Individual recognition of food in bulk by using 3D model of food

Yuya Otsu *, Tokuo Tsuji **, Tatsuhiro Hiramitsu, Hiroaki Seki
Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan

*Email: *otsuy0525@stu.kanazawa-u.ac.jp, **tokuo-tsuji@se.kanazawa-u.ac.jp*

Abstract

In this paper, we propose a method of individual recognition of food in bulk by using 3D model of food. First, color images and depth images of them are generated by using 3D model of food and physics engine of simulator. Then, color and depth composite images are created by converting two channels from color images and one channel from depth images. In the experiments, the accuracy of individual recognition of food in bulk with color and depth composite images are shown to compare the accuracy with only color images.

Keywords: Instance segmentation, Color space, 3D model

1. Introduction

In the food manufacturing industry, individual recognition of food in bulk on an image plays a crucial role in quality control and determining the gripping positions for topping handlers. Instance segmentation has been proposed as a method of image recognition and it generally uses supervised learning. In supervised learning, it is necessary to prepare many training data to improve learning accuracy. And, since training data is usually prepared manually, it is burdensome to prepare many training data. Therefore, as an efficient training data generation method for food recognition, such as generating composite images of food by combining multiple images of individual food [1][2] have been proposed. However, accurate segmentation of individual's region with only RGB image is difficult in the case of single type food because the boundaries between individual food become ambiguous when it is stacked in bulk.

To solve this issue, we propose a method to automatically generate many color images and depth images of food in bulk in a short time using 3D models of food and physics engine of simulator, and to generate composite image that include both color and depth information for individual recognition of food in bulk. Firstly, food models captured by a 3D scanner are placed on a virtual space and obtain color image and depth image by shooting from a virtual viewpoint. Then, RGB channels of the color image are converted to the Luv color space, and the two channels (u and v) are reduced to a single channel using principal component analysis (PCA). Finally, color and depth composite images are created by combining L channel, the new channel obtained PCA, and the single channel of depth image.

In this study, we propose the individual recognition of food in bulk with color and depth composite images created using 3D model of food. In experiments, we confirmed the accuracy of individual recognition of food in bulk with color and depth composite images to compare the accuracy with only color images.

2. Methodology

The flow of the color and depth composite image generation is shown in Fig. 1. The RGB channels of the color image generated by using 3D models of food in virtual space are converted to the Luv color space. The channel obtained by reducing the dimensionality of the two channels, u and v, is referred to as C, the depth image generated by using the distance information from the camera to each food is referred to as D, and the combined channels of color and depth are referred to as LCD.

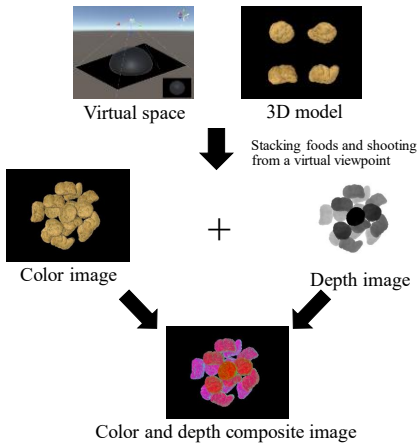


Fig. 1 Flow of color and depth composite image generation.

2.1. Generating 3D model of food by 3D scanner

In this study, we assume a chicken nugget stacked and create 3D models of four different shapes of chicken nuggets. A SCANDIMENTION SOL 3D scanner [3] (Fig. 2) is used for this purpose. Each chicken nugget is placed on the scanner's rotating platform and rotated inside a light-shielding tent while being scanned with a laser to capture the mesh and texture. The 3D models of the chicken nugget obtained are shown in Fig. 3.

Fig. 2 SOL 3D scanner [3]

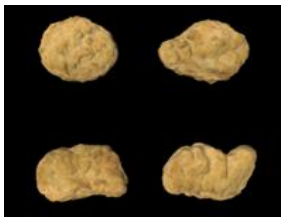


Fig. 3 3D models of chicken nuggets

2.2. Generating color image using virtual space

Color image of food in bulk is generated by arranging the 3D models of food in a virtual space using a physics engine of simulator and capturing them from virtual viewpoints. In this study, Unity [4] is used as a virtual space. A transparent hemispherical mold is placed in the Unity space. 3D models of food are randomly generated within the mold and stacked in bulk along the mold. An example of the Unity space used for stacking food and the color image generated is shown in Fig. 4.

2.3. Generating depth image

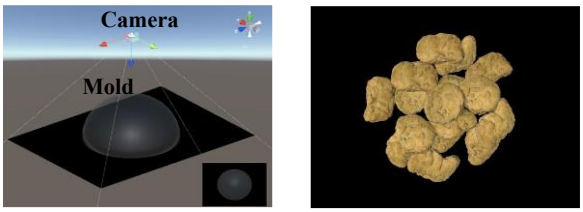
After stacking the food in the Unity space, the distance information from the camera to each food is obtained. The acquired values are normalized by setting the distance from the camera to the top of the mold as 0 and the distance from the camera to the ground as 1. Then the normalized values are applied to the materials of each food and the scene is captured to generate a depth image. An example of depth image generated is shown in Fig. 5.

2.4. Dimensionality reduction method for color channel

The RGB channels of the color image generated in section 2.2 are converted to the Luv color space. The Luv color space is defined to equalize color differences and is expressed using a three-dimensional orthogonal coordinate system with L (lightness), u, and v (chromaticity) as its axes [5]. After converting to Luv, the two channels, u and v, are reduced to a single channel using principal component analysis, resulting in the C channel. Fig. 6 shows the uv color map of the color image and the results of dimensionality reduction. Also, Fig. 7 shows the grayscale image with the L channel and the C channel obtained by dimensionality reduction.

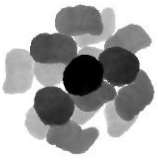
2.5. Compositing color and depth channels

By combining the L and C channels generated in Section 2.4 with the D channel generated in Section 2.3, a composite image of color and depth (LCD image) is obtained. An example of the generated LCD image is shown in Fig. 8.



Unity space Color image

Fig. 4 Generate method of color image



Depth image

Fig. 5 Example of depth image of food in bulk

3. Food recognition using composite images as training data

This section describes the specific details of the instance segmentation used for food recognition and the generation of training data using composite images.

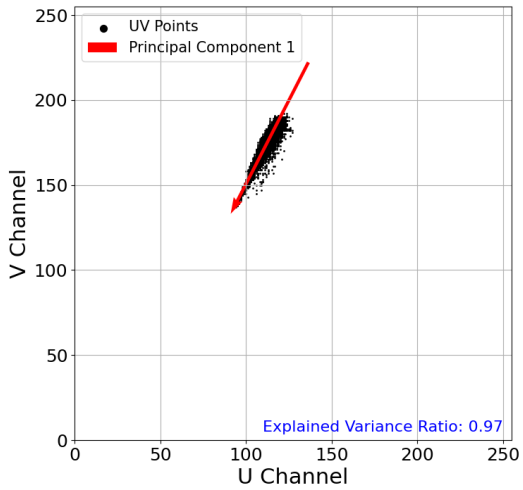
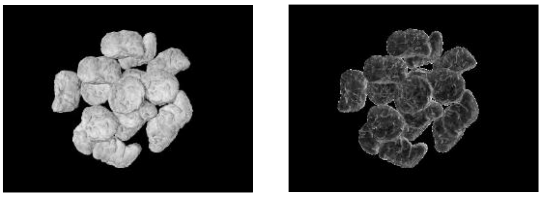
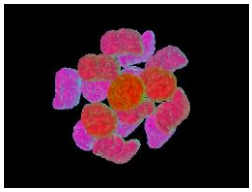


Fig.6 Example of uv color map and result of principal component analysis



L channel image C channel image

Fig.7 Example of L channel image and C channel image (gray scale)



LCD image

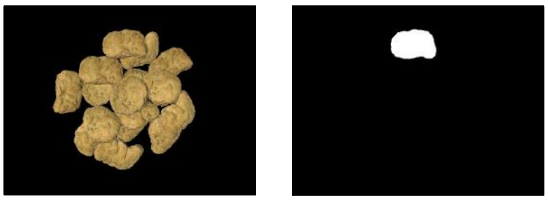
Fig.8 Example of color and depth composite image

3.1. Instance segmentation

Instance segmentation is a process of detecting the class of each object and extracting its region pixel by pixel. The machine learning used in this study is supervised learning. It is a method that uses pre-labeled input data for training and builds a model that can predict labels for unknown input data. In this study, we use YOLO-v7[6], a deep learning model for object detection and segmentation.

3.2. Generation of training datasets using composite images

In the dataset for instance segmentation used in this study, the area information of each object in the image called the COCO format [7] is often represented by the contour information of them. In this method, 3D models of food are stacked in the Unity space, and a mask image is generated for each food models by changing the texture of them to white and that of the background to black. An example of a mask image corresponding to the color image is shown in Fig. 9.



Color image Mask image

Fig.9 Example of color image and mask image

4. Evaluation experiments and results

4.1. Experiment Preparation

4.1.1 Types of food for experiment

Assuming a stacking of chicken nuggets in bulk, four chicken nuggets of four different shapes each are prepared.

4.1.2 Capture of validation images (color images and depth images)

The camera was an Intel RealSense D435, 50 cm away from the ground, under fluorescent light, and with forward light. 10 images were taken as validation images. As in section 2.3, the depth images are normalized by setting the distance from the camera to the top of the stack to 0 and the distance from the camera to the ground to 1, and the depth images are generated in grayscale. An example of the image of a chicken nugget stacked and the normalized depth image are shown in Fig. 10.

4.1.3 Generation of training images (color images)

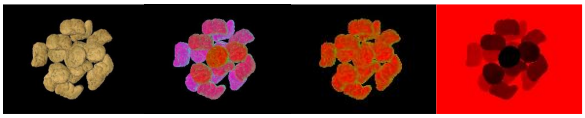
The same number of 3D models of food used to take the validation images are randomly arranged in a virtual space and the color image is generated by shooting from the same point of view as the location where the validation image was taken.

4.1.4 Creation of COCO-format dataset for validation

Since the validation images are photographs and cannot be used with the method in 3.2, the annotation tool Roboflow [8] is used to create the dataset.

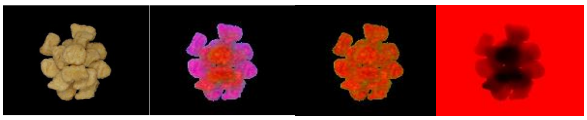
Color image Depth image

Fig.10 Example of color image and depth image for test data



RGB LCD LC D

Fig.11 Example of training data



RGB LCD LC D

Fig.12 Example of validation data

4.2. Experimental conditions

For the stacked chicken nuggets, learning is performed using the following four types of composite images as the training data. The learning conditions are summarized in Table 1. For the LC images and D images, a 3-channel image with 0 in the empty channels is used as the training data. An example of the datasets used in each learning for the experiment are shown in Fig. 11 and Fig. 12. The common training conditions are: model: YOLO-v7[6], number of training data: 500, number of validating data: 10, number of epochs: 300, batch size: 16

4.3. Evaluation method

The value of AP (Average Precision) when IoU (Intersection over Union) is set as the threshold is used to evaluate the performance of each learning model. IoU is a value between 0 and 1 that represents the degree of overlap between the predicted region of an object and the ground truth region. And, AP is the integral value of the P-R curve, which plots Precision against Recall.

Table 1. Each learning condition

	Training data and validation data
Learning A	RGB images
Learning B	LCD images
Learning C	LC images
Learning D	D images

4.4. Experimental results and discussion

The results of AP for each learning are shown in Fig. 13. AP (50) represents the AP at an IoU threshold of 0.5, AP (50-95) shows the average of AP when the IoU threshold is varied from 0.5 to 0.95. Also, for learning A and B, a part of the model's predictions on the validation images is shown in Fig. 14. Fig. 13 shows that the model learned on the LCD image is 8.8% more accurate than that on the RGB image for AP (50), confirming that the model learned with the LCD image is more accurate than that with the RGB image. Similarly, for AP (50-95), the LCD image was 15.4% more accurate than the RGB image, confirming that the model learned on the LCD image was more accurate than the model learned on the RGB image, even when the IoU threshold was increased.

Also, Fig. 14 shows that the output from the model learned with the RGB images segmented the overlapping parts as the same individual, whereas the output from the model learned with the LCD images improved the segmentation accuracy. It is thought that the difference in depth caused by the overlapping of foods was learned by the LCD image, which improved the recognition accuracy. These results confirm the effectiveness of using color and depth composite images as training data for individual recognition of a single type of food in bulk.

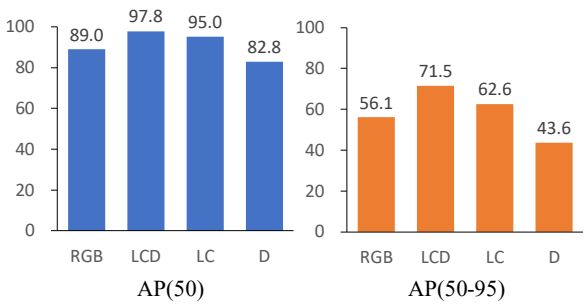


Fig.13 Result of AP for each learning condition

Answer RGB LCD

Fig.14 Result of segmentation for learning condition A and B (part of food)

5. Conclusion

In this study, we proposed a method of generating a lot of color image and depth image of food in bulk automatically by using 3D models of food and generating color and depth composite image. And the effectiveness of using the composite images as training data was demonstrated by using them as training data for instance segmentation.

References

1. T. Yamabe, T. Ishichi, T. Tsuji, T. Hiramitsu, and H. Seki, “Training Data Augmentation for Semantic Segmentation of Food Images Using Deep Learning,” International Conference on Artificial Life and Robotics, 2022
2. T. Ishichi, T. Yamabe, T. Tsuji, T. Hiramitsu, and H. Seki, “Ingredient segmentation with transparency,” IEEE/SICE International Symposium on System Integration (SII), 2023, pp. 1-5.
3. "SCANDIMENSION,SOL3DScanner,"https://scandimension.com/products/sol-3d-scanner
4. "Unity," https://unity.com
5. Y. Ohno, “CIE fundamentals for color measurements,” In NIP & Digital Fabrication Conference, Society of Imaging Science and Technology, vol. 16, 2000, pp. 540-545.
6. C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 7464-7475.
7. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, ... and C. L. Zitnick, “Microsoft coco: Common objects in context,” In Computer Vision–ECCV 13th European Conference, vol. 5, 2014, pp. 740-755
8. "Roboflow,"https://roboflow.com

Authors Introduction

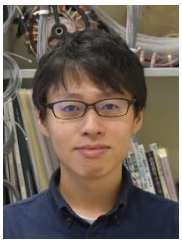
Mr. Yuya Otsu

He received his B.S. degree in engineering from Kanazawa University, Japan, in 2023. He is currently a master’s degree student in the Division of Frontier Engineering, Kanazawa University. His research interest includes food recognition with machine learning

Dr. Tokuo Tsuji

He received his BS, MS, and doctoral degrees from Kyushu University in 2000, 2002, and 2005, respectively. He worked as a research fellow of Graduate School of Engineering, Hiroshima University, from 2005 to 2008. He worked as a research fellow of Intelligent Systems Research Institute of National Institute of Advanced Industrial Science and Technology (AIST) from 2008 to 2011. From 2011 to 2016, he worked as a research associate at Kyushu University. From 2016, he has been working as an associate professor at Institute of Science and Engineering, Kanazawa University. His research interest includes multifingered hand, machine vision, and software platform of robotic systems.

Dr. Tatsuhiro Hiramitsu



He is assistant professor of Institute of Science and Engineering, Kanazawa University. He received Dr E. degrees from school of engineering, Tokyo Institute of Technology, Japan, in 2019. His research interest is in the soft structure mechanisms for robotic systems. He is a member of the Japan Society of Mechanical Engineers (JSME), the Robotics Society of Japan (RSJ), and Institute of Electrical and Electronics Engineers (IEEE).

Dr. Hiroaki Seki

He received his Ph.D. in precision machinery engineering from the University of Tokyo in 1996. He is currently a professor of Institute of Science and Technology in Kanazawa University. His research interests include novel mechanism and sensor system in robotics and mechatronics.