

Robotic vision system for random bin picking with dual-arm robots

Sangseung Kang, Kyekyung Kim, Jaeyeon Lee and Joongbae Kim

Electronics and Telecommunications Research Institute, 34129, Daejeon, Korea

Abstract. Random bin picking is one of the most challenging industrial robotics applications available. It constitutes a complicated interaction between the vision system, robot, and control system. For a packaging operation requiring a pick-and-place task, the robot system utilized should be able to perform certain functions for recognizing the applicable target object from randomized objects in a bin. In this paper, we introduce a robotic vision system for bin picking using industrial dual-arm robots. The proposed system recognizes the best object from randomized target candidates based on stereo vision, and estimates the position and orientation of the object. It then sends the result to the robot control system. The system was developed for use in the packaging process of cell phone accessories using dual-arm robots.

1 Introduction

Automation has been most successfully used in manufacturing systems, and can be an effective solution in part-packaging processes at industrial sites because such processes are simple and repetitive tasks [1]. To achieve such automation, the solution applied must be able to provide reliable detection for the target parts during the packaging process. Further, it is important to operate the automated picking function of randomized parts stored in bins. Although an improvement in the accuracy of bin picking has already been extensively studied, estimating the position and pose of the target objects remains a difficult issue. Therefore, the development of applicable and reliable methods for solving the problems inherent to industrial bin picking is required.

There has been a large body of research on vision systems used for bin picking applications. Such vision systems use 2D or 3D feature information for object detection. Single-view based systems are affected by the environmental conditions of the handled objects, including the lighting reflections and an overlapping of the target objects. Rahardja and Kosaka [2] propose a vision-based bin picking architecture that utilizes simple visual cues as triggers toward recognition and pose estimation of target objects. Rodrigues et. al. [3] build a multi-light imaging system and develop a data-driven method for pose estimation using random ferns to map the patch appearance into pose hypotheses votes. Liu et. al. [4] use a multi-flash camera that extracts robust depth edges and a shape-matching algorithm called fast directional chamfer matching.

Most 3D vision systems exploit 3D object models for image interpretation. Such approaches include the use of

3D range sensors based on line laser scanners, time-of-flight cameras, or structured light-based area scanners. Boughorbel et al. [5] combined video and range images. They generate the 3D model of parts and objects in the bin and find the geometry of the bin contents. Fuchs et. al. [6] approach a bin picking system by combining an impedance controlled lightweight robot and a time-of-flight camera. They used the camera for fast modelling of the dynamic environment and for localizing the bin plus the objects therein. Scharstein and Szeliski [7] propose a method for acquiring high-complexity stereo image pairs with pixel-accurate correspondence information using structured light. Papazov et al. [8] present a 3D object recognition and pose estimation approach for grasping procedures using a Kinect sensor.

Model-based vision systems are expensive and require predefined models of the handling objects. Low-cost sensors such as a Kinect are attractive alternatives to expensive 3D sensors, but do not include reliable object detection owing to a poor quality of the input depth data [9].

In this paper, we present a robotic vision based bin picking system for industrial robotics applications. The proposed system is used in the packaging process for cell phone accessories using dual-arm robots. It recognizes the best picking object from randomized target objects based on stereo vision, and sends the position and orientation information of the object to the robot control system. The robot system conducts a packaging operation using a pick-and-place task.

The developed robotic vision system performs detection and recognition functions of the target object in multiple bins and provides the information for sequential executions of the pick-and-place operations to the robot control system. We set up a vision based random bin

picking test environment and performed experiments on the proposed method.

The remainder of this paper is organized as follows. Section 2 presents the method of object detection and recognition used by our robotic vision system for randomized bin picking based on stereo vision. Section 3 describes the selection of the candidate objects to be picked up by the proposed system and the related components. Section 4 presents the experimental results obtained from the position and pose estimation experiments. Finally, some concluding remarks are given in Section 5.

2 Object detection

For the task of picking a random object from a bin, the bin picking system comprehends all aspects, from identifying the objects to choosing the applicable candidates. The system determines the optimal approach path to pick the object, and controls the robot movements. It requires vision sensors, operating in conjunction with a computer for processing the sensed data. For a different approach [10], we developed a vision-based bin picking system using a single high-resolution camera. The system estimates the pose and distance of the object using geometrically transformed parameters of the local features. In this paper, we introduce a stereo vision based approach to detect a particular object from a bin and identify its position and orientation. Figure 1 shows our vision system used for bin picking, and randomized target objects. The right side of the figure shows our system along with a ring light device.

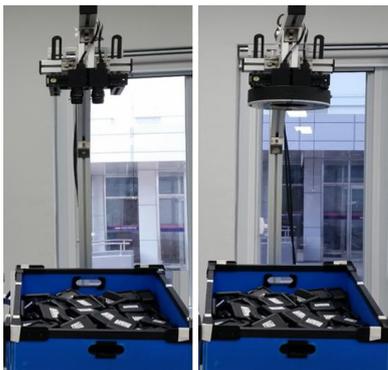


Figure 1. Vision system and randomized target objects.

It is important to apply the optimal illumination to acquire high-quality images of the target objects. In the process of applying illumination, it is necessary to consider the influence of the illumination in accordance with the pattern and direction of the light. We estimate the optimal angle formed by the surface of the object and the incident light, and attempt to solve the problem caused by the incidence of shade in a particular lighting direction. We also use a diffuser plate and a polarizing filter along with the lighting itself. Polarization is a general descriptor of light, and contains information on the reflecting objects; in addition, based on its wave property, light can oscillate with more than a single

orientation [11]. The application of a polarization filter and a diffuser increases the resolution and contrast for the same light source, and minimizes the occurrence of specular highlights arising from the surface. Sample cases of applying illumination are shown in Figure 2. One is a case of an overexposure of lighting, and the other is an appropriate illumination using a lighting device with a polarizing filter.

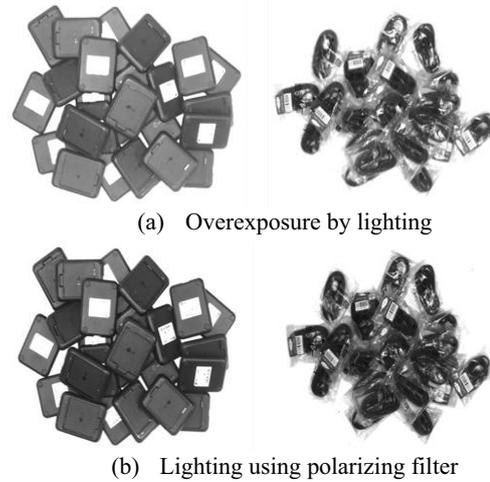


Figure 2. Sample case of applying illumination.

Two different images of the target objects are captured using a stereo camera; the amount of noise is then reduced through a pre-processing procedure, and the images are converted into a binary form using different threshold values. The system detects the contours of the objects from the filtered images after applying morphological filters [12], and selects the candidate objects through an analysis of the contour size. A binary mask is generated from the detected contours, and SIFT feature points [13] are extracted. The feature points are extracted from limited local areas using the binary mask described above, and around less than 100 feature points are extracted within the each local area. This has the advantage of improving both the accuracy and speed.

After the feature point detection for each view is conducted, a matching function is applied by comparing the extracted feature points from both the left and right camera images [14], and the feature matching results that are deemed unsuitable based on the predefined conditions are removed. For example, the difference in scale of the feature points must be less than a single step, and the matching points must maintain a one-to-one relationship. If two corresponding points are obtained between the left and right stereo images, the system can calculate the three-dimensional spatial point using the parameter information from the stereo cameras such as optical centre, focal length, and baseline between the cameras. Let $p_l=[x_l, y_l]^T$ be the left matching points, and $p_r=[x_r, y_r]^T$ be the right matching points, the three-dimensional spatial point $P=[x_s, y_s, z_s]^T$ can then be calculated as follows.

$$x_s=(b(x_r-x_c))/d, y_s=(b(y_r-y_c))/d, z_s=fb/d \quad (1)$$

The disparity is represented as $d=x_l-x_r$, where $[x_c, y_c]^T$ is the primary optical centre, f is the focal length of a camera, b is the baseline or distance between two cameras. X is the X-axis and Z is the optical axis of a camera. Both cameras must have the same focal length. Figure 3 shows the three-dimensional reconstruction geometry based on stereo triangulation.

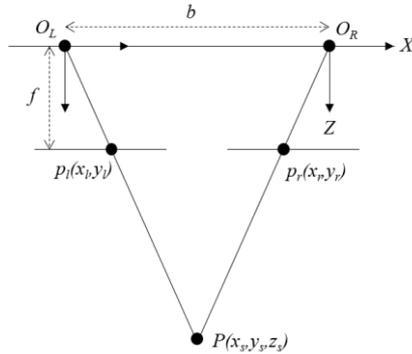


Figure 3. Three-dimensional reconstruction geometry.

The equation of a plane is determined using the reconstructed three-dimensional spatial points. Through the calculation of the centroid from the four corner points, the system estimates the position of the target object. It also estimates the orientation of the object based on a calculation of the normal vector of the plane. Let $ax+by+cz+d=0$ be the equation of a three-dimensional plane, and $P=[a, b, c, d]^T$ be the plane vector, the plane equation can be rewritten as $[x \ y \ z \ 1][a \ b \ c \ d]^T=0$. If there are n three-dimensional points, the linear system can be expressed through an extension to an n -dimensional matrix. The plane vector can then be calculated using singular value decomposition.

The system estimates the pose of the object. Let the four corners of the detected rectangular area be p_1 through p_4 , vectors e_1 through e_4 can then be calculated using the intrinsic camera parameter. The centroid of the detected rectangular area is used as the reference picking point for the robot system. For the given plane, the crossing point can be calculated using the plane equation, the normal vector of the plane, and the normal vector of e_1 through e_4 . The centroid can be calculated by dividing the sum of the four crossing points.

The orientation of the object is determined based on the x-, y-, and z-axis vectors of the coordinate system at a corner point of the detected rectangular area.

3 Picking

In general, good objects for picking may be considered to be at higher positions. Therefore, the system will choose the best object among the picking candidate objects based on the distance from the camera. To reduce any errors, the candidates are also prioritized based on additional comparisons of the scale, orientation and aspect ratio of the targets. Let D_n be the normalized distance, O_n be the normalized orientation, S_n be the normalized scale factor of the detected object, the probable cost is given by

$$p(\omega_i) = D_n \cdot \delta + O_n \cdot \rho + S_n \cdot \lambda \quad (2)$$

where δ is the distance weight factor, ρ is the orientation weight factor, and λ is the scale weight factor.

The weight factors are appropriate values determined through a large number of experiments. For all detected objects, the best object to pick up was selected as the object with the lowest cost.

$$\omega_N = \arg \min(p(\omega_i)) \quad (3)$$

An end effector is a device for picking up objects through robotic manipulation, and serves various purposes. It is attached to the end of a robotic arm and is a part of an industrial robot for interaction with the work environment. The selection of an end effector depends on the application of the industrial robot. Robot grippers are widely used, and support grasping functions between the gripper and the object to be grasped [15]. In general, a gripper can be mechanical, but suction-based grippers are also available. Our dual-arm robot system uses vacuum suction-type end effectors. Therefore, it is important to accurately estimate the position of the centroid of the object. This position is used as the picking point for the robot system. The normal vector of the detected rectangular area can be calculated using the plane equation. The direction of the normal vector is used as the picking orientation for the robot system. The sample results of position and pose estimation for robotic picking are shown in Figure 4.



Figure 4. Pose estimation and picking direction.

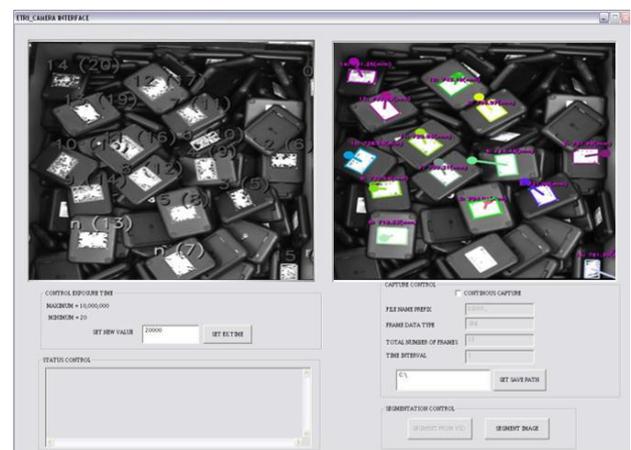


Figure 5. Robotic vision client system.

The vision system converts the result of the vision coordinate system into the value of the robot coordinate

system, and sends the result information to the dual-arm robot control system. The robot system conducts the picking operation based on the position and orientation information of the target object. The implemented robotic vision client system is shown in Figure 5. The left side of the image viewer area shows a stereo matching and classification result, and the right side shows a measurement result with a camera coordinate system.

4 Experiments

We set up a random bin picking test environment using a robotic vision system, and conducted experiments on the proposed method. In the experimental study, charging cradles and travel adapters, which are different cell phone accessories, were used for the objects. The objects were randomly contained in a bin, and were placed in such a way that they partially occluded each other. The objects were located approximately 800 mm away from the camera.

We applied the object detection and position estimation functions, and compared the measured distance and actual distance values. In the distance measurements taken from the input images, the offset value is applied to correct the actual distance. Samples of the experimental results for the charging cradles and travel adapters are shown in Table 1.

Table 1. Experimental results.

Distance	Measure	Error	Distance	Measure	Error
807.2	807.1	0.1	815.8	814.5	1.3
828.4	828.8	0.4	829.2	827.7	1.5
837.2	836.6	0.6	811.6	813.4	1.8
829.6	828.3	1.3	826.7	827.5	0.8
810.4	811.2	0.8	818.7	820.3	1.6
827.1	824.9	2.2	828.5	827.6	0.9
812.2	812.1	0.1	830.3	830.8	0.5
810.8	809.8	1.0	827.3	826.2	1.1
Avg. Case 1		0.81	Avg. Case 2		1.19

Table 2. Distance accuracy.

Test Case	Avg.Err.1	Std.Dev.1	Avg.Err.2	Std.Dev.2
1	0.81	0.70	1.19	0.44
2	0.76	0.59	1.02	0.83
3	0.67	0.33	0.75	0.61
4	0.93	0.62	1.26	0.64
5	1.09	0.43	1.16	0.97
Avg.	0.85	0.53	1.08	0.70

In the experiment on the charging cradles, the mean distance error was 0.81, and the standard deviation was 0.7, for eight detected objects with priority. The mean distance error for the travel adapters was 1.19, and the standard deviation was 0.4. The distance accuracy for the travel adapters was relatively low because they are smaller in size and wrapped in plastic. We repeated the experiment on the charging cradles and travel adapters

five times. The results of this experiment are shown in Table 2.

As demonstrated through the experiment results above, the average error rate in the distance to the target object was around 1 mm. The reference distance was obtained using a laser distance measurer, which has a measurement accuracy of 1 mm. Therefore, even when assuming that an error of up to 2 mm may occur, there was no problem for the robot to pick up the object in our bin picking environment. The developed bin picking system with a dual-arm robot is shown in Figure 6.



Figure 6. Random bin picking.

5 Conclusion

One of the most difficult problems in industrial robot applications is bin picking. Although many different solutions to the problems inherent to bin picking have been proposed, a general solution remains unavailable. In this paper, we presented a robotic vision system for randomized bin picking based on stereo vision. The proposed system detects candidate objects from randomized objects, estimates the position and orientation of the target candidates, and determines the best object to pick up for a pick-and-place task of a robot. We developed a pilot system for use in the packaging process of cell phone accessories using dual-arm robots. More research is needed to apply practical solutions to many different real-world situations.

Acknowledgement

This work was supported by the R&D program of MOTIE & KEIT [10038660, Development of dual-arm robot system based on multi-robot cooperation for cell manufacturing process of IT products]

References

1. C. Harrison, "A review of automation in manufacturing illustrated by a case study on mixed-mode hot forging," *Manufacturing Review* (2014)
2. K. Rahardja and A. Kosaka, "Vision-Based Bin-Picking: Recognition and Localization of Multiple Complex Objects Using Simple Visual Cues," *Proc. of the Intl' Conf. on Intelligent Robots and Systems*, 1448-1457 (1996)
3. J. Rodrigues, J. Kim, M. Furukawa, J. Xavier, P. Aguiar, T. Kanade, "6D pose estimation of textureless shiny objects using random ferns for bin-

- picking," *Proc. of the Intl' Conf. on Intelligent Robots and Systems*, 3334-3341 (2012)
4. M. Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T. Marks, R. Chellappa, "Fast Object Localization and Pose Estimation in Heavy Clutter for Robotic Bin Picking," *The International Journal of Robotics Research*, **31**(8): 951-973 (2012)
 5. F. Boughorbel, Y. Zhang, S. Kang, U. Chidambaram, B. Abidi, and A. Koschan, "Laser ranging and video imaging for bin picking," *Assembly Automation*, **23**(1): 53-59 (2003)
 6. S. Fuchs, S. Haddadin, M. Keller, S. Parusel, A. Kolb and M. Suppa, "Cooperative Bin-Picking with Time-of-Flight Camera and Impedance Controlled DLR Lightweight Robot III," *Proc. of the Intl' Conf. on Intelligent Robots and Systems*, 4862-4867 (2010)
 7. D. Scharstein and R. Szeliski, "High-Accuracy Stereo Depth Maps Using Structured Light," *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (2003)
 8. C. Papazov, S. Haddadin, S. Parusel, K. Krieger, D. Burschka, "Rigid 3D Geometry Matching for Grasping of Known Objects in Cluttered Scenes," *International Journal of Robotics Research*, **31**(4): 538-553 (2012)
 9. S. Park and S. Lim, "Template-Based Reconstruction of Surface Mesh Animation from Point Cloud Animation," *ETRI Journal*, **36**(6): 1008-1015 (2014)
 10. K. Kim, S. Kang, J. Lee, J. Kim, "Vision Based Bin Picking for Industrial Robot," *Advanced Science and Technology Letters*, **64**: 14-17 (2014)
 11. A. G. Andreou and Z. K. Kalayjian, "Polarization Imaging: Principles and Integrated Polarimeters," *IEEE Sensors Journal*, **2**(6):566-576 (2002)
 12. P. Maragos, "Morphological Filtering for Image Enhancement and Feature Detection," *Image & Video Processing Handbook (2nd ed.)* (2004)
 13. D. Lowe, "Object recognition from local scale-invariant features," *Proc. of the Intl' Conf. on Computer Vision*, 1150-1157 (1999)
 14. D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, **47**(1): 7-42 (2002)
 15. G. Fantoni, M. Santochi, G. Dini, K. Tracht, B. Scholz-Reiter, J. Fleischer, T.K. Lien, G. Seliger, G. Reinhart, J. Franke, H.N. Hansen, A. Verl, "Grasping devices and methods in automated production processes," *CIRP Annals - Manufacturing Technology*, **63**(2): 679-701 (2014)