

A Comparative Study on Multi-object Tracking Methods for Sports Events

Sungwon MOON*, Jiwon LEE*, Dowon NAM*, Howon KIM*, Wonjun KIM**

*ETRI(Electronics and Telecommunications Research Institute), Korea

**KONKUK University, Korea

moonstarry@etri.re.kr, ez1005@etri.re.kr, dwnam@etri.re.kr, hw_kim@etri.re.kr, wonjkim@konkuk.ac.kr

Abstract— Due to the rapid growth of machine learning technology, there is a need for research to automatically recognize objects and analyze their behavior in various fields, as is the case with sports. Currently, a system for detecting and tracking multiple objects in a sporting event is not accurate enough. Since most of the services depend on the manual operation of an experienced operator, it is necessary to develop a real time tracking technique for detecting the position of an object. In this paper, we propose an algorithm for multi - object tracking in a sporting event by presenting the results of comparing the performance of existing algorithms for multi-object tracking.

Keywords— Multi object tracking, Visual tracking, Sports events, Video processing, KCF

I. INTRODUCTION

Recently, the improvement of machine learning technology represented by deep learning has dramatically improved the performance of classifying objects and motions in videos, and the demand for recognizing and analyzing objects in images is increasing. In the sports field, there is a need for a technique of automatic multi-object tracking in a sports game in order to automatically recognize and analyze the motion of objects such as players, referees, and balls in a sports game as shown in Figure 1. However, sports objects are fast and irregular in their movement, so it is difficult to track them, and therefore they are analyzing the contents of sports games through the manual operation of skilled operators.

Existing real-time single-object tracking has been studied extensively and shows a high level of tracking success rate, but multi-object tracking remains at object detection level. In addition, it is difficult to predict the performance of object tracking in a sporting event, since previous works apply multi-object tracking technique with the videos which objects are moving in certain direction such as CCTV.

In this paper, we suggest a suitable algorithm for multi-object tracking in sports events by applying algorithms with high performance on multi-object tracking such as MEDIANFLOW[1], BOOSTING[2], MIL(Multi instance learning)[3], KCF(Kernalized correlation filter)[4] proposed in the past to track multi-objects in sporting events and comparing their performance.



Figure 1. Example of multiple objects in soccer video

This paper is organized as follows. We explain the previous works in section II. Comparative experimental results are given in section III and concluding remarks are given in section IV

II. RELATED WORKS

In the case of MEDIANFLOW, tracking is based on the trajectory of the point, and tracking performance is fast due to low object tracking complexity and accurate when the median of the object is suitable for specifying the object. However, it shows low tracking success rate in images where the object shows irregular motion or the median of the object is not suitable to specify the object.

BOOSTING and MIL are tracking methods through object detection, and both have commonality in object detection and tracking through learning. In the case of MIL, it continuously updates the model to be tracked, which is good for object detection, but consumes a lot of resources for updating. BOOSTING is a method that combines learning data with a number of weak but specialized methods to derive high-level learning results, and uses various elements such as LBP (Local Binary Pattern) and color for object tracking.

Since KCF requires less resources than a learning-based approach even though it shows high tracking success rate, KCF is recognized as a state-of-the-art in many fields. KCF tracing object based on color characteristics by using correlation filter without learning or model update.

Most of the existing methods have been used in a single



Figure 2. ISSIA soccer videos

object tracking environment, and experiments have been performed on objects with a large number of feature points, such as faces, or objects with a large difference in color, such as tennis ball in the room.

III.COMPARATIVE EXPERIMENTAL RESULTS

The images used in the experiment are ID 1 and ID 3 images of ISSIA soccer video [5] in Figure 2, 13 players are tracked in ID 1, 10 players are tracked in ID 3, Performance was compared through the number of failed objects, which can be seen in Table 1. The CPU used in the experiment is Intel i7-6500U CPU@2.50GHz and also implemented OpenCV 3.1.0 in Visual Studio 2013 environment for implementation.

In the case of using the MEDIANFLOW algorithm, if the movement of the object is constant, the tracking of the object is performed at a high speed. However, if the object suddenly changes the movement, it cannot follow the trajectory and the white uniform and the line in the soccer field are often not distinguished. This was remarkable in the ID 1 video, in which the athletes changed rapidly, and showed a success rate of 7.7% for the corresponding video.

The MIL method uses a method of updating the object model in the video, but in many cases, the players of each team track the wrong player due to the nature of the sports game wearing the same uniform. As a result, the existing model was mistakenly updated as the model of the corresponding player, which showed a very low tracking success rate unlike the single object tracking. In addition, we confirmed the slowest of the four algorithms due to the nature of MIL, which uses a lot of resources in each model update.

TABLE 1. PERFORMANCE COMPARISON BETWEEN MULTI-OBJECT TRACKING ALGORITHMS

Method	ID 1		ID 3	
	Tracking Speed (fps)	Tracking Accuracy (%)	Tracking Speed (fps)	Tracking Accuracy (%)
KCF	45.41	84.6	63.22	90
MIL	1.46	31.7	1.97	20
BOOSTING	7.60	76.9	9.54	80
MEDIAN FLOW	19.33	7.7	21.77	40

BOOSTING, which uses weak learning compared to MIL but utilizes various characteristics, showed better performance even though similar method was used, and it showed a higher tracking success rate while tracking about 5 times faster than MIL. However, it is vulnerable to interference between objects in the image, which shows a low tracking success rate in the ID 1 image test, where there is a large number of inter-object crossings.

The KCF algorithm, which tracks objects through associations with existing objects based on color, showed the best performance for tracking multiple objects in a sporting event. We have traced all objects in most frames without failure and confirmed that they detect and track objects about 30 times faster than MIL.

We have confirmed through experimental results that there are some cases where algorithms that show excellent performance when tracking single objects are not working properly when applied to sports videos. These results seem to be due to the characteristics of sports videos in which similar objects move fast and irregularly. In addition, it has been confirmed that it is difficult to apply the learning - based detection / tracking algorithm to real - life environment because the speed of the learning - based detection / tracking algorithm is very slow.

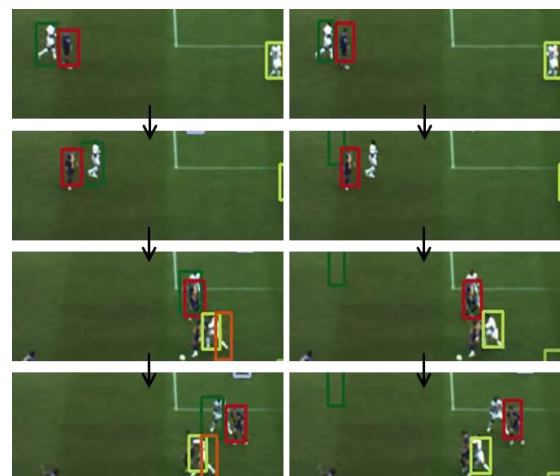


Figure 3. Tracking results with KCF(Left) and BOOSTING(Right) in case of frequently switching positions between players

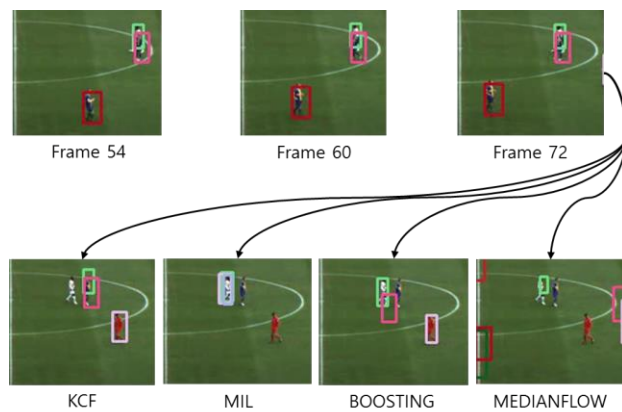


Figure 4. Tracking failures

IV. CONCLUSIONS

In this paper, typical object tracking techniques experimented with conventional single object tracking are tested in a multi-object environment in sports images. Through the experimental results, it is expected that the KCF technique and the improved method can be used to create a multi-object tracker for sports images with high speed and high accuracy.

As shown in Figure 4, in some cases, all existing trackers fail to track. We will try to develop robust automatic tracker to track the players, referees, and balls in sports videos by using color based feature like KCF.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2016 (R2016030044, Development of Context-Based Sport Video Analysis, Summarization, and Retrieval Technologies)

REFERENCES

- [1] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. "Forward-backward error: Automatic detection of tracking failures." in Pattern Recognition (ICPR), 2010 20th International Conference on, pages 2756–2759. IEEE, 2010.
- [2] Helmut Grabner, Michael Grabner, and Horst Bischof. "Real-time tracking via on-line boosting." in BMVC, volume 1, page 6, 2006.
- [3] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. "Visual tracking with online multiple instance learning." in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 983–990. IEEE, 2009.
- [4] Joao F. Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. "High-speed tracking with kernelized correlation filters." in Pattern Analysis and Machine Intelligence, IEEE Trans. on, 37(3):583–596, 2014.
- [5] T. D'Orazio, M.Leo, N. Mosca, P.Spagnolo, P.L.Mazzeo, "A semi-automatic system for ground truth generation of soccer video sequences." in Proc. IEEE International Conference on Advanced Video and Signal Surveillance, Sept. 2009.



Sungwon Moon received his B.S and M.S degrees of computer science from KAIST, Korea in 2010 and 2012. Since 2012 he has been a researcher in the Sportainment Section, SW.Content Reserch Laboratory, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea. His research interests are Digital Watermarking, Video Forensic, and Video Processing.



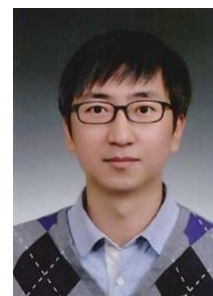
Jiwon Lee received the B. S. degree in Computer Engineering from Kyungpook National University, Republic of Korea, in 2008, and the Ph. D. degree in Computer Science from Korea Advanced institute of Science and Technology (KAIST), Republic of Korea, in 2013. Since 2013 he has been a senior researcher in the Sportainment Section, SW.Content Reserch Laboratory, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea. His research interests include multimedia security, image/video watermarking, and image/video processing.



Do-Won Nam received the B.S. degree in Computer Science from Korea Advanced Institute of Science and Technology (KAIST) in 1996 and M.S. degree in Information Technology from Pohang University of Science of Technology (POSTECH), Korea in 1998. He is working as a principal researcher in the Electronics and Telecommunications Research Institute (ETRI) since 2001. His research interests include data mining, digital rights management, digital cinema system and sports video analysis.



Howon Kim received his BS degree in electronic engineering from Kyungpook National University in 1997, and his MS and PhD degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST) in 1999 and 2004, respectively. From Jul. 2004 to Jan. 2006, he worked at LG Electronics (CTO) as a senior researcher. Since Feb. 2006, he is currently working at Electronics and Telecommunications Research Institute (ETRI) as a principal researcher and the director of Sportainment Research Section. His current research interests include image-based 3D modeling, human motion tracking and augmented reality



Wonjun Kim(Korea, 1980. 09. 06.) received the B.S. degree in electrical engineering from Sogang University, Seoul, Korea, the M.S. degree from the Department of Information and Communications, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, and the Ph.D. degree from the Department of Electrical Engineering, KAIST, in 2006, 2008, and 2012, respectively. From Sept. 2012 to Feb. 2016, he was a Research Staff Member in Samsung Advanced Institute of Technology (SAIT). Since Mar. 2016, he has been with the Department of Electronics Engineering, Konkuk University, Seoul, Korea, where he is currently an Assistant Professor. His research interests include image and video understanding, computer vision, pattern recognition, and biometrics, with an emphasis on saliency detection, face and action recognition, multiple object tracking, etc.