IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INTELLIGENT VIDEO ANALYTICS AND SURVEILLANCE SYSTEM USING MACHINE LEARNING

[1]Priya P. Morey, [2]Prof. Dr. S. M. Deshmukh, [3]Prof. Dr. N. N. Khalsa

[1]Student M.E.(Digital Electronics), [2]Professor and Head, [3]Associate Professor

[1]Department of Electronics and Telecommunication Engineering

[1]Prof. Ram Meghe Institute of Technology & Reaserch, Badnera, India

Abstract: Surveillance video system framework assumes a huge part in the security frameworks of current urban communities since it can detect suspicious objects and abnormal behavior early instead of the late location by people. Though, the development in surveillance, a gigantic measure of information is produced which make the analytics, storage and retrieval of the information perplexing. Each private or public region today is liked to be under surveillance to guarantee significant degrees of safety. Subsequently the surveillance occurs nonstop, information accumulated thus is immense and requires a great deal of manual work to go during each time of the recorded recordings that is the reason manual perception by security faculty for extended periods isn't plausible nor proficient. There is a requirement for a smart and intelligent framework which can conquer this trouble via computerizing this interaction. It should alarm the security staff when doubtful items in the edge are recognized. It ought to likewise restrain fake alerts. If this is accomplished, crafted by surveillance will be upgraded and the existence of safety recruits can be utilized else-where more workforce is required. In this project, we presented a system which can detect anomalous behavior and alarm the user on the type of anomalous behavior. The model has been trained with videos belonging to anomalous and normal behavioral classes with help of machine learning approach using MATLAB.

Keywords: **Video Processing, Machine Learning, Surveillance System, Intelligent Video Surveillance**

I. Introduction

Video surveillance systems have grown in sophistication over the past two decades, becoming ubiquitous in both privately owned venues (malls, stadium) and public spaces (transportation hubs, city streets). Live monitoring and recorded video serve to deter crime, solve crimes, provide evidence for prosecution, and act as a force multiplier for otherwise limited numbers of security staff. While the number and sophistication of CCTV recording technologies (cameras, servers, network infrastructure, and display walls) has grown, the review of video remains a heavily manual task [1] [5]. It is impossible for a limited number of operators to continually monitor all feeds across a camera network, complicated further by known human-factors issues of visual overload, vigilance, change blindness, and task interruption. In many security settings, such as operation centers of critical infrastructure (airports, rail systems), video is used in a reactive sense. An operator is cued in some way, typically via report of a suspicious person or item, a trip-wire alarm, or automated alert from the video system itself. The operator then needs to find relevant cameras, locations, and time ranges in order to investigate the event [8]. Surveillance is a process of monitoring an area, to keep track of any activities. Every public or private area is preferably kept under surveillance to enhance security measures.

However, Closed Circuit Television (CCTV) cameras produce a huge amount of video data. Effective surveillance can reveal anomalous events in these videos. Detecting crime occurrences, explosions, accidents, etc. are some of the critical tasks in video surveillance. It is challenging to manually find anomalies from the tremendous amount of data available for surveillance purposes. The occurrence of anomalous events that are of interest for monitoring purpose would be very low compared to normal events. Hence, manually analyzing this data to find out anomalous events is a very tedious task and can be labor expensive [12-15]. Anomaly detection in surveillance is a process of monitoring an area to detect any patterns of behavior which show a deviation from the patterns of normal behavior. Conventional methods which use manual observation by security personnel 24x7 is not an efficient or feasible way for surveillance. Suppose if any person carries a knife or a gun in a crowded place, it can go undetected because of the inability of the security personnel to see such objects in the crowd through the naked eye.

Traditional video surveillance systems use real time monitoring mechanism which results into wastage of memory space and could become difficult to find out footage which shows theft occurred. This passive monitoring records surveillance video 24x7 and thus leads to excessive cost. Conventional approaches based on browsing time/camera based combinations are inadequate. Conventional video surveillance systems are not completely automatic. These systems do require human operator who interprets the acquired information and controls the evolution of the events in a surveyed environment. These human operators are responsible for noticing the unusual situations and discovering theft. The requirement of human operators induces to wastage of human efforts and subsequent rise in effective cost. Orthodox video surveillance systems require human hours or specialized software to discover the theft transpired and may require revisiting again and again to identify appropriate footage for investigation. Thus there was a need of a system which would eventually overcome this issue, a system which would be able to count the number of people in the frame and distinguish between safe and suspicious objects. This would lower the burden on the security personnel and make the process of surveillance more thorough and efficient [16-18]. The resulting system with machine learning will not only help in automating the surveillance process but will also help to efficiently minimize human errors and negligence in security aspects.

II. LITERATURE REVIEW

Work presented by various authors is portrayed as an action to determine, identify and scrutinize all publications or researches explores or distributions in a particular territory to introduce the response to each examination question is done and given beneath.

A proposed system developed [1] which can detect anomalous behaviors and alarm the user on the type of anomalous behavior. The dataset used is UCF Crime dataset. The classification, using a CNN-LSTM model achieves an accuracy of 85%. An algorithms proposed [2] for surveillance monitoring student behavior in the laboratory and identified student behavior with activities while learning disorders. The system will alert the Lecture to recognize abnormal behavior of students. A low-cost and efficient approach [3] presented that integrates the use of computational object recognition to perform fully-automated identification, tracking, and counting of human traffic on camera video streams. The implementation provides automated video analytics for medium crowd density monitoring and tracking, eliminating labor-intensive tasks traditionally requiring human operation, with results indicating great reliability in real-life scenarios.

An add-on suite of analytic tools, FOVEA described [4] for the forensic review of video in large-scale surveillance systems. FOVEA is designed to integrate with existing surveillance systems. Tools can be applied to any video stream in an on-demand fashion without additional hardware. A system developed [5] captures video stream, computes the input and the system alerts are generated in real time, which means no additional sensors would be required. The project will benefit to reduce cost of road surveillance system and complete automation of road surveillance system. The detecting a crime which takes place in an enclosed environment and catch the theft in an efficient manner [6]. Moreover, proposed system does not waste its memory by recording the activity unnecessarily. Hence saves lot of wastage of memory of hard disk.

The main goal is to automate identifications of signs of aggression and violence in real-time, which filters out irregularities from normal patterns by utilizing different Deep Learning models (CNN and RNN) to identify and classify levels of high movement in the frame [7]. A framework presented [8] to use the surveillance video to extract useful information such as detection of trucks, their registration number/ownership identification, count of incoming and outgoing trucks, and count of loaded or empty trucks. A review summarizes [9] the information based on abnormality detection, features, system framework and methodology, image acquisition, sample specification, performance analysis and project funding.

A novel intelligent processing and utilization solution to big surveillance video data based on the event detection and alarming messages from front front-end smart cameras is presented [10]. Experimental results reveal that our proposed approach can reliably pre pre-alarm security risk events, substantially reduce storage space of recorded video and significantly speed up the evidence video retrieval associated with specific suspects.

III. SYSTEM IMPLEMENTATION

The block diagram of proposed system is shown in figure 1.

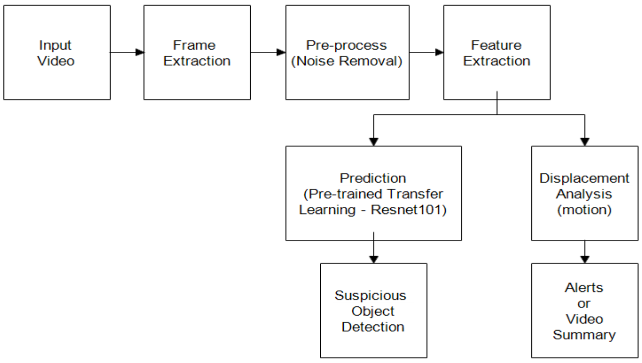

Fig 1: Block Diagram of Proposed System

The proposed work is implemented on Intel Core processor i5, 8GB RAM Laptop configuration and operating system is windows 10. MATLAB R2018a software was used to write the programming code. In this we used image processing, video processing toolbox and statistics and machine learning toolbox. The input video is taken from google, YouTube, surveillance CCTV footage video and some crowed places videos are downloaded.

3.1 Input Video

Basically visual multimedia source that combines a sequence of images to form moving picture is called a video. The input video is taken from CCTV camera and some crowed places videos and normal videos that are without crowd are downloaded from YouTube and uploaded on the given system. Input videos taken in .mp4, .avi and .wmv format.

3.2 Frame Extraction

Frame extraction from video which is sequence is a visual multimedia source that is combination of sequence of images or frames per second to form a moving picture that is video. Sequence is a series of scenes that form a distinct narrative unit, which is usually connected either by a unity of location or a unity of time. Here we used the different frames of the video sequence for further processing.

3.3 Pre-process (Noise Removal)

Pre-processing in which noise removal is performed which aims to remove or attenuate pixels with in the frame, that are not of interest. Noise removal applied on sequence of images, will blur the sequence evolution, smoothing out the temporal variation, like rapid variation in illumination or movement of an object. Adaptive median filter is used for noise removal in frames. Basically adaptive is the process of iterating operation in median filter process. Median filter works on a rectangular region. It changes the size of images during the filtering operation depending on certain conditions as listed below. Each output pixel contains the median value in the 3-by-3 neighborhood around the corresponding pixel in the input images. Zeros however, replace the edges of the images. The output of the filter is a single value, which replaces the current pixel value at (x,y) the point on which S is centered at the time. The following notation is used.

$$Z_{min} = \text{Minimum pixel value in}$$

$$Z_{max} = \text{Maximum pixel value in}$$

$$Z_{med} = \text{Median pixel value in}$$

3.4 Feature Extraction

Feature extraction a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector. The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in lower dimensionality space. The first step aims at finding descriptive visual features of the crowd flow in the observed scene. We decided to exploit the Blob features for detecting local visual features.

3.5 Video Segmentation

Video i.e. frame or image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. In this system image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. BLOB stands for Binary Large object and refers to a group of connected pixels in a binary image. The term "Large" indicates that only objects of a certain size are of interest and that "small" binary objects are usually noise. Blob detection methods are aimed at detecting regions in a digital image that differ in properties, such as brightness or color, compared to surrounding regions. Informally, a blob is a region of an image in which some properties are constant or approximately constant; all the points in a blob can be considered in some sense to be similar to each other. The most common method for blob detection is convolution. The purpose of BLOB extraction is to isolate the BLOBs (objects) in a binary image. A BLOB consists of a group of connected pixels. Whether or not two pixels are connected is defined by the connectivity, that is, which pixels are neighbors and which are not.

3.5 Displacement Analysis of Motion

The crowd behavior in the observed scene is detected by carrying out a statistical analysis by finding MSE i.e. mean square error between consecutive frames on the data collected over a temporal window. As shown in Figure 5, given in input a set of binary blob images (Fig. 5.2a), a 3D-Grid of size m×n×w (Fig. 5.2b) is used to generate a gray scale activity map (Fig. 5.2c). The width m and the height n of the grid are the same of the input image, while the depth corresponds to the length of the temporal window.

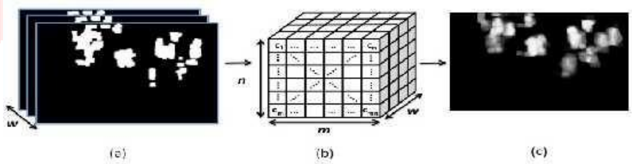

Fig 2: Activity Map Computation: a) A set of blob images is collected over a time window w. b) 3D-Grid of size m×n×w is used to record the time persistence of each pixel. c) The activity map is obtained by clustering the data in the 3D- Grid.

3.6 Prediction (Pre-trained Transfer Learning – Resnet 101)

Pre-trained modelling which is mainly based deep neural network i.e deep learning which is again subset of machine learning, image classification network that has already learned to extract powerful and informative features from natural images and use it as a starting point to learn a new task. The majority of the pre-trained networks are trained on a subset of the Image Net database. These networks have been trained on more than a million images and can classify images into 1000 object categories, such as keyboard, coffee mug, pencil, and many animals. Using a pre-trained network with transfer learning is typically much faster and easier than training a network from scratch. Pre-trained networks have different characteristics that matter when choosing a network to apply to your problem. The most important characteristics are network accuracy, speed, and size. Choosing a network is generally a tradeoff between these characteristics.

ResNet-101 is a convolutional neural network that is 101 layers deep. A residual neural network (ResNet) is an artificial neural network (ANN) of a kind that builds on constructs known from pyramidal cells in the cerebral cortex. Residual neural networks do this by utilizing skip connections, or shortcuts to jump over some layers. Typical ResNet models are implemented with double- or triple- layer skips that contain nonlinearities (ReLU) and batch normalization in between. ResNet network uses a 34-layer plain network architecture inspired by VGG-19 in which then the shortcut connection is added.

3.7 Alert / Video Summary

After sequentially finishing all processes of system model, interpreting that anomaly is in the crowd or not, if any anomaly is in the present scenario it will give result in the form of notification. This notification will be given to the respective security of that crowded place or area. Also major events from video are stored in terms of frames snapshots in output folder with respective time duration of video. So that it is easy for inspection to analyze the event rather than understanding or visualizing whole video.

3.8 Suspicious Object Detection

As we have use pre-trained model which is already trained on n number of objects so as per the accuracy of deep neural network model, it predicted the objects in video frame to detect it is suspicious or normal object. In suspicious object, various objects are considered like gun, clavier, knife etc.

IV. RESULTS AND DISCUSSION

In experimental evaluation, numerous videos are tested for normal or anomaly conditions, in which first video frames extracted shown in figure 3, 7, 10 and then video segmentation is applied with feature extraction as shown in figure 4, 8, 11. After analyzing features, proposed system will predict the results for anomaly or normal conditions as shown in figure 5, 9 and 12. And if any suspicious object is found in any video frame then it's predict its label as object name on video as shown in figure 13.

Fig 3: Extracted input frames of sample video1

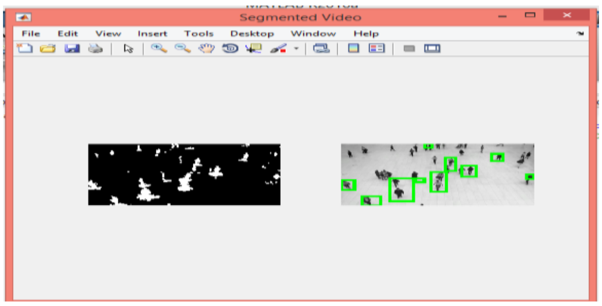

Fig 4: video segmentation and feature extraction

Fig 5: Predicted result of video1

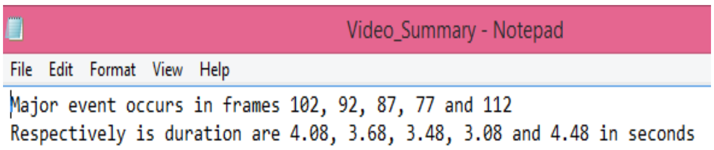

Fig 6: Video summery in text file

Fig 7: Extracted input frames of sample video2

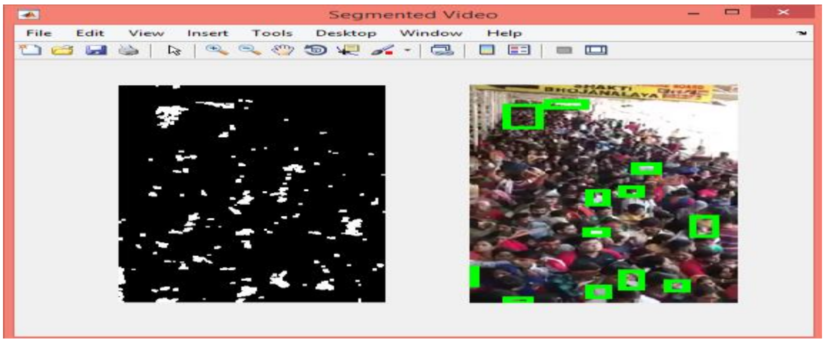

Fig 8: video segmentation and feature extraction

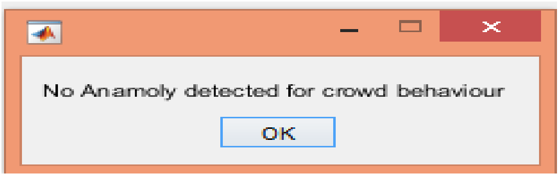

Fig 9: Predicted result of video2

Fig 10: Extracted input frames of sample video3

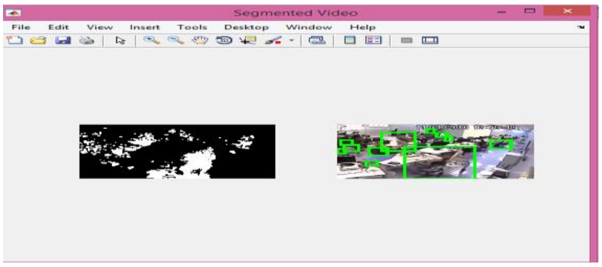

Fig 11: video segmentation and feature extraction

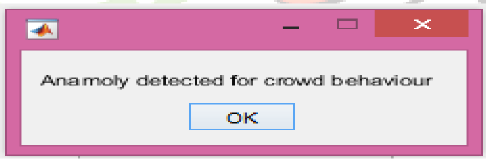

Fig 12: Predicted Result of video3

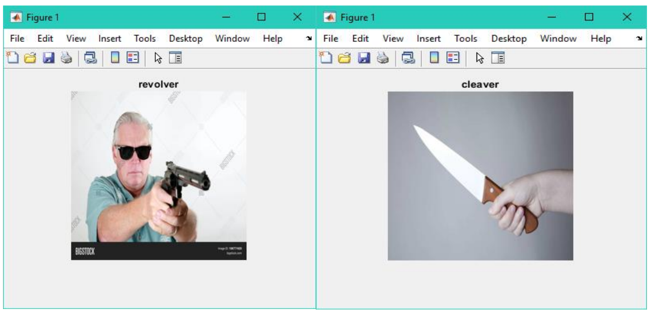

Fig 13: Predicted output of suspicious object in video frames with its predicted label

V. CONCLUSION AND FUTURE SCOPE

In this work, video processing with machine learning algorithm will developed to identifying the anomalies in surveillance video typically crowd scenario. In the first part of the system feature extraction is applied to extracted frames from video sequence. Then developed algorithm pick the suspicious region in the surveillance crowd video. Accordingly, the activity detection and activity analysis classified between the normal and abnormal activities and result is given in the form of notification.

With the increase in public gathering and need of public safety, the scope of improvement in the automatic anomaly detection systems in our country and all over the world is immense. A lot of improvement could be done to our current system. We can improve the system in future by avoiding false target detection and it can be avoided by using pre-trained deep-learning algorithms. Also can improve the feature extraction to improve the accuracy of the system. And can make a portable device based on IOT.

REFERENCES

[1] M.R., M. Makker and A. Ashok, "Anomaly Detection in Surveillance Videos," 2019 26th International Conference on High Performance Computing, Data and Analytics Workshop (HiPCW), Hyderabad, India, 2019, pp. 93-98, doi: 10.1109/HiPCW.2019.00031.

[2] Siruruang Phatchuay, and Mahasak Ketcham, "The Surveillance System for Lab Security based on Image Processing", Int'l Conference on Advanced Computational Technologies & Creative Media (ICACTCM'2014) Aug. 14-15, 2014 Pattaya (Thailand), http://dx.doi.org/10.15242/IIE.E0814542

[3] K. H. Cheong et al., "Practical Automated Video Analytics for Crowd Monitoring and Counting," in IEEE Access, vol. 7, pp. 183252-183261, 2019, doi: 10.1109/ACCESS.2019.2958255.

[4] M. DeAngelus, R. Duarte, Z. Elko and J. Thornton, "On-demand Forensic Video Analytics for Large-Scale Surveillance Systems," 2019 IEEE International Symposium on Technologies for Homeland Security (HST), Woburn, MA, USA, 2019, pp. 1-7, doi: 10.1109/HST47167.2019.9033004.

[5] G. Desai, V. Ambre, S. Jakharia and S. Sherkhane, "Smart Road Surveillance Using Image Processing," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5, doi: 10.1109/ICSCET.2018.8537279.

[6] Mandar Shriram Munagekar, "Smart Surveillance system for theft detection using image processing", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 05 Issue: 08 | Aug 2018

[7] Virender Singh, Swati Singh, Pooja Gupta, Real-Time Anomaly Recognition Through CCTV Using Neural Networks, Procedia Computer Science, Volume 173, 2020, Pages 254-263, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.06.030.

[8] A. S. Patel, O. P. Vyas and M. Ojha, "Vehicle Tracking and Monitoring in Surveillance Video," 2019 IEEE Conference on Information and Communication Technology, Allahabad, India, 2019, pp. 1-6, doi: 10.1109/CICT48419.2019.9066256.

[9] Sikandar, Tasriva & Ghazali, Kamarul & Rabbi, Mohammad. (2018). ATM crime detection using image processing integrated video surveillance: a systematic review. Multimedia Systems. 2018. 10.1007/s00530-018-0599-4.

[10] Shao, Zhenfeng et al. "Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data." IEEE Transactions on Big Data 4 (2018): 105-116.

[11] V. C. M. Vishnu, M. Rajalakshmi, and R. Nedunchezhian, ''Intelligent traffic video surveillance and accident detection system with dynamic traffic signal control,'' Cluster Comput., vol. 21, no. 1, pp. 135–147, 2018.

[12] C. Ma, D. Liu, X. Peng, L. Li, and F. Wu, ''Traffic surveillance video coding with libraries of vehicles and background,'' J. Vis. Commun. Image Represent., vol. 60, pp. 426–440, Apr. 2019.

[13] D. Karthikeswaran, N. Sengottaiyan, and S. Anbukaruppusamy, ''Video surveillance system against anti-terrorism by using adaptive linear activity classification (ALAC) technique,'' J. Med. Syst., vol. 43, no. 8, p. 256, 2019

[14] V. Tsakanikas and T. Dagiuklas, ''Video surveillance systems-current status and future trends,'' Comput. Elect. Eng., vol. 70, pp. 736–753, Aug. 2018.

[15] D. Frejlichowski, K. Gościewska, P. Forczmański, and R. Hofman, ''Application of foreground object patterns analysis for event detection in an innovative video surveillance system,'' Pattern Anal. Appl., vol. 18, no. 3, pp. 473–484, 2015.

[16] B. Cui, J. Cui, and D. Yong, "Intelligent security video surveillance system based on deviance technology," Journal of Mechanical Strength, vol. 42, no. 1, pp. 1–12, 2016.

[17] V. Khatri, "Intelligent video surveillance using soft biometrics," Xitsonga Lila You Shavian/System Engineering Theory and Practice, vol. 35, no. 5, pp. 12-13, 2015.

[18] N. Babaguchi, A. Cavallaro, R. Chellappa, F. Dufaux, and L. Wang, "Special issue on intelligent video surveillance for public security and personal privacy," IEEE Transactions on Information Forensics & Security, vol. 16, no. 1, pp. 8–15, 2017.

[19] http://web.ipac.caltech.edu/staff/fmasci/home/astro_refs/Digital_Image_Processing_2ndEd.pdf.

[20] Digital Image Processing by S. Jayraman, S. Esakkirajan and T. Veerakumar.

[21] A Comparative Study of Three Image Matching Algorithms: Sift, Surf, and Fast https://pdfs.semanticscholar.org/3f67/3caedafdfea1c5b4e77118792b6a22fa4998.pdf.