Pose Detection for Flexible-Indefinite Objects using Pseudo-Bone Data

Yuma Yoshimoto
Department of Creative Engineering, National Institute of Technology (KOSEN), Kitakyushu College,
5-20-1, Shii, Kokura-Minami-ku, Kitakyushu-shi, Fukuoka, Japan

Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan

E-mail: yoshimoto@kct.ac.jp
https://researchmap.jp/yoshimoto116

Abstract

This paper proposes a method for recognizing the poses of a flexible-indefinite object. Some flexible-indefinite objects, such as fried shrimp, differ between individuals. Therefore, it is difficult to estimate the pose of these foods by point cloud fitting or other methods. We propose "pseudo-skeleton" data for these objects. Pseudo-skeleton data consists of "key-points," which are joints, and "bones," which connect between key-points. For example, fried shrimp are given 3 key-points; "head," "belly," and "tail." In addition, the bones that connect them are given. In the experiment, objects pose recognition model based on the human pose recognition model trains pseudo-skeleton data of fried shrimp. We confirmed that the model estimates the poses in the images.

Keywords: Pseudo-skeleton, Flexible-indefinite objects, Pose estimation model, Key-points and bones

1. Introduction

The population decline in recent years has introduced robots into factories. The industrial field is especially true. However, the food industry, such as making lunch boxes, is handled because the robots need to grasp some flexible-indefinite objects, such as fried shrimp. It is not easy to estimate the posture of flexible-indefinite objects.

In a lunch factory, food ingredients are stacked in bulk. To recognize industrial objects stacked in bulk, the robots generally fit the design data, such as computer aided design (CAD) data, to objects in bulk. However, we cannot prepare the design data of food ingredients. The method is not suitable for the situation. As a method for such cases, camera data is input into a neural network, and the grasping point is estimated directly. However, this method is unsuitable for cases such as foods that are required to dish up. The reason is that robots cannot estimate the posture of foods using the method.

In this study, we propose a method for estimating the posture of flexible-indefinite objects such as food ingredients using pose estimation neural network models for humans. In the methods, the structures of flexible-indefinite objects are described, such as "pseudo-skeleton," using a combination of key-points and bones.

2. Related Works

2.1 *Skeleton Data*

Skeleton data consists of "key-points," which are joints, and "bones," which connect between key-points. Figure 1 shows an example of human skeleton data. In the data, head, hand, arm, and etcetera, are key-points. In addition, bones connect the key-points.

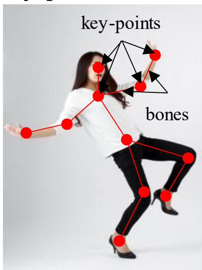

Fig. 1. Human skeleton data

2.2 *Pose Estimation Neural Network Models*

Pose estimation models are generally used to estimate the human pose. These methods are two types: bottom-up type [1] and top-down type [2]. Bottom-up type first search all key points in an image. Then, there are matched and connected for each person. The top-down type first detects people using an object detection algorithm. Then, the models estimate key-points and bones for each person. The latter is more robust to occlusion and motion blur than the former. The latter is used in this paper.

3. Proposed Methods

In this study, we apply the pose estimation method, which is generally used for humans, to flexible-indefinite objects, such as food ingredients. In the method, the "pseudo-skeleton" of these objects are defined for estimation poses. Fig. 2. Fried shrimp's pseudo-skeleton shows an example of this method. In the figure, the proposed method is applied to fried shrimp. In this example, "head," "belly," and "tail" are defined as key-points. The connections between the head and belly and between the belly and tail are defined as bones.

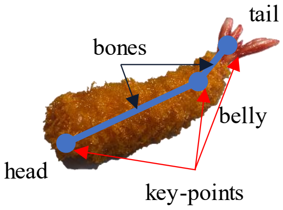

Fig. 2. Fried shrimp's pseudo-skeleton

4. Experiments

4.1 *Setup*

The experiment was conducted to evaluate the proposal. We prepared 325 pairs of fried shrimp's images, pseudo-skeleton data, and bounding boxes. Figure 3 shows an example of datasets. The 300 pairs were used as training data, and 25 pairs as test data. In addition, a combination of ResNet50 [3] and Top-down Heatmap Simple Head was used as a pose estimate model [2].

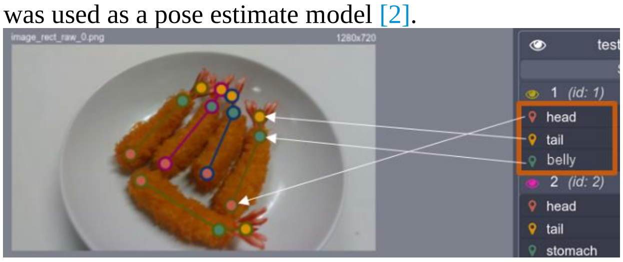

Fig. 3. Datasets

4.2 *Results and Discussion*

Table 1 shows the experimental results. We calculated the mean average precision (mAP), which is used to confirm dataset effectiveness. The same model for estimating the human poses, is included as a comparison [2]. The results show that the proposed method can estimate the pseudo-skeleton of flexible-indefinite objects at the same level as the human skeleton. Figure 4 shows an image of the estimate results. The result shows that the pseudo-skeleton of a fried shrimp can be inferred.

Table 1. Experimental results

Training Data	mAP
Fried Shrimp's Pseudo-Skeleton (Ours)	**0.795**
Human Skeleton [2]	0.724

Fig. 3. Estimate results

5. Conclusion

We propose a pseudo-skeleton method for the pose estimate of flexible-indefinite objects in the paper. In the proposed method, the "pseudo-skeleton" of these objects are defined for estimation poses. Experimental results show that the model, which was trained pseudo-skeleton, can estimate the posture of flexible-indefinite objects at the same level as a human skeleton. In the future, we will incorporate this system into a robot to verify grasping.

References

1. J. Li, S. Bian, A. Zeng, C. Wang, B. Pang, W. Liu, C. Lu, "Human Pose Regression with Residual Log-likelihood Estimation," 2021 International Conforence on Computer Vision (ICCV 2021), pages 11025-11034, 2021.
2. B. Xiao, H. Wu, Y. Wei, "Simple Baselines for Human Pose Estimation and Tracking," 2018 European Conference on Computer Vision (ECCV 2018), pages 466-481, 2018.
3. K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pages 770-778, 2016.

Authors Introduction

Asst. Prof. Yuma Yoshimoto

He received his B.Eng. degree from National Institute of Technology (KOSEN), Maizuru College, Japan, in 2016. He received his M.Eng. and D. Eng. degrees from Kyushu Institute of Technology, Japan, in 2018 and 2021, respectively. And he was JSPS researcher, in 2019 - 2021. He was a post-doctoral researcher at the Kyushu Institute of Technology, Japan in 2021-2022. Currently, he is assistant professor at the National Institute of Technology (KOSEN), Kitakyushu College, Japan. His research interests include deep learning, robot vision and digital hardware design. He is a member of IEICE, IEEE.

Prof. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001. He received the M.Eng. and the PhD degree from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of 21st century center of excellent program at Kyushu Institute of Technology, from 2006 to 2007. He was an Assistant Professor of Tokyo University of Agriculture and Technology, from 2007 to 2013. He is currently an Associate Professor in the Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interest includes hardware/software complex system, digital hardware design, neural networks, soft-computing and home service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI and RSJ.