Planning Multiple Observations for Specular Object Recognition

Keith D. Gremban and Katsushi Ikeuchi

School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213

Abstract

Specular objects present a special challenge for computer vision systems. The most prominent features of specular objects are the specularities, which are highly variable and dependent on local object geometry. In order to unambiguously recognize specular objects, more information is required. One approach to obtaining additional information is to use multiple observations from different viewpoints. This paper presents initial results in planning observations for specular object recognition.

1 Introduction

Specular objects, which are those with shiny, mirror-like surfaces, present a special challenge for computer vision systems. Most computer vision systems perform object recognition on the basis of the information contained in a single image. Implicit in this paradigm is the assumption that the three-dimensional structure of the object is unambiguously represented by the features detectable in the image, which is not true in the specular domain.

Specularities in optical images are often the most prominent features. However, specularities are strictly local phenomena that are very unstable, and may not even be present in all images from arbitrary viewpoints. Therefore, specularities are generally considered to provide insufficient information for object recognition. More generally, there are a number of sensor systems that are "impoverished" in that a single observation does not provide complete information about object geometry. It is important to establish techniques to recognize objects using specular reflections and other kinds of exotic or unconventional sensors.

Our approach to overcoming the problem of insufficient information is to utilize multiple sensor observations obtained from different viewpoints, and combine the information from each observation using knowledge of the 3D structure of the objects. Using multiple observations, two objects that appear the same from one viewpoint can be distinguished as long as there is at least one viewpoint from which the observed appearances are different. For complex sets of objects, more than a single additional observation may be required to resolve ambiguity.

2 Related Work

Ikeuchi and Kanade [5] described a general approach to automatic recognition program generation that emphasized the importance of modeling sensors as well as objects. Hong et al [3] extended the work of Ikeuchi and Kanade by optimizing the object recognition code generated.

Sato, et al [8] presented a model based system for recognition of specular objects. However, the system did not adequately deal with the problems of ambiguity that result from different aspects with similar appearances.

Hutchinson and Kak [4] addressed the problem of resolving ambiguous sensor information. Their approach involved making plans dynamically, based on the system's current information about the scene.

Liu and Tsai [7] described a system for recognizing objects from multiple observations. Recognition was performed by matching features extracted from 2D silhouettes against the features extracted from model objects using a fixed set of camera views.

3 Planning Multiple Observations

Rather than consider the most general case, we restrict the problem to the simplest possible interesting case, that in which the object and sensor each have only a single degree of freedom of motion. In particular, the object position is fixed, as is the distance between the sensor and object. Both the object and the sensor can rotate about the same axis. Hence, the system as a whole has a single degree of freedom. This case is simple, but is still similar to a variety of application problems, and, as will be seen, is still fairly complex to solve.

Specularities are visible within relatively small ranges of viewpoints. It is convenient to group together viewpoints into sets, which we call aspects. An aspect, as formally defined in [6], is a set of topologically equivalent object appearances. In this paper, we relax the definition somewhat

This research was supported by the Defense Advanced Research Projects Agency (DOD) and monitored by the Avionics Laboratory, Air Force Wright Aeronautical Laboratories, Aeronautical Systems Division (AFSC), Wright-Patterson AFB, Ohio 45433-6543 under Contract F33615-87-C-1499, ARPA Order No. 4976, Amendment 20.

and define an aspect as a contiguous set of poses which yield similar object appearances. That is, the appearances of the object in a single aspect are more-or-less the same, and two poses yielding qualitatively different appearances belong to different aspects.

The specific problem that we consider is that of *aspect resolution*, which is the problem of determining which aspect is being observed. Aspect resolution yields a rough solution to the localization problem, since it restricts the possible poses to those consistent with observing a particular aspect. Aspect resolution can also be used to solve object identification problems.

We perform aspect resolution by combining the results of aspect classifications from multiple viewpoints. Planning of multiple observations is based on knowledge of the geometric extent and relations between aspects as well as on the uncertainty in object pose. The representation that we employ for specifying the geometry and relations between object aspects is called an *aspect diagram*. In the restricted planar domain we are dealing with, the aspect diagram is like a pie chart. The set of object poses can be represented as a circle in the plane containing the object and sensor. The aspect diagram divides the circle into wedge-shaped regions, each wedge representing an aspect.

Figure 1 illustrates a specular cone and the aspect diagram resulting when the cone axis lies in the plane of the viewing circle. The aspects form three classes: class-0, with a circular specularity; class-1, which lacks any detectable specularity; and class-2, with a specular streak.

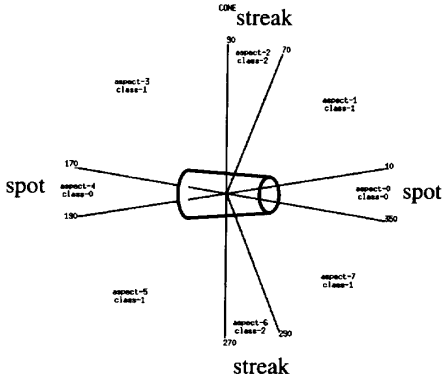

Figure 1: Aspects and Aspect Diagram

3.1 The Observation Function and Observation Graph

Let θ_0 represent the displacement of the object from its reference pose. We define the *observation function* that relates the angular displacement ψ of the sensor to the observed aspect class with respect to the object coordinate system:

$\Omega_{\theta_0} : (\Psi \rightarrow \Gamma)$

where Ψ denotes the set of viewing positions, and $\Gamma = \{\gamma_1 \ldots \gamma_n\}$ denotes the set of aspect classes.

In object localization problems, one of the goals is to determine the object coordinate system with respect to some world coordinate system. In the case being considered here, the relationship between object and world coordinates can be specified by a single parameter, θ, which is the angular displacement of the object from its reference pose. Ideally, localization can be performed by inverting the observation function, and using the sensor position to determine θ. However, the observation function may not yield a unique sensor position, and multiple observations may be necessary.

To combine multiple observations, it is necessary to be able to predict possible observations on the basis of known observations. For a known displacement of world to object coordinates, the observation function can be used. We need a way to make predictions given uncertainty in object pose; this can be done by adding an extra dimension to the observation function. The general observation function becomes:

$\Omega: (\Psi \times \Theta \rightarrow \Gamma)$

where Ψ and Γ are defined as above, and Θ denotes the set of possible object poses. The graph of the observation function is two-dimensional, and we refer to it as the *observation graph*. The observation graph corresponding to the object in Figure 1 is illustrated in Figure 2 for $0° \leq \psi, \theta < 360°$.

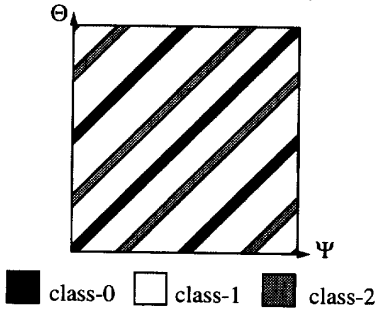

Figure 2: 2D Observation Graph

The observation function can be restricted in various ways to define observation functions for various subsets of ψ and θ. The *aspect-restricted* observation function is defined by:

$$\Omega_{\alpha_i}(\psi, \theta) = \Omega(\psi, \theta)\,|_{\Phi(0,\theta) = \alpha_i}$$

Intuitively, the aspect-restricted observation function Ω_{α_i} represents the collection of observations that could result if α_i was the aspect underlying the observation made at the sensor base position ($\psi = 0°$).

Similarly, the *class-restricted* observation function is defined by:

$$\Omega_{\gamma_j}(\psi, \theta) = \Omega(\psi, \theta)\,|_{\Omega(0,\theta) = \gamma_j}$$

The class-restricted observation function represents the situation that results when a particular aspect class has been observed, but the particular aspect is unknown.

Figure 3 illustrates aspect and class restrictions of the observation graph of Figure 2.

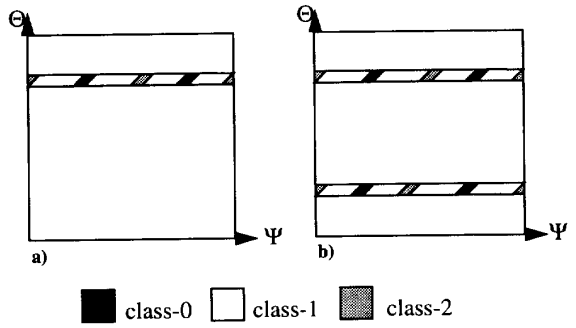

Figure 3: Restrictions of the Observation Graph
a) Aspect Restriction: $\Omega_{\text{aspect-6}}(\psi,\theta)$
b) Class Restriction: $\Omega_{\text{class-2}}(\psi,\theta)$

The goal of aspect resolution is less ambitious than that of object localization. Rather than determining an exact value of θ_0, aspect resolution determines a range of orientations (θ_l, θ_h) and an aspect α_i, such that aspect α_i is guaranteed to underlie any observation in the interval.

The tool which can be applied to perform aspect resolution is a sensor observation, which yields only the class of the underlying aspect. We define the *observation-restricted* observation function by:

$$\Omega_{\psi_0, \gamma_j}(\psi, \theta) = \Omega(\psi, \theta) \mid_{\Omega(\psi_0, \theta) = \gamma_j}$$

Figure 4 illustrates a restriction of the observation graph of Figure 2 for the observation $\Omega(80°, \theta)$ = class-0.

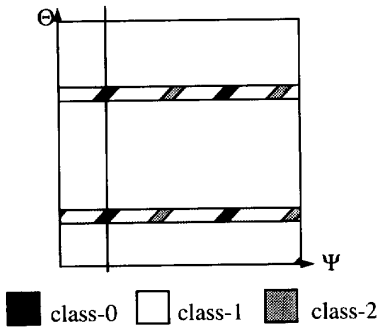

Figure 4: Observation Graph for Truncated Cone Restricted to $\Omega(80°, \theta)$ = class-0

Each observation restricts the observation function. If the relationships between multiple observations are known, then multiple restrictions can be simultaneously applied to the observation function. Graphically, the operation is that of intersecting the strips resulting from the individual operations. When the multiple restrictions result in a subset of the aspect-restricted observation function for a single aspect, then aspect resolution has been completed.

A procedure for using the observation function to perform aspect resolution now suggests itself:

construct the observation function for the object
perform a sensing operation
construct the observation-restricted observation function corresponding to the observation
while (the observation-restriction is not a subset of any aspect-restriction) do

    select a new sensor position and move the sensor
    perform a sensing operation
    construct a new observation-restriction

3.2 Planning with the Observation Graph

We assume that the first observation defines the 0° sensor position, so all moves will be made with respect to this position. Each additional observation made results in an additional restriction of the observation graph. What is needed, then, is a way to select sensor positions, or moves, in such a way as to eventually guarantee a restriction of the observation graph to some subset of an aspect restriction. The problem is that there is potentially an infinite number of moves.

An examination of the structure of observation graphs simplifies the move selection problem. The observation function is discontinuous, with the discontinuities occurring at the boundaries between different observation results. Let the points along the border of a restricted observation graph at which the observation function is discontinuous be referred to as *nodal points*. The linear structure of the observation graph guarantees that the set of observations possible is the same between any two nodal points; only the relative extent of the observations will vary.

Since the information between nodal points is constant, only nodal points need be considered when selecting sensor moves. Additional moves can be investigated by examining the observations possible at the new nodal points resulting from each new restriction. Figure 5 illustrates the procedure.

While Figure 5 depicts only two of the possible moves starting from $\Omega(0°, \theta)$ = class-2, it illustrates the tree structure of the collection of possible move sequences. We call this tree the *observation tree*. Each leaf node represents the resolution of some aspect. Planning for aspect resolution consists of searching through the observation tree, looking for subtrees in which every aspect is resolved. We call such a subtree a *resolving subtree*.

In practice, despite the use of nodal points to define moves, there are many possible moves at every node of the observation tree. The result is that observation trees are far too large to search exhaustively, so heuristic search is required.

The resolving subtree that results from search is called a *resolution tree*, since it specifies sequences of moves for resolving any aspect. Typically, several moves are required. Figure 6 illustrates a resolution tree for the truncated cone. The res-

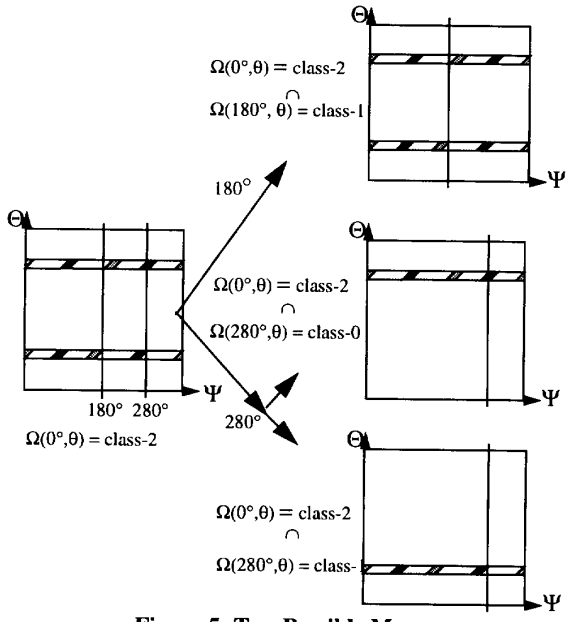

Figure 5: Two Possible Moves

olution tree shown has three types of nodes. *Move nodes,* represented by circles and labeled "M-", denote sensor moves; the move position is indicated in degrees relative to the initial position. *Congruent-set nodes*, represented by squares and labeled "C-", denote sets of congruent aspects. The observed class is indicated within the square, as well as the labels of the congruent aspects. *Resolved nodes*, also represented by squares but labeled "R-", denote aspects that have been resolved. Note that all leaf nodes are resolved nodes.

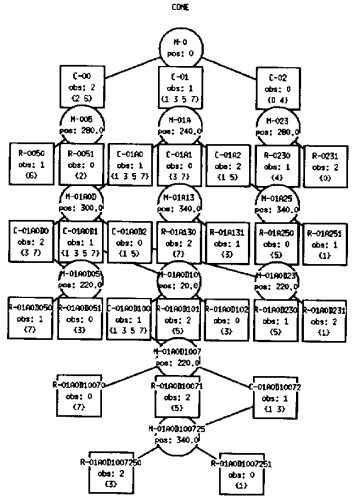

Figure 6: Resolution Tree

3.3 Higher Degrees of Freedom

Abstractly, there is little difficulty in extending our previous results to higher degrees of freedom. The observation function, Ω, becomes a scalar function of two vector-valued variables, $\vec{\psi}$ and $\vec{\theta}$, which define sensor position and object pose, respectively: As before, the higher-dimensional observation graph can be partitioned into disjoint regions corresponding to either aspect-restrictions, class-restrictions, or observation-restrictions. Planning for aspect resolution is again a search through the tree of possible sensor moves to find a subtree in which every leaf node is a subset of an aspect restriction.

Several representations for the three degree of freedom observation graph have been derived. Each representation has its own advantages and disadvantages in terms of size, complexity, and robustness. These representations are described in [2].

4 Experiments

In this section, we report the results of applying the multiple observation methodology in the domain of specular objects. Most of the experiments were performed on synthetic data, although experiments on actual specular objects were also conducted.

4.1 Synthetic Specular Domain

For each of the objects used, a CAD-like model of the object was constructed. The model was used as the input to an appearance simulator [1] which generated images of the object, corresponding to the images that would be obtained by a co-located light-source and sensor. Images were generated at fixed intervals around the object. Thresholding was applied to each image to extract the specularities, and then each specularity was analyzed to compute the feature values. Aspects were defined by the number of detected specularities. Each aspect was then characterized by averaging the values of the features for each sensor position within that aspect. Finally, an aspect classification tree was generated. All of the above tasks were performed autonomously.

Each trial consisted of selecting an object pose (rotation) at random, and feeding the rotated object model to the system as input. The initial rotations were selected at random from the interval [0.00, 360.00), with a resolution of 0.01°. The trial proceeded by generating, processing, and classifying an image of the object. Then, the process was halted if a leaf node was reached (meaning that resolution was completed), or if the classification failed (meaning that the feature set matched no known aspect). Otherwise, the simulated sensor was rotated with respect to the model as specified by the resolution tree, and the image generation, processing, and classification steps repeated. Three results were possible for each experiment: correct resolution, incorrect resolution, or error (unresolvable).

To explore the effect of model fidelity on resolution accuracy, the sampling interval used to construct each model was varied. Intervals of 1°, 2°, 4°, 6°, 8°, and 10° were used. 100 resolution trials were executed for each of the models created.

Our initial experiments were conducted using *ell*, an object composed of four spheres connected to form an "L" shape. The aspect diagram for *ell* with a 1° sampling interval is shown in Figure 8, along with the specularities that characterize each aspect. The resolution tree that was generated for this model is shown in Figure 8. The darkened path through the tree is the resolution path taken during one of the trials. The results of the experiments with *ell* are shown in Table 1. The experiment was repeated for several other objects composed of spheres, with similar results in all cases.

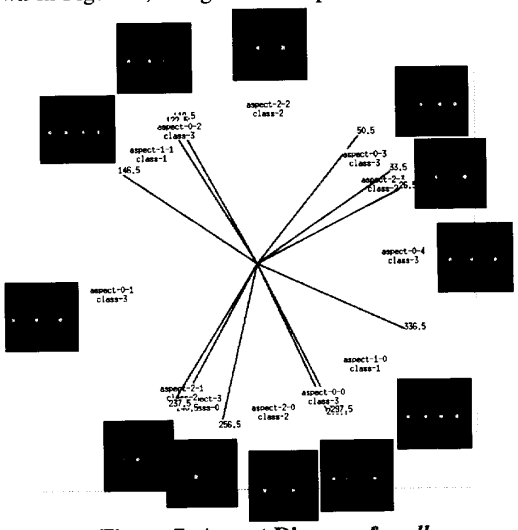

Figure 7: Aspect Diagram for *ell*

Table 1: Aspect Resolution Results for *ell*

sampling interval	1°	2°	4°	6°	8°	10°
total tests	100	100	100	100	100	100
correct	100	99	94	87	86	73
incorrect	0	1	4	11	10	15
unresolvable	0	0	2	2	4	12

We also applied our methodology to a fairly complex object - a stylized jet. The aspect diagram for the jet model for a 1° sampling interval is shown in Figure 9. The corresponding resolution tree is shown in Figure 10. The results are shown in Table 2.

An analysis of the errors proved to be interesting. The most common error was due to something we refer to as a *border effect*. The aspect diagram depicts sharp boundaries between aspects. In reality, the boundaries are somewhat blurred.

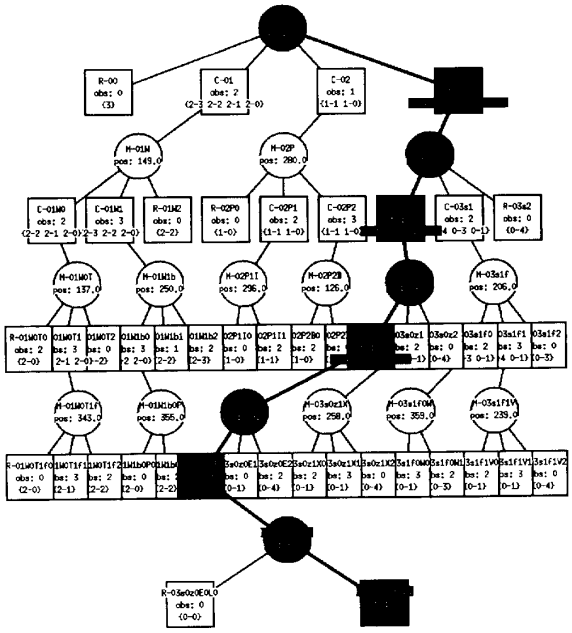

Figure 8:Resolution Tree for *ell* with Resolution Path

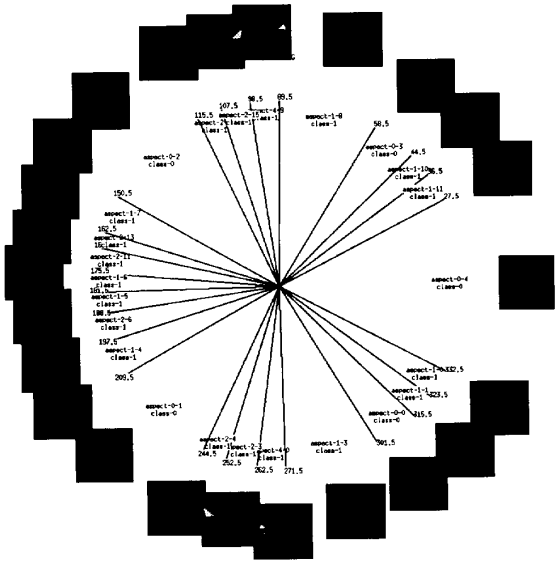

Figure 9: Aspect Diagram for Jet Model

Table 2: Aspect Resolution Results for Jet Model

sampling interval	1°	2°	4°	6°	8°	10°
total tests	100	100	100	100	100	100
correct	99	97	69	46	26	34
incorrect	1	3	11	24	12	9
unresolvable	0	0	20	30	62	57

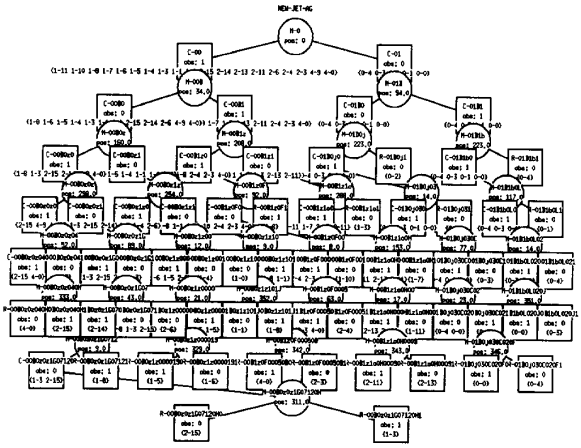

Figure 10: Resolution Tree for Jet Model

Specularities are unstable features, but they do not instantaneously wink out of existence. Finer sampling means that each specularity will be more finely sampled, and that the gradual effects of appearing and disappearing will be more exactly modeled. Table 1 clearly shows that incorrect resolutions increase with increasing sampling interval.

Unresolvable cases often result when sampling is too coarse. Since specularities appear and disappear over small ranges of sensor motion, specularities can be missed entirely by sampling around the region in which a specularity appears. As expected, the number of unresolvable cases increases with increasing sampling interval.

4.2 Actual Specular Domain

In addition to the simulation experiments, experiments with actual objects were performed in the laboratory. Several sphereworld objects were constructed and painted glossy black. For each experiment, the object was placed on a black rotation table, in front of a black background. A pose for the object was selected, and the rotation table moved to the appropriate position. The camera was fixed, and a bright point source was co-located with the camera.

The first set of trials involve a "T" shaped object, *tee*. On each trial, resolution was correctly performed.The second set of trials involved the object *ell*. Again, on each trial, resolution was correctly performed.

5 Discussion

This paper has presented an approach for specular object recognition that relies on the use of multiple observations from different viewpoints to resolve any ambiguity in scene interpretation. The results show that the multiple observation strategy can be very accurate, and is in fact limited only by the accuracy with which decisions can be made about individual observations.

The domain selected for this research was that of specular images. However, the general approach of planning with uncertainty is applicable in many other domains; essentially, the multiple observation strategy can be employed whenever a single observation may not be sufficient for unambiguous object recognition.

The major limitation of the multiple observation approach discussed here is the dependence on aspect classification at each step: if any single step is in error, the whole sequence of observations will result in an error. To remedy this shortcoming, it is necessary to perform some sort of probabilistic reasoning that will not discard alternatives completely, but instead will combine evidence and yield the best answer given all the available information.

6 Acknowledgments

Takeo Kanade provided many useful comments and encouragements. The authors would also like to thank the members of the Vision and Autonomous Systems Center of the Robotics Institute of Carnegie Mellon University for their valuable comments and discussions.

7 References

[1] Fujiwara, Y., Nayar, S., and Ikeuchi, K. *Appearance simulator for computer vision research.* **CMU-RI-TR-91-16,** Carnegie Mellon University (1991).

[2] Gremban, K. and Ikeuchi, K. *Planning multiple observations for object recognition.* **CMU-CS-92-146,** Carnegie Mellon University (1992).

[3] Hong, K. S., Ikeuchi, K., and Gremban, K. D. *Minimum cost aspect classification: A module of a vision algorithm compiler.* **Proc. 10th Int. Conf. on Pattern Recognition,** pp. 65-69 (1990). A longer version is available as **CMU-CS-90-124,** School of Computer Science, Carnegie Mellon University (1990).

[4] Hutchinson, S. A. and Kak, A. C. *Planning sensing strategies in robot work cell with multi-sensor capabilities,* **IEEE Transactions on Robotics and Automation,** vol. 5(6):765-783 (1989).

[5] Ikeuchi, K. and Kanade, T. *Automatic generation of object recognition programs.* **Proc. of the IEEE,** vol 76(8):1016-1035 (1988).

[6] Koenderink, J. J. and van Doorn, A. J. *The internal representation of solid shape with respect to vision.* **Biological Cybernetics,** 32:211-216 (1979).

[7] Liu, C. and Tsai, W. *3D curved object recognition from multiple 2D camera views.* **Computer Vision, Graphics, and Image Processing,** 50:177-187 (1990).

[8] Sato, K., Ikeuchi, K., and Kanade, T. *Model based recognition of specular objects using sensor models.* **CVGIP: Image Understanding,** 55(2):155-169 (1992).