A Novel Efficient Method for Covered Face Detection in ATM Video Surveillance System

Suvarna Nandyal[1]
[1]Professor, Department of Computer Science & Engineering
[1]Poojya Doddappa Appa College of Engineering, Kalaburagi, India

Sanjeevkumar Angadi[2]
[2]Assistant Professor, Department of Computer Science & Engineering
[2]Nutan College of Engineering & Research, Pune, India

Abstract—
Face identification in video is a difficult and fascinating topic, especially when it's used in Automated Teller Machines (ATM). Covering one's face with items such as a mask, scarf, or sunglasses is a popular illegal conduct in ATM robbery. As a result, reducing robberies and other crimes can be accomplished by deploying ATM security cameras to detect covered faces. In this paper, a new method is presented for detecting covered faces in ATM surveillance system. To achieve precise foreground, a new and expedient foreground extraction method is recommended. In this research work, we propose binarization with Otsu's Thresholding method to find the intensity value of pixel that subtracts background and foreground from a given input image. To determine whether the human face is covered or not, we then locate the face using the Histogram of Oriented Gradient (HOG) approach.The results of our experiments demonstrate that our algorithm achieves a high detection rate while maintaining a low false negative rate.

Keywords—Suspiciousbehavior,Face detection, Otsu's, HOG, Video Surveillance
DOI Number: 10.48047/Nq.2022.20.17.Nq880285 **Neuroquantology 2022; 20(17):2199-2207**

I. Introduction

Automatic Teller Machines (ATMs) are well known to have a significant impact on current economic activity.It offers a quick and simple solution for banks to perform cost-effective transactions with their customers. Unfortunately, it also offers a simple means for thieves to obtain illegal money. A customer will be required to enter their bank card information and a password when using an ATM. Additionally, each ATM has a surveillance system that records the consumer's face information for security purposes. However, when a fraudster uses an ATM to conduct an illegal transaction, they typically hide their faces to avoid having their face data captured by the monitoring system. Therefore, it won't fulfil the purpose of the surveillance system.

In the past, the usage of ATM services grew quickly because they provided clients with a more convenient option to withdraw money at any time (24*7 hours).The banking industry has adjusted to the rise of digital transactions. However, a lot of security issues plague the ATM service, which has an impact on the entire banking industry. Due to ATM crimes such card skimming, cash trapping, and shoulder surfing, secure financial transactions are not guaranteed. The rate of theft and robbery is rising every year, according to a new report. Due to a lack of security, frauds occur frequently at ATM centres. According to historical data, 5500 frauds were reported annually [1].These frauds, which are getting more and more sophisticated every day, are made possible by security flaws. During ATM crimes, thieves typically conceal their faces with accoutrements like masks, scarves, and sunglasses [2][3].With developments in computer vision techniques, surveillance cameras can be utilized to detect masked faces and provide an immediate warning signal during any ATM theft.

With CCTV surveillance applications, significant tracking and moving object detection attempts have been made in the past, but obstacles like poor video quality, lighting, and occlusion may make progress difficult. In the proposed work, such automatics covered face detection in CCTV footage at ATM Centre is implemented.

Two cascading steps—face detection and anomaly detection—can be used to decompose covered face detection in videos.There are numerous sophisticated facial detection algorithms. Some face detection techniques in images are introduced in [4].The four primary types of current face detection techniques include knowledge-based approaches, feature-based methods, template matching-based techniques, and appearance-based techniques.Each category has benefits and drawbacks of its own.Methods based on knowledge [5] Despite having a modest processing cost, [6] has a poor detection accuracy; Skin colour [7] and facial features [8] are examples of feature-based techniques that are simple to use and perform well, but they cannot be employed in the complex settings; When there are many templates, template-matching-based approaches [9] take a long time; appearance-based approaches [10] are more accurate but more difficult.Additionally, it has been noted that the [11] method of radial symmetry performs well when the symmetrical qualities of the eyes, nose, ears, and mouth are obvious.However, many current algorithms are ineffective at accurately spotting obscured faces.

Furthermore, the video presents difficulties.Face detection in a video stream should be real-time, so the method must first be efficient and cost-effective.Second, the detecting system is perplexed by the fact that the pose, size, and number of human faces can change continuously in a matter of seconds.Third, the background of video sequences is always changing.All during the day, the lighting changes, and anything could briefly blend into the background.

In this research, we offer a novel and successful two-step method for covered face detection to address the existing challenges. (1) Binarization with Otsu's thresholding method is proposed for calculating the intensity value of pixel;(2) Histogram of Oriented Gradient (HOG) is used for face detection based on bounding box.

The following is the layout of the paper: The second section summarizes related face detection techniques. The proposed methodology is summarized in Section III. In section IV, the implementation specifics are defined, accompanied by a conclusion and future work in section V.

II. LITERATURE SURVEY

Although there have been many studies on the use of covered face detection techniques in ATM surveillance, most of the techniques failed to consider the actual ATM surveillance captured scene (camera view, camera-user distance), as well as face covering accessories.

Sang Min Yoon et al., presented a methodology for covered face detection in ATM [12]. To effectively identify and minimize the dimension, we first trained the features of the normal faces and the occluded faces wearing sunglasses or masks using Principal Component Analysis and Support Vector Machines. In this paper, the training was carried out on 3200 normal face images with variations in illumination and expression, as well as the 2900 and 4500 partially obscured face images with 60*25 and 60*35 resolutions that wear sunglasses or masks. The investigations based on the Purdue University Face DB yield the 95.2% and 98.8% partially obstructed face detection ratio after face detection and the 2.5% and 0% false alarm ratio.

An occluded face detection technique based on the YCbCr elliptical model was proposed by Sun Hongxing et al. [13]. The connected area of the target faces was then picked in external ellipses using the CamShift target tracking technique.Results indicate that in terms of detection accuracy rate, false alarm rate, and detection and tracking speed, the algorithm achieves nearly perfect results.

Using an omni-directional vision sensor (ODVS) and computer vision technology, Yiping Tang et al. describe a novel intelligently monitoring system for the automated teller machine (ATM) [14]. By using the perspective algorithm, the omnidirectional image is unwrapped into several distinct key surveillance sections, and the faces of ATM users and the suspected actions of the voyeur who is behind them are detected using the Gaussian skin-color model of face detection and the Kalman filter of face tracking algorithms. According to the experimental findings, the system has the advantages of a broad detection range, high intelligence, high efficiency storing and searching, and high robustness. Thus, it offers a novel means of stopping financial crimes.

Facial fraud discrimination utilizing facial feature recognition and classification based on the AdaBoost and a neural network is proposed by Inho Choi et al. [15]. The two eyes, the lips, and the face are all detected by the suggested method using the AdaBoost detector. We utilize a neural network to classify detection findings as normal or abnormal eyes and lips.

An efficient face occlusion detection method is presented by Daw-Tung Lin et al. [16] to support security personnel in surveillance by offering both useful data for additional video indexing applications and crucial hints for criminal investigation.To combine the dividing blobs, a Straight-Line Fitting (MSLF) algorithm is developed.At a rate of up to 20 frames per second, the suggested detection system can detect objects with 100% and 96.15% accuracy for non-occlusive and occlusive detection, respectively.

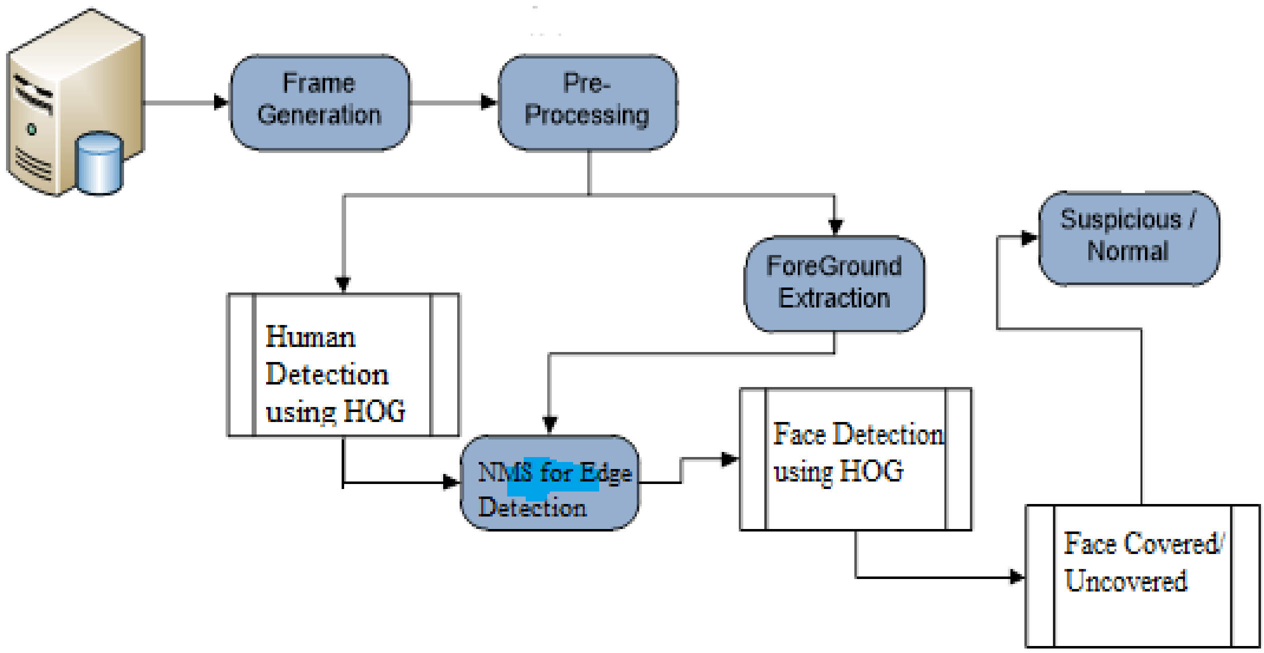
III. SYSTEM OVERVIEW

If any suspicious incident occurs inside ATM environment, the proposed system will use videos collected from surveillance cameras to track whether the human face is covered or not for analyzing suspicious event and send a message to the suitable authority.

A. Architecture of the Proposed System

The proposed architecture includes phases such as video capture, pre-processing, human identification, Binarization of image, and Face detection.Figure 1 depicts the overall layout of the system architecture.

Fig 1: Block Diagram of Covered Face Detection inside ATM



Following are the steps carried out during the detection of covered face of a human that can lead to further suspicious activities:

Step 1: **Background Model & Preprocessing:** The background image changes based on the design of required background model. Gray-Scale Conversion and Resize function is applied to each input frames from a selected video.Background subtraction is employed to identify foreground objects.The background image is removed from the selected frame image to obtain the objects visible in the foreground image.

Step 2: **Human Detection:** Histogram of Oriented Gradient (HOG) is used for human detection and Non-maximum Suppression (NMS) algorithm is used for human edge detection, the result of this stage will be helpful for human face detection inside ATM cabin. 2201

Step 3: **Binarization of Image:** Binarization is done on each input image using Otsu's Thresholding method, that returns a single intensity threshold of two separate classes, foreground, and background.

Step 4: **Face Detection:** Histogram of Oriented Gradient (HOG) method is used to classify whether the customer face is covered or not. This step's outcome will aid in classifying additional activity as suspicious or normal.

B. Binarization with Otsu's Thresholding:

To binarize an image based on the pixel intensities, image thresholding is utilized. Such a thresholding technique typically requires a threshold and a grayscale image as inputs [17]. It produces a binary image.When a pixel's intensity in the input image exceeds a threshold, the associated output pixel is labelled as white (foreground), and when it is less than or equivalent to the threshold, the corresponding pixel position is labelled as black (background).

Image thresholding is a pre-processing procedure that is utilised in numerous applications. In medical image processing, for instance, you might use it to locate a natural catastrophe in satellite images or detect a tumour in a mammography.Simple thresholding has the drawback that the threshold value must be manually specified. We can manually test a threshold's performance by attempting various settings, but this is time-consuming and may not work in practise.Therefore, we require a method for computing the threshold automatically. Auto thresholding is well-exemplified by the Nobuyuki Otsu technique, which bears his name.

In our research work, to process threshold value, an automatic global thresholding algorithm is used as follows:

1. Human Blob image is selected as input image
2. Binarization of image is done, to hold the pixel values
3. A threshold value T is considered, here T value is 4.
4. Replace image pixels into white in those regions, whose scale Factor (value=1.1) is greater than T.

The technique analyses the image histogram and segments the items by minimising the variance for each class.This method typically yields the correct outcomes for bimodal images.Two distinct peaks can be seen in the histogram of this image, each of which represents a different range of intensity values [17].

The fundamental strategy involves splitting the image histogram into two clusters using a threshold that is determined by minimising the weighted variance of these classes, which is represented by$\sigma_u^2(v)$

The complete computation equation is as follows:

$$\sigma_u^2(v) = u_1(v)\sigma_1^2(v) + u_2(v)\sigma_2^2(v) \quad (1)$$

where the probabilities of the two classes are expressed as $u_1(v)$, $u_2(v)$respectively, and are divided by a threshold v, whose value falls between 0 and 255, inclusive.

For each pixel value in the two distinct clusters C_1, C_2, the probability P is determined using the cluster probability functions written as follows:

$$u_1(v) = \Sigma_{j=1}^{t} P(j) \quad (2)$$

$$u_2(v) = \Sigma_{j=t+1}^{I} P(j) \quad (3)$$

Intensity function $f(x, y)$ with gray-level values can be used to portray the image, it should be emphasised.The number of pixels with a specific gray-level j is denoted by j.The image generally contains n pixels.As a result, the probability that gray-level j will occur is:

$$P(j) = \frac{n_j}{n} \quad (4)$$

C. Face Detection Method:

To detect and identify visual objects, histograms of oriented gradients are frequently employed in computer vision, pattern recognition, and image processing (i.e., faces).Because we require a strong feature set to discriminate and discover faces under challenging lighting backgrounds, wide range of poses, etc., we propose to employ HOG descriptors. These feature sets outperform the ones currently used for face detection.HOG is like shape context, the SIFT descriptor, and the edge orientation histogram.To enhance the effectiveness of the detector, they are calculated on a large grid of cells that overlap local contrast histogram normalisations of image gradient orientations [18].Because of the dispersion of local intensity gradients, this feature set performs exceptionally well for various shape-based object classes (such as face detection), even without precise information of the related gradient [19].

Count the occurrences of edge orientations in an image near vicinity before extracting HOG descriptors.This implies that the histogram of edge orientations is calculated for each of the image's small, connected areas, known as cells (for example, size 9).The histogram channels are dispersed over either $o° - 180°$ or $o° - 360°$based on whether the gradient is unsigned or signed.By accumulating a measure of local histogram energy over the connected regions, histogram counts are normalised to account for illumination. The results are then used to normalise all cells in the block (for example, size 2), and the combination of these histograms produces the HOG descriptor as shown in figure 2.

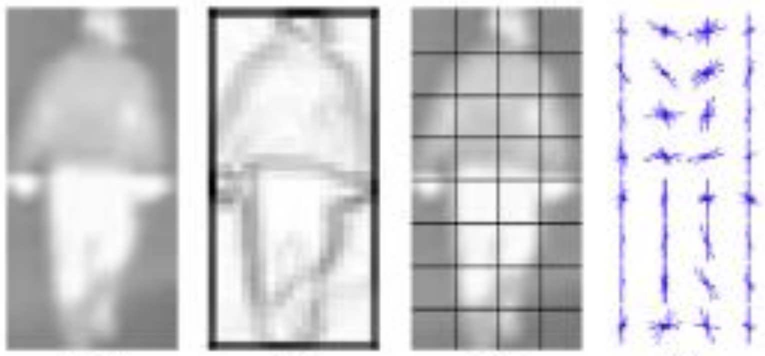

(a) (b) (c) (d)

Fig 2: Images from the several steps of creating a Histogram of Oriented Gradients feature vector. (a) The original image of a pedestrian, scaled to 20x40 pixels; (b) the gradient image; (c) the image segmented into cells of 5x5 pixels, resulting in 4x8 cells; and (d) the resulting HOG descriptor for the image, showing the gradient orientation histograms in each cell [20].

The Lowe's Scale Invariant Feature Transformation methodology [21] uses orientation histograms in conjunction with local spatial histogramming and normalising, and these two techniques have been employed extensively in various ways. Shape Context analyses the forms of the cell and block; at first, it only employed edge pixel counts without orientation histogramming. Figure 3 depicts the flow chart of face detection.

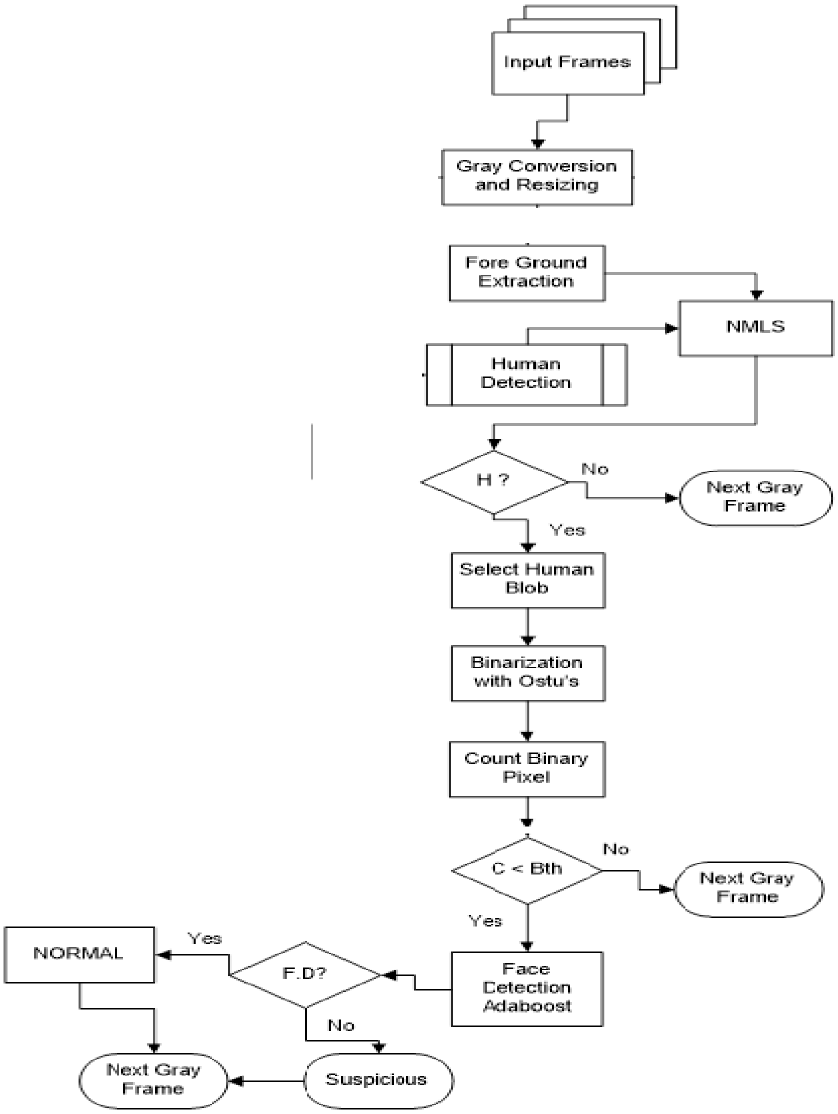


Fig 3: Covered face detection flow chart

IV. RESULTS AND DISCUSSION

The proposed technique is used on a database [22] of ATM surveillance-related activity.The required object is extracted from an image using a provided input using background subtraction [23].For tracking a covered or uncovered face in an input image, a human must be detected first.Human identification is done using the HOG and NMS algorithms [24].To locate important information in an image, an NMS mask is built for further phases.To calculate the intensity of a pixel based on threshold value, a binarization with Otsu's thresholding method is used. After human is successfully identified, human face is tracked to understand whether the face is covered or uncovered. For face detection, HOG algorithm is used. Based on result of face detection phase, an alert message is generated to higher authorities, if the face is found to be covered inside ATM cabin, which is considered as suspicious event.

A. Experimental results of face detection inside ATM cabin:

To produce the intended outcome for a given piece of video data, vector prediction and measurement are considered. Frame extraction begins by considering several factors, including the video format, frame rate, and length of an input video.Here, RGB24 and a frame rate of 15.016 seconds are chosen for the video format. There were 794 input images used to generate the images.

Figure 4 shows the results of the proposed approach by considering the video sample 1. Figure 4 (a)show the input image, Figure 4 (b) detected human and score of input image, Figure 4 (c) NMS mask image, Figure 4 (d) human detection, 4 (e) face detection, and Figure 4 (f) output image.

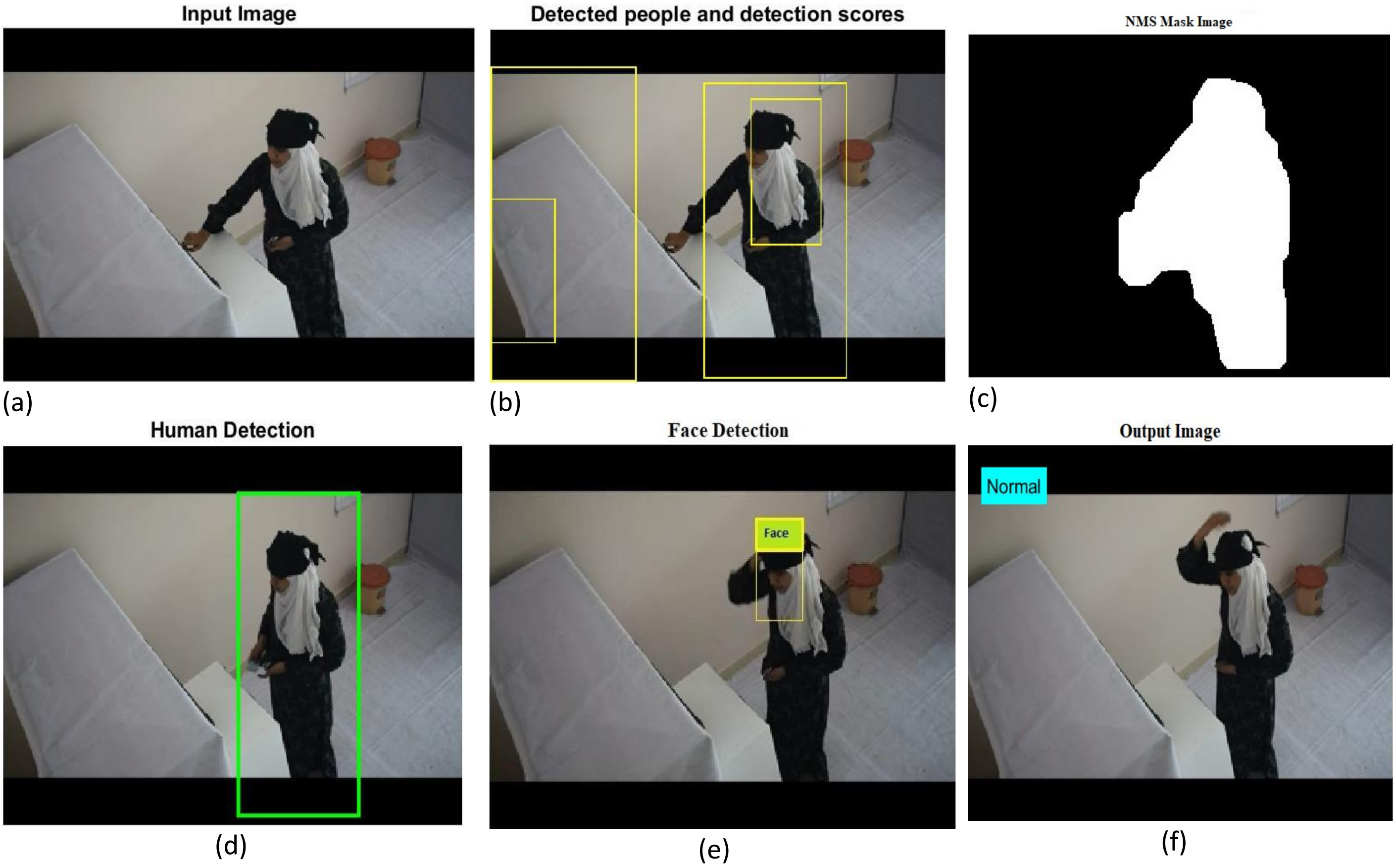
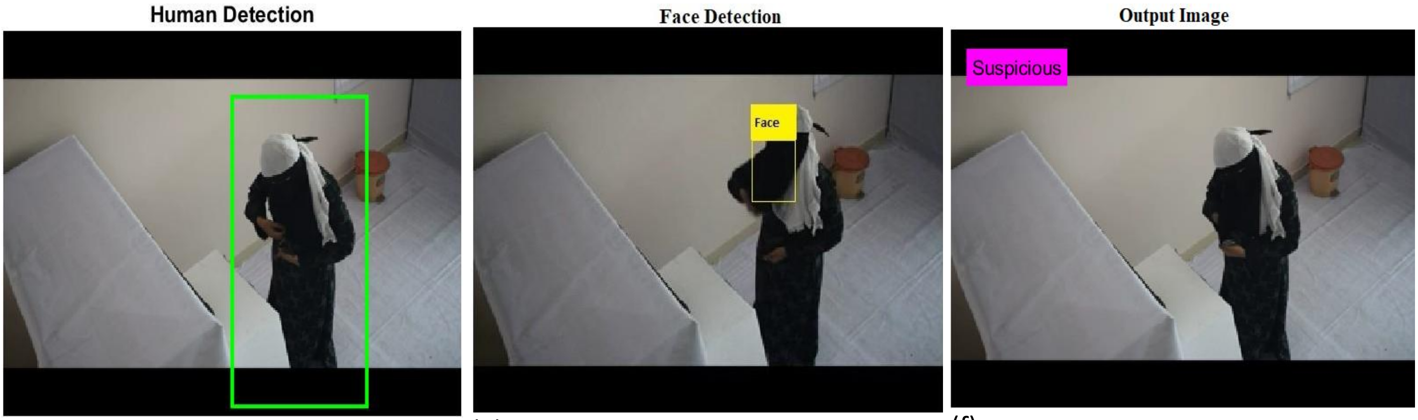


Fig. 4: Face Covered Recognition (a) input image (b) Detected customer and Score of input image, (c) NMS mask image, (d) human detection, (e) face detection, (f) output image (Normal)

Figure 5 shows the results of the proposed approach by considering the video sample 1. Figure 5 (a)show the input image, Figure 5 (b) detected human and score of input image, Figure 5 (c) NMS mask image, Figure 5 (d) human detection, 5 (e) face detection, and Figure 5 (f) output image.



(d) (e) (f)

Fig. 5: Face Covered Recognition (a) input image (b) Detected customer and Score of input image, (c) NMS mask image, (d) human detection, (e) face detection, (f) output image (suspicious)

The sequence of results shows detection of covered face inside ATM cabin. Based on number of test cases considered, the detection accuracy is good compared to the ground truth value. Table 1 shows the comparative analysis between proposed algorithm and ground truth value. Figure 6 represents the graphical analysis of table 1.

Table I: Comparative analysis based on Ground Truth Frames for Face Covered Detection

Test-Cases	**Ground Truth**	**Face Detection using HOG**
Test_1_Customer Face Covered _Video 1	458	412
Test_2_Customer Face Uncovered _Video 1	365	335
Test_3_Customer Face Covered _Video 2	725	708
Test_4_Customer Face Uncovered _Video 2	795	752

V.CONCLUSION

Many individuals in today's world are aware of the significance of CCTV footage, but they normally only use these recordings for forensic purposes after a crime or incident has occurred. The proposed technique has the benefit of preventing crime before it even begins. CCTV footage is being monitored and analyzed in real time. If the analysis's findings suggest that an unwanted event is probable, it will issue an order to the appropriate authorities to act.

The experimental result demonstrate that proposed method suggested in this paper performs well when it comes to covered face detection. The proposed method has produced improved accuracy when compared to the ground truth value while considering a variety of test cases for covered and uncovered faces in an ATM environment.

REFERENCES

[1] S.Menaga, Yamili.A, Rekha.P,Tamilarasi.R, "Internet of Things Based ATM Secure Monitoring", IJIRCCE, 2017, Vol. 5, Issue 3.

[2] Graczyk M. Masked men steal lobby ATMs from 5 Houston Marriotthotels. Daily Herald, Houston. 2017. http://www.daily herald.com/artic le/20171 214/news/31214 9823. Accessed 17 Dec 2017.

[3] Bromwich JE. A smash-and-grab heist in Pennsylvania: maskedmen steal an A.T.M. The New York Times, Pennsylvania. 2016.https ://www.nytim es.com/2016/12/01/us/a-smash -and-grab-heist-in-pennsylvania-masked-men-steal -an-atm.html. Accessed 20Dec 2016.

[4] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: Asurvey", IEEE Transactionson Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34–58, 2002.

[5] G. Yang and T. S. Huang, "Human face detection in a complexbackground," Pattern recognition, vol. 27, no. 1, pp. 53–63, 1994.

[6] C. Kotropoulos and I. Pitas, "Rule-based face detection in frontal views,"IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997, vol. 4, pp. 2537–2540.

[7] L. Yang and A. Waibel, "A real-time face tracker",3rdIEEE Workshop on Applications ofComputer Vision, pp. 142–147.

[8] K. C. Yow and R. Cipolla, "Feature-based human face detection," Imageand vision computing, vol. 15, no. 9, pp. 713–735, 1997.

[9] I. Craw, D. Tock, and A. Bennett, "Finding face features," in ComputerVisionECCV'92. Springer, 1992, pp. 92–96.

[10] K.-K. Sung and T. Poggio, "Example-based learning for view-basedhuman face detection," IEEETransactions on Pattern Analysis and Machine Intelligence,vol. 20, no. 1, pp. 39–51, 1998.

[11] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points ofinterest," IEEE Transactionson Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 959–973, 2003.

[12] Yoon, S.M., Kee, S.C, "Detection of Partially Occluded FaceUsing Support Vector Machines", IAPR workshop on machinevision applications (MVA), Nara-ken New Public Hall, Nara,Japan, pp. 546–549 (2002).

[13] Hongxing, S., Jiayi, W., Peng, S., Zou, X, "Facial area forecast,and occluded face detection based on the YCbCr elliptical model", In: Mechatronic Sciences, IEEE International Conference on Electric Engineering and Computer(MEC), 2013,pp. 1199–1202.

[14] Tang, Y., He, Z., Chen, Y., Jinyi, W, "ATM Intelligent Surveillancebased on Omni-directional Vision" IEEE World Congresson Computer Science and Information Engineering, vol. 4, pp. 660–664 (2009)

[15] Choi, I., Kim, D, "Facial Fraud Discrimination Using Detection and Classification", Advances in Visual Computing, Lecture Notes in Computer Science, vol 6455, https://doi.org/10.1007/978-3-642-17277-9_21.

[16] Lin, D.T., Ming-Ju, L, "Face occlusion detection for automatedteller machine surveillance", In Pacific-Rim symposium on imageand video technology, pp. 641–651 (2006).

[17] https://learnopencv.com/otsu-thresholding-with-opencv/

[18] O. Déniz, G. Bueno, J. Salido, and F. De la Torre. Face recognition using histograms of oriented gradients. Pattern Recognition Letters, 32(12):1598–1603, 2011.

[19]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, volume 1, pages 886– 893.

[20] R. O'Malley, E. Jones, and M. Glavin. Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation. Infrared Physics & Technology, 53(6):439– 449, 2010

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004

[22] Angadi Sanjeevkumar and Nandyal Suvarna, "Database Creation for Normal and Suspicious Behaviour Identification in ATM Video Surveillance", Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, http://dx.doi.org/10.2139/ssrn.3835113.

[23] S. Nandyal and S. Angadi, "Adaptive Background Generation Method for Automated Teller Machine (ATM) with an Integrated Video Monitoring System", 2020 IEEE International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), 2020, pp. 1-5, doi: 10.1109/TEMSMET51618.2020.9557436.

[24] Angadi Sanjeevkumar and Nandyal Suvarna, "Human Identification Using Histogram of Oriented Gradients (HOG) and Non-Maximum Suppression (NMS) for ATM Video Surveillance", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), 2021, Volume-9, Issue-3, pp. 1-10 https://doi.org/10.21276/ijircst.2021.9.3.1.