Active Learning Based Pedestrian Detection in Real Scenes

TaoYang JingLi QuanPan Chunhui Zhao Yiqiang Zhu
College of Automatic Control, Northwestern Polytechnical University, Xi'an, China , 710072
{yangtaonwpu , jinglinwpu}@163.com , {quanpan, zhaochunhui}@nwpu.edu.cn

Abstract

This work presents an active learning based method for pedestrian detection in complicated real-world scenes. Through analyzing the distribution of all positive and negative samples under every possible feature, a highly efficient weak classifier selection method is presented. Moreover, a novel boosting architecture is given to get satisfied False Positive Rate (FPR) and False Negative Rate (FNR) with few weak classifiers.

A unique characteristic of the algorithm is its ability to train special cascade classifier dynamically for each individual scene. The benefit is that the trained classifier will only focus on the differences between the positive samples and the limited negative samples of each individual scene, thus greatly reduce the complexity of classification and achieve robust detection result even with few classifiers.

A real-time pedestrian detection system is developed based on the proposed algorithm. The system produces fast and robust detection results as demonstrated by extensive experiments which use video sequences under different environments.

1. Introduction

Pedestrian detection, tracking and activity analyzing is an essential component of automatic video surveillance system. The most common approach to detect and track pedestrians is to first detect them using background subtraction, and then establish correspondence from frame to frame to find the tracks of the objects [1]. Despite its popularity, background subtraction [1,2] based methods still lack the robustness to handle specific events such as shadow, reflection in water, camera shaking in the wind, partial occlusion and bad weathers.

Recently, another type of method based on visual detection has got lot of attention. Approaches in this class use machine learning to construct a detector from a large number of training samples. The detector is scanned over the entire input image in order to find a pattern of intensities, which is consistent with the target object. Local features [3,4] or part detectors[5,6] are used here.

The methods above use a static offline trained detector in different scenes, although similar algorithms and systems works well for face detection, their performances are not satisfied in pedestrian detection due to changes in human body pose, clothing, low resolution and various complicated background. Thus how to make pedestrian detection really work is a serious problem that must be conquered.

Instead of training a static detector, we tackle in this paper the problem of build an active learning based boosting cascade detector online in various dynamic scenes. The main advantage of this approach is that the trained classifier will mainly focus on the differences between the negative samples of each individual scene and the positive samples, thus greatly decrease the complexity of classification problem in boosting.

Novel contributions of this paper include: (1) presenting an online updated detector that integrate background modeling and boosting training together to realize a very low false positive rate under different scenes, (2) designing a fast weak classifier selection approach which greatly reduce the selection time, and (3) given a new architecture in boosting training which achieves satisfied FPR and FNR directly with few weak classifiers.

This paper is organized as follows. Section 2 describes our active learning based pedestrian detection algorithm. Experimental results are shown and discussed in section 3. We finish the paper with the conclusions and some ideas of further work in section 4.

2. Active learning based pedestrian detection

2.1 System Overview

We integrate the background modeling and boosting cascade into one framework. The flowchart of our approach is shown in Figure 1. In the initially offline training step, the approach build a basic pedestrian detector by training method (see section 2.2, 2.3). In the online detection step, the background is estimated dynamically. Although lots of background modeling methods may integrate into our system [1,2,7], considering the processing speed of the overall system, we adopt one of our previous works [7], a two-level background maintenance algorithm, for real-time background estimation. Detections in the background are taken as false alarms. Once the number of false alarms is a higher than a threshold, detected sub images in the background image will be used as negative samples to train additional levels in the basic cascade detector (Figure 1, right part). Here a highly efficient weak classifier selection

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

algorithm is presented (see section 2.2) to build weak classifier pool for boosting. In addition, a new architecture is developed which achieves satisfied FPR and FNR directly with few weak classifiers. During the boost training process, if the current FNR is higher than a threshold T_nr, weak classifiers with lowest FNR will be added. Once the FNR satisfied, the threshold in final strong classifier is automatically selected to reduce the false positive rate as much as possible while maintain the desired detection rate after training (see section 2.3).

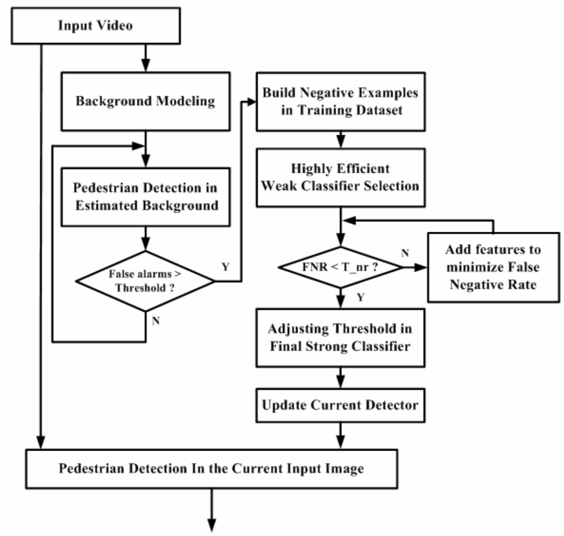


Figure 1. Pedestrian detection system diagram.

2.2 Fast weak classifiers selection

During the online pedestrian detection process, the estimated background is used to evaluate the performance of current trained detector. Once the false alarms in the background image is higher than a certain threshold, the detected regions will be taken as negative samples, meanwhile additional classifier levels will be trained and added into the existing cascade detector. Boosting with all weak classifiers has heavy computational burden and doesn't suitable for online update application. To increase the processing speed, it's reasonable to build a weak classifier pool with lowest errors. Here we present a novel fast weak classifier selection approach, which achieves the same selection result while greatly reduces selection time.

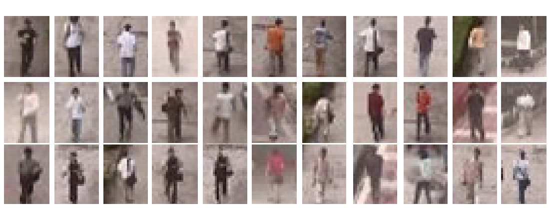


Figure 2. Samples of pedestrian images used for training

We used ten of the sequences to create a training set which has 4430 positive samples. Each positive example is a 24x32 pedestrian image (Figure 2). Rectangles features [4] are selected for its powerful representative ability and efficiency in conjunction with integral image.

Let $f = \{f(1),...,f(F)\}$ the feature set, $p_m^+(t)$ and $p_m^-(t)$ the feature value probability density function of the mth feature $f(m)$ in positive sample set S^+ and negative sample set S^- separately (1).

$$p_m^+(t) = k^+ / n^+, \qquad p_m^-(t) = k^- / n^- \tag{1}$$

where k^+ and k^- represents the number of the positive and negative samples that has feature value t, in this paper, the entire feature value scope is $t \in [0,6000]$. Then the distribution function $F_m^+(t)$ and $F_m^-(t)$ of feature value can be computed by (2).

$$F_m^+(t) = \sum_{T \le t} p_m^+(T), \quad F_m^-(t) = \sum_{T \le t} p_m^-(T) \tag{2}$$

A weak classifier h_m consists of a feature $f(m)$, a threshold θ_m and a parity λ_m indicating the direction of the inequality sign [5].

$$h_m = \begin{cases} 1 & if \quad \lambda_m f(m) < \lambda_m \theta_m \\ 0 & \quad otherwise \end{cases} \tag{3}$$

To build a weak classifier h_m, we need to find a suitable threshold θ_m with lowest error. Instead of searching all possible values in the entire feature value space $[0,6000]$ which has very heavy computational costs, we directly find the precise searching scope $[ts\min, ts\max]$ through analyzing the distribution function $F_m^+(t)$ and $F_m^-(t)$. To ensure the performance of boosting, the correct classification rate of selected weak classifier must be better than 60%. Let T_{FNR} and T_{FPR} denotes the maximal value of false negative rate and false positive rate of the selected weak classifier, and the equation (4) must be satisfied.

$$(T_{FNR} \cdot n^+ + T_{FPR} \cdot n^-)/(n^+ + n^-) \le 0.5 \tag{4}$$

Because the goal of boosting is to construct a strong classifier that can realize less than 1% false negatives and 50% false positives, to ensure the low false negatives, we add more constraints on the threshold of FNR in feature selection, and set T_{FNR} as 30% and T_{FPR} as 50% in the experiments. Once T_{FNR} and T_{FPR} are selected, the precise searching scope $[ts\min, ts\max]$ in feature value can be computed by equation (5) and (6).

$$ts\min' = \begin{cases} \min_t \{t \mid F_m^-(t) > 1 - T_{FPR}\} & if \quad \lambda_m = 1 \\ \min_t (t \mid F_m^+(t) > 1 - T_{FNR}) & if \quad \lambda_m = -1 \end{cases} \tag{5}$$

$$ts\max' = \begin{cases} \min_t \{t \mid F_m^+(t) > T_{FNR}\} & if \quad \lambda_m = 1 \\ \min_t (t \mid F_m^-(t) > T_{FPR}) & if \quad \lambda_m = -1 \end{cases} \tag{6}$$

Only if $\lambda_m = 1$ and $ts\max' > ts\min'$ or $\lambda_m = -1$ and $ts\max' > ts\min'$, we set $ts\min, ts\max$ as (7)

$$ts\min = ts\min',\ ts\max = ts\max' \tag{7}$$

Based on the distribution function $F_m^+(t)$ and $F_m^-(t)$ and selected searching scope $[ts\min, ts\max]$, the threshold θ_m for feature p_m can be computed by (8).

$$\theta_m = \begin{cases} \underset{t}{\operatorname{argmin}}\left(\dfrac{F_m^+(t)\cdot n^+ + (1-F_m^-(t))\cdot n^-}{n^+ + n^-}\right) & if \quad \begin{matrix}\lambda_m = 1 \\ \& t \in [ts\min, ts\max]\end{matrix} \\ \underset{t}{\operatorname{argmin}}\left(\dfrac{(1-F_m^+(t))\cdot n^+ + F_m^-(t)\cdot n^-}{n^+ + n^-}\right) & if \quad \begin{matrix}\lambda_m = -1 \\ \& t \in [ts\min, ts\max]\end{matrix} \end{cases} \tag{8}$$

Note that compared with tradition selection methods that need to compute the classify error of every single feature with all possible threshold in feature value space, the proposed method (8) can directly find most suitable threshold θ_m for feature p_m from the distribution function $F_m^+(t)$ and $F_m^-(t)$, and select a weak classifier with lowest error for all positive and negative samples. Moreover, finding the precise search scope (5),(6),(7) can make the threshold selection process more efficient and effective.

2.3 Direct Adaboost training

The key of designing each level of cascade pedestrian detector is to achieve a very high detection rate (higher than 99%) and a moderate false positive rate (less than 60%). Although the false positive rate is not difficult, the detection rate is hard to accomplish. Viola-Jones method [4] reduces the FNR and FPR by adding the features with minimum weighted error over the whole train set and then increase the detection rate by adjust the threshold μ in the strong classifier (9) for all layers of the cascade.

$$h = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t \geq \mu \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases} \tag{9}$$

where $\alpha_t = \log 1/\beta_t$ [4]. This method[4] is an indirect way to get high detection rate. During their training process, large numbers of weak classifiers are used to perform a high detection rate but also a very low false positive rate which is not necessary.

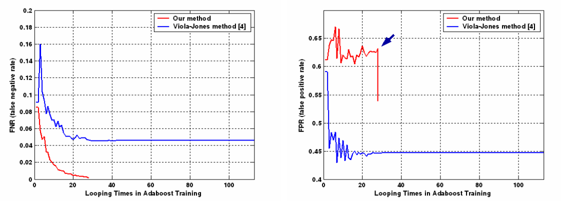
In this paper, we present a novel direct training approach. In the initial loops of training, classifiers with lowest FNR will be selected from weak classifier pool P (built by method in section 2.2) to realize high detection rate quickly. Once the current ensemble's FNR is less than 0.2%, we try to enlarge the threshold μ to reduce FPR while maintain FNR. Since the $\sum_{t=1}^{T} \alpha_t > 0$ in boost training, we modified (9) to (10).

$$h = \begin{cases} 1 & \psi \geq \mu \\ 0 & otherwise \end{cases} \tag{10}$$

Where $\psi = \sum_{t=1}^{T} \alpha_t h_t / \sum_{t=1}^{T} \alpha_t$. Through analyzing the distribution of ψ in positive train set S^+, we can directly get the optimal threshold $\mu_{\max}$ by (11).

$$\mu_{\max} = \min_{\psi}\{\psi \mid F(\psi) > 0.002\} \tag{11}$$

where k^{ψ} represents the number of the positive samples that has feature value ψ, $F(\psi) = \sum_{T \leq \psi} p(T)$ is the distribution function and $p(\psi) = k^{\psi} / n^+$ is probability density function of variable ψ .



a b

Figure 3. FPR and FNR curve of our method an [4].

We contrast the FNR and FPR curve of our method and Viola-Jones [4] during training a certain level in the cascade pedestrian classifier (Figure 4). The training set of for Figure 3 is built by 4430 positive samples and 4430 negative samples. Both methods select features from the same weak classifier pool P. Note that our method only uses 28 weak classifiers to get satisfied FNR as 0.0015 for all 4430 positive samples (Figure 3.a, red line), while Viola-Jones [4] method (Figure 3.a, blue line) doesn't reach this target even with more than 100 weak classifiers. In addition, through optimal threshold $\mu_{\max}$ selection by (11), the FNR of our method drop from 63.3% to 54.3% (Figure 3.b, blue arrow) while maintain the FNR after training.

3. Experimental Results

We used ten sequences to create a training set, and labeled 4430 pedestrian images as positive samples (shown in Figure 2). The size of each example is 24x32. The video sequences are captured in various complicated real scenes under different weather conditions. The pedestrian detection system has been implemented on Pentium 4 3.0Ghz, 768M computer. The system can detect pedestrian at 7 frames/second over a 320x240 pixel image without using object motion information, and the processing speed reaches 20 frames/second with foreground segmentation

- **Comparison of computing time**

In this experiment, 4430 positive samples and 4430 negative samples are used for training. The entire feature space contains 138,600 features, and for each feature there are 6000 possible thresholds. Thus the total number of possible weak classifiers is 138,600x6000. It costs 8 hours 17 minutes and 13 seconds for searching in the entire feature space, and 1 hour 22 minutes and 52 seconds for searching in the scope of minimal and maximal feature value in each feature. In contrast, the proposed method mentioned in section 2.2 only spends 5 minutes and 58 seconds to finish weak classifier selection, and the computational time is only 1.2% of method in the third row of Table 1. Finally, 937 weak classifiers have been selected as feature pool by the proposed method and the boosting time is 4 minutes and 12 seconds. As a result, the total time for build a new stage in the cascade classifier is only 10 minutes and 10 seconds in our system, which is fast enough for online classifier update application due to slow scene changes.

◆ **Pedestrian detection in real scenes**

Figure 4 shows the processing steps of the proposed active learning based pedestrian algorithm. For the input image of a new scene (Figure 4.a), the outputs of the basic detector has lots of false alarms (Figure 4.b). When the background of input video is estimated (Figure 4.c), detection results on the estimated background (Figure 4.d) are taken as negative samples for training additional stages of cascade classifier by the proposed algorithm. Figure 4.e displays the multiple detections by updated classifier, and Figure 4.f shows the final results after detections integration and filtering.

We report the detection results of the system under real scenes (road, parking lot, crossing) in Figure 5. The video sequences are captured in different seasons (winter (Figure 5.a) and summer (Figure 5.b,c,d)) with different weather conditions(snow(Figure 5.a), sunshine (Figure 5.b,c), rain (Figure 5.d)). The proposed algorithm successfully handles many serious problems like heavy snow(Figure 5.a), shadows (Figure 5.b,c), water reflections (Figure 5.d), multiple close persons (Figure 5.b), that most background subtraction based pedestrian detection methods may fail in these conditions. Moreover, it's hard to those detection methods to train a static detector offline that can perform high detection rate and low false alarms in all those scenes above, while based on fast weak classifier selection and training, our method updated the classifier for each individual effectively and efficiently, so as to achieve satisfied detection results under various scenes.

4. Conclusion

We have presented a novel active learning based pedestrian detection algorithm which combines background modeling and boosting training to build a robust detector for each individual scene online.

A new fast weak classifier selection is presented which achieves same selection results while greatly reduce the feature selection time (computational time is only 1.2% of traditional selection method). A new architecture of boosting training is given. Compared with Viola-Jones method [4], the proposed architecture can reach the high detection rate (higher than 99%) and moderate false positive rate (less than 60%) directly with fewer classifiers. Post processing techniques as multiple detections integration and temporal smoothing and are used to filter false alarms. Experiment results on a large and complex dataset are satisfied. In the future we would like to integrate the approach with objects tracking algorithm to improve tracking performance under difficult conditions.

Acknowledgements

The work presented in this paper was sponsored by Aeronautics Foundation (#03D53032) and National Natural Science Foundation (#60372085) of China .

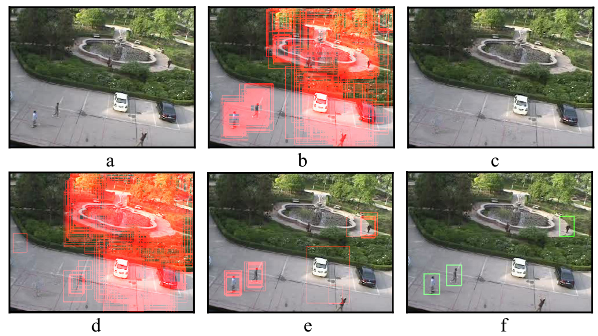


Figure 4. Online learning of pedestrian detection.

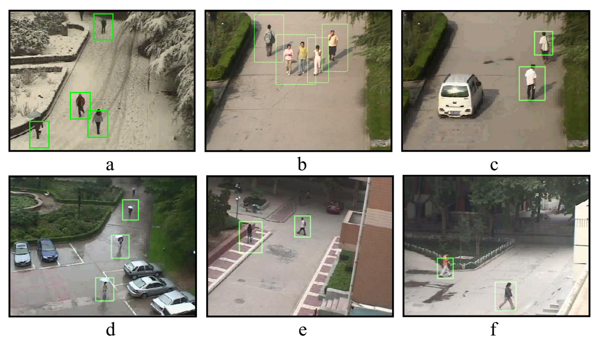


Figure 5. Pedestrian detection results in complicated real scenes

References

[1] Collins et al. A System for Video Surveillance and Monitoring. VSAM Final Report, *Technical report CMU-RI-TR-00-12*, Carnegie Mellon University, May, 2000.

[2] Rita Cucchiara, Costantino Grana,Massimo Piccardi, Andrea Prati, Detecting Moving Objects, Ghosts, and Shadows in Video Streams, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 10, pp.1337-1342, October 2003.

[3] Viola, P., Jones, M.J., Snow, D., Detecting Pedestrians Using Patterns of Motion and Appearance, *Proceedings of the IEEE International Conference on Computer Vision* , Vol 2, pp.734-741, October 2003.

[4] Viola, P. and Jones, M.J., Rapid object detection using a boosted cascade of simple features. *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference(CVPR'01)*, Vol 1, pp.511-516,2001.

[5] Mikolajczyk, K. , Schmid, C. and Zisserman, A., Human detection based on a probabilistic assembly of robust part detectors, *Proceedings of the European Conf. on Computer Vision*, Vol 1, pp.69-81, 2004.

[6] Bo Wu, Nevatia, R, Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors, *Proceedings of the IEEE International Conference on Computer Vision,* Vol 1, pp.90- 97, Oct 2005, Beijing , China.

[7] Tao Yang, Stan Z.Li ,Quan Pan, JingLi. Real-time Multiple Object Tracking with Occlusion Handling in Dynamic Scenes. *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, Vol 1, pp.970-975, San Diego, CA, USA.