A Robust Object Recognition Method Using Surface Images of 3D Objects

Kentaro Kawamura
Kyushu Electric Power Co. , Inc
kentarou_kawamura@kyuden.co.jp

Katsushi Ikeuchi
Institute of Industrial Science,
the University of Tokyo

Abstract

Robots which automatically perform outdoor work require robust and accurate environments. The authors devised a new object recognition method, comparing surface images of a target object with 2D surface image models stored in a database. First, range data is segmented into several regions to obtain stable surface images. Second, a surface image is compiled from each stable region. The surface image is comprised of a series of depth values. The depth value is defined as the height from a tangential plane on the center point of a region to points on a geodesic circle. Third, the compiled surface image is compared with prototypical surface images, so as to identify the classification of the region. Lastly, the object's pose can be precisely obtained by applying the 3D template matching technique with the use of a 3D object mesh model. From these processes, the proposed method enables robust recognition of a complex-shaped object. Through the simulation applying the synthesized 3D model, as well as the outdoor experiment, the validity of the proposed method can be obtained. The method is expected to allow robot operators to perform distribution work, such as pole and insulator replacement, more easily and rapidly.

1 Introduction

A recent trend has seen the accelerated use of robots at home and in public facilities; and much R&D has been made into industrial robots and entertainment robots. The robots currently used at assembly factories have data regarding the pose and classification of assembly parts built into their mechanism in advance, resulting in a limitation of work tasks. In order for flexible and complicated work to be performed by robots, advanced object recognition technology is indispensable. Various methods used to recognize target objects have been discussed in previous papers. Most of these methods are for recognition in a structured indoor environment. However, general recognition techniques that only recognize the characteristics of a target object's shape cannot precisely identify complicated shapes. In an outdoor situation, the environment (e.g. illuminant, robot's pose towards the target object) has unlimited variations. It is difficult for outdoor work robots to recognize an unstructured and complicated object using conventional methods. The authors propose a new method that can accurately recognize target objects with complex shapes in various, cluttered environments. This method features the use of range data segmented into several regions to obtain stable surface images with less error, and the application of 3D object mesh models stored in the computer. The algorithm devised for this method was simulated at first, and its validity subsequently confirmed after examining the recognition rate using this method. The theory and procedures involved in this proposed method are described herein, along with verifications made through experiments.

2 Object Recognition Method Applying 3D Range Data

A laser range sensor with enhanced capabilities and lower costs will enable a wider applicable range of object recognition methods using 3D data. Research was made into various object recognition methods using different techniques [1][2][3][4][5][6], for instance the Eigen space method [7][8][9] and the Spin image method using distance histograms [10]. However, some of these require much time for recognition processing and a huge reference database. Although other methods of recognizing objects exist, such as by identifying a 3D object's surface using several parameters [11][12] or using the region segmentation technique [13][14], these methods are not suited for recognizing objects with complicated shapes.

We have proposed a similar method using range images in a previous paper [15]. Satisfactory recognition was made when applied to simple shapes; however, recognition remained hindered when the shapes were complicated. The Spin image, a technique whereby the planes of an object are

rotated, was presented by Johnson, et al [10]. This method featured the projection of the range and height from one given point of the measured data to another point as a 2D image for object recognition. While this technique renders a fairly robust recognition, it does not include rotation data because the 2D image obtained is based on the distribution histogram. Thus, high calculation costs were necessary to determine the pose (rotation direction) of the object.

To rectify this problem, we focused on using characteristics of the object's surface, which do not vary for translation and rotation, and devised a method based on surface images. The surface image is defined as a distribution of heights from the tangential plane to points on a geodesic circle. By incorporating the techniques of region segmentation and temperate matching, the method enables the extraction of a stable portion from the range image, contributing to robust recognition. The proposed method's procedural flow is described below:

1. Segment range data into several regions applying planar fitting technique.
2. Compile surface images applying the range data, starting from the center point of each region.
3. Calculate the similarity of the compiled surface images with the image models pre-registered in the database.
4. Correct the pose of the target object with the template matching method, using a 3D object model to minimize errors and improve recognition precision of the object's pose.

2.1 Segmentation of range data

In compiling the surface image, selecting each region's optimal point of the measured range data is crucial. While it is effective to determine points on or near edges, or with great curvature, the direction of the normal of these points may be unstable. A more stable point is extracted by applying region segmentation. The range data measured is divided into several regions by employing planar fitting, resulting in a comparatively stable group of points. Minor regions, such as noise, are removed. The region growth method [16] is thus incorporated into the proposed method.

2.2 Compilation of surface image at each region

After identifying the optimum points, the point (r_{mid_i}) around the center of each region is calculated and an individual surface image is compiled from this point. As shown in Fig. 1 (a) and (b), the surface image is obtained by plotting the height between the compiled tangential plane and another point. The surface image is defined as being a 2D plane, as illustrated in Fig. 1 (c) with the point (r_{mid_i}) as the origin. Respective pixels are sought by the equation below; thus, intensity is indicated.

$$I_{i,j} = \overline{h(u_k, v_l)} \quad (1)$$

Where

$$u_k : u_{i-1} \leq u_k < u_i \quad (2)$$

$$v_l : v_{j-1} \leq v_l < v_j \quad (3)$$

$$\overline{h(u_k, v_l)} : \text{average value of } h(u_k, v_l) \quad (4)$$

The direction (normal vector of r_{mid_i}) of the tangential plane at the point (r_{mid_i}) contains errors. The more the distance from (r_{mid_i}) increases, the more errors its height measurement contains. In addition, the surface image is arbitrary when in a rotating direction and therefore is difficult to handle. To reduce the influence of such errors, the surface image is developed in a polar coordinate system. From this, the height of the point located further from (r_{mid_i}) can be leveled, compared to that of the point (r_{mid_i}). The curvature around the center point of each region is identified as the characteristic quantity. The direction with the greatest curvature is defined as having a rotating angle of 0 degrees. By developing the image surface within the polar coordinate system on the base of this direction, the correlation of an object with characteristic curvature can be easily and quickly calculated.

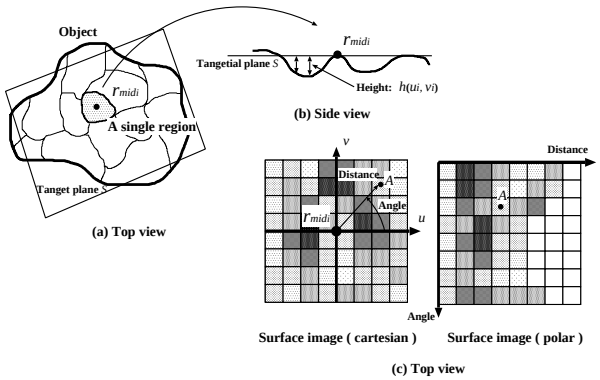


Figure 1. Compilation of surface image

The proposed method needs far fewer points to compile surface images compared to the Spin image method which requires many points to compile spin images because rotation data to determine translation and rotation factors are lacking. Since the maximum direction of curvature may not be obtained in a stable manner, depending on the shape of the surface, the correlation operations are repeated by sliding the surface image models against the surface images, in order of greater curvature. This results in surface image matching. If there is little difference in the scale of curvature, such as for a plane, it becomes necessary to calculate

the correlation for 36 sliding times (for instance, at 10 degree intervals) of the rotating direction. However, the experiment also showed that when distribution equipment had comparatively greater curvature, a maximum of 8 sliding times was required.

2.3 Similarity calculation of surface image

The correlation coefficient can be used to judge whether the surface image obtained through the measured range data is identical to the database model. The linear correlation coefficient (Pr:Person's r) was incorporated into the proposed method. Calculation of this coefficient is formulated as follows:

$$P_r = \frac{\sum_i (h_{r_i} - \bar{h_r})(h_{m_i} - \bar{h_m})}{\sqrt{\sum_i (h_{r_i} - \bar{h_r})^2} \sqrt{\sum_i (h_{m_i} - \bar{h_m})^2}} \quad (5)$$

Where

$$i = 1, \ldots N \quad (6)$$

Here, the value N is the quantity of pixels in which range data exists in both the surface image and the database model. The value h_{r_i} indicates the intensity of the surface image, and h_{m_i} indicates that of the database model. Values $\bar{h_r}$ and $\bar{h_m}$ indicate the average value of h_{r_i} and h_{m_i}, respectively.

2.4 Robust matching of surface image with its database model

A satisfactory recognition of the classification and pose of the object is not always obtained, even when the surface image compiled from the center point of the range data is obtained, and even if a higher correlative surface image model is successfully found from the database. This is because the calculation for the correlation coefficient is derived from only a single given point obtained from a region, and because the normal vector includes errors. To solve this, we proposed finding a correlation method for surface images at several points. Here, the manner in which these several points are selected is crucial. While selection at random is effective to an extent, it is difficult to select several points on the same target object when several other objects exist in the work environment. To obtain several points on the same object, it is necessary to perform many calculations for the correlation coefficient.

Here, we applied the local correlation coefficient in order to obtain several points on the same object. These points were identified from the higher local correlation coefficient calculations between the database surface image model and the surface image compiled from the range data which had the highest correlation coefficient. The center point of the local region or the sub-segmented area (3×3 pixels) is defined as the point in the second stage (named as a 2nd candidate point) in compiling the surface image. When the local region selected has a higher correlation coefficient, the range data must be contained within both the surface image and its database model. There is a high probability that the point obtained at the 1st stage and the 2nd candidate point will appear on the same object. By comparing the correlation coefficient of the surface image at the 2nd candidate point with that of the database model at the 2nd candidate point, a robust matching of surface images can be obtained. The procedure for selection of second order points is illustrated in Fig. 2. In the experiment, we compiled the surface images by selecting three 2nd candidate points.

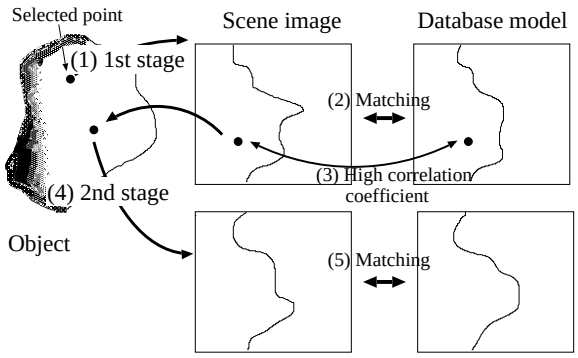


Figure 2. Robust matching of surface image

2.5 Calculation for object pose

There is a great probability that a 3D object model, corresponding to the database surface image model, exists in the measured range data when the highest correlation coefficient value is obtained. The pose of the model is calculated using an affine transformation.

$$(\vec{e_0}, \vec{t_0}) \xrightarrow{\text{Affine transformation}} (\vec{e_1}, \vec{t_1}) \quad (7)$$

Where
- e_0: local coordinate system of the surface image model selected
- t_0: origin of coordinates of the surface image model
- e_1: local coordinate system of the surface image based on the range data
- t_1: origin of coordinates of the surface image based

Next, the pose of the 3D object model calculated by the affine transformation is plotted towards the voting space to determine the real pose, as shown in Fig. 3. The data of

the object classification and pose identified at each region are plotted towards the voting space, and those exceeding pre-set threshold values are selected as candidates for comparison with the stored range data. Multiple candidates exist for rotational symmetrical objects, making it difficult to determine the classification and pose of these objects within the voting space. To solve this issue, surface images similar to the candidates registered in the database were compiled.

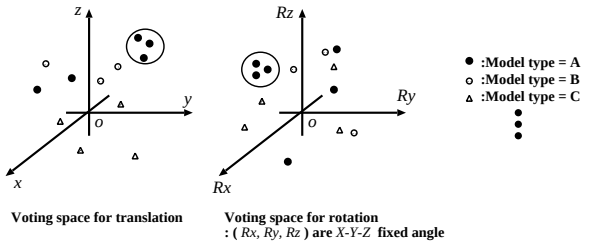


Figure 3. Voting space for seeking translating factor and rotating factor

It is possible to find the classification and pose of the 3D object model from the voting space. However, they are not precise because the surface images are created from dispersed data. For better precision, the pose of the model projected into the range data must be corrected by applying the temperate matching technique (3DTM method) [17]. This enables full recognition of the object's pose.

3 Experiments

3.1 Simulation using geometric models

A simulation is conducted using synthesized data and several 3D objects compiled in the computer in order to clarify the characteristics of the robust object recognition method proposed by the authors using surface images of 3D objects. Four different CAD objects were developed in the computer, and the surface images for each object were compiled in advance. In addition, sensor models were prepared to simulate the group of 3D points found when a laser range sensor measured a target object. Data from the four CAD objects observed from various directions were used for the simulation.

We examined the influence of measurement noise attributable to resolution (e.g. tolerable error, performance) of the sensor, as well as the error of the target object's shape, by mixing noise into the 3D data. The noise mixed in had a normal distribution with an error average of zero. We examined its standard deviation, which was (σ).

The CAD models used for simulation are shown in Fig. 4. Seven viewing directions were used, as shown in Fig. 5. They comprised 6 directions out of the 12 peaks of a regular icosahedron, and 1 frontal viewing direction. When the target object is a perfect sphere, different viewing directions do not affect the data as the object looks the same from any point, resulting in a single direction. Fig. 6 shows several example images with various noise levels. Table 1 shows the simulation results, where (a) indicates the number of successful results achieved when using a seven-view direction (although a single viewing direction was used for spherical objects), which implies complete recognition for all seven directions. The value (b) indicates the estimated pose of the object obtained by calculating the surface images, and (c) indicates the error between the corrected poses in the 3DTM. The cylinder's Y-axis and sphere were not taken into account for the experiment. The noise applied had three patterns:σ=5[mm], 10[mm], and 15[mm]. The success of recognition was confirmed visually after 3DTM was carried out. A favorable simulation result was obtained for the ellipsoid. However, for the cylindrical and octagonal objects, the more mixed error increased, the more difficult it became to identify the difference between the two shapes. As a result, recognition was less successful. At the same time, however, satisfactory noise robustness was achieved, considering that the repetitive measurement error of the range sensor used for the experiment was less than 2 [mm].

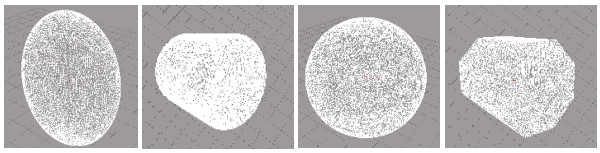


Figure 4. CAD models used for simulation

3.2 Experiment in an outdoor environment

Kyushu Electric has employed distribution work robots to execute replacement of electric wires and insulators mounted on poles, as illustrated in Fig. 7. The operator remotely controls the robots from the ground, applying 3D laser range sensors and CCD cameras [19]. The series of work tasks involved are difficult for linemen who lack relevant expertise. To make robot operation for overhead work more efficient and less time consuming, it is essential that even operators without special skills can manipulate the robot. By applying the method proposed by the authors, operators will be relieved of the complicated task of ma-

Table 1. Simulation results

Model		Scanned data with noise			
		No noise	5 mm	10 mm	15 mm
Ellipsoid	(a)	7	7	7	6
	(b)	6.95	7.04	7.74	11.02
	(c)	1.54, 7.96, 4.41	3.45, 7.25, 3.98	6.96, 15.66, 14.35	3.64, 7.18, 4.05
Cylinder	(a)	7	6	5	4
	(b)	1.96	1.97	5.60	6.94
	(c)	16.23, -, 10.41	20.14, -, 18.01	21.04, -, 10.23	16.54, -, 3.22
Sphere	(a)	1	1	1	0
	(b)	6.71	8.68	2.79	-
	(c)	-	-	-	-
Octagonal cylinder	(a)	7	4	4	2
	(b)	6.59	11.09	5.63	13.32
	(c)	3.72, 1.94, 6.10	13.14, 6.57, 11.09	15.07, 2.52, 8.13	3.29, 1.21, 0.49

(a): Number of successful times (ellipsoid, cylinder and octagonal cylinder: 7 trials; sphere: 1 trial)
(b): Initial translation error [mm]
(c): Initial rotation error (X-Y-Z fixed angle [deg.])

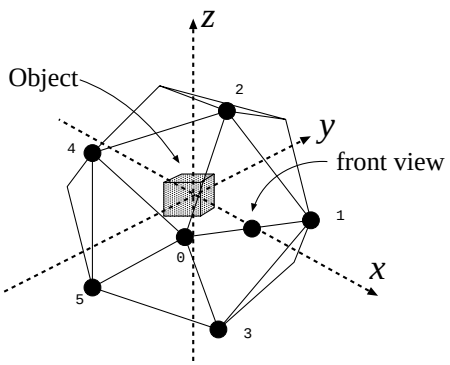
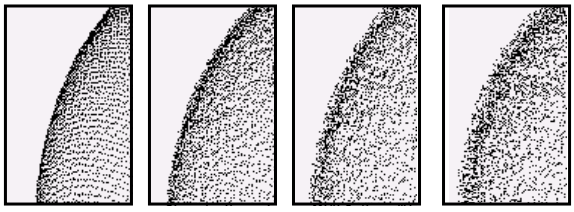


Figure 5. Viewing directions used for simulation



Sigma = 0 **Sigma = 5** **Sigma = 10** **Sigma = 15**

Figure 6. Examples of noise mixed in (for an ellipsoid)

nipulating the robot, and the tasks outlined above will be completed more quickly. This method is expected to be incorporated within fully automatic distribution work robots, which are currently being developed by Kyushu Electric. Insulators mounted on poles are targeted as particularly important recognition objects, as they need to be located and worked on frequently.

3.2.1 Compilation of surface images and 3D objects

In the experiment, one of the laser range sensors manufactured by Pulstec was used. This sensor is capable of providing range data outdoors, and is effective when applied to outdoor tasks such as distribution work. The insulators used for the experiment were actual service insulators with a length of 20 – 35 [cm] and a width of approximately 10 [cm]. The experiment set out to compile 3D geometric models of distribution equipment and the surface images of each of these. The CAD system may be used to compile these models and images, but complicated shapes such as insulators are not applicable. To solve this problem, the range data of an object from several viewing directions was obtained and triangle patches were produced. The final model was obtained by aligning and integrating the triangle patches. [18]

The surface images were compiled in the manner described in section 2.2 by executing region segmentation at each point. No surface images were produced for minor regions to minimize the volume of data to be stored. Even though the data of minor regions were not stored, recognition accuracy was not affected because of the number of regions with large areas. In addition, rotating angles for plotting the surface image onto the polar coordinates were

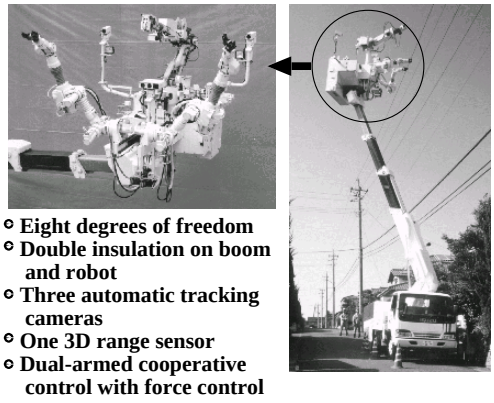
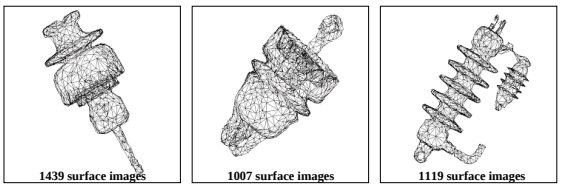


Figure 7. Semi-automatic hot-line work robot

determined as 10 [deg.], and the distance from the center point was set at 5 [mm] spans. These specifications were defined from the resolution of the laser range sensor and the surface image model size found from the experiment results.

3.3 Recognition using the proposed method

Three types of 3D objects were used in the experiment, as illustrated in Fig.8. All were identical to the types of distribution equipment in service. Fig. 9 shows six image scenes on which recognition was experimented. Study was made in different theoretical situations, such as when part of an object was shielded or when several objects were present within a single datum.



(a) Type I (b) Type II (c) Type III

Figure 8. Three types of 3D objects (insulators)

Fig. 10 shows the relation between the surface image and the curvature of the range data of scene 1. The surface images were rearranged according to maximum curvature direction in order to compile the surface images at several points from the central point of each region obtained by the region segmentation. In the left figure, the area around point P is similar to a cylinder in shape. The surface images are rearranged automatically, setting the point where

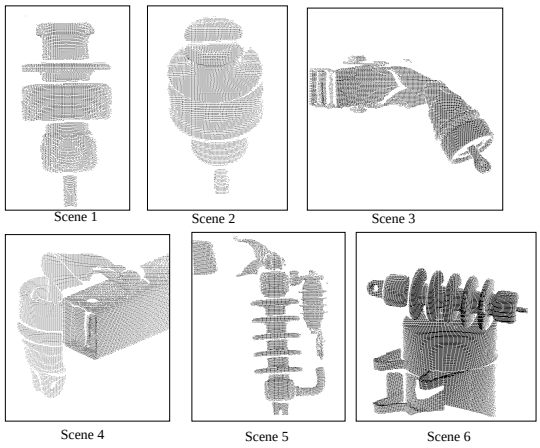


Figure 9. Six image scenes based on range data

the curvature is the largest at 0 [deg.], shown in the figures on the right. Then, the surface image was matched with the database model at several points to obtain a robust matching performance.

Fig. 11 shows an example of scene 1. In the first stage, a stable point (point P of Fig. 11 (a)) from the region segmentation was selected and the surface image was matched with a database model. Next, a point with a correlation coefficient higher than the database model was selected as the 2nd candidate point. In the experiment, we pre-set the extraction of three 2nd candidate points at this stage. The points indicated as A, B and C on Fig. 11 (a) and (b) were each defined as 2nd candidate points. Thus, four points in total were used for the matching process. The matching result determined a surface image compiled from a point of the 1st stage; thus the classification and pose of the target object were selected. Then the 3DTM method was applied to determine whether the classification and pose of the target object obtained using surface images had been correctly identified. For this process, we paid attention in particular to distance errors in corresponding points from the 3D object and range data. The use of 3DTM allowed us to accurately recognize the target object's pose. Fig. 11 (c) illustrates an object whose 3D object obtained from the surface image matching has been superimposed onto the range data. A favorable recognition performance was achieved.

The results of the experiment with different scenes are indicated in Table 2. Successful recognition was found for each scene. The table shows the difference between the target object pose obtained from surface image matching, and the pose corrected by the 3DTM process for different scenes. As it was difficult to find the exact pose of the target object, we carried out visual observation to find out whether

the work robot would be able to perform its task.

The use of 3DTM, together with region segmentation, contributes to more accurate recognition of a target object's pose. In addition, error in the distance between the 3D object obtained by the 3DTM and the range data measured contributes to verifying whether the data result obtained from surface image matching is correct.

The time required for recognition is about 50 [sec] (Pentium III 866MHz) on average. To calculate the correlation coefficient of the surface image, 1.5 times this average time are required if maximum curvature data are not used.

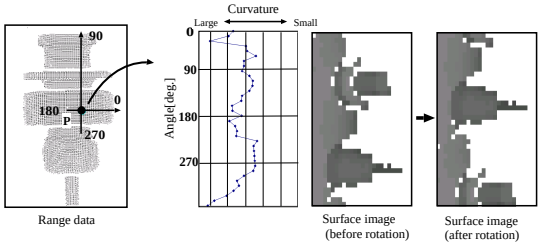


Figure 10. Relation between surface image and curvature (for scene 1)

Table 2. Recognition Results

Scene No.	Results			
	translation error [mm]	rotation error [deg.] x	y	z
1	3.74	0.32,	3.41,	2.52
2	5.79	0.15,	1.06,	14.58
3	9.69	1.14,	1.64,	1.55
4	5.30	2.65,	0.69,	2.17
5	22.77	7.01,	0.312,	2.58
6	18.81	6.12,	20.36,	1.92

Rotation error : X-Y-Z fixed angle [deg.]

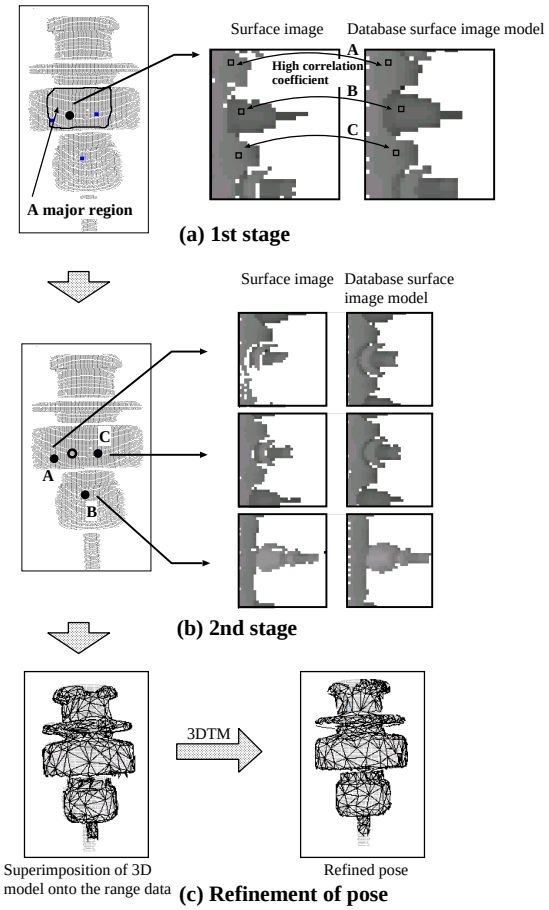


Figure 11. Selected points for the second matching stage

4 Conclusions

We have proposed a new object recognition method, not only applicable for distribution work robots but also for robots employed in various other fields. Focus was made on object recognition using the surface images of the target object together with region segmentation, to reduce errors of the normal on tangential planes and thus identify the surface image. The proposed method is summarized below:

- Characteristics of the surface shape of the target object are used.
- A 2D surface image is compiled by transforming to polar coordinates according to the distance between the 3D range data and the tangential plane.
- The region segmentation method is used to compile stable surface images.
- The correlation coefficient is calculated using plural surface images at the 1st and 2nd stages, enabling robust recognition.

The series of simulations revealed that the proposed method has robustness against noise. The results of the experiment achieved satisfactory recognition even for an object with a complicated shape, such as distribution equipment. At present, recognition time of about one minute is required, but research is being made into shortening the recognition time required for searching the surface images. In the future, the proposed method will be incorporated into

distribution work robots. Additional focus will be placed on further improvement of the method's recognition rate.

References

[1] R. J. Campbell and P. J. Flynn, "A Survey Of Free-Form Object Representation and Recognition Techniques," *Trans. IEEE. CVIU*, vol. 81, pp. 166-210, 2001.

[2] F. Stein and G. Medioni, "Structual indexing: Efficient 3-D object recognition," *Trans. IEEE. PAMI*, vol. 14, no. 2, 1992.

[3] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3D object recognition," *Int. J. Computer Vision*, pp. 63-85, 1997.

[4] K. Ikeuchi, "Recognition of 3-D Objects Using the Extended Gaussian Image," *Proc. of the Intern. Joint Conf. on Artificial Intelligence*, pp. 595–600, 1981a.

[5] P. J. Flynn and A. K. Jain, "BONSAI:3-D Object Recognition Using Constrained Search," *Trans. IEEE. PAMI*, vol. 13, no. 10, 1991.

[6] G. Barequet and M. Sharir, "Partial surface and volume matching in three dimensions," *Trans. IEEE. PAMI*, vol. 19, pp. 929–948, 1997.

[7] M. Turk and A. Pentland ,"Face recognition using eigenfaces," *IEEE Conf. CVPR*, pp. 586–591, 1991.

[8] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Computer Vision*, vol 14, pp. 5-24, 1995.

[9] Kohtaro Ohba and Katsushi Ikeuchi, "Detectability, Uniqueness, and Reliability of Eigen-Windows for Robust Recognition of Partially Occluded Objects," *Trans. IEEE. PAMI*, Vol. 19, No. 9, pp.1043-1048, 1997

[10] A. E. Johnson. and M. Herbert, "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes," *Trans. IEEE. PAMI*," Vol. 10, No. 5, May, 1999.

[11] Ernest M. S. and Shang Y. W. , "Surface Parameterization and Curvature Measurment of Arbitrary 3-D Objects: Five Principal Methods," *Trans. IEEE. PAMI*, Vol. 14, No. 8, Aug, 1992.

[12] Stein F. , Medioni G., "Structual Hashing: Three Dimensional Object Recognition, *Trans. IEEE. PAMI*, Vol. 14, No. 2, Feb, 1992.

[13] R. Hoffman and A. K. Jain, "Segmentation and Classification of Range Images, " *Trans. IEEE. PAMI*, vol. PAMI-9, no. 5, pp. 608–620, 1987.

[14] P. J. Besl and R. C. Jain, "Segmentation Through Variable-Order Surface Fitting, " *Trans. IEEE. PAMI*, vol. 10, no. 2, pp. 167–192, 1988.

[15] K. Kawamura, K. Hasegawa, Y. Someya, Y. Sato and K. Ikeuchi, "Robust Localization for 3D Object Recognition Using Local EGI and 3D Template Matching with M-Estimators," *IEEE Conf. on Robotics and Automation*, pp. 1848–1855, 2000.

[16] M. Oshima and Y. Shirai, "Object Recognition Using Three-Dimensional Information, " *Trans. IEEE. PAMI*, vol. PAMI-5, no. 4, pp. 353–361, 1983.

[17] M. D. Wheeler and K. Ikeuchi, "Sensor Modeling, Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition," *Trans. IEEE. PAMI*, vol. 17, no. 3, pp. 252–265, 1995.

[18] Greg Turk and Marc Levoy, "Zippered polygon meshes from range images," *Proc. of SIGGRAPH 94*, pp. 311–318, 1994.

[19] M. Nakashima, H. Yakabe, M. Maruyama, K. Yano, K. Morita and H. Nakagaki, "Application of Semi-Automatic Robot Technology on Hot-Line Maintenance Work," *IEEE Conf. on Robotics and Automation*, pp. 843–850, 1995.