A Deep Learning System that Learns a Discriminative Model Autonomously Using Difference Images

Ryodai Hamasaki
Graduate school of Science and Engineering, Saga University
Saga, Japan
19704013@edu.cc.saga-u.ac.jp

Koichi Nakayama
Graduate School of Science and Engineering, Saga University
Saga, Japan
knakayama@is.saga-u.ac.jp

ABSTRACT

In our proposed method, an object can be detected from time-series images taken by two or a few cameras. When one of the cameras detects that the object has moved, a system locates that object from the images taken by the other camera(s) by using the timestamps. Then, a position data of that object can be obtained autonomously without taking a lot of photographs of that object and teaching data to the system in advance. Moreover, the system passes the position data and a label of that object to YOLO, which is a learning model for discriminating objects. YOLO learns the data and become possible to indicate the label of the object. The results of an experiment showed that this system could discriminate the moving object only by two cameras without the previously prepared teaching data.[1]

CCS CONCEPTS

• **Computing methodologies** → **Machine learning;** *Machine learning algorithms*

KEYWORDS

Background difference method, YOLO

ACM Reference format:

G. Gubbiotti, P. Malagò, S. Fin, S. Tacchi, L. Giovannini, D. Bisero, M. Madami, and G. Carlotti. 2019. SIG Proceedings Paper in word Format. In *Proceedings of ACM GECCO conference, Prague, Czech Republic, July 2019 (GECCO'19)*, 4 pages.
DOI: 10.1145/ 3319619.3326887

1 INTRODUCTION

People with dementia are apt to forget when and where they put down their glasses, wallet, and other objects. Recently, it has become possible to attach tags to important objects so that their owners can locate them using a smartphone. However, such tags cannot be attached to all objects. Furthermore, it is desirable to know when the objects were moved from an initial position to their current position so that people with dementia can be shown how this happened. To do this, it is necessary to obtain data regarding when and to where the objects without tags were moved. Generally, when a model of an object is created, many images are photographed from many angles to be used as teaching data. This method is not realistic. Therefore, we propose a method that can detect the object in three-dimensional space from the time-series images taken by two or a few cameras, which operate only if the object moves. Several background difference methods can be used to detect a moved object in time-series images. BackgroundSubtractorMOG [1] is a Gaussian Mixture-based Background/Foreground Segmentation Algorithm. A detection system must be able to distinguish between moved shadows and moved objects. In this method [1], the shadow is considered when there are differences in both the chromatic and the brightness components when a computational color model is used [2].

In our system, if one of the cameras tells the system that an object has moved, the system can use the timestamps of the images taken by the other camera(s) to locate that object. If only the object has moved, a model of that object can be created autonomously without the need for teaching data.

In this paper, we examine a model that detects the location of a moved object using the time-series images taken by two cameras and uses deep learning to discriminate the object. The system passes the label and the position data of the object to You Only Look Once (YOLO) [3], which is a model for detecting objects that uses teaching to learn to discriminate them. Finally, a system must be able to detect the appearance and/or disappearance of an object using only two cameras in the room.

2 EXPERIMENT

2.1 Development Environment

The Python programming language was used to develop the system. We also used Open Source Computer Vision (OpenCV), an open-source library that includes the

GECCO'19, July 13-17, 2019, Prague, Czech Republic
 978-1-4503-6748-6/19/07…$15.00
DOI: 10.1145/ 3319619.3326887

BackgroundSubtractorMOG algorithm, and Keras, a neural network library.

2.2 Procedure

Fig. 1 shows the procedure whereby the system detects and discriminates an object taken by plural cameras (two cameras in this paper). The system uses the BackgroundSubtractorMOG algorithm to detect the moved object and surrounds the object with a rectangle (bounding box). Fig. 2 shows a background image. Fig. 3 shows that the image of an object has been added to the background. This object is a moved object because it has moved into this image. Fig. 4 shows the results of the detection. The white part in the figure represents the moved object. Then, the system creats a rectangular bounding box around the moved object using OpenCV, as Fig. 5 shows.

The system identifies the same object on time-series images taken by the other cameras (another camera) using the timestamps. Then, the system obtains coordinates of the bounding box (position data of the object) and lists the path in which the images are saved, the coordinates of the bounding box, and the object's label (dataset) in a table. Moreover, the system obtains additional position data of assuming reversed the images at the x-axis and y-axis and both the x and y-axes. Then, the system lists the dataset in the table. Namely, the system can obtain the position data of the object from multiple viewpoints.

After repeating this flow n times, the dataset is delivered to YOLO as teaching data. The system learns the coordinates of the bounding box and the labels of the objects using the deep learning of YOLO. Then, the system becomes able to detect and discriminate the object.

2.3 Dataset for Learning and Evaluation

In this experiment, two fixed cameras recorded the image of the object, which was the one shown in Fig. 3. The flow for creating the teaching data (see Fig. 1) was repeated 40 times (n = 40) for each camera. In addition, the coordinate data which were assumed to reverse at the x-axis, the y-axis, and at the x-and-y axis were prepared. Therefore, including these coordinates of reversed images, the two cameras provided 320 set data. The deep learning used 224 of the set data for the learning data and 96 for the validation data. The number of learning times was 100 epochs.

In addition, to evaluate the learning model, we prepared 20 set data for a test. The precision-recall measure was employed to evaluate the learning model. However, in this experiment, the precision indicates a rate that the bounding box detected by the learning model overlapped with the correct bounding box in pixels. The recall indicates a rate of the bounding box detected by the learning model in the correct bounding box. F-measure is a weighted harmonic mean of precision and recall. These metrics are expressed as follows;

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}, \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}, \quad (2)$$

$$F - measure = \frac{2\ Precision * Recall}{Precision + Recall}, \quad (3)$$

where *True Positive* indicates a number of pixels of the bounding box that the learning model correctly designated as a positive, *False Positive* indicates a number of pixels of the bounding box that the learning model incorrectly designated as a positive, *False negative* indicates a number of pixels of the bounding box that the learning model did not designate as a positive but should have been designated as a positive.

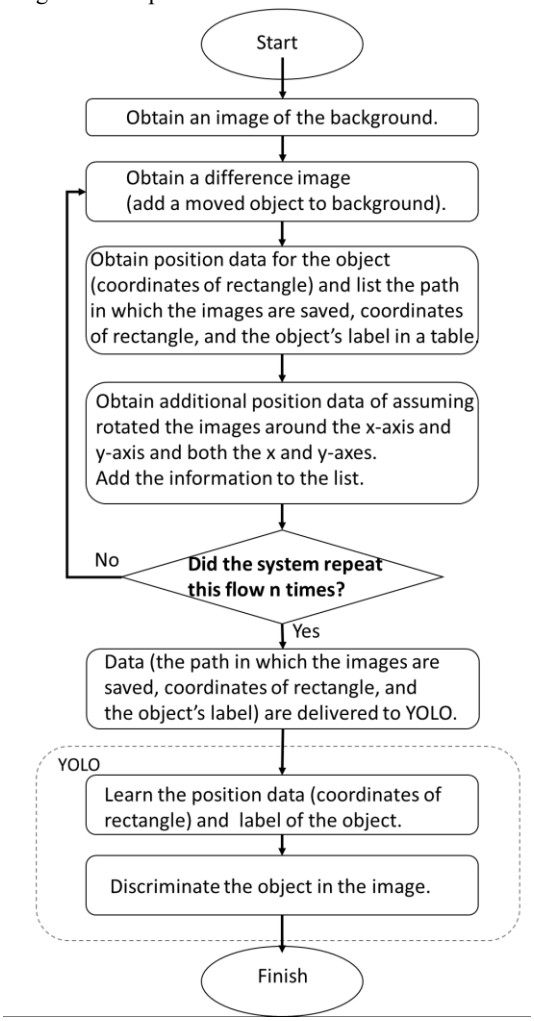

Figure 1: Flowchart of the system detecting a moved object.

2

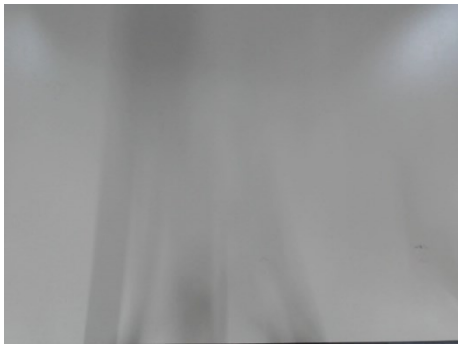

Figure 2: Background.

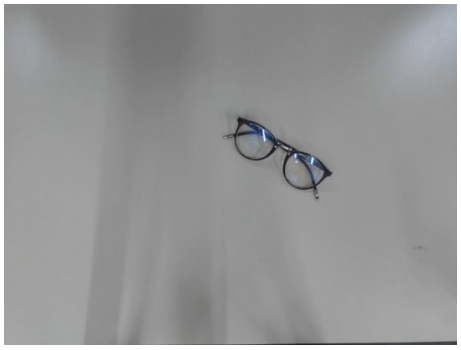

Figure 3: Object added to the back ground.

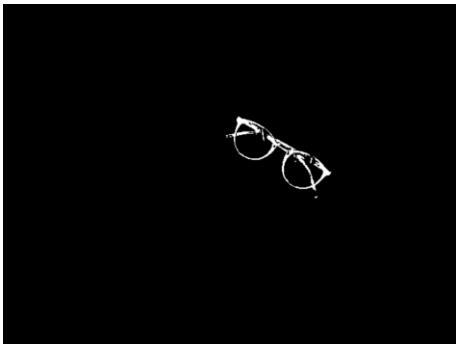

Figure 4: Background difference image.

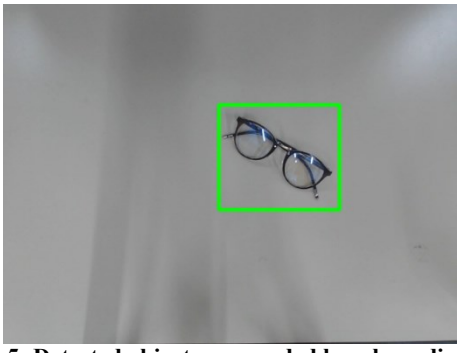

Figure 5: Detected object surrounded by a bounding box.

2.4 Results and Discussion

The average precision/recall/F-measures were 0.91, 0.90, and 0.91, respectively. The results indicate that the system's model efficiently detected the object.

Generally, when a system discriminates an object, it is necessary to create the teaching data by accumulating a large number of photographs from many angles in advance. However, in our system, the object could be discriminated only by two cameras without a preliminary teaching data. Namely, if two cameras are set in the room, the appearance and/or disappearance of an object will be detected.

4 CONCLUSIONS

In this paper, we proposed a model that detects and discriminates a moved object in three-dimensional space using the images taken by two cameras and the background difference method. The timestamps on the images were used to identify the movement of the object. Then, a model of the object was created without the use of a large number of photographs as teaching data, and the system used deep learning to successfully discriminate that object. In the near future, we will conduct an experiment using this system in a typical room.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 17H01950.

REFERENCES

[1] P. KaewTraKulPong, and R. Bowden. 2002. An improved adaptive background mixture model for real-time tracking with shadow detection. *In Video-based surveillance systems*. Springer, Boston, MA, 135-144.

[2] T. Horprasert, D. Harwood, and L. S. Davis.1999. A statistical approach for real-time robust background subtraction and shadow detection. *In IEEE ICCV'99 FRAME-RATE WORKSHOP*.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. 2016. You only look once: Unified, real-time object detection. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 779-788.