A Novel Algorithm for Fast Human Detection Based on HOG

CHUNXIANG LI, JIAHONG XU and ZHEJI SHEN

ABSTRACT

Human detection has been appear for several years and has greatly effect our daily life for it is the base of surveillance system, intelligent control and auto operation, which has changed our life more convenient. However, as the complexity of human detection algorithm, the system above cannot response in time or interact normally with human. In this paper, we focus on the algorithm based on Histogram of Oriented Gradient (HOG) and proposed a novel method to accelerate the process of this algorithm by reducing the regions that needless for useless calculation.

KEYWORDS

Human detection; Histogram of Oriented Gradient, Redundancy reducing.

INTRODUCTION

With the development of technology, computers have played an important role in our daily live and work. Due to the efficiency and automatic operation of computer, it is competent to perform intelligent work like certification work that base on fast human detection and face detection. Human detection has been widely used in many public scenes such as surveillance system, people counting, intelligent control, and smart vehicles base on driving record for preventing accident. However, the wide variability in appearance of human due to the height, weight, clothing, pose and disunity of the complex scenes challenge this work for the follow reason. Firstly, the features that should be enough to cover all type of scenes are difficult to extract from the data source. Secondly, the great amount of assemble all the appearances and backgrounds affect the efficiency of features extracting algorithm, large space and time are necessary to store and handle the appearance and background information; Thirdly, the other factors, such as light, camera angle and so on, influence the sharpness which could help expose useful features or hide important information.

Recently, researchers have overcome some difficulties. In processing the image, there is integral processing and fractional processing that had been proposed, while the integral refers to process and analysis the whole area of the image and the fractional processing refers to divide the image into several small sub regions and analysis the image through the correlation between regions. For the feature, Gavrila...
& Philomin used edge orientation histograms to perform the characteristic but it is not robust enough and is greatly influenced by the noisy of background [1]. Papa Georgiou has use Haar wavelets as descriptor to be extracting pedestrian [2]. Shashua et al. [11] divided the characteristic of human into 13 regions and calculate the scale-invariant feature transform(SIFT) features to classifier pedestrian by using Ada Boost. In [3], Dalal & Triggs proposed an algorithm using a dense of Histogram of Oriented Gradient (HOG) which can extract the feature of human as a vector faster, and they use Support Vector Machine (SVM) for the classifier training. This algorithm attracted much attention because it provider robust description of detection window. Though, the method base on HOG [3] shows lot of shortcomings such as the large calculation and it is not capable enough for intelligent control. Base on the HOG, Zhu et al. speed up the detecting method by combining the cascade of rejecter approach and designed a large set of blocks for each size of image. Then Xiaoyu Wang enriched the feature set by combining HOG and Local Binary Pattern (LBP) and using the mean shift algorithm to improve correct ratio of detection. Base on HOG-LBP, Won-Jae et al. presented a novel algorithm by reducing the regions that would be selected by AdaBoost classifier to speed up the process of detecting [6]. After that, Aili Wang et al. presented a novel algorithm by combining HOG with LBP histogram Fourier which efficiently enhance the rate of the true positive and reduce rate of false positive [7]. In [8], Jain Stoble & Sreeraj proposed a method using hybrid HOG and BO (Block Orientation) feature which can reduce the number of false positive. Then, Sumati et al. proposed an algorithm using HOG, HOC (Histogram of Color), HOB (Histogram of Bar) and BO [9], it improved the accuracy further.

The algorithm that searchers had proposed mainly deal with the feature that human could be and the classifier that can distinguish the image whether it contain human. However, the algorithm above base on HOG cost much time to calculate the regions that contain only the sky or other background that is "smooth". In this paper, we focus on the algorithm base on HOG that using a linear SVM as a classifier. To speed up the algorithm and enhance the rate of true positive, firstly, we smooth the image with the mean shift algorithm. Secondly, based on image that smoothed before, we calculate the variance of block with difference scales and then obtain several regions where is impossible that human appear. Thirdly, we choose the candidate regions that would be a human base on the result of mean shift and the regions selected at the second step. Fourthly, we extract the feature in a small area from the candidate regions and determine the small area whether it would be a part of human by using one of the linear SVM classifier. Finally, by clustering the small area, we get the regions that might contain a human, and get the result from another SVM classifier. The rest of this paper is organized as follows: In section 2, we review the related algorithm that had been proposed before. Then we present our method in section 3. In section 4, the experiment results would be shown. And in section 5, we drew a conclusion of our work.

**RELATED WORK**

**Description of HOG**

Dalal & Triggs [3] had proposed the detecting algorithm based on histogram of oriented gradient (HOG) since 2005. It can extract the feature of the appearance and
shape of images without other information of correlation gradient and correct position of edge. By using Gamma correction, the feature extracting would be affected less. The algorithm of HOG can be described as follow:

A. To reduce the influence of lighting, the first step of algorithm is normalization the image by using Gamma correction. In the texture intensity of the image, the contribution of the local surface exposure is larger, so this correction can effectively reduce the local shadow and illumination changes.

B. Calculate the ladder of the image. By using the derivative operation, the silhouette and texture information would be captured. Furthermore, the influence of illumination would be weakened.

C. To provide a coding for the local image region, while maintaining the weak sensitivity of the pose and appearance of the human object in the image. We divide the image window into several small regions called "cells". Then the gradient histogram or edge direction of all pixels in each cell is accumulated. Finally, the basic histogram of the orientation is mapped to a fixed angle, and the final features are formed.

D. In order to compressing light, shadow, and edges, normalization is required. Generally, each cell is shared by seven different blocks, but its normalization is based on different blocks, so the results are not the same. Therefore, the characteristics of a cell will appear in the final vector multiple times with different results. After we normalize, the block descriptor is called the HOG descriptor.

F. By collecting the HOG descriptor of detection windows, the set of HOG feature would be obtained, which would be used for SVM.

Here we interpret the extraction of HOG. The detection windows of size 64 ×128 pixels are divided into 105 overlapping blocks. Each blocks of size 16×16 are divided into 2×2 cells of size 8×8 pixels. Each cell consists of a 9-bin Histogram of Oriented Gradients (HOG) and each block contains a concatenated vector of all its cells. Thus, each block can be represented by a 36(2×2×9) − D feature vector, which would be normalized to an L2-norm vector. For the reason that blocks are overlapping each other, each detection windows of size 64×128 pixels are covered and represented by 7×15 blocks, which means there is a total of 3780 features per detection window. These features are then used to train a linear SVM classifier.

The normalization scheme L2-norm can be represented as:

\[ v \rightarrow \frac{v}{\sqrt{\| v \|^2 + \varepsilon^2}} \]  

(1)

Where \( v \) is the features vector before normalized, \( \| v \|=k \) is k-norm (k=1, 2,..), and \( \varepsilon \) is a small constant.

However, these features are extracted from a fixed size block (105,64×128), which is restricted to a single scale and unsuitable for a big image, it is, these features cannot be use in multiple scales. Thus, Zhu et al. proposed a fast-human detecting based on the algorithm mentioned above. They spread the features vector from 105 blocks to 5031 block that contained each size of block and is suitable for multiple scale detection window. To speed up the algorithm, they replaced the L2 normalization with L1 normalization for reducing the calculation which can be represented as (2). And they used AdaBoost algorithm to improve the results of detection.
\[ v \rightarrow v/(||v|| + \varepsilon) \] (2)

**SVM Classifier**

SVM is widely used in object recognition for it can find the separate hyper, which allows us to maximize the distinction between object and no-object regions. Thus, it is suitable to drive it as a classifier. The HOG feature vectors of positive and negative samples that extracted above are used to train a linear SVM. The SVM detector obtained from training is represented by:

\[ f(x) = \beta + w^T \cdot x = \sum_{i=1}^{105} (\beta_i + w_i^T \cdot x_i) = \sum_{i=1}^{105} f_i(x_i) \] (3)

Where \( x \) is the features vector extracted above that contained 105 vectors of blocks (exactly contained 3780 element) and \( \beta_i \) is the constant bias and \( w_i \) is the weighting vector of the linear SVM of the \( i \) th block.

Zhu et al. construct rejection cascade to accelerate the detection process that contained several levels of cascade and each cascade contained several weak linear classifier (SVM).

**Mean Shift**

An image can be represented as a vector with the length of \( p \) in a two-dimensional grid and each point represents a pixel. It is a grayscale when \( p=1 \). It shows a color map when \( p=3 \). Coordinates in grid representation of the spatial information of the images. We consider the spatial information and the color (or gray level) information of the image, and form a \( p+2 \)-dimension vector \( x=(x',x^c) \), where \( x' \) represents the coordinates of the grid points. And \( x^c \) represent the vector feature of the grid point [10].

We use kernel function to estimate the distribution, which has the following form:

\[ K_{h_s,h_r} = \frac{C}{h_s^2 h_r^p} k \left( \frac{\|x\|}{h_s} \right) k \left( \frac{\|x^c\|}{h_r} \right) \] (4)

Where \( h_s, h_r \) dominate the smooth resolution and \( C \) is a normalization constant.

We use \( x_i \) and \( z_i, i=1,2, n \) to represent the original image and smooth image. The algorithm of Mean Shift can be for image smoothing specific steps is as follows:

Initialize \( j=1 \), and let \( y_{i,1} = x_i \).

Use Mean Shift algorithm to calculate \( y_{i,j+1} \) until it weakens or disappear, and get the value.

Let \( z_j = (x^s_j, y^r_{i,j+1}) \)
Figure 1. (a) the gray values of each pixel of the original image. (b) the Mean Shift moving path of each pixel point. (c) the gray values of each pixel point after smoothing. (d) the result after segmentation.

PROPOSED METHOD

Based on the algorithm that had been proposed. We improve the method that Dalal & Triggs proposed considering the real situation that human in a real image is greatly different from the background with less edge region and the human body has distinguishable characteristic from complex background. For the first issue, we use mean shift to smooth the image and reduce regions with a simple calculation as follow:

A. Use a sliding window to obtain the regions of size 16×16 from image.

B. Calculate the variance of the pixels.

By comparing the variance of pixels, it is easy to distinguish the simple background from other object that contained more edge regions for the variance of simple background is nearly equal with zero.

After obtaining the regions that would not be a human, we use the classifier that Won-Jea Park et al. proposed before to select the block from the rest regions. Difference from their method, we use the input features vector that contained 5031 blocks. Thus, the SVM detector obtained from training is presented as (5). And the essential training of the cascade is shown as follow:

a. Train 250 linear SVM using Pos and Neg samples for random block.

b. Add the best SVM and update $w_i^T$ for each block.

c. evaluate Pos and Neg by current strong classifier

d. decrease $T_{weak−b}$

$$f(x) = \beta + w^T \cdot x = \sum_{i=1}^{5031} (\beta_i + w_i^T \cdot x_i) = \sum_{i=1}^{5031} f_i(x_i)$$  (5)
The method above reduced the regions that would be calculated and selected several candidate windows with multiple scales from the rest regions. And the candidate windows is probably a part of human. By clustering the candidate windows, we obtain the detection window that most likely a human there. To determine whether there is a human, the third step we use a fast-human detection using cascade that Zhu et al. presented. The result of detection shows whether the detection window is a background. Based on the method, we greatly reduce the calculation of extracting the HOG features with 5031 blocks, and partly, the HOG features in the third step had been already calculated in the second step.

EXPERIMENT RESULTS

We implement our method mentioned above and perform the experiment on INRIA data set, which contained 2417 positive training samples and 1269 negative training samples that are images without human. In order to reducing the time, we trained the classifier with 1000 positive samples and 500 negative samples. And we test the algorithm with image with multiple scales.

From the Fig 2 and Table 1, we found that the method we proposed can accelerate the process by reducing the number of features that is needless for extracting and using the candidate windows with less calculation to scan the image. However, our method is of lower accuracy ratio for the reason that it is based on the algorithm that Zhu et al. proposed, and we adopt their algorithm as part of our method. And our method would probably miss the regions with intense noise such as multiple human that influence each other.

![Figure 2. Comparison of FPPW among HOG, Fast human detection and our method.](image)

**TABLE 1. COMPARISON OF PROCESSING TIME.**

<table>
<thead>
<tr>
<th></th>
<th>HOG</th>
<th>Fast human detection</th>
<th>Our method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing time</td>
<td>2sec</td>
<td>152ms</td>
<td>116ms</td>
</tr>
<tr>
<td>Time ratio (%)</td>
<td>100</td>
<td>7.6</td>
<td>5.7</td>
</tr>
</tbody>
</table>
CONCLUSION

In this paper, we proposed a human detection method to accelerate the process of extracting the HOG features by the searching regions and it run efficiently. The main treatment of reducing the regions is reducing the needless regions and obtaining the candidate windows that a human may appear. The method we proposed provided a novel approach to achieve fast human detection based on the block. Though, the result of our method shows a significant shortcoming in accuracy, the advantage could not be ignored in speed. In future work, we would improve the accuracy ratio of this method.

REFERENCES

7. Aili Wang; Shiyu Dai; Mingji Yang; Yuji Iwahori, A novel human detection algorithm combining HOG with LBP histogram Fourier, 10th International Conference on Communications and Networking in China (China Com), 2015, pp. 793-797.