Acceleration of training dataset generation by 3D scanning of objects

Yushi Abe, Yutaro Ishida, Tomohiro Ono, Hakaru Tamukoh
Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Fukuoka-ken, Japan
E-mail: abe.yushi109@mail.kyutech.jp, ishida.yutaro954@mail.kyutech.jp, ono.tomohiro342@mail.kyutech.jp, tamukoh@brain.kyutech.jp
http://www.lsse.kyutech.ac.jp/english/

Abstract

A semi-automatic dataset generation system is effective to prepare a training dataset for object recognition in a personal residence. However, a semi-automatic that method requires significant manual processing to capture images of household objects. Therefore, we apply three-dimensional object scanning to eliminate manual processing and speedup dataset generation. Experimental results demonstrate that the proposed method can generate the dataset 40 minutes faster than a comparable previous method that did not require manual processing.

Keywords: dataset generation, CNN, Home Service Robot, Image recognition

1. Introduction

Recently, home service robots have attracted increasing attention due to the needs of an aging society with a declining birth rate. Home service robots can reduce the burden on people engaged in various types of repetitive and predictable work, such as waiters and housekeepers. The Kyushu Institute of Technology's home service robotics development team participated in the RoboCup@Home[1] League and the World Robot Challenge (WRC)[2] using the human support robot[3] developed by Toyota Motor Corporation. Object recognition technology is required for home service robots and recently using neural network-based deep learning[4] for image recognition has attracted attention. For example, You Only Look Once (YOLO)[5] provides real-time, high-precision object recognition.

To recognize objects in a personal living environment, a dataset that matches the environment must be trained in advance. However, deep learning generally requires a significant amount of data and manual generation of the dataset is time consuming.[6] In addition, manual annotation errors are an issue.

In this study, our objective is to realize a training dataset generation system that can easily recognize domestic objects with high accuracy and thereby reduce the time required to generate the dataset.

2. Previous research

2.1. *Semi-automatic Dataset Generation System*

Our robotics development team has been developing a semi-automatic dataset generation system[7] and using it in competitions.

Figure 1 shows an overview of object data capture using the previously developed semi-automatic dataset generation system. The system uses two RGB-D cameras to capture object images from different orientations. The object's background permeates using chroma key processing to make the background monochromatic. The dataset preparation flow is shown in Figure 2. After capturing the object

© The 2020 International Conference on Artificial Life and Robotics (ICAROB2020), Jan. 13-16, B-Con Plaza, Beppu, Oita, Japan

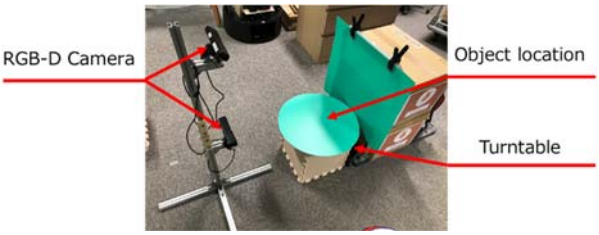

Fig. 1. Overview of object data capture system

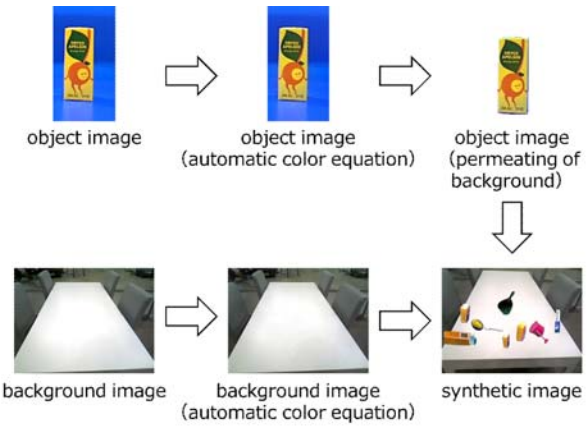

Fig. 2. Dataset generation process flow for the previous system

and background images separately, the automatic color equalization algorithm[8] is applied. Then, the background of the object images permeates. A synthesized image is generated automatically by randomly selecting an object image that has been permeated and synthesized it with the background image. Annotation information is automatically created simultaneously with the synthesized image.

2.2. *Problems*

The previous system uses chroma key and a fixed camera. Consequently, it has the following problems.

(i) The background may not permeate well depending on the direction of light or shadows.

(ii) When objects must be placed in multiple patterns, more time is required to capture the images.

The background color is filtered out by HSV conversion based on the pixel value. The threshold value for filtering is determined by empirical rules. Therefore, as described in problem (i), it may not be possible to completely filter out the background or

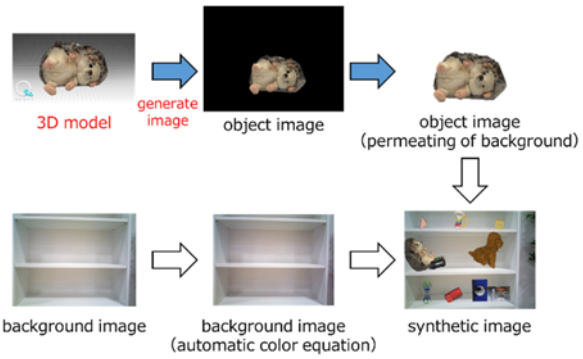

Fig. 3. Dataset generation process flow for the proposed system

the object image was broken. In such a case, the threshold must be adjusted, and the process needs to be repeated. However, threshold adjustment may also fail. In that case, the background must be removed one by one manually. Regarding the second problem (ii), since there are six possible placements on the paper pack, as shown in Figure 2, it is necessary to capture images of all six patterns. The time requires to capture all six patterns increases the time required to generate the dataset. In this paper, we propose a new method for the system used to capture object data.

3. Reducing Image Capture Time Using Three-Dimensional Scanning

In this paper, we propose a method to reduce the time required to capture images by using three-dimensional (3D) scanning. We used the *QLONE*[9] application, which can generate 3D models from scans captured by an iPhone camera. Scans are obtained by placing an object on a dedicated marker and rotating the object. Because the bottom of an object cannot be captured, two patterns of 3D models are created by inverting the top and bottom of each object. Thus, each object requires two or fewer shooting patterns. Multi-viewpoint images are automatically generated from the captured 3D models. At this time, the filtering process can easily be performed by making the background a different color from the color of the object. The dataset generation process flow of this system is shown in Figure 3. Note that only the image capture method differs from the existing system; thus the modified part is represented by the blue arrows in Figure 3.

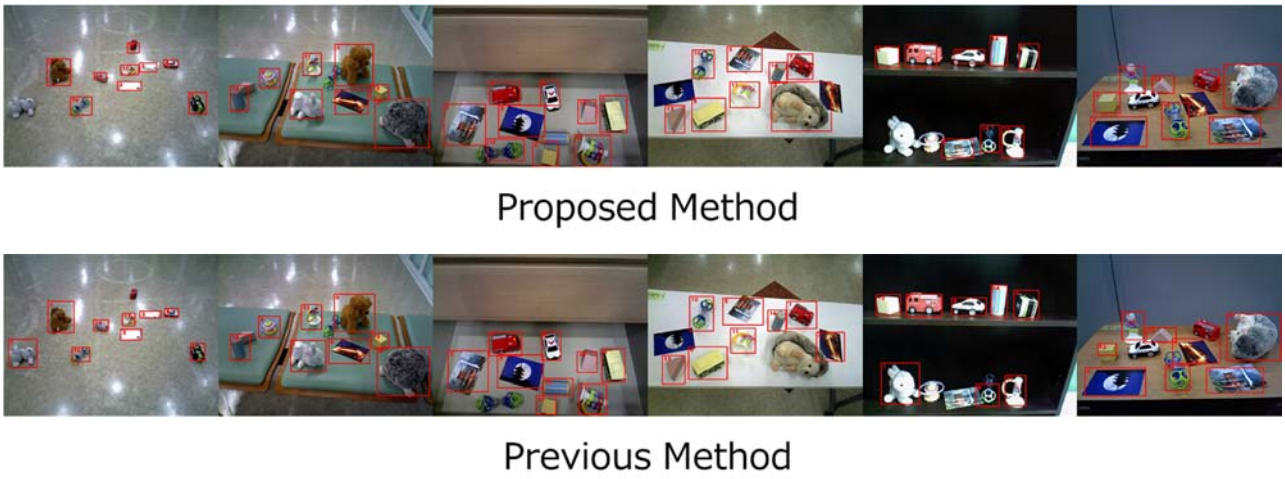

Fig. 5. Recognition result

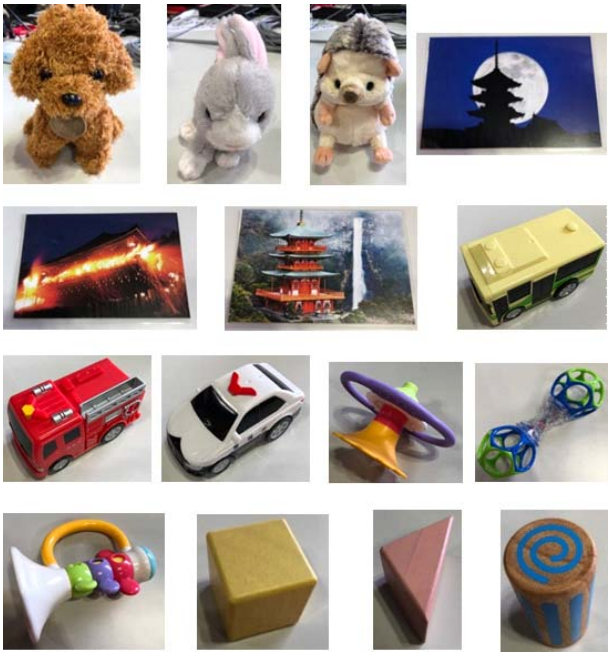

Fig. 4. Objects in WRC

4. Experiments

We compared the time required to generate the dataset using the proposed method and the existing method and the recognition accuracy of a trained YOLO model.

4.1. *Experiment environment*

The experimental environment was as follows.

- Total number of objects: 15 toys (used in WRC's Tidy Up task) shown in Figure 4
- Background image for synthesize: 18 types * 17 (shot in advance)
- Object recognition: YOLOv2
- PC: Intel Core i7-8700K, DDR4 32GB, NVIDIA GTX 1080

4.2. *Comparison of processing speed*

We compared the time duration from capturing the object images until all backgrounds have been filtered out. In the existing system, the total number of patterns to be placed was 39 for 15 objects. Shooting required approximately two minutes per pattern. Therefore, it took approximately 80 minutes to capture the image data, and it took approximately 70 minutes for all backgrounds to be completely removed. Thus, the total time required was approximately 150 minutes. On the other hand, the proposed system requires four minutes per pattern. One pattern is sufficient for objects that do not change even if the image is inverted vertically. Thus, it only takes approximately 100 minutes a total of 26 patterns. In addition, it takes approximately 10 minutes to generate images from 3D models and filter out the background; thus, the total time required is approximately 110 minutes. Note that the filtering process does not need to be done entirely manually. The above results demonstrate that the proposed system was

Table 1. Comparison of performance

Evaluation index	Proposed method	Previous method
Number of data	12600	15600
mAP (%)	60.72	64.77

able to generate a dataset at a faster speed than the existing system.

4.3. *Comparison of recognition accuracy*

The number of datasets generated was approximately 15,600 for the existing method and approximately 12,600 for the proposed method. We compared the recognition accuracy on YOLOv2 that was trained for up to 10,000 epochs using the dataset generated. The recognition results are shown in Figure 5. The results of comparing the proposed method with existing methods are shown in Table 1. We used mean Average Precision (mAP) as the comparison metric. The mAP value obtained by the proposed method was four points less than that of the previous method.

5. Conclusion

In this paper, we have proposed a dataset generation system using 3D scans. We generate a dataset of objects that were used in the WRC and to train YOLOv2. As a result, compared to the previous method, using 3D scans, dataset generation time can be reduced. In addition, the mAP obtained by the proposed method was only four points less than that of the previous method, which is not a significant drop in accuracy. The proposed method to generate images from 3D models can be used to generate datasets for home service robots.

We applied *QLONE* which is iPhone's 3D scanning application. *QLONE* requires a human operator; therefore, the process cannot be completely automated. A future challenge is to construct a system that enables the generation of 3D models and datasets without any human intervention.

Acknowledgements

This research is supported by the New Energy and Industrial Technology Development Organization (NEDO) and JSPS KAKENHI grant number 17H01798.

References

1. T. Wisspeintner, T. van der Zant, L. Locchi, and S. Schiffer, “RoboCup Home: Scientific Competition and Benchmarking for Domestic Service Robots,” Interaction Studies, pp.392-426, 2009.
2. H. Okada, T. Inamura and K. Wada, “What competitions were conducted in the service categories of the World Robot Summit?,” Advanced Robotics, 2019.
3. T. Yamamoto, K. Terada, A. Ochiai and F. Saito, “Development of Human Support Robot as the research platform of a domestic mobile manipulator,” ROBOMECH Journal, 2019.
4. G. E. Hinton, S. Osindero and Y. W. Teh, “A fast learning algorithm for deep belief nets,” Neural Computation, vol.18, no.7, pp.1527-1544, 2006.
5. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Real-Time Object Detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.779-788, 2016.
6. J. Hestness, S. Narang, N. Ardalani, G. Diamos, H.Jun, H. Kianinejad, M. M. A. Patwary, Y. Yang, and Y. Zhou, “Deep Learning Scaling is Predictable, Empirically,” arXiv:1712.00409, 2017.
7. Y. Ishida and H Tamukoh, “Semi-automatic Dataset Generation for Object Detection and Recognition and its Evaluation on Domestic Service Robots,” Journal of Robotics and Mechatronics, vol.32, no.1, 2020.
8. A. Rizzi, C. Gatta and D. Marini, “A new algorithm for unsupervised global and local color correction,” Colour Image Processing and Analysis. First European Conference on Colour in Graphics, Imaging, and Vision (CGIV 2002), pp.1663-1677, 2003.
9. “QLONE, the all in one tool for 3D Scanning,” https://www.qlone.pro/