Object Recognition Algorithm based on Deep Learning

Yong-Hwan Lee* and Youngseop Kim
1Dept. of Digital Contents, Wonkwang University, Korea
(E-mail: hwany1458@empal.com)
2 Dept. of Electronics and Electrical Engineering, Dankook University, Korea
(E-mail: wangcho@dku.ac.kr)

ABSTRACT Object tracking is an exciting and interesting research area in the field of computer vision, and its technologies have been widely used in various applications such as surveillance, military, and augmented reality. This paper proposes and implements a novel and more robust object recognition and tracking system to localize and track multiple objects from input images, which estimates target state using the likelihoods obtained from multiple CNNs. As the experimental result, the proposed algorithm is effective to handle multi-modal target appearances and other exceptions.

Keywords: Object Recognition, Deep Learning, Convolutional Neural Network (CNN)

1. INTRODUCTION

Since object tracking technology based on computer vision is commonly used in many fields, many recognition and tracking system have been developed and researched to monitor several circumstances [1]. As well as the market of mobile device are expanding, various technology of object recognition and tracking are applied to personal device, such as mobile phone and drone. The task of tracking an unknown object in a video can be referred to as long-term tracking [1] or model free tracking [2]. The goal of such systems is to localize the object (some references call it as target object) in a generic video sequence, when an object is given only the bounding box that defines the wanted object in the first frame. Recognition and tracking objects are challenging technology because the tracking system has to deal with changes of the object appearance, illuminations, occlusions, out-of-plane rotations, and real-time processing requirements.

This paper implements a robust object recognition system to localize and track multiple objects from input images, which estimates target state using the likelihoods obtained from multiple convolutional neural networks (CNN) [3]. The CNNs are maintained in a tree structure and updated online along the path in the tree. Since each path keeps track of a separate history about target appearance changes, the proposed algorithm is effective to handle multi-modal target appearances and other exceptions such as short-term occlusions and tracking failures. In addition, since the new model corresponding to the current frame is constructed by fine-tuning the CNN that produces the highest likelihood for target state estimation, more consistent and reliable models are to be generated through online learning only with few training examples.

2. THE PROPOSED SCHEME

The overview of the proposed recognition and tracking system is shown in Figure 1, which has a simple architecture which is based on set of cyclically interconnected modules. Each module deals with specific type of input data that is elaborated to provide appropriate data to the next module. First, the system is initialized by feeding video from a camera view of mobile phone. The first step of the proposed approach is distinguishing foreground objects from stationary background. To achieve this, we use a combination of adaptive background subtracting and image pre-processing to create a foreground pixel map at every frame. The object recognition module extracts feature points and identifies a tracking object by using SURF and Locality Sensitive Hashing (LSH) algorithm. The output of the extracted object information is required by the object tracking module which traces the interest object using feature points, descriptors,
sub-window of the object, and its histogram. This module should redetects the object using the learned object information, when the interest object is missing in the tracking.

![Diagram of the proposed object recognition scheme](image)

**Figure 1.** Diagram of the proposed object recognition scheme

We implement the proposed system on iOS platform, using Objective-C and C++ with Xcode 8.1 on iMac running MacOS Sierra (10.12.1), 4GHz Intel Core i7, and 32GB DDR3 memory. To evaluate a performance of the proposed recognition and tracking method, we use 3 sample images, which are trained offline by training image data, shown in the figure. Experimental tests focus on target object modelling and matching with the trained object, and these are concentrated on how well does it detect the target object and how robust does it on appearance changes, changes scale, and covered object. Figure 2 shows an example screenshot of the execution results, running on the change of rotation and scale for interesting object.

![Example of the detected and tracked object from input image](image)

**Figure 2.** Example of the detected and tracked object from input image

### 3. CONCLUSION

This paper proposes a novel and more robust object recognition and tracking system to localize and track multiple objects from input images, which estimates target state using the likelihoods obtained from multiple CNNs. The CNNs are maintained in a tree structure and updated online along the path in the tree. Since each path keeps track of a separate history about target appearance changes, the proposed algorithm is effective to handle multi-modal target appearances and other exceptions such as short-term occlusions and tracking failures.

### ACKNOWLEDGMENT

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2018R1A2B6008255).

### REFERENCES