Recognition and Grasping Objects from 3D Environment by Combining Depth and Color Stereo Image in the Mobile Picking Robot System

Trong Hai Nguyen[1,3], Jong Min Oh[1], Dae Hwan Kim[1], Sang Kwun Jeong[2], Hak Kyeong Kim[1], and Sang Bong Kim[1(✉)]

[1] Department of Mechanical Design Engineering, Pukyong National University, Busan 48547, South Korea
haintmitu@yahoo.com, ohmin018@gmail.com, {kimdh2599, hakkyeong, kimsb}@pknu.ac.kr

[2] Department of Automation System, Korea Polytechnic Colleges, Jinju 52766, Republic of Korea
sk20140745@kopo.ac.kr

[3] Hutech High Technology Research Institude, Hochiminh, Vietnam

Abstract. This paper proposes recognition and grasping objects from 3D environment by combining depth and color stereo image in the mobile picking robot system. To do this task, the followings are done. Firstly, an image processing system including Kinect camera sensor is described. Secondly, RGB color map and new depth map for image inpainting are obtained using Kinect SDK mapping function to align RGB image with depth image. Thirdly, the new depth map are segmented to distinguish between the image background and the objects that should be recognized. The feature colours are generated based on colour histograms. Euclidean distance is used to measure the similarity between the feature vectors computed from the colour image and the feature vectors stored in a database. Fourthly, by converting RGB map and new depth map into 3D point clouds, an algorithm for localizing handle-like grasp affordances is proposed. The main idea is to search the point cloud for neighborhoods that satisfy handle-like grasp affordances and can be grasped by the end-effector of the manipulator. Finally, the effectiveness of the proposed algorithms is verified by using experiment. The experimental results show that the mobile picking robot successfully detects an object and finds its grasping points with an acceptable small error.

Keywords: Mobile robot · Object recognition · Object localization · Grasping object · 3D point cloud

1 Introduction

A robotic manipulation of objects typically involves object detection/recognition and grasping control. In the detection of the object, it is very similar with the basic process of pattern recognition: source data acquisition, preprocessing, feature extraction, classification training, and object detection. There are some popular features extraction:

V.H. Duy et al. (eds.), *AETA 2017 - Recent Advances in Electrical Engineering and Related Sciences: Theory and Application*, Lecture Notes in Electrical Engineering 465,
https://doi.org/10.1007/978-3-319-69814-4_34

color feature [1], edge feature [2], and texture features [3]. These features have their own advantages and limitations, so in reality they are often used in combination. This representation is often only suitable for "simple objects" (e.g., circles, crosses, rectangles, etc. in 2D or cylinders, cones in the 3D case). Robust robot grasping in novel and unstructured environments is an important research problem that has many practical applications. A key sub-problem is localization of the objects or object parts to be grasped. Localization is challenging because it can be difficult to localize graspable surfaces on unmodelled objects. Moreover, even small localization errors can cause a grasp failure. For RGBD image, several learning algorithm [4, 5] have shown promise in handling incomplete and noisy data and variations in the environment as well as grasping novel objects. Fischinger et al. [4] presented a learning approach for grasping unknown objects in a basket. In [5], a new rectangle representation algorithm based on the Kinect camera was proposed to localize the grasping point of the object. For 3D point cloud image, Papazov et al. [6] utilized a Kinect stereo camera sensor to acquire depth images of the scene. Bley et al. [7] proposed another approach of grasp selection by fitting learned generic object models to point cloud data. Choi et al. [8] proposed a Hough voting-based approach that extended point-pair features, which was based on oriented surface points, by boundary points with directions and boundary line segments. In most of these mobile manipulation demonstrations, the handled objects are well-separated. However, they are complex and expensive.

To solve that problems, this paper proposes recognition and grasping objects from 3D environment by combining depth and color stereo image in the mobile picking robot system. Recognizing of the object is obtained by using Kinect camera sensor based on Euclidean distance. For grasping object, an algorithm for localizing handle-like grasp affordances is proposed. Finally, the effectiveness of the proposed algorithms are verified by experiment. The experimental results show that the mobile picking robot successfully reaches the goal point with an acceptable small error.

2 System Description

Figure 1 shows the workspace of a picking robot system. The work space consisting of manipulator platform, stereo camera and a horizontal table with an object at the manipulator workspace. The Kinect sensor is placed on the table with 1 m height.

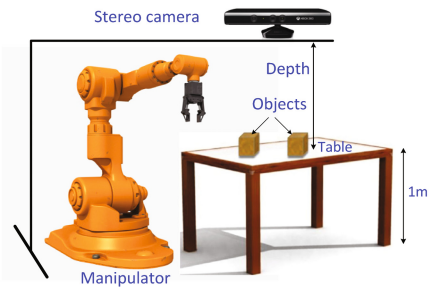


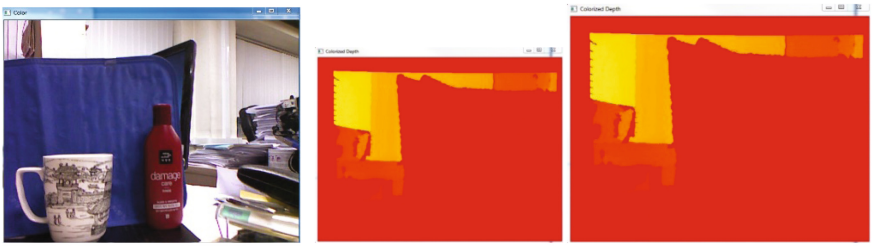
Fig. 1. Workspace of a picking robot system

3 Proposed Algorithm

3.1 Mapping RGB Image into Depth Map

The transformation of the color frame has to be performed because an original RGB color image has higher resolution than an original depth map. This transformation is achieved based on a mapping function of the SDK, which enables to map a corresponding color to a corresponding pixel in depth space for RGB color image to complete the lost depth information from the original depth map.

Figure 2 shows the result of original RGB image (a) original depth map (b) new depth map converted by a mapping function SDK (c).



a. Original color map b. Original depth map c. New depth map

Fig. 2. Mapping RGB image into depth map

The original color map is obtained from RGB camera with resolution of 640 × 480 pixels. The original depth map is obtained from the IR camera with resolution of 512 × 424 pixels and the new depth map has resolution of 640 × 480 pixels which enables to map a corresponding color (640 × 480 pixels) to a corresponding pixel in depth space.

3.2 Object Recognition

The segmentation of disparity maps or depth maps into objects and image background constitutes a problem that cannot be solved without additional constraints about the scene. A basic assumption is that objects located nearer to the camera show higher disparity values than the disparity values for the pixels representing the image background. The disparity values computed for the image background should be close to zero. Feature vectors were generated for each object in the database and for each segmented object in the image. Euclidean distance was used to measure the similarity between the feature vectors computed from the color image and the feature vectors stored in the database. A value close to zero indicates a high similarity between the feature vectors if the Euclidean distance is used. As opposed to this, a value close to one indicates a high similarity between the feature vectors if the scalar product is applied. Figure 3 show results for detecting a AETA book 2015.

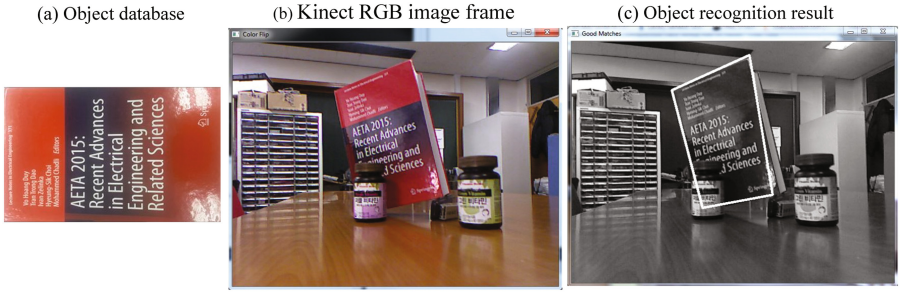


Fig. 3. Results for object recognition

3.3 Grasping Object

Process of localizing grasp affordances are as follows:

[Step 1] Randomly sample N neighborhoods in the point cloud.

Starting with the uniform case, given a pair of sample points $p_i = (x_i, y_i)$ and $p_j = (x_j, y_j)$, the points' correlation, $E(p_i, p_j)$, is assumed to decrease with the Euclidean distance, d_{ij}, between the two sampled points, p_i and p_j, as follows:

$$E(p_i, p_j) = \sigma^2 e^{-\lambda d_{ij}} \tag{1}$$

where $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, σ and μ are scale factor.

Based on their linear estimator, subsequently put forward the following representation for the mean square error, i.e., the deviation from the "ideal" image resulting from estimation error after the N^{th} sample as follows:

$$\varepsilon^2(p_0, \cdots, p_{N-1}) = \iint (\sigma^2 - U^T R U) dx dy \tag{2}$$

where $R_{ij} = \sigma^2 e^{-\mu d_{ij}}$, $U_i = \sigma^2 e^{-\mu\sqrt{(x_i - x)^2 + (y_i - y)^2}}$ for all $0 \le i, j \le N - 1$, (x, y) is arbitrary point.

[Step 2] For each neighborhood, fit an implicit surface in three variables to points in the local neighborhood. A quadratic surface can be described by $f(\mathbf{c}, \mathbf{x}) = 0$ as follows:

$$\begin{aligned} f(\mathbf{c}, \mathbf{x}) = {} & c_1 x_1^2 + c_2 x_2^2 + c_3 x_3^2 + c_4 x_1 x_2 + c_5 x_2 x_3 + c_6 x_1 x_3 + c_7 x_1 \\ & + c_8 x_2 + c_9 x_3 + c_{10} = \mathbf{c}^T l(\mathbf{x}) \end{aligned} \tag{3}$$

where $l(\mathbf{x}) = \left[x_1^2, x_2^2, x_3^2, x_1 x_2, x_1 x_3, x_2 x_3, x_1, x_2, x_3, 1\right]^T \in R^{10}$, $\mathbf{c} = [c_1, c_2, \cdots, c_{10}]^T \in R^{10}$.

$\mathbf{c}$ denotes the parameters of the quadratic surface and $\mathbf{x} = [x_1, x_2, x_3] \in R^3$ denotes the Cartesian coordinates of a point on the quadratic surface.

From Eq. (3), the following can be obtained:

$$\min \sum_{i}^{n} f(\mathbf{c}, \mathbf{x}^i)^2 = \mathbf{c}^T M(\mathbf{x}^i)\mathbf{c} \tag{4}$$

where $\mathbf{x}^1, \cdots, \mathbf{x}^n \in R^3$, $M(\mathbf{x}^i) = \sum_{i=1}^{n} l(\mathbf{x}^i) l(\mathbf{x}^i)^T \in R^{10\times 10} \equiv M$,

To avoid the trivial solution $\mathbf{c} = 0$, according to Taubin's method, by setting $\|\nabla f(\mathbf{c}, \mathbf{x}^i)\|^2 = 1$, the following can be obtained:

$$\mathbf{c}^T N(\mathbf{x}^i)\mathbf{c} = 1 \tag{5}$$

$$N(\mathbf{x}^i) = \sum\nolimits_{i=0}^{n} l_x(\mathbf{x}^i) l_x(\mathbf{x}^i)^T + l_y(\mathbf{x}^i) l_y(\mathbf{x}^i)^T + l_z(\mathbf{x}^i) l_z(\mathbf{x}^i)^T \equiv N$$

where ∇ is gradient, and $l_x(\mathbf{x}^i)$, $l_y(\mathbf{x}^i)$, $l_z(\mathbf{x}^i)$ denote the partial derivatives of $l(\mathbf{x}^i)$ with respect to x_1, x_2, x_3 as follows:

$$\begin{cases} l_x(\mathbf{x}^i) = \frac{\partial l(\mathbf{x}^i)}{x_1} = [2x_1,\ 0,\ 0,\ x_2,\ x_3,\ 0,\ 1,\ 0,\ 0,\ 0]^T \equiv l_x \\ l_y(\mathbf{x}^i) = \frac{\partial l(\mathbf{x}^i)}{x_2} = [0,\ 2x_2,\ 0,\ x_1,\ 0,\ x_3,\ 0,\ 1,\ 0,\ 0]^T \equiv l_y \\ l_z(\mathbf{x}^i) = \frac{\partial l(\mathbf{x}^i)}{x_3} = [0,\ 0,\ 2x_3,\ 0,\ x_1,\ x_2,\ 0,\ 0,\ 1,\ 0]^T \equiv l_z \end{cases} \tag{6}$$

From Eq. (3), the following is

$$\nabla f(\mathbf{c}, \mathbf{x}^i) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix}^T \begin{bmatrix} 2c_1x_1 + c_4x_2 + c_6x_3 + c_7 \\ 2c_2x_2 + c_4x_1 + c_5x_3 + c_8 \\ 2c_3x_3 + c_5x_2 + c_6x_1 + c_9 \end{bmatrix}^T = \begin{bmatrix} l_x^T\mathbf{c} \\ l_y^T\mathbf{c} \\ l_z^T\mathbf{c} \end{bmatrix} \tag{7}$$

From Eq. (7), the following is

$$\left\|\nabla f(\mathbf{c}, \mathbf{x}^i)\right\|^2 = \nabla f(\mathbf{c}, \mathbf{x}^i)^T \nabla f(\mathbf{c}, \mathbf{x}^i) = \begin{bmatrix} \mathbf{c}^T l_x & \mathbf{c}^T l_y & \mathbf{c}^T l_z \end{bmatrix} \begin{bmatrix} l_x^T\mathbf{c} \\ l_y^T\mathbf{c} \\ l_z^T\mathbf{c} \end{bmatrix} = \mathbf{c}^T N\mathbf{c} = 1$$

Equation (5) is reformulated as the generalized Eigen decomposition $M(\mathbf{x}^i)\mathbf{c} = \lambda\mathbf{c}$ and $\mathbf{c}^T N(\mathbf{x}^i)\mathbf{c} = 1$.

By setting $c_1^2 + c_2^2 + \cdots + c_{10}^2 = 1 \rightarrow \mathbf{c}\mathbf{c}^T = 1 \rightarrow \mathbf{c}\mathbf{c}^T N(\mathbf{x}^i)\mathbf{c} = \mathbf{c} \rightarrow N(\mathbf{x}^i)\mathbf{c} = \mathbf{c}$.

Therefore,

$$\begin{aligned} &M(\mathbf{x}^i)\mathbf{c} = \lambda\mathbf{c} \rightarrow M(\mathbf{x}^i)\mathbf{c} = \lambda N(\mathbf{x}^i)\mathbf{c} \\ &\quad \rightarrow \left(M(\mathbf{x}^i) - \lambda N(\mathbf{x}^i)\right)\mathbf{c} = 0 \Leftrightarrow \det\left(M(\mathbf{x}^i) - \lambda N(\mathbf{x}^i)\right) = 0 \ \textit{with}\ \mathbf{c} \neq 0 \end{aligned}$$

where λ is eigenvalue with eigenvector $\mathbf{c}$ of $M(\mathbf{x}^i)$.

The eigenvector with the smallest eigenvalue provides the best-fit parameter vector.

[Step 3] To fix the axis of the cylindrical shell to lie along the axis of minor principal curvature, the magnitude and the direction of the curvature of the quadratic surface are needed. The eigenvectors of the shape operator describe the principal are directions of the surface and its eigenvalues describe the curvature in those directions. This can be calculated for a point, **x**, on the surface by taking the Eigenvalues and Eigenvectors of:

$$\left(I - N(\mathbf{x})N(\mathbf{x})^T\right)\nabla N(\mathbf{x})$$

where $N(\mathbf{x})$ denotes the normal vector of the quadratic surface. It is calculated by differentiating and normalizing the implicit surface: $N(\mathbf{x}) = \frac{\nabla f(\mathbf{c},\mathbf{x})}{\|\nabla f(\mathbf{c},\mathbf{x})\|}$.

[Step 4] Project each neighborhood onto a plane orthogonal to the axis calculated in step 3. Fit a circle to points in plane.

[Step 5] Linear search for a gap between inner shell and outer shell of the cylindrical shell in order to let robot hand fit in

- The gap contains no points
- The radius of the inner cylinder is less than the diameter of the robot hand.

4 Experimental Results

Physical size of the gripper is 40 × 1 0 mm, Figs. 4, 5, 6, 7 and 8 show localization and grasping results of the proposed method using 5 objects such as spire bottles, cleaning

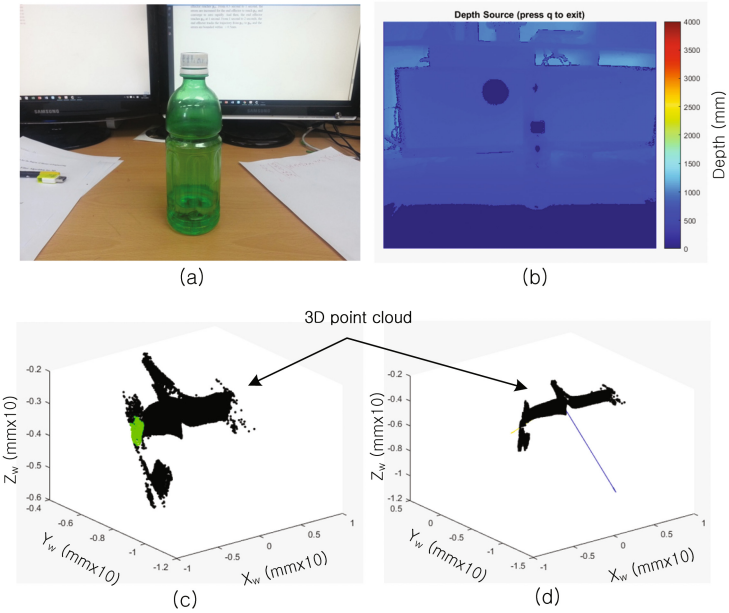


Fig. 4. Localization and grasping result for spire bottles

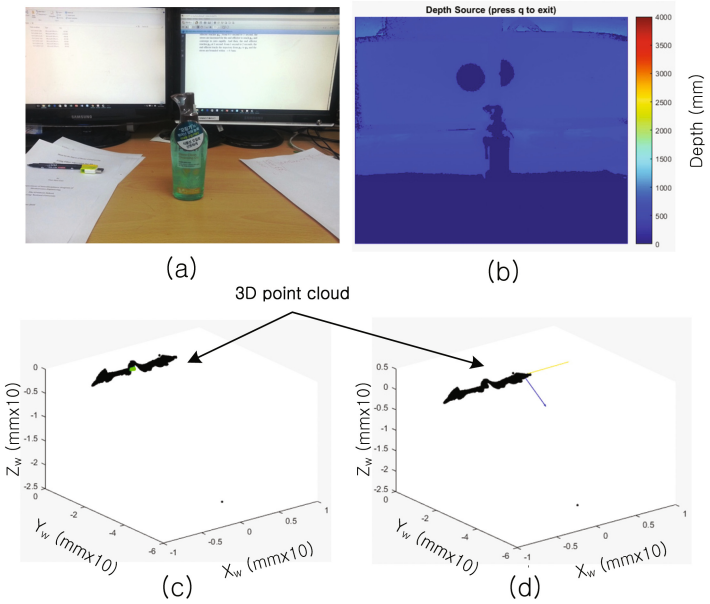


Fig. 5. Localization and grasping result for cleaning oil

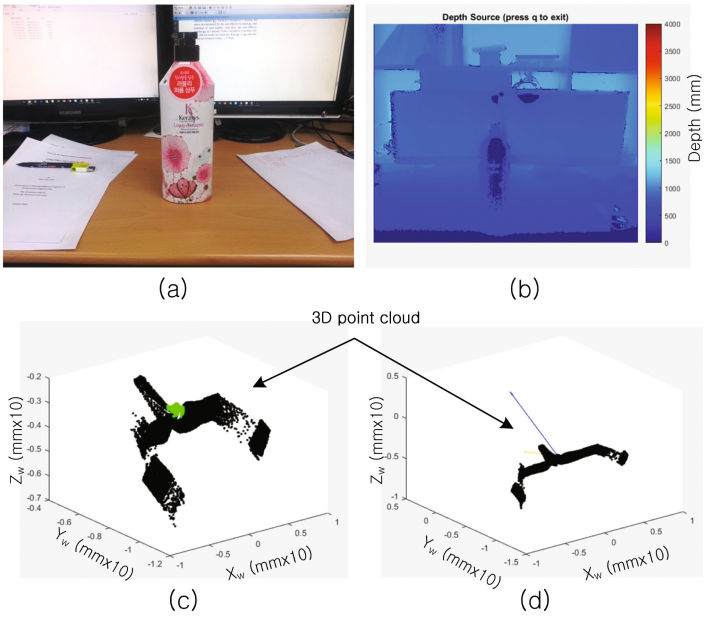


Fig. 6. Localization and grasping result for shampo bottles

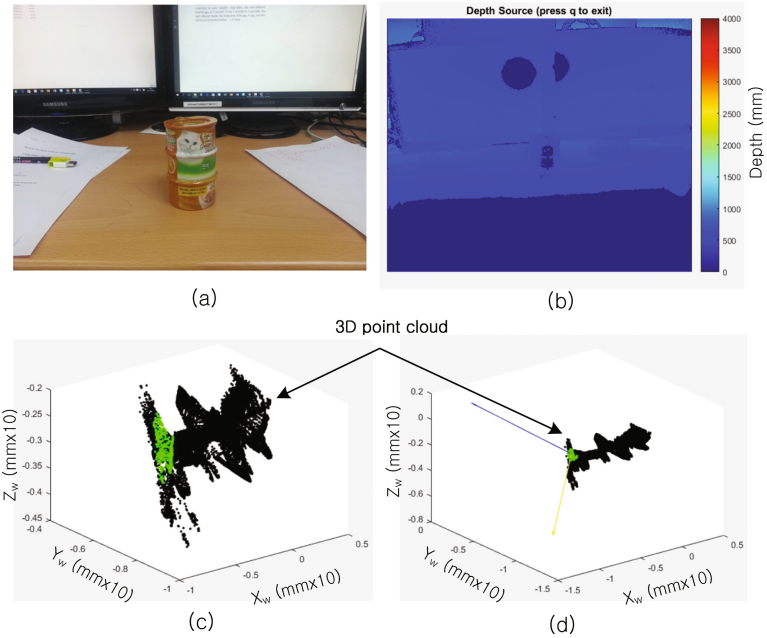


Fig. 7. Localization and grasping result for food pack

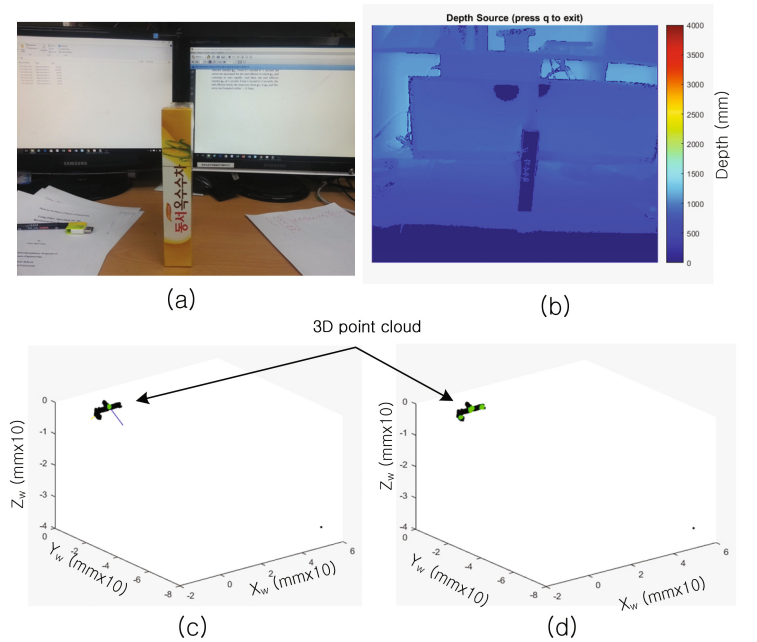


Fig. 8. Localization and grasping result for bisquy

Table 1. Rectangle representation results

Object	Grasped on 1st try	Grasped on 2nd try	Grasped on 3rd try	Remarks
Spire bottles	2/5	3/5	3/5	Transparent
Cleaning oil	2/5	4/5	4/5	Transparent
Food pack	3/5	4/5	5/5	Non-transparent
Bisquy box	4/5	5/5	5/5	Non-transparent
Shampoo bottle	4/5	5/5	5/5	Non-transparent
Average (%)	60%	84%	88%	

oil, shampoo bottles, food pack and bisquy box. In each result, RGB image (a), depth Image (b), localization (c) and grasping direction shown as arrow (d) are shown.

Objects were placed such that a significant number of points on the handle were visible to the Kinect range sensor. Out of the 5 grasp trials for each object, Table 1 shows the number of successful grasps performed on the first try, the second try, and the third try. The experimental results show that the proposed algorithm successfully detects an object on the first try approximately 60%, and 88% by the third try, it had nearly perfect grasp success for no transparent object.

5 Conclusion

This paper proposed a new approach to recognition and grasping objects from 3D environment by combining depth and color stereo image in the mobile picking robot system. An image processing system including Kinect camera sensor was described. Recognizing of the object is obtained by using Kinect camera sensor based on Euclidean distance. For grasping object, an algorithm for localizing handle-like grasp affordances is proposed. The experimental results showed that the proposed algorithm successfully detected an object with accuracy 60% for the first try and 88% for the third try and found the grasping points with accuracy 100% for no transparent object.

Acknowledgments. This work was supported by the Materials and Components Technology Development Program of MOTIE/KEIT. [10063273, Development of Picking Tool for Logistic Robots to Automate Picking Process of Atypical Parcels].

References

1. Chen, W.T., Liu, W.C., Chen, M.S.: Adaptive color feature extraction based on image color distributions. IEEE Trans. Image Process. **19**(8), 2005–2016 (2010)
2. Canny, J.F.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
3. Verma, B., Kulkarni, S.: Texture feature extraction and classification. In: International Conference on Computer Analysis of Images and Patterns, pp. 228–235 (2001)

4. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
5. Nguyen, T.H., Jeong, S.K., Kim, H.K., Kim, S.B.: A method for localizing and grasping objects in a picking robot system using kinect camera. In: Proceedings of 2016 International Symposium on Advanced Mechanical and Power Engineering, ISAMPE, pp. 178–180 (2016)
6. Papazov, C., Haddadin, S., Parusel, S., Krieger, K., Burschka, D.: Rigid 3D geometry matching for grasping of known objects in cluttered scenes. Int. J. Robot. Res. **31**(4), 538–553 (2012)
7. Bley, F., Schmirgel, V., Kraiss, K.F.: Mobile manipulation based on generic object knowledge. In: Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (2006)
8. Choi, C., Taguchi, Y., Tuzel, O., Liu, M.-Y., Ramalingam, S.: Votingbased pose estimation for robotic assembly using a 3D sensor. In: Proceedings of IEEE International Conference on Robotics and Automation (2012)