

Surface-Graph-Based 6DoF Object-Pose Estimation for Shrink-Wrapped Items Applicable to Mixed Depalletizing Robots

Taiki Yano¹^a, Nobutaka Kimura²^b and Kiyoto Ito¹^c

¹Research & Development Group, Hitachi, Ltd., Kokubunji, Tokyo, Japan

²Research & Development Division, Hitachi America, Holland, Michigan, U.S.A.

Keywords: Object Recognition, 6DoF Pose Estimation, Depalletizing, Shrink-Wrapped Item.

Abstract: We developed an object-recognition method that enables six degrees of freedom (6DoF) pose and size estimation of shrink-wrapped items for use with a mixed depalletizing robot. Shrink-wrapped items consist of multiple products wrapped in transparent plastic wrap, the boundaries of which are unclear, making it difficult to identify the area of a single item to be picked. To solve this problem, we propose a surface-graph-based 6DoF object-pose estimation method. This method constructs a surface graph representing the connection of products by using their surfaces as graph nodes and determines the boundary of each shrink-wrapped item by detecting the homogeneity of the edge length, which corresponds to the distance between the centers of the products. We also developed a recognition-process flow that can be applied to various objects by appropriately switching between conventional box-shape object recognition and shrink-wrapped object recognition. We conducted an experiment to evaluate the proposed method, and the results indicate that the proposed method can achieve an average recognition rate of more than 90%, which is higher than that with a conventional object-recognition method in a depalletizing work environment that includes shrink-wrapped items.

1 INTRODUCTION


Mixed depalletizing is the process of unloading multiple types of items that are stacked on pallets or in cages, and there is a great need to automate this heavy workload process using robots. Many researchers have proposed various systems and methods to achieve robot automation (Nakamoto et al., 2016; Eto et al., 2019; Doliotis et al., 2016; Aleotti et al., 2021; Stoyanov et al., 2016; Katsoulas and Kosmopoulos, 2001; Katsoulas et al., 2002; Kirchheim et al., 2008; Kimura et al., 2016). Size and six degrees of freedom (6DoF) object-pose estimation, which accurately determines the size, position, and orientation of an object, plays an essential role in enabling robots to unload a variety of items (Poss et al., 2019; Mitash et al., 2020; Monica et al., 2020; Fuji et al., 2015; Yano et al., 2022).


One of the challenges in size and 6DoF object-pose estimation for mixed depalletization is estimating the boundaries of shrink-wrapped items. Shrink-wrapped items, often seen in depalletizing operations,

consist of multiple products such as plastic bottles wrapped together in a transparent wrap. Because it is difficult to measure this wrap area with a distance sensor, only the top surface of each product in the shrink-wrapped item can be measured, making it difficult to identify the boundaries of the entire shrink-wrapped item. Since the size and arrangement of each item is generally not known in advance in mixed depalletizing operations, it is difficult to correctly estimate the 6DoF pose of each shrink-wrapped item, especially when such items are placed adjacent to each other. This incorrect recognition must be avoided because it may lead to incorrect grasping, such as picking two items at the same time.

To achieve shrink-wrapped item recognition, we propose a surface-graph-based object-pose recognition method that estimates the 6DoF pose and size of each shrink-wrapped item (Fig.1). Focusing on the fact that the same type of products are packed closely together in a shrink-wrapped item and the distance between them is almost constant, we detect the boundaries of the shrink-wrapped items by constructing a surface graph, where the product surfaces are nodes and the distances between their centers are the edge lengths. Since there are usually gaps or mis-

^a <https://orcid.org/0000-0001-9433-0569>

^b <https://orcid.org/0000-0001-5248-5108>

^c <https://orcid.org/0000-0002-2243-5756>

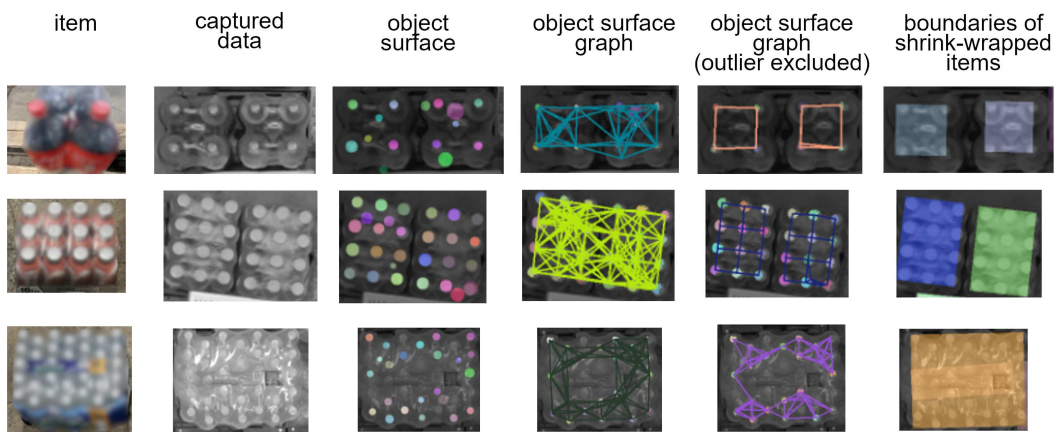


Figure 1: Surface graph-based 6DoF object-pose estimation for shrink-wrapped items. Columns show results of each processing stage in estimating boundaries of shrink-wrapped items (with each image blurred). Focusing on fact that same type of products are packed closely together in shrink-wrapped item and distance between them is nearly constant, proposed method constructs surface graph representing connection of products, and estimates region of each shrink-wrapped item by determining homogeneity of edge length, which corresponds to distance between centers of products. This makes it possible to correctly estimate boundaries of variety of shrink-wrapped items.

alignments between the shrink-wrapped items, there will be a difference between the distances among the products "within" one shrink-wrapped item and the distances among the products "between" multiple shrink-wrapped items. Therefore, the boundaries of shrink-wrapped items can be estimated from the differences in the edge length of this surface graph. We also developed a recognition-process flow that can be applied to various objects by appropriately switching between using a conventional edge-based object-recognition method for boxed items and the proposed method for shrink-wrapped items.

To compare the performance of the proposed method with a conventional edge-based object-recognition method, we conducted experiments to simulate a mixed depalletizing process using a robot arm. The results indicate that the proposed method could correctly recognize 6DoF poses of both boxed and shrink-wrapped items and achieve an average recognition rate of more than 90%, which is higher than that of the only the conventional method. This recognition accuracy validates the applicability of the proposed method to mixed depalletizing robots that handle shrink-wrapped items.

2 RELATED WORK

There has been much research on 6DoF object-pose estimation for depalletization. Before we describe the proposed method, we discuss conventional object-recognition methods. These methods are categorized as local feature matching, deep-learning-based object recognition, and edge-based boundary detection.

Local Feature Matching. One of the most commonly used methods for recognizing the 6DoF pose of an object is a local-feature-matching method (Tang et al., 2012; Lowe, 2004). Such a method generates and matches local features representing pattern information between a three-dimensional (3D) object model created in advance and captured images and has high recognition performance for objects with patterns. However, it is not suitable for cases in which it is difficult to create a 3D object model in advance, such as in mixed depalletizing operations. When the same pattern is seen repeatedly, as with shrink-wrapped items, local features from the 3D model and captured images obtained at different locations can potentially be matched, which may result in incorrect 6DoF pose estimation. It is also difficult to recognize items without patterns, such as cardboard boxes.

Deep-learning-based Object Recognition. Current object-recognition methods using deep learning can automatically learn useful features for estimating the boundaries and/or 6DoF pose of items on the basis of a large amount of data prepared in advance (Xiang et al., 2018; Kehl et al., 2017; He et al., 2020; Wang et al., 2019; Mitash et al., 2020; Yang et al., 2022; Gou et al., 2021). However, when handling new items that differ significantly in appearance from those used for training, a large amount of training data must be prepared again, and this learning cost is an issue in the field of logistics, where products are frequently replaced.

Edge-based Boundary Detection. Conventionally, edge-based boundary detection has been widely used

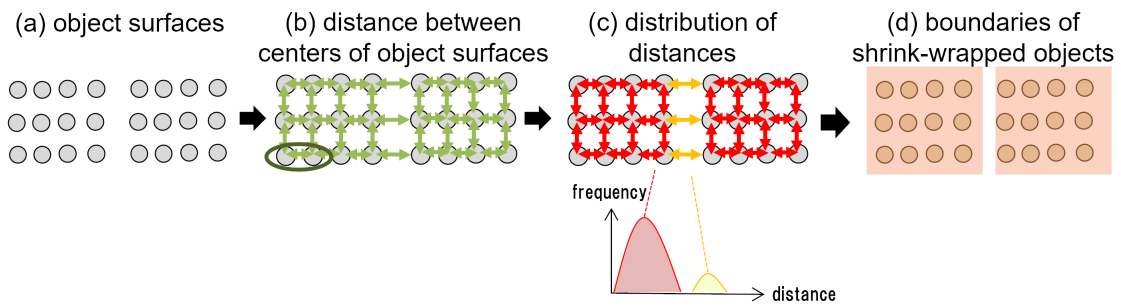


Figure 2: Boundary-detection procedure for shrink-wrapped items. First, graph connecting each detected surface and distribution of distances between surface centers are obtained, and edges corresponding to infrequent center-to-center distances are excluded from graph. This enables estimating boundaries of shrink-wrapped items.

as 6DoF object-pose estimation for boxed items (Katsoulas and Kosmopoulos, 2001; Katsoulas et al., 2002; Monica et al., 2020; Naumann et al., 2020; Stein et al., 2014). By estimating the edges on the basis of the luminance gradient or degree of change in the normal direction, it is possible to recognize the boundary of each item and estimate the 6DoF pose. However, due to the difficulty in detecting the edges of wraps that cannot be measured, there is a risk of detecting not each shrink-wrapped item but its internal products as a single item. The proposed method is similar to this type of method, but differs in that it estimates the boundaries of shrink-wrapped items on the basis of the regularity of distances between products.

3 METHODS

3.1 Surface-Graph-Based 6DoF Object-Pose Estimation for Shrink-Wrapped Items

This section describes the proposed method for recognizing shrink-wrapped items (Fig. 2). As mentioned in 1, when a shrink-wrapped item is captured with a distance sensor, only the top surface of each product in the shrink-wrapped item is measured, making it difficult to determine the boundary of each shrink-wrap item. Therefore, the proposed method constructs a surface graph that represents the connection of each product in shrink wrapped items and uses this graph to estimate the boundaries of shrink-wrapped items and their 6DoF poses and sizes.

After extracting the product surfaces by the process described in 3.2.1, we connect the surfaces that are within a certain distance, forming a surface graph that represents each surface as a node and its connections as edges (Fig.2(b)). We associate the distance

between the centers of the surfaces (referred to as the center-to-center distance) with the corresponding edges. We then obtain the distribution of the center-to-center distances on the basis of all this information (Fig. 2(c)). In this distribution, center-to-center distances with high frequency are considered edges connecting products within shrink-wrapped items, while center-to-center distances with low frequency are considered edges connecting products from different shrink-wrapped items. Therefore, we exclude edges corresponding to low-frequency center-to-center distances from the graph and identify the subgraphs that represent shrink-wrapped items to detect the boundaries between shrink-wrapped items (Fig. 2(d)). We exclude edges corresponding to center-to-center distances that are more than a certain distance from the median of the distribution. Finally, for the group of surfaces within the estimated shrink-wrapped-item region, we carry out plane fitting using random sample consensus (RANSAC) (Fischler and Bolles, 1981; Holz et al., 2015) and estimate the 6DoF pose and size of the bounding rectangle surrounding the region.

When the wrap is translucent and partially measured, many noisy surfaces are included other than the surface corresponding to the product, and the regularity of the center-to-center distance is buried in the noise. This has a significant negative impact on the boundary-detection procedure described above (Fig. 3). To avoid this problem, we carry out plane fitting using RANSAC for the surface nodes belonging to the graph formed in Fig. 2(b) and exclude nodes that are more than a certain distance from the plane in advance.

3.2 Developed Recognition-Process Flow for both Boxed and Shrink-Wrapped Items

This section describes our developed recognition-process flow for recognizing items in various pack-

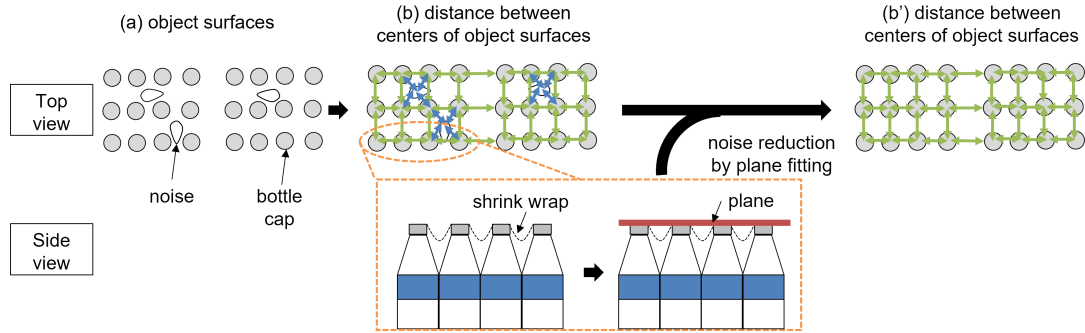


Figure 3: Noise reduction in boundary-detection procedure. Carrying out plane fitting on surfaces and excluding surfaces that are more than certain distance from the plane reduce impact of noise on regularity of center-to-center distance.

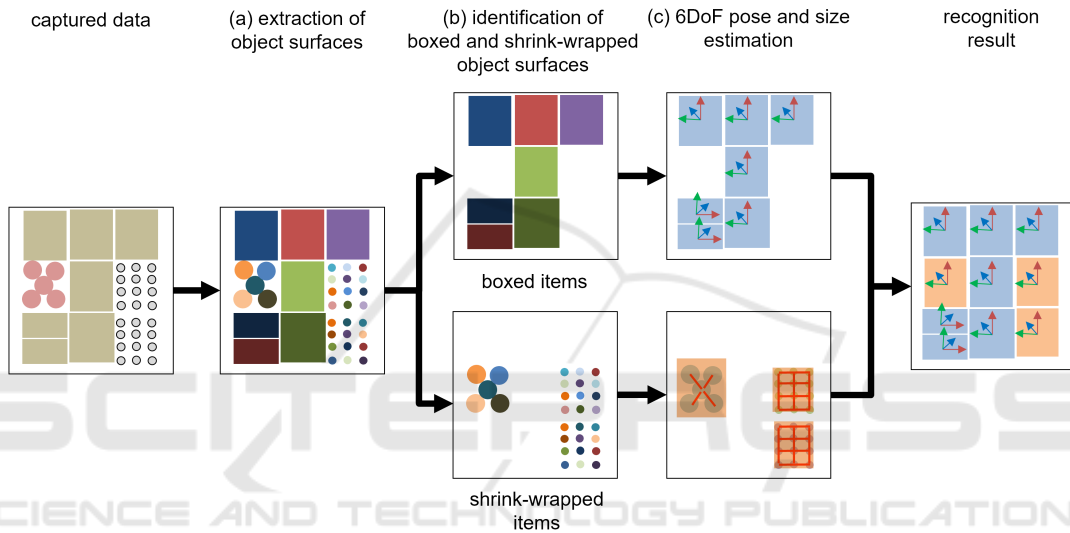


Figure 4: Developed recognition-process flow. Given scene image and point cloud, our recognition-process flow first estimates whether each region obtained by segmentation is surface of boxed or shrink-wrapped item. Then, for boxed item, its size and 6DoF pose are estimated, and for shrink-wrapped item, its boundary, size, and 6DoF pose are estimated.

aging forms. This flow first estimates whether the extracted surfaces are boxed or shrink-wrapped items, respectively. If it is a boxed item, its size and 6DoF pose are obtained with the same procedure as a conventional method, and if it is a shrink-wrapped item, its size and 6DoF pose are obtained with the proposed method on the basis of the regularity of the distance between the items. These processes mainly follow the three steps shown in Fig. 4.

3.2.1 Extraction of Object Surfaces

The surface of each item is first extracted by dividing the captured image into regions (Fig. 4(a)). Note that for shrink-wrapped items, the surface of each product inside the shrink-wrapped item is extracted at this step rather than the surface of the entire shrink-wrapped item. Since the distance and brightness information changes significantly at the boundary of the surface of each item, it is effective to use this information to seg-

ment the area. However, when items are densely piled up, the distance between the gaps of the items may not be measured correctly; thus, multiple items may be recognized as one object surface. If there is a pattern on the item, the luminance gradient near the pattern also increases, making it difficult to distinguish between the pattern and object boundary. Therefore, we estimate an edge region via the following values v that emphasizes the boundary by combining the magnitudes of both types of information,

$$v_{ij} = l_{ij}^{2\alpha} * c_{ij}^{2(1-\alpha)} \quad (1)$$

where i, j are the x and y coordinates of the pixel, l is the magnitude of the vector consisting of luminance gradients in the x and y directions, c is the value of curvature at each pixel, and α is a parameter that adjusts the ratio between the luminance gradient and curvature. We then apply the watershed algorithm (Vincent et al., 1991) on the region bounded by the

identified edges to obtain segmented regions, which are regarded as the object surfaces.

3.2.2 Identification of Boxed and Shrink-Wrapped Object Surfaces

Next, we identify whether each of the extracted surfaces mentioned in 3.2.1 is the surface of a boxed item or that of each product in a shrink-wrapped item (Fig. 4(b)). While most items are large rectangular shapes and can be recognized with conventional methods, shrink-wrapped items containing plastic bottles or cans, which have small circular top surfaces, are difficult to be recognized with such methods. Therefore, the extracted surfaces larger than the threshold are identified as the surfaces of boxed items, and those smaller than the threshold are identified as the surfaces of the products in shrink-wrapped items. The circle extraction process proposed in (Yuen et al., 1990) is further applied to the surfaces that belong to products in a shrink-wrapped item, and only the area overlapping the detected circle area is considered as its surface area. This process eliminates small noise areas.

3.2.3 6DoF Pose and Size Estimation

In this step, we obtain the 6DoF poses and sizes of boxed and shrink-wrapped items. Surfaces belonging to products in the shrink-wrapped item mentioned in 3.2.2 are estimated for their 6DoF poses and sizes using the method described in 3.1. Since there is no need to obtain a surface graph for boxed items, we carry out planar fitting directly on each surface that belongs to a boxed item, as mentioned in 3.2.2, to obtain its 6DoF pose and size. In this study, we calculated the 6DoF poses of all items except shrink-wrapped items using planar fitting, but it is also possible to estimate the 6DoF pose by applying the RANSAC algorithm assuming cylindrical or conical shapes.

4 EXPERIMENTS AND RESULTS

4.1 Dataset

We conducted an experiment to evaluate the performance of the proposed method. The experiments were conducted under the assumption of recognition in a mixed-depalletizing operation, as shown in the left figure of Fig. 5. We used TVS4.0, a 3D vision head with 2 cameras and an industrial projector, with a resolution of 1280 x 1024. An example of an actual image is shown in the right figure.



Figure 5: Experimental setup. Left figure shows experimental environment and right figure shows example scene image (with each image blurred).

We evaluated the recognition rate of the proposed method for a total of 175 scenes, which were captured by changing the set of items in the captured image. The 25 types of items to be recognized are shown in Fig. 6. There existed overlap between boxed and shrink-wrapped items, i.e., one item can belong to both categories. Since the purpose of this experiment was to evaluate the recognition accuracy of the size, position, and orientation of each item, we did not evaluate the category-identification accuracy. The depth distance between the camera and top surface of the item was 2.4–3.2 m. We compared the proposed method with a conventional edge-based boundary-detection method.

4.2 Definition of Success Rate

We defined a successful recognition as the correct estimation of the 6DoF pose of an item without occlusion (i.e., part of the item is hidden by other items and cannot be seen) and used the ratio of scenes in which the item was successfully recognized as the recognition rate. The success or failure of the 6DoF pose was evaluated in accordance with the evaluation method (Hodañ et al., 2016) and defined as success when the intersection over union score (IoU) of the 2D bounding rectangle of the estimated 6DoF pose projection was 0.5 or more.

4.3 Results and Discussion

Figure 7 shows the recognition results of the proposed method. The proposed method correctly extracted each surface area (Fig. 7(a)) and correctly classified the area into either boxed item or shrink-wrapped item in accordance with its size (Fig. 7(b)). The surfaces that were recognized as shrink-wrap items were correctly divided into two shrink-wrapped-item areas on the basis of the distribution of the center-to-center distances (Fig. 7(c)). For the surfaces that were recognized boxed items, the correct 6DoF pose of each item was successfully estimated by fitting a plane to a set



Figure 6: Target items. Each item is classified into eight categories, and corresponding item is shown in each column.

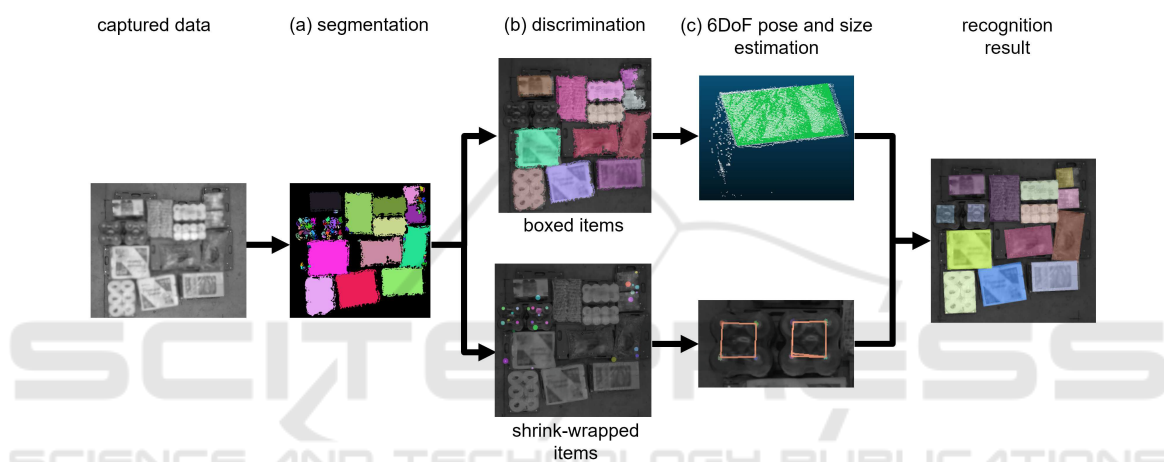


Figure 7: Results of proposed method. Each image shows result of each processing stage with proposed method (with each image blurred). Proposed method recognized items in variety of packaging forms by dividing the items into boxed and shrink-wrapped and applying different methods. It also correctly recognized boundaries of shrink-wrapped items even when such items of the same type were adjacent to each other by estimating boundaries on basis of regularity of distances between products.

of points corresponding to the estimated item area, as with the conventional method (Fig. 7(c)).

Table 1 shows the recognition rate for each item shown in Fig. 6 using the evaluation criteria defined in 4.2. The conventional method had a high recognition rate for boxed items, while the recognition rate for objects with complex shapes such as bottles significantly decreased. The proposed method had a high recognition rate for boxed items as well as bottles, with an average recognition rate of more than 90%, which is higher than that of the conventional method. This shows that the proposed method is capable of recognizing both boxed and shrink-wrapped items.

Figure 8 shows the results of failed recognition with the proposed method. When the shrink-wrapped items were closely aligned without any misalignment, there was insignificant difference in the distance between the products within the shrink-wrap items and

between the shrink-wrap items. Therefore, the proposed method recognized two shrink-wrapped items as one large shrink-wrapped item. When the boundary of the shrink-wrapped items cannot be determined by only the distance between products, it is necessary for the robotic system to grasp the edge of the shrink-wrapped item and shift it to create a gap then recognize the boundary again in the scene where the boundary is clear.

The selected parameters can affect the performance of the proposed method. Figure 9 shows the difference in recognition results when recognition is executed with two different parameters in the recognition process. The parameter that correctly detects the boundary of a large boxed item and the parameter that detects a very small area, such as a plastic bottle cap, are often different. Therefore, there are potentially cases in which one fixed set of the pa-

Table 1: Recognition rate.

object	Edge-based boundary detection (Conventional method)	Proposed method
Case	97.0% (32 / 33)	97.0% (32 / 33)
Box	88.1% (214 / 243)	87.7% (213 / 243)
Boxes	71.0% (233 / 328)	70.4% (231 / 328)
Grains	97.7% (208 / 213)	97.7% (208 / 213)
Packs	95.4% (313 / 328)	95.4% (313 / 328)
Paper rolls	97.4% (150 / 154)	97.4% (150 / 154)
Cylinders	95.7% (400 / 418)	95.7% (400 / 418)
Bottles	67.6% (294 / 435)	91.3% (397 / 435)
All	85.7% (1844 / 2152)	90.3% (1944 / 2152)

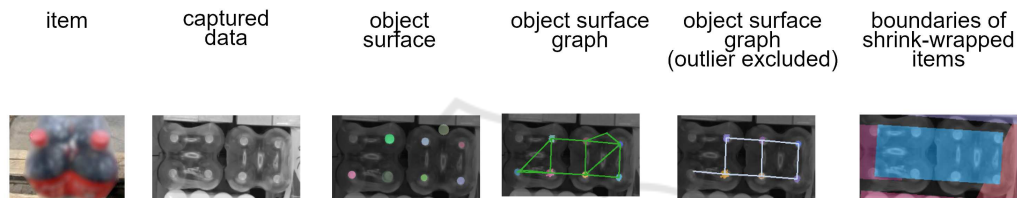


Figure 8: Results of proposed method (failure case). Columns show results of each processing stage in estimating boundaries of shrink-wrapped items. (with each image blurred). When shrink-wrapped items are closely aligned without misalignment, it is difficult to estimate boundaries of shrink-wrapped items on basis of regularity of distances between products alone.

rameters cannot be utilized to recognize all types of items simultaneously. However, we believe that such a problem can be solved by incorporating an automatic local-parameter adjustment process that selects locally optimal parameters for each detected surface.

5 CONCLUSION

We proposed a surface-graph-based 6DoF object-pose estimation method that enables 6DoF pose and size estimation of shrink-wrapped items for mixed depalletizing processes. The proposed method constructs graphs that connect the product surfaces corresponding to each shrink-wrapped item on the basis of the distribution of the center-to-center distance of the products in the shrink-wrapped item, enabling accurate boundary detection of shrink-wrapped items and the estimation of their sizes and 6DoF poses, which are difficult when using conventional object recognition methods. To correctly recognize each item even in an environment where boxed and shrink-wrapped items are mixed, we also developed a recognition-process flow that switches between an edge-based boundary-detection method for boxed items and the proposed method for shrink-wrapped items depending on the size of the item surface.

In the experiment that simulated mixed depalletizing operations, we confirmed that the proposed method could correctly recognize 6DoF poses of both boxed and shrink-wrapped items and achieve an average recognition rate of over 90%, which is higher than that of a conventional method.

By applying the proposed method to recognize individual items then having the robot pick and unload them, an effective robot-picking system can be implemented in a depalletizing work environment even when boxed and shrink-wrapped items are mixed.

Our future work will include incorporating a robot-assisted object-displacement process for cases in which shrink-wrapped items are in close contact with each other, which makes it difficult to distinguish the boundary of each item. We also aim to improve the recognition accuracy of the proposed method by selecting locally optimal parameters for each item in the scene.

ACKNOWLEDGEMENTS

We are grateful to Mr. Takaharu Matsui for his useful discussions with us. We are also grateful to Mr. Kento Sekiya and Mr. Koichi Kato for their cooperation in implementing the software for the proposed method.

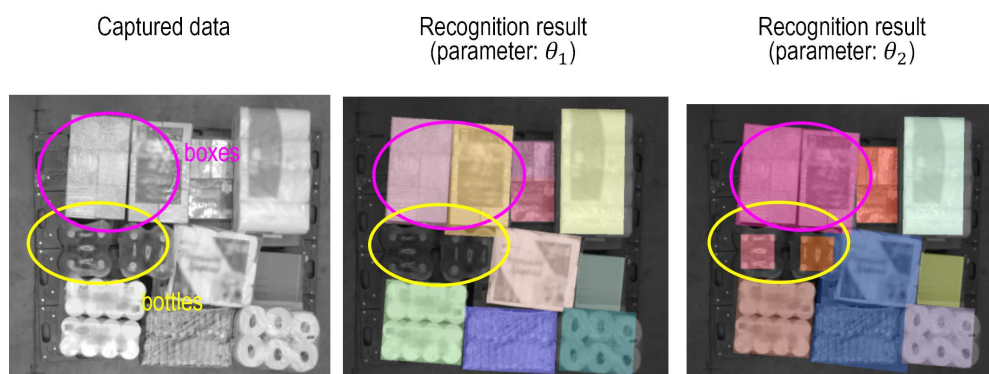


Figure 9: Difference in recognition performance due to parameter setting. Each image is recognition result for same scene but with different parameters (with each image blurred). Parameters that correctly capture boundaries of large boxed items and those that detect very small areas, such as plastic bottle caps, often differ, and there are cases in which it is not possible to set parameters that allow all objects to be recognized simultaneously.

We are also grateful to Mr. Daisuke Katsumata and Mr. Tsubasa Watanabe for their cooperation in obtaining the experimental data.

REFERENCES

- Aleotti, J., Baldassarri, A., Bonfè, M., Carricato, M., Chiaravalli, D., Di Leva, R., Fantuzzi, C., Farsoni, S., Innero, G., Rizzini, D. L., Melchiorri, C., Monica, R., Palli, G., Rizzi, J., Sabbatini, L., Sampietro, G., and Zaccaria, F. (2021). Toward future automatic warehouses: An autonomous depalletizing system based on mobile manipulation and 3d perception. *Applied Sciences (Switzerland)*, 11(13).
- Doliotis, P., McMurrugh, C. D., Criswell, A., Middleton, M. B., and Rajan, S. T. (2016). A 3D perception-based robotic manipulation system for automated truck unloading. *IEEE International Conference on Automation Science and Engineering*, 2016-Novem:262–267.
- Eto, H., Nakamoto, H., Sonoura, T., Tanaka, J., and Ogawa, A. (2019). Development of automated high-speed depalletizing system for complex stacking on roll box pallets. *Journal of Advanced Mechanical Design, Systems and Manufacturing*, 13(3):1–12.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- Fuji, T., Kimura, N., and Ito, K. (2015). Architecture for recognizing stacked box objects for automated warehousing robot system. In *Proceedings of the 17th Irish Machine Vision and Image Processing conference*.
- Gou, L., Wu, S., Yang, J., Yu, H., Lin, C., Li, X., and Deng, C. (2021). Carton dataset synthesis method for domain shift based on foreground texture decoupling and replacement. *arXiv preprint arXiv:2103.10738*, (Xi-aoping Li).
- He, Y., Sun, W., Huang, H., Liu, J., Fan, H., and Sun, J. (2020). PVN3D: A deep point-wise 3D keypoints voting network for 6DoF pose estimation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 11629–11638.
- Hodaň, T., Matas, J., and Obdržálek, Š. (2016). On evaluation of 6d object pose estimation. In *European Conference on Computer Vision*, pages 606–619. Springer.
- Holz, D., Ichim, A. E., Tombari, F., Rusu, R. B., and Behnke, S. (2015). Registration with the point cloud library: A modular framework for aligning in 3-d. *IEEE Robotics & Automation Magazine*, 22(4):110–124.
- Katsoulas, D., Bergen, L., and Tassakos, L. (2002). A versatile depalletizer of boxes based on range imagery. *Proceedings - IEEE International Conference on Robotics and Automation*, 4(May):4313–4319.
- Katsoulas, D. K. and Kosmopoulos, D. I. (2001). An efficient depalletizing system based on 2D range imagery. *Proceedings - IEEE International Conference on Robotics and Automation*, 1:305–312.
- Kehl, W., Manhardt, F., Tombari, F., Ilic, S., and Navab, N. (2017). SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-October:1530–1538.
- Kimura, N., Ito, K., Fuji, T., Fujimoto, K., Esaki, K., Beniyama, F., and Moriya, T. (2016). Mobile dual-arm robot for automated order picking system in warehouse containing various kinds of products. *2015 IEEE/SICE International Symposium on System Integration, SII 2015*, pages 332–338.
- Kirchheim, A., Burwinkel, M., and Echelmeyer, W. (2008). Automatic unloading of heavy sacks from containers. *Proceedings of the IEEE International Conference on Automation and Logistics, ICAL 2008*, (September):946–951.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.
- Mitash, C., Wen, B., Bekris, K., and Boularias, A. (2020). Scene-level pose estimation for multiple instances of densely packed objects. In *Conference on Robot Learning*, pages 1133–1145. PMLR.

- Monica, R., Aleotti, J., and Rizzini, D. L. (2020). Detection of Parcel Boxes for Pallet Unloading Using a 3D Time-of-Flight Industrial Sensor. *Proceedings - 4th IEEE International Conference on Robotic Computing, IRC 2020*, pages 314–318.
- Nakamoto, H., Eto, H., Sonoura, T., Tanaka, J., and Ogawa, A. (2016). High-speed and compact depalletizing robot capable of handling packages stacked complicatedly. *IEEE International Conference on Intelligent Robots and Systems*, 2016-Novem:344–349.
- Naumann, A., Dorr, L., Ole Salscheider, N., and Furmans, K. (2020). Refined Plane Segmentation for Cuboid-Shaped Objects by Leveraging Edge Detection. *Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, pages 432–437.
- Poss, C., Ibragimov, O., Indreswaran, A., Gutsche, N., Irrenhauser, T., Prueglmeier, M., and Goehring, D. (2019). Application of open Source Deep Neural Networks for Object Detection in Industrial Environments. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, pages 231–236.
- Stein, S. C., Schoeler, M., Papon, J., and Worgotter, F. (2014). Object partitioning using local convexity. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (June):304–311.
- Stoyanov, T., Vaskevicius, N., Mueller, C. A., Fromm, T., Krug, R., Tincani, V., Mojtahedzadeh, R., Kunaschk, S., Mortensen Ernits, R., Canelhas, D. R., Bonilla, M., Schwertfeger, S., Bonini, M., Halfar, H., Pathak, K., Rohde, M., Fantoni, G., Bicchi, A., Birk, A., Lilienthal, A. J., and Echelmeyer, W. (2016). No More Heavy Lifting: Robotic Solutions to the Container Unloading Problem. *IEEE Robotics and Automation Magazine*, 23(4):94–106.
- Tang, J., Miller, S., Singh, A., and Abbeel, P. (2012). A Textured Object Recognition Pipeline for Color and Depth Image Data. *2012 IEEE International Conference on Robotics and Automation*, pages 3467–3474.
- Vincent, L., Vincent, L., and Soille, P. (1991). Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598.
- Wang, H., Sridhar, S., Huang, J., Valentin, J., Song, S., and Guibas, L. J. (2019). Normalized object coordinate space for category-level 6d object pose and size estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2642–2651.
- Xiang, Y., Schmidt, T., Narayanan, V., and Fox, D. (2018). PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. *arXiv preprint arXiv:1711.00199*.
- Yang, J., Wu, S., Gou, L., Yu, H., Lin, C., Wang, J., Wang, P., Li, M., and Li, X. (2022). SCD: A Stacked Carton Dataset for Detection and Segmentation. *Sensors*, 22(10).
- Yano, T., Hagihara, D., Kimura, N., Chihara, N., and Ito, K. (2022). Parameterized B-rep-Based Surface Correspondence Estimation for Category-Level 3D Object Matching Applicable to Multi-Part Items. In *2022 IEEE 18th International Conference on Automation Science and Engineering*, pages 607–614.
- Yuen, H., Princen, J., Illingworth, J., and Kittler, J. (1990). Comparative study of hough transform methods for circle finding. *Image and vision computing*, 8(1):71–77.