Automatic Video surveillance for theft detection in ATM machines: An enhanced approach

Rupesh Mandal
Computer Science and Engineering
Sikkim Manipal Institute of Technology
Sikkim, INDIA
Email Id: rupeshmandal29@gmail.com

Nupur Choudhury
Department of Computer Science Engineering & IT
Assam Don Bosco University
Guwahati, INDIA
Email Id:nupur.choudhury@dbuniveristy.ac.in

Abstract – **This paper deals with the development of an application for automation of video surveillance in ATM machines and detect any type of potential criminal activities that might be arising with the system which would considerably decrease the inefficiency that are existing in the prevalent systems. An advanced digital Image processing technique along with the combination of computer vision and unsupervised machine learning techniques would be utilized which would create phenomenal results in the detection of the activities and their categorization. The proposed system makes efficient utilization of vector graphics and lists out an effective algorithm which comprises of methodologies like background modelling, subtraction, identification of salient objects, tracking of those objects and finally ending up with the detection and identification of the necessary action for the prevention of such type of activities. The proposed system also indulges in semi-supervised learning techniques and uses matching techniques like pose clustering and consistency in order to train the system and develop it to be an automated system as a whole. Since the captured video is fragmented into smaller frames and then the vector graphics and image processing techniques are implemented, the entire mechanism takes place in real time decreasing the time complexity to a great extent making the system an efficient mechanism to prevent such anti-social activities.**

Keywords – — background modelling; background subtraction;gaussian; k-gaussion; machine learning; semi-supervised learning.

I. INTRODUCTION

The present world scenario witnesses extensive usage of automated Video surveillance systems which plays a vital role in our day to day lives in order to enhance protection and security for individuals and infrastructure. Tracking and detection of objects is an essential component in various traffic monitoring systems, biometrics and security infrastructures, safety monitoring, various web applications and recognition of objects for mobile devices etc. in addition to playing important roles in medical domain, analysis of sports and others. One major application area of this process is the detection of robbery. In this paper the primary focus will be in the field of detection of suspicious activities or crime in an ATM (Automatic Teller Machine) which is basically a profitable bank service which enables financial transactions in public spaces where the machines are a replication of the bank clerks and tellers. Although several researches are going on in the field of ATM crime detection, however the utilization of the crime detection system is scarcely observed due to lack of efficiency and processing in the existing crime detection systems. Hence the idea of creating such a system was conceived after relative observations of the real life incidents that are happening in and around the globe. The increasing proliferation of the ATM frauds which involves activities like usage of mobile phones, multiple access in the same time, robbery, fights and vandalism is a matter of concern which would be tackled by the proposed system to enable secure financial transaction during anytime of the entire day and night.

II. BACKGROUD

A.ATM (Automatic Teller Machine)

It is a computerized machine built over efficient telecommunication system which enables financial institution along with the combination of financial transactions in a public domain which is primarily responsible for the cash dispense procedure as well as checking of account balance etc. This machines possess different structures in different nations worldwide. USA during 1969 witnessed the first ATM. They contribute nominally towards the positive currency growth which do not have a very robust effect. ATMs are considered to be more profitable bank service as it is a prime attraction for most of the non-bank customers. Its main structure comprises of a Central Processing Unit, a Pin pad, Secure Cryptographic processor, magnetic chip card, vault and the function keys.

B. Threats in ATM

Generally ATM frauds can be categorized into 3 main categories: Logical attacks, frauds related to cards and currency and physical attacks. However the crimes and threats corresponds to.

- Personal identification number threats
- Electronic data interception
- Fraudulent electronic transactions

978-9-3805-4421-2/16/$31.00

- Theft of money.
- Burglary and vandalism in ATMs
- Multiple access and Physical attacks

III. RELATED WORKS

Over the recent years human being detection in video surveillance systems is fairly gaining popularity due to its wide range of applications that involves vital processes like detecting abnormal events, characterization of gaits in humans, to count individuals in crowds, identifying people, classifications based on gender, detection of fall for the elderly people etc. Generally the various scenes that are a result of the video surveillance system is composed of very low resolution. However static camera captures scenes with minimum change in the background scenarios wherein the outdoor surveillance has to detect object that are in a larger scope. Some of the existing systems depend upon the human observers which would perform real time activity detection which leads to limitations like the difficulty related to simultaneous monitoring in the displays of the surveillance systems [1]. This calls for an automation of the video surveillance for analysis of human motion and has creates a research attraction on the field of pattern recognition and computer vision. The entire process comprises of 2 primary processes: Object detection and classification. The former can be carried out by processes like background subtraction, optical flow followed by spatio-temporal filtering. The first process, background subtraction is extremely popular for detection of objects where pixel by pixel or block by block fashion is considered to find the difference between the background and the current frame while detecting moving objects. In addition to this various other approaches include Gaussian mixture [2-5], non-parametric background process [6-9], temporal differencing [10-12], warping background model [13] and hierarchical background [14] models. In another attempt, optical flow based object detection technique is used [10, 15-16] which uses moving object flow vectors are used in definite intervals of time which detects the objects in motion in image sequence.

The second process, object classification process is generally classified into 3 categories: shape- based classification, classification based on motion and texture. The first category which is based on information based on shape of the moving objects initially describes the vital information like boxes, points, blobs etc. Succeeding this step is a process which behaves like a template matching phase [10, 15, 17-20]. However in addition to this the shape based approach leads to lot of misconception as the entire process might lead to several observed viewpoints which depicts inaccuracy in distinguishing the various articulations of the human body and confuses them with other moving objects. However this limitation was nullified by partial template matching [18]. Moreover, methods based on texture like HOG (histograms of oriented gradient[22] is also used for human region detection along with the combination of Support Vector Machine(SVM) and detection of high dimensional features based on the edges of the image objects.

This type of researches involves numerous studies which makes use of existing datasets which has been generated publicly for the purpose of training and evaluation of the processing. Various datasets like KTH human motion dataset [21], Weizmann human action dataset [24] and INRIA XMAS multi-view dataset consists of 6, 10 and 11 activities respectively. In a similar approach, different datasets are used [25] for various issues related to vision-based approach in Performance Evaluation of tracking and Surveillance (PETS) which evaluates based on a evaluation framework using various objectives and specific datasets. Gait recognition based research is done by the Institute of Automation, Chinese Academy of Sciences (CASIA) which has generated the CASIA Gait Database.

Due to the existence of various intra class variables like numerous objects, different poses, different lighting conditions it involves a great deal of effort for the classification of an unknown class into a general category. Although there are various available methods, even then they require a lot of intervention from humans which involves construction of models [31]. Extraction of features [29] etc. which are not that efficient as they cannot be utilized for multiple purposes. In another attempt operation on raw pixels [30] is done by Convolutional Neural Network (CNN) which presents remarkable performance on various tasks in image processing [33]. In addition to it the adjustment in the connection weights is performed by Back Propagation (BP) algorithm [32] which contributes trivial issues like time constraints, local minima etc. CNN also requires enormous computations and datasets for training purpose which is used for tuning purpose. Another significant approach was taken known as Local receptive fields based extreme learning machine (LM-LRF) [28] which performs the same task by randomly generating the input weights and analytically calculating the output weights thereby generating a deterministic and relatively simpler solutions. In addition to it computations and samples for training are also reduced to a great extent due to random generation.

IV. METHODOLOGY

The primary approach for the proposed system primarily consists of the following 5 steps.

1. Background Registration
2. Background subtraction
3. Object Tracking
4. Feature Extraction
5. Pattern Matching.

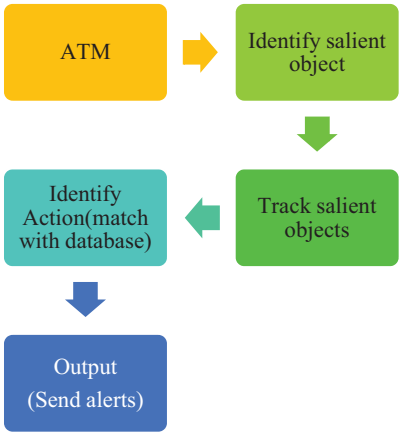

Fig. 1. Proposed System workflow

A. Background Registration

Initially the information from the frame difference is accumulated and a reliable background image is constructed using a background registration technique. The moving object region is then separated from the background region by comparing the current frame with the constructed background image. Finally, a post-processing step is applied on the obtained object mask to remove noise regions and to smooth the object boundary.

B. Background Subtraction

The background subtraction was carried out using the Pixel-Wise Local information based approach [35] where the neighboring spatial distribution of each pixel was utilized Gabor features are extracted using Gabor filters with multiples scales and orientation which generates various feature vectors. The representation based on Gabor filters for an input image is as follows:

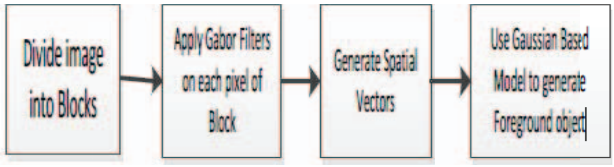

Fig. 2. Foreground Image detection using QR decomposition

$$G(x, y, \mu, v) = F(x, y) * \varphi\, x, y(z) \quad (1)$$

Here * = convolution

F(x, y) = input image

φx, y(z) = Gabor filter with v = {0,1…p-1} p scale

μ = {0, 1…q-1} orientation being q [35]

Spatial feature vector is obtained for pixel(x, y) by using equation (1) which is shown below

$$SfV(x,y)=(G(x,y,0,0),.....,G(x,y,0,p-1)\ ,G(x,y,1,0).....,G(x,y,q-1,p-1))T \quad (2)$$

Equation (6) represents K-Gaussian mixture model which generates the spatial vector of each of the pixels (SV1, SV2, SV3 …SVt}.

The equation for pixel SVt+1 is given by

$$(SVt+1-\mu i,\ t)/\sigma i,\ t<2.5) \quad (3)$$

If equation (4) is satisfied by SVt+1 it represents that this pixel can be marked as foreground which shows at least one matched model. K Gaussian is then sorted in a descending order which generates B distribution.

$$B=arg\ min\ (\Sigma k=1 bwk > T) \quad (4)$$

Here T = threshold, first B distribution = background model.

K Gaussian distribution at time t is given by

$$P(Xt)=\Sigma Ki=1 wi,t..\ \eta(Xt,\mu i,t,\Sigma i,t) \quad (5)$$

η = probability density function [36].

C. Object Tracking

This process estimates the trajectory of the object in the plane of image as and when it moves around in scene. The tracker is responsible for assigning various consistent labels in different frames of the video. Various object centric information like the orientation, area, shape etc. is also provided by the tracker which can be further made simple by imposing various constraints to the object motion and appearance like velocity and constant acceleration which can be based on prior knowledge regarding the object and its identification.

Object representation: The objects in the proposed system can be represented using their appearances and shapes. Various shape representations include articulated models which basically represents objects comprising of body parts along with the joints where the parts relationships are governed by kinematic motion parameters like angle etc. and can be represented using cylinders, ellipses etc. on the other hand skeletal models can be extracted by application of medial axis transformation to the object silhouette. Some of the major representations used proposed system are:

Probability density which includes parametric (Gaussian) and a mixture of Gaussians or nonparametric (histograms).

Templates which are represented by simple geometric shapes or silhouettes that is capable of spatial as well as information related to appearance in a single view which makes it efficient for tracking objects which do not involve much change in positions.

Active appearance models and Multiview appearance models where the former is responsible for shape (depicted by a set of landmarks) and appearance. The later deals with different views of the objects by generating subspaces from the views obtained which involves Principle Component Analysis (PCA), Independent Component Analysis (ICA) that can be used both for shape and appearance. Some other approaches can also be used which involves SVM and Bayesian networks.

D. Feature Selection For Tracking

Feature selection represents the uniqueness so that the objects can be easily distinguished in the feature space. It closely resembles object representations. In the proposed system both histogram based appearance representations as well as contour

based representations are considered for features and the following visual features are used for distinguishing purpose.

Color: (histogram based appearances): the proposed system utilizes L*u*v* and L*a*b* color spaces along with the combination of HSV (Hue, Saturation, Value) since RGB dimensions are highly correlated for tracking.

Edges: (Contour based representations): Canny edge detector is used to consider this type of feature representation which detects the edges that are responsible for generation of strong changes in the intensity of the image. These edges are depicted as features in the near future.

E. Pattern Matching

The system is trained with all the possible actions of theft and safety. All the frames are matched with the available dataset using hamming distance algorithm. The features are matched in the form of decision tree. All the actions are matched based on the features matched in decision tree.

V. RESULTS AND DISCUSSIONS

Here in this system we provide a video file as an input. The system the process the video and indicates the action performed by the person, whether it's a safe or theft action. At first the screen shots of the original video frame is shown. Then the segmented the segmented image of those video frame is represented. These segmented frame is the matched with the trained dataset. Based on this the final result will be displayed whether the action performed is a theft action of a safe action. The video is taken with a 5 megapixel mobile phone's camera. The ATM machine shown in the video is a dummy model made of cardboard and the video is taken in a room.

Fig. 3. Image frames form orginal video (*Safe Action*)

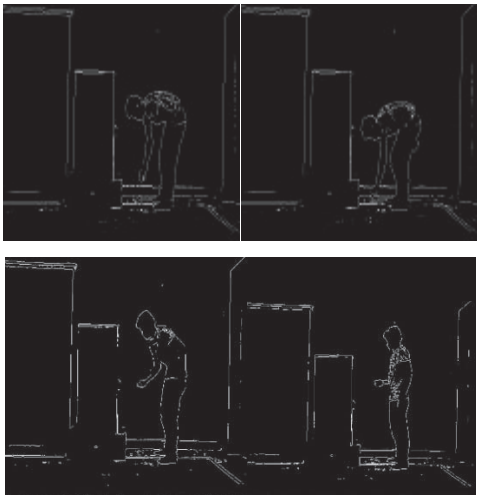

Fig. 4. Segmented image frames of the original video (*Safe Action*)

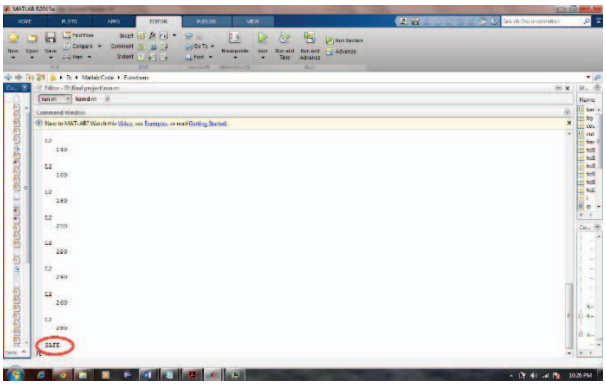

Fig. 5. Output of the system showing 'SAFE' action

Fig. 6. Image frames form orginal video (*Theft Action*)

Fig. 7. Segmented image frames of the original video (Safe Action)

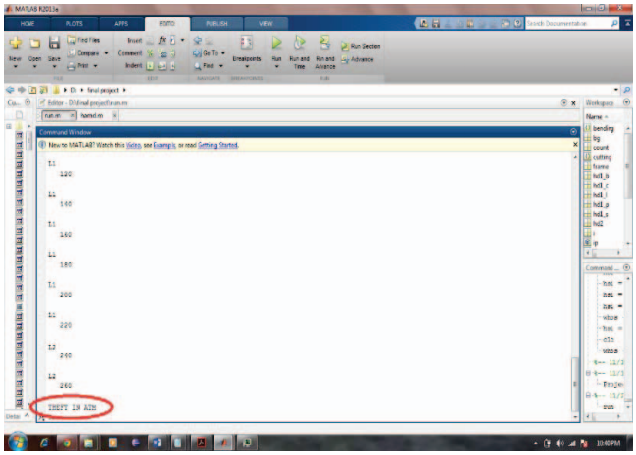

Fig. 8. Output of the system showing 'SAFE' action

VI. CONCLUSION AND FUTURE SCOPE

In this paper a new framework is proposed which will be in the current world. The existing security for the present days ATM are facing a number of threats. Among these threats thefts in the ATM system is most common one. This research mainly focus on building an intelligent framework that provide high class security to the existing ATM system. The video stream obtained from the surveillance camera in the ATM room if feed to this framework. Based on the activities performed by the customer in the ATM room the system will report suspicious activities instantaneously. The main focus area of this research is to prevent crime in ATM by using image processing and machine learning techniques. By implementing the proposed system in the existing ATM a robust atmosphere is created, which evaluate each and every actions inside the ATM.

REFERENCES

[1]. N Sulman, T Sanocki, D Goldgof, R Kasturi, "How effective is human video surveillance performance?", 19th International Conference on Pattern Recognition, (ICPR 2008) (IEEE, Piscataway, 2008), pp. 1–3.

[2]. C Stauffer, W Grimson, "Adaptive background mixture models for real-time tracking", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1999) (IEEE, Piscataway, 1999), pp. 246–252.

[3]. YL Tian, RS Feris, H Liu, A Hampapur, M-T Sun, "Robust detection of abandoned and removed objects in complex surveillance videos". Syst. Man Cybern. Part C Appl. Rev. IEEE Trans. 41(5), 565–576 (2011).

[4]. DS Lee, "Effective Gaussian mixture learning for video background subtraction". IEEE Trans. Pattern Anal. Mach. Intell. 27(5), 827–835 (2005).

[5]. C Ridder, O Munkelt, and H. Kirchner, "Adaptive Background Estimation and Foreground Detection Using Kalman-Filtering", Proc. Int'l Conf. Recent Advances in Mechatronics, ICRAM 95, pp. 193–199 (1995).

[6]. W Kim, C Kim, "Background subtraction for dynamic texture scenes using fuzzy color histograms", Signal Process. Lett. IEEE 19(3), 127–130 (2012).

[7]. A Lanza, "Background subtraction by non-parametric probabilistic clustering", 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (IEEE, Piscataway, 2011), pp. 243–248.

[8]. A Elgammal, D Harwood, L Davis, "Non-parametric model for background subtraction", 6th European Conference on Computer Vision - Part II (ECCV '00) (Springer, London, 2000), pp. 751–767.

[9]. L Li, W Huang, IY-H Gu, Q Tian, "Statistical modeling of complex backgrounds for foreground object detection", IEEE Trans. Image Process. 13, 1459–1472 (2004).

[10]. J Candamo, M Shreve, DB Goldgof, DB Sapper, R Kasturi, "Understanding transit scenes: A survey on human behavior-recognition algorithms", IEEE Trans. Intell. Transp. Syst. 11(1), 206–224 (2010).

[11]. F-C Cheng, S-C Huang, S-J Ruan, "Scene analysis for object detection in advanced surveillance systems using Laplacian distribution model", Syst. Man Cybern. Part C Appl. Rev. IEEE Trans. 41(5), 589–598 (2011).

[12]. D-M Tsai, S-C Lai, "Independent component analysis-based background subtraction for indoor surveillance", IEEE Trans. Image Process. 18(1), 158–167 (2009).

[13]. T Ko, S Soatto, D Estrin, "Warping background subtraction", 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010) (IEEE, Piscataway, 2010), pp. 1331–1338.

[14]. S Chen, J Zhang, Y Li, J Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction", IEEE Trans. Ind. Inform. 8(1), 118–127 (2012).

[15]. J Xiaofei, L Honghai, "Advances in view-invariant human motion analysis: a review", IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 40(1), 13–24 (2010).

[16]. A Efros, A Berg, G Mori, J Malik, "Recognizing action at a distance", Ninth IEEE International Conference on Computer Vision (ICCV 2003) (IEEE, Piscataway, 2003), pp. 726–733.

[17]. L Wang, X Geng, C Leckie, R Kotagiri, "Moving shape dynamics: a signal processing perspective", IEEE Conference on Computer Vision and Pattern Recognition(IEEE, Piscataway, 2008), pp. 1–8.

[18]. M Singh, A Basu, MK Mandal, "Human activity recognition based on silhouette directionality", IEEE Trans. Circuits Syst. Video Technol. 18(9), 1280–1292 (2008).

[19]. FZ Eishita, A Rahman, SA Azad, A Rahman, "Occlusion handling in object detection", Multidisciplinary computational intelligence techniques: applications in business, engineering, and medicine. IGI Global. (2013), doi: 10.4018/978-14666-1830-5.ch005.

[20]. B Wu, R Nevatia, "Detecting and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors", Int. J. Comput. Vision (IJCV), 75(2), 247–266 (2007).

[21]. C Schuldt, I Laptev, B Caputo, "Recognizing human actions: a local SVM approach", 17th International Conference on Pattern Recognition (IEEE, Piscataway, 2004), pp. 32–36.

[22]. N Dalal, B Triggs, “Histograms of oriented gradients for human detection”, IEEE Conference on Computer Vision and Pattern Recognition (IEEE, Piscataway, 2005), pp. 886–893.
[23]. D Weinland, R Ronfard, E Boyer, “Free viewpoint action recognition using motion history volumes”. Comput. Vision Image Understanding (CVIU) 104(2–3), 249 257 (2006).
[24]. M Blank, L Gorelick, E Shechtman, M Irani, R Basri, M Blank, L Gorelick, E Shechtman, M Irani, R Basri, “Actions as space-time shapes”, 10th IEEE International Conference on Computer Vision (IEEE, Piscataway, 2005), pp. 1395–1402.
[25]. PETS, Performance Evaluation of Tracking and Surveillance. http://www.cvg.rdg.ac.uk/slides/pets.html. Accessed on 17 Nov 2013.
[26]. P Turaga, R Chellappa, VS Subramanian, O Udrea, “Machine recognition of human activities: a survey”, IEEE Trans. Circuits Syst. Video Technol, 18(11), 1473–1488 (2008).
[27]. Yoshua Bengio, Aaron Courville, and Pascal Vincent. “Representation learning: A review and new perspectives”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8):1798 1828, 2013.
[28]. Guang-Bin Huang, Zuo Bai, L.L.C. Kasun, and Chi Man Vong. Local receptive fields based extreme learning machine. Computational Intelligence Magazine, IEEE, 10(2):18–29, 2015.
[29]. Iasonas Kokkinos and Alan Yuille. Scale invariance without scale selection,IEEE Conference on Computer Vision and Pattern Recognition pages 1–8. IEEE, 2008.
[30]. Yann LeCun, Fu Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting, International Conference on Computer Vision and Pattern Recognition, volume 2, pages II–97–104, 2004.
[31]. B. Leibe and B. Schiele,“Analyzing appearance and contour based methods for object categorization”, Computer Vision and Pattern Recognition, IEEE Computer Society Conference, volume 2, pages II–409–15, June 2003.
[32]. David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, “Learning representations by back-propagation errors”, Nature, 323:533–536, 1986.
[33]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. “Going Deeper with Convolutions”, ArXiv e-prints, 2014.
[34]. Masashi Nishiyama, Osamu Yamaguchi, and Kazuhiro Fukui. “Face recognition with the multiple constrained mutual subspace method”, Audio-and Video-Based Biometric Person Authentication, pages 71–80. Springer, 2005.
[35]. Zhong Wei, Shuqiang, Qingming Huang” pixel-wise local information based background subtraction approach ”, ICME 2008 IEEE, 2008.
[36]. Shih-Chieh Wang, Te-Feng Su and Shang-Hong Lai, ”Detecting Moving Object From Dynamic Background with Shadow Removal”ICASSP2011 IEEE, 2011