Volume : 4 | Issue : 6 | June 2015 ISSN - 2250-1991

Research Paper Engineering

Evaluation of Kinect Vision Sensor for Bin-Picking Applications

Improved Component Separation Accuracy with Combined Use of Depth Map and Color Image

Shigeharu Miyata

School of Engr., Hiroshima, Dept. of Robotics, Kindai U.,Hi-gashi-Hiroshima, JAPAN

ABSTRACT

This report describes a problem involved with use of Kinect depth maps for robot picking of randomly stacked components, and also a solution to this problems. When Kinect is installed above stacked parts and processing is performed using only the obtained Kinect depth map information, there are cases when individual small metal components cannot be separately identified. So that the robot can reliably pick up a single component in these cases, this report demonstrates that in areas where the system incorrectly identifies multiple components as a single component, the addition of color image information and blob analysis of the color image results in accurate separation of the individual components, allowing a single item to be identified for picking.

KEYWORDS 3D vision sensor, Depth map, Blob analysis, Color image, Labeling

I. INTRODUCTION

In a production workplace for industrial products, the processes from supply of product components to the production line to finishing machining are almost entirely automated. However the supply of components to the machining line is still performed using a conventional parts feeder, and at present it is still necessary for an operator to load the components into the feeder manually. With earlier production styles in which large quantities of each product were produced, components could be easily supplied to the production line by a parts feeder. However in order to support the present style in which multiple products are produced in small lots of each, consideration is being given to constructing systems that use a robot to pick components from a bin and supply them to the line, and many corporations are working to commercialize such a system. On the road to this commercialization, many studies are being carried out related to random bin picking systems [1]-[10]. However due to reasons such as limitations on the items that can be picked or the need to prepare multiple robot hands, it is impossible to say that robot picking is significantly more advantageous than the alternative of using a parts feeder and creating special jigs. Moreover because the vision systems which recognize the component position, shape, direction, andother elements are expensive, there are reasons why use of such systems in the workplace may be problematic.

A 3D vision system is essential in order to use a robot for picking components. Microsoft recently released the Kinect vision sensor for its Xbox gaming system. Because Kinect is low-priced (can be purchased for several hundred dollars), and allows color images and depth maps to be easily acquired, there have been studies concerning its use in robot picking [11]-[13]. Microsoft provides Kinect for Windows SDK free of charge with Kinect. This kit allows programs to be developed for a variety of purposes. It was decided to also use the Microsoft Kinect as the 3D vision sensor for this study in order to construct a system for robot picking of components stacked in bins.

When a robot is used to pick stacked components, the positions of the components and the directions they are facing are calculated based on the depth map obtained from Kinect. At this time, when multiple components are stacked randomly to approximately the same height from the bottom of the bin and are adjacent to one another, the system may incorrectly identify multiple components as a single component. In these cases, the robot hand may be unable to grasp the component securely or, in the worst case, the hand may be damaged by contact with the components. There is the risk that such incidents may result in severe problems that will stop the production line.

Based on the above, this report first shows that when only the Kinect depth map is used to identify individual stacked components, there are cases when the system will incorrectly identify multiple components as a single component. Next, as a solution to this problem, this report shows that when the color image that is obtained from Kinect at the same time is also used and blob analysis is performed, then in the cases when multiple components were earlier identified incorrectly as a single component, the system is able to separate the individual components and reliably select a single component as the target for picking.

In Section 2, we describe an overview of a measurement system using Kinect as the 3D vision sensor. Section 3 describes the method for separating and identifying individual components through combined use of the depth map information and color image information in areas where multiple components were incorrectly identified as a single component, and shows the test results. Section 4 describes our conclusions.

II. OVERVIEW OF THE 3D VISION SENSOR

We used Kinect in combination with software that was developed using the Windows software development kit (Kinect for Windows SDK). Because it uses the latest technologies including voice recognition, motion sensors, and skeleton tracking in order to allow manipulation of a virtual space by hand or body movements, Kinect is commonly used as an intuitive interface. The Kinect specifications are as follows.

[Hardware] Sensors: color camera × 1, depth camera × 1 (infrared), microphone × 4; port: USB 2.0; color camera resolution: 640×480/1280×960; depth sensor resolution: 320×240/640×480/80×80; depth sensor effective range: 0.4 m-3.0 m (near mode), 0.8 m-4.0 m (normal mode); light source: LED

Fig. 1 shows the Kinect external appearance and the processing screen of the software that was developed for processing the Kinect depth maps and color images. The left side of the processing screen shows the depth map and the right side shows the color image. The depth information is output at bottom right, making it possible to check the depth at the location selected with the mouse. The area indicated by the yellow box in the depth map is the area for measurement. The green area within it shows the candidate components for picking. This screen shows an example of when individual components are not separated and correct identification does not occur.

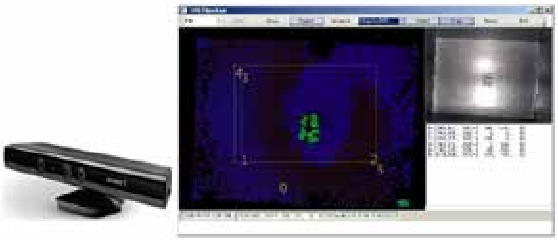


Figure 1: Kinect external appearance and processing screen.

Figure 2: Robot component picking.

III. IDENTIFICATION OF THE PICKING POSITION

A. When only the depth map information is used

Fig. 2 shows the test conditions for robot picking of stacked components based on Kinect image information. The components are piled in a bin of width 790 mm × length 630 mm × height 570 mm.

The component positions and directions are calculated based on the depth map obtained from Kinect. At this time, when multiple components are stacked randomly at approximately the same height from the bottom of the bin and are adjacent to one another (lying flat), the system may incorrectly identify multiple components as a single component, causing robot picking to fail.

Figure 2: Robot component picking.

Fig. 3 shows the conditions when the component ends are in contact at an angle, the conditions when components contact each other in a T shape, and the conditions when components are in contact in longitudinal and parallel rows.

Components ends are in contact at an angle.

In contact in a T-shape.

In a longitudinal row.

In parallel rows.

Figure 3: Component positions that prevent the robot from grasping the component.

Fig. 4 shows the depth map when the components are arranged in contact in longitudinal and parallel rows on the same plane. The yellow lines on the components indicate the longitudinal direction of the components as calculated from the image. The robot grasps the components at close to the center of this line. As you can see from the figure, because the calculated position is a position where 2 components are in contact with each other, it will not be possible to grasp the component correctly.

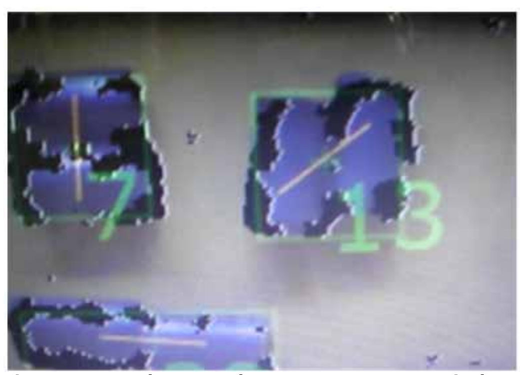


Figure 4: Depth map when components are in longitudinal and parallel rows.

B. **When only the depth map information is used**

Fig. 5 shows cases when the components were arranged on the same plane with the ends in contact at an angle and in a

T-shape, and the component positions were calculated using only the depth map information. The case 6 shows that although the components are correctly separated and identified as individual components when the ends are in contact, for cases 1 and 2, the 2 components are incorrectly identified as a single component.

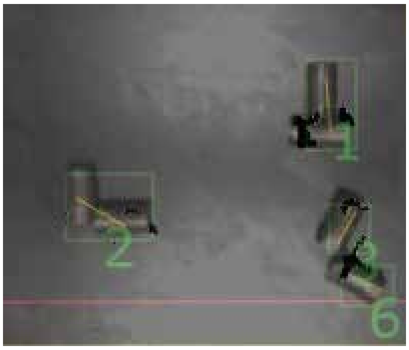


Figure 5: Examples of components with ends in contact at angles and in T-shapes that are each incorrectly identified as a single component.

Fig. 6 shows reproduced conditions where 6 components are in contact in longitudinal or parallel rows and recognition is difficult. The top 3 and bottom 3 components are each incorrectly identified as a single component.

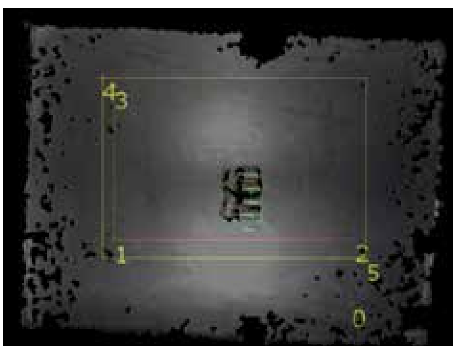


Figure 6: Example of 6 components arranged in longitudinal and parallel rows that are difficult to identify. The top 3 and bottom 3 components are each incorrectly identified as a single component.

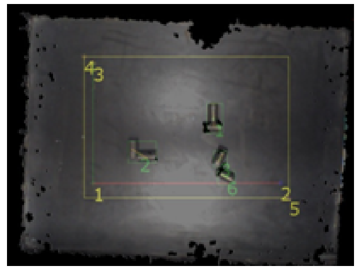
C. When only the depth map information is used

This section shows that in the cases of incorrect recognition in Fig. 5 and Fig. 6, it is possible to separate the individual components by performing blob analysis of a color image that is obtained at the same time as the depth map as described below.

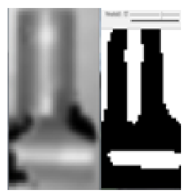
When the color image obtained from Kinect at the same time is also used and blob analysis [14] -[16] is performed, then in the earlier cases where multiple components were incorrectly identified as a single component, the components can be separated into individual components and a single component can be reliably selected for picking. Fig. 7 shows the algorithm used for this purpose.

Figure 7: Image processing algorithm for cases when multiple components are incorrectly identified as a single component.

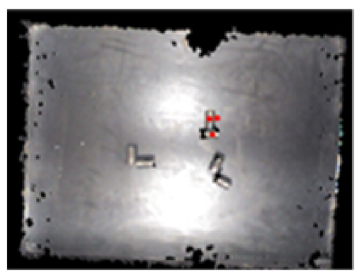
From the picking candidates selected by the green boxes shown in Fig. 5 and Fig. 6, the target for picking is decided and an ROI is set according to preset evaluation standards. The average of the pixel values within the ROI is found, and a binarized image is created using 1.15 times this average as the threshold value. (This factor was determined based on our experience.) From this binarized image, blob detection is performed and the feature value is calculated. Of the multiple blobs that exist, the labeling process is performed only for the blobs which have the area feature value that was set in order to create the labeled image. This shows that the labeled areas are separated component areas. At the same time we obtain a mask image that can be used for the separation process. Of the multiple separated components, the component which is closest to the center of the ROI becomes the picking target. The robot should be set to adopt a position parallel to the longitudinal axis of the selected component.



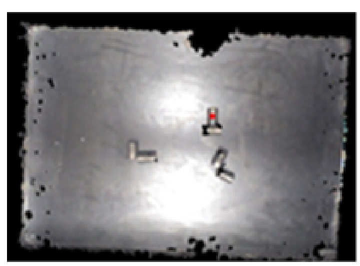
(8)-a.Depth map.



(8)-b.ROI.(8)-c.Binarized image.



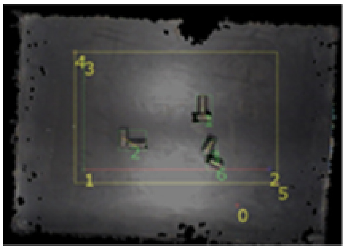
(8)-d. 3components individually identified.



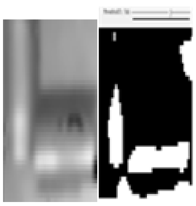
(8)-e. Finally a single component is selected for picking.

Figure 8: Processing results when 2 components are in contact in a T-shape.

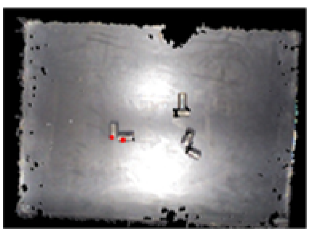
First we will explain Fig. 8. In Fig. (8)-a, the components in T-shape contact and labeled No. 1 have been selected as the picking target. In Fig. (8)-b, the areas corresponding to these component areas in the color image have been extracted. This is a binarized image with 1.15 times the average pixel value of the extracted area set as the threshold value. When blob analysis is performed, 3 blobs are detected. Finally, the component located closest to the center of the ROI, positioned vertically in the image, is selected as the picking target.



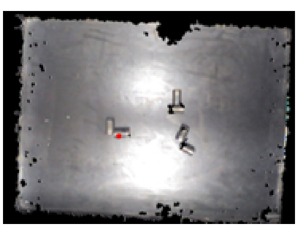
(9)-a.Depth map.



(9)-b.ROI.(9)-c.Binarized image.

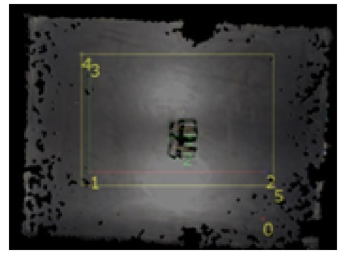


(9)-d. 2 components individually identified.

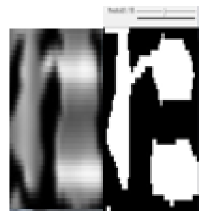


(9)-e. Finally a single component is selected for picking.

Figure 9: Processing results when 2 components are in contact in an L-shape.
Fig. 9 shows the processing results for 2 components in contact in an L-shape.
Fig. 10 shows the processing results for the top 3 components labeled as 1.
Fig. 11 shows the processing results for the bottom 3 components. These results show that by adding the color image information to the depth map processing results and performing blob analysis, in the end it is possible to select a single component as the target for picking.



(10)-a. Depth map.

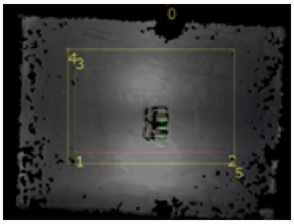


(10)-b.ROI.(10)-c.Binarized image.

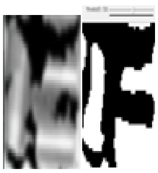
(10)-d.3 components individually identified.

(10)-e. Finally a single component is selected for picking.

Figure 10: Processing results for a cluster of the top 3 components.

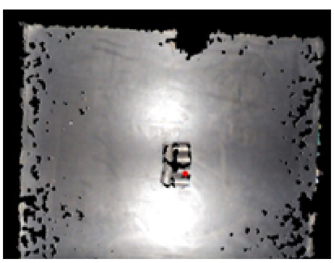


(11)-a. Depth map.



(11)-b. ROI. (11)-c Binarized image.

(11)-d. 3 components individually identified.



(11)-e. Finally a single component is selected for picking.

Figure 11: Processing results for a cluster of the bottom 3 components.

IV. CONCLUSIONS

This report describes a solution to a problem which occurs when Kinect depth maps are used for robot picking of stacked components.

If the component positions and directions are calculated based on the depth map obtained from Kinect, then when multiple components are stacked randomly to approximately the same height from the bottom of the bin and are adjacent to one another, the system may incorrectly identify multiple components as a single component. As a solution to this problem, this report demonstrates that it is possible to separate the multiple components into individual components by using the color image that is obtained from Kinect at the same time and performing blob analysis.

For blob analysis, the average pixel value for the ROI was calculated and binarization was performed using 1.15 times this average as the threshold value. (This factor was determined based on our experience.) This threshold value setting is a key point for blob analysis. We performed the process using a setting which we had identified from experience, however it will be necessary to set a suitable value with consideration for the surrounding light environment. The best method for calculating the threshold value remains an issue for the future.

After multiple candidate components were separately identified by blob analysis, when finally deciding the component for picking, it was decided to select the component which was closest to the center of the ROI. Future studies of a more rational method will be necessary in the future.

ACKNOWLEDGMENT

This study was conducted as a part of the 2013 Program to Support Companies for the Creation of New Business. We would like to thank President Shitanaka of Sigma Corporation who provided great support for this study and to everyone who we consulted with at the Hiroshima Prefectural Technology Research Institute.

REFERENCES

[1]E. Al-hujai and A Sood, “Range image segmention with applications to robot bin-picking using vacuum gripper,” IEEE Trans. On systems, man, and cybernetics, vol. 20, no. 6, pp. 1313-1325, Novm./ Dec. 1990. | [2]S. Kristensen, S. Estable, M. Kossow and R. Brosel, “Bin-picking with a solid state range camera,” Elsevier Robotics and Autonomus Systems, vol. 35, pp. 143-151, 2001. | [3]A. Zuo, J. Z. Zhang, K. Stanly and Q. M. J. Wu, “A hybrid stereo feature matching alorithm for stereo vision-based bin picking,” Int. Jour. Of Pattern Recgnition and artificial Intelligence, vol. 18, no. 8, pp. 1407-1422, Dec. 2004. | [4]O. Skotheim, J. T. Thielemann, A. Berge and A. Sommerfelt, “Robust 3D object localization and pose estimation for random bin picking with the 3DMaMa algorithm,” Porc. Of SPIE 7526, Three-Dimensional Image Processing(3DIP) and Applications, 75260E, Feb. 2010. | [5]A. Morales, T. Asfour and P. Azad, “Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands,” Proc. Of the 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 5663-5668, Oct. 9-15, Beijing, China, 2006. | [6]P. Piccinini, A. Prati and R. Cucchiara, “Real-time object detection and localization with SIFT-based clustering,” ELSEVIER Image and Vision Computing, vol. 30, pp. 573-587, 2012. | [7]I. Effenberger, J. Kuhnle and A. Vert, “Fast and flexible 3D object recognition solution for machine vision applications,” Porc. Of SPIE 8661, Image Processing: Machine Vision Applications VI, 86610N, March, 2013. | [8]D. Kraft, L.P. Ellekilde and J. A. Jorgensen, “Automatic grasp generation and improvement for industrial bin-picking,” Springer Tracts in Advanced Robotics, vol. 94, pp. 155-176, 2014. | [9]D. Holz, M. Nieuwenhuisen, D. Droeschel, J. Stuckler, A. Berner, J. Li, R. Klein, S. Behnke, “Active recgnition and manipulation for mobile robot bin picking,” Springer Tracks in Advanced Robotics, vol. 94, pp. 133-153, 2014. | [10]G. Sansoni, P. Bellandi, F. Leoni and F. Docchio, “Optoranger: a 3D pattern matching method for bin picking application,” ELSEVIER Optics and Lasers in Engineering, vol. 54, pp. 222-231, 2014. | [11]A. R. Fugl, A. Jordt, H. G. Peterson, M. Willatzen, R. Koch, | “Simultaneous estimation of material properties and pose for deformable objects from depth and color images,” Springer Lecture Notes in Computer Science, vol. 7476, pp. 165-174, 2012. | [12]A. Wilson, “Kinect API makes low-cost 3-D imaging systems attainable, Vision Systems Design, October 1, 2012. | [13]R. Mojtahedzadeh, A. Bouguerra and A. J. Lilienthal, “Automatic relational scene representation for safe robotic manipulation tasks,” Proc. Of 2013 IEEE/RSJ Int. Conf. on Intelligent Robotics and Systems(IROS), pp.1335-1340, Nov. 3-7, Tokyo, Japan, 2013. | [14]C. Damerval and S. Meignen, “Blob detection with wavelet maxima lines,”, IEEE Signal Processing Letters, vol. 14, no. 1, Jan., 2007. | [15]F. Wang, X. Ren and Z. Liu, “A robust blob recgnition and tracking method in vision-based multi-touch technique,”, Proc. Of 2008 Int. Symposium on Parallel and Distributed Processing with applications, pp. 971-974, 2008. | [16]P. C. Ravoor, R. S and S. R Rupanagudi, “Novel algorithm for finger tip blob detection using image processing,” Int. Jour. Of Signal Processing Systems, vol. 1, no.1, pp. 11-16, June, 2013. |