Human detection using Histogram of oriented gradients and Human body ratio estimation

Conference Paper · August 2010

DOI: 10.1109/ICCSIT.2010.5564984 · Source: IEEE Xplore

Yunli Lee
Sunway University

Human Detection using Histogram of oriented gradients and Human body ratio estimation

Kelvin Lee, Che Yon Choo, Hui Qing See, Zhuan Jiang Tan, Yunli Lee
Faculty of Information & Communication Technology
Universiti Tunku Abdul Rahman (UTAR)
MALAYSIA.
wing426_elite@yahoo.com.hk, joshua5367@hotmail.com, hq_88@hotmail.com, kirakyoto@yahoo.com, leeyl@utar.edu.my

***Abstract*- Recent research has been devoted to detecting people in images and videos. In this paper, a human detection method based on Histogram of Oriented Gradients (HoG) features and human body ratio estimation is presented. We utilized the discriminative power of HoG features for human detection, and implemented motion detection and local regions sliding window classifier, to obtain a rich descriptor set. Our human detection system consists of two stages. The initial stage involves image preprocessing and image segmentation, whereas the second stage classifies the integral image as human or non-human using human body ratio estimation, local region sliding window method and HoG Human Descriptor. Subsequently, it increases the detection rate and reduces the false alarm by deducting the overlapping window. In our experiments, DaimlerChrysler pedestrian benchmark data set is used to train a standard descriptor and the results showed an overall detection rate of 80% above.**

Index Terms: Human detection, Histogram of Oriented Gradients (HoG), Support Vector Machine (SVM), background subtraction, features extraction, human body ratio estimation, local region sliding window classifier.

I. 1. Introduction

Human detection in images and videos has become an increasingly important research area in computer vision and pattern recognition community during the past few years after the researches on faces [1, 2, 3]. As an active research topic, its potential application can be implemented in video surveillance in dynamic scenes, driving assistance system, content-based retrieval, etc. However, detecting humans in images and videos is still a challenging task owing to their variable appearance caused by variety of clothes, shadows, articulation and illumination situations, and unpredictable poses that they can adopt.

In this paper, we presented a human detection system using Histogram of Oriented Gradients (HoG) which is originally by Dalal and Triggs [4] in 2005 and our approach human body ratio estimation. Histogram of Oriented Gradients is used to classify moving objects into human and non-human categories. We also used the human body ratio estimation method to constraint the objects' characteristics and this helps to filter non-human object fading in and out of the scene.

Our paper is organized as shown below: Reviews the related works; Section 2 details the system architecture, implemented methods and definition of Dalal and Triggs's human detection algorithm [4]; Section 3 presents the human datasets; Section 4 demonstrates the experimental results and finally Section 5 draws the conclusion for this paper.

A. Related Works

In human detection, Dalal and Triggs [4] proposed the Histogram of Oriented Gradients and it reported impressive results on human detection. The factor that makes their works notable is the use of HoG descriptors which outperform the previous feature sets for human detection. However, their human detection approach did not address the problem of processing speed. Qiang Zhu *et al.* [5] used the discriminative power of HoG and integrated the cascade-of-rejectors approach to perform a faster and accurate human detection system. They adopted the integral image representation and a rejection cascade to boost up the computation speed and their human detection system succeed to process 5-30 frames per second. Hui-Xing Jia *et al.* [6] proposed a novel real-time human detection system based on Viola's face detection framework and the HoG feature pool. They used the real-time properties of face detection and the discriminative power of HoG to detect human and vehicles. Xiaoyu Wang *et al.* [7] proposed a novel human detection system by combining HoG and Local Binary Pattern (LBP) as the feature sets in their system which is able to detect human with partially occlusion. W.Zhang *et al.* [8] proposed a multi-resolution framework to reduce the computational cost. Schwartz *et* al. [9] proposed a human detection approach using partial least square analysis, which adopts edge-based features with texture and color information to provide a richer descriptor for detection.

II. Methods

A. Human Detection System Architecture

In this section, the system architecture of our human detection system is presented (Figure 1). This human detection system is using Histogram of Oriented Gradients [4] as the main feature/descriptor set to locate humans in the video scenes. DaimlerChrysler pedestrian datasets [10] is used as the training samples for our system.

978-1-4244-5539-3/10/$26.00 ©2010 IEEE

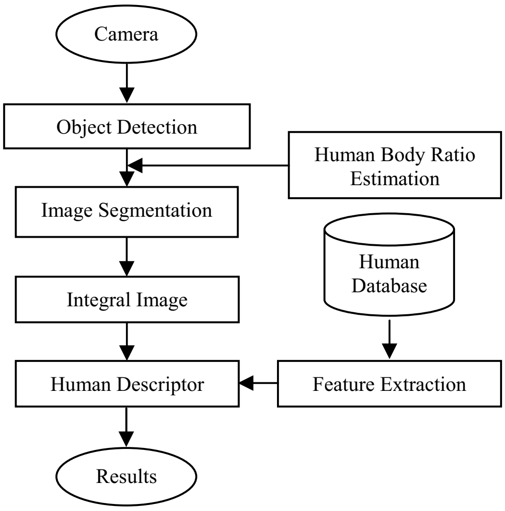


Figure 1. System architecture for Human Detection System.

In this paper, we have to achieve the following qualities: I) to develop a human detection system using HoG by Dalal and Triggs [4] with human body ratio estimation technique and locate human in the video scenes; II) to do research on Histogram of Oriented Gradients (HoG) to detect human.

B. Image Preprocessing

The preprocessing stage of this research includes image acquisition, background subtraction and image segmentation. The framework is shown as below:

- Image Acquisition – an Internet Protocol (IP) Camera is installed and used to capture videos and images for further processing.
- Background Subtraction – subtraction between background image ($background_t$) and foreground image ($frame_t$) is computed to obtain the difference pixel values between both images.
- Image Segmentation – the difference pixel values are thresholded using a threshold value ranging from 0 to 255. The pixels with values lower than the threshold value will be set to 0 whereas pixels with values greater than the threshold value (Th) will be set to 1 and labeled as the motion region (Equation 1).

$$Motion\ Region = |frame_t - background_t| > Th \qquad (1)$$

Image segmentation is used to filter out the unnecessary noises and extract the more concentrated regions, in order to ease the human classification stage which will thus speed up the process.

C. Human Body Ratio Estimation

This technique is simple and widely used in checking the size ratio of the moving object and determines whether it is similar to human body ratio. A normal human standing at a corner usually has the body ratio of 3:7 (width: height) and a normal walking human usually has the body ratio of 5:7. The difference of width and height is insufficient to calculate the weight of the moving blobs. In addition, as the 64 128 detection window is the minimum size of the window, we use the size of the detection window to constraint the size of the moving blobs whereas $0.25 < K < 0.51$.

$$K = \frac{Width_t}{Height_t, Height_{HoG}},\ K = human\ body\ ratio \qquad (2)$$

For instance, if the height of the detected moving blob is greater than the height of the detection window $Height_{HoG}$, our system will pass the current blobs to classification stage; but if the height of the detected moving blobs is smaller than the height of the detection window, the system will ignore it and assume it as noises.

D. Dalal and Triggs Algorithm

In Dalal and Triggs Algorithm, each detection window is separated into cells of size 8 x8 pixels and each group of 2 2 cells for the usage of integration of all into a block in a sliding fashion. Each cell consists of a 9-bin Histogram of Oriented Gradients (HoG) and each block contains a conterminous vector of all its cells where each of them is represented by a 36 – D feature vector that is normalized to an L2 unit length. Each 7 15 blocks is representing a 64 128 detection window which contains 3780 features. A linear SVM classifier is trained by using these features. [4]

E. Integral Image/Histogram of Oriented Gradients

The usage of sliding window classification [9] is always criticized as being too resource and computationally expensive. The integral image/histogram [5, 8, 9] is an intermediate representation for fast location evaluation of Haar-wavelet type features, named as rectangular filter. Somehow, it is able to increase the efficiency of the parallel computing hardware to alleviate the speed problem. Previously, this filter is commonly used in face detection system and that later extended to a human detection system [6]. Within the framework of the integral image/histogram, the extraction of the features for sliding window has a constant complexity. According to [2, 5, 7, 11, 12], sliding window classifiers using the integral image method is able to increase the computation speed by several folds.

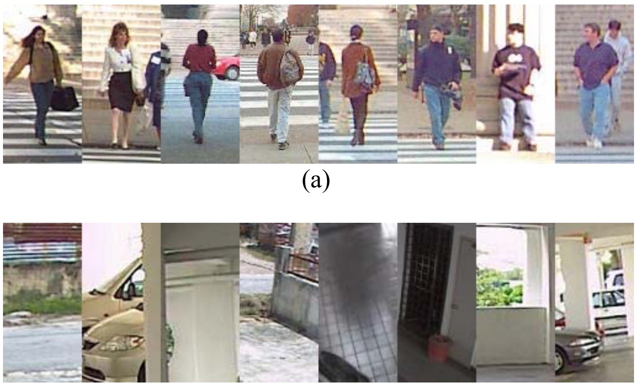
Porikli [13] used the “Integral Histogram” to efficiently compute histograms over arbitrary rectangular image regions. Qiang Zhu et al [5] proposed a fast way of calculating the HoG Feature. Firstly, they discretize each of pixel’s orientation into 9 histogram bins. Process of computation and storing of an integral image for each bin of HoG are to increase the efficiency in computing the Hog for any rectangular image region.

F. 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. Somehow, it is sufficient in the context of a human detection problem.

III. Dataset Preparation

To build a strong human detection system, training samples are needed. Since DaimlerChrysler pedestrian dataset [10] is provided, we used it to train our human detection system. The human images are aligned to a base resolution of 64 128. Different angles of human can be obtained from DaimlerChrysler pedestrian dataset. In addition, we have to capture some non-human objects (false positive) to train the human detection system. The positive samples and negative samples are shown in Figure 2.



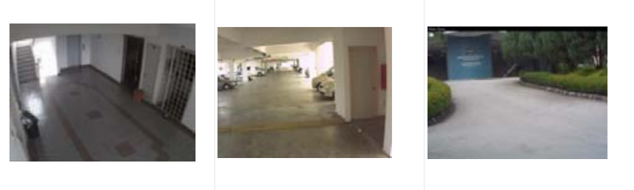
(a)

(b)

Figure 2. (a) Positive Training Samples and (b) Negative Training Samples.

IV. Experimental Result

We implemented Dalal and Triggs algorithm [4] using DaimlerChrysler pedestrian datasets [10] as training samples into our human detection system. This system is executed in Visual Studio 2008 with OpenCV library and synchronized on a 2.0GHz Core 2 Duo Laptop. The frames are in 420 360 pixels resolution and they are sampled at a rate of 15 frames per second. In order to verify and evaluate the performance of our human detection system, it is tested in different environments. Figure 3 shows the testing environments, which include residential area (Dataset A), condominium car park (Dataset B), and campus main entrance (Dataset C and Dataset D).



Residential Area

Condominium Car Park

Campus main Entrance

Figure 3. Testing environments; (A) Residential Area, (B) Condominium Car Park, (C, D) Campus Main Entrance.

At the first stage of our system, moving objects passing through the scene are completely extracted through background subtraction technique. After that, human body ratio estimation technique is applied in order to determine whether the particular moving blob has the size ratio of a human. Once it is verified to have human body ratio, image segmentation is implemented to crop out the local region to further justify it. With the possible human candidates segmented, our system will classify them as human or non-human objects.

From Table 1, Dataset A and Dataset B have higher detection rates of 93.75% and 95.58% respectively as compared to Dataset C and Dataset D. The number of testing samples for Dataset A and Dataset B is lesser and only one person appears in the video scene. In Dataset A, the brightness of the residence environment is lower and it caused the background subtraction unable to extract the motion region accurately. Among all the datasets, Dataset B achieved the highest detection rate of 95.58% due to the higher brightness of the testing environment which helps the background subtraction to efficiently extract the motion region.

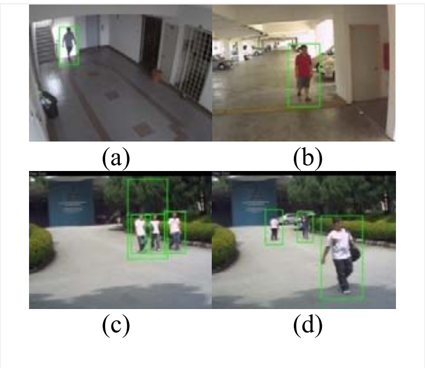
From Table 1, Dataset C has the lowest detection rate of only 72.17%. In this Dataset, there are too many people walking overlapping each other which caused a lot of occlusion. In addition, the outdoor environment has a higher brightness and cause a lot of noises. Even though we use a threshold value of 25 for background subtraction but it is unable to minimize the noises in the scene. To overcome the noises problem, we applied a 3 3 cell size of masking technique to erode the noises. After that, we decided to apply a threshold value of 23 for background subtraction in Dataset D to test our human detection system. With the improvements, the detection rate in Dataset D has been increased to 76%. However, occlusions in this dataset remain as the main problem for our human detection system to detect human efficiently.

TABLE I. Numbers of testing samples, threshold value, detection rate and false positive rate for each dataset.

	Number of Testing Samples	Threshold Value	Detection Rate	False Positive Rate	Precision Rate
Dataset A	240	31	0.9375	2.52e-3	0.9298
Dataset B	240	27	0.95833	2.67e-3	0.9274
Dataset C	600	25	0.72167	1.926e-2	0.797
Dataset D	495	23	0.76	1.62e-2	0.803

We did a comparison on HoG using features with the cell size of 4 4, 8 8, 16 16 and 32 32. We found that 8 8 cell works best because of its patterns that are sufficient to discriminate in these testing environments (Dataset A, Dataset B, Dataset C and Dataset D). In our observation, although 32 32 cell provides more patterns than 16 16 cell, it has much smoothing over the histogram bins. By testing our system in Dataset C and Dataset D, we found that 4 4 cell has the best patterns among all, but it takes the longest time to process images.

In addition, we found that if the size of motion region is smaller than the size of the detection window, the system is hard to identify the motion objects due to the problem of insufficient information of pixel values. Figure 4 shows some typical results of our human detection system. Figure 5 shows the Detection Error Tradeoff (DET) curves of Dataset A, Dataset B, Dataset C and Dataset D respectively.


(a) (b)

(c) (d)

Figure 4. Figure shows the sample detection results of our human detection system in different environments. (a) is obtained from Dataset A, residential area. (b) is obtained from Dataset B, Condominium Car Park, (c) and (d) are obtained from Dataset C and Dataset D respectively, Campus Main Entrance.

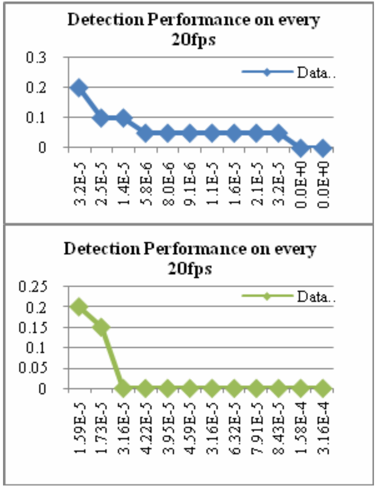
According to Dalal & Triggs paper [4], the Detection Error Tradeoff (DET) curves on a log-log scale is used to quantify detector performance using the miss rate and the false positive per window (FPPW). It is defined as using Equation 3,

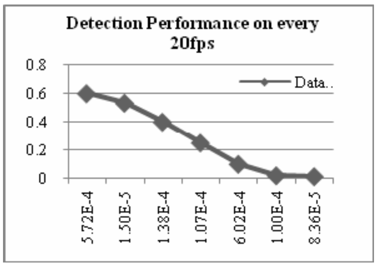
$$DET = \frac{1 - Recall\ Rate}{False\ Positive\ Per\ Window\ (FPPW)} \quad (3)$$

Where

$$Recall\ Rate = \frac{Total\ of\ True\ Positive\ regions}{Total\ of\ labeled\ Positive\ regions} \quad (4)$$

$$FPPW = \frac{False\ Positive\ Rate}{Number\ of\ Detection\ Window} \quad (5)$$





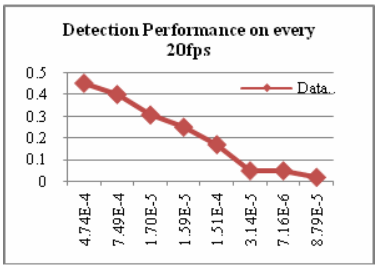


Figure 5. Figure shows the detection performance on every 20fps for Dataset A, Dataset B, Dataset C and Dataset D respectively.

V. Conclusion And Future Works

A. *Conclusion*

In this paper, we present a human detection system using motion extraction and Histogram of Oriented Gradients features (HoG). Implementation of motion detection enables our system to rapidly segment the motion region and have a more focusing observation on the moving region. The HoG features enable our system to classify the motion region as human or non-human. The highest detection rate achieved by our system is 95%. Regarding the lowest detection rate of 72%, it is due to the insufficiency of the stage of image preprocessing algorithm to normalize the contrast of the images. The result shows that our system is able to detect human with the average detection rate of 80% above.

B. *Future Work*

For future works, gamma correction can be implemented into our system in order to provide a better contrast of images for the further processing purpose. Besides, Local Binary Pattern approach can be employed to overcome the occlusion problems. In the end, we would like to emphasize on building our human detection system successfully using human body ratio estimation and histogram of oriented gradients.

References

[1] Biswaroop Palit, Rakesh Nigam and Keren Perlmutter. Spectral Face Clustering. ICCV, 2009.

[2] C. H. Lampert, M. B. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient.

[3] W. T. Ho and Y. H. Tay, "On Detecting Spatially Similar and Dissimilar Objects Using AdaBoost", Proceedings International Symposium on Information Technology 2008, 2, Page(s):899-903.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. CVPR, 2005.

[5] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In CVPR, pages 1491–1498, 2006. subwindow search. In CVPR, 2008.

[6] Hui-Xing Jia and Yu-Jin Zhang. Fast Human Detection by Boosting Histogram of Oriented Gradients. In ICIG 2007.

[7] Xiaoyu Wang, Tony X.Han and Shuicheng Yan. An HOG-LBP Human Detector with Partial Occlusion Handling. In ICCV 2009.

[8] W. Zhang, G. Zelinsky, and D. Samaras. Real-time accurate object detection using multiple resolutions. In ICCV, 2007.

[9] William Robson Schwartz, Aniruddha Kembhavi, David Harwood, and Larry S.Davis. Human Detection Using Partial Least Squares Analysis. In ICCV 2009.

[10] "Daimler Pedestrian Classification Benchmark Dataset," 2006, Dataset available at: http://www.gavrila.net/Research/Pedestrian_Detection/Daimler_Pedestrian_Benchmarks/Daimler_Pedestrian_Class__Benc/daimler_pedestrian_class__benc.html

[11] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. International Conference on Computer Vision (ICCV), 2003.

[12] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. In CVPR, pages 1–8, 2007.

[13] F. Porikli. Integral histogram: A fast way to extract histograms in Cartesian spaces. CVPR, 2005.