HOG APPROACH FOR HUMAN NOVELTY ACTIVITY DETECTION

Abdul-Lateef Yussiff*, Suet-Peng Yong, Baharum B. Baharudin, Muhammad Azeem Abbas[2]
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, 31750, Tronoh, Perak, Malaysia
[2]University Institute of Information Technology, PMAS-Arid Agriculture University Rawalpindi, Pakistan.
*Corresponding author's e-mails: azeem.abbas@uaar.edu.pk, ayussiff@gmail.com,

ABSTRACT: *Surveillance system is one of the most active research areas in the computer vision, there has not been much improvement in automating the surveillances due to human behavior being versatile and it is not easy to capture every single human behaviour abnormalities. We have addressed the most fundamental task of every computer vision project which is detection human in images and surveillance. We adopted a sliding window technique based on Histogram of Oriented Gradient (HOG) as features in detecting human in images. Our Method achieves 56.41% F-score and 88.71% precision when applied to Bank security image dataset.*

Key words: Human Activity Detection; Human Detection; Novelty Detection; Histogram of Oriented Gradient (HOG); Bank Security.

INTRODUCTION

Recent series of Automatic Teller Machine (ATM) heist in Malaysia, especially in Kuala Lumpur (KL) and its environ whereby about 1.17 million of Malaysian Ringgits was stolen [1] has called for review of the current surveillances' operations in Malaysia. In Banks, automated bank machines are primary targets for criminal acts [2]. The identified weakness of the surveillances at the ATM scene is common to many other surveillance systems; firstly, the current surveillances are passive type that can only be used for aftermath forensic analysis. Secondly the manpower required for monitoring the surveillances is very expensive, thus one person need to monitor several surveillance cameras simultaneously. Study shows that human visual attention drops significantly below acceptance levels within 20 minutes, even when experts are assigned to the task of visual monitoring of the surveillance system[2, 3]. Furthermore, the probability of immediate reaction to an event captured by a current surveillance camera network is estimated at 1 out of 1,000 [2]. To overcome these limitations, it will be helpful to have a real-time automated surveillance system which can screen the video streams and report any potential abuse during the operations.

More recently, government organizations, schools and businesses, are using video surveillance as a tool to enforce security. The recent widespread of high quality but cheap surveillance cameras and the availability broadband wireless networks, has made installation of group of cameras to enforce security become technically and economically realistic. Video surveillance's effectiveness and response to an event is mainly determined not by the technological capabilities of the device but by the alertness of the operator observing the camera system [4].

The questions are; how do we keep our financial cash points (ATM) much more secured? Is it possible to model object activities in video surveillance? One possible solution is to use novelty detection technique of human activities in surveillance video images i.e., to categorize images in terms of acceptable and allowable behaviors or abnormal usage of the ATMs. This technique is well suited to problems in which there are many examples of 'normal' data but lack or scarcity of 'abnormal' data or behaviors [5].

Using novelty technique in video surveillance analysis for detection of human activities is a pipeline of three tasks; the first and is the detection of interesting objects, then tracking of the detected moving objects from one frame of video to another, and finally is the high level semantic analysis of the object tracks to recognize behaviour. Detecting object in computer vision is a fundamental and a necessary task hence deserves some attention in the field. An efficient object detector has a ripple effect on all other tasks.

This paper focuses on human detection in video surveillances for an efficient, real time human novelty detectors using Histogram of Oriented Gradients. Related work is discussed in section 2, material and method is discussed in section 3, section 4 is the discussion and results and finally, conclusions in section 5.

RELATED WORKS

Novelty detection (also known as one-class classification, anomaly detection or outlier detection) classifies input test data as 'normal' or 'abnormal' with respect to the model of normality [5, 6]. Security and surveillance-based industries are looking for ways to detect novel events in very large streams of increasingly non relevant video data [7]. Diehl and Hampshire looked at novel event that occurs within a video and classify the new events based on previously labeled events. Stauffer at al. focused on real time event detection by learning the general patterns of activity that can be found in a scene [8]. Owens et al. was looking at detection of pedestrian unusual behavior in video-based surveillance system using the hierarchical neural networks [9]. Hearing et al., tracked and detected event in wildlife hunt videos [10]. Yong et al. used novelty detection technique coupled with scene semantic to identify abnormalities or novel scenes in the wildlife images [11]. Wu [12] and Banerjee et al. [13] described how to detect human body and motion detection as one of the steps of building a bank intelligent system.

Earlier research works on people detection focused mainly on background subtraction [14-17]. High sensitivity to background changes and illumination, and unsuitability to high density of persons are among the drawbacks of this approach. Background subtraction techniques generally determined foreground object from the video and then group it into categories like human and non-human depending on color (skin color), contour, shape, or motion and others. There have been paradigm shifts whereby, greater percentage of current methods are based on machine learning which uses discriminative classifier on images [18].

Among the sliding window detection technique, the work of Papgeorgiou and Poggio [19] adopted Haar wavelets representation coupled with a polynomial SVM as the machine learning classifier. As described by the authors, an image is mapped from pixels space to an over-complete dictionary of Haar features which provides definitive features of the pattern which is capable of expressing the class structure of the object of interest. Gavrila and Philomin [20] used chamfer distance to compare edge images to an exemplar dataset. Viola et al. [21] built on the Haar-like wavelets in order to handle space time information for human detection. The authors combined appearance and motion features using Cascade of Adaboost of weak learner classifiers to select optimized features that have higher performance contribution among all the features.

The other side of human detection approach using machine learning detects each part separately and concludes detection of human if some or all of its parts are presented in a geometrically feasible configuration. Corvee and Bremond [22] use hierarchical tree of HOG descriptors coupled with sliding window of 48 by 96 to identify each individual human part and combination of body parts to handle occlusion cases. The authors claimed a performance improvement over the HOG descriptor alone. Ioffe and Forsyth [23] model parts as projections of straight cylinders and bars, then propose efficient technique to incrementally aggregate these segments into a complete body based on the probability of likelihood. Mikolajczyket al. [24] represent parts as co-occurrences of local orientation features that capture the parts appearance as a spatial layout.

One of the first early methods with promising good performances was the cascade of Haar-like features proposed by Viola-Jones [25]. The technique adopted simple Haar wavelet filters as feature and a cascade of weak learner Adaboost classifiers to achieve near real time result of detecting human in images. The number of features obtained from this process was much higher than using the intensities features making the computation time to be longer during training, hence to boost the speed of detection, the authors used an integral image to compute Haar feature. Although, the technique was successful for face detection; however, the performance could not be replicated for the person detection in images. The Histogram of Oriented Gradient feature, proposed by Dalal and Triggs [26], proved very effective for person detection and it is one of the recent techniques researchers are employing in pedestrian detection.

MATERIALS AND METHODS

We crawled 376 images of people using ATM from Google images and other sites from the web. Fig. 1 shows the sample images in our database. Fig. 2 shows the properties of the average images. 63% of image dataset have only one person in the image. Figure 3 shows the people density per image.

Fig. 1: Sample Images in the database

File Size	= 7231
Format	= jpg
Width	= 201
Height	= 251
Bit per pixel	= 24
Color Type	= True color
Number of Samples	= 3
Coding Method	= Huffman
Coding Process	= Sequential

Fig. 2: Average Image properties

According to Sheikh et al. [27] "Accurate detection of objects in surveillance is a fundamental precursor to a stable tracking". The algorithm adopted for this work took inspiration from HOG [26] with modifications to detect human in ATM video surveillance environments.

The algorithm used is from Dalal and Triggs[26] Histogram of Oriented Gradient (HOG) for detecting object in images, which took inspiration from SIFT[28] and an Integral channels[29] for fast and efficient calculations of the gradient and orientations. The descriptors are used to classify object into human or non-human using a single sliding window approach, a kind of binary classification using SVM as classifier. The first step is the computation of gradients of an image, centered derivate mask [-1 0 1] and $[-1\ 0\ 1]^T$ were used on the image object in both x and y direction. Then the gradient magnitude and directions (orientations) are calculated. Each gradients are then classified into a one of the predefined nine bins orientation (0 180] degrees inclusive with respect to a localized blocks. A block contributes 36 features, so a single window with 105 blocks will contributes a total of 3780 histogram descriptors. The detection task is performed by scanning the given input image with a single window at various scales and positions, and classifying each window as human or non-human.

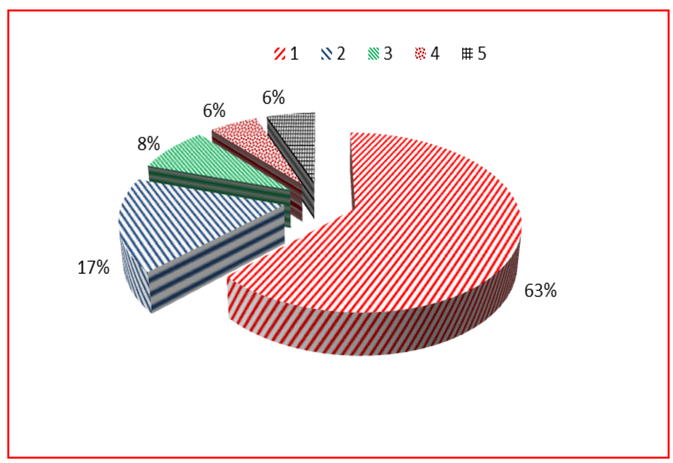

Fig. 3: Pie Chart of People Density per Image

RESULTS AND DISCUSSION

The designed algorithm was evaluated against the Bank security dataset. It was evaluated based on the number of persons that appear in the image. The experimental setup consists of 376 Google images with 533 persons in the image database. From Figure 3, 63% of the sampled images contain exactly one person in the image, 17% have two people in the image, 8% have 3 people in the image, 6% each have four people and more than four respectively. Figure 4 is the processing time it takes to process image. Figure 5 is the test input image and Figure 6 is the intermediate output result showing the gradient histogram of the image. Figure 7 shows the Human as detected by the framework.

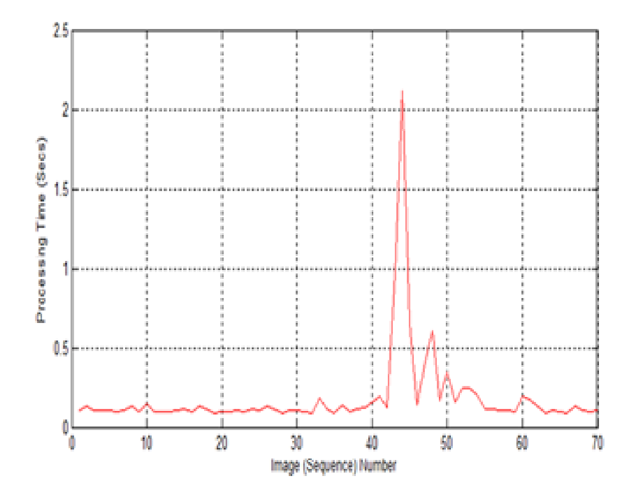

Fig. 4: Graph of Processing Time per Image

Fig. 5: Input Image to Histogram of Oriented Gradient

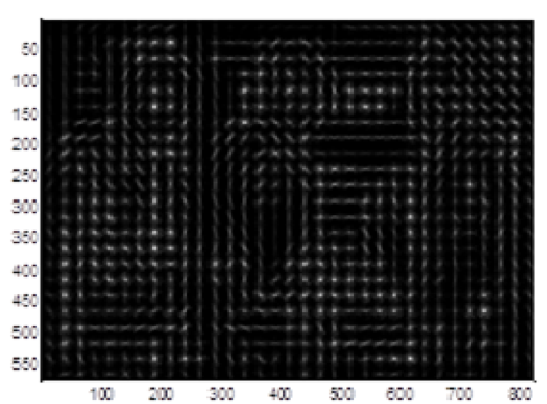

Fig. 6: Intermediate Output Graph of Histogram of Oriented Gradient

Fig. 7: Output from the HOG Detector

CONCLUSION

The algorithm implemented can detect human in both still images and dynamic images. Precision is 88.71, Recall of 41.35 and F-Score of 56.41 when applied to Bank Security image dataset. This model outperformed Viola- Jones human detection for this dataset. Our next goal is to increase the performance to the allowable limit that can further be used for the Novelty in Human behaviour detection and also semantic analysis of human behaviour.

REFERENCES

[1] Bernama, "Special team to probe ATM heist," in *The Star Online*, ed.

[2] V. Gouaillier and A. E. Fleurant, "Intelligent video surveillance: Promises and challenges," *Technological and commercial intelligence report, CRIM and Technôpole Defence and Security*, vol. **456**, p. 468, 2009.

[3] L. Ovsenik, A. Kolesarova, and J. Turan, "Video Surveillance Systems," *Acta Electrotechnica et Informatica*, vol. **10**, pp. 46-53, 2010.

[4] M. Shah, O. Javed, and K. Shafique, "Automated visual surveillance in realistic scenarios," *Multimedia, IEEE*, vol. **14**, pp. 30-39, 2007.

[5] D. A. Clifton, S. Hugueny, and L. Tarassenko, "Novelty detection with multivariate extreme value statistics," *Journal of Signal Processing Systems*, vol. 65, pp. 371-389, 2011.

[6] J. Kang, M. Aurangzeb Ahmad, A. Teredesai, and R. Gaborski, "Cognitively Motivated Novelty Detection in Video Data Streams," *Multimedia data mining and knowledge discovery*, pp. 209-233, 2007.

[7] C. P. Diehl and J. Hampshire, "Real-time object classification and novelty detection for collaborative video surveillance," in *International Joint Conference on Neural Networks*, pp. 2620-2625,. 2002.

[8] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. **22**, pp. 747-757, 2000.

[9] J. Owens, A. Hunter, and E. Fletcher, "Novelty detection in video surveillance using hierarchical neural networks," *Artificial Neural Networks—ICANN 2002*, pp. 140-140, 2002.

[10] N. Haering, R. J. Qian, and M. I. Sezan, "A semantic event-detection approach and its application to detecting hunts in wildlife video," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. **10**, pp. 857-868, 2000.

[11] S. P. Yong, J. D. Deng, and M. K. Purvis, "Novelty detection in wildlife scenes through semantic context modelling," *Pattern Recognition,* 2012.

[12] Y. Wu, "Research on bank intelligent video image processing and monitoring control system based on OpenCV," pp. 211-214. 2009.

[13] P. Banerjee and S. Sengupta, "Human motion detection and tracking for video surveillance," in *National conference for Communication*, 2008.

[14] T. Horprasert, D. Harwood, and L. S. Davis, "A robust background subtraction and shadow detection," in *Proc. ACCV,* pp. 983-988, 2000.

[15] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, pp. 627-630,. 2000

[16] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground–background segmentation using codebook model," *Real-time imaging,* vol. **11**, pp. 172-185, 2005.

[17] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *Image Processing, IEEE Transactions on,* vol. **13**, pp. 1459-1472, 2004.

[18] D. M. Gavrila, "The visual analysis of human movement: A survey," *Computer vision and image understanding,* vol. **73**, pp. 82-98, 1999.

[19] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision,* vol. **38**, pp. 15-33, 2000.

[20] D. M. Gavrila and V. Philomin, "Real-time object detection for “smart” vehicles," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, pp. 87-93, 1999

[21] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 734-741,2003, .

[22] E. Corvee and F. Bremond, "Body parts detection for people tracking using trees of histogram of oriented gradient descriptors," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on,* pp. 469-475,. 2010,

[23] S. Ioffe and D. A. Forsyth, "Probabilistic methods for finding people," *International Journal of Computer Vision,* vol. **43**, pp. 45-68, 2001.

[24] K. Mikolajczyk, C. Schmid, and A. Zisserman, "Human detection based on a probabilistic assembly of robust part detectors," in *Computer Vision-ECCV 2004*, ed: Springer, pp. 69-82,. 2004.

[25] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. **1**: pp. I-511-I-518, 2001.

[26] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 886-893, 2005,

[27] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. **27**, pp. 1778-1792, 2005.

[28] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, pp. 1150-1157, 1999.

[29] P. Dollár, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *British Machine Vision Conference*, 2010.