Suspicious Behavior Detection Using Man Machine Model with Integration of Virtual Reality

Jitin Bahri

Computer Science, Amity University
Amity Rd, Sector 125, Noida, Uttar Pradesh
201301
jitin.bahri1@student.amity.edu

Dr. Rakesh Garg

Computer Science, Amity University
Amity Rd, Sector 125, Noida, Uttar Pradesh
201301
rgarg2@amity.edu

***Abstract*— Video or CCTV surveillance plays an important role in the security of any place whether it is residential areas, industries, public spaces like shopping malls, museums and other monuments, banks, offices, building sites, warehouses, airports, railway stations, etc. It will help in preventing theft and damage to manufactured goods and products as well as manufacturing equipment, having complete and recorded production accident data, having complete and recorded production accident data, monitoring every stage of the manufacturing process, and prevention and analysis of any type of crime. But the current systems rely too much on humans monitoring the feeds from these videos which are prone to some problems like reduced attention and fatigue during long stretches of monitoring. So, there is a need for a system where these humans are aided by the machines in the monitoring process. The system proposed and implemented in this study would help to overcome this problem by aiding the man with smart machines and neural networks. Also, the system works on video camera feed instead of static camera shots which would help in capturing the sequential information that may be missed when using static images. And along with this, a model has been proposed using neural network technology that can automatically identify the individuals exhibiting the suspicious behavior from live camera feed input. At last research has been done on whether to use virtual reality on live video or CCTV surveillance or not based on various research papers published in similar domain.**

Keywords*— *Surveillance Camera; Suspicious behavior; convolutional neural network; CCTV.

I. INTRODUCTION

Today police and various other security forces rely on video surveillance or CCTV surveillance system for better security and easy monitoring. It has been proven that CCTV surveillance is one of the most important tools for security and one of the most effective method to prevent dangerous event that can cause damage to life and property in large public spaces like bus terminal, railway stations, metro stations, schools, etc.

Most of the CCTV or video surveillance systems that exist today work on the following principle:

1. They record all the footage and stored in some central space or in cloud.
2. If any mishaps happen, they refer the footage and evaluate.
3. Finally, the theft or culprit get caught.

Basically, this system is based on offline video evaluation after an event like robbery, murder, burglary, etc. occurred.

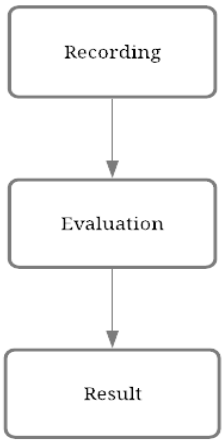

Fig. 1 Offline video evaluation

Another way in which CCTV or video surveillance is work by monitoring lively through camera feed and security personnel. Still this method has some limitations such as mental or visual fatigue as they have to keep on monitoring the screen and this reduced the overall accuracy in the surveillance system. Research published by "RTI International" [1] for "Science and Technology Directorate, US Department of Homeland Security" focus on "Transportation Security Administration (TSA)" on two fronts namely "Body detection visual search" and "X-ray visual search". Main motive is to find the characteristics which will help the fronts and traits of the trained personnel from one group or team will help on the other fronts. Analysis like Regression analysis, one way variance analysis and the correlation evaluation of the relation between different traits on the two fronts were done. Based on the evident from research it is well understood that visual and mental fatigue

were found to be playing a major role in the accuracy and performance of the teams. It is also observed that more focus needed on reducing the visual and mental fatigue of the security personnel.

Very less effective field view was observed, this led to failure in detection of various crimes. In these cases, automatic surveillance system helps greatly. Many researches have been done in this particularly field but most of them are limited to a static camera image as suggested by the research done by Tripathi el al. [2]. They put together all the reviewed researches that are done in the field of detecting suspicious activities using video surveillance and summarized it. Activities which are taken into consideration are fire detection, detection of abuse, collisions on roads and unlawful traffic parking, fall detection, robbery identification and Identification of abandoned items. Discussed topics which includes vast techniques like activity analysis and recognition, object classification, extracting attributes, identifying objects depending on tracking or non-tracking methodologies and extracting items in the foreground. It was also concluded that no system at present (at the time of publishing) were able to detect in 100% accuracy with zero false detection rate. In robbery field, identification and abandoned item identification, most of the work focus on static images and there was only less usage of videos. In the field of detection of abuse or violence, the same problem working static camera shots persist along with very less accuracy. Henceforth it was concluded that there is a very important need for video surveillance and automatic detection of these crimes.

The purpose of this research is to detect or predict any kind suspicious behaviour of person in order to prevent the events like robbery, murder etc. Second important purpose of this research is to improve the accuracy of CCTV surveillance.

Currently most of the CCTV or video surveillance is based on the offline evaluation after the occurrence of events like robbery, murder, burglary, etc. This model only helps in identifying culprit but it is not useful if you want to prevent the event. This research will solve this issue by Real-Time monitoring with the help of man-Machine Model.

Another major problem with the current video or CCTV surveillance is that usually all the surveillance is done by the security personnel and there are limitation like visual and mental fatigue. This reduces the accuracy over time. This research will solve this problem by implementing man-machine model, i.e., the surveillance can be either manual or fully automated in case if the security personnel want to take a break or having fatigue.

At last, this research will also be discussing whether to integrate Virtual reality in CCTV or video surveillance or not. This research is based on previous research work on Virtual reality with all the pros and cons of implementing virtual reality in video surveillance.

This paper is composed as follow:

A literature review will be described in next section. The proposed approach for the detection of Suspicious behavior will be highlighted in Section 3. Section 4 exhibits the experimental results and a comparison with other research works will be discussed. Finally, a conclusion which highlight contributors will be resumed.

II. Related Work

There is substantial amount of research on the video surveillance using human behavior analysis has been done on all the scales using many computer-based models, an extensive review of those researches has been consolidated in this section.

Bermejo, E [10] proposed a video-surveillance hardware and software prototype which detects dangerous real time events and alerts the human operator. In addition to this alarm system has been added to it which alerts in prior by the suspicion detected. So basically, this paper helps in three types of suspicious behavior detection:

1) Trespassing
2) Riots
3) Fights

Mossad Ben Ayed [11] proposed and algorithm which detects specific kind of behavior (ATM suspicious behavior detection) with highest accuracy by using a technique called Data driven error correcting output code. They made a algorithm based on image processing and this reduced false alarm rate and recognition rate increased but this proposal is limited to high resolution video and static camera. If not the accuracy reduces.

Thi Thi Zin [12] proposed an integrated framework which uses the techniques like multiple background modelling technique, high level motion feature extraction method and embedded Markov chain models are integrated for detecting suspicious behaviors in real time video surveillance systems. Integration of the computed features and the time probability of embedded Markov chain the suspicious behaviors in the video surveillance is analysed.

P. Kamala [13] proposed fully automated multiple surveillance camera and intelligent monitoring system they detect the suspicious individual and the alert will be sent directly to the human operator using Wi-Fi, This way the focus of the human operator directly shifts towards the culprit by using automated pan tilt zoom camera.

Mohannad Elhamod [14] proposed framed that processes video received from a fixed colour in a particular location. The framework first obtains the 3-D object level information by detecting the suspicious behavior using blob to object modelling. Here the focuses only is in certain kind of

behaviors for example fighting, fainting, meeting and walking together.

Lee et al [5] proposed a model named “ArchCam” for detecting behaviour that is considered suspicious inside the ATM. There are two components to their system: detection of an object in the shape of the belt through region split and merge technique and the second one is the detection of climbing or squatting activities. It was assumed that belt-shaped object can be used by the criminal to remove the ATM and carry that out to somewhere to steal money. Squatting and climbing were assumed to be dangerous because the criminal can weld off or bomb the base of the ATM to remove it. Both of these are included in the system and proposed to be used besides the traditional system for surveillance of videos.

proposed a system that has been designed to detect the user’s suspicious behavior in internet banking. System is designed in such a way it can recognize and classify the suspicious ones based on the intensity. This one is not based on video surveillance but based on banking system. They aim to detect suspicious behavior in internet banking system.

Duber Matinez Torres [17] proposed a context online learning scheme where from surveillance camera suspicious behaviors were detected. This approach uses incremental learning process which helps to detect suspicious behavior from surveillance camera feed with reduced training dataset by incremental learning. Experiments were conducted and the results concluded that proposed method was able to incrementally learn from very less initial dataset and achieved a performance which was similar to batch-type trained with all data simultaneously and outperformed 5 states of the art algorithm over violence detection.

Ben et al [3] suggested detecting the sudden changes in the trajectory of a person by calculating the "theta" parameter. When the value of theta exceeds a set threshold. This change is reported and the rectangle changes color to red. Red rectangle specifies the person can be considered suspicious based on the sudden movement. But this approach has a limitation that it does not include the automatic object detection. The person monitoring the scene have to manually draw the rectangle to let the system know about the interest points.

Bouma et al [4] proposed the concept of strong and weak tags. Strong tags signify greater threat probability whereas weak tags signify the small probability of a security-related incident. The number of strong tags required for causing an alarm would be much lesser than the number of weak tags. When using the system, a trade-off must be made between the 4 numbers of false hits of suspicious individuals and ignored suspicious individuals. If the accuracy of the system is increased, then the number of false hits also increase and accuracy goes down when it is tried to reduce the false hits. Using multiple operators for monitoring a single area is also suggested to increase the accuracy of prediction.

Shao et al [6] proposed a system consisting of three main components: fast evidence data recovery storage, intelligent camera monitoring, and smart prior intimation for suspicious incidents. For prior intimation, association analysis was done in which the camera continuously captures the video footage and extract useful information out of it and store in the database. This information will then undergo correlation analysis to detect any suspicious activities. Use of “Multi-point association analysis” is also suggested where instead of a single camera, multiple camera networks is used for detecting any suspicious behaviors as crimes many times consist of a series of events rather a single standalone event.

Hommes et al [7] suggested a new approach for surveillance system of using temporal “Evaluation of objects and multivariate (EWMA) control charts” along with multi-point analysis. The research focused on showing very less and relevant information to the security personnel monitoring the footage from security cameras. This would help in decreased false positive rates which reduces the stress and irritation of security personnel monitoring the camera footage. But it has the limitation of decreased detections and it would be helpful scarcely populated areas.

J. Iskander [22] research proposed experiment in which people of different age group were allowed to use head mount display for different durations and results were provided. They concluded more break time is needed in HMD to provide better accuracy and focus whereas in 2D displays such as monitors and displays the break or recovery time is very less comparatively

Maskazu Hirots [23] experimented this research with full focus on objective and visual fatigue experienced by individuals before and after performing a visual task while using head-mounted display for virtual reality and 2-D display. This research gave similar results as that of paper published on" Subjective and objective evaluation of visual fatigue caused by continuous and discontinuous use of HMDs".

III. Proposed Approach for Suspicious Behavior Detection

The proposed approach is composed of nine steps.

1) Creating classes. Classes for both normal/safe behavior and suspicious/danger will be created. Based on these two classes, the input video feed for supervised learning to train the model will be divided. Both two classes will have equal number of video feeds and equal number of frames per video.

2) Providing test data to class. After Class creation input data is divided into two classes i.e., normal/safe behavior and danger/suspicious behavior. This input

data will be used to train our model to identify difference between these two classes. The input data will have equal number of recorded video feed of same time and each video will have equal number of frames for both the classes.

3) Pre-processing the input data. Input video feed for the two classes will be converted into images by converting video into frame of images and all the frames will be reshaped into smaller size to improve the processing speed for real-time detection. This part of pre-processing alone will save a lot of hardware resources in real-time suspicious behavior detection. The image matrix is flattened into a single array which helps in creating deep dense layers for neural network model.

4) Transforming frames into features. Frames will be again divided into two classes i.e., normal/safe behavior and danger/suspicious behavior. It will give us the final values as in the number of frames and the number of classes.

5) Feeding data in dense neural network. The data from the transformation will be fed to a dense neural network with one hidden layer and two dense layers. Each dense layer will have a drop out factor to avoid overfitting

6) Training with 10-15 epochs. After the dense neural network is fed with the pre-processed data for supervised learning, it is allowed to run for 10-15 epochs. Model will be saved in the local server of online server where the model is trained. This file will be used later when we run it for real time suspicious behavior detection.

7) Checking the accuracy. Each epoch will provide the accuracy at its end point of time. Accuracy should not go to 100 percent as it means the model is overfitted.

8) Creating custom GPU server to test in real-time. While working on project it was found out that to execute real-time suspicious behavior detection fast and efficiently, graphical processing unit is required and to take observations and test the accuracy on real-time, a server needs to be created to get the accurate results. So, to test this model a commonly used GPU server architecture has been used in which we have included connecting many GPUs to A single motherboard through riser cards. Then configuring all the GPU, installing to their respective drivers.

9) Checking Real-time accuracy. Model will now run on the GPU server specifically designed for it. This server contains many GPUs which will not only improve the accuracy but also the speed of the whole model.

These nine phases of proposed approach is shown in the figure 2.

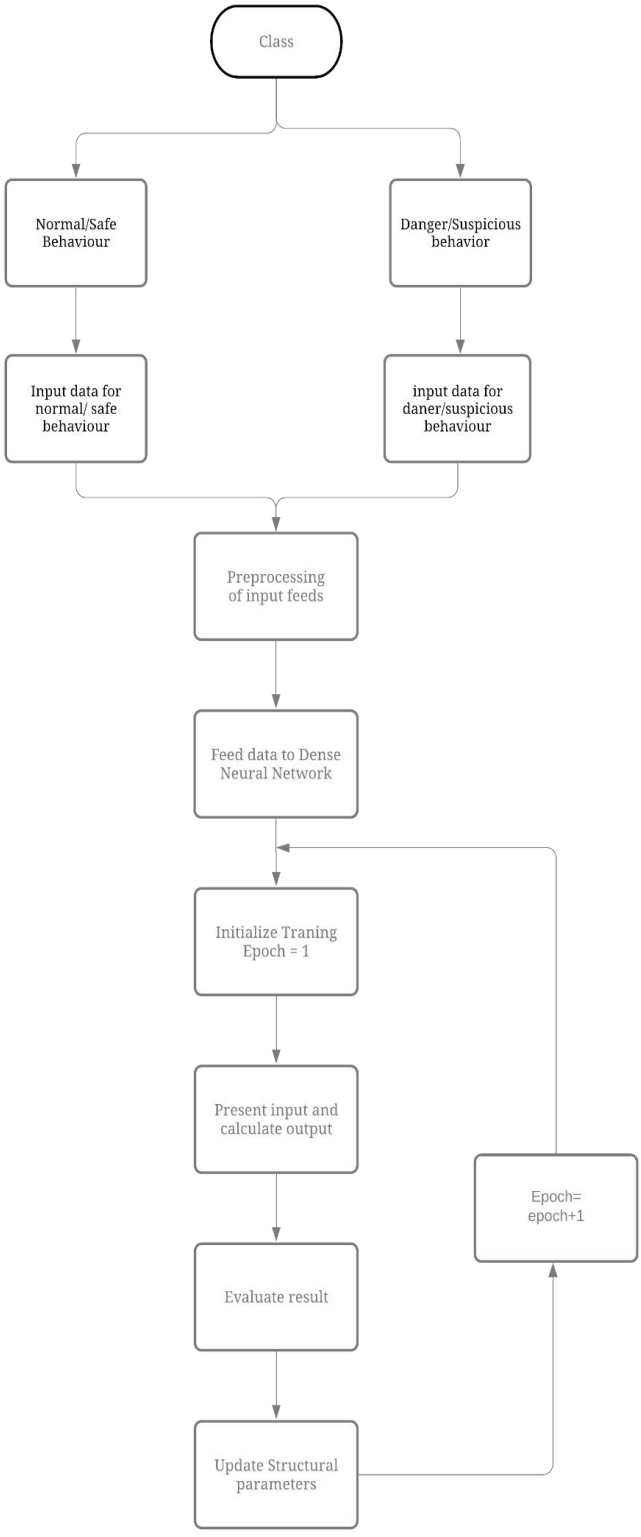

Fig. 2 Suspicious behavior detection flow diagram

Algorithm steps of the proposed approach for suspicious behavior detection

1. Create a videoCapture object

2. **If** *camera* = closed
3. **Return** "camera is closed"

4. Default resolution of the frame are obtained.
5. Convert the resolution from floating point to integer.

6. Define the codec
7. *Codec* = MJPG

8. Store video
9. Restore video capture object

10. Preprocessing videos

11. Create frame preprocessor object

12. *Resize_frame* = 224 x 224
13. Flatten the image

14. *Frame_preprocessor_weights* = 'imagenet'

15. Transform frames to features

16. Define the model with number of dense layers and dropout factors in each dense layer

17. *EPOCHS* = 20
18. *HIDDEN_SIZE* = 64

19. Initialize dense layers with drop out factor

20. model.add(**Dense**(HIDDEN_SIZE, input_shape=(X.shape[1],)))
21. model.add(**Dense**(HIDDEN_SIZE//2))
22. model.add(**Dropout**(0.2))
23. model.add(**Dense**(HIDDEN_SIZE))
24. model.add(**Dropout**(0.25))
25. model.add(**Dense**(len(CLASSES), activation='softmax'))

26. model.compile(loss='categorical_crossentropy',
27. *optimizer*='rmsprop',
28. *metrics*=['accuracy'])

29. *x_train*, *x_test*, *y_train*, *y_test* = train_test_split(X, y, random_state=42)

30. model.fit(x_train, y_train,
31. batch_size=32, epochs=EPOCHS,
32. validation_split=0.2)

33. model.save(MODEL_PATH)

34. Test the model

IV. Experiment Results

GPU server was designed and created to get the accurate and fast results. To test this model a commonly used GPU server architecture was used which included 6 NVIDIA GTX 1060 with 6gb memory to a single motherboard through riser cards. Gigabyte motherboard was used as it can easily accommodate 6 GPU. SSD was installed to further increase the speed. Image of the final GPU server created is shown in figure 4.

Fig. 3 Suspicious behavior detection flow diagram

The Pre-processing time for 4 videos each of 100 frames using CPU took 48 seconds. We increased the pre-processing time by both our algorithm and custom-built server for it. Pre-processing time for same 4 videos each of 100 frames using our server only took 3.8 seconds. Results are shown in figure 4.

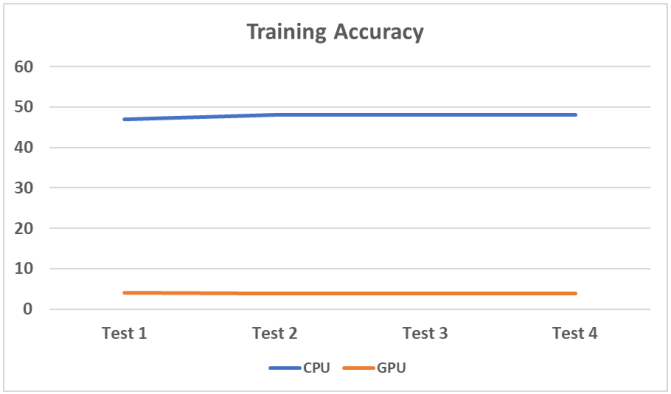

Fig. 4 Comparison of pre-processing time

The training accuracy for the first 3 epochs came out to be 96.46 % in our CPU model whereas the training accuracy came out to be 96.82 % on our GPU server. This finding was interesting as there was very little change in training accuracy as expected. Earlier assumption was that faster the hardware more should be the accuracy but later result came out to be almost same for training accuracy. Main difference was in real time accuracy. Results for the three epochs for both CPU model and GPU model is shown in figure 5.

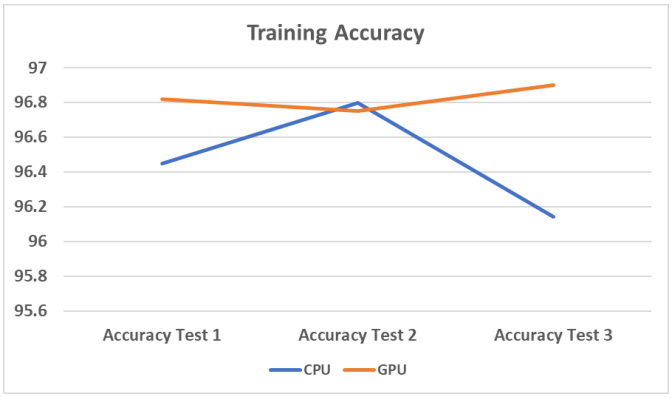

Fig. 5 Comparison of training accuracy

In real-time testing with home laptop with webcam, 50 observations have been taken by behaving normal and suspicious in front of web-cam. Out of 50 time the model was able to detect the behavior correctly 48 times (96% accurate) but the response time was 1.2 seconds means that system display the kind of behavior on screen after 1.2 seconds. In real-time testing with GPU server, 50 observations have been taken by behaving normal and suspicious in front of web-cam. Out of 50 time the model was able to detect the behavior correctly 48 times (96% accurate) but the response time was 80ms means that system display the kind of behavior on screen after 80ms. This is an interesting finding; accuracy of model was not dependent on the quality of hardware but only thing that got improved was response time in real-time scenario. Comparison of Real-time response time of both home pc and custom build server is shown in figure 6.

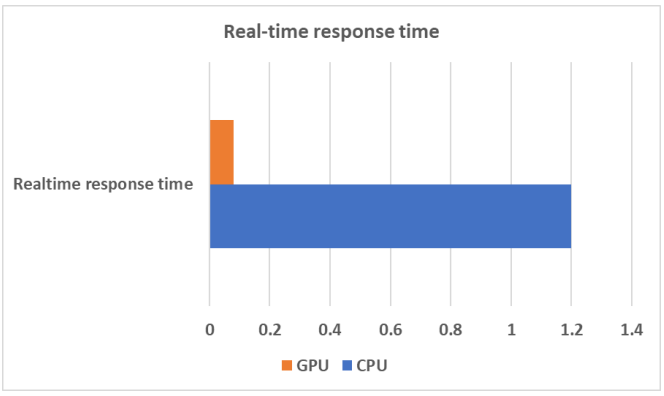

Fig. 6 Comparison of response time

Table 1 illustrates the computation time spent by different suspicious behavior recognition algorithm. According to the execution time results shown in the table 2, the model proposed in this paper has the lowest execution time. This proves that the proposed model supports the real-time exigence. Proposed method is compared with 3 papers. Schindler et al. [24] has the execution time of more than 0.2 s, Angela et al, [25] has the execution time of 0.1 s, Sadanand et al. [26] has the execution time of more than 0.2 s. The method proposed in this paper has the lowest execution of 0.08 s.

TABLE I. COMPARISON BETWEEN DIFFERENT SUSPICIOUS BEHAVIOR DETECTION METHODS BASED ON REAL TIME EXECUTION TIME

	Execution time (s)
Schindler et al. [24]	0.2 >
Angela et al. [25]	0.1
Sadanand et al. [26]	0.2 >
Our method	**0.08**

V. Discussion on Integration of Virtual Reality

According to research “A Review on Ocular Biomechanics Models for Assessing Visual Fatigue in Virtual Reality” which was published in 2018 [22] and another research paper “Comparison of visual fatigue caused by head-mounted display for virtual reality and two-dimensional display using objective and subjective evaluation, Ergonomics” [23] which was published in 2019 visual fatigue caused by using HMDs were compared with results with 2D display (Smartphone) both subjectively and objectively to study the difference of visual fatigue caused by different display devices. Binocular fusion maintenance (BFM) was measured using a binocular open-view Shack–Hartmann wavefront aberrometer equipped with liquid crystal shutter. BFM and total subjective eye symptom were not significantly different between HMDs and 2D displays.

Even with the short rest in between usage, we conclude that both the device could induce significance visual fatigue after long term usage. However, the in HMDs it is due to vergence distance and accommodation distance, whereas in 2D it is caused due to long-term focus without depth accommodations. Based on this comparison it can be concluded that the visual fatigue was less severe in HMDs than 2D displays. But recovery of smartphones is better than those of HMDs. After an exposure to visual stimuli, the accommodation and vergence response take longer time to return back to their resting values. Also, the recovery time depends on the magnitude of the stimulus and duration.

In 2Ds short rest could lessen the severity and in HMDs had a reversely subjective conclusion. It was observed that rest during the VR games increase the severity of visual fatigue symptoms which may be due to frequent switch between virtual world and real world. This switch or adaption would increase fatigue and bring discomfort which brought more conflicts with recovery in HMDs. However, in general short breaks in between made positive impact on the recovery in both HMDs and 2D display. This means that resting increased the recovery speed and reduced any impairment of

vision that it could cause. So its seen that in HMDs resting could discomfort to users but on the other hand it helped in recovery speed. So, their balance between these two conflicts should be found.

It is also observed that visual fatigue increased with increase in watching time with HMDs and the severity kept on increasing gradually after 20 minutes of usage. This has been concluded that users should always use HMDs less than 40 minutes to lessen the severity. From this research it is clear that VR is nor suitable for CCTV surveillance.

VI. Conclusions

Based on the result analysis, it was conclusive that the preprocessing time for the model was far less on custom made GPU server then normal home laptop. The difference of time between the normal system and Custom GPU server was 44.2 seconds for small sample data. This means the GPU system that was built for this purpose was 12.68 times faster than the normal system. This will result in faster training of any new suspicious behavior that user wants to feed in the model for training in future.

Training accuracy in both the systems were almost similar with the difference of 0.36%. Accuracy was higher in GPU server by a margin of only 0.36%. Little to no effect on the accuracy of the model on both systems. Similar results were found when the model was run on both the system for real-time accuracy testing. There was no difference in the accuracy of the model in real-time, both the accuracy was 96% but the model when run on GPU server was able to detect the suspicious or normal behavior with much faster speed. In normal system there was a lag of 1.2 seconds before the result was displayed on screen while the lag was reduced to only 80ms in GPU server. This proves that GPU server was able to detect the behavior 15 times faster than the normal system

References

[1] Behavior Detection Visual Search Task Analysis Project," 2018.

[2] R. K. Tripathi, A. S. Jalal, and S. C. Agrawal, "Suspicious human activity recognition: a review," Artif. Intell. Rev., vol. 50, no. 2, pp. 283–339, 2018, doi: 10.1007/s10462-017-9545-7.

[3] Hopfield, J. J. (1982). "Neural networks and physical systems with emergent collective computational abilities". Proc. Natl. Acad. Sci. U.S.A. 79 (8): 2554–2558. Bibcode:1982PNAS...79.2554H. doi:10.1073/pnas.79.8.2554. PMC 346238. PMID 6953413.

[4] "Neural Net or Neural Network - Gartner IT Glossary". www.gartner.com.

[5] Very deep Convolutional networks for large-scale image recognition. (2014, September 4). ResearchGate. https://www.researchgate.net/publication/265385906_Very_Deep_Convolutional_Networks_for_Large-Scale_Image_Recognition

[6] Kuhlman, Dave. "A Python Book: Beginning Python, Advanced Python, and Python Exercises". Section 1.1. Archived from the original (PDF) on 23 June 2012.

[7] "About Python". Python Software Foundation. Retrieved 24 April 2012., second section "Fans of Python use the phrase "batteries.

[8] [8] "Get Ready to Hear a Lot More About 'XR'". Wired. 1 May 2019. ISSN 1059-1028. Retrieved 29 August 2020.

[9] Psotka, Joseph (1 November 1995). "Immersive training systems: Virtual reality and education and training". Instructional Science. 23 (5): 405–431. doi:10.1007/BF00896880. S2CID 60705937

[10] Bermejo, E. (2010, April 25). Security System Based on Suspicious Behavior Detection. core.ac.uk. https://core.ac.uk/download/pdf/41788246.pdf.

[11] Suspicious behavior detection based on DECOC classifier. IEEE Xplore. (n.d.). https://ieeexplore.ieee.org/document/8314926.

[12] Zin, T. T., Tin, P., Hama, H., & Toriu, T. (n.d.). An integrated framework for detecting suspicious behaviors in video surveillance. NASA/ADS. https://ui.adsabs.harvard.edu/abs/2014SPIE.9026E..14Z/abstract.

[13] Kamala, P. (2015, January 1). [PDF] Automated Intelligent Surveillance using Human Behavior Analysis in Shopping Malls: Semantic Scholar. [PDF] Automated Intelligent Surveillance using Human Behavior Analysis in Shopping Malls | Semantic Scholar. https://www.semanticscholar.org/paper/Automated-Intelligent-Surveillance-using-Human-in-Kamala/ea4e8555de4160b1b42feb41711ee68b8fdf97ad.

[14] Automated Real-Time Detection of Potentially Suspicious Behavior in Public Transport Areas. IEEE Xplore. (n.d.). https://ieeexplore.ieee.org/document/6384750.

[15] [15]Wai-Kong Lee, Chun-Farn Leong, Weng-Kin Lai, Lee-Kien Leow, Thiah-Huat Yap, ArchCam: Real time expert system for suspicious behaviour detection in ATM site, Expert Systems with Applications, Volume 109, 2018, Pages 12-24, ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2018.05.014. (https://www.sciencedirect.com/science/article/pii/S0957417418302975)

[16] Montazer, G., & Saroukhani, L. (2009, January 10). Design and implementation of a fuzzy expert system for suspicious behavior detection in e-banking system. jor.iranaict.ir. http://jor.iranaict.ir/browse.php?a_id=228&sid=1&slc_lang=en.

[17] Duber Martinez Torres, Humberto Loaiza Correa, Eduardo Caicedo Bravo, Online learning of contexts for detecting suspicious behaviors in surveillance videos, Image and Vision Computing, Volume 89,n2019, Pages 197-210, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2019.07.006. (https://www.sciencedirect.com/science/article/pii/S0262885619301052)

[18] A. M. Ben et al., "Suspicious Behavior Detection of People by Monitoring Camera," in 2016 5th International Conference on Multimedia Computing and Systems (ICMCS), 2016, doi: 10.1109/ICMCS.2016.7905601.

[19] H. Bouma et al., "Behavioral profiling in CCTV cameras by combining multiple subtle suspicious observations of different surveillance operators," 2013, no. May, doi: 10.1117/12.2015869

[20] Z. Shao, J. Cai, and Z. Wang, "Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data," IEEE Trans. Big Data, vol. 4, no. 1, pp. 105–116, 2017, doi: 10.1109/tbdata.2017.2715815.

[21] S. Hommes, R. State, A. Zinnen, and T. Engel, "Detection of abnormal behaviour in a surveillance environment using control charts," 2011 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2011, pp. 113–118, 2011, doi: 10.1109/AVSS.2011.6027304.

[22] J. Iskander, M. Hossny and S. Nahavandi, "A Review on Ocular Biomechanic Models for Assessing Visual Fatigue in Virtual Reality," in IEEE Access, vol. 6, pp. 19345-19361, 2018, doi: 10.1109/ACCESS.2018.2815663

[23] Masakazu Hirota, Hiroyuki Kanda, Takao Endo, Tomomitsu Miyoshi, Suguru Miyagawa, Yoko Hirohara, Tatsuo Yamaguchi, Makoto Saika, Takeshi Morimoto & Takashi Fujikado (2019) Comparison of visual fatigue caused by head-mounted display for virtual reality and two-dimensional display using objective and subjective evaluation, Ergonomics, 62:6, 759-766, DOI: 10.1080/00140139.2019.1582805

[24] K. Schindler, and L.J.V. Gool, “Action snippets: how many frames does human action recognition require?”, In: CVPR, 2008.
[25] A. Yao, J. Gall, and L. Van Gool, “A hough transform-based voting framework for action recognition”, In: CVPR, 2010.
[26] S. Sadanand, and J.J. Corso, “Action Bank: a high-level representation of activity in video”, In: CVPR, 2012.
[27] Cheoi, K.J. Temporal Saliency-Based Suspicious Behavior Pattern Detection.Appl.Sci.2020,10,1020.https://doi.org/10.3390/app10031020
[28] Verma, K.K., Singh, B.M. & Dixit, A. A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. Int. j. inf. tecnol. (2019). https://doi.org/10.1007/s41870-019-00364-0