Pedestrian Detection on Moving Vehicle Using Stereovision and 2D Cue

Yang Yang[1,2], Jingyu Yang[1], and Dongyan Guo[1]

[1] School of Computer Science and Technology,
Nanjing University of Science and Technology
[2] School of Computer and Information Engineering, Henan University

Abstract. We present a novel approach for pedestrian detecting on moving vehicle which equipped with low-cost cameras. Our approach is working in a framework which combines two-dimensional human body characteristics and three-dimensional information such as parallax and distance. By constructing a SPM (surface parallax map), it calculates parallax of object which do not belong to the road plane such as human body and obstacles. After recording the scores of all road area, an occlusion image is created, in which high density area indicates people's most likely appearance. Then a SVM (support vector machine) classifier is trained to classify pedestrian and non-pedestrian windows in candidate area. We also propose an algorithm to maintain SPM in real time. We evaluate our approach on real data which are taken from crowded city areas, the efficient and accurate results are demonstrated.

Keywords: stereo vision, support vector machine, pedestrian detection, hog, surface parallax map.

1 Introduction

Nowadays, Along with the development of auto industry, car ownership in China is rising straight up. By the end of 2011, this number reached 104 million. But there was another fact behind this number, 6.2 million people were killed in traffic accidents in 2011. In this decade, a new study field called ADAS [1] (advanced driver assistance systems) attracted more and more attention which can reduce accidents by adding auxiliary equipment on vehicle. The ADAS is designed to prevent or at least reduce the harm from people in traffic accidents. Pedestrian detection system is a typical ADAS, and it is a very challenging task because irregular movement of human and the complex background where people appear have increased the difficulty of detecting.

Computer vision is widely used in the pedestrian detection systems, which make conclusion simulating human analysis. By processing the data acquired from video capture device, such systems perceive the traffic area to determine the location of the pedestrian. Because of the complexity of pedestrian characteristics and background in real traffic, only 2D information is insufficient to provide enough clues. Nowadays, more and more pedestrian detection system

J. Yang, F. Fang, and C. Sun (Eds.): IScIDE 2012, LNCS 7751, pp. 466–474, 2013.

introduced stereo vision. By extracting scene depth information, stereo vision makes a significant contribution to detecting pedestrian. After candidate areas which contain people have been determined, detection algorithms use classifiers to verify the windows containing pedestrian or not. Classifiers commonly use 2-D feature as human characteristics such as Harr [2], EOH [3] and HOG [4], etc. We present pedestrian detecting method combined 3-D information of scene with 2-D human features, which can detect pedestrians and obstacles accurately and efficiently.

2 Related Work

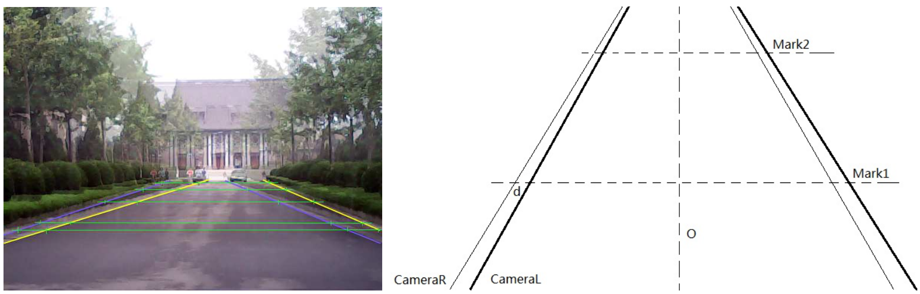
ROIs (region of interest) are hot zones with pedestrian appearing. ROI extraction is a common pre-operation in pedestrian detection system, which can avoid exhaustive searching from the whole image region and identify an accurate location of the pedestrian. Solid objects can be extracted as pedestrian candidates with disparity segmentation and fixed-size window is used to remove non-pedestrian [5]. Image is segmented into sub-image object candidates using disparities discontinuity, then a connected-component grouping operator is applied to find the pedestrian regions with smoothly varying range [6]. GAVRILA uses a feature-based, multi-resolution stereo algorithm to generate stereo-based ROI [7]. A multi-resolution approach is used to perform stereo analysis by finding correspondences on coarse level that can be recursively refined, and in order to reduce the computational cost Franke uses edge feature to match local correlation [8]. David presents a three module system in literature [1] which uses 3D information to estimate the road plane parameters and select ROIs, uses Real AdaBoost combined Haar wavelets and EOH to classify ROI windows. 3D information is extracted using stereo reconstruction algorithm which finds pairs of left-right correspondent edge points and maps them into the 3D world, then these 3D points are grouped into objects such as human body according to their distance to each other [9]. In order to reduce the search space, only edge points in left-right images are matched in stereo reconstruction algorithm. A 3D lane model is introduced to distinguish between road surface and target points by projecting those points in three specified planes [10]. In this paper, an efficient onboard pedestrian detection method is proposed. It equips with low-cost cameras, calculates occlusion image according to SPM, extracts ROIs from occlusion image by analysing the parallax pattern, adopts SVM training by using HOG feature to classify pedestrian or non-pedestrian.

3 SPM-Based ROI Generation

Within stereo vision the disparity map is generally used to calculate depth information of scene. We propose a variant of disparity map by restricting matching strategy on a specific plane–road surface to improve the efficiency of matching. This variant is so called Surface Parallax Map. Parallax is the object's different position in left-right images. Consider pedestrians and other non-road objects are located between cameras and road surface, some parts of the road surface must be occluded. By analysing occlusion pattern, locations of object on road can be calculated.

3.1 Surface Parallax Map

Traditional pedestrian detection methods filter feature points by considering their coordinates. Using our methodpedestrians within range of 50 meters from vehicle can be detected by using low-cost cameras. By introducing GCP (Ground control point) in matching algorithm, accuracy of matching can be improved. GCP was first proposed by AARON [11] in order to improve sensitivity of DSI (Disparity-space image). In this paper, GCPs are used to help synthesizing disparity of pixels on road surface into a SPM. The theory of SPM is described as follows, as shown in Fig. 1.



(a) (b)

Fig. 1. Surface Parallax Map Initialization. (a) A frame in video 1 with manually calibrated GCP. (b) Sketch of GCPs and Control Lines.

Fig.1a shows an overlap image which has obvious parallax. The original images are captured by two fixed-angle cameras at same time. As shown in Fig.1b, O is a line which passes the midpoint of two optical centers and is perpendicular to the base line. Markers are GCPs on road surface, and control lines lying at right and left side pass their GCPs respectively. All the control lines are a set of parallel lines in road plane. These bold lines which called CameraL belong to left camera and CameraR belong to right camera, d is the parallax in horizontal direction in current position. When cameras are calibrating, two optical axes of cameras must be restricted in the same plane in order to avoid parallax in vertical direction. Considering driving area in front of vehicle within a certain distance can be approximated as a plane, parallax and distance from camera to target are linear correlation. That is, parallax and ordinate of target in image are linear correlation. Because the camera equipped can only provide resolution of 320X240, details provided by pixels are limited in far distance. In order to produce visible parallax at 50 meters, we set the two optical axes intersecting behind the camera.

SPM algorithm is described in formula 1, Parallax d_y is given by:

$$d_y = SPM(y) = \frac{1}{n}\sum_{i=1}^{n} \frac{(y-\Delta y)\frac{1}{m}\sum_{j=1}^{m} d_j}{y_i - \Delta y} \ . \tag{1}$$

Where y is the ordinate in image, the origin of image is on the upper left corner, Δy is the ordinate of the horizon, m is the number of control lines, the minimum value is 2, n is the number of GCPs on every control line, y_i is ordinate of ith GCP.

By giving d_y it adopts cost function to calculate occlusion image. Literature [8] introduces classical SSD (Sum-of-squared Differences) and SAD (Sum-of-absolute Differences). Literature [12] evaluates window-based matching technique called ZSAD (Zero-mean Sum of absolute Differences), NCC (Normalized Cross Correlation) and ZNNC (Zero-mean Normalized Cross Correlation). We compare the performance of these cost functions in experiment, the results are demonstrated in Section 5.

In general, SSD gives a better representation of difference between feature points than SAD. ZSAD subtracts the mean intensity of the window form each intensity inside the window before computing the sum of absolute differences. As described in formula 2, where I_L and I_R are pixel values at coordinate (x, y) in left and right images, Ω is the neighborhood of current pixel, using rectangle with longer horizontal edges, $\overline{I}(x, y)$ expresses mean intensity of neighborhood Ω.

$$OI_{ZSAD}(x,y) = \sum_{(x,y\in\Omega)} (I_L(x,y) - \overline{I}_L(x,y) - I_R(x-d_y,y) + \overline{I}_R(x-d_y,y)) \ . \tag{2}$$

$$\overline{I}(x,y) = \frac{1}{n}\sum_{(x,y\in\Omega)} I(x,y) \ . \tag{3}$$

NCC compensates for gain changes and is statistically the optimal method for dealing with Gaussian noise. ZNCC introduces zero mean in NCC and can compensate for differences in both gain and offset within the correlation window. Fig. 2 are the occlusion image using cost function mentioned above which have obvious occluded area on road surface. In Section 4, we train a SVM to classify pedestrian or non-pedestrian area from occlusion image.

3.2 Real-Time SPM Correction

In Section 3.1, we propose a hypothesis that the driving area in front of vehicle is always belong to the same plane. However, this assumption is not fully satisfied in actual road. For example, the angle between camera orientation and road surface changes when vehicle going uphill or downhill. Fig.3 shows the details, where Z orientation is the forward direction of car, as the road surface changes, driving direction will change around three axes correspondingly. Because cameras are fixed to the vehicle, camera rotating around Y and Z axis does not affect SPM,

(a) (b)

Fig. 2. Occlusion image generated by using SPM. (a) The 150 frame in video 2. (b) Occlusion image of this frame.

but rotating around X axis will influence results of formula 1 . In this paper, a real-time road surface fitting algorithm is proposed which using angle between road and horizontal to correct SPM output.

In the previous literature, stereovision-based road fitting technique is divided into three categories. The first category extracts road surface information by detecting the lane and roadside. A lane model is proposed in literature [9] [10] [13], surface parameters are calculated by lane model which derives from road geometry. The second category detects road feature points, calculates 3D coordinates of these points by using stereo vision technique, then, fits road surface. A RANSAC based fitting is applied over 2D barycenters intended for removing outlier cells, road surface parameters are computed by means of a least squares fitting over all 3D points contained in inlier cells[1]. A RANSAC based technique is used for finding the best fitting plane to those points which belong to the road [14]. The third category apply SFM (Structure from motion) based technique to fit road plane. By recording contact points between road and wheels when the car is moving, road plane is reconstructed between roadside limit lines[15]. The limitations of such methods is that only passed area can be rebuilt. The proposed method first matches edge feature points in images to obtain point cloud of road area, then projects those points to YOZ plane and analyses the angle between road and horizontal, uses formula 8 to correct SPM output. Since the upper part of the image is usually occupied by meaningless background such as trees and buildings, our matching algorithm only consider the lower half of images, which is the road region. Because of the perspective effect of road in images, texture of road surface is oblique. The results of matching horizontal or vertical edge feature separately are not satisfying. In this paper, we use Canny operator as edge detector. By finding the local maximum of gradient, Canny algorithm can generate continuous edges which are conducive to matching road feature points. In experiment,$\sigma = 1$, T_H is decided by the image itself, $T_L = 0.5T_H$.

Matching algorithm of feature points adopts Cost Function used to calculate occlusion image in Section 3.1. As shown in Fig.4, the 3D coordinates of feature points are projected to YOZ plane, β expresses the angle between road and horizontal, the line with minimum β and passing maximum points is selected as the road surface projection.

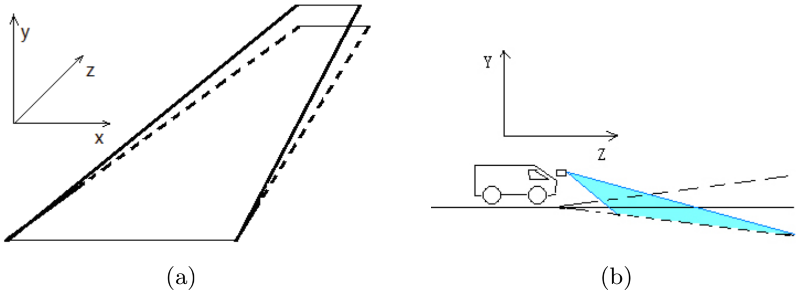


Fig. 3. SPM affected by Angle between road surface and horizontal. (a) Driving direction of vehicle. (b) Side view of uphill and downhill.

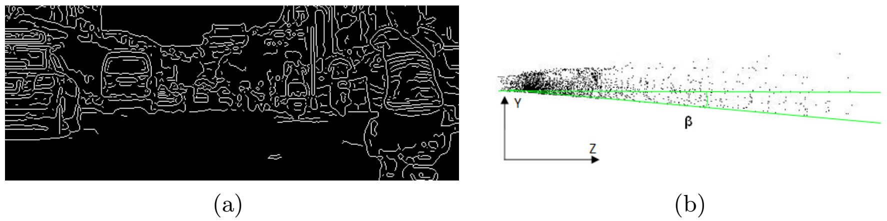


Fig. 4. Edge feature based road fitting. (a) Edge feature of this frame. (b) Feature points projected to YOZ plane, β is the angle between road and horizontal.

After correcting the formula of parallax d_y is modified as:

$$d_y = SPM(sy) = \frac{1}{n}\sum_{i=1}^{n}\frac{(sy-\Delta y)\frac{1}{m}\sum_{j=1}^{m} d_j}{y_i - \Delta y} \ . \tag{4}$$

Where s is the correction factor, H is the height from camera to road surface, L_i is the distance from each GCP to the camera center, L is the road surface distance from current target to camera, β is the angle between road and horizontal which is positive uphill and negative downhill.

$$s = \frac{\tan\alpha}{\tan\alpha + \tan\beta} \ . \qquad \tan\alpha = \frac{H}{L} \ . \qquad L = \frac{1}{n}\sum_{i=1}^{n}\frac{y-\Delta y}{y_i - \Delta y}L_i \ . \tag{5}$$

4 Pedestrian Verification in ROI

Many human features are proposed to classify detecting windows as pedestrian or non-pedestrian in previous literatures. Some detecting methods exhaustively search the entire image and cost most of the time. Under the premise of ROIs, detecting can be concentrated in the most likely area. GAVRILA proposes a shape-based detector and a texture-based classifier in literature [7]. David uses a Real

AdaBoost classifier combined with HW(Haar wavelets) and EOH (Edge orientation histograms) to locate pedestrians. In our approach, cameras are calibrated using Zhangs technique [16], uses triangulation to calculate size of the detecting window on ROIs position, a line SVM classifier [17] is trained using HOG as pedestrian feature. During the training step, the size of window is 128x64 pixels, cell is 8x8 pixels, histogram is calculated in every cell, a block is made up by 2x2 cells, L2-Hys normalize is implemented in every block. Non-maximum suppression is used on classification results to determine the accurate location of pedestrians.

Fig.5 shows the result images.

(a) (b)

Fig. 5. Final results and candidate ROIs. (a) is a frame from video 3. (b) are the candidate ROIs proposed by analysing Occlusion pattern.

5 Experiment

In experiments, We evaluate our approach on four challenging video sequences which are captured from the urban traffic. The experiment platform is equipped with Intel Core2 2.0GHz CPU and 2.0GB RAM. The main purpose of this approach is to detect pedestrian in front of the vehicle. Although our system is equipped with low-cost cameras, it can detect targets at the range of 50 meters with viewing angle of 50 degrees. The five cost functions mentioned in Section 3.1 are evaluated in experiments in both execution time and disparity between ROI and Neighborhood. ZSAD is chosen for its best overall performance which can process three frames in one second. Specialized commercial vision system and 3D reconstruction software have been used to extra 3D information of the region in front of the host vehicle [1]. On the contrary, our method generates ROI by matching parallax differences of road surface. Under the premise of using the same image capture device, the performance of SPM based method is significantly better than stereo matching based method [1] [5]. Experiment details are shown in Table 1. Since our method only detects pedestrian in ROI, it can improve the detection speed than method adopting exhaustive strategy [4].Part of experiment results are shown in Fig.5. Occlusion image based ROI extraction provides candidate regions. At the same time, some non-pedestrian regions are also proposed. After the candidate regions are verified by HOG feature based SVM, the false detection is wiped out.

Table 1. Performance of ROI generation algorithm using SPM and stereo match

Algorithm	Max hit rate	Max effective distance	Processing time
SPM-based	96%	50 m	3 fps
Stereo-match-based	60%	15 m	0.4 fps

6 Conclusion

In this paper, we have presented a novel approach for pedestrian detection from moving vehicles. Our approach relies on the difference between parallax of pedestrian projecting to the road plane and the parallax of road itself. Because of SPM that we propose in section 3, the ROIs extraction algorithm based on parallax comparison is efficiently performed. A linear SVM classifier is trained with HOG features to label the selected ROIs as pedestrians or non-pedestrians. In addition, we have proposed an algorithm to maintain SPM by fitting road plane in real time.

We evaluate our approach on several video sequences taken in urban traffic. The experiments demonstrate that our approach is able to obtain reliable detecting results, even with low-cost cameras. In this paper, we have focused on detecting pedestrians. This can be extended to other object categories for which have different parallax from road plane, above or below.

References

1. Geronimo, D., Sappa, A.D., Ponsa, D., Lopez, A.M.: 2d-3d-based on-board pedestrian detection system. Computer Vision and Image Understanding 114(5), 583–595 (2010)
2. Papageorgiou, C., Poggio, T.: A trainable system for object detection. International Journal on Computer Vision 38(1), 15–33 (2000)
3. Levi, K., Weiss, Y.: Learning object detection froma small number of examples: the importance of goodfeatures. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, CO, USA, pp. 53–60 (2004)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1(38), pp. 886–893 (2005)
5. Soga, M., Kato, T., Ohta, M., Ninomiya, Y.: Pedestrian detection with stereo vision. In: Proceedings of the IEEE International Conference on Data Engineering, Tokyo, Japan (2005)
6. Zhao, L., Thorpe, C.E.: Stereo and neural network based pedestrian detection. IEEE Transactions on ITS 1(3), 148–154 (2000)
7. Gavrila, D., Munder, S.: Multi-cue pedestrian detection and tracking from a moving vehicle. International Journal on Computer Vision 73(1), 41–59 (2007)
8. Franke, U., Joos, A.: Real-time stereo vision for urban traffic scene understanding. In: Proc. of the IEEE Intelligent Vehicle Symposium, Detroit, USA, pp. 273–278 (2000)

9. Nedevschi, S., Schmidt, R., Graf, T., Danescu, R., Frentiu, D., Marita, T., Oniga, F., Pocol, C.: High accuracy stereo vision system for far distance obstacle detection. In: Proc. of IEEE Intelligent Vehicles Symposium, Parma, Italy, pp. 161–166 (2004)
10. Nedevschi, S., Danescu, R., Marita, T., Oniga, F., Pocol, C., Sobol, S., Graf, T., Schmidt, R.: Driving environment perception using stereovision. In: Proc of IEEE Intelligent Vehicles Symposium, Las Vegas, USA, pp. 331–336 (2005)
11. Bobick, A.F., Intille, S.S.: Large occlusion stereo. International Journal of Computer Vision 33(3), 181–200 (1999)
12. Hirschmuller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans. Pattern Analysis and Machine Intelligence 31(9), 1582–1599 (2009)
13. Danescu, R., Sobol, S., Nedevschi, S., Graf, T.: Stereovision-based side lane and guardrail detection. In: Proceedings of the IEEE International Conference on Intelligent Transportation Systems, Toronto, Canada, pp. 1156–1161 (2006)
14. Sappa, A., Geronimo, D., Dornaika, F., Lopez, A.: On-board camera extrinsic parameter estimation. Electronics Letters 42(13), 745–747 (2006)
15. Cornelis, N., Cornelis, K., Gool, L.V.: Fast compact city modeling for navigation pre-visualization. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2006)
16. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)
17. Yang, Y., Yang, J.Y.: Pedestrian detection based on compound feature. Journal of Image and Graphics 17(5), 671–675 (2012)