

Multi-Object Tracker Using Kernelized Correlation Filter Based on Appearance and Motion Model

Kwang-Yong Kim*, Jun-Seok Kwon**, Kee-Seong Cho*

*Smart Media Research Group, ETRI(Electronics and Telecommunications Research Institute), Daejeon, Korea

**School of Computer Science and Engineering, Chung-Ang University, Seoul, Korea

kwangyk@etri.re.kr, jskwon@cau.ac.kr, kscho@etri.re.kr

Abstract—The objective of this study is to determine a tracking method using kernelized correlation filter based on object's appearance and motion model used to track multi-object. This system largely consists of 4 modules: motion model, background subtraction, hijacking handling and occlusion handling. Lab colour model is applied to subtract background, and histogram of oriented gradient (HoG) is used to extract object feature. If occlusion among objects occurs, we use a method that tracks again after removing the overlapping objects in consideration of the depth between objects: The head of the closer object is being taken from a camera positioned below the head of the distant object. Thus, among occluded tracking objects, we find that the most upper located object is considered as the furthest object in captured camera image. If hijacking among objects is occurred, it has been solved by removing the overlapping region of the bounding box between two objects that maintain their relative positions for a period of time. These results indicate that this method may allow a solution for tracking of multi-object to be more robust to real-world tracking environments.

Keywords— Multi-object tracker, Background subtraction, Kernelized correlation filter, Occlusion handling, Hijacking handling

I. INTRODUCTION

A technology for detecting multiple moving objects and tracking them is one that has been widely used in the security field to track the object of interest in the last few years. We will examine the trend of recent papers relating to the proposed object detection and tracking.

Kumar et al. proposed background modelling using Gaussian Mixture Model (GMM) for blob extraction of multiple human motions and multiple people tracking using kalman filter for tracking multiple objects for real world surveillance scenarios [1]. However, there can be a possibility where one object will be hidden by another object in the field of view of any one of the cameras. This problem is referred to as occlusion. Thus, this proposed system has a difficulty problem to track human in the video sequences under occlusion in the dynamic background. G. Rao et al suggested the detection using median approximation technique and done tracking based on template matching using kalman filter estimation technique for single object in a sequence of frames [2]. Nevertheless, there also remains a problem to track object the video sequences when happen to done occlusion. Prabhakar et al presented object tracking system using concept

of frame differencing for detection and done technique of object template matching for tracking [3]. Mishra et al proposed an approach for tracking multiple objects in single frame in which the centroid of objects are taken as central component using some structure information [4].

Gavit et al suggested object tracking system which is used hand shaking between camera based on multiple cameras and used the block matching algorithm [5]. Ranipa et al dealt with the object detection using the background subtraction technique and distance metric learning based Bayesian algorithm. They also proposed a technique of tracking using background subtraction [6].

However, above-mentioned researches have problems in which object of interest cannot be tracked and can be missed, when occlusion or hijacking occurs among objects [7][8]. We propose a robust multi-object tracking for occlusion and hijacking problems using kernelized correlation filters based on object shape and motion model.

The rest of the paper is organized as follows: it is explained the configuration and procedure for multi-objects detection and their tracking from the video at Section 2; it is shown the experiment results at Section 3; and finally it is concluded a conclusion at Section 4.

II. PROCEDURE FOR MULTI-OBJECTS DETECTION AND THEIR TRACKING

Our visual tracker is based on the kernelized correlation filter (KCF) method [9]. The original KCF method, however, have several drawbacks and it cannot be directly used for our visual tracking problem. Main drawback of the KCF method is that it is developed for the single object tracking problem. As another drawback, the method is weak to real-world tracking environments. To handle aforementioned drawbacks, we extend the KCF method and make a new tracking method that can track multi objects. And we improve the original KCF tracker to be more robust to real-world tracking environments. For this, the KCF method is combined with four additional components, namely, motion model, background subtraction, occlusion handling, and hijacking handling.

A. Motion model

Given the target position at time t , the original KCF tracker finds a new position of the target at time $t+1$ by searching a

certain region, namely search window, centered on the target position at time t . If the target moves abruptly across consecutive frames, however, the search window may not cover the true target position at time $t+1$ and thus the tracker fails to track the target. To solve this problem, we use the motion model. The motion model calculates the average velocity of the target and moves the search window according to the estimated velocity, as shown in figure 1. To get the velocity, we consider positions of the target during recent 10 frames.

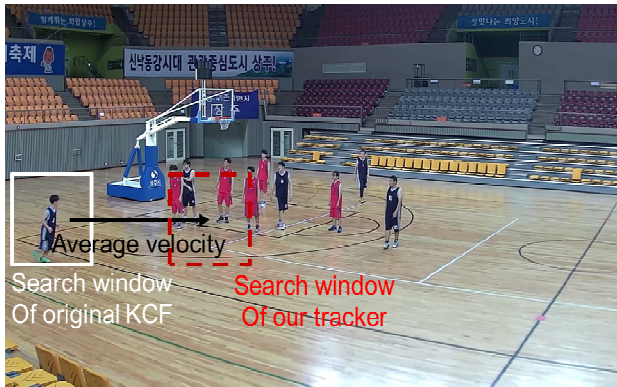


FIGURE 1: MOTION MODEL The search window (i.e. white rectangle) centered on the target position at time t does not include the true target position at time $t+1$. Our search window (i.e. red rectangle) shifted by the motion model covers the true target position at time $t+1$.

B. Background Subtraction

At the initial frame, we are given the background image. Because the camera does not move in our tracking environment, we can extract only the foreground region by subtracting an input image by the background image, as shown in figure 2. To remove noises and small regions, we apply morphologic operations, namely, erode and dilate operations into the subtracted image.



FIGURE 2: BACKGROUND SUBTRACTION The foreground region is extracted by subtracting an input image by the background image

C. Occlusion Handling

During visual tracking, a target is frequently occluded by other objects. When occlusions occur, most parts of the occluded target are invisible. Especially when the occluded target appearance is very similar to that of the occluding objects, the tracker mistakes the occluding objects as the occluded target. Before tracking the occluded target, to solve this problem, our method removes the occluding objects from the image, as shown in the right-hand side of figure 3.

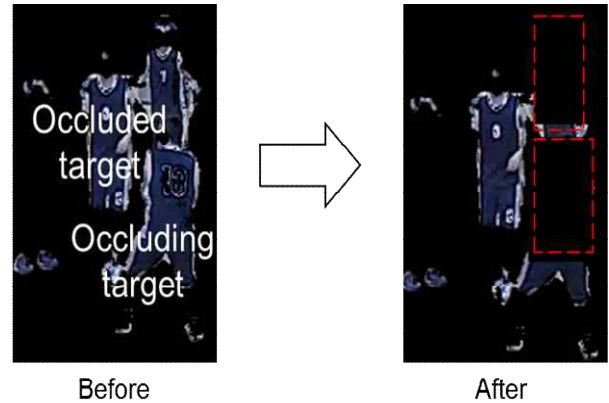


FIGURE 3: OCCLUSION HANDLING The original KCF method uses the observation as shown in left-hand side of the figure. Our tracker modifies this observation as shown in right-hand side of the figure, where there are not objects anymore, which have similar appearance.

D. Hijacking Handling

When two objects that have similar appearance are close enough, the trajectory of one object can be lost to the other object. We call this phenomenon as the hijacking problem. Unfortunately, the hijacking problem is inevitable during visual tracking. Hence we do not try to prevent the tracker from hijacking but try to recover the tracker from hijacking. If the relative positions of two bounding boxes are consistent over time, the trajectory is considered hijacked, as shown in figure 4. In this case, our tracker removes the common area of two bounding boxes, which forces the tracker to search nearby regions that contain the missed target.

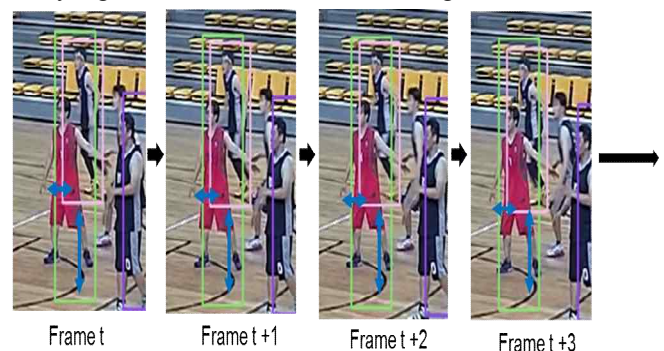


FIGURE 4: HIJACKING HANDLING The relative position (depicted in blue arrows) of green and pink bounding boxes are consistent over time. Hence one trajectory is hijacked by the other trajectory.

III. EXPERIMENT

The proposed tracker consists of four main components, namely, motion model, background subtraction, occlusion handling, and hijacking handling. Hence it is important to analyse which components of our tracker make whole tracking system to be successful.

A. Analysis on Background Subtraction

Figure 5 shows the tracking results before and after adding the background subtraction component. As demonstrated by the figure, our method accurately tracked the target depicted by the green bounding box, while the original KCF tracker failed to track the target.



FIGURE 5: ADVANTAGE OF BACKGROUND SUBTRACTION.

B. Analysis on Occlusion Handling

Figure 6 demonstrates the advantage of using the occlusion handling in visual tracking. Without the occlusion handling, IDs of two targets (depicted as violet and green bounding boxes) were switched during the occlusion, as shown in left-hand side of the figure. With the occlusion handling, on the other hand, two targets were accurately tracked, as shown in right-hand side of the figure.

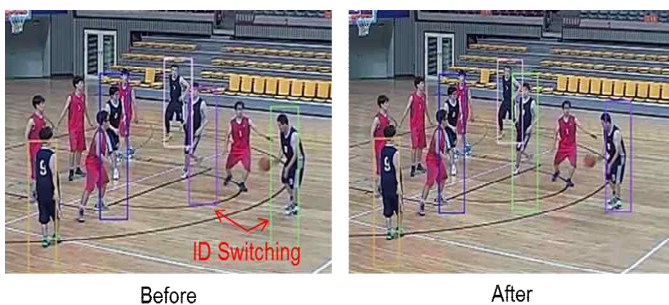


FIGURE 6: ADVANTAGE OF THE OCCLUSION HANDLING.

C. Analysis on Hijacking Handling

Figure 7 shows the hijacking handling help improve the tracking accuracy qualitatively. As shown in the left-hand side of the figure, the green bounding box was hijacked by the pink bounding boxes in the case of the original KCF method. On the other hand, our tracker that is combined with the hijacking handling successfully tracked the target, as shown in the right-hand side of the figure.

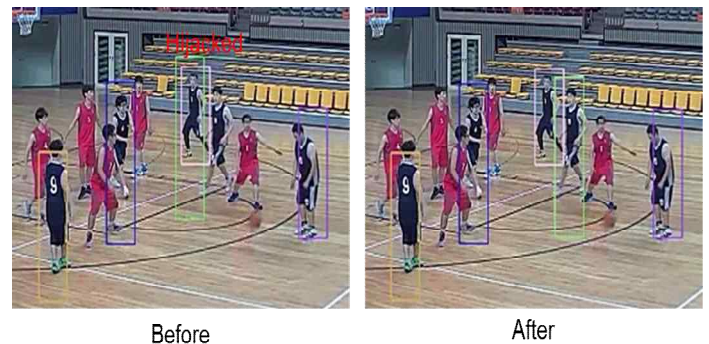


FIGURE 7: ADVANTAGE OF THE HIJACKING HANDLING.

IV. CONCLUSION

There have been various researches to track them without missing multi-object of interest in a video since a long time ago. However, these researches have problems in which object of interest cannot be tracked and can be missed, when occlusion or hijacking occurs among objects in real-world tracking environments.

To manipulate these difficulties, we improve the original KCF tracker and make a robust tracking method that can track multi-object in real-world tracking environments. Experimental results show that we can build robust multi-target object trackers by using four main components, namely, motion model, background subtraction, occlusion handling, and hijacking handling.

ACKNOWLEDGMENT

This work was supported by ICT R&D program of MSIP/IITP. [R0101-16-293, Development of Object-based Knowledge Convergence Service Platform using Image Recognition in Broadcasting Contents]

REFERENCES

- [1] A. Kumar, and C. Sureshkumar, "Multiple Human Tracking in Surveillance Videos", *The SIJ Transactions on Computer Science Engineering & its Applications(CSEA)*, vol. 2, no. 6, pp. 202-207, Sep. 2014.
- [2] G. Rao, and C. Satyanarayana, "Object Tracking System Using Approximate Median Filter, Kalman Filter and Dynamic Template Matching", *I.J. Intelligent Systems and Applications*, vol.6, no.5, pp. 83-89, 2014.
- [3] N. Prabhakar, V. Vaithyanathan, and A. Sharma, "Object Tracking Using Frame Differencing and Template Matching", *Research Journal of Applied Sciences, Engineering and Technology*, vol. 4, no. 24, pp. 5497-5501, 2012.
- [4] R. Mishra, M. Chouhan, D. Nitawwre, "Multiple Object Tracking by Kernel Based Centroid Method for Improve Localization", *I.J of Advanced Research in Computer Science and Software Engineering*, vol.2, no.7, pp. 137-140, 2012.
- [5] L. Gavit, R. Sanghavi, M. Parab, and M. Sardey, "Object Tracking Using Multiple Cameras", *I.J of Technology Enhancements and Emerging Engineering Reaearch*, vol. 2, Issue 5, ISSN 2347-4289. pp. 36-41, 2014.
- [6] P. Ranipa, and K. Naina, "Real Time Moving Object Tracking In Video Processing", *International Journal of Engineering Research and General Science*, vol. 3, Issue 1, January-February 2015.
- [7] J. Kim, J. Choi, S. Lim, S. Park, J. Kim, and C. Ahn, "Development of Scent Display and Its Authoring Tool", *ETRI Journal*, vol. 37, no. 1, pp. 88-96, February 2015.

- [8] H. Kim, S. Lim, and S. Yu, "Fast Reference Frame Selection Method Based on Best Reference Frame Index Correlation", *ETRI Journal*, vol. 36, no. 1, pp. 179-182, February 2014.
- [9] J. F. Henriques, R. Caseiro, P. Martins, J. Batista, "Exploiting the Circulant Structure of Tracking-by-detection with Kernels", *ECCV* 2012.



Kwang-Yong Kim received the B.S. degree in 1991, M.S. degree in 1993, Ph.D. degree in 1998, all from Chungnam National University, Daejeon, Korea, Computer Engineering Department. He is currently Principal researcher at Smart Media Research Group, Broadcasting & Media Research Laboratory, ETRI (Electronics and Telecommunications Research Institute). He has been currently working at ETRI as a research staff since 2000. His research interests include human detection and tracking, convergence broadcasting service, and the application of fuzzy logic and artificial neural networks.



Jun-Seok Kwon received the B.S. degree in 2006, M.S. degree in 2008, Ph.D. degree in 2013, all from Seoul National University, Seoul, Korea, Electronics & Electrical Engineering Department. He is currently Assistant Professor at School of Computer Science and Engineering, Chung-Ang University. He has been working at Chung-Ang University as a assistant professor since 2016. His research interests include computer vision, deep learning, object tracking, and visual surveillance.



Kee-Seong Cho received the B.S. degree in 1982, M.S. degree in 1984, all from Kyungpook National University, Daegu, Korea, Electrical Engineering Department. He is currently Project Leader at Smart Media Research Group, Broadcasting & Media Research Laboratory, ETRI (Electronics and Telecommunications Research Institute). He has been working at ETRI as a research staff since 1984. His research interests include cloud computing and network, image recognition, and convergence service, cloud computing and network, convergence broadcasting service, and mobility.