A 3-Dimensional Object Recognition Method Using SHOT and Relationship of Distances and Angles in Feature Points

Hiroyuki Kudo
Department of Information system Science, Graduate School of Engineering, Soka University
Hachioji-Shi, Tokyo, Japan
e17m5212@soka-u.jp

Kazuo Ikeshiro
Department of Information system Science, Graduate School of Engineering, Soka University
Hachioji-Shi, Tokyo, Japan
ikeshiro@soka.ac.jp

Hiroki Imamura
Department of Information system Science, Graduate School of Engineering, Soka University
Hachioji-Shi, Tokyo, Japan
imamura@soka-u.jp

ABSTRACT

In recent years, a human support robot has been receiving attention. This robot is required to perform various tasks to support humans. Especially the object recognition task, which is important when people request the robot to transport and rearrange objects. Object recognition methods, especially using the 3D sensor are also receiving attention. As conventional object recognition methods using 3-dimensional information, Signature of Histogram of OrienTations (SHOT) is commonly used. SHOT performs highly accurate object recognition since SHOT descriptor is represented by 352 dimensions. However SHOT misrecognizes objects which have the same feature but which are not the same objects and if there is occlusion in the 3-dimensional object. As a solution, I would like to propose the object recognition method with high quality by using the positive part of SHOT.

KEYWORDS

Cognitive system, 3D object, SHOT descriptor, List matching, Human Support Robot

1 INTRODUCTION

In recent years, human support robots have been receiving attention [1] [2].

Especially, objects recognition task is important in case that people request the robots to transport and rearrange an object. We consider that there are five necessary properties to recognize in domestic environment as follows.

(1) Robustness against occlusion
(2) Fast recognition
(3) Pose estimation with high accuracy
(4) Coping with erroneous correspondence
(5) Recognizing objects in a noisy environment

Firstly, the robots need the robust recognition for occlusion because occlusion occurs between different objects in domestic environment. Secondly, the robots need to recognize a target object fast to achieve required tasks fast. Thirdly, the robots need to estimate a pose of a target object with high accuracy to manipulate a target object. Fourthly, the robots need to cope with erroneous correspondence to recognize objects which have the same feature in a local region but which are not the same object. For example, a cube and a rectangular they both have same future points in their vertex, but aspect ratio is totally different.

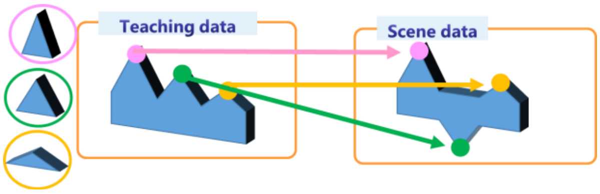


Figure 1. Mismatching in the local regions by using SHOT

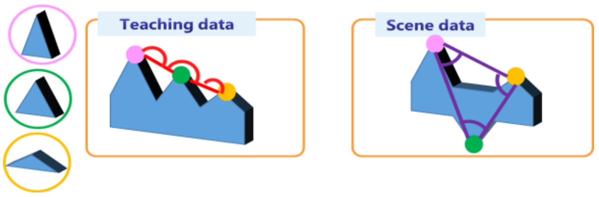


Figure 2. The relationship between the feature points by previous research

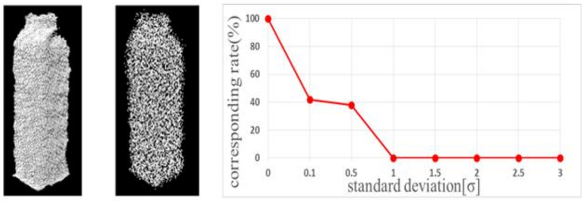


Figure 3. The changing of the correspondence rate

Table 1. Properties of conventional methods and the proposed method

	Robustness against occlusion	Fast recognition	Pose estimation with high accuracy	Coping with erroneous correspondence	Recognizing objects in a noisy environment
SHOT	×	○	○	×	○
Previous research	○	○	○	○	×
Proposed method	○	○	○	○	○

Finally, the robots need to recognize an object which has some noises.

As conventional object recognition methods using 3-dimensional information, Signature of Histogram of OrienTations (SHOT) is commonly used [3] [4].

SHOT focuses on the local region and expresses the relationship between the point of interest and the surrounding points as SHOT descriptor in a histogram. SHOT performs highly accurate object recognition since SHOT descriptor is represented by 352 dimensions. Therefore if there is some noise, value of shot descriptor is hardly interfered. But SHOT misrecognizes objects which have the same feature but which are not the same objects, because SHOT only focuses on Local feature points to match objects as shown Figure 1. Therefore if an object has same features in local, SHOT incorrectly recognizes as same objects.

Therefore, to compensate for the defect of SHOT, our laboratory has developed the previous research for the object recognition by Maehara et al [5]. The previous research used some high curvature points in regions for feature points. Furthermore the previous research generates a list by listing relationships of distances and angles between feature points and matches lists as shown Figure 2. Thereby, the previous research estimates a pose of a target object with high accuracy and copes with erroneous correspondence by using not only the feature points but also relationships between feature points.

However, the previous research does not satisfy recognizing objects in a noisy environment. Figure 3 shows the corresponding rate when we added Gaussian noise according to standard deviation, and shows the result of matching the scene data with noises with its original data. In the Figure 3, we added noises on the data of the pack. As you can see, corresponding rate is gradually decreasing, therefore the previous research is easily interfered by noise. The calculation method of corresponding rate will be shown in the section 2.7.

Table 1 shows properties of these methods. As I mentioned, two of the method do not satisfy all the properties. To satisfy all the properties of recognition, we propose a 3-dimensional object recognition method by using SHOT and relationships of distances and angles in feature points. We use the positive parts of both SHOT and previous research. As our approaches, firstly, to have the robustness against noises, the proposed method uses SHOT in region to extract feature points. SHOT focuses on the local region and expresses the feature amount in the histogram as SHOT descriptor when extracting feature points. For this reason, it is conceivable that they are less likely to interfere with noise since feature points are determined by the values of the histogram. Furthermore, the proposed method generates the list of distances and angles between extracted corresponding points that SHOT descriptors are matched. In addition, the proposed method matches lists which are generated in the model data and scene data.

2 PROPOSED METHOD

2.1 Flow of The Proposed Method

In this section, we describe about an overview of the proposed method based on its processing flow as shown Figure 4.

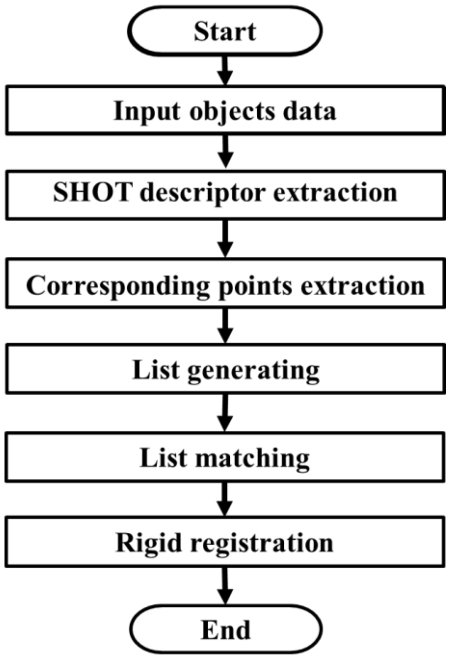
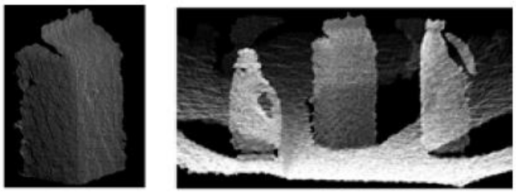


Figure 4. The flow of the proposed method



(a). model data **(b). scene data**

Figure 5. The input data

2.2 Input Object Data

Firstly, the proposed method inputs target objects as a teaching data and a scene data as shown Figure 5.

2.3 SHOT Descriptor Extracting

To extract feature points, the proposed method uses SHOT (Signature of Histogram of Orientations). The surface features of the three-dimensional model can be described with unique and repeatability by using SHOT. It expresses the relationship between the point of interest and its surroundings by histograms. Since SHOT descriptor is expressed in 352 dimensions, SHOT is the method that can extract feature points with high accuracy. In this section, we explain about how to extract a SHOT descriptor. To extract a SHOT descriptor, we use an isotropic spherical grid that encompasses partitions along the radial, azimuth and elevation axes, as sketched in Figure.1. Since each volume of the grid encodes a very descriptive entity represented by the local histogram, SHOT can use a coarse partitioning of the spatial grid and hence a small cardinality of the descriptor. In SHOT, the angle of the dot product of the normal vector of the reference point and the normal vector of the point of each grit is represented by histograms.

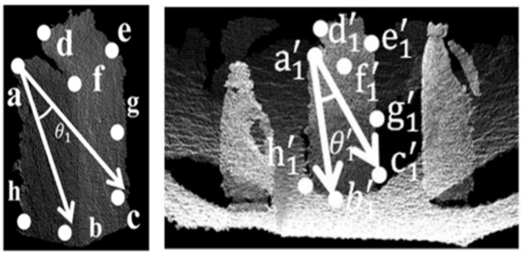
2.4 Corresponding Points Extraction

To match the SHOT descriptor, the proposed method matches each scene feature against all model features. Furthermore, the proposed method computes the ratio between the nearest neighbor and second best. If the ratio is below a threshold, a corresponding is established between the scene feature and its closest model feature.

2.5 List Generating

In the list generating process, the proposed method generates the list of distances and angles between extracted corresponding points as relationships of these points.

At this time, the proposed method extracts the combination of three points as much as



(a). In the model **(b). In the scene**

Figure 6. Overviews of matched points

Table 2. The LIST in the model

number	corresponding point①	corresponding point②	corresponding point③	distance between point ① and ②	distance between point ① and ③	angle
1	a	b	c	l_{ab}	l_{ac}	θ_1
2	a	b	d	l_{ab}	l_{ad}	θ_2
3	a	b	e	l_{ab}	l_{ae}	θ_3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
56	f	g	h	l_{fg}	l_{fh}	θ_{56}

Table 3. The LIST in the scene

number	corresponding point①	corresponding point②	corresponding point③	distance between point ① and ②	distance between point ① and ③	angle
1	a'	b'	c'	$l_{a'b'}$	$l_{a'c'}$	θ'_1
2	a'	b'	d'	$l_{a'b'}$	$l_{a'd'}$	θ'_2
3	a'	b'	e'	$l_{a'b'}$	$l_{a'e'}$	θ'_3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
56	f'	g'	h'	$l_{f'g'}$	$l_{f'h'}$	θ'_{56}

possible in the corresponding points as shown Figure 6, Table 2 and Table 3. In recognizing the object, the proposed method is able to be less mismatching of list elements since the lists we generated are including all the corresponding points by matching SHOT descriptor, furthermore if there is a point that is mismatched by SHOT matching, the proposed method is able to exclude that point from matching targets due to the difference in the three points relationship.

2.6 List Matching

In the list matching process, the proposed method matches the list of the model data and the list of the scene data. As shown in Figure 6, Table 2 and Table 3, a list has distances and an angle as element. Then, the proposed method matches between list number 1 of the model and all the lists of the scene data. Furthermore, in the proposed method, the list with the smallest difference of between the sum of distance between point① and point②, distance between point① and point③ and angle, which is less than the threshold is subjected to matching.

2.7 Rigid Registration

To recognize the target object in the scene data, the proposed method applies the rigid registration to the teaching data. Firstly, the proposed method fits the teaching data to the matched object in the scene by calculating the optimum rotation matrix R and the translation vector t from associated corresponding points. Secondly, the proposed method calculates a corresponding rate M between a teaching data and the matched object by using

$$score = \sum_{i=1}^{N} f\left(\min\{dist_{ij} | 1 \le j \le L\}\right),$$

$$f(x) = \begin{cases} 1 \ (x \le \boldsymbol{th}_c) \\ 0 \ (x > \boldsymbol{th}_c), \end{cases} \qquad (1)$$

$$dist_{ij} = \|\boldsymbol{p}_i - \boldsymbol{q}_j\|,$$

$$M = \frac{score}{L} \cdot 100 \quad . \qquad (2)$$

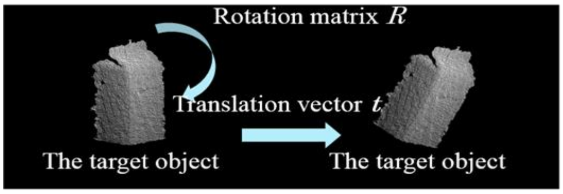


Figure 7. Illustration of the rigid registration

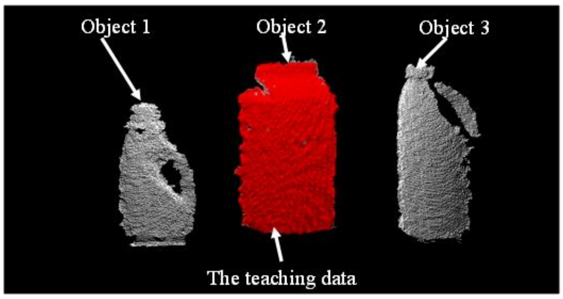


Figure 8. The result of the rigid registration

Where, N is the number of points of the teaching data. L is the number of points of the object in the scene. $\boldsymbol{p}_i$ is matched point of the fitted teaching data. $\boldsymbol{q}_j$ is matched point of the matched object in the scene. The proposed method counts a number of $\boldsymbol{p}_i$ which are within a threshold $\boldsymbol{th}_c$ which is 1 [mm] of $\boldsymbol{q}_j$ by the equation (1) as a score. And then, the proposed method calculates the corresponding rate M based on the score by equation (2). Finally, the proposed method selects a clustered object which has the highest corresponding rate.

3 EXPERIMENTS

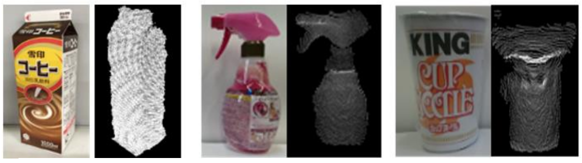
In this section, to evaluate effectiveness of the proposed method, we compare the proposed method with the previous research about five properties mentioned in section 1 as follows.

(1) Robustness against occlusion
(2) Fast recognition
(3) Pose estimation with high accuracy
(4) Coping with erroneous correspondence
(5) Recognizing objects in a noisy environment

3.1 Object Recognition in Occlusion Scene

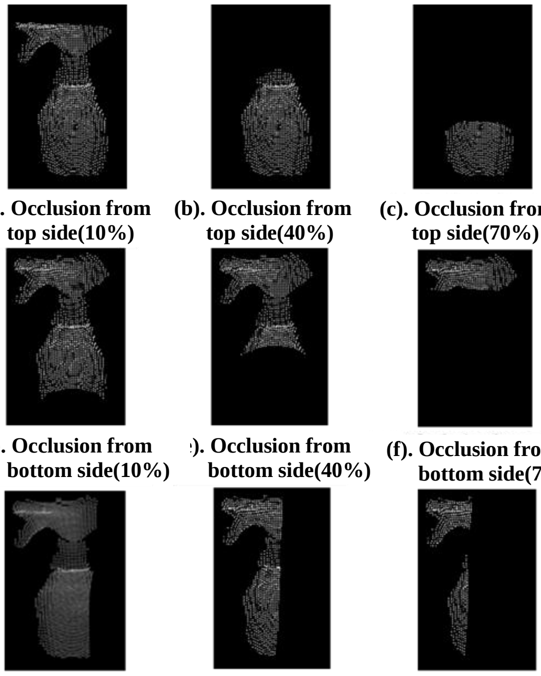
In this experiment, we compared the proposed method with the previous research and SHOT to evaluate about three properties as follows.

(1) Robustness against occlusion



(a). Pack **(b). Spray** **(c). Cup noodle**

Figure 9. Overviews of objects and 3-dimensional data of objects in the experiment



(a). Occlusion from top side(10%) **(b). Occlusion from top side(40%)** **(c). Occlusion from top side(70%)**

(d). Occlusion from bottom side(10%) **). Occlusion from bottom side(40%)** **(f). Occlusion from bottom side(70%)**

(g). Occlusion from right side(10%) **(h). Occlusion from right side(40%)** **(i). Occlusion from right side(70%)**

Figure 10. The examples of occlusion scene of the spray

(2) Fast recognition

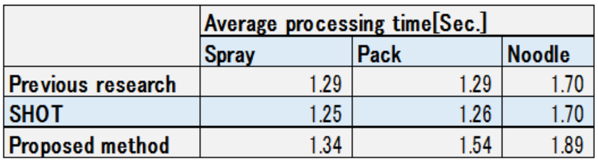
(3) Pose estimation with high accuracy

We used three actual objects as recognition targets which are usually in domestic environment and obtained those 3-dimensional data with Kinect as shown in figure 9. In Figure 9, (a) shows a pack, (b) shows a spray and (c) shows a cup noodle.

To generate occlusion scenes, we delete part of each 3-dimensional object data from 3-directions (top, bottom and right side) by 10% each of point number of each 3-dimensional object data as shown in figure 10.

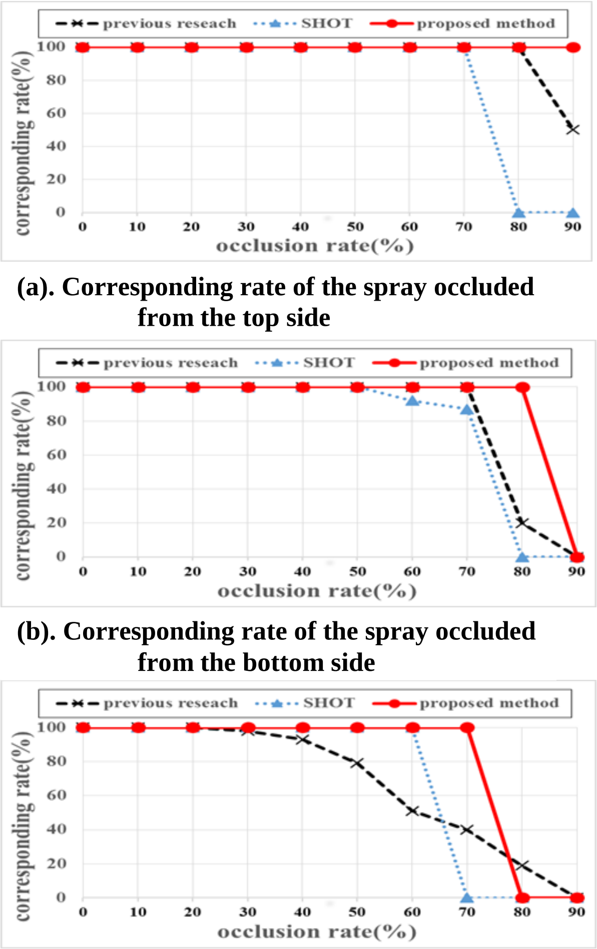
To evaluate a pose estimation accuracy of a target object, we use the corresponding rate M between the target object fitted by using the optimum rotate matrix R and the translation vector t mentioned in the rigid registration process (section 2.7). To calculate the corresponding rate M as a pose estimation accuracy, we use the equation (1) and (2) with th_c which is 1 [mm]. In case that, the corresponding rate is high, that means methods estimate the pose of a target object with high accuracy. On the contrary, in case that, the corresponding rate is zero, which means methods mismatch the target object. The reported processing time is obtained using Intel(R) Core(TM) i5 3.1GHz with 8.0 GB of main memory.

Table 4. The result of average processing time



	Average processing time[Sec.]		
	Spray	Pack	Noodle
Previous research	1.29	1.29	1.70
SHOT	1.25	1.26	1.70
Proposed method	1.34	1.54	1.89

Figure 11 shows the result of occlusion scenes for the spray object. As shown Figure 11, the proposed method was able to recognize objects nearly equal to the previous



(a). Corresponding rate of the spray occluded from the top side

(b). Corresponding rate of the spray occluded from the bottom side

(c). Corresponding rate of the spray occluded from the right side

Figure 11. The result of the spray occluded from 3 directions

research and SHOT in occlusion scene. In addition, as shown in Table 4, a processing time of the proposed method was nearly equal to the SHOT and the previous research. From these results, we consider that the proposed method has the robustness against occlusions because the proposed method is able to match the feature points of the target object and the feature points of unoccluded scene data by using the SHOT. In addition, we consider that the proposed method is able to estimate a pose of a target object with high accuracy because the proposed method uses is not only the corresponding points but also relationships between corresponding points. In this paper, I only show the result of the spray, however we got the same result in other objects.

3.2 Recognizing objects in a noisy environment

In this section, to evaluate effectiveness of the proposed method in recognizing objects in a noisy environment. We prepared same objects with the first experiment. To generate noisy scenes, we added some Gaussian noise on the scenes. Figure 12 shows the changing of the scene data when we added noises.

To evaluate a pose estimation accuracy of a target object, we use the corresponding rate M same as section 3.1.

Figure 13 and Table 5 show results about accuracy and processing time of the proposed method, SHOT and the previous research. As shown the result, we consider that the accuracy and processing time of the proposed method are equal to or more than these of the previous research and SHOT. Although I show only the result of the spray here, the pack and the Cup noodle were able to obtain equivalent result.

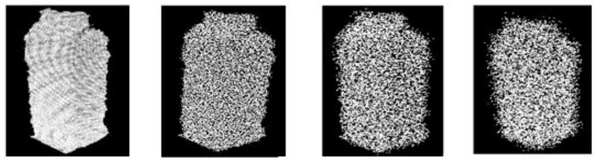

(a). 0.001[σ] (b). 0.002[σ] (c). 0.003[σ] (d). 0.005[σ]

Figure 12. The result of the spray occluded from 3 directions

Table 5. The result of average processing time

	Average processing time[sec.]		
	Spray	Pack	Noodle
Previous research	3.20	2.75	2.62
SHOT	1.24	1.09	1.12
Proposed method	1.82	0.91	1.24

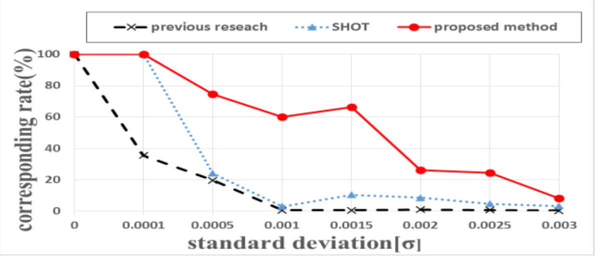


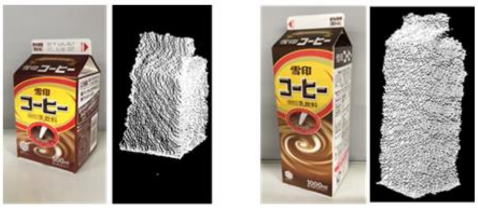
Figure 13. The result of the spray of the pose estimation in a noisy environment

From these results, we consider that the proposed method has the robustness against noises, because the proposed method uses SHOT to generate feature points, SHOT is hardly interfered with noises due to the high dimensionality of SHOT feature quantities.

3.3 The Experiment in Recognition of Objects Which Have the Same Feature but Which are not the Same Object

To qualitatively evaluate about coping with erroneous in the proposed method, we compared the proposed method with the SHOT. As target objects which have the same feature in a local region but which are not the same object, we prepared a 500ml-pack and a 1000ml-pack as shown in Figure 14.

We generated the teaching data from the 1000ml-pack and applied the proposed method, the SHOT to a scene data in 500ml-pack. Figure 15 and Figure 16 show the results of the SHOT, Figure 17 shows the result of the proposed method. As shown in these results, erroneous correspondence occurred in the SHOT and it misrecognized the 1000ml-pack as the 500ml-pack.



(a). 500-ml pack (b). 1000-ml pack

Figure 14. Overviews of objects and 3-dimensional data of objects in the experiment

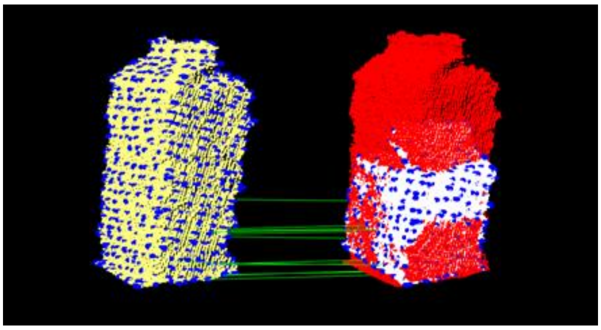


Figure 15. The result of SHOT

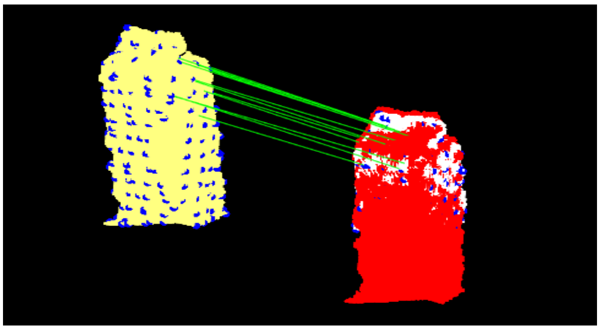


Figure 16. The result of SHOT

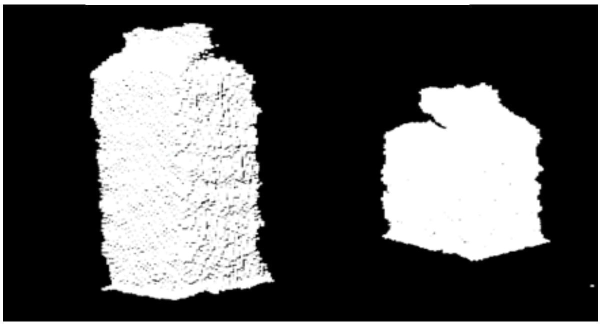


Figure 17. The result of the proposed method

Since the feature quantities of the upper part and the lower part of the pack are exactly the same, when matching is performed, different results are obtained each time depending on a threshold value.

Conversely, the proposed method recognizes that they are the different object since there is no corresponding lines, which means the relationship of distances and angles between points is no corrected. From these results, the proposed method is able to cope with erroneous correspondence and is more effective than SHOT.

4 CONCLUSION

In this paper, we proposed the 3-dimensional object recognition method which has five properties as follows using relationships of distances and angles in feature and SHOT descriptor points for the human support robot.

(1) Robustness against occlusion
(2) Fast recognition
(3) Pose estimation with high accuracy
(4) Coping with erroneous correspondence
(5) Recognizing objects in a noisy environment

In experiments about five properties, we saw the proposed method is more effective than the SHOT and the previous research. Summarizing the above, the proposed method extracts the matching candidate points using SHOT, focuses on the relationship between the candidate points, and eventually uses the list to match.

As a result, it becomes possible to recognize objects that could not be recognized by conventional methods that have been used up to now with high accuracy.

However, the proposed method cannot recognize same shape objects which has different texture because the proposed method only uses SHOT descriptor which are calculated from a shape data of objects. Therefore, as future works, I improve the proposed method by using not only a shape data of objects but also color features in a future work.

REFERENCES

[1] S.Sugano, T.Sugaiwa and H. Iwata,"Vision System for Life Support Human-Symbiotic-Robot," The Robotics Society of Japan, 27 (6), pp. 596-599, 2009.

[2] Y.Jia.,H.Wang.,P.Sturmer, N.Xi, "Human/robot interaction for human support system by using a mobile manipulator," ROBIO, pp. 190-195, 2010.

[3] F.Tombari and S.Salti,"Unique signatures of histograms for local surface description", ECCV, pp. 356-369, 2010.

[4] F.Tombari,S.Salti and L.D. Stefano, "A Combined Texture-Shaped Descriptor for Enhanced 3D Feature Matching", ICIP, pp.809-812, 2011.

[5] S.Maehara,H.imamura,K.Ikeshiro," A 3-Dimensional Object Recognition Method Using SHOT and Relationship of Distances and Angles in Feature Points.",DIPDMWC2015,2015