IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

AN INTELLIGENT VIDEO SURVEILLANCE SYSTEM FOR DETECTING ANOMALOUS BEHAVIOR OF HUMAN

Dr. Ganesh V Patil, Associate Professor, Shradha Swami, DYPCET, Kolhapur

ABSTRACT:

Abnormal detection refers to as a result of unusual or exceptional striking activity that may reveal a person or group of people to danger in particular context. There has been an increase in the number of offensive or disruptive activities recently. Because of this, security has been given the utmost importance. If a single person is watching a live CCTV screen more than 20 minutes it can leads to loss 95% of his/her eye vision. Human monitoring of video is very laborious task and requires a work force and their constant attention. To automate this process here we are creating a computer vision-basedsurveillance system to notify users. Video surveillance system plays an important role in keeping indoors and outdoors environment secure. Various surveillance methods were introduced for security purposes. The current advancement in computer vision and machine learning has outstanding role in enabling such intelligent framework. Different algorithms that are specially designed for building smart vision framework seek to understand and build correct semantic inference from observe motion cause by target. This paper addresses an overview most recent research algorithm that focuses on abnormal behavior detection in surveillance application.In this paper we have used YOLO technique and TensorFlow's MoViNettechnique to detect these abnormal activities.

Keywords: *Video surveillance, Abnormal Detection, Learning Methods, YOLO, MoViNet.*

INTRODUCTION:

Nowadays, because of the wide spread of CCTV, there is considerable research effort to improve analysis for surveillance videos together with machine learning technique for the seek of autonomous analysis of such data resources. Although video capturing devices are extremely common in today's world, available human resources to observe an analyse the video clips are very limited and mostly cheap[4].In French, the word "Sur" means "From above" and "Veiller" means "to watch". Surveillance means monitoring moments, activities, and behavior in order to manage, control and protect people [10].

Outlier detection is broadly studied topic which has been utilize in many areas like Surveillance environment, human activity recognition systems, abnormal activity recognition systems, entertainment environments, healthcare systems [4]. In order to detect abnormal activities, it is necessary to determine all possible activities that can be performed. [1]. An abnormal activity is defined as a result of extra-ordinary and unusual activity that may reveal a person or group of people to danger in particular context.Abnormal activity occurs very rarely.

Image processing research in video surveillance is one of the most active areas and computer vision. Firstly, video surveillance started with analogue CCTV system and sensors and motion detectors to gather information and to monitor people. This human monitoring of surveillance is very laborious task. To overcome this problem many solutions was added. As a result of machine learning and deep learning.

Technologies play an important role in security surveillance in developed countries Video surveillance cameras in these countries are equipped with advanced technology. The image -based technique are widely used in surveillance [5]. 2008 Mumbai Terrorist attack is observe as major event in the Indian history that has impacted the video surveillance market in India very strongly [9]. Various govts. are funding the areas of security specially surveillance. A multi-institution project on video surveillance and monitoring (VSAM) was funded by DARPA (Defence Advanced Research Project Agency).[8].

A proposed system integrates object detection and deep learning techniques to detect abnormal behaviour of humans automatically without human intervention.

Benefits of Video Surveillance:

- Availability- With the developments in surveillance system we can implement this system everywhere like school, colleges, parks, hospitals, old age home.

- Real-time monitoring- Installation of CCTV and It is common practice for most organizations to monitor their employees continuously. Countries having a population of millions every person is captured by camera many times a day. A lot of videos are generated and stored. Since, constant monitoring of videos requires a work force and constant attention. Real-time monitoring system is designed to automatically detect suspicious activities.

RELATED WORK:

Principal component analysis (PCA) is the first face recognition technique for intruder detection, and it uses the eigenface produced by PCA in the face image to identify faces [13]. Other illustrative techniques include the use of face image reconstruction using the discrete cosine transform (DCT) coefficient to enhance PCA performance [14], the combination of PCA and linear discriminate analysis (LDA) [15], and the use of PCA and LDA for face image reconstruction at distances between 1 and 5 m [16]. The use of convolution neural networks (CNN) to extract features with local binary patterns (LBP) and categorise them is another way [17]. The last technique involves using face expression change, illumination change, and altering angle images to recognise faces with Restricted Boltzmann Machines (RBM) of deep learning [18].

The two main categories of fire detection techniques are sensor-based and image-based. The method for using sensors is further broken down into two categories: the method for using a device made by combining 8 sensors of AMS MOX sensors, PID sensors, and NDIR CO2 sensors et [21] and the method for using a device made by using the values obtained by using temperature sensors, smoke sensors, and CO2 sensors as parameters of the Fuzzy logic [20].

The two types of fall detection techniques are sensor-based and image-based techniques. The two methods for employing sensors are the method for using 3-axis accelerometer sensors and pressure sensors [23] and the method for attaching accelerometer sensors to a body to utilise sensor values and acceleration values [22]. The method for recording typical acceleration patterns as Activities of Daily Living (ADL) and compare them with ADL based on real-time nearest neighbour rule (NNR) is another way to use the accelerometer sensors of mobile phones [23].

Intelligent video surveillance system

The methods for employing computers and small devices are two divisions of the intelligent video surveillance system for sending an alarm when an abnormal situation occurs in video. The computer-based intelligent video surveillance system has been studied and performs a variety of detections. However, they have high power consumption, installation and maintenance expenses, and privacy leaks, making them unsuitable for usage in a real environment. An object tracking system with multiple cameras is an example of a traditional computer-based intelligent video surveillance system. Real-time streaming protocol (RTSP) is used to send the audio and video encoding standard (AVS)-encoded videos from IP cameras with digital signal processors (DSP) to the IP network. In order to carry out distributed processing, the IP network uses GPU [24].

Video surveillance safeguards both the business and the employees. Setting up a surveillance system either inside or outside the office helps catch thieves who steal or damage company property. The camera footage will assist them in their investigation and in locating the offender. Finding someone or a group of persons who are about to attack a resident of a home [28]–[29] and being able to employ the In order to protect our privacy and safeguard our assets, there is an increasing need for user-friendly security solutions.[25] Shahad et al. [26] demonstrate a complex event processing-based intelligent surveillance system that uses data correlation to identify intrusions. On the basis of recognising of patterns in data sequences acquired from door sensors and security cameras, the engine also uses four classifiers to forecast the occurrence of occurrences. An autonomous surveillance system based on video that can monitor a scene in real time and spot abnormalities is suggested by Al-Nawashi et al. [27] for use in academic settings. The three modules of this system, which use a support vector machine type classifier, are pre-processing, detection of anomalous human behaviour, and content-based image retrieval phase.

PROPOSED WORK:

The proposed approach is focused on quickly identifying anomalous activity. Using a state transition table, which contains all potential states, the abnormal actions are reported. The system is taught to categorise the activities carried out by people and report anomalies.

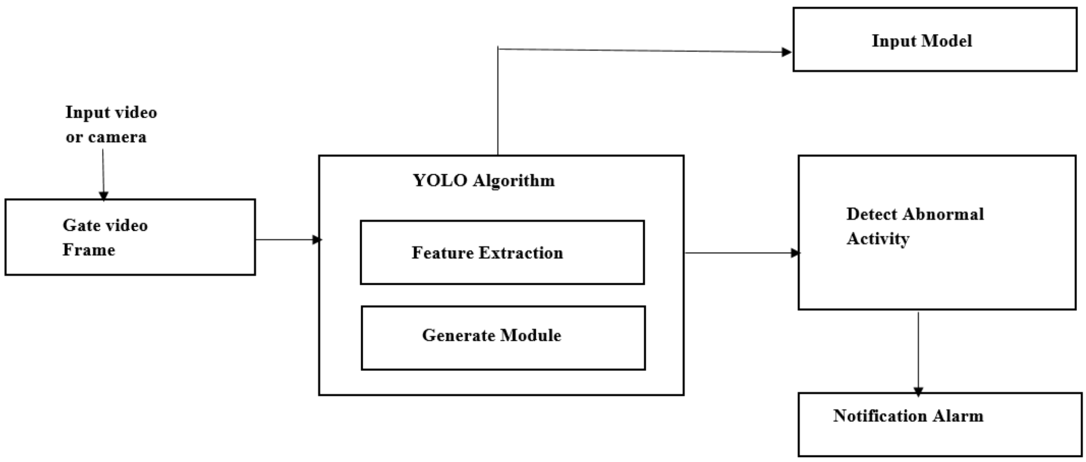

Fig. 1. Architecture Diagram of proposed system.

The methods rely on a kinetic model for action recognition, a mechanism for object detection utilising image processing, and YOLO for feature extraction. The strategy is intended to increase the true classification rate, which will increase the system's reliability.

The suggested system carries out the various phases of the objective in the control room. The following are these phases:

A.Camera/ Video Plugin

In places like hospitals, parks, and school grounds, cameras are installed. Through the network, this camera is connected to the control room. The suggested system's surveillance area is defined by the camera's field of view. The sensed data is captured, gathered, and then continually communicated to the application, known as the receiver, which is located at the hospital or any other remote location for assistance if any abnormal activity occurs. Continuous monitoring and comparison of captured data with various characteristics, traits, and patterns are performed.

B. Feature Extraction and Frame Generating

Here, we break up videos into frames in order to extract features more effectively. The video will be more fluid the higher the number of frames. The YOLO algorithm seeks to accurately and quickly detect objects.

The YOLO algorithm consists of 24-layer, convolutional neural network for feature extraction and two fully connected layers for predicting the probabilities.

YOLO:

The YOLO algorithm which stands for "You Look Only Once", is a popular object detection algorithm that uses CNN to identify objects in images and video frames. It's known for its speed and accuracy and it is used in variety of applications YOLO is a single-stage detection method that moves across the network in a single step while handling object detection and categorization. Additionally, YOLO is thought to be more efficient in terms of accuracy and handling speed. Since its original introduction in 2016 [7], the YOLO model family has continued to develop.

A single Convolutional Neural Network is used by the YOLO algorithm to generate predictions. An S x S grid of cells is first created from the input image. Predicting B bounding boxes and their associated class probabilities is the responsibility of each cell. Based on their confidence score, the anticipated bounding boxes are subsequently filtered. In order to eliminate any unnecessary bounding boxes, non-maximum suppression is then used [31–32].

The architecture works as follows:

- Before passing the input image through the convolutional network, it is resized to 448x448.

- To create a cuboidal output, a 3x3 convolution is used after a 1x1 convolution to reduce the number of channels.

- With the exception of the final layer, which employs a linear activation function, the activation function used internally is ReLU.

- A few extra methods, such batch normalisation and dropout, respectively regularise the model and stop it from overfitting.

C. Train Dataset and Create Model

In this phase we are working on extracted features from the frames using YOLO algorithm. We have used TensorFlow's MoViNetspretrained model for action recognition.

MoViNets:

MoViNets, a feature of TensorFlow, presents a three-step method to increase computational effectiveness while significantly lowering the peak memory utilisation of 3D CNNs. A family of CNNs called MoViNets handle video streams effectively and predict outputs accurately with a small fraction of the latency of CNN video classifiers.

MoViNet is an efficient and lightweight neural network architecture developed by google. It is designed for real-time video understanding and can be used for task such as action recognition, object detection and semantic segmentation.

On a number of sizable video action detection datasets, the model has proven to be quite accurate and effective. It accomplishes this in three important steps:

1. Create a search space for the video network for the neural architecture to produce effective and varied 3D CNN designs.
2. The Stream Buffer approach is used to isolate memory from the length of the video clip, allowing CNNs to include streaming video of any length with decreased memory use.
3. A process of assembly that increases precision without reducing efficiency.

The Kinetic-600, a collection of a sizable and high-quality set of URL links to 650,000 video clips, is used to train MoViNets. The dataset comprises of human-annotated video clips that show 400/600/700 different human activity classifications, such as interactions between people and objects. MoViNets can recognise 600 human movements thanks to training, including bowling, robot dancing, and playing the trumpet. It can also instantly categorise video streams that were recorded with a contemporary smartphone.By minimising their drawbacks, MoViNets enables users to take advantage of the advantages of 3D video classifiers and 2D frame-based classifiers. It accomplishes this using a hybrid methodology with causal convolutions in place of 3D CNNs.

For temporal data, a type of convolution called causal convolutions is employed to guarantee that models cannot deviate from the ordering in which the data was modelled. This enables users to use a Stream Buffer to cache intermediate activations across frames. The method duplicates the input activations of all 3D operations that the model outputs and feeds them back into the model at the beginning of the subsequent clip. As a result, MoViNets are able to accept input one frame at a time. This lowers the peak memory consumption without sacrificing accuracy. Given that the model in 3D CNNs processes every frame of a video clip at once, this consumes a substantial amount of memory.

Additionally, Neural Architecture Search (NAS), a widely utilised method for automating the creation of artificial neural networks, is employed to look for effective model configurations. On video datasets throughout network width, depth, and resolution, it looks for model configurations. It develops a collection of action classifiers that produce predictions that transition smoothly over time based on the content of the frames.

Kinetics-600 Datasets:

The Kinetics-600 Dataset, which contains over 480K films from 600 different action categories, is a sizable action recognition dataset. For training, validation, and test sets, the 480K videos are segmented into 390K, 30K, and 60K segments, respectively. Each video is a 10-second action moment clip annotated from the original YouTube video. The Kinetics-400 dataset has been expanded [30].

D.Detect Anomalous Behaviour

By identifying the defined actions to correspond with the relevant area, anomalous activity can be discovered. Finding all potential actions that can be undertaken from the current activity allows for the detection of abnormal activity.

E. Alert to User

Whenever defined activity which is categorized as abnormal is occur system buzzesan alarm to the remote area or to the control room.

RESULT ANALYSIS:

The proposed system is implemented in python and the system is trained using YOLO and TensorFlow's movinet. The experimental setup uses the dataset collected from online. Around 480K videos from 600 different action genres make up the dataset. For training, validation, and test sets, the 480K videos are segmented into 390K, 30K, and 60K segments, respectively. From the data received from the camera help classifying the activity with the help of feature extracted. The model is trained using pre-trained model movinet.

After completion of the training and loading of the CNN using YOLO weight; videos are used to test efficiency and performance of GUI. As expected, videos were perfectly operated on each object in the scene and independently detected. The model was able to detect objects, an abnormal activity in live video streaming as well as already uploaded/loaded with good accuracy. Surveillance on an IP camera streaming as well as uploaded video resulted in detecting motion with perfect bounding boxes and labelling around the object.

The study result is shown in following figure:

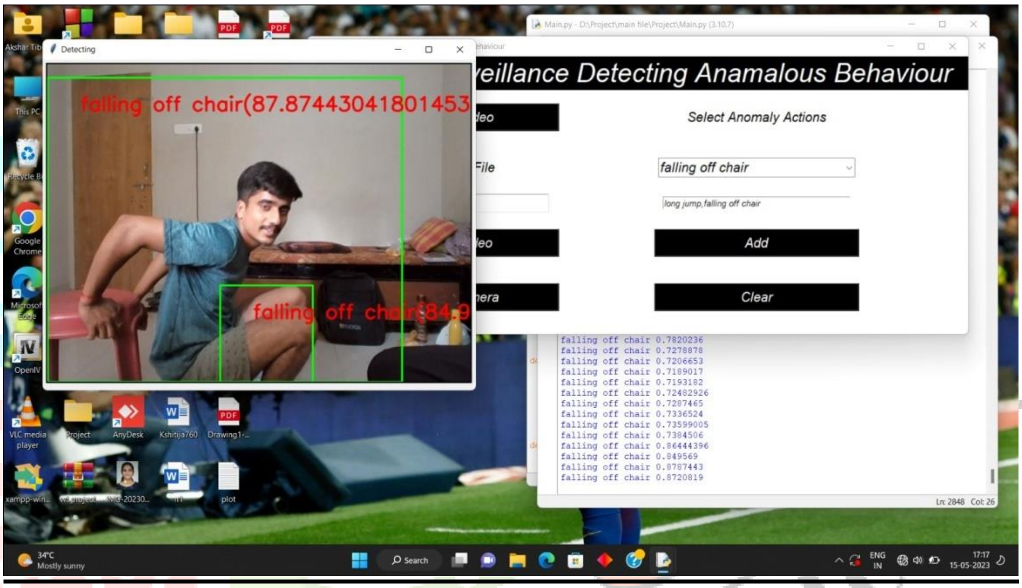

Fig.2. Prediction using proposed model

CONCLUSION:

Intelligent video surveillance using the YOLO algorithm for object detection obtained phenomenal performance. Once models were trained on available data the system was able to identifying objects and classes as well as detecting motions. The proposed system significantly contributes to situation awareness. Smart surveillance systems can respond instantly through real-time video analysis. Our technology detects the intrusion and notifies an appropriate party so that the intrusion may be dealt with.

REFERENCES:

1. Kshitija Deshmukh, Aishwarya Kokane, Shweta Konde, Guide- Gauri Virkar. "Abnormal Activity Detection from Video using SVM Algorithm.". International Research Journal of Engineering and Technology. Volume: 07.
2. Shian-Ru Ke, Hoang Le Uyen Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-HeeYoo."A Review on video-Based Human Activity Recognition.". MDPI (Computers 2013)ISSN 2073-431x.
3. NorSurayahaniSuriani, Aini Hussain and Mohd AsyrafZulkifly. "Sudden Event Recognition: A Survey.".MDPI (Sensors 2013) ISSN 1424-8220.
4. Ahlam Al-Dhamari, Rubita Sudirman, NasrulHumaimi Mahmood."Abnormal Behavior Detection In Automated Surveillance Videos:Review".Journal Of theoretical and Applied InfromationTechnology:ISSN 1992-8645.
5. Tarun Kumar, Sanjiv Kr. Pippal, Aishwarya Mishra AlloraDudi, Vinod Chaudhary."A Video Surveillance System For Unmanned Surveillance of Cantonment Boundry".Intenational Journal Of Innovative Technology and Exploring Engineering:ISSN 2278-3075, Volume:09.
6. JinSu Kim, Min-Gu-Kim and Sung Bum Pan"A study of Implementation of Real Time Intelligent Video Surveillance System Based on Embedded Module".EURASIP Journal on Image and Video Processing.OpenAccess.
7. Mehmet Sirin Gunduz.GultekinIsik"A New YOLO -Based Method For Real-Time Crowd Detection From Video and Performance Analysis Of YOLO Models"Journal of Real -Time Image Processing(2023).
8. Liang Wang,WeimingHu,Tieniu Tan."Recent DevelopmentinHuman Motion Analysis". National Laboratory of PatternRecognition Institute Of Automation, ChineseAcademy Of Sciences ,Beijing,P.R.China,100080.
9. CMR's India Video Surveillance Report(2011-16)
10. PreethiVennam,PramodT.C,ThippeswamyB.M,Yong-Guk Kim and Pavan Kumar B.N."Attacks and preventive measures on video surveillance systems:Areview".AppliedSciences.MDPI(2021).
11. SumonGhosh,PrashamShah,AdityaGhadage,VaibhavSanghavi,Dr.Vaqar Ansari."Suspicious activity detection (2022)"Intenational Journal Of Innovative Technology and Exploring Engineering:ISSN2321-9653.
12. Mrs.NaziyaFarhat,MohammadNasiruddian."Computer vision based surveillance system (2022)"IJRASET,ISSN:2321-9653,Volume:10
13. M.A. Turk, A.P. Pentland, Face recognition using eigenfaces. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 586–591 (1991) J.Z.
14. He, Q.H. Zhu, M.H. Du, Face recognition using PCA on enhanced image for single training image. IEEE International Conference on Machine Learning and Cybernetics, 3218–3221 (2006)
15. J. Li, B. Zhao, H. Zhang, Face recognition based on PCA and LDA combination feature extraction. IEEE International Conference on Information Science and Engineering, 1240–1243 (2009)
16. H.M. Moon, D.J. Choi, P.K. Kim, S.B. Pan, LDA-based face recognition using multiple distance training face images with low user cooperation. IEEE International Conference on Consumer Electronics, 7–8 (2015).
17. H. Zhang, Z. Qu, L. Yuan, G. Li, A face recognition method based on LBP feature for CNN. IEEE Advanced Information Technology, Electronic and Automation Control Conference, 544–547 (2017)
18. J. Zeng, Y. Zhai, J. Gan, A novel sparse representation classifcation face recognition based on deep learning. UICATC-ScalCom, 1520–1523 (2015).
19. G.J. Jang et al., Recognition of fre levels using fuzzy reasoning. IEEE International Conference on Industrial Mechatronics and Automation. 2, 557–560 (2010)
20. A. Solozano et al., Fire detection using a gas sensor array with sensor fusion algorithms. ISOCS/IEEE InternationalSymposium on Olfaction and Electronic Nose, 1–3 (2017)
21. U. Lindemann et al., Evaluation of a fall detector based on accelerometers: a pilot study. Med. Biol. Eng. Comput. 43(5), 548–551 (2005)

22. Z. Liu, Y. Song, Y. Shang, J. Wang, Posture recognition algorithm for the elderly based on BP Neural Networks. IEEE Chinese Control and Decision Conference, 1446–1449 (2015)
23. C. Medrano et al., Personalizable smartphone application for detecting falls. IEEE-EMBS International Conference on Biomedical and Health Informatics, 169–172 (2014)
24. S. Zhang et al., On the design and implementation of a high defnition multi-view intelligent video surveillance system. IEEE International Conference on Signal Processing, Communication and Computing, 353–357 (2012)
25. Bandi Narasimha Rao, Surveillance Camera using IoT and Raspberry Pi, Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020) IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2
26. Tan Zhang, AakankshaChowdhery, Paramvir Bahl, Kyle Jamieson, Suman Banerjee "The Design and Implementation of a Wireless Video Surveillance System" University of Wisconsin-Madison, Microsoft Research Redmond, University College London
27. C M Srilakshmi1, Dr M C Padma2 "IOT BASED SMART SURVEILLANCE SYSTEM" International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 05 | May -2017
28. Mrs. Prajakta Jadhav1, Mrs. Shweta Suryawanshi2, Mr. Devendra Jadhav3 "Automated Video Surveillance eISSN: 2395 -0056 | p-ISSN: 2395-0072" International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 05 | May -2017
29. Pawan Kumar Mishra "A study on video surveillance system for object detection and tracking" Nalina.P, Muthukannan .
30. Ang Li, Joao Carreira etc. "The AVA- Kinetics localise human actions video dataset".
31. "Algorithmo You Look Only Once".https://g.co/kgs/w1nW9u.
32. "Introduction to YOLO algorithm for object detection".https://www.section.io/engineering-education/introduction-to-YOLO-algorithm-for-object-detection/
33. "MoViNet for streaming action recognition".https://www.tensorflow.org/hub/tutorials/movinet