Presented at the DLSU Research Congress 2014
De La Salle University, Manila, Philippines
March 6-8, 2014

Application of Histogram of Oriented Gradient in Person Detection from Aerial Images

Alghie Marie Garcia*, Ma. Andrea Rufino, Louie Carlo Sangalang, John Arcy Teodoro, And Joel Ilao†
Computer Technology Department, College of Computer Studies
De La Salle University, Manila, Philippines
*alghie_garcia@dlsu.edu.ph, †joel.ilao@delasalle.ph

Abstract: With the steady increase in the frequency of occurrences and intensity of natural calamities in recent years, the use of technology for disaster prevention and mitigation is becoming more prominent. Aerial surveillance is one way of monitoring the progression of events in disaster-stricken areas, one important aspect of which is conducting a search for survivors who are in immediate need of rescue and relief. Computer vision algorithms can be used to aid in the detection, recognition, and tracking of humans in aerial videos recorded using unmanned aerial vehicles (UAVs). In this paper, Histogram of Oriented Gradients (HOG) is used for the human detection algorithm. This technique uses feature descriptors that generalize objects under different conditions (i.e. perspective, pose and illumination), in order to make the classification task easier. A Support Vector Machine (SVM) is employed for the image classification task. The developed system's performance is presented using confusion matrix, recall, precision, and F-score metrics.

Key Words: HOG; SVM; person detection; aerial images; image processing

1. INTRODUCTION

Automation has proven to be very helpful in accomplishing tasks that were otherwise done manually, more efficiently. When a major disaster strikes, rescuers need to locate survivors as quickly as possible in order to prevent fatality; this underscores the importance of using technology that can aid in the search process. A number of techniques have been explored in the past with the aim of improving systems used for person detection. These studies, however, are fewer and limited in performance, compared with studies on vehicle tracking systems. The reason for this is that humans, compared to vehicles, have movable limbs and parts that may not be visible from a given camera perspective, especially one taken with a high altitude observation platform. Furthermore, a human body is deformable and will appear in many different forms, causing brute force search methods to generate many false detections (Reilly et al., 2010).

A number of researches have been conducted on human detection systems. Jia & Zhang (2007) were able to develop a real-time human detection system by integrating the work of Viola and Jones (2001) and Histogram of Oriented Gradients (HOG) features. By substituting the Haar features with the HOG features, the system keeps the speed advantage of Viola and Jones' object detection framework, as well as the discriminative power of HOG features on

human detection. Experiments demonstrated that the system achieves a better accuracy at nearly the same speed as the original Haar features for human detection. Chua et al. (2012) presented a pedestrian detection system that uses Histogram of Oriented Gradients (HOG) as the feature descriptor, and Adaboost and Linear Support Vector Machines (SVM) as classifiers. The entire system was tested and evaluated in publicly available databases, such as the INRIA database which consists of 170 images and 3007 relevant pedestrian photos, and personally acquired videos. Experiments showed that the system is 20% faster compared to OpenCV's default detector (Chua et al., 2012).

The occurrence of the gradient orientation of an image is counted in HOG as a feature descriptor. An image is divided into cells. For each cell, the histogram of gradient directions for member pixels is built. Neighbouring cells are further grouped in blocks. For each block, the histogram of gradients of all member cells is also computed. Normalization of gradient vectors then happens by dividing them with their respective magnitudes, and the combination of the histograms would then be considered as descriptions for the detection.

This study aims to create a system that can identify the objects detected from aerial images as persons or not. This is for easier tracking of different people in different scenes for situations such as rescue operations.

The paper is organized as follows: Section 2 will describe the methodology used for person detection in aerial images. Section 3 presents the results of performance tests, while Section 4 contains the study's conclusion and suggestions for future work.

2. METHODOLOGY

2.1. Video Acquisition

Video Acquisition will start by using the obtained aerial video as the input. The video may be in .mp4 or .avi format. Frame extraction is used to convert the videos into a sequence of images. Fig. 1 shows the flowchart of the Video Acquisition module.

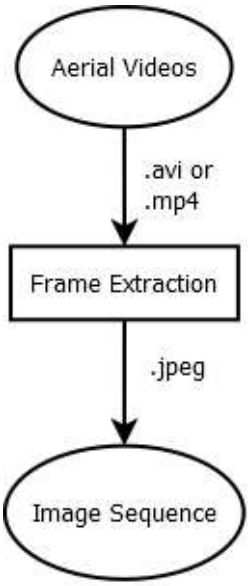

Fig.1. Image Acquisition

2.2. Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG) is used in evaluating the well-normalized local histograms of image gradient orientations in a dense grid. The appearance and shape of an object can often be characterized rather well by the distribution of its local gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. HOG is implemented by dividing the image window into small spatial regions or cells. The HOG features are computed using the following steps:

Gradient Computation

The colour or intensity data of the image is filtered using the kernels [-1,0,1] and $[-1,0,1]^T$. The result of the filtering can be used to compute the magnitude and orientation of every pixel.

Orientation Binning

The image is divided into non-overlapping cells with equal dimensions. The histogram of

orientations is computed for each cell. The magnitude of the gradient determines the weight of the vote of each pixel. The shape of the cells can either be rectangular or radial. The histogram bins are evenly spread over 0 to 180 degrees if the gradient is unsigned and 0 to 360 if the gradient is signed. The size of the cell used is 8x8 pixels. The size of the block is 2x2 cells.

Descriptor Blocks

In order to locally normalize the gradient information, cells are grouped into overlapping blocks. The blocks used in computation consist of 2x2 cells while a cell consists of 8x8 pixels. The blocks also have a 50% overlap. Even though the cells will appear multiple times in the final descriptor, it will be normalized by a different set of neighbouring cells in the next step.

Block Normalization

A normalization scheme is applied within each block. The histograms of four cells in one block are concatenated into a vector with 36 components[1]. This vector is normalized by dividing it by its magnitude. Since contrast changes are more likely to occur in smaller regions of an image, breaking the image into blocks and normalizing those blocks one by one is done in order to make the image invariant to contrast changes more, rather than normalizing over the entire image.

A feature vector of each cell's histogram entries is formed after performing the four successive steps.

2.3. Support Vector Machine

SVM uses "training sets" as its reference in classifying persons in an image. A training set is a set of images that are already labelled as positive or negative in containing a person in them. The HOG features from the training sets are used to train the SVM for person detection. When the test sets are inputted, its HOG features are now compared to the features in the trained SVM to determine if the corresponding image contains a person or not. 220 images are used for the training set, with 100 images labelled as positive (i.e. contains a person) and 120 labelled as negative (i.e. does not contain a person).

2.4 Performance testing

In testing the system's performance, recall, precision, and F-score are used. Precision and recall are measures used in retrieving information to measure the performance of information retrieval system. These measures were intended originally for set retrieval, but current researches mostly assume a ranked retrieval model where the search returns results in order of their estimated likelihood of relevance. The traditional F-score is the harmonic mean of precision and recall. Refer to Fig. 2 for the system flowchart.

[1]Each of the four histograms have nine (9) bins.

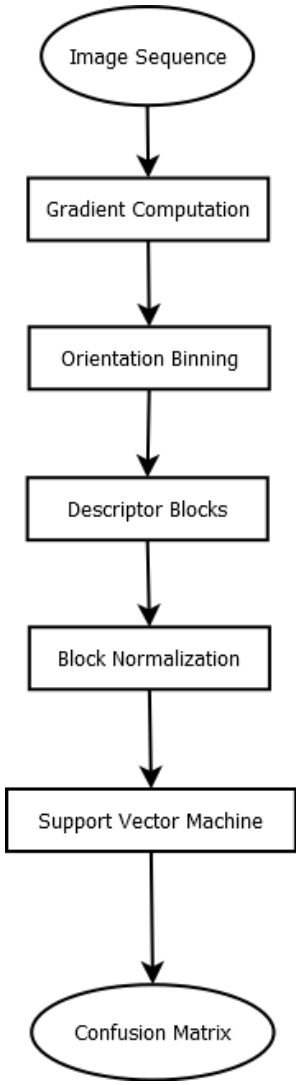

Fig. 2. System Flowchart

3. RESULTS AND DISCUSSION

An image database was created for the training and testing of the system. The images were taken from videos obtained from MF Mediahouse (Aerial videos, n.d.) using frame extraction followed by manual cropping. Each image is a 64x128 jpeg file containing either a person or any other object of interest. The images were manually cropped from the frames extracted from the video.

The SVM was trained with 120 images labelled as negative and 100 labelled as positive. It was tested to classify 100 images, with 50 images containing a person and are predicted to be positives. The other 50 images do not contain a person and are expected to be classified as negatives. Table 1 presents the actual results of the detection process while Table 2 shows the computed recall, precision and F-score based on these results.

Fig. 3. Sample Images of True Positives

Fig. 4. Sample Images of True Negatives

Table 1. Results of HOG and SVM

		Predicted	
		Positive	Negative
Actual	Positive	39	11
	Negative	5	45

Table 2. Recall, Precision, and F-score

Performance Metric	Value
Recall	0.78
Precision	0.89
F-score	0.83

The system showed satisfactory results as 78% of the positives images were classified as true positives and 22% as false positives. On the other hand, 90% of the negative images were classified as true negatives and 10% as false negatives.

Aside from noise being the main contributor of errors in the classification process, lack of training images is another factor that limits the performance of classifiers that use the HOG descriptor. This is because HOG is not a rotation-invariant feature descriptor, and does require multiple shots and orientations of different scenarios in order to classify correctly. In both scenarios, there are but a few training images that have similar orientation in relation to the false negatives and false positives.

4. CONCLUSIONS

The Histogram of Oriented Gradients as an overall feature descriptor has proven to be suitable in detecting people in aerial images, since it can describe an object without the need to detect smaller, individual parts of a person (i.e. the face), which may not be visible in a given image frame. Additionally, the HOG features of an image are not affected by varying lighting conditions.

In this paper, an implementation of the Histogram of Oriented Gradients in aerial images has been described. A dataset was built for training an SVM classifier, which tested fairly using an image test set.

Possible extensions to this study include expanding the dataset used by the SVM with additional images. The amount of errors in the detection process can be further reduced by adding more training images whose orientations are similar to the images that resulted to false positives and false negatives. The use of pre-processing algorithms to eliminate or reduce the amount of noise in the images will also improve the detection process.

5. REFERENCES

Aerial videos. (n.d.). Retrieved from http://www.mfmediahouseproduction.com/mfmediahouseproduction/AERIAL_VIDEOS.html

Brehar, R., & Nedevschi, S. (2011, August 25-27). A comparative study of pedestrian detection

methods using classical Haar and HoG features versus bag of words model computed from Haar and HoG features. IEEE. doi:10.1109/ICCP.2011.6047884

Chua, A., Dadios, E., Hilado, S., Gan Lim, L., Marfori, I. & Sybingco, E. (2012). Vision based pedestrian detection using Histogram of Oriented Gradients, Adaboost& Linear Support Vector Machines. IEEE Xplore. doi: 10.1109/TENCON.2012.6412236

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. IEEE. doi:10.1109/CVPR.2005.177

Jia, H. & Zhang, Y. (2007).Fast Human Detection by Boosting Histograms of Oriented Gradients.doi: 10.1109/ICIG.2007.53

Reilly, V., Solmaz, B. & Shah, M. (2010).Geometric Constraints for Human Detection in Aerial Imagery. Computer Vision – ECCV 2010 6316, 252-265. doi: 10.1007/978-3-642-15567-3_19