ONLINE CO-TRAINING RANKING SVM FOR VISUAL TRACKING

Pingyang Dai, Kai Liu, Yi Xie, Cuihua Li

Xiamen University
Computer Science Department
Xiamen, China

ABSTRACT

Online learned tracking is widely used to handle the appearance changes of object because of its adaptive ability. Learning to rank technique has attracted much attention recently in visual tracking. But the tracking method with online learning to rank suffers from the error accumulation problem during the self-training process. To solve this problem, we propose an online learning to rank algorithm in the co-training framework for robust visual tracking. A co-training algorithm combined with ranking SVM collects features and unlabeled data for training. Two ranking SVMs are built with different types of features accordingly and dynamically fused into a semi-supervised learning process. This semi-supervised learning approach is updated online to resist the occlusion and adapt to the changes of object's appearance. Many experiments on challenging sequences have shown that the proposed algorithm is more effective than the state-of-the-art methods.

Index Terms— Visual tracking, online co-training, ranking SVM

1. INTRODUCTION

As a major topic in computer vision and related fields, visual tracking in video sequence has been widely used in many applications such as surveillance, vehicle navigation and augmented reality to human-computer interaction. Visual tracking has been investigated for more than ten years, but the design of a robust tracker is still an open problem. The robust tracker shall adapt to various cases in real world environment such as sudden illumination changing, varying view points, occlusion and clutter background.

Generally, this tracking problem can be divided into two categories: generative methods and discriminative methods. Generative tracking methods learn a model to represent the appearance of an object. These methods formulate tracking as finding the most similar object appearance to the traget model. The IVT method [1] utilizes an incremental subspace model to adapt the changes of appearance. This method performs well when the target object encounters illumination changes and pose variations. A tracking framework [2] was proposed that incrementally learns a low-dimensional covariance tensor representation, which efficiently adapts to appearance changes. Adam et al. [3] proposed a fragment-based method to handle occlusion. In this method, the idea of voting has been exploited for tracking where the template of the object is represented by a set of local patches.

Instead of building a model to describe the appearance of an object, discriminative tracking methods formulate tracking as a classification problem. The trained classifier is used to find a decision boundary that can effectively separate the object from the background. To handle appearance changes, the classifier is updated incrementally over time. The ensemble tracker [4] trains an ensemble of weak classifier online to distinguish the object from background. Each weak classifier is a linear hyperplane in a feature space composed of histogram of gradient orientations and R, G, B colors. Collins et al. [5] proposed a method to adaptively select color features that effectively discriminate the object from background in each frame. To alleviate the drift problem, Babenko et al. [6] proposed a tracking method based on the online multiple instance learning (MIL) method which puts all ambiguous positive and negative samples into bags to learn a discriminative model. And Grabner et al. [7] proposed a semi-supervised online boosting method to handle the drift problem.

However, the main problem of above discriminative method is the error accumulation during the self-training process. The trained classifier updates itself by the classification results. When each time the classifier updates, an error might be introduced which results in a tracking error. So the tracker is not robust to outliers and may drift or fail in target tracking. To solve this problem, we propose an online tracking algorithm which considers the learning to rank problem into the co-training framework. Co-training is a typical semi-supervised learning algorithm. In this framework, two different types of features such as Haar-like features and HOG features are collected and used to learn the ranking SVM accordingly. Two ranking SVMs are online updated

This research work was supported by the National Natural Science Foundation of China Under Grant No.61373077, Doctoral Program of Higher Specialized Research Fund (No.20110121110020), the Natural Science Foundation of Fujian Province of China Under Grant (No.2013J01257,2013J05101) and Shenzhen City Special Fund for Strategic Emerging Industries (J-CYJ2012).

and dynamically fused into a final classifier by co-training. The final ranking SVM outputs the classified result in which the target is ranked higher than others around it. Compared with other discriminative trackers, our online semi-supervised learning approach promotes each ranking SVM classifier that uses the information from other features, thus leading to be more robust.

The rest of this paper is organized as follows. Section 2 gives a briefly overview of the related work. The proposed Online Co-traing ranking SVM algorithm is presented in Section 3. Finally, experimental results are presented in Section 4, and the last section draw a conclusion.

2. RELATED WORK

Discriminative method formulates object tracking as a binary classification which aims to separate the target object from the background. As a discriminative method, learning ranking function has been a major issue in machine learning. Among rank learning methods, ranking SVM has been favorably applied to various applications in computer vision. The RankBoost algorithm with L1 regularization was proposed by Yang et al. [8] for facial expression recognition. Huang et al. [9] proposed a transductive learning framework with hypergraph ranking in image retrieval. Recently, Bai et al. [10] tried to deal with the tracking problem as ranking problem and proposed a tracker based on the ranking SVM. Furthermore, they formulated the tracking process as a weakly supervised ranking problem and proposed an online Laplacian ranking SVM [11] for visual tracking. However the online discriminative method is self-training. Each time the classifier updats, tracking error would be introduced and accumulated which would finally result in drift or tracking failure.

Recent approaches formulate the discriminative tracking methods as semi-supervised learning to tackle the above problem. Co-training method proposed by Blum and Mitchell [12] is a principled semi-supervised method for learning in data with two feature representations. Yu et al. [13] proposed a Bayesian undirected graphical model for con-training. A nonstationary co-training kernel for Gaussian process classifiers is introduced in order to avoid alternating view optimizations. A co-training framework [14] was proposed to train two SVMs trackers online with color histogram features and HOG features.

3. TRACKING WITH ONLINE CO-TRAINING RANKING SVM

3.1. Ranking SVM

Learning to rank is a kind of machine learning technique used to build a ranking model which has received increasing attention. Among existing learning to rank methods, support vector machine are used widely in building ranking model. Ranking SVM has been proposed by Herbrich et al [15] and applied to various applications for its prominent performance on learning a ranking function.

Assume that the input space and output space are $\mathbf{R}^n$ and $\mathbf{R}$. Let input set $\mathbf{X} \subseteq \mathbf{R}^n$, where n denotes the dimension of the input space. And label set $\mathbf{Y} = \{y_1, y_2, \dots, y_k\} \subset \mathbf{R}$ is the output set. The ranking function is define as

$$\mathbf{H} : \mathbf{X}^n \rightarrow \mathbf{Y}, \quad (x_i, x_j) \in \mathbf{X} \tag{1}$$

where (x_i, x_j) are two objects to be rank. If x_i is preferred to x_j, it can represented as $x_i \succ x_j$. Then the original training set can be transformed into a pair-wise training set as

$$(x_i - x_j, y), \quad y = \begin{cases} +1 & \text{if } x_i \succ x_j \\ -1 & \text{otherwise} \end{cases} \tag{2}$$

where y is the label of the training pair. Assume there exists a linear ranking function $H(x) = \mathbf{w}^T\mathbf{x}$, then the preference relations between instances can be calculated by

$$x_i \succ x_j \Leftrightarrow H(x_i) > H(x_j) \tag{3}$$

$$H(x_i) > H(x_j) \Rightarrow w^T x_i > w^T x_j \Rightarrow w^T(x_i - x_j) > 0. \tag{4}$$

The ranking problem with training data is transformed to the binary classification problem with training data pairs. And a SVM model is generated to solve the binary classification problem. The primary problem of ranking SVM is formulated as

$$\begin{aligned} &\min_{w,\eta} \frac{1}{2}\|w\|^2 + C\sum \xi_{i,j} \\ &\quad s.t.\ \forall (i,j) \in \{(x_i, x_j) : y_i > y_j\} : w^T(x_i - x_j) > 1 - \xi_{i,j}, \\ &\qquad \xi_{i,j} > 0. \end{aligned} \tag{5}$$

where C is the parameter that controls the trade-off between training error and margin size. Training ranking SVM is a quadratic optimization problem to balance the maximization of ranking margin and the minimization of the ranking error of the training pairs. This problem can be expressed and solved in its dual form

$$\begin{aligned} &\max_{\beta} \sum_{i,j} \beta_{i,j} - \frac{1}{2}\sum_{i,j}\sum_{u,v} \beta_{i,j}\beta_{u,v}(x_i - x_j)^T(x_u - x_v) \\ &\quad s.t.\ 0 \leq \beta \leq C. \end{aligned} \tag{6}$$

The result ranking function can be written as

$$\begin{aligned} H(x) =&(w^*)^T x = \sum_{i,j} x^T(x_i - x_j) \\ w^* =& \sum_{i,j} \beta^*_{i,j}(x_i - x_j). \end{aligned} \tag{7}$$

Figure 1 illustrates an overview of the learning of ranking SVM.

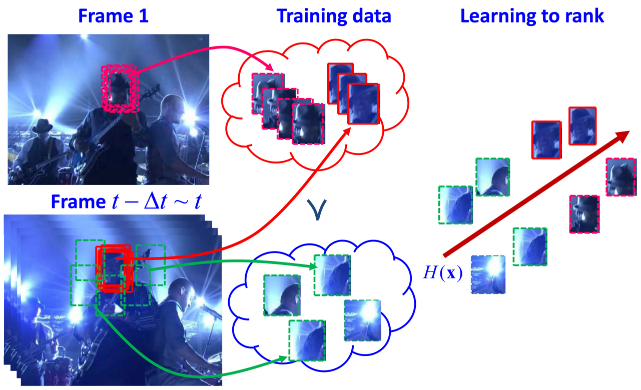

Fig. 1. Illustration of ranking SVM learning.

3.2. Co-training

Co-training has been proved to be an effective semi-supervised classification framework. It was formulated as a PAC-style learning formally [12]. This framework assumes that each example is described by at least two different feature sets that provide different, complementary information about the instance. It is proved that co-training can find an accurate decision boundary by given a small quantity of labeled data.

In the algorithm we proposed, two types of features such as Haar-like features and HOG features are considered to provide different and complementary information about the instance. Two feature sets formed with these features are used to learn the ranking SVM model accordingly. These classifiers then go through unlabeled instances, label these instances and add the error predictions into the label set of the other classifier. Then two ranking SVMs are updated simultaneously and dynamically fused into a final ranking SVM. This final classifier outputs the classified result and the location of target.

3.3. Discriminative tracker using CoRSVM

In our approach, Image patches are extracted from initial frames and several recent frames to construct the labeled training sets and then every instance is represented by a vector of Haar-like features and HOG features. The way we construct the training set has a certain similarity to the work [10, 11]. The labeled training set consist of two sets X_t^h and X_t^l. It is assumed that the patches around the location of object within a couple of pixels should be ranked higher than other patches. Then we can difine $X_t^h = \{x : \|l(x) - l^*\| \leq r\}$ and $X_t^l = \{x : \alpha \leq \|l(x) - l^*\| \leq \beta\}$, where x is an image patch and l^* is the object location in frame t. α and β control the area of sampling and $1 \leq r \leq \alpha$. We generate pairs of instances by sampling frames from X_t^h and X_t^l then construct the training set $X_t \equiv \{(x_i, y_i), (x_j, y_j) : y_i > y_j\}$. Then the training set is represented by Haar-like features and HOG features respectively and form the training set X_t^{haar} and X_t^{hog} accordingly. Afterwards, the ranking function $H(x)^{haar}$ and $H(x)^{hog}$ are learned though Eq.6 with X_t^{haar} and X_t^{hog}. Then we get two classifier C^{haar} and C^{hog}.

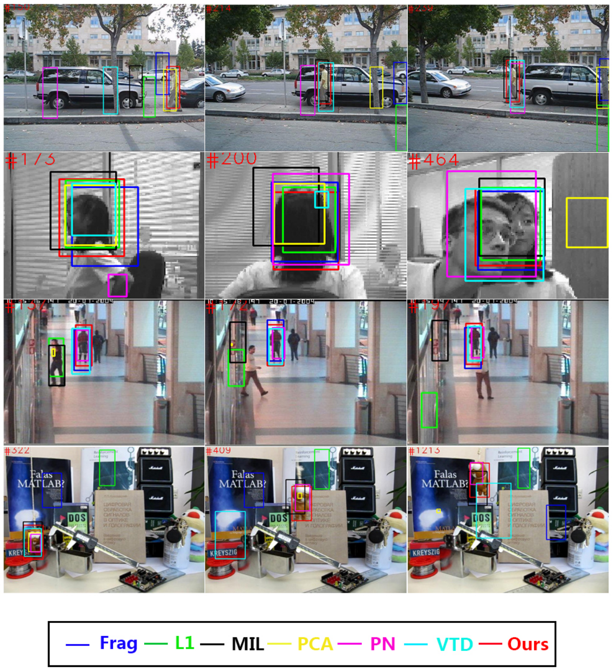

Fig. 2. Qualitative comparison results with six algorithms.

In order to fuse trained classifiers into a final classifier, we evaluate the classifiers in the labeled sample set and calculate the errors [16]. The error is defined as

$$\begin{aligned} error &= \frac{(N - Y_+) - (M - Y_-)}{2} \\ Y_+ = \sum_{i=1}^{N} sign(C(V_{i+})), &\quad Y- = \sum_{i=1}^{M} sign(C(V_{i-})) \end{aligned} \tag{8}$$

where N and M are the number of positive and negative samples accordingly, V_{i+} is the i^{th} positive, V_{i-} is the j^{th} negative sample. Once the errors have been computed, we assign weight to each classifier by

$$\begin{aligned} w_{haar} &= \frac{1 - error_{haar} + \epsilon}{error_{haar} + error_{hog} + \epsilon} \\ w_{hog} &= \frac{1 - error_{hog} + \epsilon}{error_{haar} + error_{hog} + \epsilon} \end{aligned} \tag{9}$$

where ϵ is some small constant used to avoid the divide by 0. Afterwards, the score of the final classifier C can be calculated by

$$H(x) = H(x)_{haar} * w_{haar} + H(x)_{hog} * w_{hog}. \tag{10}$$

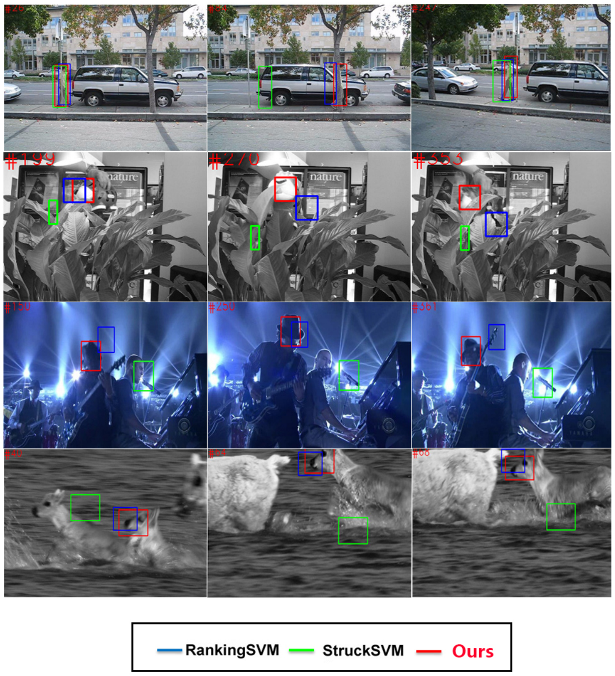

Fig. 3. Qualitative comparison results with two algorithms.

Then, the patch with score by $H_t(x)$ is calculated and we use a greedy strategy to get the object location:

$$l_t^* = l(\arg\max_{x \in X_t} H_t(x)) \tag{11}$$

4. EXPERIMENTS

We compare our algorithm with 6 state-of-the-art tracking algorithms denoted as: Frag [3], L1 [17], MIL [6], PCA [1], P-N [18] and VTD [19]. Moreover, we compare our algorithm with two kind of SVM based tracking algorithms, Ranking SVM [10] and Struck SVM [20].

Our proposed algorithm is implemented in C++. In our experiments, all parameter settings are fixed. In training stage, the location of the target is labeled by the ground truth in the first 5 frames and by the trained classifier in 5 most recent frames. We set sample parameters $r = 2$, $\alpha = 8$, $\beta = 30$, and the search radius $\gamma = 30$.

Firstly, we evaluated proposed algorithm compared with 6 algorithms on 4 challenging sequences: Davidoutdoor, Girl, Caviar1 and Lemming. The tracking results shown in Figure 2 demonstrate that the proposed tracker performs well against the other state-of-the-art algorithms. Secondly, the proposed algorithm was compared with two kind of SVM based algorithms on 4 sequences: Davidoutdoor, Tiger2, Shaking and Animal. Our tracker outperforms other SVM based methods. The compared tracking results are shown in Figure 3. And the results of quantitative comparison shown in Figure 4 also verify that our tracker consistently produces a small distance error than other trackers in most of cases.

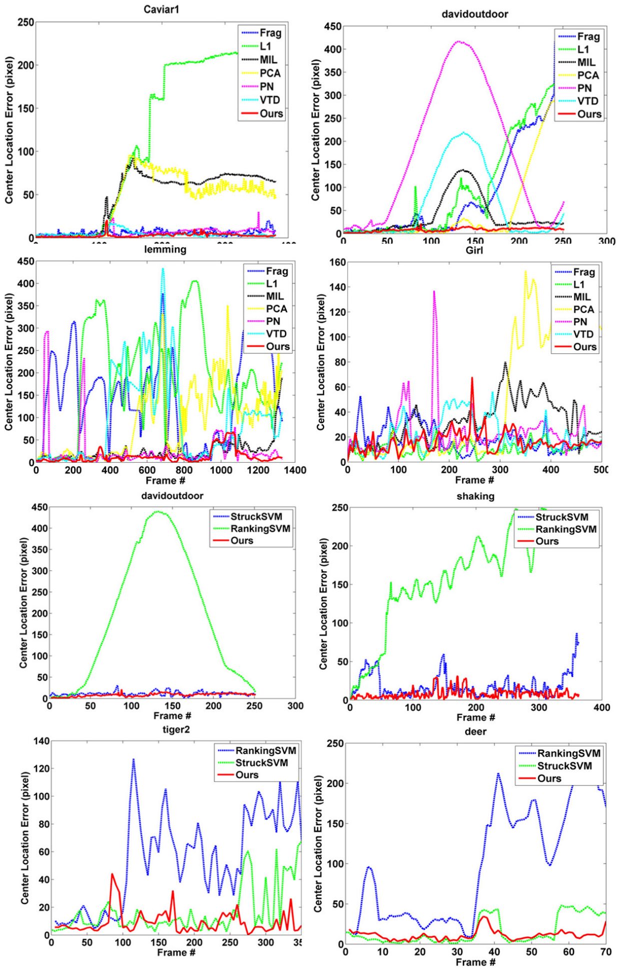

Fig. 4. Quantitative comparison results.

5. CONCLUSION

In this paper, we have formulated tracking as an online semi-supervised learning problem and proposed an online learning to rank algorithm in the co-training framework. In this approach, two ranking SVMs are built with different types of features accordingly. Moreover, they are dynamically fused into a co-training process. This semi-supervised approach is updated online to resist the occlusion and adapt to the changes of object's appearance. The results of the experiments launched on challenging sequences have demonstrated that our proposed algorithm is more robust and effective than several state-of-the-art algorithms.

6. REFERENCES

[1] David A. Ross, Jongwoo Lim, Ruei-Sung Lin, and Ming-Hsuan Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, 2008.

[2] Yi Wu, Jian Cheng, Jinqiao Wang, and Hanqing Lu, "Real-time visual tracking via incremental covariance tensor learning," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 1631–1638.

[3] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Computer Vision and Pattern Recognition, 2006*, 2006, vol. 1, pp. 798–805.

[4] S. Avidan, "Ensemble tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 2, pp. 261–271, 2007.

[5] R.T. Collins, Yanxi Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1631–1643, 2005.

[6] B. Babenko, Ming-Hsuan Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, 2010.

[7] Helmut Grabner, Christian Leistner, and Horst Bischof, "Semi-supervised on-line boosting for robust tracking," in *ECCV 2008*, vol. 5302 of *Lecture Notes in Computer Science*, pp. 234–247. 2008.

[8] Peng Yang, Qingshan Liu, and D.N. Metaxas, "Rankboost with l1 regularization for facial expression recognition and intensity estimation," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 1018–1025.

[9] Yuchi Huang, Qingshan Liu, Shaoting Zhang, and D.N. Metaxas, "Image retrieval via probabilistic hypergraph ranking," in *Computer Vision and Pattern Recognition, 2010*, 2010, pp. 3376–3383.

[10] Yancheng Bai and Ming Tang, "Robust visual tracking via ranking svm," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 517–520.

[11] Yancheng Bai and Ming Tang, "Robust tracking via weakly supervised ranking svm," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 1854–1861.

[12] Avrim Blum and Tom Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 92–100.

[13] Shipeng Yu, Balaji Krishnapuram, Rómer Rosales, Harald Steck, and R. Bharat Rao, "Bayesian co-training," in *NIPS*, 2007.

[14] Feng Tang, S. Brennan, Qi Zhao, and Hai Tao, "Co-tracking using semi-supervised support vector machines," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1–8.

[15] R. Herbrich, T. Graepel, and K. Obermayer, "Support vector learning for ordinal regression," in *ICANN*, 1999, vol. 1, pp. 97–102.

[16] Feng Tang, S. Brennan, Qi Zhao, and Hai Tao, "Co-tracking using semi-supervised support vector machines," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1–8.

[17] Xue Mei and Haibin Ling, "Robust visual tracking and vehicle classification via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 11, 2011.

[18] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 7, 2012.

[19] Junseok Kwon and Kyoung-Mu Lee, "Visual tracking decomposition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June.

[20] S. Hare, A. Saffari, and P. H S Torr, "Struck: Structured output tracking with kernels," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 263–270.