Human Detection Based on Integral Histograms of Oriented Gradients and SVM

Yahia Said, Mohamed Atri, Rached Tourki
Laboratory of Electronics and Microelectronics (EμE)
Faculty of Sciences Monastir, 5000 TUNISIA
said.yahia1@gmail.com, mohamed.atri@fsm.rnu.tn, rached.tourki@fsm.rnu.tn

Abstract — This paper presents a method for human detection in video sequence. The Histogram of Oriented Gradients (HOG) descriptors show experimentally significantly out-performs existing feature sets for human detection. Because of HOG computation influence on performance, we finally choose a more better HOG descriptor to extract human feature from visible spectrum images based on OpenCv and MS VC++. We realized an image descriptor based on Integral Histograms of Oriented Gradients (HOG), associated with a Support Vector Machine (SVM) classifier and evaluate its efficiency.

Keywords- Human detection; Image descriptor; Integral HoG; SVM, OpenCv.

I. INTRODUCTION

The automatic objects recognition in images is one of the most difficult problems in computer vision. It is at the same time a paramount stage for implementation of several current applications which require a high level images interpretation. Consequently, there is an increasing interest on this research field in the last years and a vast literature. This paper fits more particularly in this context. Indeed, the problem of objects recognition is to decide if a specific object or an object pertaining to a preset object category exists in an image. We can pose the problem of the recognition like a pairing between the model of the target object and a whole of descriptors which we extract from the test image.

After the development of successful face detection algorithms, human detection was the logical next step. However, the pattern recognition field applied to the human detection is faced with nature even of the human. Various variabilities enter in account: variability on the scale, of posture and appearance. Indeed, the human can be positioned with various sites in the image, under various aspects and point of view. The first need is a robust feature set that allows the human form to be discriminated clearly, even in cluttered backgrounds under difficult illumination.

Dalal and al. [1] presented a human detection algorithm with excellent detection results. Their algorithm uses a dense grid of Histograms of Oriented Gradients (HOG) to represent a detection window. The Histogram of Oriented Gradient (HOG) [1] descriptors provide better performance than other existing feature sets. Each stage of the HOG computation has much influence on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks are all important for good results.

Zhu and al. [2] then showed that the combination of the cascade of rejecters approach and the HoG features led to a fast and accurate human detection system. The features were HOGs with variable block-size. To select the best blocks for the detection, from a large possible block set, AdaBoost was used. Furthermore, the integral image representation was used and the rejection cascade significantly speeded up the computation.

In this paper, we propose a single frame human detection system which follows the path of Dalal and al.[1] and Zhu and al.[2] This detection system is based on Integral Histogram of Gradients descriptor combined with Support Vector Machines [3] for the recognition stage. SVM is a classification method, which has proved to be very efficient in such case of high dimensional data. It has been developed to detect a human centered in a 128 × 64 single image. The implementation of our system was based on MS VC++ and OpenCv [4] which is an Open source computer vision library, originally developed by Intel, specialized in the real time image processing.

This paper is organized as follows: section II describes related works in human detection systems. In section III, we detail the single frame detector and we give more precision about the Integral HOG descriptor and its parameters. Results are discussed in paragraph IV and, finally, section V ends the paper with some final remarks.

II. RELATED WORKS

Many different approaches for human detection have been proposed in the literature. Papageorgiou and al. [5], described the object shape in an image using wavelet decomposition. The coefficients obtained make it possible to describe the presence of contours in the image. This type of description obtains very good results, because it characterizes the general object shape, without being related to the objects appearance.
A powerful method for pedestrians’ detection was presented by Shashua and al. [6]. This method uses histograms of gradient orientation. The gradient makes it possible to determine the brightness variations in the image and thus contours. The histograms are not calculated on the whole image, but correspond to local areas. The image is then divided in various areas located in specific zones of the pedestrian: head, legs, arm. Each area generates a histogram, the whole of the histograms forming the image’s characteristic data. The results are related to the pedestrian’s position in the image and do not authorize significant variations in its posture.

Lowe [7] uses a description with orientation of the gradient around interest points. The research of the interest points is carried out of multi-resolution, which makes it possible to define a variable size for the vicinity interest points. The size depends on the scale to which the interest point is detected. The image is then characterized by a whole of local histograms. By regarding these data as “bags of words”, we lose then the data space component, which reduces the method relevance.

The method introduced by Dalal and al.[1], is founded on a description of the image using histograms of gradient orientation. These histograms are calculated on the whole image, but are defined on a local area. The image description is a vector containing all the calculated histograms. This vector brings, on the one hand, local information for each area and makes it possible, on the other hand, to preserve a space fitting, the vectors being calculated systematically by the same process.

A complete system for pedestrian detection based on HOG descriptors was introduced by Suard and al. [8]. Their system deals with stereo infrared images. The system first locates positions of interest within the larger view field where humans could possibly be located. Then normal Support Vector Machine classifiers operate on the HOG descriptors taken from these smaller positions of interest to formulate a decision regarding the presence of a pedestrian. This approach gives good results for window classification, and a preliminary test applied on a video sequence proves that this approach is very promising.

Zhu and al. [2] presented an algorithm to significantly speed up human detection using HOG descriptor methods. They use HOG descriptors in combination with the cascade of rejecters’ algorithm. The most important point of their approach was that features were selected from a large set of multiple size blocks, locations and aspect ratios using AdaBoost as a feature selection.

Zhang and al. [9] optimized the HOG features to achieve an accurate human detection system. They don’t normalize the input detection windows but resize the cell and block by same ratio. Simulation results show that their approach method was effective.

III. OVERVIEW OF THE METHOD

A. Integral HOG Feature

We give, below, followed steps to calculate the integral HOG descriptor:

- Image brightness normalization

The entry images of our system are initially converted into gray level. They are then normalized according to the brightness information (Figure 1).

- Gradient computation

The gradient computation is a critical stage in the descriptors formation. The accuracy of the calculated orientations, and the histograms, depends on this stage and the results are thus closely related to the method employed to calculate the image gradient. The fast calculation of the gradient can be made, for example, by masks of simple derivation 1-D (centered [- 1, 0, 1] and not centered [- 1, 1]), by the 2-D Sobel operators, or by the Deriche recursive operators. In our method, we use the Sobel algorithm. It is the one of the simplest operators which gives however correct results. For a given image I, horizontal and vertical gradient of each pixel (x, y), are calculated as shown in equation 1 and equation 2 (Figure 2).

\[
G_x(x, y) = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * I(x, y)
\]  
(1)

\[
G_y(x, y) = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I(x, y)
\]  
(2)

In each pixel (x,y), the approximations of the horizontal and vertical gradients are combined as follows, to obtain an approximation of the gradient value:

\[
G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}
\]  
(3)
The orientation of gradient is calculated as follows:

$$\theta(x, y) = \arctan \left( \frac{G_y(x,y)}{G_x(x,y)} \right)$$

(4)

**Histograms computation**

Each detection window is divided into sized cells $8 \times 8$ pixels and for each cell we compute the histogram of gradients by accumulating votes into bins for each orientation. Votes are weighted by the magnitude of a gradient. We used the ‘integral image’ representation to compute efficiently the HoG of each cell.

The integral image [10], detailed in Figure 3, is an intermediate representation for images which allows a rapid computing of rectangular features.

With the integral image, we can compute the sum of the elements within a rectangular region by using 4 image access operations.

Porikli [11] suggested a similar method to efficiently compute histograms over arbitrary rectangular image regions, called 'Integral Histogram' (see Figure 4).

Based on the last methods, we discretize each pixel’s gradient orientation into 9 bins, then compute and store an integral image for each histogram bin. The HoG for any rectangular region then can be computed by $9 \times 4 = 36$ image access operations, 4 for each of the 9 bins. Process of computation and storing of an integral image for each bin of HoG are to increase the efficiency in computing the HOG for any rectangular image region.

For a block of $2 \times 2$ cells, we calculate the Integral HOG descriptor for each cell, the vectors obtained are assembled in a single 1-D vector of 36 components and is L1-normalized. This normalized vector corresponds to a block Integral HOG descriptor.

An image of $64 \times 128$ pixels contains $7 \times 15$ blocks with multiple overlapping. We assemble the normalized vectors obtained for all blocks in a single 1-D vector of 3780 components. This is our final Integral HOG descriptor.

**B. SVM Classifier**

Vladimir Vapnik proposed the Support Vector Machine (SVM) in 1995 [3]. Since this time, the SVM have been largely used for the pattern recognition, the regression and the density estimation.

Given a training data set $(x_k, y_k) \in R^n \times \{-1, 1\}$, where $x_k$ are the training examples Integral HOG feature vector and $y_k$ the class label, the SVM require the solution of the following optimization problem:

$$\min_w b, \xi \quad \frac{1}{2} w^T w + c \sum_{k=1}^n \xi_k$$

Subject to

$$y_k (w^T x_k + b) \geq 1 - \xi_k \quad \xi_k \geq 0$$

(5)

The method consists in mapping $x_k$ in a high dimensional space owing to the function $\phi$. Then SVM looks for an optimal decision function of the form $f(x) = w \cdot \phi(x) + b$.

The kernel function is $(x_j^T x_i) = \phi(x_i)^T \cdot \phi(x_j)$.

The basic kernels are:

- Linear:
  $$k(x_i, x_j) = x_i^T x_j$$

(6)
• Polynomial:
  \[ k(x_i, x_j) = (y \cdot x_i^T x_j + r)^d, \quad y > 0 \quad (7) \]

• Radial Basis Function (RBF):
  \[ k(x_i, x_j) = \exp \left( -\gamma |x_i - x_j|^2 \right), \quad y > 0 \quad (8) \]

• Sigmoid:
  \[ k(x_i, x_j) = \tan(y x_i^T x_j + r) \quad (9) \]

In our experiments, we use Linear kernel to classify the feature data of train set.

IV. EXPERIMENTS AND RESULTS

We evaluated our system with two different static human data sets:

• The MIT pedestrian database [12]: contains 709 standing human images taken in city streets. The set only contains images featuring the front or back of human figures and contains little variety in human pose.

Figure 4 shows some sample images from the MIT pedestrian data set.

• The INRIA database [13]: proposes examples of human in varied postures, with different appearances and under very diverse scenes and luminosity conditions.

Figure 5 shows some sample images from the INRIA static person data set.

For the test, we use 1178 images as positive examples, and 452 images as negative examples from INRIA data set and 709 images from MIT data set.

This system is executed in Visual Studio 2008 with OpenCV library and synchronized on a 2.0GHz Dual-core Laptop. The test result on test set, compared to [1] and [9], is shown in table 1.

V. CONCLUSION

In this paper, we present a human detection system using Integral Histogram of Oriented Gradients features combined with SVM classifier. This implementation is characterized by the improvement of time computing without performances degradation.
REFERENCES


