Triple Online Boosting Training for Fast Object Detection

Ning Sun(✉), Feng Jiang, Yuze Shan, Jixin Liu, Liu Liu, and Xiaofei Li

Engineering Research Center of Wideband Wireless Communication Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
sunning@njupt.edu.cn

Abstract. In this paper, we present an online co-training method called Triple Online Boosting Training (TOBT). TOBT comprehensively combines the advantages of online boosting and tri-training to accomplish the function of online learning and sample validation at the same time. With the help of the novel feature extraction scheme named Fast Feature Pyramid (FFP) reported recently, we develop a real-time online method for multi-scale object detection. This method is proposed for detecting all-sized instances of a certain class in entire image, which is different from other online detectors for tracking purposes. Various experiments based no benchmark datasets and real videos demonstrate the efficacy of the proposed method in the respect of processing speed and stability for changing object appearance and scenarios.

Keywords: Object detection · Fast feature pyramid · Online boosting · Tri-training

1 Introduction

Sliding windows detectors are the most popular and powerful approaches in the field of object detection. This kind of detector is widely applied in the various environment, especially in the dynamical background scene, such as mobile or vehicle-mounted video system [1]. There are two main challenges of sliding windows based object detector involving real-scene automotive application. The first problem is computational cost for extracting the heterogeneous features to build multi-scale feature pyramid. The second one is the adaptive ability about variability of target appearance due to the change of illumination and different pose.

In decades, many research efforts have been devoted to improve the performance of object detection in the respect of processing speed [2, 3]. For instance, the groundbreaking work is the method of Viola and Jones [4], which develops a real-time (about 15 fps) face detector by boosting Haar like feature and integral image representation. After that, many successful sliding windows based methods are proposed for detecting multifarious target in various scenes, such as HOG, DPM and ChnFtrs and so on. However, these methods are unable to meet the need of real-time application due to heavy computational load. Recently, Dollár et al [5]

H. Zha et al. (Eds.): CCCV 2015, Part II, CCIS 547, pp. 70–79, 2015.
DOI: 10.1007/978-3-662-48570-5_8

presents an ingenious speed-up scheme named Fast Feature Pyramid (FFP) for multi-scale feature extraction in 2009. Therefore, a more polished version of FFP was published in literature [6]. The key idea of FFP is that finely sampled pyramids may be obtained inexpensively by extrapolation from coarsely sampled ones, which prominently decreases the time-consuming of feature extraction procedure in the multi-scale object detection.

In order to deal with the problem about the generalization and stability of detector in real scene, the effective way is online learning mechanism. There are a lot of works focused on the online learning algorithm [7-10], most of which are successfully implemented in tracking application [11]. When we exploit continuous learning mechanisms in single task of object detection, no prior knowledge about size and position of targets help us to restrict the scope of sliding windows. The issue of automatic validation for image samples has to be addressed. Co-training [12] is a semi-supervised learning paradigm which trains two or more learners respectively from different views and lets the learners label some unlabeled examples for each other. Most previous theoretical analyses on co-training are based on the assumption that each of the views is sufficient to correctly predict the label. However, this assumption can hardly be met in real applications due to feature corruption or various feature noise [13-16]. Tri-training proposed by Zhou et al [14] in 2005 neither requires the instance space described with sufficient and redundant views nor does it put any constraints on the supervised learning algorithm.

The key contributions of this work can be summarized as follows:

1. We present novel online semi-supervised leaning algorithm called Triple Online Boosting Training (TBOT), which simultaneously achieves online learning of new features and sample images identification.
2. A fast multi-scale online object detection framework is developed through an asynchronous interactive mechanism with TOBT algorithm and fast feature pyramid scheme.
3. We design several experiments about pedestrian detection and tank detection based on the benchmark and real videos to evaluate performance of the proposed method. Experiments demonstrate that the proposed method achieves better performance than some previous object detection methods.

The remainder of this paper is organized as follows. We give overview of the proposed method and a more description of TBOT in Section 2. Experimental results on four different data sets compared to existing approaches are given in Section 3. Conclusions are presented in the last section.

2 The Proposed Method

2.1 Overview of Our Method

As shown in Fig. 1, the proposed method mainly consists of two modules: (1) fast detection (FD), (2) online verification and training (OVT). The detector in FD module is a binary classifier updated by module of OVT, which exploits the fast feature pyramid scheme to accelerate the procedure of multi-scale feature extraction to about

five times the speed of traditional ones [2]. And, the image results regarded as object by fast detector are saved and delivered to the module of OVT. In OVT, we present a co-training algorithm called Triple Online Boosting Training (TOBT) to accomplish the function of online learning and sample validation at the same time. In TOBT, the image patches are input as ambiguous samples, and three pre-trained online boosting classifiers are initialized to start the tri-training scheme. Then, one ambiguous sample is labeled for a classifier if the other two classifiers agree on the labeling in each round. The iterative process continues until the three classifiers do not change. Finally, we build a cascaded classifier as fast detector using the weak hypotheses of the best one of the above-mentioned classifiers.

It should be pointed out that the operation of FD module and the OVT module is not synchronous. This is mainly due to the obvious different between the computational cost of fast detector and the one of another. And, it is not necessary that frequently updated the fast detector in a short time.

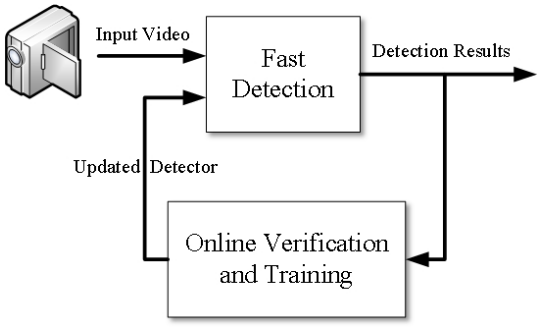

Fig. 1. The flowchart of the proposed method

2.2 The Fast Detector

The speed-up of the fast detector in our method is depended on the scheme of Dollár's Fast Feature Pyramid. When we built feature pyramid by this scheme, only feature data of one scale per octave is required to be computed precisely. And, the feature data of rest scale can be approximated by the ones of intermediate scales with minor loss in accuracy according to the exponential power law. In the respect of feature representation, we also follow the Aggregated Channel Features (ACF) in [6], which has been proven effective in general object detection method and suitable for FFP scheme.

2.3 Triple Online Boosting Training

The online learning is the pivotal function of our method. Meanwhile, it is very important to the performance of online learning that design a reliable verification module to identify the input image samples. We build a semi-supervised leaning algorithm named Triple Online Boosting Training (TOBT) to simultaneously achieve online learning and sample images identification, which comprehensively combine the benefits of online boosting and tri-training.

Table 1. Pseudo code of Triple Online Boosting Training

when TOBT is triggered by new input ambiguous image samples

1. **Input**: pre-trained classifiers $H_i^P, i = 1,2,3$; ambiguous image samples A from fast detector; initial parameters $e_i^{'} = 0.5, l_i^{'} = 0$

Verification and online training phase:

2. **repeat until** none of $H_i^P, i = 1,2,3$ changes

3. **for** $i = 1$ *to* 3 **do**

4. $L_i = NULL$; $Flag_i^u = FALSE$; $e_i = MeasureError(H_j \,\&\, H_k), (j,k \neq i)$

5. **if** $(e_i < e_i^{'})$ **then for** every $x \in A$ do

6. **if** $(H_j(x) == H_k(x), (j,k \neq i))$ **then** $L_i = L_i \cup \{x, H_j(x)\}$

7. **end for**

8. **if** $(l_i^{'} == 0)$ **then** $l_i^{'} = \lfloor e_i / (e_i^{'} - e_i) + 1 \rfloor$

9. **if** $(l_i^{'} < |L_i|)$ **then if** $(e_i |L_i| < e_i^{'} l_i^{'})$ **then** $Flag_i^u = TURE$

10 **else if** $(l_i^{'} > e_i / (e_i^{'} - e_i))$ **then**

11 $L_i = SubSample\ (L_i, \lceil e_i^{'} l_i^{'} / e_i - 1 \rceil), Flag_i^u = TURE$

12 **end for**

13 **for** $i = 1$ *to* 3 **do**

14 **if** $(Flag_i^u == TURE)$ **then** $H_i = OnlineBoosting(L_i)$, $e_i^{'} = e_i$, $l_i^{'} = |L_i|$

15 **end for**

16 **end repeat and choose**
$H_c = \arg\max(sum(H_i(x_m) == Labl(L_i(x_m))), i = 1,2,3$

Building cascaded detector phase:

17 $H_s = sort(H_c)$; $L = 0$;

18 **repeat until** $H_s == NULL$

19 $\omega = 0$

20 **for** $h_n \in H_s$ **do**

21 **if** $\omega > T$ **then** $\omega = 0$; $L = L + 1$; **break**

22 **else** $\omega = \omega + \alpha_n$; $H_s = H_s - h_n$; $H_L = H_L + h_n$

23 **end for**

24 **end repeat**

25 **output** $H_{final} = \bigcup_{m=1}^{L} H_m$

In the following, we describe the TOBT algorithm in detail. The TOBT is composed of two parts: (1) verification and online phase, (2) building cascaded detector phase. The pseudo code of TOBT is shown in Table 1. First of all, three classifiers, which are previously trained using online boosting [7] on three different sub-set of one training data

set, and the training samples of the detection resulted from FD module in a certain period of time, are both prepared for starting the TOBT algorithm. In the verification and online phase, we mostly follow the scheme of Zhou's [14] tri-training expect that we use the validated image set L_i and online boosting algorithm to update corresponding classifier instead of using the original label sample set plus set L_i and any offline learning. After that, the best performance classifier H_c is chosen for building cascaded detector. At the beginning of building cascaded detector phase, the week hypothesis $h_n \in H_c$ is sorted in descending order based on the weight $\alpha_n = \ln((1-\varepsilon_n)/\varepsilon_n)$ of h_n, where ε_n is the error of h_n. Then, first layer of cascaded is combined with the best part of week hypotheses in H_s. The number of week hypothesis in each layer is limited by a pre-defined factor T. The process is repeated by filling the next layers with the remaining week hypothesis. After the levels are completed, their concatenation forms the cascade detector and updates the classifier in fast detection module.

3 Experiments and Discuses

We design several experiments on benchmark dataset and real scene video to test the performance of the proposed approach and compare the results with other the-start-of-are object detection methods: HOG [17] and ACF. The implementation of all test methods are based on C++ and OpenCV library except for the training of ACF, which uses directly the Matlab code of Dollár's toolbox[18]. And all experiments execute on the image processing server with Intel Xeon E7 4820 CPU and 32GB memory.

3.1 Person Detection

In this section, three detectors are training to detect persons for testing. Experiments are run on image sequence S3-T7-A View4 in PETS2006 and a real scene video named LIBRARY captured from the D1 resolution camera installed in the library of our campus (Fig.2). All three detectors are trained on the INIRA dataset [17].

Fig. 2. Test video in the pedestrian detection experiments. Top: test video PETS2006. Bottom: test video LIBRARY

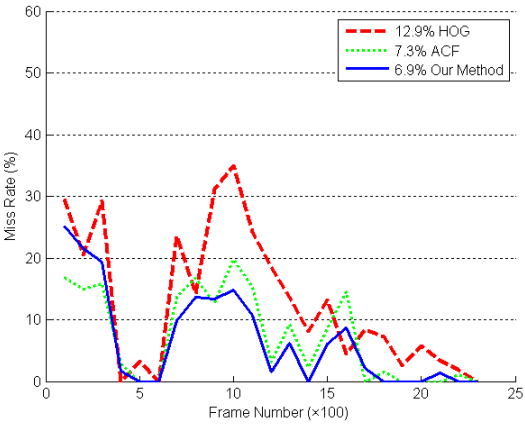

Fig. 3. Detection results on video PETS2006

Three detectors are run on the PETS2006 test video under the condition of one false positive per frame. According to our method, we trigger the TOBT algorithm to update fast detector every 100 frames. So, the average of miss rate per 100 frames was calculated to evaluate the accuracy of three detectors. As shown in the Fig.3, the average miss rate (See the legend in the top left corner of Fig.3) of our method is superior to one of other methods in the mostly period of test video, although the performance of our method is worse than that of ACF and HOG in few trigger periods at the beginning of video PETS2006. It is proved that the TOBT algorithm proposed in our method can continuously learn the new feature of pedestrian in the test video to improve the discriminability of the fast detector. Secondly, ACF based detectors (Our methods and ACF) achieve the outperforming detection results in the comparison with the HOG detectors, especially in the present of object partial occlusion. This result demonstrates that the multi-channel features are more discriminative that HOG feature, which is consistent with the study in literature [2].

The second test of pedestrian detection is applied on the real video called LIBRARY, which is consists of 1458 frames with the size of 720*576. The proposed method is also compared with detectors based on HOG and ACF under the same condition as the first test. In Fig.4, it can be found that the curves follow the similar trend as Fig.3, and the reduction of miss rate by the proposed method is more distinct than one in the first test. It is also shown that the TOBT algorithm can effectively improve the adaptability of detector for various application environments. Another similarity trend shown in Fig.3 and Fig.4 is that the miss rate of our method is higher than that in ACF in the first one or two round. This is caused by the different depth choice of decision tree in our method and ACF detector. We use one level decision tree (decision stump) in Online-boosting algorithm but two level depth decision tree in ACF. This kind of choice makes the proposed method faster in detection processing, which can be found in Fig.5.

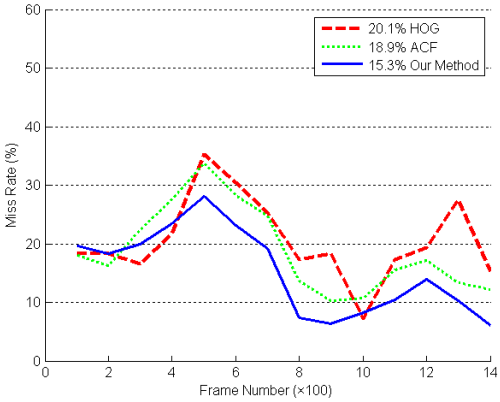

Fig. 4. Detection results on video LIBRARY

Moreover, we record the run-time of three detectors in the above-mentioned two tests. In Fig.5, we plot average miss rate versus runtime for three detectors. With D1 resolution video, the proposed method can achieve the detection speed at 28.5 fps, faster than the one of ACF at 22.2 fps and the one of HOG at 3.9 fps. It is mainly due to the reason that the weak hypothesis of the proposed method is the decision stump classifier, which is more efficient than the two level of decision tree of ACF detector.

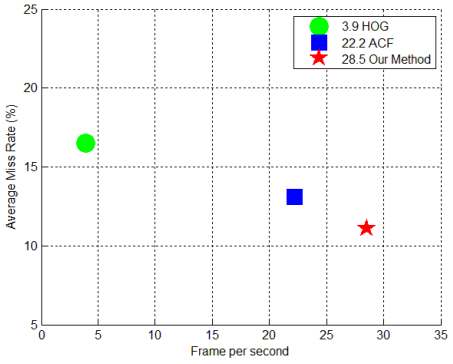

Fig. 5. Time versus miss rate of three detector

3.2 Tank Detection

After the experiments of person detection, we apply our method to detect tanks in the video, which is something more challenging task. Different from pedestrian detection, it is usually hard to gather the sufficient image samples of military equipment for training, just like main battle tank, belonging to enemy. The same as the above section, we compare the proposed method with HOG and ACF. The training samples and test video are both collected from internet. The training set is consist of 1370 image of

tanks around the world except for the German Leopard 2 tank and 1500 negative samples randomly bootstrapped from landscape images (Fig.6). The test data is a 2089 frames video of Leopard 2 tank named TANK with D1 resolution (Fig.7). This experimental setting can assess the ability of online learning of the proposed method more effectively.

Fig. 6. The positive and negative samples of tank detection. Top: positive samples. Bottom: negative samples.

Fig. 7. The test video of tank detection

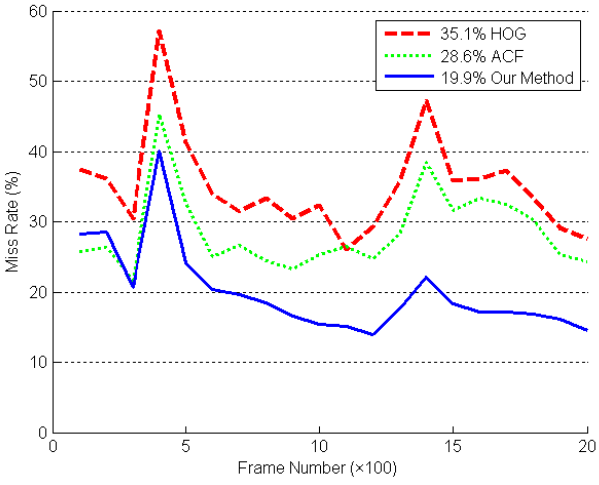

Fig. 8. Detection results on video TANK

Similar as the result of pedestrian detection experiments, the proposed method rapidly achieves the best accuracy of detection after three updating operation. And, the average miss rate of our method is lower than one of ACF above 30% (Fig.8). In the respect of run-time, the fps of three detectors keeps unchanged because that computational cost of detection is invariable with the sliding windows scheme under the same resolution video.

4 Conclusions

In this paper, the Triple Online Boosting Training (TOBT) algorithm combined with online learning and co-train is proposed for incrementally learning new features from autonomously invalidated image patches of detection results. Through an asynchronous interactive mechanism with TOBT algorithm and fast feature pyramid scheme, we build a real-time online method for universal object detection. Then, we design several experiments of pedestrian detection and tank detection based on the benchmark and real videos to evaluate performance of the proposed method and the other two the-state-of-art detectors. The comparison of experimental results prove the effectiveness of the proposed method in the field of accuracy and run-time.

Acknowledgments. This work was supported by the National Nature Science Foundation of China (61471206, 61401220) and Natural Science Foundation of Jiangsu province (BK20141428, BK20140884 and BK20140896).

References

1. Geronimo, D., Lõpez, A.M., Sappa, A.D., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. IEEE Trans. Pattern Anal. Mach. Intell. **32**(7), 1239–1258 (2010)
2. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian Detection: An Evaluation of the State of the Art. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 743–761 (2012)
3. Benenson, R., Omran, M., Hosang, J., Schiele, B.: Ten years of pedestrian detection, what have we learned? In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014 Workshops. LNCS, vol. 8926, pp. 613–627. Springer, Heidelberg (2015)
4. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 511–518 (2001)
5. Dollár, P., Belongie, S., Perona, P.: The fastest pedestrian detector in the west. In: Proc. British Machine Vision Conf. (2010)
6. Dollár, P., Appel, R., Belongie, S., Perona, P.: Fast Feature Pyramids for Object Detection. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1532–1545 (2014)
7. Grabner, H., Bischof, H.: On-line boosting and vision. In: Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, pp. 260–267 (2006)
8. Chang, W., Cho, C.: Online Boosting for Vehicle Detection. IEEE Trans. Systems, Man, and Cybernetics-Part B: Cybernetics **40**(3), 892–902 (2010)
9. Qi, Z., Xu, Y., Wang, L., Song, Y.: Online multiple instance boosting for object detection. Neurocomputing **74**, 1769–1775 (2011)

10. Visentini, I., Snidaro, L., Foresti, G.: Cascaded online boosting. Journal of Real-time image processing **5**(4), 245–257 (2010)
11. Wu, Y., Lim, J., Yang, M.-H.: Online object tracking: a benchmark. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2411–2418 (2013)
12. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proc. the 11th Annual Conference on Computational Learning Theory, pp. 92–100 (1998)
13. Goldman, S., Zhou, Y.: Enhancing supervised learning with unlabeled data. In: Proc. the 17th International Conference on Machine Learning (2000)
14. Zhou, Z., Li, M.: Tri-Training: exploiting unlabeled data using three classifiers. IEEE Trans. Knowledge and Data Engineering **17**(11), 1529–1541 (2005)
15. Xu, J., He, H., Man, H.: DCPE co-training for classification. Neurocomputing **86**, 75–85 (2012)
16. Li, M., Zhou, Z.: Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Trans. Systems, Man and Cybernetics **37**(6) (2007)
17. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2005)
18. Piotr's Computer Vision Matlab Toolbox. http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html