

Development of Structured Light based Bin Picking System Using Primitive Models

Jong-Kyu Oh, KyeongKeun Baek, Daesik Kim, and Sukhan Lee

Abstract— As a part of factory automation, bin picking systems perform pick-and-place tasks for randomly oriented parts from bins or boxes. Conventional bin picking systems can estimate the pose of an object only if the system has complete knowledge of the object (e.g., as a result of the geometric features of the object being provided by an image or a computer-aided design model). However, these systems require the features visible in an image to calculate the pose of an object, and they require additional setup time for an operator to register the reference model every time that the workpiece changed. In this article, we propose a structured light based bin picking system that makes use of primitive models that involve a small amount of prior knowledge. To obtain a reliable 3D range image for comparison with conventional systems, we use a structured light sensor with gray-coded patterns. With the 3D range image, the pose of the object is estimated with the use of primitive segmentation, rotational symmetric object modeling, and recognition. Through experiments that involve an industrial robot, we validated that the proposed method could be employed for a bin picking system.

I. INTRODUCTION

APPLICATIONS of industrial robots have increased to improve the flexibility of factory automation systems. The work performed by industrial robots has broadened from spot welding to assembly, which requires more sophisticated control. The material handling process involves an especially high degree of difficulty, and it can have a noticeable impact on production output rates and costs; fixtures can be removed and thus save the amount of working space required in high-value manufacturing areas. Thus, there are increasing demands from manufacturers for a reliable sensing system and a system integration technology so that the robot system can be applied in factories.

In the area of part feeding automation, some research has looked at bin picking systems that perform pick-and-place tasks for randomly oriented parts from bins or boxes [1][2].

Ban proposed the laser-vision-based bin picking system [3]. The proposed laser-vision sensor consists of a projector that projects a cross-type pattern and a camera to capture the image. This sensor calculates the X and Y values and the roll of the object with the use of 2D image processing; it then computes the Z value, the pitch, and the yaw by analyzing a projected image with a laser pattern. However, this system

J.- K. Oh, K. Beak, D. Kim, and S. Lee are with the Intelligent System Research Center of Sungkunkwan University, Gyeonggi, Korea (e-mail: {jkoh, kk.beak, daesik80, lsh}@ece.skku.ac.kr).

J.- K. Oh is also with the Electro-Mechanical Research Institute of Hyundai Heavy Industries Co. Ltd, Gyeonggi, Korea

has a narrow sensing range, because the 3D position of the object is calculated only with regard to the surface on which the laser slit is projected.

Bin picking systems that make use of stereo vision have been widely researched. A study by Rahardja calculated the position and normal vector of randomly stacked parts with the use of a stereo camera [4]. However, this system required unique landmark features, which are composed of seed and supporting features to identify the target object and to estimate the pose of the object.

To overcome the disadvantages of conventional bin picking systems, Schraft proposed a pose estimation method that made comparisons between the computer-aided design (CAD) model of a workpiece and the range image captured by a laser scanner [5]. However, this system has disadvantages in that the accuracy and speed of the pose estimation depend on the number of registered models in the related database.

To tackle the aforementioned problems, in this article, we propose a structured light based bin picking system that makes use of primitive models. The term *primitive models* refers to basic 3D shapes, such as a plane, a cylinder, a cone, and a sphere. Gray-coded patterns are employed to obtain a reliable range image, and the pose of an object is estimated with the use of primitive segmentation, rotational symmetric object modeling, and recognition. Because it makes use of the geometric primitives of target objects, this system can estimate the pose of the objects without having complete prior information about the objects (e.g., 2D geometric features, CAD models). Although all parts of the object are not visible or measurable, the system can estimate the positioning of those objects.

The organization of this article is as follows. Section 2 describes the configuration of the proposed structured light based bin picking system. Section 3 presents the operation procedure of the proposed system, including the structured light technique, the sensor calibration, and the pose estimation method. Experimental results are given in Section 4, and Section 5 is the conclusion.

II. CONFIGURATION OF THE STRUCTURED LIGHT BASED BIN PICKING SYSTEM

For the precise manipulation of complex objects in a bin, we configured a vision-guided robotic system with two parts, as shown in Fig. 1. One part is the personal-computer-based structured light vision system, which includes an IEEE 1394

PCI board, an IEEE 1394 camera, a lens, and a projector. The other part is the robot system, which includes a robot controller. The robot controller and the vision system are connected via RS232C or Ethernet.

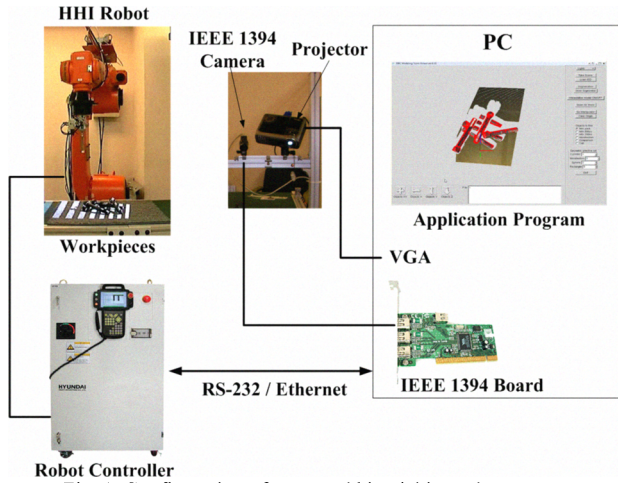


Fig. 1. Configuration of proposed bin picking robot system

III. OPERATION PROCEDURE OF THE PROPOSED SYSTEM

Fig. 2 represents the operation procedure of the proposed system. It consists of the preparatory stage, which includes the initial setup of the 3D vision system and the definition of the robot's tasks, and the operation stage, during which the system automatically performs the pick-and-place task in the workspace by means of measured pose information.

During the preparatory stage, the operator has to calibrate the sensor, register the 3D pose of the reference object, and assign the trajectory of the robot for the task.

During the operation stage, the structured light sensor illuminates a gray-coded pattern on parts in a bin. The point cloud can be obtained after finding a corresponding pair of points between the projector and the camera, and then triangulates them.

The pose of randomly stacked objects is estimated with the use of 3D edge detection, object segmentation, and primitive modeling. The 6-DOF variance between the reference object and the current workpiece is transmitted to the robot controller. Finally, the robot modifies its trajectory. The primary functions of the proposed vision system are the acquisition of a 3D range image and the estimation of the pose of the object to be grasped; these functions are explained in detail in the following paragraphs.

A. 3D Range Image Acquisition

Several approaches have been developed to reconstruct objects and environments, such as stereo vision, focusing technique, and light and shadow analysis [6][7].

Although these approaches—especially stereo vision—are popular, in the case of the passive approach (which only makes use of a camera), the accuracy and reliability of the 3D reconstruction results are relatively low.

The structured lighting technique is reliable for 3D reconstruction; it makes use of triangulation with a

corresponding pair of points between the projection device and the camera [8].

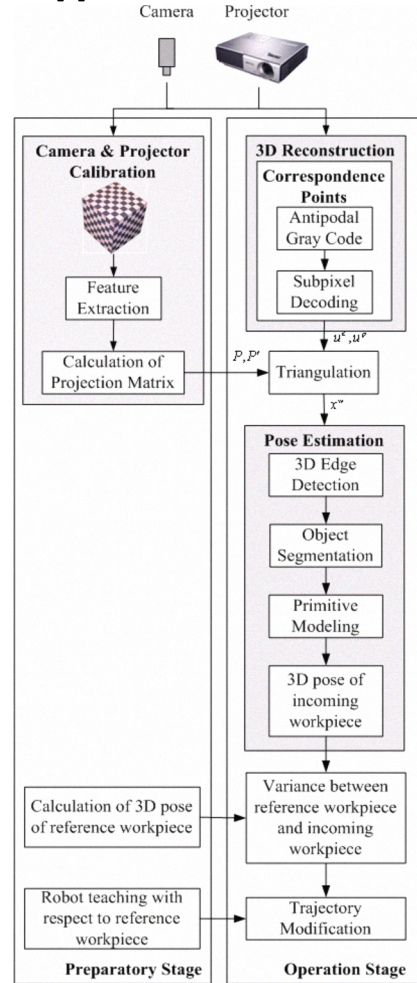


Fig. 2. Diagram of proposed bin picking system

1) Structured Light System

Our structured light system consisted of a DLP projector and a camera, as shown in Fig. 3. The projector illuminated the coded patterns, and the camera captured several patterned scenes. We employed the antipodal Gray code, which is the variant of the conventional Gray code, for our structured light system [9] and we improved the decoding method for finding the corresponding pair of points to obtain more accurate results.

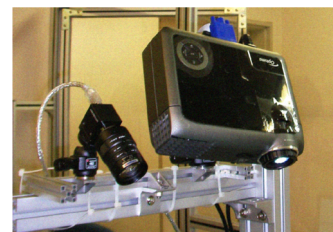


Fig. 3. Structured light system

2) Camera/Projector Calibration

A camera/projector calibration is required to establish the relationship between the 3D coordinates and their corresponding 2D image coordinates. Under perspective

projection, 3D point $\mathbf{X}=[X \ Y \ Z \ 1]^T$ in space is projected to an image point $\mathbf{x}=[u \ v \ 1]^T$ in the image plane via a projection matrix \mathbf{P} as follows, where \mathbf{P} consists of the intrinsic parameter matrix \mathbf{K} , the rotation matrix \mathbf{R} , and the translation vector \mathbf{T} [7].

$$\mathbf{x} = \mathbf{K}[\mathbf{R} \ | \ \mathbf{T}]\mathbf{X} = \mathbf{P}\mathbf{X} \quad (1)$$

To obtain the \mathbf{P} matrix, we needed a calibration rig that could provide 3D position information in an unknown environment, so we used a calibration rig composed of square patterns, as shown in the left-hand image of Fig. 4.

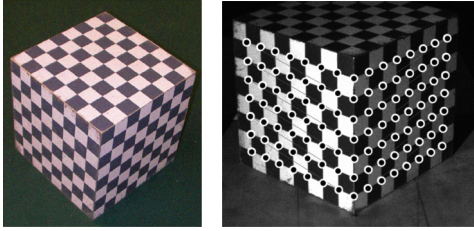


Fig. 4. Calibration rig (left) and the detected corner points (right)

The procedure for the camera/projector calibration was performed as follows. First, we put the rig within a common field of view in the structured light system, and we acquired the image without illuminating any light from the projector. Next, the corner points of the image were extracted, and the camera calibration was performed by mapping from the known corner points in 3D space to the corner points in image plane. The right-hand image of Fig. 4 shows the distribution of detected corner points in the calibration program.

After calibrating the camera, we illuminated the coded-pattern to the calibration rig, captured the pattern, and decode the coded-pattern for calibrating the projector. Because we had already extracted the corner positions of the image and obtained the decoded pattern positions at those points, the relationship between the known corner points of the calibration rig and the projector's pattern position could be computed. If the accuracy of the corner position in the camera image is under a sub-pixel, the corresponding pattern position may not be obtained directly. In this case, we computed the pattern position by interpolating the pixel around the corner position.

The camera/projector calibration was performed just once during the preparatory stage. After the camera/projector calibration process, we acquired the projection matrices of the camera and the projector.

At this point, the transformation between the calibration rig and the robot was computed from more than three corresponding points represented in each coordinate frame. As shown in Fig. 5, we taught the corner positions of the calibration rig to the robot's end-effector.

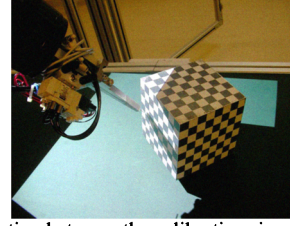


Fig. 5. Transformation between the calibration rig and the robot can be computed by teaching the robot more than three corresponding points.

Then, we can compute the rotation and translation using a unit quaternion, which may be converted in closed form to get a 3×3 conventional rotation matrix [10].

3) Reconstruction

If the projection matrices of the camera and the projector are given and if the pair of corresponding points in the camera and the projector image plane is given, we can determine the coordinate of the scene point that corresponds with the points in the two-image plane.

The relationship between point \mathbf{X} in the scene and the corresponding points \mathbf{x} and \mathbf{x}' in the camera and the projector image coordinates can be written as follows, where \mathbf{P} and \mathbf{P}' denote the projection matrix of the camera and the projector, respectively:

$$\mathbf{x} \cong \mathbf{P}\mathbf{X} \quad (2)$$

$$\mathbf{x}' \cong \mathbf{P}'\mathbf{X} \quad (3)$$

Equations (2) and (3) are expressed in terms of homogeneous coordinates:

$$\mathbf{M}\mathbf{X} = 0 \quad (4)$$

where

$$\mathbf{M} = \begin{bmatrix} p_{11} - up_{31} & p_{12} - up_{32} & p_{13} - up_{33} & p_{14} - up_{34} \\ p_{21} - vp_{31} & p_{22} - vp_{32} & p_{23} - vp_{33} & p_{24} - vp_{34} \\ p'_{11} - u'p'_{31} & p'_{12} - u'p'_{32} & p'_{13} - u'p'_{33} & p'_{14} - u'p'_{34} \\ p'_{21} - v'p'_{31} & p'_{22} - v'p'_{32} & p'_{23} - v'p'_{33} & p'_{24} - v'p'_{34} \end{bmatrix}, \quad (5)$$

$$\mathbf{X} = [X \ Y \ Z \ 1]^T. \quad (6)$$

Thus, we can estimate the scene point through singular value decomposition (SVD)-related techniques. The four elements of the last column of \mathbf{V} obtained by the SVD of \mathbf{M} (i.e., $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T$) are the homogeneous coordinates of \mathbf{X} .

B. Pose Estimation

The traditional industrial workpieces in the manufacturing assembly line contain geometric features such as the plane, the cylinder, the cone, and the sphere; some of them have rotational symmetric characteristics through the principal axis.

Most pose estimation methods of vision-guided robotic systems make use of the complete knowledge of the parts, such as geometric features in an image and CAD models [11][12]. In this case, every time that the workpiece changes, an operator must register the features of the new object in the database, which reduces the productivity of the manufacturing line. Moreover, if the features of the object cannot be extracted, the pose of the object cannot be

calculated.

To overcome the problems of conventional methods, we propose a new pose-estimation algorithm that calculates the pose of an object by fitting the geometric primitives (as mentioned previously), thus extracting the rotational symmetric features of industrial parts without requiring the prior complete knowledge of the workpiece.

The process of pose estimation for the rotational symmetric object is divided into three phases. First is the process that segments the primitives by using detected 3D edges from a range image; the second process identifies rotational symmetric primitives among the separated primitives and creates a model; the third phase involves the estimation of the variance of the pose of the object by making comparisons between the reference object in the database and the modeled object.

1) Object Segmentation

To grasp a part among the stacked objects, the target object must be extracted. In this study, object segmentation is determined with the use of the 3D edges and the normal vector map. When compared with the 2D edges of an image, the 3D edges can segment out the object from a background or neighboring objects, regardless of texture, because this system makes use of the difference of depth rather than the difference of intensity [6].

To determine the 3D edge, we apply Wani's 3D edge-detection method, which identifies three types of edges: the fold edge (FE), the boundary edge (BE), and the semistep edge (SE) [13]. The FE is the edge that corresponds with pixels where normal surface vector discontinuity occurs. BE is formed by pixels that have at least one immediate neighbor pixel that belongs to the background, that occurs where one object obstructs another object or a part of itself. To identify the 3D edges, normal vectors are calculated in whole areas of a scene with the use of the cross-product of a directional vector [6]. With the use of the difference of normal and the difference of depth, the three kinds of 3D edges are extracted. Among the three types of 3D edges, FE and BE are needed to distinguish objects from the background and from neighboring objects [14]. Using the FE and the BE, we were able to find the boundary of the surface that encloses the primitives in the range image. Finally, we divide the surface that is contained with the equivalent primitive features in the input image. Fig. 6 shows the results of 3D edge detection.

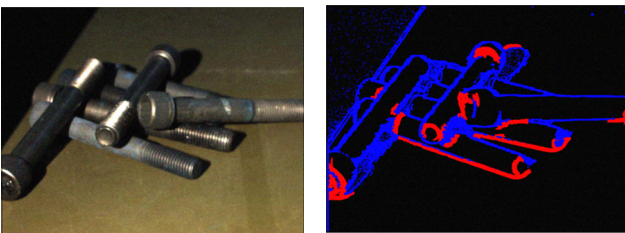


Fig. 6. Input image (left) and 3D edge image (right). Red, Fold edge; blue, boundary edge.

2) Primitive Object Modeling

After recognizing the primitive objects through the previous process, it is necessary to perform an additional modeling process to detect the rotational symmetric object. In this study, we used Baik's modeling method [14].

As shown in Fig. 7, if two points are on the surface of an object at the same height, the vectors whose directions are opposite to the ones of the normal vectors of two points will meet at a point on the axis of the object at the same distances from the surface.

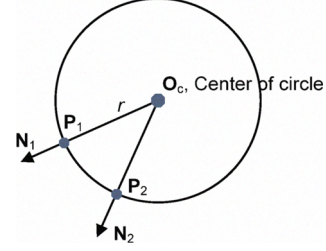


Fig. 7. Features of a rotational symmetric object

We then calculated the center points at several heights to find the equation of the axis.

We defined two points on the surface of an object as $\mathbf{P}_1=[x_1 \ y_1 \ z_1]^T$ and $\mathbf{P}_2=[x_2 \ y_2 \ z_2]^T$ and two normal vectors on \mathbf{P}_1 and \mathbf{P}_2 as $\mathbf{N}_1=[n_{x1} \ n_{y1} \ n_{z1}]^T$ and $\mathbf{N}_2=[n_{x2} \ n_{y2} \ n_{z2}]^T$, respectively. The equations of the normal vectors are shown here:

$$\mathbf{N}_1 = \begin{bmatrix} \frac{x_{p1} - x_c}{r} & \frac{y_{p1} - y_c}{r} & \frac{z_{p1} - z_c}{r} \end{bmatrix}^T \quad (7)$$

$$\mathbf{N}_2 = \begin{bmatrix} \frac{x_{p2} - x_c}{r} & \frac{y_{p2} - y_c}{r} & \frac{z_{p2} - z_c}{r} \end{bmatrix}^T \quad (8)$$

The point $\mathbf{O}_c=[x_c \ y_c \ z_c]^T$ is a point on the axis of the object, and r is the radius of the object. Because we know the values of \mathbf{P}_1 , \mathbf{P}_2 , \mathbf{N}_1 , and \mathbf{N}_2 , we can derive the following equations from Equations (7) and (8) and obtain the three radiuses r_x , r_y , and r_z :

$$x_c = (n_{x1}x_2 - n_{x2}x_1)/(n_{x1} - n_{x2}),$$

$$y_c = (n_{y1}y_2 - n_{y2}y_1)/(n_{y1} - n_{y2}), \quad (9)$$

$$z_c = (n_{z1}z_2 - n_{z2}z_1)/(n_{z1} - n_{z2}),$$

$$r_x = (x_1 - x_c)/n_{x1} = (x_2 - x_c)/n_{x2},$$

$$r_y = (y_1 - y_c)/n_{y1} = (y_2 - y_c)/n_{y2}, \quad (10)$$

$$r_z = (z_1 - z_c)/n_{z1} = (z_2 - z_c)/n_{z2}.$$

If both $|r_x - r_y|$ and $|r_y - r_z|$ are within the threshold (e.g., 2 mm), we can consider the point $\mathbf{O}_c=[x_c \ y_c \ z_c]^T$ to be a point on the axis.

In practice, we randomly selected two points, \mathbf{P}_1 and \mathbf{P}_2 ; the first one is on the surface of the object, and the second one is every other point on the surface of the object. We then computed the point \mathbf{O}_c . By iterating the process several times (e.g., for 30% of all of the object's surface points), we can obtain the set of points that are considered to be on the axis.

If there are more than a certain number of acquired points (e.g., 30 points), then the object is rotation symmetric.

By calculating the covariance matrix from the set of points and performing SVD, we can get three directional vectors from three unique variables. The vector that corresponds with the largest value of the unique variables becomes the vector of the initial axis.

There can be error in the initial axis as a result of noise. To fix this error, we set up the axis in parallel with the Z-axis (as shown in Fig. 8), divided the axis by 1 mm, obtained cross-sections of the object in the X-Y plane, and gathered the points from all cross-sections.

The center of a circle that is fitted on the points of each cross-section becomes a point on the new axis. By gathering the center points of all of the cross-sections and fitting them into a line, we obtained an accurate model of the rotation symmetric object.

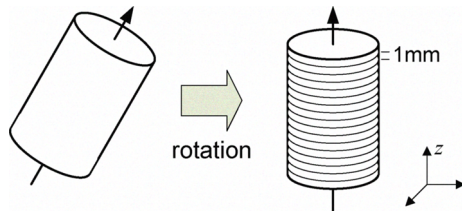


Fig. 8. Correction of the principal axis

3) Finding the Position of Rotation Symmetric Objects

The results of the modeling of rotation symmetric objects are the direction of the axis, the points on the axis, and the radius of each cross-section. The direction of the axis shows the position of the object in 3D, and the center point of the axis shows the location of the object.

To determine if the object is the target for bin picking, the object should be compared with the information stored in the database. The information stored in the database is the radius of each cross-section, which means that whether the object is the target or not can be determined by comparing the radius information of each cross-section.

Equation (11) is used to compare the values of the radius. \mathbf{R}_i is the set of radius values of the cross-sections, and \mathbf{R}_m is the radius of the modeling object. If Equation (11) is satisfied, we can confirm that the object is the target for bin picking.

$$\text{Threshold} \geq \text{Average}(|\mathbf{R}_i - \mathbf{R}_m|) \quad (11)$$

4) Variance between Reference Object and Modeling Object

The measured 3D pose of the reference object is registered in the database. During the operation stage, the shifted data between the 3D pose of the reference workpiece and the measured 3D pose of the incoming workpiece is transmitted via RS-232 or Ethernet to the robot controller. According to the shifted data, the robot completes the task even though the workpieces are placed in an arbitrary fashion.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

For our experiments, the Optoma EX330 projector and the PGR flea2 IEEE 1394 digital camera were used for the structured light system. The resolution of the projector was 1024×768 , and the resolution of the camera was 640×480 . The position of the camera was about 20 cm away from the projector. There were 256 code strings in a horizontal direction and 192 code strings in a vertical direction that were used for pattern projection. The distance between the camera/projector and the objects was about 0.6 m to 1 m (minimum and maximum distances). A HYUNDAI HA020 robot was used for the bin picking task, and it picked an M12-sized bolt from randomly stacked environments.

After the camera/projector calibration process, we put the reference object on the pallet. The position and pose of the object were registered in database by means of the structured light sensor, as shown in Fig. 9. We then taught the trajectory of the robot for the bin picking task with respect to the reference object. Fig. 10 shows the modeling result of the reference object.

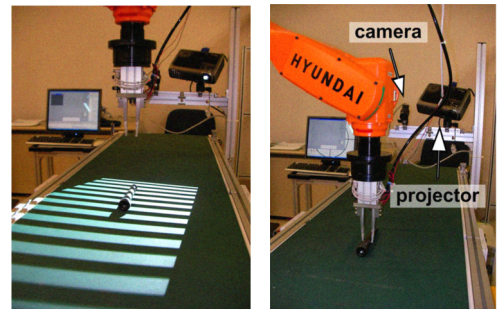


Fig. 9. The position and pose of the reference object is registered (left), and the trajectory of the robot is taught during the preparatory stage (right).

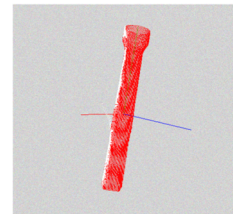


Fig. 10. Modeling result of the reference object

During the operation stage, we randomly scattered the M12-sized bolts on the pallet, as shown in Fig. 11. The structured light sensor then acquired the 3D range image and calculated the pose of the workpieces. The right-hand image of Fig. 11 shows the 3D reconstruction image.

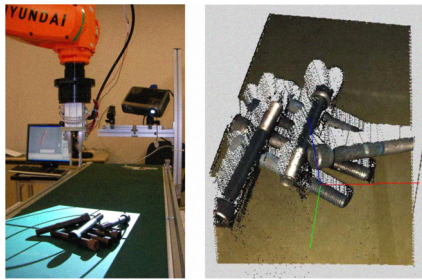


Fig. 11. Randomly piled objects (left) and modeling results (right)

B. Qualitative Results

Fig. 12 explains the process of segmenting the target object and estimating its position. The left-hand image of Fig. 12 represents the primitives that were detected with the proposed object segmentation method; the gray regions are segmented objects, and the white regions are planes. The modeling results of the randomly scattered objects are shown in the right-hand image of Fig 12.

The target object to be grasped was determined by the quantity of modeling. In most cases, the top object among the stacked objects was selected as the object to be picked up.

After the object to be picked up was selected, the variance between the pose of the reference object and the selected object was transmitted via the RS232C. Finally, we examined whether the robot could successfully grasp the bolt. Fig. 13 shows the robot picking up an object from among the randomly piled bolts.

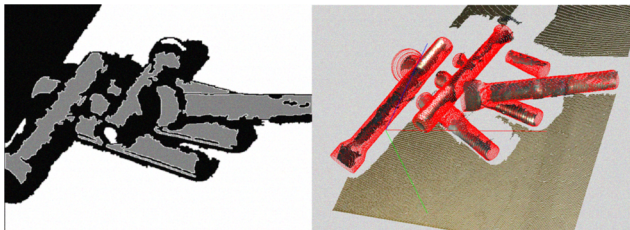


Fig. 12. Object segmentation(left) and primitive object modeling(right)

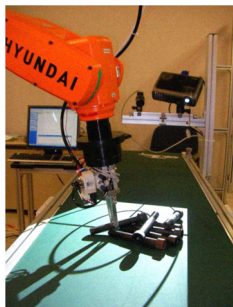


Fig. 13. The robot successfully picks up a part from among the randomly piled objects.

C. Quantitative Results

To estimate the repeatability of the proposed method, we compared the error of the rotational symmetric object's radius and orientation with the average value of 30 measurements. Tables 1 and 2 show the repeatability of the position and orientation for the proposed method, respectively.

TABLE 1. REPEATABILITY OF THE SPECIMEN RADIUS (UNIT: mm)

	Radius
repeatability	± 0.15

TABLE 2. REPEATABILITY OF THE SPECIMEN ORIENTATION (UNIT: DEGREE)

	x	y	z
repeatability	± 0.482	± 0.549	± 1.260

Through the qualitative and quantitative experiments performed with the use of the industrial robot, we verified that this system could be employed for certain material handling and automotive plant assembly applications.

V. CONCLUSION

In this article, we proposed a structured light based bin picking system using primitive models. To obtain a reliable range image, we made use of a structured light sensor that consisted of a camera and a DLP projector, which illuminated 8-bit antipodal Gray-coded patterns. In contrast with the conventional pose estimation method for a bin picking system, which requires complete knowledge of the object (e.g., geometric features in an image, CAD models), we estimated the pose of the object with known geometric primitive information through primitive segmentation, rotational symmetric object modeling, and recognition.

The experimental results obtained with the use of an industrial robot confirmed that the proposed system could be employed for a bin picking system. In the near future, we will study the pose estimation method for complex objects with free-form surfaces, and we will look methods to determine multiple grasp points without collision.

ACKNOWLEDGMENT

This research was performed for the Intelligent Robotics Development Program, which is one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge Economy of Korea. This research was also supported by Hyundai Heavy Industries as a part of an industry-academic cooperative research program.

REFERENCES

- [1] W. Iversen, "Vision-guided Robotics : In Search of the Holy Grail," *Automation World*, pp. 28-31, February 2006.
- [2] W. Hardin, "Vision enables freestyle bin picking," *Vision System Design*, vol. 12, no. 6, Jun 2007.
- [3] K. Ban, F. Warashina, I. Kanno and H. Kumiya, "Industrial Intelligent Robot," *FANUC Tech. Rev.*, vol. 16, no. 2, pp 29-34, August 2003.
- [4] K. Rahardja and A. Kosaka, "Vision-based Bin-Picking: Recognition and Localization of Multiple Complex Objects using Simple Visual Cues," *Proc. of the IEEE/RSJ international Conf. on Intelligent Robots and Systems*, vol. 3, pp. 1448-1457, November 1996.
- [5] R. D. Schraft and T. Ledermann, "Intelligent picking of chaotically stored objects," *Assembly Automation*, vol. 23, no. 1, pp. 38-42, 2003.
- [6] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2003.
- [7] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

- [8] B. Jähne and H. Haußecker, *Computer Vision and Applications: A Guide for Students and Practitioners*, Academic Press, 2000.
- [9] D. Kim, M. Ryu, and S. Lee, "Antipodal Gray Codes for Structured Light," *Proc. of the IEEE International Conference on Robotics and Automation*, pp. 3016-3021, May 2008.
- [10] B. K. P. Horn, "Closed-Form Solution of Absolute Orientation Using Unit Quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629-642, 1987.
- [11] N. Fujiwara and T. Onda "Three -Dimensional Circle Detection and Location of Pipe Joints for Bin-Picking Tasks" *Proc. of the IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, vol. 2, pp. 1216-1221, 1998.
- [12] C. R. Hema and M. P. Paulraj, "Segmentation and Location Computation of Bin Objects," *International Journal of Advanced Robotic Systems*, vol. 4, no. 1, pp. 57-62, 2007.
- [13] M. A. Wani and B. G. Batchelor., "Edge-Region-Based Segmentation of Range Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, March 1994.
- [14] K. -K. Baek and Y. -C. Park, "Object/Environment Segmentation based on Generic Model for 3D Object Recognition and Modeling," *Korean CAD/CAM Conference*, pp. 414-418, 2008.