FEATURE-BASED ROI GENERATION FOR STEREO-BASED PEDESTRIAN DETECTION

Maral Mesmakhosroshahi, Maziar Loghman, Joohee Kim

Illinois Institute of Technology
Dept. of Electrical and Computer Engineering, Illinois, USA

ABSTRACT

Region of interest (ROI) generation is an important step in stereo-based pedestrian detection systems. In this paper, we propose an ROI generation method by fusing the color and depth information obtained from a stereo camera mounted on a vehicle. In our proposed method, a feature-based method which uses contour properties of the image is used to find the ROIs. In our feature-based ROI extraction method, we extract four features which are contour density, maximum area, maximum perimeter and matching score. Then we create a feature vector from these features and classify them using SVM. ROIs are then classified into the pedestrian and non-pedestrian classes using Histogram of Oriented Gradients (HOG)/Linear SVM. We have tested our proposed method on the Daimler dataset and experimental results show that our proposed method has a 96.5% accuracy for 1 false positive per frame and outperforms existing monocular and stereo-based methods.

Index Terms— Pedestrian detection, ROI generation, ADAS

1. INTRODUCTION

Pedestrians are the main and most vulnerable participants in any transportation system. Over the past decades, road accidents have become a key cause of fatalities and in most of the cases the human negligence is the main cause of the accidents [1, 2, 3]. In order to improve the road safety, numerous contributions have been made by the scientific community and automobile industry during the past several years which result in several intelligent on-board systems. Such systems which aim to anticipate potential accident situations in order to avoid severe injuries are referred to as advanced driver assistance systems (ADAS).

In recent years, ADAS has attracted an extensive amount of interest from the computer vision community. The main objective in the related research is to improve driving safety by supporting the driver in hazardous driving situations which require immediate reactions. Pedestrian detection system is one of the particular types of ADAS which aims to detect the presence of static and moving pedestrians in a specific region around the moving vehicle. The system will warn the driver regarding the location and state of the pedestrian or activate the braking system in order to avoid a collision. There are several challenges in any pedestrian detection scenarios such as variety of pedestrians appearance and size, illumination variation and real time processing speed.

Various methods have been developed for detecting pedestrians using monocular cameras installed in a vehicle [4]. A review on recently developed methods for different steps of on-board pedestrian detection is presented in [5]. The simplest method to find pedestrians is an exhaustive search [6]. In these methods, fixed or variable sized windows are defined to search for pedestrians at all locations using histogram of oriented gradients and optical flow. However, exhaustive search based methods are computationally expensive and select many irrelevant regions as ROIs. As a result, these approaches do not fulfill real-time requirements and increase the potential number of false positives. In order to reduce the computational complexity and reduce the search space, several methods that combine ROI generation and object classification have been proposed.

In recent years, various attempts have been made to improve the performance of pedestrian detection by taking advantage of stereo vision systems for ROI generation and pedestrian detection [7]. Dense depth maps are used for extracting information about the size, distance and geometric features of the objects and generating the candidate regions. A stereo-based system is presented in [8] that uses the disparity image for ROI generation and pose estimation. In our previous works [9, 10], we used depth layering and skeleton extraction for ROI generation and classified ROIs using the HOG descriptor and linear SVM. In [11], ROIs are generated by scanning the depth image with windows related to the maximum extent of pedestrians. In [12], stixel world is used as an attention stage to improve the recognition performance and reduce the computational complexity of the object recognition systems. A fast pedestrian detection system is presented in [13]. This pedestrian detection system uses the stixel world for ground plane estimation and candidate generation.

In this paper, we propose an ROI generation algorithm for stereo-based pedestrian detection. The focus of the proposed method is to increase the accuracy of pedestrian detection while reducing the search space for defining pedestrian candidates. In the proposed method, four features are calculated by extracting the contour information of the color image. The calculated features are contour density, maximum area, maximum perimeter and matching score. These features are then classified using linear SVM.

The rest of the paper is organized as follows. In Section 2, the proposed ROI generation algorithm is explained in detail. In Section 3, the simulation results are provided and the paper is concluded in Section 4.

2. PROPOSED METHOD

In this section, we describe our proposed ROI generation method. Our proposed method consists of three main steps which are candidate window generation, feature extraction and feature training and classification. Fig. 1 shows the block diagram of the proposed method.

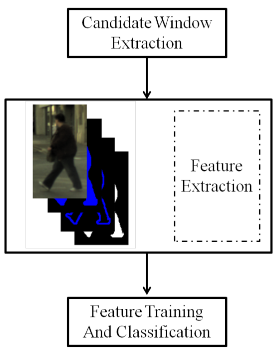

Fig. 1. Block diagram of the proposed ROI generation method.

2.1. Candidate Window Generation

To find ROIs, we first need to extract candidate windows from the image. To do this, we use the boundary of the ground plane extracted using the ground plane estimation method in [14] to locate the candidate windows. Considering the stereo correspondence, we can conclude that the size of the object in pixel is proportional to the depth value. Hence, we can estimate the height and width of the detection window at a boundary point using (1),

$$\begin{bmatrix} h \\ w \end{bmatrix} = \frac{I_{db}}{I_{dmax}} \times \begin{bmatrix} h_1 \\ w_1 \end{bmatrix}, \tag{1}$$

where the initial detection window size is $w_1 \times h_1$, I_{db} is the depth value of the pixel on the boundary of the ground plane and I_{dmax} is the maximum depth value. To ensure that pedestrians with different poses are detected, for each boundary pixel (i, j) we extract three detection windows at $(i, j - \frac{w}{2})$, (i, j), and $(i, j + \frac{w}{2})$.

2.2. Feature Extraction

In this step, our goal is to create a feature vector for each candidate window to be able to differentiate the windows that contain a pedestrian from the non-pedestrian windows. To do this, we introduce four features based on the shape information extracted from the image. To find these features, we first create a binary image by finding the edges in the image and retrieve contours from this binary image. After retrieving the contours, we extract the following features from the contours:

2.2.1. *Contour density*

The first feature we use for shape analysis is the contour density. To find the contour density, we first extract all of the contours from the extracted detection window. We then find the contour density (ρ_B) for the bounding box B using (2),

$$\rho_B = \frac{\sum_{i=1}^{N} g(c_i)}{P_B}, \tag{2}$$

where P_B is the perimeter of a bounding box, N is the total number of contours (c_i) in B and $g(.)$ is a binary function which can be computed by (3).

$$g(c_i) = \begin{cases} 1 & , c_i \in B \\ 0 & , otherwise. \end{cases} \tag{3}$$

Contour density shows whether the window contains an object or if it is a texture less area.

2.2.2. *Maximum area*

The second feature we use is the maximum area. To find this feature, we use the contours extracted from the detection window and find the contour with maximum area using (4),

$$MA_B = \max_{i=1}^{N} \frac{A_{c_i}}{A_B}, \tag{4}$$

where A_B is the area of the bounding box and A_{c_i} is the area of the contour c_i. Maximum area determines how large the object inside the window is. Pedestrians and tall objects are expected to have a high maximum area. Therefore, higher the MA_B, higher is the probability that the window contains a tall object.

2.2.3. *Maximum perimeter*

Maximum perimeter is the next feature used for shape analysis. To find this feature, we first find the contour with the maximum perimeter using (5),

$$MP_B = \max_{i=1}^{N} \frac{P_{c_i}}{P_B}, \tag{5}$$

where P_B is the perimeter of the detection window and P_{c_i} is the perimeter of the contour c_i.

2.2.4. *Matching score*

Matching score is a measure of comparison between two images which shows how similar the two images are. To find the matching score, we first find three orders of central moments from the image using (6),

$$m_{ij} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \overline{x})^i (y - \overline{y})^j f(x, y), \tag{6}$$

where $f(x, y)$ is the intensity value at pixel location (x, y) for an $M \times N$ image f. $\overline{x} = \frac{M_{10}}{M_{00}}$, $\overline{y} = \frac{M_{01}}{M_{00}}$ and $M_{ij} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^i y^j f(x, y)$.

Using the central moments, we then find the seven Hu invariants [15]. These values are proved to be invariant against the image scale, rotation and reflection except the h_7 whose sign is changed by reflection. Hu invariants can be used to find the matching score. To calculate the matching score (MS) between two contours c_i and c_j, (7) can be used.

$$MS(c_i, c_j) = \sum_{k=1,\cdots,7} \left| \frac{1}{m_k^{c_i}} - \frac{1}{m_k^{c_j}} \right|. \tag{7}$$

In (7), we have:

$$m_k^{c_i} = sign(h_k^{c_i}) \cdot log(h_k^{c_i}), \tag{8}$$

where $h_k^{c_i}$ is the Hu moments of the contour c_i.

In our proposed method, we use a binary pedestrian template to find the matching score. The template is resized to the size of the candidate window. Fig. 2 shows the pedestrian template we used in our experiments. We use this template to determine the similarity of each of the ROIs to the pedestrian shape. We do this by finding the matching score between the ROIs and the pedestrian template.

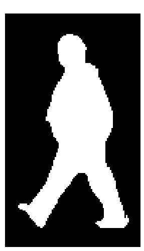

Fig. 2. Pedestrian template used for computing the matching score.

2.3. Feature Training and Classification

Up to here, we have computed four features for each detection window extracted from the frame. Each of these features expresses a characteristic which can differentiate pedestrians from the non-pedestrian windows. Contour density helps us find windows that contain an object. Maximum area and maximum perimeter together can differentiate tall objects from the other objects and matching score shows how the object in the window is similar to a pedestrian.

However, using each of these features can produce so many false positives. The number of detection windows considered as ROI is minimized by creating a feature vector which contains the values of all the features described in Section 2.2. These features need to be classified to find the windows that possibly contain a pedestrian.

Among the existing classification methods, we use Linear SVM as our classifier to find the final ROIs. To do this, we first create two different sets of image samples. The positive set contains images that have a pedestrian. The negative set contains images from different pedestrian and non-pedestrian samples. To find the negative set, we first apply our ground plane estimation method on a set of negative images. We then extract detection windows from these images and find the four features for each window. Among the extracted windows, we keep the ones which have feature values less than a user-defined threshold.

Using the positive and negative set, we train the Linear SVM and then use the trained model to find the ROIs on the test data. To find these ROIs, after classifying the windows, we keep the windows with the SVM score greater than a threshold. We set this threshold to a value that reduces the number of false negatives.

3. EXPERIMENTAL RESULTS

For the performance evaluation of the proposed method, we use the Daimler Stereo Pedestrian Detection Benchmark Dataset [8]. Our system runs on a quad-core $3.07GHz$ Intel Xeon CPU with AMD graphic card. The system components are implemented with C++ using OpenCV 2.4.8 library in Linux environment. For fast stereo correspondence, we use a simple block matching algorithm among several available stereo matching algorithms [16, 17]. The ROIs extracted from the proposed ROI generation algorithm contain both pedestrian and non-pedestrian windows. Therefore, in order to find pedestrians among them, we need to apply an object classification algorithm. From the existing object classification algorithms, we use HOG/Linear SVM [6] which has the best performance in pedestrian classification.

3.1. Simulation Results

To reduce the computational complexity, the input images are first downsampled by two in the horizontal and vertical directions. For depth map clustering, the quantization level is set to 15 and the value and distance thresholds in ground plane estimation are set to 60 and 30, respectively. In order to refine the ground plane, morphological closing and opening are applied to fill the holes. Also, the initial size of the detection window for the depth value of 255 is set to 125×250.

3.1.1. *ROI generation*

In this experiment, the average number of ROIs per frame and the detection rate of the ROI generation algorithm are computed using the unclassified ROIs extracted from the 21790 frames in the test dataset. Table 1 shows the performance of the proposed ROI generation algorithm.

Table 1. Performance evaluation of the proposed ROI generation method for 21790 test frames in the Daimler dataset.

Method	Detection rate (%)	#ROIs
Proposed ROI generation.	98.8	993
Objectness measure [18]	83.68	1000
Edgebox [19]	86.94	1000

Pedestrians missed in the ROI generation step cannot be recovered later. Therefore, the miss rate of this step must be as low as possible. As can be seen in Table 1, the detection rate of our proposed ROI generation is 98.8% which means that the probability of missing pedestrians is very low, while reducing the computational complexity significantly as compared to the exhaustive search.

3.1.2. *Pedestrian detection*

In this experiment, we compared the performance of our proposed method with the existing ROI generation and pedestrian detection methods. Fig. 3 shows the performance comparison between our method and existing methods introduced in [8, 13, 18, 19].

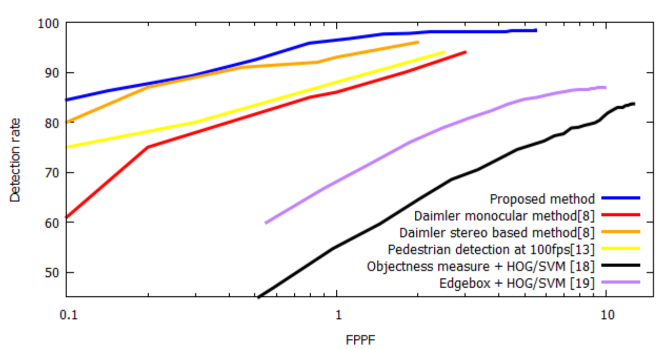

Fig. 3. Comparison of ROC curves between the proposed method and the methods in [8, 13, 18, 19].

Results show that our system outperforms most of the existing pedestrian detection and ROI generation methods and provides competitive results with the Daimler stereo based method. In addition to accuracy, we have calculated the computation time of the system. Fig. 4 shows the detection rate for 1 FPPF and computation time comparison between our method and existing methods introduced in [8, 13, 18, 19]. The proposed method outperformes the existing methods in term of detection rate while maintaining a reasonable computaion time. The reason for better performance of the proposed method is that we estimate the size of the pedestrians and generate more accurate bounding boxes while in the existing methods different predefined aspect ratios are used to generate the bounding boxes which might be irrelevant to size of the pedestrians.

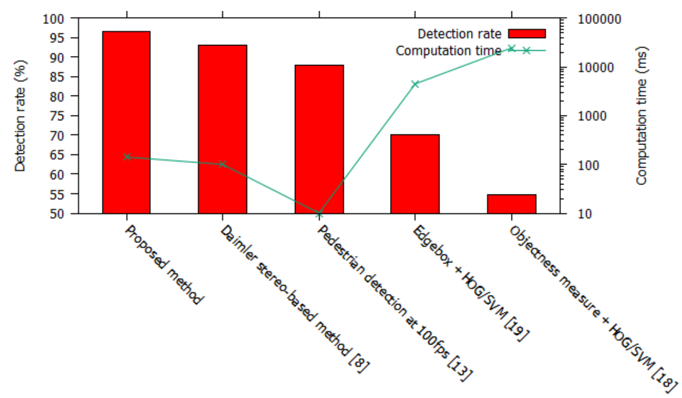

Fig. 4. Comparison of detection rate and computation time between the proposed method and the methods in [8, 13, 18, 19].

4. CONCLUSION

In this paper, an ROI generation method for stereo-based pedestrian detection has been proposed. In the proposed method, a feature-based method has been used to find ROIs from these windows. In this method, contour properties such as contour density, maximum area and perimeter and a matching score has been used as features which have been classified using Linear SVM. ROIs are then classified using HOG/Linear SVM. Simulation results show that our proposed method outperforms existing monocular and stereo-based methods.

Acknowledgment

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).

5. REFERENCES

[1] D. Gavrila, P. Marchal, and M.-M. Meinecke, "SAVE-U, Deliverable 1-A: Vulnerable Road User Scenario Analysis," in *technicalreport, Information Soc. Technology Programme of the EU*, 2003.

[2] W. Jones, "Building Safer Cars," in *IEEE Spectrum*, vol. 39, no. 1, pp. 82–85, Jan. 2002.

[3] United Nations-Economic Commission for Europe, "Statistics of Road Traffic Accidents in Europe and North America," 2005.

[4] P. Dollar, Ch. Wojek, B. Schiele and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, Aug. 2011.

[5] D. Geronimo, A. M. Lopez, A. D. Sappa and T. Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance Systems," In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239–1258, Jul. 2010.

[6] N. Dalal, "Histograms of Oriented Gradients for Human Detection," In *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893, Jun. 2005.

[7] D. F. Llorca, M. A. Sotelo, A. M. Hellin, A. Orellana, M. Gavilan, I. G. Daza, and A. G. Lorente, " Stereo regions-of-interest selection for pedestrian protection: A survey," In *Transportation Research Part C: Emerging Technologies*, vol. 25, pp. 226–237, Dec. 2012.

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

[9] M. Mesmakhosroshahi, J. Kim, Y. Lee and J-B. Kim, "Stereo based Region of Interest Generation for Pedestrian Detection in Driver Assistance Systems," In *Proceedings of the International Conference on Image Processing (ICIP)*, pp. 3386–3389, 2013.

[10] J. Kim and M. Mesmakhosroshahi, "Stereo-based Region of Interest Generation for Real-time Pedestrian Detection," In *Peer-to-Peer Networking and Applications*, pp. 1–8, Sep. 2013.

[11] Ch. G. Keller, Th. Dang, H. Fritz, A. Joos, C. Rabe, and D M. Gavrila, " Active Pedestrian Safety by Automatic Braking and Evasive Steering," In *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1292–1304, Dec. 2011.

[12] M. Enzweiler, M. Hummel, D. Pfeiffer and U. Franke, " Efficient Stixel-based Object Recognition," In *Proceedings of the Intelligent Vehicles Symposium (IV)*, pp. 1066–1071, Jun. 2012.

[13] R. Benenson, M. Mathias, R. Timofte, L. Van Gool, "Pedestrian Detection at 100 Frames per Second," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2903–2910, Jun. 2012.

[14] M. Mesmakhosroshahi, K. H. Chung, Y. Lee and J. Kim, "Depth gradient based region of interest generation for pedestrian detection," In *Proceedings of the International SoC Design Conference*, Nov. 2014.

[15] M. -K. Hu, "Visual pattern recognition by moment invariants," In *IRE Transaction on Information Theory*, vol. 8, no. 2, pp. 179-187, Feb. 1962.

[16] D. Scharstein and R. Szeleiski, ""A Taxonomy and Evaluation of Dense Two Frame Stereo Correspondence Algorithm," In *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, May 2002.

[17] M. Loghman and J. Kim, "SGM-based Dense Disparity Estimation Using Adaptive Census Transform," In *Proceedings of the IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 592–597, Dec. 2013.

[18] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," In *IEEE Transactions Pattern Analysis Machine Intelligence*, vo. 34, no. 11, pp. 189–202, 2012.

[19] C. L. Zitnick and P. Dollár, "Edge Boxes: Locating Object Proposals from Edges," In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014.