Application of Cortical Learning Algorithms to Movement Classification towards Automated Video Forensics

By
Abdullah Alshaikh

A Thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy at
Staffordshire University, Stafford, United Kingdom

February 2019

Declaration

I, Abdullah Alshaikh, declare that all work submitted is my own, that any other external material used within my report my report is documented and referenced fully.

Date: ....

Signature:.......................................

Abstract

The need for proper and acceptable video forensics process is necessary due to the proliferation and advancement of video technologies in all aspects of our life, hence the legal system has heavily invested in this area. Also, the classification of the movements of objects detected from a video feed is an essential module to automate the forensic video process. Recently, new bio-inspired machine learning techniques have been proposed in the attempt of mimicking the function of the human brain. Hierarchical Temporal Memory (HTM) theory has proposed new computational learning models, Cortical Learning Algorithms (CLA), inspired from the neocortex, which offer a better understanding of how our brains function.

This research aims to study the requirements of video forensic investigation and Police procedures to propose a new semi-automated post-incident analysis framework and to investigate the application of the CLA to movement classification towards an automated video forensics process.

This research starts by reviewing the research related to police practices for video forensics as well as various proposed video forensic frameworks. Then a Questionnaire targeting CCTV/Video Forensics practitioners has been developed to capture the requirements for automating the video forensics process, this has been followed by proposing a new post-incident analysis framework. Then, a literature review covering state-of-the-art movement classification algorithms has been carried out. Finally, a novel CLA-based movement classification algorithm has been proposed and devised to classify the movements of moving objects in realistic video surveillance scenarios, and the test results have been evaluated.

Tests applied on twenty-three videos have been conducted to detect movement anomalies in different scenarios. Additionally, in this study, the proposed algorithm has been evaluated and compared against several state-of-the-art anomaly detection algorithms. The proposed algorithm has achieved 66.29% average F-measure, with an improvement of 15.5%compared to the k-Nearest Neighbour Global Anomaly Score (kNN-GAS) algorithm. The Independent Component Analysis-Local Outlier Probability (ICA-LoOP) scored 42.75%, the Singular Value Decomposition Influence Outlier (SVD-IO) achieved 34.82%, whilst the Connectivity Based Factor algorithm (CBOF) scored 8.72%. The proposed models, which are based on HTM, have empirically portrayed positive potential and had exceeded in performance when compared to state-of-the-art algorithms.

List of Publications

This research activity has led to several publications in international journals. These are summarized below:

1. AlShaikh, A. and Sedky, M., (2015). Post Incident Analysis Framework for Automated Video Forensic Investigation. International Journal of Computer Applications, 129(17), pp. 38-44.
2. Alshaikh, A., & Sedky, M. (2016). Movement Classification Technique for Video Forensic Investigation. International Journal of Computer Applications, 135(12), pp. 1-7.
3. AlShaikh, A. and Sedky, M., (2017). A Cortical Learning Movement Classification Algorithm for Video Surveillance. International Journal of Computer Applications 6(12), pp. 501-510.
4. AlShaikh, A. and Sedky, M., (2019), Application of Cortical Learning Algorithms to Movement Classification, International Journal of Computer Applications 8(3), pp. 58-65.

Acknowledgements

First of all, I would like to thank God Almighty for reconciling me in my whole life. Then I would like to express my thanks to the honourable man who is made of great knowledge and ethics, my supervisor Dr Mohamed Sedky. He has made great contributions to this research project since it was only an idea until it developed and became what it is now. I would also like to thank the great University of Staffordshire and the great professors for their great role in supporting research and researchers for the advancement of science not only in the UK but throughout the world, also to my wonderful mother and father who blessed me every step of my life.

My precious wife and beloved children who accompanied me throughout the course of my life.

I would also like to thank all the officials responsible for the university library that provided me the material necessary for this research thesis.

I thank all those who stood beside me.

List of Contents

List of Figures

List of Tables

Abbreviations

AI	**A**rtificial **I**ntelligence
ANN	**A**rtificial **N**eural **N**etwork
ACPO	**A**ssociation of **C**hief **P**olice **O**fficers
BPNN	**B**ack **P**ropagation **N**eural **N**etwork
CART	**C**omputer **A**nalysis and **R**esponse Team
CCTV	**C**losed **C**ircuit **T**elevision
CNN	**C**onvolutional **N**eural **N**etworks
CLA	**C**ortical **L**earning **A**lgorithm
DFF	**D**igital **F**orensics **F**ramework
DFRWS	**D**igital **F**orensic **R**esearch **W**orkshop
DTI	**D**igital **T**echnology **I**nternational
DVRs	**D**igital **V**ideo **R**ecorders
FBI	The **F**ederal **B**ureau of **I**nvestigation
FOV	**F**ield **o**f **V**iew
HDD	**H**ard **D**isk **D**rive
HTM	**H**ierarchical Temporal Memory
IDI	**I**ntegrated **D**igital **I**nvestigation
IDIP	**I**ntegrated **D**igital **I**nvestigation **P**rocess
IoT	**I**nternet **o**f **T**hings
LDA	**L**inear **D**iscriminant **A**nalysis
NIJ	**N**ational **I**nstitute of **J**ustice
NuPIC	**Nu**menta **P**latform for **I**ntelligent **C**omputing
NN	**N**euron **N**etworks
NVRs	**N**etwork **V**ideo **R**ecorders
OCFA	**O**pen **C**omputer **F**orensic **A**rchitecture
PRNU	**P**hoto **R**esponse **N**on-**U**niformity
RDS	**R**eference **D**ata **S**et
RMIs	**R**ecurrent **M**otion **I**mages
SDR	**S**parse **D**istributed **R**epresentations
SSA	**S**mart **S**urveillance **A**nalysis
SSF	**S**mart **S**urveillance **F**ramework
SP	**S**patial **P**ooler

SVM	**S**upport **V**ector **M**achine
SWGDE	**S**cientific **W**orking **G**roup on **D**igital Evidence
TWGDE	**T**echnical **W**orking **G**roup on **D**igital **E**vidence

Chapter 1

Introduction

1.1 Context

The amount of data captured from Closed Circuit Television (CCTV) cameras in the UK every day is more than what resides on the entire Internet (Draper, 2018). Nevertheless, this is only one type of digital evidence that includes covert video, photography and audio (Draper, 2018). The task of retrieving, analysing and storing this information is becoming a major challenge (Dee and Velastin, 2008).

According to the UK Home Office and the Association of Chief Police Officers (ACPO) National CCTV Strategy Document, CCTV should have the same performance standards as other crime scene evidence (Draper, 2018; Popoola and Wang, 2012). Police services require an effective and efficient means to recover and analyse large quantities of CCTV footage.

The use of CCTV cameras in the UK is rapidly increasing. According to Draper (2018), the number of CCTV cameras in London increased by 72% between 2012 and 2015, making Londoners one of the most meticulously monitored city occupants in the world. The Digital Technology International (DTI) group indicates that despite the growing investment in CCTV surveillance systems, today more than 98% of footages go unseen due to the high cost of skilled monitored staff (Creative, 2017; iCetana, 2017). Even when a video is monitored, there might be operator fatigue, which might lead to important security events being overlooked (Draper, 2018; iCetana, 2017). The lack of efficient monitoring often leads to a poor return on investment and sub-optimal security outcomes (Creative, 2017).

According to gov.uk (2017), in today's legal system, computer evidence is continuously being used. Over half of the cases in courts today rely on computer evidence.

However, a startling fact is that circa 85% of these cases make up the caseloads – cases awaiting to be heard and determined (ncsc.gov.uk, 2018) Another more shocking fact is that out of these cases, only 2% will be rightfully determined (gov.uk, 2017) wrongful conviction is dominant in the majority of the cases involving computer evidence, an aspect attributed to the analysts limited ability to review and analyse the videos and the lack of a standard framework for video classification, asncsc.gov.uk posits (2018). This, therefore, calls for the development of a standard video analysis framework that utilizes advanced analytic tools (Council of the Inspectors General on Integrity and Efficiency, 2012). The use of video analysis tools for post-incident analysis dictates the application of well-established digital forensics rules into the new video forensic framework. Steps of forensics include: Preparation, Collection, Examination, Analysis, and Reporting.

A video analytic system consists of many modules like motion detection, object classification, object tracking, and movement classification. One key module is the movement classification module. In this module, the movements of detected objects are recorded and compared to differentiate between normal and abnormal movements i.e. to infer anomalies (AlShaikh and Sedky, 2016). State-of-the-art movement classification techniques rely mainly on rule-based classification techniques, where the abnormalities in the video are traced and reported to the user (iCetana, 2017; FutureTech (Conference) and Park, 2012). Recently, machine learning techniques have been introduced to learn normal movements and hence to identify abnormal movements.

Artificial Intelligence (AI) and Neuron Networks (NN) have been heavily used to solve the problem of movement classification (AlShaikh and Sedky, 2016). However, they still suffer from various limitations such as their limited scope of operations (Council of the Inspectors General on Integrity and Efficiency, 2012). In the attempt of mimicking the function of a human brain, learning models inspired from the neocortex have been proposed, which offer a better understanding of how our brains function (AlShaikh and Sedky, 2016; Creative, 2017; iCetana, 2017). Recently, new bio-inspired learning techniques have been proposed and have shown evidence of superior performance over traditional methods (Li, C.-T and IGI Global, 2013; AlShaikh and Sedky, 2016). In this regard, the Cortical Learning Algorithms (CLA) inspired from the neocortex are more favored.

1.2 **Project Aim**

The aim of this research is to study the requirements of video forensic investigation and police procedures to propose a new semi-automated post-incident analysis framework and to investigate the application of the cortical learning algorithms for movement classification towards automated video forensics.

1.3 **Objectives**

In order to fulfil the aim of this research, the objectives of this investigation are to:

1. Critically review research relating to police practices for video forensics. Through a secondary research, this information will be acquired from already gathered data relating to Police practices for video forensics.
2. Carry out a literature review covering existing video forensics frameworks. This will be achieved through both primary and secondary researches.
3. Propose a new post-incident analysis framework.
4. Conduct a systematic review of state-of-the-art movement classification algorithms.
5. Devise a novel movement classification algorithm based on neocortex inspired learning techniques e.g. CLA.
6. Conduct experiments to test the devised movement classification algorithm,
7. Evaluate the test results.

1.4 Scope of the Investigation

In digital forensic, there are a series of data sources to be considered such as text, video, image and audio files, as Figure 1-1 shows. Following methodical tactic in pinpointing possible evidence from the data collected, which is known as the examination steps, that are carried out in order for the significance of evidence to be determined and conclusions to be drawn in the analysis steps. This study primarily focuses on video forensics. It will mostly concentrate on video evidence collection, cloning, extraction, conversion, and analysis before generating the report.

The video analysis will involve tracking the object, detecting changes, detecting the event and visualising it, to help in movement classification. Movement classification will be accomplished via the use of cutting-edge machine learning algorithms, like the Neural Networks and Artificial Intelligence (AlShaikh and Sedky, 2016; FutureTech (Conference) and Park, 2012). Rule-based methods or model-based methods will further be used to analyse the trajectory of moving objects and determine whether the movements are normal or abnormal.

This study focuses on the application of new bio-inspired machine learning algorithms in detecting abnormalities in objects' movements in the context of video forensic. Noteworthy, the study is limited to video forensics. Computer generated evidence, images, and mobile phone-based evidence are beyond its scope.

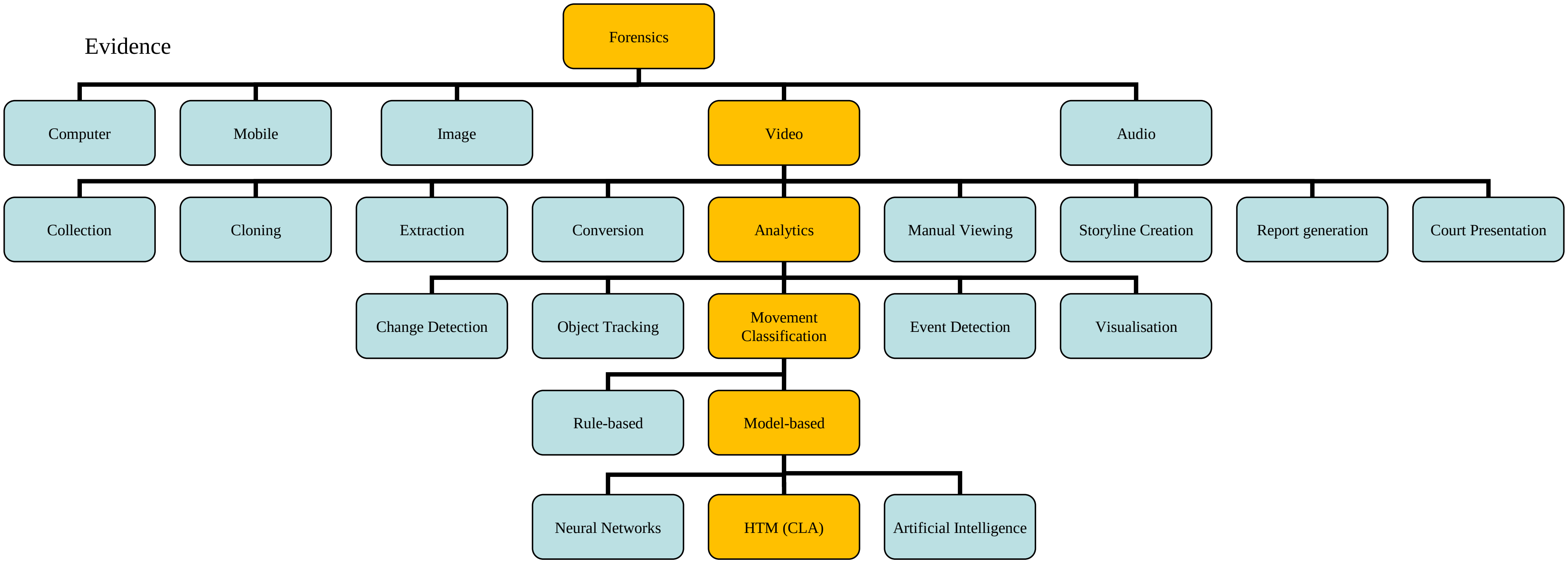


Figure 1-1: Scope of this investigation

1.5 **Research Approach**

The methods of conducting investigation in this research are obtained from using a standardized research method model. The research onion model has been considered, according to Saunders, Lewis, and Thorn hill (2007) who proposed the model as the illustration of research project stages in association with the research philosophy, data collection and strategies involved. This is further supported by Saunders, Lewis, and Thorn hill (2007), who indicate that having a clear research model streamlines the research process as it provides the research with clear guidelines on what to do. The research onion model has been applied as the method of investigation in this study.

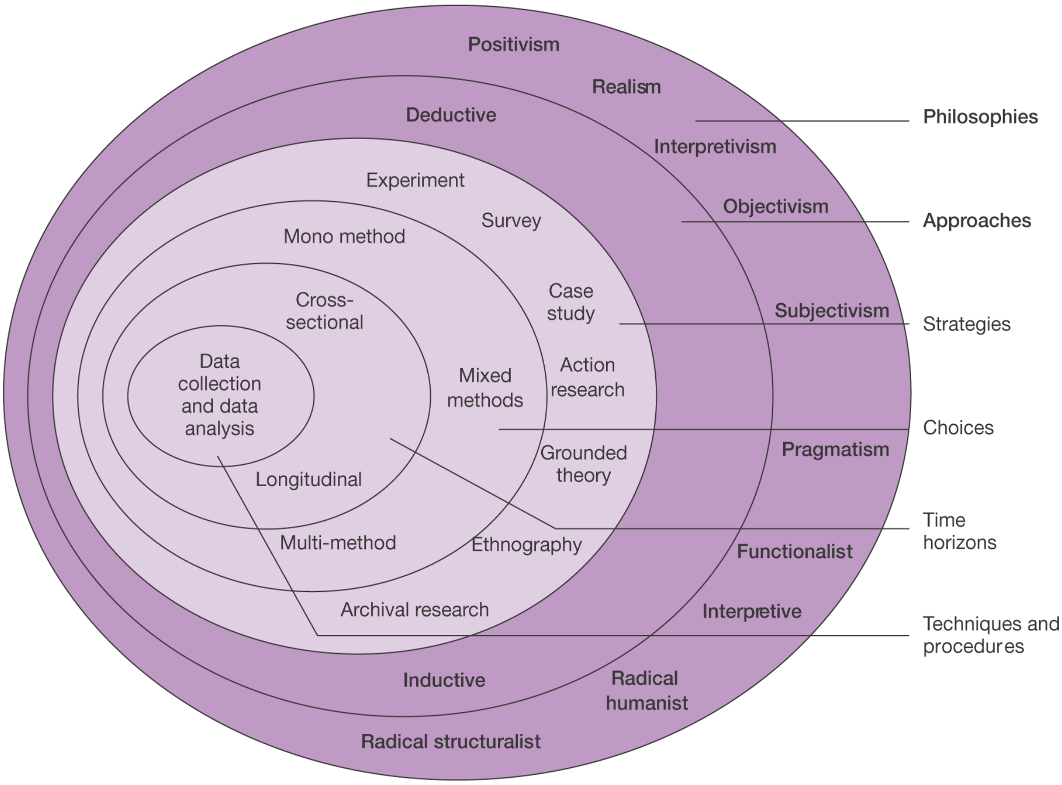


Figure 1-2: The research onion model (Saunders, Lewis and Thorn hill 2007)

In Figure 1-2, the study focuses on techniques and procedures, which are directly related to data collection and data analysis in the innermost layer of the onion model. In a research project, the choice of data types can be classified into qualitative, quantitative or both in some cases. The quantitative research method, according to (May, 2011), is about large number datasets that can be measured and analysed through statistical means. The qualitative research method focuses on interpreting the reality, which is associated with the inductive approach from the onion model in Figure 1-2. According to Robson (2002), adopting an inductive approach allows the researcher to develop theory from an analysis of the existing data. Through the analysis of existing video forensics, allows the researcher to make solid conclusions regarding movement classifications and the efficiency of applying machine learning to forensic evidence analysis (Bryman and Bell, 2007; Perumal, Norwawi and Raman, 2015). In this study, research data were collected by quantitative means through the interview, and large collection of datasets were obtained to test and evaluate the proposed technique.

The study will also utilize both primary and secondary data collection methods. Before collecting primary data, the researcher will first collect secondary data by undertaking an extensive review of relevant literature about the subject of the application of Cortical Learning Algorithm (CLA) to movement classification towards automated video forensics as suggested by Perumal, Norwawi, and Raman (2015). Search terms and phrases to be used will include “video forensics”, “digital forensics”, “video forensic frameworks”, “semi-automated video forensic framework”, “cortical learning algorithms”, and “post-event forensic analysis” (Hargreaves and Patterson, 2012). This review will help in learning different forensic models and related algorithms and their implementation as well as capabilities in video forensic investigation (Li, C.-T and IGI Global, 2013). As such, the review of existing relevant literature will help to provide a framework on which the semi-automated post-incident analysis framework to be proposed will be anchored on.

For primary data, interviews, questionnaires, and experimentation will be utilized. The interview will involve cross-examining about 30 people comprising of forensic video users, law enforcement agents, and experts such as forensic analysts to determine what is lacking and the need for new measures. The interview will be structured implying that it will be methodical, and a set of pre-determined questions will be used during interview sessions that will be expected to last for about 20 minutes per group – the main groups are forensic evidence users, forensic analysts, and law enforcement agents (Chris and David, 2012). AlShaikh and Sedky (2016) emphasizes that for interviews in video forensic research to provide information that is reliable and informative, the research should use unrestricted prompts. As such, the interviews for this project will use open-ended prompts like "tell me more about that" as well as "tell me, what happened?" to stimulate the interviewees to reveal more about the problems associated with the existing post-incident analysis frameworks and, thus, help in highlighting loopholes that will be addressed by the post-incident analysis framework that will be proposed (AlShaikh and Sedky, 2016: 39).

Similarly, a self-administered questioner containing questions relating to the need to automate post event analysis will be prepared and dispatched to Police officers, CCTV investigators, technicians, engineers and civilians in the Kingdom of Saudi Arabia. Through the questionnaires, the researcher will be able to identify more weaknesses of the existing video forensic frameworks and develop an innovative movement classification technique, and the semi-automated video forensic framework that will be specially designed to be used in criminal investigations (DFRWS, 2001).

An experiment will also be conducted to test and compare the effectiveness of the proposed framework as advised by Hargreaves and Patterson (2012). The selected dataset is VIRAT DATASET RELEASE 2.0, which will be used to evaluate, analyse and validate the proposed technique. This will be done with the primary objective of determining the HTM algorithm's ability to capture evidence, when objects are moving like this - HTM algorithm - is an important aspect on which the framework to be proposed will be centered on.

1.5.1 Evaluation of the proposed post-incident analysis framework

The evaluation of the proposed post-incident analysis framework will be carried out through a questionnaire to evaluate its validity. The questionnaire will target experts from police forces and practitioners as advised by AlShaikh and Sedky (2016) and the Council of the Inspectors General on Integrity and Efficiency report (2012). This is carried out to appraise the proposed framework and to assess its fitness for purpose amongst experts and professionals (Creswell, 2003; Newman, 2003). The questionnaire is designed with practical knowledge of forensic investigation practices in mind, and in conformity with the research aim and objectives.

1.5.2 Evaluation of the proposed movement classification technique

Considering the need for improved video analytic systems for the detection and classification of events in video feeds, various benchmark datasets are available in public domain (Li, C.-T and IGI Global, 2013; Chris and David, 2012). For example, the i-Lids, Imagery Library for Intelligent Detection Systems, datasets developed by the UK Home Office (i-Lids, 2013) and VIRAT dataset, a large-scale benchmark dataset for event recognition in surveillance video, developed by DARPA, Defence Advanced Research projects agency (Oh et al., 2011). These datasets are captured from realistic surveillance scenarios. This project will use some real datasets to test and evaluate the effectiveness of the devised algorithm.

The performance evaluation of how well the developed algorithm can differentiate between an unusual movement and a normal movement will be based on the ground truth provided by the used dataset (Creswell, 2003; DFRWS, 2001; Newman, 2003).

Objective evaluation will be adopted, in this project, which is targeted at comparing the output of the proposed algorithm to that of the ground truth and then measure how well it is able to determine such unusual movements accurately.

1.6 **Research questions**

The primary research gap that this thesis expects to fill is the lack of a standard framework for video forensics. The gap remains even though the current legal system profoundly relies on automated video forensics (Council of the Inspectors General on Integrity and Efficiency, 2012). Due to the existence of this gap, many crime culprits continue to walk scot-free after committing severe crimes, while some other people are erroneously convicted based on poor video evidence (AlShaikh and Sedky, 2016). Apart from wrongful conviction – that means justice is not served, the millions of pounds that governments have invested in digital forensics continue going to waste as a result of lack of a standard framework for video forensic (Council of the Inspectors General on Integrity and Efficiency, 2012). Similarly, despite the continued existence of advanced machine learning algorithms, movement classification remains a significant problem in video forensic (Future Tech (Conference) and Park, 2012).

Taking this into perspective, this research strives to answer the following research questions:

What are the main features of a semi-automated post-incident analysis framework for video forensic?

Can the CLAs contribute to a solution for the movement classification problem?

1.7 **Contribution to knowledge**

The research presented in this project has produced a novel bio-inspired movement classification technique based on HTM theory. This work has further contributed to the field of video forensic by proposing a new post-incident analysis framework linking critical literature reviews relating to Police practices for video forensics, existing video forensics frameworks and state-of-the-art video analytic technologies.

Currently, the information regarding current standards for video forensics in the UK and Saudi Arabia is limited. This study adds to the current standards by showing how best to analyse forensic evidence, captured from a CCTV source, and how to classify targets' movements. Similarly, in video forensic and crime investigation practices, this study provides a well-organised basis through the proposed semi-automated post-incident analysis framework. In academia, the outcome of this study is expected to offer insight and a new direction for further research in the field of forensic investigation.

1.8 **Thesis structure**

The structure of the thesis shown in Figure 1-3 is as follows:

Chapter 1: gives an introduction that covers the context, aims, objectives, and scope of the investigation, research approach, and contributions to knowledge, research questions and the structure of the thesis.

Chapter 2: presents a literature review that covers the area or video forensic, covering post-incident analysis application, forensic video process, Police forces practice, legal requirements in video forensic and video forensic development.

Chapter 3: focuses on the proposed semi-automated video forensic framework, including some discussion about other video forensic frameworks. It presents a questionnaire for the video forensic practice, which serves as the basis for the proposed framework. The analysis of the results of the questionnaire is presented as a detailed workflow.

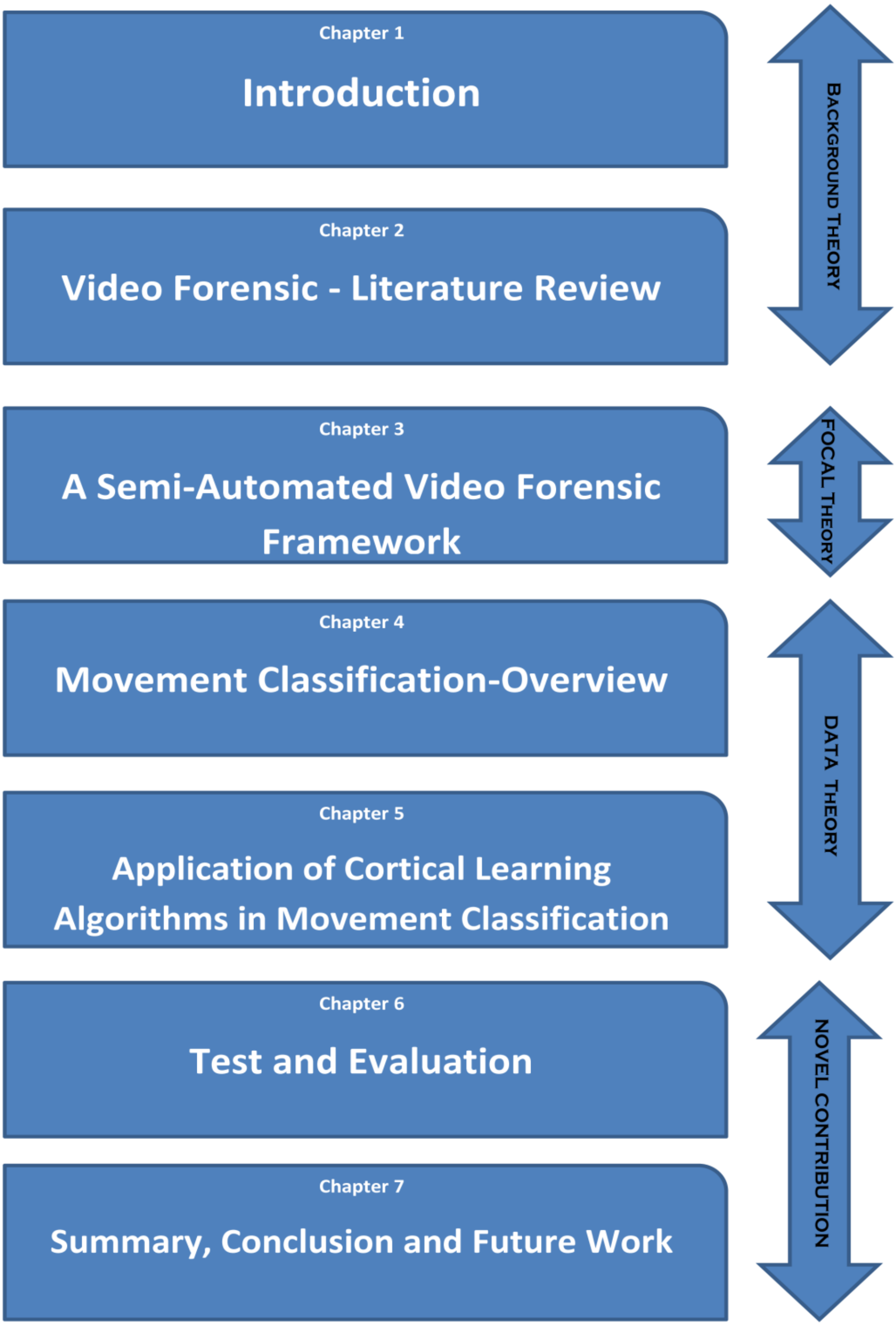


Figure 1-3: Structure of the thesis

Chapter 4: presents an overview of the movement classification within the video analytic and post-incident analysis arenas. In this chapter, problem definition, challenges for outdoor surveillance as well as related literature for movement classification techniques are discussed. The discussion is extended to artificial intelligence, neural networks and CLA.

Chapter 5: devises the application of cortical learning algorithms to movement classification. It introduces the HTM theory; it progresses with the discussion of CLA, its components and the choice of CLA. In this Chapter, the requirements, implementation of the devised movement classification technique is presented.

Chapter 6: tests and evaluates the devised movement classification algorithm. The Chapter starts by discussing the evaluation methodologies. VIRAT video dataset is presented including the experiments setup, data preparation, evaluation and test results.

Chapter 7: presents the summary and conclusion, which concludes the thesis and highlights future work. It summaries the proposed semi-automated video forensic framework, the devised movement classification algorithm and the application of CLA in movement classification.

Chapter 2

Video Forensic

2.1 Introduction

This Chapter presents the review of related literature within the scope of the research investigation. Forensics as the main subject, movement classification for video analytic, and the HTM are the focus of this review. It also includes the critiques of the various existing digital forensic models and highlights the significant positioning of the current research.

The use of video forensic has become an important mean of crime investigation. In Police departments, various devices have been adopted for video forensic purposes. Recently, there has been a significant increase in the use of CCTV cameras, mobile phones and portable cameras to capture activities that might be very useful in detecting, investigating and bringing criminal acts to justice using video forensic evidence, by Police. Consequently, the analysis and investigation of these video feeds require a systematic approach or framework to make the captured forensic information adequately fit for its purpose and to ensure that criminal acts do not go unpunished. This can be achieved using a standardised approach to ensure accurate analysis and interpretation of the video feeds. In the literature, different proposed frameworks are discussed.

2.2 Background

Forensic is the term given to an investigation of a crime using scientific means. It is also used as the name of the application of scientific knowledge to legal matters (Knowles et al. 2015). However, recently, forensic science is used to investigate nearly all crime scenes. With the advancement of science, most forensic science techniques are standard, this is a necessary part of criminal investigation.

The Criminal Justice System (CJS) is focusing on presenting evidence in the courtroom with a video linking technology. In this sense, the magistrate needs no paper documents as a form of evidence in the courtroom. Video evidence can be presented on a display screen: - presenters can present their evidence using their own devices; advocates can also display their evidence on their screens (Rowland 2014). Champion and Edgar (2013) noted that the prison reforms trust and prisoners' education trust are using computers, telephones, video conferencing tools, wing-based PC terminals, e-readers, and Internet technologies to transform rehabilitation in this sector.

The importance of technology improvement is taken as serious issues as can be seen that, the Ministry of Justice is investing up to £375m in court and tribunal technology reform. The Treasury has agreed on a one-off package of investment averaging up to £75m per annum over the five years from 2015/2020, which is currently used to deliver more efficient and effective courts and tribunal administration for all users and deliver significant savings (Ahamad and Hawkins 2016). The Home Office is also investing £2m as part of the Police Innovation Fund. To move towards paperless evidence reporting, the CJS is investing £160m in the future of the digital courtroom, moving away from the paper-based systems.

The digital video refers to the capturing, manipulating and storage of moving images that can be displayed on a computer screen (Buch et al. 2011). The word digital refers to a system based on discontinuous events as opposed to analogue. Before the digital era, to display, analogue video images on a computer, the video signal had first to be converted from analogue to digital (Vaughan, 1998). Hence, a camera and a microphone capture the picture and sound of a video session and send analogue signals to a video capture adapter board. The board only captures half of the number of frames per second that movies use to reduce the amount of data to be processed. Second, there is an analogue-to-digital converter chip on the video capture adapter card, and it converts the analogue signals (waves) to digital patterns (0s and 1s). Third, a compression/decompression chip or software reduces the data to a minimum necessary for recreating the video signals (White, 1999).

Video Forensic tools are usually used to respond to an incident in cases of criminal investigations. Most cases, when crimes are committed, surveillance cameras or even individuals at such scenes may record such a crime. As a forensic investigator, such evidence is handed over for investigation, there must be a careful examination of the video evidence first to justify the integrity of the video and also to extract useful information that will aid in responding to such incidence. With this development, the need for digital video forensic as legal evidence as well as detection of forgery is necessary and vital.

The legal system is working hard to capture video evidence which can be admissible in the court of law; hence there exists every need to develop research methods that handle video evidence and their admissibility in the court of Law. Therefore, in addition to the already stated composition of this Chapter, Section 2.3 below includes a review on the video forensics literature, covering digital forensic, forensic video process, post-incident analysis application, Police forces practice and the legal requirements of video forensic.

2.3 **Digital Forensic**

Traditionally, the digital forensic process begins with the collection, duplication, and authentication of every piece of digital media before examination; these first three phases of the digital forensic process are by far the costliest (Cantrell et al., 2012).

Computer evidence is becoming a routine part of criminal cases with nearly 85% of current caseloads involving digital evidence (Meadaris, 2006). Computer crimes are on the rise, and unfortunately, less than two percentage of the reported cases result in a conviction (Baryamureeba and Tushabe 2004).

Computer forensics can be traced back to as early as 1984, when The Federal Bureau of Investigation (FBI) laboratory and other law enforcement agencies begun developing programs to examine computer evidence. Research groups, like the Computer Analysis and Response Team (CART), the Scientific Working Group on Digital Evidence (SWGDE), the Technical Working Group on Digital Evidence (TWGDE), and the National Institute of Justice (NIJ), have since been formed in order to discuss the computer forensic science as a discipline, including the need for a standardised approach to examinations. Some previous work in the literature concludes that computer and network forensics frameworks consist of three basic components that Kruse et al. (2002) refer to as the basic building blocks in computer forensic investigations. These are: acquiring the evidence, while ensuring that the integrity is preserved; authenticating the validity of the extracted data, which involves making sure that it is as valid as the original and analysing the data, while keeping its integrity. Some process models that put the three factors into consideration include the Forensics Process Model (Kohn 2013).

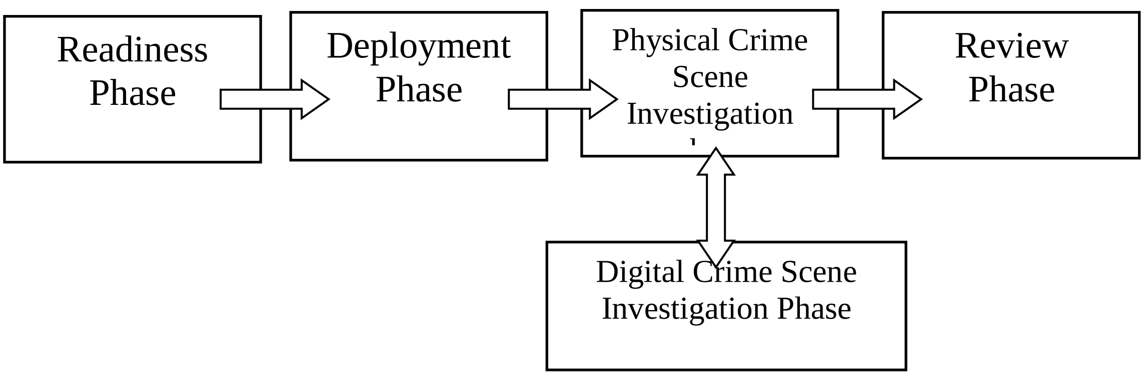


Figure 2-1: IDIP process Model (Brian C & Eugene, 2003)

The Abstract Digital and the Integrated Digital Investigation Process IDIP proposed in (Carrier 2003) organised the process into five phases: Readiness, Deployment, Physical Crime Scene Investigation, Digital Crime Scene Investigation and the Review phase. All these phases have specific roles to play in ensuring reliable digital data forensic evidence (Carrier and Spafford 2004).

The objective of the Readiness phase is to ensure operations and infrastructure that can fully support an investigation, while the Deployment phase is to provide a mechanism for an incident to be detected and confirmed. The main objective of Physical Crime Scene Investigation phase is to collect and analyse the physical evidence and reconstruct the actions that took place during the incident.

The goal of the Digital Crime Scene Investigation phase is to collect and analyse the digital evidence that was obtained from the physical investigation phase and or through any other future way round, and finally, the Review phase reviews the whole investigation and identifies areas for improvement if necessary. This proposal was later enhanced and came up with Enhanced Digital Investigation Process (EDIP), presented in (DFRWS, 2001), which separates the investigations at the primary and secondary crime scenes, while depicting the phases as iterative instead of linear. The EDIP model is based on the IDIP model and expands the deployment phase in the IDIP model to include the physical and digital crime investigations, while introducing a new phase dedicated to tracing back to the computer (the primary crime scene) that was used as a tool to commit the offence. In the dynamite phase, the investigation is conducted at the primary crime scene, with the purpose of identifying the potential culprits. The phase comprises of 4 sub-phases, namely, Physical Crime Scene Investigation, Digital Crime Scene Investigation, Reconstruction and Communication (Yusoff et al. 2011).

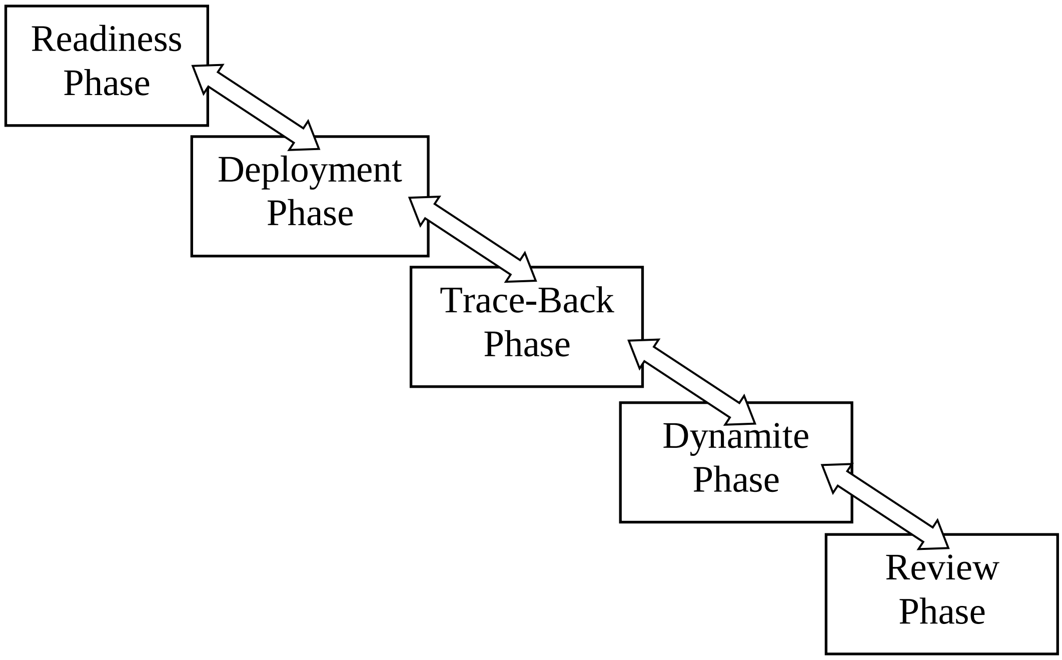


Figure 2-2: Enhanced Digital Investigation model EIDIP (DFRWS, 2001)

Proliferation and advancement of high Digital technology in all aspects of our life and desire of needs for optimising time and cost of doing things has pushed humans to deeply depend on digital data for decision making. The need for proper and acceptable forensic process is necessary. However, the computer crime culprit may walk Scot-free,or an innocent suspect may suffer negative consequences (both monetary and otherwise) (Bobick and Davis 2001) simply on accounting of a forensic process or investigation that was inadequate or improperly conducted (Baryamureeba and Tushabe 2004). Computer-related crime is on the rise and skipping one aspect of the forensic process or step may result in an incomplete or inconclusive result of an investigation that may affect interpretations and conclusions in a court of law.

In today information world, digital evidence has become a popular aspect of the forensic process. In the next sections, digital evidence definition, characteristics and roles are discussed.

2.3.1 Digital Evidence

There are many definitions explaining what is meant by digital evidence. In Cole et al., (2007), digital evidence is defined as the evidence in the form of information provided to the court of law with the aim of attesting that a crime has been committed.

Unlike Cole et al., (2007), Carrier and Spafford (2004) argued that digital evidence represents digital data that sustenance or contest a digital events proposition or the existence of digital data. In this definition, the compositions of an evidence that is not only capable of being presented in a court of law but may also possess investigative value. Evidence can be collected from vandalising or stealing of intellectual property, scam or related vices associated with digital gadgets. In their report, Perea et al. (2009) presented digital evidence as any data that can offer a substantial connection between the crime victim and the cause of that crime. However, the current study defines digital evidence as digital information or data packet conveyed or stored through a digital device that back or disprove a statement about the digitally related occurrence or happening that offers a connection between the crime victim and the cause of a crime event.

2.3.2 Characteristics of Digital Evidence

Digital evidence is delicate by nature. It can be destroyed, damaged or changed due to inappropriate management or investigation. It is easily copied and modified, and not easily kept in its original state. Precautions should be taken to document, collect, preserve and examine digital evidence (Carrier, 2003). Buttressing this point is research carried out by Summer (2009) which argued that data from computers can be accurately preserved and presented and, like all other evidence, digital evidence must be admissible, authentic, accurate, complete and convincing. Digital evidence is different from all other evidence in that it can change from moment to moment within a computer and along the transmission line, digital evidence can easily be altered without a trace and can be changed during evidence collection (Ademu et al. 2012). The main problem is to determine how an expert can measure the reliability of digital evidence. Digital evidence is data of investigative value that are stored on or transmitted by a digital device. Therefore, digital evidence is "hidden" evidence in the same way that deoxyribonucleic acid (DNA) or fingerprint evidence is hidden. In its natural state, digital evidence cannot be known by the content in the physical object that holds such evidence. Investigative reports may be required to explain the examination process and any limitations (Pollitt, 2007).

The digital devices such as those shown in Figure 2-2 may contain potential evidence that relates to criminal activity. The majority of digital devices contain data that could be lost if not handled properly. Examples of other digital devices are audio recorders, answering machines, cables, GPS devices, telephones, pagers, chips, digital organisers, copy machines, scanners, dongles, wireless access points and fax machines. Potential evidence can also be found on multiple computers connected to each other or the central server in a computer network.

2.3.3 Role of Digital Evidence

The major goal in an investigation is to relate the crime to its executor by uncovering compelling links between the offender, victim and crime scene. If the evidence suggests that the suspect committed a crime or violated an organization's policy, the investigator begins a case, which is a collection of evidence that can be presented in the court or to interested parties, and for an internal hearing in an organisation (Ademu, Imafidon, and Preston 2012). A witness may identify a suspect, but evidence of a person's involvement is usually more compelling and reliable. Previous scholars argue that anyone or anything penetrating a crime scene takes something of the scene with them and leaves something or a trace behind. In the physical world, a criminal might unconsciously leave fingerprints or hair at the scene and take fibre from the scene. Similar to categories of evidence in the traditional forensic sense, digital equipment and their attributes can be grouped into classes and individual groups. Printers, fax machines, scanners and all-in-one office devices may leave discernible artefacts that lead to common class characteristics allowing the identification of a particular device, e.g. Canon, Epson, etc. Even though individual characteristics could be sometimes rare, it should be possible to identify through detailed analysis (Ademu, Imafidon, and Preston 2012). The investigator must evaluate the evidence thoroughly and document the chain of evidence, or chain of custody, which is the route the evidence took from the time evidence was found until the case is presented.

2.3.4 Legal Requirements

Digital forensic as a discipline comprises information assurance and is perhaps one most closely defined by legal requirements and one whose growth and evolution are informed and guided by case law, regulatory changes, and the ability of cyber lawyers and digital forensics experts to take the products of forensic tools and processes to court.

The tension between privacy rights and law enforcement need to search and seize digital evidence sometimes mirrors, and frequently extends, the extent of tensions inherent in the rules of evidence. Technology is present in every aspect of modern life. At one time, a single computer filled an entire room. Today, a computer can fit in the palm of our hand. Criminals are exploiting the same technological advances which are driving forward the evolution of society. Today, virtually every business and personal document is prepared on a computer and mobile, hand-held devices. It is the use of specialised techniques for recovery, authentication and analysis of electronic data when a case involves issues relating to reconstruction of computer usage, examination of residual data, and authentication of data by technical analysis or explanation of technical features of data and computer usage.

Computer and digital forensics are useful for the detection and investigation of a crime committed on computers, computer networks, the Internet and other digital devices with the intent of giving digital evidence in law courts and tribunals (NITDA, 2014). It is also the professional extraction and handling of potential electronic evidence from any digital device or digital storage media to assist investigators, prosecutors, and the trier of fact (Judges, magistrates and members of tribunals) in a criminal justice system in arriving at the right judgment in litigation. In July 2011, Nigeria as an African Country, signed into law her Evidence Act, 2011, which recognises electronic, digital and computer-generated evidence.

No doubt that this singular act can transform our legal and judicial systems. As electronic evidence grows in both volume and importance in criminal and civil courts, judges and magistrates need to fairly and justly evaluate the merits of the offered evidence. To do so, prosecutors, investigators, judges and magistrates need a general understanding of the underlying technologies and applications from which forensic evidence is derived and the appropriate standards that must be met.

There is a need for standard documents aimed principally for the police officers, law-enforcement and security agents, military officers, prosecutors, anti-corruption agencies, regulatory agencies, other public-sector investigators and private sector investigators working for their organisation and those working in conjunction with law enforcement. However, some work in the literature indicates that every investigative process that reaches the point where specific competency questions are answered, digital evidence must survive the threshold test posed by the Scottish Executive (2003) of its competency as a class of evidence. The Court further clarified that the admissibility inquiry must focus "solely" on the expert's "principles and methodology," and "not on the conclusions that they generate. So, digital forensic evidence proposed for admission in court must satisfy two conditions: it must be (1) relevant, arguably a very weak requirement, and (2) it must be "derived by the scientific method" and "supported by appropriate validation. Digital forensic is, of course, highly technical, and therefore grounded in science, computer science, mathematics, physics, and so forth. It is also a discipline that requires knowledge of engineering, particularly electrical, mechanical and systems engineering. Moreover, applying the science and engineering in specific investigations is a complex process that requires the professional judgment that is sometimes more art than science.

2.3.5 Legal Evidence

As technology advances, the need for dealing with digital evidence increases, to achieve general forensic process and procedural principles applied to actions taken, Investigators examining digital evidence and activity relating to the seizure, examination, storage, or transfer of digital evidence should be documented, preserved, and available for review. The digital forensic process is a recognised scientific and forensic process used in digital forensics investigations (Kohn 2013). Digital forensics process is defined as some steps from the original incident alert through to reporting of findings. The process is predominantly used in computer and mobile forensic investigation and consists of three steps: acquisition, analysis and reporting.

Digital media seized for investigations usually referred to as an "exhibit" in legal terminology. Investigators employ the scientific method to recover digital evidence to support or disprove a hypothesis, either for a court of law or in civil proceedings. Various types of techniques are used to recover evidence, usually involving some form of keyword searching within the acquired image file; either to identify matches to relevant phrases or to parse out known file types. Individual files e.g. graphic images have a specific set of bytes which identify the start and end of a file if identified a deleted file can be reconstructed. Many forensic tools use hash signatures to identify notable files or to exclude known (benign) ones; acquired data is hashed and compared to pre-compiled lists such as the Reference Data Set (RDS) from the National Software Reference Library. On most media types including standard magnetic hard disks, once data has been securely deleted it can never be recovered. Solid State Disk (SSD) Drives are specifically of interest from a forensic viewpoint because even after a secure erase operation some of the data that was intended to be secure-erased persists on the drive. Once evidence is recovered, the information is analysed to reconstruct events or actions and to reach conclusions, work that can often be performed by less specialist staff (Reith et al., 2002).

2.4 **The Forensic Models**

In the past, different researchers in this field of study have introduced various digital forensic models. The details of these have been equally reported in the literature. Valjarevic and Venter (2012) reported on the first digital forensic model initiated by Ashcroft while working with the United States National Institute of Justice. The idea behind the initiation was based on the crime scene investigation process related to the electric field which eventually became a guideline for the responders who does not know the development. Over a period, the model has been adopted by the law enforcement agencies for the digital evidence identification and protection. The model has three phases comprised of evidence collection, examination and analysis. The evidence collection phase starts by performing a thorough search around the crime scene. The process of examination put together the evidence collected from the previous phase as transparent and identifies its source as well. The outcome of the examination phase is then analysed in the analysis phase, which reports and draws outcomes of all previous phases and the information that was collected in the entire process. However, the only constraint of this model is that it remains unclear and is not explained properly.

Reith, Carr and Gunsch (2002) expanded the Ashcroft (2001) model, clarifying the digital forensic investigation model by adding some phases to the process. They incorporated the traditional approach of accumulating the evidence to simplify the model. The first phase of this model is to identify the occurred incident and its type and provide all the assistance to achieve the goal of this phase. The second phase prepares procedures and methods that would be applied in other forensic model phases. Likewise, the second phase serves as preparatory guides for any needed search warrants to gather the evidence (Mushtaque et al. 2015). The third phase is to devise appropriate approaches and processes which will be adopted in the fifth phase of evidence gathering. The fourth phase of preservation is to preserve all the components and devices potentially containing the relevant evidence. After securing the evidence containing devices and components, the fifth phase of the collection is used to unify the procedures to record the physical scene. The sixth phase is to examine, which treats with the finding of the relevant suspect of the crime that was committed. The seventh phase is to analyse the importance of items on which the inspection has been performed. Presentation of all phases associated with this model is the one that comes at the end of analysis while the concluding phase is involved with the returning process for the sources and devices of digital evidence to the real owner after the accomplishment of the task of forensic investigation (Mushtaque et al. 2015). However, the only setback in this new phase is the similarity between the second and the third phase of the model.

In 2003, Carrier and Spafford developed another model of digital forensic investigation. They named the model Integrated Digital Investigation Process (IDIP) and it was adopted as another guideline for the forensic examiners to perform a digital forensic investigation and gather the evidence. This model was also organised into five phases. The objective of the readiness phase, which is the first phase, was to ascertain the actions and the given procedures of actions are adequate to help and support the process of investigation. For this reason, the first phase is basically used to prepare for the rest of the phases of the investigation process (Mushtaque et al. 2015). The second phase is the deployment phase: it supplies a system to the forensic examiners through which they could become capable of detecting an incident and then certify it.

The third phase is about the gathering and examining the physical evidence from the crime scene and go through the keen observation of the acts that were associated with the incident. The fourth phase is a sequel to the third phase but it deals with the examining and gathering of digital evidence which was obtained by the physical crime scene investigation phase. The remaining process used in phase four is similar to the third phase of this model. The fifth and final phase is to review the entire analysis that was performed during previous phases of the digital forensic investigation process and then underlined those areas where the room for improvement exists.

However, this model was found to be not appropriate as the deployment phase only handles the certification of the event, and also makes it difficult to swiftly validate the digital crime before proper investigation is being conducted (Mushtaque et al. 2015).

Khan, Kock and Memon (2010) reported the model proposed by Ciardhuain (2004) for digital forensic investigation. The model, however, does not have detailed phases as well as proper guidelines, and so it has not widely been used and sees no limelight compared to the previously developed forensic models. In the same vein, the model proposed by Perumal et al. (2009); Ruibin, Yun and Gaertner (2005) for the investigation and evidence collection could not be widely recognised. This is because they lack an unclear and detailed explanation of their process. Also, the model has not been categorised into separate phases.

According to Ruibin, Yun and Gaertner (2005) the model for investigation process proposed by Ademu Imafidon and Preston (2012) are regarded as the most recent guideline with a significant comprehensive report of phases and the dissemination of the whole investigation process into individual phases. According to Mushtaque et al. (2015), the model tagged the Systematic Digital Forensic Investigation Model (SRDHM) containing 11 phases to perform investigation. The details of the 11 phases are recorded in Mushtaque, Paquistao and Umer (2015). The findings of the model revealed that even though there is room for improvement, the lesson learned can be incorporated into forensic investigations as a result of all the phases' review and analysis.

More definitively, the following sections discuss the existing digital forensic frameworks and models in details.

2.4.1 DFRWS Investigative Model

Palmer (2001) discussed that the first Digital Forensic Research Workshop proposed a general-purpose digital forensic investigation model. The workshop aimed to allow means for knowledge sharing among professionals and academia based on digital forensic science. This collection of people consists of individuals from groups such as civilian, law enforcement agents as well as military personnel whose work is related to the use of forensic methods and procedures to discover criminal evidence through digital means (Aremu, Imafidon and Preston 2012).

The group created a consensus document that drew out the state of digital forensics at that time. Among the group's conclusions was that digital forensics a process with some agreed steps. The framework introduces digital investigation phases. The phases defined by the framework serve to categorise the activities of an investigation into groups along which a list of techniques where provided. They outlined phases such as identification, preservation, collection, examination, analysis, presentation and decision (Palmer 2001). As shown in Figure 2-3, the framework is represented as a table, the grey boxes at the top of their matrix, which is the column, is identified by the group as fundamental phases, and each row contains techniques although many will debate the forensic nature of each step of the process. This can be called an enhanced model of the DOJ model since it includes stages that were absent in the existing model, like the presentation stage. The overall benefit of DFRWS is that it is the first large-scale organisation that is anchored by academia instead of law enforcement; this is welcoming development because it defines and helps concentrate on the right path of the scientific community concerning the challenge of digital forensics, but the DFRWS model is just a foundation for further research (Aremu, Imafidon and Preston 2012).

2.4.2 Computer Forensics Investigation Process

Pollitt (2007) proposed an approach where digital evidence can be investigated in a manner that the result will be scientifically reliable and legally acceptable. The author compared and mapped the computer forensic process to the admission of documentary evidence in a court of law. Four different steps are identified as a guide to the admission of any evidence into court.

2.4.3 The Scientific Crime Scene Investigation Process Model

According to Ashcroft (2001), the US National Institute of Justice (NIJ) published a process model. The Technical Working Group for Scientific Crime Scene Investigation is completely designed as a procedure for improving the collection process. The first responder uses the document as a guide, and it is expected to be used by law enforcement and other responders with duty of guarding an electronic crime scene (Aremu, Imafidon and Preston 2012). The procedures involve recognition, collection, preservation, transportation and storage of digital evidence. The model consists of four phases, and starts with the Collection phase, which comprises the following; to search, recognise, collect, and document electronic evidence. The process proceeds with the Examination phase that assists in making the evidence detectable and clarifies its source and importance. It involves showing concealed and hidden information and the appropriate documentation (Aremu, Imafidon and Preston 2012). This is followed by the Analysis, which involves studying the product of the examination for its importance and probative value of the case. Then finally, reporting which involves writing a report, outlining the examination process and information obtained from the whole investigation.

2.4.4 Integrated Digital Investigation Process (IDIP) Model

Carrier and Spafford (2004) proposed a model, which is based on previous work with the purpose of combining the different available investigative processes into one integrated model. They introduce the idea of the digital crime scene which refers to the virtual environment developed by software and hardware where digital evidence of a crime or incident exits. The process started with readiness phases that require the physical and operational infrastructure to be ready to support any future investigation. The phase is an ongoing phase of the entire life-cycle of an organisation. After the Readiness phase is Deployment phase, which provides a mechanism for an incident to be detected and confirmed. The other phases introduced are Physical Crime Scene Investigation, Digital Crime Scene Investigation and finally, Review Phase where the whole investigation processes are reviewed to identify areas of improvement that may result in new procedures or training requirement (Carrier and Spafford, 2004).

2.4.5 Abstract Digital Forensic Model

Reith et al. (2002) proposed a model, known as the Abstract Digital Forensic model. The foundation of this model is using the ideas from traditional forensic evidence collection approach as practised by law enforcement (e.g. FBI). The authors argued that the suggested model could be dubbed as an enhancement of the DFRWS model because it is motivated inspired by it. The model includes provisions for tool preparation and the dynamic formulation of investigative methods (Aremu, Imafidon and Preston 2012). The model comprises nine components as follows:

- Identification – This recognises an incident from indicators and determines its type. This component is important because it has an impact on other steps, but it is not explicit within the field of forensics.

- Preparation – This involves the preparation of tools, techniques, search warrants and monitoring authorisation and management support.
- Approach strategy – This is formulating procedures and approach to use to maximise the collection of untainted evidence while minimising the impact on the victim.
- Examination – is a detailed, organised search of evidence concerning to the alleged crime. This emphasis on the finding and locating likely evidence.
- Analysis – This defines importance and probative value to the case of the examined product.
- Presentation - This concludes the summary and explanation.
- Returning Evidence – Physical and digital property returned to the proper owner
- Preservation – This involves the isolation, securing and preservation of the state of the physical and digital evidence.
- Collection – This is where the physical scene is recorded and duplicated digital evidence using standardised and recognised measures (Aremu, Imafidon and Preston 2012).

The three important phases introduced in this model were Preparation, Approach Strategy and Returning of Evidence. In Preparation, phase activities such as preparing tools, identifying techniques and getting management support were carried out. Approach Strategy was introduced with the objective to maximise the acquisition of unaltered evidence. In ensuring that evidence is securely reverted to the rightful owner or property disposed, the Returning Evidence phase was introduced. The future technologies and the technical details required to analyse them forensically can be instantiated to provide a standard methodology for providing electronic evidence (Reith et al. 2002). This will improve the science of forensics because it consists of the basis for examining new digital technology while simultaneously providing a common framework for law enforcement and the judicial system to work practicably within a court of law (Aremu, Imafidon and Preston 2012).

Case-Relevance Information Investigation

Rubin et al. (2005) recognised the necessity for computer intelligence technology in the existing computer forensic framework. It proposes an automatic and efficient framework to provide the case-relevance information by joining the existing computer forensic to the computer intelligence technology. The researcher clarified that for computer intelligence professional to provide more support in the investigation procedures and better knowledge reuse within multiple cases and sharing in computer forensics. The initial idea that the authors present is the view of 'Seek Knowledge', and this is the investigative evidence that determine the data analysis. Another perception defined by the researchers is the idea of Case-Relevance. They used this notion to describe the differences between forensics and computer security, likewise defining degrees of case relevance (Aremu, Imafidon and Preston 2012).

The authors argued that the major problem faced in the deployment of computer intelligence in digital forensics is the lack of Standard Test Dataset and Evaluation Criteria. Some ideas have been given to the formalization of the test and evaluation activities of a different product. It is very urgent to establish a formal and repeatable test dataset and evaluation environment for the data analysis phase. The authors emphasize that computer intelligence is extremely computational intensive and need large volume of data for training and testing.

2.5 **Video forensic process**

Videos forensic are becoming more popular and accessible through the various media technology advances which enable users to capture, manipulate and store video data in efficient and inexpensive ways. With the increasingly efficient compression formats and easiness of integrating videos in web pages, more people can enjoy producing and publishing videos in the digital world (Fisher and Schroeder, 1999).

Video forensic is a video recorded as digital data which can be stored, manipulated and edited on a computer. Video forensic differs from analogue video in some important ways: Video cameras are smaller and lighter than VHS camcorders and have higher picture quality. The key difference, however, is the ease with which digital video can be edited. This enables users to produce a video of a higher standard in a shorter time (Geradts et al. 2000). The video is also easier to share via the Internet and integrate with other ICT applications, such as presentation software. With this development, the need for digital or analogue video forensic as legal evidence as well as detection of forgery is necessary and important. This study thus aims to propose a post-incident analysis framework that can be used for video forensic investigation. Video forensics is considered as a process, and the steps or phases of this process consist of:

Preservation: This means ensuring the evidence being gathered does not and cannot change. Essentially, this can be considered as maintaining the 'integrity' of the evidence. It also means making sure that proper process is followed throughout.

Identification: This is the process of identifying which particular artefacts will be acquired during the evidence collection activities.

Extraction: This is the process of removing the evidence components that you will further analyse. This may be analysing a single file or files or could be an entire volume.

Documentation: This is making sure that the entire process is documented chronologically so that a third party could analyse the steps and following the same processes would yield identical findings. However, depending on whether the forensic examiner is working for the prosecution or defence, may conclude something different from these findings (Bird 2012).

According to Bird (2012), the forensic process can also be considered to include Preparation, Collection, Examination, Analysis, and Reporting. In this classification the preparation phase is included as illustrated in the Figure 2-3:

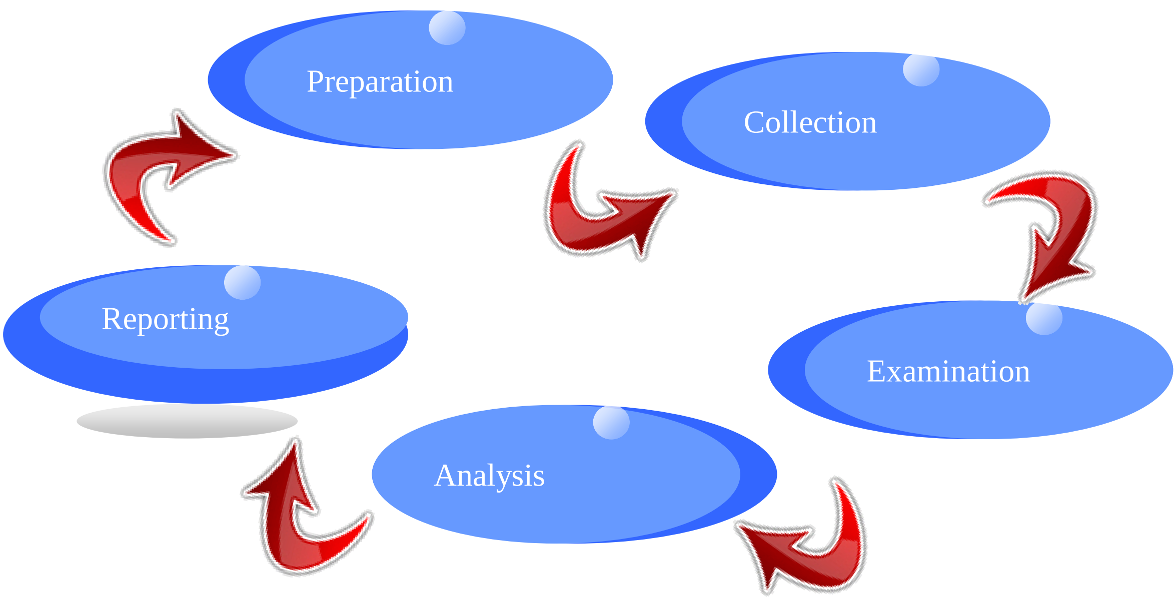


Figure 2-3.The Video Forensic Process (Bird 2012)

An explanation of the forensic process as presented by Bird (2012) can be seen below:

- **Preparation:** Proper preparation should be done on a particular investigation before such an investigation can be carried out.

- **Collection:** Data should be identified and acquired from all the relevant sources. This procedure should always preserve the integrity of the data. The data (evidence) collection should be carried out in an acceptable format.
- **Examination:** Automated and manual methods should be used to examine the collected data to assess and extract data of particular interest for that particular situation. In all of this process, the integrity of the data should always be maintained.
- **Analysis:** After examination, the results should be properly analysed by using well-documented techniques and methods to get the information that is useful in addressing the questions that were the main reason for the collection and examination.
- **Reporting:** A report of the results of the analysis should be written down. This report should include issues like actions that have been taken, why such actions were taken, findings made from the actions taken and recommendations for improvements to policies, guidelines and other aspects of forensic process amongst other issues (ACPO 2012).

There is another important framework which is targeted at powers bestowed on a police officer, which is referred to as Police and Criminal Evidence Act 1984 (PACE). Of importance to this research is the PACE Code F which deals with visual recordings with sounds of interviews with suspects (PACE 1984). These recordings could cover videos which could be used as evidence in the court of law. Video footages are admissible as evidence in the UK, USA and most European countries, when not tampered with. Hence, when using videos as evidence, this should be done under some guidance. Thus, the need for including the guidance mentioned above to be followed when carrying out research relating to video analysis. To analysis a video for forensic purposes, the above-described forensic steps and PACE Code F must be properly taken into consideration.

Video content analysis is targeted at carrying out forensics in a video scenario where investigators aim at knowing what went wrong? Who committed such a crime? Etc. Hence, video analytics is a key tool for video forensics.

2.5.1 Police Forces practice

Video forensics as a discipline demands specially trained personnel, support from management, and the necessary funding to keep a unit operating, and this can only be attained by constructing a comprehensive training program for examiners, which is mostly police personnel by providing sound digital video evidence recovery techniques, and a commitment to keep any developed unit operating at maximum efficiency (Dee and Velastin, 2008). The ability of Police personnel to follow the standard approach to video classification involves three major stages (Dee and Velastin, 2008): Participation in the legal system and testifying in court is associated with poorer mental health outcomes, especially when the experiences are particularly stressful for the individuals concerned (Quas et al., 2005).

It is difficult to estimate the time required to complete a full investigation. Information gained during investigative interviews thus plays a crucial role in the investigation of a crime. Fortunately, more than 30 years of research on crime, interviewing shows how investigative interviews should, and should not, be conducted (Quas et al., 2005).

The most reliable information is obtained when interviewers use open-ended prompts for information such as “tell me what happened?” and “tell me more about that;” such prompts also yield information that is most likely to be accurate.

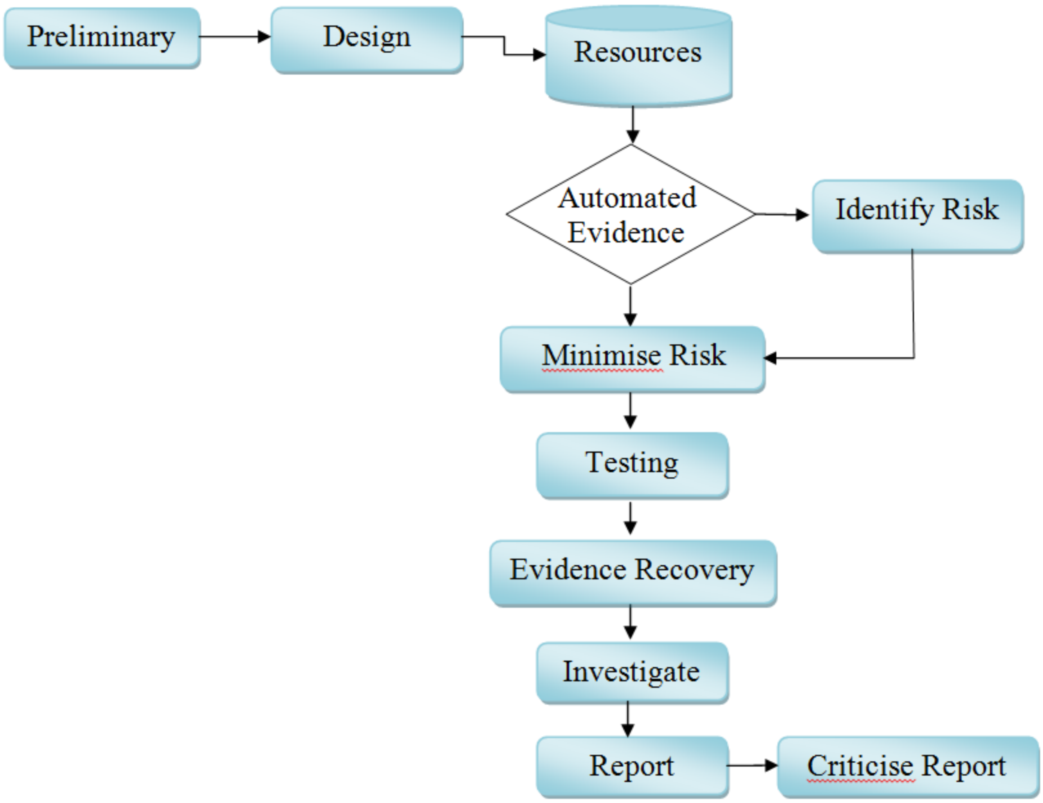


Figure 2-4: Visual Design Approach. (Sternberg et al., 2001).

Professionals have also translated research findings into guidelines for interviewers such as the Memorandum of Good Practice, Achieving Best Evidence, the NICHD Protocol and the Guidance for Interviewing Child Witnesses and Victims in Scotland. The survey yielded further support for suggestions that investigative interviews with children should be electronically recorded (Costello and Wang 2005). Furthermore, the quality of interviews must be independent and regularly checked to ensure that standards are achieved and maintained. Make an initial assessment of the type of case being investigated. The interviewer or investigator should systematically follow the following (Sternberg et al., 2001).

1. Determine a preliminary design or approach to the case.
2. Create a detailed design.
3. Determine the resources you need.

4. Obtain and copy evidence.
5. Disk drive.
6. Identify the risks.
7. Mitigate or minimise the risks.
8. Test the design.
9. Analyse and recover the digital evidence.
10. Investigate the data you recovered.
11. Complete the case report.
12. Critique the case

2.6 **Development in Video Forensics**

Several methods such as camera-based, coding-based, and geometrical/physical inconsistencies methods are used to assist in video forensics. With consideration of camera-based video forensics, some artifacts left behind are exploited for both camera identification and tampering detection (Wang and Farid 2007). Photo Response Non-Uniformity (PRNU) fingerprint technique was proposed by Mondaini (2007) which aids in detecting a different kind of attacks. Other works in this area include the use of the noise acquisition device to detect tampered regions in static scenes carried out by Kobayashi et al. (2010).

Even though there are several types of research in this area generally, the method suggested in this research works better when the video under consideration is uncompressed. Practically, most videos recovered from recording devices are usually compressed, and as such, this method may not be effective and applicable to compress video. Considering coding-based video forensics, forensic experts exploit the presence or irregularities in coding artefacts to assist in detecting tampering in videos. Research done by Wang & Farid (2009) exploited tampering detection, focusing on the assumption of double and single compression of videos. Several other works are done in this area, but it should be noted that most of the assumptions carried out by researchers do not apply in real life situations and up till now, there is no clear standard as to what extent using coding-based techniques could assist in presenting evidence in the courtroom.

For detection, considering geometry or physical lighting properties of a crime scene, it is very difficult to justify as whether such a scene is consistent or not. Several algorithms have been developed over the years considering this method including (Su, Zhang, and Lui 2009), which considers "ghost shadows" and Conotter et al. (2011) who is concerned with three-dimensional parabolic trajectory with considering of objects in a video. The methods mentioned above are useful in handling particular tasks. However, there are no defined patterns as to how these methods may assist in presenting the evidence recovered from such videos in the court of law. This project proposes a video analytics-based video forensic framework showing how the evidence could be analysed in order to be acceptable in the court of law.

2.7 **Conclusion**

Digital evidence must be properly admissible, precise, authenticated and accurate to be accepted in the court of law. Because of the fragile nature of digital evidence, the process must be handled properly and carefully.

The Enhanced Integrated Digital Investigation Process is capable of describing the development right from the point when the initial infrastructure is put in place, to investigate when an incident is reported through what it called trace back phase. A detailed digital forensic process provides important assistance to forensic investigators in gathering evidence admissible in the court of law. In the current study, it is concluded further that, there is a need to have a standard guideline for investigators when it comes to video forensic.

The video forensic community needs a structured framework for rapid development of standard operational procedures that can be tested effectively and validated quickly. This is because speed and accuracy are significant to both development and the outcome of a video or digital forensic examination. Digital forensic practitioners can benefit from the iterative structure proposed in this research to build a forensically sound case and for the development of consistent and simplified forensic guides on a digital forensic investigation that can be a guideline for a standard operational procedure and a model for developing future technology in the digital forensic investigation. Furthermore, video forensics such as imagery analysis, facial mapping and vehicle comparison, etc. are among the various video forensic services. Looking at the imagery analysis where a recovered image from different sources including CCTV, mobile phones, and the likes are analysed need also to be included for easy access and analysis of the recorded data.

In the next chapter, a semi-automated video forensic framework proposed in this study is presented. This is in response to the outcome of selected research questionnaire results and analysis. A review of the video forensic framework is also presented while a comprehension discussion of the proposed framework is given.

Chapter 3

A Semi-Automated Video Forensic Framework

3.1 Introduction

This chapter is an attempt to relate video forensics for evidence analysis and presentation in the court of law. The concept of video analytic is of significance due to the increased number of devices that are capable of recording and producing videos, increasing the probability of video evidence that can be used for a trial.

Previously, video analytic was subject to investigator viewing, but this is subject to human error. The majority of video forensic solutions are implemented in forensic practice, often for academic and law enforcement purposes. However, there are several challenges that this field faces, such as not following the right approach in carrying out video forensic activities and as such, making the evidence not admissible in the court of law.

This Chapter aims to describe the field of video forensics and point its benefits to the law enforcement agencies e.g., Police. A new semi-automated evidence analysis framework is proposed and if followed it could assist the courtroom to accept evidence from videos more easily.

3.2 Video and digital forensic frameworks - review

The previous Chapter presented a number of published models and frameworks around video, and digital forensic, many of the research outputs fundamentally use the concept or ideas derived from traditional methodologies, known as a physical forensic evidence collection of digital evidence strategy as practiced by police or any law enforcement agents. Such frameworks have previously been examined by Reith et al. (2002) for digital forensics. The authors argued that the proposed model can be termed as an enhancement of the Digital Forensic Research Workshops (DFRWS) 2001. Therefore, their model involves nine components such as:

Identification– it recognises an incident from indicators and determines its type. This component is important because it impacts other steps, but it is not explicit within the field of forensic.

Preparation – it involves the preparation of tools, techniques, search warrants and monitoring authorization and management support.

Approach strategy – formulating procedures and an approach to use to maximise the collection of untainted evidence while minimising its impact on the victim.

Preservation – it involves the isolation, securing and preserving the state of the physical and digital evidence.

Collection – This is to record the physical scene and duplicate digital evidence using standardised and accepted procedures.

Examination – An in-depth systematic search of evidence relating to the suspected crime. This focuses on identifying and locating potential evidence.

Analysis – This determines the importance and probative value in the case of the examined product.

Presentation - Summary and explanation of conclusions. Returning Evidence – Physical and digital property returned to the proper owner.

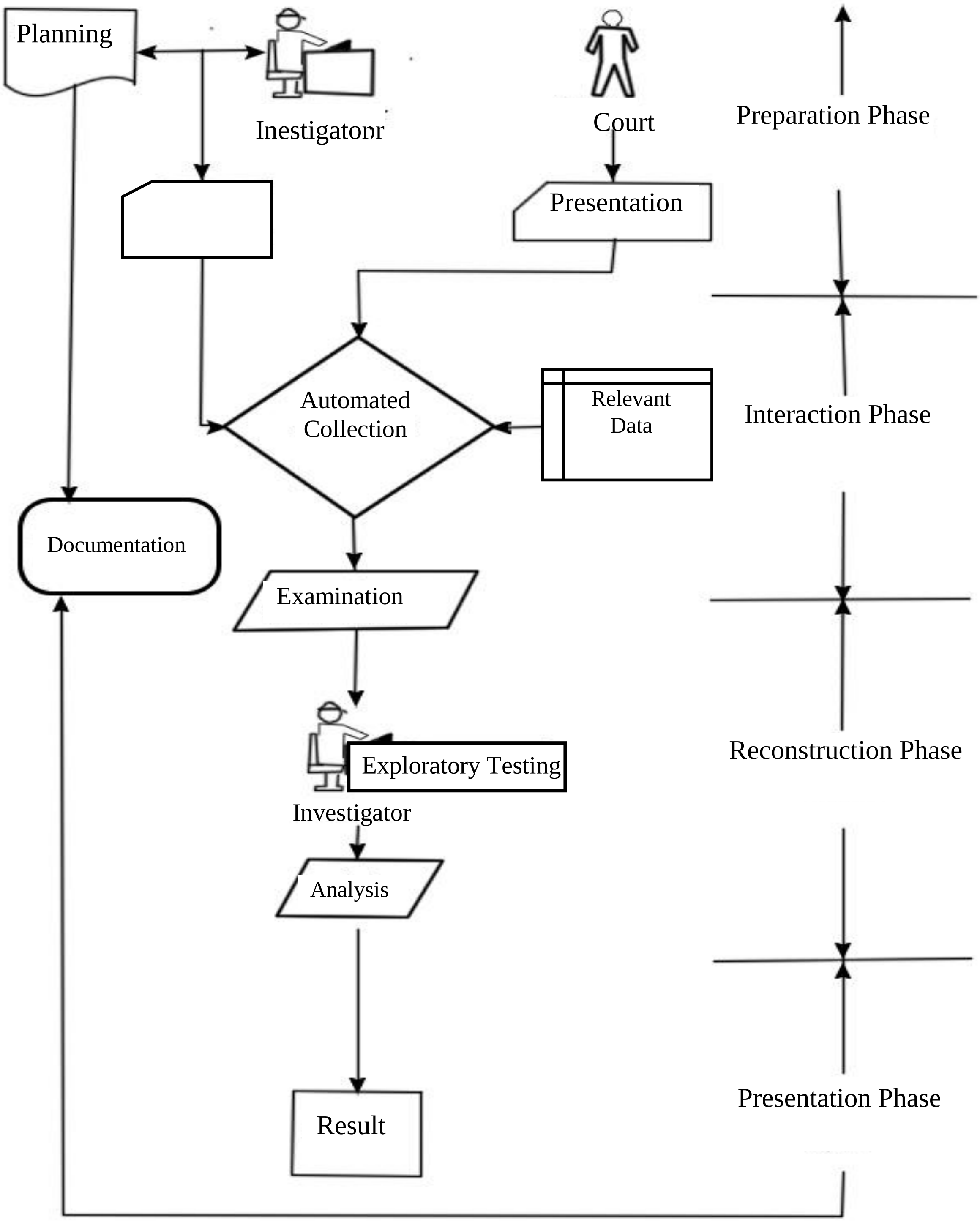


Figure 3-1: Digital forensic investigation Model (Ademu et al., 2012)

Other digital forensic frameworks have been proposed by researchers in the past, in some of these related works, frameworks such as Open Computer Forensic Architecture

(OCFA), Smart Surveillance Analysis (SSA), and Digital Forensics Framework (DFF) have been proposed, which will be discussed further in the next sections.

The Open Computer Forensics Architecture (OCFA), according to Schatz, Bradley and Clark (2006), is very popular distributed open-source Cyber forensic framework. It builds on Linux platform and makes use of PostgreSQL database for record storage purpose. It was developed for automating virtual forensics manner by the Dutch National Police business enterprise (Prasanthi 2016). The OCFA is also referred to by Raghavan (2013), as an integrated forensic architecture that can only integrate forensic image formats in the form of EnCase, RAW, and EWF for file system investigation.

Smart surveillance technology has become a critical component in security infrastructure (Wang and Farid 2006). It is used in pattern recognition and computer vision for information analysis of situated sensors (Collins et al. 2002; Bobick and Davis 2001; Bosch et al. 2008). Bosch et al. (2008) grouped smart surveillance into three broad categories such as; real-time alerts, automatic forensic video retrieval, and situation awareness. The real-time alerts category has two types of alerts, which are classified as user-defined and automatic unusual activity alerts. Smart Surveillance Analysis (SSA) is a framework that helps in the analysis sensor data which generate events of interest in the investigation environment. Wang and Farid (2006) applied the use of SSA to analyse the security requirements of an airport, by beginning with threat models. The SSA has been identified as isolated application that offer certain set of functionalities. However, even though it delivers some degree of value to forensic investigators, it fails to address the comprehensive security requirements. Nazare et al. (2014) proposed a version of smart surveillance technology, called Smart Surveillance Framework (SSF). The significant contribution of this framework is its consideration for problem-solving in a sequence rather than individually. The framework was therefore delivered to allow for implementing the results to the identified problem as a sequence of processing modules that communicate through shared memory (Nazare et al. 2014).

Digital forensics frameworks has been an important subject of study in academia within a relatively short period (Pollitt 2007). It is one of the fundamental way in which researchers try to understand the science behind forensic and to develop models, which reflect observation for forensic analysis. Reith, Carr and Gunsch (2002) considered a number of published Digital For Fun (DFF), and proposed a framework, which includes Identification, Preparation, Approach Strategy, Preservation, Collection, Examination, Analysis, Presentation, and Returning Evidence. The purpose of their proposed framework was to track the traditional forensic evidence collection techniques as practised by law enforcement. The physical investigation process was mapped with digital investigation process by Carrier and Spafford (2003) in their review of related digital forensic frameworks. They proposed a framework, called the Integrated Digital Investigation Process (IDIP), which comprises of 17 phases categorised into five different groups: Readiness, Deployment, Physical Crime Scene Investigation, Digital Crime Scene Investigation and Review Phases. Stephenson (2003), the Digital Forensic Research Workshop (DFRW) framework processes were considered as a ‘class’ while each action in the processes, were considered as an ‘element of the class. In this review, the six classes were further defined as the investigative processes. The number of the process were extended to 9 steps tagged as the End-to-End Digital Investigation Process (EDDI) through which coloured Petri Net modelling techniques were used to develop the process formal presentation. Carrier (2003) focused on abstraction layers that constitute forensic examination. The abstraction layer was identified as having two inputs and two outputs. Furthermore, the work classified forensic tools as presentation and translation tools. The translation tools read data by applying a set of rules before passing it to the presentation tools, which display the information in a way to be understood by the users. This was however proposed as analysis tool requirements. Contrary to Stephenson's approach (2003) of the DFRW framework, Mocas (2003) identified multiple digital forensic contexts in his related work review. Among the contexts identified are the military, law enforcement, as well as business security. However, in each of the context, a common process was identified, whereby one or more precipitating events initiated an examination, which was constrained by external forces and that specific outcomes could be defined as a specific sub-set of the desired outcome.

Baryamueeba and Tushabe (2004) proposed a modified version of the Integrated Digital Investigation (IDI) Model, which was introduced by Carrier and Spafford (2003). With the addition of two phases, known as trace back and dynamite. Their proposed model looks to separate the investigation into a primary and the secondary crime scenes, which was aimed at preventing discrepancies by reconstructing two crime scenes simultaneously. The work of Beebe and Clark (2004) provide a significant review of the already proposed digital forensics models, which are a single tier in nature. They, therefore, proposed multi-tiered process trends, which include sub-tasks for data analysis phase using an approach called Survey, Extract and Examine) (SEE). Ultimately, their proposed model was aimed at introducing objective-based concepts where analysis of tasks can be selective. Ruibin, Yun and Gaertner (2005) adopted Beebe and Carrier's model to introduce new concepts. The new concept encourages the notion of seek knowledge as well as the reuse of knowledge as investigation evidence that drive the analysis of data.

However, the identified shortcomings of the previously proposed digital forensic frameworks or models have led to the quest for continuous research for a more robust framework that is expected to add more positive outcome to forensic examination and analysis more than the existing ones. For this reason, section 3.3 presents questionnaire in which the outcome of its analyses has been used to develop the framework proposed in this study.

3.3 **Police forces investigation - Questionnaire**

The research questionnaire tends to address some issues about movement classification towards automated video forensic framework and will be shared and answered by Police officers, CCTV investigators, technicians, engineers and others in the United Kingdom and the Kingdom of Saudi Arabia. The questionnaire has a total of twenty-six (26) questions as illustrated in table 3-1 below. The respondents have been made aware that the questionnaire would only be used for research purposes i.e. to improve the present CCTV forensic tools for optimal performance; the respondents, with their full consent, were requested to answer the questions to the best of their ability and knowledge. The questionnaire, designed as a web link, was divided into three different links that are then converted into one questionnaire and powered by monkey survey. It is designed in such a way that the first link of the questionnaire has a total of ten (10) questions with its link[1]. Examples of the questionnaire are what your role is? The experience you have and the type of applications you normally used to acquire evidence just to mention a few. The second link[2] also has a total of ten (10) questions: the questionnaire here seeks to know some of the issues and challenges normally faced when using an automatic system to review CCTV footage, tools, equipment and software used for the analysis of the evidence. This will also allow the researcher to know the current state of the art forensic frameworks that are widely in use; the issue is faced when manually viewing CCTV

[1]https://www.surveymonkey.co.uk/r/LYDVMZJ

[2]https://www.surveymonkey.co.uk/r/7QKM9JF

footages, ways of authenticating videos footages to be sure that it has not been altered and finally the type of training that would be needed for the respondents. The third link seeks to address the importance of the feature in video forensic tools and what effect would the use of automated forensic tools in reviewing a month worth of data regarding cost reduction, time-consuming and preparation to court. This link[3] has a total of six (6) questions.

3.3.1 Video forensics practice – questionnaire

In table 3-1 below, the 26 questions designed for the research questionnaire about the video forensics practice are presented with the corresponding purposes of asking each question to the selected group of respondents.

Table 3-1: Questionnaire's questions

1	What is your job role? This question addresses which job roles the highest number of response as it has indicates those that are frequently using the CCTV application to detect and advice where necessary. This can tell if whether the participant is an engineer, a technician or a police investigator.
2	How many work experience (in years) do you have on CCTV area? Making decision normally depends on the number of years a staff put into an organization as it will indicate that he/she has good experience of video forensic investigation. This question will be used to make an informed decision as if the number of years is high then the respondent is highly experienced than those that their number of years is low.
3	Do you use CCTV or other recording evidence for the purpose of criminal prosecution? This question will address if the respondent uses CCTV for capturing activities or other recording evidence for criminal prosecution as it will assist in knowing which among the recording evidence has been used and why is the respondent using it instead of CCTV, which is the current state of the art.

[3]https://www.surveymonkey.co.uk/r/25YXMJ6

4	Where do you get most of your evidence from? This question will tell how those evidences are gathered and advice on how best and accurate evi could be achieved.
5	Out of these steps, which steps do you carry out after you have received the evidence? (s more than one if required) . This question deals with what do the respondents do with the received evidence.
6	. Which of these issues/challenges you face when manually viewing CCTV footages? . This question will highlight the negative or difficult aspects that are involved in reading the CCT footage manually to come up with a better framework that will address the difficulties.
7	Which video enhancement technique do you employ to get the best result? [Tick all that apply] This question will tell which among the mentioned video enhancement techniques are frequentl for a better result.
8	How do you authenticate the video recording to ensure it had not been altered? [Tick all apply] This question will address how well a recording is received without any evidence of tampering an what way do the respondent detects alteration if any.
9	. What training did you undergo in order to make a positive identification of a pers object? ining is the best practice to be up to date in terms of identifying an object or a person in a CCTV tage. It equally helps in advising which training is the best to attend.
10	What is your current practice regarding automatic reviewing of CCTV footage? Does the respondent have any experience regarding the usage of the automatic reviewing of CCT footage? This will indicate if he/she is using it regularly for reviewing recorded video footage.
11	. Which of these issues/challenges do you face when using automatic systems to review Y CCTV footage? This question will address the shortcomings of the automatic review of CCTV footage with the a identifying the highest challenge and proper solution to the challenges.
12	Which of the following equipment/software/tools do you use to analyse the evidence

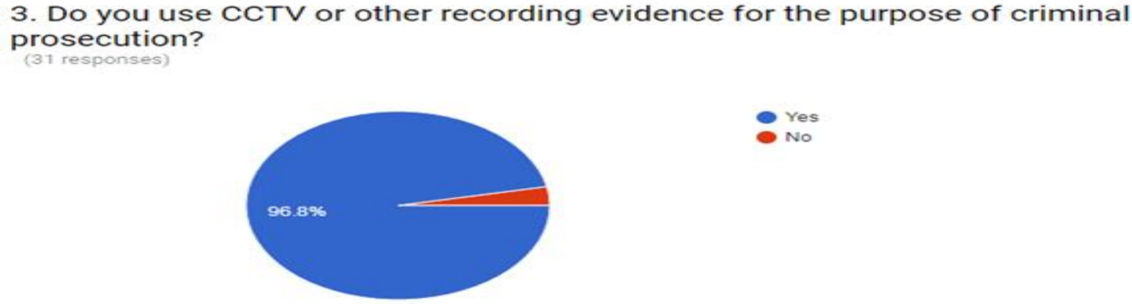
	received? Knowing which tools that are frequently used will assist in making the framework better as it wil indicate simple and well-planned tools that are easy to learn and possibly maintained.
13	Identify any of the listed problems you have faced when using video analytic software? Listing the problems faced by the respondents will go an extra mile to see how those problems w tackled by the proposed framework.
14	Which of these issues/challenges do you experience with your current video analytic sys This question will indicate the problems that are faced mostly with the use of the current video an systems and link those problems with the type of the software they are using and advice where appropriate.
15	Which of the following current CCTV/Video forensic framework did you know/use? It tells us which framework; the respondent is aware of to make sure that he/she is up to date rega the state of the art frameworks.
16	How do you ensure that there are no issues with the courts in accepting your CCTV/Vid forensic reports? This question would tell how the respondents are legally presenting their evidence in the court of following accepted forensic practices that whatever you presented can equally be checked forensi by any other investigator and achieved the same result.
17	What are the criteria you follow in selecting forensic video analytics? This question will tell how the respondent selects the best forensic video analytical tool based on needs and sophistication of the analysis that will be involved in tracing and detecting abnormal behaviour in a footage.
18	What are your expectations/requirements/dreams for the future capabilities of CCTV/Video analy It will assist in knowing how well we will provide a necessary framework that can satisfy the respondent needs.
19	Has the way CCTV is used changed in the time you have been a CCTV Investigator/Police Officer/Sergeant/Engineer/Technician? This question attempts to find out if there are changes in the field of CCTV video forensic right f inception to date to inform the proposal fora good and acceptable framework.

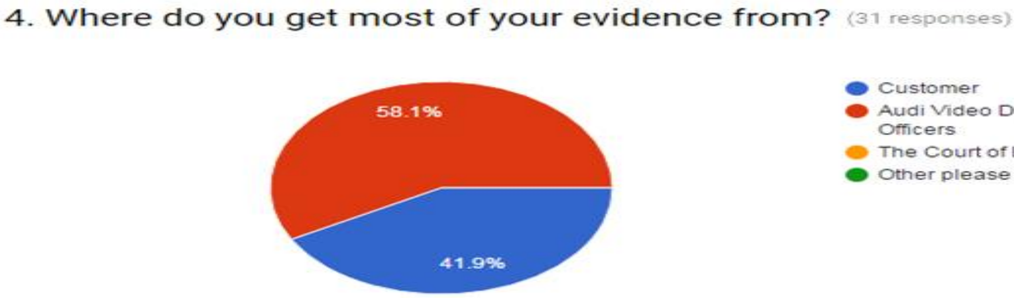
3.3.2 Analysis of the results of Questionnaire CCTV/Video Forensics Evidence

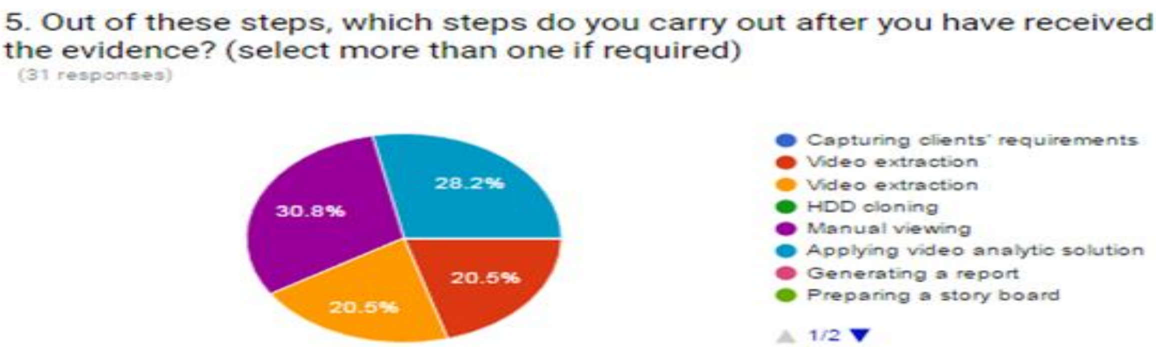
1. What is

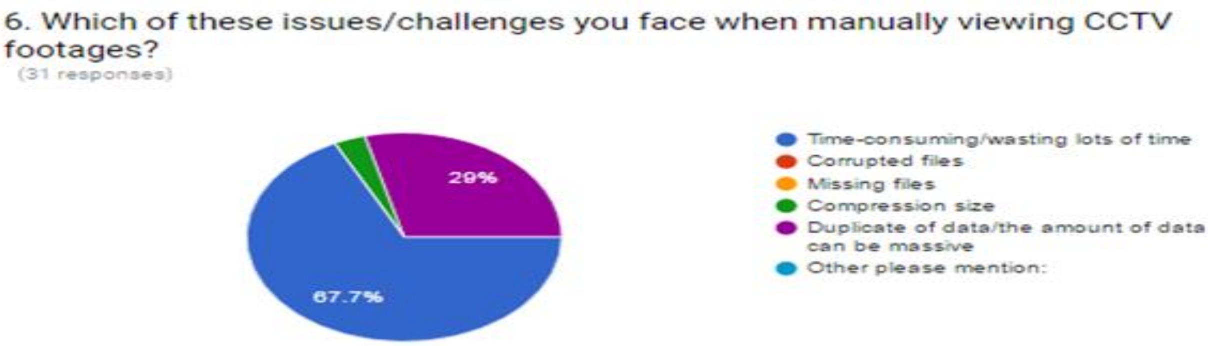
2. How ma
(31 responses)

20	What features are you looking for from your video forensic tools? This question deals with the features that the respondents deemed it necessary to be part o forensic tools for investigation so that it eases their work and allows them time to analyse and easily if those features are taking into consideration.
21	Rate the importance of these features (from 1to 4 where one is less important, and four most impo This question measures the importance of the features on a scale of 1 to 4 to deliver an acceptable framework.
22	Do you think using an automated video forensic tool to review one-month worth of video footage would reduce (analysis time/cost/the number of forensic investigators/the time to court? This question would inform the main features to be included in the proposed CCTV forensic tool solve the highlighted problems.
23	How long does it take for you to analyse 24hours of video? It will address the time it takes to analyse a 24 hours video to inform how long does investigation hours video takes (by the respondents) and how an analytic tool can speed up this process.
24	What are the challenges that you face when analysing a video? This will identify the type of issues and challenges that are normally faced during the analysis video by the respondents.
25	What could be developed to facilitate the analyses of video in order to make the analyses les consuming more accurate and less stressful for investigator? This will let us know what the respondents want to be developed that will assist them in investigation and that will solve most of their challenges and issues they face using the present analytic tools.
26	In your opinion would an automated video analytic tool be accepted by the court of law if not, wh This question deals with how the respondents think about the evidences they should presenti court of law, are the evidences convincing and can show clearly what the judges are looking for.









In general, there are 31 responses to the questions were asked. For question 1, out of the responses, the job role of 51.6 % is CCTV investigator. Also, 45.2 % of the responses are from police officers while 3.2% covers the rest that is not listed. Most of the responses from the CCTV investigators appear to be obvious since they frequently use the CCTV application for crime detection and advice.

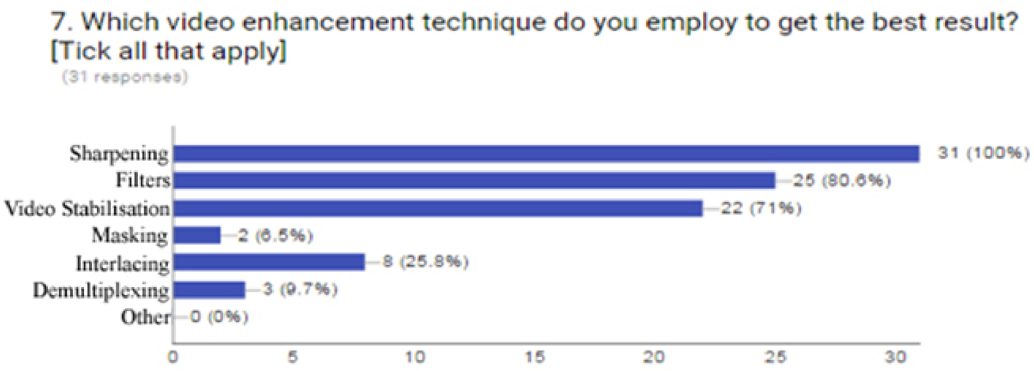
The focus of question 2 is the responses for the number of years of experience about CCTV. Most Responses is recorded for 7-10 years of experience at 45.2 % while the least is three years at 3.2 %. A respondent with six years and those with more than 11 years of experience are 12.9 % and 38.7% respectively. The implication is that majority of the respondents are highly experienced in video forensic investigation.

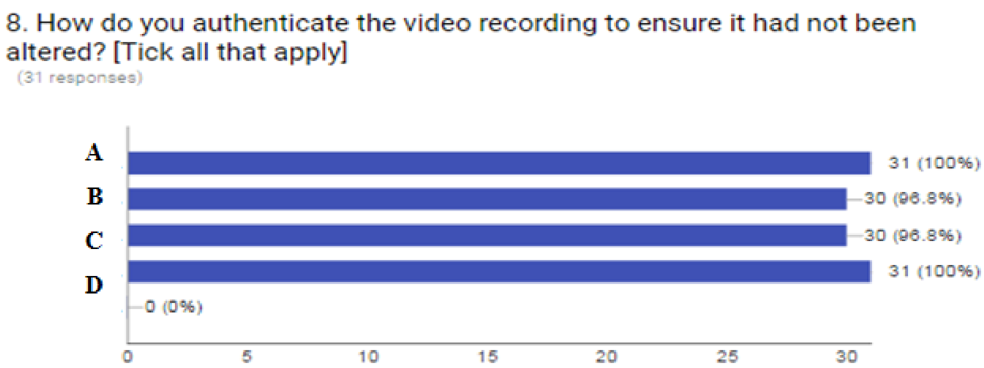
Question 3 which was asked to determine the use of CCTV application for criminal evidence, there is 96.8% usage of CCTV over other sources of recording evidence. This established the use of CCTV as it is the state-of-the-art technology for crime investigation which involves video recording evidence.

Question 4 posed the sources of evidence used in criminal investigation. While 41.9% of the evidence was obtained from the customer, the majority of it was through audio-video sources with 58.1% coverage.

In question 5, some steps were given in which the crime evidence can be analysed. Among these, video extraction and the application of video analytic solution carries major parts with 20.5 and 20.2 % respectively over manual viewing.

Question 6 deal with issues with a manual viewing of CCTV footages, which turned out to be time-consuming and time wasting at 67.7 % of the total responses.





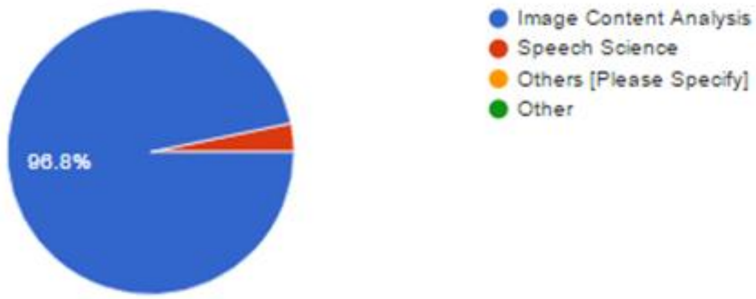
A=Examining the ambient sound for abrupt changes
B=Examining the volume and tone for clues as to the distance and spatial within a scene
C=Examining the lighting condition to estimate time of the day or environmental condition
D=Examining the presence of an unnatural waveform on the video signal which may indicate it had been altered or edited
E=Others

To determine the most commonly used video enhancement technique, question 7 asked respondents of their opinion. Video sharpening has been widely used from all responses with 100 % usage. This is followed by filters with 80.6 %, and masking has only two users out of 31 translating to 6.5 %.

The question of authenticating the video recording to ensure no alteration has been given five options for respondents to choose from. The responses clearly showed that examining the ambient sound for abrupt changes and the examination of the presence of an unnatural waveform on the video signal has largely been adopted by all 31 respondents. However, only one respondent has not used the B and C options, which has 96.8% usage each.

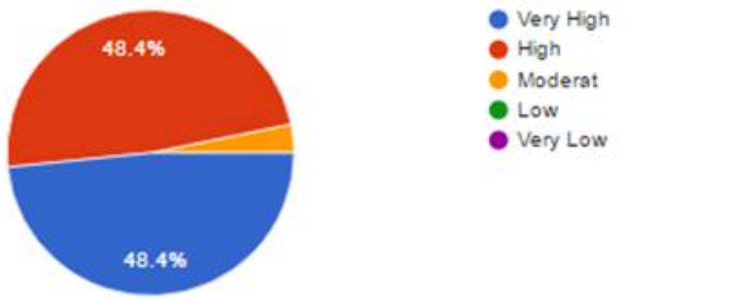
9. What training did you undergo in order to make a positive identification of a person or object?

(31 responses)



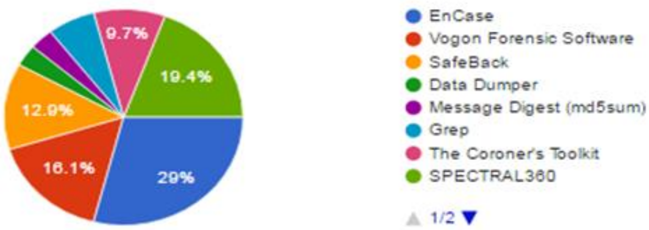
10. What is your current practice regarding automatic reviewing of CCTV footage?

(31 responses)



11. Which of these issues/challenges do you face when using automatic systems to review Your CCTV footage?
(31 responses)

12. Which of the following equipment/software/tools do you use to analyse the evidence received?
(31 responses)



13. Identify any of the listed problems you have faced when using video analytic software?
(31 responses)

14. Which of these issues/challenges do you experience with your current video analytic system?
(31 responses)

In response to question 9, 96.8 % of the respondents have undergone image content analysis training to make a positive identification of a person or object when analysing CCTV footages. Only 3.2 % responded to have undergone speech science training.

The current practice regarding automatic CCTV footage review has been asked in question 10. There is almost a division in the responses as 48.4 % each either very high or high, while 3.2 % has moderate practice.

The issues or challenges encountered with the use of automatic systems review of CCTV footage is sought in question 11. The corrupt issue files seem to be the major problem with 48.4%, followed by missing files with 25.8% and compressed size with 19.4 %. Data duplicating is recorded as the least problem, having only 6.4% of the responses.

The software packages or tools used in the video evidence analysis is asked in question 12. The Encase software gained more popularity usage with 29% among others such as data dumper with 19.4 %, Vogon Forensic Software which has usage of 16.1% while Grep, Message Digest, The Coroner's toolkit shares the rest of the responses.

Among the problems faced by respondents in the process of analysing video evidence is learning new software where 38.7 % of them admitted to in question 13. Whereas 22.6 % of them believe that the no existence of not enough functions or features is a problem they faced while the least problem is missing events with 9.7% responses.

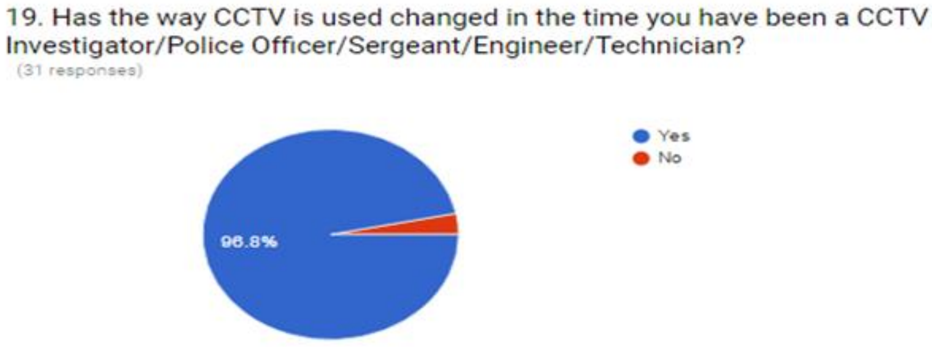
In question 14, when asked about the issues of challenges they experience currently with their video analytic system, 29% opted for the fact that it is time-consuming, 38.7% think it is difficult to watch a video file for long period, and 32.3 % find it difficult to capture all important information as a result of lack of concentration.

15. Which of the following current CCTV/Video forensic framework did you know/use?
(31 responses)

16. How do you ensure that there are no issues with the courts in accepting your CCTV/Video forensic reports?
(31 responses)

17. What are the criteria you follow in selecting forensic video analytics? Please choose one or more option)
(31 responses)

18. What are your expectations/requirements/dreams for the future capabilities of CCTV/Video analytics?
(31 responses)



To identify the type of CCTV/Video forensic framework respondents is most familiar with, question 15 was asked. As a result, some framework options suffice from their response. The Open Computer Forensic Architecture (OCFA) is popular with 35.5% of the respondents, 29 % of them has used or known about the Smart Surveillance Analysis (SSA) framework. The Digital Forensics Framework (DFF) is common among 25.8% respondents while only 9.7% have come in contact with FFmpeg framework. This shows the majority usage and familiarity of OCFA framework among others.

Question 16 explored the means which respondents used in ensuring the courts accept the provided CCTV/Video forensic reports. The majority of them which represents 35.5% ensure that images are playable and capture what has been reported.

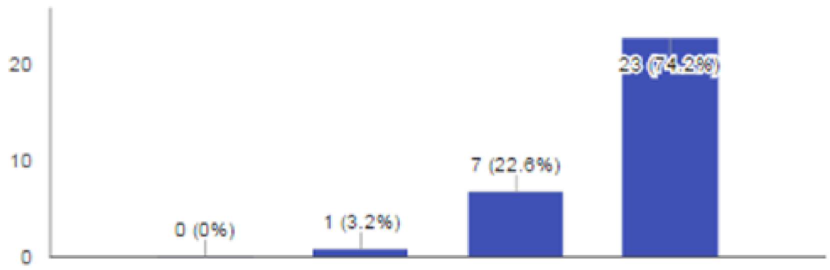
The same percentage of respondents adopts recognition accuracy criteria in selecting their forensic video analytics as posed in question 17.

It is important to know how well to provide a framework that would satisfy respondent needs. Therefore, question 18 was asked about their expectation, dreams and requirements for future CCTV/Video analytics capabilities. Moreover, so, 58.1% of the respondents are in support of Managed Video as a Service (MVaS), 38.7% for Supports for Video IP networks Supports for open standards ONVIF and PSIA while the tiny few of 3.2 % opted for Support for megapixel resolution as their respective expectation or requirements.

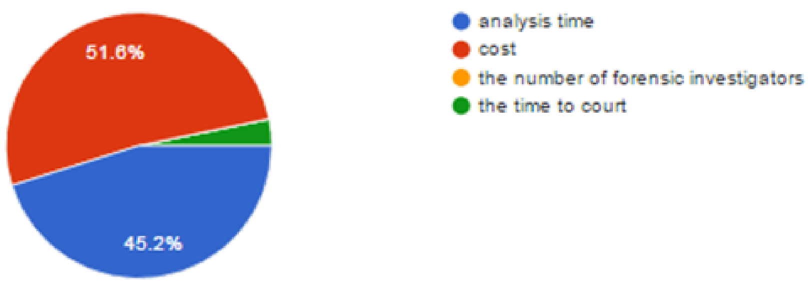
Most interestingly, the way in which CCTV is used has not changed since the time 96.8% of the respondents have been using it as an investigator, police officer, technician, sergeant or engineer.

Question 20 was asked about the features that the respondents are most interested in video forensic tools. In response, 41.9 % are most interested in movement detection, 32.3 % prefer image enhancement, and 19.4% opted for video stabilisation while the rest are interested in tracking.

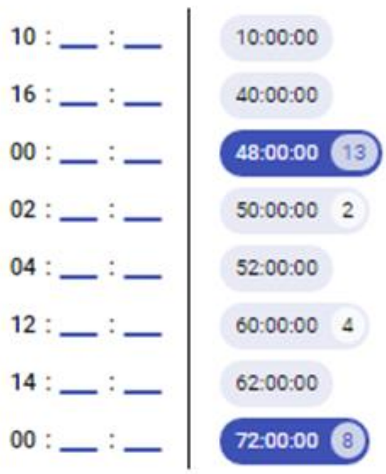
21. Rate the importance of these features (from 1 to 4 where 1 is less important, and 4 is most important)
(31 responses)



22. Do you think using an automated video forensic tool to review one month worth of video footage would reduce?
(31 responses)



23. How long does it take for you to analyses 24 hours of video? (8 responses)



24. What are the challenges that you face when analysing a video?

31 responses

Time Constraint	13/31
Blurry images	5/31
Unclear images	5/31
Not Enough Time	3/31
Camera Angles And Lighting	1/31
Time Constraint And Blurry Images	1/31
Blurry Images And Time Constraints	1/31
Unstable Images	1/31
Images Not Clear	1/31

25. What could be developed to facilitate the analysis of video in order to make the analysis consume less time and to be more accurate and less stressful for investigators?

31 responses

Better Tools	5/31
Good Software	5/31
Good Forensic Tools	3/31
Behaviour Analysis of Images	1/31
Good Video Analytic Software	1/31
Better Video Analytic Software	1/31
Good Tools	9/31
Good Analysis Tools	1/31
Better Analysis Tools	1/31
Good Forensic Equipment	1/31
Good Forensic Software	1/31
Upgraded Softwares	1/31
Better Softwares	1/31

26. In your opinion would an automated Video analytic tool would be accepted by the court of law if not why?

31 responses

Yes	28/31
Yes, as long as it can be explained and replicated	1/31
Yes it Should be Accepted	1/31
It Should be accepted	1/31

In question 21, it is important to deliver an acceptable framework that works; this was aimed to be achieved by asking respondents to rate the importance of the features. The rating is expected to demonstrate the most useful features or less useful features that have generally been agreed on by the respondents who are familiar with these features. Out of the 31 responses, 20 opted for 4 rating which signifies that these features are most important while no rating for less important. The result suggests there are more important features which are deemed significant. Meanwhile, it is equally important to identify the main features to be included in the proposed CCTV forensic tool to solve the highlighted problems. For this reason, question 21 was asked to know the effect of using an automated video forensic tool to review one-month worth of video footage, regarding analysis time, cost, time to court and the number of forensic investigators. The response reveals two major splits, one for cost and the other for analysis time where 51.6% of the respondents believed the cost would be reduced while 46.2% thought it would be analysis time that would be reduced, and the rest opted for time to court. Question 22 prompted question 23 for the time taken to analyse 24 hours of video. Only eight responses were received for this question in term of analytic tool to speed up the process. In question 24, the researcher aims to identify the type of challenges that can be faced during video analysis. The response from 31 respondents reveals that time constraint is the major issue, with 13 out of 31 agreeing to it. Blurry and unclear images problem was supported by each of 5 respondents while only one respondents each believed unstable images, camera angle and lighting, images not clear and the combination of two from the list are the main challenge. In other to deliver a framework that consumes less time which is more accurate and less stressful for a forensic investigator, question 25 was asked. The 31 respondents were asked what could be developed to facilitate the analysis. Five each believed better tools and good software is significant in achieving it, while nine respondents thought good tools need to be developed. Question 26 was meant for the respondents to give their opinion on whether or not an automated video analytic tool would be acceptable in the court of law. The response showed that 28 respondents out of 31 believed that it should without contemplating while the remaining three also agree with a bleep.

3.4 **Proposed framework**

Although there are other digital forensic frameworks proposed by other researchers in the past, the extensive review of these related frameworks reveals a considerable gap for improvement, regarding achieving a clear and concise understanding of the difficult crimes and evidence that are associated with the recent forensic inputs and investigation strategies. The need for a standard guideline for investigators contributes to the development of the detailed digital forensic process presented in this study to assist forensic investigators in gathering evidence admissible in the court of law. This aspect is found to the lacking in the previous framework proposed. Based on the responses and analysis of the above questions, the proposed framework in this study is presented. Currently, forensic research focuses mainly on identification, individualization and association at the source level. Even though a forensic expert aims at achieving any of these, considering video forensics, the court of law finds it difficult accepting evidence which they are not so sure how it was handled, gives a description of our proposed framework with the associated steps. The eleven (11) steps are explained in figure 3-2:

Figure 3-2: Video forensic framework.

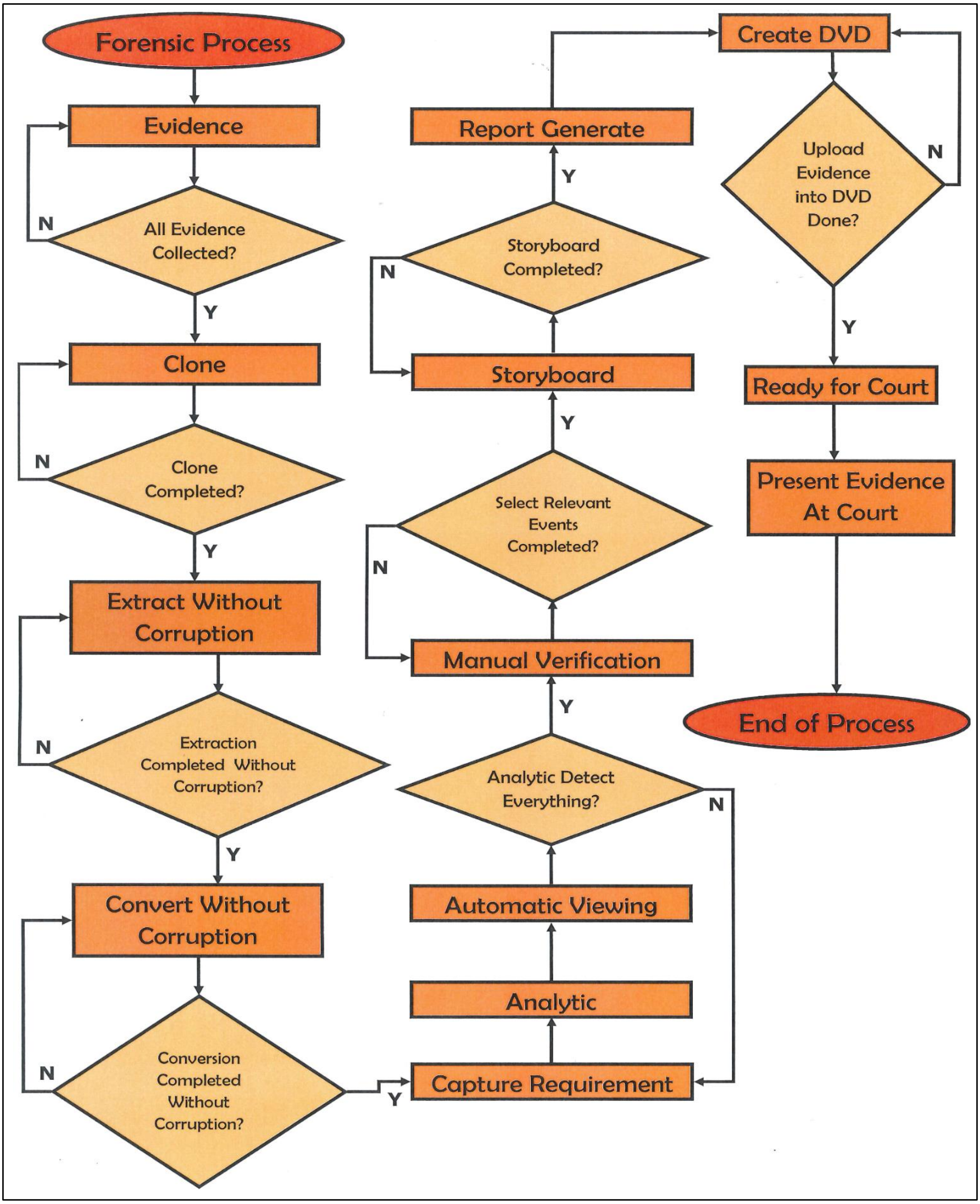


Figure 3-3: Summary for Forensic Flowchart

3.4.1 Evidence collection

Devices that contain video footages should be identified and acquired from all the relevant sources. Digital Video Recorders (DVRs) and Network Video Recorders (NVRs) serve as a source of input evidence. For this input to be used as legal evidence, which is the focus of this research, it is important that the recording device include embedded proof of authentication. Recordings such as DVR4C cannot be manipulated/altered without being noticed, and as such can guarantee the authenticity or integrity of the recording. Most DVRs can perform the above tasks successfully. The evidence is very important in crime investigation since, without it, it is difficult to make a conclusive verdict in the court of law. In the context of this research, the devices mentioned are used to capture footages of the crime scene in which relevant recording called digital evidence is collected. Digital evidence is, therefore, a collection of very useful details or sources closely related to a certain event of interest.

3.4.2 Hard Disk Drive (HDD) cloning

Cloning the HDD means making an exact copy of the evidence collected in step 1 and saving to other storage media. This is done to avoid issues of manipulating the original evidence on the original HDD. The original HDD can be cloned as many times as requiring without manipulating its contents. All experiments must be performed on the cloned HDD. The process of cloning is an important one in forensic investigation to maintain the integrity of the collected evidence. A forensic investigator is required to use the original source footage to clone the replica of the original evidence. The tools or equipment used in this process has to be the same for both cloned and original evidence. This is to ensure there is no mismatch or problem of device compatibility during analysis experiment process and recovery from HDD. Cloning the original evidence is highly encouraged, as supported by ACPO Guideline, to keep the source secure. When using the cloned HDD recording, to analyse crime situation, or experiment, it can be accidentally altered, damaged or deleted, meaning that the cloned evidence is destroyed, but the original evidence can be cloned again to continue the investigation. Although cloning can be done by just a simple copy and paste method of video evidence from one location to another, Kalker (2001) suggests a more technical approach to successful cloning. This is because video files are usually larger file sizes which can take a longer time to be copied.

3.4.3 Video extraction

Retrieving the required evidence from the cloned HDD is carried out in this step. This is done to have access to the content on the cloned HDD. DVRs and NVRs can archive videos to a USB flash drive, external Hard Disk Drive (HDD), or other storage devices. This recording is usually in a digital format. Archiving the video and audio must be done consistently. It is also important that appropriate software for the video file format be present before extracting the video evidence. This is to avoid corrupting or losing the file in the process of trying to open in a different format. Because the video footage is by nature large file which takes a longer time to download, cloned or extracted from the HDD.

3.4.4 Video conversion

The video and audio must be converted to the right format for both viewing and analysis. Most of the DVRs and NVRs store videos in their proprietary formats. There is every needs to convert the collected evidence from the DVR/NVR proprietary format to a standard format to match the requirement of the video analytic tool. The video conversion helps the investigator to easily perform the analysis process.

3.4.5 Requirements capturing

Police officers define the events to be detected from the video footages, by specifying the area(s) as well as events that they want the video analytic tool to consider. For example, a suspect wearing red clothes and a green hat, or a criminal walking or running in a particular direction or hanging around for the video length of time.

3.4.6 Automated video analysis

A careful analysis of the evidence is important. This step automates the investigation process. A video analytic tool can generate a report detailing the start and end of each event. The list of events that happened per day, or per week, or per month, or even per year will be properly reported. Failure to properly analyse the evidence entails a miss of the target for the evidence sought for.

3.4.7 Report Generation

Through the start and end time, the list of events that will be detected would be generated for report creation. For instance, when a suspicious appears on CCTV footage, the recording of their events in the region commences, the recording stops as long as the suspicious action ends. The law enforcement agent would be asked by the investigator to elaborate the work and reveal the recording with its analysis on the required region ;the generated report includes records of each software and camera where the events have occurred. By using multiple cameras which have held in different positions with same events, therefore to show them together at the same time to ensure the evidence is not missing rather than having one camera in a part of the area. Once the police investigator has checked all detected events, the he/she would pick which ones he/she is interested in and the most relevant to the case at hand. Essentially, event lists would be generated by the system to be reported e.g. Camera 2 and event one from start time to end time.

Finally, the report generation software generates video clips for each event identified by the police investigator from the start time to the end time.

3.4.8 Manual verification

This is the identification of occurrences that are directly connected with the crime and have been chosen by the crime investigator. It is important for a crime investigator to carefully examine the list of video footages that have been selected through analysis and establish the footages that are suitable for case hearing and appropriate for presentation in support as evidence. In order not to take laws into their hands, the police officers manually verify and select relevant events from the detected events by the video analytic tool. Irrelevant events are not considered after careful examination by police experts.

3.4.9 Building a storyboard

An investigation storyboard gives a sequence of events, typically with some directions and dialogue, representing the patterns of events. The storyboard is presented in court by the investigator, and so it is an important aspect of crime scene investigation. It required creating a video combination and separated with tagged-in details on each video evidence. This enables the court to follow through each step of the certain crime scene in a sequential and timely manner to give a better understanding which makes it easier for the verdict delivery.

3.4.10 Report generation

A report of the results of the analysis should be written down. This report should include issues like actions taken, why such actions were taken, findings made from the actions taken and recommendations for improvements to policies, guidelines and other aspects of forensic process amongst other issues.

3.4.11 DVD Creation

Immediately the storyboard finished with information and video clips, then all related video evidence of the suspect scene on to a DVD would be uploaded by the investigator.

3.4.12 Court presentation

The evidence in a DVD must be provided to the court. This helps the court to be sure that forensic expert and police officers are not formulating stories. The court has the right to get a clone of the DVD and give it to another expert to double check the evidences.

3.3 The very clever quality of the proposed framework

In this section, the proposed framework is explained by using the word SMART to explicitly discuss its quality. This is to expand the qualities and possible potential of the proposed framework in addition to those already highlighted in the sections before. In the next sub-sections, the meaning of each word of SMART is used to formulate each quality offered by the proposed framework, after taking insights from some of the questionnaire responses.

Why should new framework be S (Simple)?

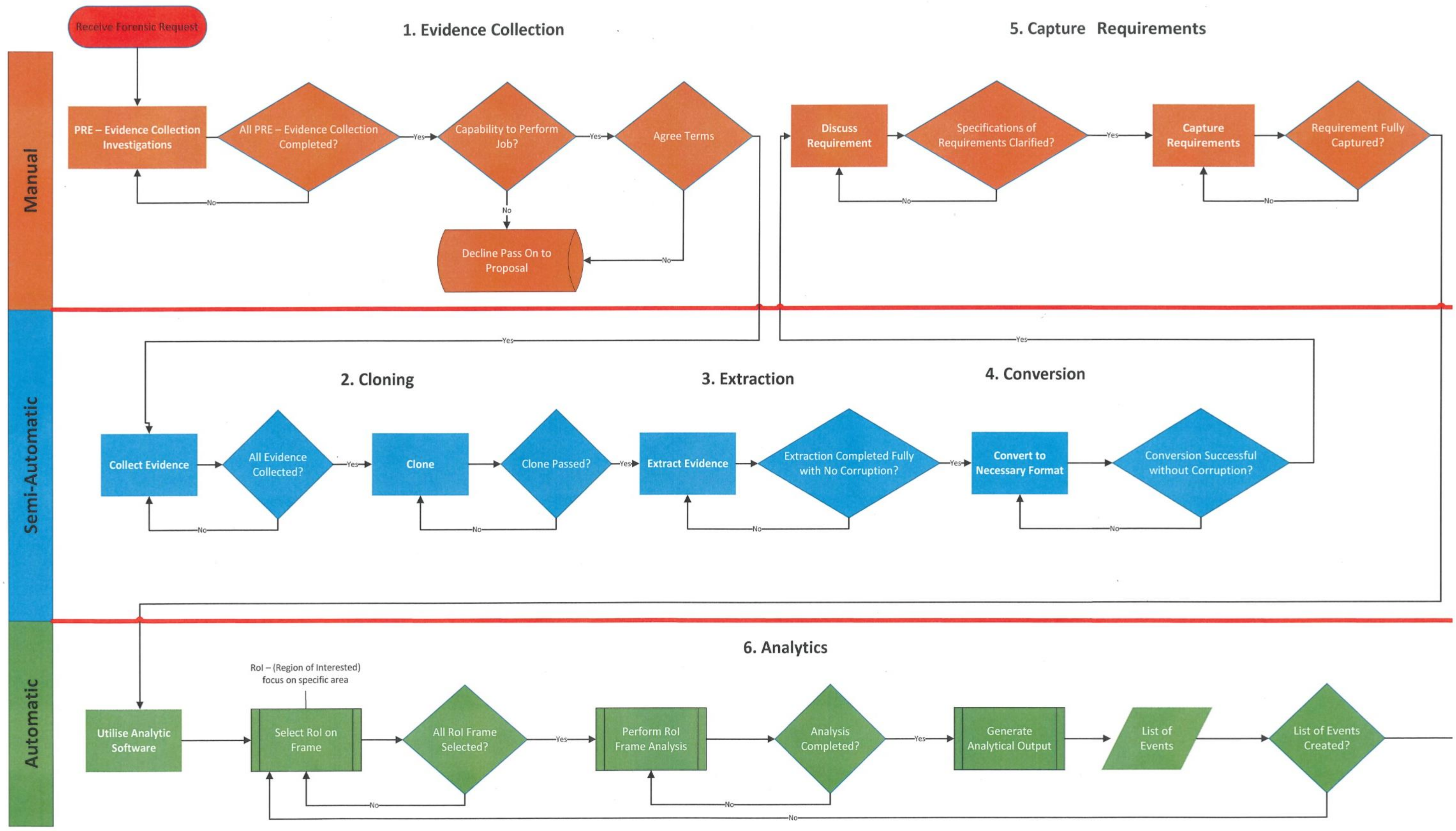
The first letter of the SMART quality is ‘S’ which implies ’Simple’. The framework that is fit for purpose should be simple to apply and easy to manage the process to use this framework. The idea of ‘Simple’ framework is to present that which is not complex and cumbersome to learn and utilize.

Why should new framework be M (Meaningful)?

The letter's’ M is the next which means ‘Meaningful’. The sources and data evidence collected and the way they are being represented and stored must be meaningful in support of the specified crime investigation.

Why should new framework be A (Accurate)?

The letter ‘A’ stands for ‘Accurate’ in the SMART quality is used to describe the framework quality. In forensic evidence and crime investigation, data accuracy is very crucial for the collected evidence to be acceptable to further help in the crime case investigation. The proposed framework must make sure that all procedures and processes are accurately documented without errors. Accurate recording of all evidence would assist in the case where an individual within the forensic team leaves the job,and so someone new can carry without loss of information.



Manual
Semi-Automatic
Automatic
Receive Forensic Request
1. Evidence Collection
PRE – Evidence Collection Investigations
All PRE – Evidence Collection Completed?
Capability to Perform Job?
Agree Terms
Decline Pass On to Proposal
5. Capture Requirements
Discuss Requirement
Specifications of Requirements Clarified?
Capture Requirements
Requirement Fully Captured?
2. Cloning
3. Extraction
4. Conversion
Collect Evidence
All Evidence Collected?
Clone
Clone Passed?
Extract Evidence
Extraction Completed Fully with No Corruption?
Convert to Necessary Format
Conversion Successful without Corruption?
6. Analytics
RoI – (Region of Interested) focus on specific area
Utilise Analytic Software
Select RoI on Frame
All RoI Frame Selected?
Perform RoI Frame Analysis
Analysis Completed?
Generate Analytical Output
List of Events
List of Events Created?
Yes
No

Figure 3-4(a): Detailed Workflow for the Proposed Framework.

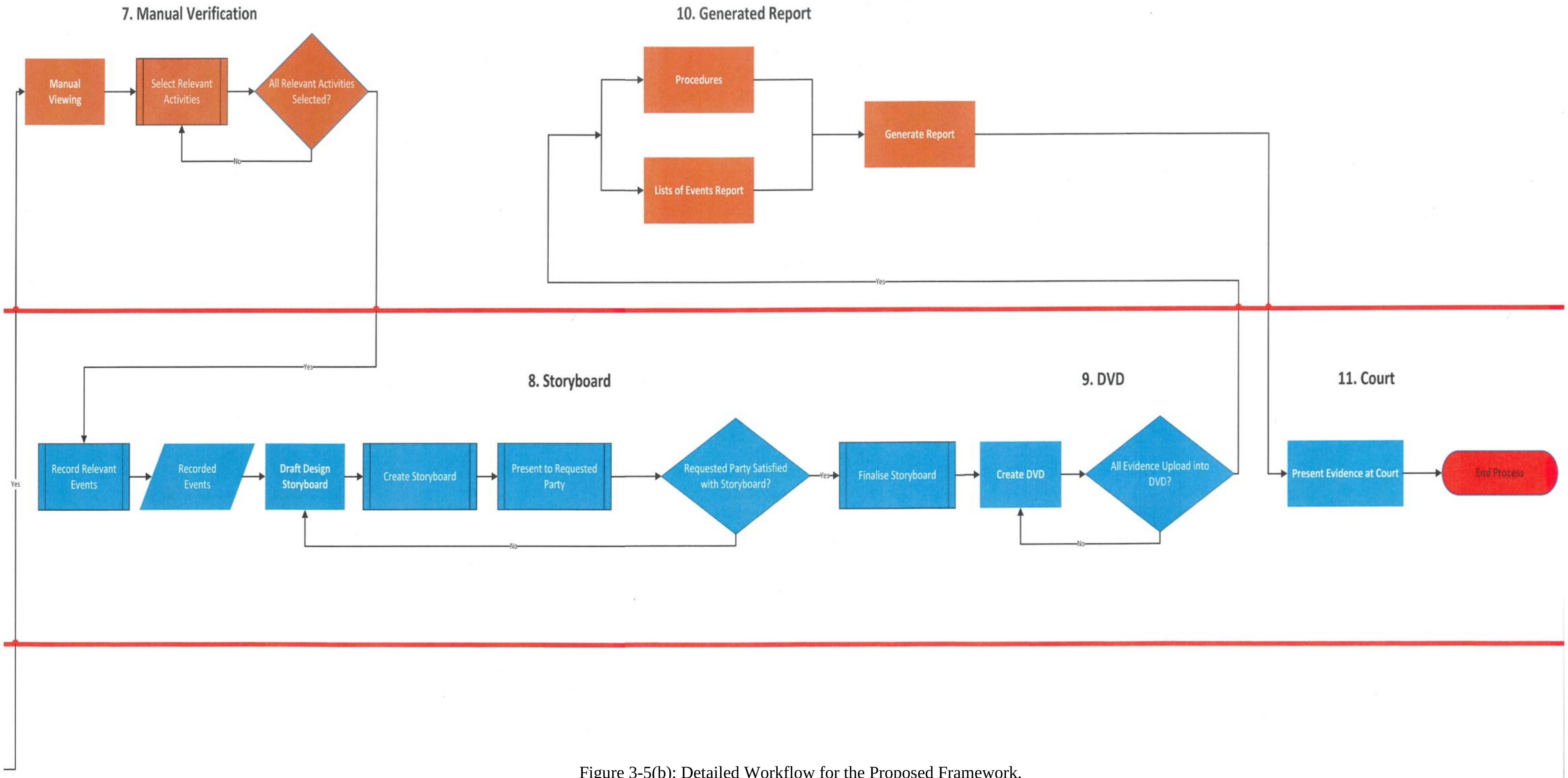


Figure 3-5(b): Detailed Workflow for the Proposed Framework.

Why should new framework be R (Reliable)?

The fourth letter of the framework quality is 'R' which refers to 'Reliable'. The framework should be dependable and consistent to make sure it is applied correctly for the expected functions. It must also be stable to avoid changes being made without reasons. Some questionnaire respondents mentioned that for the forensic evidence to be acceptable by forensic experts and court of justice, it is important that evidence be reputable and reliable. It must be convincing and trustworthy without discrepancies. The risk of this evidence been altered can decrease the level of trust as some personnel handling the evidence can carelessly misrepresent it and therefore weaken the potency of it.

Why should new framework be T (Tenacious)?

Lastly, the 'T' which means 'Tenacious', which is used to describe the quality a successful framework should possess. The new proposed framework should be firm, robust and persistently applicable, surpassing the already existing ones regarding many questions it can answer and many problems it can solve. Based on the questionnaire respondents, semi-automatic forensic investigation framework should be robust regarding the way data are handled and be free of errors

3.5 Conclusions

Forensic investigators can easily be faced with hundreds, if not thousands of hours of video to review. This situation is made more complex by the likelihood that viewer exhaustion means that notable events could be missed. Automatic analysis of videos, video analytic helps to solve this problem. However, the use of video analytic tools for post-incident analysis dictates the application of well-established digital forensics rules into the new video forensic framework.

In this Chapter, a new semi-automated post-incident investigation framework was proposed that could assist in using the output generated from a video analytic tool as evidence in the court of law. Even though the area of video forensic is attracting great attention, researchers should be awakened the need to relate this area to evidence analysis and presentation to the court of law. This will assist the development of new video forensic methods and trigger the improvement of existing ones. The conclusion is drawn from three perspectives of this chapter; the related work, the questionnaire, and the proposed framework. From the related works, the various academic researches about video or digital forensic framework have been extensive, with many frameworks been proposed and some modified. The idea of proposing a new framework is, however, to respond to the identified gap in the existing literature were little, or no attention has been given to the highlighted problems of forensic evidence in this study. The current related literature draws attention to the need and importance of creating a more detailed and digital evidence-oriented framework that would be extensive through pre-event video analysis tool. From the questionnaire's point of view; the outcome of the data collected through questionnaires has been significant in identifying the issues surrounding the existing digital forensics frameworks and their application by forensic investigators. The questionnaire posed a strategic measure through which the proposed framework has been developed. It was used to explore and understand the requirements and functionalities as well as device a new framework that befit the highlighted forensic evidence issues.

The proposed framework is presented to emphasis the ten procedure steps which collectively enhance the purpose of creating a framework that generates analyses and presents digital evidence that is admissible in the court of law. The newly proposed framework in this study is considered SMART, simple, meaningful, accurate, reliable and tenacious, the qualities which are deemed essential from legal perception.

In the next chapter, the movement classification is discussed, which includes the overview, different classification techniques and their analysis.

Chapter 4

Movement Classification-Overview

4.1 Introduction

The major challenge of video forensics is the sheer volume of information that requires to be analysed. It can take an investigator for about 7 to 8 hours to review 24 hours of recorded videos, to find specific incident. A video analytic system can automatically extract events of interest from the 24 hours of recorded video in 25-45 minutes (Adams and Ferryman, 2015). The operator can then scan the list of identified events and decide which to keep as evidence. Manual review of video data is not only expensive but is also highly unproductive and delays responses to events. By automating the analysis of recorded videos, video analytic significantly reduces the time and cost of video forensics.

In this chapter, the emerging field of video analytic is introduced, as well as its capability in fulfilling the requirements of video forensic. This chapter will look into the use of video analytic technologies, which have gained much attention especially in the context of the security of the community. The purpose of the intelligent visual surveillance is also discussed herein with much focus on the challenges that are posed when it comes to handling different behaviours. Another critical aspect that is considered and discussed herein includes the real environment as detailed in surveillance regarding the area that is covered. Video analytic is very important due to the fact that a lot of devices capable of producing videos ranging from handheld phones to video cameras. As the world is turning digitised, human beings are becoming less sensitive even to the videos they upload on social media such as Facebook, Instagram, etc. Police officers, most of times, manually analyse videos from surveillance cameras to detect anomalies (AlShaikh and Sedky, 2015). The manual analysis of such videos may pose challenges such as missing a very important event by the investigator because he/she gets bored or being distracted. The process of investigating video footages usually involves much manpower. At each point in time, at least, one investigator is supposed to physically view such video footage on a screen and manually analysing it if the need arises. This is a time-consuming operation, and is very costly, because, the staff involved in the investigation needs to be paid for the services rendered. Also, human beings are easily distracted and are sometimes inefficient.

Considering an investigator who is manually viewing one screen showing two video footages, after 10 minutes, the investigator may miss 45% of the events, and after 22 minutes, the investigator may miss 95% of the events (Dee and Velastin, 2008). The big challenge comes in when during the mess that occurs due to the inefficiency of the investigator leads to an event that goes unnoticed. Thus, there exists a need for proper consideration of research in the direction of automating video forensic investigation and the development of a proper framework to handle such sensitive issues. The use of video analytic tools for post-incident analysis dictates the application of well-established digital forensics rules into the new video forensic framework. A video analytic system consists of many modules; e.g. change/object detection, object classification, object tracking and movement classification. One key module is the movement classification module. In this module, the movements of detecting objects are recorded and compared to infer anomaly. State-of-the-art movement classifications rely mainly on rule-based classification techniques, where the abnormalities in the video are traced and reported to the user (Zhang et al., 2015). Such techniques attempt to learn normal movements to identify abnormal movements.

The adaptive multi-modal background subtraction method will be discussed in this chapter as well. The method that is used to deal with the slow changes in illumination is discussed as well as Recurrent Motion Images (RMI) which is associated with the detection of repeated motion and thus assisting with object detection and classification. Classification is also discussed here in based on how it could be done. In this context, it is determined that it can be done using texture-based, motion-based and shape-based features. The techniques that are used in this case are also discussed in this chapter.

Movement classification is also discussed in the chapter with much focus on the problem at hand. The objectives of the outdoor surveillance technologies are also discussed in this chapter. It is indicated that one of the main objectives of these technologies is the detection and capturing of the objects that move in the field of view of CCTV camera (Yaseen, Anjum, Rana and Hill, 2018). Notably, this technology is being used by many agencies in the community including the security organs that monitor the national security. More details about video analytic systems will also be discussed in this chapter.

Artificial Intelligence (AI) and Neural Networks (NN) have been heavily used to solve the problem of movement classification. However, they still suffer from various limitations such as their limited scope of operations. In the attempt of mimicking the function of a human brain, learning models inspired by the neocortex has been proposed which offer better understating of how our brains function. Recently, new bio-inspired learning techniques have been proposed and have shown evidence of superior performance over traditional techniques. In this regard, Cortical Learning Algorithms (CLA) inspired from the neocortex are more favoured. The CLA processes streams of information, classify them, learning to spot the differences, and using time-based patterns to make predictions (Agrawal and Franklin, 2014). In humans, these capabilities are largely performed by the neocortex. Hierarchical Temporal Memory (HTM) is a technology modelled on how the neocortex performs these functions. HTM offers the promise of building machines that approach or exceed the human level performance for many cognitive tasks (Numenta, 2011).

Under movement classification, the challenges of outdoor surveillance will also be discussed. These challenges are mainly faced by the implementers and the system designers. There is also the challenge of the varying appearance of the objects that are being monitored by the system. Some of the challenges include the issue of the sensor resolution which is considered to be finite. In the same context, some solutions to these issues are discussed such as the deployment of multiple cameras. It is also indicated that it is important to record the movement of the objects that are recorded by the system.

Related literature on the movement classification techniques is also discussed in this chapter. For example, the CCTV are discussed with findings from various studies on the same being detailed in next section. However, more focus is accorded to movement classification. Modelling scene behaviours are also detailed here in with focus being given to how the classification of the normal behaviour and abnormal behaviour is done. Generally, scene behaviour is discussed in detail. Artificial Intelligence is also discussed in this chapter with attention on the automation of activities that are associated with human thinking. Different applications of AI are included as well. Neural networks and Cortical Learning Algorithms are also discussed with much focus on their relation to movement classification. For instance, the Cortical Learning Algorithm (CLA) is discussed in the context of how it is applicable. Generally, this chapter will seek to determine the challenges and application of the movement classification and what can be done to make things better through the realization of the challenges that are being experienced in this context.

4.2 Video analytic for post-incident analysis

As a discipline, vide forensics demands specially trained personnel, as well as very supportive management. It also requires that the process be funded adequately to keep the unit operational. To accomplish this, it is necessary to have adequate training for the personnel that handle such data. Video recovery techniques are critical since they determine the end product of the forensics process. There is a need for efficient analytic techniques.

The use of video analytic technologies has gained much attention in the research community and the global security around the world (Popoola and Wang, 2012). The purpose of intelligent visual surveillance in most cases is to detect, recognise, or learn interesting events that seem to constitute some challenge to the community or area of the target (Lavee and Thuraisingham, 2007). These challenges posed by defining and classifying events as unusual behaviour (Hara et al., 2002), abnormal behaviour (Lee et al., 2006), anomaly (Feng and Weinong, 2006) or irregular behaviour (Zhang and Liu, 2007).

When considering the real environment and trying to relate the way objects interact in surveillance covered area, it is not so easy interpreting every activity correctly.

However, much effort has been made in respect to smart video surveillance server. Video processing and computer vision technologies are usually carried out by the smart surveillance server to automatically segment moving objects (blobs), localise the segmented blobs, classify them, identify them, track their positions, and automatically classify their movements (Sedky et al., 2005). This process is based on what the user requests and these activities are inserted into an active scene, as

Figure 4-1 shows. The majority of automatic video analysis methods are based on background analysis that aims at segmenting moving objects by distinguishing between foreground and background areas in video sequences, based on building and updating a background model (Brutzer et al., 2011).

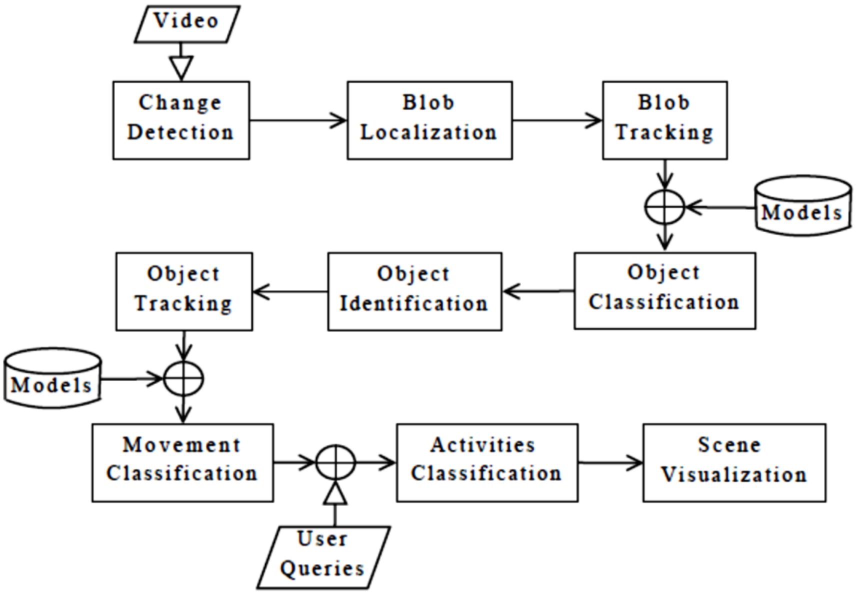


Figure 4-1: The internal structure of smart video surveillance serves (Sedky et al., 2005)

Cluttered environments that contain so many moving objects pose a challenge for many anomaly-detection algorithms. However, in real life cases, these are the kind of scenarios we meet when considering movement classification in video surveillance (Popoola and Wang, 2012).

There have been several proposed methods to tackle issues relating to surveillance cameras/videos to include the work carried out by Wren et al., (1997) which is targeted at locating interesting bodies by using uni-modal background model; the overall diagram is presented as shown in Table 4-1.

An adaptive multi-modal background subtraction method was proposed by Bloisi, Pennisi and Iocchi (2017) which handles slow changes in illumination, repeated motion from background clutter and long-term scenes. Temporal Templates are important when considering the movement classification problem. Researchers at Liu, Ma and Fu (2017) proposed the motion history and motion energy which includes different forms of temporal templates. There have been several researchers studying Recurrent Motion Images (RMIs) which could detect repeated motion, and hence it has been successful in object detection and classification (Alshaikh and Sedky, 2016). Shilpa and Sunitha (2016) recently carried out exhaustive literature concerning surveillance videos. They pointed out that, classification could be done using shape-based, texture-based or motion-based features. They also noted that shape-based methods are moderately accurate, computationally low and operate as a simple pattern-matching approach that can be applied with appropriate templates. This kind of technique, however, does not work well in dynamic situations and finds it difficult locating internal movements. Considering motion-based techniques, the accuracy is also moderate, computationally high, and does not require predefined pattern templates (St-Charles, Bilodeau and Bergevin, 2015). The challenges with these techniques are that they find it difficult identifying a non-moving human. When considering texture-based methods, they noted that the accuracy is high, computation, and these kinds of techniques provide improved quality, even though there is an additional computational time (Bloisi, Pennisi and Iocchi, 2017).

The different researches that have been conducted on the subject matter have shed light on the issue at hand. The different strengths and weaknesses have been improved upon with time. Some of the approaches that have been researched on include Pfinder which has been used successfully in different applications such as the wireless interfaces, video databases, and low-bandwidth.

Table 4-1: Video Analytic and Post Incidence Solution Development

	Researcher(s)	Strengths	Weaknesses
1	Wren et al. (1997)	This approach, called Pfinder, has been successfully used in a wide range of applications such as wireless interfaces, video databases, and low-bandwidth coding. The system uses a multiclass statistical model of colour and shape to obtain a 2D representation of the head and hands in a wide range of viewing conditions. Pfinder has the ability to compensate for small, or gradual changes in the scene	Pfinder expects the scene to be significantly less dynamic, and it cannot compensate for large, sudden changes in the scene. Another limitation, related to the dynamic scene problem, is that the system expects a single user to appear in the camera field of view at one time.
2	Ivanov and Bobick, (1999)	The main contributions of this Solution are the ability to extend the parsing algorithm to handle multi-agent interactions or concurrent events within a single parser, efficiently increment parsing scheme; This framework can demonstrate the results on real surveillance data.	The inability of this solution to address more accurate modelling of the environmental issues.
3	Sebastian et al. (2010)	In this method identification of the main challenges of background subtraction in the field of video surveillance was achieved.	In this work some aspects of background subtraction methods like time and space complexity as well as the number of parameters to be tweaked for a particular sequence were not discussed in detail.
4	Hara et al. (2002)	Fulfil the remaining requirement such as technology neutrality and wide user community applicability	Focused on traditional main stream computer and network forensic in which hard disks are primary digital devices are analysed additional work is needed

			across various layers of abstraction.
	Researcher(s)	**Strengths**	**Weaknesses**
5	Xu (2003)	The architecture and mechanisms proposed in this research are designed to solve the security challenges involving control system design, configuration, operation, monitoring, and maintenance were achieved.	The trust mechanisms between different entities (clients, repositories, smart devices) are also under investigation, and detailed security analysis of the effectiveness is currently under development to this research group.
6	Tu, et al. (2007)	A crowd segmentation method enabling the tracking of individuals through dense environments such as retail and mass transit sites are discussed.	A major challenge is to achieve a state where algorithms degrade gracefully and do not become overwhelmed by circumstances not anticipated by system developers.
7	Tian, et al. (2008)	The IBM smart surveillance system (S3) is one of the few advanced surveillance systems which provides not only the capability to automatically monitor a scene but also the capability to manage the surveillance data, perform event-based retrieval, receive real-time event alerts.	The S3 system has been deployed and tested in limited application environments.
8	Baudry et al. (2009)	This work proposes a framework for the forensic analysis of video content, for example, those illegal copies of feature films proliferating in peer-to-peer networks or on bootleg DVDs. Compact and discriminate feature vectors, so-called “fingerprints”.	The need to enhance the temporal and spatial fingerprints to increase the registration accuracy without degrading copy detection rate.
9	Màrmol and Sevillano (2016)	The Quick Spot which is a video analytics solution was developed to help with spot detection when street parking. This was a	There are times when the system fails though not many times. The computational complexity of the

		successful project	solutions is also quite high.
10	Zouaoui et al. (2015)	Fusion algorithm that functions based on the Dempster-Shafer's theory is used to analyse asynchronous detection outputs as well as compute the degree of belief for the events.	The data on violent events that were to be used in the study was not easy to come by. As such alternatives were sought which could have compromised the findings
11	Grobler, et al. (2010)	The purpose of this theoretical framework is to provide management with a holistic view of what to consider when preparing the organisation for forensic investigations.	There should be an integrated management model to enable management to demonstrate good corporate governance, ensure that policies, processes, procedures, trained staff, and technology are available to successfully investigate incidents and to have CDE available when required.
12	Cantrell, et al. (2012)	This work proposes the use of pre-examination techniques commonly referred to as digital triage. Pre-examination techniques can assist the examiner with intelligence that can be used to priorities and lead the examination process	The computer profiling stage is probably the area of this model that will be the most difficult to develop.
13	Karpathy et al. (2014)	This work presents an extensive empirical evaluation of conventional neural network (CNN) on large-scale video classification using a new dataset of 1 million YouTube videos belonging to 487 classes.	The system's performance is not particularly sensitive to the architectural details of the connectivity in time.
14	Yeh et al., (2017)	The Pfinder model is discussed in the paper together with the Gaussian probability density function.	It is noted that even those systems that are determined to be advanced also suffer from cavity issues. These include the false detection as well as some deficiencies.

15	Verma et al., (2018)	The article details a project taken up for the development of surveillance system that will record people real-time. Matlab is used to develop the system. Motor rotation is taken into consideration as a critical aspect of surveillance and motion detection.	It is determined that the system not be able to follow objects as they make their way through certain paths. There is no definite and object-oriented policy on how the system is to establish contact and maintain contact with specific objects.
16	Khanam and Deb, 2017	The paper proposes a specific luggage detection and classification framework. The possibility of occlusion taking place in the environment is also mitigated against. It is recommended that a strong classifier be generated through the boosting of the Support Vector Machine (SVM).	The presence of similar false objects in several places along the path of an object serves to be a major deterrent to proper and effective surveillance.
17	Divya, Shalini, Deepa and Reddy, 2017	Advanced Motion Detection (AMD) is noted to be a strategic solution in video surveillance. It is recommended as a good option when it comes to the detection of moving objects. The AMD algorithm also facilitates background subtraction. The system also makes it easy to examine the background based on the simplified computational complexity.	The system does not work in offline situations. The system has to receive instructions on whom to track. As such, the system is not totally automated.

On the same note, there are some weaknesses in the approach which is an assumption that the scene as to be less dynamic. The other weakness on the same note is the expectation by the system that it is only one individual that is to be in front of the camera at a time.

The other positive aspects of the different reviewed solutions are quite comprehensive and serve the purpose. Some of these solutions help to deal with the issue of background subtraction. They also help to deal with the issues of technology neutrality. However, it has also come out clearly that the background subtraction has to put into consideration the time and space complexity. It would also be important to consider the parameters that need to be tweaked for a particular sequence.

Solutions to video forensic analysis could come in handy in the fight against illegal videos selling and distribution. Through such solutions, it would be possible to determine the fingerprints in such videos. However, it is important to test the provided solutions in different scenarios to determine just how effective the solutions are. It is also clear that when it comes to forensic analysis of a view is being undertaken it would also be important to enhance the temporal and spatial fingerprint as well. Overall, the solution presented was substantially effective though in need of some more attention.

Advanced motion detection has been determined as a strategic method of detecting moving objects. In this case, the surveillance is facilitated using an AMD algorithm. The algorithm starts to track an individual or an object once the system’s user has specified the suspicious object or person. The system has been determined to be efficient. However, there is the issue of having to receive instructions from the user on what the system needs to track. This is indicative that the system is not totally automated.

The AMD algorithm makes use of the background subtraction method with the intention of creating a reliable background model that will facilitate better object tracking. Once more occlusion seems to be a major issue that must be dealt with in surveillance. Solutions to the issue include the development of a strong classifier that is generated through the boosting of the Support Vector Machine (SVM).

4.3 Movement classification - Problem definition

In recent years, the development of outdoor surveillance technologies has captured the interest of both researchers and practitioners across the globe. The objective of these technologies is to detect the presence of objects that are moving in the field of view of a CCTV camera(s) for national security, traffic monitoring in big cities, homes, banks and market safety applications (Lambie et al., 2017). However, detecting, classifying and analysing the movements of objects were traditionally a manual job performed by humans in which the guaranty of absolute attention over time by a human on duty remains small, especially in practical scenarios.

In general, in an attempt to develop a video analytic system that can detect and classify the presence of objects moving in its field of view, that system must be able to:

a) Classify detected objects into various categories.

b) Track the detected objects over time.

c) Classify their movements.

It is also interesting to note that, each of the above tasks poses its challenges in term of design and implementation.

4.3.1 Object Tracking and Detection

There are different methods that have been proposed as the best for object tracking and detection as the video is being recorded. In this context, it is possible to cluster together the edge and the corner features. These are clustered together to form an object. These are then tracked together. There are other alternatives to track where the snake contours can be utilized. In such cases, the contours can be used to detect the outline of an object. They can also be used to determine the track that is taken by the object as it crosses frames (Wang et al., 2017). Variations of the optical flow, as well as blob trackers, can be used in certain situations. These are applied where the background model and the foreground have been learned. In such cases, the background subtraction techniques that are mentioned are region or background based.

In surveillance, there is a great need to detect suspicious objects. Such objects include luggage that has been left where it should not like e.g.at an airport. Due to this need, several systems have been developed with this capability. These systems are engineered to detect objects that have been resting in a given location for a given time (Wang et al., 2017). The system, in this case, considers the background as it is in the video. Histogram methods are utilized in such cases. The methods measure to determine whether there have been changes to the background which indicate that a new object has been introduced into the scene and whether it is static.

In surveillance, there is a rule that an object should be tracked for a long period. The object should also be tracked through various conditions. In such conditions arises the challenges that are faced when tracking an object. Such challenges include challenges to do with lighting where an object must be tracked even during the illumination changes. Challenges also arise where an object must be tracked through a dynamic and cluttered background. Shadows do not make things any easier (Wang et al., 2017).

Due to the demanding nature of the process of tracking objects, different methods have been utilized with the effort to make it easier. For instance, the *prior* model is utilized at times. The prior involves the development of a model object that has a close resemblance to the object that is being tracked. For example, a model of a human being can be used as well as a model of a car. These models provide the surveillance system with the average dimensions and shapes associated with the particular objects that are being tracked. Armed with this information, it is easier for the system to track the objects without being distracted. (Zhang, Zhang & Chen, 2017)

In other instances, objects can be tracked using different methods. For instance, human beings can be tracked based on the models of the learned textural appearance. Contour tracking can also be employed. There is a cardboard model of the human shape that can be used when tracking people. This model is considered to be strong. This method can be used where the person is going through partial occlusion. It can also be applied when they are in groups. This method can also be used to determine whether the individual that is being tracked has picked up any object. This is a critical aspect of object tracking. The trackers should adopt to different conditions and situations.

4.3.2 Object Classification

Object classification has been a major issue when it comes to surveillance. It has been a major issue for computer vision as well. There are various solutions to these challenges. Often, the footage recorded and stored during a surveillance process has poor resolution (Anjum et al., 2016). In such cases, the objects that are being tracked might have very few pixels in the frames. In such a case, there will most likely be lack of information. As such, the course colour histogram methods can be applied in each frame (Anjum et al., 2016). Also, a strategic solution to the issue would be to track the video over a long time to create a model for the motion of the object being tracked.

VSAM employs two algorithms used for classification. Notably, both algorithms have to employ some training. These algorithms are the neural network and the Linear Discriminant Analysis (LDA).

The neural network is trained on area and blob shapes. The algorithm can discriminate between human groups, clutter, vehicles and humans. The second algorithm can run on vectors of up to 11 dimensions. These dimensions include blob height, position, and width and imagery features as well. In this case, the features of the image within the blob are taken into consideration. Interestingly, these algorithms are noted to have an average success rate of around 90%. However, it seems as if the LDA algorithm has better capability to discriminate compared to the neural network. The reason for this conclusion is that LDA can include more features (Wan, Wang, Guo and Wei, 2018). Nonetheless, both classifies are operated under single frames. However, the results that are acquired from previous frames are cached to smooth.

4.3.3 Movement classification challenges for outdoor surveillance

There are many hurdles faced by outdoor surveillance system designers and implementers. The first step toward automated activity detection is detection, tracking and classification of moving objects in the field of view of CCTV cameras; another challenge is that sensor resolution is finite, and it is impractical for a single camera to observe the complete area of interest. Therefore, multiple cameras need to be deployed. Also, the detected objects are context-dependent, but for a general surveillance system any independently moving object such as a vehicle, animal or a person are deemed to be interesting, but detecting and classifying these objects is a difficult problem because of the dynamic nature of object appearances and viewing conditions in practical scenarios (Alshaikh and Sedky, 2016).

Other challenges are that appearance of objects can vary considerably over time depending on the camera view, like front view or side view; the dynamic nature of lighting condition, relative distance to cameras can also affect the true appearance of an object over time. After tracking objects, it is important to record their movements over time, and the problems arise because of the need to estimate the trajectory of an object as the object moves around a scene, in modelling this kind scenario that requires where the object is in the image at each instant in time. The whole idea is to keep tracking and continuously observing an object concerning their size, motion and shape that vary over time. In a realistic environment such as shopping centres, busy street or ports, some objects are likely to change shape and size while moving. Another scenarios challenge in outdoor surveillance is the issue of occlusion, in a real-life situation, an object may be blocked by another object or structure or even a shadow of another object that leads to a discontinuity in real time observation, detection and classification of moving objects toward effective and automatic video forensic analysis. Detecting and classifying moving objects under occlusion is very difficult because of accurate positioning, shape, size and motion or velocity of an occluded object cannot be determined which could affect the result of forensic video analysis.

Robustness has been a major issue when it comes to surveillance. There is a need for the surveillance systems to be made more robust. This is associated with most of the issues to do with movement classification. For instance, to effectively monitor goods left in an airport lounge, the system should have a low false-negative rate. On the other hand, there will need to have excellent system performance. Such performance should be maintained when there is difficulty and the operating conditions are fluctuating (Verma, Zhang and Stockwell, 2017).

The variations in illumination have also been presenting problems in the system. This is a major problem when it comes to the monitoring of outdoor scenes. In such cases, the need to isolate the shadows might arise. The changes that might occur due to the changing position of the sun and the appearance or lack of the moon at night also have a great impact. Robustness is also affected by the changes that might occur in the indoor scenario. In such a case, changes might arise as a result of the changes in the light-source. Occlusion might also occur. Other factors that might be sources of challenges are a reflection, saturation, doors opening etc. (Verma, Zhang and Stockwell, 2017). The problems listed above should be avoided in commercial systems as much as possible.

4.3.4 Camera Handoff

Often, when several cameras are used to track an object, it is assumed that the field of view (FOV) will overlap. This is a serious restriction as far as surveillance is concerned. It is more severe when the area that is being monitored is a wide area. It also becomes more serious if there is occlusion of the lines of sight. When there are several FOVs between cameras, ambiguity is avoided by using the camera handoff algorithm (Zhang, Zhang & Chen, 2017). A handoff algorithm is one that is generated by moving the serving station of the camera while the cell boundary position is kept constant (Wu, 2010). The handoff algorithm helps to keep track of the objects as it is being tracked in between cameras (Akintola, 2015). Another solution to the FOV issue is the Bayesian approach. The approach is used to handle situations where the FOVs do not overlap. For the system to operate, it has to be facilitated with some allowable paths. A good example of how this is applied is in the traffic camera systems. The surveillance of traffic should be done by cameras that do not have overlapping FOV (Akintola, 2015). The reason for this notion is that traffic will usually follow well-defined routes. More so, these routes extend over long distances. It is critical to note that it is almost impossible to monitor all every inch of each road. In such a situation, it is recommended that the cameras be placed in certain intervals throughout the roads. The recommended distance between the cameras is 2 miles (Akintola, 2015). As the cars enter the zone where there is camera, they are identified based on the positional information of the car as well as the appearance of the vehicle. This information is relayed based on the views that have been made previously by other cameras where the vehicle had passed in the past.

4.3.5 Efficiency

As far as surveillance analysis is concerned, the software that is being used is critical regarding determining how efficient the process will be. Such software come in handy, especially in real-time surveillance situations. The software should be designed to take action as they are getting information from the surveillance equipment. The central processing unit (CPU) and computer memory serve a great role in terms of the efficiency of such systems (Akintola, 2015). Nonetheless, the efficiency of such systems is inhibited by network bandwidth bottleneck. Usually, when surveillance is being done, the videos are located in specific locations. Such videos would then have to be moved to the location where the software that is issused for analysis is being run from.

When the videos are being moved over the network, there is need to control the network usage to facilitate efficient movement. In this case, the activities that are being undertaken in a frame will be key in determining whether it will be allowed access to the network or not. Those frames with lots of activity are allowed most of the network resources. To accomplish this, there is need to sensor motion in cameras. This is accomplished by the Sensor Processing Units that are located in VSAM (Akintola, 2015).

4.3.6 Mining of Surveillance Videos

Movement classification is also associated with how the videos are mined. After the occurrence of an event, the video has to be mined to gather information concerning that occurrence. This is done through the analysis of the videos that have been stored that were recorded earlier on. Public cameras are designed to be ubiquitous in that as one goes about their business in a town with several cameras, it is very likely that they will be recorded by several cameras that are spread out in the town/city.

Whenever there is a security breach or occurrence that requires the security teams to consider reviewing CCTV footage, they do so manually. The security personnel get to go through the footages using their own eyes searching for any clues as to what happened. Several challenges present themselves in this case since the personnel are limited to the number of videos that they can watch at any given time (Feris et al.2015). The task is very intensive and requires many people to view some videos. In such a case, it would be better if a system was developed and utilized that could analyse the videos based on given description of the object that is being searched for. A system might also be required to search in the footage for instances where specific events took place, e.g. violence.

In case there are too many cameras that are linked to the video repository, this might present an issue when the process of video analysis is concerned. In such a case, it would not be easy to track an individual or an object through all different videos.

With a proper technique for movement classification, this issue should be easy to deal with.

4.4 Related Literature of Movement Classification Techniques

Over the years, various researchers have suggested that activity analysis in video surveillance will be the most important area of research when considering video surveillance research (Popoola and Wang 2012) and (Collins et al. 2002). Popoola and Wang (2012) further stated that even though there are many CCTV cameras to capture abnormalities or unusual behaviours, resources to monitor and analyse such captured video footage is limited. Most of the existing movement classification algorithms use a simple rule-based technique (Mathie et al. 2004). Some researches in video surveillance focused on action recognition, body parts recognition and body configuration estimation (Aggarwal 2011). There has been a recent advancement in researching about semantic descriptions of humans in challenging unconstrained environments (Poppe 2010).

The focus of this research is however on movement classification and when considering movement classification techniques, the question “what is going on in a scene” is considered (Popoola and Wang 2012). In this sense, there must be a clear definition of what is considered normal/usual and abnormal/unusual. Abnormalities are defined as actions that are fundamentally different in appearance or action done at an unusual location, at an unusual time (Varadarajan and Odobez 2009). When considering anomaly detection algorithm, detecting the spot and where anomalies occur with little to no false alarm is of great emphasis (Popoola and Wang 2012).

When considering modelling scene behaviour, statistically based methods are currently used instead of rule-based methods, which use already defined rules to classify the normal behaviour from abnormal behaviour that was previously used (Ivanov and Bobick 1999). Statistical based methods are believed to achieve a more robust solution to get useful information in behaviours of a considered scenario (Popoola and Wang 2012). This kind of method is based on either learning from normal behaviour and then using such criteria to differentiate between normal behaviour and abnormal behaviour or the process of learning and detecting normal and abnormal behaviour is done automatically. The challenge with this kind of approach is that human beings behave abnormally in different ways, and as such systems may trigger wrongly, hence a false alarm. Much work has been done on using machine learning techniques to train some algorithms on normal behaviours and abnormal behaviours so that such algorithms could effectively differentiate these two cases. Unfortunately, it is a hard task to exhaust all the abnormal behaviour to be carried out in real life scenarios (Amershi et al., 2014).

Several attempts have been in place for movement classification, using different techniques such as pattern recognition (Weiming et. al. 2004 and Iwashita et. al. 2013), artificial intelligence (Iwashita et. al. 2013, Agrawal and Mena 2013), and neural network techniques (Karpathy et. al. 2014). In the next section, some selected studies of some of these different movement classification techniques are discussed.

4.4.1 Artificial Intelligence

The whole concept of Artificial intelligence (AI) is centred on four approaches which are namely: thinking humanly, acting humanly, thinking rationally and acting rationally (Russell and Norvig 2010). Haugeland in 1985 looked at AI as an exciting new effort to make computers think, i.e. machines possessing minds in the fullest and literal sense (Haugeland 1985). Bellman looked at AI to be automation of activities that can associate with human thinking, activities such as decision-making, problem-solving, learning amongst many other activities that could be performed by the real human brain (Bellman 1978). Considering acting humanly, Kurzweill looked at AI as the art of creating machines that perform functions that require intelligence when performed by human beings (Kurzweill 1990). In the same direction, Rich and Knight in 1991 looked at AI to be the study of how to make computers do things at which, currently, people are better (Rich and Knight 1991). In respect to thinking rationally, AI is viewed as the study of mental faculties using computational methods (Charniak and McDermott 1985). In this same direction, Winston considered AI to be the study of the computations that make it possible to perceive reason and act (Winston 1992). When considering acting rationally, Poole et al. (1998) looked at AI as the study of the design of intelligent agents. In this same light, Nilsson (1998) defined AI as being concerned with intelligent behaviour in artefacts (Nilsson 1998).

AI has several applications to include:

1. Robotic vehicles: AI techniques have been deployed to successfully move a robotic car called STANLEY and won the 2005 DARPA Grand Challenges.
2. Speech recognition: It is now possible with the help of AI to be guided by automated speech recognition and dialogue management system.
3. Autonomous Planning and Scheduling: NASA’s Remote Agent program successfully carried out an autonomous planning program to control scheduling of operations for a spacecraft.
4. Game planning: Development of Atari Games (Guo et al., 2014).
5. Spam fighting: Classification techniques have been successful is separating authentic messages from spam messages, hence saving our inbox messages from spam emails.

6. Logistic planning: AI techniques have successfully been deployed in many applications to generate planning of events that would have taken several weeks or months in just minutes.

   Robotics: Robots have been deployed as vacuum cleaners for home use, handling of hazardous materials, clear explosives and identify the locations of snipers.

7. Machine Translation: With the help of AI techniques, a person can speak from one language and can be conveniently translated into another language (e.g. From Arabic to the English language) (Russell and Norvig 2010).

According to Alzou'bi et al. (2014), AI can simplybe applied to calculations in integral and differential calculus, electrical circuit theories, logical mathematics, and game playing fields. It can also act in new applications as akey technology, such as to telephone systems where it applied to recognise the speech, the banking systems, where AI is applied to detect attempted credit card fraud, and software systems where AI response to problems and present appropriate advice.

The academic resources available show that AI techniques have already had numerous applications such as in combating cybercrimes. For example, neural networks are applied to intrusion prevention and detection systems, but there are also proposals for using neural networks in “Denial of Service (DoS) detection, spam detection, computer worm detection, malware classification, zombie detection, and forensic investigations. Also added to that AI techniques such as Data Mining, Heuristics, Artificial Immune Systems (AISs) and Neural Networks have also been applied to new generation anti-virus technology (Charniak, and McDermott 1985). Nevertheless, applications of AI to movement classification remains less concern to researchers, therefore; this study is aimed at addressing it. Likewise, various attempts on movement classification by researchers in the past using neural network techniques are discussed in section 4.4.2 below.

4.4.2 Neural Networks

The idea of Neural Networks (NN) is concerned with learning the kind of behaviours that could be achieved when by attaching several neurons, therefore, it is likened to be closely related to the brain (Hawkins and Blakeslee 2004). NN had the hope that by correctly linking a bunch of neurons together will correctly improve the challenges that were not solved in AI.

Artificial Neural Networks (ANNs) has been an active research area since the 1980s and has made a huge success. Not with standing it has several challenges such as selection of the structure and the parameters of the networks, selection of the learning samples, selection of the initial values, the convergence of the learning algorithms amongst several other challenges (Ding. et al. 2013). In the 1990s, tooptimise the design and parameters of ANNs, Evolution Algorithms (EAs) came into play to achieve this purpose (Whitley 2001). However, Ding et al. (2013) noticed that both EA algorithms and ANNs are the theoretical results of applying biological principles to science research. Hence, for now, even the combination of EA algorithms and ANNs has not been able to achieve intelligent solutions close to that of a human being.

It is very clear that NN and AI techniques are very important and have several applications that cut across several important areas of life. However, Hawkins and Blakeslee (2004) noticed that in the mid-1980s, artificial intelligence systems were failing in some applications and hence people started thinking of alternative ways to solve problems and in an effective way that it will not fail. Nilsson (1998), viewed this assessment quite differently by postulating that by aiming to achieve real human-level AI could imply the swap of operations that humans carry out for pay, which could be automated thus involving the task of building a special purpose system, this lead to the interest in Neural Networks, which were an improvement over AI and was in a way built on architecture to depict real nervous systems (Hawkins and Blakeslee 2004).Furthermore the application of NN to movement classification is also an area that needs to be investigated more broadly. According to Saravanan and Sasithra (2014), the integration of neural network and artificial intelligence is significant in that, it teaches the system the execution of a task, rather than using computational programming system to perform certain tasks. The inclusion of AI in NN is to perform such tasks, which is referred to as artificial neural network (ANN). In ANN, Saravanan and Sasithra (2014) discussed that it contains numerous artificial neurons that are linked by explicit network architecture. The main objective of the neural network is to transform the inputs into substantial outputs that are fit for purpose.

Movement classification in ANN is considered one of the most dynamic studies and areas of application. However, the major drawback of applying ANN is identifying the most suitable learning, transfer function and grouping of training for data sets or images classification with increasing number of classified sets and features. The diverse functions combination and its effect while applying ANN as classification technique is examined in Saravanan and Sasithra (2014) study and the accuracy of functions have been analysed using different datasets and image inputs. Considering different application datasets and images, various ANN integrated techniques have been proposed. Elazary and Itti(2010) proposed a technique that was applied to the classification of IRS-1D satellite images. Also, the study suggested the fitness of Back Propagation Neural Network (BPNN) for remote sensing image classification based on three steps. The first step measures the first order histogram with feature extraction. The next step, classification based on BPNN is done, while the final step compares the results with the maximum likelihood classification (MLC) technique. The statistical features in this study are based on the first-order distribution measure like the standard-deviation, mean, kurtosis, energy, skewness, and entropy. The network comprises three layers. The input layer is fed with extracted features which contains 18 neurons in the classification of IRS-1D satellite images, six classes were used, and the back propagation neural network was trained on these classes. The whole image was classified using this trained network. ANN is used by the IRS-1D satellite images classification of IRS data to select a suitable training method. There are various training algorithms for feed forward networks. The gradient of the performance function is used by all the algorithms to find out how to fiddle with the weights to decrease the performance. The back-propagation technique determines the gradient. This gradient performs computational backwards through the network. An innovative classification technique was devised in Gijsberts et al. (2014) for neck movement patterns using a flexible BPNN. The major aim is the estimation of the prognostic ability of a BPNN using neck movement variables as input. In this study of the selected test case of a collection of three-dimensional (3D) neck movement data from 59 samples were subjected to testing proposed system. The result of the BPNN in an actual calculation for 84% of the control subjects, revealing the suitability of a BPNN for motion characteristic predictions. ANN was used in the study of Wang and Farid (2006) that proposed remote sensing images classification. The ANN is referred to as a significant part of AI which has been used extremely in the study of remote sensing classification. Using a case study of a complicated sensing classification, supervised classification is carried out on the training samples. The multilayer feedforward network is discussed in Choromanska et al. (2015) as classification technique which divers from many traditional statistical classifications due to its dependency on conditions and assumptions. This pose problem in which NN is regarded as a solution since they are non-linear, self-adaptive and universal function approximators Apart from AI, NN, and ANN, Cortical Learning Algorithms are another power tool that has been attempted for classification, some of the records of this are discussed in the next section.

4.4.3 Cortical Learning Algorithms

The Cortical Learning Algorithm (CLA) is a result of an attempt to model the complex and structural nature of neocortex and capture the algorithmic properties and characteristics of the neocortex. It is biologically proven that neocortex is the seat of intelligent thought in the human or mammalian brain. Intelligent properties such as vision, movement, hearing, touching, etc. are all performed by this intelligentsia, these cognitive tasks that are largely performed by the neocortex of humans are very difficult to design in real life scenarios.

There are many things humans find easy to do that computer are currently unable to do (Numenta, 2011). This task includes vision, hearing, touching, movement, understanding spoken language and planning capabilities that are largely performed by the human neocortex. Despite extensive research that was carried out previously only few results were achieved in modelling higher-level cognitive functions like Hierarchical Temporal Memory (HTM). Human capabilities largely depend on their behaviour that influences what they perceive. Almost all human actions change what they sense (Numenta, 2011). Sensory input and motor behaviour are intimately entwined. For decades the prevailing view was that a single region in the neocortex, the primary motor region, was where motor commands originated in the neocortex. Over time it was discovered that most or all regions in the neocortex have a motor output, even low-level sensory regions. It appears that all cortical regions integrate sensory and motor functions

Cerebral context is part of the human brain that constitutes about 85% of brain total mass and is responsible for higher level cognitive functions (Numenta, 2011). The different parts of the neocortex, whether they are responsible for vision, hearing, touch, or language, all work on the same principles (Jeff Hawkins,). The cells in a region of cortex can learn and recall sequences of patterns, which is an essential element for forming invariant representations and making predictions (Jeff Hawkins,). Cortical is derived because of the function or condition of the cerebral cortex in the human brain. In an HTM network, Sparse Distributive Representation(SDR) is used to learn a large number of spatial patterns and temporal sequences. Training data in the form of an input stream is presented to the network, and a model of the statistical structure of the training data is built. Unlike models for static pattern recognition, HTM accounts for spatial and temporal variability in the input data. It accomplishes this by learning sequences of commonly occurring input patterns in an unsupervised manner. The prior versions of the HTM algorithms differ significantly from the HTM CLA. Prior versions of the algorithm used Markov chains and Bayesian Belief Propagation. In CLA versions, novel input patterns were compared to the subset of stored input patterns, and the likelihood over the set of stored input patterns was calculated. The likelihood over the set of stored patterns became the input to the temporal learning component of a node in which a Markov graph of temporal transitions is learned by building a first-order transition matrix. The Markov graph is then partitioned to form Markov chains. The likelihood over the spatial input pattern is used to compute the single most probable Markov chain given the current evidence.

4.5 Analysis

It is very clear that AI techniques are very important and have several applications that cut across several important areas of life. However, Hawkins and Blakeslee (2004) noticed that in the mid-1980s, artificial intelligence systems were failing in some applications and hence people started thinking of alternative ways to solve problems and in an effective way that it will not fail. Nilsson (1998) viewed this assessment quite differently by postulating that by aiming to achieve real human-level AI could imply the swap of operations that humans carry out for pay, which could be automated thus involving the task of building a special purpose system, this lead to the interest in Neural networks, which were an improvement over AI and was in a way built on the architecture to depict real nervous systems (Hawkins and Blakeslee 2004).

Hawkins and Blakeslee (2004) however noticed that NN could not meet very three important criterion which the brain had. These are:

a) In real cases, brains process a rapidly changing stream of information and not the static flow of information.
b) The feedback connections which dominates most connections in the neocortex were not understood
c) Any theory that wishes to imitate the brain should take the physical structure of the brain into consideration and as such neocortex is never a simple structure (Hawkins and Blakeslee 2004).

The whole process of neural networks was concerned with a static input pattern being converted to a static output pattern which is far from what the actual brain processes. As time has evolved, NN has evolved too, but none of the techniques associated with neural networks has cared to incorporate the architecture of neocortex into NN.

AI and NN have so much focused on the fact that, intelligence lies in the behaviour that a program or neural network produces after a given input is processed. However, intelligence is not all about acting or behaving intelligently, but is also about knowing exactly what goes on in your head. Thus, understanding what goes on in one's head will greatly assist in developing machines that are more intelligent (Hawkins and Blakeslee 2004). To be able to extract the real intelligence, to build intelligent machines, it will need to extract intelligence from nature's engine of intelligence which is the neocortex (Hawkins and Blakeslee 2004). Nilsson (1998) argued for the development of general purpose educable system that can learn and be taught to perform any of the high-volume jobs that humans perform. In general Hawkins and Blakeslee (2004) noted that to build a system that behaves like the brain, there should be an intake of the stream of changing information, recognition of patterns in such a way that, there is no prior knowledge about the input source, making accurate predictions and react correctly.

This could only be achieved with the help of understanding how the neocortex works, which is the part of the brain that handles higher functions in human beings such as conscious thoughts and language processing. This approach is based on modelling the structure of the neocortex and how it works. However, approaches like AI is built upon the idea of a neural network, which in essence, NN does not behave in the same way as the brain thinks, and this is not what is considered intelligence. The HTM is different form NN in that there is no need to carry out a backpropagation since the HTM is usually being updated as the information flows for the first time. According to (Clark 2014), NNs cannot produce systems that can have intelligent behaviour; this approach is thought to be implemented using the Cortical Learning Algorithm (CLA). This approach is usually made up of six very important components which include:

1. online learning from streaming data,
2. hierarchy of memory regions,
3. sequence memory,
4. sparse distributed representations,
5. all regions are sensory and motor, and
6. Attention.

The CLA processes streams of information, classify them, learning to spot differences, and using time-based patterns to make predictions.

In humans, these capabilities are largely performed by the neocortex. Hierarchical Temporal Memory (HTM) is a technology modelled on how the neocortex performs these functions. HTM offers the promise of building machines that approach or exceed human level performance for many cognitive tasks. Almost all the most important activities carried out by mammals are controlled by the neocortex such as vision, hearing, touch, movement, language, and planning (Hawkins et al. 2010).

HTM models neurons which are arranged in columns, in layers, in regions, and in a hierarchy. HTM works by a user specifying the size of a hierarchy and what to train the system on, but how the information is stored is controlled by HTM. In general, the HTM is a hierarchical organisation and is basically time-based. The HTM consists of the region which is the main unit of memory, and it also comprises feedback connections which make the hierarchy diverges as one descends the hierarchy.

In all of these, the time has a major role as it plays a very important role in learning, inference and prediction. A CLA algorithm learns the temporal sequence from the stream of input data; even though it is difficult to predict what patterns may likely follow the next. This HTM algorithm is very important because it covers what is believed to be the building block of the neural organisation in the neocortex (Hawkins et al. 2010).

4.6 Conclusion

It has come out clearly, from this chapter, which vide analytic for post-incident analysis is quite useful for forensic purposes. Even with the intelligent surveillance that is intended to be accomplished by the surveillance equipment, there is a difference in capability when it comes to the analysis of different videos due to the environment that they were recorded at. Nonetheless, there are several methods that have been proposed as a possible solution to the issue. One method that is discussed herein is the adaptive multi-modal background subtraction method. This method handles slow changes in illumination, repeated motion from background clutter and long-term scenes. This is just one of the different methods of motion-based techniques.

The different methods of video analytics have been noted to be facing certain challenges. In the development of a video analytic system, the main functions that the system is expected to perform include the classification of the detected objects, tracking of the detected objects over time, and classification of the object's movements. The challenges to this endeavour include the challenge associated with the sensors, occlusion, lighting issues, angle issues, etc. The literature on the main issue at hand which is movement classification is also included in the chapter. What emanates from the literature review is that studies and researches conducted in the past on the same were fruitful most of the time though the proposed solutions were marred with some challenges. Nonetheless, it is clear that there is progressive improvement on the issue. More so, the video analytic technologies that are discussed herein are strategically engineered in order to try and meet the video forensic requirements.

There has been some research in using AI and NN techniques to solve problems of movement classification. However, AI and NN have shortcomings of classifying what is abnormal based on the training it previously acquired from the inputs (Hawkins and Blakeslee 2004). However, Bio-inspired computational models have proved to be successful over conventional methods. Although ANNs remain of the most active classification techniques in various applications and researches, it is not without flaws as mentioned earlier. One effective method is known as Back Propagation Neural Network (BPNN), which has shown more accuracy when compared with maximum likelihood method. However, this work is, therefore, focusing on the study of the CLA through the investigation of the possibility of applying it to movement classification, to develop a novel movement classification technique.

The problem with the movement classification in video analytics is discussed in detail. To understand the issue better, the literature on the subject is discussed with a further focus on the weaknesses and strengths of these research studies. Seemingly, there are challenges that are associated with video forensics as far as information analysis is concerned. The analysis of the video in different contexts depends on the method that is used. Here in, there are discussions on the different researches that have been conducted in the past seeking to shed more light on video analytic especially in its movement classification task.

Artificial Neural Networks (ANNs) has been noted to be an active research area since the 1980s and has made a huge success. It is also indicated herein that it faces several challenges such as selection of the structure and the parameters of the networks, selection of the learning samples, selection of the initial values, the convergence of the learning algorithms amongst several other challenges (Ding. et al. 2013).

This Chapter concludes that, previous research in building intelligent machine similar to human capabilities remain limited, even though literature evidence indicate that despite series of effort made by AI researchers and recent effort made by Artificial neural network researchers to build viable algorithms for achieving human-like performance, they still suffer fundamental flaws, as all the existing solutions fail to adequately address what intelligence is or what it means. However, it was also concluded that more efforts are required to address how the brain works such as remembering fast events. Although human brains are made of neurons; therefore, the brain is a neural network that is the fact. For the understanding of how neurons interact which will lead to the emergence of properties of intelligence, understanding this connection of neurons still remains a problem that was unsolvable with AI which means replicating the correct connections between populations of neurons. Another important conclusion that can be drawn from the studies on the application of the Cortical Learning Algorithm in movement classification.

Chapter 5

Application of Cortical Learning Algorithms to Movement Classification

5.1 Introduction

This chapter discusses the use of Cortical Learning Algorithms for movement classification. It gives details about the Hierarchical temporal memory (HTM), and its components and choices of CLA were discussed. Furthermore, the proposed movement classification technique is presented with its requirements and implementation. The chapter concludes with a recap which further leads to the next chapter.

The studies of how the human brain works with the aim of imitating and training a machine to intelligently behave and computationally act as the human brain for ease of analysis is of great importance to the research community, looking at an enormous task that the brain goes through in the study of various events that might occur or has occurred (Hawkins, Ahmad, and Dubinsky, 2011).

HTM is a memory-based system that models the architectural details of neurons that, as such, is one of the categories of neural network (NN) (Ahmad and Hawkins, 2015). HTM model's neurons, which are arranged in columns, layers as well as regions of hierarchy. However, HTM as a network is trained on lots of time-varying data and relies on storing a large set of object patterns and sequences.

HTM is a machine learning technology capturing the algorithmic and structural properties of the neocortex. Simply put, the neocortex is the intelligence thought in a mammal's brains performing high-level touch, language, hearing, vision, movement, and planning thus neocortex implements a wide range of specialized algorithms (Flannelly, 2017). HTM, therefore, gives a framework of how neocortex functions as its programming require training by exposure to a stream of sensory data. It is a system based on memory, and its networks are trained with lots of time-varying data which relies on stored sequences and patterns and has a restrictive memory (Branch

2006). HTM is inherently time-based and is organized in a hierarchal manner that allows the specification of the size of the hierarchy and what to train the system on as the information distribution is being controlled by HTM. The user identifies what to train the system on, and the size of the hierarchy as the HTM only controls how and where the information is stored.

The benefit of the hierarchy in the neocortex is efficiency perhaps significantly reduces training time for the system and memory usage because patterns learned at each level of the hierarchy are reused when combined in novel ways at higher levels. At the lowest level of the hierarchy, the human brain stores information about a point. However, the point is the small unit of many objects and its properties such as movements in the world. Also, the lower level pattern is recombined at middle levels or regions into more complex objects such as cars, houses, street, etc.

To learn a new high-level object the brain does not need to re-learn its components, that is one advantage of the hierarchical model in the neocortex of the human brain and also that is one of its capabilities. As another example, consider that when you see a new car, you do not need to relearn whether if it can run, stop and reverse behaviour, so also if, for the first time in your life-time you see a new animal with wild teeth you do not have to relearn whether if it can bite and can also eat with, etc., you can actually predict the behaviour of that animal.

Sharing representations in a hierarchy is also leading to a generalization of expected behaviour. The hierarchy enables a new object in the world to inherit the known properties of its sub-components. How much can a single level in an HTM hierarchy learn? Alternatively, rather how many levels in the hierarchy are necessary? There is a trade-off between how much memory is allocated to each level and how many levels are needed. Fortunately, HTMs automatically learn the best possible representations at each level given the statistics of the input and some resources allocated (Fan, Sharad, Sengupta and Roy, 2016). If you allocate more memory to a level, that level will form representations that are larger and more complex, which in turn means fewer hierarchical levels, may be necessary. If you allocate less memory, a level will form representations that are smaller and simpler, which in turn means more hierarchical levels may be needed. Up to this point, it has been describing difficult problems, such as vision inference, ("inference" is similar to pattern recognition). However, many valuable problems are simpler than vision, and a single HTM region might prove to be sufficient. For example, to apply an HTM to predict where a person browsing a website is likely to click next. This problem involves feeding the HTM network streams of web click data. In this problem there is little or no spatial hierarchy, the solution mostly required discovering the temporal statistics, i.e., predicting where the user would click next by recognizing typical user patterns. The temporal learning algorithms in HTMs are ideal for such problems. Hierarchies reduce training time, reduce memory usage, and introduce a form of generalization; this study argues that such capabilities of learning algorithms would tackle and simplify solving the problem of movement classification of an object.

Cortical Learning Algorithms (CLAs) comprises an effort by researchers in Numenta Inc. to design a model that can perceptually and computationally analyse neocortex learning in the brain (Hawkins et al. 2011; Agrawal & Franklin 2014). The cortical learning algorithm is utilized as a part of the second implementation of a designed framework for perceptual learning called Hierarchical Temporal Memory (HTM) (Agrawal & Franklin 2014; George 2008; George & Hawkins 2009; Hawkins et al. 2009). As further highlighted by Agrawal and Franklin (2014) that the algorithm (CLA) typically functions on a set of data structure, and the two of them together to accomplish some level of spatial and temporal pattern recognition. The data structure utilized is a gathering of segments of cells, called a locale. A cell in a section is a neuron-like substance, which makes associations with different cells, and totals their action to decide its state of initiation.

5.2 Challenges and Requirements of Movement Classification

As noted in the previous chapter, there are several challenges to movement classification. The main issues are associated with the hurdles that the outdoor surveillance system designers have to deal with. Notably, these issues must also do with the detection, tracking as well as the classification of the moving objects that are captured by the surveillance equipment (Alshaikh and Sedky, 2016). The other main challenges, which were discussed in the previous chapter is the fact that the sensor resolution that is used by the cameras are finite and thus cannot be able to tack all angles and everything that is under the complete area considered to be of interest. To deal with these issues, it was recommended that there is need to install several cameras which would be placed in a way that will enable them to capture different angles and positions of the moving target.

The other main challenge that is experienced in movement classification has to do with the appearance of the objects. Notably, many times the object will change position and angles and thus is considered to be dynamic (Alshaikh and Sedky, 2016). In this context, the camera and surveillance system are expected to be dynamic as well when it comes to the tracking of the moving object. The movements of the object should also be tracked and even their trajectory estimated. Overall, if these roles are enacted as indicated in the previous chapter, it will make it easy to capture and trace the movement of an object through different surveillance equipment.

As noted in the previous chapter as well, there is the challenge of occlusion that takes place when surveillance is being done. This is when other structures or objects or even shadows block the object that is being observed. Because of such blockage, the discontinuity of such observation is disruptive to effective analysis and video forensics. In this case, it is usually not easy to accurately position the object. It is not easy as well to determine the shape the size and the velocity of the object, and as such, the overall quality of the forensic video analysis is affected.

5.3 The rationale for The Proposed Algorithm

The rationale behind the Cortical Learning Algorithms (CLA) is based on the capabilities of the algorithm which enable it to solve the movement classification issues at hand. The main reason what the Cortical Learning Algorithms (CLA) is considered to be a good option is based on the ability of the algorithm to analyse neocortex learning in the brain both conceptually and perceptually (Agrawal and Franklin, 2014). The CLA also functions on a data structure that strategically makes it a good option simply because it can accomplish some level of spatial and temporal pattern recognition (McLean, 2014). The data structure that is used is based on a segment of cells that are referred to as a locale. The algorithm brings the idea of real-life scenario closer to be a reality in surveillance.

5.4 Hierarchical Temporal Memory

Hierarchical Temporal Memory (HTM) is a machine learning technology that models how higher-level capabilities of the human neocortex brain is and how it performs tasks such as visual pattern recognition, understanding language recognizing object's movement, etc. The neocortex is the seat of intelligent thought in the mammalian brain. High-level vision, touch, hearing, planning, movement, and language are all performed by the neocortex (Leake, Xia, Rocki and Imaino 2015). Given such a diverse suite of cognitive functions, one might expect that the neocortex would implement an equally diverse site of specialized neural algorithms. However, this is rather untrue as neocortex displays an extraordinary unchanging pattern of neural circuitry. It is suggested by biological evidence that the neocortex implements a common set of algorithms to perform many different intelligence functions. The neocortex ability is understood through the theoretical framework provided by HTM (Kirtay et al., 2016).

This study proposes an object movement classification technique based on the concept of this technology depicted in Figure 5-1. However, HTM provides a type of neural network that is a memory-based network (Tian et al., 2018). This is fundamentally trained on a lot of time-varying data and relies on storing a large set of patterns and sequences. Information flow is always in a distributed fashion (Demirkus, Precup, Clark and Arbel, 2015).

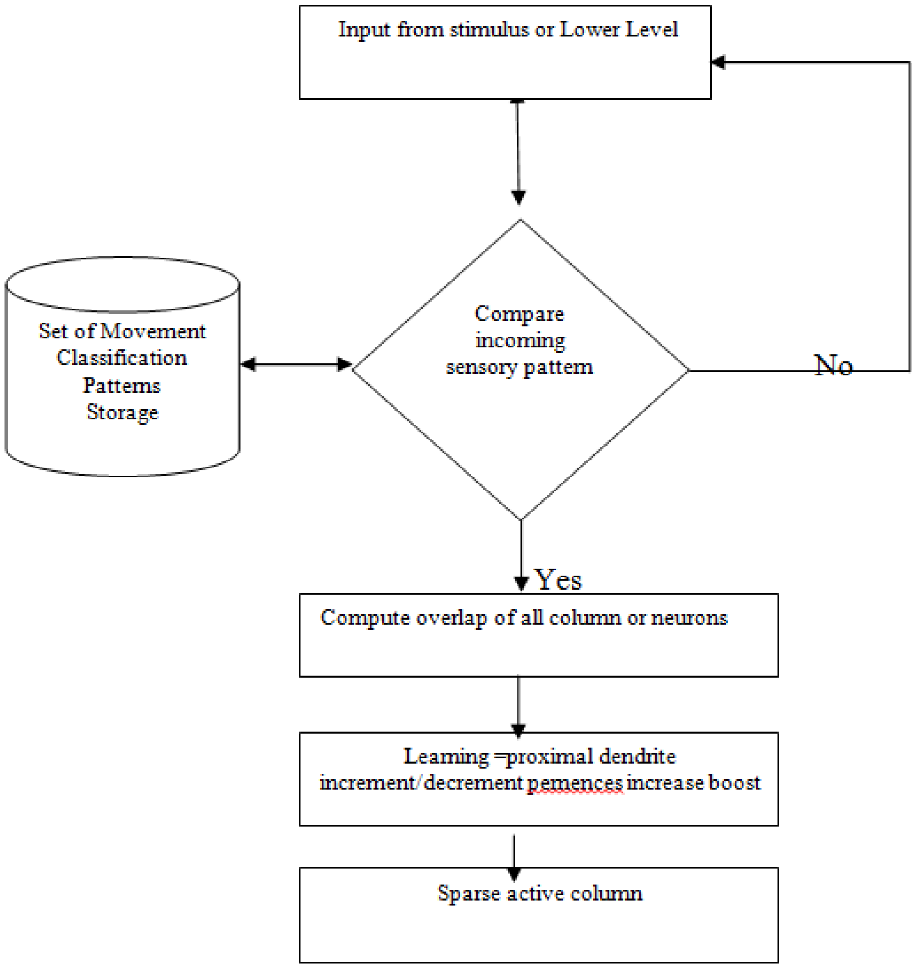


Figure 5-1: A novel movement classification technique based on HTM

HTM consists of some regions that represent a level of hierarchy in the network (Streat, 2016). As you ascend in the hierarchy, there is always convergence of information due to feedback so while as you descend the information diverges in the hierarchy.

Travelling up and down in the hierarchy, spatial and temporal resolution diverges and converges. At the lowest level of an HTM network, the input patterns of object identification and movement classification are constantly changing, much like the incoming sensory stimuli humans receive (Mnatzaganian, 2016). Cell activation patterns are more stable because information is transferred up and down in the hierarchy in predictable sequences. The brain constantly compares incoming sensory patterns and stores a model of the world that is largely independent of how it is perceived under changing conditions.

According to Hole (2016), the HTM builds sparse, invariant representations of pattern sequences representing repeated structures in the input stream. The algorithm learns which patterns are likely to follow each other, thus learning to predict future patterns. When the HTM receives a novel pattern, it tries to match it to stored patterns. Because inputs never repeat in the same way, invariance of the stored sequences is vital to the ability to recognize inputs. Also, time plays a crucial role in HTM. Predictions can only be made by a sequence of earlier received patterns. Sometimes it is enough to know the previous pattern most recently received while at other times it is also necessary to know patterns received earlier. The ability to predict using variable-length sequences of patterns is due to the variable order memory of HTM. It is important to note that HTM does not understand the meaning of patterns; it only knows what patterns are likely to follow particular observed patterns.

5.5 Cortical Learning Algorithms (CLA) and its Components

According to Purdy (2016), the Cortical Learning Algorithms (CLA) constitute effort made by Numenta Incorporation to design a model that can analyse neocortex learning in the brain both computationally and perceptually. The CLA is utilized as a part of the second implementation of a designed framework for perceptual learning called Hierarchical Temporal Memory (HTM). The algorithm, CLA, functions on a set of data structure, and the two of them together accomplish some level of spatial and temporal pattern recognition (Galetzka, Strüngmann, and Weber, 2014). The data structure utilized is a gathering of segments of cells, called a locale. A cell in a section is a neuron-like substance, which makes associations with different cells, and totals their action to decide its state of initiation. It is biologically proven that neocortex is the seat of intelligent thought in the human or mammalian brain (Horschig et al., 2015). Intelligent properties such as vision, movement, hearing, touching, etc. are all performed by this intelligent seat, this cognitive task that is primarily performed by the neocortex of humans are challenging to design in real life scenarios. Several thousands of neurons can be contained in a normal neuron. Most of the neurons are noted to be on distal dendrites. The remaining neurons are determined to be on proximal dendrites. For a long time, it had been considered that learning was about the weakening and strengthening of the effect that the synapses have regarding weight. In this context, it is noted that all synapses are stochastic. As such when the synapses are released, they do not release neurotransmitter in a reliable manner. It is also important to note that the synapses contained in the HTM cells have binary weight. It is also important to consider that the synapses that are contained in the HTM cell have a permanence value. This value is a scalar value. This value is usually adjusted as the learning process is underway.

The number of the valid synapses are not fixed especially those that are contained on the distal and proximal dendrite. The synapses are contained in the HTM cell.

5.5.1 CLA Components

The CLAs can be divided into four components: spatial pooler, encoder, temporal pooler, and classifier as Figure 5-2 shows.

Figure 5-2: A Process Flow (Balasubramaniam, Krishnaa, and Zhu, 2015)

Encoder: HTM gives an adaptable and naturally precise system for settling expectation, grouping, and oddity location issues for a wide scope of information sorts (Hawkins and Ahmad, 2015). HTM frameworks require information contribution to the type of Sparse Distributed Representations (SDRs) (Ahmad and Hawkins, 2016). SDRs are not quite the same as standard PC representations, for example, ASCII for content, in that significance is encoded straightforward into the representation. An SDR comprises of a vast exhibit of bits of which most are zeros, and few are ones. Every piece conveys some semantic meaning so if two SDRs have more than a couple covering one-bits, then those two SDRs have comparable implications. Any information that can be changed over into an SDR can be utilized as a part of an extensive variety of uses utilizing HTM frameworks. Therefore, the initial step of utilizing an HTM framework is to change over an information source into an SDR format utilizing what we call an encoder. The encoder changes over the local configuration of the information into an SDR that can be bolstered into an HTM framework. The encoder is in charge of figuring out which yield bits ought to be ones, and which ought to be zeros, for given information esteem in such a route as to catch the essential semantic qualities of the information.

Spatial pooler: The open field of every section is a settled number of information sources that are arbitrarily chosen from a much bigger number of hub data sources. In light of the info design, a few segments will get more dynamic information values

(Yang et al. 2009). Spatial pooling chooses a consistent number of the most dynamic sections and inactivates (represses) different segments in the region of the dynamic ones. Comparable information designs tend to actuate a steady arrangement of sections (Zhuo et al., 2012).

There are various goals that are accomplished by the spatial pooler which overlap. These goals are used to determine how the spatial pooler learns and operates.

- The first goal involves the use of columns. Usually, there are a fixed number of columns in an HTM region. These columns are used to represent the common patterns that are contained in a region. Some of the objectives in this context include the reassurance that the columns are being used to represent something that is regarded as useful (Moya and Rojas, 2015). This should be despite the number of columns that are present. It is important that a situation where there are some inactive columns is avoided. To accomplish this, the columns are tracked down. As this is being done, their activity is compared to that of their neighbours. If the inactivity level is determined to be too low, the activity level of the columns input is increased. This motivates the columns to change and increase its activity. As a result, the column ends up being one of the winning columns. As such, the columns compete with their immediate neighbours.

- There is also the goal to maintain the desired density. It is important that a sparse representation of the inputs be formed by a region. Those columns that do not have many inputs are inhibited by those that have. The columns receptive field's size is determined to be proportional to the radius of inhibition (Byrne, 2015). Only some of the columns that are determined to be active are classified as being winners within the radius of inhibition. Usually, in the radius of inhibition, the columns are arranged in 2D formation.

- The other goal is to avoid the trivial patterns. It is critical that only the non-trivial patterns are represented by the columns (Byrne, 2015). It is possible to accomplish this using the minimum threshold. The threshold for the different columns that have been classified to be active can be set. For instance, the threshold can be set at 40. In such an instance, the column would have to have 40 active synapses for it to be active. When there is a threshold that has been set, trivial patterns are avoided since there is a guarantee of a specific level and trait in the columns.
- The other main goal is to avoid extra connections. As the column is forming connections, it should be controlled. Notably, the initial connections might not overwhelm the column. However, with time, the column could start responding to unrelated patterns (Byrne, 2015). To avoid this, the permanence value of the different synapses that are not contributing to the winning column is reduced. When this is done, then the non-contributing synapses are controlled. In such an ideal situation, the column only gets to represent limited input patterns.
- The other goal is associated with the self-adjusting of the receptive fields. It is important to note that the real brain is designed to be plastic. The neocortex is noted to be flexible and can change in certain situations. For instance, it can be able to learn to represent different things as it is reacting to change (Byrne, 2015). Interestingly, when a part of the neocortex is unable to perform its duties, it is replaced by other parts. The other parts come in to accomplish what it failed to do. The system is designed to be self-adjusting to mitigate issues that arise.

The spatial pooler functions are designed to accomplish the following: Firstly, it starts with a specific number of bits which represent the sensory nerve. It is also tasked with the role of assigning the columns to the regions which are expected to receive the input. In this context, the columns have a fixed number of dendrites. These dendrites, in turn, have specific potential synapses. The potential synapses are then assigned permanence values. These values are then used to determine whether a synapse will be determined to be valid or not.

It is the responsibility of the spatial pooler to determine the exact number of active synapses which are connected to the active input bits. Thisis done for every given input. The pooler also functions to facilitate the multiplication of all the active synapses. Upon boosting, a fixed number of columns that are contained in the inhibition radius are disabled. This is accomplished by those columns that have the highest rate of activations. The spread of the input bits is what determines the inhabitation radius. The result of the process is the active columns that are sparse.

The final function of the spatial pooler is the adjustment of the permanence values. This is done with all the potential synapses. During the adjustment, those synapses which are aligned with the active input bits, their permanence values are increased. On the other hand, for those that are aligned with those that have inactive input, their values are decreased. Due to the changes that are made to the permanence values, some synapses might change from being valid to invalid. The opposite can also happen.

Temporal Memory: Temporal pooling has been a dynamic region of research for HTMs and Numenta for quite a long while. The significance of temporal pooling and the general objectives of temporal pooling have been to a great extent predictable (Melis et al. 2009). Be that as it may, the expression "temporal pooler" has been utilized for various diverse executions and looking through the code, and past documentation can be to some degree confounding. The first CLA whitepaper utilized the term temporal pooler to depict a specific usage. This usage was unpredictably tied in with succession memory. Thus the succession memory and transient pooling were both alluded to as "temporal pooling," and the two capacities were perplexed (Perea et al. (2009) and Zhituo et al. (2012)).

The temporal pooler checks the active columns to determine whether there are cells which have assumed a predictive state in the column. If found, the pool functions to activate such cells. Finally, the cells that will be active will be a resemblance to the input regarding the prior input. The pooler also makes sure that the active cells that are connected to synapses. This is done for every dendrite segment. If it happens that the number is exceeding the threshold, then the dendrite is considered to be active. As for those cells that do not have active dendrites, they are set to remain inactive. The prediction of such a region is the collection of those cells that are in a predictive state.

When a dendrite segment is activated, the pooler modifies the permanence values for the various synapses. The synapses that are affected are those that are associated with that particular segment. The permanence of those synapses which are connected with the active cells is also increased by the pooler. However, it also serves to decrease the permanence of those synapses which are connected to inactive cells. Nonetheless, the changes that are made to the synapses are considered to be temporary changes. The pooler also serves to modify the synapses that are located on the segments which are fit to be made active. Such segments are those that have already been trained enough and can achieve an active state.

There are instances when the status of a cell can change from active to inactive. In such an instance, if the change takes place as a result of feed-forward input, then the potential synapses that are associated with that particular cell are traversed. In such cases, the temporary marks that have been put on the synapses are removed. As such, it is only those synapses that made the correct predictions as far as feed-forward activation is concerned that is updated. When there is a change in a cell where the cell changes from either inactive to active or vice versa, the permanent changes that had been marked previously are changed. In such cases, the permanence changes are marked as active.

Classifier: HTM-CLA plans to learn and speak to structures and groupings in light of memory predictions. In any case, the classifier used to construe the arrangement yield from HTM-CLA are a long way from palatable. Classifiers utilized as a part of the Numenta Platform for Intelligent Computing (NuPIC) structure are Classifier, e.g. H-DS Classifier, H-MSC Classifier, and CLA Classifier (Balasubramaniam et al. 2015). Some classifiers are also proposed by Balasubramaniam et al. (2015). The principal technique is based on dot similarity (H-DS) Classifier and the second strategy is H-MSC Classifier given Mean-Shift Clustering. It intends to make the classifiers in HTM-CLA more productive and powerful.

5.6 The Choice of CLA

Lavin& Ahmad (2015) states that a significant part of the world's information is spilling, time-arrangement information, where oddities give noteworthy data in basic circumstances; illustrations possess large amounts of spaces, for example, fund, IT, security, medicinal, and vitality. However, distinguishing inconsistencies in gushing information is a troublesome assignment, obliging indicators to prepare information continuously, not clusters, and learn while all the while making expectations. This is also applicable to movement classification techniques in video forensic. As it is stated by these researchers, the difficulties of identifying or locating anomalies are quite challenging,

The advantages of the CLA includes its accuracy as it chooses the main stage for automatic sleep classification and this gives a lower rate of error. For example, CLA has been applied to determine the number of distinct features present in a set of binary images data and to classify unknown input, 66% accuracy has been obtained with distortions on binary images using given parameters. (Zhang, Zhang & Chen, 2017) High processing speed which is influenced by the hierarchy in the structure of HTM is also an advantage. Although a large number of patterns was determined in their training data, the pattern count size increases the processing speed of a single pattern. Only the number of patterns to be checked affects the processing speed.

The CLA requires no inference as they can learn from any new input pattern continuously thus it is not necessary to have separate inference mode (Zhang, Zhang & Chen, 2017). HTM matches the streaming inputs to the previous fragments of sequences learned; the recognized sequences are therefore continuously predicted. CLA being an online learning algorithm, needs no pre-processing and requires less training time for example, CLA has been applied to solve the problem of classifying Electrocardiogram (ECG) samples into sick and healthy groups discriminating subsequence eliminated in the signal after supervision which could otherwise be done by human supervisor Balasubramaniam, Krishnaa & Zhu (2015). It is advantageous that it is the CLA's variable length contextual memory can detect future anomalies.

5.7 Proposed Novel Movement Classification Technique

This study proposes the application of HTM for movement classification. The proposed algorithm, which is based on the CLA that learns to predict a sequence of movement, will learn some events, represented by a sequence of movements and then it will be tasked to differentiate between an event similar to the ones that have been learnt and events that have not been learnt. This would be a desirable property since, post-incident analysis, the detection of abnormal movement is required. A slightly erroneous copy of the learned sequences will be presented to the algorithm, which will recover quickly after any unexpected or suspicious movement patterns. In the section below, the requirements for the proposed technique are presented.

5.7.1 Requirements

A set of requirements for the proposed movement classification technique are:

1. Spatial and temporal object movement classification input pattern (feed-forward). Feeding the HTM network stream of movement trajectories discovering the temporal statistic to predict how a specific object type moves.
2. Normal trajectory patterns to be learned and efficient storage for storing a representation of learnt patterns.
3. A defined inhibit radius that defines the area around a column that actively inhibit.
4. Provision of scalar value which indicates the connection state of potential synapses. This value indicates the synapses are not formed if the value is below the threshold otherwise it is valid if it is above the threshold.
5. A set of model parameters.

5.8 Implementation

The implementation of the proposed movement classification algorithm follows the following steps:

Model creation: the model is created by running swarm to define the model parameters and adjust for any modifications if required. The swarm starts by running the Permutations function to automatically generate multiple prediction experiments that are permutations of a base experiment via the CLA engine. The type of inference is to be specified, e.g., 'Temporal Anomaly,' the encoder settings as well as Spatial Pooler and Temporal Pooler parameters.

The encoder parameters include the type of the encoder, e.g. 'Scalar,' 'Adaptive Scalar,' 'Category' or 'Date' encoders. *N* the number of bits to represent input and *w* the width or the number of 'On' bits (1's). The Spatial Pooler parameters include: Column Count, number of cell columns in the cortical region, number of Active Columns Per Inhibition Area, maximum number of active columns in the SP region's output (when there are more, the weaker ones are suppressed), potential Percentage, the percent of the columns' receptive field is available for potential synapses, synapse Permanence Connected, the default connected threshold. Any synapse whose permanence value above the connected threshold is a "connected synapse," meaning it can contribute to the cell's firing.

The Temporal Pooler parameters include Column Count, number of cell columns in the cortical region (same number for Spatial Pooler), cells Per Column, the number of cells (i.e., states), allocated per column, max Synapses Per Segment, maximum number of synapses per segment, max Segments Per Cell, maximum number of segments per cell, initial Perm, initial Permanence, permanence Increment, permanence increment and permanence Decrement, permanence decrement, min Threshold, minimum number of active synapses for a segment to be considered during search for the best-matching segments, activation Threshold, Segment activation threshold, a segment is active if it is greater than this threshold.

Learning: the model starts by enabling learning, this indicates the training state. After this stage, the model has learned normal activities and is ready to perform prediction and anomaly detection.

Anomaly detection: When the preparation is finished, the learning is disabled, and the model is exchanged in the anomaly detection state. The model can perform detection of normal and abnormal activities.

5.9 Spatial Pooler Pseudocode

Three phases take place in the sequence. The first phase involves the computation of the overlap using the current input. This is done for all the columns. The second phase involves the computation of the winning columns. The process is undertaken after the inhibition has taken place. The third phase involves the updating of the synapses permanence. The internal variables are also updated as well. Before any inputs are received into the pooler, a list is developed of the possible synapses for all the columns.

The list of the potential synapses includes some sets of inputs that are selected randomly. The sets are acquired from the input space. A specific synapse is assigned to each input. The synapse is also assigned a permanence value that is acquired randomly. The values used are chosen using two criteria. First, they should be in small ranges that should be above the minimum permanence value. The results are that such synapses can then be connected. The connection is done after training has taken place. Secondly, it is important to note that the columns have a natural center that is over the input region. It is also critical to consider that the permanence values are noted to have a bias against the centre.

Phase 1: Overlap

The first phase of the pooler involves the calculation of the overlap vector. The overlap is the synapses that are determined to have active inputs. These are then multiplied by the boost. If the resultant value is below the set min Overlap, then its overlap score is determined to be zero. Here is a pseudo code for the mentioned process.

`Pseudo code1: Spatial Pooler: Calculation of the overlap vector

```
For c in columns
    overlap(c) = 0
for s inconnected Synapses(c)
    overlap(c) = overlap(c) + input(t, s.sourceInput)
    if overlap(c) < minOverlap then
        overlap(c) = 0
    else
overlap(c) = overlap(c) * boost(c)
```

The second phase is the inhibition phase. In this phase, the columns that should remain as winners after inhibition has taken place are calculated here. The number of columns that are classified as winners is controlled by the desired Local Activity parameter. This is detailed in the pseudocode below.

Pseudo code2: Spatial Pooler: Inhibition

```
For c in columns
    for c in columns
    minLocalActivity = kthScore(neighbors(c), desiredLocalActivity)
    if overlap(c) > 0 and overlap(c) ≥ minLocalActivity then
    active columns(t). append(c)
```

The third phase is the learning phase. At this phase, the permanence values for all the synapses are updated. This is done where necessary. The inhibition radius and the boost as well are also updated. Below is the implementation of the main learning rule. Notably, the increment of permanence values is done only by winning columns where the synapses are active. If they are not, then the values are decreased. Boosting is implemented in the pseudo code three below. These mechanisms are implemented to facilitate the learning process. The code below shows that there is re-computation of the inhibition radius. The pseudocode is as follows:

Pseudo code3: Spatial Pooler: Learning

```
for c in activeColumns(t)
for s inpotential Synapses(c)
        if active(s) then
          s. permanence += permanenceInc
          s. permanence = min(1.0, s. permanence)
else
    s. permanence -= permanenceDec
    s.permanence = max(0.0, s. permanence)
for c in columns:
    minDutyCycle(c) = 0.01 * maxDutyCycle(neighbors(c))
    activeDutyCycle(c) = updateActiveDutyCycle(c)
    boost(c) = boostFunction(activeDutyCycle(c), minDutyCycle(c))
    overlapDutyCycle(c) = updateOverlapDutyCycle(c)
if overlapDutyCycle(c) < minDutyCycle(c) then
   increasePermanences(c, 0.1*connectedPerm)
 inhibitionRadius = averageReceptiveFieldSize()
```

5.10 Temporal Pooler Pseudocode

The first phase of the temporal pooler is used to calculate the different cell's active states. Using the pseudo code 4, the cell that is supposed to be active is determined. Firstly, those cells that predict the bottom-up input are made active (Hawkins, Ahmad, and Dubinsky, 2011). On the other hand, if this was unexpected, then all cells in that column will be made active. This is accomplished in the inference stage, which is in phase 1. Below is the pseudo code for these processes.

Pseudo code 4: Temporal Memory–Calculation of the Cell's Active States

```
for c in activeColumns(t)
        buPredicted = false
for i = 0 to cellsPerColumn - 1
if predictiveState(c, i, t-1) == true then
s = getActiveSegment(c, i, t-1, activeState)
if s. sequenceSegment == true then
buPredicted = true
activeState(c, i, t) = 1
        if buPredicted == false then
for i = 0 to cellsPerColumn - 1
activeState(c, i, t) = 1
```

After the calculation of the active state, which is carried out in the first phase, the predictive state of the cells is calculated in the second phase. Below is the pseudocode for the second phase.

Pseudo code 4: Temporal Memory - Calculation of Active Cells

```
for c, i in cells
 for s in segments (c, i)
 if segmentActive(c, i, s, t) then
predictiveState(c, i, t) = 1
```

As far as combined inference learning is concerned, there are two phases as well. The first phase includes the calculation of the active State of the cells that are in the winning column (Hawkins, Ahmad, and Dubinsky, 2011). In this process, a learning cell is selected from each column. The logic that is employed here is that the cells that will become active will be based on the prediction that was made as far as the bottom-up input is concerned. On the other hand, the learning cell is selected from cells that were chosen to learn State, and the segment became active (Hawkins, Ahmad, and Dubinsky, 2011). The pseudo code for this process is as detailed below.

Pseudo code 5: Temporal Memory – Learning

```
for c in active Columns(t)
        bu Predicted = false
lc Chosen = false
for i = 0 to cellsPerColumn - 1
 if predictive State(c, i, t-1) == true then
s = getActiveSegment(c, i, t-1, activeState)
if s. sequenceSegment == true then
```

```
buPredicted = true
activeState(c, i, t) = 1
if segmentActive(s, t-1, learnState) then
lcChosen = true
learnState(c, i, t) = 1
        if buPredicted == false then
for i = 0 to cellsPerColumn - 1
activeState(c, i, t) = 1
        if lcChosen == false then
        I,s = getBestMatchingCell(c, t-1)
LearnState(c, i, t) = 1
sUpdate = getSegmentActiveSynapses (c, i, s, t-1, true)
sUpdate.sequenceSegment = true
segmentUpdateList.add(sUpdate)
```

It is in the second phase that the predictive state of the different cells is calculated.

The pseudo code in this phase is as follows:

Pseudo code6: Temporal Memory –Calculation of Predictive Cells

```
for c, i in cells
for s insegments (c, i)
if segmentActive(s, t, activeState) then
predictive State(c, i, t) = 1
activeUpdate = getSegmentActiveSynapses (c, i, s, t, false)
```

```
segmentUpdateList.Add(activeUpdate)
predSegment = getBestMatchingSegment(c, i, t-1)
 predUpdate = getSegmentActiveSynapses(
c, i, predSegment, t-1, true)
segmentUpdateList.add(predUpdate)
```

In the third phase, those segments updates that had been queued are implemented. This is done once the feed-forward input is acquired. At the same time, the cell has to be determined to be a learning cell. This is accomplished in pseudo code seven below. On the other hand, the segments are negatively reinforced whenever the cell stops carrying out the prediction.

Pseudo code7: Temporal Memory – Segment Updates

```
for c, i in cells
if learnState(s, i, t) == 1 then
adaptSegments (segmentUpdateList(c, i), true)
segmentUpdateList(c, i). delete ()
else if predictiveState(c, i, t) == 0 and predictiveState(c, i, t-1)==1
then
adaptSegments (segmentUpdateList(c,i), false)
 segmentUpdateList(c, i).delete()
```

5.11 Conclusion

As much as HTM networks differ from classic computing, general purpose computers can be used to model them, and the key functions of sparse distributed representations, time and hierarchy incorporated.

The application of HTM for movement classification is proposed in this study. The proposed algorithm as contained in the study is based on the CLA that learns to predict a sequence of movement. It is also noted that these will learn some events, represented by a sequence of movements and then it will be tasked to differentiate between an event similar to the ones that have been learnt and events that have not been learnt. As such, this is considered to be a desirable property. The reason for this conclusion is since, for post-incident analysis, the detection of abnormal movement is required.

The HTM hierarchical organisation aids in efficiency, where memory usage and training time is reduced since at each hierarchy level the patterns learned are reused. Prediction by HTM informs whether a new input is expected or not as each of its regions detect novelty and stores sequence of patterns, as stored sequences are matched with stored inputs. In HTM, the concept of time is crucial in learning, prediction, and inference because nothing will be inferred from a sensory reading without using its history in time. Therefore, HTM suits best learning from a stream of data during training.

The Hierarchical Temporal Memory (HTM) which is a machine learning technology is also discussed. More importantly, it is determined that the neocortex ability is understood through the theoretical framework provided by HTM. On the other hand, the components of the CLA are also discussed in the chapter. These are the spatial pooler, encoder, temporal pooler, and classifier.

Various versions of HTM cortical learning algorithms have been implemented. CLA is thus suited for processing various data streaming online, learning unsupervised, dealing with unaligned sequential data and noise as the patterns encode a temporal relation which captures global information about events.

The Cortical Learning Algorithms (CLAs) that comprises an effort by researchers in Numenta Inc. to design a model that can perceptually and computationally analyse neocortex learning in the brain is the proposed solution. The CLA is said to be utilized as a part of the second implementation of a designed framework for perceptual learning called Hierarchical Temporal Memory. The challenges to movement classification are also detailed in this chapter. These include occlusion and other challenges. The rationale for CLA is also included herein.

In this chapter, a bio-inspired movement classification technique based on HTM theory was proposed. The proposed technique starts by learning about the events, after the training phase, learning is stopped and then it performs prediction. The technique can perform classification of normal and abnormal activities when the learning is finished.

The proposed movement classification technique and its requirements include the spatial and temporal object movement classification input pattern (feed-forward), normal trajectory patterns to be learnt and efficient storage for storing a representation of learned patterns and a defined inhibit radius that defines the area around a column that actively inhibit among others.

Chapter 6

Test and Evaluation

6.1 Introduction

This Chapter tests and evaluates the proposed CLA movement classification algorithm. It presents the evaluation methodologies as well as the dataset used in this study,i.e. VIRAT video dataset. It includes the experimental setup, test results and evaluation. It also concludes with a summary of the chapter.

Currently, the discovery of what is happening in a scene can be seen by automatic scrutiny of activities included in a video. In the previous chapter, Neocortex algorithms have been identified as a solution to the movement classification problem. Movement classification is a key component of automated video forensic systems. However, the required performance of such algorithms differs depending on the target scenario, and on the characteristics of the monitored scene (Demirkus, Precup, Clark and Arbel, 2015).

Due to the diversity of video surveillance scenarios and the increasing development of movement classification algorithms (Zouaoui et al., 2015), an automatic assessment procedure is desired to compare the results provided by different algorithms (Eigenraam and Rothkrantz, 2016). This objective evaluation compares the output of the new algorithm with the ground truth, obtained manually, and measures the differences using objective metrics (Alshaikh and Sedky, 2016). The key challenge here is the presence of a decision process parameter(s) (threshold(s)) that influence the outcome of the algorithms. So, such test should assist in both optimal selecting parameters for the tested algorithm and measure its accuracy when applied on different scenarios for a given scene.

This chapter aims to test the movement classification algorithm proposed in Chapter 5, targeting the requirements of the post-incident analysis application discussed in Chapter 2. An objective criterion, which indicates the difference between the detected anomaly score, after applying a threshold, and a reference activity will be used. A public domain movement classification dataset, VIRAT, will be used in this chapter to test the performance of the proposed algorithm. VIRAT video dataset includes training and testing datasets, total of 152 videos, captured from three different scenes.

Each video compromises a scenario that represents a set of activities. VIRAT training dataset includes manually annotated activities, i.e. ground truth (Wang et al., 2017).

Section 6.2 discusses the performance metrics which will be used in the threshold selection criterion and objective evaluation throughout this chapter. The experiments setup the dataset used in this chapter are introduced in Section 6.3. Optimal algorithm's decision rules (thresholds) that provide the best performance for each algorithm are chosen in Section 6.4. In the same chapter also, the performance metrics are applied to objectively evaluate the proposed algorithm. In Section 6.5, a conclusion of this chapter is detailed.

6.2 Evaluation Methodologies

Scene-independent and scene-adapted learning recognitions are the two evaluation modes that are used for testing datasets. Scene-independent has a trained event detector on the scene which is not included in the test. In this case, the test clips are used during the test process. On the other hand, the scene-adapted learning recognition is applied to the clips that are to be used for the training processes (Wan et al., 2018). In the latter case, the test clips are not used during the process.

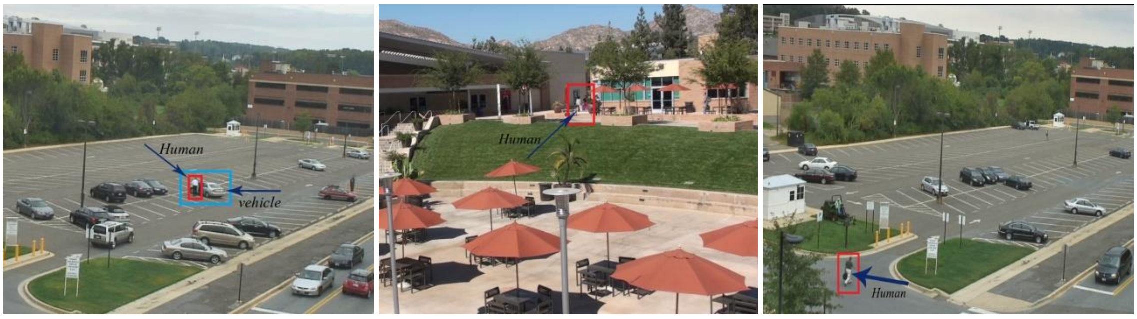
Evaluation approaches involve functional scene recognition, and multi-object tracking which is ground truth-based annotations providing important basis to perform large-scale evaluations and determine real-world performance measures. Various metrics are devised for the evaluation of movement classification algorithms (Anjum et al., 2016).

The Cortical Learning Algorithm is applied in this case. The CLA processes streams of information, classify them, learns to spot differences, and uses time-based patterns to make predictions (Byrne, 2015).

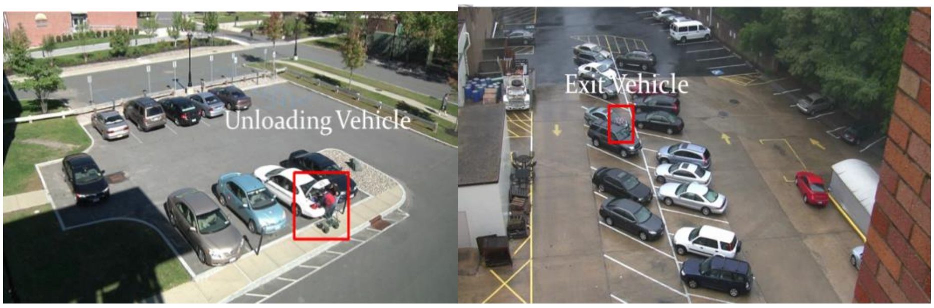
6.2.1 VIRAT Video Dataset

Most post-incident analysis cases target outdoor scenarios. Not all publicly available movement classification and action recognition datasets represent realistic real-world surveillance scenes and/or scenarios as they, generally, contain short clips that are not representative of each action performed (Khanam and Deb, 2017). Some of them provide limited annotations which comprise event examples and tracks for moving objects, and hence lack a solid basis to carry out evaluations in large-scale. Moreover, performance on current datasets have been flooded, and therefore a new more complex and large dataset must be introduced to enhance progress (Khanam and Deb, 2017).

VIRAT video dataset is a large-scale dataset facilitates the assessing of movement classification algorithms. The dataset was designed to be natural, realistic, and challenging for video surveillance domains stipulated to its background clutter, resolution, human event/ activity categories and diversity in scenes than existing action recognition datasets (Moon et al., 2015).



a. **A human getting into a vehicle** b. **A human entering a facility** c. **A human carrying a bag**



d. **A human is accessing a trunk** e. **A human exiting a car**

e. A human is entering a car e. Not known

Figure 6-1: Snapshots from VIRAT video dataset.

According to Moon et al. (2015), VIRAT video dataset distinguishing characteristics are as:

- **Realism and natural scenes:** The data is collected in natural scenes by showing people in standard contexts performing normal actions, with cluttered and uncontrolled backgrounds. Also, all actions performed by the directed actors were minimised, and in actual sense, they were being performed by the general population.
- **Diversity:** The data is collected from multiple sites through a variety of camera resolutions and viewpoints were included, while actions are performed by many different people (Moon et al., 2015).
- **Quantity:** Various types of human-vehicle and human interaction are included with a large number of examples (>30) per action class.
- **A wide range of frame rates and resolution:** Many applications operate across a wide range of temporal and spatial resolutions such as video surveillance. Therefore, the dataset is designed purposely to capture the ranges, (with 2–30Hz) frame rates and 10–200 pixels in person-height.
- **Ground and Aerial Videos:** Both aerial videos and ground camera videos are collected and released as part of VIRAT Video Dataset.

VIRAT video dataset includes a total of Eleven scenes that were captured by stationing high definition cameras and encoded in H.264. Each scene contains many video clips, and each clip has many activities. The file name format is unique which makes it easier for the identification of videos that are from the same scene using the last four digits that indicate collection group ID and scene ID. Figure 6-1, shows snapshots taken from the VIRAT dataset that presents Three sample activities. This project uses VIRAT video dataset to perform the quantitative evaluation for the proposed movement classification algorithm.

VIRAT video dataset is divided into two parts, training and testing datasets. The training dataset contains video scenes with several categories of human and vehicle activities recorded by stationary cameras, in a surveillance setting, in scenes considered realistic. Six object categories are included, unknown, person, car, another vehicle, other object and bike. Seven activities are presented, unknown, loading, unloading, opening trunk, closing trunk, getting into a vehicle and getting out of the vehicle.

6.2.2 The identification model

The dataset Virat that would be included will be frame number, the object type, the object ID, Bounding Box X position, y position, height and width. The frame number is a crucial aspect of identification as it is used to identify an object. The object type dataset would include "classes" such as person, bike or car. The object ID dataset includes shape size, variations in time, previous location, and velocity that coincide with the desired object. The Bounding Box uses data such as X position, y position, height and width to determine the size of the box which is used to define the boundary of an object which is usually represented by the centre (bx, by), rectangle height (bh), rectangle width(bw).

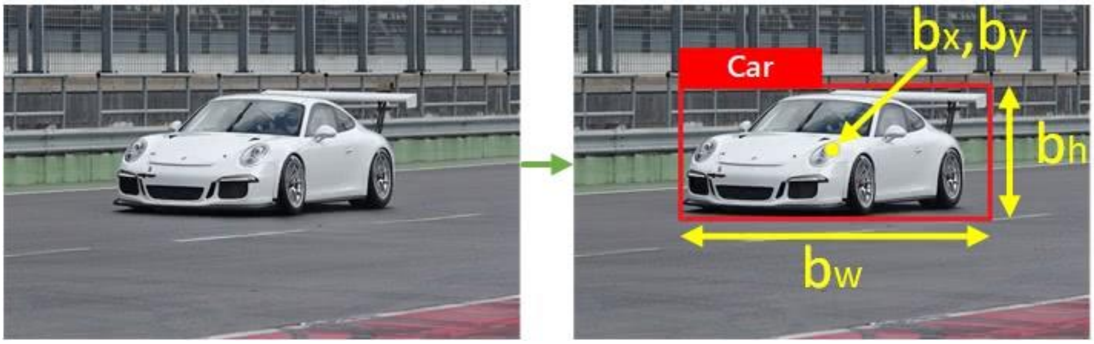


Figure 6-2: Identification model for a car

6.2.3 Annotation Standard

There is a total of twelve different types of events which are either fully annotated or partially annotated. The fully annotated videos have Thirteen event types labelled from 0 to 12 while the partial annotation has Seven event types labelled from 0 to 6. Event, activity, is represented as the set of objects involved with the temporal interval of interest, e.g. “PERSON loading an OBJECT to a VEHICLE” and “PERSON unloading an OBJECT from a VEHICLE”. All this is clearly shown in the recorded videos. A person or object are annotated if they are within the vicinity of the camera and the dataset stops recording a few seconds after the object is out of the vicinity of the camera.

The training dataset includes two sets of annotation files that describe a. the objects and b. The events depicted in the videos. Samples of the event annotation files and the object annotation files are shown in Table 6-1 and Table 6-2. These annotation files were generated manually and represent the ground truth. The training includes 66 videos representing Three scenes.

Table 6-1: Sample of VIRAT training dataset event annotation file

Event ID	Event Type	Event length	Event start frame	Event end frame	bbox X_lt	bbox Y_lt	bbox Width	bbox Height	Number of objects involved
1	5	172	3670	3841	670	454	267	228	2
2	5	217	10413	10629	985	406	209	204	2
3	2	66	10068	10133	891	357	202	128	3
4	6	131	9614	9744	983	399	226	211	2
5	6	112	15812	15923	1220	378	241	126	2
6	5	151	17522	17672	1253	380	198	126	2
7	5	217	10413	10629	985	406	209	204	2

Event Type: unknown=0, loading=1, unloading=2, opening trunk=3, closing trunk=4, getting into vehicle = 5, getting_out_of_vehicle = 6.

Table 6-2: Sample of VIRAT training dataset object annotation file

Object ID	Duration of object	Frame number	bbox X_lt	bbox Y_lt	bbox Width	bbox Height	Object Type
1	385	3495	157	659	76	132	1
1	385	3496	162	658	76	132	1
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
1	385	3838	747	498	73	97	1
1	385	3839	747	498	73	97	1
3	4732	0	613	469	254	189	2
3	4732	1	612	468	255	190	2
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

Bbox X_lt: Bounding Box Left Top Corner X-index

Bbox Y_lt: Bounding Box Left Top Corner Y-index

Bbox Width: Bounding Box Width

Bbox Height: Bounding Box Left Top Corner X-index

Object Type: type of object (Unknown=0, person=1, car=2, other vehicle=3, other object=4, bike=5)

6.3 Experiments setup

The experiment starts by dividing the VIRAT video dataset into two parts. The first part is used for training purposes and the second part is used for testing, 60% of the data has been used for training, and the remaining has been used for testing.

Each experiment starts by moving one event or two events, from the training dataset. For all seven experiments, the file name indicates the hidden event which has been moved to the end of the testing dataset, e.g. Event0, Event1, Event2 …, Event6.

The aim is to hide those events in training and present them in the testing to find out how the system has learnt other events.

6.3.1 Data preparation

A Matlab script has been developed to generate a file that combines information from VIRAT's object annotation files with corresponding information from VIRAT's events annotation files for each video file. A sample of the output file id shown in Table 6-3.

6.3.2 Evaluation

This evaluation is basically based on the documents from VIRAT DATASET RELEASE 2.0 accessed from VIRAT DATASET[4]. The VIRAT Video Dataset Release 2.0 is used in the analysis and evaluation of the data throughout this chapter.

The results of the HTM anomaly detection algorithm is represented by an anomaly score for each field; a field represents a movement. The anomaly scores vary between Zero and One. Where Zero represents a normal movement (ideally part of an event that has been learned) and One represents an abnormal movement. Values between

[4]http://www.viratdata.org/

Zero and One represent the anomaly score, where values close to Zero represent movements closer to normal ones and values closer to One represent movements that are closer to abnormal movements, i.e. suspicious.

Table 6-3: Sample of the generated data file

Reset	Event ID	Frame Number	Event Type	Object Type	Object ID	bbox X_lt	bbox Y_lt	bbox Width	bbox Height
int	int	int	Int	int	int	int	int	int	int
1	0	0	6	2	2	648	497	154	66
0	0	1	6	1	1	720	490	26	22
0	0	1	6	2	2	648	497	154	66
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
1	1	0	3	2	1	457	432	93	58
0	1	0	3	1	2	533	479	21	48
0	1	0	3	2	3	205	371	71	44

Precision and recall are commonplace measures in information retrieval. They are based on the comparison of an expected result and the effective result of the evaluated system. These results are considered as a set of items. As precision and recall are easily explained measures, it is useful to maintain the precision and recall structure when looking for new measures. This also ensures that measures derived from precision and recall e.g., F-measure. (Jérôme Euzenat, INRIA Rhône-Alpes Monbonnot, France).

First, the evaluation starts with the first scenario, for each record, the Precision, Recall and false-measure are calculated by comparing the resulted anomaly score with a threshold. If the anomaly score is less than the threshold, the detection is considered correct. In the case of an event that has not been shown in the training dataset, if the resulted anomaly score is greater than the threshold the result is considered correct. The True Positive, True Negative, False Positive and False Negative are considered as below:

- TP - the number of "true positives", positive Examples that have been correctly identified

- FP - the number of "false positives", negative Examples that have been incorrectly identified
- FN - the number of "false negatives", positive Examples that have been incorrectly identified
- TN - the number of "true negatives", negative Examples that have been correctly identified

This process has been repeated for threshold values between 0.1 and 0.9 with a step of 0.05 to find the maximum accuracy and hence identify the optimum threshold.
The Precision, Recall and F-measure formulas are shown below:

$$Precision = \frac{\text{True positive}}{True\ Positive + False\ Positive} \qquad \text{Equation 6-1}$$

$$Recall = \frac{\text{True positive}}{True\ Positive + False\ Negative} \qquad \text{Equation 6-2}$$

$$F - measure = \frac{2(\text{PrecisionxRecall})}{(\text{Precision+Recall})} \qquad \text{Equation 6-3}$$

6.3.3 Test Results

The table shown below explains the statistics of events presented in the seven experiments including the number of training and testing samples as well as the total number of samples for each hidden event.

Table 6-6-4: The hidden numbers of events

Hidden event	Number of training samples	Number of testing samples	Total number of samples
Event 0	59309	62726	122035
Event 1	60414	61621	
Event 2	61280	60755	
Event 3	57095	64940	
Event 4	58571	63464	
Event 5	47951	74084	
Event 6	32144	89891	

6.4 Evaluation

In this part of the experiment, an evaluation of the proposed HTM Cortical Learning Algorithm has been tested using the same dataset, Virat, to do a comparison of the performance metrics between each output of different machine learning technique.

Several anomaly detection algorithms are evaluated using Rapid Miner Studio version 8.2. Each model's anomaly score is normalised to the range 0.0 to 1.0. The higher the value is, the higher the likelihood of an anomaly occurring.

6.1.1 k-Nearest Neighbour Global Anomaly Score (kNN-GAS):

k-NN Global Anomaly Score algorithm (GAS) calculates anomaly scores using the nearest k neighbours. Ramaswamy et al. (2000) proposed that the outlier score as the average distance nearest k^{th} neighbour. The core algorithm spend 99% of the execution computing all K Neighbours while the rest of the time is used for storage. The algorithm results are shown in Figure 6-4.

Table 6-5 Performance metrics for the k-NN algorithm

	Precision	Recall	F-measure
kNN-GAS	72.2%	39.17%	50.79%

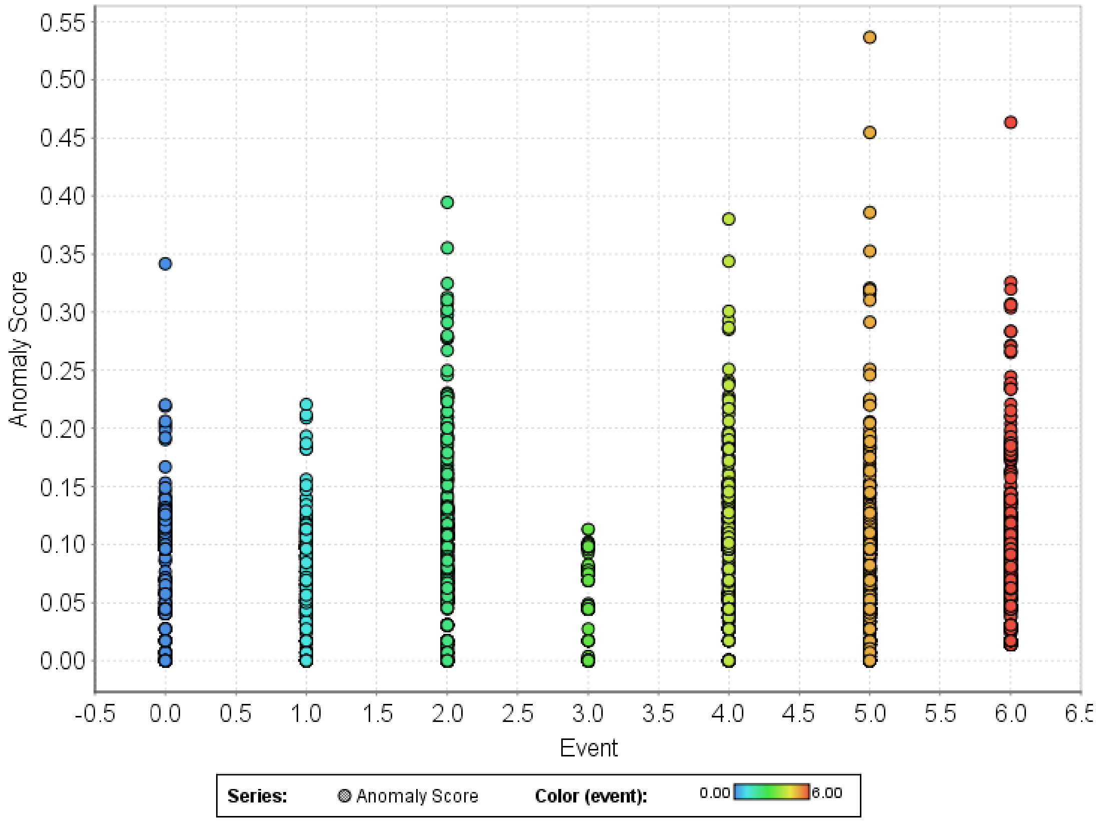


Figure 6-3: A sample result for the k-NN algorithm for event 0

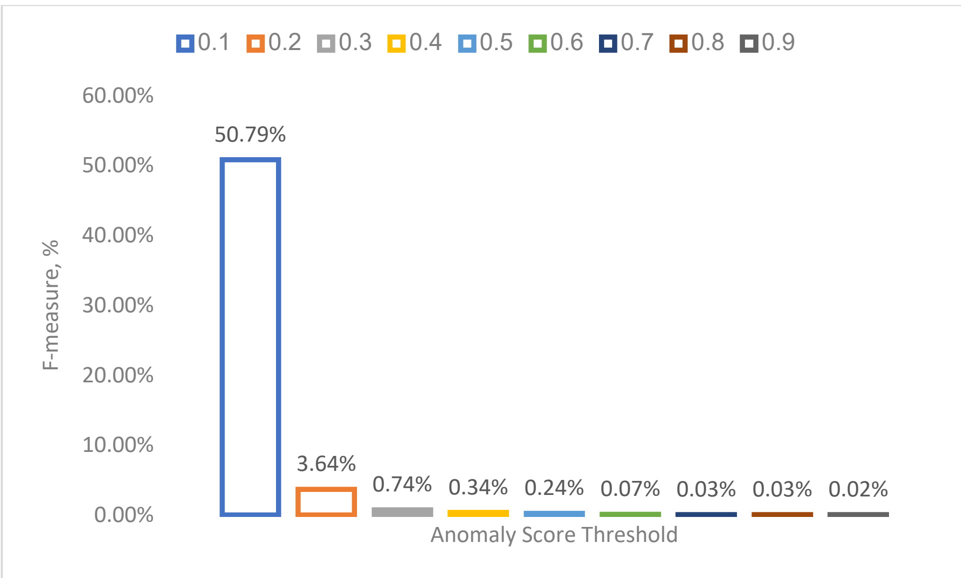


Figure 6-4 : Average F-measure for the k-NN algorithm

6.1.2 Connectivity-Based Outlier Factor

According to Maslov (2006), Connectivity-Based Outlier Factor (CBOF)algorithm determines performance of data using queries, but its effectiveness is affected by sensor-generated time sequences. CBOF is used to determine bounds for the k-Nearest Neighbour KL-CBOF algorithm. This algorithm determines the lowest and highest bound of the multivariate data. Amongst the different models available for testing bounds for the k-Nearest Neighbour, this analysis uses only CBOF models to focus on the changes in bounds. This algorithm involves the rearrangements of the order of the N time domain samples through the counting in binaries that have been flipped from left to right. After the bit reversal sorting stage of the "CBOF" algorithm, the next step is the finding of the frequency spectra which belongs to the 1-point time domain signals at the end of the last decomposition phase. This is a very easy process since the frequency spectrum of a 1-point signals is equal to itself, and therefore there is virtually nothing to be done at this stage. Also, it should be noted that the final 1-point signals are no longer time domain signals but rather, a frequency spectrum.

Table 6-6 Performance metrics for the CBOF algorithm

	Precision	Recall	F-measure
CBOF	39.67%	6.55%	11.24%

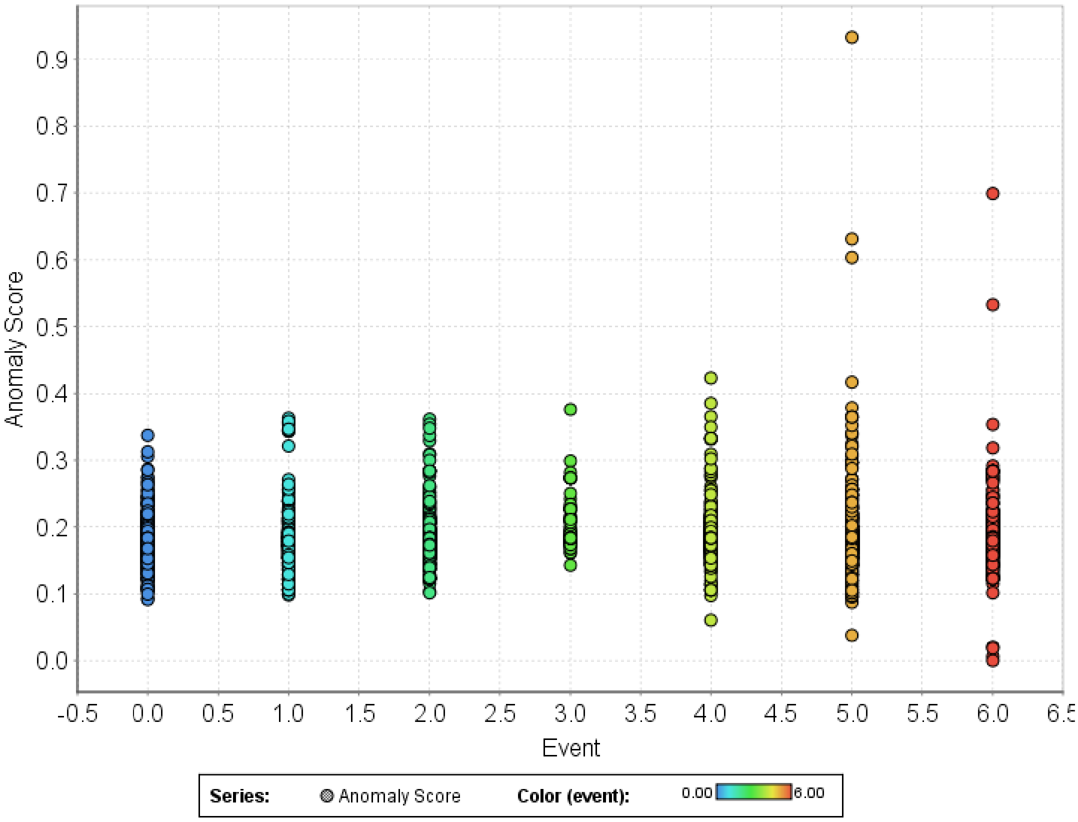


Figure 6-5 : A sample result for the CBOF algorithm for event 0

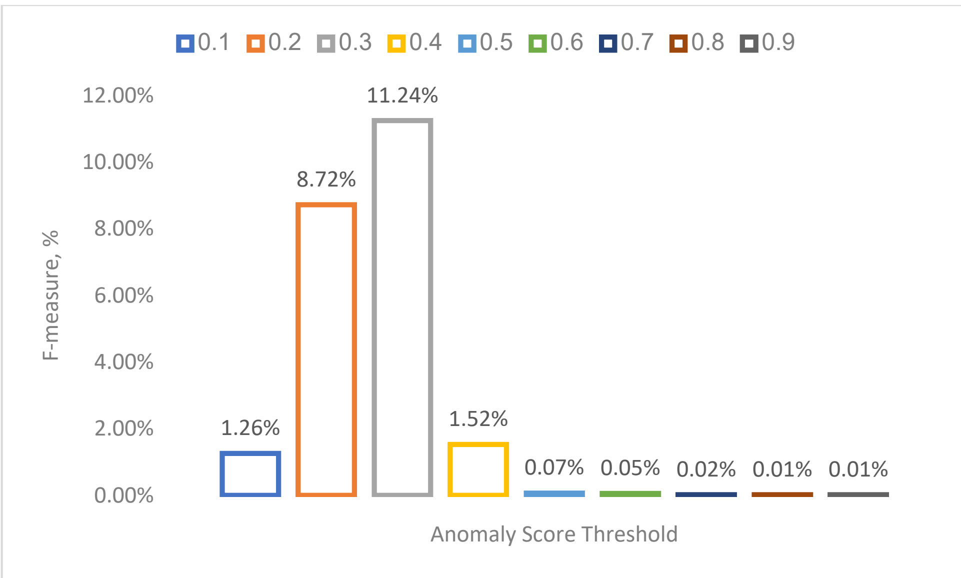


Figure 6-6 : Average F-measure for the CBOF algorithm

6.1.3 Singular Value Decomposition Influence Outlier (SVD-IO):

According to Breunig el al, 2000), COF is a modification of LOF algorithm is used to handle outliers that are not of low-density patterns. The outliers have an outlier score of more than 1. According to Tang *et al.*, (2002), there is no difference between COF and LOF algorithms except density calculation. Whereas LOF, the distances are determined using a hypersphere centred on a data point. The COF calculates the distance incrementally.

Table 6-7 Performance metrics for the SVD-IO algorithm

	Precision	Recall	F-measure
SVD-IO	54.42%	25.60%	38.42%

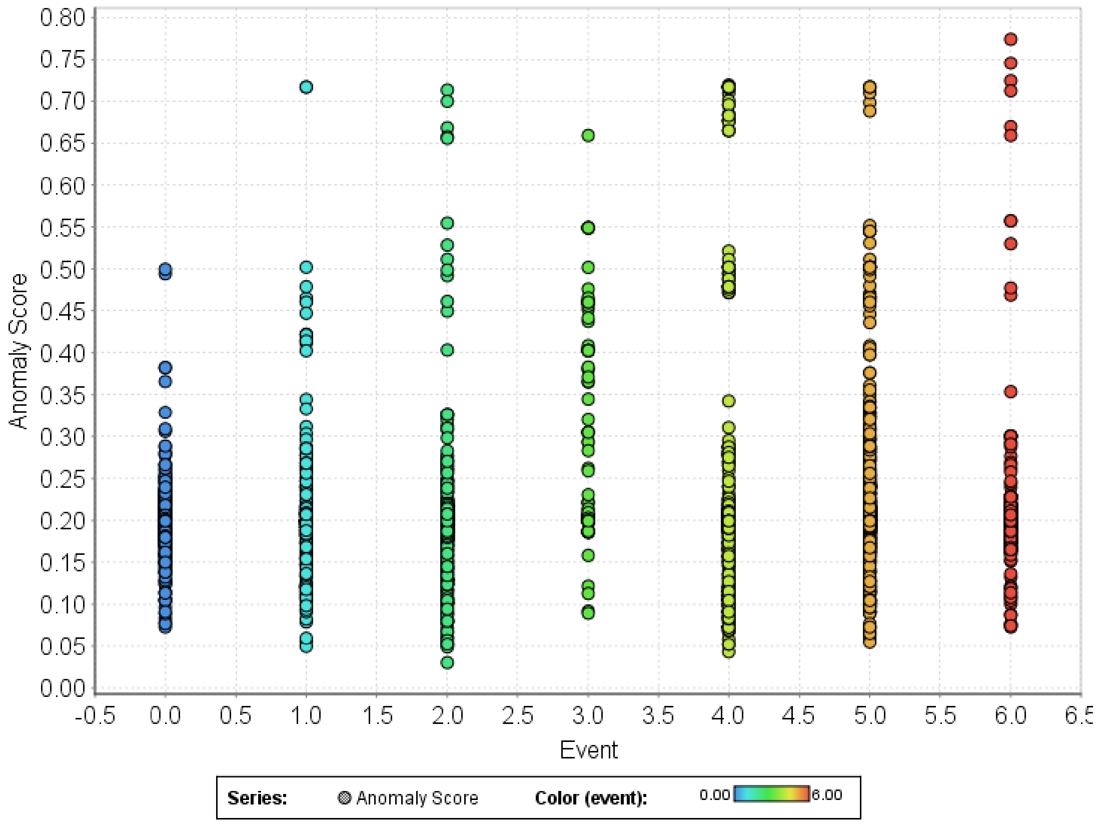


Figure 6-7 : A sample result for the SVD-IO algorithm for event 0

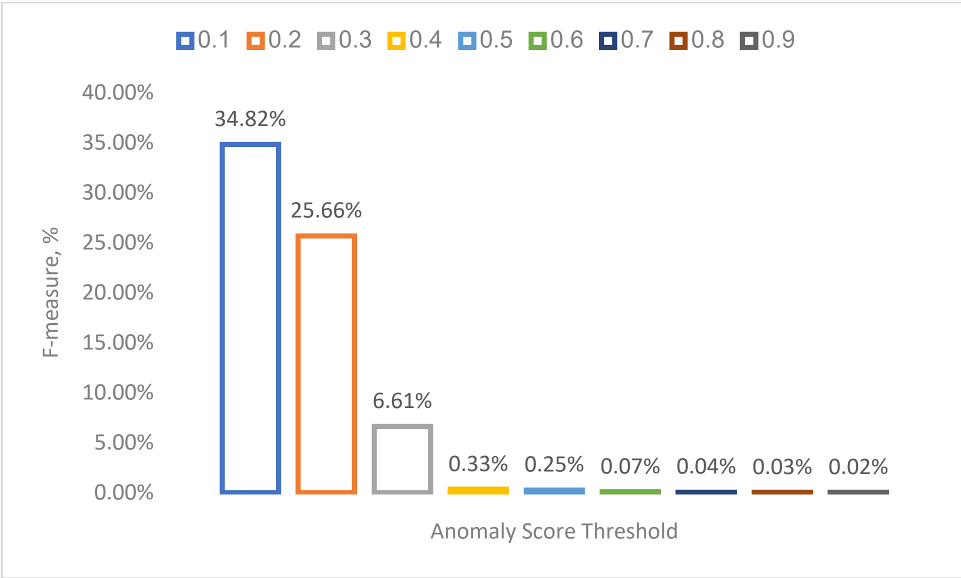


Figure 6-8: F-measure for the SVD-IO algorithm

6.1.4 Independent Component Analysis - Local Outlier Probability (ICA-LoOP):

According to Comon (1994), Independent Component Analysis (ICA) is an algorithm that is used to compute hidden factors within sets of statistical data. The algorithm generates a model that will be to hidden factors of multivariate data that is data from the big database. The model generated by the algorithm is assumed to linear and has unknown latent variables. This is usually applicable to Audio Processing, Array processing and medical data analysis. Most of this data is non-gaussian and mutually independent, thus fit for ICA LOP. The following shows Independent Component Analysis of a signal.

Table 6-8 Performance metrics for the ICA-LoOP algorithm

	Precision	Recall	F-measure
ICA-LoOP	42.58%	42.93%	42.75%

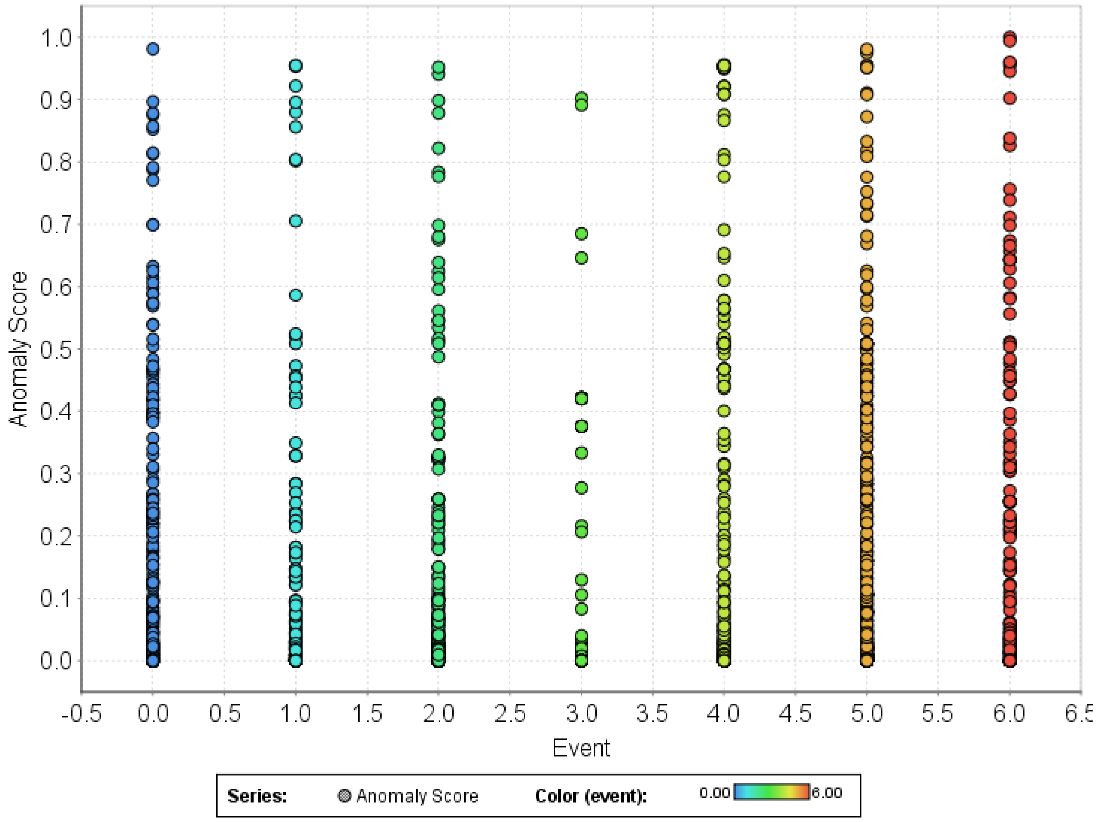


Figure 6-9 : A sample result for the ICA-LoOP algorithm for event 0

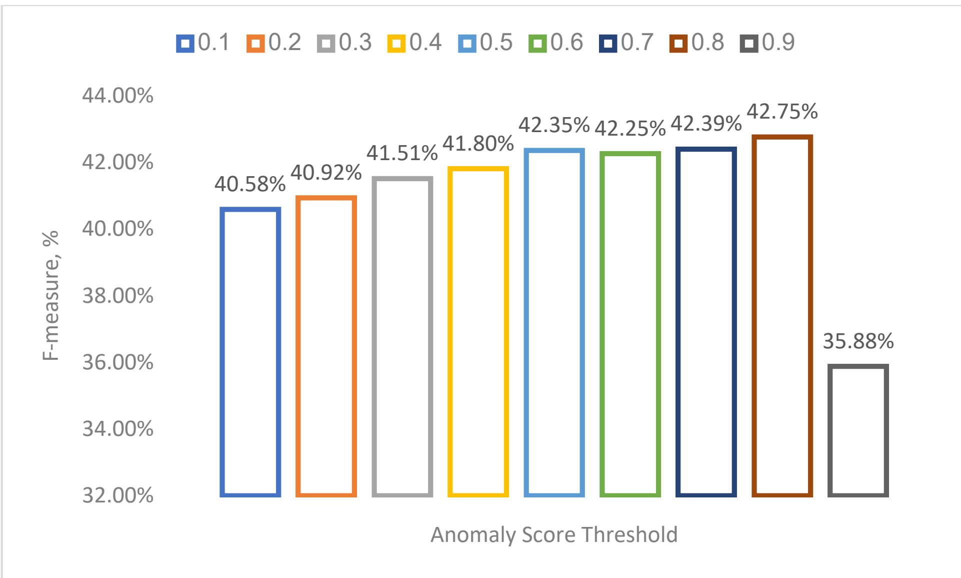


Figure 6-10 : Average F-measure for the ICA-LoOP algorithm

6.1.5 Proposed Algorithm

The proposed HTM algorithm has been run with the datasets to evaluate its performance. The results are shown below:

Table 6-9 Performance metrics for the proposed algorithm

	Precision	Recall	F-measure
Proposed Algorithm	61.54%	71.84%	66.29%

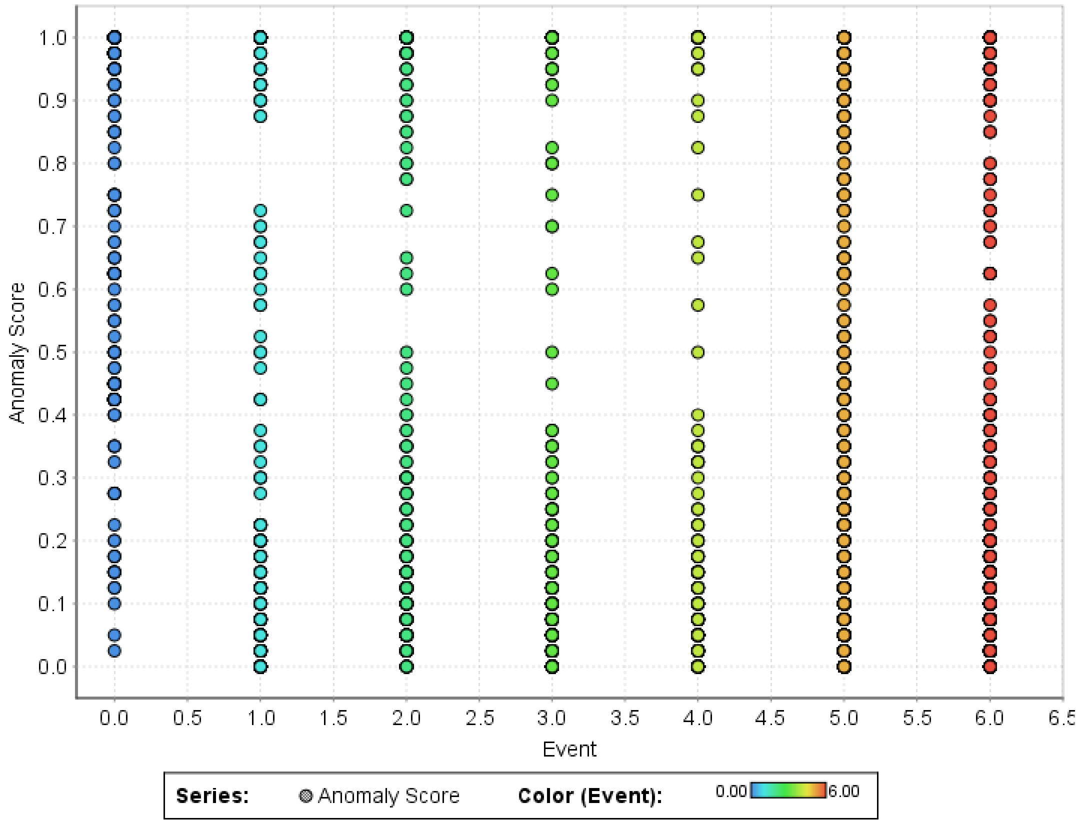


Figure 6-11 :A sample result for the proposed algorithm for event 0

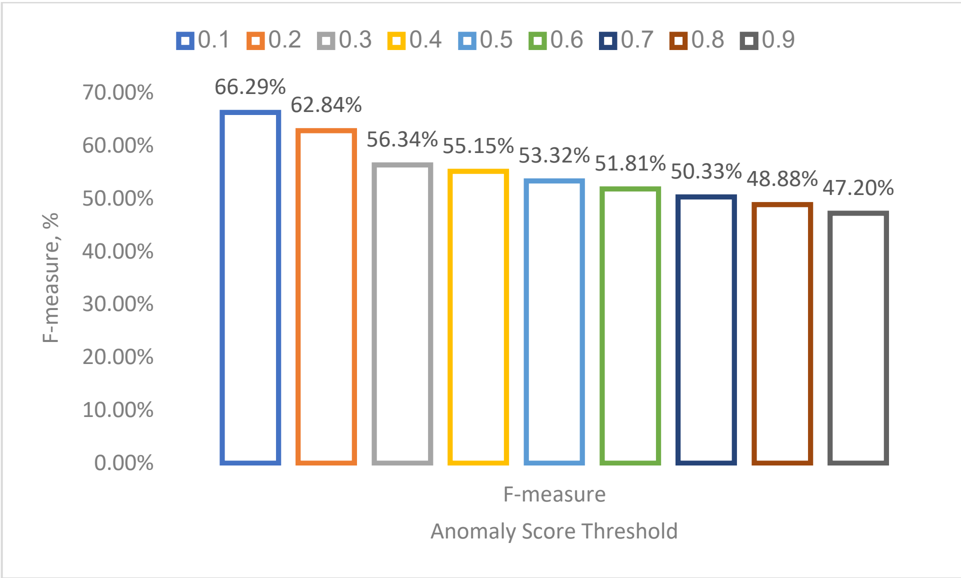


Figure 6-12 : Average F-measure for the proposed algorithm

6.2 **Objective Evaluation**

In this section, the objective evaluation metric defined in Section 6.2 is used to compare and rank the proposed algorithm with the state-of-the-art anomaly detection algorithms. These results complement the illustrative visual comparison results in the previous Section 6 and those presented in Chapter 5. The evaluation of the accuracy of anomaly detection is carried out using Precision, Recall and the overall metric F-measure. The optimal threshold values obtained from the analysis carried out in Appendix B.

Tests applied on twenty-three videos have been conducted to detect movement anomalies in different scenarios. Additionally, in this study, the proposed algorithm has been evaluated and compared against several state-of-the-art anomaly detection algorithms. The proposed algorithm has achieved 66.29% average F-measure, with an improvement of 15.5% compared to the k-Nearest Neighbour Global Anomaly Score (kNN-GAS) algorithm. The Independent Component Analysis-Local Outlier Probability (ICA-LoOP) scored 42.75%, the Singular Value Decomposition Influence

Outlier (SVD-IO) achieved 34.82%, whilst the Connectivity Based Factor algorithm (CBOF) scored 8.72%. The proposed models, which are based on HTM, have empirically portrayed positive potential and had exceeded in performance when compared to several other algorithms.

Table 6-10 Performance metrics for all algorithms

	Precision	Recall	F-measure
Proposed Algorithm	61.54%	71.84%	66.29%
kNN-GAS	72.2%	39.17%	50.79%
ICA-LOP	42.58%	42.93%	42.75%
SVD-IO	54.42%	25.60%	34.82%
CBOF	39.07%	4.91%	8.72%

Table 6-11:Comparisons of performance accuracy, precision, recall, F-measure

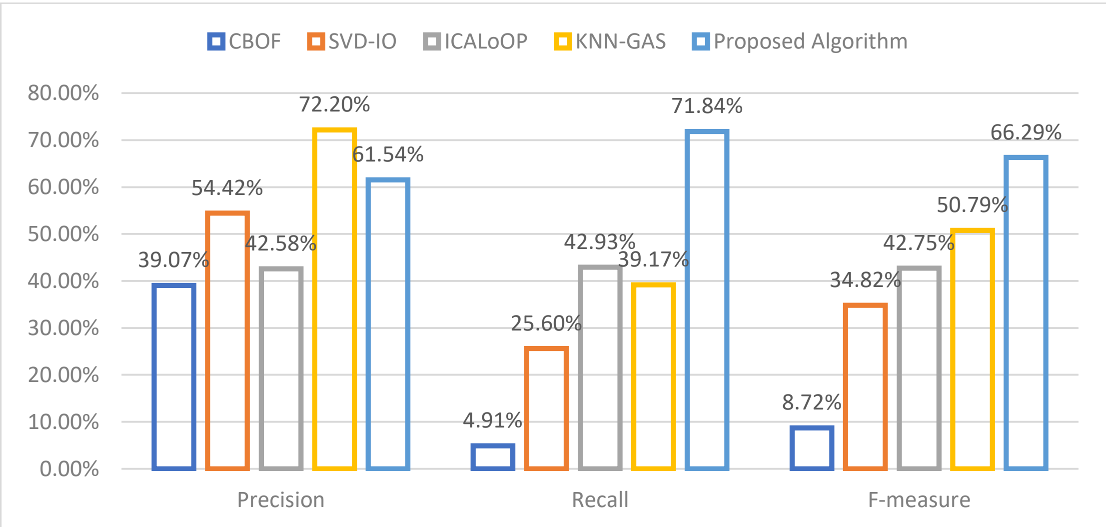


Figure 6-13 :Comparisons performance accuracy, precision, recall, F-measure

6.4 Conclusions

Currently, the discovery of what is happening in a scene can be seen by automatic scrutiny of activities included in a video. The literature shows, different algorithms that have been proposed to identify a solution to the movement classification problems. However, the required performance of such algorithms differs depending on the target scenario, and on the characteristics of the monitored scene.

Due to the diversity of video surveillance scenarios and the increasing development of movement classification algorithms, an automatic assessment procedure is desired to compare the results provided by different algorithms. This objective evaluation compares the output of the algorithm with the ground truth, obtained manually, and measures the differences using objective metrics. There are various datasets for activity and human action recognition, though older datasets provide limited ground truth classification to manual annotation at a simpler level, most of the modern datasets, in this case, VIRAT Video Dataset, gives high-quality ground truth.

This chapter has tested the proposed movement classification algorithm and evaluated its accuracy. Several experiments have been carried out to calculate the optimum anomaly threshold for each algorithm. the average achieved average F-measure for the proposed algorithm was 66.29%, with an improvement of 15.5% compared to the k-Nearest Neighbour Global Anomaly Score (kNN-GAS) algorithm.

Chapter 7

Summary, Conclusion, Limitations and Future Work

7.1 Summary

Automatic analysis of videos for forensic purposes has been subject to technology evolution. There have been various developments over the subjects that have compromised human error in post-incident analysis. The concept of video analytics is more significance due to the increased number of devices that are capable of recording and producing videos, increasing the probability of video evidence that can be used for a trial.

Video forensic aims to retrieve CCTV evidence by extracting video information from different sources, digital and analogue. This can be achieved by introducing intelligent tagging of data for fast analysis, review, and the guarantee of evidential quality data at every part of the process to deliver evidential quality information as and when required.

The UK legal system has been investing heavily in CCTV cameras hoping to capture evidence that will help in post-incident analysis. However, in most cases, the analysis has been challenging due to the ostensible fact that while technology has improved, there has not been a standard post-incident analysis framework that incorporates such technologies. Therefore, there is a need for an automated video forensic investigation tool and a proper development of a framework that can address the sensitive issues associated with this application domain.

As one of the most important applications of image processing, understanding and computer vision, computer-based video analytic has recently received significant attention, particularly during the past decade. The need for analysing video information captured by CCTV cameras has grown and is becoming a crucial issue. This growth is augmented by the extensive range of commercial and law enforcement applications, and by advances in signal processing and computer vision techniques, along with the rising uptake of CCTV installations, and the increasing availability of cheap computational power which makes real-time systems for video processing feasible. Video analytic makes the entire process of intelligence gathering, and evidence is building quicker and more efficient. Video analytic technologies are used for intelligence and analysis during an active investigation, presentation in an interview room or at trial, or as part of appeals process or judicial review.

With the use of an automated video analytic tool, searching through 24 hours of video data for events that take 7-8 hours from an investigator can be done in 25-45 minutes. For instance, using an automated video analytic tool, in 2-3 hours, an investigator can filter through 30 days of video footage to find and receive a continuous video clip of all qualifying video data. This same search without video analytic would require sifting through 720 hours of video manually.

Classifying the movements of objects detected from a video feed is a key module to systematize or rather automate the forensic video process. Machine learning techniques have been heavily proposed to solve the problem of movement classification. However, they still suffer from various limitations such as their limited ability to learn data streams or data with temporal behaviour. Recently, new bio-inspired machine learning techniques like HTM and its implementation CLA has been proposed, which is inspired by the neocortex. HTM offers a better understanding of how our brains function. This report shows that CLA would make movement classification more effective and minimize the hustle involved in collecting evidence from CCTV footages. With this in mind, the project aimed to study the requirements of video forensic investigation and Police procedures to recommend a new semi-automated post-incident analysis framework and to explore the application of the CLA in movement classification. This research fundamentally focused on finding answers to the following questions: Can the CLAs contribute to a solution for the movement classification problem? And What are the main features of a semi-automated post-incident analysis framework for video forensic?

7.1.1 Video Forensic

Video Forensic is an important part of digital forensics which can be defined as the application of analysis and examination procedures to collect, and reserve evidence from relevant computing devices in ways that are seemingly suitable for presentation in a law court. The predominant aim of the entire process of video forensics, therefore, is to perform a vigilant and thorough investigation while simultaneously maintaining a chain of evidence that shows who exactly is to be blamed for a certain crime and what transpired before, during, or after the crime. As such, video forensic tools are usually used to respond to an incident in cases of crime investigations.

The process of investigating video footages usually involves much manpower. At each point in time, at least one investigator is supposed to be physically viewing such video footage on a screen and manually analysing it if the need arises. This is a time-consuming operation, and it is very costly, because, the staff involved in the investigation needs to be paid for the services rendered. Also, human beings are easily distracted and sometimes inefficient. Considering an investigator who is manually viewing one screen showing two video footages, after 10 minutes, the investigator may miss 45% of the events, and after 22 minutes, the investigator may miss 95% of the events. The big challenge comes in when during the mess that occurs due to the inefficiency of the investigator leads to an event that goes unnoticed. Thus, there exists a need for proper consideration of research in the direction of the automated video forensic investigation and a proper development of a framework to handle such sensitive issues.

The Abstract Digital and the Integrated Digital Investigation Model popularly called IDIP organized the digital forensic process into five phases: Readiness phase, Deployment phase, Physical Crime Scene Investigation phase, Digital Crime Scene Investigation phase and the Review phase. All these phases have specific roles to play in ensuring reliable digital data forensic evidence. This proposal was later enhanced and came up with Enhanced Digital Investigation model EIDIP, which separates the investigations at the primary and secondary crime scenes while depicting the phases as iterative instead of linear.

The review of the video forensic literature shows that the link between computer vision algorithms capabilities and requirements of video forensics as an application of video analytic is still not clear. There is, therefore, a need to have a better understanding of capabilities of video analytic techniques versus requirements of video forensics. This thesis has introduced a generic model for a video analytic system and has reviewed the practical requirements of video forensics and has related the capabilities of video analytic techniques to video forensics' requirements.

7.1.2 Proposed Semi-Automated Video Forensic Framework

The automatic analysis of videos saves forensic analysts from the need to review innumerable long videos for evidence analysis. This is accomplished via the use of automated forensic digital analytic tools (AlShaikh and Sedky, 2016). However, the use of these digital analytic tools requires the establishment of digital policies into the new forensic structures. Even though the area of video forensic is attracting great attention, researchers should be awakened of the need to relate this area to evidence analysis and presentation to the court of law.

Over the years, there has been a wide range of use of physical forensic evidence. In most cases, it involves the collection of digital evidence especially as it relates to scenes of different crimes. This idea has been presented and published in some frameworks in digital forensics.

To capture the requirements of Police investigators, a survey questionnaire was used. The questionnaire contained 26 questions that were divided into Three categories. The first category can be considered as the screener. It contained Ten questions aimed at introducing the respondent and gauging their experience in handling crime and crime investigation. The second category as well contained Ten questions. The rationale behind these questions was to explore the problems and challenges customarily faced when using an automatic system to review CCTV footage, tools, equipment and software used for evidence in video forensics. Category Three contained Six questions that explored the importance of the features in video forensic tools and to an extension, the effect that the tools would have when used as automated forensic tools in reviewing video evidence. The questionnaire was shared through the Survey Monkey and was answered by Police officers, CCTV investigators, technicians, engineers and others in the Kingdom of Saudi Arabia. The responses from the questionnaire have been analysed and have informed the proposal of a new semi-automated post-incident framework that could assist in using the output generated from a video analytic tool as evidence in the court of law.

Before moving to the proposed framework, below is a brief review of the results obtained from the questionnaire. Firstly, out of the 31 responses received, 96.8% of the participants asserted that there is an upsurge in the usage of CCTV over other sources of recording evidence and in fact favoured its use in crime evidence collection. Circa 67.7% of the respondents indicated that manual viewing of CCTV footages is time-consuming and exasperating. 100% of the participant indicated to have used video sharpening as a video enhancement technique, 80.6% admitted having ever used filters, and only 6.5% (Two respondents) had at one point used masking, showing that video sharpening is one of the most utilized video enhancement technique and can be used even in a semi-automated framework. 96.8% of the participants affirmed to have undergone image content analysis training to help them accurately identify a person or object the process of analysing CCTV footages while only 3.2% had speech science training. While the survey revealed that some of the people involved in crime investigation use automated video analysis process, it emerged that most of the times, files get corrupted thus hampering the analysts' ability to conduct a comprehensive and objective video analysis – 48.8% of the respondents identified corrupt files as a major problem, followed by missing files at 25.8% and compressed size with 19.4%. Most of the respondents did not consider data duplication as having any negative effect on the video analysis process - only 6.4% of the responses identified it as a problem. Similarly, with regards to the aspects that people are concerned with in any tool that is designed to be used in forensic video analysis, it emerged that 41.9% are concerned with movement detection, image enhancement was at 32.3 %, while 19.4 % opted for video stabilisation. Since most people are interested in movement detection, the proposed framework considered this as a way of ensuring that it meets the demands and preferences of the framework users. The need for an automated video analytic framework was assessed by question Twenty-Six in the question. It emerged that Twenty-Eight of the Thirty-One believed that it is a high time to have automated video analysis tools. This support for automated forensic tools is overwhelming and can, in fact, be attributed to the paucities of manual video forensic analysis. They supported the development of automated video analysis tools as they consider it to be leeway to better video analysis and the presentation of correct evidence in the courts of law. These responses significantly assisted in the development of a new video forensic methods and also prompted the enhancement of the prevailing ones.

The proposed framework can be summarised as follow:

Evidence collection: DVRs and NVRs serve as a source of the input signal. The recording devices have to include entrenched proof of validation. The DVRs and NVRs will be used to get video evidence the strategically placed CCTV. DVRs and NVRs were selected as the spring of input evidence because they cannot be influenced, changed or manipulated (AlShaikh and Sedky, 2016). Any attempt to alter them will easily be noticed. This ensures that the recordings to be used in the analysis are authentic – free from any human manipulation.

Hard Disk Drive (HDD) cloning: Cloning the HDD means making an exact copy of the evidence collected in step 1 and saving to other storage media. This is done to avoid issues of manipulating the original evidence on the original HDD. With a cloned HDD, various experiments can later be performed without impacting on the evidence on the original HDD – this hints that experiments are done using the cloned HDD.

Video extraction: Retrieving the required evidence from the cloned HDD is carried out in this step. This is done to access the content of the cloned HDD. DVRs and NVRs can archive videos to a USB flash drive, external Hard Disk Drive (HDD), or other storage devices (Carner, 2013). However, the archiving must have been done properly so that retrieving the information becomes easy. When it is easy to extract the videos for forensic examination, the entire video analytic process will be fast-tracked as the necessary evidence will be obtained as and when required by the Forensic Analysts.

Video conversion: The video and audio must be converted to the right format for both viewing and analysis.

Most of the DVRs and NVRs store videos in their proprietary formats (AlShaikh and Sedky, 2016). There is every need to convert the collected evidence from the DVR/NVR proprietary format to a standard format to apply the video analytic tool.

Requirements capturing: Police officers define the events to be detected from the video footages, by specifying the area(s) as well as events that they want the video analytic tool to take into consideration. For instance, the police may command the video to focus on a suspect in a black suit and a yellow hat for a certain duration or simply concentrate on a suspect moving towards the south or north.

Automated video analysis: This step automates the investigation process. A video analytic tool can generate a report detailing the start and end of each event. The list of events that happened per day, or per week, or per month, or even per year will be appropriately reported. This must be done cautiously, and with much care since if the video analytic tool fails to properly scrutinise the evidence, the analytic process will be extraneous as important evidence that would have helped in steering the case in a court of law will be missed.

Manual verification: In order not to take laws into their hands, the police officers manually verify and select relevant events from the events detected by the video analytic tool. In the proposed framework, the work of the police will be easier as instead of swotting hundreds of footages; they will become simply the video analytic tool for footages that are relevant to the crime under consideration.

Building a storyboard: An investigation storyboard gives a sequence of events, typically with some directions and dialogue. Since the proposed framework has a storyboard, it will be easy for the Forensic Analysts to decipher the events in a crime scene by reviewing the videos and lip-reading the persons in the footage, if the moving objects are people (Carner, 2013).

Report generation: A report of the results of the analysis should be written down. This report should be comprehensive enough to inform those reviewing the results of the forensic evaluation. It should, accordingly, include issues like actions taken, why such actions were taken, findings made from the actions taken and recommendations for improvements to policies, guidelines and other aspects of forensic process amongst other issues.

Court presentation: The evidence recorded and stored on a DVD must be presented before the court handling the criminal case under review. With the forensic evidence in the DVD, the court will be certain that the forensic expert and/or police officers are not formulating stories and will be at liberty to countercheck the forensic expert and/or police officers' assertions against the evidence that was captured by the CCTV. This will be helpful and a step towards justice delivery.

7.1.1 Movement Classification

Movement classification is one of the most important areas in video analytic. However, manually detecting, classifying and analysing moving objects do not guarantee outright precision. It is not so easy interpreting every activity correctly when considering the real environment and trying to relate the way objects interact in surveillance covered area.

The challenges posed by defining and classifying objects' behaviours as normal or abnormal movements can be tackled using video analytic technologies (Savoldi and Gubian, 2009). The objective of video analytic technologies is to detect the presence of objects that are moving in the camera's field of view and classifying their movements for security, traffic monitoring and safety applications. There are many hurdles faced by video analytic systems that impede their ability to perform accurately (Carner, 2013). This study presented a review of movement classification techniques and algorithms, which can tackle the challenges of realistic and practical outdoor surveillance scenarios.

There has been much research in using AI and NN techniques to solve problems of movement classification. However, AI and NN have shortcomings of classifying what is abnormal based on the training it previously acquired from the inputs. However, bio-inspired computational models have proved to be successful over conventional methods, e.g. deep neural networks outperform traditional object classification methods.

The review of movement classification concluded that, previous research in building intelligent machine similar to human capabilities remains limited, even though literature evidence indicates that despite series of effort made by AI researchers and recent effort made by artificial neural network researchers to build viable algorithms for achieving human-like performance suffers fundamental flaws, as all the existing solutions fail to adequately address what intelligence is or what it means (Carner, 2013).

It was also concluded that more efforts are required to address how the brain works such as remembering fast events. Although human brains are made of neurons - the brain is a neural network -, an understanding of how neurons interact, which will lead to the emergence of properties of intelligence, remains a problem that was unsolvable with AI. The implication of this is that replicating the correct connections between populations of neurons remains a challenge (Savoldi and Gubian, 2009). The HTM theory is predominantly intended to accomplish predictive abilities that are inspired by the intelligence of human beings. The theory is used based on the hierarchical and Neocortex structure of the regions that are in it. The hierarchical approach used in classification and recognition are mainly popular based on hierarchical neural network and the modern recognition models.

Even though the HTM networks usually differ from the classic computing, the general function of the computers always model them. Moreover, the hierarchical organization of the HTM networks in most cases facilitates the improvement in efficiency. The usage of the memory and also the training time is reduced. Therefore, at each of the hierarchy levels, the patterns learned are always reused. Also, the prediction of the HTM usually determines whether there is the expectation of a new input since each of the HTM regions usually detects the sequence pattern of the storage because the stored sequences always match with the stored inputs (Savoldi and Gubian, 2009).

7.1.3 Application of Cortical Learning Algorithms in Movement Classification

This thesis argues that understanding and correctly modelling some important properties of all the regions of cortex in the human brain will provide a formidable, intelligent machine that covers all human capabilities and proposes a novel bio-inspired movement classification technique that tends to achieve an efficient and effective automated video forensics analysis work.

Under the HTM networks, the aspect of time is quite necessary for predicting, learning and also for inference. The aspect of time is considered essential since there is nothing which can be inferred from the sensory reading without making use of its history of time. Hence, the HTM is most suitable for learning during training when using a stream of data (Investigation, 2008). Different versions of the learning algorithms of the HTM have been implemented, whereby, in this research, the CLA has been used in processing data through online learning without supervision and also dealing with the nonaligned sequential data.

The algorithm description limits its capabilities through fixing the output values at single nodes. The hierarchical structure that is associated with the HTM algorithm has some significant effects on the speed used in processing (Savoldi and Gubian, 2009). There are many patterns that have been discovered in the process of training data, but the processing speed associated with the single pattern can grow linear. The number of patterns that need to be checked affects the processing speed (Savoldi and Gubian, 2009; Carrier and Spafford, 2004). However, there is usually a relationship that can have been existing between the speed of processing and number of patterns.

This work has studied the CLA through the investigation of the possibility of applying it to movement classification, to develop a novel movement classification technique.

A novel bio-inspired movement classification technique based on the theory of the HTM has been proposed. This classification technique is usually used to classify the normal activities from the abnormal ones.

The evaluation was grounded on data obtained from the VIRAT DATASET RELEASE 2.0. The data were divided into training and testing groups; the detection algorithm ranged between Zero and One where algorithms close to one represents movements that are closer to abnormal while those closer to zero represent movements that are almost normal. The HTM algorithm was compared against the anomaly score threshold for various events. For accuracy event Zero, the best optimal threshold for the HTM algorithm was 90% to 95%, 90% for event One, 85% for event Two, 70% for event Three, 80% for event Four, 20% for event Five, and below 10% for event Six. For the Seven experiments that were carried out to calculate the optimum anomaly threshold, the average achieved accuracy is 85%, and the optimal threshold is 0.714. Considering that event Five and Six had a low number of training samples and a higher number of testing samples, it is clear that the HTM algorithm is likely to detect movements as abnormal when recalling them is grim and has a high accuracy in detecting normal movements. From this, it was concluded that the intelligence

7.2 **Conclusion**

In chapter two of this project, several related problems that arise through the use of forensic evidence were reviewed, several proposed frameworks were identified as well as the development of structured frameworks for video forensic investigation. Through the findings in Chapter Two, the project developed in Chapter Three, by proposing a semi-automatic video analysis framework. In Chapter Four, the different methods of video analytics have been noted to be facing certain challenges. Towards the development of a video analytic system, the main functions that the system is expected to perform include the classification of the detected objects, tracking of the detected objects over time, and classification of the object's movements. This discussion of the HTM in chapter Five is determined by the ability of neocortex to be understood through the theoretical framework provided by HTM. In Chapter Six, it is evident that the required performance of different algorithms differs based on the target scenario as well as the characteristics of the monitored scene.

Overall, digital evidence must be properly admissible, precise, authenticated and accurate to be accepted in the court of law. Because of the fragile nature of digital evidence, the process must be handled properly and carefully. A detailed digital forensic process provides important assistance to forensic investigators in gathering evidence admissible in the court of law, thus, there is need to have a standard guideline for investigators. The use of video analytic tools for post-incident analysis dictates the application of well-established digital forensics rules into the new video forensic framework.

This research introduces a new semi-automated post-incident framework that could assist in using the output generated from a video analytic tool as evidence in the court of law. The proposed framework can be summarised as follow: Evidence collection, Hard Disk Drive (HDD) cloning, video extraction, video conversion, requirements capturing, automated video analysis, manual verification, storyline creation, report generation and court presentation.

Notably, the proposed semi-automated post-incident analysis framework is closely related to the existing frameworks for various reasons. Firstly, the proposed framework relies on recording devices – cameras (Investigation, 2008). The implication of this is that just like other frameworks, it will be affected by the Crime Scene Investigation (CSI) effect (Zjalic, 2017). This implies that the quality of the footages will significantly depend on the configuration of the CCTV system. If cameras used, have a higher resolution, upon zooming, finer details about the moving object will be obtained. However, if the resolution of the used camera is low, the proposed semi-automated post-incident analysis framework will not provide any significant evidence since the moving objects will be blurred. This is common in many forensic video evaluation frameworks.

Moreover, the proposed semi-automated post-incident analysis framework comprises many of the ideas contained in the physical crime scene investigation framework proposed by Reith et al. this model is commonly called the RCG02 framework. For instance, the forensic evidence collection phase and the presentation phase are very similar (Carrier and Spafford, 2004). However, the other phases of the proposed framework are somewhat different from those in the existing framework as they are more intuitive and also provides flexibility as the Forensic Analyst may decide how to store the evidence in DVDs or other storage devices such as flashbacks.

The current framework may also be compared to the existing frameworks by object identification (Carrier and Spafford, 2004). In the existing frameworks, the Forensic Analysts obtains the footage of an unknown object and tries to parallel it against a standard reference. As such, it can be hypothesised that in the existing frameworks, the video forensics involves looking for evidence, pinpointing it, and then reconstructing the events (Carrier and Spafford, 2004). This is summed by the figure below:

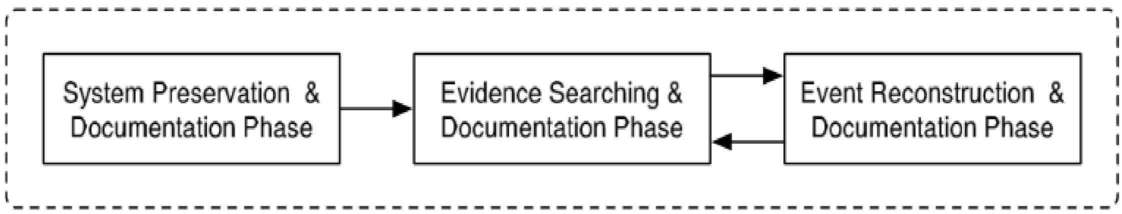


Figure 7-1: Sequence of the existing digital forensic investigating framework (Source Carrier and Spafford, 2004: 6).

The proposed semi-automated post-incident analysis framework makes the process more logical by evading the reconstruction of events and instead replacing it with automated analysis of footage of interest. This makes it easy to deduce meaning from the forensic investigation process, an aspect that made it very comparable to the MXSERVER System that was developed by Tygart Technology and is currently being used by law enforcement agencies in various parts of the world (Tygart Technology Inc., 2014). The value proposition of MXSERVER is similar to that of the proposed semi-automated post-incident analysis framework – save analysts and investigators from the hustle that comes with having to view a myriad of images and videos to pinpoint and extract the videos that capture a person of interest (Tygart Technology Inc., 2014). Per se, MXSERVER automatically classifies captured images so that the moving objects – presumably people – can be identified via facial recognition. MXSERVER rapidly goes through millions of footages in the databases and brings up those of interest so that Forensic Analysts can easily view and preview various characteristics of the person of interest (Tygart Technology Inc., 2014). This makes it very comparable to the proposed semi-automated post-incident analysis framework.

Because this project was inspired by the lack of a standard framework for video forensic, the proposed semi-automated post-incident analysis framework fills a significant gap in research. This framework will give Forensic Analysts an easy task when evaluating videos for forensic evidence.

This research studied the application of cortical learning algorithms to movement classification and reviewed HTM-based techniques developed for the efficient automation of post-incident analysis and video forensics.

The proposed movement classification algorithm has been tested using Virat dataset. Tests applied on twenty-three videos have been conducted to detect movement anomalies in different scenarios. Additionally, in this study, The proposed algorithm has been evaluated and compared against several state-of-the-art anomaly detection algorithms. The proposed algorithm has achieved 66.29% average F-measure, with an improvement of 15.5% compared to the k-Nearest Neighbour Global Anomaly Score (kNN-GAS) algorithm. The Independent Component Analysis-Local Outlier Probability (ICA-LoOP) scored 42.75%, the Singular Value Decomposition Influence Outlier (SVD-IO) achieved 34.82%, while the Connectivity Based Factor algorithm

(CBOF) scored 8.72%. The proposed models, which are based on HTM, have empirically portrayed positive potential and had exceeded in performance when compared to several other algorithms. Such performance indicates that there is a chance that the technique will be a success, in the context of video forensic.

The conclusions, from this project, are stated below:

1. The proposed bio-inspired learning technique is suitable for online learning and prediction of a sequence of movements,
2. The technique can be presented with slightly erroneous copy of sequence and can recover quickly after unexpected or suspicious movement patterns.
3. It is also capable of predicting a part of a sequence; this would be a desirable property since real-world data is likely to be noisy and dynamic,

This report concludes that Neocortex inspired learning techniques are applicable for movement classification aspect of analysis and can be used as a part of a video forensic framework. The proposed technique has been evaluated against state-of-the-art machine learning techniques such as HTM, KNN-GAS, ICA, CBOF, SVD-IO, Logistic Regression, Fully Connected NN, SVM, Random Forest works best when it is easy to detect the sequence of events (normal movements).

7.3 **Research Contributions**

This research project offers some contributions to knowledge as follows:

- The major contribution of this research is the development of an innovative movement classification technique based on HTM theory.

- A new post-incident analysis framework tagged semi-automated video forensic framework has also been proposed to assist in crime investigation and court evidence.
- A significant revelation of knowledge gap through critical literature reviews relating to police practices for video forensics, existing video forensics frameworks and state-of-the-art movement classification techniques. An addition to the information based on current standards for video forensics used in the UK and Saudi Arabia which is limited. This offers insight and opportunities for further research in the field of forensic investigation.

7.4 **Limitations**

This project was momentously limited to the video forensic analytics. For future work, this can be extended to study the repercussion of audio evidence and its integration with the semi-automated post-incident analysis framework proposed in this thesis.

Another limitation is that only Seven experiments were undertaken to determine the optimum anomaly threshold of the HTM. These experiments were significantly low and undertook more experiments would potentially have altered the optimum anomaly scores and the conclusions of this project. During the testing that was done using data from Virat, there is a possibility that the HTM recorded unnecessary moving objects that affected its ability to recall objects and classify their movements as normal or abnormal (Investigation, 2008).

For further research, it is essential that the forensic investigator has suitable and more sophisticated equipment to avoid limitations that might unavoidably interact with the results of findings, testing and analysis (Investigation, 2008).

The small sample size used in the questionnaire is also another limitation. The project utilized a sample of Thirty-One participants. While the questionnaire was quite comprehensive with Twenty-Six carefully crafted questions, the fact that only Thirty-

One responses have received puts into question the validity of the findings (Newman, 2003). Considering the importance and sensitivity of the issue under exploration, such a small sample number may not provide findings that are generalizable. Considering this, future researchers should consider using bigger samples – about 200 people. Moreover, they should also consider using a hardcopy Self-administered Questionnaire (SAQ) instead of the online questionnaire that was used in the current project (Newman, 2003). Robson (2002) accentuates that in online research, respondents often skip questions that they consider to be difficult and may also fail to submit or even see the questionnaire further lowering the response rate. The use of hardcopy SAQ will enhance the response rate as emphasized by Robson (2002).

Despite these limitations, the work presented in this report lends itself well to be implemented in real crime investigation providing court evidence through video forensic. However, the evolving potential of the field of the study meant there are opportunities for further research in this area (Savoldi and Gubian, 2009). As a matter of suggestion, it is expected that the future work for improving video investigation, a fully automated video forensic framework could be explored, rather than just the semi-automated proposed in this study.

The aspect of automation of video forensic framework is aimed at jettisoning the involvement of humans which can be more expensive, time-consuming, a higher percentage of error and most important manipulation of useful facts. It is, however, promising for the interested researcher to take further the current level of this research approach to even better performance as advised by Kaur, Kaur, and Khurana (2016). Also, the dataset used to obtain results from experiments can be extended further. Although the current research has applied the use of Virat dataset, more datasets that follow similar methodology can be considered with more complex scenarios. The additional dataset is expected to incorporate numerous pattern changes to test results and generate more significant results that would expand understanding of the forensic images under investigations.

7.5 Recommendations for Future Work

This project aims to study the requirements of video forensic investigation and Police procedures to propose a new semi-automated post-incident analysis framework and to investigate the application of the CLA to movement classification towards an automated video forensics process. The following recommendations for research are based on the study findings:

1. In a bid to show the performance of the proposed approach in a generic framework and especially in the case of using a sample of large size, also to optimize the classification process and solve the problem of exploration. More research is needed to propose improvements to the proposed approach so that it can be perfected even in the case of large samples. Example: to use the genetic algorithm that belongs to the family of evolutionary algorithms.

   Their goal is to obtain an approximate solution to an optimization problem, when there is no exact method (or the solution is unknown) to solve it in a reasonable time. Genetic algorithms use the notion of natural selection and apply it to a population of potential solutions to the given problem. The solution is approached by successive "leaps", as in a branch & bound procedure, except that they are formulas that are sought rather than values. In our case the genetic algorithm must be used to select the samples and then solve the classification problem in a reasonable time even in the case of a large sample.
2. Research to explore the classification of object movements detected from a video stream. Additional research is needed to provide a real-time video-based classification approach i.e. to provide an essential module for automating the real-time court video process.

3. A number of performance measures have been identified, generating motivations aiming to develop new methods of classification of the video such as an automatic classification method without the intervention of the expert.

Appendix A Evaluation Models

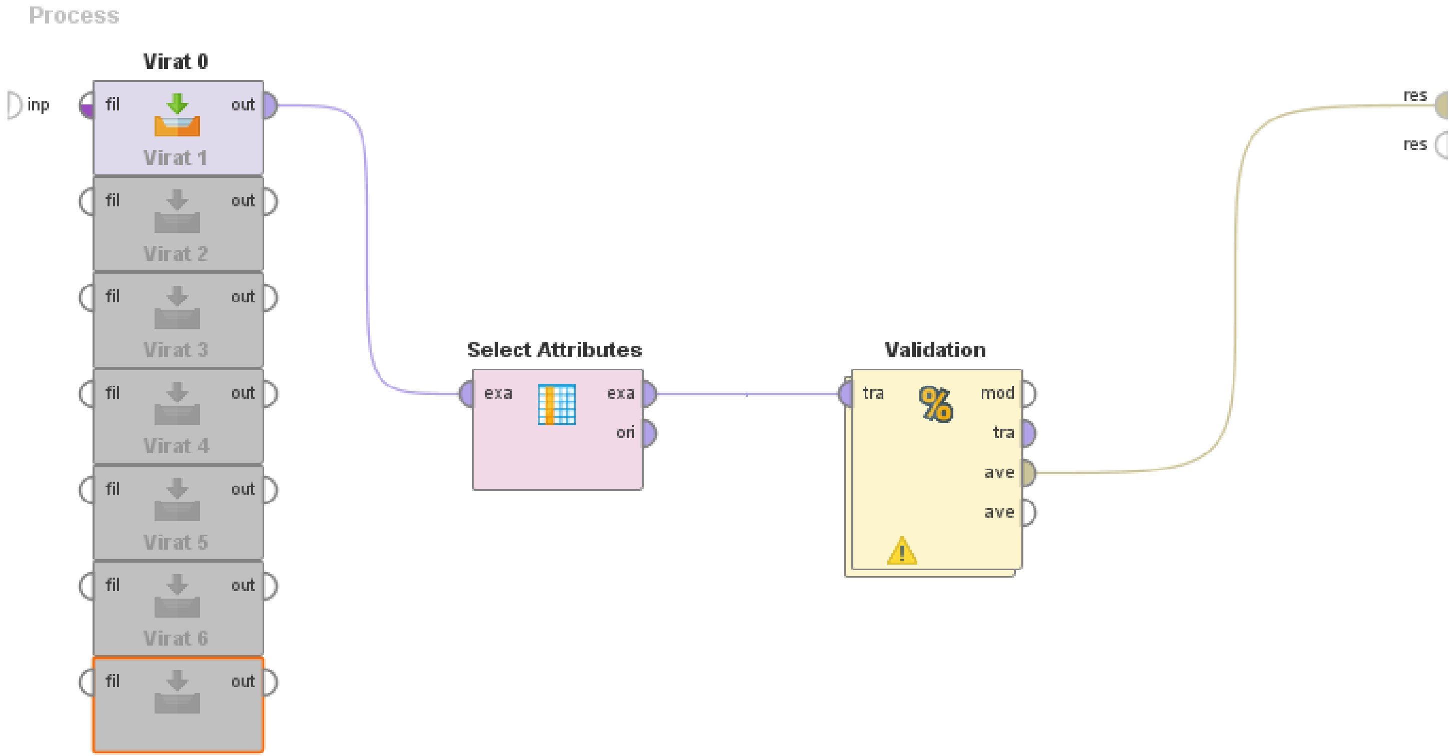


Figure A-1: Rapidminer validation block diagram

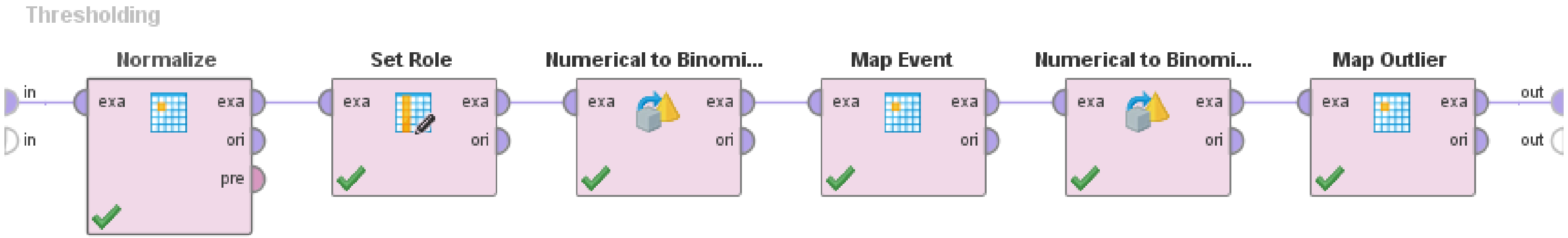


Figure A-2: Thresholding

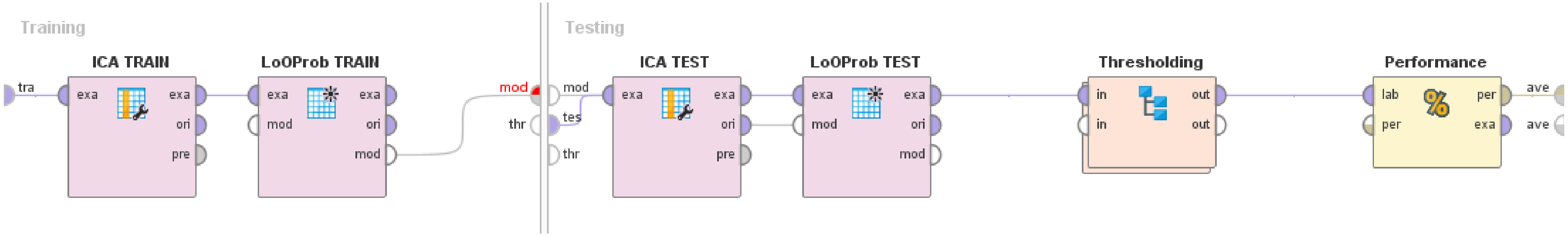


Figure A-3: Rapidminer training and testing block diagram - ICA Local Outlier Probability

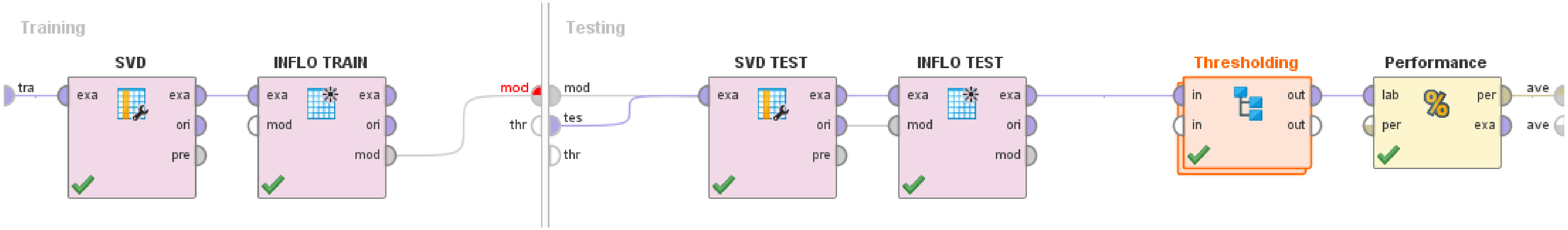


Figure A-4: Rapidminer training and testing block diagram - SVD Influenced Outlierness

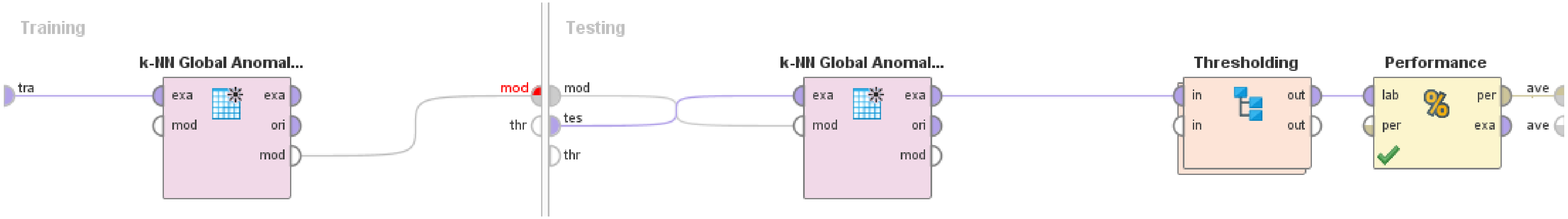


Figure A-5: Rapidminer training and testing block diagram – k-NN Global Anomaly Score

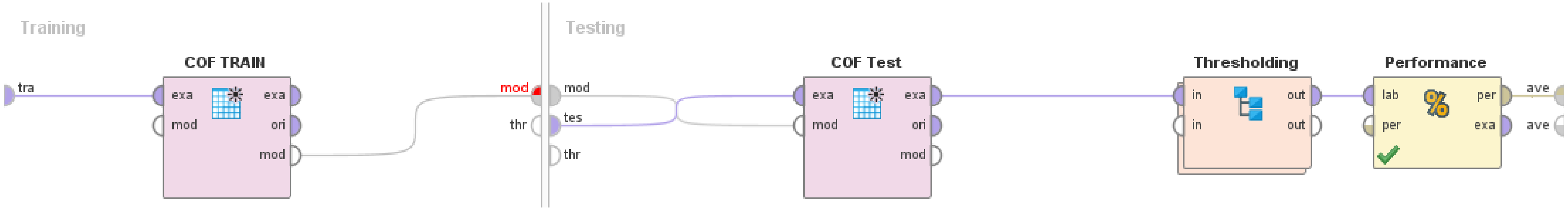
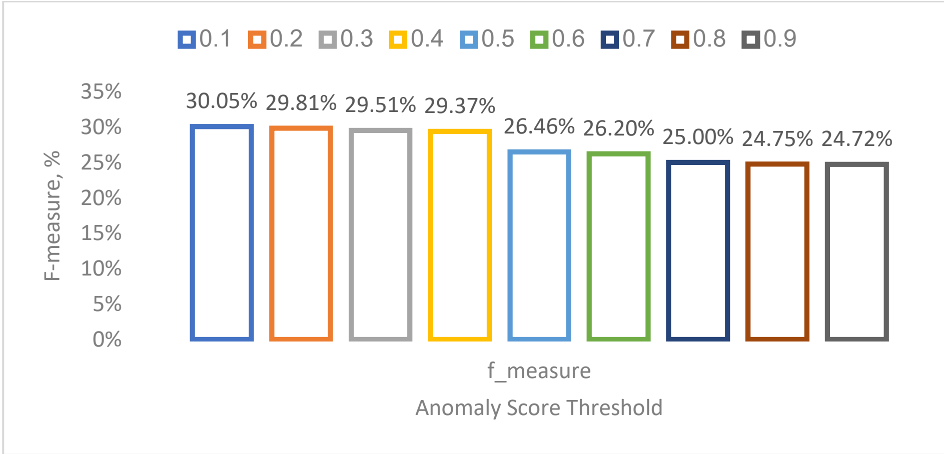


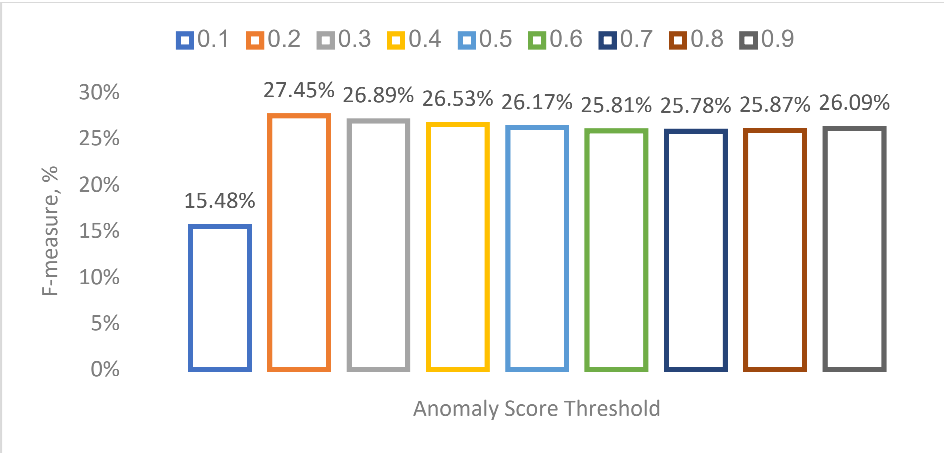
Figure A-6: Rapidminer training and testing block diagram – Connectivity-Based Outlier Factor

Appendix B Results - Optimum Threshold Selection Experiments

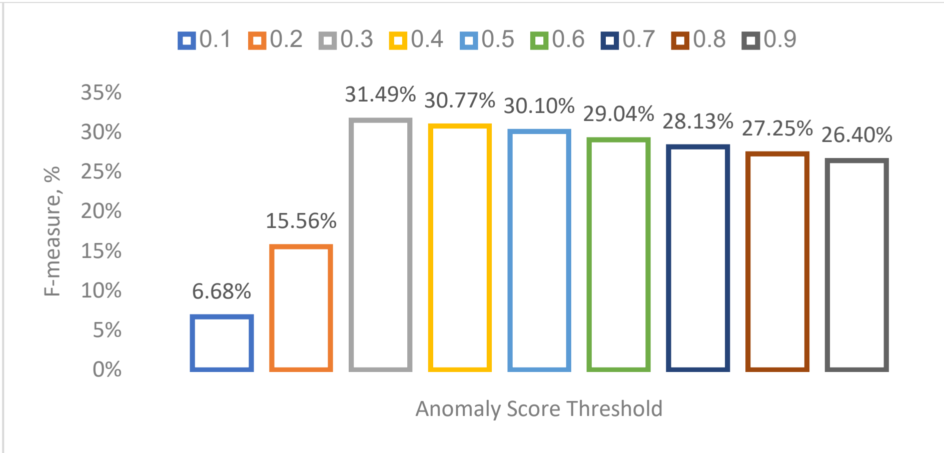
Proposed Algorithm:



Event 0



Event 1



Event 2

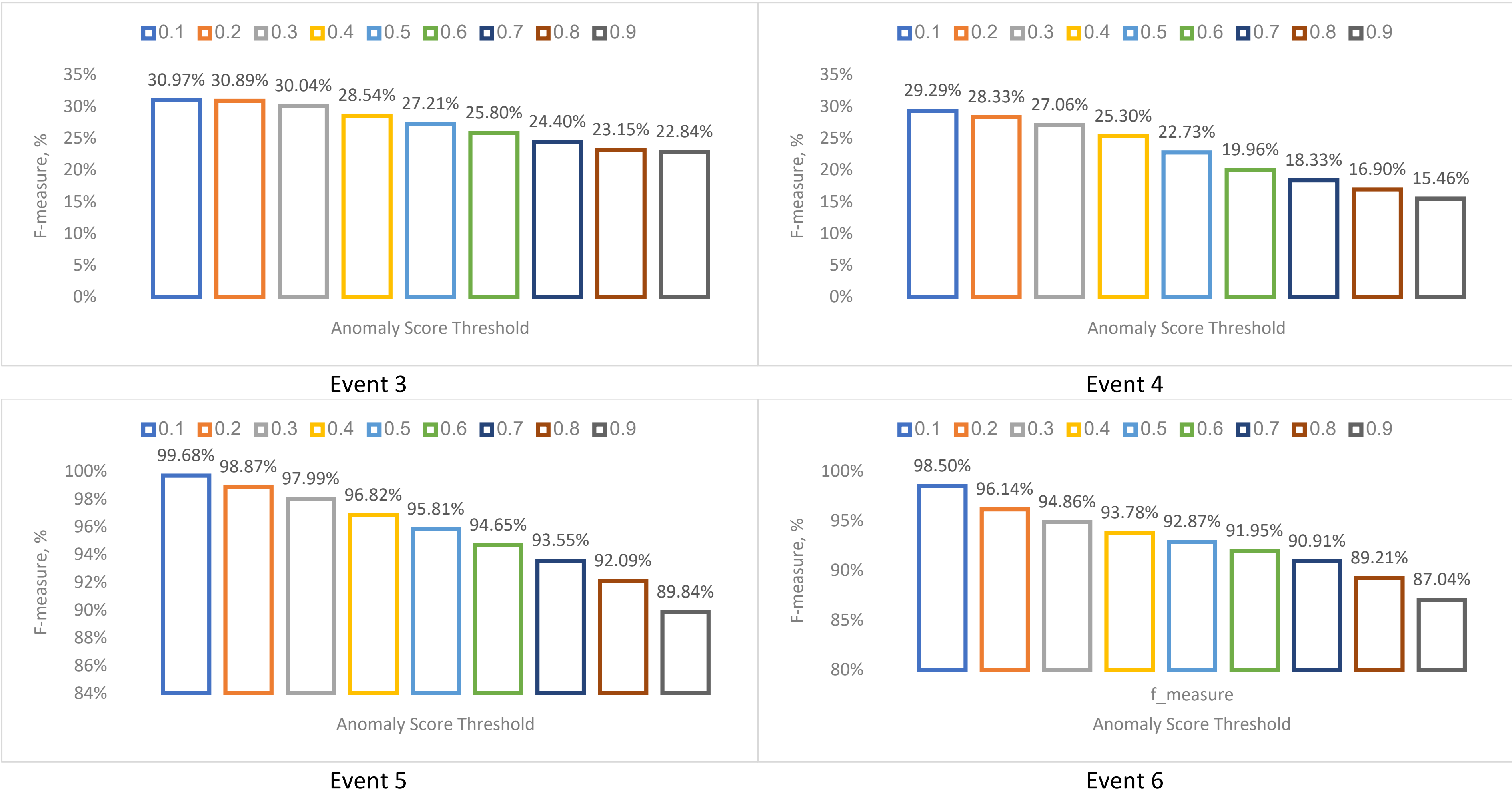
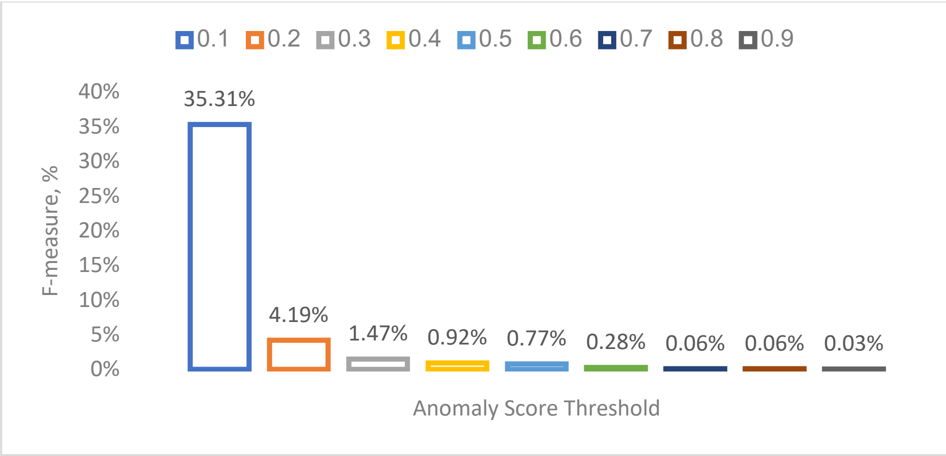


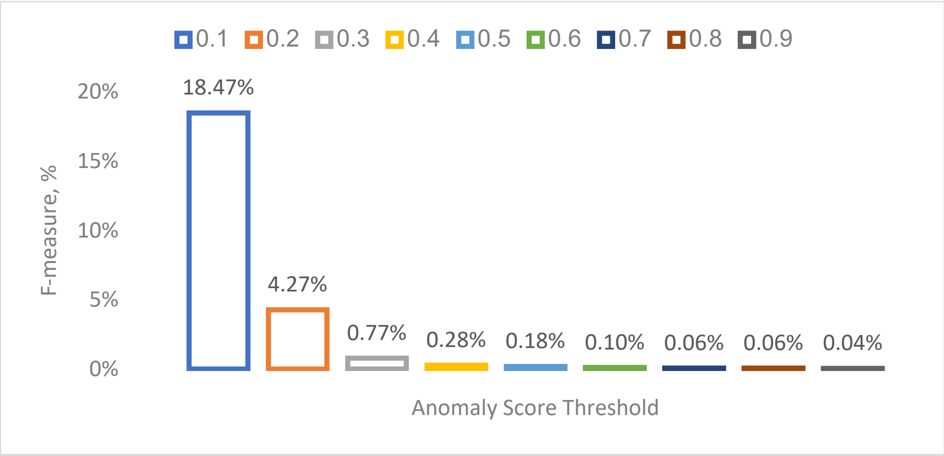
Figure B-1: Optimum threshold selection – Proposed algorithm

k-NN Global Anomaly Score

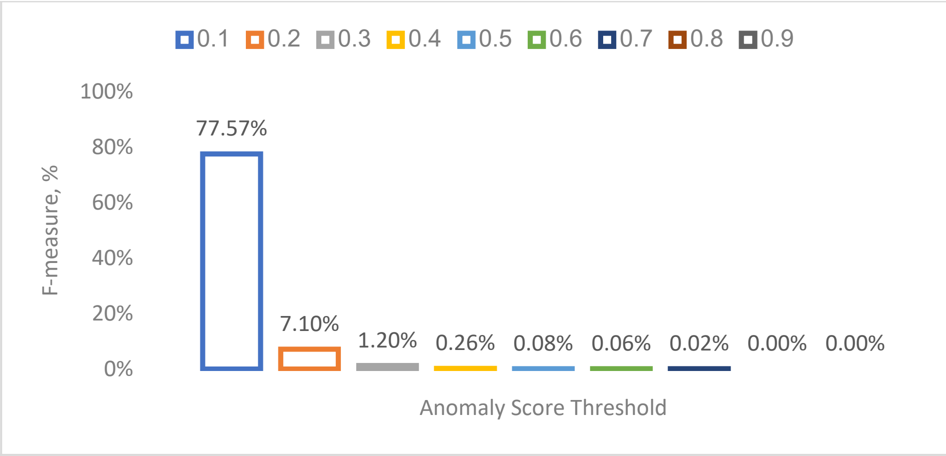


Event 0

Event 0



Event 1



Event 2

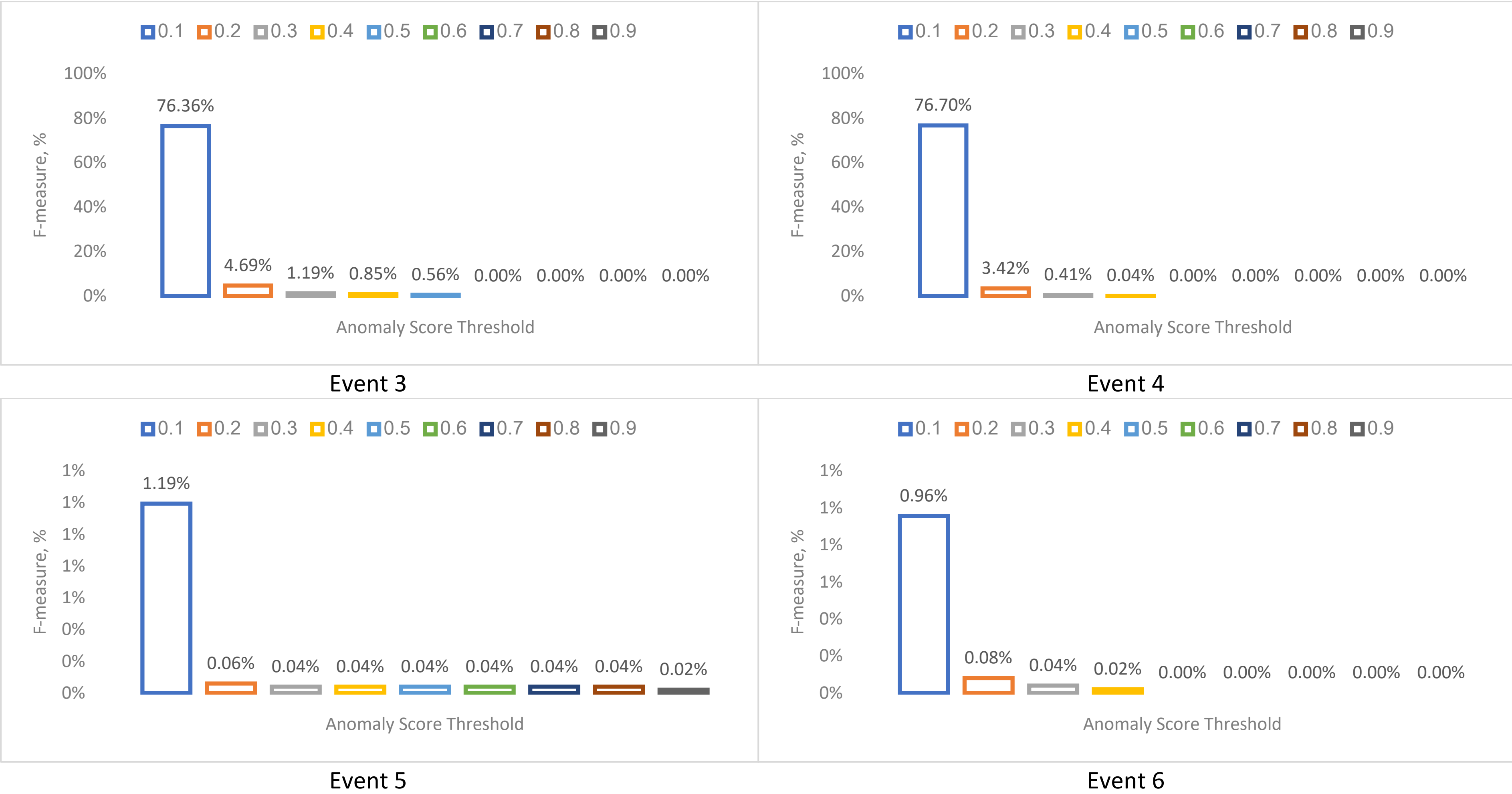
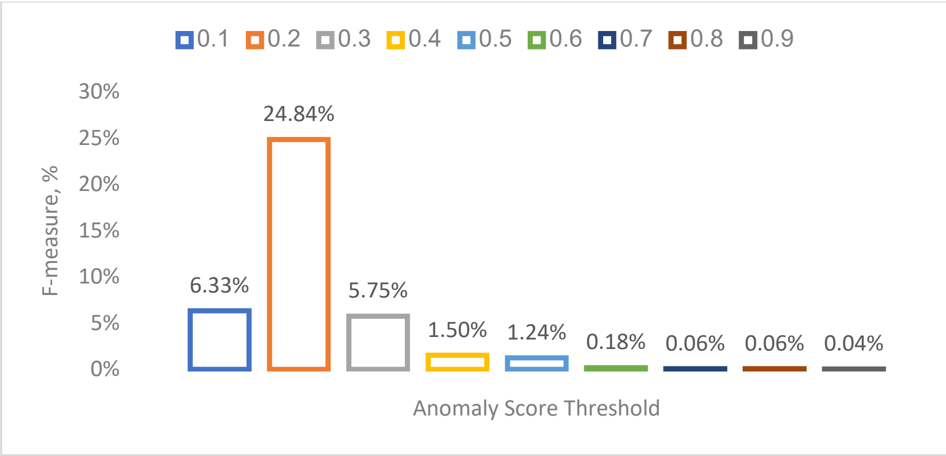
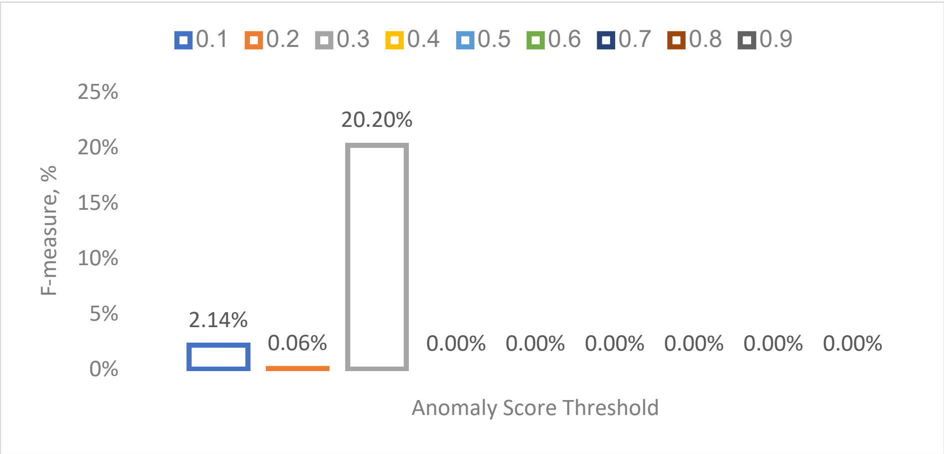


Figure B-2: Optimum threshold selection – k-NN Global Anomaly Score

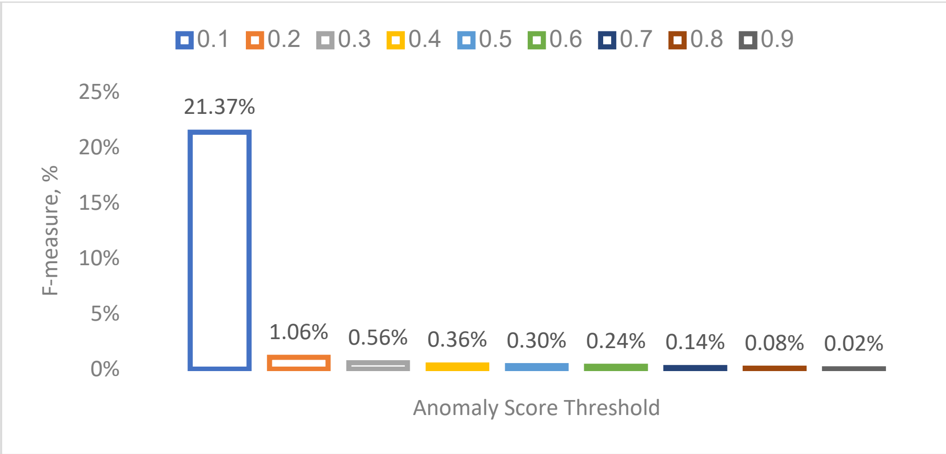
Singular Value Decomposition Influence Outlier (SVD-IO):



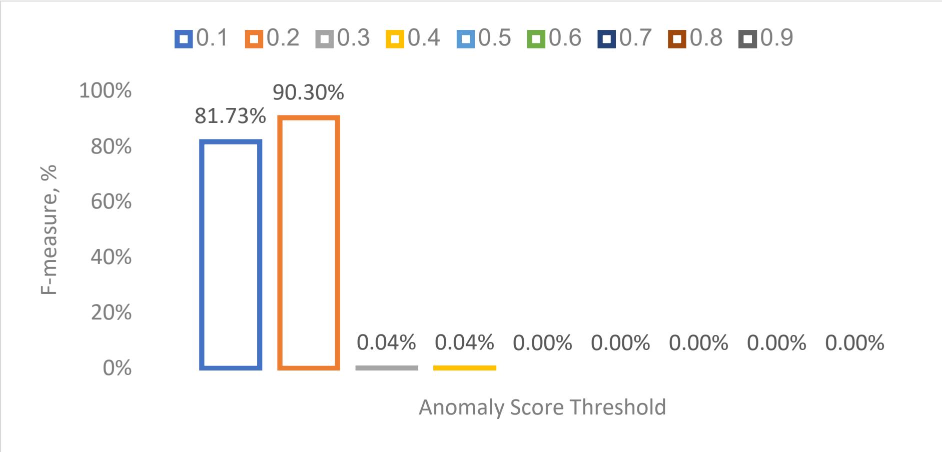
Event 0

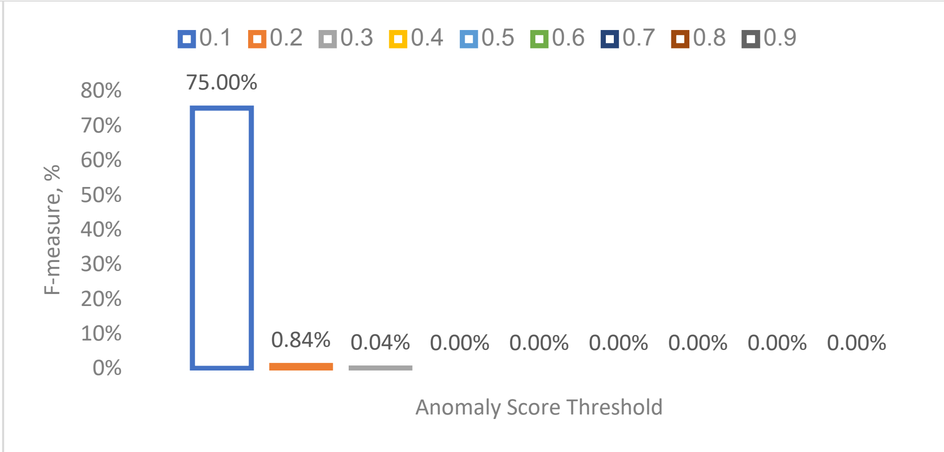


Event 1



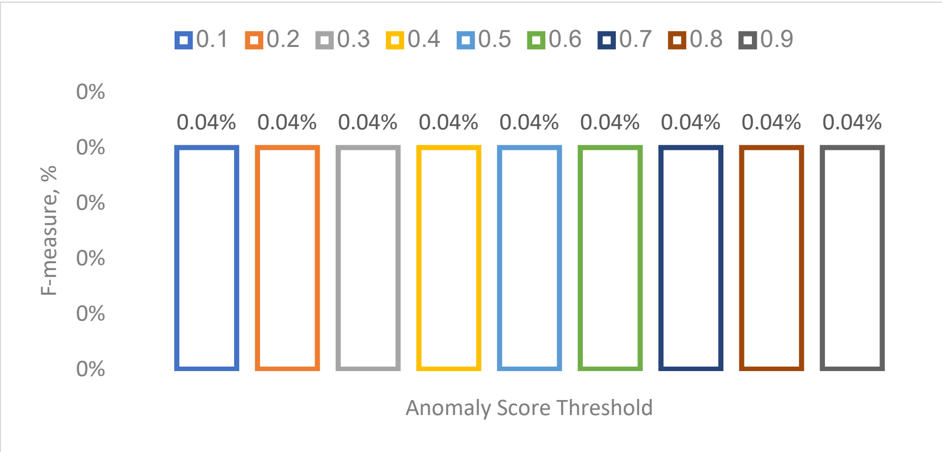
Event 2

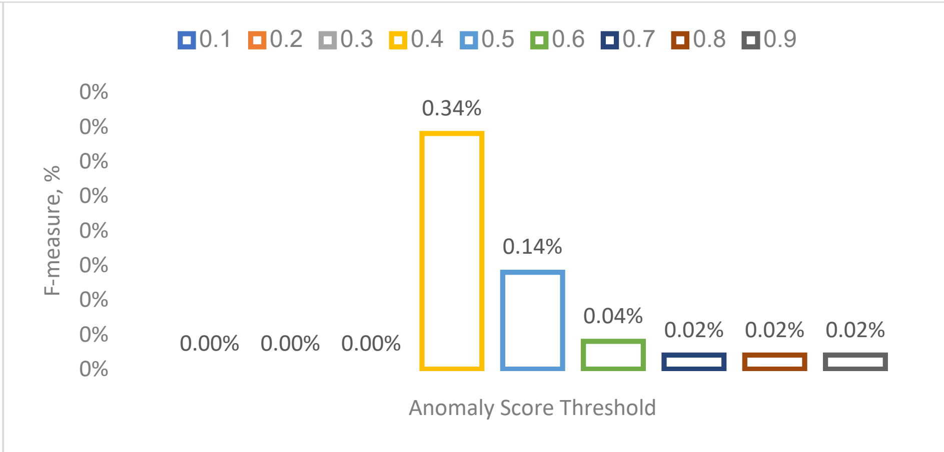




Event 3

Event 4



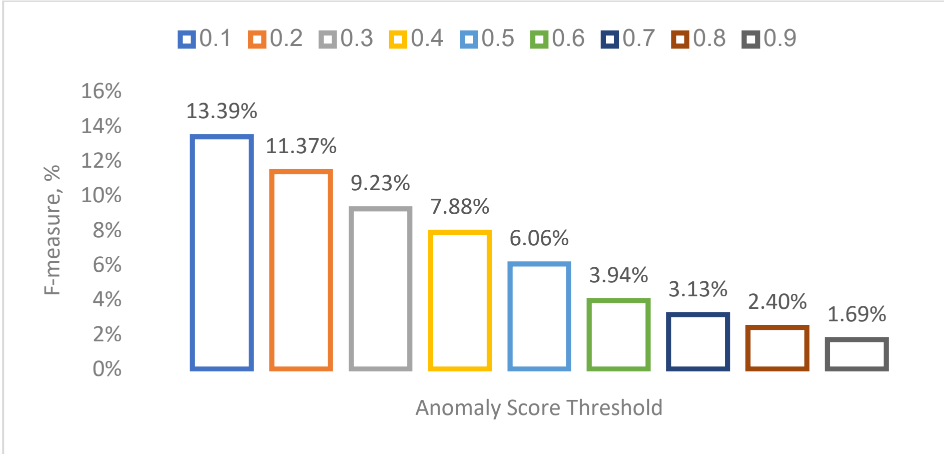


Event 5

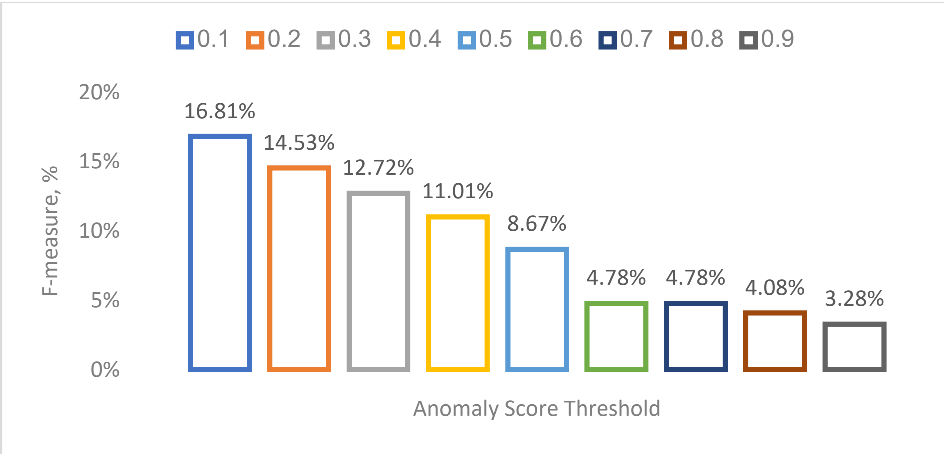
Event 6

Figure B-3: Optimum threshold selection – SVD Influence Outlier

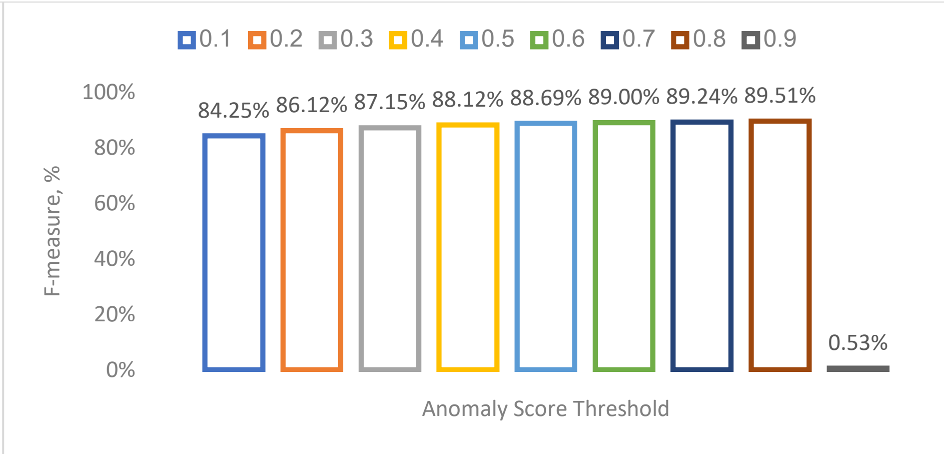
Independent Component Analysis – Local Outlier Probability (ICA-LOP):



Event 0



Event 1



Event 2

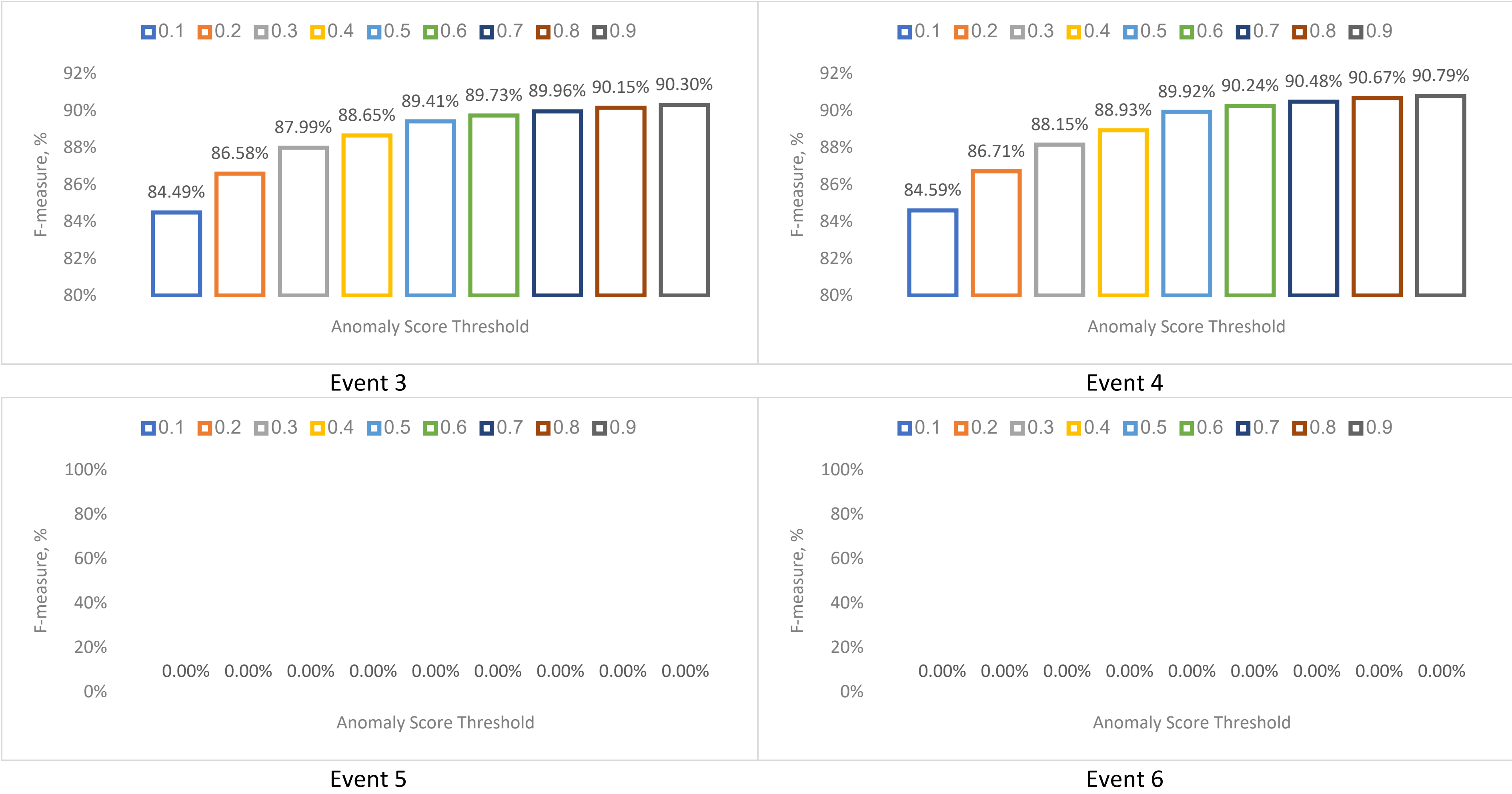
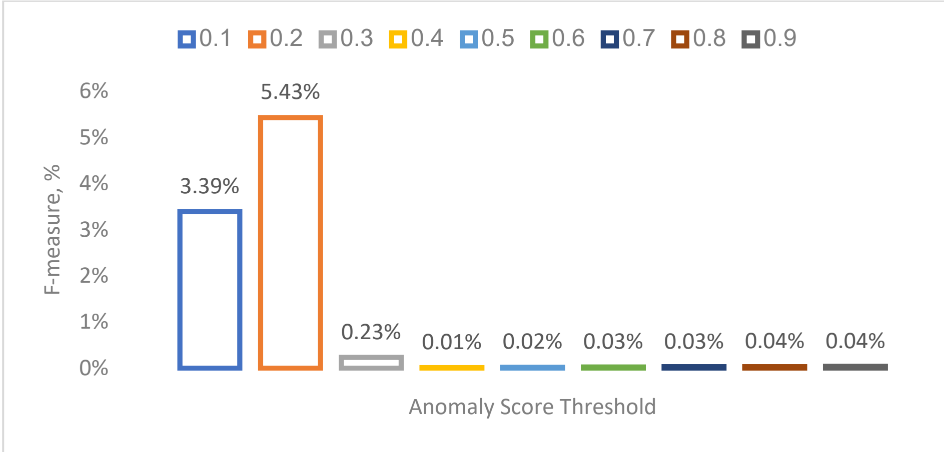
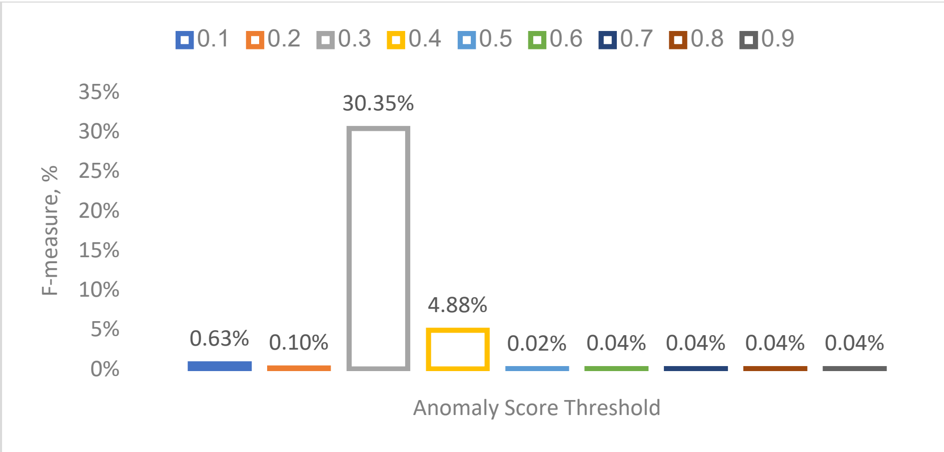


Figure B-4: Optimum threshold selection – ICA – Local Outlier Probability

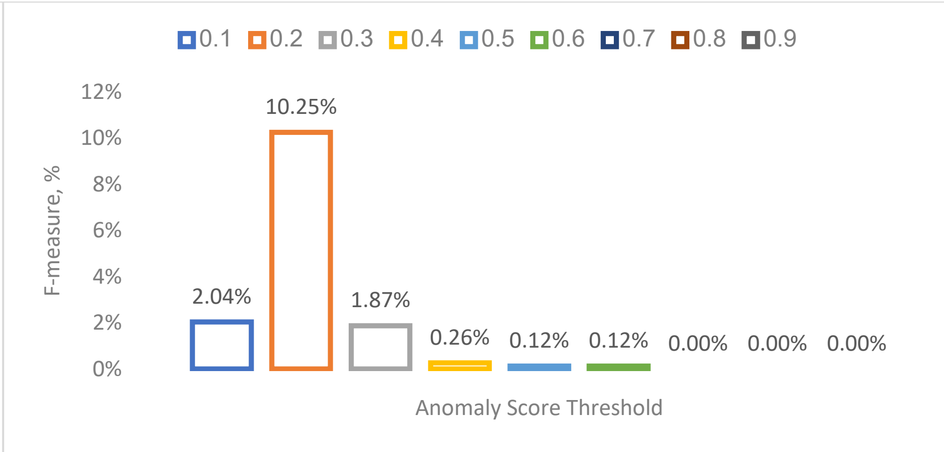
Connectivity-Based Outlier Factor (CBOF):



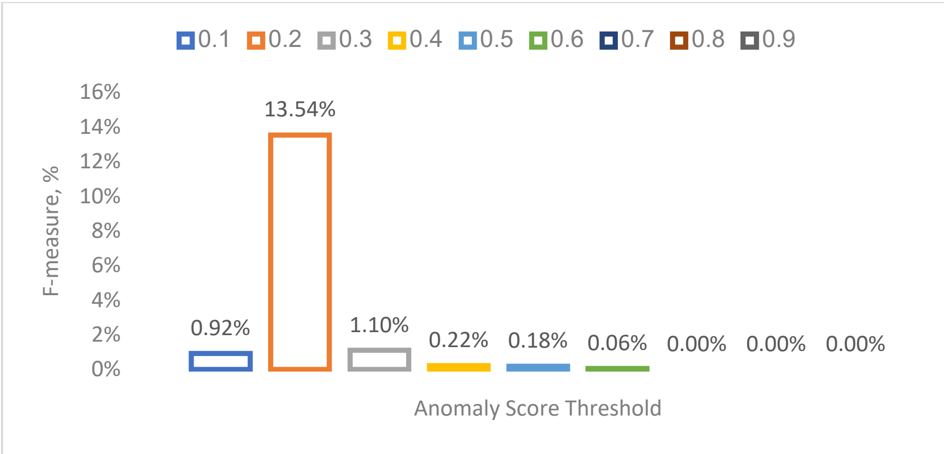
Event 0



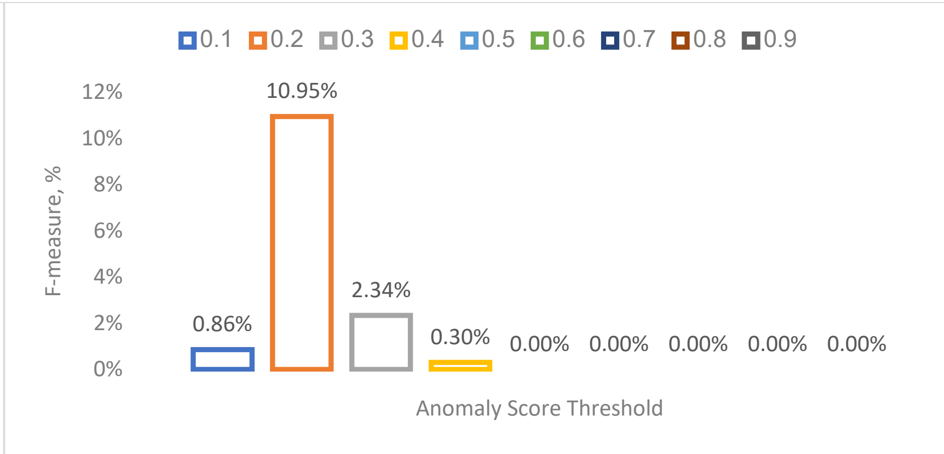
Event 1



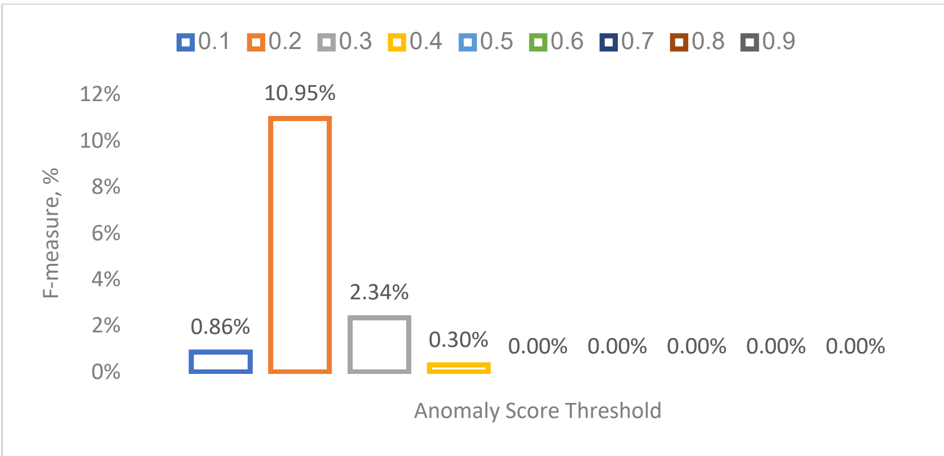
Event 2



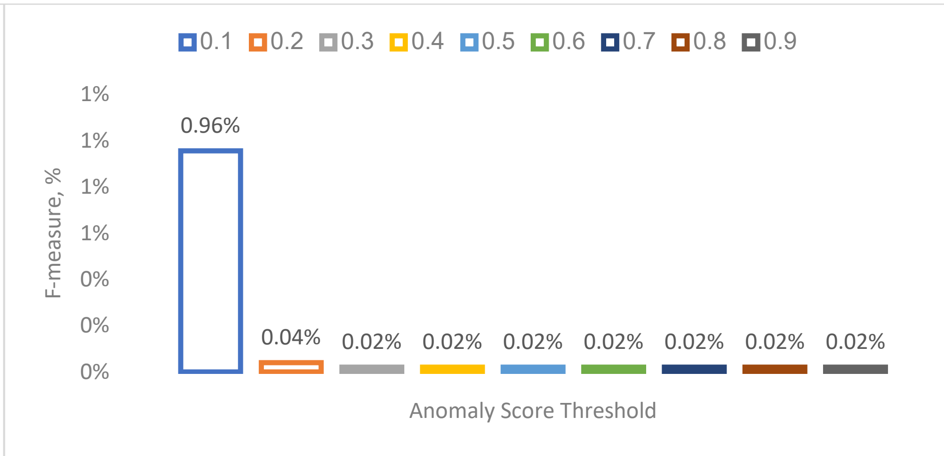
Event 3



Event 4



Event 5

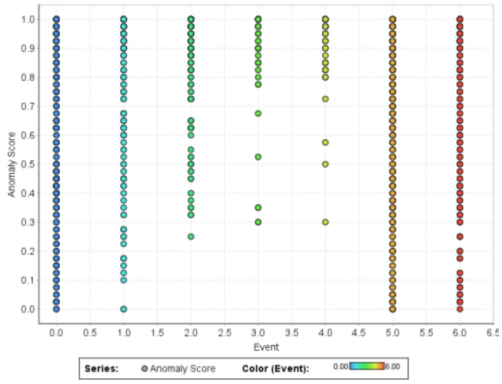


Event 6

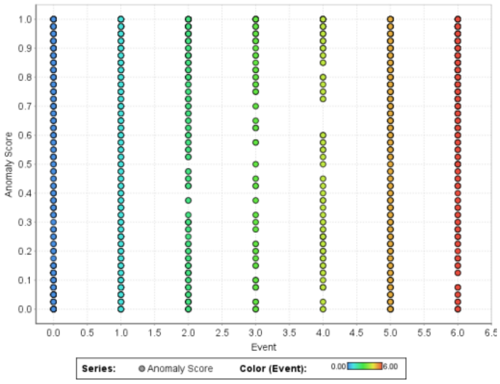
Figure B-5: Optimum threshold selection – Connectivity-Based Outlier Factor

Appendix C Results – Data Distribution

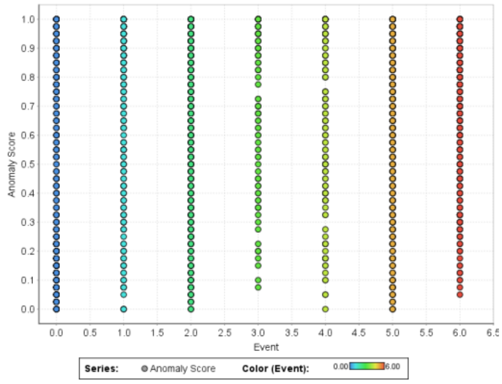
Proposed Algorithm:



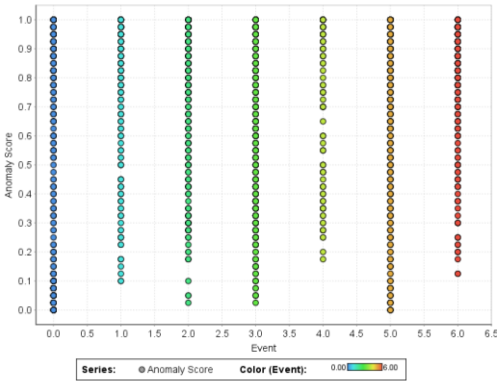
Event 0



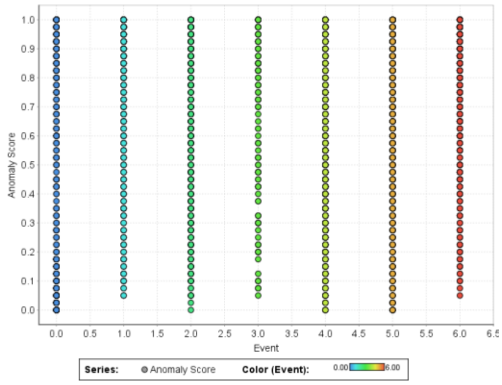
Event 1



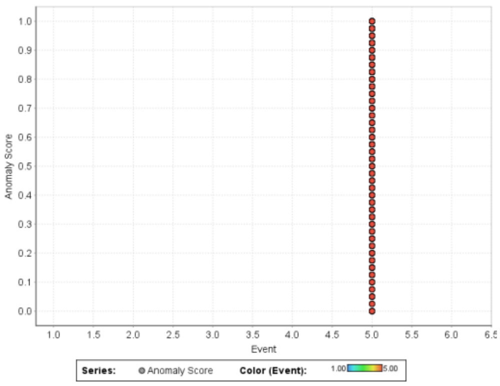
Event 2



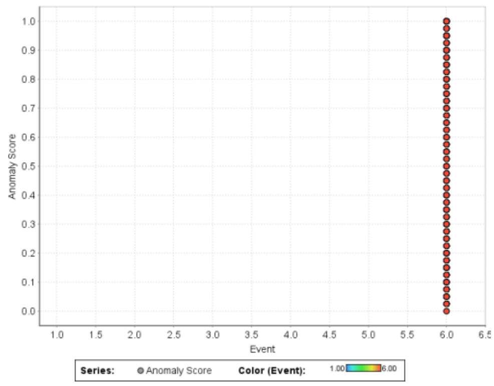
Event 3



Event 4



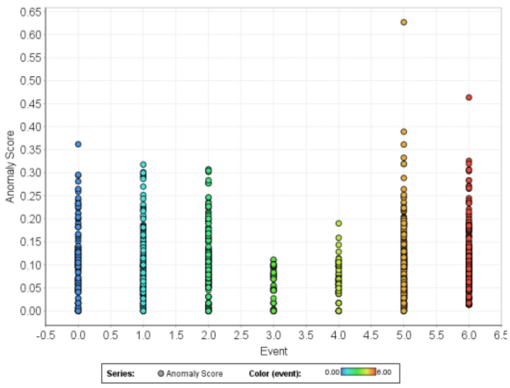
Event 5



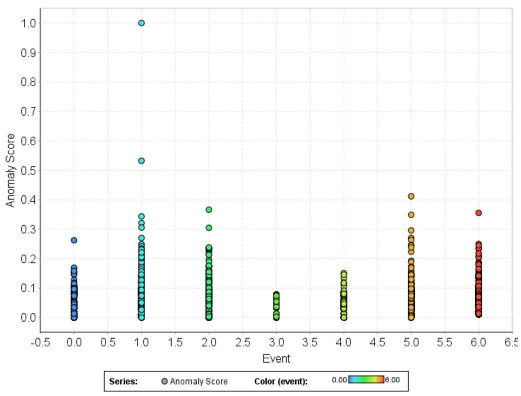
Even 6

Figure C-1: Scatter diagram – Proposed algorithm

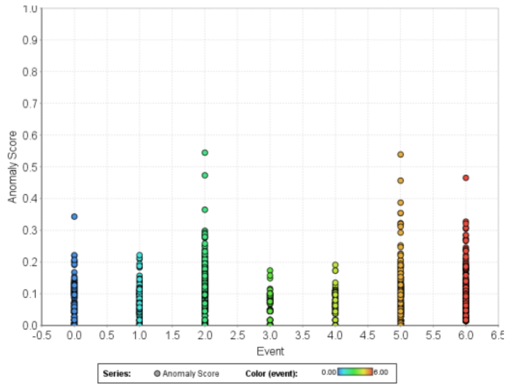
k-NN Global Anomaly Score



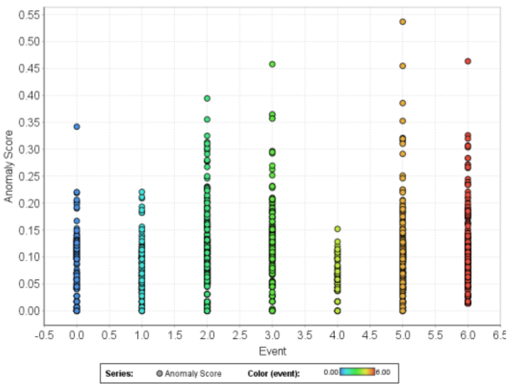
Event 0



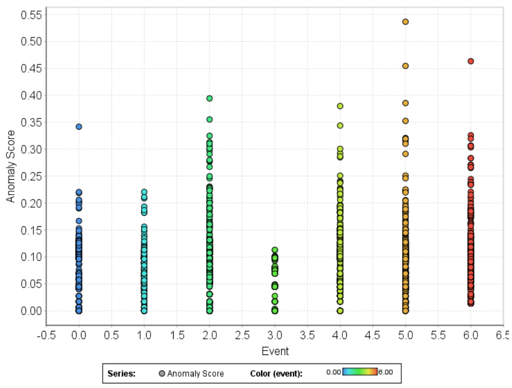
Event 1



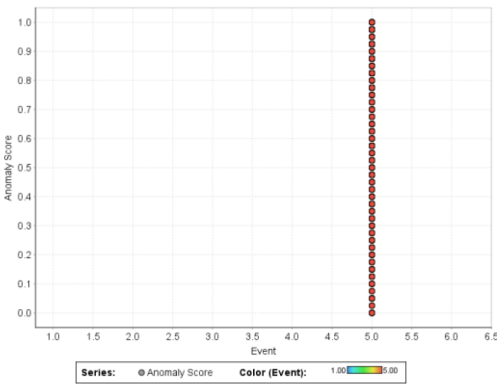
Event 2



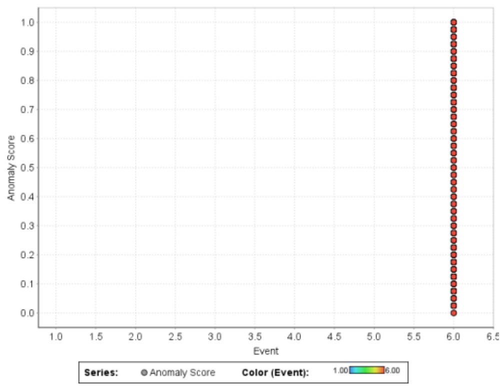
Event 3



Event 4



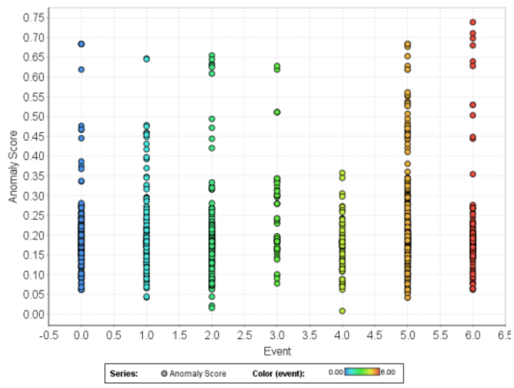
Event 5



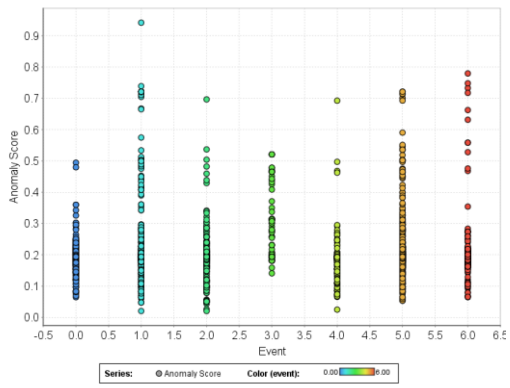
Event 6

Figure C-2: Scatter diagram – k-NN Global Anaomaly Score

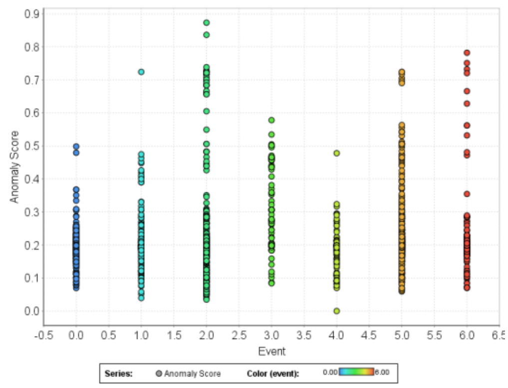
Singular Value Decomposition Influence Outlier (SVD-IO):



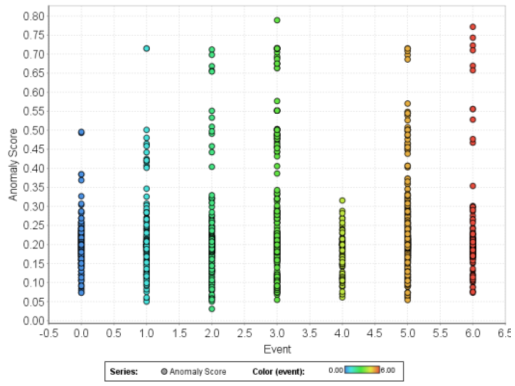
Event 0



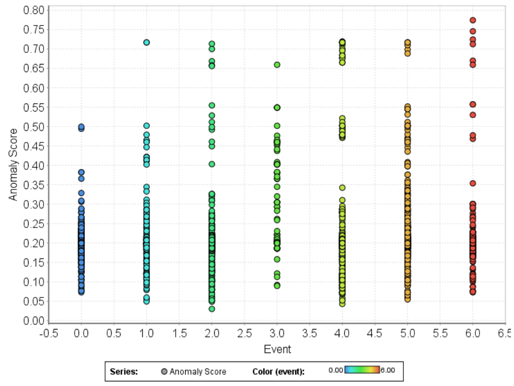
Event 1



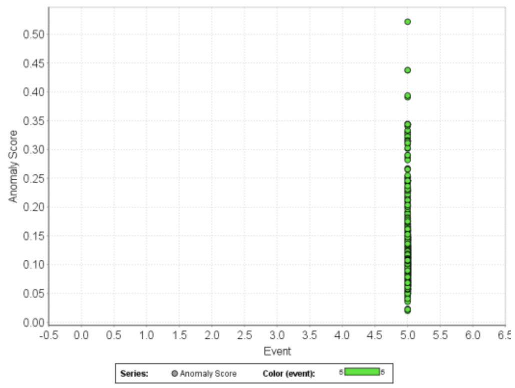
Event 2



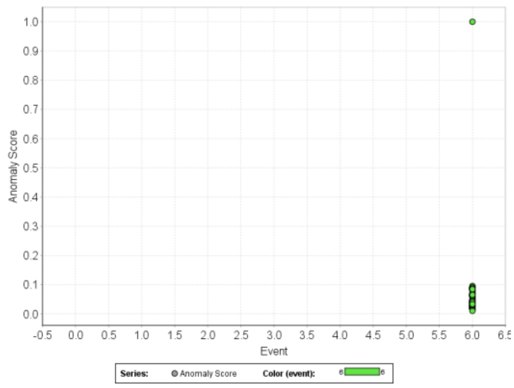
Event 3



Event 4



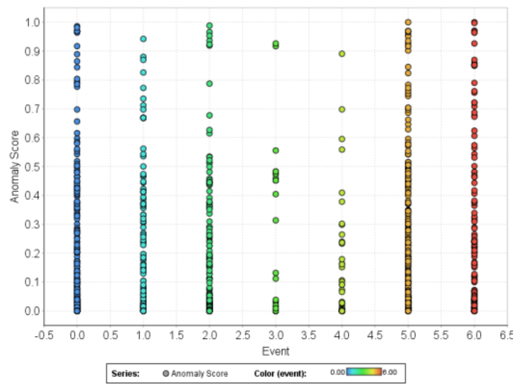
Event 5



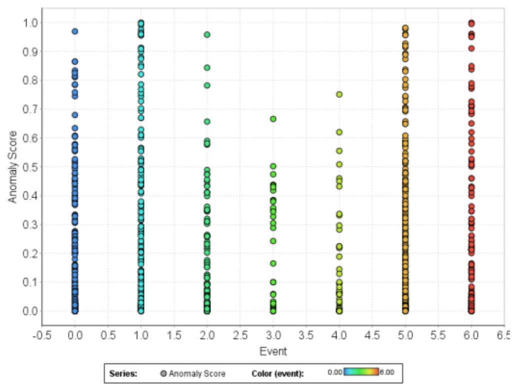
Event 6

Figure C-3: Scatter diagram – SVD Influence Outlier

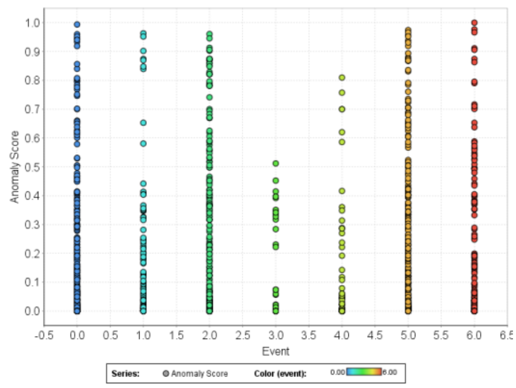
Independent Component Analysis – Local Outlier Probability (ICA-LOP):



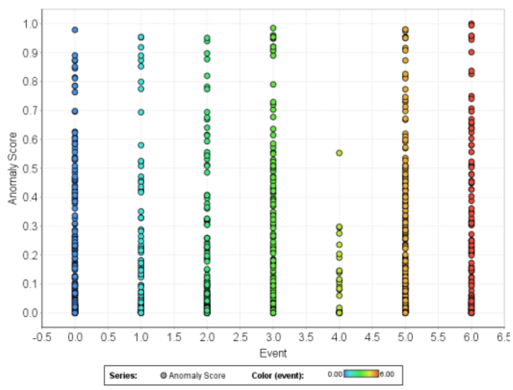
Event 0



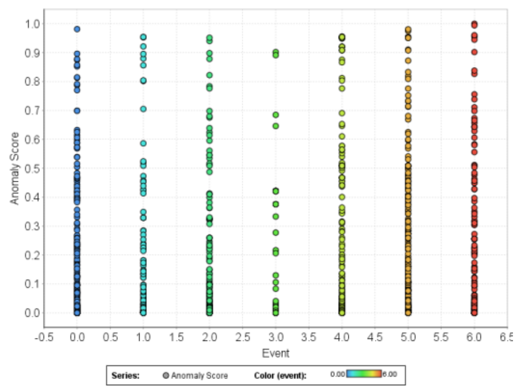
Event 1



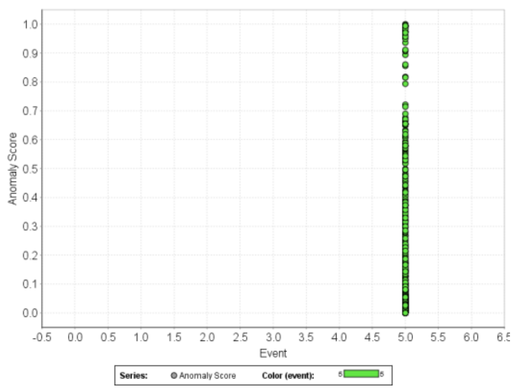
Event 2



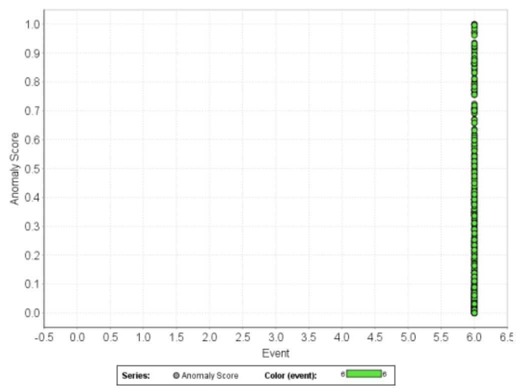
Event 3



Event 4



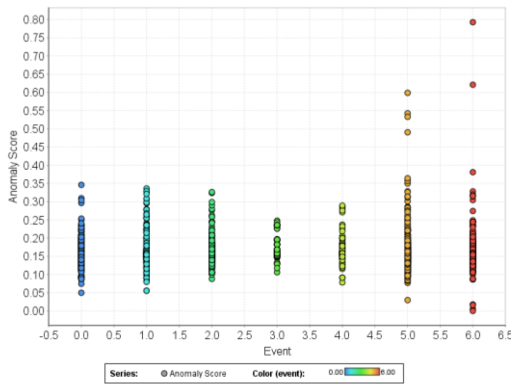
Event 5



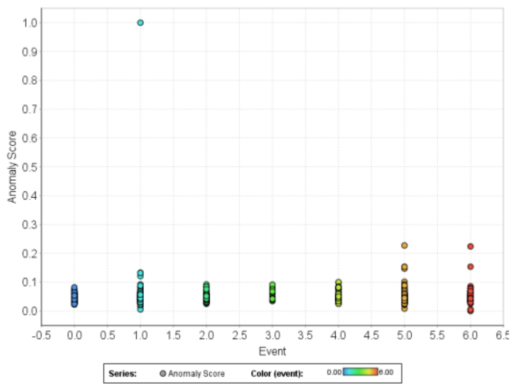
Event 6

Figure C-4: Scatter diagram – ICA – Local Outlier Probability

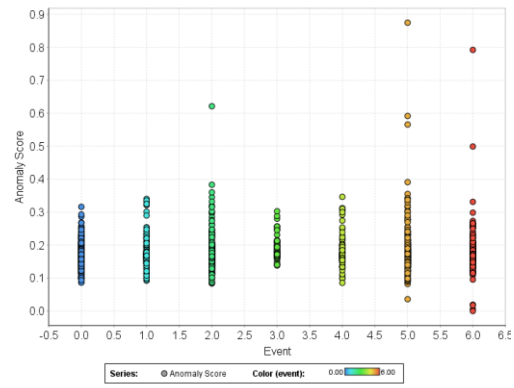
Connectivity-Based Outlier Factor (CBOF):



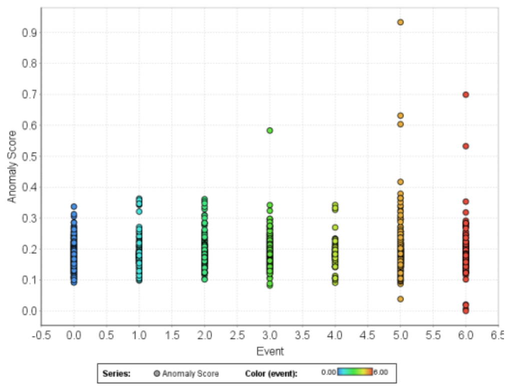
Event 0



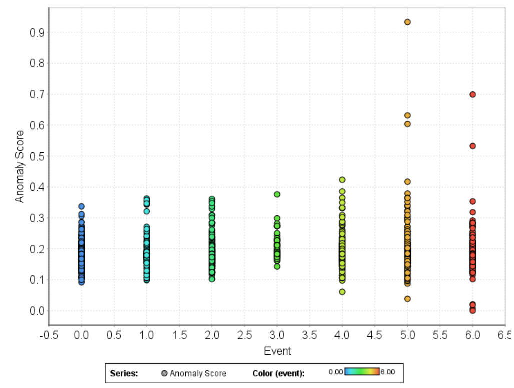
Event 1



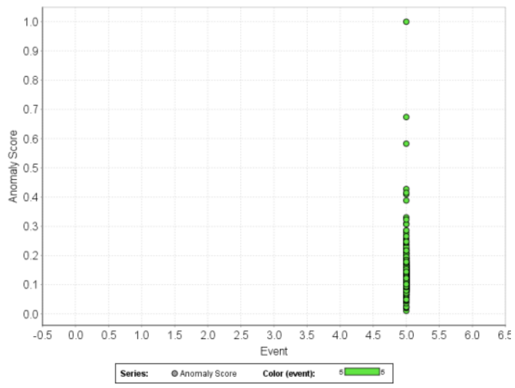
Event 2



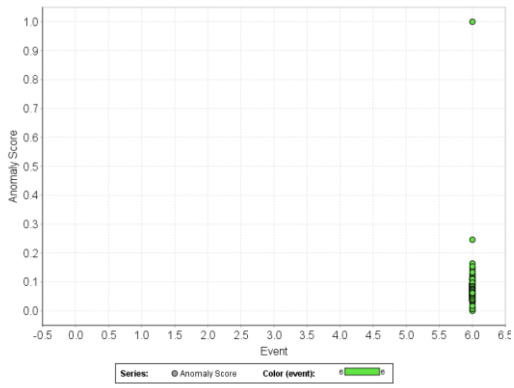
Event 3



Event 4



Event 5



Event 6

Figure C-5: Scatter diagram – Connectivity-Based Outlier Factor

References

ACPO (2012). Good Practice and Advice Guide for Managers of e-Crime Investigations (‘Managers’ Guide’) [online] available from <http://www.acpo.police.uk/documents/crime/2011/20110301%20CBA%20ACPO%20managers_guide_v10.1.4%20for%20ecrime%20investigations_2011.pdf>

Adams, A.A. and Ferryman, J.M., (2015). The future of video analytics for surveillance and its ethical implications. *Security Journal*, *28*(3), pp.272-289.

Ademu, I.O., Imafidon, C.O. and Preston, D.S., (2012). A new approach ofdigital forensic model for digital forensic investigation. *Int. J. Adv. Comput. Sci. Appl,* 2(12), pp.175-178.

Aggarwal, J.K., (2011). Motion analysis: Past, present and future. In Distributed Video Sensor Networks (pp. 27-39). Springer London.

Agrawal, D. and Meena, N., (2013). Performance comparison of moving object detection techniques in video surveillance system. The International Journal of Engineering and Science (IJES), 2(01), pp.240-242.

Agrawal, P. and Franklin, S., (2014). December. Multi-layer cortical learning algorithms. *In Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2014 IEEE Symposium on* (pp. 141-147). IEEE.

Ahmad, S. and Hawkins, J., (2015). Properties of sparse distributed representations and their application to hierarchical temporal memory. *arXiv preprint arXiv:1503.07469*.

Ahmad, S., & Hawkins, J. (2016). How do neurons operate on sparse distributed representations? A mathematical theory of sparsity, neurons and active dendrites. *arXiv preprint arXiv*:1601.00720.

Akintola, K., (2015). Real-time Object Detection and Tracking for Video Surveillance. *VFAST Transactions on Software Engineering*, *4*(2), pp.9-20.

Alshaikh, A. and Sedky, M., (2015). Post Incident Analysis Framework for Automated Video Forensic Investigation. *International Journal of Computer Applications*, *129*(17).

Alshaikh, A. and Sedky, M., (2016). Movement Classification Technique for Video Forensic Investigation. *International Journal of Computer Applications*, *135*(12).

Alzou'bi, S. Alshibly, H., and Al-Ma'aitah, M. (2014). Artificial Intelligence in Law Enforcement, A review. *International Journal of Advanced Information Technology* (IJAIT) 4 (4), 1- 9

Amershi, S., Cakmak, M., Knox, W.B. and Kulesza, T., (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), pp.105-120.

Anjum, A., Abdullah, T., Tariq, M., Baltaci, Y. and Antonopoulos, N., (2016). Video stream analysis in clouds: An object detection and classification framework for high performance video analytics. *IEEE Transactions on Cloud Computing*.

Amer, M. & Goldstein, M. (2012). Nearest-neighbor and clustering based anomaly detection algorithms for rapidminer. In *Proc. of the 3rd RapidMiner Community Meeting and Conference (RCOMM 2012)* (pp. 1–12).

Augustine et al (2005). Agile project management: steering from the edges. Communications of the ACM. [Online]. Available from: http://delivery.acm.org/10.1145/1110000/1101781/p85-augustine.pdf?ip=194.66.175.89&id=1101781&acc=ACTIVE%20SERVICE&key=BF07A2EE685417C5%2E7DBDC0B8C84E833A%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=680986131&CFTOKEN=83922823&__acm__=1433603461_7122f3ed8183a0608d27a2a96113f91e. (p. 86).

Ashcroft, J. (2001) 'Electronic crime scene investigation: A guide for first responders'. *NIJ Guide NCJ*

Balasubramaniam, J., Krishnaa, C. G., & Zhu, F. (2015). Enhancement of Classifiers in HTM-CLA Using Similarity Evaluation Methods. *Procedia Computer Science,* 60, 1516-1523

Baryamureeba, V. and Tushabe, F., (2004), The enhanced digital investigation process model. *In Proceedings of the Fourth Digital Forensic Research Workshop August.* (pp. 1-9).

Baudry, S., Chupeau, B. and Lefèbvre, F., (2009). A framework for video forensics based on local and temporal fingerprints. In 2009 16th IEEE *International Conference on Image Processing (ICIP)* (pp. 2889-2892).

Beebe, N. and Clark, J. (2004). 'A Hierarchical, Objectives-Based Framework for the Digital Investigations Process', *DFRWS Baltimore,* MD.

Bellman, R., (1978). An introduction to artificial intelligence: Can computers think?. Thomson Course Technology.

Bird, R., (2012). Lecture notes on Computer Forensic Methods [online] available from (Rowland 2014).

Bloisi, D.D., Pennisi, A. and Iocchi, L., (2017). Parallel multi-modal background modeling. *Pattern Recognition Letters, 96*, pp.45-54.

Bobick, A.F. and Davis, J.W., (2001). The recognition of human movement using temporal templates. *IEEE Transactions on pattern analysis and machine intelligence,* 23(3), pp.257-267.

Bosch, A., Zisserman, A. and Muñoz, X., (2008). Scene classification using a hybrid generative/discriminative approach. *IEEE transactions on pattern analysis and machine intelligence,* 30(4), pp.712-727.

Branch, H.O.S.D., (2006), June. Imagery Library for Intelligent Detection Systems (i-LIDS). In Crime and Security, 2006. *The Institution of Engineering and Technology Conference on* (pp. 445-448). IET.

Brutzer, S., Höferlin, B. and Heidemann, G., (2011). Evaluation of background subtraction techniques for video surveillance. *In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference* on (pp. 1937-1944). IEEE.

Bryman, A. and Bell, E. (2007). *Business Research Methods*, 2nd edition, England: Oxford University Press.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). Lof: Identifying density based local outliers. SIGMOD Rec., 29(2), 93–104.

Buch, N., Velastin, S.A. and Orwell, J., (2011). A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems,* 12(3), pp.920-939.

Byrne, F., (2015). Encoding reality: Prediction-assisted cortical learning algorithm in hierarchical temporal memory. *arXiv preprint arXiv:1509.08255*.

Cantrell, G., Dampier, D., Dandass, Y.S., Niu, N. and Bogen, C., (2012). Research toward a partially-automated, and crime specific digital triage process model. *Computer and Information Science*, 5(2), p.29.

Carner, D. (2013). Media forensics - Audio, video and image files: Structure, hash tags, metadata and tamper detection. *Journal of Forensic Research*, *s2*(01). doi:10.4172/2157-7145.s1.009

Carrier, B. (2003). 'Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers', *International Journal of Digital Evidence*, 1 (4).

Carrier, B. and Spafford, E.H., (2004). Getting physical with the digital investigation process. *International Journal of digital evidence*, 2(2), pp.1-20.

Carrier, B. and Spafford, E. (2004). *An Event-Based Digital Forensic Investigation Framework.* From the proceedings of The Digital Forensic Research Conference DFRWS 2004 USA Baltimore.

Champion, N. and Edgar, K., (2013) Through the gateway: How computers can transform rehabilitation. *Prison Reform Trust and Prisoners' Education Trust*.

Charniak, E. and McDermott, D., (1985*).* Introduction to AI. *Reading* (Mass.): Addison.

Choromanska, A., Henaff, M., Mathieu, M., Arous, G.B. and LeCun, Y., (2015), February. The loss surfaces of multilayer networks. In *Artificial Intelligence and Statistics* (pp. 192-204).

Chris, D. and David, D. (2012). A New Approach of Digital Forensic Model for Digital Forensic Investigation. *International Journal of Advanced Computer Science and Application*s, 2(12). doi:10.14569/ijacsa.2011.021226.

Ciardhuain, S. (2004). 'An Extended Model of Cybercrime Investigations'. International Journal of Digital Evidence 3 (1), 1- 22

Clark, J (2014) Meet the man building an AI that mimics our neocortex – and could kill off neural networks [online] available from <Clark, J (2014) Meet the man building an AI that mimics our neocortex – and could kill off neural networks >

Cole, E. Krutz, R. Conley, J. Reisman, B. Ruebush, M. Gollman, D. Reese, R. (2007) *Network Security Fundamentals*. USA: Wiley

Collins, R., Lincoln, R., and Frank, M. (2002). 'The Effect of Rapport in Forensic Interviewing'. Humanities & Social Science Papers 4 (1), 1-16

Conotter V, O'Brien J, Farid H. (2011). Exposing digital forgeries in ballistic motion. *IEEE Trans Inf Forensics Security* 2011; 7:283–96

Costello, C.J. and Wang, I.J., (2005), December. Surveillance camera coordination through distributed scheduling. *In Proceedings of the 44th IEEE Conference on Decision and Control* (pp. 1485-1490). IEEE.

Comon, P. (1994). Independent component analysis—a new concept? Signal Processing, 36:287–314.

Council of the Inspectors General on Integrity and Efficiency. (2012). Quality Standards for Digital Forensics. Available at: http://www.crime-scene-investigator.net/quality-standards- for-digital- forensics.html (Accessed 31 March, 2018).

Creative, B. (2017). Scene Anomaly Detection - DTI Group. Available at:http://www.dti.com.au/product-features/video-data-analytics/scene- anomaly-detection/.

Creswell, J. W. (2003). *Research design: Qualitative, Quantitative and mixed method approach*.(2nd ed.). Thousand Oaks, CA: Sage Publications

Dee, H.M. and Velastin, S.A., (2008). How close are we to solving the problem of automated visual surveillance?. *Machine Vision and Applications*, 19(5-6), pp.329-343.

Demirkus, M., Precup, D., Clark, J.J. and Arbel, T., (2015). Hierarchical temporal graphical model for head pose estimation and subsequent attribute classification in real-world videos. *Computer Vision and Image Understanding*, 136, pp.128-145.

DFRWS. (2001). *DFRWS Technical Report: A Road Map for Digital Forensic Research*, Utica, New York.

Ding, S., Li, H., Su, C., Yu, J. and Jin, F., (2013). Evolutionary artificial neural networks: a review. *Artificial Intelligence Review*, 39(3), pp.251-260.

Divya, P.B., Shalini, S., Deepa, R. and Reddy, B.S., (2017). Inspection Of Suspicious Human Activity In The Crowdsourced Areas Captured In Surveillance Cameras.

Draper, R. (2018). They Are Watching You'and Everything Else on the Planet. Available at: https://www.nationalgeographic.com/magazine/2018/02/surveillance-watching-you/.

Dua, S. and Du, X., (2016). *Data mining and machine learning in cybersecurity*. CRC press.

Eigenraam, D. and Rothkrantz, L.J.M., (2016), May. A smart surveillance system of distributed smart multi cameras modelled as agents. In *Smart Cities Symposium Prague (SCSP), 2016* (pp. 1-6). IEEE.

Elazary, L. and Itti, L., (2010). A Bayesian model for efficient visual search and recognition. *Vision research*, 50(14), pp.1338-1352.

Executive, S., (2003). Guidance interviewing child witnesses and victims in Scotland. Edinburgh: Author.

Euzenat, J., 2007, January. Semantic Precision and Recall for Ontology Alignment Evaluation. In IJCAI (Vol. 7, p. 348353).

Fan, D., Sharad, M., Sengupta, A. and Roy, K., (2016). Hierarchical temporal memory based on spin-neurons and resistive memory for energy-efficient brain-inspired computing. *IEEE transactions on neural networks and learning systems*, *27*(9), pp.1907-1919.

Feng P. and Weinong, W (2006) ‘Anomaly detection based on the regularity of normal behaviors’ in Proc. *1st Int. Symp. Syst. Control Aerosp. Astronautics,* Jan.19–21, 2006, pp. 1041–1046.

Feris, R., Bobbitt, R., Pankanti, S. and Sun, M.T., (2015), August. Efficient 24/7 object detection in surveillance videos. In *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on* (pp. 1-6).

Fisher, B. and Schroeder, U.(1999).[online] available from: http://www7.tomshardware.com/video

Flannelly, K.J., (2017). Brain Evolution and Emotions. In *Religious Beliefs, Evolutionary Psychiatry, and Mental Health in America* (pp. 71-78). Springer, Cham.

FutureTech and Park, J. J. (2012). Future information technology, application, and service: FutureTech 2012. Volume 1. Dordrecht: Springer.

Galetzka, M., Strüngmann, L. and Weber, C., (2014). Intelligent Predictions: an empirical study of the Cortical Learning Algorithm. *University of Applied Sciences Mannheim.*

George, D. and Hawkins, J., (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS computational biology*, 5(10), p.e1000532.

George, D., (2008). How the brain might work: A hierarchical and temporal model for learning and recognition (Doctoral dissertation, Stanford University).

Geradts, Z. and Bijhold, J. (2000). Forensic Video Investigation. *Multimedia Video-Based Surveillance Systems*, 3-12. doi:10.1007/978-1-4615-4327-5_1

Gijsberts, A., Atzori, M., Castellini, C., Müller, H. and Caputo, B., (2014). Movement error rate for evaluation of machine learning methods for sEMG-based hand movement classification. *IEEE transactions on neural systems and rehabilitation engineering*, *22*(4), pp.735-744.

Grobler, C.P., Louwrens, C.P. and von Solms, S.H., (2010), February. A multi-component view of digital forensics. In Availability, Reliability, and Security, 2010. *ARES'10 International Conference on* (pp. 647-652). IEEE.

Guo, X., Singh, S., Lee, H., Lewis, R.L. and Wang, X., (2014). Deep learning for real-time Atari game play using offline Monte-Carlo tree search planning. In *Advances in neural information processing systems* (pp. 3338-3346).

Hara, K., Omori, T. and Ueno, R., (2002). Detection of unusual human behavior in intelligent house. *In Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on* (pp. 697-706). IEEE.

Hardt, M., Recht, B. and Singer, Y., (2015). Train faster, generalize better: Stability of stochastic gradient descent. *arXiv preprint arXiv:1509.01240*.

Hargreaves, C. and Patterson, J. (2012). An automated timeline reconstruction approach for digital forensic investigations. *Digital Investigation*, 9, S69-S79. doi:10.1016/j.diin.2012.05.006.

Haugeland, J., (1985). Artificial intelligence: The very idea. *MIT press*.

Hawkins, J. and Blakeslee, S., (2004). On intelligence (Adapted ed.).

Hawkins, J., & Ahmad, S. (2015). Why Neurons Have Thousands of Synapses, A Theory of Sequence Memory in Neocortex. *arXiv preprint arXiv*:1511.00083.

Hawkins, J., Ahmad, S. and Dubinsky, D., (2010). Hierarchical temporal memory including HTM cortical learning algorithms. Techical report, Numenta, Inc, Palto Alto http://www. numenta. com/htmoverview/education/HTM_CorticalLearningAlgorithms. pdf.

Hawkins, J., Ahmad, S. and Dubinsky, D., (2011). Hierarchical Temporal Memory Including HTM Cortical Learning Algorithms, 0.2. Techical report Numenta. *Inc., September* Palto Alto

Hawkins, J., George, D. and Niemasik, J., (2009). Sequence memory for prediction, inference and behaviour. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 364(1521), pp.1203-1209.

Horschig, J.M., Smolders, R., Bonnefond, M., Schoffelen, J.M., Van den Munckhof, P., Schuurman, P.R., Cools, R., Denys, D. and Jensen, O., (2015). Directed communication between nucleus accumbens and neocortex in humans is differentially supported by synchronization in the theta and alpha band. PloS one, 10(9), p.e0138685.

Hu, W., Tan, T., Wang, L. and Maybank, S., (2004). A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3), pp.334-352.

Hole, K.J. (2016) Anti-fragile ICT Systems, Simula Springer Briefs on Computing 1

iCetana. (2017). iMotionFocus. Intelligence Video Surveillance System. Available at:https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwifmIDCmZHaAhXMsY8KHbxnAYUQFggtMAE&url=http%3A%2F%2Fwww.agilitygrid.com%2Fwpcontent%2Fuploads%2F2017%2F02%2FiCetana_Overview_0413.pdf&usg=AOvVaw1zI0tA5rqsjxylxDA5Ktdq (Accessed 29 March, 2018).

Investigation, E., (2001). Electronic Crime Scene Investigation: A Guide for First Responders, US Department of Justice. NCJ.

Investigation, F. B. (2008). *FBI Handbook of Crime Scene Forensics*. New York: Skyhorse Publishing, Inc.

Ivanov, Y.A. and Bobick, A.F., (1999). Recognition of multi-agent interaction in video surveillance. In Computer Vision, 1999. *The Proceedings of the Seventh IEEE International Conference on* (Vol. 1, pp. 169-176).

Iwashita, Y., Ryoo, M.S., Fuchs, T.J. and Padgett, C., (2013). Recognizing Humans in Motion: Trajectory-based Aerial Video Analysis. In BMVC 1 (3) p. 6.

Javed, O. and Shah, M., (2002). Tracking and object classification for automated surveillance. *In European Conference on Computer Vision* (pp. 343-357). Springer Berlin Heidelberg.

Jose, M. Chaquet, Escarmona, J., and Antonio. F.C., (2013). A survey of video datasets for human action and activity recognition. Computer Vision Image Understanding. pp. 423-424.

Kalker, T. (2001) Considerations on watermarking security. *Multimedia Signal Processing, 2001 IEEE Fourth Workshop on,* 201-206

Kapuscinski, T., (2010). Using hierarchical temporal memory for vision-based hand shape recognition under large variations in hand's rotation.*In International Conference on Artificial Intelligence and Soft Computing* (pp. 272-279). Springer Berlin Heidelberg.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L., (2014). Large-scale video classification with convolutional neural networks. *In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1725-1732).

Kaur, M., Kaur, N. and Khurana, S. (2016). A Literature review on Cyber Forensic and its Analysis tools. *IJARCCE*, *5*(1), 23-28. doi:10.17148/ijarcce.2016.5106. Available at: https://www.ijarcce.com/upload/2016/january16/IJARCCE%206.pdf

Khan, A., Kock W.U. & Memon, N. (2010). Digital Forensics and Crime Investigation: Legal Issues in Prosecution at National Level, Fifth International Workshop on Systematic Approaches to Digital Forensic Engineering, IEEE

Khanam, T. and Deb, K., (2017). Baggage Recognition in Occluded Environment using Boosting Technique. *KSII Transactions on Internet and Information Systems (TIIS)*, *11*(11), pp.5436-5458.

Kirtay, M., Falotico, E., Ambrosano, A., Albanese, U., Vannucci, L. and Laschi, C., (2016). Visual Target Sequence Prediction via Hierarchical Temporal Memory

Implemented on the iCub Robot. In *Conference on Biomimetic and Biohybrid Systems* (pp. 119-130). Springer, Cham.

Knowles, S. F., Hearne, J., and Smith, I. (2015) 'Physical restraint and the therapeutic relationship'. *The Journal of Forensic Psychiatry & Psychology* 26 (4), 461-475

Kobayashi, M., Okabe, T. and Sato, Y., (2010). Detecting forgery from static-scene video based on inconsistency in noise level functions. *IEEE Transactions on Information Forensics and Security,* 5(4), pp.883-892.

Kohn, M.D., Eloff, M.M. and Eloff, J.H.P., (2013). Integrated Digital Forensic Process Model. C*omputers & Security*, 38, pp. 103–115.

Kruse II, W.G. and Heiser, J.G., (2002). Computer forensics: incident response essentials. Pearson Education.

Kuehne, H., Jhuang, H., Garrote, E., Poggio, T. and Serre, T., (2011). HMDB: a large video database for human motion recognition. *In 2011 International Conference on Computer Vision* (pp. 2556-2563).

Kumar. R (2014). Research Methodology. 4th ed. Singapore: SAGE. (p. 14).

Kurzweil, R., Richter, R. and Schneider, M.L., (1990). The age of intelligent machines (Vol. 579). Cambridge: MIT press.

Lambie, E.S., Wilson, T.M., Brogt, E., Johnston, D.M., Ardagh, M., Deely, J., Jensen, S. and Feldmann-Jensen, S., (2017). Closed circuit television (CCTV) earthquake behaviour coding methodology: analysis of Christchurch Public Hospital video data from the 22 February Christchurch earthquake event. *Natural hazards*, *86*(3), pp.1175-1192.

Lavee, G., Khan, L. and Thuraisingham, B., (2007). A framework for a video analysis tool for suspicious event detection. *Multimedia Tools and Applications*, 35(1), pp.109-123.

Lavin, A. and Ahmad, S., (2015). Evaluating Real-Time Anomaly Detection Algorithms--The Numenta Anomaly Benchmark.*In Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on* (pp. 38-44).

Leake, M., Xia, L., Rocki, K. and Imaino, W., (2015) Effect of Spatial Pooler Initialization on Column Activity in Hierarchical Temporal Memory. In *AAAI* (pp. 4176-4177).

Lee, C.K., Ho, M.F., Wen, W.S. and Huang, C.L., (2006). Abnormal event detection in video using n-cut clustering. *In 2006 International Conference on Intelligent Information Hiding and Multimedia* (pp. 407-410). IEEE.

Li, C.-T, and IGI Global. (2013). Emerging digital forensics applications for crime detection, prevention, and security. Hershey, PA: IGI Global (701 E. Chocolate Avenue, Hershey, Pennsylvania, 17033, USA.

Liu, L., Ma, S. and Fu, Q., (2017). Human action recognition based on locality constrained linear coding and two-dimensional spatial-temporal templates. In *Chinese Automation Congress (CAC), 2017* (pp. 1879-1883).

Lotz M (2013). 'Waterfall Vs. Agile: Which Is the Right Development Methodology For Your Project?'. Segue Technologies. N.p. [Online]. Available from: http://www.seguetech.com/blog/2013/07/05/waterfall-vs-agile-right-development-methodology.

Lycos Tech Glossary. (1999). [online] available from < http://webopedia.lycos.com/Multimedia/Video/video.html> [June 15 2015]

Màrmol, E. and Sevillano, X., (2016). QuickSpot: a video analytics solution for on-street vacant parking spot detection. *Multimedia Tools and Applications*, *75*(24), pp.17711-17743.

Mathie, M.J., Celler, B.G., Lovell, N.H. and Coster, A.C.F., (2004). Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, 42(5), pp.679-687.

May, T. (2011). *Social Research: Issues, methods and process* (4th edn). Maidenhead, UK: McGraw-Hill Education.

McGivern. Y (2006) The Practice of Market and Social Research. 2th ed. Essex: Pearson Education. (p. 57, 58).

McLean, M.R., (2014). Concurrent learning algorithm and the importance map. In *Network Science and Cybersecurity* (pp. 239-250). Springer, New York, NY.

Meadaris, K., (2006). Grants to help develop ways to improve digital evidence collection. Purdue University.

Melis, W.J., Chizuwa, S. and Kameyama, M., (2009). Evaluation of the hierarchical temporal memory as soft computing platform and its VLSI architecture. *In Multiple-Valued Logic, 2009. ISMVL'09. 39th International Symposium on* (pp. 233-238).

Mnatzaganian, J.W., (2016). A Mathematical Formalization of Hierarchical Temporal Memory's Spatial Pooler for use in Machine Learning. Rochester Institute of Technology.

Mushtaque, K., Ahsan, K., and Umer, A. (2015). 'Digital Forensic Investigation Models: An Evolution Study'. Journal of Information Systems and Technology Management 12(2), 233- 244

Mocas, S. (2003) “Building Theoretical Underpinnings for Digital Forensics”, International Journal of Digital Evidence Winter 2003, 1, (4).

Mondaini, N., Caldelli, R., Piva, A., Barni, M. and Cappellini, V., (2007). Detection of malevolent changes in digital video for forensic applications. In Electronic Imaging 2007 (pp. 65050T-65050T). *International Society for Optics and Photonics.*

Moon, J., Kwon, Y., Kang, K. and Park, J., (2015). ActionNet-VE Dataset: A Dataset for Describing Visual Events by Extending VIRAT Ground 2.0. In *Signal Processing, Image Processing and Pattern Recognition (SIP), 2015 8th International Conference on* (pp. 1-4).

Moya, F.F. and Rojas, F.J.T., (2015). Object Tracking Based on Hierarchical Temporal Memory Classification.

Nazare, A. C.,Santos, C. E., Ferreira, R., and Schwartz, W. R. (2014). Smart surveillance framework: A versatile tool for video analysis. *Proceedings of IEEE Winter Conference on Applications of Computer Vision* (WACV 2014), pp. 753-760

National Institute of Standards and Technology: Information Technology (2013). Information Technology. Retrieved from https://www.nist.gov/topics/information-technology

Newman, W. L. (2003). *Qualitative and quantitative approaches* (5th ed.). New York: Pearson Education Inc.

Nilsson, N.J., (1998). Human-level artificial intelligence? Be serious! AI magazine, 26(4), p.68.

Nimura, Y., Hayashi, Y., Kitasaka, T. and Mori, K., (2015). Automated torso organ segmentation from 3D CT images using structured perceptron and dual decomposition. In *Medical Imaging 2015: Computer-Aided Diagnosis* (Vol. 9414, p. 94143L). International Society for Optics and Photonics.

NITDA. (2014). National Information Technology Development Agency. [online] available from < www.nitda.gov.ng >

Numenta (2011) Hierarchical Temporal Memory including HTM Cortial Learning Algorithms. [online] available from <https://numenta.org/resources/HTM_CorticalLearningAlgorithms.pdf>

Oh, S., Hoogs, A., Perera, A., Cuntoor, N., Chen, C.C., Lee, J.T., Mukherjee, S., Aggarwal, J.K., Lee, H., Davis, L. and Swears, E., (2011). A large-scale benchmark dataset for event recognition in surveillance video. *In Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on (pp. 3153-3160).

PACE, (1984). Police and Crime Evidence Act 1984 (PACE) codes of practice [online] availablefrom <https://www.gov.uk/police-and-criminal-evidence-act-1984-pace-codes-of-practice>

Palmer, G. L. (2001). A Road Map for Digital Forensic Research. Technical Report DTR- T0010-01, DFRWS. *Report for the First Digital Forensic Research Workshop* (DFRWS)

Pan, F. and Wang, W., (2006). Anomaly detection based-on the regularity of normal behaviors. *In 2006 1st International Symposium on Systems and Control in Aerospace and Astronautics* (pp. 6-pp).

Paul, M., Haque, S.M. and Chakraborty, S., 2013. Human detection in surveillance videos and its applications-a review. EURASIP Journal on Advances in Signal Processing, 2013(1), p.1.

Perea, A.J., Meroño, J.E. and Aguilera, M.J., (2009). Application of Numenta® Hierarchical Temporal Memory for land-use classification. *South African Journal of Science*, 105(9-10), pp.370-375.

Perumal, S., Norwawi, N. M. and Raman, V. (2015). Internet of Things (IoT) digital forensic investigation model: Top-down forensic approach methodology. *2015*

Fifth International Conference on Digital Information Processing and Communications (ICDIPC). doi:10.1109/icdipc.2015.7323000.

Pollitt, M.M., (2007). An Ad Hoc Review of Digital Forensic Models. In Proceedings of 2nd International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE 2007). pp. 43–54.

Poole, D.L., Mackworth, A.K. and Goebel, R., (1998). Computational intelligence: a logical approach (Vol. 1). New York: Oxford University Press.

Popoola, O.P. and Wang, K., (2012). Video-based abnormal human behavior recognition—a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* (Applications and Reviews), 42(6), pp.865-878.

Poppe, R., (2010). A survey on vision-based human action recognition. *Image and vision computing*, 28(6), pp.976-990.

Prasanthi, B. V. (2016) 'Cyber Forensic Tools: A Review' *International Journal of Engineering Trends and Technology (IJETT)* 41(5); 266- 271

Purdy, S., (2016). Encoding data for HTM systems. *arXiv preprint* arXiv:1602.05925.

Quas, J.A., Goodman, G.S., Ghetti, S., Alexander, K.W., Edelstein, R., Redlich, A.D., Cordon, I.M., Jones, D.P. and Haugaard, J.J., (2005). Childhood sexual assault victims: Long-term outcomes after testifying in criminal court. *Monographs of the Society for Research in Child Development*, pp.1-139.

RADACK, S., (2009). Forensic techniques: helping organizations improve their responses to information security incidents.

Raghavan, S. (2013) Digital Forensic Research: Current State of the Art. *CSI Transactions on ICT,* 1(1), 91–114

Reith, M., Carr, C. and Gunsch, G., (2002). An Examination of Digital Forensic Models. *International Journal of Digital Evidence*, 1(3),1-12.

Rich, E. and Knight, K., (1991). Artificial intelligence. McGraw-Hill, New.

Robson, C. (2002) Real World Research. 2nd Edition, Oxford: Blackwell.

Rowland, J., (2014). ‘Delivering a Modern Criminal Justice Services through Technology’ Criminal Justice System

Ruibin, G., Yun C., and Gaertner, M. (2005). 'Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework', *International Journal of Digital Evidence Spring*, 4, (1).

Russell, S.J. and Norvig, P., (2002). Artificial intelligence: a modern approach (International Edition).

Ramaswamy, S., Rastogi, R., & Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. In ACM Sigmod Record, volume 29 (pp. 427–438).: ACM.

Saravanan, K. and Sasithra, S. (2014). Review on Classification Based on Artificial Neural Networks. *International Journal of Ambient Systems and Applications (IJASA)*, 2(4), 11- 18

Saunders, M., Lewis, P. and Thornhill, A. (2007). Research Methods for Business Students, 4th Ed, Essex: Prentice Hall.

Savoldi, A. and Gubian, P. (2009). Embedded Forensics. *Handbook of Research on Computational Forensics, Digital Crime, and Investigation*, 396-423. doi:10.4018/978-1-60566-836-9.ch017

Sebastian, P. Voon, Y. V. V., Comley, R. (2010). Colour Space Effect on Tracking in Video Surveillance. *International Journal on Electrical Engineering and Informatics* 2(4), 298- 312

Schatz, Bradley, and Andrew J. Clark. (2006) "An open architecture for digital evidence integration.": 15-29.

Sedky, M.H., Moniri, M. and Chibelushi, C.C., (2005). Classification of smart video surveillance systems for commercial applications. *In IEEE Conference on Advanced Video and Signal Based Surveillance, 2005*. (pp. 638-643). IEEE.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. and Poggio, T., (2007). Robust object recognition with cortex-like mechanisms. IEEE transactions on pattern analysis and machine intelligence, 29(3), pp.411-426.

Shilpa, P.H. and Sunitha, M.R., (2016). A Survey on Moving Object Detection and Tracking Techniques. *International Journal of Engineering and Computer Science, 5*(5).

Sommer, P., (2009). Directors and corporate advisors' guide to digital investigations and evidence.

Sprenger, M., Schemm, S., Oechslin, R. and Jenkner, J., (2017). Nowcasting Foehn Wind Events Using the AdaBoost Machine Learning Algorithm. *Weather and Forecasting*, *32*(3), pp.1079-1099.

Stapleton J. (1997). DSDM – Dynamic Systems Development Method. 1st ed. The method in Practice, person education. [Online]. Available from: https://books.google.co.uk/books?hl=en&lr=&id=kp656t4p7soC&oi=fnd&pg=PR11&dq=DSDM+(Dynamic+Systems+Development+Method)&ots=TxGdPvZPTj&sig=cvs-zpzOXGCb7ByryTSbBd2X6dI#v=snippet&q=%20Stapleton&f=false.

Stauffer, C. and Grimson, W.E.L., (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8), pp.747-757.

St-Charles, P.L., Bilodeau, G.A. and Bergevin, R., (2015). Subsense: A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, *24*(1), pp.359-373.

Stephenson, P. (2003). 'Modeling of Post-Incident Root Cause Analysis', *International Journal of Digital Evidence*, 2 (2).

Sternberg, K.J., Lamb, M.E., Davies, G.M. and Westcott, H.L., (2001). The memorandum of good practice: Theory versus application. *Child abuse & neglect*, 25(5), pp.669-681.

Streat, L.G., (2016). A Scalable Flash-Based Hardware Architecture for the Hierarchical Temporal Memory Spatial Pooler. *Rochester Institute of Technology*.

Su, Y., Zhang, J., and Lui, J. (2009). 'Exposing Digital Video Forgery by Detecting Motion-Compensated Edge Artifact'. 2009 International Conference on Computational Intelligence and Software Engineering, 2009. CiSE 2009. 1- 4

Tang, J., Chen, Z., Fu, A. W.-C., & Cheung, D. W. (2002). Enhancing effectiveness of outlier detections for low density patterns. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 535–548).: Springer.

Tian, Y., Kong, Y., Ruan, Q., An, G. and Fu, Y., (2018). Hierarchical and Spatio-Temporal Sparse Representation for Human Action Recognition. *IEEE Transactions on Image Processing,* 27(4), pp.1748-1762.

Tian, Y.L., Brown, L., Hampapur, A., Lu, M., Senior, A. and Shu, C.F., (2008). IBM smart surveillance system (S3): event-based video surveillance system with an open and extensible framework. *Machine Vision and Applications,* 19(5-6), pp.315-327.

Tu, P.H., Doretto, G., Krahnstoever, N.O., Perera, A.A., Wheeler, F.W., Liu, X., Rittscher, J., Sebastian, T.B., Yu, T. and Harding, K.G., (2007). An intelligent video framework for homeland protection. In Defense and Security Symposium (pp. 65620C-65620C). *International Society for Optics and Photonics*.

Tygart Technology Inc. (2014). Crime Scene Post Event Analysis | Tygart Technology. Available at: http://www.tygart.com/products/solution-application/law-enforcement/

UK Government Website. (2017). Criminal justice system statistics quarterly: September 2017. Retrieved from https://www.gov.uk/government/statistics/criminal-justice-system-statistics-quarterly-september-2017

UK Government Website. (2017). Criminal justice system statistics quarterly: September 2017. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/681336/criminal-justice-statistics-quarterly-sept-2017.pdf

UK Government Website. (2018). NIS Directive guidance and Cyber Assessment Framework. Retrieved from https://www.ncsc.gov.uk/

Valjarevic, A. & Venter, Hein. (2012). Harmonised Digital Forensic Investigation Process Model, IEEE

Van, L.P., De Praeter, J., Van Wallendael, G., De Cock, J. and Van de Walle, R., (2015). Performance analysis of machine learning for arbitrary downsizing of pre-encoded HEVC video. *IEEE Transactions on Consumer Electronics*, *61*(4), pp.507-515.

Varadarajan, J. and Odobez, J.M., (2009). Topic models for scene analysis and abnormality detection. In Computer Vision Workshops (ICCV Workshops), *2009 IEEE 12th International Conference on* (pp. 1338-1345). IEEE.

Vaughan, T., (1998). Multimedia: Making it Work (4th ed.). Berkeley, CA: McGraw-Hill.

Verma, B., Zhang, L. and Stockwell, D., (2017). *Roadside Video Data Analysis: Deep Learning* (Vol. 711). Springer.

Verma, G., Gautam, S., Agarwal, R., Saxena, S. and Verma, D., (2018). Implementation of Smart Video Surveillance System Using Motion Detection Technique. In *Sensors and Image Processing* (pp. 65-72). Springer, Singapore.

Wan, H., Wang, H., Guo, G. and Wei, X., (2018). Separability-Oriented Subclass Discriminant Analysis. *IEEE transactions on pattern analysis and machine intelligence*, *40*(2), pp.409-422.

Wang, J., Amos, B., Das, A., Pillai, P., Sadeh, N. and Satyanarayanan, M., (2017). A Scalable and Privacy-Aware IoT Service for Live Video Analytics. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 38-49). ACM.

Wang, W. and Farid, H., (2009). Exposing digital forgeries in video by detecting double quantization. *In Proceedings of the 11th ACM workshop on Multimedia and security* (pp. 39-48). ACM.

Wang, W. and Farid, H., (2006). Exposing digital forgeries in video by detecting double MPEG compression. *In Proceedings of the 8th workshop on Multimedia and security* (pp. 37-47). ACM.

Wang, W. and Farid, H. (2007). Exposing digital forgeries in interlaced and deinterlaced video. IEEE Trans Info Forensics Security 2 (4), 38-49

Wang, Y., Yang, D., Shi, Y., Wang, Y. and Chen, W., (2017). Resource quality prediction based on machine learning algorithms. In *Systems and Informatics (ICSAI), 2017 4th International Conference on* (pp. 1541-1545). IEEE.

Weiming, H. Tan, T., Wang, L. and Maybank, S. (2004) 'A Survey on Visual Surveillance of Object Motion and Behaviors' *Ieee Transactions On Systems, Man, And Cybernetics* 34 (3).

Weston, J., (2014). Support Vector Machine. *Tutorial, http://www. cs. columbia. edu/~ kathy/cs4701/documents/jason_svm_tutorial. pdf, accessed*, *10*(0), pp.0-5.

White, R., 1999. How Computers Work. Indianapolis, IN: QUE

Whitley, D., (2001). An overview of evolutionary algorithms: practical issues and common pitfalls. *Information and software technology*, 43(14), pp.817-831.

Winston, P., (1992). Learning by building identification trees. *Artificialintelligence*, pp.423-442.

Wren, C.R., Azarbayejani, A., Darrell, T. and Pentland, A.P., (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on pattern analysis and machine intelligence,* 19(7), pp.780-785.

Xu, T., Agrawal, A., Abashin, M., Chau, K.J. and Lezec, H.J., (2013). All-angle negative refraction and active flat lensing of ultraviolet light. *Nature*,497(7450), pp.470-474.

Wu, C. (2010). A Handoff Algorithm of Dynamic Decisions with Extenics for Wireless Cellular Networks. *Journal of Electrical and Computer Engineering*, *2010*, 1-5. doi:10.1155/2010/656593

Xu, Y., Korba, L., Wang, L., Hao, Q., Shen, W. and Lang, S., (2003). A security framework for collaborative distributed system control at the device-level. In Industrial Informatics, 2003. INDIN 2003. *Proceedings. IEEE International Conference on* (pp. 192-198). IEEE.

Yang, J., Yu, K., Gong, Y. and Huang, T., (2009). Linear spatial pyramid matching using sparse coding for image classification. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 1794-1801).

Yaseen, M.U., Anjum, A., Rana, O. and Hill, R., (2018). Cloud-based scalable object detection and classification in video streams. *Future Generation Computer Systems*, *80*, pp.286-298.

Yeh, C.H., Lin, C.Y., Muchtar, K., Lai, H.E. and Sun, M.T., (2017). Three-Pronged Compensation and Hysteresis Thresholding for Moving Object Detection in Real-Time Video Surveillance. *IEEE Transactions on Industrial Electronics*, *64*(6), pp.4945-4955.

Yusoff, Y., Ismail, R. and Hassan, Z., (2011). Common phases of computer forensics investigation models. *International Journal of Computer Science & Information Technology (IJCSIT)*, 3(3), pp.17-31.

Zhang Y. And Liu, Z (2007) 'Irregular behavior recognition based on trading track' in Proc. *Int. Conf. Wavelet Anal. Pattern Recog., 2007*, pp. 1322–1326

Zhang, J., Su, Y. and Zhang, M., (2009). Exposing digital video forgery by ghost shadow artifact. In Proceedings of the First ACM workshop on Multimedia in forensics (pp. 49-54). ACM.

Zhang, J., Zhang, Y., & Chen, B. (2017). Out-of-Focus Projector Calibration Method with Distortion Correction on the Projection Plane in the Structured Light Three-Dimensional Measurement System. *Sensors*, *17*(12), 2963. doi:10.3390/s17122963

Zhang, T., Chowdhery, A., Bahl, P.V., Jamieson, K. and Banerjee, S., (2015), The design and implementation of a wireless video surveillance system. In *Proceedings of the 21st Annual International Conference September on Mobile Computing and Networking* (pp. 426-438). ACM.

Zhang, X., Zhang, J. and Zhong, J., (2017). Skill Learning for Intelligent Robot by Perception-Action Integration: A View from Hierarchical Temporal Memory. *Complexity*, *2017*.

Zhituo, X., Hao, R. and Hao, W., (2012). A Content-Based Image Retrieval System Using Multiple Hierarchical Temporal Memory Classifiers. *In Computational*

Intelligence and Design (ISCID), 2012 Fifth International Symposium on (Vol. 2, pp. 438-441).

Zhuo, W., Cao, Z., Qin, Y., Yu, Z. and Xiao, Y., (2012). Image classification using HTM cortical learning algorithms. *In Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 2452-2455).

Zjalic, J. 2017. The CSI Effect? Expectations Vs Limitations. Available at: https://articles.forensicfocus.com/2017/11/30/the-csi-effect-expectations-vs-limitations/

Zouaoui, R., Audigier, R., Ambellouis, S., Capman, F., Benhadda, H., Joudrier, S., Sodoyer, D. and Lamarque, T., (2015). Embedded security system for multi-modal surveillance in a railway carriage. In *Optics and Photonics for Counterterrorism, Crime Fighting, and Defence XI; and Optical Materials and Biomaterials in Security and Defence Systems Technology XII* (Vol. 9652, p. 96520C).