Australian Journal of Basic and Applied Sciences, 7(8): 632-640, 2013
ISSN 1991-8178

People Detection Enrichment for Abnormal Human Activity Detection

Abdul-Lateef Yussiff, Suet-Peng Yong, Baharum B. Baharudin

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, 31750, Tronoh, Perak, Malaysia

Abstract: A vibrant branch of research in computer vision that has attracted a lot of attention for decades is the human activity understanding from video. A means for accurately locating humans in image or a video is a prerequisite to the process of understanding human activities or action. This work's focus is on investigating the use of people detectors for video surveillance in Financial Banks premises so that it can eventually be used for abnormal human activity detection. An integrated framework which is made up of histogram of oriented gradient descriptors and Haar integral features is proposed thus, it is a union of Full body detector and Upper body detector. The proposed framework gives an improvement over the state of the art when applied as a case study to bank security. The technique obtained an F-score of 65.83 and precision of 73.83 and recall of 59.40 percentage points.

Key words: Histogram of Oriented Gradient (HOG); people detection; video surveillance; Bank Security; Abnormal human Activities.

INTRODUCTION

The Human detection is attracting attention from the computer vision and its related research community not only because of its potential benefits but also due to the several promising applications that it is an integral part of them. Among several promising applications of people detection is People surveillance which has emerges as one of the vibrant research area in the last decade especially since the event of September 11, USA, London bombing, UK, Madrid bombing, Spain, just to mention a few. Surveillance has always been a critical component in guaranteeing security at banks, airport and correctional institutions. The aim of most video-based surveillance is usually to identify and monitor humans, for security purposes, in scenes such as airports, train stations, supermarkets, etc. The widespread of high quality but cheap surveillance cameras and the availability of broadband wireless networks, made installation of group of cameras to enforce security become technically and economically realistic. But video surveillance's success and response to an event is not determined by the technological capabilities of the equipment rather determined by the vigilance of the operator monitoring the surveillance system (Shah, Javed, & Shafique, 2007). Recently, Boston marathon bombing in the USA has further shifted the focus back to real time detection of abnormal human activities. Hence, the current trend of research in the computer vision field.

The abnormal human activities can be grouped into a stack of three critical tasks in video surveillance analysis; at the lower layer is the detection and classification of interesting objects, the middle layer implements tracking of the detected moving objects from one frame of video to another, and finally, the high level analysis of the tracked object to recognize human behaviour. According to Yao and Odobez (Yao & Odobez, 2008) "Detecting humans in images and videos is one of the important challenges in computer vision" Good abnormal activities detector is directly dependent on the performance at detection layer. Poor performance at the lower stages will have ripple effects on the higher stage; so much attention needs to be paid to this stage so that accurate and good performance will be produced.

The problem of human detection being an active area of research in computer vision can simply be stated as: given an image or video sequence, identify all objects that are humans. In spite of several research efforts, the current performance of detector in video or images is still far from the ideal that could be deployed under most realistic environments, due to the inherent difficulties related with the human body and the environmental condition in which human are found. The foremost difficulty in building a robust object detector is the amount of variation in images and videos.

To achieve an improve performance for human detection in Bank environment, people detection based on Histogram of Oriented Gradient (HOG) and upper body detection based on Haar wavelets. To the best of our knowledge little or no research has been conducted for human detection in a Bank environment. Our experiments show that integrated descriptors improve the performances of the detector. The rest of this paper is organized as follows. Related works is presented in Section 2. Section 3 presents the methodology, Section 4 described the experiment and results and lastly, the conclusion and future works in Section 5.

Corresponding Author: Abdul-Lateef Yussiff, Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, 31750, Tronoh, Perak, Malaysia

Related Works:

Earlier research works on people detection in video sequence focused mainly on background subtraction (Haritaoglu, Harwood, & Davis, 2000; Horprasert, Harwood, & Davis, 2000; Jabri, Duric, Wechsler, & Rosenfeld, 2000; Javed & Shah, 2006; Kim, Chalidabhongse, Harwood, & Davis, 2005; Li, Huang, Gu, & Tian, 2004; Wren, Azarbayejani, Darrell, & Pentland, 1997; Zhao & Nevatia, 2003) in order to reduce the search area in video sequence. In background subtraction, human/object is assumed to be moving. 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 in a pixel-wise manner. High sensitivity to background changes and illumination, and unsuitability to high density of persons, static camera assumptions, reference model requirement are among the drawbacks of this approach.

There have been shifts in recent years whereby, larger fraction of current methods are based on machine learning which uses discriminative classifier (SVM, Adaboost, Neural Networks, etc) on images(Gavrila, 1999). Research works on using machine learning in human detection can be classified into two different approaches as reported in (Gavrila, 1999) namely single window detection and part-based approach. Within each technique, different authors propose different descriptors and different classifiers to tackle the problem.

Under the umbrella of single detection windows technique, the work of Papgeorgiou and Poggio(Papageorgiou & Poggio, 2000) adopted Haar-like feature 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(Gavrila & Philomin, 1999) used chamfer distance to compare edge images to an exemplar dataset. Viola *et al*. (Viola, Jones, & Snow, 2005) built on the Haar-like wavelets in order to handle space time information for human detection.

The alternative side of human detection approach using machine learning detects each part separately and concludes detection of human if parts are presented in a geometrically feasible configuration. Corvee and Bremond (Corvee & Bremond, 2010) use hierarchical tree of HOG descriptors coupled with sliding window of 48×96 to identify each individual human part and combination of body parts to handle occlusion cases. Ioffe and Forsyth (Ioffe & Forsyth, 2001) 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. Mikolajczyk*et al*. (Mikolajczyk, Schmid, & Zisserman, 2004) represent parts as co-occurrences of local orientation features that capture the parts appearance as a spatial layout. Feature selection and training were done in Adaboost classifier.

One of the first such early methods with promising good performances was the cascade of Haar-like Features - robust basis functions set which denotes neighbouring intensity difference in the regions, proposed by Viola-Jones (Viola & Jones, 2001). The number of features obtained from this process was much higher making the computation time to be longer during training, hence to boost the speed of detection, the authors used an integral image (Viola & Jones, 2001) to compute Haar feature. The two important properties of Haar feature is that, the value of Haar function remains same for different scale of the image. Also, the calculation of Haar-feature using Integral Image remains the same for the different size of image. Although, the technique was successful for face detection with performance of 93.7% detection rate; however, the performance could not be replicated for the person detection in images.

The Histogram of Oriented Gradient feature, proposed by Dalal and Triggs (Dalal & Triggs, 2005), proved very effective for person detection and it is one of the recent techniques researchers are employing in pedestrian detection. The pioneer used HOG descriptors and linear SVM classifier to detect object in an image. HOG descriptor depicts the distribution of intensity of the gradient depending on its orientation based on the predetermined 9 bins orientations. The HOG descriptors took inspiration from the SIFT (Lowe, 1999), a scale invariant feature transformation for detecting images of various sizes. Wang and Lien (Wang & Lien, 2007) extended the HOG to obtain rotation invariant HOG in order to detect people with different sizes, orientations, and complex cluttered background. Another variant of Histogram of oriented gradient proposed by Piotr et, al. (Dollár, Belongie, & Perona, 2010) used integral channel feature, that is made up of sums over local rectangular regions of multiple image channels, obtained using linear and non-linear transformations of the input image to speed up the pedestrian detection in an image.

Covariance features (Tosato, Farenzena, Cristani, & Murino, 2010; Yao & Odobez, 2008) combined with background subtraction technique, was shown to be an effective and fast human detector with promising results in video surveillance. Yao and Odonbez's (Yao & Odobez, 2008) work uses a cascade of LogitBoost classifiers on features mapped from the Riemanian manifold of local region covariance matrices derived from input image features. Tuzel *et al*.(Tuzel, Porikli, & Meer, 2008) proposed a method for classifying points lying on a connected Riemannian manifold using the geometry of the space. The drawback of those methods is that it is time intensive to calculate the covariance matrices of features and also edge-based features is difficult to capture under low contrast conditions.

Sim *et al*. (Sim, Rajmadhan, & Ranganath, 2008) used the cascade of classifier similar to that Viola-Jones for head detection. The results from the first classifier serve as input to the second classifier, based on color bin images features. The main goal is to eliminate or reduce false positives of the Viola-Jones head detector. Although, the goal was achieved, but that also lower the detection rate as compared to that of Viola-jones technique.

Methodology:

The schematic diagram of framework is shown in Figure 1. The detection framework uses Histogram of Oriented Gradients as feature/descriptor set and Haar-like integral image features to locate humans in the video scenes. The framework inludes: pre-processing, detector framework, and the post-processing module. The development environment is Matlab. Matlab provides an extensive library package support in the Computer vision System Toolbox.

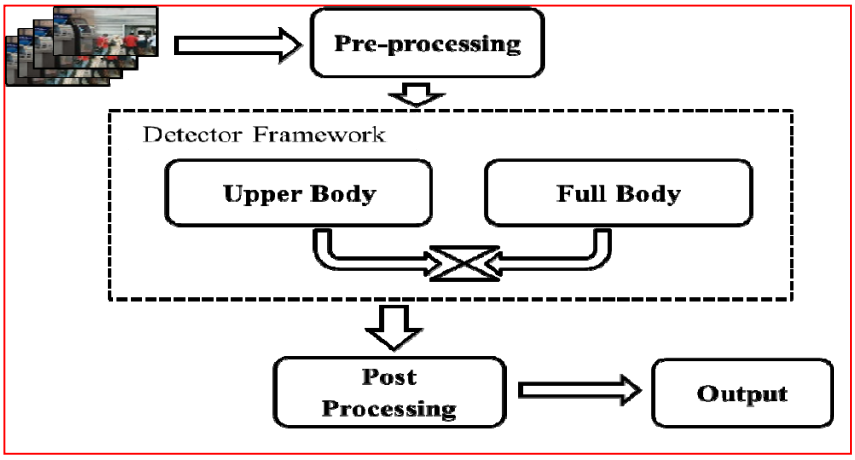


Fig. 1: System Overview.

Data Preparation and Image Source:

The case study for this paper is based on Bank Security. Videos were crawled from YouTube and the relevant videos of people using ATM are manually sorted out from the irrelevant video clips. The video clips have variable frame rate and resolution. The downloaded videos were converted and saved in .avi files using ffmpeg (Tomar, 2006), an open source software. Crawled YouTube videos lacked uniform background and this makes it difficult to apply background subtraction technique for the image segmentation process. Furthermore, due to limited number of downloaded videos, a total of 167 image data of people using ATM were downloaded from Google™ images.

Figure 2 shows the sample images in the database and Figure 3 is the pie chart of people density per image sample. From Figure 3 About 63% of the images in the database have only one person operating the ATM at a time, 17% of the sample images have 2 persons per image, 8% have three people, 6% have four persons in the image and about 6% have at least five people in the ATM surveillance environment. As it is expected from the result breakdown of people density per image, majority of people using ATM prefers to do that alone most probably because of security reason of protecting their bank Personal Identification Number (PIN).

Fig. 2: Sample Images of people using ATM in the Database.

Pre-Processing:

All video and images are initially converted to a standard format and sizes then Gamma correction are applied to the RGB input image. Images are resized to 300 × 250 pixels. The resized images are passed as input to the detector framework.

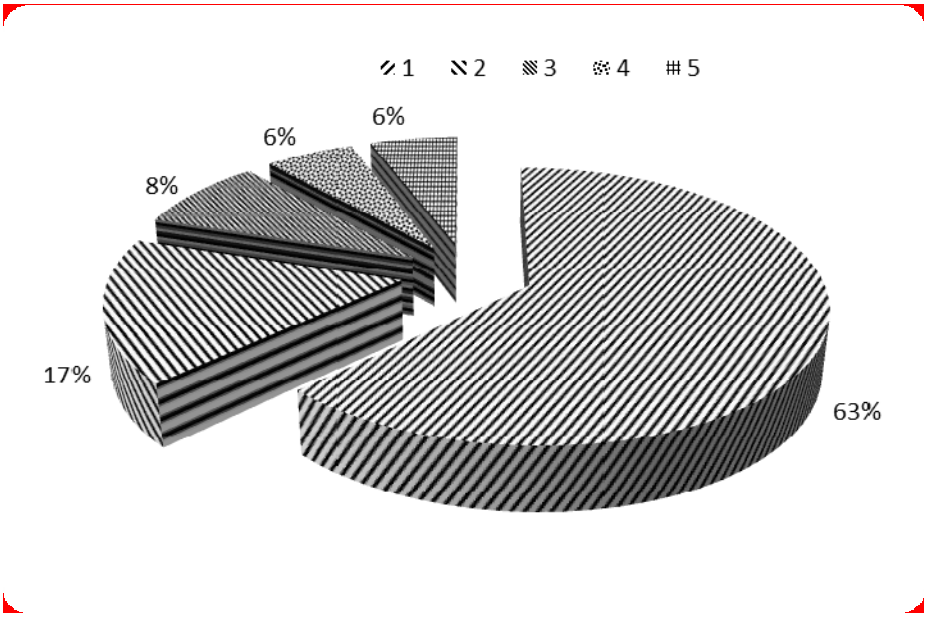


Fig. 1: Pie Chart Showing Average Number of People per Image.

Detector Framework:

The current state of the art detectors' performance on our dataset is not encouraging so a synergized detector is proposed. This framework integrates HOG detector and Viola-Jones face detector. The comparison of each detectors and the framework is discussed in the result section.

A. Full Body:

The algorithm used is from Dalal and Triggs(Dalal & Triggs, 2005) Histogram of Oriented Gradient (HOG) for detecting object in images, which took inspiration from SIFT(Lowe, 1999) and an Integral channels(Dollár *et al*., 2010) for fast and efficient calculations of the gradient and orientations. The algorithm asumes human is upright and fully visible in an image. First, is the feature extraction, extracting good descriptors for classification is the backbone of every machine learning classifiers. The descriptors are used to classify object into human or non-human using a single sliding window approach of size 64 × 128, a kind of binary classification using SVM as classifier.

The first step towards feature extraction is the computation of gradients for each channel of a RGB color image. To calculate the gradient and orientations, centered derivate mask [-1 0 1] and $[-1\ 0\ 1]^T$ were used on the image object in both x and y direction. Next, the gradient orientation range is mapped into 9 bins, with the value between 0 and 180 which are evenly seperated at equal intervals of 20 degrees.Then every pixels votes for the bin where its gradient orientation weighted by its gradient magnitude with respect to a localized block. The normalized histogram of the orientations are concatenated together as descriptors for each block. A block of an image is made up of 16 × 16 pixels. Each block is further divided into four cells consisting of 8 × 8 pixels. For each 8 × 8 pixel cells, a 9 HOG descriptors is obtained, thus a block will contributes a 36 HOG descriptors. The default scaning window of 64 × 128 have 105 blocks (7 blocks across the row and 15 blocks down the column). The total sum of features from the detection scanning window is 3780. The detection window depicts the area location of which can be found in the given input image. 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. Figure 4 depicts the sample input image and its graph visualization of the descriptors. The descriptor is what is used for both training and detection of human in an image.

a) Input Image b) HOG Graph Visualization

Fig. 4: (a) Input RGB image and (b) Visualization graph of the Histogram of Orientated (HOG) descriptors of the input image.

B. Upper Body:

Upper body, the head inclusive is the most visible part of the human body even in the occluded scenes. There have been many attempts by researchers to detect pedestrians through head detection. The Viola and Jones' integral image feature(Viola & Jones, 2001) proposed for face detection algorithm was adapted for this module. Matlab™ provide an efficient and fast implementation of this work so with little modification it was adopted for this module. We used a scale factor to accommodate different sizes of scanning windows. In Viola-Jones algorithm, the area around the region of the eyes are assumed to be generally darker than other regions of the face. This technique of detecting upper body in images combines four key concepts, namely, simple rectangular features, called Haar-like features, Integral Image for fast feature detection, AdaBoost classifier for feature selection, and cascaded of AdaBoost classifier to combine many features while quickly rejecting non-upper body part of the image. So the feature is similar to an adjacent pair of square wave. i.e. one dark and the other is light in color, which is computed by finding the difference between the sum of pixels in the darker and lighter region, and also the difference must be greater than a predefined threshold. In order to extract Haar feature at every single image location and scales, An integral image; the sum of pixels above and left of the pixel at a given location *(i, j),* beginning from the top left pixel to the right, then down till all the area of the image are covered with the detection window.

Mathematically, if we let *p(i,j)* be the pixel value of an image at a given location *(i,j)*, and *s(i, j)* be the cumulative sum and *I(i,j)* be the integral image, then

$$I(i,j) = I(i-1,j) + s(i,j) \tag{1}$$

$$s(i,j) = s(i,j-1) + p(i,j) \tag{2}$$

This integral image can be obtained in constant time and this gives considerable speed advantage over all other features that have been proposed for detecting faces. AdaBoost was used in the selection of Haar features which are good discriminant features among numerous candidates features. AdaBoost perfectly combines several weak classifiers in a cascading fashion to make a strong classifier. The algorithm goes over the image many times to search for upper Body of different scales.

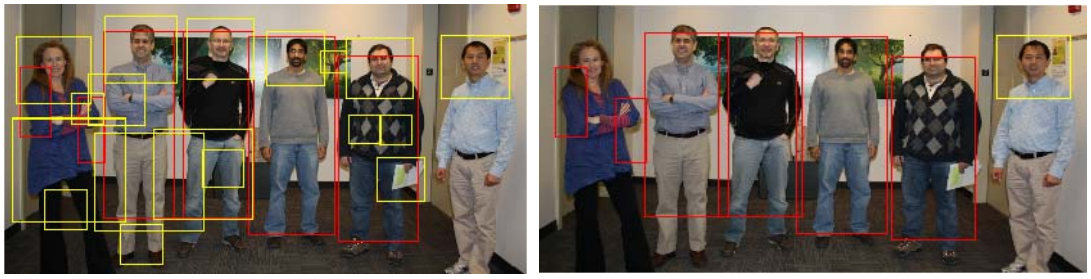
Post-Processing:

People detection with multiple detection algorithm produce multiple bounding boxes on the same object in an image, so the bounding boxes need to be fused together. The goal of this process is to improve the quality of the result and to eliminate multiple detection marker. Since two different detectors are used, there is higher chance of duplicate bounding boxes, hence Cover and Overlap algorithm is used to further processed the combined results to remove multiple bounding boxes on the same person. In this algorithm we compare each bounding boxes from one detector and check if it overlaps with any of the bounding boxes in the other detector. Then, we calculate area of overlaps between each pair of the bounding boxes result set. We next select one of the boxes with overlap and also satisfy the inequality equation 3.

Let R_{fb} be the set of bounding box obtained from the *Full Body* detector and R_{ub} be the *Upper Body* bounding box set. *Then*, $\forall x \in R_{fb}; \forall y \in R_{ub}$

$$Area(x \cap y) > \beta \tag{3}$$

β is a user defined value which represents the overlapping (tightness) of the bounding box of the two detectors. In the experiment, we set the value of β to be 0.5. The effect of post-processing can be seen by compairing the output (a) and (b) in figure 5. All redundant bounding boxes have been eliminated.



a) Actual output b) Post-processing result

Fig. 5: (a) Output detection result before post-processing and (b) Result after post-processing where by multiple detection markers have been eliminated.

RESULT AND DISCUSSION

The experiment was conducted using Intel i5 core duo CPU and each core frequency is 2.50 GHz. System memory of 10GB and NVDIA getforce 610M. MATLAB 2012a for Microsoft windows was used for the programming development. The Full body training dataset comes from INRIA dataset (Dalal & Triggs, 2005), which is made up of 1000 positive (human) images and 500 negatives (non-human) images. The testing dataset consists of 167 crawled Google images with scenes of people using ATM. The average image resolution is 300 × 250 pixel. The evaluation metric used shown in equations 4, 5 and 6 are based on standard performance measurement used in computer vision and image retrievals which are precision, recall and F score.

$$\prod = \frac{\sum_{k=1}^{n}\sum_{i=1}^{p} TP_{ki}}{\sum_{k=1}^{n}\sum_{i=1}^{p}\left(TP_{ki} + FP_{ki}\right)} \tag{4}$$

$$\rho = \frac{\sum_{k=1}^{n}\sum_{i=1}^{p} TP_{ki}}{\sum_{k=1}^{n}\sum_{i=1}^{p}\left(TP_{ki} + FN_{ki}\right)} \tag{5}$$

$$\Phi = 2\left(\frac{\Pi \times \rho}{\Pi + \rho}\right) \tag{6}$$

Where, k is the iterator for the image sequence in the database, *i* is the current image being processed, *TP* is the true positive, *FP and FN* are the false positive and false negative respectively as certified by the ground truth. *n* is the total number of images in the database. *π and ρ* are the precision and recall respectively, while Φ is the F-score measurement.

Figure 6 shows side by side processing time for all the detectors. The processing time for HOG Human detector is the lowest of the processing times. The F-score measurement of 56.41% precision of 88.71% and recall of 41.35%, as shown in Table I. Although the processing times are lower but more work needs to be done to improve the F-score.

Table I: Performance (%) of Detectors in terms of Precision, Recall and F-Measure.

Detector	Precision	Recall	F-Score
UB5	46.94	34.59	39.83
UB10	55.32	19.55	28.89
HOG Full Body	88.71	41.35	56.41
Integrated Framework	73.83	59.40	65.83

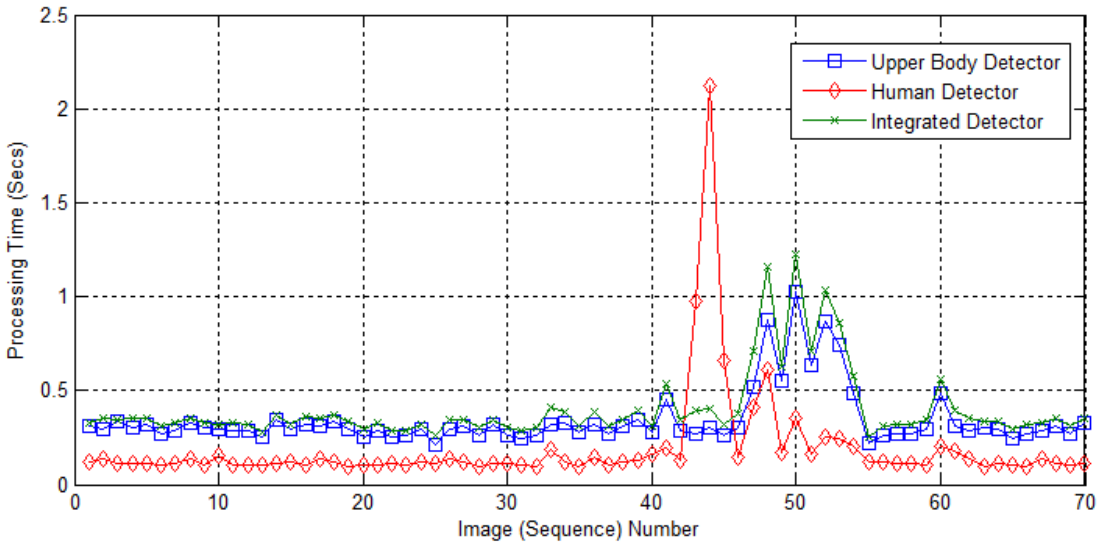


Fig. 2: Graph comparison of Processing Time for each Detectors.

Performance of Upper Body Detector:

The experiment was tested on the dataset with different scale factor settings while keeping all other parameters such as minimum window size constant. The detector scans the input image several times at different scales and finds the rectangular regions in the image classified as positives by the cascade of AdaBoost classifier then returns those regions as a sequence of rectangles. Reason for varying the scale factors of the detection windows is to find the optimum scale factor suitable for integration with HOG Human Detector to produce best performance for our dataset. The scale Factors that was used are; 5% and 10%. For example, a scale of 1.05 implies 5% increase in the detection window size in each pass over the image to allow human of various size to be detected. Minimum detection window size parameter was set to 20× 25. Refer to Figure 6 for processing times and Table I for the precision, recall and F-score values The 5% scale factor detector outperformed the 10% scale factor in terms of all the performance metrics.

Performance of Integrated Framework:

The processing time and performances are shown in Figure 6 and Table I respectively. The precision is 46.94, recall is 34.59 and F-Score of 39.83 was obtained for the upper body detector. Full body detector has precision of 88.71, recall of 41.35 and F-Score of 56.41, while the proposed technique has precision of 73.83, recall of 59.40 and F-Score of 65.83 percentage points. The F-score improves to 65.83% compared to other standalone detectors, hence an improved performance from the baseline performance for each detector. See Table I for the side by side comparison of all the detectors. Detection output of the integrated framework is shown in Figure 7 with the detected region marked in rectangular shape. The rectangular bounding boxes marked in red color depicts the output from the standalone body detector while the yellow marked rectangular bounding boxes depicts the people that was not detected by the full body detector but were detected by the upper body detector.

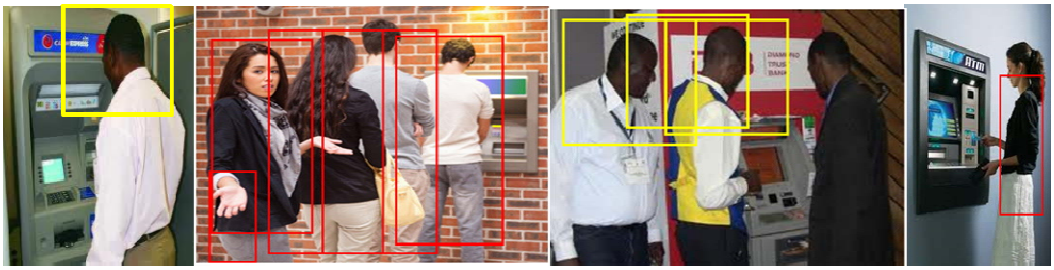


Fig. 3: Sample output of the integrated Detector.

Conclusion:

It is shown that, existing detection algorithms performance is not very good in detecting human in this type of images because most of the algorithms assumed human always faces camera. When the people are not facing camera, it becomes difficult to detect people. The main contribution of this framework is enrichment of human detection in images and video surveillance that integrates two state of the art works; HOG descriptors detection and Haar like integral image for upper body detection in our case studies — Bank Security. The proposed and implemented framework shows that result outperformed both the standalone HOG detector and standalone Viola-Jones Upper body detector. The framework can be used to locate human even in a crowded scenes. Our intention is to consider using parallel processing algorithm in the future work to see effect of parallelization on performance and processing time. With this algorithm, F measure has increased to about 65.83% from 56.41% for the Bank security dataset. The result obtained so far undoubted will serves as the foundation for the next phase of the research — people tracking in video surveillance.

REFERENCES

Corvee, Etienne, Bremond, Francois, 2010. *Body parts detection for people tracking using trees of histogram of oriented gradient descriptors.* Paper presented at the Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on.

Dalal, Navneet, Triggs, Bill, 2005. *Histograms of oriented gradients for human detection.* Paper presented at the Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.

Dollár, Piotr, Belongie, Serge, Perona, Pietro, 2010. *The fastest pedestrian detector in the west.* Paper presented at the British Machine Vision Conference.

Gavrila, Dariu M., 1999. The visual analysis of human movement: A survey. *Computer vision and image understanding,* 73(1): 82-98.

Gavrila, Dariu M., Philomin, Vasanth, 1999. *Real-time object detection for "smart" vehicles.* Paper presented at the Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on.

Haritaoglu, Ismail, Harwood, David, Davis, Larry S., 2000. W< sup> 4</sup>: real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 22(8): 809-830.

Horprasert, Thanarat, Harwood, David, Davis, Larry S., 2000. *A robust background subtraction and shadow detection.* Paper presented at the Proc. ACCV.

Ioffe, Sergey, Forsyth, David A., 2001. Probabilistic methods for finding people. *International Journal of Computer Vision,* 43(1): 45-68.

Jabri, Sumer, Duric, Zoran, Wechsler, Harry, Rosenfeld, Azriel, 2000. *Detection and location of people in video images using adaptive fusion of color and edge information.* Paper presented at the Pattern Recognition, 2000. Proceedings. 15th International Conference on.

Javed, Omar, Shah, Mubarak, 2006. Tracking and object classification for automated surveillance *Computer Vision—ECCV 2002* (pp. 343-357): Springer.

Kim, Kyungnam, Chalidabhongse, Thanarat H, Harwood, David, Davis, Larry, 2005. Real-time foreground–background segmentation using codebook model. *Real-time imaging,* 11(3): 172-185.

Li, Liyuan, Huang, Weimin, Gu, Irene Yu-Hua, Tian, Qi, 2004. Statistical modeling of complex backgrounds for foreground object detection. *Image Processing, IEEE Transactions on,* 13(11): 1459-1472.

Lowe, David G., 1999. *Object recognition from local scale-invariant features.* Paper presented at the Computer vision, 1999. The proceedings of the seventh IEEE international conference on.

Mikolajczyk, Krystian, Schmid, Cordelia, Zisserman, Andrew, 2004. Human detection based on a probabilistic assembly of robust part detectors *Computer Vision-ECCV 2004* (pp. 69-82): Springer.

Papageorgiou, Constantine, Poggio, Tomaso, 2000. A trainable system for object detection. *International Journal of Computer Vision,* 38(1): 15-33.

Shah, Mubarak, Javed, Omar, Shafique, Khurram, 2007. Automated visual surveillance in realistic scenarios. *Multimedia, IEEE,* 14(1): 30-39.

Sim, Chern-Horng, Rajmadhan, Ekambaram, Ranganath, Surendra, 2008. A Two-Step Approach for Detecting Individuals within Dense Crowds *Articulated Motion and Deformable Objects* (pp. 166-174): Springer.

Tomar, Suramya, 2006. Converting video formats with FFmpeg. *Linux Journal,* 146: 10.

Tosato, D., M. Farenzena, M. Cristani, V. Murino, 2010. *Part-based human detection on Riemannian manifolds.* Paper presented at the Image Processing (ICIP), 2010 17th IEEE International Conference on.

Tuzel, Oncel, Porikli, Fatih, Meer, Peter, 2008. Pedestrian detection via classification on riemannian manifolds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 30(10): 1713-1727.

Viola, Paul, Jones, Michael, 2001. *Rapid object detection using a boosted cascade of simple features.* Paper presented at the Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.

Viola, Paul, Jones, Michael, J., Snow, Daniel, 2005. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision,* 63(2): 153-161.

Wang, Chi-Chen Raxle, Lien, Jenn-Jier James, 2007. AdaBoost learning for human detection based on histograms of oriented gradients *Computer Vision–ACCV 2007* (pp: 885-895): Springer.

Wren, Christopher Richard, Azarbayejani, Ali, Darrell, Trevor, Pentland, Alex Paul, 1997. Pfinder: Real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 19(7): 780-785.

Yao, Jian, Odobez, Jean-Marc, 2008. *Fast human detection from videos using covariance features.* Paper presented at the The Eighth International Workshop on Visual Surveillance-VS2008.

Zhao, Tao, Nevatia, Ram, 2003. *Bayesian human segmentation in crowded situations.* Paper presented at the Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on.