DATA LEVEL OBJECT DETECTOR ADAPTATION WITH ONLINE MULTIPLE INSTANCE SAMPLES

Bobo Zeng[⋆†] *Guijin Wang*[⋆] *Zhiwei Ruan*[⋆] *Xinggang Lin*[⋆]

[⋆] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[†] Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

ABSTRACT

In object detection, the offline trained detector's performance may be degraded in a particular deployed environment, because of the large variation of different environments. In this work, we propose a data level object detector adaptation method to new environments. By recording a small amount of offline data, it's fully compatible with offline training method and easy to implement. We re-derive an efficient MILBoost by eliminating line search in optimization and introduce it to collect online multiple instance samples, which don't require strict sample alignment. Experiment results with the human detector on public datasets illustrate the effectiveness of the proposed adaptation method. The adapted detector has good adaptation ability, while maintaining its generalization ability as well.

***Index Terms*—** detector adaptation, multiple instance samples, MILBoost, object detection

1. INTRODUCTION

In object (e.g. faces, humans) detection, learning based methods have demonstrated good performance. Typically, a detector is trained in offline with labeled training samples and some learning algorithm (e.g. AdaBoost). The performance of the detector depends largely on the representativeness of the training samples. Though increasing the training set can make it more representative, it's impossible to collect all the test data encountered in the deployed environment, since the training and test data often have disparities due to different viewpoints or scenes. Retraining the detector by adding samples in the test data is also infeasible, as the offline training is time-consuming with several hours or days.

To solve the above problem, online adaptation of offline trained generic detector to the test scenes is a good choice. There are two main challenges in the detector adaptation. The first is the effectiveness of the adaptation method. The adapted detector should not only have good adaptation ability on the new environment for a better performance, but also maintain the generalization ability to avoid over adaptation. Over adaptation makes the detector become worse for the possible changes of the new scene (such as entering of a car). Many online learning methods are proposed for detection or tracking. Online boosting is applied in [1, 2] for online feature selection on a group of selectors. Gradient-based feature selection approach [3] is proposed to update the weak classifiers using gradient descent. These approaches don't consider the offline data in feature updating, so they may cause over adaptation. Realizing this, Zhang et al. [4] use Taylor expansion to parameterize the loss function for offline data, so the weights of weak classifiers learned with DiscreteBoost is adjusted with both offline and online data. The method has good performance on their human dataset, but extending it to other boosting method is not straightforward, which limits its application.

The second challenge is adaptation efficiency. Fast and easy adaptation without much human labor is a desirable property for deploying the detector to a new environment. The most cumbersome part is the sample collection. In offline training, positive samples are labeled and aligned carefully and costly. While in the online case, such expensive labeling is infeasible for the end user, so online labeling should be minimal. Many methods [5, 6] utilize co-training to select samples in an semi-supervised way. Detected objects of one detector are used to update the other detector. This requires availability of two independent detectors with different visual cues. Also, it doesn't solve the online sample misalignment. Viola et al. [7] proposes MILBoost which doesn't require strict alignment by collecting multiple instances around a labeled sample. But the complexity is increased with the large number of instances.

In this paper, we present an effective and efficient detector adaptation method. The main contribution is two-fold. Firstly for effectiveness, we propose a data level detector adaptation method. A small amount of offline data is recorded and used in the online adaptation to prevent over adaptation. The online adaptation is fully compatible with the offline training and widely applicable to the boosting based method. Secondly for efficiency, we re-derive MILBoost under GradientBoost [8] framework to eliminate the slow line search in optimization. Besides, by firstly introducing MILBoost into adaptation, multiple instance samples can be collected and pruned with the offline detector itself, instead of two independent detectors in co-training. The end user only needs to remove

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

some false positives, so the manual labeling is easy without strict alignment.

The rest of the paper is organized as follows. Section 2 presents our detector adaptation method. Section 3 covers the online multiple instance samples with improved MILBoost. Experiment results is given in Section 4 and the paper is concluded in Section 5.

2. DATA LEVEL DETECTOR ADAPTATION

The Viola-Jones detector [9] is a seminal framework for many object detectors, including our human detector [10], which consists of cascaded strong classifiers. The strong classifier $C^l(x)$ (l is stage index in the cascade, we will omit it for convenience) has the format

$$C(x) = \text{sign}\left[H(x)\right] = \text{sign}\left[\sum_{t=1}^{T} h_t(x) - b\right] \qquad (1)$$

where $h(x)$ is the weak learner and b is the threshold. $H(x)$ is learned with some boosting algorithm (such as RealBoost, GentleBoost [11] or GradientBoost [8]) by minimizing the cost $\mathcal{L}(H(x), \mathcal{X})$ over the training set $\mathcal{X} = \{(x_1, y_1), ..., (x_N, y_N)\}, x_n \in \mathbb{R}^D, y_n \in \{-1, +1\}$. Weak learner $h(x)$ can be denoted in detail as $h(f(x), a)$ where $f(x)$ is the feature vector and a is the weak learner's parameter. For example, in decision stump based weak learner [9, 10], $f(x)$ is a scalar feature and a is the regression values of the stump. In SVM based weak learner [12], $f(x)$ is a 36D feature vector and a is the SVM coefficients. In offline boosting round t, the optimal weak learner $h_t(f(x), a)$is learned by selecting the best feature $f(x)$ from feature set $\mathcal{F}_t$ and the parameter a on the data $\mathcal{X}$.

Online adaptation can be achieved by adjusting $f(x)$ or a. But the adaptation should be very careful in order not to damage the offline detector's generalization ability. In [1, 2, 3], the feature $f(x)$ is changed to a new one with only online data. The new feature may be only discriminative for online data and the detector has high risk to be over-adapted.

In our adaptation method, we don't change the feature $f(x)$ as it's selected in offline with huge training data. Maintaining the features help to keep the detector's generalization ability. Also, we use both offline and online data in the adaptation to further prevent over adaptation. The proposed data level adaptation method records $f(x)$'s value of all the offline samples. The online adaptation has the identical routine with offline training, except the feature selection is removed. In online boosting round t, the optimal weak learner $h_t(f(x), a')$ is learned by obtaining the new parameter a' on both the online data $\mathcal{X}_{online}$ and recorded offline feature data $\{f_t(x), x \in \mathcal{X}_{offline}\}$. The recorded offline data is small. Take our offline human detector [10] as an example. It has about 1,000 scalar features, which is trained with 10,000 samples. The total data needed to record is about 38M.

The proposed method has four advantages. Firstly, offline selected discriminative features are kept unchanged and the feature data of all offline samples is used in adaptation, so the risk of over adaptation is greatly reduced. Secondly, it's highly compatible with the offline training method, so it's widely applicable to many existing offline methods and easy to implement. Thirdly, no complex parameters such as the learning rate are needed for tuning. Fourthly, the adaptation time is fast as no feature selection is required.

3. ONLINE MULTIPLE INSTANCE SAMPLES

Given the adaptation method, the next step is online sample collection, which should be manually easy so as not to involve much human labor. Unlike the general boosting algorithms, MILBoost [7] doesn't require the positive samples being strictly aligned. The original MILBoost needs line search in the optimization and slow, so we re-derive a improved MILBoost algorithm. Then we apply it to the online detector adaptation, by collecting multiple instances around the location of an object. To reduce the number of instances, we also use the offline detector itself to do instance pruning.

In MILBoost, the training set is $\{(X_1, y_1), ..., (X_N, y_N)\}$, where $X_i = \{X_{i1}, ..., X_{im}\}$ denotes a bag containing m instances and $y_i \in \{-1, +1\}$ is the bag's label. Different bags can have different numbers of instances. Given the set, the goal is to learn a model $H(x)$ the same as the general boosting algorithms. The instance probability and bag probability is defined as

$$p_{ij} \equiv p(y_{ij} = 1|x_{ij}) = \frac{1}{1 + \exp(-2H(x_{ij}))} \qquad (2)$$

$$p_i \equiv p(y_i = 1|X_i) = 1 - \prod\nolimits_j (1 - p_{ij}) \qquad (3)$$

Eq.(3) is a noise OR model meaning the bag is positive if at least one instance in the bag is positive. We use the negative log likelihood loss function

$$\begin{aligned} &\mathcal{L}(H) = \sum_{i=1}^{N} l(H) = \\ &-\sum_{i=1}^{N} (\mathbf{1}(y_i = 1)\log p_i + \mathbf{1}(y_i = -1)\log(1 - p_i)) \end{aligned} \qquad (4)$$

We solve the MILBoost following GradientBoost framework [8] and the pseudoresponse is derived as

$$\tilde{y}_{ij} = -[\frac{\partial \mathcal{L}(y_i, H(x_{ij}))}{\partial H(x_{ij})}] = \begin{cases} \frac{2p_{ij}(1-p_i)}{p_i} & \text{if } y_i = 1 \\ -2p_{ij} & \text{if } y_i = -1 \end{cases} \qquad (5)$$

In boosting round t, the update is

$$H_t(x) = H_{t-1}(x) + \rho_t h_t(x, a) \qquad (6)$$

h_t is obtained by fitting to $\{\tilde{y}_{ij}\}_{ij}$ and ρ_t can be obtained by line search to minimize $\mathcal{L}(H)$ as in [7]. But line search is slow and we seek to optimize it directly in the case when the weak learner h_t is a R-terminal regression tree. The tree has the form $h(x) = h(x, \{b_r, R_r\}_1^R) = \sum\limits_{r=1}^{R} b_r \mathbf{1}(x \in R_r)$, where $\{R_r\}_1^R$ and $\{b_r\}_1^R$ is the disjoint tree regions and its regression values respectively. Then update Eq. 6 becomes

$$H_t(x) = H_{t-1}(x) + \sum_{r=1}^{R} \gamma_{rt} \mathbf{1}(x \in R_{rt}) \qquad (7)$$

where $\gamma_{rt} = \rho_t b_{rt}$. We do separate updates in each terminal region R_{rt}

$$\gamma_{rt} = \arg\min_{\gamma} \sum_{x_{ij} \in R_{rt}} l(y_i, H_{t-1}(x_{ij}) + \gamma) \qquad (8)$$

No closed-form solution exists for Eq. 8, and similar to GradientBoost derivation we approximate it by a Newton-Raphson step. After derivation, the final result is

$$\gamma_{rt} = \sum_{x_{ij} \in R_{rt}} \tilde{y}_{ij} / [\sum_{x_{ij} \in R_{rt}, y_i = 1} 2\tilde{y}_{ij}(p_{ij} + \frac{p_{ij}}{p_i} - 1) + \sum_{x_{ij} \in R_{rt}, y_i = -1} |y_{ij}| (2 - |y_{ij}|)] \qquad (9)$$

Note the above equation equals to the GradientBoost when the bag contains a single instance, which means it can be smoothly applied to the offline data in the adaptation in Sec. 2, as the offline samples can be viewed as 1-instance bags.

In online adaptation, firstly objects should be localized for generating positive samples. Due to MILBoost, manually annotation with heavy human labor is not necessary. We use the offline trained detector to localize the objects and only the false positives are removed manually. So the annotation work is dramatically reduced. Exhaustively collecting all the instances around an object spatially in all neighbor scales will result in too many instances, and we use the offline detector to do instance pruning. Only the instances classified to be positive by the detector with high confidence value are retained.

4. EXPERIMENTS

We test the proposed detector adaptation method on the human detection task, in terms of the adaptation ability and generalization ability. The offline detector is trained as described in [10] on the Inria human training dataset [13] with general boosting. Then we perform the detector adaptation on two scenes taken from PETS2006 [14], and each has two videos for training and testing (SceneA: S4-T5-A-4 for training, S3-T7-A-4 for testing. SceneB: S4-T5-A-1 for training,S3-T7-A-1 for testing). The adapted detectors' adaptation ability on the two scene is evaluated on the respective testing video firstly. Then the adapted detectors' generalization ability is evaluted on the Inria human testing dataset.

In SceneA and SceneB, we test two adaptation methods called onlineMILSample and onlineNoMILSample. OnlineMILSample collects multiple instance samples from the training video in a 20 frames interval and about 200 bags (averagely a bag contains 8 instances) are collected. OnlineNoMILSample only use one instance with maximum confidence (this equals GradientBoost). The miss rate vs. FPPI(False Positives Per Image) results are showed in Fig. 1, with the offline detector as a comparison. From the results on both scenes, both online adaptation methods gain significant improvements compared to the offline detector, especially the onlineMILSample method (3%-7% on SceneA, 8.5%-10% on SceneB). Also, onlineMILSample is better than onlineNoMILSample, which illustrates the effectiveness of multiple instance samples collection. We also compare our method to Zhang's method [4], which also consider the offline training data by taylor expansion. Directly comparison is infeasible since our offline detector is not learned with DiscreteBoost, so it cannot be adapted with their method. We compare the relative improvement between the adapted detector and offline detector. Zhang's best result selected from different learning rates gives an improvement from 1% to 5%, which is inferior to our improvement. Our adaptation method also has high efficiency with only 30ms to update a weak classifier.

Good generalization ability is also important for an adapted detector, since the scene is not strictly fixed, and still unseen data in the scene is expected. The result on the Inria human testing set is showed in Fig. 2, in terms FPPW (False Positive Per Window) [13]. It illustrates the adapted detectors also have similar performance with the offline detector (especially in the actual working region around 10^{-4} FPPW), which means their generalization ability is not damaged.

5. CONCLUSIONS

We have proposed a data level object detector adaptation method which can be widely applied to boosting based detectors. The method is fully compatible with offline training method and easy to implement. We also introduce a re-derived MILBoost with multiple instance samples to the online adaptation process, which doesn't require the samples to be strictly aligned and is ideal for online sample collection. Experiment results with the human detector on public datasets illustrate the adapted detector has both good adaptation ability, while maintaining its generalization ability as well.

Currently, our method still needs small human labor to remove some false positive in the sample collection. In the future, we will combine MILBoost with co-training to fully automatically collecting online samples.

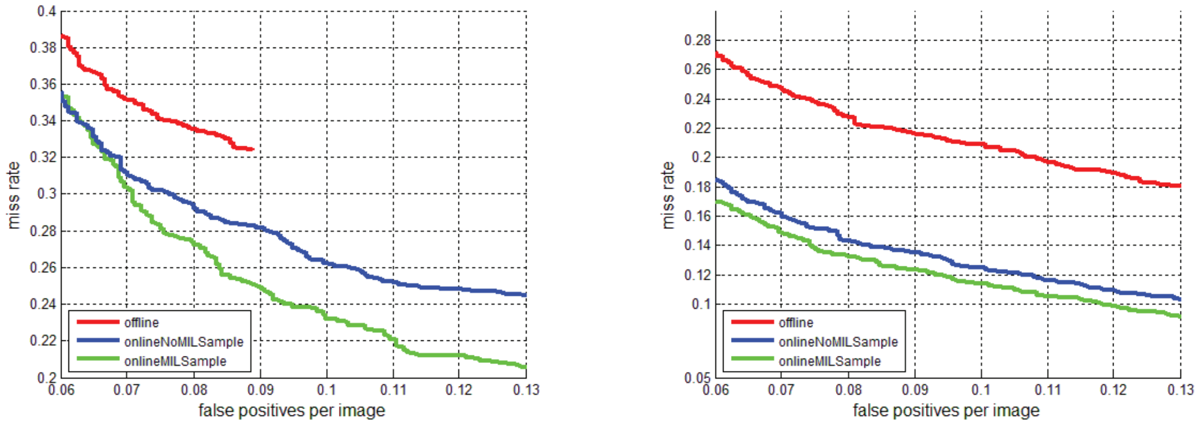

Fig. 1. The performance of the adapted human detectors with onlineMILSample and onlineNoMILSample on the test video. **Left**: SceneA results. **Right**: SceneB results.

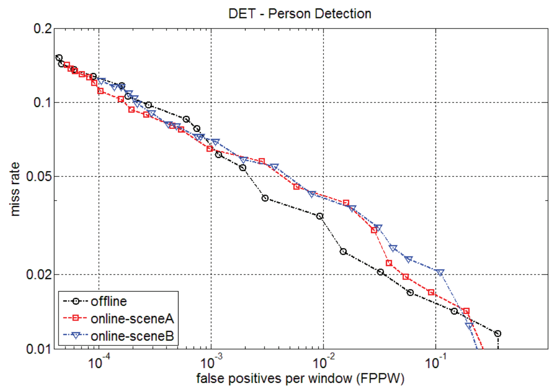

Fig. 2. The performance of the adapted human detectors in SceneA and SceneB on Inria human test set.

6. REFERENCES

[1] H. Grabner and H. Bischof, “On-line boosting and vision,” in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, pp. 260–267.

[2] C. Leistner, A. Saffari, P.M. Roth, and H. Bischof, “On robustness of on-line boosting - a competitive study,” in *ICCV Workshops, 2009 IEEE 12th International Conference on*, 27 2009-oct. 4 2009, pp. 1362 –1369.

[3] X. M. Liu and T. Yu, “Gradient feature selection for online boosting,” *2007 Ieee 11th International Conference on Computer Vision, Vols 1-6*, pp. 660–667 3027, 2007.

[4] Zhang Cha, R. Hamid, and Zhang Zhengyou, “Taylor expansion based classifier adaptation: Application to person detection,” in *CVPR 2008*, pp. 1–8.

[5] A. Levin, P. Viola, and Y. Freund, “Unsupervised improvement of visual detectors using cotraining,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 626–633 vol.1.

[6] Omar Javed, Saad Ali, and Mubarak Shah, “Online detection and classification of moving objects using progressively improving detectors,” in *CVPR’05*, 2005, pp. 696–701.

[7] P. Viola, J. Platt, and C. Zhang, “Multiple instance boosting for object detection,” *Advances in neural information processing systems*, vol. 18, pp. 1417, 2006.

[8] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[9] P. Viola and M. Jones., “Rapid object detection using a boosted cascade of simple features,” in *CVPR*, 2001.

[10] Xinggang Lin Bobo Zeng, Guijin Wang and Chunxiao Liu, “A real-time human detection system in video,” in *Technical Report*, 2011.

[11] J. Friedman, T. Hastie, and R. Tibshirani, “Additive logistic regression: A statistical view of boosting,” *Annals of Statistics*, vol. 28, no. 2, pp. 337–374, 2000.

[12] Zhu Qiang, Yeh Mei-Chen, Cheng Kwang-Ting, and S. Avidan, “Fast human detection using a cascade of histograms of oriented gradients,” in *CVPR, 2006*.

[13] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 1, Proceedings*, pp. 886–893 1223, 2005.

[14] “http://www.cvg.rdg.ac.uk/pets2006/,” .