STEREO BASED REGION OF INTEREST GENERATION FOR PEDESTRIAN DETECTION IN DRIVER ASSISTANCE SYSTEMS

Maral Mesmakhosroshahi⋆*, Joohee Kim*⋆*, Yunsik Lee*†*, and Jong-bok Kim*††∗

⋆ Illinois Institute of Technology, Dept. of Electrical and Computer Engineering, Illinois, USA.
†Korea Electronics Technology Institute (KETI), Bundang-gu, Sungnam-si, Gyungki-do, South Korea.
††Sane System, 439 DongAhn-Gu, AnYang-Si, Gyungki-Do, South Korea.

ABSTRACT

Pedestrian detection is one of the major goals in advanced driver assistance systems (ADAS) which has become an active research area in recent years. In this paper, we present a stereo based pedestrian detection system by fusing the depth and color data provided by a stereo vision camera on a moving platform. The proposed method uses an adaptive window for region of interest (ROI) generation using dense depth map. The extracted candidates are then applied to a Histogram of Oriented Gradients (HOG) feature descriptor to refine ROIs and Support Vector Machine (SVM) is used to classify them into pedestrian and non-pedestrian classes. The system is tested on a stereo based DAS dataset and results show that our system is able to detect pedestrians with different scales and illumination conditions and in presence of partial occlusion.

***Index Terms*—** Advanced driver assistance system, ROI generation, pedestrian detection, stereo vision, depth map

1. INTRODUCTION

In an effort to reduce driver errors in being the major cause of traffic accidents, there is a lot of research being conducted into the development of advanced driver assistance systems (ADAS). ADAS is a system aimed at helping the driver in tasks such as pedestrian and vehicle detection, traffic sign recognition, lane detection, etc.. Such systems are classified into two categories based on the nature of the sensor used to acquire data. In the first category, active sensors such as lasers and radars are used and in the other category, passive vision based sensors are used. A major drawback of active sensors is that colors and patterns cannot be detected, making object recognition very difficult. For this reason, passive sensors are preferred because image processing techniques can be used for interpreting and processing the information.

One of the main goals of driver assistance systems is to locate and track pedestrians in the path of the vehicle. Pedestrian detection is a challenging task due to several complexities such as cluttered backgrounds (especially in urban areas), illumination variation, weather condition, non-uniform shapes and postures of pedestrians and camera movement. Most of the pedestrian detection algorithms follow two major steps which are region of interest (ROI) generation and pedestrian classification. In some cases, a tracking and refinement step is also added to improve the detection rate. Fig. 1 shows the general block diagram of pedestrian detection.

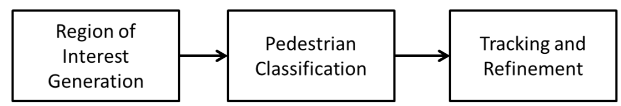


Fig. 1. Block Diagram of Pedestrian detection.

Several methods have been proposed for detecting humans with a static camera, however, pedestrian detection from a moving platform is a more complex task. Early pedestrian detection methods used a monocular camera mounted on a moving car and used techniques such as contour extraction and chamfer distance [1]. A pedestrian detection method using a moving camera is proposed in [2] which focuses on sudden detection of partially visible pedestrians in a three level approach. In [3] an on-board pedestrian detection and warning system is described using a monocular camera. In this algorithm, crosswalks are detected by image processing methods and optical flow is used for detecting moving objects on the crosswalk.

However, pedestrian detection and tracking done using monocular cameras and only RGB data is not reliable. Of late, researchers have started using stereo vision for designing driver assistance systems. Using stereo vision gives the opportunity to take advantage of object depth for detection and tracking. Early stereo-based approaches for object detection using moving cameras were designed for robotic vehicles such as NASA's Mars rover [4] that uses 3-D optical flow for detection. In [5], a GPU-based system is proposed for pedestrian detection in mobile robots using stereo vision that uses edge properties to extract ROIs. In this system, HOG and SVM is used for classifying pedestrians. In applications such as driving assistance systems, we have to deal also with the high speed of the car. In the past few years, a few approaches have been proposed for stereo-based pedestrian detection using a camera on a moving vehicle. A real time stereo-based system is proposed in [6] that uses ROI detection by projecting the stereo data into a polar-perspective map and classifies objects by computing geometric features. Another stereo-based method is proposed in [7] which uses Viola-Jones detection method [8] for ROI generation in left and right image and performs tracking based on correlation. The stereo-based system presented in [9] is a stereo modification of the system introduced in [1]. A method based on dense stereo is proposed in [10] that uses dense depth information for both detection and classification. A 2D-3D based pedestrian detection algorithm is presented in [11] that selects ROIs by estimating the road plane parameters. A Real AdaBoost and a combined set of Haar wavelets and edge orientation histograms are used in this algorithm for classification. In [12] also dense stereo and road and camera parameters are used for ROI generation.

∗This work was supported by the Technology Development Program for Commercializing System Semiconductor funded By the Ministry of Trade, industry & Energy (MOTIE, Korea). (No. 10041126, Title: International Collaborative R&BD Project for System Semiconductor).

However, some of these methods face problems in case of illumination and scale variation and in presence of occlusion. In this paper, we propose a pedestrian detection algorithm by combining the ROI generation using dense depth map and HOG feature descriptor in the color image to create a system that is invariant to scale and illumination and can handle occlusion and multiple objects.

The structure of the paper is as follows: In Section 2 the proposed pedestrian detection algorithm is explained in detail. In Section 3 we present experimental results and in Section 4 the paper is concluded.

2. PROPOSED ALGORITHM

In this paper, we propose a new pedestrian detection algorithm using both color and depth values from a stereo vision based DAS. As discussed above, lots of stereo based methods are available, however, a few of them use dense depth map in the detection process. In our proposed algorithm, we take advantage of the dense depth to present a novel method for ROI generation which is one of the most important steps in detecting pedestrians since missing any pedestrian in this step cannot be recovered.

In traditional methods, possible pedestrian locations were found by exhaustive search. However, exhaustive search based techniques incur huge computational complexity and have several drawbacks in detecting pedestrians with different scales, illumination conditions and in presence of occlusion. Also, because of the camera motion in DAS applications, scale of each object is changing frame by frame and it is very important to have a scale invariant ROI generation method to be able to detect objects in each frame. Motivated by these observations, we present an adaptive windowing method for ROI generation using dense depth. The detected ROIs are applied to a feature descriptor such as Histogram of Oriented Gradients for classification. Fig. 2 shows an overview of our system.

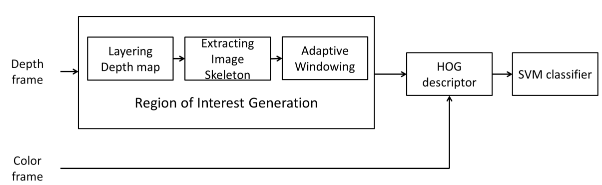


Fig. 2. Overview of our proposed algorithm.

2.1. ROI generation using dense depth

A depth map is an image in which the value of each pixel is related to the distance of that point from the camera in real world. Therefore, these depth values can be used to estimate the average size of the human at each distance. In our proposed method, we first divide the depth map into several layers and define rectangular windows with sizes varying for each depth layer based on the depth values. Then, by comparing the number of pixels that fall into the window with a predefined threshold, we decide whether the region contains a pedestrian or not. Our ROI generation method can be divided into three main steps of layering depth map, extracting skeleton for each depth layer and defining an adaptive window for each layer and sliding it on the skeleton of the objects in each layer.

2.1.1. Layering depth map

At first step, a depth image is divided into several sub-images where each sub-image contains objects with the same depth values. To do this, depth histogram is obtained and pixel values that belong to each histogram bin constitutes a sub-image. Fig. 3 (c) shows an example of sub-images extracted from a depth image.

2.1.2. Extracting image skeleton

In exhaustive search based methods, all of the pixels are considered for searching but most of these pixels are not necessary for ROI decision. To reduce the complexity, we consider the pixels on the skeleton of the image only. For each sub-image, skeleton or medial axis is extracted and the search window is slided on these pixels only. Image skeleton, $S(X)$, can be extracted using morphological operations given in Eqns. (1) and (2):

$$S(X) = \bigcup S_n(X) \tag{1}$$

$$S_n(X) = (X \ominus nB) - (X \ominus nB) \circ B, \tag{2}$$

where $\ominus$ and $\circ$ are morphological erosion and opening, respectively, $n = 0, 1, \cdots N$, X is the input image, B is the structuring element and

$$nB = \underbrace{B \oplus \cdots \oplus B}_{n\ times}, \tag{3}$$

where $\oplus$ is the morphological closing. Fig. 3 (d) shows the extracted skeletons for each sub-image.

2.1.3. Adaptive windowing

Since the perceived size of an object is most strongly influenced by the object's distance from the camera, we determine the detection window size for ROI search in each sub-image based on the depth value. Specifically, we first set the detection window size of the first depth layer which is closest to the camera as $height/2 \times width/6$ considering the approximate shape of a human. Then, we adjust the height and width of the detection window in the n-th depth layer using Eqns. (4) and (5).

$$h_n = \frac{D - d_n}{D - d_1} \times h_1, \tag{4}$$

$$w_n = \frac{D - d_n}{D - d_1} \times w_1, \tag{5}$$

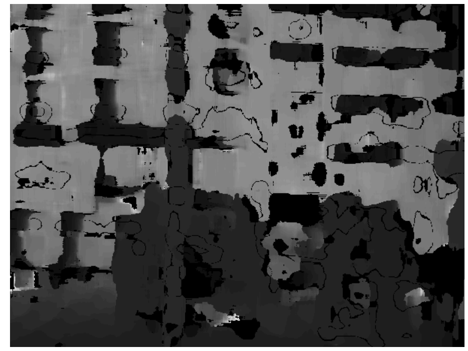
where d_1, h_1 and w_1 are the depth value of the first layer and height and width at that layer, D is the maximum depth range (255 in our experiments), and d_n is the depth value in the n-th depth layer.

Then, we slide the detection window along the skeleton of the objects and calculate the amount of overlap between the detection window and a candidate region. If the amount of overlap is larger than $1/3$, the region is identified as a candidate region of interest. Fig. 4 shows several examples of the detected ROIs.

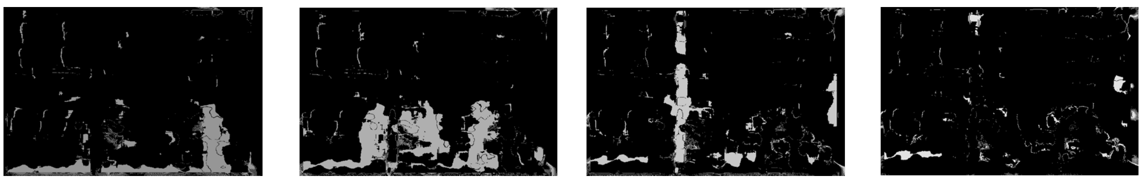
3. EXPERIMENTAL RESULTS

To test the performance of our proposed method, we used a stereo vision based DAS dataset provided by [11]. We used the classification and system packages of this dataset which provide both training and test sets and several video sequences. Their test set contains 250 frames with 640×480 pixels and range of 0 to $50m$ captured by a stereo vision camera from a moving vehicle in different illumination conditions and with different size and number of pedestrians.

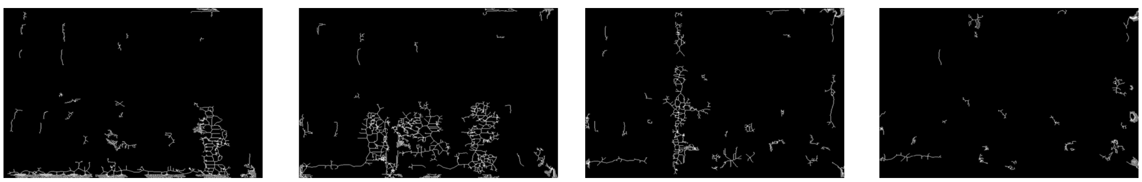
(a)



(b)



(c)



(d)

Fig. 3. Proposed ROI generation: (a) Color image, (b) depth map, (c) depth layers obtained using the depth histogram, and (d) skeleton of the objects in each depth layer.

Their training set contains 7649 negative and 1015 positive samples. Since the depth maps in the dataset have a poor quality and a limited range, we applied image inpainting and linear contrast stretching to improve the quality of the depth maps.

Fig. 4. An example of detected ROIs.

3.1. Object Classification

3.1.1. Histogram of Oriented Gradients

Since ROIs contain both pedestrian and non-pedestrian regions, an object classification step is necessary to classify these regions. Among several silhouette based and appearance based methods, we use Histogram of Oriented Gradients (HOG) [13] for classifying ROIs since it shows the best results in human detection. HOG is a feature descriptor inspired by Lowe's SIFT descriptor [14] and is based on calculating orientation gradients in sub-blocks of a region. In HOG descriptor, the region is divided into smaller cells and orientation histograms are calculated for each cell. The histograms can be created by computing the orientation of gradients to form the histogram bins and its magnitude as a vote for the bin using Eqns. (6) and (7).

$$m(x, y) = \sqrt{L_x^2 + L_y^2} \tag{6}$$

$$\theta(x, y) = \arctan \frac{L_y}{L_x}, \tag{7}$$

where m and θ are magnitude and orientation at each pixel and L_x and L_y are gradients in x and y direction.

These cells are then grouped into larger regions called blocks and histograms of the blocks are normalized to improve illumination invariance. The normalized blocks are then concatenated to make the final feature vector.

3.1.2. Support vector machine

We use a linear SVM to classify the feature vectors. To do this, SVM is first trained using 1000 pedestrian and 1000 non-pedestrian frames. Based on this support vector, ROIs extracted from test frames are classified into pedestrian and non-pedestrian classes.

ROIs extracted from the 250 frames of the test set using our ROI detection algorithm is then classified by this classifier. In the implementation of the ROI detector, the size of the initial bounding box is set to 240×107 and the threshold for the filled area in the bounding box is set to $1/3$. For the HOG descriptor, the size of blocks and cells are set to 3×3 and 9×9, respectively, and 9 bins are used for the histograms.

The ROC curve is plotted to show the final results which is shown in Fig. 5. Comparing the ROC of our proposed method with that of the HOG based pedestrian classification method proposed in [11], we can see that our method outperforms the state-of-the-art method except in a small interval. The major reason for this improvement is because of using depth maps for ROI generation which are invariant to illumination conditions. Our system is also scale invariant because of using adaptive windows for extracting ROIs. We have also performed another experiment by changing the initial size of the adaptive window and the results show that by reducing the window size, the detection rate is decreased.

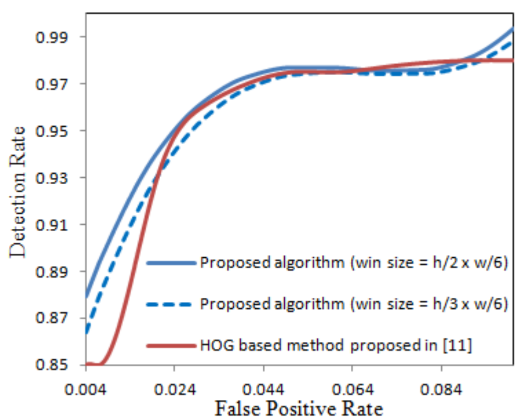


Fig. 5. ROC curve comparison between the proposed method and the HOG based method proposed in [11].

4. CONCLUSION

In this paper, a ROI generation method has been proposed for pedestrian detection in advanced driver assistance systems. The proposed method exploits both color and depth information obtained by a stereo camera system to improve the detection performance under illumination variations and in presence of partial occlusions. A depth-based adaptive windowing method has been proposed to detect pedestrians with different scales. A HOG descriptor and a linear SVM are used to classify the ROIs into pedestrian or non-pedestrian classes. Simulation results show that the proposed ROI generation method improves the performance of pedestrian detection in challenging environments.

5. REFERENCES

[1] D. M. Gavrila, "Pedestrian Detection from a Moving Vehicle," in *European Conference on Computer Vision*, pp. 37–49, 2000.

[2] Y. Xu and D. Xu and S. Lin and T. X. Han and X. Cao and X. Li, "Detection of Sudden Pedestrian Crossings for Driving Assistance Systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 42(3), pp. 729–739, 2012.

[3] P. Raksincharoensak, Y. Sakai, I. Shimizu, M. Nagai, D. Ulbricht, and R. Adomat, "Development of an On-Board Pedestrian Detection System Using Monocular Camera for Driver Assistance Applications," in *FISITA 2012 World Automotive Congress*, pp. 341–350, 2013.

[4] A. Talukder, S. Goldberg, L. Matthies, and A. Ansar, "Real-time detection of moving objects in a dynamic scene from moving robotic vehicles," in *Int. Conf. on Intelligent Robots and Systems*, pp. 1308–1313, Oct. 2003.

[5] B. D. Nam, S. I. Kang, and H. K. Hong, "Pedestrian Detection System based on Stereo Vision for Mobile Robot," in *Frontiers of Computer Vision.*, pp. 1–7, Feb. 2011.

[6] M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. Matthies, "Results from a real-time stereo-based pedestrian detection system on a moving vehicle," in *International Conference on Robotics and Automation (ICRA 2009)*, 2009.

[7] B. Bhowmick, S. Bhadra, and A. Sinharay, "Stereo Vision Based Pedestrians Detection and Distance Measurement for Automotive Application," in *Int. Conf. on Intelligent Systems, Modeling and Simulation*, pp. 25–29, Jan. 2011.

[8] P. Viola, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition, (CVPR 2001)*, vol. 1, pp. 511–518, 2001.

[9] C. G. Keller, M. Enzweiler, and D. M. Gavrila, "A New Benchmark for Stereo-based Pedestrian Detection," in *Intelligent Vehicles Symposium*, pp. 691–696, Jun. 2011.

[10] C. G. Keller, M. Enzweiler, M. Rohrbach, D. F. Llorca, C. Schnorr, and D. M. Gavrila, "The Benefits of Dense Stereo for Pedestrian Detection," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 1096–1106, 2011.

[11] D. Geronimo, A. D. Sappa, D. Ponsa, and A. M. Lopez, "2D-3D Based on-board pedestrian detection system," in *Computer Vision and Image Understanding (Special Issue on Intelligent Vision Systems)*, vol. 114, pp. 583–595, May 2010.

[12] C. G. Keller, D. F. Llorca, and D. M. Gavrila, "Dense Stereo-based ROI Generation for Pedestrian Detection," in *Proceedings of the 31st DAGM Symposium on Pattern Recognition*, pp. 81–90, 2009.

[13] N. Dalal, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, (CVPR 2005)*, vol. 1, pp. 886–893, Jun. 2005.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. Journal of Computer Vision*, vol. 2, pp. 91–110, Winter 2004.