

A Novel Method for Day Time Pedestrian Detection

Reena Kumari Behera, Jiji Gangadharan, Krishnan Kutty, Smita Nair, and Vinay Vaidya
KPIT Technologies Ltd.

ABSTRACT

This paper presents a vision based pedestrian detection system. The presented algorithm is a novel method that accurately segments the pedestrian regions in real time. The fact that the pedestrians are always vertically aligned is taken into consideration. As a result, the edge image is scanned from bottom to top and left to right. Both the color and edge data is combined in order to form the segments. The segmentation is highly dependent on the edge map. Even a single pixel dis-connectivity would lead to incorrect segments. To improve this, a novel edge linking method is performed prior to segmentation. The segmentation would consist of foreground and background segments as well. The background clutter is removed based on certain predefined conditions governed by the camera features. A novel edge based head detection method is proposed for increasing the probability of pedestrian detection. The combination of head and leg pattern will determine the presence of pedestrians. The extracted segments are merged to form the complete pedestrian based on the evident leg and head pattern. The algorithm is tested on real time images and provides good detection capability. The accuracy of the above method can be further improved by using a classifier on the segmented region.

CITATION: Behera, R., Gangadharan, J., Kutty, K., Nair, S. et al., "A Novel Method for Day Time Pedestrian Detection," *SAE Int. J. Passeng. Cars – Electron. Electr. Syst.* 8(2):2015, doi:10.4271/2015-01-0319.

INTRODUCTION

With increase in the number of vehicles on the road, the pedestrian safety has become a new challenge area for the Automotive OEMS. The statistics released by the World Health Organization (WHO) [14] is alarming with more than 2, 70,000 pedestrians losing their lives on the road. The pedestrians constitute to about 22% of the road accidents globally. The insight shows a crucial need for pedestrian safety system on vehicle that could warn the drivers well ahead of time. Sensors play an important role when it comes to detecting obstacles in the surrounding. Commonly used sensors on the vehicle are LIDAR, RADAR, ultrasound and camera. As compared to various sensors, vision-based systems are gaining significant importance due to their lower cost and advantages as compared to other sensors.

LITERATURE STUDY

Camera based pedestrian detection is a challenging problem because of various poses and clothing of pedestrians, which needs to be handled under varying illumination and environmental conditions. Generally, there are two main approaches for vision based pedestrian detection, the whole and the part based approach. In the whole body detection, the pedestrian is detected as a whole object where as in part based; the detection process is concentrated on parts like head, torso, arms, legs etc. The general process for detection constitutes of pre-processing, foreground segmentation, object classification and tracking. Pre-processing includes exposure correction, dynamic ranging, noise removal etc. to provide a better input for further

processing. Foreground segmentation extracts possible candidate ROI by eliminating background and sky region. This restricts the search ROI, thereby reducing the processing time and false positives.

Literature study into the existing methods shows that the histogram of oriented gradients (HOG) [1] based search is one of the simplest and most widely used pedestrian detection techniques. However, since the entire image is scanned at various scales, the process is extremely slow. Saliency based method [2] uses 2D features such as gradient, color, intensity, edge etc., to extract object segments. Since the method is highly dependent on the selected features, human detection is not much efficient. Stereo-based foreground segmentation is another way to eliminate background. Numerous methods [3] [4] [5] are available in literature and one of the major assumptions in most of these techniques is that pedestrians possess a vertical structure at a specific depth. Grubb et al. [3] proposed the use of v-disparity representation to find vertical and horizontal planes to extract candidate ROIs. Stereo-based plane fitting is another method used to find different planes. M Soga et al. [4] used disparity map analysis with Pedestrian Size Constraint (PSC) to extract better ROIs. Multimodal stereo methods that make use of different spectrums like visual and thermal infrared [5] can provide better accuracy even in complex scenes with multiple people at different depths and various levels of occlusion.

Most of the algorithms are invariant to illumination changes. Time complexity is more for the stereo-based algorithms and the detection rate is very less in non-textured regions. Motion based segmentation is another method to extract ROIs. Elrein et al. [6] used optical flow and inter frame motion to identify moving obstacles. Both motion

and stereo are utilized in some methods to get the best foreground. Heinrich et al. [7] proposed a method that makes use of motion analysis and depth information to get minor grey value changes thereby enabling early detection of moving obstacles. Leibe et al. [8] presented a structure from motion-based method that is used to extract segments from the ground plane. Object classification aims at classifying pedestrian or non-pedestrian segments in a given candidate ROI. Numerous methods based on number of features with different classifiers are available in literature for whole body and part based detection. Papageorgiou et al. [9] used Haar wavelets to train quadratic SVM with different poses of pedestrians (front, rear etc). Jones et al. [10] proposed Haar like features with Adaboost cascade classifiers for classification. HOG, edgelets, shapelets are other mostly used features with classifiers like Adaboost, SVM, Logitboost etc. In part-based approach, classification of different parts are combined to define the whole pedestrian. Mohan et al. [11] used Haar Wavelets with a quadratic SVM to classify four different parts of the body (head, legs, left arm and right arm). Wu et al. [12] proposed a part based detection system with four body parts (full-body, legs, torso and head-shoulder); the feature used is edgelet with Adaboost.

Tracking is used to predict future pedestrian positions in order to avoid false detections and to reduce the computational time. Kalman filters [13] [14], particle filters [15], stereo-odometry based trackers [16] are some of the commonly used methods for tracking.

OVERVIEW

As compared to the above work, the presented work exploits the fact that pedestrians have strong edge features, especially in the leg and the torso region. An on board forward facing camera captures the scene ahead of the vehicle. The extracted image is first segmented and the background clutter is removed from the scene. The pedestrians are detected from the foreground-segmented regions based on leg and head detection criteria. Figure 1 (a) provides an overview of the described method -

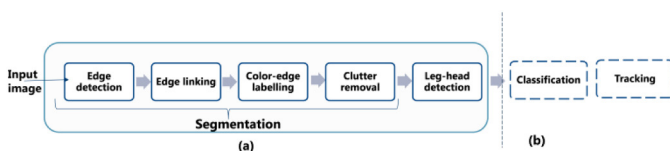


Figure 1. Block diagram of pedestrian detection system, (a) consisting of segmentation and detection stage, considered in the proposed method (b) classification and tracking stage.

Segmentation is a crucial step in any detection and tracking based system. Accurate segmentation of the regions would reduce the overhead on the following detection stage. In the proposed method, the segmentation is done based on the edge map. To further enhance the segmentation stage, disconnected edges are linked via a novel edge-linking step. The segmentation is a two-step process, wherein edge image is scanned using a bottom-up approach and segments are generated based on certain statistical measure. These segmented regions are further grouped based on the spatial connectivity. The extracted segments are further classified into background and foreground clusters based on camera features and the vertical edge

features. The final segmented region is assumed to consist of segments belonging to pedestrians. Segments that match leg features are extracted as probable pedestrian segments. Head detection on the segmented foreground image further adds weightage to the detection stage.

The paper is organized as follows: The section on Segmentation provides details on edge detection, linking and the color-edge labelling stages. The background clutter removal forms a part of the color-edge labelling stage within Segmentation. The Detection stage provides details on the head and leg detection from the extracted segments. The results are presented in each section followed by Summary and Conclusion.

SEGMENTATION

The pedestrians have very strong edge based features and within a region defined by the two edges, the pedestrian clothing color has some spatial relationship. Here we combine both the color and edge data to form the segments for detecting pedestrians. Since edge is crucial information, any breaks in the edge data caused by thresholding can lead to unwanted segments. This is handled by an edge linking technique to fill in the broken gaps. With the edge and color information, the segments are labeled and provided with unique code. The later stage is followed by grouping the segmented regions and removal of background data. The segmentation consists of the following steps - edge detection, edge linking and color-edge based labelling.

Edge Detection

To form probable segments, the color data is considered between two edges in a row. For this reason, the edges detected should be strong, smooth and of one pixel width. A comparison of various edge detection techniques such as Sobel, Prewitt and Canny shows that the Canny detector provides good detection, localization and single response to a particular edge [17]. Two main highlights of the Canny detector implemented for our pedestrian detection is as follows:

The image gradient is computed using the following centered mask in both the x and y directions

Centered Mask:

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$Magnitude(|\nabla F|) = \sqrt{F_x^2 + F_y^2}$$

$$orientation \theta = \tan^{-1} \left(\frac{F_y}{F_x} \right)$$

(1)

Where,

F_x is gradient along x direction,

F_y is gradient along y direction,

θ is the gradient orientation.

A comparison of various kernels in [1] shows that the above 1-D centered mask provides best results for pedestrian images. The gradient computation is followed by non-maximal suppression and hysteresis thresholding stage.

To obtain optimal threshold values for hysteresis thresholding stage, the threshold values are taken as a factor of the distribution of the edge pixels. The upper threshold TH is considered to be 0.2, such that 20% of the total pixels above TH are retained. The lower threshold TL is a factor of the high threshold value. TL is set to 0.9, which is 90% of the high threshold value. For our tests, the ideal range for TH is between (0.15-0.25) and for TL is between (0.8-0.9). As per Canny detection [15], the gradient values greater than TH are retained while lower than TL are removed. The gradient values between TH and TL are retained based on the connectivity with the high threshold pixel.

Even after getting optimal threshold values, thresholding introduces gaps in the edge map. Since the segmentation is based on the edge map, even a single pixel gap could lead to unwanted segments. This can be handled using edge linking methods.

Edge Linking

Practical issues with any edge detection method are that it causes missing links in the edge image. This leads to issues while identifying object boundaries. The local edge-linking method extends edges by finding the most compatible edge in its neighborhood.

Let (x, y) be the center pixel value. The main criteria for establishing similarity are:

- Magnitude of the gradient vector $|\nabla F(x, y)|$
- Direction of the gradient vector $\theta(x, y)$

Edge pixel (x_0, y_0) , in the 3×3 neighborhood is similar to pixel (x, y) if,

$$|\nabla F(x, y) - \nabla F(x_0, y_0)| < M \quad (2)$$

$$|\theta(x, y) - \theta(x_0, y_0)| < \alpha \quad (3)$$

Where,

M is the threshold for similarity in gradient magnitude.

α is the threshold for similarity in gradient direction.

The value for M is set to 25. The threshold value for α is maintained at 20. If both the above conditions are satisfied, the pixel (x_0, y_0) is set and is linked to pixel (x, y) .

Since the edge pixels are scanned from left to right and bottom to top, to ensure connectivity in the forward direction, following steps are adopted in the local edge linking method:

- For a center pixel (x, y) , a 3×3 neighborhood is monitored. A pixel is identified as disconnected, if all the pixels in the four forward direction are zero. The pixels satisfying the disconnected patterns in Figure 2. are considered for edge linking.

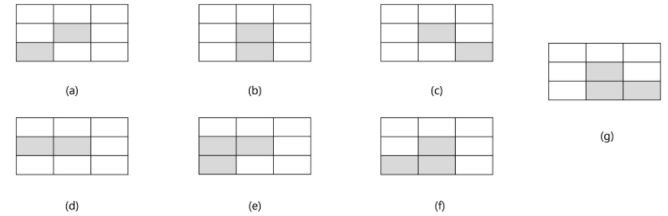


Figure 2. (a-f) Different disconnected patterns considered for edge linking. The colored positions in a 3×3 window shows existence of the edge pixels.

- As discussed above, a similarity check is performed between the center pixel and the four forward positions in 3×3 window. Accordingly, the corresponding neighboring pixel, with shortest distance is set.
- If the above condition is not satisfied, instead of ignoring the pixel, choose a neighboring pixel, based on the orientation value of the center pixel (x, y) .
- The above step of force fitting a neighboring pixel is performed only if the magnitude of the corresponding neighboring pixel is greater than a threshold. In our case, the threshold is maintained at 10.
- As required, the above force fitting is performed for 3 times continuously. Even after three attempts, if the similarity criteria is not satisfied, then pixel linking is discontinued and the force fitted links are removed.
- For every disconnected pixel, the edges are iteratively extended based on the steps mentioned earlier. A maximum threshold of 5 is maintained for every iteration. Also the disconnected link is visited during the scanning process. This ensures that long disconnected edges can be filled (Figure 3).

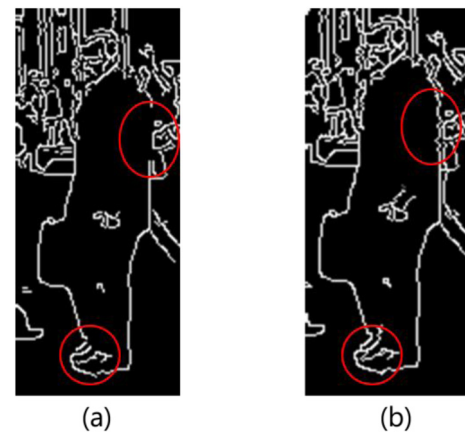


Figure 3. (a) Shows the circled regions (red) that undergo edge linking process, (b) shows the output of edge linking stage