

An Efficient Vision-based Object Detection and Tracking using Online Learning

Byung-Gyu Kim^{1,*}, Gwang-Soo Hong², Ji-Hae Kim¹, Young-Ju Choi¹

Abstract

In this paper, we propose a vision-based object detection and tracking system using online learning. The proposed system adopts a feature point-based method for tracking a series of inter-frame movement of a newly detected object, to estimate rapidly and toughness. At the same time, it trains the detector for the object being tracked online. Temporarily using the result of the failure detector to the object, it initializes the tracker back tracks to enable the robust tracking. In particular, it reduced the processing time by improving the method of updating the appearance models of the objects to increase the tracking performance of the system. Using a data set obtained in a variety of settings, we evaluate the performance of the proposed system in terms of processing time.

Key Words: Object detection, object tracking, semi-supervised learning, real-time processing

I. INTRODUCTION

Recently, high-performance processing device, and the interest in the object tracking algorithm for automated image analysis in the spread of high quality cameras is large [1]. In particular, the intelligent monitoring system, video compression and editing, robot vision, the importance in the computer vision field, because the spotlight in various fields such as the augmented reality, it is emphasized. However, the object tracking performance is significantly affected by changes in illumination, the pose of the object, object size, and etc. Also it is a difficult problem because of the accurate positioning objects with low resolution and low quality image. To solve this problem, various studies have recently been actively conducted.

Object tracking areas has been studied greatly progressed in three directions [1]. Object tracking is a method for tracking an object by using the first feature such as a characteristic of Gradients, Color, and Texture. Also the combination of multiple features can be employed. The representative methods are typically a Mean Shift, and the Cameshift [1], [2], [3].

Monte Carlo approach using the second statistical data tracks an object using a mathematical algorithm for simulating the behavior. Typically, the particle filter and the Kalman filter are similar. In the case of a particle filter using Monte Carlo method, it measures repeatedly the posterior probability of the state to track the strong object because it keeps the track of the object based on the measured posterior probability.

Over time through the study of learning based on the object-tracking method as the third, the more accurate data learning changes in the varying objects according to the movement. The object tracking method of an exemplary learning-based is Tracking Learning Detection (TLD) [4]. This paper is organized as the following: In the next Section, the proposed algorithm will be introduced. In Section 3, we will discuss the result of experiments briefly. Finally, we will make a concluding remark in Section 4.

II. PROPOSED ALGORITHM

TLD method and the proposed method are also known as PN Tracker that tracks the objects using optical flow-based

Manuscript received December 6, 2017; Revised December 10, 2017; Accepted December 12, 2017. (ID No. JMIS-2017-0045)

Corresponding Author (*): Byung-Gyu Kim, Dept. of IT Engineering, Sookmyung Women's University, Seoul, Korea, Tel: 82-2-2077-7293, E-mail: bg.kim@sm.ac.kr.

¹Dept. of IT Engineering, Sookmyung Women's University, Seoul, Korea, E-mail: {yj.Chio, jh.Kim}@ivpl.sookmyung.ac.kr, bg.kim@sm.ac.kr

²Dept. of Computer Engineering, SunMoon University, Asan, Korea, E-mail: gs.Hong@ivpl.sookmyung.ac.kr

tracker and Random Ferns detector [4]. It initializes both the tracker and detector after the first frame in video, and looks for a tracker to pursue a target. If you succeed in even trace detector to find the target, thereby it learns the detector by utilizing the training data. If there is failure to find an object tracking detector again until you stop tracking the target initialize the tracker to the detected position to have a way to back track.

Object detection capability to track moving objects with respect to the correct object is also to be correct. Because the shape change with the sensitive area in a strong local and scale change estimate for the shape change with the scale and proposes using the TLD data of existing data (HOG, Haar, LBP).

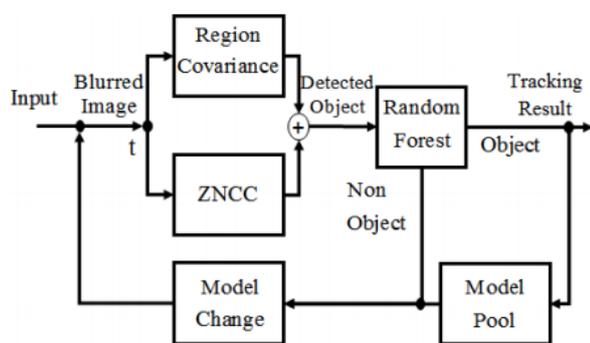


Fig. 1 Structure of the proposed algorithm.

Figure 2 shows a system flow chart. In a typical structure for learning TLD it is to collect data in order to track a specific part without using the learned data. In this structure, it is learning how to detect features to existing pedestrians, vehicles, etc. and faces bond TLD module is tailored to the generally available modules. The most important point is usually the step of combining the training set and the training of the existing TLD made using feature detection.

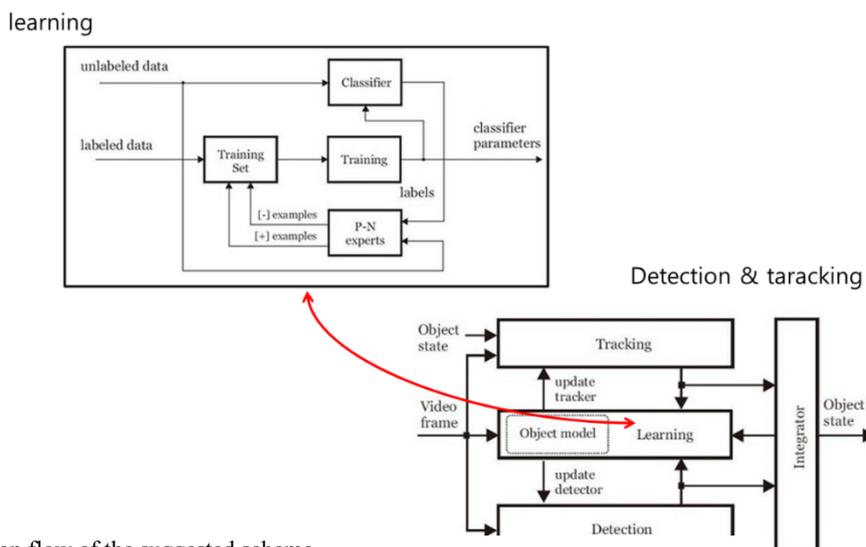


Fig. 2. Operation flow of the suggested scheme.

The long-term tracking problem addressed by TLD can be constrained when applied to faces. In the original formulation, the entire detector was learned online, starting from a single frame. An efficient classifier (randomized forest) was used to represent the decision boundary between the object and its background. In the case of face tracking, building the entire detector is not necessary since a range of face detectors is readily available. The learning therefore consists of building a validator that decides whether a face patch corresponds to the target or not. The validation is significantly less time demanding than the face detection since only a fraction of candidates have to be verified. On the other hand, its precision has to be high in order to avoid confusing two different identities. Face recognition is in general very challenging and a large number of sophisticated methods have been designed already. Here we show that a very simple validator and a learning method works very well when tracking a face in cluttered background.

We adopt a frontal face detection which demonstrated state-of-the-art performance and runs at 20 frames per second (FPS). On the top of the detector we incorporated a validator, a module that analyzes a face patch and outputs a confidence that the patch corresponds to the specific face. The validator is realized by a collection of examples. This collection is initialized by a single example in one frame and it is extended during tracking by inserting more examples.

III. PERFORMANCE ANALYSIS

Figure 3 shows the experimental results by the sequence. From the results, we can see that it has advantage in terms of processing speed. Especially, when Tracker and Detector are integrated at the same time, the processing time is better

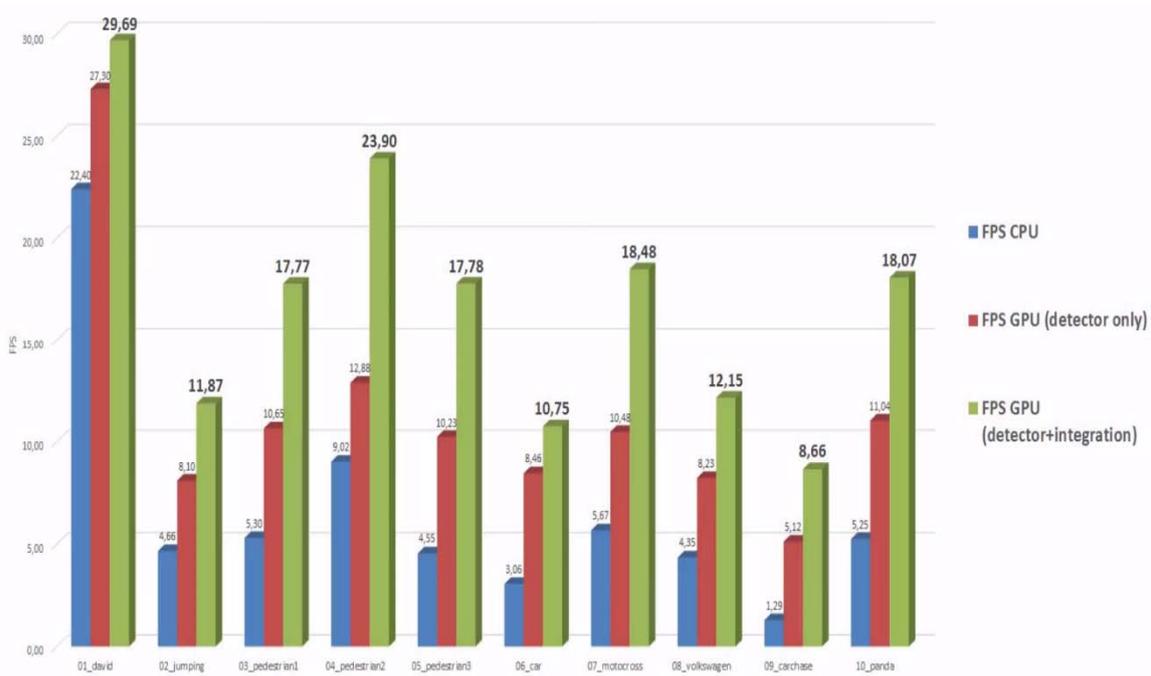


Fig. 3. Results of face detection and tracking.

performance than when the Detector is only operated. In case of the GPU platform, the processing time was good with only Detector.

In our experiment, the multi-tracking was applied in the CPU platform. In terms of processing time, about 4.88ms was observed on CPU and approximately 7.21ms was measured on the GPU. From the results, we can see up-to 29.69 FPS. It means this scheme can support a real-time processing system.

IV. CONCLUSIONS

In this paper, in conjunction with the training set, we made a general characteristic of the TLD framework which can be applied by different characteristics. The proposed scheme can be expected to be adopted for wide range of applications. In addition, except for the multi-tracking experiments, the proposed scheme confirmed real-time processing. This algorithm will be very useful to make a real-time gesture recognition and control in smart car technology.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B04934750).

REFERENCES

- [1] Alper Yilmaz et al., "Object tracking: A Survey," *ACM Computing Surveys*, Vol. 38, No. 4, Article 13, December 2006.
- [2] D. Comaniciu, V. Ramesh and P. Meer, P, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," *IEEE Computer Vision and Pattern Recognition*, Vol II, pp.142-149, 2000.
- [3] P. Hidayatullah and H. Konik, "CAMSHIFT improvement on multi-hue object and multi-object tracking," *3rd European Workshop on Visual Information Processing (EUVIP)*, pp. 143-148, 2011.
- [4] Zdenek Kalal, Krystian Mikolajczyk and Jiri Matas, "Tracking-Learning-Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL. 6, NO. 1, pp.1409-1422, 2010.

Authors



Byung-Gyu Kim has received his BS degree from Pusan National University, Korea, in 1996 and an MS degree from Korea Advanced Institute of Science and Technology (KAIST) in 1998. In 2004, he received a PhD degree in the Department of Electrical Engineering and Computer Science from Korea Advanced Institute of

Science and Technology (KAIST). In March 2004, he joined in the real-time multimedia research team at the Electronics and Telecommunications Research Institute (ETRI), Korea where he was a senior researcher. In ETRI, he developed so many real-time video signal processing algorithms and patents and received the Best Paper Award in 2007.

From February 2009 to February 2016, he was associate professor in the Division of Computer Science and Engineering at SunMoon University, Korea. In March 2016, he joined the Department of Information Technology (IT) Engineering at Sookmyung Women's University, Korea where he is currently an associate professor.

In 2007, he served as an editorial board member of the International Journal of Soft Computing, Recent Patents on Signal Processing, Research Journal of Information Technology, Journal of Convergence Information Technology, and Journal of Engineering and Applied Sciences. Also, he is serving as an associate editor of Circuits, Systems and Signal Processing (Springer), The Journal of Supercomputing (Springer), The Journal of Real-Time Image Processing (Springer), Helyion Journal (Elsevier), and International Journal of Image Processing and Visual Communication (IJIPVC). He also served as Organizing Committee of CSIP 2011 and Program Committee Members of many international conferences. He has received the Special Merit Award for Outstanding Paper from the IEEE Consumer Electronics Society, at IEEE ICCE 2012, Certification Appreciation Award from the SPIE Optical Engineering in 2013, and the Best Academic Award from the CIS in 2014. He has been honored as an IEEE Senior member in 2015.

He has published over 150 international journal and conference papers, patents in his field. His research interests include software-based image and video object segmentation for the content-based image coding, video coding techniques, 3D video signal processing, wireless multimedia sensor network, embedded multimedia communication, and intelligent information system for image signal processing. He is a senior member of IEEE and a professional member of ACM, and IEICE.



Gwang-Soo Hong received his BS and MS degrees in the Department of Computer Engineering from SunMoon University, Korea, in 2011 and 2013, respectively. In 2013, he joined the Department of Computer Engineering for pursuing his PhD degree at SunMoon University. His research interests include

image coding algorithms, video coding technique and depth estimation algorithms in stereo vision.



techniques.

Ji-Hae Kim received her BS degrees in the Department of IT Engineering from Sookmyung Women's University, Korea, in 2017. In 2017, he joined the Department of IT Engineering for pursuing her MS degree at Sookmyung Women's University. Her research interests include video signal processing, facial expression recognition, emotional expression, and video coding



Young-Ju Choi received her BS degrees in the Department of IT Engineering from Sookmyung Women's University, Korea, in 2017. In 2017, he joined the Department of IT Engineering for pursuing her MS degree at Sookmyung Women's University.

Her research interests include image and video coding techniques, deep learning-based video recognition, and intelligent/perceptual video coding algorithms.