F-040

Vehicle-borne Pedestrian Detection by Fusing Horizontal Laser Data and Video Images

Quanshi Zhang†, Huijing Zhao‡, Xiaowei Shao†, Ryosuke Shibasaki†, Hongbin Zha‡

1. Abstract

This paper describes a system that uses horizontal laser data to improve image-based pedestrian detection. Our image-based pedestrian detection is an extension of Depoortere et al.'s work. We build a 4-layer cascade of classifiers using both Haar wavelet features and Histograms of Oriented Gradients (HOG) features. Our main contribution is to use horizontal laser data to estimate probable positions of pedestrians on the ground plane, which limits pedestrians' searching range in an image to a small area and certain scales. We attempt to apply a laser-based SLAM system on our intelligent vehicle to further reduce the pedestrians' searching area to regions of moving objects and newly observed objects. Our system has significantly lower false positive rate and computation cost, compared to image-based pedestrian detection. To the best of our knowledge, this is the first trial to combine the laser-based moving object detection with image-based pedestrian detection.

2. Introduction

Pedestrian detection in real-world environment has many applications, such as surveillance and automatic driver assistance systems. At the same time, it is one of the most challenging object detection tasks due to large intra-class variability in poses and clothes' textures, and occluding problems.

Most previous approaches of pedestrian detection are image-based methods. Haar wavelet features and Histogram of Oriented Gradients (HOG) features are used as full-body features to detect pedestrians [1], [2], [3]. Silhouette matching [6], [7], [8] and part detectors [9], [10], [11] are also used in pedestrian detection from static images. Bastian Leibe et al. give an approach to detect pedestrians in crowded environment by fusing part detectors and silhouette matching [12]. Some approaches extract instant motion from two neighboring images in video sequence as additional features in pedestrian detection [13]. There are also some approaches in pedestrian detection fusing other sensing technologies, such as the stereo [15], laser scanner [16], [17], and radar [18].

Many ordinary image-based approaches have to do brute searching for pedestrians candidates all over the image at every possible scale [1], [2], [3], [6], [7], [8], [9]. To reduce the large computation brought by numerous candidates, the coarse-to-fine classification methods are applied in detection work, such as the cascade of classifiers in Haar-based and HOG-based approaches [4] , [5], and template hierarchy in silhouette-based approaches [7], [8]. These methods reduce the detection computation of pedestrian-unlike candidates. Papageorgiou reduces the number of candidates by focusing pedestrians' searching range on general motion area extracted from short-term video sequence [14]. His method reduces false positive rate as well.

Our final goal is to use horizontal laser data to improve the image-based pedestrian detection on an intelligent vehicle. One horizontal laser scanner is used to sense the environment in front of the vehicle. Laser points show obstacles' positions in the laser coordinate system. We use laser data to estimate the probable pedestrian positions, and project them on images to avoid brutely searching for pedestrian candidates all over the image. Depth information of laser points is used to limit the searching at some certain scales. The laser-based SLAM propounded by H.Zhao et al. [19] is applied as preprocessing to further cut down background points, and only keep laser points on moving objects and newly observed objects.

Because of high dimensions of image features and large variability in appearance between different frames, image-based region matching and tracking cost large computation and their reliability is suspicious. Thus, they are usually not applied as preprocessing in pure image-based pedestrian detection. However, the condition is different for laser data. Laser data have comparatively lower dimensions (361 dimensions here), and only describe the distance between laser scanner and objects in the laser plane. Map matching and moving-object detection and tracking do not cost much computation in our laser-based SLAM system and we use it as preprocessing of our image-based pedestrian detection. This is the first trial to combine the laser-based moving object detection with image-based pedestrian detection.

The main characteristics of this system are given below:

(1) Laser data reduce pedestrian candidates' searching range to limited area in images and certain scales. It significantly reduces the computation cost.

(2) The static background and area with no obstacles are detected and removed from searching range by SLAM when pedestrian detection is done on a mobile platform, which greatly reduces false positive rate.

(3) The laser-based SLAM can run at about 10Hz. Its computation cost of each frame is independent to the image size.

The paper is structured as follows. Section 3 introduces our intelligent vehicle and sensors. Section 4 presents our basic image-based pedestrian detection task. Section 5 proposes a novel approach which uses horizontal laser data to improve image-based pedestrian detection. Section 6 presents our experiments. Section 7 states conclusion and future work.

3. Intelligent Vehicle

Figure 1 shows our intelligent vehicle. A laser scanner (LMS291 from SICK) is mounted at the front of the vehicle, monitoring a wide angle (0°–180°, and 0.5°/point) of the environment with a scanning rate of 37.5Hz. The software

† The Center for Spatial Information Science, University of Tokyo.

‡ The State Key Lab of Machine Perception, Peking University.

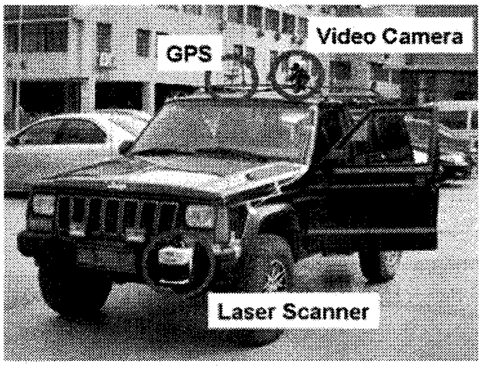

Fig. 1. The intelligent vehicle and sensors

sampling rate is set to 10Hz considering the computation efficiency of the laser-based SLAM. A video camera is mounted at the top of the vehicle to do image-based pedestrian detection and it is calibrated with the laser scanner to combine horizontal laser data with image-based pedestrian detection.

4. Image-Based Pedestrian Detection

Our image-based pedestrian detection uses both Haar wavelet features and HOG features. They are sensitive to intensity differences, not absolute brightness, and more robust to variable illumination conditions. Compared to Haar wavelet features, HOG features achieve better performance. The size of the pedestrian-candidate image clip is normalized to 128×64 pixels. The Haar wavelet feature extraction is based on the work of Papageorgiou et al. [1]. We use 3 oriented wavelet templates—vertical, horizontal, and diagonal—at two scales, 32×32 and 16 × 16 pixels, (template overlap = 3/4) to extract 1326 Haar wavelet features, and then manually select 29 most discriminating features from them. The HOG feature extraction is based on the work of Dalal and Triggs [3]. The SHIFT features are extracted from cells of 8×8 pixels based on 4 orientation bins spaced over 0°–180° ("unsigned" gradient). Features are normalized on blocks of 3×3 cells (we only take the central cell's response of one block as features). The overlap between neighboring two blocks are 3/4. Finally, we get 1860 HOG features.

Our image-based pedestrian detection is an extension of Depoortere et al.'s work [4]. They build a 3-layer cascade of classifiers using Haar wavelet features. We add HOG features to the cascade as the 4th layer. Our coarse-to-fine cascade of classifiers is configured as follows:

(1) the SVM on the basis of 29 Haar wavelet features, with a linear kernel
(2) the SVM on the basis of 29 Haar wavelet features, with a RBF kernel
(3) the SVM on the basis of 1326 Haar wavelet features, with a RBF kernel
(4) the SVM on the basis of 1860 HOG features, with a RBF kernel

5. Combination with Laser Data

This section describes the brute searching for pedestrians in image-based pedestrian detection as well as the method of using horizontal laser data to reduce searching range in images.

5.1 Brute searching for pedestrians

Without information from other sensors beside the camera, such as the laser scanner and the radar, it is very hard to extract reliable prior knowledge of possible positions and scales of pedestrians in a high speed in complex environment. Many pervious image-based pedestrian detection approaches do brute searching for pedestrian candidates at all possible positions and scales in images. The brute searching in our image-based pedestrian detection is shown below.

The size of our video images is 576×720 pixels. We search for sub-windows of pedestrian candidates at 17 scales in total. The scale decreases from 458×229 pixels, 406×203 pixels, 360 ×180 to 64×32. Each scale is 1.125 times smaller than the former one, and the height is set to twice of the width. Sub-windows shift by 0.125 times of the height in vertical direction or 0.125 times of the width in horizontal direction. Thus, we totally cut 46304 sub-windows of pedestrian candidates from one images. Then, We normalize these sub-windows of different scales to 128×64 pixels and extract Haar wavelet features and HOG features from these normalized sub-windows. In the end, We use coarse-to-fine cascade of classifiers to do detection work.

5.2 Optimization based on Laser Points

We only want to detect pedestrians on the ground plane in front of the vehicle in the application of the driving assistance system, which is the qualification for us to use horizontal laser points to reduce searching range of pedestrian candidates. The optimization work consists of the following three steps, projection of laser points on images, ideal center and scale's estimation of sub-windows and searching range determination (Figure 2).

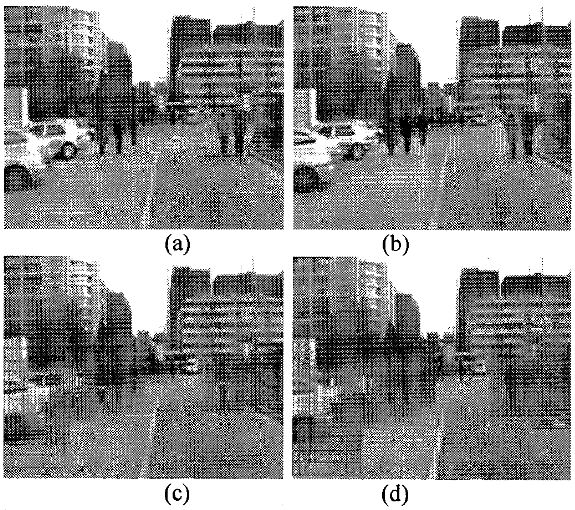

Fig. 2. Searching range reduction using raw horizontal laser data. (a) The initial image. (b) Horizontal laser points (red points) are projected on the image. (c) Ideal centers (green points) and scales (red rectangles) of laser points (red points). (d) The searching range: "active" sub-windows.

5.2.1 Projection of Laser Points on Images

The video camera is calibrated with the laser scanner to project laser points in the laser coordinate system into the image coordinate system. Figure 2(b) shows the projection of laser points on images. Some laser points with monitoring angles near 0° or 180° can not be shown on images as they are projected beyond boundaries of images. When there is no object within the maximum monitoring distance in some sensing directions, the corresponding laser points are labeled as infinite points and invalid.

5.2.2 Ideal Center and Scale Estimation of Sub-windows

The ideal scale of a sub-window is formulated as follows.

$$\frac{H_{\text{scale}}}{H_{\text{image}}} \propto \frac{h}{H} = \frac{h}{d(\tan\alpha + \tan\beta)} \qquad (1)$$

where, H_{scale} denotes the ideal scale in height. H_{image} denotes the height of images. h denotes the height of the pedestrian. d denotes the depth of the pedestrian in the direction paralleled with the optical axis. $-\beta$ and α denotes boundaries of the camera's angle range. (Figure 3)

H_{image}, α and β are constants. h is set to the average height of pedestrians manually. We set the width of scale to $W_{\text{scale}} = H_{\text{scale}}/2$. We assume that $H_{\text{laser}}/H_{\text{scale}} = \gamma$ is a constant, and estimate the ideal center of the sub-window in the position $H_{\text{scale}}/2 - H_{\text{laser}} = (1/2-\gamma)H_{\text{scale}}$ higher than the laser point, where H_{laser} denotes the height of horizontal laser scanner (Figure 4). Figure 2(c) shows the corresponding ideal scales and center points of all laser points. We ignore some too far laser points whose ideal scales are smaller than the minimum scale used in brute pedestrian searching—64 × 32 pixels.

5.2.3 Searching Range Determination

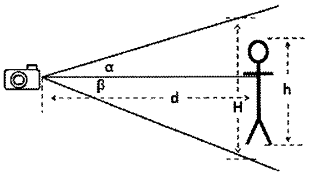

Fig. 3. The calculation of the ideal scale

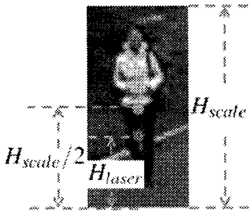

Fig. 4. The calculation of the center point

We should not only search the ideal sub-window at the estimated scale and center for each laser point, considering possible systematic errors and measurement errors. The fact that laser points may not lie on the perpendicular bisector of pedestrians, may bring errors in center estimation. The assumption that the height of pedestrians h is unique—the average height of pedestrians—may also bring large errors in scale and center estimation. Thus, to overcome these problems, we select the ideal sub-window's nearest 27 sub-windows among all of those in brute searching as "active" sub-windows.

We select three nearest scales used in the brute searching to the ideal one and then choose the nearest sub-windows in position for each scale. We shift each sub-window up, down, to the left and to the right, by one step, and get 9 sub-windows. Thus, we get 27 sub-windows at three scales in total for one laser points, and then set them as "active" sub-windows.

With the help of horizontal laser data, our searching range for image-based pedestrian detection is reduced to these "active" sub-windows (Figure 2(d)). Two neighboring laser points may share several "active" sub-windows, but one "active" sub-window is only examined once in detection work. The computation cost to process one image is O(N), where N denotes the average number of laser points projected on images, which is independent on the size of images.

5.3 Combination with Laser-Based SLAM

We use the laser-based SLAM system propounded by H.Zhao et al. [19]. It does simultaneous localization and mapping as well as detection and tracking of moving objects on an intelligent vehicle. It processes about 10 scans per second.

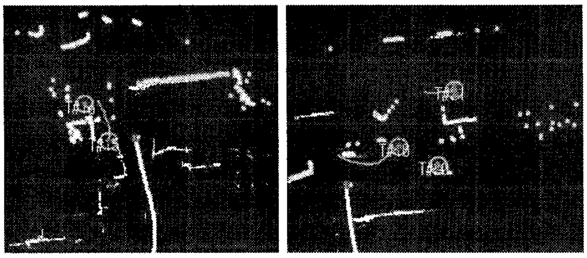

Fig. 5. Our laser-based SLAM. There are the trajectory of the intelligent vehicle (yellow curve), trajectories of moving objects (orange curves), the background detected in former scans (white points), static objects in current scan (green points), unknown objects in current scan (blue points).

Figure 5 shows our laser-based SLAM. Where, only static object will be updated to background in later scans, and moving objects are subtracted from background. Unknown objects (green points) are usually objects newly observed by horizontal laser scanner and SLAM system needs information of more scans to judge whether they are static or mobile. We limit our searching range to the area near the moving objects and unknown objects.

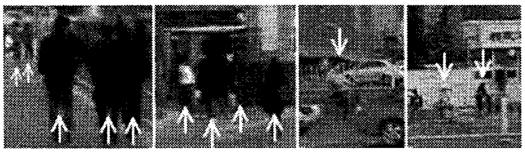

Fig. 6. Difficult cases for detection in our video. Arrows show pedestrians in the video.

6. Experiments

6.1 Video Data

We have two video sequences of street scenes with many pedestrians recorded by the video camera on our intelligent vehicle (Section 3) in different places. We create a training set for image-based pedestrian detection using one video sequence recorded in a sunny day, and do the detection task on the other video sequence recorded in a cloudy day.

The quality of video sequences is not good and challenging for image-based pedestrian detection. As both the two video sequences used in training and detection are taken on our running intelligent vehicle, the quality of video images may suffer more or less from shaking of the video camera. Due to the sampling rate, the video camera can only get clear image record

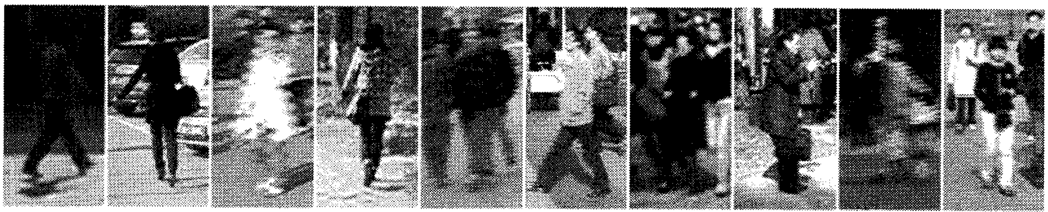

Fig. 7. Some positive samples from our training sample set. Pedestrian are always upright, but with much blurredness, partial occluding and a large variability of illumination and poses.

of low-speed objects, so the quality of video images decreases dramatically when our vehicle runs fast or turns around. Some clips in training set normalized from very small scales are very blurred. Illumination changes significantly from place to place, as there are many shadows of buildings and trees in the street. There are many occluding cases among our positive training samples. Figure 6 shows some difficult cases for detection in our video.

The training set contains 882 positive samples and 2151 negative samples. Positive samples are clips of pedestrians cut from video images, and negative samples are clips of background. Each clip is normalized to 128×64 pixels. Figure 7 shows some of sour positive samples.

	Image-based detection	Combination with raw laser data	Combination with laser-based SLAM
Detection rate	63.7%	50.8%	36.5%
False positive rate	56.6%	10.2%	5.6%
Sub-window number	46304.0	685.7	371.4
Time cost (ms)	14429	413	264

Table 1. Experimental result

The video used in detection lasts 26.4 seconds, containing 661 frames. There are 25 pedestrians including very far ones, blurred ones, ones riding bicycles and ones appearing in only a few frames. We uniformly select one from every five frames to do both the image-based pedestrian detection and detection with help of raw laser data (without SLAM). Due to the efficiency of SLAM (about 10Hz), our laser-based SLAM uniformly gives outputs for 246 corresponding video frames, and we do the pedestrian detection combining with SLAM on these frames.

6.2 Laser-based SLAM

We collect the horizontal laser data at the same time when we take the video used in detection. Figure 8 shows the 2D map of background and moving objects detected by our laser-based SLAM. There are 26 trajectories of moving objects in total, but some of them are very tiny due to occluding problems and sensing accuracy. Not all the pedestrians are detected as moving objects, but its detection rate does not affect our detection work much as we only use SLAM to cut static objects and background from our searching range. We only focus our searching range near laser points of moving objects and unknown objects (newly observed objects). Figure 9 shows reduction of searching range. We only focus on laser points on moving objects and newly observed unknown objects (Figure 9.1(a), 9.2(a)). We put corresponding sub-windows into the cascade of classifiers to select pedestrians, and ignore other points and their corresponding sub-windows (Figure 9.1(b), 9. 2(b)).

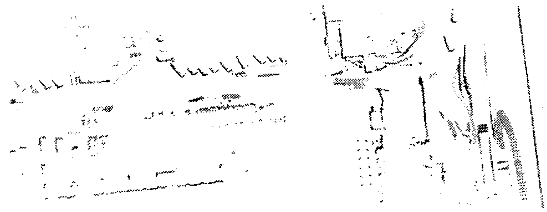

Fig. 8. The 2D map of background and moving objects detected by the laser-based SLAM. There are the trajectory of the intelligent vehicle (yellow curve), background (black points), trajectories of moving objects (curves with various colors).

6.3 Pedestrian Detection Results

Table 1 shows the quantitative results of different pedestrian detection systems. We do not put the sub-windows very far from our vehicle whose corresponding ideal scales are smaller than the minimum boundary—64×32, into our image-based detector, as we want to pay more attention on those near pedestrians in the application of the driving assistance system. However, those very far pedestrians may still be detected in the pure image-based system with brute searching. To make the comparison between different systems meaningful, we ignore all these very far pedestrians detected in the three systems in statistics.

The low detection rate in all the three detection systems is due to our very blurred video images with extremely small contrast between pedestrians and environment. Our image-based pedestrian detector performs much better in MIT and INRIA pedestrian databases.

Combination with laser data significantly reduces false positive rate from 56.6% to 10.2% and 5.6%, but detection rate decreases from 63.7% to 50.8% and 36.5% as well. Figure 9.1(c) and Figure 9.2(c) intuitively show how false detections are reduced step by step in our three detection systems.

The detection rate of pure image-based detection is higher than it of combination with raw laser data by about 13%. The reason is that searching range is limited to "right" positions and "right" scales in systems combining with laser data, but in pure image-based system, pedestrian may be detected not only in these "right" range, but also in other "not so right" range with brute searching. In some cases, left or right side and legs of a pedestrian may be detected as pedestrians. In other cases, a pedestrian may be detected in the center of a so huge sub-window that it is even taller than 3 times of the pedestrian's height. However, whether these detections in "not so right" range should be counted as correct detections is controversial.

There are two reasons for the low detection rate of the system combining with laser-based SLAM. One is that some pedestrians may be identified as static objects in a few scans if they are observed very far and walking slow by the laser scanner. Due to measurement error of the laser scanner and motion of the intelligent vehicle, laser-based SLAM is not very sensitive to low-speed motion, though it can easily detect and track high-

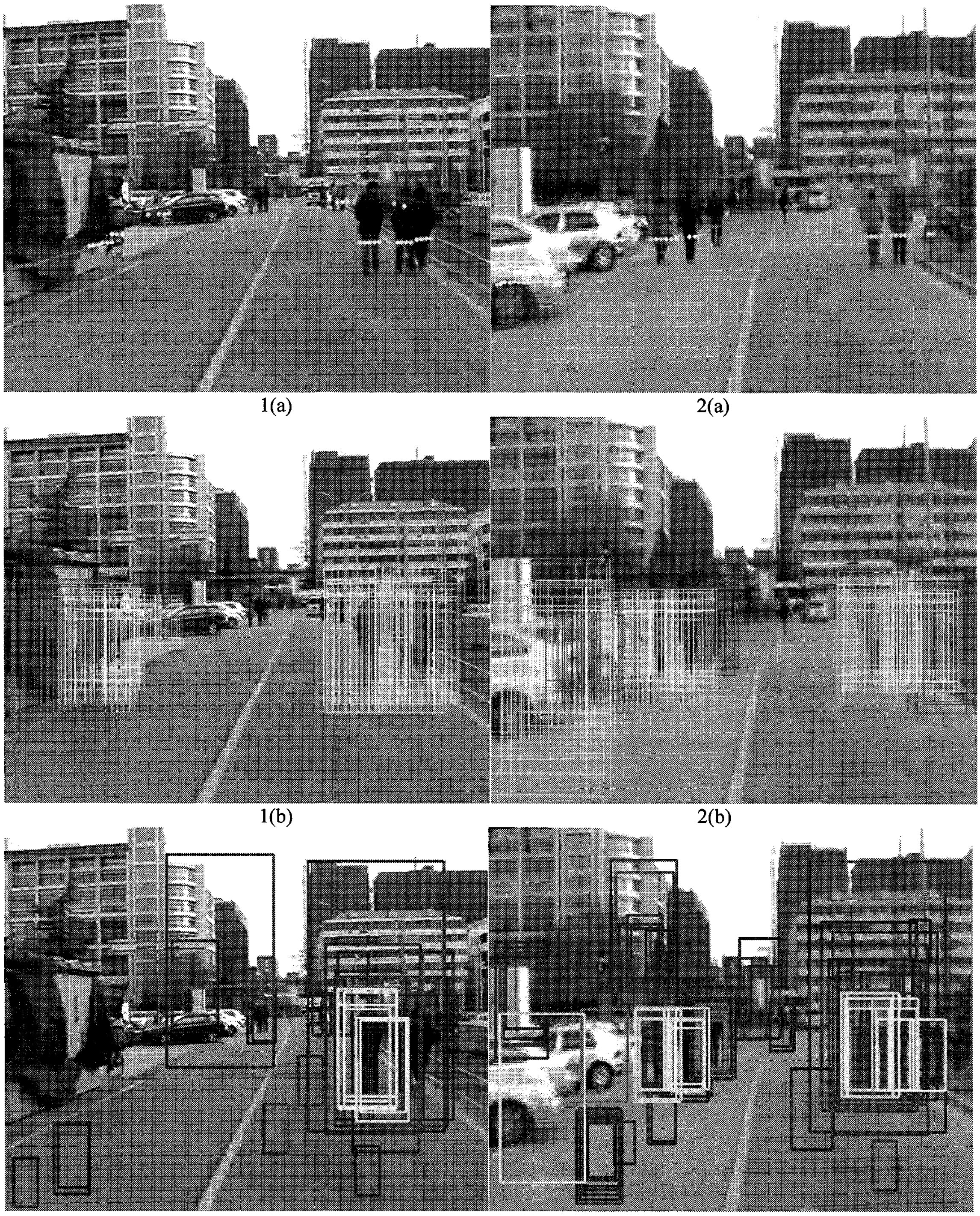

Fig. 9. Pedestrian detection combining with laser-based SLAM. 1(a),2(a) Laser-based SLAM separate laser points on moving objects and unknown (newly observed) objects (green points) and laser points on static objects and background (red points). Only green points are used in detection work. However, some very far pedestrians may be judged as static ones temporarily due to measurement error of laser scanner. 1(b),2(b) Searching is done among sub-windows of moving objects and unknown objects (green). Sub-windows of static objects and background (red) are ignored. 1(c),2(c) All the sub-windows are detection results of pure image-based pedestrian detection. Combination with raw laser points reduces blue sub-windows. Further combination with laser-based SLAM reduces red sub-windows and only leave green ones.

speed objects such as cars and bicycles. However, in this case, most of them can be identified as unknown objects or moving objects in later scans by the SLAM and be detected. The other is that if one pedestrian keep standing still for a while, he will be judged as a static object by SLAM and will not be put into image-based detector to do further examination.

Combination with raw laser data dramatically reduces the number of sub-windows used in the image-based detection by 98.5%, and reduces time cost by 97.1%. Combination with laser-based SLAM further reduces the number of sub-windows by 45.8%, and reduces time cost by 36.1%. As the cascade of classifiers in image-based pedestrian detection distributes pedestrian-like sub-windows more classification computation than other sub-windows, the time cost does not linearly decreases with the decrease of sub-window number.

7. Conclusion and Future Work

In this paper, we represent a novel system that uses laser data to improve image-based pedestrian detection. With the help of laser data and laser-based SLAM, both the computation cost for processing each video frame and false positive rate significantly decrease. However, detection rate decreases as well.

Combination with laser-based SLAM is just our primary trial to use high-level information of laser data. There are many problems with its reliability and stability, as we just simply use the output of SLAM in our pedestrian detection. We will find more robust approaches to fuse laser-based SLAM and image-based pedestrian detection, such as using the pedestrian detection result to assist the moving object detection in SLAM.

The tracking of objects in SLAM provides motion information for pedestrian detection and makes it possible for us to detect one pedestrian in continuous several frames. We can also apply the shape information extracted from either 2D laser data or 3D laser data in pedestrian detection in the future.

Reference

[1] Papageorgiou, C., Evgeniou, T., Poggio, T., “A trainable pedestrian detection system” IEEE Conference on Intelligent Vehicles, (1998) 241–246

[2] Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., Poggio, T., “Pedestrian detection using wavelet templates” IEEE Conference on Computer Vision and Pattern Recognition (1997) 193–199

[3] Dalal, N., Triggs, B., “Histograms of oriented gradients for human detection” IEEE Conference on Computer Vision and Pattern Recognition, vol. 2 (2005) 886–893

[4] Depoortere, V., Cant, J., Van den Bosch, B., De Prins, J., Fransens, R., Van Gool, L., “Efficient pedestrian detection: a test case for svm based categorization” Workshop on Cognitive Vision (2002)

[5] Zhu, Q., Avidan, S., Yeh, M.C., Cheng, K.T., “Fast human detection using a cascade of histograms of oriented gradients” IEEE Conference on Computer Vision and Pattern Recognition, vol. 2 (2006) 1491–1498

[6] Felzenszwalb, P.F., “Learning models for object recognition” IEEE Conference on Computer Vision and Pattern Recognition (2001) 1056–1062

[7] Gavrila, D.M., “Pedestrian detection from a moving vehicle” European Conference on Computer Vision (2000) 37–49

[8] Gavrila, D.M., Philomin, V., “Real-time object detection for ’smart’ vehicles” IEEE International Conference on Computer Vision (1999) 87–93

[9] Mohan, A., Papageorgiou, C., Poggio, T., “Example-based object detection in images by components” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23 (2001) 349–361

[10] Opelt, A., Pinz, A., Zisserman, A., “Incremental learning of object detectors using a visual shape alphabet” IEEE Conference on Computer Vision and Pattern Recognition (2006) 3–10

[11] Leibe, B., Schiele, B., “Interleaved object categorization and segmentation” British Machine Vision Conference (2003) 759–768

[12] Leibe, B., Seemann, E., Schiele, B., “Pedestrian detection in crowded scenes” IEEE Conference on Computer Vision and Pattern Recognition (2005) 878–885

[13] Viola, P., Jones, M., Snow, D., “Detecting pedestrians using patterns of motion and appearance” International Conference on Computer Vision (2003) 734–741

[14] Papageorgiou, C., “Object and pattern detection in video sequences” Master’s thesis, MIT (1997)

[15] Zhao, L., Thorpe, C.E., “Stereo- and neural network-based pedestrian detection” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1 (2000) 148–154

[16] Zhao, H., Zhang, Q., Chiba, M., Shibasaki, R., Cui, J., Zha, H., “Moving object classification using horizontal laser scan data” IEEE International Conference on Robotics and Automation (2009) 2424–2430

[17] Frerstenberg, K.C., Lages, U., “Pedestrian detection and classification by laserscanners” IEEE Intelligent Vehicles Symposium (2002)

[18] Milch, S., Behrens, M., “Pedestrian detection with radar and computer vision” Conference on Progress in Automobile Lighting (2001)

[19] Zhao, H., Chiba, M., Shibasaki, R., Shao, X., Cui, J., Zha, H., “Slam in a dyanmic large outdoor environment using a Laser scanner” IEEE International Conference on Robotics and Automation (2008)