Localization of freely curved pipes for bin picking

Yasuyo Kita and Yoshihiro Kawai
Intelligent Systems Research Institute
National Institute of Advanced Industrial Science and Technology (AIST)
AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
Email:{y.kita, y.kawai}@aist.go.jp

***Abstract*—A method for detecting the position and orientation of freely curved pipes is proposed for the purpose of bin picking. We assume that the geometrical model of the object of interest is given and the bin of objects can be observed using a widely available low-cost 3D sensor. As freely curved pipes do not have distinguishable feature points, typical feature-based matching methods cannot be applied directly in this case. Although the iterative closest point (ICP) algorithm [1] is a suitable strategy for registering a 3D model with free-form surfaces to its observed 3D data, two difficulties arise when applying the ICP algorithm in the case where the background is filled with objects with the same shape. One of these is the initial estimation of the position and orientation of an object from among similar shaped point clouds. The other difficulty is robust point matching between the 3D model and the observed data belonging to a single object. To solve these issues, we investigated two key ideas: 1) estimation of pipe orientation by checking the shape of the cross-section of the observed 3D data while changing cutting directions of the sections; and 2) point matching based on ridge lines with respect to the camera direction. Experiments using actual observed data show the effectiveness of the proposed method.**

I. INTRODUCTION

Many studies have focused on localization of an object for bin picking (e.g. [2],[3],[4],[5],[6]), which is required in various situations, such as requesting parts from stock to assemble into a product. Although both two-dimensional (2D) and three-dimensional (3D) observations have been used for this purpose, recently, along with the popularization of low-cost 3D sensors such as the Kinect, 3D point clouds have become easier to obtain at low cost and with little complexity. In fact, there has been much interest from small- and medium-sized enterprises to build a low-budget automatic bin-picking system using such sensors. In many industrial scenarios including such requests, target objects are rigid (non-articulated, non-deformable) and the geometrical model of the object is available in advance. Thus, in this paper, our research is based on these conditions.

For bin picking, it is essential to detect and localize objects lying on top of other objects, thereby allowing them to be picked out first. This is generally a non-trivial task since an object must be localized against a cluttered background with many of the same features.

Most existing methods perform this localization by aligning the model with the observed data. These methods can be classified into three types according to the features used to match the model with the observed data, namely: 1) salient feature points (e.g., corners and marks); 2) shape primitives (e.g., spheres and cylinders); and 3) generic features.

978-1-4673-7929-8/15/$31.00

If objects of interest have distinguishable feature points, the first approach works in a relatively simple and fast way. For example, Rahardja et al. [3] detected circular holes on a target object by using ellipse fitting in images captured by stereo cameras. However, since the target is observed from only one direction in bin picking, this strategy is applicable only when enough features can be viewed from any direction.

In the case where objects of interest consist of basic primitive shapes like spheres and cylinders, geometrical characteristics of such primitives are reliable clues to determine the position and orientation. Nieuwenhuisen et al. [7] used subgraph matching, where nodes of the graph are instances of simple shapes such as cylinders and planes.

The third approach calculates more generic features on the surface. Kirkegaard et al. [8] used differential geometrical features that characterize the surface in a small region around a point of interest. However, features based on differential geometry are often unstable in practical scenes, especially in the case where 3D point clouds are obtained using a low-resolution 3D sensor. Avoiding the use of differential features, Buchholz et al. [9] used four-dimensional features that are fixed by a pair of points on a surface, the distance between the two points, and three angles fixed by the normal of the points. Matching these features between the model and observed data produces many hypotheses of the location. These hypotheses are checked and refined by the iterative closest point (ICP) algorithm [1]. The ICP algorithm iteratively finds the best matching of two surfaces based on correspondences of the closest pairs of points.

Although this algorithm is theoretically applicable to arbitrary shapes, in reality, it is not that straightforward because too many false hypotheses can be produced depending on the shapes. For example, in their work [9], Buchholz et al. had to remove a planar area that produces many similar features.

If the target object consists of free-form curved surfaces, the first and second approaches cannot be used since the objects do not contain both salient feature points and shape primitives. The third approach is applicable, but is expected to require high computational time owing to the large number of similar features. Freely curved pipes, as illustrated in Fig. I, are one such example.

In this paper, we propose a method that can work for such pipes by using the characteristics of the pipe shapes effectively.

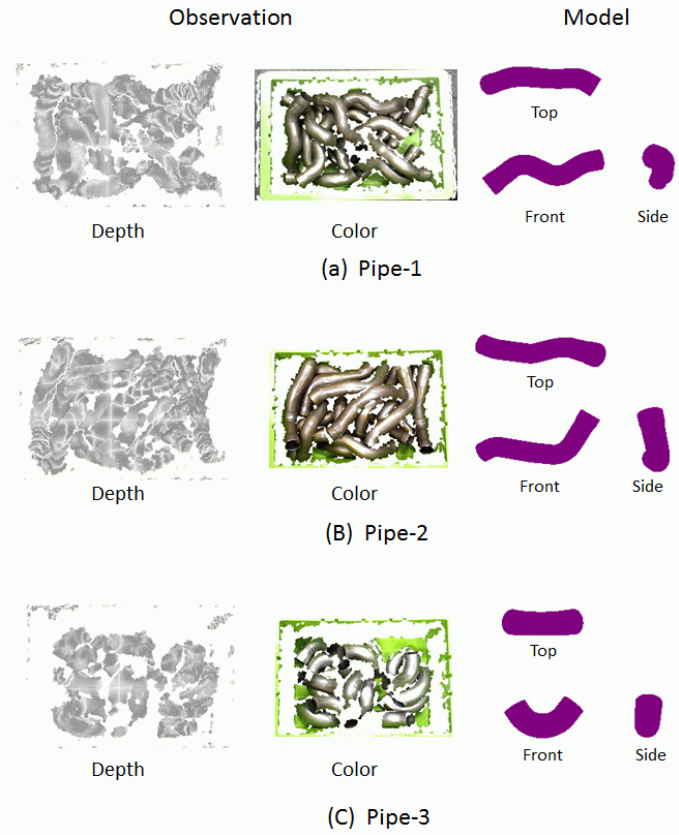


Fig. 1. Example of a model and the observations

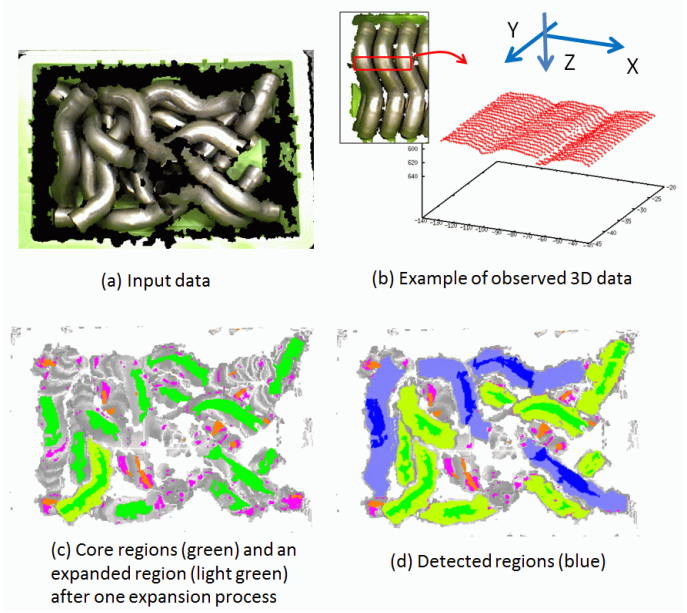


Fig. 2. Preprocessing: segmentation of pipes lying on others

Our main contribution is twofold. First, the proposed method obtains an initial estimation of the position and orientation of a pipe for the ICP algorithm based on imperfect segmentation results of the pipe. It estimates the principal axis of the pipe by checking the cross-section of the segmented region while changing the cut direction of the section. Second, the method enables robust and fast point correspondences in the ICP process by using ridge lines with respect to the camera direction instead of all the points on the surface. The newly proposed feature for point matching, called the Z-ridge, has advantages such as good observability and robustness to noise.

In Sections II, III, and IV, we describe, respectively, the preprocessing and both these proposals. In Section V, experimental results using actual observations are discussed.

II. Preprocessing

A. Observed data

As previously noted, we assume the use of a widely available low-cost 3D sensor for observing a pile of objects in a box. Fig. I shows examples of observations taken by the ASUS Xtion and the model of pipes dealt with in this study. Here, the distortion of the original 3D point clouds obtained by the sensor have already been corrected by a sensor calibration method [10]. Although color information of the 3D points is also obtained, we use only the depth information. In the depth images of Figs. I and 2, 3D points are represented by gray dots. Fig. 2(b) shows a different view of part of the observation with the axes of the 3D coordinates we set. The 3D sensor is set directly above the bin so that the view direction, the Z-axis of our coordinates, coincides with the direction of gravity.

Observed 3D points are restored in a Z-buffer image, the pixel width of which is 1 mm, zimg[j][i]; the Z-coordinate of a 3D point (X,Y,Z) is stored at zimg[round(Y)][round(X)], where function $round(x)$ returns the integer closest to real number x. Using this Z-buffer image, the normal direction of each point is calculated in advance by plane fitting using the points stored in the point's neighboring pixels.

B. Segmentation of pipe regions from the background

When picking up an object from the bin, the situation in the bin changes and must be observed again before the next action. Therefore, we set our goal as localization of only those pipes that are easily picked up, that is, pipes lying on top of other pipes. To extract regions corresponding to such pipes, first, core regions that are expected to belong to the uppermost pipes are extracted and then, these core regions are expanded into the neighboring pixels. The resultant regions with sufficient length to cover the entire shape are selected as possible pipe regions. The algorithm flow is shown in Algorithm 1 with the details given below:

1. Extraction of core regions
   Regions with upward normals are extracted as core regions. First, pixels with a normal (N_l) that is close to the vertical direction $(N_v) = (0, 0, -1)$, are extracted, using the condition $N_l \cdot N_v > N_{th1}$. Then, the region

Algorithm 1 Extract candidate regions

1. Extract connected regions with upward normals as core regions (area condition: $A_n \geq A_{th}$)
2.
for Each core region, n **do**
  2.1 Expand the region
  while Neighboring pixels satisfying the condition below exist **do**
    if The normal of pixel i, neighboring any boundary pixel of the region N_i, is close to the normal of boundary pixel j, N_j,
    $(N_i N_j > N_{th1})$ **then**
      Merge the pixel into the region
    end if
  end while
  2.2 Calculate the furthest distance between points in the region, l_{max}
  if $l_{max} > aL$ **then**
    Select the expanded region as a region corresponding to a pipe lying on others
  end if
end for

composed of pixels connected with 8-connectivity to each other, A_n, is selected if it is larger than the fixed area threshold A_{th}, $(A_n > A_{th})$. The green region in Fig. 2(c) shows an example of the result.

2. Expansion of core regions
   The core regions are expanded to points surrounding these regions if the normal direction of the points change continuously from the boundary of the region. The light green region in the lower left of Fig. 2(c) shows the result after expanding one of the core regions.
3. Selection of uppermost pipes
   If a pipe lies on top of other pipes, most of the upper parts of the pipe can be observed. This can be determined using the longest 3D distance between the points in the region. The four light purple regions in Fig. 2(d) show an example of the result, with the threshold for the selection set to $aL, a = 0.7$ and L is the longest distance between the points furthest from each other in the 3D model.

Please note that each resultant region corresponds to the pipe that includes the core region before expansion.

III. Estimation of Position and Orientation

If the region corresponding to one pipe is only segmented, the principal axis of inertia of the 3D data of the region gives a close estimate of that of the corresponding pipe. However, segmentation can sometimes fail as shown in Fig. 3(b). Here, because the pipes are in close contact with each other, the region swells out into neighboring pipes (light-purple region), while expanding the core region colored blue. As a result, the principal axis of the region is different from that of the pipe

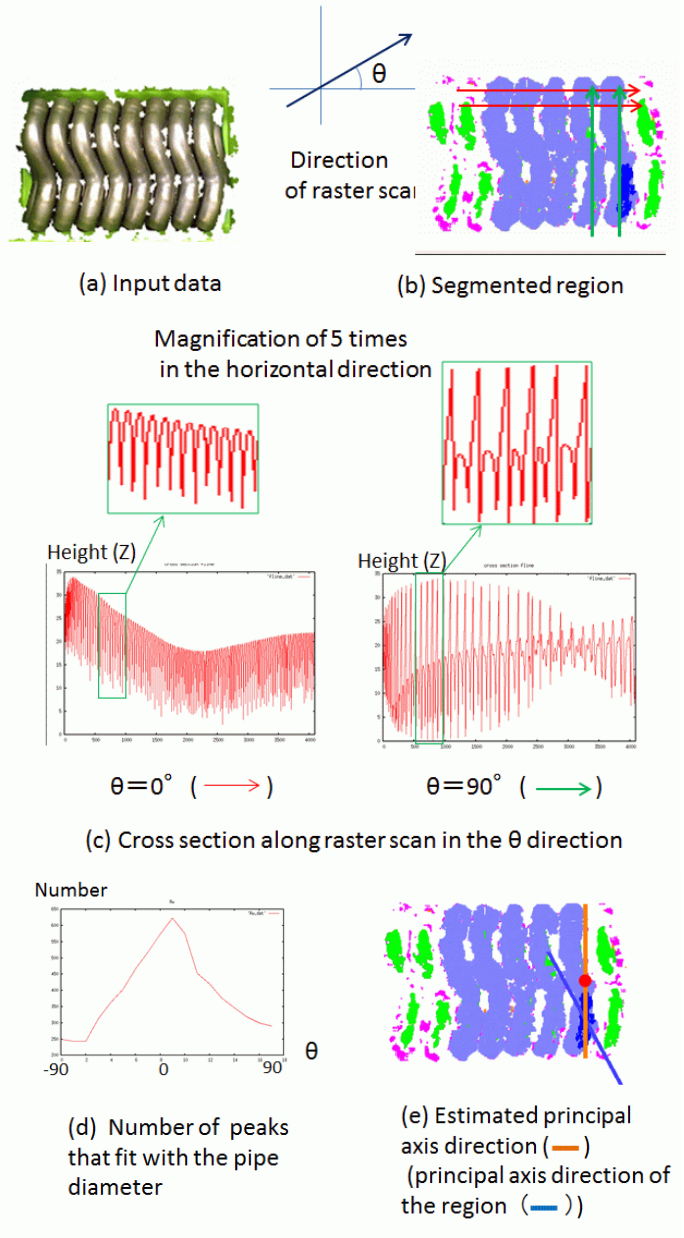


Fig. 3. Estimation of orientation

corresponding to the core region, as shown by the blue line in Fig. 3(e).

To estimate the correct orientation even in such cases, we adopt a different method to estimate the orientation, based on the characteristics of the pipe shape. First, the depth of the region is raster scanned into a one-dimensional signal with the direction of the scan, θ. Red (green) arrows in Fig. 3(b) show an example for $\theta = 0°(90°)$. During the scan, the values of the pixels in the segmented region only are included in the signal; that is, the signal comprises a sequence of cross-sections. The two graphs in Fig. 3(c) show the resultant signals for $\theta = 0°$ and $90°$, respectively. When the region is scanned in a direction perpendicular to the principal axis

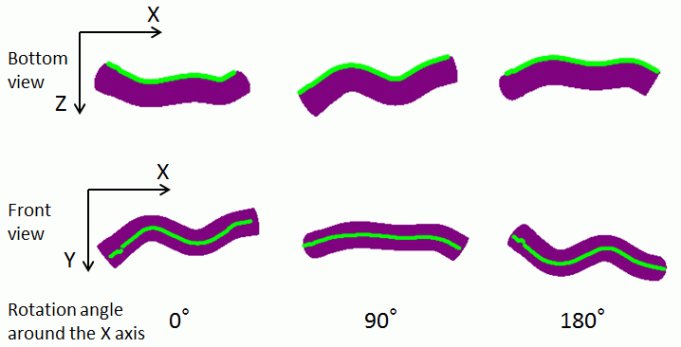


Fig. 4. Z-ridge of pipe-1

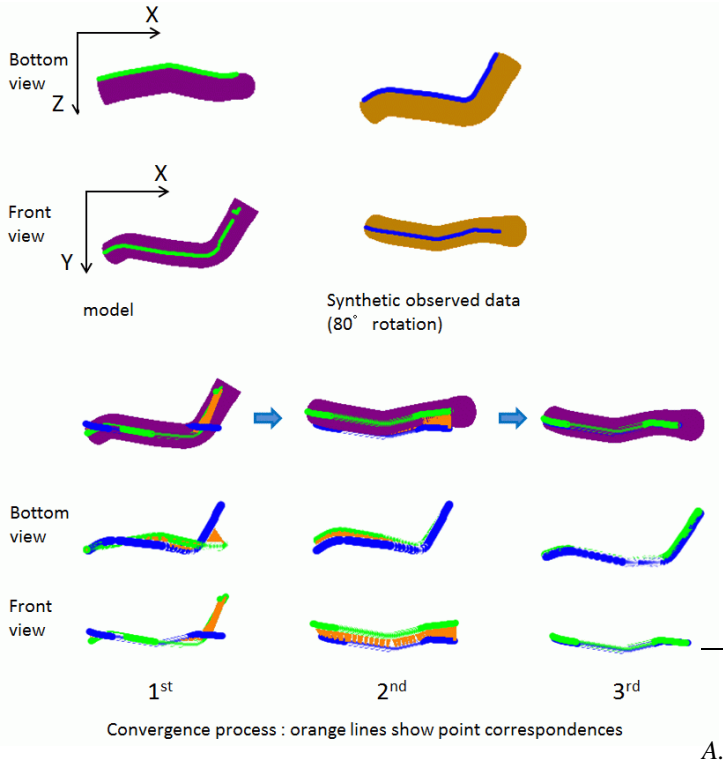


Fig. 5. Example of ICP process using Z-ridge (pipe-2)

of the corresponding pipe, the signal should include cross-sections, the width of which is close to the pipe diameter, R, and the narrowest of all the directions. Since our 3D sensor, like most vision sensors, cannot measure a surface with the normal almost perpendicular to the camera direction, the narrowest width is not the same as the diameter, R, but a little smaller. Taking this into consideration, we count the number of cross-sections in each signal with width $0.9R \sim 1.0R$. Fig. 3(d) shows the transition of the number relative to θ. The angle giving the maximum number is detected as the direction perpendicular to the principal axis of the pipe. The orange line in Fig. 3(e) shows the resultant direction.

This process determines an angle around the Z-axis in the XY-plane only. However, considering that most pipes lying on others are almost horizontal, we think we can ignore angles around the X- and Y-axes. Once this direction has been determined, a line segment is drawn in the direction passing the centroid of the core region and with length equal to the width of the observed area in the direction. The orange line in Fig. 3(e) shows an example. The center point of the line segment (the red point in Fig. 3(e)) and its direction is used to determine the initial position and orientation of the model in the ICP registration process.

Algorithm 2 Register model to candidate regions

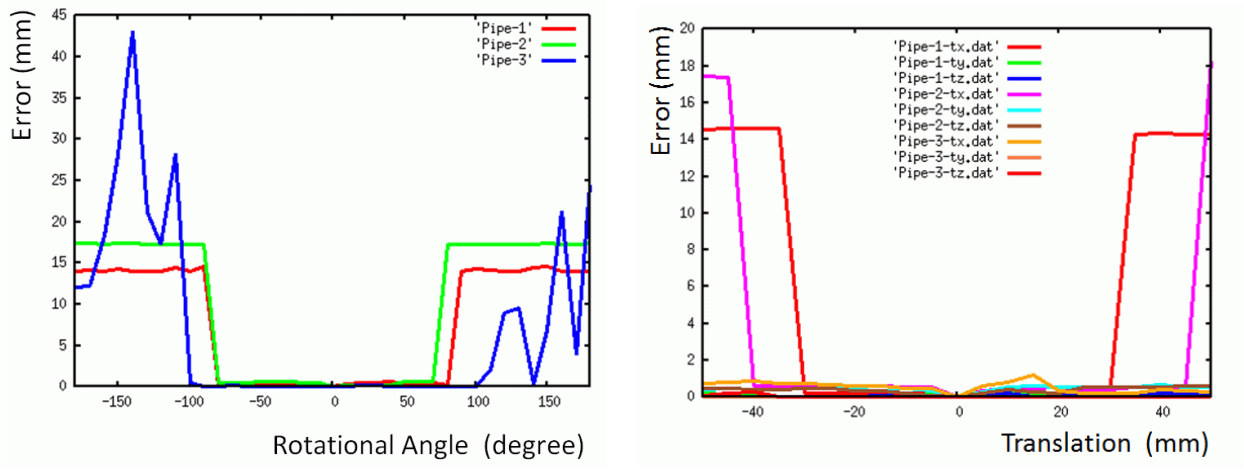
for Each candidate region, n **do**
  1. Estimate principal axis direction
  2. Detect Z-ridge points P_o
  for Initial states, m ($1 \sim 6$) **do**
    2.1 Set model at the state
    2.2 Register the model to the region
    while until convergence **do**
      2.2.1 Calculate Z-ridge points P_m from the model at the current state
      2.2.2 Find closest point pairs between P_m and P_o
      2.2.3 Transform the model based on the least square error
    end while
    2.3 Record the converged state, $\hat{m}$, with the minimum residual error, $R_{\hat{m}}$
    if Residual error, $R_{\hat{m}} < R_{min}$ **then**
      $R_{min} = R_{\hat{m}}$, $M = \hat{m}$
    end if
  end for
  3.
  if $R_{min} < R_{th}$ **then**
    Locate the model at state M
  end if
end for

IV. Registration by ICP using Z-ridge

A. Feature for point correspondences

When registering a 3D model to its 3D observed point cloud within the framework of ICP, in general, points for matching are sampled regularly from all the model surfaces that are visible from the current view direction. Here, assuming that all surface points are not necessarily required to fix the position and orientation of pipes, we propose using a kind of ridge line instead of using points from the entire surface. Concretely, points that are local minima of the Z-coordinates along the direction of the Y-axis under the condition that the object (or observed 3D data corresponding to the object) is placed so that its principal axis coincides with the X-axis. The green lines in Fig. 4 show some examples. Henceforth, we refer to the line as a Z-ridge.

The Z-ridge of an object changes according to the rotation angle around the X-axis as shown in Fig. 4. So, while the orientation of the model does not coincide with the actual



(a) (b)

Fig. 6. Convergence characteristics

orientation of the observed pipe, the Z-ridge calculated from the model and that calculated from the observed data do not correspond exactly with each other. Nevertheless, we found that the Z-ridge still works as a feature for matching within the framework of the ICP algorithm, where iterative calculations using not exact, but close point correspondences cause the model to converge to the correct state. Fig. 5 shows an example of the convergence process using artificially created observed data (yellow points) by rotating the model (dark purple) by 80 degrees around the X-axis. Here, orange lines illustrate correspondences between the modelfs Z-ridge points (green points) and the observed Z-ridge points (blue points). Several iterations, during which Z-ridges of the model are recalculated at each state, align the model in the correct position and orientation.

Using Z-ridge points instead of all the surface points has three advantages and one disadvantage. First, since the ridge compactly represents the 3D shape, it offers an effective way of decreasing the number of points for point matching and reduces computational time. Second, the Z-ridge suffers from less occlusion than all the surface points owing to its good observability from the camera. Third, the Z-ridge tends to cause fewer incorrect point correspondences than all the surface points since it is distant from the points of other pipes even when the pipes are in close contact with each other. The disadvantage is that using the Z-ridge may cause registration inaccuracy. We reconsider these characteristics in the discussion in Section V.

B. Necessary number of initial states

As we can estimate the principal axis direction using the method described in Section III, the remaining degrees of freedom about the orientation are the two opposite directions along the axis and the rotation around the axis. The number of initial states needed to converge to the correct state in this situation depends on the shapes. The required number can be estimated in advance by means of a test using synthetical observation data like that shown in Fig. 5. Fig. 6(a) shows the residual errors after convergence to the synthetic observation data created by rotating the model around the principal axis while increasing the rotation angle at 10-degree intervals. As can be seen, the model converges to the correct orientation within $\pm 80^\circ$, $\pm 70^\circ$, and $\pm 100^\circ$ for pipe-1, pipe-2, and pipe-3, respectively. These results indicate that trials from three initial states, $0^\circ, 120^\circ, 240^\circ$ can converge to the correct orientation.

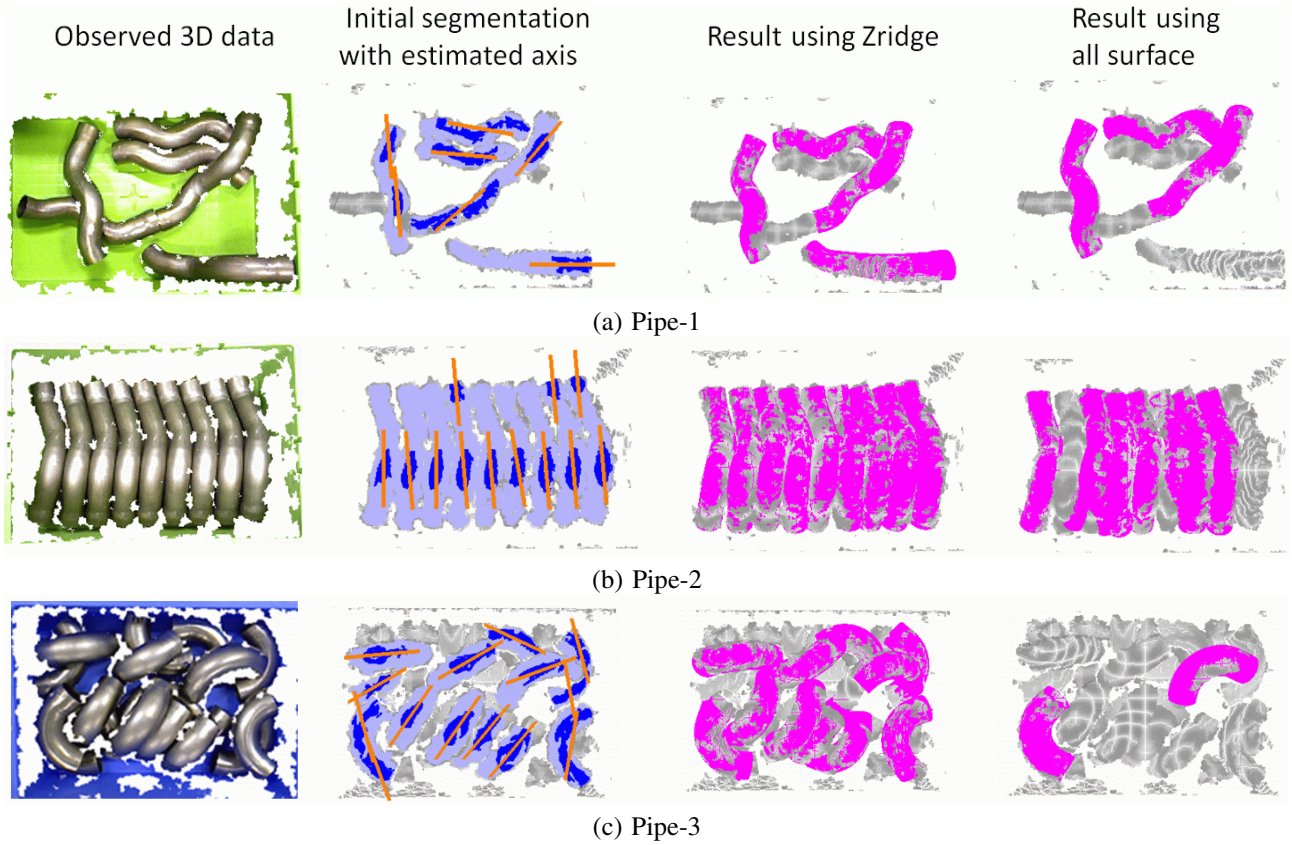
Similarly, the allowable translation difference for correct convergence can be tested to determine how dense we should set the initial positions. Fig. 6(b) shows the residual errors after convergence to the synthetic observation data created by moving the model by 5 cm in the X-, Y-, and Z-directions, respectively. As can be seen in all the experiments except those in the X-direction, the model converges to the synthetic data in all cases. In the case of X-translation of pipe-1 (pipe-2), translation within 3 cm (4 cm) is acceptable. Because, in our case, translation errors are expected to be within these limits, correct convergence is expected from one initial estimation position only.

Of course, these results are for cases without noise. However, addition of Gaussian noise to the synthetic data has a negligible effect on the results. Therefore, we determine the number of initial states for the experiments in Section V based on these results. Concretely, for each candidate region, six (2x3) states are used as initial states for the ICP processes: two opposite directions along the principal axis and three rotation angles in each direction.

The algorithm flow corresponding to Sections III and IV is shown in Algorithm 2.

V. Experiments

Experiments were conducted using 10 observations of each pipe captured by an ASUS Xtion sensor, as shown in Fig. I, giving a total of 30 observations. As noted in Section I, we



(a) Pipe-1

(b) Pipe-2

(c) Pipe-3

Fig. 7. Comparison of the results of Z-ridge points and all points

used only the depth information and not RGB. The 3D model of each pipe is given in stl format.

The number of pipes lying on top of other pipes was manually counted in all images: 43, 40, and 71 for pipe-1, pipe-2. and pipe-3, respectively. Of these, the number of pipes localized by the proposed method were 40/43, 35/40, and 52/71, respectively. In all our experiments, whether the model converges to an observed region was determined based on the average residual distance between corresponding pairs of the model and the observed data. A manually set threshold was used to judge this.

Fig. 7(a), (b), and (c) shows examples of the results for pipe-1, pipe-2, and pipe-3, respectively. The resultant regions of Algorithm 1 are shown as purple regions with a blue core region in the second column of Fig. 7 (and Fig. 9). The orange lines show the estimated direction of the principal axis. Most of these angles are within 10 degrees of the manually obtained ground truth.

To clarify the effect of using the Z-ridge instead of the whole visible surface, results using points sampled regularly from all visible surfaces are shown in the fourth column of Fig. 7 (and Fig. 9). Whereas ICP using surface points often failed to find the correct localization, ICP using the Z-ridge mostly yielded correct localizations. For example, in Fig. 7(a), all four pipes not lying under any other pipes, were correctly localized.

Fig. 8 illustrates the reason for this difference using the case of the third pipe from the left in Fig. 7(b). The orange line in Fig. 8(a) shows the principal axis estimated for the blue core region. The convergence process from the initial state closest to the actual state of the six initial states derived from the principal axis, is shown in Fig. 8(b) and (c). Here, blue and green points represent the observed surface/Z-ridge and model surface/Z-ridge, respectively, while the orange lines show point correspondences between the model and observed data. Registration using all surfaces was contaminated by incorrect correspondences with the 3D points of the neighboring pipes. On the contrary, registration using the Z-ridge succeeded in finding most desirable correspondences by eliminating the effect of those 3D points. The rightmost column shows the resultant location.

Regarding pipe-1 and pipe-2, the positions and orientations were almost correctly obtained as shown in Fig. 7(a) and (b). For pipe-3, however, deviations from the correct orientation occasionally appeared as shown in Figs. 7(c) and 9(b). The reason for this may be that the shape of pipe-3 closely matches part of a torus, which has shape invariance with respect to the rotation of the center axis of the torus. Another possible factor is that only small parts of the pipe were observed depending on the view direction. These factors pose difficulties not only for the proposed method, but for all methods.

Fig. 9(a) and (b) shows examples of pipe-2 and pipe-3 that include mis-localization. In the case of Fig. 9(a), strong

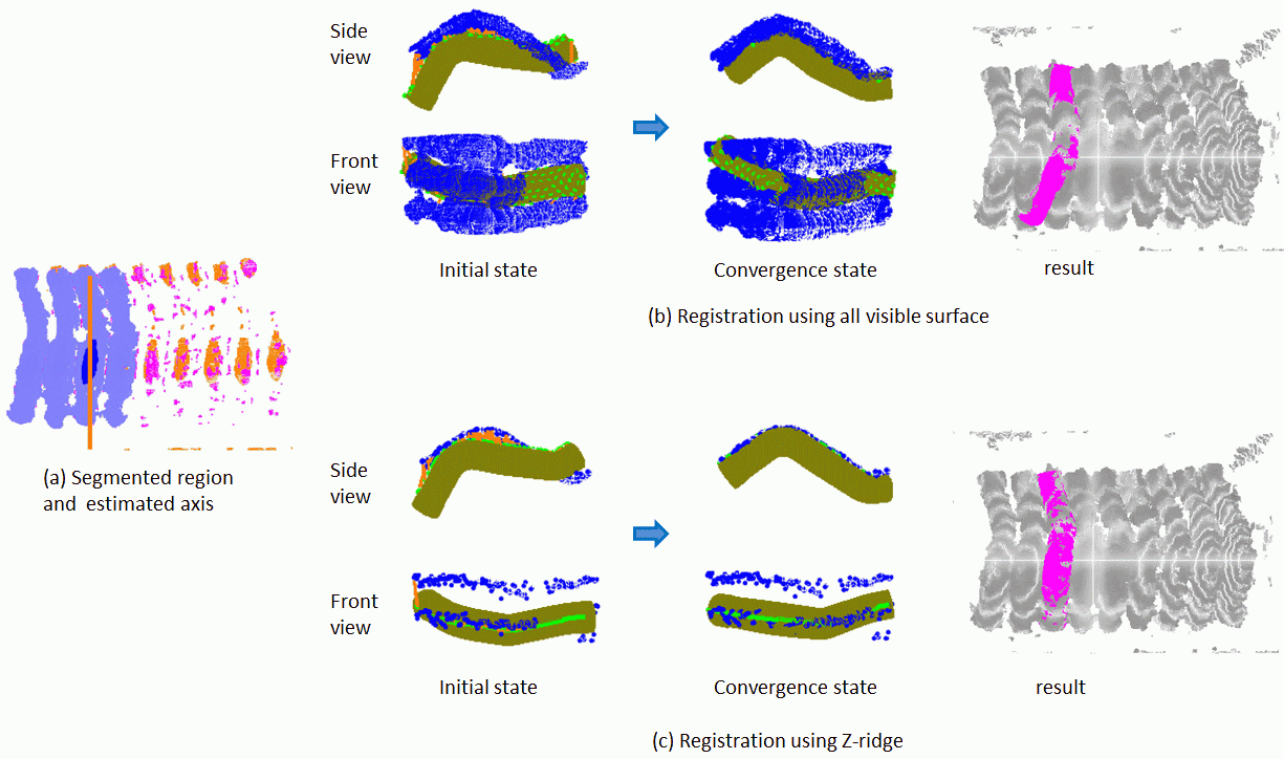


Fig. 8. Registration processes for the third pipe from the left in Fig. 7(b)

specular reflection contaminated the original 3D observation causing a lack of 3D data for certain pipes demarcated by red dashed circles. The preprocessing process failed to extract certain parts of two pipes that were incorrectly located.

In the case of Fig. 9(b), one pipe demarcated by a blue dashed circle was located upside down at a different position. This occurred because the Z-ridges of the two objects were unfortunately connected. This can be avoided by checking the length of observed Z-ridges.

VI. Conclusion

A method for localizing freely curved pipes for the purpose of bin picking has been proposed. Free-form shaped objects lying in a pile of the same objects are generally difficult to localize, mainly because there is no particular feature that specifies exact point-to-point correspondences between the model and the observed data. In this paper, we showed that a ridge line with respect to the camera direction (Z-ridge) can be used to find point correspondences within the framework of the ICP algorithm. A Z-ridge has some good characteristics; one of these is its good observability arising from the fact that the normal at the ridge is parallel to the view direction. In addition, for smoothed convex objects, the Z-ridge can be determined for any view of the object. Another advantage is that the Z-ridge is less affected by outliers since the line is generally set apart from background points. Experimental results using actual observations showed that these advantages are evident in practice and work well.

One of the problems in the experiments was that strong specularity tends to contaminate the 3D data around the Z-ridge. Another issue is that, in the case of strongly curved pipes, often only small parts of the pipes can be observed; these are not sufficient to determine the exact location of the pipes.

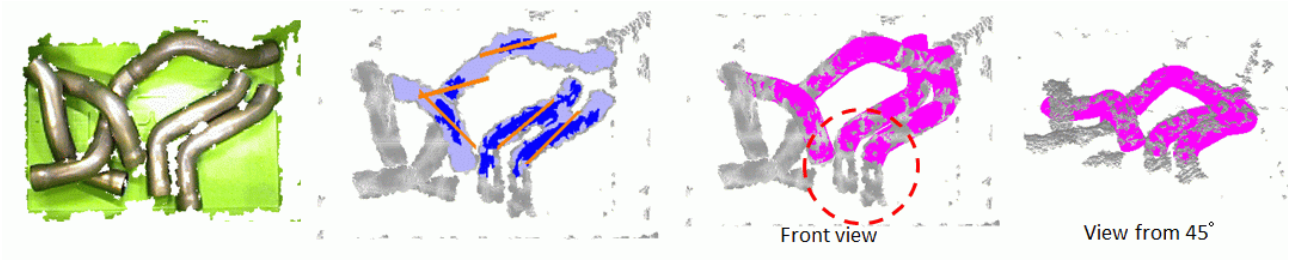
Although we deal only with pipes having an almost uniform diameter in the experiments in this paper, the strategy should be extendable to pipes with more complicated conditions. Extension of the shape generality is one of our future research topics. We are also constructing a complete picking and placement system by combining the proposed method with advanced planning and manipulation technologies [11].

Acknowledgments

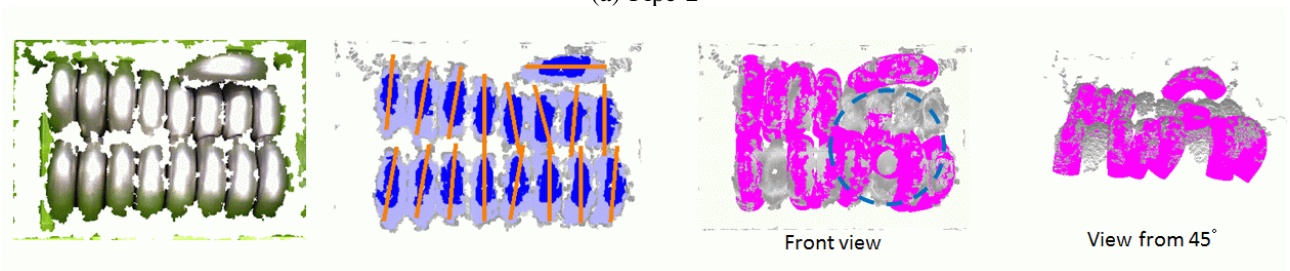
We are grateful to all members of the project entitled "Development of a robot system for random picking to support middle- and small-scaled companies" for their support in various aspects including discussions, data acquisition, and calibration.

References

[1] P. J. Besl and N. D.Mckay: "A method for registration of 3D shapes", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, No.2, pp. 239–256, 1992.



(a) Pipe-2



(b) Pipe-3

Fig. 9. Examples including incorrect localization

[2] D. G. Lowe: "The viewpoint consistency constraint", *International Journal of Computer Vision*, Vol. 1, No. 1, pp. 57–72, 1987.

[3] K. Rahardja and A. Kosaka:"Vision-based bin-picking: Recognition and localization of multiple complex objects using simple visual cues", In *Proc. of Int. Conf. on Intelligent Robots and Systems (IROS 1996)*, pp. 1448–1457, 1996.

[4] J. J. Rodrigues, J. Kim, M. Furukawa, J. M. F. Xavier, P. M. Q. Aguiar and T. Kanade: "6D pose estimation of textureless shiny objects using random ferns for bin-picking", In *Proc. of Int. Conf. on Intelligent Robots and Systems (IROS 2012)*, pp. 3334–3341, 2012.

[5] M-Y Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T. K. Marks, and R. Chellappa:"Fast object localization and pose estimation in heavy clutter for robotic bin picking", *International Journal of Robotics Research*, Vol 31, No. 8, pp. 951–973, 2012.

[6] K. Harada, K. Nagata, T. Tsuji, N. Yamanobe, A. Nakamura, Y. Kawai:"Probabilistic Approach on Object Bin Picking Approximated by Cylinderst", In *Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA '13)*, pp.3727–3732, 2013.

[7] M. Nieuwenhuisen, D. Droeschel, D. Holz, J. Stuckler, A. Berner, J. Li, R. Klein, and S. Behnke:"Mobile Bin Picking with an Anthropomorphic Service Robot", In *Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA '13)*, pp.2319–2326, 2013.

[8] J. Kirkegaard and T. Moeslund: "Bin-picking based on harmonic shape contexts and graph-based matching", In *Proc. of 18th International Conference on Pattern Recognition*, pp.18–21, 2006.

[9] D. Buchholz, M. Futterlieb, S. Winkelbach and F. M. Wahl: "Efficient Bin-Picking and Grasp Planning Based on Depth Data", In *Proc. of IEEE Int'l Conf. on Robotics and Automation (ICRA '13)*, pp.3230–3235, 2013.

[10] Y. Satoh, K. Iwata, T. Nagami, and K. Takeuchi: "Compensation of Distortion in Depth Data Obtained from RGB-D Camera", In *Proc. Of ViEW2013*, , IS1–F6, 2013 [In Japanese].

[11] K. Harada, T. Yoshimi, Y. Kita, K. Nagata, N. Yamanobe, T. Ueshiba, Y. Satoh, T. Masuda, R. Takase, T. Nishi, T. Nagami, Y, Kawai, and O. Nakamura:"Project on Development of a Robot System for Random Picking", In *Proceedings of IEEE/SICE International Symposium on System Integration*, pp. 583–589, 2014.