Proceedings of the 1998 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
Victoria, B.C., Canada • October 1998

Hough-space-based Object Recognition Tightly Coupled with Path Planning for Robust and Fast Bin-picking

Ayako Takenouchi, Naoyoshi Kanamaru and Makoto Mizukawa

Autonomous Robot Systems Laboratory
NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo 180, Japan
e-mail: {ayako, nao, mizukawa}@nttarm.hil.ntt.co.jp

Abstract

The proposed bin-picking method combines object recognition with path planning. To avoid conflicts between the assumptions of the elemental techniques needed for bin-picking, object recognition is combined with path planning by using environmental information. To achieve this combination, a Hough transform, which records the model-to-image matches in a Hough space, is used to estimate the pose. The matches represent the arrangement of the objects, so they can be regarded as environmental information for path planning. To reduce the number of recognition errors and the object-detection time, a pair of object features that reduces the number of invalid votes in the Hough space is used for the Hough transform. Simulated path planning showed that using a Hough space to represent the environmental information improves the ability to plan a safe path for the manipulator. Simulated object recognition showed that using a pair of features makes the process faster and reduces the number of invalid votes. The pose estimation and safe path planning ability were confirmed by an experiment on casting objects using a range finder and a robot.

1 Introduction

We are developing an automated method for bin-picking that combines object recognition with path planning. "Bin-picking" means to pick an object from a large number of similarly shaped objects. It is used to supply parts to the production line of a factory. Although there has been a strong demand for a technique to automate bin-picking in the industrial field, such a technique has not been fully developed yet.

The delay in the automation of bin-picking is due to two problems. Automated bin-picking needs several elemental techniques, including object recognition, path planning for the robot to grasp the part, and modeling of the part. However, while much research has been done on these elemental techniques, not all of them are suitable for bin-picking. The other problem is that simply combining these elemental techniques is not sufficient for bin-picking.

The suitability problem is related to the assumptions of the elemental techniques. An example suitability problem is path planning for a manipulator by using environmental information. Many studies have focused on how to express the environmental information to achieve safe and fast path planning. One well-known approach to creating environmental information is to use a configuration space[1]. Other approaches use algebraic equations to represent the boundaries of obstacles in order to represent the free space[2], or a heuristic potential field that gives obstacles a high potential. Many studies have looked at planning a path for a robot having many degrees of freedom. Kondo[3] proposed using hierarchical cell decomposition for planning the paths of robots having six degrees of freedom. However, he assumed that exact environmental information was known although such information cannot be obtained due to sensor errors and blind spots.

The combining problem is related to incompatibility between the input and output of the elemental techniques. An example combining problem is combining object recognition with path planning. From the object-recognition side, many studies have focused on recognizing partially visible objects. Bolles et al.[4]

presented a technique for recognizing partially visible industrial parts in a bin by using range and intensity information. They modeled the workpiece as a small number of local surface features that are helpful in distinguishing individual parts, and developed a formula for determining an object's location from these features. They verified their formula by using an object model. More sophisticated approaches (e.g., [5] and [6]) attempt to recognize partially visible objects by using local features, making it possible to recognize complex-shaped objects from partial information. However, these techniques recognize only a few objects, while path planning needs the arrangement of objects around the target object.

In addition, several works have focused on bin-picking. Horn and Ikeuchi[7] used binocular vision to determine the grasp configuration of the manipulator used to pick the object. Dessimoz et al.[8] proposed a model for determining which shape to grasp. Wang et al.[9] proposed a model-based vision system for picking up twisted tubes from a bin by using a 3-D structured light to detect the graspable segments of the tubes. The robot places the grasped tube on a bottom-lit table and determines its pose by using monochrome vision. Because these approaches use an end-effector, which is suitable for objects that require skillful grasping, and because the shapes of the objects are limited, they make the bin-picking problem easier than the general state.

For bin-picking of general objects, we must consider combining two or more techniques, such as object modeling, object recognition, and path planning. We propose a method that combines object recognition and path planning. The key to connecting object recognition with path planning is environmental information. Because obtaining complete environmental information is difficult, we define "uncertain areas", enabling us to identify a safe path based on environmental information that is incomplete and that may contain errors. To acquire information about the uncertain areas, we use a generalized Hough transform for object recognition.

A generalized Hough transform constructs a "maximum likelihood" estimate by using the Hough space-clustering technique[10]. It can calculate any arbitrary set of parameters (position, orientation, scale) for any shape 2-D object. Krishnapuram and Casasent[11] proposed a Hough transform that uses a range image to determine the position and orientation of objects. They use planes to determine the position and orientation of each object. To extract the planes, they extended the two-dimensional straight-line Hough transform into a three-dimensional plane in the range image. Hu[12] proposed using 3-D vectors to calculate the pose of an object. He added an extra vector to the 3-D edge object model. The model-to-image matches are calculated using this extra vector and a vector extracted from the object image. Only three pose parameters are estimated to reduce the amount of memory needed.

When these approaches are used to recognize only one object, the invalid votes are usually spread out, so they do not cause a recognition error. However, when they are used to recognize many similarly shaped objects, the invalid votes can create several peaks, which cause recognition errors. To reduce the number of recognition errors and the object-detection time, we need to reduce the number of invalid votes. We do this by using a Hough transform based on pairs of local object features.

2 Modeling

Bin-picking includes both object recognition and path planning for the manipulator. Object models for object recognition express the shape of the object in detail, so that the position and orientation of the object can be calculated accurately. Object models for path planning express the space that the object occupies, so that potential collisions with the manipulator can be detected quickly. These differences make it difficult to define one object model that can handle both processes. We thus use two models — one for object recognition, and one for path planning — for bin-picking (Fig. 1).

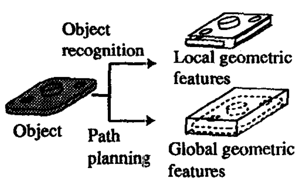

Figure 1: Object models.

Simple local geometric features of the object are used as the model for object recognition because object models that contain many local features of an object make it possible to detect an object from only a partial view. To make the object model usable for various industrial parts and to reduce the pre-processing time, simple geometric features in range image, such as flat or curved surfaces are used as the local features.

In our simulation and evaluation we used planes; each plane was defined by an equation and by the coordinates of the vertices of a rectangle covering the plane.

A boundary box was used as the object model for path planning. Object models that cover the whole object are faster, but sometimes give incorrect results. However, this can be overcome by adding a check step using a detailed object model. To reduce the processing time, we use a boundary box that covers the whole object.

3 Flow of information

The flow of information that we propose for bin-picking is shown in Fig. 2. Object recognition identifies both detected and possible objects. The environmental information is then created by using the information about both types of objects. This environmental information and the position and orientation of the detected objects are transferred to path planning process. Path planning identifies the target object and creates the robot path. The key is that the environmental information consists not only of information about the detected objects, but also about the possible objects. By taking the undetected objects into consideration, the path of the robot is made safer.

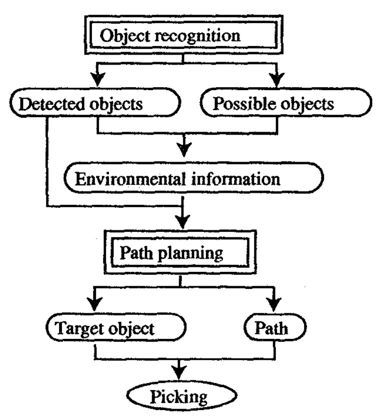

Figure 2: Flow of information.

4 Object Recognition

In bin-picking, object recognition has two roles. One is to determine the pose and orientation of the robot needed to grasp the target object. This role is equal to identifying the target object among the detected objects when the grasping points are known. The other role is to make environmental information which is used to avoid collisions in the manipulator path planning. Environmental information containing only the detected objects is not sufficient for safe path planning because it does not help to avoid collisions with undetected objects. We therefore use the generalized Hough transform to add undetected objects to the environmental information.

The generalized Hough transform stores all possible model-to-image matches, i.e., it creates parameter sets of objects as votes in a Hough space. The votes for objects that are easier to see create high peaks while the votes for objects that are more difficult to see create low peaks. Therefore, the height of each peak in the Hough space is an index of the probability of an object actually existing at the corresponding location. Therefore, a Hough space can be regarded as the configuration of objects in a workspace and as a distribution map of probable existence.

Moreover, the generalized Hough transform is useful for bin-picking. It does not require that the object features be grouped into separate objects beforehand. In bin-picking there are many identical objects in each bin, so distinguishing one object from another is difficult. It is robust because it adopts the majority for estimation. Also, it can locate several objects simultaneously, which is useful for determining which object to grasp, i.e., selecting the object that is the easiest to grasp.

However, a Hough transform has three problems. One is that it needs a lot of memory for the Hough space due to the dimensions of the parameter sets to be detected. Another is that it is time consuming because it uses majority decision making. And third, is that it may result in recognition errors due to invalid votes when there are many similarly shaped objects in a workspace. The first problem can be avoided if only a few parameters are detected by the Hough transform, as done by Hu[12]. The effect of the other problems can be reduced by decreasing the number of invalid votes.

To detect only a few parameters with the Hough transform, we divided six parameters into two independent parameter sets — three for position and three for orientation. To reduce the number of invalid votes, we use features that can determine a finite number of parameter sets composed of three pose parameters. Using these features limits the number of possible orientations, so each vote is represented by a dot, rather than by a line in our Hough space. Not only does this reduce the number of votes, it also reduces the amount of calculation. Using the Hough transform with a set of local geometric features to estimate the pose enables objects to be recognized faster with a reasonable amount of memory.

To restrict the orientation of an object by using simple geometric features such as planes and curved surfaces that are a part of a cylinder, we use a combination of them. Because the normal vector of a plane or the axis direction of a cylinder restricts the orientation freedom to one degree, their combinations are sufficient to restrict the object pose. We compare each combination of features extracted from the range image with the object model, and if the combination matches the model, it is assumed to be that of one object. The object's orientation is then calculated using a pair of feature vectors.

In summary, as shown in Fig. 3, we first extract all features from the range image. After calculating the vector for each feature, all two-feature combinations are compared to the object model. The matching combinations are assumed to represent objects, and the orientations of these objects are calculated using the normal vectors or axes; a vote is then recorded in the Hough space. Non-matching combinations are regarded as features of different objects and are discarded. After all combinations are compared, the peaks in the Hough space are searched for, and the parameter sets of each peak are obtained. These sets represent the possible orientations of the objects. The model to feature matches from two votes in one peak are used to calculate the possible position of object. To distinguish the objects having the same orientation, clustering of the position parameter sets was used to estimate the position of the objects.

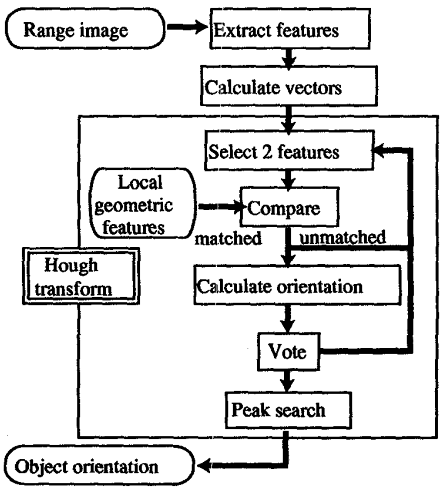

Figure 3: Process flow for calculationg object orientation.

5 Combining object recognition with path planning

Our approach focuses on combining the elemental processes that compose bin-picking. This requires transferring information from one process to another, for example, passing the environmental information obtained in object recognition to path planning. By combining object recognition with path planning, motion of the manipulator is made safer because the areas having uncertainty in the environmental information are identified explicitly.

One technique for making uncertain areas in the environmental information is to add a uniform "safety margin" around each object in the workspace[13]. However, the information obtained by object recognition is limited to that for the detected objects, and the missing information concerning undetected objects cannot be compensated for. Our approach aims to identify all risky spaces caused by undetected objects. To identify these uncertain areas, all possible objects are added to the environmental information. To calculate both detected and possible objects, the Hough space is used.

As described in Sec. 4, the Hough space stores all possible orientations as determined by the various combinations of extracted local geometric features. The height of each peak represents the probability that there is an object having that parameter set. The probability of each peak is used to distinguish between reliable areas and uncertain areas when creating the environmental information. High peaks are used to identify the detected objects, and low peaks are used to identify the possible objects.

Both detected and possible objects are expressed as boundary boxes in the environmental information. The boundary boxes representing the high peaks indicate reliable areas, and the boundary boxes representing the low peaks indicate uncertain areas.

However, many low peaks due to voting errors can appear in a Hough space. To distinguish low peaks representing objects from these error peaks, the position given by each vote for each peak is checked for conflict. There is little or no conflict between the votes for a low peak representing an object, while there is much conflict between the votes for a peak due to voting errors. After identifying the valid low peaks, the possible objects are calculated.

To calculate the possible objects, we first determine the orientation of each possible object represented by a low peak. Because each vote in the Hough space represents the correspondence of two extracted planes to two planes described in the local geometric model, each vote restricts the corresponding object to be located parallel to the intersection of the two planes. This means that each vote in the Hough space limits the positional freedom to one degree. Next, the extracted planes are compared with the rectangles described in the object models. This process restricts the uncertainty about where each object is located. Finally, we make set of representative positions at the same interval within the restriction of the object position along the axis of positional freedom. The possible objects indicated by each low peak are the objects that have the correct orientation and a corresponding set of representative positions.

For a safe path planning, we assumed that there are objects not only represented by the detected objects, but also represented by the possible objects. By this way, the manipulator can avoid the collision against undetected objects.

6 Path planning

Planning the path of a manipulator usually requires two processes. One is compiling a list of the detected objects that is used to select the target object. This list ranks the detected objects in order of their grasping feasibility. It is used to identify a candidate target object to grasp. The other is planning and verifying the ability of the manipulator to grasp the candidate target object. If the ability is verified, the candidate target object becomes the selected target object.

To determine the object rankings, the area around each object, their pose, and their position relative to the robot should be considered. However, the importance of these factors changes when the objects, the robot, or the arrangement of the bin-picking system changes, so a general method for making the list would be difficult to create. We therefore use a simple heuristic. The ranking is based on which detected object is closest ot the top of the bin.

To plan the grasping motion of the manipulator, a process to search for a safe path is needed. However, searching path for a robot having six degrees of freedom requires searching in six-dimensional space. Searching in high-dimensional-space is time-consuming, and the identification of a safe path is not guaranteed. Therefore, we restrict the manipulator to move in a straight line and to maintain the same orientation while grasping an object. Under these conditions, the object orientation and position are sufficient for determining the path of the manipulator.

Although these limitations mean that some objects having a feasible configuration for being grasped may be judged to have an unfeasible configuration, our objective is not to propose a new method for path planning, but to combine the elemental processes into a complete bin-picking process.

7 Simulation

We simulated basic object-recognition to evaluate the performance of our method. The object we used is shown in Fig. 4. All of its planes, except the base plane, were used as local geometric features for object detection. The input arrangement is shown in Fig. 5; boundary boxes are used to represent the configuration of each object. We placed 20 objects in the workspace. Self hiding was considered, but mutual hiding was not, i.e., every object was at least partially visible. In our simulation, 97 planes were seen and used as features for object recognition.

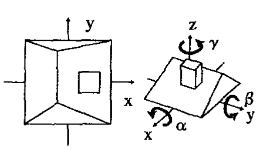

Figure 4: Object used in simulation.

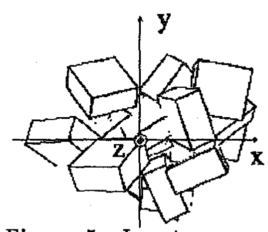

Figure 5: Input arrangement for simulation.

The corresponding votes made in a Hough space are shown in Fig. 6, where each axis represents the rotation of the objects around their own axes (see Fig. 4). The size of each cube in the space represents the number of votes for that parameter set. The total number of two-features combinations from 97 features is ${}_{97}C_2$. Although this is a large number, we reduced the number of invalid votes by matching each pair of features to the object model. This reduced the number of votes in the Hough space to only 608, i.e., 13% of ${}_{97}C_2$. Because this process retains the valid votes, it also reduces the number of errors in recognition and speeds up the recognition process.

When the threshold for peak extraction was set to five, 12 of the 20 objects were detected correctly. Three of the undetected objects formed low peaks in the Hough space, and the others were hard to detect due to self hiding.

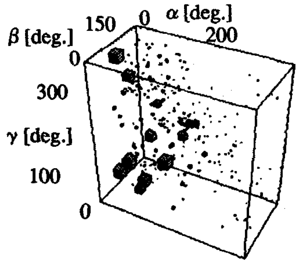

Figure 6: Votes for each object in a Hough space.

The time needed to calculate the orientations and positions of all the objects from the vectors of the local geometric features was about two seconds using an SGI Indy workstation, with a 200-MHz processor and 64MB RAM.

We also evaluated the combination of object recognition and path planning. The heuristics for grasping objects automatically were "the manipulator approaches the objects from above" and "the manipulator grasps the top object." This means that the grasping path was predefined within the global geometric model. Differences in environmental information thus affected the grasping decision.

When the environmental information was completely known (Fig. 7), the target object was correctly selected and grasped. Figure 8 shows the result when the environmental information came only from the detected objects. The white boundary boxes represent detected objects and the gray boundary box represents an undetected object. A lack of information caused incorrect grasping; motion to the target object was blocked by an undetected obstacle. Figure 9 shows the result of our approach, which uses information from both detected objects and identification of other possible objects. The gray boundary boxes show the other possible objects. Although we could have used a global geometric model for these uncertain objects, we used discrete geometric models for expediency. Because we also included the uncertain objects, the problem shown in Fig. 8 was overcome, and the correct object was grasped.

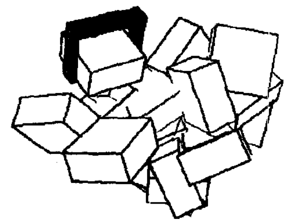

Figure 7: Result when environmental information was completely known.

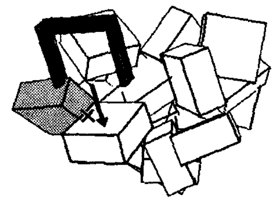

Figure 8: Result when environmental information came from detected objects only. (The gray box represents an undetected object.)

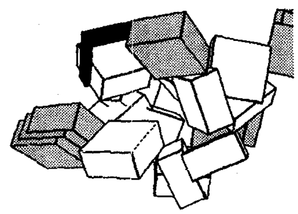

Figure 9: Result when environmental information came from detected objects (white boxes) and identification of other possible objects (gray boxes).

8 Experiment

We performed a basic bin-picking experiment using the casting object shown in Fig. 10. Though casting objects are used in industrial field, it is difficult to recognize them because they have construction error and burr. Moreover, their metal like surface are difficult to measure accurately by vision sensors.

The system shown in Fig. 11. A range finder (PULSTEC:TDS-1500) was set opposite to a robot (YASKAWA:Motoman) with a parallel-jaw gripper. Fourteen planes on the object were used as local geometric features for object recognition. The two ends of the large cylinder shape were chosen for the grasping points.

A range image showing the arrangement of the three objects used in the experiment is shown in Fig. 12. Object 1 and 2 were partially hidden, but object 3 was hidden from the range finder because it was below the other objects. Object 1 was above object 2, so it was assumed to be easier to grasp than object 2. However, moving the manipulator to grasp object 1 was not feasible because it would collide with object 3. Object 3 could not be detected because it was hidden so the collisions cannot always be detected when the environmental information comes only from inforthe environmental information comes only from information about the detected objects.

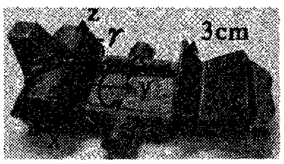

Figure 10: Object used in experiment.

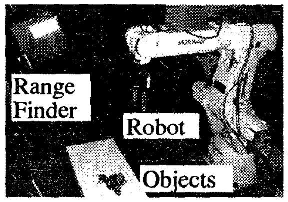

Figure 11: System used for experiment.

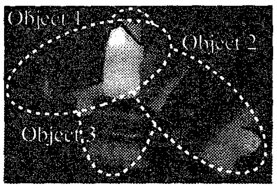

Figure 12: Input arrangement for experiment.

Figure 13: Extracted planes.

Figure 13 shows the planes extracted from the range image. Although the range image had noise, 37 planes could be extracted because simple geometric features, such as a plane, are generally extracted robustly. The votes corresponding to the arrangement in Fig. 12 in a Hough space are shown in Fig. 14. There were 3875 votes, which is about 3% of the maximum number of votes, i.e., ${}_{37}C_2 \times_{14} C_2$. Figure 15 shows the result of position clustering from the high peaks in the Hough space. The peaks representing objects 1 and 2 produced clusters having 13 and 33 members respectively. The invalid peaks had fewer members. Although the normal vector of each extracted plane had an error, objects 1 and 2 were detected because the Hough transform estimates the parameters by majority. Object 3 is represented as a low peak, so possible objects representing objects are recognized. The time needed to detect the objects from the range image was about 15, and the time needed to calculate the environmental information was about 30.

The environmental information used for path planning is shown in Fig. 16. The white boundary boxes represent detected objects, and the gray boundary boxes represent possible objects. In path planning, the target object is selected by checking the motion of the manipulator. First, object 1 was selected as the candidate target object because it was above object 2. However, the motion of the manipulator to grasp object 1 was found not to be safe because the manipulator would collide with the possible object representing object 3. Therefore, object 2 was selected as the candidate target object, then set as the selected target object after the motion of the manipulator was found to be safe. By considering the possible object, object 2 was set as the target object, and safe planning was achieved.

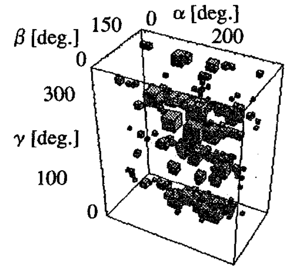

Figure 14: Votes for each object in a Hough space.

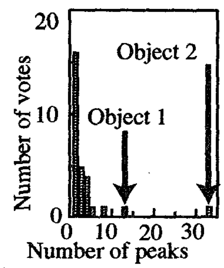

Figure 15: Distribution map of number of votes.

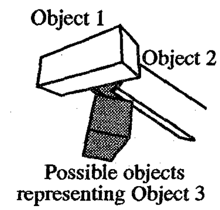

Figure 16: Environmental information.

9 Conclusion

We have presented an approach to bin-picking that combines object recognition with path planning. A generalized Hough transform is used for object recognition. The Hough space represents the obtained environmental information. The orientations and positions of undetected (i.e, uncertain) objects are added to the environmental information as possible objects and used in planning the manipulator path. The enhanced reliability of path planning by this approach was confirmed by both simulation and experiment.

We also presented an approach for reducing the time and memory needed for processing the generalized Hough transform. In our approach, we use a pair of object features to calculate the parameter sets. This reduces the number of votes recorded in the Hough space from each feature, thus reducing recogniin the Hough space from each feature, thus reducing recognition time. It also reduces the number of invalid votes, which cause recognition errors. Because only the orientation of each object is detected by the Hough transform, less memory is needed for the transform. Simulation showed that the Hough transform can detect object orientations quickly and needs only a reasonable amount of memory. The effectiveness of the presented method was confirmed by the experiment using real casting objects.

References

[1] T. Lozano-Perez, "Spatial planning: A configuration space approach", *IEEE Trans. Computers*, Vol. C-32, No. 2, pp. 108-120, 1983.

[2] J. Latombe, "Robot motion planning", *Kluwer Academic Publishers*, 1991.

[3] K. Kondo, "Motion Planning with Six Degrees of Freedom by Multistrategic Bidirectional Heuristic Free-Space Enumeration", *IEEE Trans. Robotics and Automation*, Vol. 7, No. 3, pp. 267-277, 1991.

[4] R.C. Bolles, P. Horaud, and M.J. Hannah, "3DPO: A Three-Dimensional Part Orientation System", *Int. J. Robotics Res.*, Vol. 5, No. 3, pp. 3-26, 1986.

[5] K. Rahardja and A. Kosaka, "Vision-Based Bin-Picking: Recognition and Location of Multiple Complex Objects Using Simple Visual Cues", *Int. J. Robotics Res.*, Vol. 2, pp. 1448-1457, 1996.

[6] K. Ohba and K. Ikeuchi, "Recognition of the Multi Specularity Objects for Bin-picking Task", *Proc. IROS*, Vol. 2, pp. 1440-1447, 1996.

[7] B.K.P. Horn and K. Ikeuchi, "The Mechanical Manipulation of Randomly Oriented Parts", *Scientific American*, Vol. 251, No. 2, pp. 100-111, 1984.

[8] J. Dessimoz, J.R. Birk, and R.B. Kelley, "Matched Filters for Bin Picking", *IEEE Trans. Pattern Anal. Machine Intel.*, pp. 686-697, 1984.

[9] S. Wang, R. Cromwell, A. Kak, I. Kimura, and M. Osada, "Model-Based Vision for Robotic Manipulation of Twisted Tubular Parts: Using Affine Transforms and Heuristic Search", *IEEE Proc. Int. Conf. Robotics and Automation*, pp. 208-215, 1994.

[10] D.H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes", *Pattern Recognition*, Vol. 13, No. 2, pp. 111-122, 1981.

[11] R. Krishnapuram and D. Casasent, "Determination of Three-Dimensional Object Location and Orientation from Range Image", *IEEE Trans. Pattern Anal. Machine Intel.*, Vol. 2, No. 2, pp. 1158-1167, 1989.

[12] G. Hu, "3-D Object Matching in the Hough Space", *IEEE Trans. Systems, Man, and Cybernetics*, No. 3, pp. 2718-2723, 1995.

[13] T. Lozano-Perez, "Automatic Planning of Manipulator Transfer Movements", *IEEE Trans. Systems Man and Cybernetics*, SMC-11, No. 10, pp. 681-698, 1981.