A FAST AND ROBUST PEDESTRIAN DETECTION FRAMEWORK BASED ON STATIC AND DYNAMIC INFORMATION

Tao Xu, Hong Liu, Yueliang Qian and Zhe Wang

Key Laboratory of Intelligent Information Processing &&
Beijing Key Laboratory of Mobile Computing and Pervasive Device
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190
{xutao, hliu, ylqian, wangzhe01}@ict.ac.cn

ABSTRACT

With the powerful development of pedestrian detection technique based on sliding-window and machine-learning, detection-based tracking systems have become increasingly popular. Most of these systems rely on existing static pedestrian detectors only despite the obvious potential motion information for people detection. This paper proposes a novel pedestrian detection framework fusing static and dynamic features. Motion cue is firstly used to detect potential pedestrian regions. Secondly, static detector scans potential regions to get candidate pedestrian detections. Final detection results are improved by removing false detections based on their motion distribution. The proposed framework significantly raises detection speed and detection performance. Static detector of pedestrian in this paper is trained by AdaBoost with simplified HOG feature (1HOG). Additionally, we introduce a detection-window-pyramid based scanning strategy for quickly extracting 1HOG features. The experimental results on several public data sets show the effectiveness of the proposed approach.

Index Terms—detection-based tracking, pedestrian detection, dynamic information, sliding-window strategy

1. INTRODUCTION

Due to the development of sliding-window based and machine-learning based pedestrian detection technique, detection-based tracking methods have gained increasing attention since they are more robust in complex scene [4]. Pedestrian detection in every or interval frames of the video is the basis of detection-based tracking methods. Results of detection are used for data association in following tracking process.

The detection performance and speed are bottlenecks of detection-based tracking approaches in which most researches mainly focus on how to associate detection responses of multi-objects. They prefer to adopt existing detectors [4,5] for pedestrian detection, such as HOG detector [1] or part-based ISM detector [18]. These detectors may result in a great many false alarms in complex scenes. Though many false alarms can be removed based on the following object matching between frames, an excess of false alarms may cause a huge search space for data association in return. Moreover, pedestrian detection is time-consuming using HOG [1] or ISM detector [18], which makes the whole speed of detection-based system very slow. For instance, Brendel [5] spends more than 40s to detect and track one frame. Without considering detection time, Cheng [19] and Breitenstein [4] only give tracking time. To track pedestrian in real-time, quick and robust pedestrian detection is important and necessary. While few methods focus on fast and robust pedestrian detection in detection-based tracking systems, some researchers use hardware such as GPU [13] or multi-thread [20] to accelerate. But they only use static features of pedestrian [1, 2, 3]. Pedestrian's dynamic features have already been proved to be very important for human visual understanding. On the other side, sliding-window based scanning method scans the whole frame without considering the potential pedestrian regions.

We focus on studying fast and robust pedestrian detection method in video for detection-based tracking systems. A pedestrian detection framework which fuses static and dynamic features is proposed. The contributions of this paper are as follows:

(1) We propose a pedestrian detection framework based on static and dynamic features. We firstly use Motion Cue (MC) to extract potential pedestrian regions. Secondly, only potential regions are scanned by Static Detector (SD). Final detection results are improved based on their Motion Distribution (MD). In this paper, our framework is called MSM (MC+SD+MD) for short.

(2) We introduce a detection-window-pyramid based scanning (WPS) strategy for quickly extracting features.

We test proposed method on public data sets: INRIA [1], PETS [9] and TUD [12]. The experimental results show that MSM framework provides fast and robust pedestrian detection result. The WPS strategy also contributes to increasing detection speed and detection performance.

2. RELATED WORK

Most of detection-based tracking methods use existing static detectors for pedestrian detection, such as HOG detector [1], edgelets detector [7] and part-based ISM detector [18]. Dalal [1] proposed HOG feature to characterize gradient orientation distribution in a rectangular block. This feature

978-0-7695-4711-4/12 $26.00 © 2012 IEEE
DOI 10.1109/ICME.2012.66

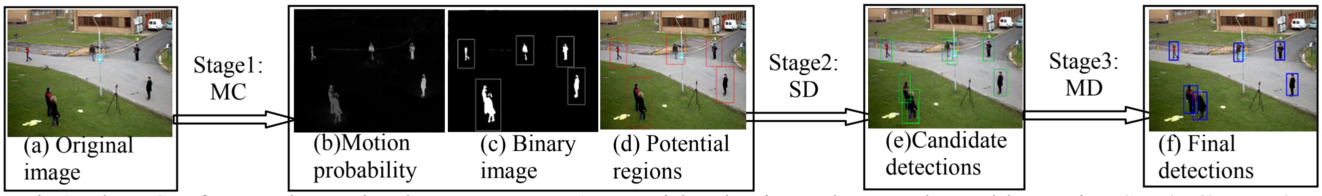


Fig.1. The MSM framework contains three stages: stage1, potential pedestrian regions are detected by Motion Cue (MC); stage2, candidate pedestrian detections are detected by Static Detector (SD); stage3: detection results are refined by Motion Distribution (MD). (rectangles in (c)(d) are potential pedestrian regions; rectangles in (e) are candidate detections; rectangles in (f) are final detections).

was more than three thousands dimensions and a linear SVM was used for training the detector by which the pedestrian detection was very slow. For increasing the detection speed, Zhu [2] simplified the HOG feature [1] to a 36-dimensional vector and trained a cascaded detector by AdaBoost, but the weaker classifiers were also trained by SVM. Hou [3] proposed the EHOG feature and directly used it to train a tree-structure detector by Vector Boosting.

The methods above focused on simplifying static features and used cascaded methods to increase the detection speed. Dynamic features of the pedestrian in video are also important information. Some existing methods [8, 10] combined motion and appearance to detect or track pedestrian by using dynamic detectors and static detectors which required similar movement of pedestrian in training and testing data. Zhang [11] used HOG detector on the whole images and then used motion information to remove false alarms. These methods [8, 10, 11] improved detection accuracy but they did not reduce detection time.

In this paper, we propose a pedestrian detection framework MSM and a novel detection-window-pyramid based scanning strategy WPS, which is described in section 3. The experimental results and conclusions are introduced in section 4 and in section 5 respectively.

3. OUR MAIN WORK

3.1 Our system framework

Fig.1 shows our MSM framework. We firstly use Motion Cue (MC) of pedestrian in video to extract the motion probability map which locates potential pedestrian regions in the original image. Secondly, only these potential regions are scanned by Static Detector (SD). Final detection results are improved based on their Motion Distribution (MD). The feature used in this paper is 1-demensional HOG (1HOG) which is simplified from popular HOG feature [1]. We use 1HOG features to train weaker classifiers and construct a cascaded static detector by AdaBoost. Our WPS strategy is used to scan potential regions.

3.2 Potential regions extraction based on Motion Cue

For tracking-by-detection approach, dynamic of pedestrian is an important feature in video. We extract dynamic regions as potential pedestrian regions to reduce detection time, which also weaken the influence of cluttered background. Many ways have been proposed to gain dynamic information [14, 15]. In our MSM framework, potential regions should be extracted quickly and most moving information should be remained in initial stage, so we adopt Gaussian mixture background model [16]

$$b(x,y)=\sum_i \alpha_i g(f(x,y),\theta_{i,x,y},\delta_{i,x,y}) \quad (1)$$

Where *(x, y)* is the location of each pixel, $\theta_{i,x,y}$, $\delta_{i,x,y}$ are the model parameters of each individual Gaussian components *g*, and *f(x,y)* is the local pixel intensity. In our method, we let $i=1$ and $\alpha_i=1$ for simplification, then the motion probability map *v(x,y)* of one frame *f(x,y)* is estimated by equation(2). δ_s is the model parameters of the simplified Gaussian components.

$$v(x,y)=1-e^{\left(-\frac{1}{2}\frac{(f(x,y)-b(x,y))^2}{\delta_s^2}\right)} \quad (2)$$

The binary motion map can be gained by equation (3):

$$I_b(x,y)=\begin{cases}1, & if\ v(x,y)>v_{th1}\\ 0, & else\end{cases} \quad (3)$$

Where $I_b(x,y)$ means the value of pixel *(x, y)* in the binary motion map; v_{th1} is the threshold that decides whether the pixel *(x, y)* is a motion one or not.

Then several potential pedestrian regions can be gained by analyzing the binary motion map. Here we expand the potential regions for getting more around pixels. Finally, based on the potential pedestrian regions of binary motion map, we get the potential regions in the original input image.

3.3 Candidate detection by Static Detector

In the second stage of MSM framework, 1HOG static detector scans potential regions to get candidate detections. Our WPS strategy is used as the scanning method.

3.3.1 The cascaded 1HOG detector (SD)

Similar to method used in [3, 21], we simplify HOG feature to 1-demensional HOG (1HOG) feature. Then we use 1HOG features to train weaker classifiers and construct a cascaded static detector by AdaBoost.

Feature extracting: Firstly we divide each detection window into variable-sized blocks [2]. Then the orientation over 0◦~180◦ is divided into 9 bins. Finally we calculate histogram of oriented gradients of 9 bins of a block. The histogram of oriented gradients of each bin of a block is called 1HOG feature in our paper, and it can be directly learned by weaker classifiers without extra reducing dimensions and SVM training [1, 2]. We neither divide the block into cells nor construct multi-dimentisional features using 1HOGs for following reasons: some small blocks of the variable-sized blocks can replace cells and useful 1HOG features will be learned automatically by AdaBoost. Although only one oriented gradient of a block is used as the feature unit for boosting learning, many oriented gradients of the block will be selected if they are distinguished features. Also 1HOG features can be extracted quickly by integral image without tri-linear interpolation.

Detector training: We train the 1HOG Static Detector (SD) on public data set INRIA [1]. The training set includes 2476 positive examples and 12180 negative examples. The size of 1HOG detector used in our experiment is 32×64 pixels. We directly use 1HOG features to train weaker classifiers and construct a cascaded static detector by AdaBoost. Based on variable-sized blocks [2], the 32×64 pixels detection window creates 22248 1HOG features. Instead of random selected features [2], we use all features for training weaker classifiers. It takes several days to train a 1HOG detector with 15 stages on the PC with 2.93 GHz CPU and 2GB memory. Each stage satisfies minimum detection rate of 0.999 and maximum false positive of 0.5.

Fig.2. shows that 1HOG detector achieves similar detection performance with Dalal's HOG detector [1]. Moreover, the detection time of 1HOG detector trained by AdaBoost is 5 times faster than Dalal's detector. Fig.3 illustrates some right detection results of 1HOG detector.

3.3.2 The Detection Window Pyramid

For scanning all positions and scales, the image pyramid (IPS) is used in most existing sliding-based detection methods [1,2,6,10]. The integral image [6] is usually adopted for quickly computing HOG [2,3] and haar-like [10] features. However, constructing integral image of HOG features is more time-consuming than Haar features. The integral image of HOG features should be recomputed for different scales of images by IPS, and it wastes a lot of time and memory. We propose a novel detection-window-pyramid scanning (WPS) method based on mapping scheme. In our method the integral image is computed only once.

Like the method of constructing image pyramid [17], the number of scales of detection window pyramid (n) is computed by equation (4):

$$n = \left(\frac{\log(S_e / S_0)}{\log(S_r)} + 1\right) \tag{4}$$

Where S_0 is the start scale of scanning window; S_r is the scale step; S_e is the end scale computed by $S_e = min\ (w_d/w_0, h_d/h_0)$, where w_d and h_d are width and height of the detection region. This region is the whole input image when MC is not used and it is the potential region when MC is used. The w_0 and h_0 are width and height of the detector.

For each scale of the pyramid $S_i = S_0 \times (S_r)^{(i-1)}$, $i = 1, 2, \ldots, n$. The size of the *i-th* scanning or detection window $W_i(w_i, h_i)$ is rescaled by $w_i = w_0 \times S_i$, $h_i = h_0 \times S_i$ which may not match our 1HOG detector size such as 32×64 in our experiment. In this case, we adopt location mapping method. The locations of 1HOG features learned by the detector will be mapped to the new locations in detection window. As Fig.4 shows, we use $B_0(x_{b0}, y_{b0}, w_{b0}, h_{b0})$ to represent the location of one of 1HOG features and use $B_i(x_{bi}, y_{bi}, w_{bi}, h_{bi})$ to represent the rescaled locations in the *i-th* detection window after mapping. In our method, B_0 is mapped to B_i by $x_{bi} = x_{b0} \times S_i$, $y_{bi} = y_{b0} \times S_i$, $w_{bi} = w_{b0} \times S_i$, $h_{bi} = h_{b0} \times S_i$. Fig.4 shows the mapping from B_0 to B_i in our detection window pyramid. 1HOG features of the detection window are extracted in the rescaled locations.

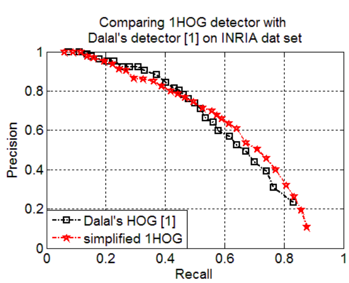


Fig.2. PR curves of 1HOG detector and Dalal's detector [1] on INRIA data set under the same condition.

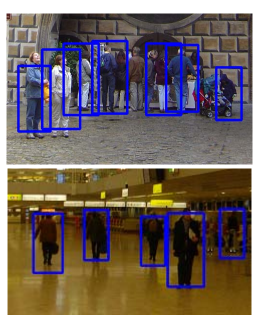


Fig.3. Some right results of 1HOG detector on INRIA data set.

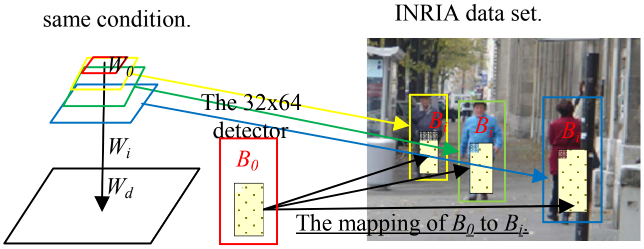


Fig.4. The detection window pyramid (WPS). (the left is the rescaled detection windows of different levels of the pyramid; the right is the original image that is scanned by the rescaled detection window.)

3.4 Improve detection results by Motion Distribution

Although candidate detections are obtained on potential regions, static detector always results many false alarms to gain a high recall rate. In this paper, false alarms are removed by analyzing Motion Distributions (MD) of outside rectangles of candidate detections. Motion probability map $v(x, y)$ is computed in the section 3.2. The MD of certain candidate detection is gained by dividing its outside rectangle region into 9 cells equally, as Fig.5 shows. The moving weight of the *i-th* candidate detection (H_i) is estimated by the equation (5):

$$H_i = \sum h_i(j), \quad j = \{0,1,2,\ldots,8\} \tag{5}$$

Where $h_i(j)$ represents the moving weight of the *j-th* cell in the *i-th* detection, and it is computed by equation(6):

$$h_i(j) = \frac{\sum_{(x,y)\in\Omega_{i,j}} w(x,y)}{S_i(j)} \tag{6}$$

Where $\Omega_{i,j}$ represents the region of the *j-th* cell in the *i-th* detection; $S_{i,}(j)$ represents the area of the *j-th* cell in the *i-th* detection; $w(x, y)$ means the motion weight of the pixel at position (x, y), and it is computed by equation (7):

$$w(x,y)=\begin{cases} 1, & if\ v(x,y)>v_{th2} \\ 0, & else \end{cases} \qquad (7)$$

Where $v(x, y)$ is the motion probability map calculated by equation (2); v_{th2} is the threshold that decides whether a pixel is moving. We let $v_{th2} > v_{th1}$ in our experiment.

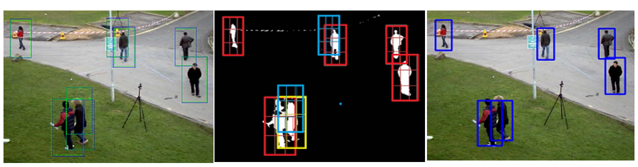
Motion information is more likely to distribute in the vertical center of a stand-up pedestrian as Fig.5 (b) shows. Our MD method is based on following two hypotheses:

(1). The H of the right detection is higher than the false detection;

(2). the vertical center cells of detection have high motion weight ranking among that of all cells.

We use the following strategy to remove false alarms from candidate detections:

> loop all the candidate detections
> {
> (a)calculate the H_i for the i-th detection
> if (H_i<H_{th}), remove the detection;
> (b)else calculate h_i for the i-th detection and get the ranking of the vertical center cell (R_1, R_4,R_7),
> if(R_4<R_{th}&&(R_1<R_{th}||R_7<R_{th})), remove the detection.
> }
> The remaining candidate detections are the final detections.



(a) Candidate detections MC+SD (1HOG) (b) Motion weight map (c) Final detections MC+SD+MD (1HOG)

Fig.5. Final improved detection results based on MD

4. EXPERIMENTAL RESULTS

4.1. Introduction of test data

The 1HOG static detector is trained and tested on INRIA as subsection 3.3.1 describes. We test proposed MSM framework and WPS strategy on PETS [9] and TUD [12] data sets without retraining 1HOG detector. The original resolution of test image is used.

The first test set is View_001 of S2_L1 sequence from PETS [9]. This video contains 795 frames with the size of 768×576 pixels. We select 160 frames one from 5 frames to construct test set, which contains 991 pedestrians. The size of pedestrian in PETS is much smaller than INRIA. This data set is captured by a static camera and all the pedestrians keep moving in the whole sequence. The background which is constructed by Gauss model with the first 200 frames is perfect for extracting a good motion probability map.

The second test set comes from the more challenging TUD crossing sequence [12] which contains 202 frames with the size of 640×480 pixels. This sequence is pretty short, so we use all the 202 frames as the test data. The size of pedestrian in TUD is bigger than PETS, and most of them only show their profiles with serious frequent occlusion. For the reasons that TUD provides a very shot sequence and most of the frames contain slowly moving cars, the background construction and motion probability map are not very well, as Fig.9 (a, c) shows.

4.2 Performance evaluation metrics

We adopt the Precision-Recall (PR) curve, Average Precision (AP) under the maximum recall rate (MR) [17] and runtime per frame (T) to evaluate the performance of our method. Dalal [17] assumes that the more overlapping detections there are in the neighborhoods of an image region the higher the probability for this region to be a true positive. We obtain points of PR curves by changing the threshold of the number of overlapping detections (density) in the cluster after non-maximum suppression. As method in [17], we measure the interpolated precision at a set of 30 equally spaced recall levels over the whole range [0, MR]. The i-th spaced recall level is defined as $p_i(r)$ which is the maximum precision for any recall rate in the interval [r_{is}, r_{ie}],

$$p_i(r)=\max_{r\in[r_{is},r_{ie}]} p(r) \ , i=1,2,...,30 \qquad (8)$$

AP is computed by:

$$AP=\frac{1}{n}\sum_{1\le i\le n} p_i(r)\,, \qquad n=30 \qquad (9)$$

4.3 Experimental results and analysis

The MSM framework provides three testing schemes as using the static detector to scan the whole original image (SD), using the static detector to scan the potential pedestrian regions (MC+SD), and using the static detector to scan the potential pedestrian regions with post-process of removing false detections based on their motion distribution (MSM or MC+SD+MD). In this paper, we make two sets of experiments on PETS and TUD data sets. Firstly, we test 1HOG detector on the whole image (SD+1HOG) by IPS and WPS methods to validate the effectiveness of WPS method. Secondly, we test the performance of our MSM framework by comparing its different combination of schemes.

The scanning methods of IPS and WPS have same scanning-density parameters: W_0, W_r , S_0 and S_r (W_0 means the size of the detector; W_r means the step of detection window scanning on the detection region; S_0 and S_r are defined in subsection 3.3.2). The size of Dalal's detector provided by OpenCV is 64×128 pixels; the size of 1HOG detector is 32×64 pixels; the other parameters are the same in all the experiment as W_r = 4×4; S_0 = 1 and S_r =1.02.

4.3.1. WPS methods with 1HOG detector on the whole test image (SD+1HOG)

To validate the effectiveness of WPS method, we test 1HOG detector by IPS and WPS methods on the original image. As big rectangles responses in Fig.8 shows, the number of false

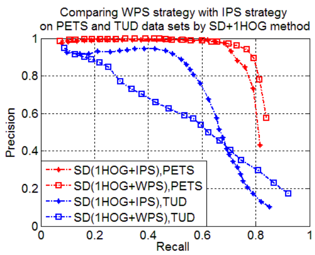


Fig.6. Results of WPS and IPS strategies on PETS and TUD by SD (1HOG)

alarms caused by background by IPS method is more than by WPS method. This is possible the reason why IPS gains a lower precision rate than WPS at the recall rates over 0.7 on both test sets, as Fig.6 shows. However, on TUD when the size of pedestrian is large in the image, as Fig.8 (c, d) shows, parts of pedestrian are more likely to be considered as responses by WPS for the small size of 1HOG detector. Thus the precision rates of WPS are lower than IPS during a range of recall rates on TUD. Detection results are probably improved by using the 1HOG detector with bigger size [3].

As Table1 illustrates, comparing with IPS method, the detection time is decreased from 12.82s to 3.61s on PETS and from 11.3s to 2.98s on TUD by using WPS method with 1HOG detector on original image. The result shows that our WPS method can reduce detection time again with similar detection performance compared with IPS.

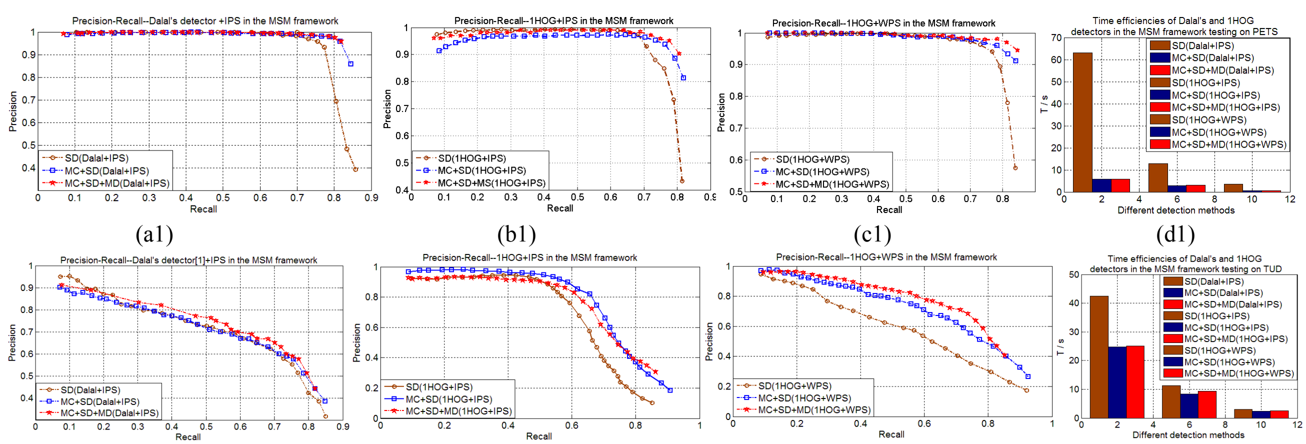
4.3.2. MSM framework

To evaluate the performance of our MSM framework with different detectors and scanning methods, we test Dalal's detector and 1HOG detector with IPS and WPS strategy under the MC+SD and MSM scheme respectively.

MC+SD: Fig.10(a, b) shows some detection results only by SD scheme on PETS, which remains a great many false alarms to get high recall rates. MC process gets potential pedestrian regions firstly. Fig.10(c) shows that most false alarms on background are removed by adding MC process. On PETS, at the recall rate of 0.8, the precision rate of Dalal's detector is increased from 0.733 to 0.97. 1HOG detector with IPS and WPS are increased from 0.608 to 0.876 and from 0.88 to 0.948 respectively. Meanwhile, the detection speed is increased, especially in scenes with a few of moving regions. As Table1 shows, on PETS detection time is reduced from 63.12s to 5.8s, from 12.82s to 2.87s and from 3.61s to 0.575s when different detectors and scanning methods are used. Comparing with 1HOG, the speed is increased more when we use Dalal's detector. It is possibly because that for detecting the same potential region the cascaded 1HOG detector should cost more time on the complex potential regions than that on background, while Dalal's SVM-based detector cost same time on potential regions and on background. On TUD, detection time is decreased less than 2 times by Dalal's detector and even less by 1HOG detector, as Table1 shows. Detection time is related with the size of potential regions. As Fig.9 (a, c) shows, background construction of TUD are not very well and pedestrians are large in this scene, which make the potential pedestrian region taking a big ratio at the original


(a) IPS on PETS (b) WPS on PETS (c) IPS on TUD (d) WPS on TUD

Fig.8. Responses before no-maximum suppression by 1HOG detector with IPS and WPS methods on the whole image



(a1) (b1) (c1) (d1)
(a2) (b2) (c2) (d2)

Fig.7. Results of MSM framework with different detectors and scanning methods on PETS (the first row) and TUD (the second row)

Table 1. The AP under MR valued 0.8 and T of all the experiments on PETS and TUD

Data set	Method	Dalal+IPS				1HOG+IPS			1HOG+WPS		
		BS[11]	SD	MC+SD	MSM	SD	MC+ SD	MSM	SD	MC+ SD	MSM
PETS	AP	0.983	0.978	0.992	0.995	0.967	0.958	0.976	0.984	0.992	0.995
	T(s/fr)	63.28	**63.12**	5.80	5.85	**12.82**	2.87	3.11	**3.61**	0.575	**0.586**
TUD	AP	0.735	0.749	0.751	0.759	0.753	0.819	0.845	0.687	0.801	0.824
	T(s/fr)	42.56	**42.46**	24.78	25.11	**11.3**	8.39	9.38	**2.98**	2.28	2.44

image. This example shows the results of MC influence detection performance and detection speed in MSM framework. For complex scenes as TUD dataset, more effective method should be used to extract motion cues, which will improve the whole detection performance.

MSM: MD strategy is used to remove some false alarms to improve detection results. Fig.10 (d) shows the final detection results after motion distribution analysis. As Fig.7 shows, comparing with MC+SD scheme, both Dalal's and 1HOG detectors are improved by adding MD. For instance, at the recall rate of 0.8, the precision rate of 1HOG detector with IPS method increased from 0.876 to 0.913 on PETS. Additionally, we compare our MSM framework with the Bayesian framework (BS) [11], as table1 shows. The results validate that our MSM framework is more robust and faster than BS under the same condition.

Table1 illustrates AP under MR valued 0.8 to summarize the overall performance of our method. The results prove that our MSM framework is robust and fast for combining different static detectors, such as Dalal's detector or 1HOG detector, with dynamic information. WPS method and 1HOG feature contribute a faster detection speed. When 1HOG detector and WPS strategy are used in the MSM framework on PETS, we achieve the faster speed of 0.586s/fr with high precision rate. If we use larger scanning windows scale, a faster detection speed will be achieved.

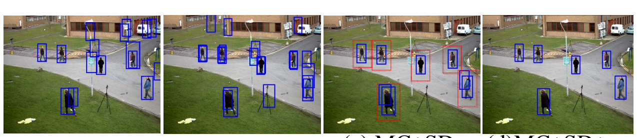
5. CONCLUSIONS

In this paper, we propose a novel pedestrian detection framework MSM fusing static detector and dynamic information. Experimental results on some public data sets show that not only the detection speed but also the detection performance is increased by proposed MSM framework especially when 1HOG detector and WPS strategy are used in our MSM framework.

In future work, we will consider more dynamic features, such as optic flow or dynamic texture features for complex scenes. Also we are planning to work on the problem of partial occlusion, which is a main drawback of global object detectors.

(a) Constructed background (b) Original image (c) Motion map (e) Final results

Fig.9. Middle results of the MSM framework on TUD



(a) Dalal's SD (b) 1HOG SD (c) MC+SD (1HOG) (d)MC+SD+ MD (1HOG)

Fig.10. Detection results of different methods on PETS (blue rectangles mean the detection results; red rectangles in (c) means the potential regions.)

6. ACKNOWLEDGMENT

The research work is supported by the National Nature Science Foundation of China No. 60802067.

7. REFERENCES

[1] Dalal N. And Triggs B.,"Histograms of Oriented Gradients for Human detection," in Proc. of CVPR, vol. 1, pp. 886-893, 2005.

[2] Zhu Q., Avidan S. et al., "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients," in Proc. of CVPR, vol. 2, pp. 1491-1498, 2006.

[3] Hou C., Ai H. Z. et al., "Multiview Pedestrian Detection Based on Vector Boosting," in Proc. of ACCV, vol. 1, pp. 210-219, 2007.

[4] Breitenstein M. D., Reichlin F. et al., "Robust tracking-by-detection using a detector confidence particle filter," in Proc. of ICCV, pp. 1515-1522, Sept. 2009.

[5] Brendel W., Amer M. and Todorovic S., "Multiobject Tracking as Maximum Weight Independent Set," in Proc. of CVPR, pp. 1273-1280, June 2011.

[6] Viola P. and Jones M., "Rapid Object Detection using a Boosted Cascade of Simple Features," in Proc. of CVPR, vol.1, pp. 511-518, 2001.

[7] Wu B. and Nevatia R., "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors," in Proc. of ICCV, vol. 1, pp. 90-97, 2005.

[8] Dalal N., Triggs B. and Schmid C., "Human Detection Using Oriented Histograms of Flow and Appearance," in Proc. of ECCV, pp. 7-13, 2006.

[9]. Available at http://pets2010.net/.

[10]. Viola P., Jones M. and Snow D., "Detecting Pedestrians Using Pattern of Motion and Appearance," in Proc. of ICCV, vol. 2, pp. 734-741, 2003.

[11] Zhang J. and Gong S. "Beyond Static Detectors: A Bayesian Approach to Fusing Long-term Motion with Appearance for Robust People Detection in Highly Cluttered Scenes," IEEE Visual Surveillance Workshop, 2006.

[12] Andriluka M., Roth S. and Schiele B., "People-Tracking-by-Detection and People-Detection-by-Tracking," in Proc. of CVPR, pp. 1-8, 2008.

[13] Prisacariu V. and Reid I., "Fast HOG - a real-time GPU implementation of HOG," in the Technical Report 2310/09, University of Oxford, 2009.

[14] Proesmans M., Van Gool L. and Oosterlinck A., "Determination of optical flow and its discontinuities using non-linear diffusion," in Proc. of ECCV, vol. 2, pp. 295-304, 1994.

[15] Gautama T. and Van Hulle M., "A phase-based approach to the estimation of the optical flow field using spatial filtering," IEEE Trans. on Neural Networks, vol. 13, pp. 1127-1136, 2002.

[16] Xiang T. and Gong S., "Beyond tracking: Modeling activity and understanding behavior," IJCV, pp. 195-204, 2006.

[17] Dalal N., "Finding People in Images and Videos," PhD dissertation, Inst. National Poly technique de Grenoble, 2006.

[18] Leibe B., Leonardis A. and Schiele B., "Robust object detection with interleaved categorization and segmentation," Int. J. Computer Vision, pp. 259–289, 2008.[19] Cheng H. K. and Nevatia R., "How does person identity recognition help multi-person tracking," in Proc. of CVPR, pp. 1217-1224, 2011.

[20] Benfold B. and Reid I., "Stable multi-target tracking in real-time surveillance video," in Proc. of CVPR, pp. 3457-3464, 2011.

[21] Jia H. X. and Zhang Y. J., "Fast Human Detection by Boosting Histograms of Oriented Gradients," in Proc. of ICIG2007, pp.683-688, 2007.