AGGRESSIVE BEHAVIOUR DETECTION SYSTEM FOR LIFT USING CONVOLUTIONAL NEURAL NETWORK

SURRYA VIKNESH A/L CHANDRA SEGAR

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH

2022

AGGRESSIVE BEHAVIOUR DETECTION SYSTEM FOR LIFT USING CONVOLUTIONAL NEURAL NETWORK

SURRYA VIKNESH A/L CHANDRA SEGAR

THIS THESIS IS SUBMITTED AS A PARTIAL REQUIREMENT TO OBTAIN DEGREE OF BACHELOR OF COMPUTER SCIENCE WITH HONOURS(NETWORK ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH

2022

NAME	**: SURRYA VIKNESH A/L CHANDRA SEGAR**
MATRIC NUMBER	**: BI18110184**
TITLE	**: AGGRESSIVE BEHAVIOUR DETECTION SYSTEM FOR LIFT USING CONVOLUTIONAL NEURAL NETWORK**
DEGREE	**: BACHELOR OF COMPUTER SCIENCE WITHHONOURS (NETWORKING ENGINEERING)**
VIVA'S DATE	**: 28TH JANUARY 2022.**

CERTIFIED BY;

1. SUPERVISOR

SIGNATURE

ASSOC.PROF.TS.DR. RAYNER ALFRED

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries, and references, which have been duly acknowledged

28 APRIL 2021
SEGAR

SURRYA VIKNESH A/L CHANDRA

BI18110184

ACKNOWLEDGEMENTS

First and all, I would like to express my thanks and gratitude to the Faculty of Computing and Informatics at the University of Malaysia Sabah (UMS) for providing me with the necessary educational foundation and experimental training to carry out my final year project.

Furthermore, I would like to convey my heartfelt gratitude to my esteemed project supervisor, Dr. Rayner Alfred, for his guidance, support, and direction during the course of this research. His honest and helpful advice has helped me weave throughout the composition of my thesis. Not to forget my examiners, Dr. Aslina and Dr. Awang Asri, who provided insightful feedback to improve my project.

My sincere gratitude goes to my course mates, housemates, and fellow friends for their moral support and conversations on this project. Finally, I'd want to convey my heartfelt thanks and affection to my parents and family members for their spiritual support and encouragement during the process of finishing my final year project. Again, I am grateful for all the help and support from everyone who helped me.

SURRYA VIKNESH A/L CHANDRA SEGAR

28 APRIL 2021

ABSTRACT

We live in a community that depends heavily on the use of CCTV cameras to maintain a high degree of surveillance. However, such a strategy is extremely problematic, as we typically use CCTV video just hours or even days after the incident has occurred (Febin,2019). It offers useful information in court but is seldom used to deter or respond to crime in real time. The explanation for this inefficiency is that the task of tracking vast amounts of CCTV footage is primarily carried out by a small number of security personnel. Fustigation, worker exhaustion, and discontinuity of observation make human monitoring ineffective.

Action Recognition is an active research area in the field of computer vision, and it has broad implications in today's world, and the detection of aggressive action is of high importance as it is directly connected to our protection and welfare. The idea of an intelligent monitoring system is to automatically identify unusual activity in surveillance videos and therefore to enable security staff to take necessary action at the right time. Aggressive behaviour action recognition has significant importance in developing automated video surveillance systems (Akti,2019). In closed spaces such as a lift, aggressive behaviour poses a significant danger to physical security and social safety.

It is therefore of considerable importance to automatically identify aggression activities from CCTV videos on the spot. The objectives of this project are ;(1) To design a deep learning based for aggressive behaviours in a lift, (2) To implement and optimize the structure of the deep learning architecture for aggressive behaviour detection and (3) To assess the performance of the proposed deep learning based aggressive behaviour detection. The proposed system has the potential to aid authorities and University management in security.

ABSTRAK

SISTEM PENGESANAN TINGKAH LAKU AGRESIF UNTUK LIF MENGGUNAKAN RANGKAIAN NEURAL BERKONVOLUSI

Kami hidup dalam komuniti yang sangat bergantung pada penggunaan kamera CCTV untuk mengekalkan tahap pengawasan yang tinggi. Walau bagaimanapun, strategi sedemikian amat bermasalah, kerana kami biasanya menggunakan video CCTV hanya beberapa jam atau bahkan beberapa hari selepas kejadian itu berlaku (Febin,2019). Ia menawarkan maklumat yang berguna di mahkamah tetapi jarang digunakan untuk menghalang atau bertindak balas terhadap jenayah dalam masa nyata. Penjelasan bagi ketidakcekapan ini ialah tugas menjejak sejumlah besar rakaman CCTV dilakukan terutamanya oleh sebilangan kecil anggota keselamatan. Fustigation, keletihan pekerja, dan ketidaksinambungan pemerhatian menjadikan pemantauan manusia tidak berkesan.

Pengiktirafan Tindakan ialah bidang penyelidikan aktif dalam bidang penglihatan komputer, dan ia mempunyai implikasi yang luas dalam dunia hari ini, dan pengesanan tindakan agresif adalah sangat penting kerana ia berkaitan secara langsung dengan perlindungan dan kebajikan kita. Idea sistem pemantauan pintar adalah untuk mengenal pasti secara automatik aktiviti luar biasa dalam video pengawasan dan oleh itu untuk membolehkan kakitangan keselamatan mengambil tindakan yang perlu pada masa yang sesuai. Pengecaman tindakan tingkah laku agresif mempunyai kepentingan yang signifikan dalam membangunkan sistem pengawasan video automatik (Akti,2019). Dalam ruang tertutup seperti lif, tingkah laku agresif menimbulkan bahaya yang ketara kepada keselamatan fizikal dan keselamatan sosial.

Oleh itu, adalah amat penting untuk mengenal pasti secara automatik aktiviti pencerobohan daripada video CCTV di tempat kejadian. Objektif projek ini ialah ;(1) Untuk mereka bentuk pembelajaran mendalam berdasarkan tingkah laku agresif dalam lif, (2) Untuk melaksanakan dan mengoptimumkan struktur seni bina pembelajaran mendalam untuk pengesanan tingkah laku agresif dan (3) Untuk menilai prestasi pengesanan tingkah laku agresif berasaskan pembelajaran mendalam yang dicadangkan. Sistem yang dicadangkan berpotensi untuk membantu pihak berkuasa dan pengurusan Universiti dalam keselamatan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Artificial Intelligence is improving rapidly with a soul purpose to make computer think like a human. Deep Learning makes this even easier by training the computer to imitate the workings of human brain in creating patterns and decision making. Deep neural learning and deep neural network are other terms for the same thing (Hargrave,2021). Current high- performance computers and plentiful of datasets paved the way for the principle of deep learning, which extracts automatically features or factors of difference that separate subjects from one another. Video surveillance data is one of the main contributors for the big data. Surveillance video analysis entails a sequence of modules such as object identification, behaviour recognition, and classification of detected behaviours into categories such as anomalous or natural (Sreenu,2019).

Violence has long been a major social issue. The increase in criminal acts in public spaces can be attributed to a variety of factors. Individual aggression, anger, and hate, as well as social and economic insecurity, are the primary causes of the rise of abuse. In recent years, the study of human activity behaviour has been illuminated by computer vision and data science. Prevention of crime and violent activities are not possible unless the brain signals are analysed and detected the specific pattern inferred the criminal thoughts in real-time (Halder & Chatterjee, 2020). But this is not achievable using current technologies. However, using deep learning-based machine learning, we can spot aggressive activity in public spaces. Surveillance cameras are now used in the majority of public places and private companies. The effective violent identification strategy will assist the government or

authorities in taking a quick and formalised approach to identifying conflict and preventing the loss of human life.

The main objective of this project is to provide a better surveillance system in UMS lifts using deep learning. This chapter consist of five sections. The first section is problem background which state the problems of UMS surveillance system.

1.2 PROBLEM BACKGROUND

University Malaysia Sabah authorities has been conducting very expensive surveillance methods such as case history-based CCTV. This will be hard to manage, and the costing will be high in server maintenance and data recording and archives. Lift surveillance usually overlooked or forgotten. These can result undetected violence in lifts such as robbery, fighting and even rape attempt. For example, in China there was a case where a woman was assaulted by a man in an apartment lift. The woman was beaten repeatedly by the man for asking him not to smoke in the lift. The security camera footage went viral in social medias and Chinese news. Yet the man is still at large (Feng. E,2016). This can be avoided if there was an automated alert system in the lift which would have alerted the authorities real- time, and they would have caught that man red handed.

An automated alert security system would be best for it. There is no need to be 24 hours monitoring since the system will detect automatically and alert the security staffs in case there is any suspicious activities in the lift.

1.3 PROBLEM STATEMENT

Most of the surveillance systems are focused on open areas like classrooms, labs and pathways so closed areas such as lifts are over-looked. UMS surveillance system can be better and more efficient by using deep learning method such as Convolutional Neural Network (CNN). The problems of UMS surveillance system are listed as below:

1) There are less surveillance systems in UMS lifts.
   Some of the lifts in UMS are not monitored by any CCTV cameras.

2) UMS uses human surveillance system which is less efficient due to their

inability to monitor all the lifts simultaneously.

Security staffs cannot monitor all the available lifts in UMS just by patrolling and monitoring CCTV feeds for 24 hours.

1.4 PROJECT QUESTIONS

1) What is the best method to design a surveillance system that can be used to detect aggressive behaviour in UMS lifts?
2) Can Convolutional Neural Network (CNN) be used to optimise the surveillance system architecture?
3) How effective is the proposed algorithm towards aggressive behaviour detection system?

1.5 PROJECT OBJECTIVES

The primary aim of this project is to support UMS authorities by improving the security monitoring system of the university. The objectives of this project are:

1) To design a deep learning-based system for aggressive behaviours in a lift. (This objective is mapped with Project Question 1)
2) To implement and 3ptimize the structure of the deep learning architecture for aggressive behaviour detection using Convolutional Neural Network (CNN). (This objective is mapped with Project Question 2)
3) To assess the performance of the proposed deep learning-based aggressive behaviour detection system. (This objective is mapped with Project Question 3)

1.6 PROJECT SCOPE

This project aims to develop a deep learning-based aggressive behaviour detection system which will help UMS authorities to detection an violence or aggressive behaviour in lift on time.

Modules	Description	Authority
User Management	• User register and log in to the system • View previous recorded CCTV feeds	Security staffs and UMS authorities personnel
Notification	• Security staffs receive notifications if there is any aggressive behaviour detected by the system	Security staffs
Admin	• Add or remove sample data such as videos	Security Division Head

Table 1.1 shows the modules used in the project

1.7 ORGANISATION OF REPORT

This report consists of 7 chapters which are Introduction, Literature Review, Methodology, System Analysis and Design, Implementation and Results, Experiment Results and Conclusion.

Chapter 1 is the introduction of the Aggressive Behaviour Detection System For Lift. In problem background aggressive behaviours which happens in lifts are not detected on time which causes a lot of security concerns and fear in many people. Its hard for the security staffs to monitor all the CCTV feeds simultaneously. Next, all the problems are gathered and 4ummarized in problem statement. Objectives and scope of project are then prepared with description of every modules which will be included into the system.

Chapter 2 is literature review. In this chapter, existing surveillance systems have been reviewed. Implementation of the existing system will be analysed then advantages and disadvantages are listed under this chapter.

Chapter 3 is methodology. This chapter explains the methodology of developing the system and the reasons choosing it. Activities involved in each stage of the project were attached. The required hardware and software to develop the system is also mentioned in this chapter.

Chapter 4 is system analysis and design is where the initial design for implementing deep learning for lift surveillance. System architecture and User Interface Design had been determined in this chapter.

Chapter 5 is implementation and results where implementation of Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) was conducted. The coding for this project started here.

Chapter 6 is experiment results. This is where the project prototype was tested in real-time for detection. The accuracy of the system was determined in the test run.

1.8 SUMMARY

In this chapter Aggressive Behaviour Detection System objectives are clearly stated. The problem statement and problem background are classified, and the scope was set for the development of the system. The report was organised according to the timeframe for developing the system.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter reviews existing systems which shares same purpose of this project. This existing helped to identify the advantages and disadvantages of using different algorithms and choose the suitable algorithm for this project.

2.1.1 Violence detection in surveillance video using low-level features

In this project, the motion regions were segmented according to the distribution of optical flow fields. Second, in motion areas, they propose to derive two kinds of low-level features that reflect the appearance and dynamics of violent behaviour. The proposed low-level features are the Local Histogram of Oriented Gradient (LHOG) descriptor extracted from RGB images and the Local Histogram of Optical Flow (LHOF) descriptor extracted from optical flow images. Third, the collected features are coded using the Bag of Words (BoW) model to delete unnecessary information and a specific- length vector is obtained for each video frame. The video-level vectors are eventually classified by the support vector machine (SVM).

As seen in Figure 1, the general flow chart of the proposed solution consists of five phases: pre- processing of video, segmentation of motion regions, extraction of low-level features, processing of features and classification/prediction. Consecutive frames are highly redundant, so there is no need to extract images frame by frame.

During the phase of video pre-processing, they extract frames from a long video sequence using a sparse temporal sampling strategy, which is called temporal segment framework. For a video clip V, it is equally divided into K segments $\{S_1, S_2, \cdots, S_K\}$, and K short fragments $\{s_1, s_2, \cdots, s_K\}$ are randomly sampled from each segment. Then, the algorithm of violence detection is designed in terms of the K short fragments. Next, they make details about the proposed approach(Zhou et al., 2018).

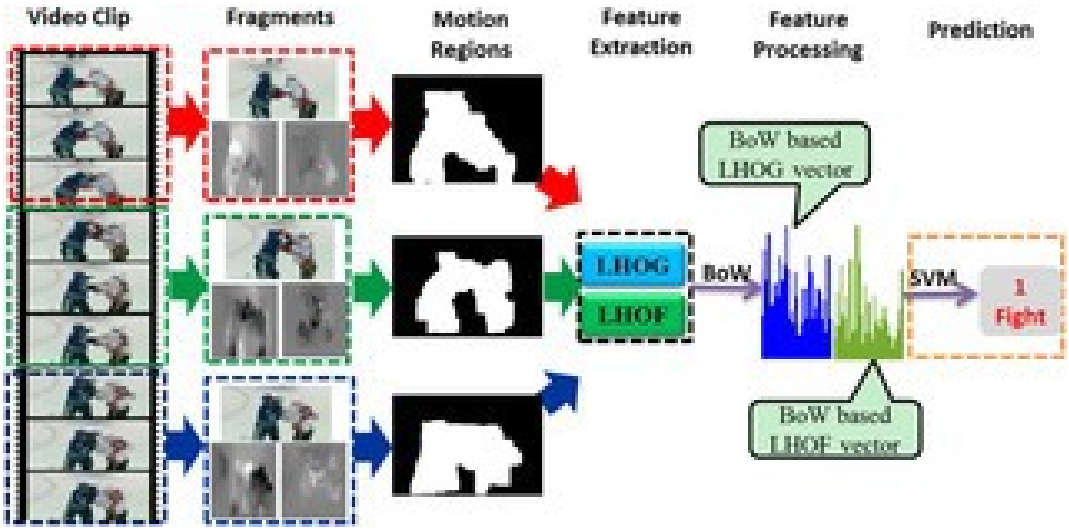

Figure 2.1: General flow chart of the proposed approach.
Five phases are involved: video preprocessing, motion region segmentation, feature extraction, feature processing and prediction

Advantages	Disadvantages
The system uses support vector machine (SVM) which is efficient with memory. The system uses Bag of Words (BoW) model to code which is easy to implement	SVM algorithm is not suitable for large data sets. SVM algorithm might not perform well if the number of features exceeds the number of training samples.

Table 2.1 shows the advantages and disadvantages of violence detection using low-level features

2.1.2 Digital triplet approach for real-time monitoring and control of an elevator security system

In this project, they used digital triplet concept which has an intelligent activity layer on top of the typical digital twin concept. This additional layer reflects the research, decision making, and progress implementation performed by engineers/technicians, who are motivated by years of expertise and experience. Figure 2 shows the concept of digital triplet.(Gichane et al., 2020)

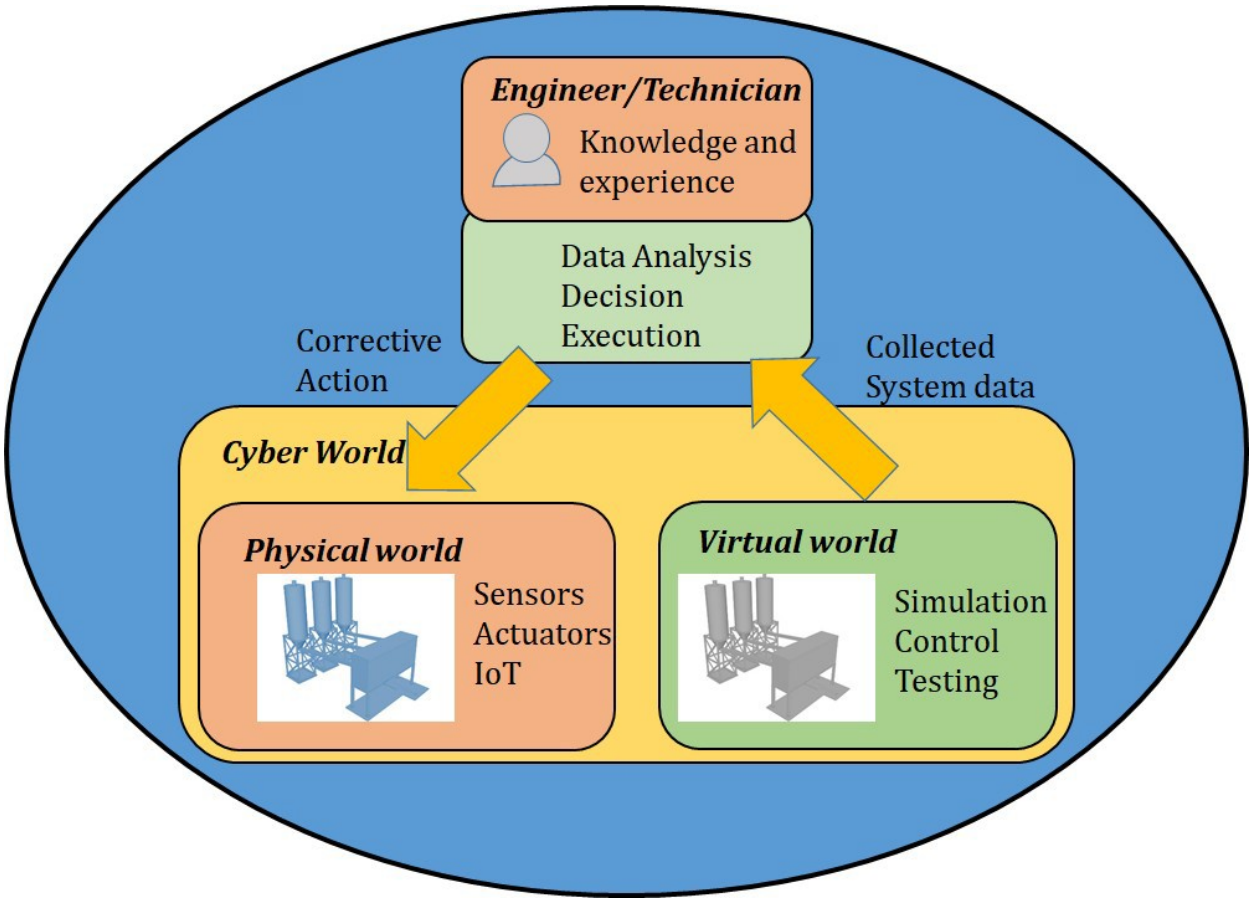

Figure 2.2: Digital Triplet Concept

This project was started by building a virtual prototype of a model elevator system that was connected to its physical equivalent and provided real-time system status information. Contact and data sharing between UA clients is accomplished by using the advanced features of the OPC- UA protocol. Next, to analyse image data obtained from a camera, an object detection module was embedded within the server. This module sensed dangerous items and took steps to stop elevator motion. The addition of a camera-based interface and a machine learning module increases the versatility

of the Digital Twin and transforms it into a Digital Triplet. The system also maintains a log based on the objects identified by the object recognition system; therefore, a user can see any prohibited items detected by the system, along with their corresponding time stamp.

Advantages	Disadvantages
1. Can detect harmful objects 2. All the video feeds are stored in a server which can be replayed anytime 3. Detects any harmful or prohibited objects in real-time	1. Too expensive to make 2. Can't detect any violence 3. Expensive to maintain

Table 2.2 shows the advantages and disadvantages of digital triplet approach for real-time monitoring.

2.1.3 Autocorrelation of gradients-based violence detection in surveillance video

This project objective was to explore spatio-temporal autocorrection of gradient-based features to detect any violence or aggressive behaviour in public places. The datasets used for this project was 1000 videos of hockey fights and no fights from the National Hockey Leagues. For crowd violence dataset they used 246 videos where half of it was violent videos and the other half was non- violent videos. Figure below shows the comparison between other algorithms.(Deepak et al., 2020)

Table 1
Detection result over Crowd Violence dataset.

Algorithm	Accuracy (%)
Histogram of Optical flow + BoW [5]	58.71
HNF + BoW [5]	57.05
MoSIFT + BoW [5]	57.09
Histograms of Optical flow + HMM(Baseline)	65.20
Motion Boundary Histogram + HMM(Baseline)	73.00
VIF + SVM [3]	82.13
Gaussian Model of Optical flow [6]	82.89
AMDN [7]	84.72
Weber local Descriptor + BoW [5]	88.16
MoIWLD + BoW [5]	88.16
Bi-Directional LSTM [8]	92.18
STACOG + KNN	**81.80**
STACOG + MBH features + SVM	**89.00**
STACOG features + SVM	**91.38**

Table 2
Detection result over Hockey fight dataset.

Algorithm	Accuracy (%)
Histograms of Optical flow + HMM(Baseline)	73.00
Motion Boundary Histograms + HMM(Baseline)	76.50
Histograms of Optical flow + BoW [5]	86.07
MoSIFT + BoW [5]	88.81
HNF + BoW + SVM [5]	89.27
Weber local Descriptor + BoW + SVM [5]	89.28
AMDN [7]	89.70
VIF + SVM [3]	90.07
MoIWLD + BoW [5]	91.80
Gaussian Model of Optical Flow [6]	92.10
Hough Forests [9]	94.60
Bi-Directional LSTM [8]	96.54
STACOG + KNN	**83.40**
STACOG + MBH features + SVM	**88.50**
STACOG features + SVM	**90.40**

Figure 2.3 shows the comparison between other algorithms

2.1.4 Convolutional Neural Network(CNN)

Because of its excellent recognition capabilities, the Convolutional Neural Network (CNN) is one of the most well-known learning algorithms in the area of computer vision. Currently, despite the fact that all networks are identical, many researchers create their own CNN infrastructure. CNN contains 4 layers which are 8 Convolutional Layer, Pooling Layer, Fully Connected Layer and Activation Layer. Convolutional layers use information from input data to generate a function map with the help of kernels, and the number of convolutional layers varies by design. The first layer of convolutional layers typically learns low-level features such as light and dark pixels, while the second layer of convolutional layers which learn horizontal and vertical edges. The next layer of convolutional layer learns more complicated functions such as ears, nose, and mouth. If the number of layers increases, the neural network learns more complex functions such as motion, object, and character recognition. This feature maps are fed into a non-linear activation function, which speeds up the CNN's ability to comprehend complex functions. Finally, one or more completely linked layers summarise this learnable content, which is then gathered into a softmax classifier. For the given input, the Softmax classifier returns the outcome probability of each class.

A recent research suggested a CNN-based deep learning system for detecting mitosis in histopathology breast cancer images using crowd-sourced learning on MICCAI-AMID13, a publicly accessible dataset of histology images from 23 patients(Albarqouni et al., 2016). Next, another research demonstrated that deep learning algorithms (CNN) are successfully used in medical imaging applications, proposing a multi-scale CNN-based method for automated segmentation of MR Images that classifies voxel into brain tissue classes(Moeskops et al., 2016).

2.1.5 Suspicious Action Detection in Intelligent Surveillance System Using Action Attribute Modelling

This research journal focuses on abnormal activities that are violent such as hitting, slapping, punching and snatching. The method proposed in this journal to model violence actions are using the Gaussian Mixture Model with Universal Attribute Model.

The motivation and application of these surveillance systems are explained in this research paper. For instance, this intelligent video surveillance system can is said to be useful in college and university campuses as it can be use to avoid unwanted violent activities such as fights. However, there are several challenges that can be faced in order to develop an intelligent system that is good for the detection of suspicious activities. Some of the challenges are changes due to illumination that occurs due to changes in weather and noise in video as these both can be a problem for video analysis. Next challenge faced is object shadow, which can appear as a problem as it might change the appearance of the object. Huge crowd and blurred objects may also pose as a challenge to detect and recognise violence. One of the biggest challenges in this is to develop a time system that processes at real time as background that are complex in videos may absorb more time for processing and tracking of the object in the video. The dataset for the violence detection in this paper consists of video sequences of scenes of fighting where is certain videos, people meet, fight and then run away. Meanwhile, in other videos, two to three people meet, and one person falls down and the second person runs away. The UCF101 or Kaggel dataset is used for this. The proposed framework for violence detection is using Universal Attribute Model (UAM) which does not depend on labelled violence videos for training. To describe the action process which is in every video, there is a need of pdf of parameters and to estimate that pdf, Gaussian Mixture Model (GMM) is used (Mudgal et al., 2021).

2.1.6 Campus Violence Detection Based on Artificial Intelligent Interpretation of Surveillance Video Sequences.

In this research paper, a video-sequenced based method is proposed to detect violence that happens in campus. Image features and acoustic features are used for the violence detection purpose. Role playing is used to gather the data of the campus violence data. From every 16 frames of video images, the 4096-dimension feature vectors are extracted. Few studies have covered on campus violence detection, thus in this study, the objective is to build campus violence databases and design a campus violence detecting method. The method used to detect physical violence is by video sequence. In this research paper, one of the methods used to detect violence in campus is by using video-based physical violence detection. The campus violence