

An Optimization Algorithm for Pedestrian Detection ahead of Vehicle

A.Y. Guo

School of Mechanical Engineer and Automation
Shanghai University
Shanghai, China

Department of mechanical and electrical engineering
Shanxi Light Industry Technology School
Taiyuan, China

M.H. Xu, F. Ran

School of Mechanical Engineer and Automation
Shanghai University
Shanghai, China

C.Q. Du

School of Automobile Engineer
Wuhan University of Technology
Wuhan, China

Abstracts—The pedestrian detection ahead of vehicle is a very important technology in automobile driving system. Histogram of Oriented Gradients (HOG) plus Support Vector Machine (SVM) is the general algorithm in the pedestrian detection. But, most of algorithms are obtained only by the Matlab or Open CV, rarely focused on the transportation to the hardware and self-developed C language library. Therefore, this paper proposed a new algorithm of cascaded Local Binary Pattern (LBP) and HOG detecting pedestrian based on the linear SVM. This algorithm can be transported to the ARM, DSP and some other embedded systems. The experiment shows the result very well.

Keywords—pedestrian detection HOG LBP ROI (region of interest)

I. INTRODUCTION

In a few years, the automatic driving system appears. The pedestrian detection ahead of vehicle based on the computer vision is one of the most important functions in automatic driving system. The information of pedestrians are obtained by the camera, which is putted in the moving vehicle. The automobile driving system can extract the pedestrian using the specific algorithm. The pedestrian detection has the great application prospect in auxiliary and pedestrian safety. So, the pedestrian detection become the research hot spot.

Significant achievements have been made in pedestrian detection. The early researches are focused on the image segmentation, edge extraction and motion detection in statistic image. With the development of technology, the mainstream view is that algorithms of the pedestrian detection are divided into two kinds. One kind is that the module building based on the background. The other kind is that the algorithm based on the statistic learning. Last years, the modeling building and pedestrian detection are fixed into the system of pedestrian detection. The Histogram of Oriented Gradients (HOG) plus Support Vector Machine (SVM) are the widely used to detect pedestrian [1][2]. However, the fact is that the research just remains on the theory. Most of algorithms are obtained only by the Matlab or Open CV, rarely focused on the transportation to

the hardware and self-developed C language library. Practical vehicular processor are usually DSP, FPGA, ARM or combinations of them. Because most of researchers finds that it is very difficult to run in the embedded system purely by Matlab or Open CV. Therefore, this paper proposed a new algorithm of cascaded Local Binary Pattern (LBP) and HOG detecting pedestrian based on the linear SVM. Considering the transplantation, the algorithm doesn't rely on any computer vision library. In order to decrease the robustness and real time of pedestrian detection very well, this paper does some improvements on the segment Region Of Interest (ROI) and objection recognition.

In this paper, we mainly focus the problems that don't rely on any library vision and give the optimized algorithm to utilize the pedestrian detection in hardware. Firstly, we give the algorithm to realize in the form of implementation diagram and the math which used. Secondly, the experiments are putted into the system to test the performance and give the conclusion.

II. SYSTEM DESIGN OF PEDESTRIAN DETECTION

In order to protect the real-time, accuracy and transportation of pedestrian detection, we use cascaded LBP and HOG characteristics based on linear SVM in this algorithm. These modules are realized using the C language, and don't rely on any vision library. The process are listed in Figure 1. As the figure 1 showed, the monocular images firstly are inputted into the module of zoom and previous warning. After the first module, the images are changed into the grayed ROIs which contains the different zoom layers. HOG and LBP features in the unit of block are stored into the buffer pool. With the moving of sliding windows, all the LBP features in effective detection area can combine the 1888 dimensions of LBP feature vector. These LBP features support the vector operator with linear SVM which are well trained using LBP features. The system will give a determine whether are the effective areas or not, according to the operations value. The next module is the objection detection module. In this module, all HOG features in objection recognition area can be scanned to

combine the 3780 dimensions of HOG features in a sliding window. These HOG features can support the vector operator with the linear trained SVM using the HOG feature. Then the system also give a decision whether it is a pedestrian area or not. The fourth module is the objection fusion and previous warning. The system fusions the inner layer and inter layer in pedestrians area, while calculating the dynamic overlap area. When the weighing of pedestrians area in fusion inner layer match the threshold, then the system will give the warning. In the last module, the pedestrians area which extracted from the image can be show in the other image.

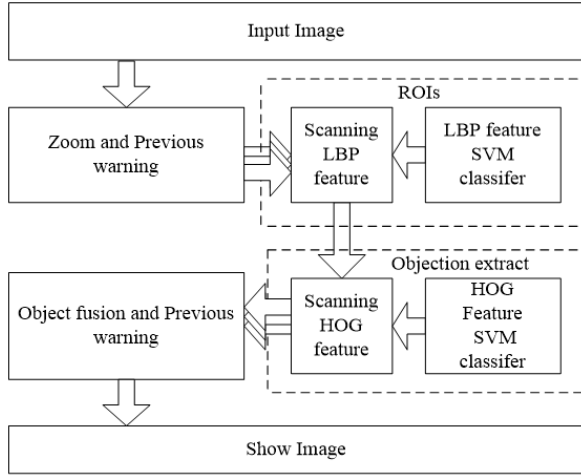


FIGURE 1. THE SYSTEM DESIGN OF PEDESTRIAN DETECTION

III. THE THEORY STUDY OF PEDESTRIAN DETECTION

A. The Module of Zoom and Previous Warning

The image of training pedestrians and non pedestrians are 64*128 pixes. Based on this, we need the same size of detection-window in order to train the SVM classifier very well. In general way, the distance between pedestrians and camera changed. Then, the image of pedestrian sometimes is large, and sometimes is small. In order to decrease the influence of size, we zoom the size by different layer. After zooming, the pedestrian in image and the size of training sample can match.

This paper we increase accuracy rate using the method of reducing the original image gradually. The experiment make a conclusion: if the zoom factor is too small, the imaging quality will fall sharply and disadvantages to detection; if the zoom factor is too large, the amount of scaling layers are too large, increase the calculation, and be disadvantages to reduce the detection time. After the comparative analysis, the three factor of image can be completely the detection-time and accuracy with the zoom factor 1,0.75 and 0.5 .

B. The Extraction of ROI

LBP is an effective texture description operator. It can extract and metric the texture information in local region in gray-level image. The advantages of LBP are invariant features and relative small amount of calculation. In this paper, we use the normal LBP.LBP features of sliding window in the effective region waiting for detection can be used to segment the ROI. This paper uses the LBP to describe the LBP features

in every zoom layer of ROIs. Assuming that the sequence of binary is clockwise, the extraction of LBP is showed in Function (1)[3].

$$LBP_{P_c} = \begin{cases} \sum_{i=0}^7 2^i * z(p_i - p_c), u \leq 2 \\ 85, u > 2 \end{cases} \quad (1)$$

The LBP features are stored in buffer pool. This buffer pool can decrease detection time when calculate the LBP feature in the unit of block temporally in the sliding windows. This module don't need to calculate any LBP block features, just combine the LBP features to 1888 dimensions vector in sliding window. The size of every LBP block is 16*16.The block of pixels can be combined with 59 dimension features. Although the LBP scanning 64*128 window include the 4*8 blocks, the minimum of sliding step is 8 pixels. So the 64*128 sliding window has the 7*15 LBP block at most. The step of sliding window don't fixed to the 8 pixels and is variant according to the result according to the previous module. If the 1888 dimension LBP features in sliding window operator the trained LBP feature linear SVM support vector, the result shows the non ROI, then the step of sliding window is 16 pixels. If the result shows the ROI, then the step of sliding window is 8 pixels. The variant of step follow the regulate of "fine-coarse-fine". [4]

C. Object Recognition

As the Fig. 2 showed, the sliding window of ROI are inputted into the module of object recognition. In the module of object recognition, the ROI which satisfied the LBP features can be deal with. The size of block is 16*16 pixels in 64*128 sliding window. Every block has the 36 dimension vectors. There are overlap between the blocks and the step of block is 8 pixels. Then, there are 3780 dimension vectors in a sliding window. Putting the trained HOG feature linear SVM classifier into the memory and according to the $y = w^T x + b$, the system can determine the numeric of y. If the y is bigger than the threshold, this sliding window is the pedestrian region. If the y is smaller than the threshold, this sliding window is the non-pedestrian region. In the module of object recognition, the most important module is the scanning HOG features. The following scrap will give some explains. The advantage of HOG characteristics is that the hog characteristics has no sensitive to the change of sunshine or position. The hog characteristics can be used to describe the information of the human body contour very well. So, it is usually used in computer vision field of pedestrian detection.

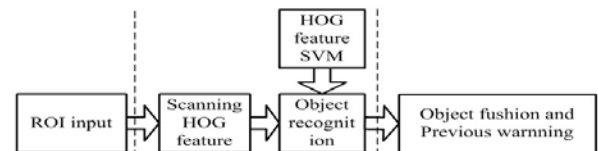


FIGURE 2. THE MODULE OF OBJECTION RECOGNITION

The HOG features are extracted with the step as figure 2 showed. The first step is the graying of sling windows. The next step is that calculating every pixels gradients. The third

step is cubic interpolation of gradient magnitude with direction and position. The ultimate, the normalized gradient magnitude of blocks, then a block generated 3780 dimensions vector.

The Dalal's experiment showed that it is best to use the undirected bin and 9 bins. This system also divided $[0,180]$ into 9 parts in $16*16$ block. There is a center pixel in every unit of a block.[2]We shadow the gradient into the 9 bins as the previous listed, normalized, then generate 36 dimension vectors in every block. According to the prior experiment, we use the angle range of $[0,180]$ with 9 undirected bins. The size of a block is $16*16$ pixels. There exists a central pixel in every $8*8$ pixels. We must put all the gradient magnitudes into the 2 adjacent bins which are the this pixels adjacent bins, four shadow center-pixels. Then we normalize the bins. By these steps, we generated 36 dimension vectors. At the meanwhile, we use the $7*7$ convolutional operator. This operators use the distance between spatial location and all gradient magnitudes of pixels to convolute.

D. The Object Fusion and Previous Warning

As the Fig.3 showed, the results of object recognition are inputted in different zoom layers. The overlap pedestrian region need to fusion. So, this paper use the dynamic overlap area to fusion and rebuilt the new pedestrian region. The strategy in different layer are serial, the order fusion of pedestrian region is that the inner layer first and the inter layer second. When the weigh after fusion of pedestrian region is bigger than the threshold, then give the driver a previous warning. If the fusion of all zoom pedestrian region, all the pedestrians will be located from the original image.[5]



FIGURE III. OBJECTION FUSION AND PREVIOUS WARNING

The key of dynamic fusion is that using the dynamic overlap fusion area as a threshold and let this threshold as a lower limit. If the overlap area is bigger than the threshold, the system generates a new pedestrian region. New pedestrian region weigh is the sum of previous two pedestrian region weigh. We set the weigh of pedestrian region to 0. Assuming that the overlap area of two pedestrian region is 70 percents the weight of two pedestrian region are 1, then 70 percents is bigger than 50 percents. So these two pedestrian region will combine a new pedestrian region and generate the starting coordinate, highth and width of the new pedestrian region. If the one pedestrian region weigh is 1 and the other pedestrian region weigh is 10, so the 70 percents is small than 10/11. So, the system will don't do any treatment to these two pedestrian region.[6]

IV. EXPERIMENT OF RESULTS

This paper uses the positive and negative samples in INRIA pedestrian library as training and testing library. In order to protect the size of all samples are $64*128$, all images must be intercepted from the large image. The system extract positive and negative samples feature vector of the training and testing set. Then, put the positive and negative sample feature vectors of training set to linear SVM. After train, the system will generate the LBP feature linear SVM classifier and HOG

feature linear SVM classifier. The third step is that we read the data of positive and negative samples using MATLAB. Compared the predicted result and the real result of annotation, the system will give the relative evaluation index. Through predict testing samples, two kinds of classifier shows the false detection and miss detection. HOG feature classifier can detect most of pedestrian. However, if the pedestrian's foot apart too large, the body's bending too large or the body are shielded, the classifier will give the false detection. If there exist trees, building or some other things in the vertical of image, the classifier will give the miss detection.

This paper use the LBP59 and optimize HOG algorithm to analysis 1000 pieces of positive and negative samples in testing set respectively. The result are compare with the fact positive and negative samples. As the table 1 showed, the predicted result in pedestrian will decrease greatly, but the miss fatal has no obviously change. However, the predicted results just are gained by the single window. If a detection object is chosen by more than one sliding windows, the predicted results will rely on the movement of sliding window and fusion of pedestrian region.

Dollár P proposed the method of multi-scale pedestrian detection in real time[7]. He use the 6fps to detect the pedestrians in the image of $640*40$ pixels. Zhu Q proposed the method of cascade hog characteristics using the size variation block, means the width and highth and images are 1:1,1:2 and 2:1 respectively. About the image of $320*240$ pixels, Zhu Q must will reach the speed of 5-30 frames[8]. Normally, the detection time of $640*480$ pixels are four times compared $320*204$ pixels. But in this paper, we just use 10-25 fps to detect pedestrians. This method works very well. In figure 4, it shows the cascaded LBP and HOG feature classifier, it shows that there decrease the miss detection of LBP feature and HOG feature and improve the accuracy of detection.

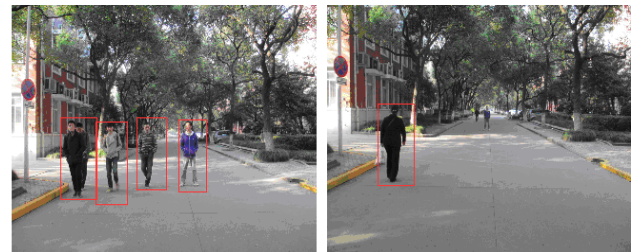


FIGURE IV. THE EFFECTIVE OF PEDESTRIAN DETECTION OF CASCADDED LBP AND HOG

V. CONCLUSION

This paper combines the typical pedestrian algorithm and optimizes the algorithm, then proposed a new pedestrian detection based on linear SVM cascaded LBP and HOG. This algorithm shows the high accuracy, fast speed, good portability, and so on, so it can be used in automatic driving system.

ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China(Grant:61376028),the 2011 Annual Shanghai absorption of imported technology and innovation

plan(11XI-15) and the Key Laboratory of Advanced Display and System Application(Shanghai University)(P200803).

REFERENCES

- [1] Dalal N, Triggs B(2005) Histograms of oriented gradients for human detection. Proc IEEE Conf Comput Vis Pattern Anal Mach Intell 34(4):743-761.
- [2] Gavrilu D, Giebel J, Munder S. Vision-based pedestrian detection: The protector system [C]. Intelligent Vehicles Symposium, 2004 IEEE, 2004, pp. 13-18.
- [3] Broggi A, Bertozzi M, Fascioli A, Sechi M. Shape-based pedestrian detection [C]. Intelligent Vehicles Symposium, 2000 IV 2000 Proceedings of the IEEE, 2000, pp. 215-220.
- [4] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002, 24(7): 971-987.
- [5] Van Gestel T, De Brabanter J, De Moor B, et al. Least squares support vector machines[M]. Singapore: World Scientific, 2002.
- [6] Chapelle O, Haffner P, Vapnik V N. Support vector machines for histogram-based image classification[J]. Neural Networks, IEEE Transactions on, 1999, 10(5): 1055-1064.
- [7] Dollár P, Belongie S, Perona P. The Fastest Pedestrian Detector in the West[C]//BMVC. 2010, 2(3): 7
- [8] Zhu Q, Yeh M C, Cheng K T, et al. Fast human detection using a cascade of histograms of oriented gradients[C]//Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, 2: 1491-1498.