ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

Detection of Malfeasance through Intellectual Eye

[1] Mylavarapu Kavya, [2] Dr.S.Venu Gopal, [3] P.Aditya sasank, [4] S.Vijay Simha Reddy

[1] [2] [3] [4] Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad, India
Corresponding Author Email: [1] kavyamylavarapu2306@gmail.com, [2] adityasasank53@gmail.com, [3] svreddy873@gmail.com, [4] s.venugopal@vardhaman.org

Abstract*— Observation security is a highly boring and la- borious task. To determine whether the caught workouts are unusual or suspicious, needs a labor force and their constant consideration. Here, we'll build a framework to automate the task of reviewing video reconnaissance. We will regularly review the camera stream to look for any suspicious or unexpected activities. There have been improvements in deep learning calculations for deep reconnaissance from prior encounters. These advancements have identified a crucial trend in meticulous reconnaissance and indicate a material rise in effectiveness. Keen observation is typically used for burglary identifying evidence, brutality detection, and assessment of explosive potential. In this project, we'll introduce a spatially-based auto-encoder that relies on a 3D convolutional neural network. The decoder then recreates the edges after the encoder section has removed the spatial and temporal data. By tracking the reproduction misfortune using Euclidean distance between the original and reproduced batches, the odd occurrences are recognized.

Index Terms*—Malfeasance, Surveillance, ImageAI, reconnais- sance.

I. INTRODUCTION

To improve public safety, surveillance cameras have prolif- erated in public spaces. Since it takes a big staff and ongoing monitoring, it is nearly impossible for authorities to determine if the situations are suspicious given the vast number of recordings that are produced and preserved for a length of time. Automation of surveillance is essential for convenient monitoring. Intelligent surveillance aids in enhancing public safety by automatically identifying crimes. This study uses a variety of Deep Learning models to identify and categorize high amounts of movement in the frame. With this project, videos are broken up into pieces. In the case of a threat, it determines how long the threat will last, and a detection alert is triggered, identifying any suspicious activity occurring at that precise moment. Two categories—threat (abnormal activity) and safe—are used to categorize the films in this research (normal activities).

Deep learning techniques are applied to address current issues, producing extraordinary outcomes in the identification and classification of activities. The CNN and RNN neural networks were the two neural networks employed in this investigation. The primary function of the fundamental neural network CNN is to extract sophisticated feature maps from recorded data. High-level feature maps are extracted to lessen input complexity. Modern object recognition models take into consideration a huge number of parameters and as a result, take much longer to fully train. For this reason, a pre- trained model is used. By initially taking into consideration the previously learned model for a set of categorized inputs, such as ImageNet, which can then be retrained using fresh weights given to various new classes, the transfer learning method would enhance this job. The CNN's output is sent into the RNN as an input. The following item in a sequence may also be predicted by the RNN. So, its primary function is that of a forecasting engine. The purpose of employing this neural network in this study is to provide meaning to the recorded series of motions and movements. This network's core layer includes an LSTM cell, followed by various hidden layers with the proper activation functions and an output layer that delivers the final 13-group categorization of the video (12 anomalies and 1 normal). To prevent and spot any suspicious behavior, this system's output is utilized to undertake real-time monitoring of various organizations' CCTV cameras.

The crime was examined using the data mining technique. If the intruder handled any weapons, the security camera is instantly noticed, which is crucial for avoiding crime. This method makes crime prevention the simplest task possible. Crimes are identified using machine learning and deep learning techniques. The difficult challenge of object detection is also one that is crucial to solving crimes. Differentiating the frame, optical flow, and background removal are the three steps of object detection. To remove the moving forepart from the orig- inal surroundings, background subtraction is performed. Facial expression The next step in crime detection is identification, which is used to recognize a criminal's face in recordings or pictures. Analyzing the culprit takes more time, but the crime is accurately detected. Face matching is the process of finding a human face, whether it be in a group or on its own. The suggested technique used video surveillance to locate the crime. The research was based on video surveillance utilized for criminal detection, and this paper will discuss the various literature studies of the researchers. It discusses all the details about crime detection as well as the challenges faced by legal and administrative entities.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

II. RELATED WORK

To maintain security at public places there are some existing methods using different methodologies.

A. Feature Selection Using Optimization Mechanism

Clustering is an exploratory data exploration and exploratory statistics discovery procedure. By utilizing several strategies, the Clustering Technique was able to cluster the data. In data mining and analysis, it is crucial. An article by Rasoul Kiani focused on the clustered crimes that occurred over several years. The Fast Miner program utilizes the outlier detection Genetic Algorithm to improve. As a consequence, the maximized and non-maximized parameters were matched to assess the effect and quality. In the recent past, detecting and suppressing crime needed years of investigation and inspec- tion. Failure clustering is the K-mode approach that is most frequently utilized. Farhad combined K-modes with Elephant Herding Optimization to solve this problem. The suggested model thus exhibits complete inaccuracy. The comparison table between the K-Mode algorithm and the suggested tech- nique is based on purity and accuracy. Many studies are now being conducted in the area of criminal detection, however, there is currently no cutting-edge technology available. CCTVs are often utilized in the neighborhood to curb crime, yet crime control hasn't improved at all. Umadevi V suggested the intention Of a crime-detecting program as a solution to this problem. It locates the crime in the cameras and notifies the organizer so they may take appropriate action. In this suggested approach, the already trained model VGGNet-19 was utilized to identify the criminal intent. Faster RCNN, also known as Fast Regional based Convolutional Neural Network, is a method that is used to create a square box around suspicious photos.

B. Crime Detection Using Facial Expression

To match a human face from pictures or videos and to classify biostatic security, facial recognition systems are utilized. Human faces vary in size, hue, and other aspects. The origins of facial data gathering have greatly expanded in recent years. C. Anitha offered a thorough analysis of data-gathering methods that may be employed for facial utterance identification systems to address this. Automated Facial Utterance Identification has been analyzed since the 1990s, and there have been improvements in recognizing faces and facial expressions. Several techniques were employed to determine the facial utterance. A study on facial specifications utilizing the Face Activity Model was presented by C. P. Sumathi. The offered research intends to demonstrate a clear analysis of rec- ognizing facial utterances. Syeda Amna Rizwan's presentation of opposing facial emotions shows how facial utterances may be recognized using several landmark identifiers and locally changed attributes. The suggested approach was divided into four categories, and the accuracy and performance of the advanced facial utterance identification are good. As a result, it applies to many consumer-relevant domains, and the face is recognized using the YCbCr color space. fundamental data on the spirit is created from images with geo-temporal documents and is then used to estimate crime. Tiafan Zhang suggested using this technique to recognize facial emotions. As a result, it improves how criminal divination works. While several studies have been conducted on the subject, it is still challenging to identify crimes using this method. In detecting crimes Facial utterance plays a significant role in detecting crime. To solve this problem, Ashraf Abbas proposed a technique that uses facial expressions during surveillance to identify crimes and identify suspects before they perform banned acts. The outcome demonstrates that guards are more valuable and successful in apprehending offenders. Matching faces demonstrates the facial identification of offenders and keeps track of people's movements inside crowds. The crime was stopped by taking the necessary precautions after observing the criminal's facial expression. Also, it is a powerful way to stop illegal activity.

III. METHODOLOGY

The suggested paradigm emphasizes a thorough specification that is utilized to spot shady conduct. The number of criminal records is rapidly rising. It is exceedingly difficult for a human to keep a watch on every location on earth to stop these illegal actions. Hence, whenever the algorithm is taught to identify suspicious activities using deep learning, we prefer to propose our model. Spatiotemporal Auto Encoders, a pre-trained deep convolution neural network, and a recurrent neural network are employed for preliminary categorization and final detection of suspicious behavior, respectively.

Initially, a live video feed received from CCTV is sent to the system. The video is then converted into frames with a pre- determined, brief time period (say 1 frame per second). These frames are sent to the 3D convolution network-based spa- tiotemporal auto-encoder encoder. The decoder reconstructs the frames after the encoder takes the spatial and temporal information. By calculating the reconstruction loss using the Euclidean distance between the original and rebuilt batch, the aberrant events are found.

The collection of these frames is what's used to categorize the CCTV stream in real-time. The single-merged feature map is provided to the 3D-CNN as input. We created an LSTM cell in this manner to reduce the training time. Using data from the UCF-Crime dataset, this 3D-CNN was trained. The Kaggle dataset for UCF-Crime is used. The 1900 clips that make up the UCF-Crime dataset are each 60 to 600 seconds long, have varying resolutions, and were captured by actual-world security cameras. The 13 true irregularities that this dataset is designed to find are cruelty, incarceration, fires, attacks, crashes, robberies, eruptions, conflict, theft, murder, theft, snatching, and vandalism. The probabilistic categorization is determined using the final Softmax layer. Every event id departing from the trained model is viewed as

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

an unusual event when using the trained model.

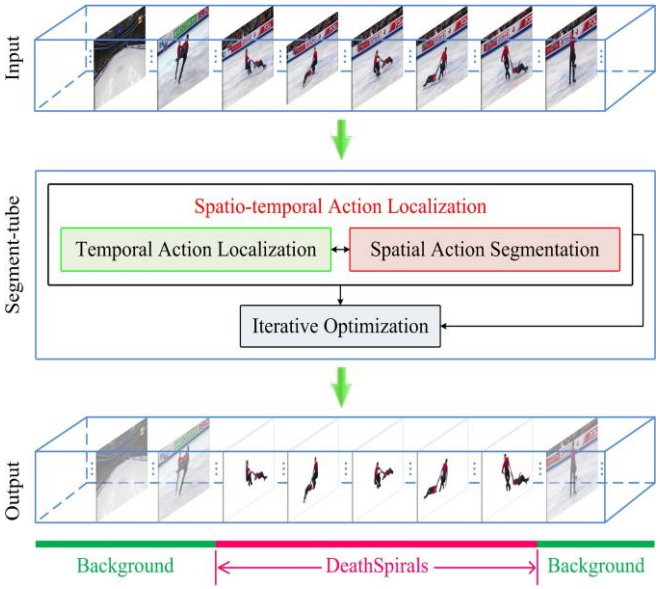

Fig. 1. Flowchart of Spatiotemporal autoencoder

A. Algorithm

1) Algorithm to generate frames::

1) void generateFrames()
2) if path not exist
3) make directory
4) else
5) remove directory
6) create new directory
7) //create a video object
8) initialize variables count and success to 0 and 1
9) while success
10) //take the values of success , count from the video object
11) if count lessthan 500
12) save the image
13) print ”image name”
14) else
15) end loop
16) increment the count

2) Algorithm to detect suspicious frames:

1) void detectActivity()
2) initialize variables count and success to 0 and 1
3) for images in the path:
4) predictions,probabilies = predictImage(images)
5) for each prediction,each probability
6) if eachprobability graterthan 80
7) increment the count
8) else if each probability lessthan 80:
9) set count 0
10) if count graterthan 10:
11) set option to 1
12) print ”The images names and its probabilities”

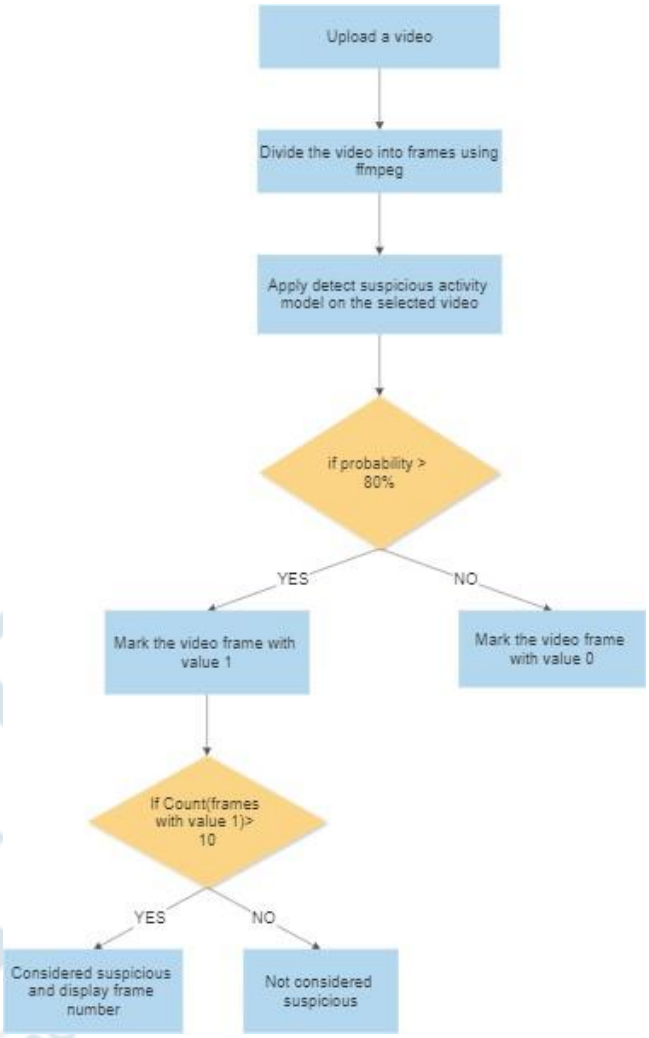

Fig. 2. Flowchart of the process

B. Generating frames

The whole video will be divided into images using Python, and these images also known as frames are given to Inception V3. Many deep learning specialists would approach video classification as if it were a series of image classifications conducted so many times let us N, where N is the overall number of images or frames in the film because videos are considered as the combination of images. By the use of video classification, which goes beyond simple picture classification, we may frequently presume that adjacent frames in a movie are linked in terms of their semantic contents. Using the temporal features of videos allowed us to improve the outcomes of our actual video classification.

C. Spatiotemporal AutoEncoder-Decoder

In our design, a spatial autoencoder is embedded with a transient autoencoder. The organization uses a video outline Yt of size H x W as input for each time step and outputs a result of a similar size that addresses the predicted future outline, Y˜t+1. We provide a thorough description of each module in the following.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

An excellent convolutional encoder-decoder is the spatial autoencoder. One convolutional layer makes up the encoder E, which is followed by tanh non-linearity, a spatial max-pooling layer, and a subsampling layer. With the exception

of the non-linearity layer, decoder D is a mirror image of the encoder and employs the closest neighbor spatial upsampling to return the output to the size of the initial information. The size of the component maps xt after the forward pass via the spatial encoder YtE→xt is d×h×w, where d is the number of highlights and h additionally, w is the level and width after downsampling, respectively. The encoding is produced at the lower-layer secret layer.

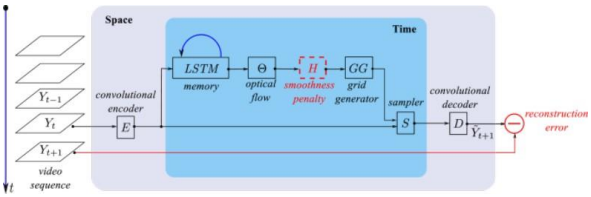

Fig. 3. Spatiotemporal video autoencoder

The transient autoencoder's goal is to identify significant changes brought on by movement, it has an idea about the future and the past so it allows it to predict the visual future. Masci (2011) developed a model spatial autoencoder in which the encoder and decoder learn separate component spaces that enable optimal information disintegration and use a sort of regularisation to prevent learning trivial planning. The decoder is required to learn about its own component space to fulfill the decay and recreate the information, using largely identi- cal activities and having a comparable number of levels of opportunity as the encoder. The encoder freely chooses upon deterioration based on its ongoing element space. Uniquely in contrast to this, the suggested ephemeral autoencoder features a decoder with limited teachable limits, whose essential task is to provide timely input to the encoder without the restriction of correcting the encoder's errors as in the spatial situation. In terms of improvement, the encoder, which is currently under more pressure to generate adequate element maps, is mostly to blame for the error made during learning.

IV. RESULT

Observation of results

TABLE I. COUNT OF SUSPICIOUS ACTIVITIES

Videos	0-50	51-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	sum
V1	0	0	0	0	0	0	0	0	0	0	**0**
V2	0	13	12	3	0	0	10	2	0	0	**40**
V3	0	0	0	0	0	0	7	3	4	1	**15**
V4	0	0	0	0	0	0	0	0	0	0	**0**
V5	0	0	0	0	4	6	3	0	0	0	**13**

For this project, we taken 5 sample videos whose results are mentioned in the table mentioned above.It contains the count of frames that are found to be suspicious.To consider the frame to be suspicious the probability must greater than the threshold value.

Thresholdvalue=80.00000%

ifprobability ≥ 80.00000% ⇒ ConsideredasSuspicious

The above table is the sample video3 which is mentioned in the table1.It has 15 anomaly frames and their probabilities.

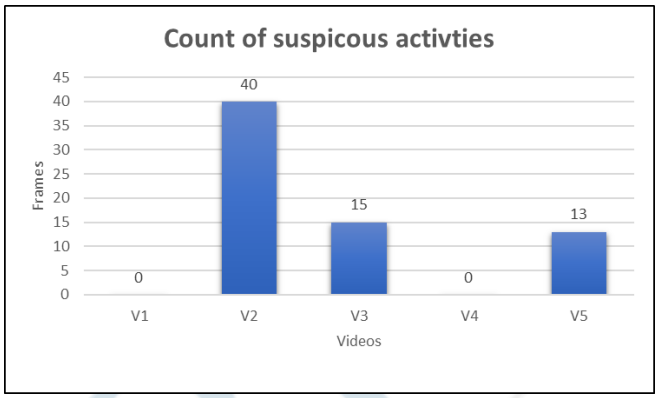

Fig. 4. Bar graph for number of suspicious frames in each video

TABLE II. PROBABILITY OF SUSPICIOUS ACTIVITIES IN V3

Frame.No	Probability
1	86.643550
2	81.815506
3	84.574853
4	81.926688
5	82.767031
6	80.778792
7	89.714965
8	86.940765
9	83.015910
10	81.221327
11	85.565089
12	89.380309
13	89.701186
14	83.363722
15	87.789232

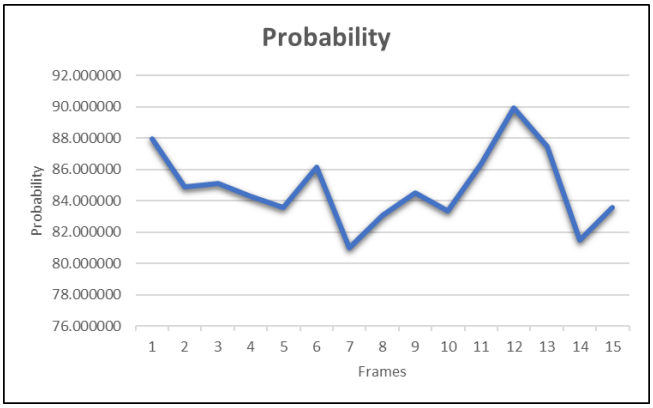

Fig. 5. Probability of the suspicious activities in V3

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 3, March 2023

TABLE III. PROBABILITY OF SUSPICIOUS ACTIVITIES IN V5

Frame.No	Probablity
1	82.24384207
2	87.34415073
3	87.20511549
4	82.2536882
5	89.57724915
6	89.19642001
7	82.5877848
8	82.00354441
9	81.08333817
10	86.14629515
11	84.94310425
12	87.68037444
13	87.86786911

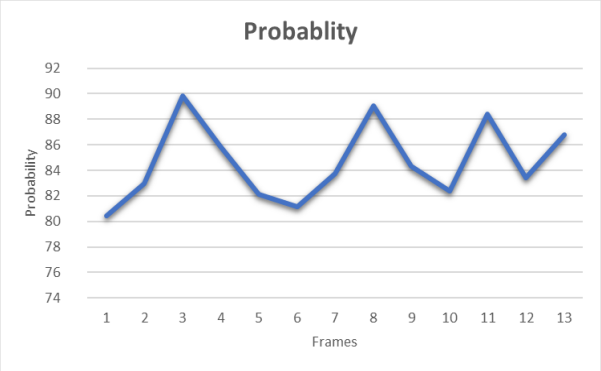

Fig. 6. Probability of the suspicious activities in V5

The above table is the sample video5 which is mentioned in the table1.It has 13 anomaly frames and their probabilities.

Here we maintain a threshold value for considering the video to be an Malfeasance which is 10.

$$countoframes \geq 10 \Rightarrow Suspiciousactivityfound$$

The above videos obey the conditions so we consider it suspicious activity.

To make working with the model easier, we created a web page that gives access to the overview. When we run the bat file it automatically opens a tab as shown in fig.

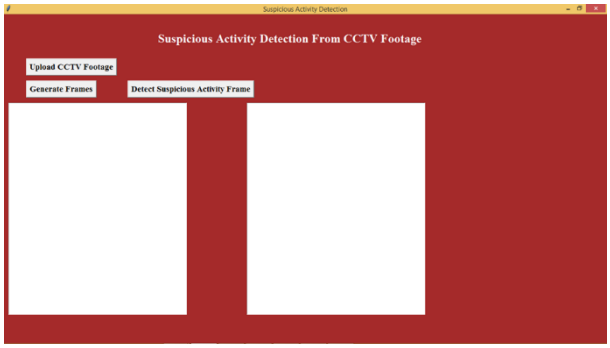

Fig. 7. Initial Webpage

We can upload the video or CCTV footage by pressing the button that says upload CCTV footage. After uploading the video by pressing a button we can generate frames.

Once the video is uploaded, to generate the frames click on the button which says 'Generate Frames'.All the generated frames are saved in a new folder named 'frames' in the project folder.

On clicking the button 'Detect Suspicious Activity Frame', all the frames containing suspicious activities are identified and mentioned in the box on the page. In the above-detected frames, suspicious activities can be found and can be taken care of by security. This helps the watchman to monitor public places more effectively.

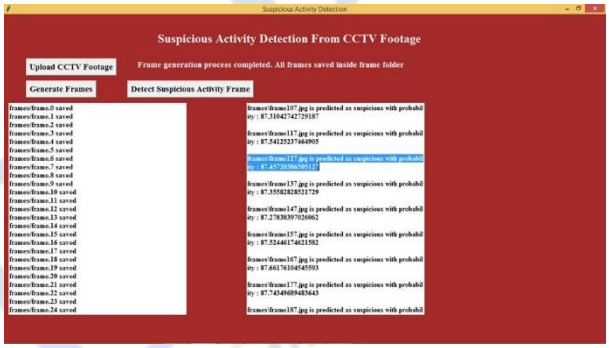

Fig. 8. Frames containing Suspicious actions

V. CONCLUSION

Nowadays the crime rate is very high so the detection of malfeasance plays a major part. By using Deep Learning from this project we can detect any malfeasance activities. Before proposing this method we referred to several other methods among the others this method become a highly accurate model. Before proposing this model we carefully observed the pros and cons carefully. It may not detect every suspicious activity but in the future, it will develop a lot and makes it easy to detect every activity from live footage.

REFERENCES

[1] B. Yang, J. Cao, R. Ni, and L. Zou, "Anomaly detection in moving crowds through spatiotemporal autoencoding and additional attention," Advances in Multimedia, vol. 2018, 2018.

[2] Y. Zhao, B. Deng, C. Shen, Y. Liu, H. Lu, and X.-S. Hua, "Spatiotem- poral autoencoder for video anomaly detection," in Proc. 25th ACM Int. Conf. on Multimedia, 2017, pp. 1933–1941.

[3] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6479–6488.

[4] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 733–742.

[5] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional

networks," in Proc. IEEE Int. Conf. on computer vision, 2015, pp. 4489–4497

[6] Fangli, Maylor K.H Leung, Mehul Mangalvedhekar, Mark Balakrishnan, "Automated video surveillance and alarm system", in Proc. IEEE Conf. on Machine Learning and Cybernetics, 2008.

[7] S Venu Gopal, N Sambasiva Rao, " An Algorithm for Simulated Routing Load while Sharing Files in Peer to Peer Systems" International Journal of Computer Mathematical Sciences (IJCMS), ISSN : 2347-8527, Volume 6, Issue 10, October 2017, Pg No: 88-93.

[8] Busaramoni Jayanth, Dr.S. Venu Gopal, "RE-ESTIMATING PROCESS- LEVEL MICROBE ESTIMATE" Palarch's Journal Of Archaeology Of Egypt/Egyptology 18(4). ISSN 1567-214x, pg: 1311-1317, Year:2021.

[9] S Venu Gopal, N Sambasiva Rao, S K Lokesh Naik " Applying Load Separation Method in Structured Peer to Peer Overlay Networks" International Journal of Engineering Science and Computing (IJESC), Vol 6 Issue No:12, ISSN: 2250-1371, 2016 / Dec, pg. No: 3748 - 3750.

[10] S Venu Gopal, N Sambasiva Rao, "Dynamic Sharing of Files from Dis- connected Nodes in Peer to Peer Systems overlay Networks", ICEEOT- 2016, IEEE Conference, ISBN:978-1-4673-9940-1.

[11] S Venu Gopal, N Sambasiva Rao, "Applying Load Separation Method in Structured Peer To Peer ", 2016 IJESC, ISSN: 2250-1371 Volume 6 Issue No. 12.