

# Vision Based Anomaly Detection System for ATM

Arpitha K<sup>1</sup>, Honnaraju B<sup>2</sup>

<sup>1</sup>MTech Student, Dept. of Computer Science and Engineering, MIT Mysore, Karnataka, India

<sup>2</sup>Associate Professor, Dept. of Computer Science and Engineering, MIT Mysore, Karnataka, India

\*\*\*

**Abstract** - The fast development in Automatic Teller Machines (ATM) has made life simple for the everyday life, except it isn't so for administrators who oversee it. ATMs are not claimed by banks rather they are outsourced to Managed service providers (MSPs) from acquiring to keeping up the machines. A few variables like the support, cash filling and security. Since the whole motivation behind an ATM benefit is to lessen manual connection and give administration to the client through mechanization, it is imperative that the security gave is likewise programmed and proficient. The exchanges through ATM's are expanding quickly alongside some different exercises because of less security. Thus through the human stance typicality and anomaly identification and Analysis utilizing SIFT (Scale-invariant feature transform) and Gabor include extraction process and SVM (support vector machine) classification technique in ATM machine focuses; the above issue proclamation can be overcome.

**Key Words:** Scale-invariant feature transform (SIFT), Gabor feature extraction, Support vector machine (SVM), Difference of Gaussian (DoG), Feature extraction.

## 1. INTRODUCTION

The quick headway in Automatic Teller machines (ATM) has made life direct for the standard day by day presence, beside it isn't so for executives who oversee it. ATMs are not attested by banks rather they are outsourced to Managed service providers (MSP) from acquiring to keeping up the machines. A few sections like the assistance, cash filling, and security and along these lines the uninvolved resources inside the ATM rooms are responsible for keeping the ATM dynamic. Normally, an ATM site contains wherever between 8 to 12 uninvolved resources which join two air course and cooling structures, two light collecting sheets, Associate in Nursing inverter/UPS, a surveillance camera and no under eight to twelve lightweight globules.

Security is one of the key worries of the edge time. The most significant issue of utilizing electronic gadgets and mechanical gathering is that the data can be hacked and spilled, security overlooked, robbery and theft. ATMs are general saving money associations which benefits especially precious for the clients. Most ATMs are open 24 hours; their zones are spread wherever on a city/town. It may be difficult to take cash from a bank, as it is equipped with high manual security and is by and large organized close populated regions of a city/town. Regardless, ATMs because of the likelihood of its association, is under a more basic danger of being burgled. Since the whole clarification behind an ATM advantage is to diminish manual connection and offer

association to the client through robotization, it is fundamental that the security gave is in like way tweaked and beneficial. The tenacious reports of burglary and theft in ATM are a making worry for the banks. These security issues are changing into a hindrance to the development of ATM association and better resolute quality. The security structure should be helpfully robotized so it can perceive bizarre conditions inside the ATM and answer to the experts or jar the machine from discharging cash. The figure 1 demonstrates the motion and posture of the human amid ordinary exchange.



**Fig 1:** Pose of human amid typical Transaction

Following are a portion of the pictures of the ATM focuses amid burglary. Figure 2 demonstrates the human posture and Gesture amid the breaking of ATM Machine. Figure 3 demonstrates the human posture amid un-mounting or breaking CC-TV camera. Figure 4 demonstrates the human posture and signal of human when the individual, assault the client in ATM.



**Fig 2:** Human posture and Gesture amid the breaking of ATM Machine.



Fig 3: Human posture on breaking CC-TV camera.



Fig 4: Human posture and Gesture when the individual, assault the person in ATM.

## 2. RELATED WORK

Sujith B [1] was proposed the strategy that will utilizes various question discovery strategy and occasion acknowledgment strategies of PC vision. Most of the time utilized systems for moving item recognition are foundation subtraction, factual strategies, transient separating and optical stream. Nithya Shree R et. all [2] was designed framing and blob recognition. This structure distinguishes chance in the territory under observation. One such perilous condition is executed, similar to a man with a blade. The proposed structure comprises of two fundamental segments: Framing and Blob recognition (FBD) for Input video preparing and (HON) Human tracking, Object recognizable proof and warning organize.

Hidetomo Sakaino [3] proposes a cost lessening strategy for the MCMC approach by taking moves, i.e., birth and demise, out of the emphasis circle of the Markov chain when diverse moving items collaborate. Gee strategy incorporates a few distinct modules so as to adapt to various broken directions. The GUI proposed thus offers an auto-portion module of images from pictures and a hand-drawing module for productive direction learning and for premium direction expansion. Zhiqian Chen, et. all, [4] show relative examination of various descriptors like Gradient based descriptor i.e. HOG(Histogram of Oriented Gradients) and

shape based descriptors like Hu Moments and Zernike Moments for powerful anomalous movement discovery. HOG is in charge of pulling back shape data of question in picture utilizing force angles and edge headings. Zernike minutes are orthogonal minutes which are compelling in picture portrayal. These are revolution invariant. Zernike minutes are developed from Zernike polynomials which are orthogonally autonomous and subsequently the picture portrayal doesn't experiences covering or repetition. Hu minutes give seven qualities as a removed component from a given picture. These minutes are invariant to interpretation, scale, and turn of a picture. Out of seven invariants, six are supreme orthogonal invariants and the seventh one is skew orthogonal invariant. Hoard gives better exactness then different descriptors. As for shape descriptors Hu minutes give better outcomes in video when contrasted with Zernike minute.

Lars Moland Eliassen, et. all, [5] proposed a calculation yields predominant exactness and vigor when utilized as a part of a stance estimation setting. General purpose computing on graphics processing units (GPGPU) that is appropriate for use in human stance estimation, and accomplishes constant execution. A diminishing calculation iteratively expels limit voxels from a question create a topologically identical skeleton. The proposed calculation ought to be contrasted with other cutting edge skeletonisation calculations. Exhibited ongoing diagram based human posture estimation approach, utilizing skeletons to recoup the subject's stance. The skeletons were created with a strategy called voxel scooping. S. Abdul Kareem et. all, [6] The utilization of hand signals as a characteristic interface fills in as a spurring power for inquire about in motion scientific categorizations, its portrayals and acknowledgment methods. Hypothetically the writing arranges motions into two sorts; static and dynamic signals. K-means calculation begins by arbitrarily finding nearest neighbor in otherworldly space. Every pixel in the information picture amass is then allotted. To the closest group focus and the bunch focus areas are moved to the normal of their class esteems. K-nearest neighbor is a strategy for grouping objects in view of nearest preparing cases in the component space.

Rishabh Agrawal, et. all, [7] proposed a technique that works by examining 3D information progressively and utilizes an arrangement of grouping guidelines to characterize the quantity of convexity surrenders into signal classes. This outcomes progressively execution and refutes the prerequisite of any preparation information. The proposed strategy accomplishes estimable execution with low processor use. The proposed technique utilizes a Creative Senz3D camera to catch both shading and profundity data, and is executed utilizing OpenCV API in C++ dialect. Hand Detection and Tracking: The Senz3d camera catches a RGB video outline alongside the related profundity information. Profundity based thresholding is performed to evacuate the foundation Hand Region Analysis: This comprises of first portioning arm from the palm locale of the hand. To accomplish this, first the fingers are dissolved far

from the hand cover and are the palm veil is made. This is finished by progressive disintegration and enlargement morphological activities. Motion Recognition: Gesture acknowledgment module is the center piece of the proposed technique. This begins by first making a shut form C1 for the hand cover. This is finished utilizing straightforward chain guess.

Hong Cheng et. all [8] was presented a survey of some recent works on hand gesture recognition using 3D depth sensors. A review of the state-of-the-art research for 3D hand gesture recognition in four aspects: 3D hand modeling, Static hand gesture recognition is clear to perceive the hand shape or stance because of the rich data in the evaluated hand display. Hand trajectory gesture recognition is to the static hand motion acknowledgment that takes a shot close by shapes; the hand direction motion acknowledgment considers the successive information of hand direction and investigates the worldly character of hand movement. Continuous hand gesture recognition was the center is to perceive the single important hand signal in the video.

Miwa takai [9] This study extracts Motion Region from moving person, and measures Motion Quantity for measuring his/her active state. And, this proposal method finds the detecting point of suspicious activity, and estimates the degree of risk of the suspicious activity. Sambarta Ray et. all, [10] proposed a framework that thinks about two strange circumstances. In the first place, it utilizes a recognition calculation to identify human faces and check the quantity of individual's exhibit inside the ATM booth. Second, it can likewise distinguish whether a man is wearing veil or not. The recognition framework utilizes Viola-Jones calculation to distinguish the face and different facial angles like the eye match, nose and mouth. The framework at that point checks the no. of individual present inside the ATM and utilizing the recognized facial component chooses whether the individual is wearing a veil.

Tadashi Ogino [11] built up the two elements extraction frameworks. One uses customary manual advances and alternate uses machine learning, i.e., a neural system. For the inconsistency recognition framework, we utilize machine learning innovation that we have produced for a digital assault discovery framework. Results affirm that the two models can distinguish atypical occasions in trial video information. Joni-Kristian Kamarainen [12] Gabor filters have kept up their prevalence in includes extraction for very nearly three decades. The first reason that draws consideration was the similitude between Gabor filters and the open field of basic cells in the visual cortex. A more commonsense reason is their accomplishment in numerous applications, e.g., confronts identification and acknowledgment, iris acknowledgment and fingerprint coordinating, where Gabor include based strategies are among the best entertainers. The determination of Gabor highlights is exquisite through the principal areas of flag handling: space (time) and recurrence. Kamarul Hawari Ghazal et. all, [13] was proposed a novel approach in include

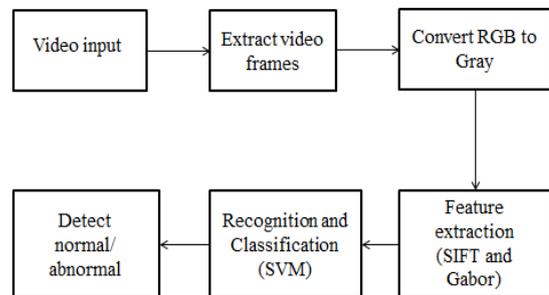
extraction to arrange restricted and expansive weed utilizing SIFT key-focuses descriptor. Specifically we dissect the SIFT key components of weed pictures and outline a calculation to extricate the element vectors of SIFT key-focuses in light of greatness and point heading. Scale Invariant Feature Transform (SIFT) has been turned out to be the most powerful nearby invariant element descriptor. Filter is an approach for distinguishing and removing nearby component descriptors that are sensibly invariant to changes in brightening, picture clamor, pivot, scaling and little changes in perspective. The SIFT calculation is generally utilized for question acknowledgment and discovery which is invariant to light changes and relative or 3D projection.

Juan Zhu et. all,[14] was proposed an acquaints scale invariant component change with the paper recognition framework. The equipment of paper recognition framework comprises of computerized flag processor and complex programmable rationale gadget. The equipment can achieve picture securing and preparing. The product of this framework utilize SIFT technique to recognize the papers. Pengwenlong Gu et. all, [15] proposed a Support Vector Machine (SVM) based Sybil attack detection method is proposed. We present three SVM kernel functions based classifiers to distinguish the malicious nodes from benign ones via evaluating the variance in their Driving Pattern Matrices (DPMs).

**3. METHODOLOGY**

The figure 5 shows the overall workflow of the proposed system. Here the video is considered as a dataset. Those videos are converted into image frames. The frames are converted from RGB to grayscale image. Extract the features from converted grayscale image using the SIFT and Gabor method. Then classify the image using SVM method. The proposed system mainly consists of two approaches

1. Feature extraction
2. Recognition and classification



**Fig 5:** Workflow of a proposed system

**3.1 Feature Extraction**

Image feature extraction in proposed system uses two methods: SIFT and Gabor methods.

The SIFT key elements of weed images and design an algorithm to extract the feature vectors of SIFT key-points based on magnitude and angle direction. For any protest there are numerous highlights, intriguing focuses on the question that can be extricated to give a "component" portrayal of the question. This depiction would then be able to be utilized when endeavoring to find the protest in a picture containing numerous different articles. Filter picture highlights give an arrangement of highlights of a protest that are not influenced by a considerable lot of the intricacies experienced in different strategies, for example, question scaling and pivot. While taking into account a question be perceived in a bigger picture SIFT picture includes additionally take into account protests in numerous pictures of a similar area, taken from various positions inside the earth, to be perceived. Filter highlights are likewise extremely flexible to the impacts of "clamor" in the picture.

To begin with, the RGB shading code pictures had experienced the fundamental preprocessing phase of shading transformation to dark and subjected to the separating procedure of Difference of Gaussian (DoG) system. The DoG is like the Laplace of Gaussian method smoothed by convolution with the Gaussian part. Scale Invariant Feature Transform (SIFT) has been ended up being the most strong neighborhood invariant component descriptor. Filter is an approach for distinguishing and removing neighborhood highlight descriptors that are sensibly invariant to changes in enlightenment, picture clamor, turn, scaling and little changes in perspective.

The DoG is like the Laplace of Gaussian procedure, in which the picture is first smoothed by convolution with the Gaussian piece of certain width  $\sigma_1$

$$G_{\sigma_1}(x, y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{x^2 + y^2}{2\sigma_1^2}\right] \tag{1}$$

To get

$$g_1(x, y) = G_{\sigma_1}(x, y) * f(x, y) \tag{2}$$

With various widths  $\sigma_2$ , a moment smoothed picture can be gotten:

$$g_2(x, y) = G_{\sigma_2}(x, y) * f(x, y) \tag{3}$$

The distinction of these two g's is known as the Difference of Gaussian (DoG) and the choice of  $\sigma$  can be eluded as usage of scale space to the picture.

$$g_1(x, y) - g_2(x, y) = G_{\sigma_1} * f(x, y) - G_{\sigma_2} * f(x, y) \\ = (G_{\sigma_1} - G_{\sigma_2}) * f(x, y) \tag{4}$$

The DoG as an administrator or convolution part is characterized as

$$DoG = G_{\sigma_1} - G_{\sigma_2} \\ = \frac{1}{\sqrt{2\pi}} \left[ \frac{1}{\sigma_1} e^{-(x^2+y^2)/2\sigma_1^2} - \frac{1}{\sigma_2} e^{-(x^2+y^2)/2\sigma_2^2} \right] \tag{5}$$

In picture handling, a Gabor channel, named after Dennis Gabor, is a straight channel utilized for surface investigation, which implies that it essentially examinations whether there are a particular recurrence content in the picture in particular bearings in a restricted locale around the point or district of investigation. Recurrence and introduction portrayals of Gabor channels are guaranteed by numerous contemporary vision researchers to be like those of the human visual framework; however there is no exact proof and no useful reason to help the thought. They have been observed to be especially proper for surface portrayal and separation. In the spatial area, a 2D Gabor channel is a Gaussian bit work adjusted by a sinusoidal plane wave. A few creators assert that basic cells in the visual cortex of mammalian brains can be demonstrated by Gabor functions. Thus, picture examination with Gabor channels is thought by some to be like recognition in the human visual framework.

At the center of Gabor filter based component extraction is the 2D Gabor filter work.

$$\psi(x, y) = \frac{f^2}{\pi\gamma\eta} e^{-\left(\frac{f^2}{\gamma^2}x'^2 + \frac{f^2}{\eta^2}y'^2\right)} e^{j2\pi fx'} \\ x' = x \cos \theta + y \sin \theta \\ y' = -x \sin \theta + y \cos \theta \tag{6}$$

In the spatial space (Eq. (6)) the Gabor filter is a mind boggling plane wave (a 2D Fourier basis function) duplicated by a cause focused Gaussian.  $f$  is the focal recurrence of the filter,  $\theta$  the revolution edge,  $\gamma$  sharpness (data transfer capacity) along the Gaussian real hub, and  $\eta$  sharpness along the minor axis (opposite to the wave). In the given shape, the angle proportion of the Gaussian is  $\eta/\gamma$ . This capacity has the accompanying diagnostic shape in the recurrence space.

$$\Psi(u, v) = e^{-\frac{\pi^2}{f^2}(\gamma^2(u'-f)^2 + \eta^2v'^2)} \\ u' = u \cos \theta + v \sin \theta \\ v' = -u \sin \theta + v \cos \theta \tag{7}$$

In the recurrence space (Eq. (7)) the capacity is a solitary genuine esteemed Gaussian focused at  $f$ . The Gabor filter in (6) and (7) is a simplified adaptation of the general 2D frame from the Gabor's unique 1D "rudimentary capacity". The simplified form authorizes an arrangement of filters self-comparable, i.e. scaled and turned renditions of each other ("Gabor wavelets"), paying little respect to the recurrence  $f$  and introduction  $\theta$ .

### 3.2 Recognition and Classification

Support Vector Machines depend on the idea of choice planes that characterize choice limits. A choice plane is one that isolates between arrangements of items having distinctive class enrollments. A schematic case is appeared in the representation underneath. In this illustration, the articles have a place either with class GREEN or RED. The isolating line characterizes a limit on the correct side of which all articles are GREEN and to one side of which all items are RED. Any new question (white hover) tumbling to the privilege is named, i.e., arranged, as GREEN (or named RED should it tumble to one side of the isolating line).

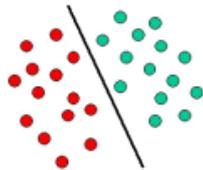


Fig 6: Classic example of a linear classifier

The figure 6 is a classic example of a linear classifier, i.e., a classifier that isolates an arrangement of items into their particular gatherings (GREEN and RED for this situation) with a line. Most characterization undertakings, in any case, are not that basic, and regularly more unpredictable structures are required keeping in mind the end goal to make an ideal division, i.e., effectively arrange new protests (test cases) based on the cases that are accessible (prepare cases). This circumstance is delineated in the representation underneath. Contrasted with the past schematic, unmistakably a full detachment of the GREEN and RED items would require a bend (which is more perplexing than a line). Grouping assignments in view of attracting isolating lines to recognize objects of various class participations are known as hyper plane classifiers as shown in figure 7. Support Vector Machines are especially suited to deal with such assignments.

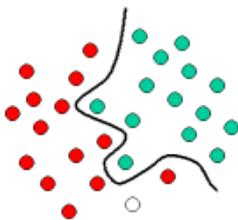


Fig 7: Classic example of hyper plane classifiers

The outline underneath demonstrates the fundamental thought behind Support Vector Machines as shown in figure 8. Here the first questions (left half of the schematic) mapped, i.e., revised, utilizing an arrangement of scientific capacities, known as kernels. The way toward modifying the articles is known as mapping. Note that in this new setting, the mapped objects (right half of the schematic) is directly divisible and, subsequently, rather than developing the mind

boggling bend (left schematic), we should simply to locate an ideal line that can isolate the GREEN and the RED articles.

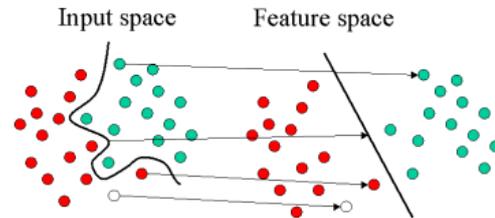


Fig 8: The essential thought behind Support Vector Machines. Here the first protests (left half of the schematic) mapped, i.e., revamped utilizing an arrangement of numerical capacities, known as kernels.

### 4. CONCLUSION AND FUTURE WORK

A new framework is proposed which will be useful for the current world. Money related exchanges of client through ATM are looked by number of robberies. Among these robberies one is connected with the presence of more than individual in ATM room. The focal point of this exploration is to propose an insightful specialist, whose capacity is consolidated with the capacity of ATM programming, to make the exchange of client more secure. Suspicious activities are expanding in the ATM areas. The point of this paper is to keep the ATM wrongdoing by utilizing PC vision strategies. This strategies utilized here create a powerful climate, which assess every minute inside the ATM.

In this paper we have exhibited an extensive near examination of SIFT and Gabor highlight descriptors to separate successful data from video to anticipate movement in ATM preface video dataset. Future extent of this paper is totally open. Examination can be utilized to recognize different uses of highlight descriptors.

### REFERENCES

- [1] Sujith B, "Crime detection and avoidance in ATM: a new framework," International Journal of Computer Science and Information Technologies, 2014.
- [2] Nithya Shree R, Rajeshwari Sah and Shreyank N Gowda, "Surveillance video based robust detection and notification of real time suspicious activities in indoor scenarios," The sixth international conference on computer science, engineering and information technology, 2016.
- [3] Hidetomo Sakaino, "Video-Based Tracking, Learning and Recognition Method for Multiple Moving Objects," IEEE Transactions on Circuits and Systems for Video Technology, VOL. 23, NO. 10, OCTOBER 2013.
- [4] Zhiqian Chen, Vishal Sanserwal, Vikas Tripathi and Monika Pandey, "Comparative Analysis of Various Feature Descriptors for Efficient ATM Surveillance Framework," IEEE International Conference on

Computing, Communication and Automation (ICCCA) 2017.

## BIOGRAPHIES

- [5] Rune Havnung Bakken and Lars Moland Eliassen, "Real-time 3D Skeletonisation in Computer Vision-Based Human Pose Estimation Using GPGPU," IEEE Conference on Image Processing Theory, Tools and Applications, 2012.
- [6] Haitham Hasan and S. Abdul Kareem, "Human Computer Interaction for Vision Based Hand Gesture Recognition: A Survey." IEEE International Conference on Advanced Computer Science Applications and Technologies, 2013.
- [7] Rishabh Agrawal and Nikita Gupta, "Real Time Hand Gesture Recognition for Human Computer Interaction," IEEE 6th International Conference on Advanced Computing, 2016.
- [8] Hong Cheng, Lu Yang and Zicheng Liu, "Survey on 3D Hand Gesture Recognition," IEEE Transactions on Circuits and Systems for Video Technology, VOL. 26, NO. 9, SEPTEMBER 2016.
- [9] Miwa takai, "Detection of suspicious activity and estimate of risk from human behaviour shot by surveillance camera" Nature and Biologically Inspired Computing (NaBIC), 2010 Second World Congress, IEEE 2011.
- [10] Sambarta Ray, Souvik Das and Dr. Anindya Sen, "An Intelligent Vision System for monitoring Security and Surveillance of ATM" India Conference (INDICON), 2015 Annual IEEE.
- [11] Tadashi Ogino, "Anomaly Detection System for Video Data Using Machine Learning", International Journal of Knowledge Engineering, Vol. 2, No. 2, June 2016
- [12] Joni-Kristian Kamarainen, "Gabor Features in Image Analysis".
- [13] Kamarul Hawari Ghazali, Mohd. Marzuki Mustafa and Aini Hussain, "Feature Extraction Technique Using SIFT Key point Descriptors", Proceedings of the International Conference on Electrical Engineering and Informatics Institute of Technology Bandung, Indonesia June 17-19, 2007.
- [14] Juan Zhu, Shuai Wang and Fanying Meng, "SIFT Method for Paper Detection System", IEEE conference on 2011.
- [15] Pengwenlong Gu, Rida Khatoun, Youcef Begriche and Ahmed Serhrouchni, "Support Vector Machine (SVM) Based Sybil Attack Detection in Vehicular Networks", IEEE conference on 2017.



Ms. Arpitha K, M.Tech student in Department of Computer Science and Engineering at Maharaja Institute of Technology- Mysore.



Mr. Honnaraju B, Associate Professor in Department of Computer Science and Engineering at Maharaja Institute of Technology- Mysore.