Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 1, Number 1; October, 2014 pp. 27-29
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/acsit.html

Real Time Human Detection using HOG Feature as Human Descriptor and Variable size Sliding Window

Smita Kakade[1], Dr. N.J. Uke[2], Niraja Jain[3]

[1]*Asst. Professor, P.D.E.A's C.O.E., Manjari, Pune*
[2]*Professor, SCOE Vadgaon (Bk., Pune Manjari)*
[3]*Asso. Professor, P.D.E.A's C.O.E., Pune*

Abstract: In this paper we propose a human detection system based on Histogram of Oriented Gradients (HOG) features. The human descriptors for training images are computed with the help of histogram of square blocks of fixed size. For testing in video sliding window technique issued assuming minimum and maximum human height. For window with size greater than training window we increase block size to get feature vector of same size. Using Support Vector Machine as a classifier we build fast human classifier with an excellent detection rate.

Keywords: SVM Classifier, HOG Features

1. INTRODUCTION

Detecting human objects in images and videos is one of the important challenges in computer vision. This is due to factors such as the large variation of appearance and pose that human forms can take. The fast detection of humans in videos recorded by a stationary camera is an essential step in many applications related to surveillance and human-computer interaction domains. Thus, human detection is the first step of the full process of these applications.

In human detection framework a feature pattern learnt by a classifiers exhaustively searched in the full image. Detecting human bodies based on appearance is much more difficult than detecting other rigid objects as cars or faces. Human bodies are non-rigid, and highly articulated. This implies that we have to deal with a high range of different poses and postures. Additionally, in human detection it is not possible to take advantage of specific textures and color information due to the variability of worn cloths.

This paper is organized as follows: section 2 explains previous work on human detection, section 3 gives an overview of the human detection framework that we built and section 4 shows the results we obtained in experimental tests. Finally section 5 discusses the conclusions.

2. PREVIOUS WORK

Human Object detection systems fall into two major categories: component-based methods and single detection window analysis.

The component-based methods detects object parts separately and checks them if they are in a geometrical natural configuration. These systems uses hierarchical detection framework. The body parts are detected in order of their importance, if one of the basic part is not detected the other parts are not searched. They are slow because they need to detect more than one component, so they are slower than single detection window based methods. But they do not handle multi-view and multi-pose cases.

The single detection window method is based on sequentially applying a classifier at all the possible sub windows in a given image. Its most important feature is his speed, while its drawback is a limited partial occlusion handling. In our system we choose to use this type of approach because of his speed.

The human detection method we use is belongs to the single detection window method based on applying a human detector for all possible subwindows in a given image [5, 6]. For instance, in [6], a SVM classifier was learned using Haar wavelets as descriptors. In [8], an efficient detector applicable to videos was built using a cascade of Adaboost classifiers relying also on Haar wavelet descriptors but extracted from spatio-temporal difference. [1] proposed a very good detector that relied on a linear SVM classifier applied to densely sampled histograms of orientation gradient (HOG). It was extended in [2] to videos using histograms of differential optical flow features in addition to HOG. Zhu at al. [4] enhanced Dalal and Triggs[1] results using integral of histograms for a fast HOG computation and Adaboost for feature selection.In the present paper, our goal is to detect humans in videos captured from stationary cameras

3. HUMAN DETECTION SYSTEM

Human Detection can be defined as classification of object as human or nonhuman. From detection image all possible subwindows are obtained assuming minimum and maximum human height. Some of the subwindows are filtered based on percentage of foreground pixels. Then we need to construct a good classifier that can differentiate between humans and not humans in a short time and with low false positive rate.

The main things that are important for human classification are: Image features extraction and learning classifier. Features extraction includes extracting the most relevant information from the available images. We want to find out an optimal image pixel representation that can underline differences between human and not human images. The choice of the feature is done throughout an analysis of the semantic significance of the feature.

We choose to use HOGfeatures because

(i) They provide a good representation of human contour,

(ii) Are invariant to illumination changes and small image movements and

(iii) Can be computed in a constant time.

HOG features are calculated by taking orientation histograms of edge intensity in a local region. The combined features are classified by using linear SVM. They showed that the grids of HOG descriptors significantly outer-performed existing feature sets for human detection.

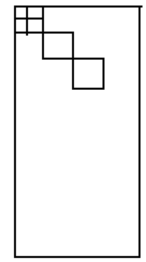
HOG feature have more discriminative power than Haar features and covariance matrix. Alternatively recent works showed that HOG (Histogram of Gradients) features can give excellent results in human detection and also that can be calculated in a constant time using the integral histogram. We use the same method of integral histogram for HOG extraction with fixed cell size for images used for learning.

Input image is converted to grayscale. We get the 2 gradient images for the x and y directions using a CVsobeloperator. We assume bin size 20 degrees and gradient (180/20 = 9), and require 9 images one for each bin. These bin images are used to calculate the integral histogram.

In bin images the magnitude and orientation of the gradient at each pixel is calculated using the sobel images in x and y direction .For every pixel in a row gradient orientation and magnitude are calculated and corresponding values set for the bin images. The bin image is selected according to the gradient values. Integral images for each of the bin images are calculated.

For training classifier we used images with size 64x128. The HOG features are computed by dividing the window into overlapping blocks. HOG vectors for each block concatenated to obtain the HOG feature vector for the window.

Block is divided into 2x2 cells each cell is of size 8x8 pixels is considered and the block is 16x16 pixels. Window of size 64x128 pixels is divided into 7x15 overlapping blocks.



HOG features for the block calculated by concatenating the vectors for each cell and then normalizing over the concatenated vector to obtain the hog features for a block.

The HOG features for training or detection window are computed by concatenating the vectors to each block.

Classifier Learning

Support vector machine (SVM) is a computer algorithm that learns by example to assign labels to objects. We trained a linear support vector machine for object classification. We used 924 positive images and 950 negative images with size 64×128 for training classifier. We first obtained Integral image for source image. From integral image HOG feature vector obtained of size:

(((window.width -cell.width * block_width)/ cell.width) + 1) *(((window.height - cell.height * block_height)/ cell.height) + 1) * 36;

Where cell.width= cell.height=8 and

block_width=block_height=2 cells

The length of the feature vector for a cell is 9(since no. of bins is 9). SVM is trained with two classes 1 for nonhuman and 2 for human object.

4. EXPERIMENTS

We trained and tested detector with MIT pedistrians database . We used 924 positive images of full body human images and 500 hundred negative images that is nonhuman images. The detector is well performing with detection rate above 80%.

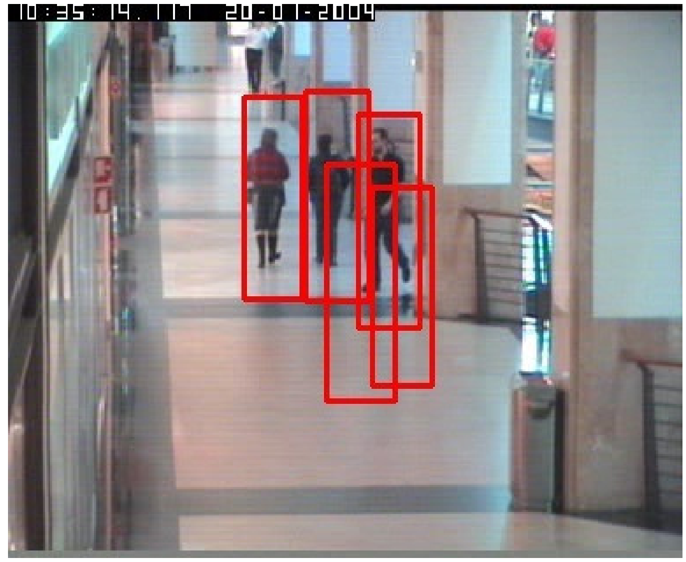
5. CONCLUSION

In this paper we presented Human Detection System with HOG features as human descriptor. We defined a new way of computing HOG features based on variable size block for testing window. The result described in previous section shows quality of approach.

But still improvements are required. We need to use database which has more human positions and situations. Also we need to shift from SVM classifier to cascade of classifier for more accuracy.

6. ACKNOWLEDGEMENT

This work is supported by grant from ISRO –UoP cell in respect of DRDO project. We hereby acknowledges the support received from ISRO-UoP cell.



Experimental Result: positive and negative results

REFERENCES

[1] N. Dalal and B. Triggs. Histograms of oriented gradients forhuman detection. In Computer Vision and Pattern Recognition,2005. CVPR 2005. IEEE Computer Society Conferenceon, volume 1, pages 886–893, June 2005.

[2] N. Dalal, B. Triggs, and C. Schmid. Human detection usingoriented histograms of flow and appearance. In Europe Conf.Comp. Vision (ECCV), volume II, pages 428–441, 2006.

[3] F. Porikli. Integral histogram: A fast way to extracthistograms in cartesian spaces. In CVPR,pages 829–836, San Diego, CA, USA, 2005.

[4] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan.Fast human detection using a cascade ofhistograms of oriented gradients. In CVPR '06:Proceedings of the 2006 IEEE Computer SocietyConference on Computer Vision and PatternRecognition, pages 1491–1498, 2006.

[5] D. Gavrila and V. Philomin. Real-time object detection for"smart" vehicles. In IEEE CVPR, pages 87–93, 1999. [6] P. Papageorgiou and T. Poggio.A trainable system for objectdetection. Int. J. of Computer Vision, 38(1):15–33, 2000.[8]P. Viola, M.J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. Int. Journal of Comp.Vision, 63(2):153–161, 2005.