

RAPID AND ROBUST HUMAN DETECTION AND TRACKING BASED ON OMEGA-SHAPE FEATURES

Min Li, Zhaoxiang Zhang, Kaiqi Huang and Tieniu Tan

National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
{mli, zxzhang, kqhuang, tnt}@nlpr.ia.ac.cn

ABSTRACT

This paper proposes a novel method for rapid and robust human detection and tracking based on the omega-shape features of people's head-shoulder parts. There are two modules in this method. In the first module, a Viola-Jones type classifier and a local HOG (Histograms of Oriented Gradients) feature based AdaBoost classifier are combined to detect head-shoulders rapidly and effectively. Then, in the second module, each detected head-shoulder is tracked by a particle filter tracker using local HOG features to model target's appearance, which shows great robustness in scenarios of crowding, background distractors and partial occlusions. Experimental results demonstrate the effectiveness and efficiency of the proposed approach.

Index Terms— head-shoulder detection, tracking, HOG

1. INTRODUCTION

Human detection and tracking can be widely used in many applications, including people counting and security surveillance in public scenes. However, full-body human detection [1] often suffers from occlusions among individuals and scenes in which people are not necessarily standing. Hence, much recent research instead is focused on the upper part of human body.

Much work has been done on head based human detection and tracking. In [2], heads are detected by contour segmentation and ellipse fitting on the moving regions, and then each detected head is tracked by a particle filter tracker. In [3], heads are located by fast Hough transform and shape matching and tracked by decentralized trackers. Qu *et al.* [4] present an automatic system integrating head detection with particle filter for real-time multi-head tracking. However, these methods have some limitations because heads are often too small to be detected in surveillance scenes and their shapes vary greatly due to different hair-styles.

As shown in Fig. 1, the head-shoulder part has a distinctive omega-like shape in almost all view angles, thus some researchers focus on head-shoulder based human detection and tracking. Wu *et al.* [5] propose a head-shoulder detection

method based on edgelet features and tracking is achieved by data association of detection results. However, this detection-based tracking method is very time-consuming. In [6], an effective head-shoulder detection method based on boosting local HOG features is proposed and shows good performance in crowded scenes. However, it needs several seconds to process a 320×240 image, which is too slow for real application.



Fig. 1. Head-shoulder samples and their omega-like shapes

In this paper, we propose a novel method for rapid and robust head-shoulder based human detection and tracking. First, based on the work in [6], a rapid detection module combining a Viola-Jones type classifier and a HOG feature based AdaBoost classifier is used to detect head-shoulders rapidly in the predefined "Entrance" zones. Then, each detected head-shoulder is tracked by a particle filter tracker using local HOG features to model target's appearance, which shows great robustness in scenarios of crowding, background distractors and partial occlusions.

The remainder of the paper is organized as follows. In Section 2, a rapid head-shoulder detection method is introduced. Section 3 describes the details of how each head-shoulder is tracked by a particle filter tracker. Experimental results and analysis are presented in Section 4. Finally, we draw our conclusions in Section 5.

2. HEAD-SHOULDER DETECTION

It is known that Haar feature based Viola-Jones type classifier can work very fast [7], but has poor classification performance [6] in head-shoulder detection. On the contrary, HOG feature based AdaBoost classifier has much better performance [6], but works very slowly. Thus, as shown in Fig. 2, we introduce a detection strategy which can tremendously increase the working speed of the head-shoulder detector proposed by [6] while keep its classification performance. First, a Viola-Jones type classifier [7] with very high detection rate (nearly

98%) as well as relatively high rate of false positives (more than 2% False Positive Per Window) is used to exclude the obvious non-head-shoulder image patches rapidly. Then, the remaining sparse image patches are classified by a local HOG feature based AdaBoost classifier with much better classification performance. Final results are obtained by averaging the positive windows around each object.

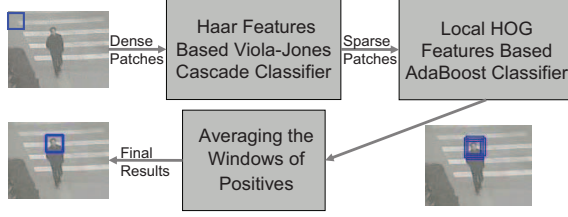


Fig. 2. The flowchart of our head-shoulder detection module

Details of how to train a Viola-Jones cascade classifier can be seen in [7]. Fig. 3 illustrates the process of local HOG feature extraction: 1) each sample is divided into 64 cells; 2) 4 adjacent cells form a block and the block stride is one cell (49 blocks in total); 3) for each cell in each block, a histogram of 8 gradient orientation bins (in 0° - 360°) is calculated and normalized within this block. Details of how to train the local HOG feature based Adaboost classifier can be seen in [6].

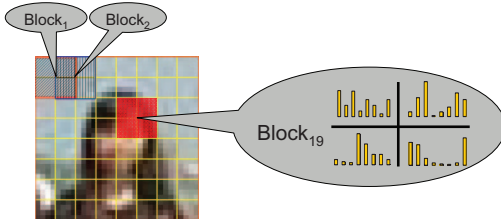


Fig. 3. Illustration of local HOG feature extraction

3. PARTICLE FILTERING BASED VISUAL TRACKING

The tracking problem of each detected head-shoulder is casted as an inference task in a Markov model with hidden state variables. The state variable X_t describes the motion parameters of the tracked target at time t . Given a set of observations $\mathcal{O}_t = \{O_1, O_2, \dots, O_t\}$, the posterior probability is formulated by Bayesian theorem as [8]:

$$p(X_t|\mathcal{O}_t) \propto p(O_t|X_t) \int p(X_t|X_{t-1})p(X_{t-1}|\mathcal{O}_{t-1})dX_{t-1} \quad (1)$$

The tracking process is governed by the observation model $p(O_t|X_t)$ and the dynamic model $p(X_t|X_{t-1})$ between two states. The sampling importance resampling (SIR) [8] particle filter is utilized to approximate the distribution over target's location.

3.1. Dynamical Model

The state variable consists of three motion parameters: $X_t = \{x_t, y_t, s_t\}$, where x_t, y_t, s_t denote x, y translation and scale at time t respectively. we use an adaptive velocity with a Gaussian noise model to describe the dynamical model: $X_t = X_{t-1} + v_t + \omega$, where v_t is the average speed of the target in the last N_v frames and $\omega \sim N(0, \Sigma)$ is a Gaussian noise.

3.2. Observation Model

The observation model $p(O_t|X_t)$ measures the probability of a sample observation belonging to the tracked target model. Usually, it involves three important aspects: object representation, distance measurement and model update.

Object Representation: refers to how to model a sample's appearance. In our work, given an image patch I_t specified by X_t , the local HOG features are extracted from this patch and used to model its appearance (O_t).

Distance Measurement: refers to how to compute the distance from a new observation to the target model. Suppose the target model at time t is $\hat{O}_t$, and the observation of any given sample (candidate target) at time t is O_t , then we use Bhattacharyya coefficient [9] to measure the similarity between the two feature vectors as:

$$S(O_t, \hat{O}_t) = \sum_{u=1}^m \sqrt{\hat{O}_t^u O_t^u} \quad (2)$$

The geometric interpretation of (2) is the cosine of the angle between the m -dimensional unit vector $(\sqrt{\hat{O}_t^1}, \dots, \sqrt{\hat{O}_t^m})^\top$ and $(\sqrt{O_t^1}, \dots, \sqrt{O_t^m})^\top$.

Using (2), the distance between any given observation O_t and the target model $\hat{O}_t$ can be defined as: $d(O_t, \hat{O}_t) = \sqrt{1 - S(O_t, \hat{O}_t)}$. Consequently, the observation model $p(O_t|X_t)$ is formulated as:

$$p(O_t|X_t) \propto \exp(-d(O_t, \hat{O}_t)). \quad (3)$$

Model Update: To track a target robustly, the target model $\hat{O}_t$ should be updated over time to adapt to appearance changes. Let $\{X_t^{(i)}, \omega_t^{(i)}\}_{i=1}^{N_s}$ is a set of particles at time t , where the weights $\omega_t^{(i)}$ are computed based on the observation model: $\omega_t^{(i)} = p(O_t^{(i)}|X_t^{(i)})$. We estimate the target state at time t by the sample with the highest weight: $\hat{X}_t = X_t^{(j)}|_{\omega_t^{(j)} = \max(\omega_t)}$. Then, target model is updated by:

$$\hat{O}_{t+1} = \beta \hat{O}_t + (1 - \beta) O(\hat{X}_t) \quad (4)$$

where β is the forgetting factor.

4. EXPERIMENTAL RESULTS

Three experiments are conducted in this section. Each test video has the resolution of 320×240 . The number of particles

N_s and the forgetting factor β for each tracker are set to 150 and 0.99 respectively. All programs are implemented in C++ and run on a computer with a P4 2.5 GHz CPU.

In the first experiment, the performance of our head-shoulder detector and that of the one proposed by [6] are compared on an open head-shoulder dataset provided in [6]. For simplicity, our detector is referred to as the "VJ+HOG" detector, and that of [6] is referred to as the "HOG" detector. The ROC curves of the two detectors are shown in Fig. 5. It can be seen that the classification performance of the "VJ+HOG" detector is almost equivalent to that of the "HOG" detector. However, the "VJ+HOG" detector works much faster than the "HOG" detector. As shown in Table 1, for a dense scan (25,000 sub windows) on a 320×240 image, the "HOG" detector consumes 8,600ms, while the "VJ+HOG" detector consumes only 290ms; for a sparse scan (2,500 sub windows), the consumed time is 840ms and 40ms respectively. Fig. 6 shows some detection results of our final detector in some photos.

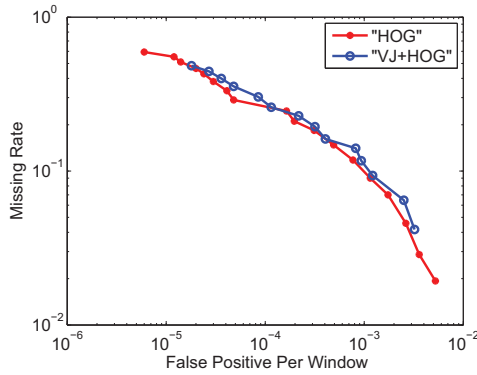


Fig. 5. Classification performance comparison of the "VJ+HOG" detector and the "HOG" detector

Table 1. Time consuming on a 320×240 image

Detector Type	Dense Scan	Sparse Scan
"HOG"	8,600 ms	840 ms
"VJ+HOG"	290 ms	40 ms

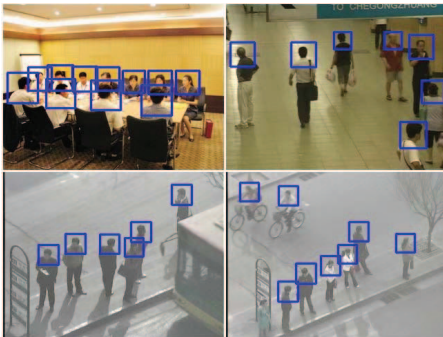


Fig. 6. Some detection results of our final detector

In the second experiment, we aim to evaluate the performance of our local HOG feature based particle filter tracker independently. Since crowding and occlusion are two most common factors that may cause tracking failures in real applications, a sequence (shown in Fig. 4) in which a person is walking in an overcrowded scene and undergoing occasional occlusions is used to test our tracker. As comparison, the results of a mean-shift based tracking algorithm [10] are also given. This method embeds a feature evaluation mechanism in a mean-shift tracking system to adaptively select the top-ranked color features for tracking, which is robust to occlusions and resistant to distractors to some extent. We refer to it as the OFS (on-line feature selection) method. As shown in Fig. 4, the color feature based OFS method fails to track the target in the overcrowded scene before frame #127, while our tracker track the target consistently correctly to the end of the test sequence even though there are some occlusions (#245, #274). This is because color features are often not distinctive in crowded scenes with many distractors. However, the local HOG features can capture some distinctive geometric structures and thus perform better in head-shoulder tracking in crowded scenes.

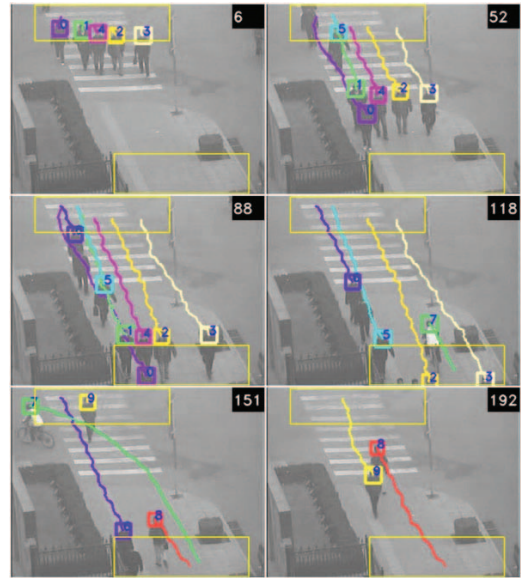


Fig. 7. Detection and tracking results of our method in a video from a real surveillance scene (the big yellow rectangles in each sub figure are the "Entrance" zones)

Table 2. Time consuming per frame in the third experiment

Module	Time
Detection	15ms
Tracking	27ms/object

In the third experiment, a challenging video from a real surveillance scene is used to test the whole proposed head-shoulder detection and tracking method. As shown in Fig. 7, head-shoulder detection is only conducted in the prede-



Fig. 4. Tracking results of our tracker (Row 1) and the OFS method[10] (Row 2) in a sequence in which the target is walking in an overcrowded scene and undergoing occasional occlusions

finer "Entrance" zones; then, every detected head-shoulder is tracked by a local HOG feature based particle filter tracker. As we can see, targets in this video are in low resolution and have very similar appearance. However, our system successfully detects 9/10 emerged targets except one whose head-shoulder shape is not observable because of merging into the torso of target No.6 (#88). All detected targets are robustly tracked until they disappear, even in cases of crowding (#52, target No.1 surpasses target No.0 and No.4 causing great shape overlap with them), background distractor (#88, the head-shoulder region of target No.3 almost merges into the region of a background object with similar color) and fast moving (#151, target No.7 is riding a bike). Moreover, our system can work very fast. As shown in Table 2, our detection module consumes only 15 ms/frame and the tracking module consumes about 27 ms/object per frame. Therefore, our method can work at more than 10 frames/second if the number of tracked targets is less than 4.

Demonstration videos can be seen in [11].

5. CONCLUSIONS

We have proposed a fast and robust method for human detection and tracking based on the omega-like shape features. There are two modules in this method: one is a head-shoulder detection module which combines a Viola-Jones type classifier and a local HOG feature based AdaBoost classifier to detect objects rapidly and effectively; the other is a particle filtering based visual tracking module using local HOG features to model target's appearance, which shows great robustness in scenarios of crowding, background distractors and partial occlusions. Since this method can rapidly detect and robustly track people's head-shoulders automatically, it can be used in many real applications.

Acknowledgement

This work is funded by research grants from the National Basic Research Program of China (2004CB318110), the

National Science Foundation (60605014, 60875021), the National Natural Science Foundation of China (60736018, 60723005) and National Laboratory of Pattern Recognition (2008NLPRZY-2). The authors also thank the anonymous reviewers for their valuable comments.

6. REFERENCES

- [1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR'05*, 2005.
- [2] Y. Jin and F. Mokhtarian, "Data fusion for robust head tracking by particles," in *IEEE Inter. Workshop on VS-PETS*, 2005.
- [3] J. S. C. Yuk, K. K. Wong, and et al., "Real-time multiple head shape detection and tracking system with decentralized trackers," in *Proc. ISDA'06*, 2006.
- [4] W. Qu, N. Bouaynaya, and D. Schonfeld, "Automatic multi-head detection and tracking system using a novel detection-based particle filter and data fusion," in *Proc. ICASSP'05*, 2005.
- [5] B. Wu and R. Nevatia, "Tracking of multiple humans in meetings," in *Proc. CVPRW'06*, 2006.
- [6] M. Li, Z. Zhang, K. Huang, and T. Tan, "Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection," in *Proc. ICPR'08*, 2008.
- [7] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. CVPR'01*, 2001.
- [8] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [9] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *Proc. CVPR'00*, 2000.
- [10] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Trans. PAMI*, vol. 27, no. 10, pp. 1631–1643, 2005.
- [11] Authors, "Demos," <http://limin81.cn/hsdemos.zip>.