, taken with different lighting conditions.) The result is used to produce an orientation histogram, which is the

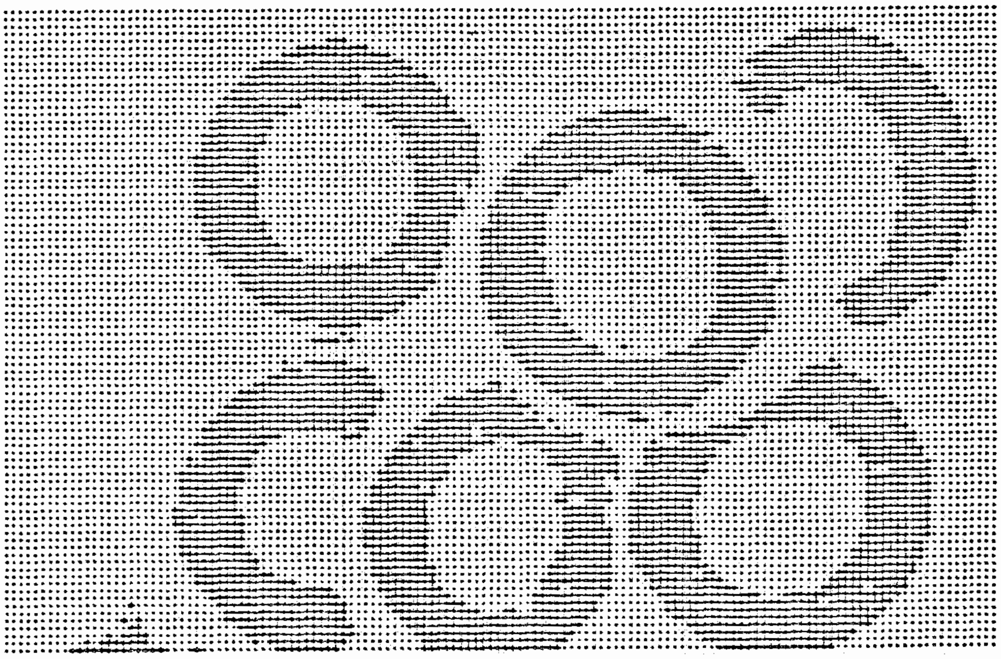

Figure 11. Isolated regions.

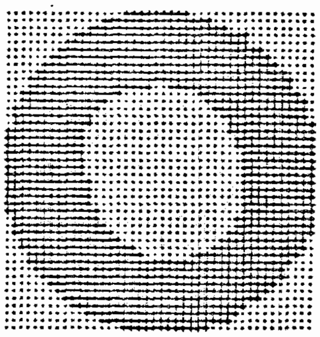

Figure 12. The target region selected among the isolated region.

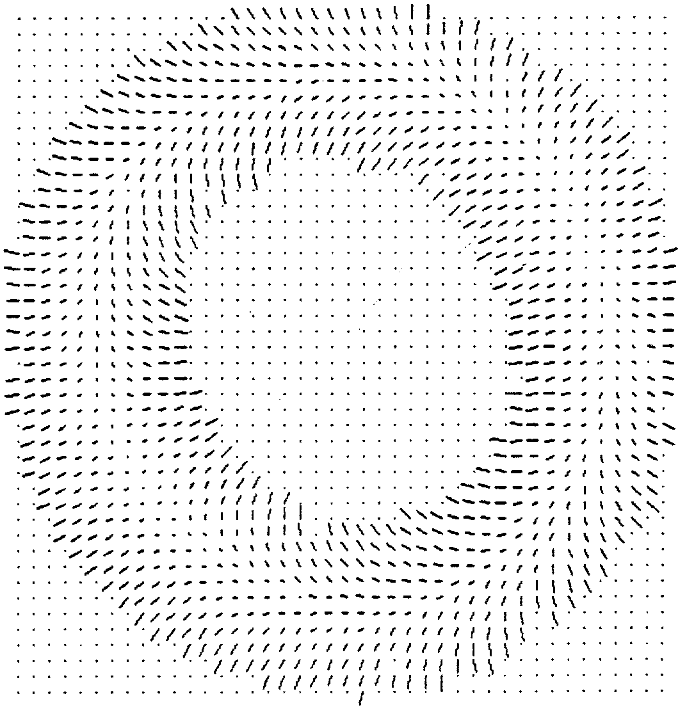

Figure 13. The detailed needle diagram over the target region.

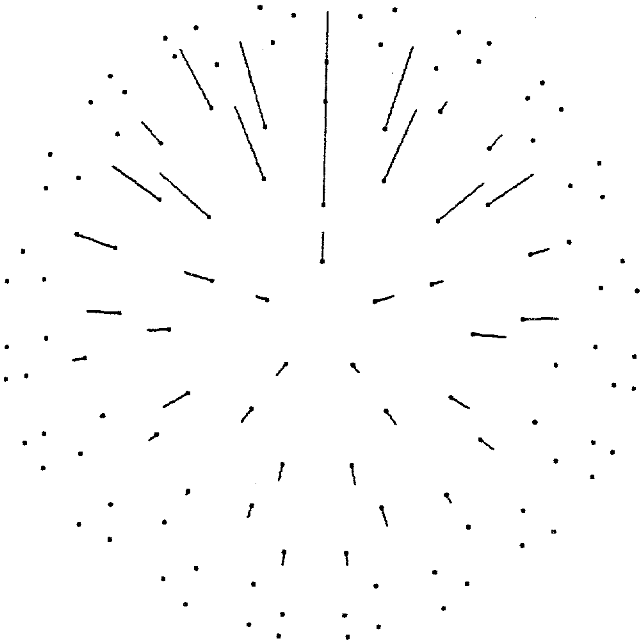

Figure 14. The EGI obtained from the needle diagram over the target region.

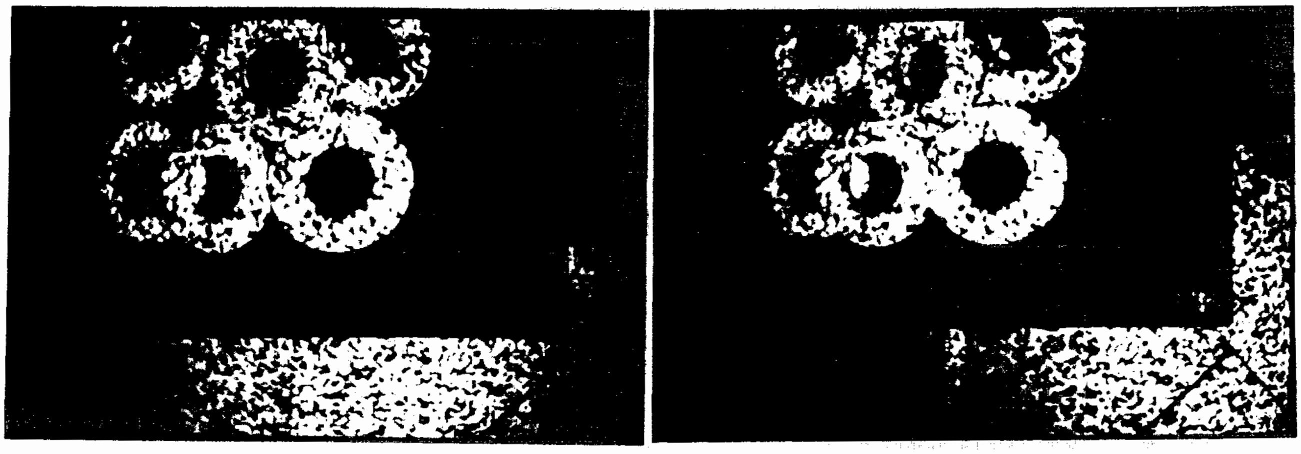

Figure 15. Stereo pair of brightness arrays with unstructured light illumination.

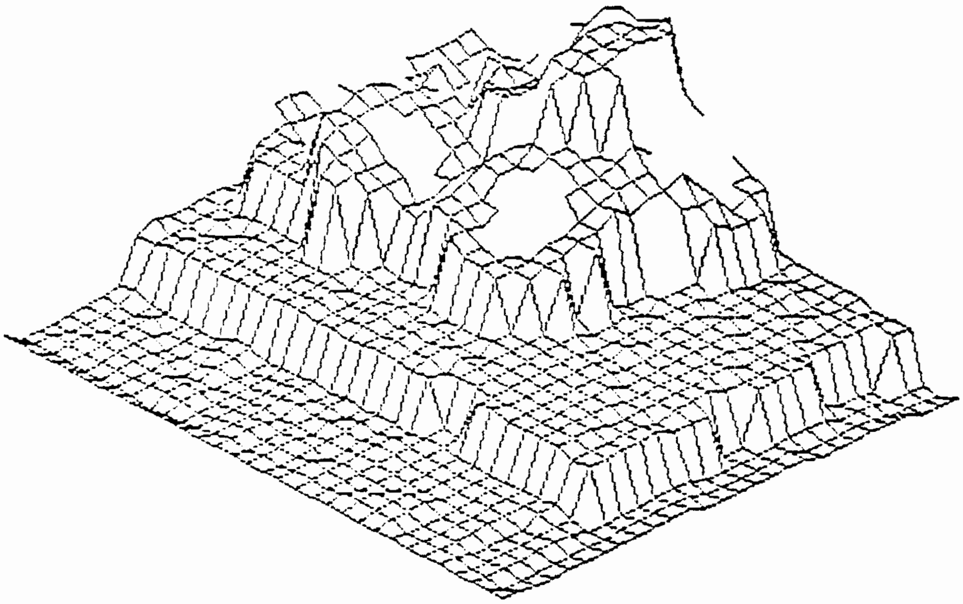

Figure 16. Output from the PRISM stereo module shown as a perspective plot.

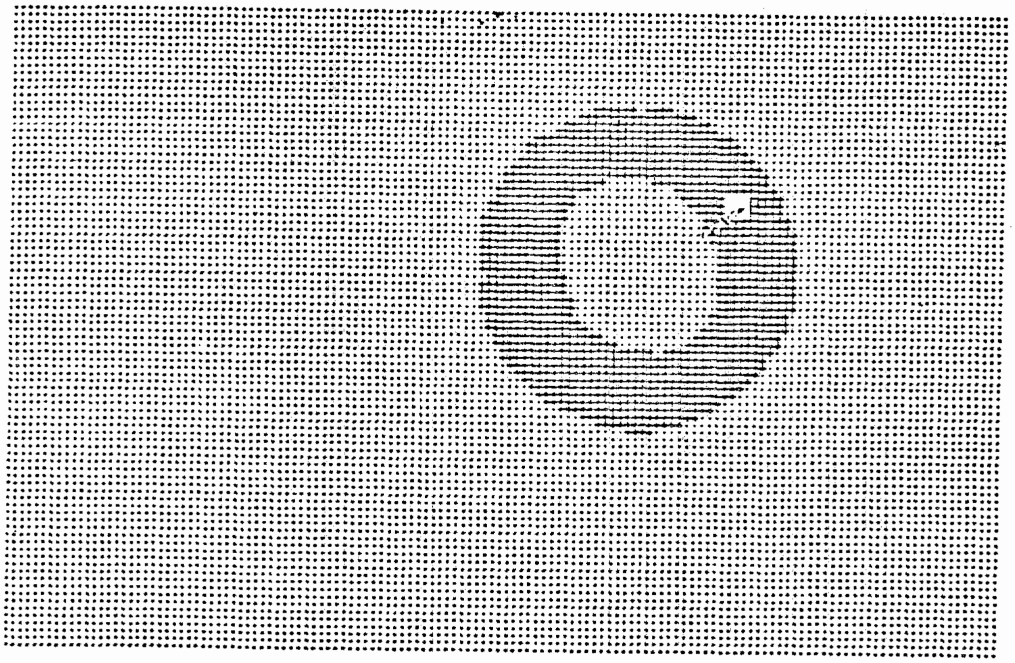

Figure 17. The pickup point.

discrete approximation of the EGI. Figure 13 shows the needle diagram produced over the target region.

(6) The EGI matching process compares the EGI obtained from the needle diagram with stored EGIs and determines the attitude of the object. Figure 14 shows the EGI obtained from data in the target region.

(7) In parallel with steps (2-6), the PRISM system produces an elevation map over the image. Figure 15 shows a pair of brightness arrays for the binocular prism stereo. Figure 16 shows the output of the PRISM stereo system as a perspective plot. A two-dimensional array containing these elevation measurements is sent to the main Lisp machine.

(8) The planner determines the pickup point by selecting the legal grasp point at the highest elevation as shown in Figure 17.

Figure 18 shows the execution of the pickup operation. Note that the manipulator approaches the doughnut shaped object directly from the initial

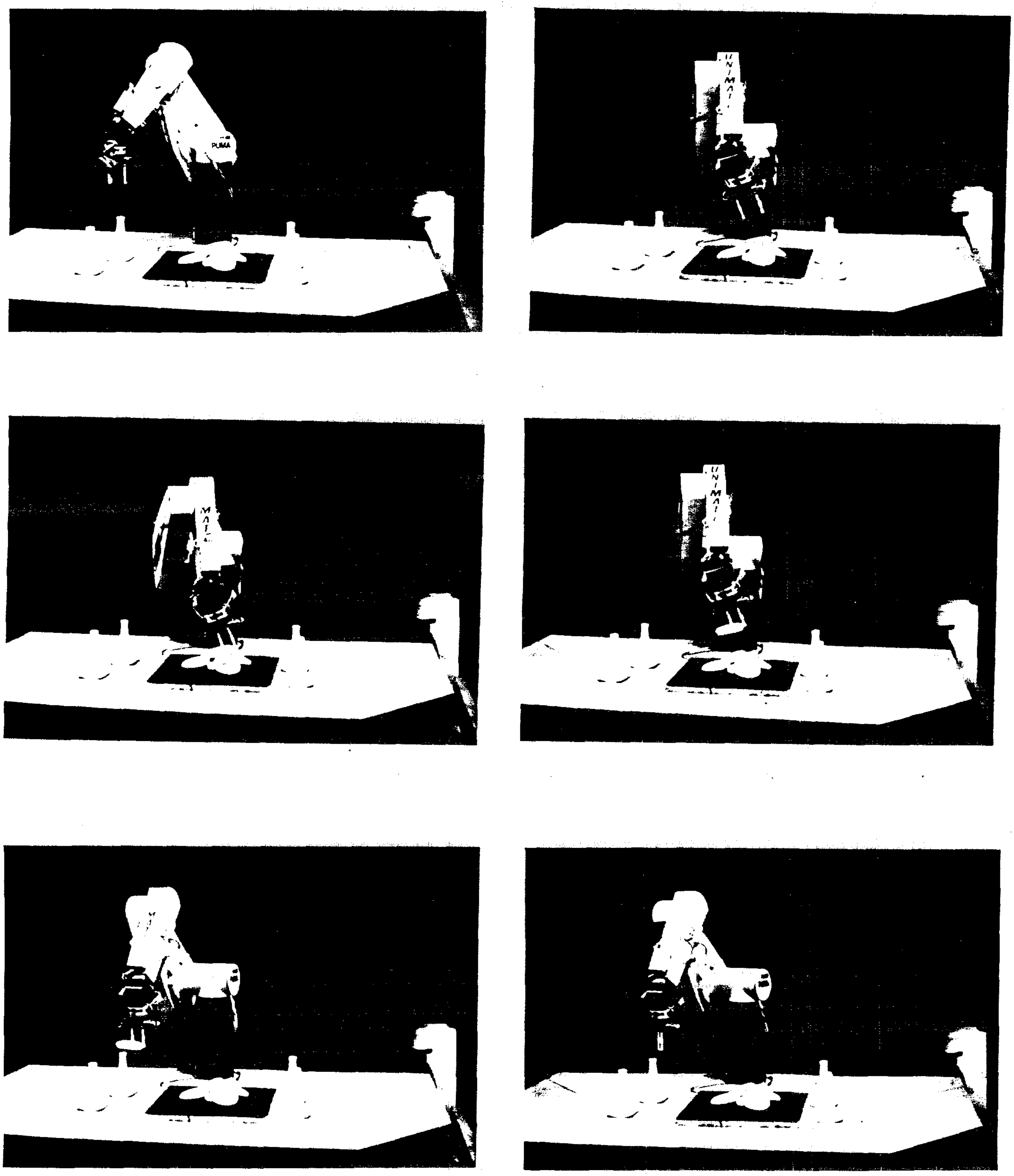

Figure 18. Pickup motion by the PUMA arm

configuration; the system described earlier required an additional arm motion [Ikeuchi&etal 83].

3.2. Strategy 2

While strategy 1 often identifies a collision free pickup point, it can easily fail, as is illustrated by the example in Figure 19. In order to insure the selection of a collision free grasping configuration, we need to take into account the height of neighboring objects. In our second strategy, we measure the finger clearance around proposed grasp points, using the elevation map provided by the PRISM module.

Our first task is to determine legal grasp configurations. In *this* strategy we model the doughnut shape as a two dimensional ring, and apply the method in Section 2.4 to this ring. Two classes of legal grasp positions are extracted, as shown in Figure 20. Since legal grasp positions of class 1 (Figure 20(a)) require too large a gripper opening, they are discarded. Legal grasp position of class 2 (Figure 20(b)), on the other hand, can be used. They can be specified by the rotation angle around the approach direction. Note that in this strategy, unlike the next one, only the direction perpendicular to the plane of the doughnut is considered to be a legal approach direction.

The next task is to determine legal grasp positions using the observed data. In this strategy, legal grasp positions occur only along the silhouette of the object. Fortunately, the silhouette of the object has already been extracted by the segmentation process.

Each legal grasp position is specified relative to the center of the target image. The direction from this center reference point also corresponds to the orientation of the line connecting the grasp points. This also gives us the rotation of the rectangular areas corresponding to the jaws of the gripper around the approach direction.

For each legal grasp position we check the corresponding rectangular regions for the distance to which the fingers can be moved past the plane of the doughnut before a collision occurs. The equation requires (x, y, z), (x_0, y_0, z_0), and a normal to the plane of the doughnut. Since the approach direction, (a, b, c), is here perpendicular to the doughnut plane, it is determined directly from the attitude of the doughnut obtained by the EGI matching process. (A common value could be used for (x_0, y_0, z_0), because the legal grasp points lies on the doughnut plane. But, we measure this value for each candidate grasp point.)

Figure 21 shows a profile of the highest points over the rectangular area of the gripper footprint with respect to the doughnut plane. (Highest here means the largest value of $a(x-x_0)+b(y-y_0)+c(z-z_0)$.) If the lowest of these values is below the doughnut plane, the gripper can pick up the doughnut using the corresponding configuration. Figure 22 shows the optimal grasp point so determined.

Figure 23 shows a pickup sequence using our second strategy on an example which would have resulted in a collision if we had used the first strategy. The program determines the configuration which has the greatest finger clearance relative to the doughnut plane. Figure 23a indicates the point selected by the first strategy and Figure 23b shows the result using the second strategy.

3.3. Strategy 3

The doughnut in the middle, in the example shown in Figure 24, could not be picked up using strategy 2. That doughnut is surrounded by obstacles and there is no position around its circumference with sufficient clearance for the fingers to get below the plane of the doughnut. In cases like this, it is still possible to find a legal grasp point, but it is necessary now to model the doughnut as a three-dimensional object. With this extension, there are three classes of legal grasp configurations to consider, namely those shown in Figure 20 and the additional one shown in Figure 25.

The legal grasp configuration can be characterized using two parameters, α and β. The first parameter, α, denotes the rotation around the axis of the doughnut and β indicates the rotation of the line connecting the grasping points relative to the plane of the doughnut. Strategy 2 corresponds to the case where β is zero. In our third strategy, we allow the gripper approach direction (a, b, c) to be specified over a range of β values relative to the attitude of the doughnuts.

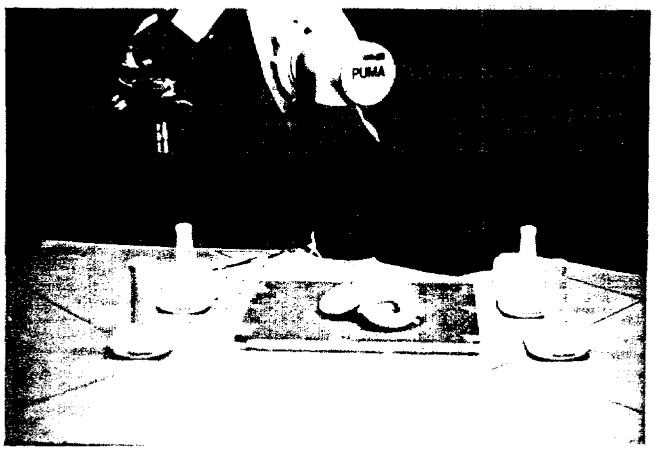

Figure 19. A more difficult case where strategy 1 would fail.

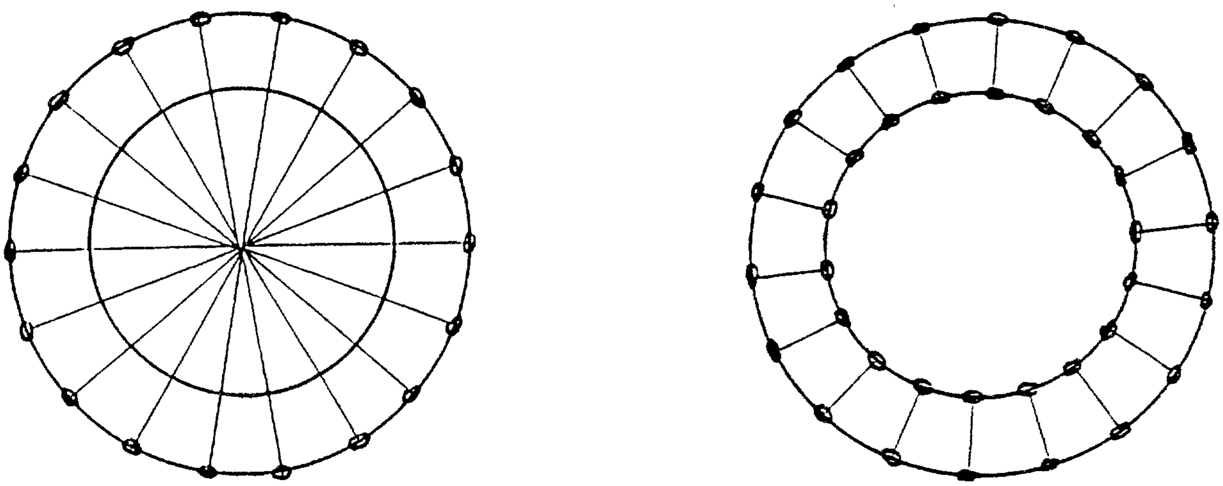

Figure 20. The legal grasp points of a doughnut.

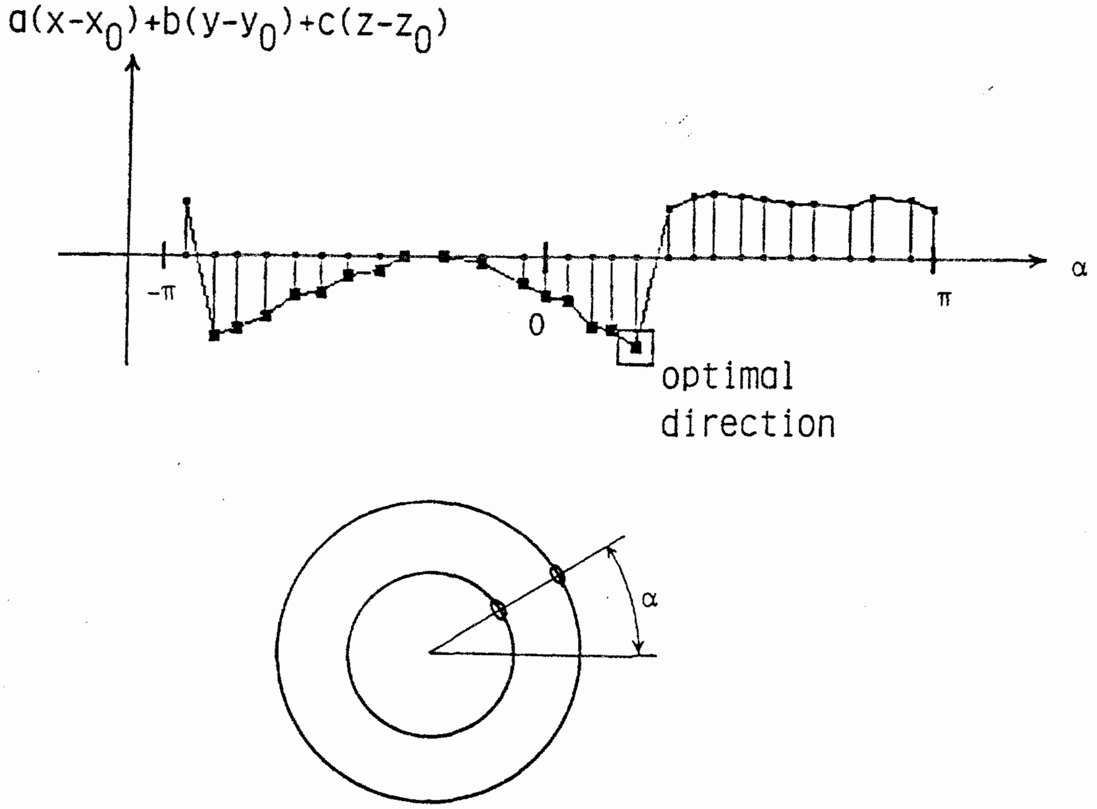

Figure 21. Profile of highest points over the rectangular area of the gripper work space, plotted as a function of rotation around the center of the doughnut

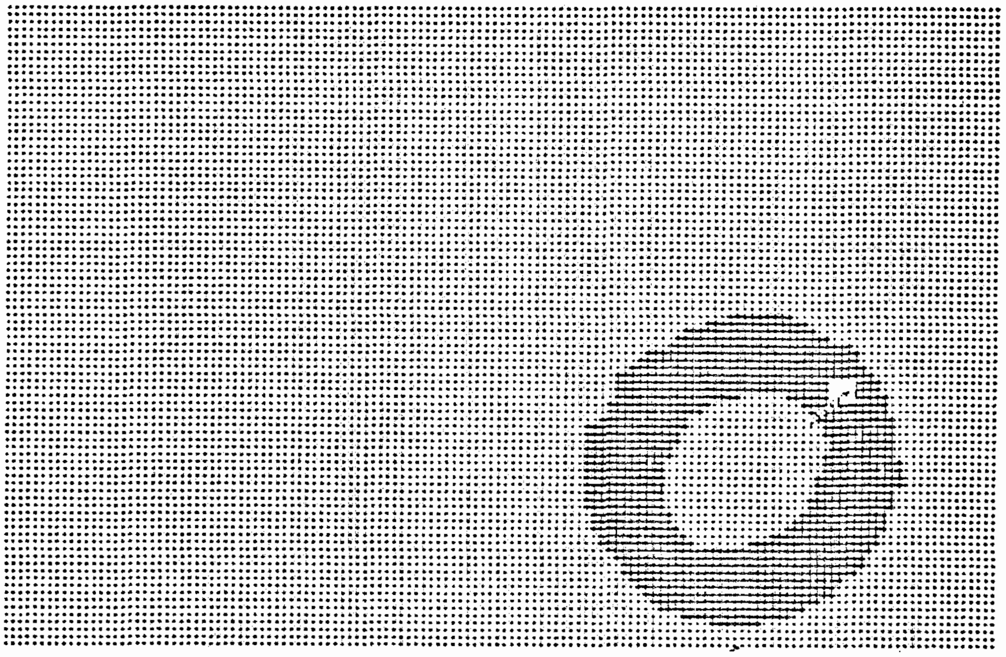

Figure 22. The grasp point determined.

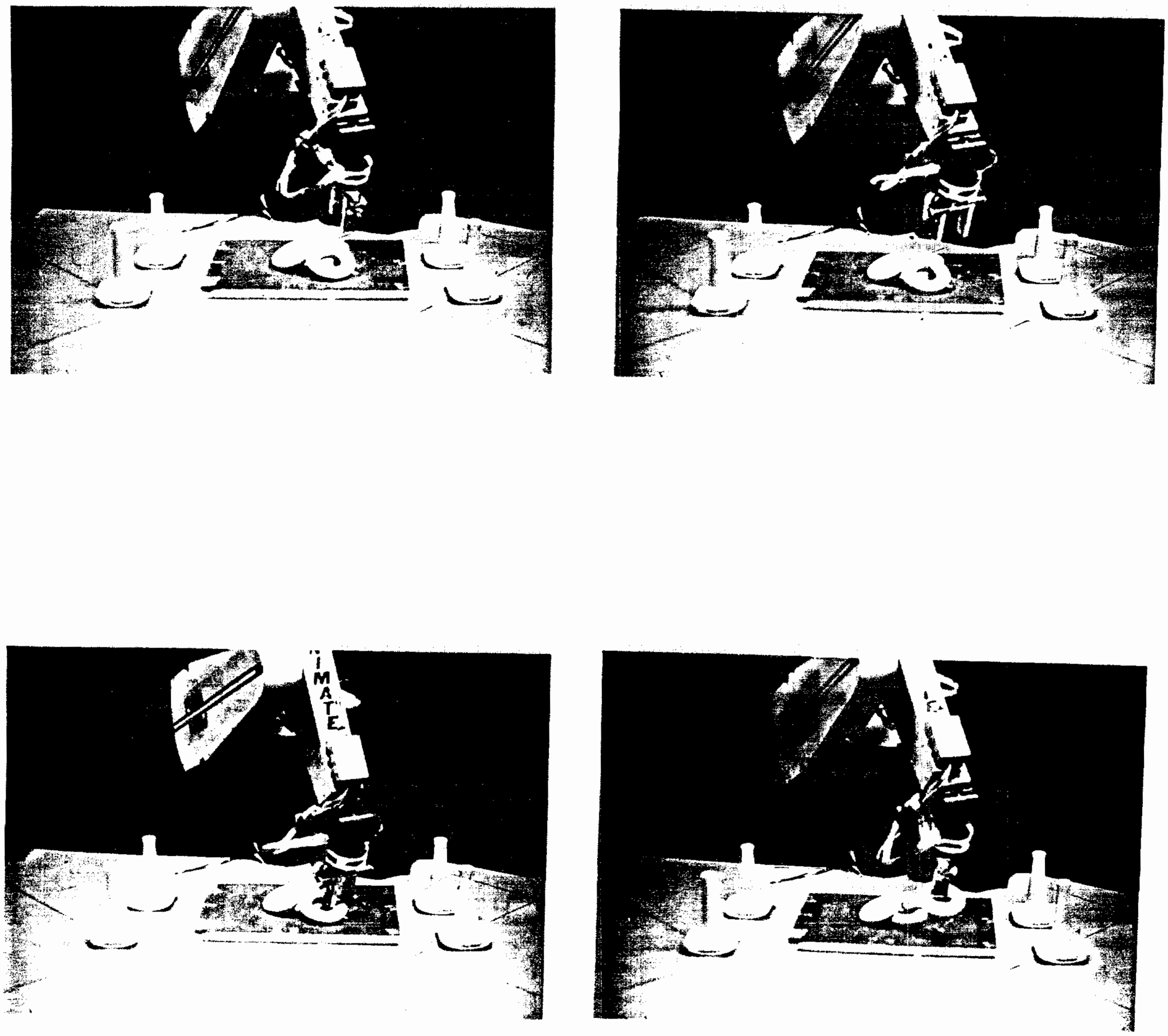

Figure 23. The pickup motion determined by strategy 2.

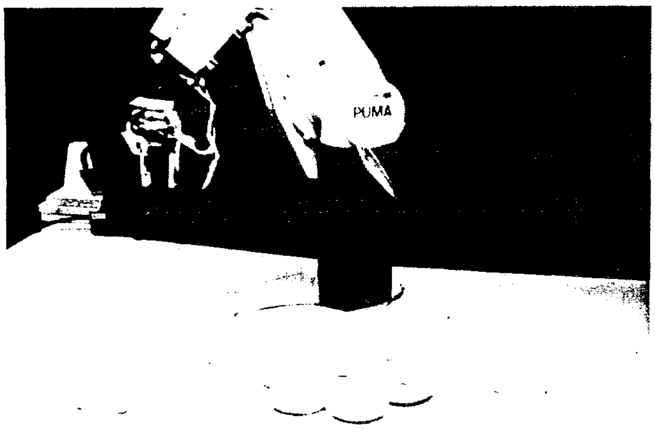

Figure 24. A situation where strategy 2 would fail.

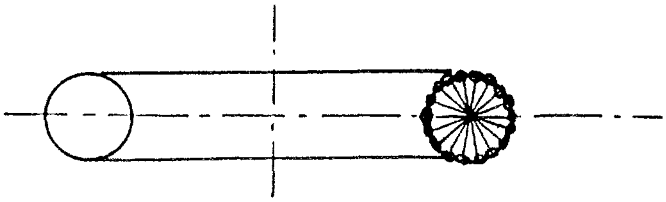

Figure 25. Additional legal grasp points for a 3-D doughnut (cross-sectional view).

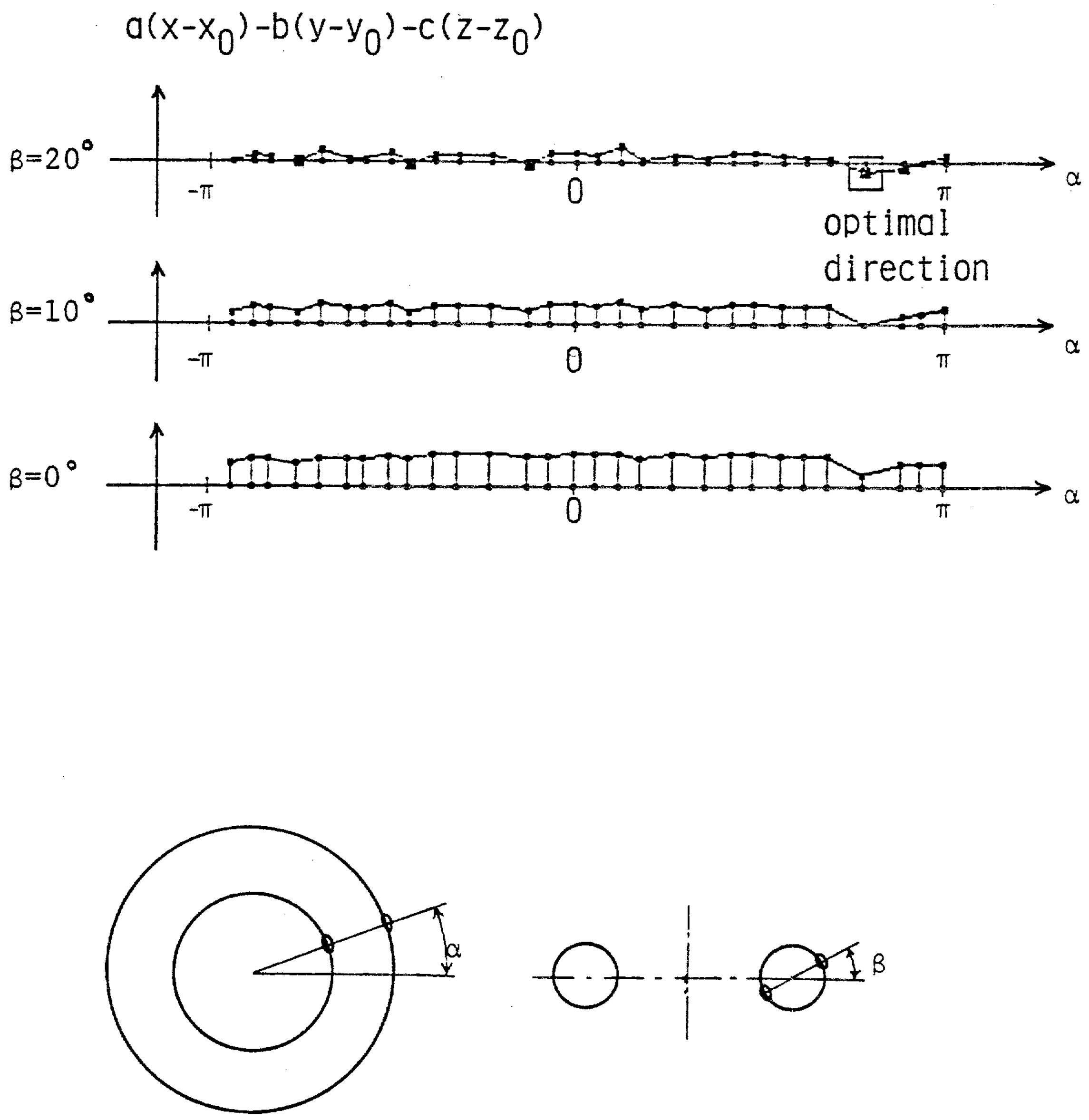

Figure 26. Obstacle height profiles around doughnuts versus α for three choices for *beta*.

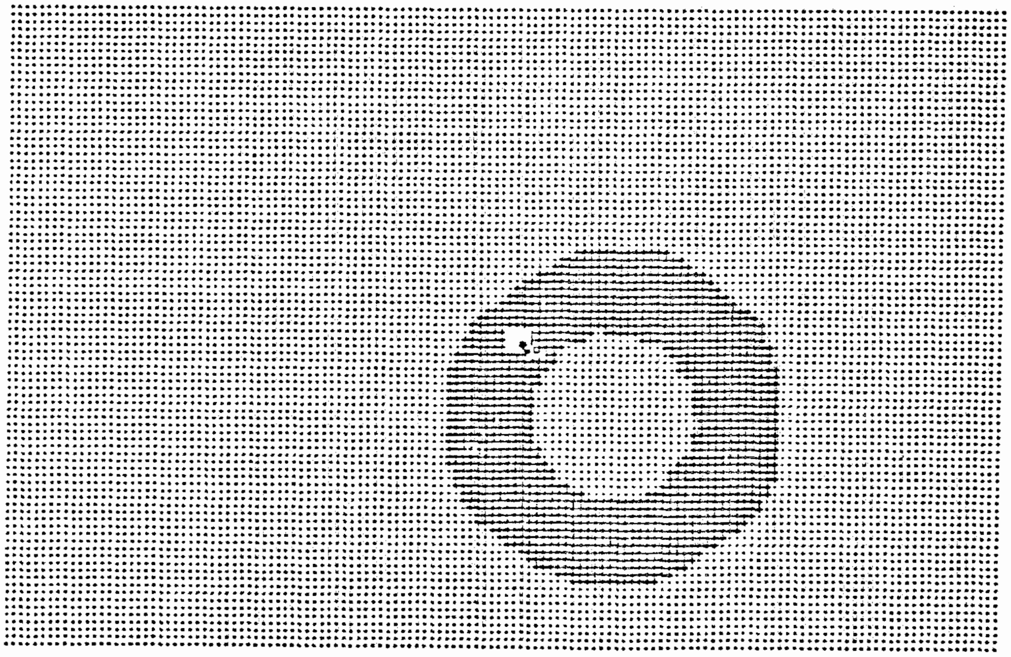

Figure 27. The pickup point selected.

Figure 26 shows the profile of the highest points with respect to the work space plane. In this case, the planner finds a collision free configuration at $\beta = 20°$. Figure 27 shows the grasp point selected.

Figure 28 shows the process of picking up a doughnut without collision. Figure 28a is the original grasp position. In Figure 28b the gripper is rotated around the axis of the doughnut by α. Then, in Figure 28c the line connecting the grasp points is rotated by β, relative to the plane of symmetry of the doughnut.

4. Summary

We have described a hand-eye system which performs bin-picking tasks. Four basic modules are used: photometric stereo, binocular stereo using the PRISM

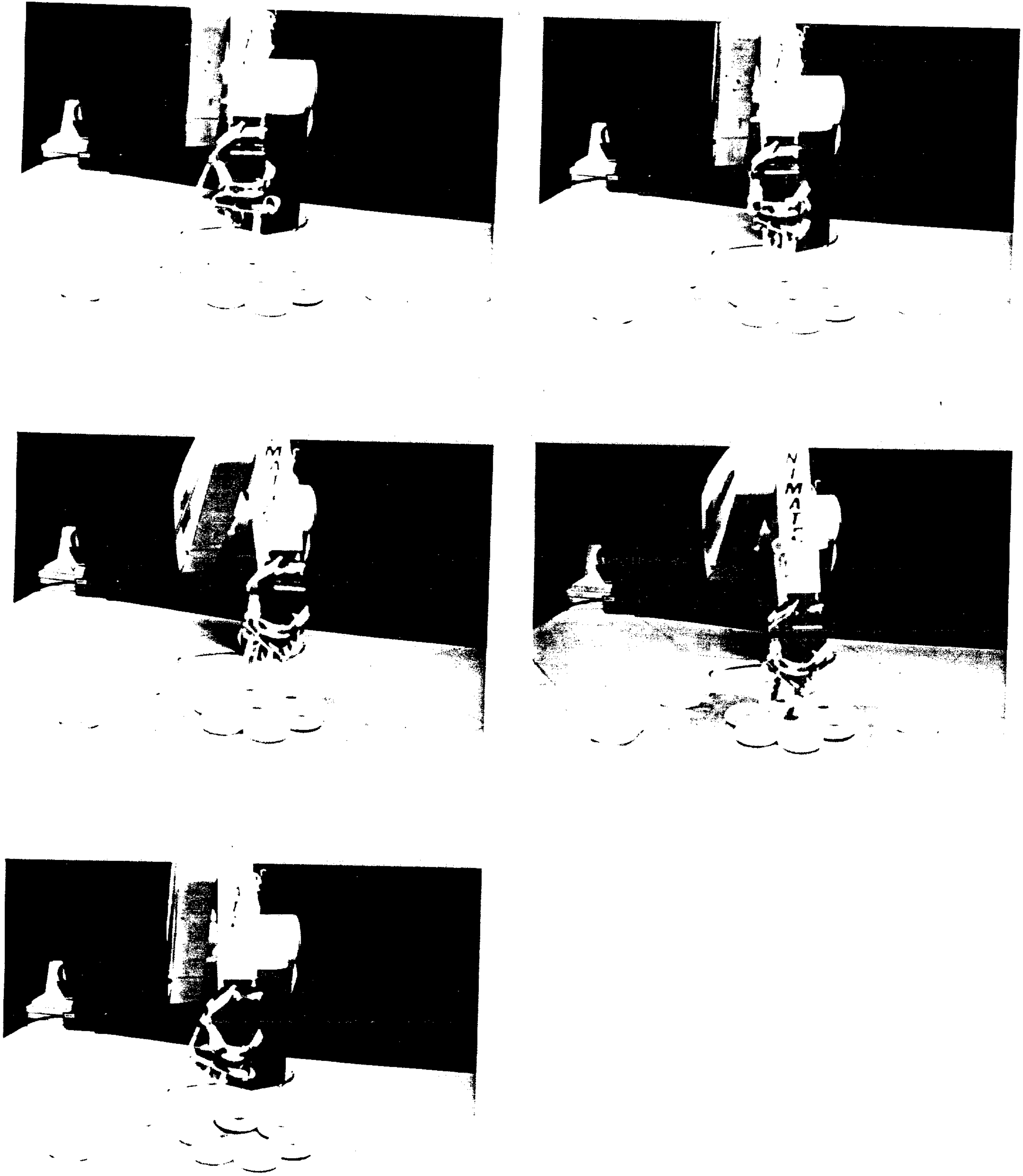

Figure 28. The pickup motion.

algorithm, extended Gaussian image matching, and collision-free configuration planning for the gripper.

Photometric stereo determines the orientation of surface patches corresponding to each picture cell, based on the brightness values in three images, obtained using different light sources. Segmentation is based on a needle diagram, the smoothness constraints, shadow areas, and mutual illumination. The attitude in space of the object is determined by comparing the orientation histogram of the object's surface with stored orientation histograms of prototypes. The orientation histogram is a discrete approximation of the extended Gaussian image. An elevation map produced by the PRISM stereo algorithm is used to determine object elevation and to check finger clearance at the proposed grasp configurations.

The system has unstacked piles of objects successfully and is able to find good pickup points in complex piles of doughnuts. The hybrid vision system cycles in less than a minute running on two (MIT) Lisp-machines—including the time for image acquisition. The entire system is written in Zeta-Lisp, a dialect of Lisp, and is compiled into "machine code" using the Zeta-Lisp compiler.

The two "low-level" vision modules produce reliable but restricted information about the visible surfaces imaged. In one case, high resolution local surface orientation measurements, and in the other, absolute height measurements at a lower spatial resolution. We have combined these two systems to produce a system that takes advantage of both, to solve a problem that neither system could solve well alone. The fine local surface orientation information allows us to locate, identify, and orient an object out of a bin of other objects. The local elevation information allows us to calculate the three-dimensional position of the target object, and to check proposed grasp points for collisions with neighboring objects. Both the photometric and PRISM stereo modules have simple kernels that can easily be adapted for use in other problems and lend themselves to high speed implementation on special purpose hardware.

5. Acknowledgment

The following people also helped in part on this project: Noble Larson constructed the CCD-TV camera interface. John Purbrick built the fingers of the gripper. Oded Feingold and John Cox made the LED sensor and collision sensor. Tom Callahan prepared the lighting stage and general setup. Without their effort, this project could not have been completed.

Tomas Lozano-Perez provided many useful comments which have improved the readability of this paper. Finally, thanks go to Ikko for preparing this manuscript and some of the drawings.

6. References

Bajcsy, R. 1980. "Three-dimensional Scene Analysis," *Proc. 5th-ICPR*, Miami Beach, pp. 1064–1074.

Ballard, D.H. and Sabbah, D. 1981. "On Shapes," *Proc. 7th-IJCAI*, Vancouver, pp. 607–612.

Brady, M. 1982. "Parts Description and Acquisition Using Vision," *Proc. SPIE*, Vol. 336, Robot Vision, pp. 20–28.

Coleman, E.N. and Jain, R. 1981. "Shape from Shading for Surfaces with Texture and Specularity," *Proc. 7th-IJCAI*, Vancouver, pp. 652–657.

Do Carmo, 1976. *Differential Geometry of Curves and Surfaces*, Englewoods Cliffs, Prentice-Hall.

Hanafusa, H. and Asada, H. 1977. "Stable Prehension by a Robot Hand with Elastic Fingers," *Proc. 7th-ISIR*, pp. 361–368.

Horn, B.K.P. 1975. "Obtaining shape from shading information," In *The Psychology of Computer Vision.*, P.H. Winston (ed.) pp. 115–155, McGraw-Hill.

Horn, B.K.P. 1977. "Image Intensity Understanding," *Artificial Intelligence*, Vol. 8, No. 2, pp. 201–231.

Horn, B. K. P., Woodham, R. J. and Silver, W. M. 1978. "Determining Shape and Reflectance using Multiple Images," *AI Memo No. 490,* Cambridge, MIT, AI Lab.

Horn, B.K.P. and Sjoberg, R.W. 1979. "Calculating the Reflectance Map," *Applied Optics,* Vol. 18, pp. 1770–1779.

Horn, B. K. P. 1983. "Extended Gaussian Images," *AI Memo No. 740,* Cambridge, MIT, AI Lab.

Ikeuchi, K 1981a. "Recognition of 3D object using Extended Gaussian Image," *Proc. 7th-IJCAI,* Vancouver, pp. 595–600.

Ikeuchi, K. 1981b. "Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method," *IEEE Trans. on PAMI,* Vol. PAMI-2, No. 6, pp. 661–669.

Ikeuchi, K. 1983. "Determining Attitude of Object from Needle map using Extended Gaussian Image," *AI Memo No. 714,* Cambridge, MIT, AI Lab.

Ikeuchi, K., Horn, B., Nagata, S., Callahan, T., and Feingold, O. 1983. "Picking up an Object from a Pile of Objects," *AI Memo No. 726,* Cambridge, MIT, AI Lab.

Lozano-Perez, T. 1976. "The Design of a Mechanical Assembly System," *AI-TR-397,* Cambridge, MIT, AI Lab.

Lozano-Perez, T. 1981. "Automatic Planning of Manipulator Transfer Movements," IEEE Transactions on System, Man, and Cybernetics, SMC-11, 681-698.

Lozano-Perez, T. 1983. "Spatial Planning: A Configuration Space Approach," IEEE Transactions on Computers, C-32, 108-120.

Marr, D. & Hildreth, E. 1980. "Theory of Edge Detection," Proc. R. Soc. Lond. B. Vol. 207, 187–217.

Marr, D. & Poggio, T. 1979. "A Computational Theory of Human Stereo Vision," Proc. R. Soc. Lond. B, Vol. 204, 301–328. Also available as MIT Artificial Intelligence Lab Memo 451.

Nishihara, H.K. 1984. "PRISM: A Practical Realtime Imaging Stereo Matcher. Optical Engineering (in press), also available as MIT Artificial Intelligence Lab Memo 780.

Nishihara, H.K. and Larson, N.G. 1981. "Towards a Real Time Implementation of the Marr-Poggio Stereo Matcher," Proceedings A.R.P.A Image Understanding Workshop, Baumann, ed., Science Applications, Inc. April.

Nishihara, H.K. and Poggio,T. 1983. "Stereo Vision for Robotics," Proceedings International Symposium of Robotics Research. Bretton Woods, N.H. to appear in M.I.T Press.

Smith, D.A. 1979. "Using Enhanced Spherical Images for Object Representation," *AI Memo No. 530,* Cambridge, MIT, AI Lab.

Silver, W. A. 1980. "Determining Shape and Reflectance using Multiple Images," *MS Thesis*, Cambridge, MIT, EECS.

Woodham, R. J. 1978 "Photometric Stereo: A Reflectance Map Technique for Determining Surface Orientation from a Single View," *Image Understanding Systems and Industrial Applications*, Proceedings SPIE 22nd Annual Technical Symposium, Vol. 155, pp. 136-143, August.

Woodham, R.J. 1979. "Reflectance Map Techniques for Analyzing Surface Defects in Metal Casting," *AI-TR-457,* Cambridge, MIT, AI Lab.

Woodham, R. J. 1980 "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Engineering*, Vol. 19, No. 1, Jan/Feb., pp. 139-144.