ON-LINE BOOSTING BASED REAL-TIME TRACKING WITH EFFICIENT HOG

Shuifa Sun[**], Qing Guo[*], Fangmin Dong, Bangjun Lei

Institute of Intelligent Vision and Image Information,
China Three Gorges University, Yichang 443002, China

ABSTRACT

In this paper, a real-time visual tracking system that delivers superior performance under difficult situations is proposed. The system is based on Histogram of Oriented Gradient (HOG) within the on-line boosting framework. For environmental adaptation, the HOG feature is calculated with blocks of random scale, position and aspect ratio which form a feature pool. The on-line boosting can then select the best distinguishable features from this pool for the robust tracking. The randomness of the blocks guarantees the existence of those features. Three experiments are conducted to highlight different characteristics of this new system. The first experiment proves the validity for the system to be able to pick out the best possible HOG features. The second experiment shows its robustness against bad illuminations and small foreground background difference. The third experiment demonstrates its advancement compared with the Haar-based state-of-the-art system. All those are offered without sacrificing the computation load.

Index Terms—Histograms of Oriented Gradient (HOG), on-line boosting, integral histogram, Random feature template

1. INTRODUCTION

It has been one of the most important while difficult problems for computer vision on tracking objects efficiently and robustly in complex environments. Impressive progress has been made by treating the tracking problem as a binary classification (between the object and the background) problem [1-5]. Avidan pioneered the way to track an object by classifying the object and the background through an off-line learning strategy based on Supports Vector Machine (SVM) [1]. But it has the drawback that it cannot adapt to possible changes of the target and the background. To tackle this problem, Collins *et. al.* proposed a tracking algorithm that combines an online feature selection scheme. Further, inspired by [2], Avidan employed Adaboost as the online feature selection algorithm and introduced it into the tracking process to realize an on-line tracking system [3]. In parallel, Grabner and Bishop proposed the on-line boosting, a feature selection framework and a real-time tracking system [4]. However, Li *et. al.* [5] have proven Haar features' limitations in the Adaboost algorithm. To achieve better tracking performance, more distinguishable features should be made available for describing the object.

In 2005, Dalal *et. al.* [6] introduced HOG (Histograms of Oriented Gradient) as feature descriptors, for human detection. It has since then become the main feature used in object detection. But the original HOG feature calculation method is too computationally expensive to be applied for practical applications. Sugano has also optimized the whole algorithm by utilizing GPU [7]. To improve on this, Zhu *et. al.* made a lot of modifications to the original pedestrian detection method [8]. The computation is dramatically decreased without sacrificing the detection rate. This improved HOG was named Improve HOG (IHOG).

In this paper, we implement a real-time tracking system based on aforementioned on-line boosting framework and IHOG features. Our contributions lie on three aspects: 1. we build a feature pool on top of HOG under on-line boosting framework. The on-line best feature selecting mechanism guarantee a higher discrimination capability to the system for separating the object out of the background; 2.We calculate various HOG features with blocks of random scale, position and aspect ratio. Those features can more completely characterize an object through various transformations; 3. The HOG calculation is carefully designed in a way that the computation is efficient enough to support the real-time tracking. Other related work is [10].

2. ON-LINE BOOSTING BASED TRACKING WITH HOG FEATURES

2.1 Previous work

On-line boosting based tracking framework is mainly composed of three parts [4]: (1) A large number of weak classifiers, which are classifiers that can distinguish positive and negative samples with classification error rate slightly less than 0.5. Every weak classifier is associated with one single object feature. All those features are selected from the pre-constructed feature pool. (2) N selectors, which are actually weak classifiers with relatively low classification error rates. (3) A strong classifier, which is obtained by weighted summation of the above N Selectors, as shown by the formula (1) and (2) in below.

$$hStrong(x) = sign(conf(x)) \tag{1}$$

$$conf(x) = \sum_{n=1}^{N} \alpha_n \cdot h_n^{sel}(x) \tag{2}$$

In the above two formulas, x is one search box for the classification; $h_n^{sel}(x)$ is one weak classification result of the

* Corresponding author, email: tsingqguo@gmail.com
** Principal Corresponding author, email: watersun@ctgu.edu.cn

search box', expressed by 1 or -1. α_n is the nth selector's weight. The greater the classification error rate is, the smaller this weight is. $conf(x)$ represents the confidence value of the search box x.

On the other hand, the HOG provides a robust description about the object's contour shape. Before calculating it, a template needs to be defined. The template involves definitions on window, block and cell. For example, for human detection, Dalal *et. al.* [6] fixed the window size at 128x64 and the block size at 16x16. Each block contains 2x2 cells. Directly applying them for object tracking would bring the following problems:

(1) HOG needs high computation load because the on-line boosting algorithm may contain hundreds or even thousands of weak classifiers and each weak classifier is associated with one HOG feature. It would be impossible for the tracking system to be done in real time if each HOG feature calculation is expensive.

(2) Sufficient kinds of HOG templates are needed to get the distinguishable features for the robust tracking instead of the fixed-size blocks (16x16) for the adaptive tracking purpose. The way for HOG to achieve global characterization is to fill the search area with densely distributed small-size blocks [8]. Apparently, this kind of global characterization is very time consuming.

2.2 On-line boosting based tracking with hog feature

Following the above analysis, we propose our tracking system in this subsection. We first introduce in detail the proposed algorithm and then present the advantages of this novel method. The algorithm is performed as follows:

(1) Initialize the weak classifiers according to the initial object information. A pool of weak classifiers is built based on the HOG features, each of which is associated to a weak classifier. All weak classifiers are subsequently evaluated for the on-line boosting algorithm to pick out features that are most effective to distinguish the object and the background. Therefore, sufficient randomness should be in the feature selection procedure when creating the initial HOG feature pool. We propose to initialize the block's size, position and aspect ratio randomly with a minimum size of 12x12 within the object area. Each block contains just four cells, as shown in Figure 1. In this way, theoretically, the created feature pool would contain any kind of blocks subject to the minimum size constraint. Thus, the blocks with strong distinguishable capability can be found for the object by on-line boosting algorithm.

(2) Based on the above template, the HOG feature is calculated for the initial state of the object and served as the reference feature. The distances between the positive/negative sample feature and the reference feature are then calculated, and the probability distribution models of those two distances are built. The distance is calculated as the Euclidean distance in below:

$$f_i(v) = D_{euc}(u,v) = \sqrt{\sum_j (u_j - v_j)^2} \tag{3}$$

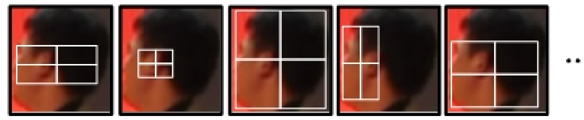


Fig. 1. The blocks in random size (from 12×12 to the size of object box), position and length-width ratio are shown in white box.

u is the reference HOG feature, v is the positive/negative sample's feature.

(3) Train the weak classifier based on the input samples. Use the updated weak classifier to classify the samples. Adopt simple threshold for the classification, and update the threshold by Equations (4) and (5).

$$h_i^{weak}(x) = sign(\mu^+ - \mu^-) \cdot sign(f_i(x) - \gamma_i) \tag{4}$$

$$\gamma_i = \left|\mu^+ - \mu^-\right| / 2 \tag{5}$$

μ^+ and μ^- are the mean values of positive and negative samples respectively. γ_i is the threshold for the ith weak classifier. Pick the weak classifier with the smallest error rate as a selector and update the samples' weights according to the new error rate.

(4) For adaptation, remove the weak classifier with the largest error rate and generate a new weak classifier from the HOG feature pool.

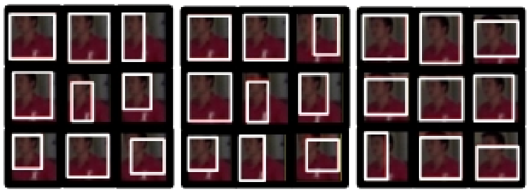
(5) Obtain a strong classifier for the next frame's object tracking by a weighted sum of all selectors, as shown in Equations (1) and (2).

(6) Update the reference feature based on the tracked object. Then return to step (2) and update the classifiers according to the new samples. Repeat these 6 steps to achieve real-time tracking.

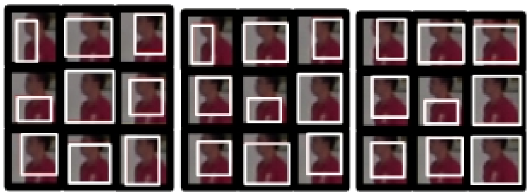
In order to accelerate the computation of the HOG features and improve the global characterization, three considerations are taken into account here based on IHOG [8]: 1. the calculation is based on the block rather than the window. Each block contains only four cells. 2. we use the integral histogram to accelerate the calculation of each cell's eigenvector. 3. L1 is employed to normalize the block. These considerations greatly improve the speed of the HOG calculation. Because each block only contains a fixed number of cells (4 in current case), the HOG calculation for one block only takes 144 additions. Therefore, no matter how big the block is, the finally formed HOG feature vector for each block always contains 36 elements.

3. EXPERIMENTS AND DISCUSSIONS

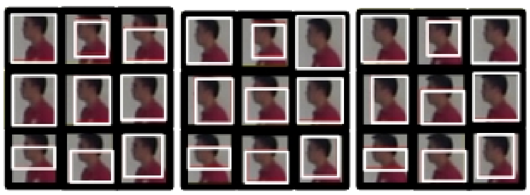
In this section, the system introduced in above is employed for tracking an object in the video data. Three different experiments are conducted. The first experiment showcases the feature selection mechanism of the online boosting with one single-environment video. The second experiment proves the robustness of the tracking by working on video data with bad illumination and small object background difference.



(a) 15^{th} frame (b) 16^{th} frame (c) 17^{th} frame



(d) 51^{th} frame (e) 52^{th} frame (f) 53^{th} frame



(g) 80^{th} frame (h) 90^{th} frame (i) 100^{th} frame

Fig.2. The black box is the object box and the HOG features of nine best selectors selected by on-line boosting in each frame are shown in white box.

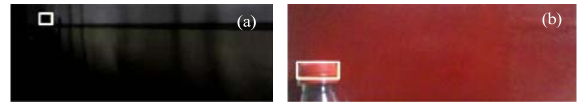
The third experiment demonstrates the advancement comparing with Haar feature based tracking by videos with illumination and scale change.

3.1 Test of HOG Feature Selection

One object is manually located for tracking. After some time, the HOG feature employed by the system for tracking is evaluated to prove the selection capability of the on-line boosting on picking the best features for robust tracking. Figure 2 shows the test results. The black box shows the tracked object. Each frame contains nine white blocks which are the first nine optimal weak classifiers selected by the current frame (The frame has updated the weak classifiers). They generate the HOG features corresponding to the first nine selectors. The following observations can be obtained from the test results:

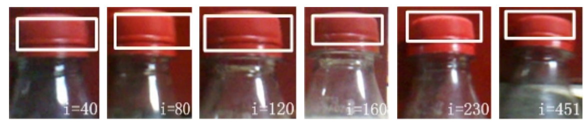
(1) Selected Features for tracking in successive frames (e.g. frames 15-17 and frames 51-53) vary slightly. Only individual feature is replaced or the feature importance is changed. This shows that the on-line boosting indeed gradually selected the best possible features for robust continuous tracking.

(2) In the case that the background differs largely with each other (e.g. compare frames 15-17 with frames 51-53). The features selected for tracking also differ very much. This shows that the on-line boosting can indeed adapt with the background to always pick up the most appropriate features for different environments.



(a) The test video under difficult illumination environment. (b) The video with the object in the environment similar.

(c)The results of tracking under difficult illumination.



(d) The tracking results in the environment similar with the object.

Fig.3. The results of tracking under difficult illumination and the environment similar with the object.

(3) The feature templates selected in 80th, 90th, and 100th frames basically do not differ too much. This indicates that the features selected by the on-line boosting have high stability when the background stays the same.

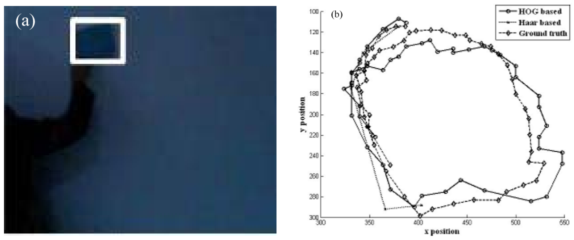
(4) All above shows that, thanks to the randomness of the template design for building the HOG feature pool, it is always guaranteed that the most appropriate tracking features are available in the pool.

3.2 Test of Robustness

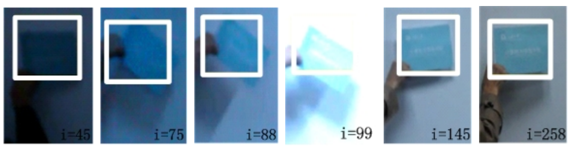
In order to test the robustness of the tracking system, we tested its performance by using one video segment with bad illumination and another video segment where the object is similar to the background, as shown in Figure 3. The tracking results in Fig. 3. (c) reveal that the system can do reliably track the human head until it is fully plunged into the darkness (as the 280th frame shows). The HOG feature characterizes the contour information of the object. Fig. 3. (d) shows the tracking process for another video. No tracking loss happened. So in the case that the object and the background have similar color, the object can be reliably tracked as long as the object has a clear outline. These two experiments demonstrates that the new system is highly robust despite of the bad illumination and the high object-background similarity.

3.3 Comparison with Haar-Based Tracking System

In order to verify the effectiveness of the proposed tracking system in practical application environments, we compare its performance with that of another Haar feature based system [4]. Two experiment videos are employed here. One is with frequent illumination variation and another is in a real surveillance situation – supermarket. Three metrics are calculated for quantitatively evaluation the performance.



(a) The test video. (b) The trajectories of HOG-based and Haar-baed algorithm and the ground truth.



(c) The frames of tracking using the HOG-based algorithm.

Fig.4. The results of tracking with illumination variation.

3.3.1 Video with illumination variation

Fig. 4. (a) shows the video tested for illumination variation. The results of Fig. 4. (b) clearly shows that the Haar feature based method lost the object in tracking while the new algorithm can continuously track it. The main reason is that the Haar feature is a statistic based on luminance and thus sensitive to illumination variation. On the other hand, the HOG feature already considers the lighting effect in the calculation process, so it can produce better tracking than the Haar feature for the object under illumination variation.

3.3.2 Video with complex background

Fig. 5. (a) shows the supermarket surveillance video. The test results in Fig. 5. (b) demonstrates that in real application environment with complex background the new system is also apparently superior than the Haar feature based system. Even when the object goes through large shape changes, the new system still can continuously track it (ref. Fig. 5(c)).

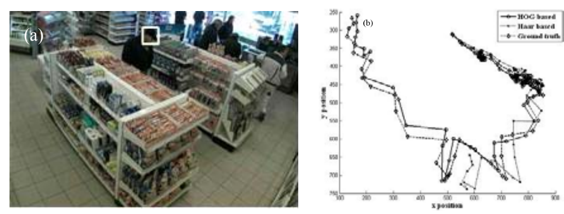
3.3.3 Quantitative evaluation

In order to evaluate the two methods quantitatively, we used the error center distance d_k together with two extended evaluation indexes: the center distance mean D_{mn}, the center distance median D_{md} [9]. The smaller the values of D_{mn} and D_{md} are, the better the tracking performance is.

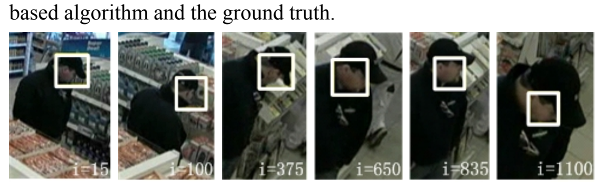
Within the scope that both methods can track the object (it is defined in this paper that d_k<50 is the scope within which the object is tracked), the values of D_{mn} and D_{md} are calculated and showed in Tab.1. The data in Table 1 show that the deviation of the HOG feature tracking results from the ground truth is far less than the deviation of which the tracking results based on Haar feature from the ground truth.

4.CONCLUSION AND OUTLOOK

This paper proposes a new on-line boosting based tracking method. By randomly selecting the scale, position and aspect ratio of the block as the feature template for calculating the HOG, an adaptive feature pool is built. The feature pool theoretically has countless feature templates, and can well represent the variations of the object appearance in all video frames. Each block in the template contains only four cells. This design greatly improves the computation efficiency. Therefore the system can be real-time. The weakness of the current realization is that the on-line boosting self-learning process will introduce incorrect information into the tracking and lead to drift of the tracking, and can eventually cause the loss of tracking. Grabner *et. al.* [11] has proposed a half supervision based on-line boosting algorithm which is formed by introducing the prior information about the classifiers into the on-line boosting framework. In the next step of work we will study further in this direction. On the other hand, the number of selectors *N* might also be learned adaptively from the feature distribution to maintain robustness through really difficult environments. This number *N* should be lower bounded to guarantee the minimum distinguish capability.



(a) The surveillance video. (b) The trajectories of HOG and Haar-based algorithm and the ground truth.



(c) The frames of tracking using the HOG-based algorithm.

Fig.5. The tracking results of the supermarker surveillance video

Tab.1. Results of three metrics

		D_{mn}	D_{md}	D_{mm}
Illumination change video	Hog-based	10.83	9.80	10.30
	Haar-based	28.12	12.04	20.08
Supermarket video monitoring	Hog-based	15.80	14.53	15.17
	Haar-based	22.40	16.78	19.69

5. ACKNOWLEDGEMENTS

This project is supported by NSFC (No.41172298, No.60972162, No.61102155, No.61105025, No.61272237, No.61272236), OYMI Research Team Plan of Hubei Province of China (No.T201002), the Graduate students scientific research innovation fund of the China Three Gorges University(No.2012CX044).

6. REFERENCES

[1] S.Avidan, “Support vector tracking,” *IEEE Trans. PAMI*, vol.26, pp.1064–1072, 2004.

[2] R.Collins, Y.Liu, M.Leordeanu, “Online selection of discriminative tracking features,” *IEEE Trans. PAMI*, vol.27, pp.1631-1643, 2005.

[3] S.Avidan, “Ensemble tracking,” *IEEE Trans. PAMI*, vol.29, 261-271, 2007.

[4] H.Grabner, M.Grabner, H.Bischof, “Real-time tracking via on-line boosting,” *Proc. BMVC 2006*, pp.47-56, 2006.

[5] Sanping Li, Zhenhan Wei, YuSen Zhang, “Adaboost-Based Face Detection Method in Hierarchical Feature Spaces,” *IJPRAI*, vol.20, pp. 435-438, 2007.

[6] N.Dalal, B.Triggs, “Histograms of oriented gradients for human detection,” *Proc. CVPR 2005*, San Diego, American, pp.886-893, 2005.

[7] H.Sugano, R.Miyamoto, Y.Nakamura, “Optimized parallel implementation of pedestrian tracking using HOG features on GPU,” *Ph.D. Research in Microelectronics and Electronics*, pp.1-4, 2010.

[8] Q.Zhu, M.C.Yeh, K.T.Cheng, S.Avidan, “Fast human detection using a cascade of histograms of oriented gradients,” *Proc. CVPR 2006*, New York, American, pp.1491–1498, 2006.

[9] C.J.Needham, R.D.Boyle, “Performance Evaluation Metrics and Statistics for Positional Tracker Evaluation,” *Proc. ICVS 2003*, Graz, Austria, pp.278-289, 2003.

[10] D.A. Klein, A. B. Cremers, “ Boosting Scalable Gradient Features for Adaptive Real-Time Tracking,” *Proc. ICRA 2011*, Shanghai, China, pp.4411-4416, 2011.

[11] H.Grabner, C.Leistner, H.Bischof, “Semi-supervised on-line boosting for robust tracking,” *Proc. ECCV 2008,* Marseille, France, pp. 234-247, 2008.