Human Body Detection using Histogram of Oriented Gradients and SVM

Kishor B. Bhangale
Department of Electronics & Telecommunication
R.M.D Sinhagad School of Engineering, Pune, India
Email: kishor.bhangale@yahoo.com

Prof. R. U. Shekokar
Department of Electronics & Telecommunication
R.M.D Sinhagad School of Engineering, Pune, India
Email: rajesh.shekokar@sinhgad.edu

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 support vector machine (SVM) classifiers. Histogram of oriented gradient (HOG) gives an accurate description of the contour of human body. Based on HOG and support vector machine (SVM) theory, a classifier for human is obtained. I have to evaluate the performances of pedestrian histogram of oriented gradients (HOG) and Support Vector Machine on INRIA human database images.

Keywords – Human Detection, Histogram of Oriented Gradients, Classification, Support Vector Machine.

I. 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 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) , HOG-LBP [1], Edgelet [2], Shapelet [3], Local Receptive Field (LRF) [4], Multi-Scale Orientation (MSO) [5], Adaptive Local Contour [5], Granularity-tunable Gradients Partition (GGP) descriptors [5], pose-invariant descriptors [7], Practical Swarm Optimization . 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 normalised gradient histograms. 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 [11]. 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 as Piecewise Linear SVM (PLSVM), Kernel SVM, and Profile SVM are options to handle this problem, but they are generally much more computationally expensive than linear methods.

II. 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 regions (cells), 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 somewhat larger spatial regions (blocks) and using the results to normalize all of the cells in the block.

We will refer to the normalized descriptor blocks as Histogram of Oriented Gradient (HOG) descriptors [1]. Tiling the detection window with a dense (in fact, overlapping) grid of HOG descriptors and using the combined feature vector in a conventional SVM based window classifier gives our human detection chain.

The HOG/SIFT representation has several advantages. It captures edge or gradient structure that is very characteristic of local shape, and it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations: translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size [15]. For human detection, rather coarse spatial sampling, fine orientation sampling and strong local photometric normalization turns out to be the best strategy, presumably because it permits limbs and body segments to change appearance and move from side to side quite a lot provided that they maintain a roughly upright orientation [4].

### III. 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.

#### A. Gamma and Color Normalization

We use exponential gamma correction function to remove the effect of ambient disturbance.

#### B. Gradient Computation

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

\[
I_x = [-1 0 1] \text{ and } I_y = [1 0 -1] \text{.}
\]

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

\[
\text{Magnitude} = |G| = (I_x^2 + I_y^2)^{0.5},
\]

\[
\Theta = \arctan (I_y / I_x)
\]

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°.

#### C. 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 0-1800 ("unsigned" gradient) or 0-3600 ("signed" gradient) [1].

To reduce aliasing, votes are interpolated bilinearly between the neighbouring 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.
D. 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.

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.

E. 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:

$$v \rightarrow \frac{v}{\sqrt{v^2 + \epsilon^2}} \times 0.5$$

where $\epsilon$ 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].

F. Detector Window

As previously mentioned, the detector window size is 64x128 pixels. Our 64_128 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.
IV. SUPPORT VECTOR MACHINE

To train our human descriptor, simple binary linear SVM is used in this research. It is a useful technique for data classification [10]. Somehow, it is sufficient in the context of a human detection problem. Training method is also very important for detection result. Reasonable training method improves result efficiently [2]. So for fair comparison of different features, the effect of training method should be considered. Support vector machine and many boosting methods, such as Adaboost, Logitboost and Gentleboost are widely used in many tasks.

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 size of sub window should be fixed. It is hard to take variable window size strategy because of the computation problem, although variable window could improve the performance efficiently. But for comparison purpose, SVM is suitable. The training time is less than boosting method. Optimization is guaranteed. The difference of performance caused by optimization can be ignored. The parameter of SVM is controllable [12]. The suitable parameter could be selected avoiding the difference caused by parameter difference.

V. CONCLUSION

In this paper, we proposed efficient implementation of human detection system using Histogram of Oriented Gradients features (HoG) and Linear Support Vector Machine algorithm. The proposed algorithm will consist of efficient detection of human in images based on detection error trade off and false positive per window.

The detection rate increases by increase in size of orientation bin and for 64 * 128 descriptor sliding window. The performance of HOG- SVM detection method is expected better than previous method for human detection algorithm like human detection using principal component analysis, human detection using Local binary Pattern, human detection using image subtraction techniques etc on the basis of detection rate and efficiency.

REFERENCES