ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 3, May 2013

Improving Histograms of Oriented Gradients for Pedestrian Detection

Ritesh Singh, T.D. Biradar, Dr. Manali Godse
M.E. Student (EXTC), Asst. Prof (EXTC), HOD (Biomedical)

***Abstract*— In this paper, we proposed a pedestrian detection system based on Hog Transform Using Open CV for smart vehicles. Two SVM classifiers on Histogram of Oriented Gradient (HOG) features are used to precisely locate pedestrians on the ROI. Experiments report over 30 time's higher speed than the state-of-the-art method and a comparable detection rate.**

***Index Terms*— Pedestrian Detections; HOG Features; Smart Vehicles.**

I. INTRODUCTION

The detection of humans in real time and especially in traffic scenarios is an important problem for artificial vision and pattern recognition. A robust solution to this problem would have various applications to autonomous driving systems, video surveillance, and image retrieval. In general, the goal of pedestrian detection is to determine the presence of humans in images and videos and return information about their position. The problem of detecting pedestrians has a high degree of complexity because of the large intra class variability, as pedestrians are highly deformable objects whose appearance depends on numerous factors:

- variability of appearance due to the size, color and texture of the clothes, or due to the accessories (umbrellas, bags etc) that pedestrians may carry;
- irregularity of shape: pedestrians may have different heights, weights
- variability of the environment in which they appear (usually pedestrians exist in a cluttered background in complex scenarios whose look is influenced by illumination or by weather conditions)
- Variability of the actions they may perform and positions they may have (run, walk, stand, shake hands etc).

The inter and intra-class variability of pedestrians makes the ask of a uniform representation extremely difficult. That is the width and height of the 2D image corresponding to perfectly framed pedestrians varies a lot. That is why the number of visual features that can be extracted for a pedestrian is not constant. Also, we may choose from a variety of features and as the number of features increases so does the complexity of the recognition algorithms. In this paper we perform a study of two types of representations used for pedestrian detection: Representation based on primitive features: we have extracted relevant visual features in the field of pedestrian detection, namely Haar and Histogram of oriented Gradient (HoG). The novelty of the paper resides in the application of the Haar and Histogram of Oriented Gradient features to the particular task of pedestrian detection in real time video.

II. OVERVIEW OF THE METHOD

METHODOLOGY

The method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. In practice this is implemented by dividing the image window into small spatial regions (.cells.), for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. The combined histogram entries form the representation. For better invariance to illumination, shadowing, etc., it is also useful to contrast-normalize the local responses before using them. This can be done by accumulating a measure of local histogram .energy. Over somewhat larger spatial regions (.blocks.) and using the results to normalize all of the cells in the block. Gradient Computation: Detector performance is sensitive to the way in which gradients are computed, but the simplest scheme turns out to be the best. We tested gradients computed using Gaussian smoothing followed by one of several discrete derivative masks. Several smoothing scales were tested including sigma=0 (none). Masks tested included various 1-D point derivatives (uncentred [..1; 1], centred [..1; 0;

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 3, May 2013

1] and cubic corrected [1;..8; 0; 8;..1]) as well as 33 Sobel masks and 22 diagonal ones .. 0 1..1 0 ; .. ..1 0 0 1 (the most compact centred 2-D derivative masks). Simple 1-D [..1; 0; 1] masks at sigma=0 work best. Using larger masks always seems to decrease performance, and smoothing damages it significantly: for Gaussian derivatives, moving from sigma=0 to sigma=2 reduces the recall rate from 89% to 80% at 10..4 FPPW. At =0, cubic corrected 1-D width 5 lters are about 1% worse than [..1; 0; 1] at 10..4 FPPW, while the 22 diagonal masks are 1.5% worse. Using uncentred [..1; 1] derivative masks also decreases performance (by 1.5% at 10..4 FPPW), presumably because orientation estimation suffers as a result of the x and y lters being based at different centers. For color images, we calculate separate gradients for each color channel, and take the one with the largest norm as the pixel's gradient vector.

III. RELATED WORK

The two-dimensional approaches to pedestrian detection scan the entire image space (process that can be very slow and prone to false detections). The development of 2D pattern analysis techniques for recognizing pedestrians has shown a strong progress. These pattern analysis techniques include Methods to detect the pedestrian shape or walking motion. Shape techniques rely on detecting spatial human features, and are the most commonly used methods. Haar wavelet transform (essentially multiple scaled edge detection) is used by [4], [5] to extract a pedestrian shape representation. A Support Vector Machine is then trained to learn a model of pedestrians in a front/rear pose, producing strong results. [6] Introduced gradient orientation histograms for pedestrian detection, and developed a recognition system using Support Vector Machines. In [7] the classification is performed based on the vertical symmetry that the human figure exhibits. [8] have developed very efficient cascade classifier to recognize image patterns. Originally created for face detection and then applied to pedestrian detection for surveillance cameras [9], these classifiers are trained by an exhaustive selection of the best weak classifiers, then combining these weak classifiers they form a strong classifier, with impressive results. Motion techniques are used to detect human walking patterns in image sequences. [10] use a model of the human gait in various poses to detect temporal patterns which resemble walking pedestrians. However this method will not detect stationary pedestrians nor any pedestrian not moving across the camera's field of view. The bag-of-words model has been used intensively in object recognition but few algorithms are implemented for the particular task of pedestrian detection in intensity images. For example [11] use the model for representing high-level concepts in images; the high-level concepts correspond to a vocabulary used for Content Based Image Retrieval. The BoW model is used by [12] to predict the presence of an object within an image and it helps to accurately segment instances of object classes in images without any human interaction.

IV. EXPLAINING HOG

DESCRIPTOR BLOCKS

In order to account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cells together into larger, spatially connected blocks. The HOG descriptor is then the vector of the components of the normalized cell histograms from all of the block regions. These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor. Two main block geometries exist: rectangular R-HOG blocks and circular C-HOG blocks. R-HOG blocks are generally square grids, represented by three parameters: the number of cells per block, the number of pixels per cell, and the number of channels per cell histogram. In the Dalal and Triggs human detection experiment, the optimal parameters were found to be 3x3 cell blocks of 6x6 pixel cells with 9 histogram channels. Moreover, they found that some minor improvement in performance could be gained by applying a Gaussian spatial window within each block before tabulating histogram votes in order to weight pixels around the edge of the blocks less. The R-HOG blocks appear quite similar to the scale-invariant feature transform descriptors; however, despite their similar formation, R-HOG blocks are computed in dense grids at some single scale without orientation alignment, whereas SIFT descriptors are computed at sparse, scale-invariant key image points and are rotated to align orientation. In addition, the R-HOG blocks are used in conjunction to encode spatial form information, while SIFT descriptors are used singly. C-HOG blocks can be found in two variants: those with a single, central cell and those with an angularly divided central cell. In addition, these C-HOG blocks can be described with four parameters: the number of angular and radial bins, the radius of the center bin, and the expansion factor for the radius of additional radial bins. Dalal and Triggs found that the two main variants provided equal performance, and that two radial bins with four angular bins, a center radius of 4 pixels, and an expansion factor of 2 provided the best performance in their experimentation. Also, Gaussian weighting provided no benefit when used in conjunction with the C-HOG blocks.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 3, May 2013

C-HOG blocks appear similar to Shape Contexts, but differ strongly in that C-HOG blocks contain cells with several orientation channels, while Shape Contexts only make use of a single edge presence count in their formulation. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

V. INTEGRAL HISTOGRAM OF ORIENTED GRADIENTS

The "Integral Image" [13] allows very fast evaluation of Harr-wavelet type features, known as rectangular filters. This led to a real-time face detection system that was later extended to a human detection system [14], using rectangular filters both in space and time. Recently, Porikli [9] suggested the "Integral Histogram" to efficiently compute histograms over arbitrary rectangular image regions. Inspired by their work, we exploit a fast way of calculating the HoG feature. First, we discretize each pixel's orientation (including its magnitude) into 9 histogram bins. We compute and store an integral image for each bin of the HoG (resulting in 9 images in our case) and use them to compute efficiently the HoG for any rectangular image region. This requires 4 × 9 image access operations.

Fig (a)

Fig (b)

This approach, while fast to compute, differs from the method suggested by Dalal ˜ Triggs and in fact is inferior to it because of the following two reasons. First, Dalal & Triggs use a Gaussian mask and tri-linear interpolation in constructing the HoG for each block. We cannot use these steps because they don't fit well into our integral image ap proach. Second, Dalal & Triggs use an L2 normalization step for each block. Again, we replace the L2 normalization with L1 normalization which is faster to compute using the integral image. The use of HoG features in the Dalal & Triggs approach was restricted to a single scale (105 blocks of size 16 × 16 pixels). Moreover, they report that using blocks and cells at multiple scales improves results somewhat while the computational cost greatly increases.We circumvent this problem by using feature selection. Specifically, for a 64 × 128 detection window we consider all blocks whose size ranges from 12×12 to 64×128. Moreover, we choose a small step-size, whiccan be any of *{*4, 6, 8*}*pixels depending on the block size, to obtain a dense grid of overlapping blocks. In total, 5031 blocks are defined in a 64 × 128 detection window, each of which contains a 36-D histogram vector of concatenating the 9 orientation bins in 2×2 sub-regions. The advantages of using a set of variable size blocks are twofold. First, towards a specific object category, the useful patterns tend to spread over different scales. The original 105 fixed-size blocks only encode very limited information. Second, some of the blocks in this large set of 5031 blocks might correspond to a semantic part in people, say human leg. A small number of fixed-size blocks is less likely to establish such mappings. Another way to view our approach is as an implicit way of doing parts-based detection using a single window approach. The most informative parts, i.e. the blocks, are automatically selected using the AdaBoost algorithm. The HoG features we use are robust to small and local changes, while the variable size blocks can capture the "global picture".

VI. TRAINING THE CASCADE

We construct rejection cascade similar to the one proposed in [13], with the following modifications. Each feature in our scheme corresponds to the 36*D* vector used to describe a block. The weak classifiers we use are the separating hyper plane computed using a linear SVM. Finally, because evaluating each of the 5, 301 possible blocks in each stage is very time consuming, we adopt a sampling method suggested by Scholkopf & Smola [12](pp. 180).They show that one can find, with a high probability, the maximum of *m* random variables, in a small number of trials. More specifically, in order to obtain an estimate that is with probability 0.95 among the best 0.05 of all estimates, a random sub-sample of size $log0.05/log0.95 \approx 59$ will guarantee nearly as good performance as if we considered all the random variables. In practice we sample 250 blocks, at random, in each round. In each level

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 3, May 2013

of the cascade we keep adding weak classifiers until the predefined quality requirements are met. In our case we require the minimum detection rate to be 0.9975 and the maximum false positive to be 0.7 in each stage.

ALGORITHM

Algorithm 1 Training the cascade:
Input: F*target*: target overall false positive rate
f*max*: maximum acceptable false positive

rate per cascade level
d*min*: minimum acceptable detection
per cascade level
Pos: set of positive samples
Neg: set of negative samples

initialize: i = 0, D*i* = 1.0, F*i* = 1.0
loop *Fi* > *Ftarget*
i = *i* + 1
fi = 1.0
loop*fi* > *fmax*
1) train 250 (%5 at random) linear SVMs using Pos and Neg samples
2) add the best SVM into the strong classifier, update the weight in AdaBoost manner
3) evaluate Pos and Neg by current strong classifier
4) decrease threshold until d*min* holds
5) compute *fi* under this threshold
loop end

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 3, May 2013

$Fi+1 = Fi \times fi$
$Di+1 = Di \times dmin$
Empty set Neg
if $Fi > Ftarget$ then evaluate the current cascaded
detector on the negative,i.e. non-human,images and add misclassified samples into set Neg.
loop end

Output: A i-levels cascade
each level has a boosted classifier of SVMs
Final training accuracy: *Fi* and *Di*

VII. CONCLUSION

The paper has presented a study and experiments on the pedestrian detection task for features extracted from two-dimensional images We have approached two directions: (a)detection based on Haar and HoG features and (b)detection implementer d on top of the codebook representation of Haar and HoG features. The codebook representation is a novel approach in pedestrian detection and we have shown that it has better performance than the usual representation. As future work we propose the extension of the method to the application on images that have a larger dimension than the training models, the introduction of other features (for example SIFT and SURF) and even the application of more complex classification algorithms.The submitting author is responsible for obtaining agreement of all coauthors and any consent required from sponsors before submitting a paper. It is the obligation of the authors to cite relevant prior work. Authors of rejected papers may revise and resubmit them to the journal again.

REFERENCES

[1] D. Lowe. Distinctive image features from scale-invariant key points. International Journal of Computer Vision (IJCV), 60(2):91–110, 2004.

[2] K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. European Conference on Computer Vision (ECCV), 2004.

[3] Papa Georgiou and T. Poggio. A trainable system for object detection. International Journal of Computer Vision (IJCV), 38(1):15–33, 2000.

[4] F. Porikli. Integral histogram: A fast way to extract histograms in Cartesian spaces. Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[5] J. M. S. Belongie and J. Puzicha. Shape matching object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 24(24):509–522, 2002.

[6] H. Schneider man. Feature-centric evaluation for efficient cascaded object detection. Conference on Computer Vision and Pattern Recognition (CVPR), 2004.

[7] Scholkopf and A. Smola. Learning with Kernels Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge, MA, 2002.

[8] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. Conference on Computer Vision and Pattern Recognition (CVPR), 2001.

[9] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance.