**A Survey on Public Safety Systems inside ATM**

1Ravi Shankar Kumawat, 2S. Iniyan

1B.Tech, 2M.E

1Computer Science Engineering

SRM Institute of Science and Technology, Chennai, India

1[ravishankar\_ku@srmuniv.edu.in](mailto:ravishankar_ku@srmuniv.edu.in), 2iniyan.s@ktr.srmuniv.ac.in

**Abstract--Automated Teller Machine (ATM) offers great convenience to people in performing financial transactions, such as cash withdrawals, deposits, transfer funds or obtaining account information at any time and without the need for direct interaction from bank staff. It was first used in 1939 and now there are more than 3 million ATM machines all over the world, thus the security of ATM should always be considered. A common method is to rob the customer amid transaction or while the staff is filling the machine with cash. The bank uses encryption techniques to ensure secrecy of transactions and other sensitive data also in some cases banks have installed biometric identification such as fingerprint, iris, palm vein patterns etc. This paper aims to summarize the techniques and work done in ATM surveillance using audio and video analysis and Machine learning techniques such as abnormal Human Behavior Recognition(AbHAR), Fall detection, Ambient Assistive Living(AAL), suspicious sound detection etc to enhance safety at ATM sites.**

*IndexTerms*—Surveillance, Video captioning, ATM, Intrusion detection, public safety, audio classification

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

1. **Introduction**

Automated Teller Machine (ATM) accepts cash request from the user, verifies the authenticity of the user to access, ensures the user has sufficient amount and dispenses the money. There are numerous cases of ATM robbery amid the ATM transactions due to the absence of security in ATMs. ATM has various security features to identify fraud detection, transactional errors or hacking but it has very few features to ensure public safety in events of a loot or robbery as can be seen in Fig. 1.1 below.



Fig. 1.1: Attacks inside ATM

This paper reviews the techniques based on Natural Language Processing (NLP), machine learning and video processing that can be applied in order to stop a physical loot/robbery from ATM systems ensuring public safety along with safeguard of physical and capital assets. There are various methods involving physical sensors, audio and video processing to identify any suspicious activities and alarm the authorities in any such scenario.

1. **DESIGN MODULES**

The figure shown below shows the modules that can be used either in isolation or with other modules in order to identify any suspicious event/activity inside the ATM.

1. Audio Processing Module: Detects the presence of suspicious sounds inside the room.
2. Video Processing Module: Detect abnormal behaviour and activities using video data.
3. Physical Sensor Module: Detects presence of Fire, High pressure, smoke etc

The data from these sensors is used as shown in Figure 2.1

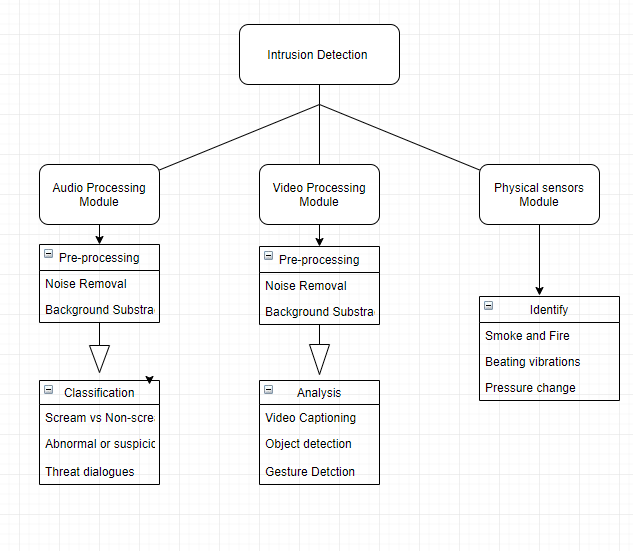
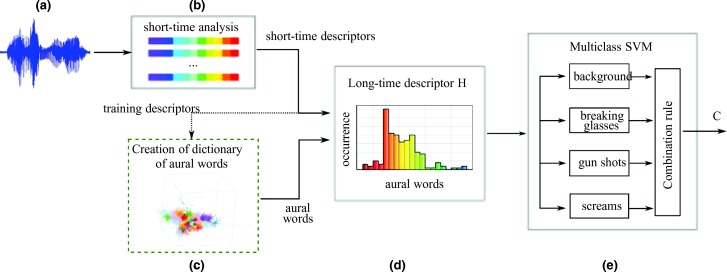


Figure 2.1: Modules of the System

1. **AUDIO PROCESSING**

Audio is a wave-like format of data where the amplitude of audio change with respect to time. Audio surveillance includes extracting features from the audio and using the features to classify different sounds as shown in Figure 3.1 (system used in [2]).



**Figure 3.1:** Overview of the system architecture. The input audio signal (a) is divided in small frames (b) and a short-time descriptor is computed for each of them. A dictionary of aural words is created during the training phase (dashed box) by means of a clustering process (c). Then the histogram of the occurrences of the aural words in a m-seconds windows (long-time descriptor H) is computed (d) and is fed to a multiclass SVM classifier (e) that performs the detection of events.

Most automated surveillance system consists of video sensors which lack the robustness and reliability required in real-world applications. Thus to tackle this issue, audio sensors are deployed either alone or in combination with video sensors. The audio processing usually involves feature extraction as shown in figure 3.2, noise reduction followed by background subtraction, event classification and situation analysis.

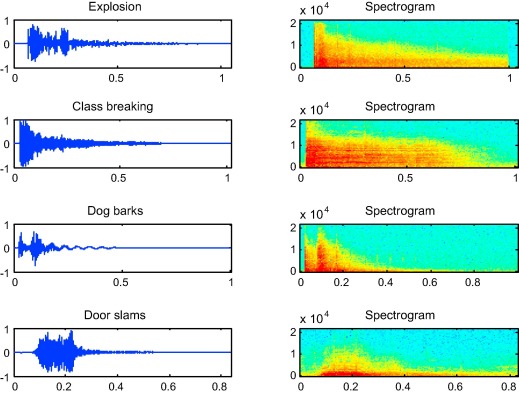


Figure 3.2: Audio waveform and spectrograms of 4 classes environmental sound

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| REF. No. | OBJECTIVE | ALGORITHM USED | DATASETS OR INPUT PARAMETERS | PERFORMANCE MEASURE USED | REMARK |
| [1] | Recognize environmental sounds | SVM classifier | GTZAN dataset ,MFCCs, scatter wavelet | confusion matrix | Retrieves lost high frequencies via consecutive wavelet convolutions and thus high accuracy. |
| [2] | Detection of audio events for surveillance applications | SVM Classifier | Mivia audio events dataset | Recognition rate , positive rate (FPR) | Robust w.r.t.to background noise and real environments |
| [3] | Acoustic event classification | NMF, SBV | FBK-Irst, UPC-TALP | Confusion Matrix | New front-end for AEC(high pass filtering),better accuracy |
| [4] | Robust acoustic event classifier using DNN | RBM, DNN | RWCP sound scene dataset | Positive classification rate | More robust and better classification |
| [5] | Audio event detection with the FPGA platform | parallel structure of SIF extraction and 1-max pooling CNN | Real World Computing Partnership (RWCP) Sound Scene Database | Accuracy, precision and CPU efficiency | FPGA increases performance by 31.9 times |
| [6] | Weakly-supervised audio classification | CNN, Temporal attention based localization method audio classifier | Manually labelled audio extracted from YouTube | F1 and Equal error along with Precision and recall | Learnable gated units help select most related features giving 57.7% F1 score |
| [7] | Audio Event Detection | Region-based Convolutional RNN(RCNN) | DCASE 2017 challenge dataset | Event-based error rate, F1 score | No post-processing needed to convert the prediction from frame level to event level |
| [8] | Weakly semi-supervised audio detection in domestic environments | Convolutional RNN | Audio extracted from YouTube | F1 score | Obtaining time coded annotations is time consuming and a bottleneck in SED systems |
| [9] | Extract DAF for acoustic event detection | multi-stream hierarchical DNN | BBC and TV audio dataset | Average F score | Shows DAF is better than both MFCC and Gabor features but DAF does not perform well in low SNR conditions |
| [10] | Sensor Management(SM) framework for IoT acoustic surveillance | SVM, team theoretic SM, LDA classifiers, Trilateration algorithm | SUSAS dataset | Confusion matrix with classification accuracy | Decision-theoretic approach can be implemented instead of team-theoretic along with optimized SVM kernels |

Table 3.1: Summary of existing works using Audio

1. **VIDEO PROCESSING**

Video processing plays a crucial role in the security system as it provides a lot of features which can be used to identify suspicious activities. Video analysis techniques such as Video Captioning, Human body tracking (position and activities), 2D and 3D tracking, Fall detection, Belt like object detection, abnormal behavior, Skeleton tracking, Motion segmentation, Real-time detection of moving objects, object classification such as shown in Fig 4.2 have been implemented for video surveillance applications. The terminologies are:

* + 1. Video Captioning: Automatically generating textual description from an artificial system using NLP.
    2. Human body Tracking: Estimating the parameters of the human body from video data as the position configuration of the tracked body change over time.
    3. Motion segmentation: Separating regions, features, or trajectories from a video sequence into coherent subsets of space and time.
    4. Object Detection: Identifying objects present in an image or video.
    5. Pose Recognition: Pose refers to computer vision techniques that detect human figures so as to identify the different parts of the body and estimate what the person is doing.



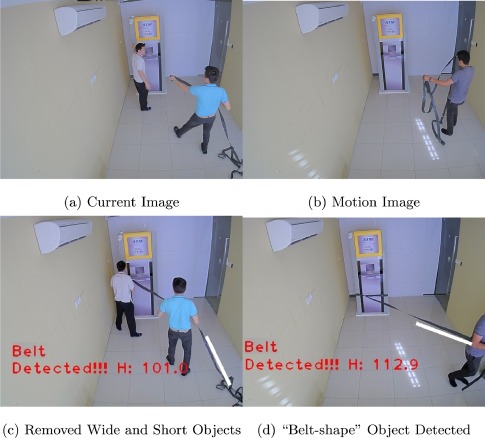


Fig 4.2: Various abnormal scenario

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| REF. No. | OBJECTIVE | ALGORITHM USED | DATASETS OR INPUT PARAMETERS | PERFORMANCE MEASURE USED | REMARK |
| [11] | Detect abnormal events at ATM site | SVM, PCA on shape and motion analysis | Handcrafted dataset | AUC, Precision, Recall | Requires Huge computational resources |
| [12] | Real-time security framework at ATM site | Random forest, Root of sum of squares(RSS) | HMDB and Caviar dataset | Accuracy, Precision, Recall, ROC and F score | Does not targets criminal activities |
| [13] | Video captioning | Attention-based LSTM model with semantic consistency | MSR-VTT, MSVD | BLUE, METEOR and CIDEr | proposed a new framework LSTMs to be extended to domain specific datasets |
| [14] | Detecting suspicious behaviour to remove or attack the ATM machine | Background subtraction, squat and climb detection algorithms | Self-recorded videos | Accuracy and computational efficiency | The computation is done at remote site by sending continuous video stream which is a overhead |
| [15] | Posture recognition for effective real-time ATM monitoring | Logistic regression on skeleton and depth based data | Self-recorded videos using Kinect 3D as in fig: peeping posture | Classification accuracy | Uses Kinect 3D camera for skeleton information along with openNI framework |
| [16] | Atuomation of ATM surveillance to detect criminal activities | Pose clustering, background modelling and subtraction, object tracking, PCA, ICA, SVM, Bayesian networks | Self-recorded videos and clips in a similar environment | Identification accuracy | Requires PC to be implemented for image processing which is not practical to be installed at ATM sites |
| [17] | Detecting suspicious behaviour in shopping malls | Kalman filtering, LSAP, SVM Occlusion management using GCH, LBP and HOG | CAVIAR dataset | Accuracy in different situations | Needs high computational resources for real-time monitoring |
| [18] | A Robust surveillance system for academic environment | Temporal differencing algorithm, SVM, Gaussian function, CBIR, OMEGA equation as filter | Photos from students with different emotions | Classification accuracy | High quality compared to other similar systems |
| [19] | Activity recognition and behavior detection for elderly people with dementia | RNN, VRNN, LSTM, GRU | Synthetic dataset, OPPORTUNITY dataset | Precision, Recall, F measure and accuracy | Robust and able to cope with imbalanced data |
| [20] | Occluded face detection for ATM surveillance | Head tracking algorithm, Bayesian framework, AdaBoost algorithm | Real world synthetic dataset | Accuracy, Detection rate, average false alarm rate, | Fast and robust for face occlusion detection |

Table 4.1: Summary of existing works using Video data

1. **PHYSICAL SENSORS**

There are various physical sensors available in ATM such as fire sensors, shock, smoke, pressure etc. which adds another layer of security against robbery/theft. Also, the money is kept in cassettes which will dye the money if incorrectly opened.

1. **CONCLUSION**

In this paper, we reviewed various techniques to enhance customer safety in ATMs by detecting suspicious criminal activities at the site. The above discussed audio, video and physical sensors can be deployed together to get better accuracy. Thus by this work, we hope to contribute to increase customer security inside ATM.

**REFERENCES**

[1] Souli, S., & Lachiri, Z , “Audio sounds classification using scattering features and support vectors machines for medical surveillance”. Applied Acoustics, 130, 270–282, 2017.

[2] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., & Vento, M. “Reliable detection of audio events in highly noisy environments”. Pattern Recognition Letters, 65, 22–28, 2015.

[3] Ludeña-Choez, J., & Gallardo-Antolín, A.“Feature extraction based on the high-pass filtering of audio signals for Acoustic Event Classification” Computer Speech & Language, 30(1), 32–42. 2015

[4] Sharan, R. V., & Moir, T. J. “Robust acoustic event classification using deep neural networks”. Information Sciences, 396, 24–32. 2017

[5] Xu, Jinwei, et al. "High performance robust audio event recognition system based on FPGA platform." *Cognitive Systems Research* 50 (2018): 196-205.

[6] Xu, Yong, et al. "Large-scale weakly supervised audio classification using gated convolutional neural network." *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.

[7] Kao, Chieh-Chi, et al. "R-CRNN: Region-based convolutional recurrent neural network for audio event detection." *arXiv preprint arXiv:1808.06627* (2018).

[8] Serizel, Romain, et al. "Large-scale weakly labeled semi-supervised sound event detection in domestic environments." *arXiv preprint arXiv:1807.10501* (2018).

[9] Li, Yanxiong, et al. "Using multi-stream hierarchical deep neural network to extract deep audio feature for acoustic event detection." *Multimedia Tools and Applications* 77.1 (2018): 897-916.

[10] Hilal, Allaa R., et al. "A distributed sensor management for large-scale IoT indoor acoustic surveillance." *Future Generation Computer Systems* (2018).

[11]Tripathi, V., Gangodkar, D., Latta, V., Mittal, A. (2015) Robust abnormal event recognition via motion and shape analysis at atm installations. Journal of Electrical and Computer Engineering, 1(10).

[12] Tripathi, V., Mittal, A., Gangodkar, D., Kanth, V. (2016) Real time security 525 framework for detecting events at atm installations. Journal RealTime Image Processing, 1-11

[13] Gao, L., Guo, Z., Zhang, H., Xu, X., & Shen, H. T. ,Video Captioning With Attention-Based LSTM and Semantic Consistency. IEEE Transactions on Multimedia, 19(9), 2045–2055,2017

[14] Lee, Wai-Kong, et al. "ArchCam: Real time expert system for suspicious behaviour detection in ATM site." Expert Systems with Applications 109 (2018): 12-24.

[15] Nar, Rajvi, Alisha Singal, and Praveen Kumar. "Abnormal activity detection for bank ATM surveillance." *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*. IEEE, 2016.

[16] Mandal, Rupesh, and Nupur Choudhury. "Automatic video surveillance for theft detection in ATM machines: An enhanced approach." *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*. IEEE, 2016.

[17] Arroyo, Roberto, et al. "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls." *Expert systems with Applications* 42.21 (2015): 7991-8005.

[18] Al-Nawashi, Malek, Obaida M. Al-Hazaimeh, and Mohamad Saraee. "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments." *Neural Computing and Applications* 28.1 (2017): 565-572.

[19] Arifoglu, Damla, and Abdelhamid Bouchachia. "Activity recognition and abnormal behaviour detection with recurrent neural networks." *Procedia Computer Science* 110 (2017): 86-93.

[20] Zhang, Tao, et al. "Fast and robust occluded face detection in ATM surveillance." *Pattern Recognition Letters* 107 (2018): 33-40.