Real-time Pedestrian Detection based on GMM and HOG Cascade

Moonyong Jin[1], Kiseon Jeong[1], Sook Yoon[2] and Dong Sun Park[1,3]

[1]Department of Electronic Engineering, Chonbuk National University, Jeonju, Jeonbuk, Korea
[2]Department of Multimedia Engineering, Mokpo National University, Jeonnam, Korea
[3]IT Convergence Research Center, Chonbuk National University, Jeonju, Jeonbuk, Korea

ABSTRACT

Most of the human detection methods are using HOG (Histogram of Oriented Gradients). In the case of fixed camera environment, it is possible to make background model using GMM (Gaussian mixture model) and easily extract motions using background subtraction. However, it is difficult to recognize pedestrians among extracted motions. In this paper, we propose an efficient coarse-to-fine pedestrian detection framework which combines motion detection and HOG cascade to make a faster pedestrian detector. Firstly, motion detection is used as the coarse detection in order to reduce the area of interest to be covered by the pedestrian detector. Then HOG cascade which detects pedestrians is executed only on the blobs or ROIs selected from the coarse detection. The experimental results on PET2009 768X576 dataset show that proposed method of which processing speed is 11.46 fps is 7.5 times faster than HOG and 2.2 times faster than HOG cascade.

Keywords: Pedestrian detection, GMM, HOG, HOG based cascade, Coarse-to-fine detection.

1. INTRODUCTION

Visual object tracking is one of the most important subjects in surveillance system [1]. We are able to obtain useful information about object through object tracking. In order to track objects in real-time, however, objects should be quickly and robustly detected in advance. Unsatisfying results of object detection affect performance of object tracking.

Papageorgiou et al [2] described a pedestrian detector based on polynomial SVM using rectified Haar wavelets. Gavrila et al [3] used extracted edge images and matched them to a set of learned models using chamfer distance. Dalal et al [4] presented a human detection algorithm with excellent detection results. Their method uses a dense grid of histogram of oriented gradients (HOG) feature that represents objects. Since this representation is proved to be powerful to classify humans using linear SVM, HOG is a quite popular method to detect pedestrians. However, their method is difficult to be real-time system because obtaining HOG feature is time-consuming. Their method can only process 768X576 images at 1.5 fps using a very dense scanning methodology.

Many researchers have extensively studied how to reduce computation time and make real-time systems. Viola et al [5] proposed a cascade structure using Adaboost with Haar-like wavelets. A cascade structure consists of many stages and each stage can reject an input window when its computed threshold is less than the predefined rejection threshold. G. Xu et al [6] combined HOG and the edge factor which is the normalized number of edge pixels in each window. When the edge factor is below the threshold, it can eliminate sub-windows. Dollar et al [7] proposed a hybrid approach which is a technique to avoid constructing such a finely sampled image pyramid without sacrificing performance. Zhu et al [8] presented a method that integrates a cascade of rejectors with HOG features to achieve a fast and accurate human detection system. Prisacariu et al [9] used an additional hardware, called a GPU, to speed up.

However, previous works have some problems in terms of detection rate and speed. Even though the Haar-wavelet feature with Adaboost is a fast method to detect objects, it achieves much lower detection accuracy than HOG feature. An Edge-based method is useful only for the background that contains uniform information, not complicated background. When a background is cluttered, its sub-windows cannot be eliminated or rejected by edge factors. It is not easy to say that Dollar and Zhu’s methods are real-time ones because they should deal with all sub-windows in one image.

In this paper, we focus on a fast method for detecting pedestrians in videos from a fixed camera environment in surveillance system for tracking by detection. In the case of a fixed camera environment, Gaussian mixture model

Sixth International Conference on Machine Vision (ICMV 2013), edited by Antanas Verikas,
Branislav Vuksanovic, Jianhong Zhou, Proc. of SPIE Vol. 9067, 906717 · © 2013
SPIE · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2051382

(GMM) can be used for background subtraction [10]. We propose an efficient coarse-to-fine pedestrian detection method based on motion detection and HOG cascade to make a faster detector. The proposed method consists of two steps: coarse detection and fine detection. In the first step, the motion detection in a frame is used for the coarse detection. From the coarse detection, we can obtain some blobs and detect pedestrian candidate regions to reduce the search region. In the second step, the HOG cascade combined with a linear SVM is used for the fine detection. It is a fast detection method and is similar in performance to HOG. By combining two steps, we are able to reduce computation time while keeping almost the same performance as HOG.

This paper organized as follows: Our proposed pedestrian detection method in a fixed camera environment is described in Section 2. The experimental results, in Section 3, show that the proposed method is able to ensure fast object detection as well as to achieve nearly the same accuracy as HOG cascade. Finally, conclusion is summarized in Section 4.

2. COARSE-TO-FINE DETECTION

Although a pedestrian detection based on Haar-wavelets with Adaboost [11] is fast and can be real-time processing, it has many false positive and miss detection windows. The detection performance of HOG is better than the method based on Haar-wavelets with Adaboost. However, we cannot apply it to a real-time system because it has heavy computations. The detection rate of HOG cascade is similar to HOG and it is faster than HOG, but this method is still too slow to be used for a real-time system.

As shown in Fig. 1, our proposed method is divided into two steps: coarse detection and fine detection. In surveillance system under the fixed camera environment, background subtraction using background model is very useful to detect ROI (region of interest) for pedestrian detection. This method is the coarse detection and can reduce time computation since there are a few search areas. Next step is the fine detection. We are able to apply HOG cascade to each ROI for pedestrian detection.

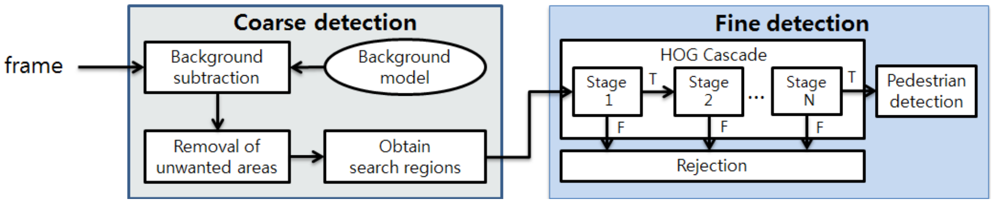

Figure 1. Coarse-to-fine detection framework

2.1 Coarse detection

In the computer vision, background subtraction is commonly used for detecting moving objects in a video from a fixed camera by taking the difference between the background model and the current frame. To get candidate regions including pedestrians, this step operates the background subtraction. Firstly, GMM [10] is used to make a background model. In this method, the background is statistically modeled on each pixel using more than two Gaussian distributions. In order to adapt to pixel value changes, we can use the GMM which updates the training set by adding new samples and discarding the old ones. Besides, we can estimate parameters such as weight, mean and covariance by EM algorithm [12].

After background modeling, the difference between the background model and the current frame is calculated. Morphological processing and connected component labeling are used for removal of unwanted areas. Morphological operations such as opening and closing are necessary, since there are many noises right after background subtraction. After that, foreground regions of each frame are grouped into components by using a connected component labeling algorithm. A size filter is used to remove small components. Each of remaining components is bounded by 2D bounding box, so we are able to obtain search regions. Fig. 2 shows an example of background subtraction and ROIs detected from the coarse detection step.

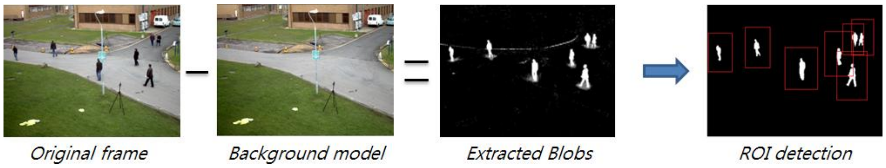

Figure 2. ROIs detected from the coarse detection step

2.2 Fine detection

Although we can reduce computation time through the coarse detection, we still need strategies to decrease more to complete given jobs within the allowable time for real-time system. In addition, we need a pedestrian detector since blobs obtained from the coarse detection are motion cues, not pedestrians. For example, if dogs, cars or other moving objects appear in the frame, their motions are detected. Hence, we need a detector using pedestrian model to distinguish pedestrians from them.

HOG is a great and popular algorithm to represent the human. Each detection window in HOG is divided into cells of size 8x8 pixels and each group of 2x2 cells is integrated into a block. Dalal [4] just used the block size fixed at 16x16 pixels and Zhu [8] used much larger set of blocks that vary in sizes. In the fine detection step, we use Zhu's method. When obtaining HOG feature vectors of blocks, integral histogram [13] is used for fast evaluation of features and Adaboost is used to choose the best blocks suited for detection and construct the rejector-based cascade which is a great method to reduce time computation (see [8] for details). As shown in Fig. 3, we apply HOG cascade as the fine detection to each ROI detected from the coarse detection step. Finally, pedestrians are detected.

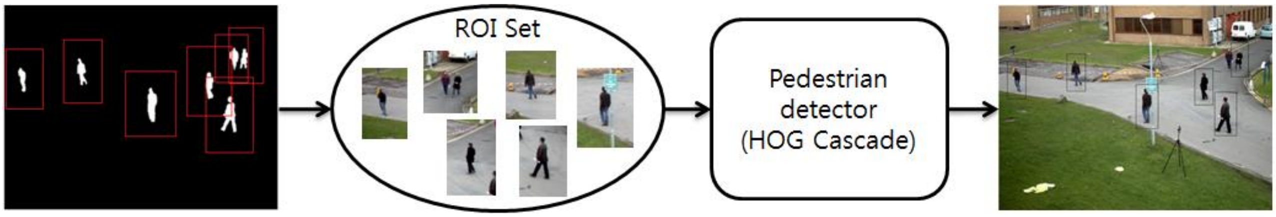

Figure 3. Visualization of coarse-to-fine detection

3. EXPERIMENT RESULTS AND ANALYSIS

We evaluate our method on the first view of the S2.L1 sequence from the PETS2009 benchmark [14]. The video of 795 frames at 7 frames per second (fps), recorded from a distant viewpoint, has become a defacto standard pedestrian benchmarking dataset and the size of each image is 768x576.

3.1 Evaluation metrics

A standard metric for evaluating multiple object detection is the Multiple Object Detection Accuracy (MODA) [15], defined as

$$N-MODA=1-\frac{\sum_{t=1}^{N_{frames}}(c_m(m_t)+c_f(fp_t))}{\sum_{t=1}^{N_{frames}}N_G^{(t)}} \tag{1}$$

Where m_t is the missed detection count, fp_t is the false positive count, c_m is the cost function for missed detects, and c_f is the cost function for false positive. According to [15], the cost functions are set to $c_m=c_f=1$.

3.2 Implementation details

HOG cascade is trained using INRIA dataset [16] after resizing the training images to 48x96 and has 15 stages and 532 weak classifiers (or blocks). To apply this cascade to each ROI, we need to set some parameters. As shown in Fig. 4, R_{width} and R_{height} are ROI ratio that changes blob size after background subtraction to ROI size. S_{ROI} is the scale factor that specifies how much ROI is reduced at each ROI scale and $n_{\min}$ is min neighbors, a parameter which specifies how many neighbors each candidate rectangle should retain.

We empirically set the proper parameters used for the coarse-to-fine detection as shown in Table 1. First, we set ROI ratio parameters while changing R_{width} and R_{height}. Second, we fixed both ROI ratio parameters and $n_{\min}$ to find S_{ROI}. Finally, we changed only $n_{\min}$ and found proper parameters for the fine detection. When we set all parameters, MODA and fps on average are 0.6356 and 11.46, respectively.

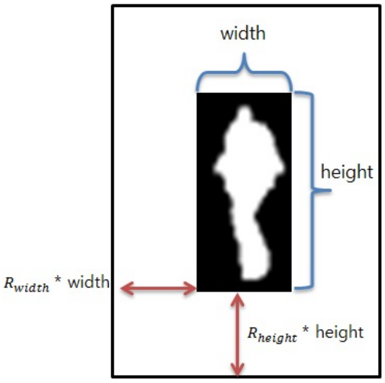

Figure 4. ROI parameters

Table 1. Proper parameters for the Coarse-to-fine detection

R_{width}	R_{height}	S_{ROI}	$n_{\min}$
1	**0.5**	**1.1**	**1**

3.3 Comparison and analysis

We compare our proposed method with previous methods: HOG, Haar-wavelets cascade, HOG cascade and some works in winter-PETS 2009 [17]. In the evaluation of average fps (average processing time), we exclude evaluation of works in winter-PETS 2009 since most of the works are for off-line object tracking and do not consider computation time.

As shown in Fig. 6, average fps of Haar-wavelet cascade method is 15.42 and it is the fastest one in our evaluation. However, MODA of this method is the worst. MODA of our proposed method is similar to HOG and Winter-PETS while our proposed method is approximately 7.5 times faster than HOG.

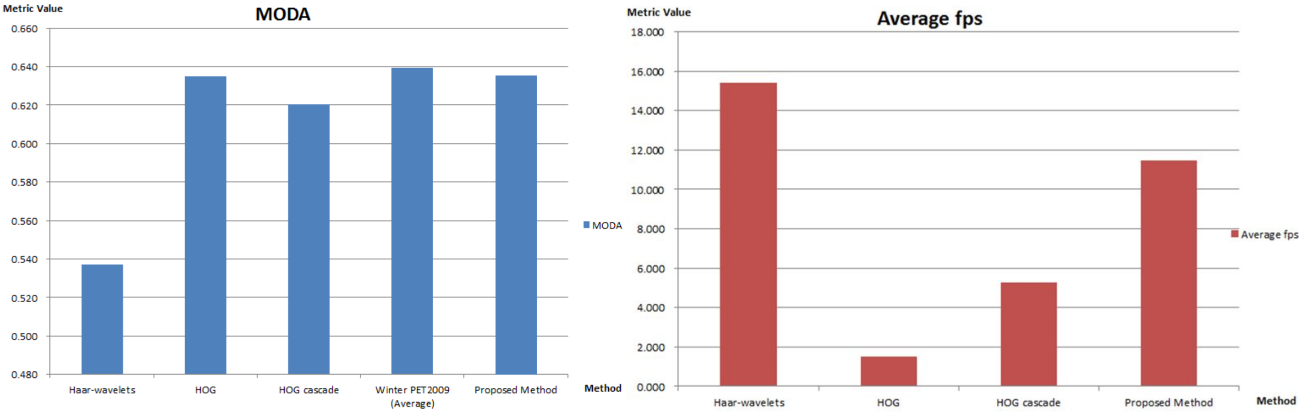

Figure 5. Comparision with previous works and our proposed method

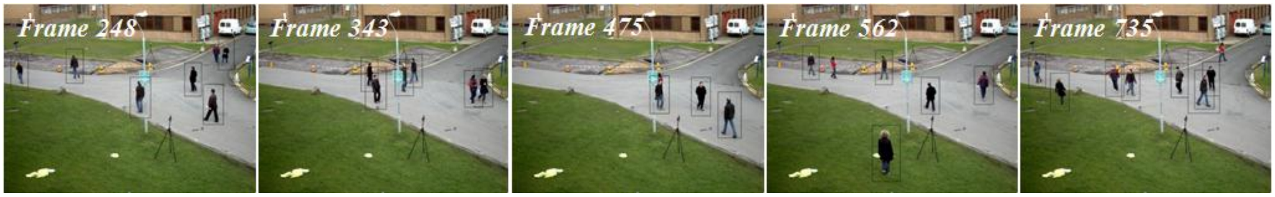

Figure 6. Pedestrian detection results in our method

4. CONCLUSIONS

We propose a real-time coarse-to-fine pedestrian detection system which is as many as 7.5 times faster than the previous methods while keeping almost the same detection performance as HOG. (Evaluation of average fps on PETS2009 dataset recorded at 7 is 11.46) This is achieved by combining two steps. First, we build adaptive background model using GMM based on pixel and subtract the background from a given frame. This step is the coarse detection to detect ROIs by filtering out some windows not including moving objects. The second step called the fine detection eventually detects pedestrians by applying HOG features-based the cascade to each ROI obtained from the coarse detection.

However, although our proposed method is allowable for a real-time system, there are still some false positive and miss-detection in our demonstration when some pedestrians are small or occluded each other. In our future work, we will study new features that represent human and develop pedestrian detector with the state-of-the-art time performance.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A2013778).

REFERENCES

[1] W. Hu, T. Tan, L. Wang and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” IEEE Transactions on Systems, Man, and Cybernetics, PART C: Applications and Reviews, vol.34, No.3, August 2004.

[2] C. Papageorgiou and T. Poggio. “A trainable system for object detection,” IJCV, 38(1):15.33, 2000.

[3] D. M. Gavrila, J. Giebel, and S. Munder. “Vision-based pedestrian detection: The PROTECTOR System,” Proc. of the IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004.

[4] N. Dalal and B. Triggs. “Histograms of oriented gradients for human detection,” In Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, California, USA, pages 886–893, 2005.

[5] P. Viola and M. Jones. “Robust Real-Time Face Detection,” IJCV 57(2), 2004.

[6] G. Xu, X. Wu, L. Liu and Z. Wu. “Real-time Pedestrian Detection Based on Edge Factor and Histogram of Oriented Gradient,” ICIA, 2011.

[7] P. Dollar, S. Belongie and P. Perona. “The fastest pedestrian detector in the west,” BMVC, 2010.

[8] Q. Zhu, S. Avidan, M. Yeh and K. Cheng. “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients,” In Proceedings of the Conference on Computer Vision and Pattern Recognition, 2006.

[9] V. Prisacariu and I. Reid. “Fast HOG – a real-time GPU implementation of HOG,” Technical Report 2310/09, University of Oxford, 2009.

[10] Z. Zivkovic. “Improved Adaptive Gaussian Mixture Model for Background Subtraction,” 17th International Conference on Pattern Recognition, Vol. 2, pp.28-31, 2004.

[11] P. Viola, M. Jones, and D. Snow. “Detecting pedestrians using patterns of motion and appearance,” International Conference on Computer Vision (ICCV), 2003.

[12] TODD K. MOON. “The Expectation Maximization Algorithm. IEEE Signal Processing Magazine, November 1996.

[13] F. Porikli. “Integral histogram: A fast way to extract histograms in cartesian spaces,” Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[14] PETS 2009 : Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2009, Available: http://pets2009.net/

[15] R. Kasturi, P. Soundararajan, J. Garofolo, R. Bowers and V. Korzhova. “Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol,” PAMI, 2009.

[16] Available : http://lear.inrialpes.fr/data. , INRIA dataset.

[17] A. Ellis, A. Shahrokni and J.M. Ferryman. “PETS2009 and Winter-PETS 2009 Results: A Combined Evaluation,” Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2009.