ONLINE REAL BOOSTING FOR OBJECT TRACKING UNDER SEVERE APPEARANCE CHANGES AND OCCLUSION

Li XU[1], *Takayoshi YAMASHITA*[2], *Shihong LAO*[2], *Masato KAWADE*[2], *Feihu QI*[1]
[1] Computer Science and Engineering Department, Shanghai JiaoTong University, P.R. China
[2] Sensing & Control Technology Laboratory, OMRON Corporation, Japan

ABSTRACT

Robust visual tracking is always a challenging but yet intriguing problem owing to the appearance variability of target objects. In this paper we propose a novel method to handle large changes in appearance based on *online real-value boosting*, which is utilized to incrementally learn a strong classifier to distinguish between objects and their background. By incorporating online real boosting into a particle filter framework, our tracking algorithm shows a strong adaptability for different target objects which undergo severe appearance changes during the tracking process.

Index Terms— online real boosting, tracking, appearance changes

1. INTRODUCTION

One of the main challenges of visual tracking is to cope with the appearance variability of target objects. Originating from the detection approach, many existing tracking algorithms try to handle the appearance variation by constructing a representation from a large training database [1, 2]. However, it is difficult to build an ideal model which adapts to all challenging situations such as poor illumination and severe occlusion.

Recently, some researchers adopted a completely different point of view to deal with variations. They considered tracking a classification problem and incrementally trained a classifier to distinguish the object from the background clutter [3, 4]. Representing an object with a constantly updated classifier, their methods naturally obtained adaptability for appearance variability. Since only a few training samples are available at each frame, the performance of these methods mainly relies on the learning ability of the online trainer. Specifically, online boosting [4, 5] is a good choice as a learning algorithm in that it has an advantage in speed over other competitors [6].

Oza et al. first proposed an online version of boosting in [5] and later Grabner et al. [4] extended their work and applied the online discrete adaboost to many vision applications. Although the online discrete boosting provides good adaptability for typical tracking tasks, it may fail when confronted with sudden appearance changes and occlusions, due to the limitation of discriminative complexity.

In this paper, we present a tracking system which is robust to severe appearance changes. First, a novel online version of real-value boosting is introduced, which entitles the strong classifier to a higher accuracy and a faster convergence speed. Then, a set of distinctive features for boosting are adopted to further improve the accuracy of the classifier. A final real-time tracker is obtained by fusing the online real boosting within the well-known probabilistic tracking framework -- particle filters. Experimental results demonstrated that our approach is much more robust for tracking objects that undergo large appearance changes.

The rest of the paper is organized as follows: Section 2 introduces the online real boosting training algorithm. Section 3 discusses the features used by the online boosting. Section 4 presents the tracking system using the online real boosting. Section 5 gives the experimental results and Section 6 concludes the paper.

2. ONLINE REAL VALUE BOOSTING

In this section, we first briefly review the real value weak classifier used for offline boosting and then we introduce the online real boosting algorithm and make a comparison to the online discrete one.

2.1. Real-Value Weak Classifier

Real Adaboost algorithms [7] use the real value confidence-rated weak classifiers which map a sample space to a real-valued confidence space. The general form of the real value weak classifier could be written as follow:

$$h(x) = \log\left(\frac{\hat{p}(y=1|x)}{\hat{p}(y=-1|x)}\right) \quad (1)$$

where x belongs to the sample space χ and $y \in \{-1,+1\}$, represents the negative and positive samples, respectively. $\hat{p}(y=1|x)$ and $\hat{p}(y=-1|x)$ are the estimated posterior probability for positive and negative samples.

In practice, Wu et al [8] proposed a look-up-table type weak classifier and partitioned sample space into several disjoint bins. The probabilities were then estimated based on the discrete sample distributions.

The real-value weak classifiers have an advantage over the Boolean classifiers in discriminative capability. Although the final strong classifier trained by discrete boosting achieves similar accuracy as the one trained by real boosting, it includes much more weak classifiers. In other words, real boosting can achieve a higher accuracy when the number of weak classifiers used is fixed, which is exactly the case in online boosting.

2.2. Online Real Adaboost

Unlike off-line training, which uses all samples to update one weak classifier, the online version has a fixed-length strong classifier and uses one arrived sample to update the entire stage.

The overall algorithm for online real boosting is depicted in Table 1. Like other online training method, we exploit a random feature subset (F_1, in Table 1) for feature selection to decrease the training time and update it after iteration (Table 1, Step 4).

For each weak classifier in the online stage, there is a feature selector working over time. Every selector maintains its own information on the features in the feature subset and updates it when a new sample is available. The feature selector then picks up the best features with respect to the cumulated error stored up to this point. Subsequently, the selector updates the sample weight according to the output of the corresponding weak classifier and passes the sample to the next selector.

The weight-update process (Eq. 7-9) boosts the online stage and enables the succeeding weak classifiers to focus on the difficult samples. It was proposed by Oza et al [5] and later modified by Grabner et al [4]. Specifically, when the cumulative error in Eq. 6 is above some threshold (0.5, for instance), we skip the weight-update steps.

Our novelties are embedded in Eq. 2-6, where the real value classifiers are applied. In each selector, we stored the cumulative information for updating the weak classifiers. $w_{n,m}^{+1}$ and $w_{n,m}^{-1}$ are the discrete distribution of the positive and negative samples, $\lambda_{n,m}^{sc}$ and $\lambda_{n,m}^{sw}$ are the cumulative weight of the correctly and wrongly classified samples.

For every feature in the feature pool, we first use the sample weight to update the cumulative distribution (Eq. 2). Then, we construct the real value weak classifier using the distribution (Eq. 3). Care should be taken when estimating the probability from the discrete distribution. For online training, especially in the initializing phase, only a few samples are available. Therefore, directly estimating probability from the distribution may lead to instable result. Our solution is to apply Gaussian smooth to the distribution. In current implementation, the length of the Gaussian filter is set to 1/5 of the feature bins. We also add a normalization to make the cumulative weight a distribution. It is necessary when the numbers of the positive and negative training samples are unequal.

Differing from the Boolean weak classifiers, the real classifier provides a real value confidence, from which we can learn about not only whether the weak classifier works on this sample but how well it works. The real confidence is then used to estimate the cumulative error (Eq. 4-6), which would later affect the weight-update steps and, behind the weight updating, the whole boosting process.

Table 1: Algorithm for online real boosting

Required: Feature pool F_1 (*M* features), F_2 (all features)

- Given sample image <x, y>, where $y \in \{-1,+1\}$
- Initialize the weight of current sample $\lambda = 1$
- For $n = 1,2,...N$ // the *n*th feature selector
  1. For $m = 1,2,...M$ // the *m*th feature in F_1

     a. Update weak classifiers

     If $F_1^m(x) \in binj$

     $$w_{n,m}^{y}(j) = w_{n,m}^{y}(j) + \lambda \quad (2)$$

     $$h_{n,m}^{weak}(x) = \frac{1}{2}\ln\left(\frac{GN\left(w_{n,m}^{+1}(j)\right)+\varepsilon}{GN\left(w_{n,m}^{-1}(j)\right)+\varepsilon}\right) \quad (3)$$

     where *GN*(.) is a normalization function followed by a Gaussian smooth.

     b. Estimate cumulative error $e_{n,m}$

     If $y \cdot h_{n,m}^{weak}(x) \ge 0$

     $$\lambda_{n,m}^{sc} = \lambda_{n,m}^{sc} + \lambda \cdot \left|h_{n,m}^{weak}\right| \quad (4)$$

     Else

     $$\lambda_{n,m}^{sw} = \lambda_{n,m}^{sw} + \lambda \cdot \left|h_{n,m}^{weak}\right| \quad (5)$$

     $$e_{n,m} = \lambda_{n,m}^{sw} / (\lambda_{n,m}^{sc} + \lambda_{n,m}^{sw}) \quad (6)$$
  2. Select the best weak $h_{n,m+}^{weak}(x)$ classifier with minimal cumulated error.
  3. Update sample weight and voting weight

     If $y \cdot h_{n,m}^{weak}(x) \ge 0$

     $$\lambda = \lambda / \left(2\left(1 - e_{n,m+}\right)\right) \quad (7)$$

     Else $\lambda = \lambda / \left(2e_{n,m+}\right)$ (8)

     $$\alpha_n = \log((1 - e_{n,m+}) / e_{n,m+}) \quad (9)$$
  4. Replace the worst feature in F_1 with a new one in F_2, reinitialize the classifier parameters.
- The final strong classifier is:

  $$h^{strong}(x) = sign(conf(x)) \quad (10)$$

  $$conf(x) = \sum_{n=1}^{N} \alpha_n h_n^{weak}(x) \quad (11)$$

2.3. Comparison

We made a comparison between the online discrete boosting and the proposed method on the face detection problem. One thousand face samples are used to train the online stages. Figure 1 shows the result of our comparison: the online discrete and real boosting achieves almost the same false alarm rate (left), while the real version has a much higher detection rate (right). Moreover, the real boosting converges more quickly as the training sample increase. Note that the online real boosting has similar computational complexity of the discrete one.

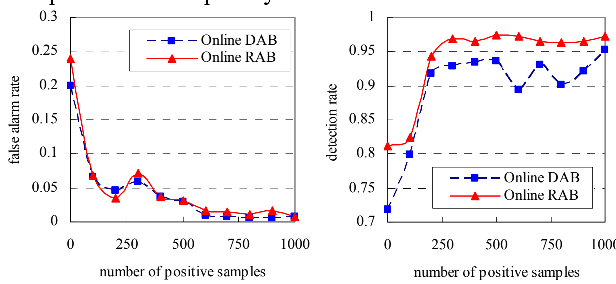

Figure 1: False alarm rate (left) and detection rate (right) as a function of the number of training samples. Curve with triangle marks represents the online real boosting.

3. FEATURES FOR BOOSTING

As the fixed number of weak classifiers limited the discriminative complexity to some extend, a straightforward technique for improving the performance is to enrich the feature type used for boosting. It is encouraged by a work of Levi et al [9], in which they reported that by combining the orientation histogram with the linear haar features, the detector obtained a higher accuracy even with a small training database.

Four different types of features are used in our system: Haar-like wavelet [10], Absolute Haar-like wavelet, edge orientation histogram (EOH) [9] and edgelet features with symmetric pair [11].

Haar and EOH features are carefully studied and proved to be very useful for object detection. Although the absolute haar-like features provide similar information of haar features (it is just the absolute value of haar), it helps the trainer converge quickly. Our reason for borrowing the edgelet features from pedestrian detection was our belief that they would provide enough cues of object contour when the targets undergo severe variation.

While the details on the performance of our final system are presented in Section 4, we launched a simple experiment to compare the effectiveness between features. We omit each type of features one by one and trained the online classifier using the rest features to demonstrate the importance of each type. We found that (Figure 2) when training the classifier without EOH features, the detection rate decreased sharply. As we use the common faces as our training samples, the edgelet features contribute a little to the overall performance. However, the computational cost of edgelet features is negligible as the edge and orientation map have already been calculated for EOH features.

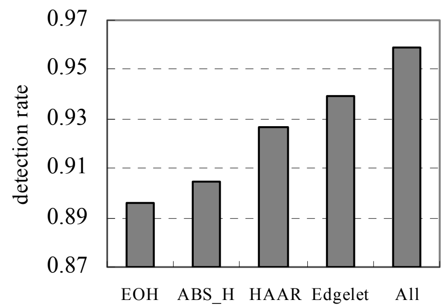

Figure 2: Effectiveness comparison between features. Each column represents the accuracy *without* using the listed type of features.

4. TRACKING

In practice a simple framework is used to apply the online real boosting to object tracking problem. We assume that the target has already been detected in the initial frame and is used to initialize the online stage. Standard particle filters [12] are utilized in our framework.

We construct a probabilistic observation model based on our online real classifier using all the features mentioned above. For every arrived frame, we evaluate current classifier on all particles (Eq. 11) and calculate the weighted mean position of the particles that have positive confidence. Then the target region is used as a new positive sample and the surrounding regions represent the negative samples. After updating the classifier, we re-sample the particle filters based on their importance weights and wait for next frames.

5. EXPERIMENTAL RESULTS

In implementation, we used 50 weak classifiers for the online stage and 64 bins for the feature look-up table. An experiment concerned about the performance of the online discrete boosting and our method using different type of features is conducted on a face database. All samples are fed into the trainer one by one. Figure 3 shows the result roc curves of the experiment: at the same detection rate, more than 50% decrease in false alarm rate was achieved by applying the real-value classifiers to the online boosting (the bottom two curves). Moreover, 50% more decrease in false alarm rate was obtained by introducing more types of features for boosting (the left-most curve).

We tested our tracker on a large number of sequences. The average running speed for a 640×480 sequence is about 25 fps on a standard PC (P4 3G, 1G memory). Although quantitative comparison with other methods is difficult to conduct, due to lack of standard test set, we made simple comparisons to an enhanced mean-shift [2] and the online discrete boosting [4]. Figure 4 showed some of the examples on our sequences and on the public data in [13]. Note that the mean-shift method (the black rectangle) is more likely to fail when the object undergoes appearance changes. The online discrete boosting (the white dotted rectangle) has adaptabilities for variation, but it also fails when sudden changes and occlusion occur. Our method (the solid white rectangle) is robust to large appearance changes in both pose and illumination, with severe occlusions.

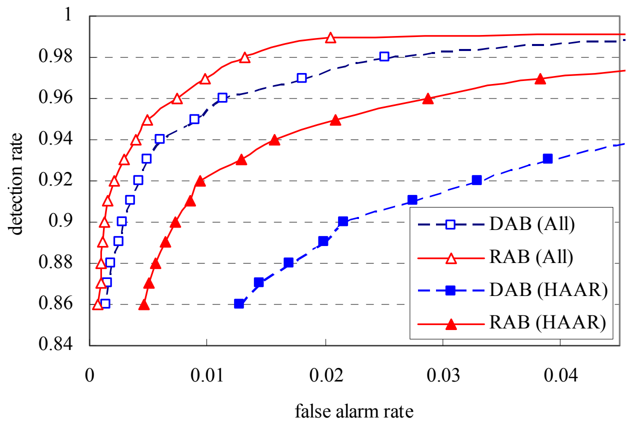

Figure 3: ROC curves of the online boosting algorithms.

5. CONCLUSIONS

In this paper, we have presented an online real boosting algorithm for real-time object tracking. As a specific version of boosting, the online real boosting has an advantage in learning speed. Moreover, the key advantage of real boosting, over its discrete counterpart, is the high accuracy it achieves, especially in the online case where only a few weak classifiers and training samples are available.

By incorporating online real boosting into the tracking system, our tracker is adaptable to severe appearance variations. Experiments demonstrate the effectiveness of the proposed method in different kinds of environments where target objects undergo large changes in appearance.

6. REFERENCES

[1] Y. Li, H. AI, C. Huang, and S. Lao, “Robust Head Tracking Based on a Multi-State Particle Filter”, In *Proc. FG*, pp.335-340, 2006

[2] T. Yang, S.Z. Li, Q. Pan, J. Li and C. Zhao, “Reliable and Fast Tracking of Faces under Varying Pose”, In *Proc. FG*, pp.421-428, 2006

[3] S. Avidan, “Ensemble Tracking”, In *Proc. CVPR*, vol. 2, pp. 494–501, 2005

[4] H. Grabner and H. Bischof, “On-line Boosting and Vision”, In *Proc. CVPR,* vol. 1, pp. 260-267, 2006

[5] N. Oza and S. Rusell, “Online bagging and boosting”, In *Proc. Artifical Intelligence and Statistics*, pp. 105-112, 2001

[6] D. Tax and P. Laskov, “Online SVM learning: From classification to data description and back”, In *Proc. Neural Network and Signal Processing*, pp. 499–508, 2003

[7] R.E. Schapire and Y. Singer, “Improved Boosting Algorithms Using Confidence-rated Predictions”, *Machine Learning,* pp. 297-336, 1999

[8] B. Wu, H. Ai, C. Huang and S. Lao, “Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost”, In *Proc. FG*, pp. 79-84, 2004

[9] K. Levi and Y. Weiss, “Learning Object Detection from a Small Number of Examples: The Importance of Good Features”, In *Proc. CVPR*, pp. 53–60, 2004

[10] P. Viola and M. Jones, "Robust real-time face detection", In *Proc. ICCV,* vol. .2, pp. 747, 2001

[11] B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors”, In *Proc. ICCV*, vol. 1 pp. 90-97, 2005

[12] S. Arulampalam, S.Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking”, IEEE Trans. *Signal Process*., vol. 50, no. 2, pp. 174–189, 2002

[13] J. Lim, D. Ross, R. Lin, and M. Yang, “Incremental Learning for Visual Tracking”, In *Proc. Advances in Neural Information Processing Systems*, pp. 793–800, 2005

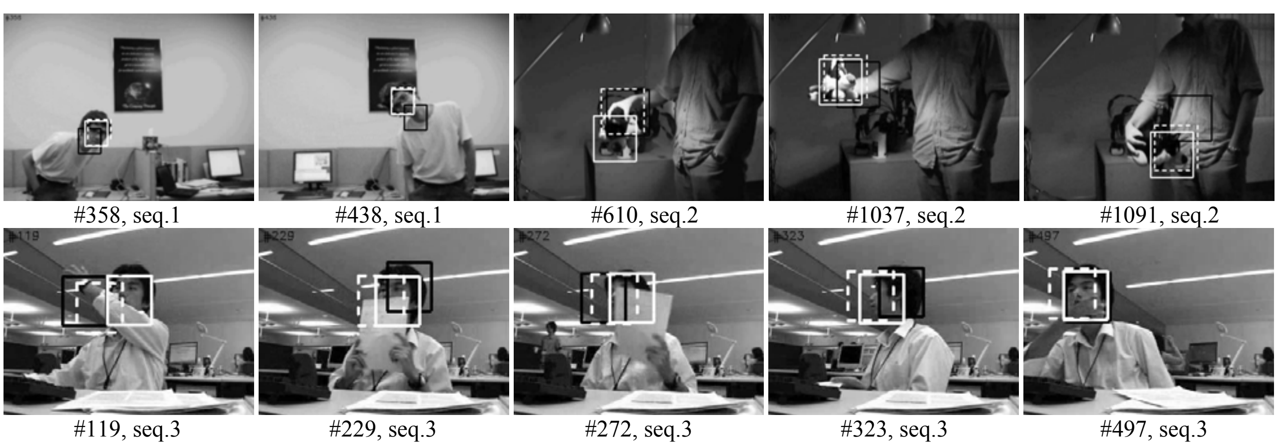

Figure 4: Tracking objects undergoing severe appearance changes. The black rectangle is based on an enhanced mean-shift [2], the white dotted rectangle is the result of the online discrete boosting and the white solid rectangle is the result of our method. Our method outperforms the others when severe changes happen. The first two sequences (seq.1, seq.2) are provided in [13].