INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Human Body Detection in Static Images Using HOG & Piecewise Linear SVM

Kishor B. Bhangale
Department of Electronics & Telecommunication
R.M.D Sinhagad School of Engineering, Pune, India
R. U. Shekokar
Professor, Department of Electronics & Telecommunication
R.M.D Sinhagad School of Engineering, Pune, India

Abstract:
Human detection is a challenging task in many fields because it is difficult to detect humans due to their variable appearance and posture. Detecting humans accurately is the first fundamental step for many computer vision applications such as video surveillance, smart vehicles, intersection traffic analysis and so on. This paper consists of efficient human detection in static images using Histogram of Oriented Gradients (HOG) for local feature extraction and linear piecewise support vector machine (PL-SVM) classifiers. Histogram of oriented gradient (HOG) gives an accurate description of the contour of human body. HOG features are calculated by taking orientation of histogram of edge intensity in a local region. PL-SVM is nonlinear classifier that can discriminate multiview and multiposture human bodies from the images in high dimensional feature space. Each PL-SVM model form the subspace, corresponding to the cluster of special view or posture of human. This paper consists of comparison of PL-SVM and several recent SVM methods in terms of cross validation accuracy.

Keywords*: Human detection, histogram of oriented gradients, classification, support vector machine*

1. Introduction

The detection of humans in images and videos especially is an important problem for computer vision and pattern recognition. A robust solution to this problem would have various applications to autonomous driving systems, video surveillance, image retrieval, robotics, and entertainment. In general, the goal of pedestrian detection is to determine the presence of humans in images and videos and return information about their position. Human detection is a challenging task in many fields because as humans are highly deformable objects whose appearance depends on numerous factors:

- variability of appearance of human 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).

In existing human detection methods, feature representation and classifier design are two main problems being investigated. Visual feature descriptors have been proposed for human detection including Haar-like features , HOG, v-HOG, Gabor filter based cortex features , covariance features , Local Binary Pattern (LBP) [19] , HOG-LBP [20], Edgelet [21], Shapelet [22], Local Receptive Field (LRF) [23], Multi-Scale Orientation (MSO) [24], Adaptive Local Contour [25], Granularity-tunable Gradients Partition (GGP) descriptors [26], pose-invariant descriptors [27], Practical Swarm Optimization[16] .

Recently, histogram of oriented gradients (HOG) and region covariance features are preferred for pedestrian detection. It has been shown that they outperform those previous approaches. HOG is a gray level image feature formed by a set of normalized gradient histograms.

Shape and appearance of object can be well defined by the distribution of local intensity gradients or intensity gradients. HOG features are calculated by taking orientation histogram of edge intensity in a local region or block [1]. A HOG feature vector represents the local shape of object, giving edge information at plural cell. For the flatter regions like ground or wall of a building, the histogram of oriented gradients has the flatter distribution. In the border between object and background, one of the elements in the histogram has larger value and it indicates the direction of edge. The total number of HOG features are more and redundant. Hog features can be applied effectively for the classification of object having specific shape or appearance such as face, humans, bicycle, motor car etc, because they are based on the information on the edge.

Linear SVM is the most popular classifier with several reported landmark works for human detection. The reasons we selection of SVM classifiers is that, it is easy to train and, unlike neural networks, the global optimum is guaranteed. The extracted features on labeled samples are usually fed into a classifier for training. However, when we need to detect multi-view and multi-posture humans simultaneously in a video system, the performance of a linear SVM often drops significantly. It is observed in experiments that humans of continuous view and posture variations form a manifold, which is difficult to be linearly classified from the negatives. An algorithm that requires multi-view and multi-posture humans to be correctly classified by a linear SVM in the training process often leads to over-fitting. Some non-linear classification methods such Kernel SVMs are options to handle this problem, but they are generally much more computationally expensive than linear methods [2]. The PL-SVM used the piecewise discriminative function to construct the non-linear classification boundary that can discriminate the multiple positive subclasses from the negative class.

PL-SVM is group of several linear SVM.

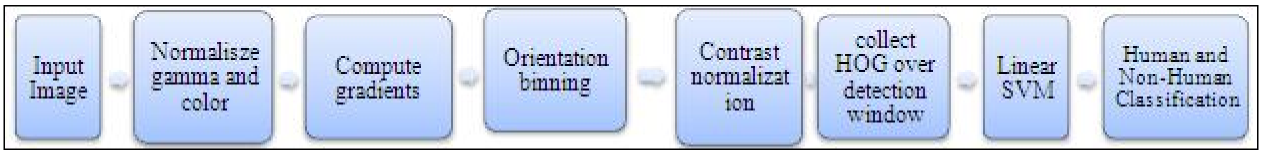

Figure 1: An overview of our feature extraction and object detection chain. The detector window is tiled with a grid of overlapping blocks in which Histogram of Oriented Gradient feature vectors are extracted. The combined vectors are fed to a linear SVM for object/non-object classification. The detection window is scanned across the image at all positions and scales, and conventional non-maximum suppression is run on the output pyramid to detect object instances, but this paper concentrates on the feature extraction process.

2. Overview of Method

Navneet Dalal and Bill Triggs algorithm on Histogram of Oriented Gradients (HoG) is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid [1]. 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 area called as 'cell', for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell [1]. 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 the larger spatial regions called as 'block'. We will refer to the normalized descriptor blocks as *Histogram of Oriented Gradient (HOG)* descriptors [1].

3. Human Detection

The detection of human body based on HOG includes the following six steps: gamma correction and normalization in RGB space, gradient calculation, statistical analysis of gradients of a cell, normalization of block, generation of vector, and classification based on SVM.

3.1. Gamma and Color Normalization

We use exponential gamma correction function to remove the effect of ambient disturbance gradients of a cell, normalization of block, generation of vector, and classification based on SVM.

3.2. Gradient Computation

For gradient computation, first the gray scale image is filtered to obtain x and y derivatives of pixels using *conv2 (image,filter,'same')* method with those kernels:

$$\mathbf{I_x = [\,-1 \quad 0 \quad 1\,]}$$

and

$$\mathbf{I_y = [\,-1 \quad 0 \quad 1\,]^T}$$

After calculating x,y derivatives (I_x and I_y), the magnitude and orientation of the gradient is also computed:

$$\textbf{Magnitude} = |\mathbf{G}| = (\mathbf{I_x}^2 + \mathbf{I_y}^2)^{\wedge 0.5}$$

$$\mathbf{\Theta = \arctan\,(I_y / I_x)}$$

One thing to note is that, at orientation calculation ***rad2deg(atan2(val))*** method is used, which returns values between [-180°,180°]. Since unsigned orientations are desired for this implementation, the values which are less than 0° is summed up with 180°.

Figure 2: (a) Original image containing human in upright position (b) Horizontal gradient of the original image (c) Vertical gradients of the original image.

3.3. Orientation Binning

The next step is the fundamental nonlinearity of the descriptor. Each pixel calculates a weighted vote for an edge orientation histogram channel based on the orientation of the gradient element centered on it, and the votes are accumulated into orientation bins over local spatial regions that we call *cells*. Cells can be either rectangular or radial (log-polar sectors). The orientation bins are evenly spaced over $0\text{-}180^0$ ("unsigned" gradient) or $0\text{-}360^0$ ("signed" gradient) [1].

To reduce aliasing, votes are interpolated bilinearly between the neighboring bin centers in both orientation and position. The vote is a function of the gradient magnitude at the pixel, either the magnitude itself, its square, its square root, or a clipped form of the magnitude representing soft presence/ absence of an edge at the pixel.

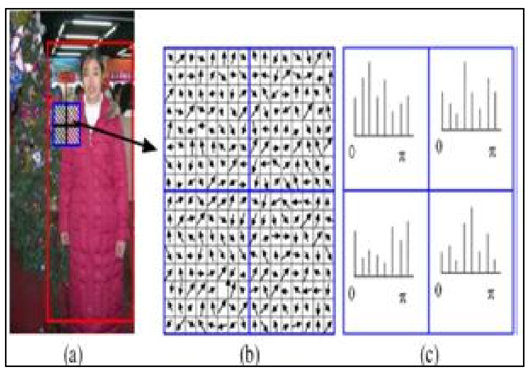

Figure 3: Histogram of orientation gradients. (a) 64 × 128 detection window (the biggest rectangle) in an image. (b) 16 × 16 blocks consists of four cells. (c) Histograms of orientation gradients corresponding to the four cells.

3.4. 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 [5]. 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.

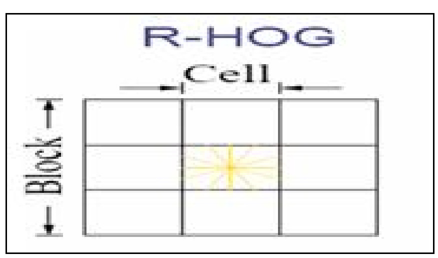

Figure 4: Rectangular HOG

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 [6]. In this implementation we have used 2x2 cell blocks of 8x8 pixel cells with 9 bin histogram channels.

3.5. Block Normalization

For better invariance to illumination and noise, a normalization step is usually used after calculating the histogram vectors. Four different normalization schemes have been proposed: L2-norm, L2-Hys, L1-sqrt, and L1-norm. This analysis used the L2-norm scheme due to its better performance:

$\boldsymbol{v} \rightarrow \boldsymbol{v} /(\| \boldsymbol{v} \|_2^2 + \varepsilon^2)$^0.5

Where ε is a small positive value used for some regularization when an empty cell is taken into account and v stands for the characterization vector [8].

3.6. Detector Window

Detector window size is 64x128 pixels. Our detection window result in 8 x 16 cells and 7 x 15 R-HOG blocks, since the blocks are overlapping. Each R-HOG block has 2 x 2 cells, which also has 1 x 9 histogram vector each. So the overall size of window is 7 x 15 x 2 x 2 x 9. Therefore, the feature vector size is of 3780. Our detection window includes about 16 pixels of margin around the person on all four sides. This border provides a significant amount of context that helps detection.

4. Support Vector Machine

Large margin classifiers have demonstrated their advantages in many vision tasks. SVM is one of the popular large margin classifiers which has a very promising generalisation capability. Linear SVM is best understood and simplest to apply [11,12].

In our experiment, SVM is used for comparison.SVM is used for training. It is effective for learning with small sampling in high-dimensional spaces. The objective in SVM is to find a decision plane that maximizes the interclass margin [12].

The PL-SVM consists of multiple linear SVMs and has the ability to do non-linear classification. In the application of the PL-SVM to human detection, each linear SVM of the PL-SVM is responsible for one cluster of humans in a specific view or posture. All the linear SVMs combined can well tackle the multi-view and multi-posture detection problem. We have proposed a PL-SVM training algorithm that can automatically divide the feature space and train the PL-SVM with the margins of the linear SVMs increased iteratively.

Figure 5: Examples of positive and negative samples from INRIA image dataset

5. Experimental Results

For the human detection the INRIA human images database is used. INRIA dataset contain single human centred images. This database consists of 2416 positive training and 912 negative training images.   It consists of 1126 positive and 300 negative testing images.

The performance of detector is evaluated for various performance parameters and cross validation accuracy is calculated. The image is split in to the 2×2 pixel cell, 4×4 pixel cell and 8×8 pixel cell. It gives better results for the 8×8 pixel cell because it better describes the orientation of gradients at the edge. The cross validity accuracy is more in 64×128 pixels detection window than other detection window because it considers the human body ratio and well fitted for the cantered image INRIA database. The comparison of the cross validation accuracy for PL-SVM is shown in the result tables.

We have compared the cross validation accuracy for the various SVM kernels and compared the results with PL-SVM. Non-linear SVM kernels perform best when the features are less. For the classification of human the features are more and linear SVM is better suited for it. But sometimes the linear SVM becomes ambiguous for some specific pose of humans and thus results in wrong classification. This can be avoided by forming the various clusters of human poses using linear SVM. In PL-SVM each cluster is dedicated to specific pose of human. Problem of linear SVM is tackled by PL-SVM to get better cross validation accuracy.

Cell size	Error rate (%)
2×2 pixel	3.15 %
4×4 pixel	1.92 %
8×8 pixel	1.43 %

Table 1: Cross validation results for various cell sizes

Bin size	Error rate (%)
4 bin (0 – 180 degree)	6.43 %
6 bin (0 – 180 degree)	4.53 %
9 bin (0 – 180 degree)	1.43 %

Table 2: Cross validation results for various orientation bin sizes

Detector window size	Error rate (%)
46 × 112 pixel	6.35 %
56 × 120 pixel	4.13 %
64 × 128 pixel	1.43 %

Table 3: Cross validation results for various detection window sizes

SVM Kernel Function	Error rate (%)
Linear - SVM	4.75 %
Polynomial - SVM	9.53 %
Quadratic - SVM	9.26 %
MLP - SVM	12.13 %
PL - SVM	1.43 %

Table 4: Cross validation results for various SVM Kernels

6. Conclusion

In this paper, we proposed efficient implementation of human detection system using Histogram of Oriented Gradients features (HoG) and Piecewise Linear Support Vector Machine algorithm. The problem of the multi-view and multi-posture detection can be tackled by PL- SVM. We have proposed a PL-SVM training algorithm that can automatically divide the feature space and train the PL-SVM with the margins of the linear SVMs increased iteratively.

7. References

1. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE Int. Conference on Computer Vision Pattern Recognition, Jun.2005, pp. 886–893.
2. Qixiang Ye, Zhenjun Han, Jianbin Jiao, Jianzhuang Liu, “Human Detection in Images via Piecewise Linear Support Vector” IEEE Transaction on Image Processing , VOL. 22, NO. 2, February, 2013.
3. Yanwei pang, He Yan, Konquio wang, “ Robust CoHOG feature Extraction in human centered inamge/ video management system”, IEEE Transaction on System, management and ybernetics, Vol. 42, No. 2, April 2012, pp. 458-468.
4. Bongjin Jun, Inho Choi, and Daljin Kim, “Local transform features and hybridization for accurate face and human detection”, IEEE transaction on pattern analysis and machine intelligence, Vo. 35, N0. 6, pp. 1423-1436.
5. Q. Zhu, S. Avidan, M. Yeh, and K. Cheng, “Fast human detection using a cascade of histograms of oriented gradients,” IEEE Int. Conf. Computer Vision. Pattern Recognition., Jul. 2006, pp. 1491–1498.
6. Fen Xu, ming Gao, “ Human detection and tracking based on HOG and particle filter”, IEEE International Congress on Image and Signal Processing, CISP-2010, pp. 1503-1507
7. Kelvin Lee, Che Yon Choo, Hui Quing See, and Yunil Lee, “ Human detection using histogram of oriented gradients and human body ratio estimation”, IEEE conference on Comp. Vision, May 2010, pp 18-22
8. R. Xu, B. Zhang, Q. Ye, and J. Jiao, “Cascaded L1-norm Minimization learning (CLML) classifier for human detection,” in Proc. IEEE International Conference on Computer Vision. Pattern Recognition., Jun. 2010, pp. 89–96.
9. P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using Patterns of motion and appearance,”, Int. J. Comput. Vis., vol. 63, no. 2, pp. 153–161, 2005.
10. O. Oladunni and G. Singhal, “Piecewise multi-classification Support vector machines,” in Proc. Int. Joint Conf. Neural Network., Jun. 2009, pp. 2323–2330.

11. S. Q. Ren, D. Yang, X. Li, and Z. W. Zhuang, “Piecewise support vector machines,” Chin. J. Comput., vol. 32, no. 1, pp. 77–85, 2009.
12. H. B. Cheng, P.-N. Tan, and R. Jin, “Efficient algorithm for localized support vector machine,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 4, pp. 537–549, Apr. 2010.
13. Junfeng Ge, Yupin Luo, and Gyomei Tei, “Real time pedestrian detection and tracking at nighttime for driving assistance system”, IEEE trans.on Intelligent Transportation Systems, vol. No. 2, pp.283-298,June. 2009.
14. Essam El-Naqa, Yongyi Yang, Miles N, Werick, Nicolas P. Galatsanos, and Robert m. Nishikawa, “ A support vector machine approach for detection of micro-calcification”, IEEE Transaction on Medical Imaging, vol. 21, Noi.12, December 2002, pp. 1552-1563.
15. Boudour Ammar, Ali Wali, Adel M. Alimi, “ Incremental learning approach for human detection and tracking”, IEEE International Conference on innovations in Information technology, 2008, pp. 128-133.
16. Sung Tae An, Jeong Jung Kim, and Ju Jang Lee, “SDAT- Simulteneous Detection and Tracking of Humans using Partical Swarm Optimization”, IEEE International Conference on Mechatronics and Automation, Aug.2010, pp. 483-488.
17. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, “Object recognition with cortex-like mechanisms,” IEEE Tranaction Pattern Anal. Mach. Intell., vol. 29, no. 3, pp. 411– 426, Mar. 2007.
18. O. Tuzel, F. Porikli, and P. Meer, “Pedestrian detection via Classification on Riemannian manifolds,” IEEE Trans. Pattern Anal. Mach. Intell.,vol. 30, no. 10, pp. 1713–1727, Oct. 2008.
19. Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, “Discriminative Local binary patterns for human detection in personal album,” in Proc. IEEE Int Conf. Comput. Vis. Pattern Recognit., Jun. 2008, pp. 1–8.
20. X. Wang, T. X. Han, and S. Yan, “An HOG-LBP human detector with partial occlusion handling,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2009, pp. 32–39.
21. B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors,” in Proc. IEEE Int. Conf. Comput. Vis., vol. 1. Oct. 2005, pp. 90–97.
22. P. Sabzmeydani and G. Mori, “Detecting pedestrians by learning shapelet features,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–8.
23. S. Munder and D. M. Gavrila, “An experimental study on pedestrian classification,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 11, pp. 1863–1868, Nov. 2006.
24. Q. Ye, J. Jiao, and B. Zhang, “Fast pedestrian detection with multi-scale orientation features and two-stage classifiers,” in Proc. IEEE 17th Int. Conf. Image Process., Sep. 2010, pp. 881– 884.
25. W. Gao, H. Ai, and S. Lao, “Adaptive contour features in Oriented granular space for human detection and segmentation,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 1786–1793.
26. Y. Liu, S. Shan, W. Zhang, X. Chen, and W. Gao, “Granularity-tunable gradients partition (GGP) descriptors for human detection,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 1255–1262.
27. Z. Lin, L. Davis, and D. Doermann, “Hierarchical part-template Matching for pedestrian detection and segmentation,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2007, pp. 1–8