Project Proposal

BY WANG ZIMENG

12011711@mail.sustech.edu.cn
Southern University of Science and Technology

1 Overview

We are now in the times that deep learning is proposed as what all you need. But mainly the traditional feature detection and image processing skills are taught. In my project, althouth there are already many fancy implement done by neural networks, I'd like to implement the human detection by famous HOG(Histogram of Oriented Gradients) feature based on the paper " Hitograms of Oriented Gradients for Human Detection".

As described in the paper " Detecting humans in images is a challenging task owing to their variable appearance and the wide range of poses that they can adopt", the HOG features gives a "excellent performance" relative to the existing feature sets.

Why do we use histogram of gradients? The gradient could characterize the appearance and shape rather well and the SIFT descriptor and Shape contexts all compute on the dense grid of uniformly spaced cells.

At last, the author uses the SVM as a baseline to classify whether there is a pedestrian in the block or not.

2 Implementation and Challenges

The basic process is in the figure below :

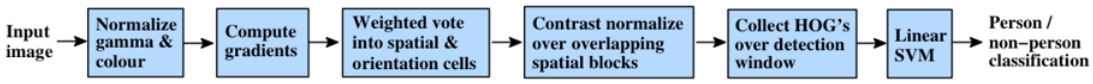

Figure 1.

2.1 Normalization & Gamma Transformation

Data is the key while dealing with problems, but the clean-up of the data is even more important. When we get an image, doing gamma transformation and normalization are necessay since the illumination may be diffcult.

$$Y(x,y) = I(x,y)^{\gamma}$$

where $I(x,y)$ stands for the intensity or brightness of pixel (x,y), and γ is the coefficient.

And the effect is conspicuous :

Figure 2. Origin

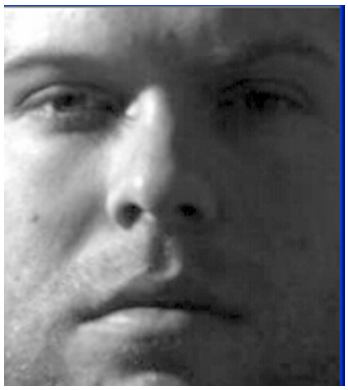

Figure 3. After transformation

Besides that, we can do a normalization to the gradients calclutaed in the steps after to eliminate the magnitude difference caused by different illumination. The *L2-norm*($v \to v/\sqrt{\|v\|_2^2+\epsilon^2}$) *L2-Hys(L2 norm followed by a clipping)* and *L1-sqrt*($v \to \sqrt{v/(\|v\|_1+\epsilon)}$) have similar effects.

2.2 Compute Gradient

Gradient computation is implemented before in the lab session, but for this part we need to calculate the angle of gradient to make a histogram.

$$G_x = I(x+1, y) - I(x-1, y)$$

$$G_y = I(x, y+1) - I(x, y-1)$$

$$G \approx |G_x| + |G_y|$$

$$\theta = \tan^{-1}\left(\frac{G_x}{G_y}\right)$$

where θ is the angle of gradient, $G_x\, G_y$ and G is gradient in x, y direction and overall gradient.

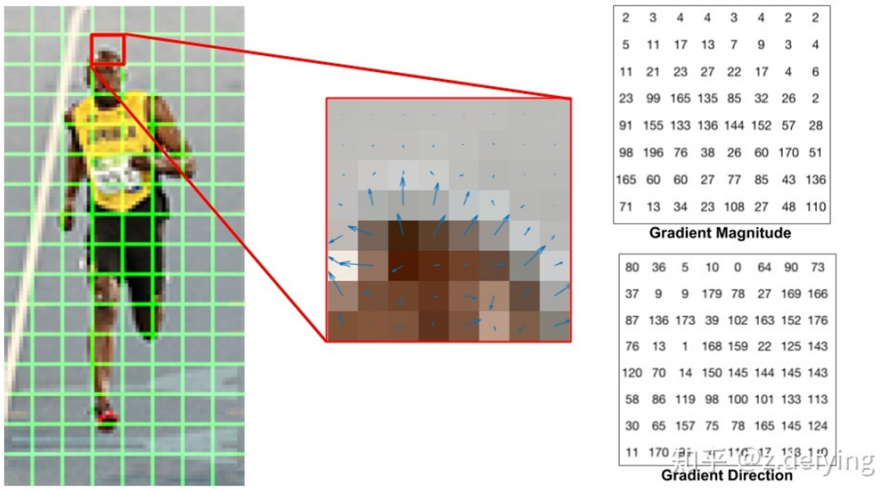

Figure 4.

2.3 Weighted vote & Orientation cells

For each image, we could divide it into many n*n pixels cell. For each cell, we could calculate the histogram of gradient and we could combine a n*n cells into a block.

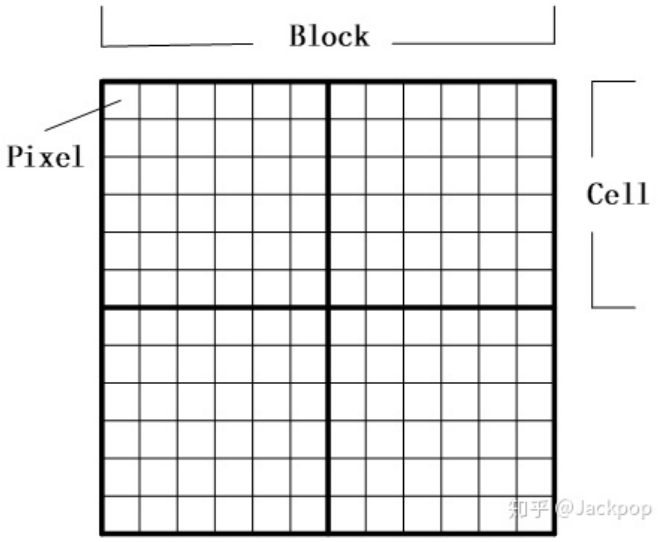

Figure 5.

The histogram of oriented gradient is calculated by a vote using bilinear interpolation. After the gradient calculation, we have magnitutde and direction. Following the result of paper, we would separate the direction in to 9 bins (0 - 180°)

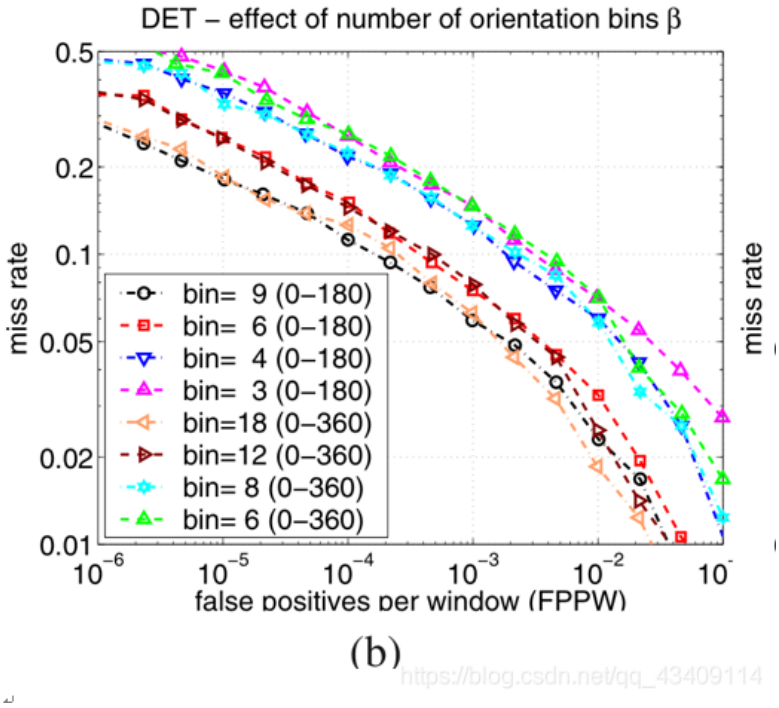

Figure 6.

The implementation image is in below. For example, the magnitude is 4 and the direction is 10°. Since there is no 10° in the bins if we divide the angle into 9, we would give 0 and 20 degree a weight of 2 seperately based on the principle of bilinear interpolation.

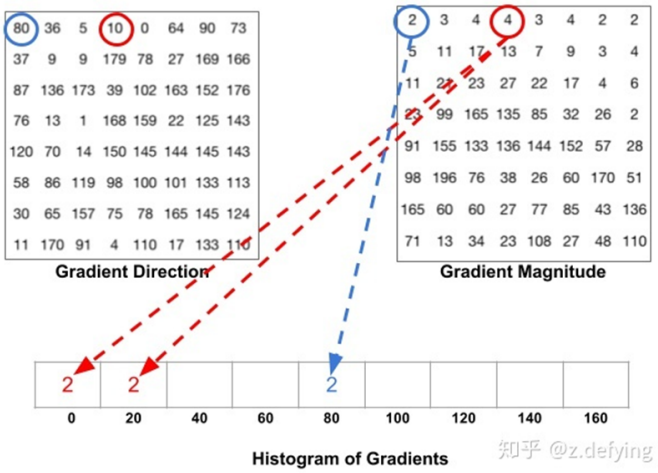

Figure 7.

Finally we will get the histogram of angle and is going to take this as the feature.

2.4 Feature construction and SVM

For each block, we could append the histogram of every cells(a 9 dimension vector) in the block together to form a 36 dimention feature (for 2*2 cell block). Use the dataset provided in the paper, we could train a support vector machine to distinguish those 36 tunnel vectors with a hyperplane.

Brief introduction about SVM

SVM is a 2-class classifier. Suppose we have a data of n dimension with label y, SVM aims to find a *n-1* dimension plane to seperate thoses n dimension vectors. We could define the geomrtric margin as

$$\gamma_i = y_i\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\cdot\boldsymbol{x}_i + \frac{b}{\|\boldsymbol{w}\|}\right)$$

we want to maximize this distance to find the optimal plane :

$$\max_{\boldsymbol{w},b}\gamma$$

where γ is

$$\gamma = \max_{i=1,2\ldots N}\gamma_i$$

such that :

$$y_i\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\cdot\boldsymbol{x}_i + \frac{b}{\|\boldsymbol{w}\|}\right) \geqslant \gamma, i = 1,2\ldots,N$$

$$y_i\left(\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|\gamma}\cdot\boldsymbol{x}_i + \frac{b}{\|\boldsymbol{w}\|\gamma}\right) \geqslant 1, i = 1,2\ldots,N$$

since $\boldsymbol{w}$ and γ is scalar, we could define :

$$\boldsymbol{w} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|\gamma}$$

$$b = \frac{b}{\|\boldsymbol{w}\|\gamma}$$

and we could get :

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geqslant 1, i = 1, 2, \ldots, N$$

maximizing γ is equal to maximize $\frac{1}{\|\boldsymbol{w}\|}$, and also equal to $\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2$. After using lagrange duality, we could get :

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) - \sum_{i=1}^{N} \alpha_i$$

$$s.t. \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

where $\boldsymbol{\alpha}$ is lagrange multiplier.

and we could get the optimal $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*)^T$

then

$$\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i$$

$$b^* = y_j - \boldsymbol{w}^* \cdot \boldsymbol{x}_j$$

Finally, use the decision function $f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}^* \cdot \boldsymbol{x} + b^*)$ to decide whether there is a human or not.

3 Expected Result

Using this method, we could see the rough outline of a object i.e. human.

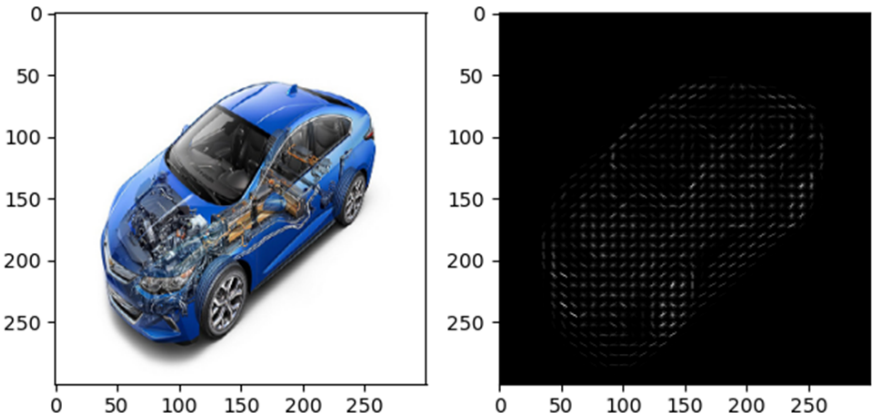

Figure 8.

And could recognize a pedestrian by SVM method.

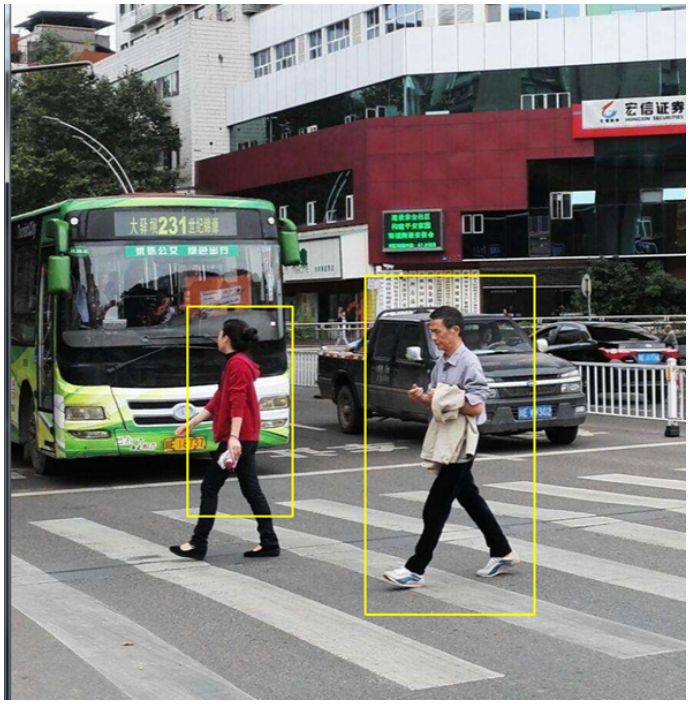

Figure 9.