AMITY UNIVERSITY

UTTAR PRADESH

Report On

Real-time suspicious behaviour monitoring system using MMVR Interface

Submitted to

Amity University Uttar Pradesh

In partial fulfilment of the requirements for the award of the degree of

Master of Technology

In

Computer Science & Engineering

JITIN BAHRI

A2300919002

SHUBHAM MINHASS

A2300919015

YASH VERMA

A2300919008

Under the guidance of

Dr. MADHULIKA

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY AMITY UNIVERSITY UTTAR PRADESH NOIDA (U.P.), APRIL 2020

DECLARATION

We, JITIN BAHRI, SHUBHAM MINHASS and YASH VERMA Enrollment No.A2300919002, A2300919015 and A2300919008 student of M. Tech (CSE) hereby declare that the project titled Real-time suspicious behaviour monitoring system using **MMVR Interface** which is submitted by us to the Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of the requirement for the award of the degree of Master of Technology in Computer Science and Engineering has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

Place: Noida

Date: 11 June 2020

JITIN BAHRI
SHUBHAM MINHASS
YASH VERMA

DECLARATION FORM

We JITIN BAHRI, SHUBHAM MINHASS, and YASH VERMA students of M. tech CSE, Enrolment No. A2300919002, A2300919008 and A2300919008 batch **2019-2021**, Department of **Amity School of Engineering & Technology**, Amity University, Noida, Uttar Pradesh, hereby declare that we have gone through project guidelines including policy on health and safety, policy on plagiarism, etc.

Place: Noida

Date: 11 June 2020

Signature: JITIN BAHRI
SHUBHAM MINHASS
YASH VERMA

CERTIFICATE

On the basis of the declaration submitted by JITIN BAHRI, SHUBHAM MINHASS and **YASH VERMA student of M. Tech. CSE, I hereby certify that the project titled "Real-time suspicious behaviour monitoring system using MMVR Interface" which is** submitted to Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of the requirement for the award of the degree of Master of Technology in Computer Science and Engineering is an original contribution with existing knowledge and faithful record of work carried out by them under my guidance and supervision.

Place: Noida

Date: 11 June 2020

Dr. Madhulika

Assistant Professor
Department of Computer Science & Engineering
Amity School of Engineering and Technology, Noida

ACKNOWLEDGEMENT

I would like to take this opportunity to express my profound sense of gratitude and respect to all those who helped me throughout our project.

This report acknowledges the intense driving and technical competence of the entire individual that has contributed to it. It would have been almost impossible to complete this project without the support of these people. I extend thanks and gratitude to, **Dr. Madhulika**, Assistant Professor**,** Department of CSE who has imparted me the guidance in all aspects. They shared their valuable time from their busy schedule to guide me and provide their active and sincere support for my activities.

This report is an authentic record of my work which is accomplished by the sincere and active support by all the teachers of my college. I have tried my best to summarize this report.

JITIN BAHRI
SHUBHAM MINHASS
YASH VERMA

M. Tech CSE

Amity School of Engineering and Technology, Noida
2019-2021

ABSTRACT

Video surveillance plays an important role in the security of any location 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 using a virtual reality interface. 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. For this purpose, an office environment has been simulated using Unity 3D where the user can mark the suspicious behaving characters. And along with this, a model has been proposed using Deep learning technology that can automatically identify the individuals exhibiting the suspicious behaviour from live camera feed input.

CONTENTS

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Today police and various other security forces rely on video surveillance systems to facilitate their work. This has proved to be a vital tool for security. Its role is much more important in large public spaces like bus terminals, railway stations, metro stations, near popular monuments, shopping complexes and malls, schools, offices, etc.

Most of the video surveillance systems that exist today are used specifically for offline video evaluation after an event like robbery, murder, burglary, etc. has already occurred. The live monitoring is still done manually by security personnel through live camera feeds. But it was observed that this approach has a few limitations. One of them is visual and mental fatigue. Research published by "RTI International" [1] for "Science and Technology Directorate, U.S. Department of Homeland Security" focus on the "Transportation Security Administration (TSA)" on two fronts namely "Body detection visual search" and "X-ray visual search". The goal was to find out what characteristics are required for both the fronts and if the traits of trained personnel from one team would be helpful on other fronts. Regression analysis, one –way variance analysis and Pearson correlation was used for evaluation of the relation between different required traits on the two fronts. Visual and mental fatigue was found to be playing a big role in the performance of both the teams as evident from the research, it was observed that much importance should be given to reducing the visual and mental fatigue of the security personnel.

Also, the effective field of view is limited which may end up in missed detections of various crimes. Automatic surveillance system plays an important in these types of situations. Many researchers have done some work in this field but most of the work is limited to static camera images as suggested by the research done by Tripathi et al. [2] They summarised and reviewed all the researches that are going in the field of detecting suspicious activity using video surveillance. The activities that were 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. Various 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 are discussed. It was concluded that no system present at the time of publishing the paper was found to be 100 percent accurate in detection with zero percent false detection rate. In the field of robbery identification and abandoned item identification, most of the work focus on static images and less work has been done in the 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. So it was stated that there was a need for video surveillance and automatic detection of these crimes.

The few important concepts related to the work are discussed below:

MMVR Interface: Man-machine virtual reality interface is a system in which man (human security monitoring personnel) and machine (the software system) coordinate with each other to identify the individuals exhibiting suspicious behaviour using virtual reality interface. Virtual reality (VR) interface is used to give a more immersive experience to the user to increase the stimulus for the human agent which help in increasing the accuracy of the system.

Unity: Unity is a software development environment that is used for the development of video games for multiple platforms like Personal computers, Sony PlayStation, Microsoft Xbox, android smartphones, etc. It is also used in the film industry and automotive industry. Some of the movies are made wholly in unity using the Cine-Machine tool in unity. Unity is also being used by the automotive industry to create 3-D models of different components and vehicles using virtual reality. It is widely being used in construction, engineering and architecture. It is being used by big companies like Alphabet and DeepMind to train their Artificial intelligence models.

C#: C-sharp is a programming language that works in many paradigms. It was developed as a part of the .NET framework by Microsoft. It is a general-purpose language that can be used in developing a variety of applications like Mobile Apps, Windows Store apps, Website development, Enterprise applications, Office applications, backend service and cloud applications, etc. It can also be used for more in-trend applications like Artificial Intelligence, Internet of Things, Blockchains, Azure cloud services, etc. C# is the most popular scripting language in the unity game engine.

Deep learning: It is a part of machine learning that comes under the wide umbrella term Artificial Intelligence. Deep learning is based on the functioning of the actual human brain inspired by Artificial Neural Networks. Deep learning work like machine learning but consists of many more layers than traditional machine learning that interpret data differently than the previous layer. Some popular applications of deep learning are colourization of black and white images, adding sounds to silent movies, automatic machine translation, object classification in photographs, automatic handwriting generation, character text generation, image caption generation, automatic game playing.

Convolutional Neural Network (CNN or Convnet): It is a deep learning algorithm that assigns some weights and biases to the various aspects of the image input and distinguishes them from each other. CNN requires very less pre-processing as compared to other classification methods. In traditional methods, the filters have to made and adjusted manually whereas CNN can learn these filters by itself with enough training. I each CNN, the image is passed through fully connected layers, pooling layers and convolution layers and mathematical functions like ‘softmax’ are applied to the image to find the probability of an image to be classified as a particular class. Some of the popular applications of CNN are object detection in images, image classification, facial recognition, image recognition, etc.

Recurrent Neural Network (RNN): RNN is a type of feedforward neural network having some memory due to which they can process sequential data. RNN is recurrent in nature that means that it can form a loop. It performs the same function for every input and the output depends upon the previous input. So inputs depend on each other instead of being independent. Applications of RNN include text summarization, image recognition, speech recognition, text classification, sentiment analysis, stock price forecasting, etc.

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 colour 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 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.

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.

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 behaviours 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.

The report consists of two parts: the first part is the implementation of MMVR interface using Unity 3D in which an office environment is built along with some non-payable characters (NPC) exhibiting random behavioural patterns to stimulate the environment of an office. Out of NPCs, some will exhibit suspicious behaviour. The user will have the ability to mark the NPC with weak or strong tags based on their behaviour. In the later part, a model is proposed to implement the video surveillance system using deep learning model. Transfer learning has been used to decrease the development cycle from months to weeks. The model can use the Pre-trained weights from Google's "Inception model" based on CNN architecture. Inception model was trained on the ImageNet dataset which consists of thousands of images classified into more than a hundred categories and sub-categories.

CHAPTER II: METHODOLOGY

An office environment has been created in which different non-playable characters (NPC) are placed to stimulate the real-life office environment virtually. The NPCs are programmed to simulate random behavioural patterns according to a regular office environment. Some of which will exhibit suspicious behaviour. Cameras are also placed at different locations to monitor the complete environment including all the rooms and halls inside the building as well as the accessible locations outside the buildings. The user interacting with this environment would be provided with a user interface (UI) having the functionality to tag any NPC as weakly suspicious marked with yellow colour and strongly suspicious marked as red. Other functionalities will include the untagging of an individual if it is marked by mistake or other similar situations, changing the active camera views, etc. The environment was made in the Unity game engine using C# as the scripting language. The character models and animations were downloaded and imported from 'Mixamo'website [8]. The map used for the development of the environment was taken from the popular multiplayer shooting video game "Counter-Strike: Global Offensive"[9].

Next, a model has been proposed to detect potentially dangerous events in real-time and alert the human operator using Deep learning techniques. Transfer learning was used to decrease training time and computational cost. The Inception model based on convolutional neural network (CNN) developed by Google could be used with some changes in the last layers which was trained on ImageNet dataset. This model would be useful for detecting human bodies which would then be classified as exhibiting suspicious or non-suspicious behaviour. Recurrent Neural Network (RNN) can also be used in the development of model due to the ability of RNN to process sequential data better than CNN as the video is a sequential form of data. Three modes of operation could be used: manual, semi-automatic and automatic. In the manual mode of operation, the person monitoring the Camera footage have to manually tag each person as suspicious. This mode poses the problem that a human agent can feel fatigued after some time due to the monotonous nature of the job and the accuracy will drop after that [11]. The semi-automatic mode comes to rescue in this situation. In semi-automatic mode, the human agent will be aided by the system in detecting the suspicious behaviour patterns. The system will detect the behaviour patterns and suggest them to the human agent. The human will validate the suggestions given by the system along with himself tagging the individuals. This will help train the model for the fully automatic mode using the supervised mode of learning. It would help the human agent in times when he is fatigued. The final mode would be the automatic mode in which the system automatically detects the crime with sufficient accuracy. In this mode, human intervention would not be necessary when it is trained sufficiently.

The concept of weak and strong tags can also be implemented in this model as suggested by Bouma et al. [10]. This concept will help in differentiating the urgent threats with less probable threats. Strong tags signify that urgent action must be taken on the individual which is strongly tagged. Whereas the weak tags signify that the person must be monitored closely for other signs of suspicion. When multiple suspicious activities are detected on the individual, the weak tag is converted to strong tag after a certain predefined threshold. Some common examples of situations where the individual must be strongly tagged would be when a person attacks any other person with a strong force, or when an individual is detected to be possessing or holding a serious weapon like gun, sword, dagger, etc. An individual would be weakly tagged when he/ she is staring at some other person, or swiftly running towards the opposite direction of movement of the crowd with a potential weapon like a bar of metal, bat, hockey, etc. or a person frequently looking over his shoulder, etc.

The proposed model is illustrated in the flowchart given below. Figure 1 illustrates the overall workflow of the model. The operator can mark a person exhibiting suspicious behaviour by the means of tags so that appropriate action can be taken to prevent any malicious event. Figure 2 illustrates the tag mechanism. Tags can be of various types based on different degree and type of suspicious behaviour but for the sake of simplicity, we can classify into two major types: weak and strong tags. Whenever a person is tagged with some strong tags, it means immediate action must be taken. But weak tags simply mean that there is a need to focus more on that person and monitor him/her closely for other signs. Weak tags add up to a certain number after which they are converted to strong tags.

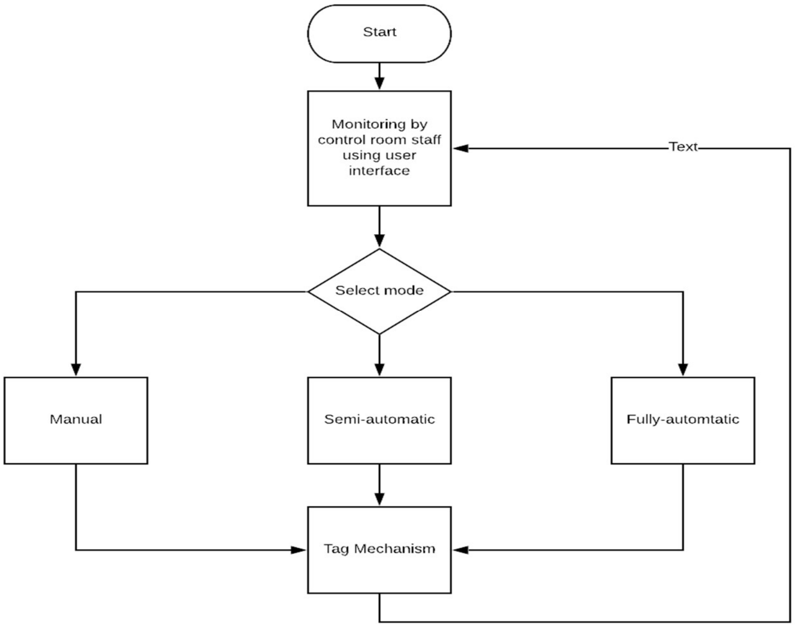

Figure 1: Overall workflow

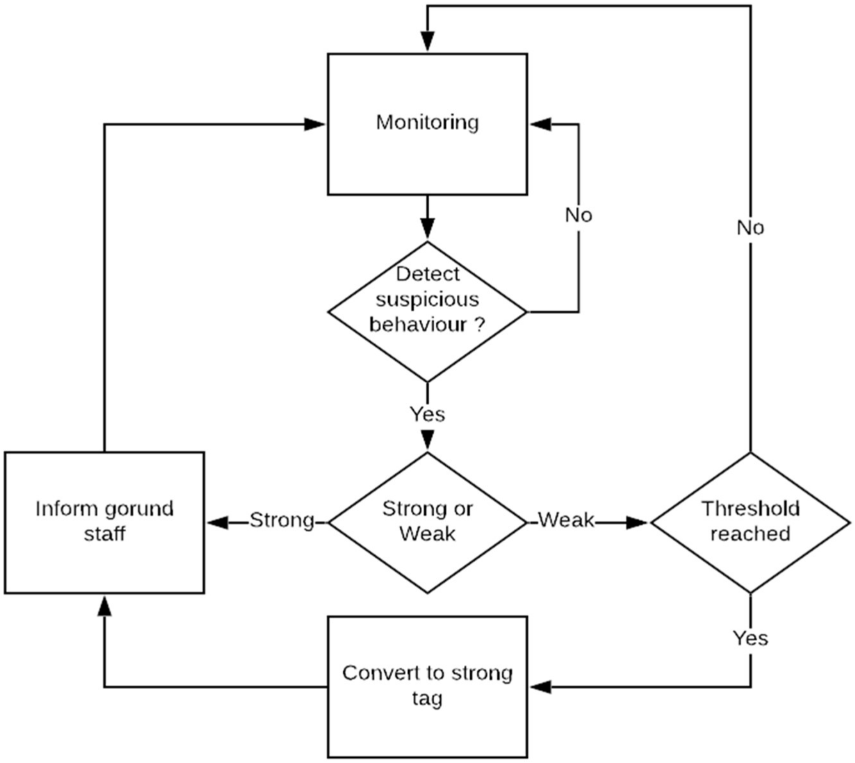

Figure 2: Tag Mechanism

CHAPTER III: RESULTS

A prototype has been built for the surveillance system using the unity game engine. Figure 3 illustrates the model of the office environment. The position of NPCs is highlighted using green colour to enhance visibility. Figure 4 illustrates a different view of the same office model with a white base to highlight the map rooms and other features.

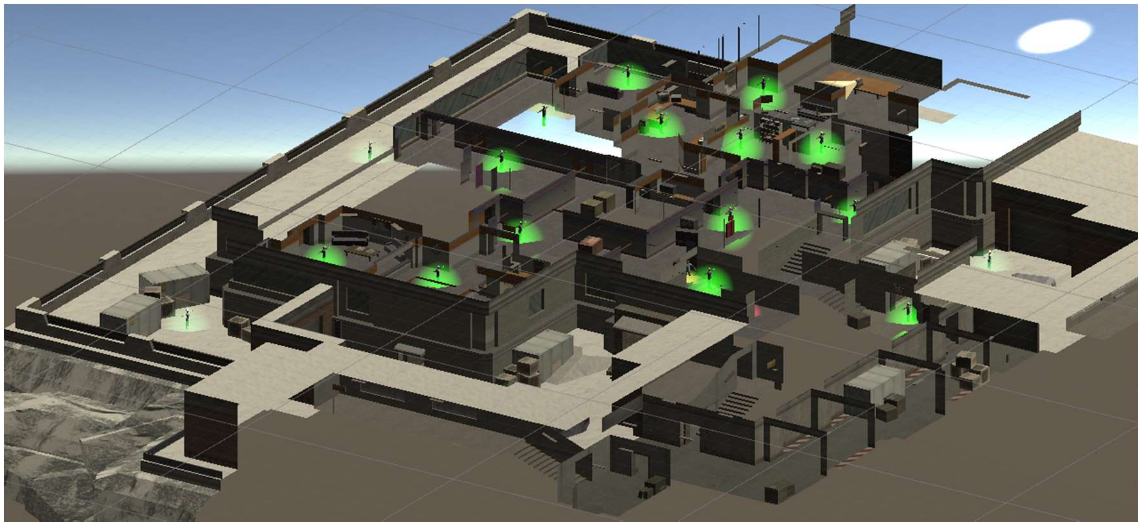

Figure 3: Map used for office environment

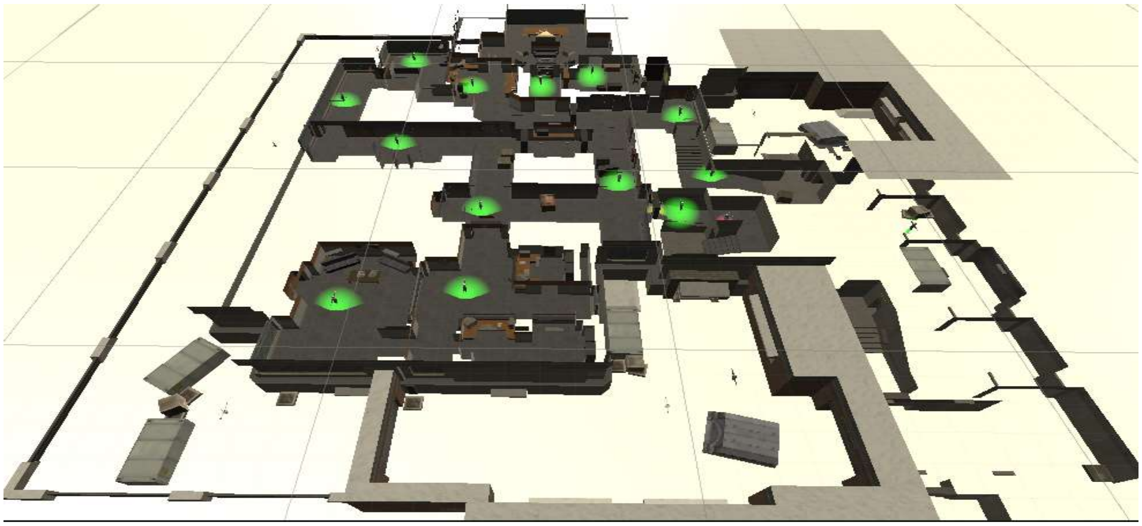

Figure 4: Top view of office map

Figure 5 illustrates different types of tags where yellow tags denote weak tags and red tags denoting the red tags.

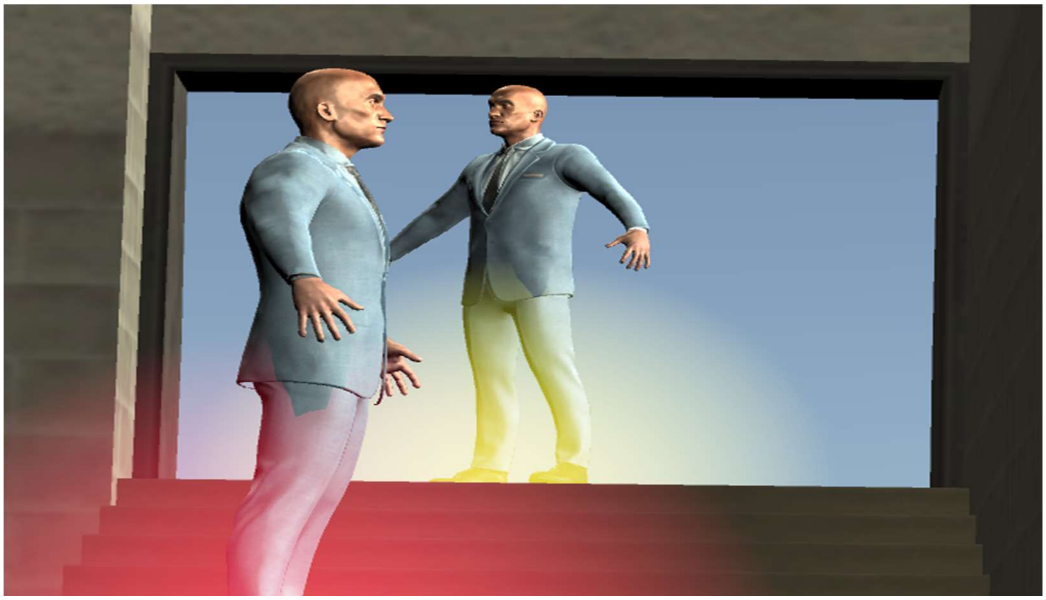

Figure 5: Types of Tags

Figure 6 illustrates the surface on which the NPCs can freely roam around to simulate the regular office environment.

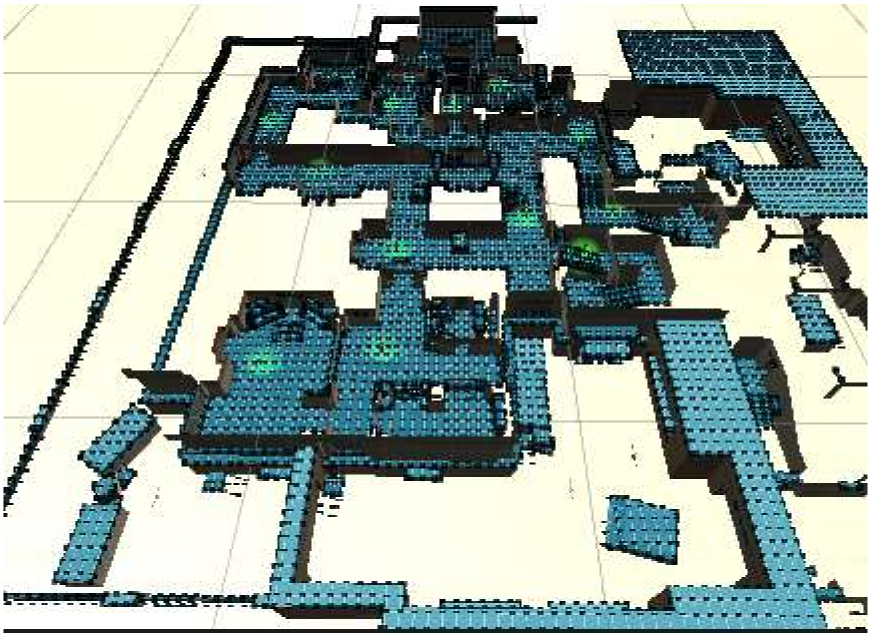

Figure 6: Navigatable surface inside buildings

A different view of the same objects and location is visible from different cameras in Figure 7a and 7b. This feature let the operator cover the whole office building as well as outside premises.

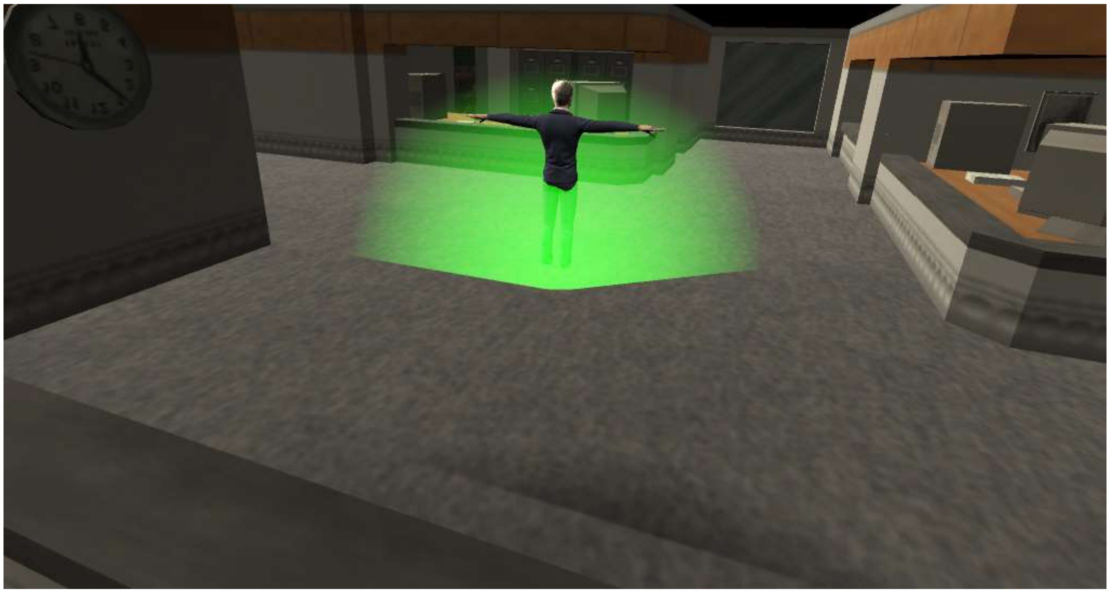

Figure 7a: Camera view 1 denoting the same room

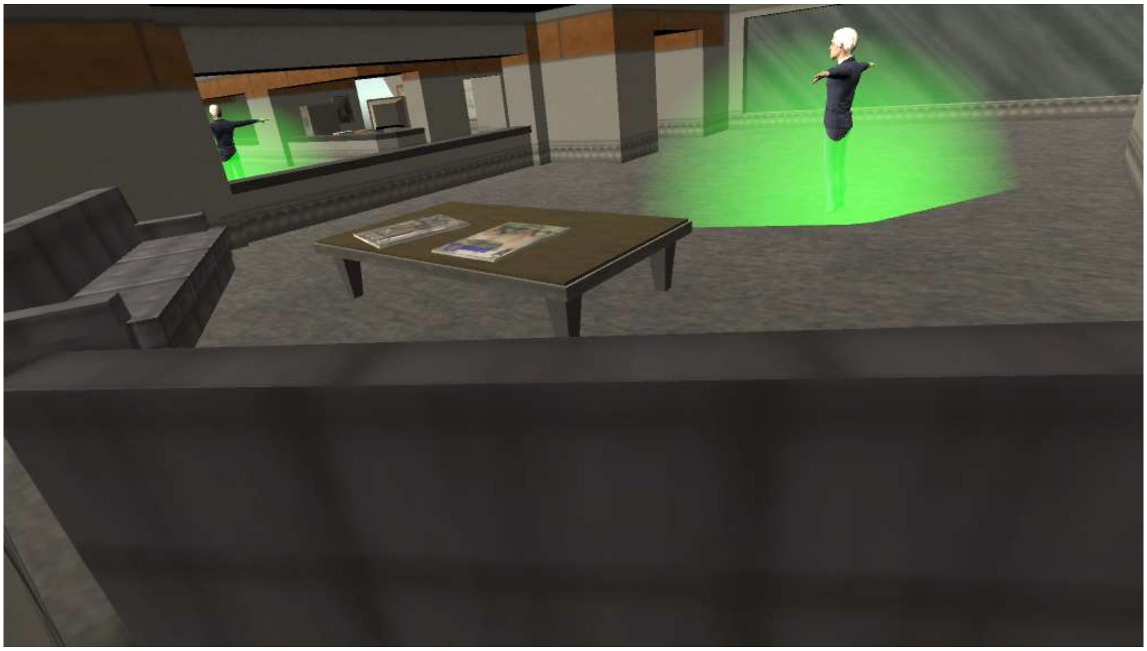

Figure 7b: Camera view 2 denoting the same room

Cameras are placed strategically throughout the map to cover the whole office environment including the open spaces within office premises and back alleys as illustrated through Figure 8 and 9.

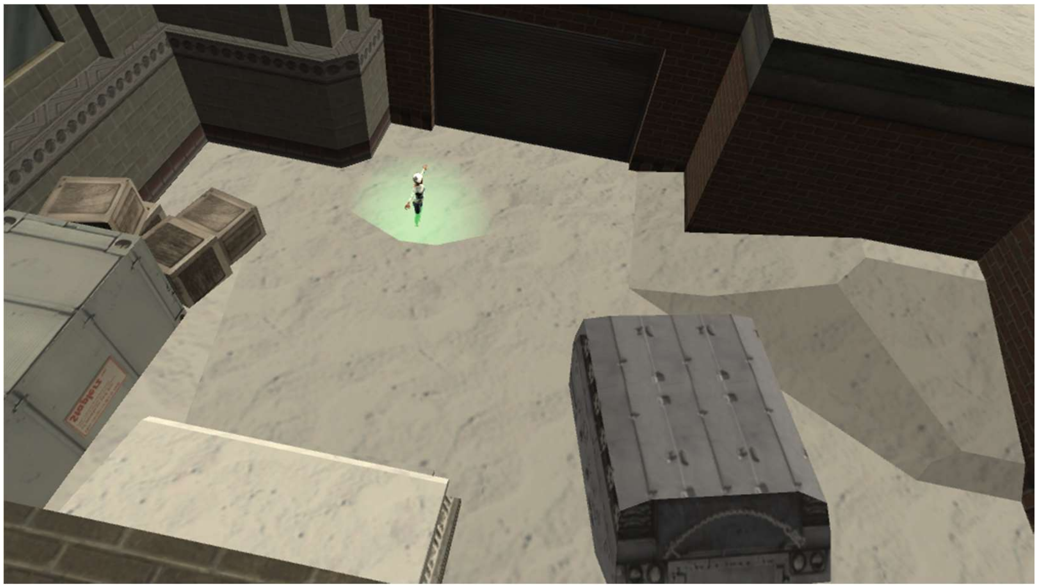

Figure 8: Camera capturing the premises

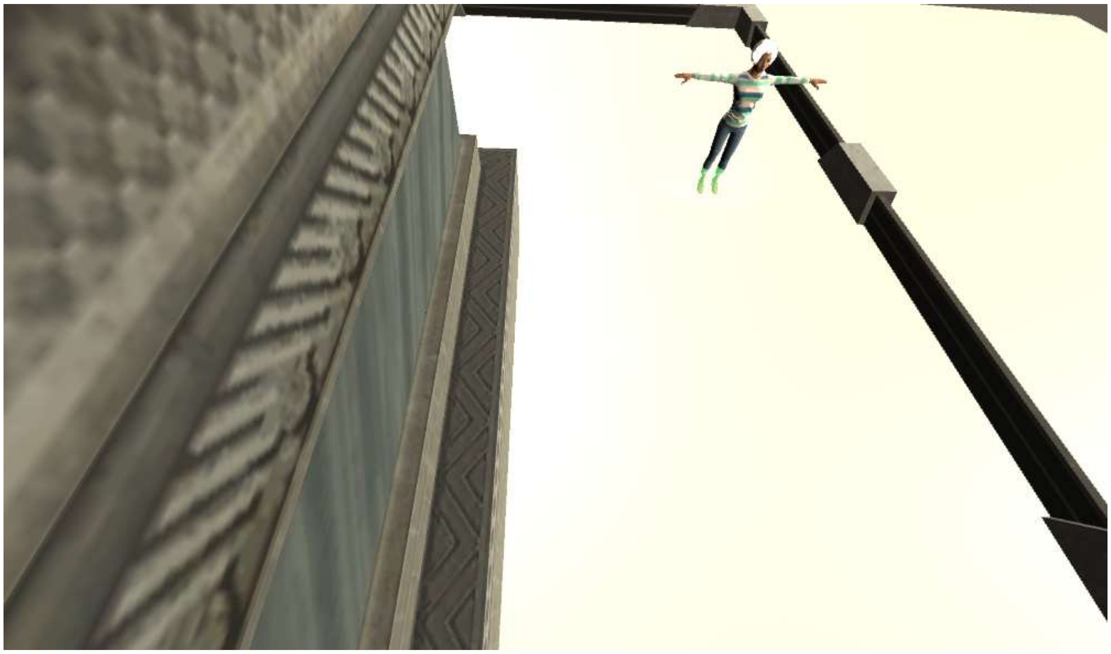

Figure 9: Outdoor camera capturing alley

CHAPTER IV: CONCLUSION AND FUTURE SCOPE

The office environment has been successfully implemented with the help of Unity 3D that can stimulate an office environment in which the user can tag the characters exhibiting suspicious behaviour with weak or strong tags. Also, the model has been proposed to automatically identify suspicious individuals from the live camera video feed. The automatic system in the current state can predict a few suspicious patterns and that too in sparsely populated areas and further development and refining are required for more accuracy and to cover large areas. Also, the system needs a human agent to monitor and validate the predictions by the system. But, in future, the system would be scaled to be able to automatically predict more suspicious behaviour patterns with high accuracy and be able to cover more population-dense areas. Also, the model developed in unity could be ported to VR Interface to get a more immersive experience.

REFRENCES

[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] 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.

[4] 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.

[5] W. K. Lee, C. F. Leong, W. K. Lai, L. K. Leow, and T. H. Yap, "ArchCam: Real time expert system for suspicious behaviour detection in ATM site," *Expert Syst. Appl.*, vol. 109, pp. 12–24, 2018, doi: 10.1016/j.eswa.2018.05.014.

[6] 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.

[7] 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.

[8] "Mixamo." [Online]. Available: https://www.mixamo.com/. [Accessed: 30-May-2020].

[9] "cs_office map." [Online]. Available: https://free3d.com/3d-model/cs-office-6260.html#. [Accessed: 30-May-2020].

Appendix: Originality Report

VR_TEAM

ORIGINALITY REPORT

3%	1%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	gettocode.com Internet Source	1%
2	"Classification and Stage Prediction of Lung Cancer using Convolutional Neural Networks", International Journal of Innovative Technology and Exploring Engineering, 2019 Publication	1%
3	David Kung, Jeff Lei. "An Object-Oriented Analysis and Design Environment", 2016 IEEE 29th International Conference on Software Engineering Education and Training (CSEET), 2016 Publication	1%
4	"Proceedings of ICRIC 2019", Springer Science and Business Media LLC, 2020 Publication	<1%
5	Submitted to University of Computer Studies Student Paper	<1%
6	Submitted to Birla Institute of Technology and Science Pilani	<1%