Pedestrian Detection Prevent Vehicle Accidents

1. Dr.I.Satyanarayana,Principal,Sri Indu Institute of Engineering&Technology (SIIET), Sheriguda, Ibrahimpatnam,Hydarabad

Emailid:umadonthagani@gmail.com

2.K.Ravali, Assistant Professor,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad,
3.D.Uma ,Assistant Professor,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad,
4.B.Rajeswari,5.Sushmita,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
5. Bhargavi,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
6.G.keerthi,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

Abstract—Pedestrian detection and driver assistance applications draws an important role on safety measure such as speed control, crash control, sensing crash and occupant location detection. The pedestrian accidents occurs due to the vulnerable traffic users such as humans, stranded or moving vehicle or other obstacles. To avoid the pedestrian accident, this paper proposes a model that can accomplish pedestrian detection automatically using Histogram of Gradient (HOG) and You Only Look Once (YOLO) algorithm. The experiments are carried out on Forward Looking Infrared Radar (FLIR) starter thermal dataset consisting of 5000 images. The HOG algorithm is implemented on these thermal image samples and is classified using Support Vector Machine (SVM) classifier. The accuracy of YOLO is calculated using intersection over union method between the ground truth and the predicted bounding box. To further improve the safety of the user an alarm is designed to alert the user on sight of pedestrians during night.

Index Terms—Histogram of gradient, Pedestrian detection, Support vector machine.

I. Introduction

Pedestrian identification has a vital task in the computer field over the last decade which increases the use of various applications such as video surveillance, robotics, and intelligent transportation systems [1]. The National Highway Traffic State Administration (NHSTA) has surveyed that pedestrian is died for every 2 hours and wounded every 8 minutes in accidents [2,3]. The main factors that affect the human detection at night are low visibility, pedestrian clothing, weather conditions and perception reaction time. The rotation of cameras and elimination of background make the situation much harder. This paper discusses the issue of finding human detection in the pedestrian region in detail. M. bertozzi et al. [2] proposed a tuned subsystem for a tetra-vision based pedestrian identification with the help of both visible and far infrared light cameras to detect the pedestrian using with classifier. In first stage, pedestrian area of the image is detected. A refinement filtering process is implemented to neglect the incorrectly detected images with no pedestrians or more than one pedestrian. A subsystem is used to find the pedestrian shape, in order to give a vote on it. This test has been executed with sample 4400 infrared images: 2200 pedestrians, and 2200 non-pedestrians. Pawel et al. [3] developed a method for pedestrian detection in low night perception infrared images. The drawback of these method are more sensitive and they offer a long distance detection but at higher cost and lower resolution. The object classification procedures are performed with low and very low resolution images. The object tests are performed with three datasets – Night-time Pedestrian Dataset (NTPD), Laboratorio de Sistemas Inteligentes / Intelligent System Lab Far Infrared Pedestrian Dataset (LSI FIR) and Ohio State University (OSU) thermal pedestrian dataset. The low resolution image sensors can be used to decrease the cost of night vision systems without loss in detection quality, it always requires more processing time. Guangyan et al. [4] developed the optimized Histogram of Gradients (HOG) to obtain a precise person detection method. The datasets used are INRIA and MIT Pedestrian for training and testing the system. The pedestrian detection method uses linear classifier which is effective only on visible spectrum images.

Joseph et al. [5] developed a method for object identification using neural network which identifies bounding boxes and class probabilities. The YOLO method can processes 155 frames per second and it can predict nearby objects. Also this model struggles with small objects that appear in same class. Compared to other state-of-the-art detection [6], You Only Look Once (YOLO) suffers in localization miscalculation but it achieves good results in detecting false positives objects with background. Further [7], YOLO9000 is a real-time object identification method that can recognize over 9000 object. Here, the experimental results show that Convolution Neural Network (CNN) has achieved 84.7% classification accuracy using inception V3 model on Flickr Material Database (FMD) [8,9].

The main objective of the proposed system is to detect the pedestrians at night using low resolution far-infrared images to produce quality image detection of pedestrians during dark. The proposed pedestrian detection system can extract features using HOG and classify them using Support Vector Machine (SVM) and also to use YOLO to detect the pedestrians with high accuracy.

The rest of the paper is organized as follows: Section II presents the overview and the proposed pedestrian detection system, Section III presents the experimental results and conclusion in Section IV.

II. Methodology

The proposed work is designed to avoid the pedestrian detection at night automatically using HOG/SVM [10] and

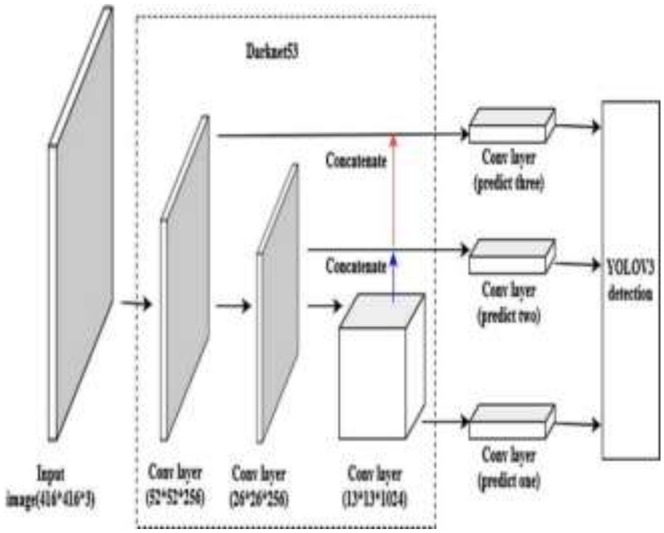

Fig. 1. YOLOv3 network structure, where the blue and red lines represent two-fold up-sampling.

YOLO [11] algorithm. It is based on following steps: image pre-processing, feature extraction and classification.

Initially, the input image of size of 416 pixels × 416 pixels are scaled usingYOLOv3. The feature extraction is carried out with a network called DarkNet-53. After the extraction, original image is changed into a feature map (i.e. size = 13 × 13). Three feature maps are generated by merging two feature maps (i.e. 26 × 26 and 52 × 52 sizes). The object detection is based on three scales, in which the feature map is changed into two neighboring scales using double up-sampling. Feature map in each cell finds three bounding boxes using three anchor boxes. Finally, the accurate bounding box is selected. Next, HOG scans the input image using detection window. The detection window is scanned the image at all positions and all scales and gradients and orientations are calculated. Each window is separated into cells and each cell finds HOG orientations for all the pixels in the cell. The histogram normalization is computed by the local histogram energy which improves brightness to standardize all cells. Finally, HOG features are gathered over the detection window and fed as an input into linear SVM for object/non-object identification is explained in Figures 1 and 2.

A. Image Pre-processing

Image pre-processing [12] is the initial step for any convolution neural networks. Images are pre-processed in batches and are explained as follows: First, the acquired images from FLIR dataset are resized in the YOLOv3 model of size of 416 pixels × 416 pixels and it is converted to array format for processing. Next, expand the dimension using batch dimension and the batches of images are sent through the network. Finally, mean subtraction is implemented by subtracting the mean RGB pixel intensity from the FLIR dataset.

B. Feature Extraction

Feature extraction in YOLO v3 is done using DarkNet-53 Network and also the features are extracted using HOG for

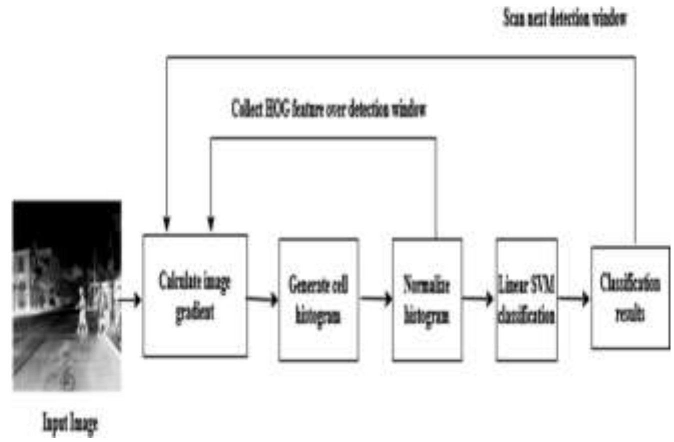

Fig. 2. HOG and SVM Classifier.

SVM Classification [13]. The pre-processed input images from the training set is fed as input to DarkNet-53 network pre-trained on ImageNet dataset with image size of 416 × 416 × 3. The Darknet-53 network in YOLO V3 is used for pedestrian detection and is calculated on a 416×416 patch of an image. Here, the processed raw input image (416, 416 and 3) is fed as input to CNN gives (13, 13, 5 and 85) dimensional output. After smoothening the last two dimensions gives the output (13, 13 and 425). Each cell in a 13 × 13 grid over the input image gives 425 numbers (i.e. 425 = 5×85 has 5 boxes predictions in each cell, corresponding to 5 anchor boxes, 85 = 5 + 80 where 5 is the boundary box (pc, bx, by, bh, bw), and 80 is used to identify the number of classes). The selection of feature extraction boxes is based on two parameters. Firstly, Score-threshold detect the box with a class has a score less than the threshold. Secondly, Non-max suppression which finds the intersection over union by neglecting overlapping boxes. Finally, the YOLO's output is extracted.

Also, the feature extraction for pedestrian detection using HOG is as follows: The HOG feature descriptor used for pedestrian identification on patch of an image (64 × 128) with aspect ratio of 1:2 and resized to 64 × 128. For each pixel within the cell 1-D Sobel vertical and horizontal operators are computed using Eq. (1) and (2).

$$G_X(y, x) = Y(y, x+1) - Y(y, x-1) \tag{1}$$

$$G_y(y, x) = Y(y+1, x) - Y(y-1, x) \tag{2}$$

Where, $Y(y, x)$ is the pixel intensity at coordinates x and y. $G_X(y, x)$ is the horizontal gradient, and $G_y(y, x)$ is the vertical gradient. After the gradient calculation, gradient magnitude and gradient angle for each of 64 pixels are determined. As a next step, for each cell HOG is created. The image is divided into 8×8 cells and histograms are formed with 9 × 1 matrixes for each cell. The histogram has a vector of 9 bins with angles 0, 20, 40, 60, 160. Consider, $Q = 9$ bins with unsigned orientation changes angles below 0° are increased to 180°.

$$G(y, x) = \sqrt{G_X(y, x)^2 + G_Y(y, x)^2} \tag{3}$$

$$\vartheta(y, x) = \arctan \frac{G_y(y, x)}{G_x(y, x)} \tag{4}$$

The various input images may have various contrast normalization can be used to reduce by normalizing the gradients using 16×16 blocks. Each 8×8 cell has a 9×1 matrix for a histogram. Therefore, four 9×1 matrices or a single 36×1 matrix is obtained. In each block, normalization is performed on the histogram vector $|v|_k$ with its k- norm for $k = 1,2$ and s is the small constant. The obtained value is further normalized into a vector of size 36×1. To find the output feature vector for the full image patch, the 36×1 vectors are merged into one big vector. The 105 blocks have a vector of 36×1 as features (sums of features $105 \times 36 \times 1 = 3780$).

$$L_1 - normnorm = \frac{v}{\sqrt{|v|_2^2 + s^2}} \quad (5)$$

$$L_2 - normnorm = \frac{v}{|v|_2} \quad (6)$$

$$L_1 sqrt - normnorm = \sqrt{\frac{v}{|v|_1 + s}} \quad (7)$$

Now, the features and the associated class labels are created for testing and training. The training data consists of the extracted features and the test data consists of the corresponding labels of the extracted features [14]. The extracted features are non-linear in nature since they were extracted from CNN are passed into linear models for classification. After implementing the classifiers, we have obtained an accuracy of 65% for HOG trained network and 73% for YOLO network.

III. Experimental Analysis

The experimental result for pedestrian object detection is executed using YOLO and HOG algorithm. The detailed explanation is given as follows,

A. Dataset

Firstly, the images collected from the FLIR dataset are pre-processed and fed into the YOLO V3 algorithm. Subsequently real time IR video captured using FLIR camera is broken down

Fig. 3. Images from the FLIR dataset.

Fig. 4. Real-time IR video file converted into frames.

into frames and fed as an input. The Figure 3 contains the images collected from FLIR dataset and the Figure 4 contains the real time images.

IV. Results and Discussion

The YOLO V3 algorithm processes both the images collected from the FLIR dataset and real time IR video captured using FLIR camera. It is converted into frames and draws boundary boxes across all the objects and images along with the name of the type of object found. The images will be left unchanged in case of no objects found. The Figure 5 contains the output of the images collected from FLIR dataset and the Figure 6 contains the output of real time images. The sizes of the images before and after transform along with the label of the objects are also found.

Figures 5 and 6 clearly shows that YOLO algorithm is used for finding the object (i.e. person) in the pedestrian region. In the experiment, the images collected from the FLIR dataset are processed and fed as an input into the HOG feature

Fig. 5. Objects detected in FLIR dataset input.

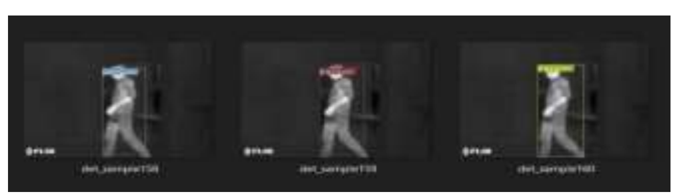

Fig. 6. Objects detected in real-time video file converted into frames.

TABLE I
TEST ACCURACY MEASUREMENT.

Algorithm	Accuracy
HOG+SVM	60%
HOG+YOLOV3	73%

descriptor. The features of the images are extracted and stored in a .csv file. The features extracted are labeled either 0 or 1. The label 0 denotes the images with no objects and the label 1 denotes the images with objects. The features extracted from the images using HOG which is stored in the .csv file is fed into the SVM classifier which gives the output in the form of SVM predictions along with the accuracy.

V. CONCLUSION

The pedestrian prediction analysis is made for both the algorithms namely YOLO V3 and HOG respectively. The pedestrian prediction experiment was carried out for the whole dataset. The experimental results show that YOLO V3 performs better than HOG with a SVM classifier with an test accuracy of 73% compared to the test accuracy of 60% with HOG and SVM algorithm is shown in Table I. Moreover, YOLO V3 also provides detection at real time. In future, these experiments will be carried out with updated versions of YOLOV3 and other detection algorithms.

REFERENCES

[1] H. Song, H. Liang, H. Li, *et al.* "Vision-based vehicle detection and counting system using deep learning in highway scenes," *European Transport Research Review*, vol. 11, pp. 51, 2019.

[2] M. Bertozzi, A. Broggi, M. Del Rose, M. Felisa, A. Rakotomamonjy, and F. Suard, "A pedestrian detector using histograms of oriented gradients and a support vector machine classifier," *Intelligent Transportation Conference*, USA, 2007.

[3] P. Pawlowski, K. Piniarski, and A. D'lcbrowski, "Pedestrian detection in low resolution night vision images," Poznan University of Technology, 2015.

[4] G. Zhang, F. Gao, C. Liu, W. Liu, and H. Yuan, "A pedestrian detection method based on SVM classifier and optimized Histograms of Oriented Gradients feature," *Sixth International Conference on Natural Computation*, 2010.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[6] J. Redmon and A. Farhadi, YOLO9000: Better, faster, stronger, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[7] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, "Pedestrian detection at 100 frames per second," ESAT-PSI-VISICS/IBBT, Belgium.

[8] S. Zubaidah, B. Rajemi, R. Hamid, F. Naim, and N. W. Arshad, "Pedestrian detection for automotive night vision using thermal camera," *7th International Conference on Computing and Convergence Technology (ICCCT)*, 2012.

[9] H. Sun, C. Wang, and B. Wang, "Night vision pedestrian detection using a forward-looking infrared camera," *International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, 2011.

[10] Y.-C. Lin, Y.-M. Chan, L.-C. Chuang, L.-C. Fu, S.-S. Huang, P.-Y. Hsiao, and M.-F. Luo, "Near-infrared based nighttime pedestrian detection by combining multiple features," *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011.

[11] W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian detection based on YOLO network model," *IEEE International Conference on Mechatronics and Automation (ICMA)*, 2018.

[12] T. Y. Han and;. C. Song, "Night vision pedestrian detection based on adaptive preprocessing using near infrared camera," *IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 2016.

[13] J. Redmon and A. Farhadi "YOLOv3: An incremental improvement," *Computer Vision and Pattern Recoginition*, 2018.

[14] K. Mizuno, Y. Terachi, K. Takagi, S. Izumi, H. Kawaguchi, and M. Yoshimoto, "Architectural study of HOG feature extraction processor for real-time object detection," *IEEE Workshop on Signal Processing Systems*, 2012.