An Object Classification Framework Based on Unmeasurable Area Patterns Found in 3D Range Images

Koichiro Matsumoto and Kimitoshi Yamazaki, *Member, IEEE*

***Abstract*— This paper describes an object detection framework. Depth images obtained from 3D range camera are used, object detection with classification into three types, which are non-transparent, partly-transparent, and transparent, are performed. We focus on image region where measurement data does not obtained, and analyze the reason how such region is produced. It enables us to reduce uncertain region of an input depth image and to provide information with viewpoint changing to obtain more advanced object information. Using the proposed framework, we implemented an application to classify above three types of objects. Non-transparent objects and partly-transparent objects were classified from a single depth image, and multi-view measurements were used to reduce uncertain data and to narrow down the existing area of transparent objects.**

I. Introduction

Robots working in real environment should have abilities to recognize various existing objects. Therefore, researches about object recognition for intelligent robot have been studied. One feasible approach to perform reliable recognition is to obtain 3D information using sensors. For instance, commonly -used sensors providing point cloud are stereo camera [1], Laser RangeFinder (LRF) [2], and their combination. Recently, 3D range camera becomes popular. It is a sensor using active light source, and produces a depth image even if measuring textureless region. This characteristic enables us to recognize distance information directory.

The purpose of this study is to estimate existence region of objects placed in front of a sensor. Target object can be classified three types: non-transparent object, partly -transparent object, and transparent object. Such function will be useful for intelligent robots working on everyday task: fetch and carry task, lost object search, obstacle avoidance etc.

A sensor we assume is 3D range camera, Microsoft Kinect [3]. It is one of optical sensors using active light source (random dot pattern). Although the sensor provides rich 3D information, it's ranging principle causes the measurement failure against reflective and transparent object because light-receiving device cannot obtain proper reflected light from their surface. Fig.1 shows an example. There are two objects: a non-transparent wooden block and a transparent plastic bottle. About the plastic bottle, we lost almost distance data. Similar problem also occurs in case of stereo camera and LRF.

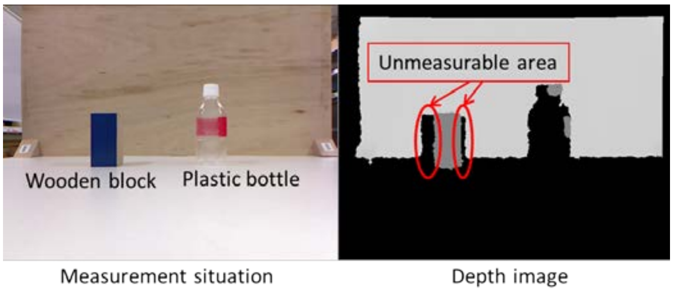


Fig.1 An example of depth image captured from Microsoft Kinect, 3D range image sensor.

On the other hand, there are some parts where we also cannot obtain distance data. Between the wooden block and a background wall, there is unmeasurable region. However, different from the case of transparent object, we can predict these areas by considering following three conditions: a distance between a sensor and a target object, a distance between the target object and a background, and an angle between optical axis and the target object.

Using this characteristic, we construct a framework to detect and classify objects. First, measurement database for a non- transparent object is prepared in advance. As mentioned above, unmeasurable areas are found both side of the object. So the size of area is registered combining with distances and an angle mentioned above. The measurement database includes the size of unmeasurable area with various distances and angles. The database is used to detect and judge non-transparent object and partly-transparent object from a single depth image. On the other hand, for transparent object, we apply multi-view measurement strategy. Using traditional technique of occupancy grid map, an area where transparent object exists can be narrowed down.

The remainder of the paper is organized as follows: Section II introduces related work. Section III explains issues and our approach. Sections IV, V and VI explain our proposed method. Section VII presents experimental results, and conclusions are presented in Section VIII.

II. Related work

Object detection using visual information has been one of important issues for intelligent robots. Many studies proposed detection methods and applied them to robot applications [4] [5] [6] [7]. In almost all of the studies, it was assumed that target objects have rich texture on their surface; thus, they only coped with non-transparent objects. In such case, partly-transparent objects can also be targeted because it is possible to detect them by image features extracted from non-transparent and textured region where are a part of the target object. Well-known image features, e.g. SIFT [8] and SURF [9], and others [10] are useful for them. On the other hand, transparent part was ignored in their work.

Koichiror Matsumoto and Kimitoshi Yamazaki are with the Mechanical Systems Engineering, Faculty of Engineering, Shinshu University, Nagano, Japan. Phone: +81-26-269-5155; e-mail: kyamazaki@shinshu-u.ac.jp.

Texture-based image features have been proposed. Filter bank is one of the major approaches [11] [12]. Texton was based on a set of results of convolution integral [13]. They can describe texture patterns without consideration of geometric repeatability of texture patterns. It has possibility to be applicable to recognize transparent or reflective surfaces; thus, applications using such filter bank technique have been proposed to recognize objects having reflective surface [14]. Because these approaches were tuned up to classify target objects into known categories, it cannot be applied to robotic application that needs information of object shape and pose.

Several researches explicitly targeted transparent object. Lysenkov et al. [11] proposed model fitting of transparent object. A 3D geometrical model was given in advance, and its 6D pose was estimated using RGB-D image. The method used unmeasurable region in a positive manner, and it is similar to our approach; however, it only targets known transparent object. We consider unknown environment where includes both transparent and non-transparent object, and recognize their existence without shape model.

III. Issues and Approach

A. Issues on 3D measurement by commonly-used sensors

Considering environment recognition by autonomous robot, it is desirable that we must install appropriate sensors in the robot. If we aim to impose manipulation task on the robot, the sensors should have a measurement ability that provides 3D information.

Owing to past studies and developments, we have various choices: stereo camera, LRF, and 3D range image sensor. Although they are easily obtainable in recent years, there is a common problem; their measurement principles did not accept to take distance information from transparent and reflective objects.

B. The things we can obtain from a region where no measurement data

Fig. 1 shows a measurement example. In the left picture, two objects, a non-transparent cube and a transparent plastic bottle, were placed in front of a wooden board. Microsoft Kinect 3D range image sensor was used. The right picture shows a captured depth image. Darker pixels mean closer distance to the camera, but black pixels mean that there was no measurement response. As mentioned in above subsection, it is well-known that such unable measurement is caused of transparent object. On the other hand, we also find another region having no distance data. The latter arises from interspace between forward and backward object.

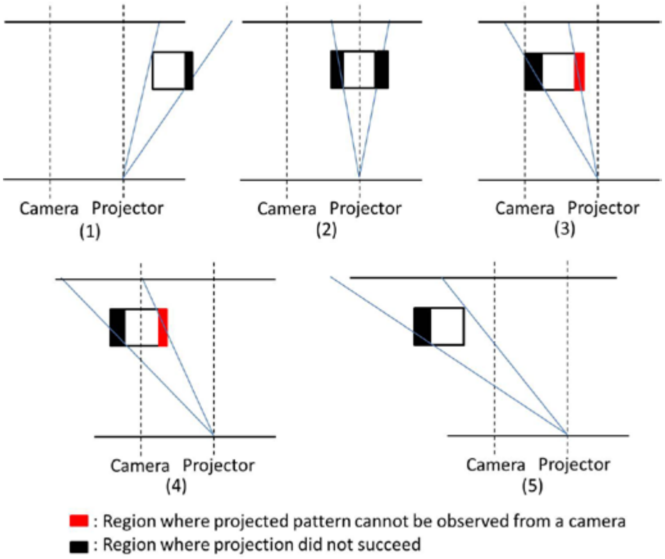

Camera Projector
(1)
Camera Projector
(2)
Camera Projector
(3)
Camera Projector
(4)
Camera Projector
(5)
🟥 : Region where projected pattern cannot be observed from a camera
■ : Region where projection did not succeed

Fig. 2 Five projection patterns we assume

Fig.2 explains several cases that we cannot obtain measurement data at sides of an object. Here, Microsoft Kinect sensor using random dot pattern [12] is assumed. The sensor projects light source from a projector and the projected pattern is captured by a camera adjacent to the projector. This formulation produces some situations that make the measurement impossible. Two main reasons are: the projected pattern cannot be observed from a camera, and the projection in itself fails because of occlusion.

The thing we focus on is that such no measurement area can be predicted if we obtain correct measurement from a target object and its background. For instance, "unmeasurable area" shown in Fig.1 is the area explained above. We can predict the size of the area if we find that front object is non-transparent. The prediction method is explained in Section IV.

IV. A Framework Using Unmeasurable Area Patterns

A. Formulation outline

We explain our framework according to probabilistic formulation. Let $P(\boldsymbol{x})$ be a probabilistic distribution of an object. $\boldsymbol{x}$ is a position of the object. The purpose of object detection will be satisfied by searching regions where high probability is obtained. However, if we cope with transparent objects and partly-transparent objects instead of merely considering non-transparent objects, it is difficult to identify their existence from a depth image.

Our framework consists of two phases; (i) the existence of non-transparent or partly-transparent object is estimated using a depth image captured from a fixed viewpoint, and (ii) the

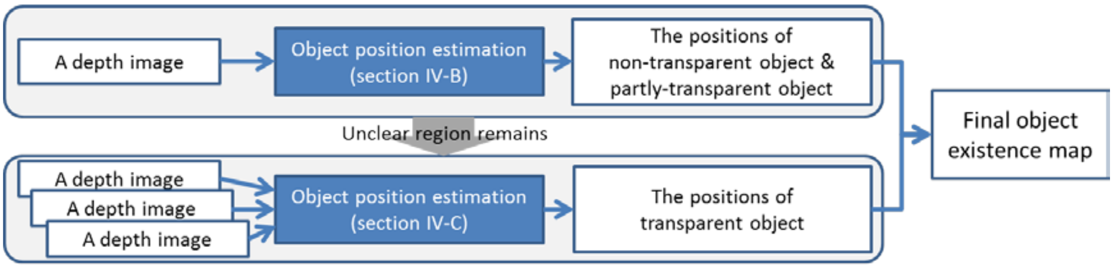

A depth image
Object position estimation (section IV-B)
The positions of non-transparent object & partly-transparent object
Unclear region remains
A depth image
A depth image
A depth image
Object position estimation (section IV-C)
The positions of transparent object
Final object existence map

Fig. 3 The proposed framework

existence of transparent object is estimated from multi-view range images. Fig.3 shows an abstract of the proposed framework.

B. *Existing probability calculation for non-transparent and partly-transparent object*

Let $\boldsymbol{z}$ be one measurement, and $P(\boldsymbol{x}|\boldsymbol{z})$ be a posterior probability using the measurement. If we target only non-transparent objects, conventional approach can be used because a certain reaction is obtained from non-transparent surface of object. However, transparent part does not return almost measurement data. On the other hand, we can also find unmeasurable region from boundary of non-transparent surfaces as shown in Fig.1. From this fact, we translate $P(\boldsymbol{x}|\boldsymbol{z})$ into another representation as follows:

$$P(\boldsymbol{x}|\boldsymbol{z}) \equiv P(\boldsymbol{e}|\boldsymbol{z})P(\boldsymbol{g}|\boldsymbol{e},\boldsymbol{z}), \quad (1)$$

where $\boldsymbol{e}$ denotes an image edge extracted from a side boundary of an object. $\boldsymbol{g}$ denotes an image region where measurement data are not obtained at the side boundary of the object. We call them “unmeasurable region” in the remains of this paper. The equation (1) means that the existing probability of a non-transparent object is calculated from a boundary edge and an unmeasurable region at the side of the edge. If the value of $\boldsymbol{g}$ measured between a target object and its background is separated from assumed value, the object may partly -transparent. On the other hand, if transparent object, we cannot obtain almost measurement data from the object. That is, this approach means that we first pick up measurable objects and estimate whether they are non-transparent and partly-transparent. Otherwise, next step for the position estimation of transparent objects is started.

As another possibility that we have a measurement result similar to the case of partly-transparent object, measurement defect may occur when the angle of object surfaces from optical axis is small. We cannot distinguish the two possibilities from one depth image. To delete such bilateral hypothesis, the following multi-view measurements are also utilized.

C *Existing probability calculation for transparent object*

Transparent object returns little measurement data; thus, using just one measurement leaves large uncertainty of the existing probability of transparent object. However, we can reduce the uncertainty by means of multi-view measurements. It enables to narrow down existing area possibility of transparent object.

The estimation principle is inspired from occupancy grid map [7]. The basic equation is as follows:

$$P(\boldsymbol{y}|\boldsymbol{z}_{1:i},\boldsymbol{a}_{1:i}) \equiv \prod_{k} P(\boldsymbol{y}_k|\,\boldsymbol{z}_{1:i},\boldsymbol{a}_{1:i})\,, \quad (2)$$

where $\boldsymbol{y}$ is a position of a transparent object, and $\boldsymbol{z}_{1:i}$ is a list of data of i times of measurement. $\boldsymbol{a}_{1:i}$ is a list of sensor poses corresponding to each of the measurements. Left side of the equation (2) shows a probabilistic distribution showing position candidates of the transparent object. However, it is difficult to obtain the distribution directly. For this reason, it is approximated to right equation. This shows that a target space divided into equally-divided small cells, and the probabilistic calculation is performed at individual cell.

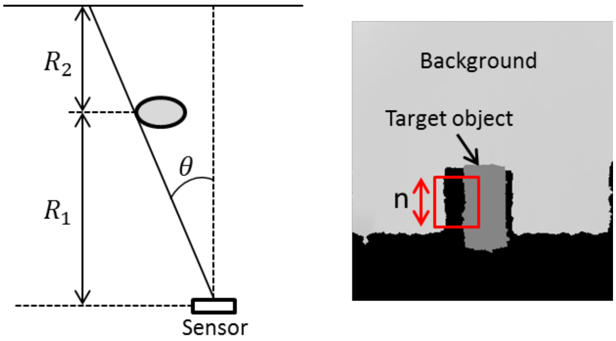


Fig. 5 Parameters for gap region estimation.

This result using one measurement is combined with other measurements captured from other sensor viewpoints. We create occupancy grid map, and calculate existing probability at each grid.

V. Detection of Non-Transparent and Partly-Transparent Object

This section describes how to implement the proposed framework. What we have to do in advance is to prepare learning data about unmeasurable region depending on two distances and one angle. In detection process, region growing algorithm is first applied, and a boundary edge and neighboring gap region are extracted. Comparing the region and learning data, probability as non-transparent object is calculated.

A. *Learning data preparation*

As mentioned in section III-B, we can predict the size of unmeasurable region if it exists between two non-transparent objects. From this fact, we prepared a learning dataset storing unmeasurable region sizes with three parameters. The left figure in Fig. 5 shows the parameters and their relationships. R_1 is a distance between camera and a non-transparent object. R_2 is a distance between the object and non-transparent background. θ is an angle between optical axis and a line that passes optical center and one side edge of the object.

In the learning data correction, these three parameters were variously changed, and the size of unmeasurable region was registered. The region was calculated at the region with n pixels height as shown in the right picture in Fig.5.

Fig.6 shows a part of the results. Horizontal axis and vertical axis indicate R_2 and the size of unmeasurable region, respectively. Each of the red or green lines shows the result of different R_1 settings. As shown in the graph, larger R_2 produces larger unmeasurable region at left of the object, however its amount of change is not linear. Thus, we use the learning data without line fitting but directly compare input value with the data by means of nearest neighbor search.

On the other hand, there was almost same area size at right side of the object despite the change of R_2. The reason is explained in the last paragraph at Section IV-B. That is, measurement defect occurred because the angle between right surface of the object and optical axis was small. In this case, only left side of region is used to calculate Eq. (1).

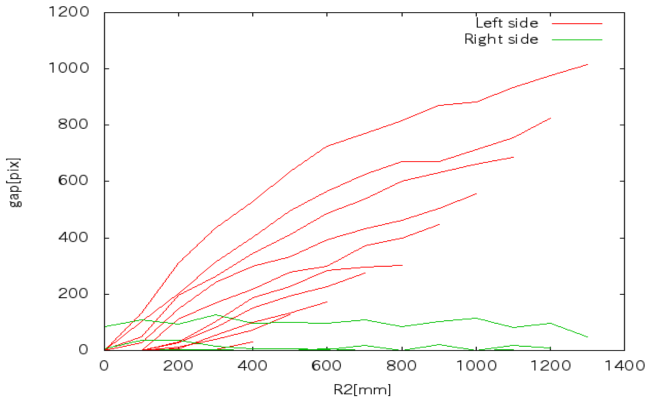


Fig. 6 Relationships between R2 and the size of unmeasurable region. Red line shows right side, and green shows left side.

B. Region growing algorithm

Region growing algorithm is applied to detect locally independent regions in a depth image. In the algorithm, a starting point $\boldsymbol{p}_0$ is first selected. From this point, neighbor points whose similarities are greater than a predefined threshold are selected as homologous points. Our similarity measure is calculated by angle difference of neighbor normal vectors.

For this process, normal vectors should be pre-calculated for all points. Therefore, a normal vector of an interest pixel $\boldsymbol{p}$ is calculated from the 3D position of $\boldsymbol{p}$ and its neighbors.

In the region growing algorithm, a region is extended if the following rules are satisfied:

$$|\,\mathrm{d(i,j)} - d(i+n, j+m)| < d_{threshold}$$

$$cos^{-1}\big(n(i,j) \cdot \mathrm{n(i+n, j+m)}\big) < \theta_{threshold},$$

where $\boldsymbol{n}_1$ is a normal vector of $\boldsymbol{p}$, and $\boldsymbol{n}_2$ is a neighbor normal vector. $(\,\cdot\,)$ indicates the inner product, and $C_{threshold}$ is a predefined threshold.

C. Calculation of $P(\boldsymbol{e}|\boldsymbol{z})$

Region growing divides a depth image into several spatial clusters. Fig. 7 shows an example; different color shows different cluster. Using these clusters, we can start to calculate $P(\boldsymbol{e}|\boldsymbol{z})$ where $\boldsymbol{e}$ denotes a 3D position of an image edge, and $\boldsymbol{z}$ denotes one sensor measurement.

To calculate $P(\boldsymbol{e}|\boldsymbol{z})$**,** each cluster is registered as a group of pixel coordinates, which are uniformed along horizontal axis. Why we make such arrangement is because our assuming Kinect sensor has alley light and optical receiver with horizontal relation. On the other hand, the 2D center position of the cluster is calculated. Next, minimum pixel at each horizontal coordinates is extracted, and the pixel is regarded as belonging to left edges. As a result, the probabilistic distribution $P(\boldsymbol{e}|\boldsymbol{z})$ has two values: 1.0 if a pixel is on image edge, and otherwise 0.0.

D. Calculation of $P(\boldsymbol{g}|\boldsymbol{e},\boldsymbol{z})$

As a result of above processing, we obtain one selected pixel group about left edge. Let v_c be a vertical coordinate of a center position of a cluster, and u_{le} be horizontal pixel coordinate of left edge. Raster scan is performed within $u_{le} - m/2 \le u \le u_{le} + m/2$ and $v_c - n/2 \le v \le v_c + n/2$, where *m* is experimentally defined as same as *n* described at section V-A. Through this raster scan, pixels whose depth value is not obtained are counted, and the result is regarded as unmeasurable image region g.

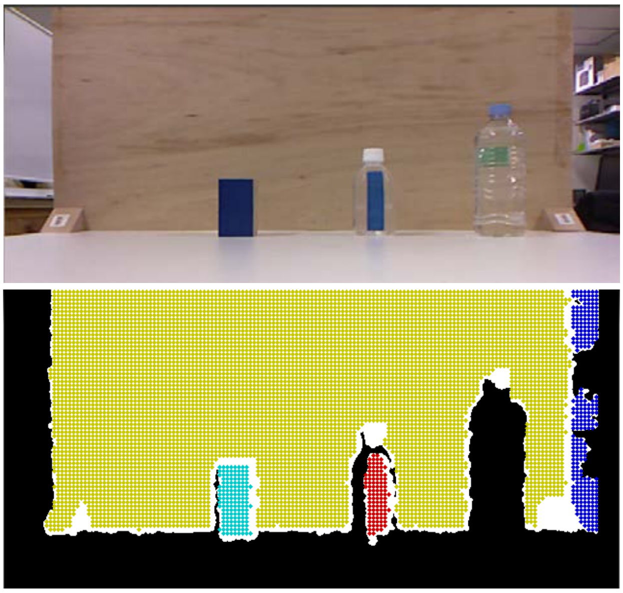


Fig. 7 region growing result.

To calculate $P(\boldsymbol{g}|\boldsymbol{e},\boldsymbol{z})$, we use previous knowledge about the size of unmeasurable image region. That is, the calculated g is compared with the knowledge to describe the likelihood as non-transparent object. The procedure is as follows: First, a target pixel $\boldsymbol{u} = (u_{le}, v_{le})$ is selected, and its 3D position is calculated. That is,

$$x_{le} = z_{le}\frac{u_{le} + u_0}{f}, \qquad y_{le} = z_{le}\frac{v_{le} + v_0}{f}, \qquad z_{le} = d,$$

where *f* denotes focal length, and *d* denotes depth value. $u_0$and v_0 are the center coordinates of the image. Next, $R_1, R_2\ and\ \theta$ are calculated using x_{le} and z_{le}. Nearest neighbor of these parameters are sought from previous knowledge. In our implementation the searching is effectively performed by using *kd*-tree that can be generated from learning data in advance. Final calculation of $P(\boldsymbol{g}|\boldsymbol{e},\boldsymbol{z})$ is obtained as distance between sought data and present unmeasurable image region. The function is defined according to normal distribution:

$$N(\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}exp\left(-\frac{(g_* - \mu)^2}{2\sigma^2}\right),$$

where μ is the size of unmeasurable image region corresponding to $R_1, R_2\ and\ \theta$. σ is experimentally defined. This equation means if expected g is obtained, the value $P(\boldsymbol{g}|\boldsymbol{e},\boldsymbol{z})$ becomes high at $\boldsymbol{x}_c$, where the 3D center position of a focusing cluster. When the probability is high, we can consider that a non-transparent object exists around $\boldsymbol{x}_c$. Otherwise, there is partly-transparent object.

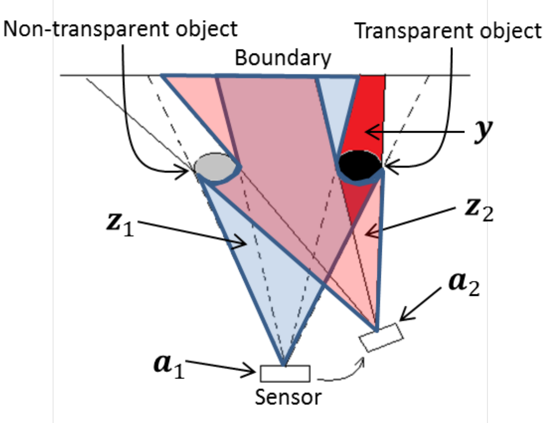


Fig. 8 Narrowing down the existing region of transparent object by two measurements. Red region has high probability.

VI. DETECTION OF TRANSPARENT OBJECT

In case of transparent object, almost no measurement data obtained. This means that one measurement data does not provide sufficient information to know the shape of transparent object. For this reason, we take an approach to multi-view measurements. The approach is inspired from volume intersection method [13], which prunes away an estimated volume by using a new measurement data incrementally. More measurement data from various views will make clear the shape of the object.

Fig.8 shows the concept. If we find a non-transparent object, there is no doubt of its position. On the other hand, we cannot recognize the position of a transparent object even the front side of it. Using multi-view measurement, such uncertainty can be reduced; thus, based on the equation (2) that is inspired from occupancy grid map, we narrow down possible volume of the transparent object.

As mentioned at section IV-B, we have another possibility at unmeasurable region in one depth image: partly-transparent object or measurement defect at a surface whose inclination is near to optical axis. Multi-view measurements are also useful to divide them. That is, after moving viewpoint, if a cluster is found instead of past unmeasurable region, there may be a non-transparent surface. Otherwise, the probability of a transparent surface remains high.

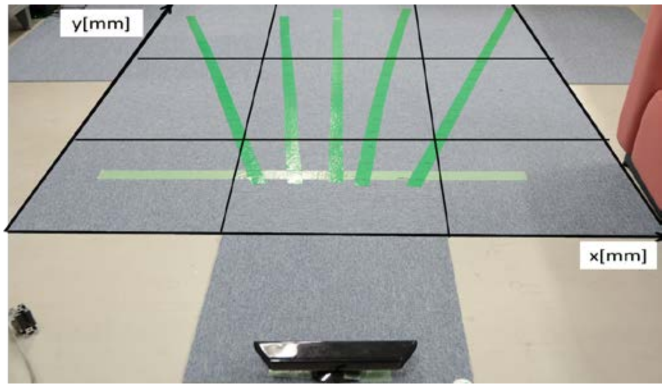


Fig.9 An experimental environment

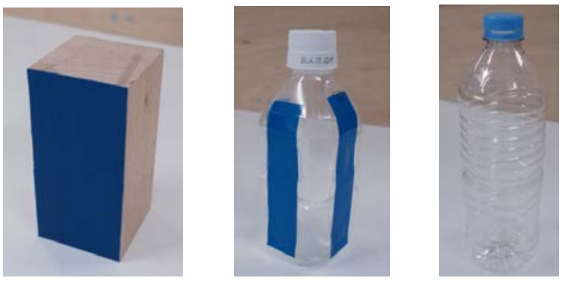


Fig.10 Target objects

VII. EXPERIMENTS

A. Experiment setting

Fig. 9 shows an experimental environment. A square shape field was divided into nine grids, each of the grid was sized 420[mm] by 420[mm]. Three objects, a wooden cube and two plastic bottles as show in Fig. 10, were prepared. A Microsoft Kinect sensor was placed on the front of the field. Green horizontal line taped in Fig. 9 shows 700 [mm] distance from the Kinect sensor.

One object was selected and placed on the field, and the wooden board was also placed on the backward of the target object. The position of the object and the wooden board were changed from lower left to upper right grid, and classification rate was calculated at each grid.

The implemented structure of the proposed framework was shown in Fig. 11. The result of clustering is used to first judgment of object classification. The output is a map that

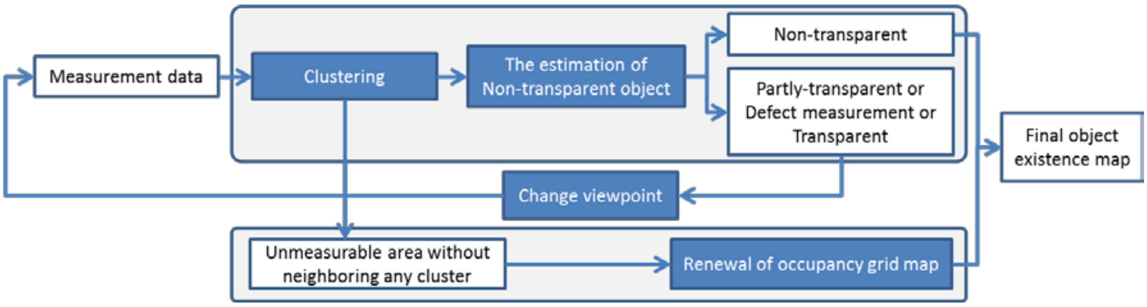
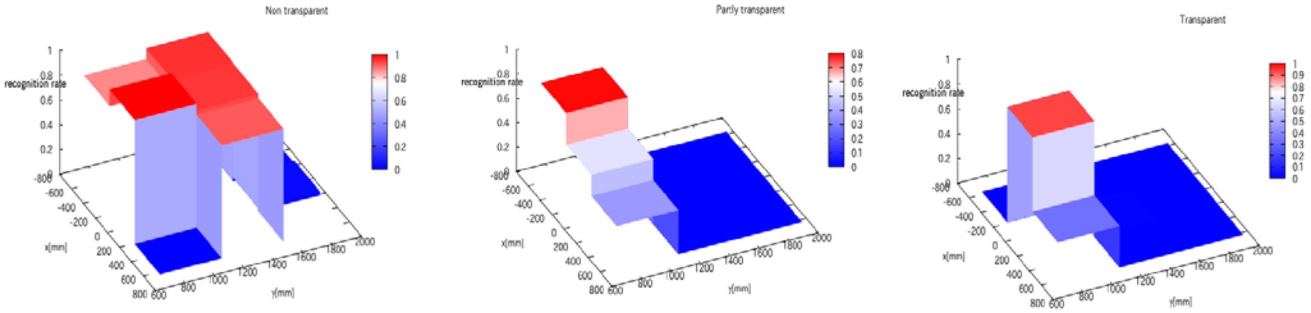

Measurement data
Clustering
The estimation of Non-transparent object
Non-transparent
Partly-transparent or Defect measurement or Transparent
Final object existence map
Change viewpoint
Unmeasurable area without neighboring any cluster
Renewal of occupancy grid map

Fig.11 A whole image of object detection procedure



Non transparent object

Partly transparent object

Transparent object

Fig.12 Success rate of classification of three types objects

described probabilistic distribution showing position of objects.

B. Results

A target object was placed on the center of each grid, and 150 frames of depth image were captured. The success rate of object classification was investigated in each grid. The resulted graphs are shown in Fig.12. Axes titled “x [mm]” and “y [mm]” are corresponding to horizontal and depth components in the experimental field, respectively. The vertical axis shows success rate. In the front centering grid, non-transparent object was correctly found with more than 90% success rate. Meanwhile, partly-transparent object was found with about 80% at left front grid, and transparent object were found with more than 90% in the front centering grid. The biased recognition rates were derived from the structure of Kinect sensor because it has a projector in its right side, and an optical receiver in its left.

Fig. 13 shows one example of multi-view measurements and their composition. A transparent plastic bottle was placed on the center of the circular track. The sensor was moved on the circular track with capturing depth images every 30 degrees. Right figure shows a result that shows a result composing several measurement data. White region shows high probability as the position of transparent object. Although some problems remained, we obtained preliminary results that were to show the possibility to estimate the shape of a transparent object.

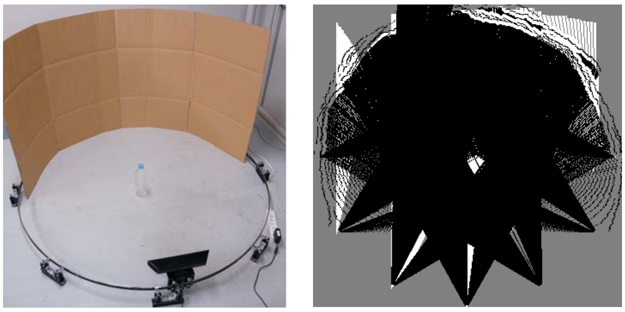


Fig.13 An experimental result by multi-view measurements

VIII. Conclusion

We proposed an object classification framework using 3D range image sensor. Unmeasurable regions were explicitly considered, we succeeded to classify three types of objects: non-transparent object, partly-transparent object, and transparent object. The procedure was divided into two phases: non-transparent and partly-transparent objects are found from one depth image, and then the existence probability of transparent object was narrowed down by using multi-view measurements. We confirmed the proposed method from some experiments.

From these results, we can mention the following three points as future works. The first issue is to improve the recognition accuracy. We still have low recognition rate as showed in Fig.12. This caused by deviation of unmeasurable area between the learning data and experimental data. To improve it, we need to consider the variance of unmeasurable area. The second issue is to extend the framework to cope with situations having multi-objects. For example, superimposed relationship is a challenging issue, and occlusion should also be considered. The third issue is that, we have to continue to develop other evaluation methods to enhance the proposed method. Application to everyday object manipulation by a real robot is our final goal.

Acknowledgment

This work is partly supported by SCOPE No.132104001, Ministry of Internal Affairs and Communications, Japan.

References

1. *Bumble Bee, Point Gray Inc.* [Online] http://www.ptgrey.com/products/stereo.asp.
2. *URG series, HOKUYO Inc.* [Online] http://www.hokuyo-aut.jp/02sensor/07scanner/utm_30lx.html.
3. *Kinect, Microsoft Corporation.* [Online] http://www.xbox.com/ja-JP/kinect/.

4. **J. Kuehnle, A. Verl, Z. Xue, S. Ruehl, J. Zoellner, R. Dillmann,.** *`6d object localization and obstacle detection for collision-free manipulation.* 2009. p. in Proc. International Conference on Advanced Robotics.
5. **K.Kitahama, K.Tsukada, F.Galpin, T.Matsubara and Y.Hirano.** *Vision-based scene representation for 3D interaction of service robots.* 2006. p. in Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems.
6. **S. Lee, H. Moradib, D. Jangc, H. Jangd, E. Kime, P M. Lef,.** *Toward Human-Like Real-Time Manipulation: From Perception to Motion Planning.* 2008. pp. Advanced Robotics, Vol. 22, Issue 9, pp. 983 -- 1005.
7. **Thrun, S. and Bücken, A.** *Integrating grid-based and topological maps for mobile robot navigation.* 1996. pp. 944–950, in Proc. of the Thirteenth National Conference on Artificial Intelligence.
8. **Lowe, D. G.** *Distinctive image features from scale-invariant keypoints.* 2004. pp. Int'l Journal of Computer Vision, vol. 60, No. 2, pp. 91-110.
9. **Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool.** *SURF: Speeded Up Robust Features.* 2008. pp. Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359.
10. **Schmid, K. Mikolajczyk and C.** *Scale and Affine Invariant Interest Point Detectors.* 2004. pp. Int'l Journal of Computer Vision vol. 60, No. 1, pp.63 -- 86.
11. **Rabaud, I. Lysenkov and V.** *Pose Estimatin of Rigid Transparent Objects in Transparent Clutter.* 2012. in Proc. of IEEE International Conference on Robotics Automation.
12. **M.R. Andersen, T. Jensen, P. Lisouski, A.K. Mortensen, M.K. Hansen, T. Gregersen and P. Ahrendt.** *Kinect Depth Sensor Evaluation for Computer Vision Applications.* s.l. : Danmark:Aarhus University, 2012.
13. **Aggarwal, W. N. Martin and J. K.** *Volumetric description of objects from multiple views.* s.l. : IEEE Transactions on Pattern Analysis and Pattern Recognition, 1987. pp. No.5, vol. 2, 150 - 158.
14. **N. Alt, P Rives and E. Steinbach.** *Reconstruction of transparent objects in unstructured scenes with a depth camera.* 2013. in Proc of International Conference on Image Processing.
15. **G. Csurka, C. Bray, C. Dance, and L. Fan.** *Visual categorization with bags of keypoints.* 2004. pp. in Proc. of ECCV Workshop on Statistical Learning in Computer Vision, pp. 59 -- 74.
16. **M. Galun, E. Sharon, R. Basri and A. Brand.** *Texture Segmentation by Multiscale Aggregation of Filter Responses.* 2003. pp. Proc. of IEEE Int'l. Conf. on Computer Vision, pp. 716-723.
17. **J. Geusebroek, A. Smeulders and J. Weijer.** *Fast Anisotropic Gauss Filtering.* 2003. pp. IEEE Trans. on Image Processing, 12(8):938-943.
18. **Z. Marton, D Pangercic, R. B. Rusu, A. Holzbach and M. Beetz.** *Hierarchical Object Geometric Categorization and Appearance Classification.* 2011. pp. in Proc. of IEEE-RAS Int'l Conf. on Humanoid Robots, pp. 365 -- 370.
19. **D. Pangercic, V. Haltakov and M. Beetz.** *Fast and Robust Object Detection in Household Environments Using Vocabulary Trees with SIFT Descriptors.* 2011. pp. IEEE/RSJ International Conference on Intelligent Robots and Systems, Workshop on Active Semantic Perception and Object Search in the Real World, pp. 25--30.
20. **B. Pitzer, M. Styer, C. Bersch, C. DuHadway, J. Becker.** *Towards perceptual shared autonomy for robotic mobile manipulation.* 2011. pp. in Proc. of Int'l Conf. on Robotics and Automation, pp. 6245 -- 6251.
21. **Malik, T. Leung and J.** *Representing and recognizing the visual appearance of materials using three-dimensional textons.* 2001. pp. International Journal of Computer Vision 43 (1): 29–44.
22. **Schmid, C.** *Constructing models for content-based image retrieval.* 2001. pp. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 2, pp. 39-45.