2019

FACIAL IDENTIFICATION FOR DIGITAL FORENSIC

AL-KAWAZ, HIBA

UNIVERSITY OF
PLYMOUTH

FACIAL IDENTIFICATION FOR DIGITAL FORENSIC

by

HIBA MOHAMMED AL-KAWAZ

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

July 2019

Acknowledgement

First of all, I would like to thank Allah and his Merciful for all He has given me throughout my life, in general, and for giving me the potential and patience to persevere and reach this stage of my PhD research, in particular. Without Him, I would not have achieved anything or even existed.

This thesis would not have been completed without the guidance, support, patient, and feedback from my Director of Studies Professor Nathan Clarke. I would like to express my special thanks and appreciation to him. Thanks must also go to my other supervisors: Professor Steven Furnell, Dr Fudong Li, who have spent a lot of time and efforts proof reading papers and my thesis, in addition to providing helpful advices throughout my studies.

I am very much owe to my family and strength in Iraq specially my mother and father who support me with their love, prayers, and continuous encouragement during my PhD research journey.

My unreserved love, thanks and appreciation must go to my wonderful husband (Mohammed) and my little angels (Hassan and Hur) who have been very patient, understanding, and inspiring to me throughout this hard journey, spending days, and nights without me. I hope the potential success of this research will compensate some of what they have missed.

I would like to thank all my friends and colleagues in CSCAN who have contributed positively in my studying life even if it was just encouragement word.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a scholarship from Republic of Iraq / Ministry of Higher Education and Scientific Research – Baghdad University.

Publication (or public presentation of creative research outputs):

1. Al-Kawaz, H., Clark, N., Furnell, S.M., Li, F. and Alburan, A., 2018, June. Advanced facial recognition for digital forensics. In 17th European Conference on Cyber Warfare and Security (pp. 11-19). Academic Conferences and Publishing International Limited.
2. Mashhadani, S., Al-kawaz, H., Clarke, N., Furnell, S. and Li, F., 2017, December. A novel multimedia-forensic analysis tool (M-FAT). In Internet Technology and Secured Transactions (ICITST), 2017 12th International Conference for (pp. 388-395). IEEE. DOI: https://doi.org/10.23919/icitst.2017.8356429
3. Mashhadani, S., Al-kawaz, H., Clarke, N., Furnell, S. and Li, F., 2018. The Design of a Multimedia-Forensic Analysis Tool (M-FAT), International Journal Multimedia and Image Processing (IJMIP), Volume 8, Issue 1. DOI: https://doi.org/10.23919/icitst.2017.8356429
4. Al-Kawaz, H., Clarke, N., Furnell, S. and Li, F., 2018. Facial-Forensic Analysis Tool. Digital Investigation, 26, p.S136, DFRWS. DOI: https://doi.org/10.1016/j.diin.2018.04.008.

Word count of main body of thesis: 41780 words.

Signed ……..

Date 29/07/2019

Abstract

FACIAL IDENTIFICATION FOR DIGITAL FORENSIC

Hiba Mohammed Al-Kawaz

Forensic facial recognition has become an essential requirement in criminal investigations as a result of the emergence of electronic devices, such as mobile phones and computers, and the huge volume of existing content. Forensic facial recognition goes beyond facial recognition in that it deals with facial images under unconstrained and non-ideal conditions, such as low image resolution, varying facial orientation, poor illumination, a wide range of facial expressions, and the presence of accessories. In addition, digital forensic challenges do not only concern identifying an individual but also include understanding the context, acknowledging the relationships between individuals, tracking, and numbers of advanced questions that help reduce the cognitive load placed on the investigator.

This thesis proposes a multi-algorithmic fusion approach by using multiple commercial facial recognition systems to overcome particular weaknesses in singular approaches to obtain improved facial identification accuracy. The advantage of focusing on commercial systems is that they release the forensic team from developing and managing their own solutions and, subsequently, also benefit from state-of-the-art updates in underlying recognition performance. A set of experiments was conducted to evaluate these commercial facial recognition systems (Neurotechnology, Microsoft, and Amazon Rekognition) to determine their individual performance using facial images with varied conditions and to determine the benefits of fusion. Two challenging facial datasets were identified for the evaluation; they represent a challenging yet realistic set of digital forensics scenarios collected from publicly available photographs. The experimental results have proven that using the developed fusion approach achieves a better facial

identification rate as the best evaluated commercial system has achieved an accuracy of 67.23% while the multi-algorithmic fusion system has achieved an accuracy of 71.6%.

Building on these results, a novel architecture is proposed to support the forensic investigation concerning the automatic facial recognition called Facial-Forensic Analysis System (F-FAS). The F-FAS is an efficient design that analyses the content of photo evidence to identify a criminal individual. Further, the F-FAS architecture provides a wide range of capabilities that will allow investigators to perform in-depth analysis that can lead to a case solution. Also, it allows investigators to find answers about different questions, such as individual identification, and identify associations between artefacts (facial social network) and presents them in a usable and visual form (geolocation) to draw a wider picture of a crime. This tool has also been designed based on a case management concept that helps to manage the overall system and provide robust authentication, authorisation, and chain of custody.

Several experts in the forensic area evaluated the contributions of theses and a novel approach idea and it was unanimously agreed that the selected research problem was one of great validity. In addition, all experts have demonstrated support for experiments' results and they were impressed by the suggested F-FAS based on the context of its functions.

Table of Contents

1 Introduction 1
1.1 Overview 1
1.2 The Research Aim 7
1.3 Thesis Structure 7
2 Biometric System 9
2.1 Introduction 9
2.2 Performance Measuring 11
2.3 Multibiometrics 14
2.4 Multibiometric Fusion Techniques 15
2.5 Biometrics in Law Environment 16
2.6 Conclusion 18
3 Digital Forensics and Facial Recognition 19
3.1 Introduction 19
3.2 Digital Forensic Science 19
3.2.1 The Digital Forensics Lifecycle 20
3.2.2 Digital Forensics Tools 23
3.3 Forensic Facial Recognition Systems 25
3.4 Facial Recognition Processing 28
3.5 Forensic Facial Recognition System Barriers 30
3.6 Conclusion 37
4 The Current State of Art 38
4.1 Introduction 38
4.2 Research Methodology 38
4.3 Facial Ageing 39
4.4 External Factors 50
4.5 Internal Issues 57
4.6 Soft Biometric Attributes 70
4.7 Commercial Face Identification Systems 74
4.8 Discussion 76
4.9 Conclusion 79
5 An Investigation Into Forensic Facial Recognition 84
5.1 Introduction 84
5.2 Experimental Methodology 85
5.2.1 Recognition Algorithms 85
5.2.2 Facial Image Datasets 87
5.2.3 Experimental Methodology 95
5.3 Experimental Results 106

5.3.1 Results of Experiment 1 106
5.3.2 Results of Experiment 2 109
5.3.3 Results of Experiment 3 113
5.4 Discussion 116
5.5 Conclusion 119
6 A Novel Architecture for Facial-Forensic Recognition 121
6.1 Introduction 121
6.2 Facial-Forensic Analysis System (F-FAS) Requirements 122
6.3 The F-FAS Architecture 124
6.3.1 Acquisition Engine 126
6.3.2 The Pre-Processing Engine 129
6.3.3 Facial Recognition Engine 135
6.3.4 Forensic Analyses Engine 140
6.3.5 Reporting 145
6.3.6 The F-FAS Management 146
6.4 The F-FAS Implementation 149
6.5 The F-FAS Prototype Samples 151
6.6 Conclusion 161
7 System Evaluation 162
7.1 Introduction 162
7.2 Research Methodology 163
7.2.1 Evaluations Questions 163
7.2.2 The Participants 164
7.3 Interviewees' Responses' Evaluation 169
7.3.1 Thoughts on the Research Problem 169
7.3.2 Thoughts on Current Facial Recognition Systems 170
7.3.3 Thoughts on the Multi-Algorithmic Fusion Experiment 171
7.3.4 The Architecture of The F-FAS realisable in Term of Digital Forensics 173
7.3.5 The F-FAS Capabilities to Reduce Effort and Time for Data Analysis .. 174
7.3.6 Effectiveness, Reliability, and Usability of the F-FAS Functions 175
7.3.7 Thoughts on the System Case Management 177
7.3.8 Strengths and Weaknesses of the F-FAS Approach 178
7.3.9 Suggested Enhancement 179
7.3.10 Further comments 179
7.4 Conclusion 180
8 Conclusion and Future Work 182
8.1 Contributions and Achievements of the Research 182

8.2 Limitations of the Research 184
8.3 Scope for Future Work..... 185
8.4 The Future of Facial Recognition in Digital Forensics 186
References 188

List of Figures

Figure 1-1: suspect 1, Tamerlan Tsarnaev. Suspect 2, Dzhokhar Tsarnaev (Klontz and Jain, 2013). 3
Figure 1-2: Three suspects in Belgium's Zaventem Airport attack in 2016 (Shoichet et al., 2016). 4
Figure 1-3: Age-progression image of missing child, Madeleine McCann (a) original image at age 4, and (b) the progression version 5
Figure 2-1: Biometric trait examples (Jain and Ross, 2015) 9
Figure 2-2: Biometric system task (Enrollment, Verification, and Identification) (Jain et al., 2004) 11
Figure 2-3: Biometric metrics factors (Clarke, 2011) 12
Figure 2-4: Example of a ROC curve (El-Abed and Charrier, 2012) 13
Figure 2-5: Cumulative Match Characteristic (zvetcobiometric, 2012) 14
Figure 3-1: The main lifecycle steps of digital forensics 21
Figure 3-2: The main steps in face recognition process 28
Figure 3-3: The image on the right is facial detection; on the left is facial normalisation (Li and Jain, 2011) 29
Figure 3-4: Some forensic facial recognition problems (Li and Jain, 2011) 31
Figure 3-5: Face aging appearance (Jain et al., 2011). 33
Figure 3-6: The two face-ageing schemes (Jain et al., 2011) 34
Figure 3-7: Two categories of occlusion: (a) dense occlusion faces, (b) sparse occlusion faces (Min and Dugelay, 2012). 35
Figure 3-8: The sketch types: (a) viewed sketches and their corresponding photographs, (b) forensic sketches and the corresponding photograph with poor quality, and (c) forensic sketches and the corresponding photograph with poor quality (Klare et al., 2011) 36
Figure 4-1: (a) full-face image, (b) periocular region normalisation, (c) illumination correction, (d) feature extraction, (e) pose correction for full face, (f) periocular region normalisation, (g) illumination correction, (h) feature extraction 43
Figure 4-2: Process of illumination-aware age progression (Kemelmacher-Shlizerman et al., 2014) 44
Figure 4-3: The idea of MDL (Sungatullina et al., 2013) 47
Figure 4-4: Facial recognition performance versus the number of PCs (Buciu, 2010) .. 51
Figure 4-5: Example of the shadow compensated images in each pose (Choi et al., 2011) 53
Figure 4-6: The recognition rates results (Ishimoto and Chen, 2009) 57
Figure 4-7: (a), (b), and (c) are input images. (d) Mosaiced face generation. (e) The final image (Singh et al., 2007) 60
Figure 4-8: The three stages of the proposal system (Cament et al., 2015) 62
Figure 4-9: The full system approach (Lee et al., 2012) 63
Figure 5-1: Example images from the CAS-PEAL-R1 dataset: (a) Accessory subset, (b) Expression subset, (c) Lighting subset, and (d) Pose subset (Gao et al., 2008) 90
Figure 5-2: Samples of the Realistic dataset used in the test set. A) Photo samples for Adam Sandler and B) Photo samples of Alec Baldwin 94
Figure 5-3: Samples of the enrolment dataset collected for the study 95
Figure 5-4: Flowchart of the Neurotechnology facial recognition algorithm 97
Figure 5-5: Flowchart of the Microsoft Face API process 99
Figure 5-6: Flowchart of the Amazon Rekognition API process 101
Figure 5-7: Multi-algorithmic fusion approach flow diagram 104

Figure 5-8: Some of testing photos that failed in acquired templates..... 112
Figure 5-9: Some of testing photos that matched incorrectly..... 112
Figure 5-10: Performance comparison between the fusion method and the other systems when using the CAS-PEAL-R1 dataset..... 115
Figure 5-11: Performance comparison between the fusion method and the other systems when using the Realistic dataset..... 116
Figure 6-1: A Novel architecture of the F-FAS..... 125
Figure 6-2: Acquisition Engine..... 127
Figure 6-3: Pre-Processing Engine levels..... 131
Figure 6-4: Facial recognition Engine..... 137
Figure 6-5: Forensic Analysis Engine..... 142
Figure 6-6: Forensic Examiner dashboard..... 152
Figure 6-7: The case dashboard..... 153
Figure 6-8: The 'add new forensic image' window..... 153
Figure 6-9: The facial recognition options window..... 154
Figure 6-10: Soft biometric option of facial recognition searching..... 155
Figure 6-11: Facial recognition results window..... 155
Figure 6-12: History facial recognition searching..... 156
Figure 6-13: Facial Social Network Visualisation..... 157
Figure 6-14: Geolocation Visualisation..... 157
Figure 6-15: Facial Modification window (ageing example)..... 158
Figure 6-16: Reporting creation window..... 159
Figure 6-17: An example of reporting design..... 159
Figure 6-18: Admin dashboard..... 160
Figure 6-19: Investigator's accounts window..... 160

List of Tables

Table 4-1: Identification accuracy (Park et al., 2010). 40
Table 4-2: Recognition rates for two database. 52
Table 4-3: Databases description (Lee et al., 2012) 64
Table 4-4: Facial image databases description (Sultana et al., 2014) 65
Table 4-5: Database description and methods used in experiments of (Liao et al., 2013) 67
Table 4-6: Accuracy of the face and biometric identification (Tiwari et al., 2012) 73
Table 4-7: Summarise the current state of art 83
Table 5-1: Comparison between the three systems used in the research 87
Table 5-2: Overview of some current facial databases 88
Table 5-3: Details of the CAS-PEAL-R1 dataset (Gao et al., 2008) 89
Table 5-4: Subsets of the CAS-PEAL-R1 dataset used in Experiment 1 91
Table 5-5: Splitting of the CAS-PEAL-R1 dataset 102
Table 5-6: Division of the Realistic facial dataset 102
Table 5-7: Experiment 1 results (using CAS-PEAL-R1 dataset) 108
Table 5-8: FPIR and FTA rates for Experiment 1 109
Table 5-9: Experiment 2 results (performance of the three commercial systems with the Realistic dataset) 110
Table 5-10: FPIR and FTA rates for experiment 2 for the Realistic dataset 111
Table 5-11: The weight values for three systems for two datasets 114
Table 6-1: Case table 129
Table 6-2: Forensic images table 129
Table 6-3: Photo evidence table 135
Table 6-4: Photo metadata table 135
Table 6-5: Face photo table 135
Table 6-6: Examples of Amazon Rekognition's prices. 137
Table 6-7: Query details of Facial Recognition Engine 140
Table 6-8: Additional Filter information for facial recognition process 140
Table 6-9: Facial recognition results 140
Table 6-10: The F-FAS' responsibilities and roles 149
Table 6-11: Investigators' details 149

1 Introduction

1.1 Overview

Facial recognition has come to play an important role in criminal investigations with the advent of electronic devices, such as the Closed-Circuit Television (CCTV), digital cameras, mobile phones, and computers. By using facial recognition, valuable information used in the detection of culprits can be extracted from images and/or videos that are found at crime scenes. According to IHS Markit Technology, approximately 245 million video surveillance systems were installed around the world in 2014 (Jenkins, 2015). However, this number has increased daily because of an increase in installed CCTV in public buildings such as hospitals, hotels, and schools and for personal use inside and outside homes, especially in capital cities where the crime rate is high. These CCTV will provide billions of hours' footage weekly, offering an enormous amount of information that can be used to track down suspects potentially when crimes occur (Jain et al., 2012). Regarding other resources, photos and videos are released daily by internet users and this rate has increased significantly since they started using social media, such as Facebook, Instagram, WhatsApp, Snapchat, and YouTube. For example, an activity usually done on social networking sites rose to 64% in 2013 from 53% in 2011. According to the UK national survey, the online image posting rate rose from 53% in 2011 to 64% in 2013 and it became the most viewed online entertainment, and surpassing listening to music (Daton et al., 2013).

A survey published by the Pew Research Center's Internet Project found that 54% of internet users upload or share photos or videos on social media applications (Duggan, 2013). These visual data have helped investigators to extract evidence

to solve ambiguity regarding criminals' identities or their relationship with other suspects.

During a digital forensic investigation, all available evidence is collected and later analysed. The analysis process will categorise files into different types then an investigator can view and examine these files. The use of photo and video files in an investigation helps investigators track suspects faces, their locations, the time, who appeared with them, and their activities (Carrier, 2003). However, this investigation type faces some drawbacks. Here are some examples:

- on April 15, 2013, two bombs exploded near the finishing line of the Boston Marathon, killing three people and injuring 264 others (Klontz and Jain, 2013). The FBI collected all CCTV videos around the location showing people who attended. Specialists then reviewed evidence. Despite the huge amount of video footage collected by the FBI around the crime scene and the fact that their photos had been saved on a USA government database, an automated face-matching system could not identify the perpetrators (Klontz and Jain, 2013). The FBI had to depend on traditional methods of identifying suspects, i.e., by asking the public for assistance or information (Gallagher, 2013). After three days of examination, two images of the suspects were released, as shown in Figure 1-1. The images were of two brothers: the elder one, Tamerlan Tsarnaev, was wearing a black hat and glasses while his brother, Dzhokhar Tsarnaev was in a white hat. These accessories were one of are as potential issues may hinder automated face recognition software from performing well. Furthermore, low-quality photos of faces and an uncooperative subject can produce images with the face in a variety of

orientations, including being occluded by other objects, which all seek to increase the matching complexity.

Figure 1-1: suspect 1, Tamerlan Tsarnaev. Suspect 2, Dzhokhar Tsarnaev (Klontz and Jain, 2013).

- In Belgium's Zaventem Airport attack on 22 March 2016, the police released photos of three suspects as shown in Figure 1-2. Two of them were brothers and were reported dead from the attack while the third member, who was wearing a hat, got away (Shoichet et al., 2016). The investigators tried to discover the third man’s identity by tracking his hat and coat on all surveillance videos. In doing so, they determined several locations where the suspect might have been. However, some issues made investigators job more complex. These issues were either related to the long distance between the camera and the person, which resulted in a low-resolution photo and illumination change or the use of accessories (glasses and hat). Furthermore, expression changes and different facial poses can cause a failure in matching facial recognition.

Figure 1-2: Three suspects in Belgium's Zaventem Airport attack in 2016 (Shoichet et al., 2016).

- There was an outbreak of rioting in London between the 6th and 10th of August 2011. Subsequently, the Metropolitan Police released hundreds of photographs of people, collected either from news crews, private mobile phones from people with recordings of the violence that unfolded, and CCTV cameras (Klontz and Jain, 2013). A few arrests come from personal footage shared on Facebook and other social media. Other arrests were dependent on witnesses to reach rioters' identities. While automatic facial recognition was used to identify suspects, few rioters' identities were discovered for a number of reasons (Hill, 2011):
    1. The low quality of photos (unclear faces);
    2. Most of the culprits were disguised using hats, glasses, masks, and scarves;
    3. The bad lighting; and
    4. Facial poses and expressions varied. All these reasons caused unsuccessful facial recognition using technology.
- According to the National Centre for Missing and Exploited Children (NCMEC), approximately 800,000 children are reported as missing every year in the United States alone (abcNews, 2013). The time factor is considered critical in missing children cases because children's facial

appearances and body sizes can change quickly; hence, as time elapses, the probability of a missing child being found decreases. An example of a missing child case is Madeleine McCann's—a 4-year-old from Leicester, UK. She disappeared from her bedroom in 2007 while on holiday in Portugal with her family (BBCHome, 2007). Police released images of what they think she looks like by using age progression software. Figure 1-3 is shown a new image for Madeleine (Telegraph, 2016). Despite this new image, the case remains unsolved.

(a) (b)

Figure 1-3: Age-progression image of missing child, Madeleine McCann (a) original image at age 4, and (b) the progression version

Identification of offenders or terrorists from a criminal watch list after the passage of time is another forensic facial recognition application that involves the factor of ageing. Criminals might commit a crime twenty years ago and, often, a database is not updated during the intervening years (Li and Jain, 2011). If such cases remain unsolved, facial ageing techniques can play a powerful role in helping police to find missing people after some time has passed.

- On March 4, 2018, the previous Russian spy Sergey Skripal and his daughter Yulia were poisoned by the nerve agent Novichok in Salisbury. The Metropolitan Police began its investigation by watching thousands of

hours of CCTV footage (Medeiros, 2018). Therefore, they used a specialist team (super-recogniser) who have super talent in recognising people when they've only seen them once. This team has the ability to recognise people even from partial captures of the face and the back side of the head (Barry, 2018). The team started the investigation by looking through CCTV footage and they depended on victims' movements. After a few days, Prime Minister Theresa May announced the investigation found evidence that the attack was by two agents of the Russian military intelligence service (Medeiros, 2018). Despite the current good facial recognition algorithms, the Metropolitan Police preferred to use the super-recognisers in their investigation.

As demonstrated by the aforementioned examples, visual evidence and facial recognition, in particular, are valuable investigative tools. Nevertheless, despite a large amount of academic and commercial research effort, automatic facial recognition still suffers from several major drawbacks in achieving accuracy, a delay in tracking suspects, and the failure to identify suspects. These reasons can be mainly categorised in two types: external or internal factors. External challenges can be found in the imaging system, such as illumination (e.g., poor lighting conditions), the facial pose for the camera (i.e., faces are not necessarily frontal to the camera), and poor sensor quality. Meanwhile, internal variations, such as facial aging, expression, and cosmetic changes (e.g. hat, glasses, makeup, etc.), are more related to individuals (Li and Jain, 2011). In addition, the huge volume of visual/image evidence that are required to be analysed can be another factor in prolonging investigation times (Jain et al., 2012). Despite that the current facial recognition tools provide identification decisions, they do not

allow an investigator to ask more complex questions of the data or visualise the data in a more meaningful way to reduce the cognitive burden on the investigator.

Therefore, a sustained and collaborative effort is needed to enhance the performance of facial recognition in the field of forensic investigation. This project will attempt to contribute to helping and supporting law enforcement in their investigation and to identify culprits in a short period and with high levels of accuracy.

1.2 The Research Aim

The aim of this research is to contribute to developing techniques to aid automatic facial recognition and analysis of persons in the context of a forensic investigation. The research will include the following objectives:

- Perform a comprehensive literature review in the forensic facial recognition domain.
- Select and evaluate a number of current facial recognition algorithms.
- Investigate and analyse fusion-based approaches seeking to improve recognition performance.
- Propose a novel facial forensic analysis tool to enable the ability to examine and analyse multimedia files in a forensically sound and effective manner.
- Design a prototype that could reflect the system architecture in reality.
- Evaluate the proposed framework.

1.3 Thesis Structure

The thesis is structured into eight chapters as follows:

Chapter 2 provides the background information and knowledge on biometric systems in terms of their techniques, performance measuring, multibiometric fusion aspects, and using biometrics in the enforcement investigation.

Chapter 3 presents a comprehensive presentation of digital forensics in terms of the baseline concept, common processes, and some of the forensic tools. In addition, it seeks to identify forensic facial recognition issues.

Chapter 4 demonstrates the current state of the art with respect to forensic facial recognition. This chapter seeks to establish knowledge of the limitations and discusses current challenges.

Chapter 5 presents a series of experiments that conduct a number of facial recognition systems based on two facial databases by describing the methodology for experiments and further presents, analyses, and discusses the results.

Chapter 6 presents the novel system architecture, including a novel framework for facial forensic analysis followed by a detailed explanation of the proposed design and suggested system implementation.

Chapter 7 provides an expert-based evaluation. Experts, including academics and practitioners, are interviewed to explore another area for the development of the proposed system, as well as to identify the strengths and weaknesses of the system.

Chapter 8 summarises the findings arising from the research, highlighting major achievements and constraints.

2 Biometric System

2.1 Introduction

The case studies of the previous chapter have shown the field of facial recognition in law enforcement is still considered an open area and needs more effort to cover. The field of facial recognition fits within the wider area of biometrics. As such, it is prudent to explore the fundamental knowledge within this domain to understand the core function and operation of such systems.

A biometric system is technology that uses attributes of humans to identify individuals. This technology has appeared sequentially based on evolution and scientific discoveries. It could be classified into physical biometrics, such as fingerprints, face, iris, and ear, chemical biometrics such as blood and DNA, and behavioural attributes such as keystroke, gait, and voice (Flynn et al., 2008). Figure 2-1 illustrates a number of biometric examples.

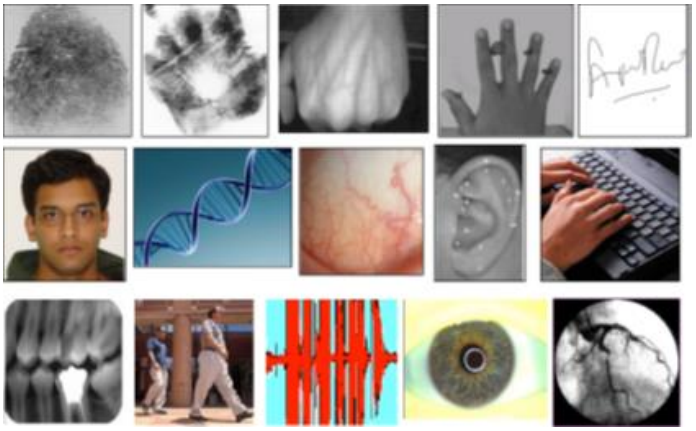


Figure 2-1: Biometric trait examples (Jain and Ross, 2015)

Most biometric traits share a number of general requirements (Jain et al., 2007):

- **Universality**: each person should have at least one biometric sample that appears in Figure 2-1 to use in the biometric identification system.

- **Distinctiveness**: a chosen biometric trait needs to have some level of unique characteristics so it can be used to discriminate users. For example, a fingerprint is unique while hair colour is not.
- **Permanence**: the ability for the biometric trait to be invariant over time.
- **Collectability**: how easily a sample of the biometric trait can be collected.

In addition, other characteristics include system performance, accuracy, and acceptability by users (e.g., some users may prefer fingerprints over iris scans because they believe scanning one's iris may be harmful). As a result, biometric traits have to meet some degree of the previously mentioned characteristics in order for them to be used in a biometric system.

The biometric system is operated in two main phases: enrolment and recognition. The enrolment phase generates the digital representation of an individual's biometric attributes and then stores it in the system database. The recognition phase typically is worked in two categories (Jain et al., 2008):

- **Identification**: aims to recognise the subject's identity by comparing a sample with all the subjects in the database (one to many).
- **Verification**: depends on verifying a subject's identity by comparing it with a single template belonging to the claimed identity (one to one).

Therefore, the identification scheme tries to determine to whom the biometric sample belongs, answering the "Who is this person?" question. In the verification scheme, it checks whether a biometric sample belongs to the person to whom it is claimed to belong. The question here is "Is this person X?" In general, identification is harder than verification because the former system performs a large number of comparisons in comparing with the latter. Figure 2-2 illustrates the difference between two types while user enrolment is applied in two tasks.

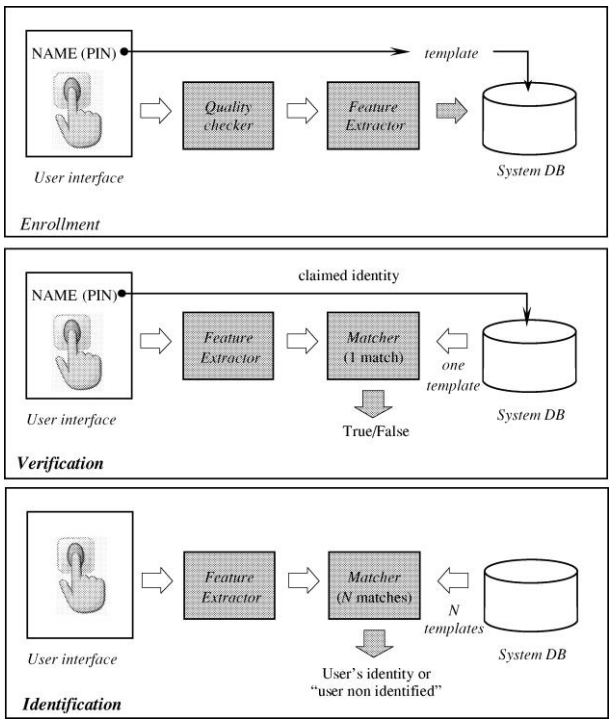


Figure 2-2: Biometric system task (Enrollment, Verification, and Identification) (Jain et al., 2004)

2.2 Performance Measuring

To evaluate a biometric system, several methods are commonly used. For verification-based biometric systems, the output is measured as either a match or non-match. Whereas biometric systems operate in the identification mode, the output is presented in an order list that contains results from the best to worst matching (DeCann and Ross, 2013). The performance metrics for the verification mode are defined as:

- **False Acceptance Rate (FAR)**: The extent to which the system accepts impostors
- **False Rejection Rate (FRR)**: The extent to which the system rejects legitimate users.

The relation between these two rates is shown in Figure 2-3: as one rate increases the other decreases. If the threshold setting is tight (i.e., requires a high

security level), it might reject more authorised users from logging into the system (i.e., high FRR) but increase the system protection. On the contrary, if the threshold setting is slack (i.e., requires low-level security), it might allow more unauthorised users to log into the system (i.e., high FAR) with high user convenience but low system security. A third metric is the *equal error rate* (EER), representing the meeting point for FAR and FRR rates (i.e., at this point, FAR an FRR are equal), and is usually used as a performance metric for comparing different biometric techniques. Furthermore, the lower EER value indicates better system performance. The verification system uses a Receiver Operating Characteristic (ROC) curve to represent the performance metrics FAR and FRR, where a plot of the rate of FAR (i.e., accepted unauthorised users) on the x-axis against the corresponding rate of FRR (i.e., rejected authorised users) on the y-axis is plotted. An illustration of a ROC curve is presented in Figure 2-4 (El-Abed and Charrier, 2012).

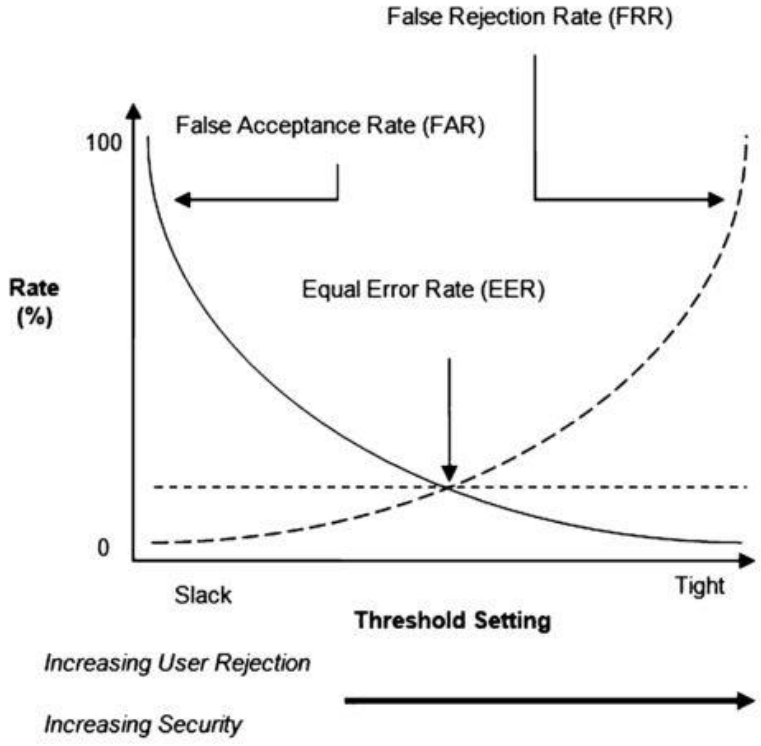


Figure 2-3: Biometric metrics factors (Clarke, 2011)

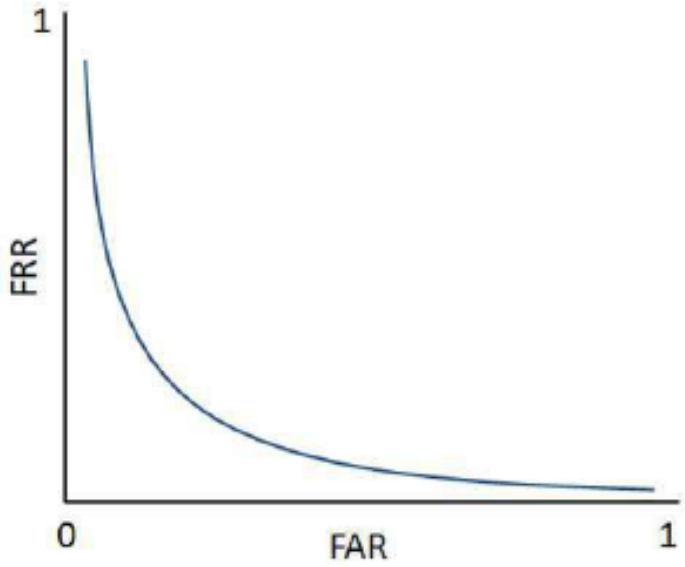


Figure 2-4: Example of a ROC curve (El-Abed and Charrier, 2012)

Biometric identification systems determine a suspect's identity as either *closed-set* or *open-set* identification (Poh et al., 2011). In *closed-set* identification, it is assumed that a suspect's biometric sample is stored in the database while for *open-set* identification needs to identify if the biometric sample is in the database or not. Generally, the real world applications operate on *open-set* principle. For example, typically, law enforcement matches the suspect's facial image with a passport database. Normally, in the identification technique, the input features are compared with all samples in the database to determine the top match. The top match indicates the largest similarity score from all matching results. The identification rate is represented by the probability of the suspect's face being mapped to an identity. Furthermore, the rank-k indicates the rates of the correct identity among the top k matches that determine as match score (Flynn et al., 2008). The performance metric for the identification technique is the Cumulative Match Characteristic (CMC), which depends on identification rate at rank-k. Figure 2-5 represents an example of the CMC where the identification rate achieved 80% at rank-1 and close to 100% at rank 80.

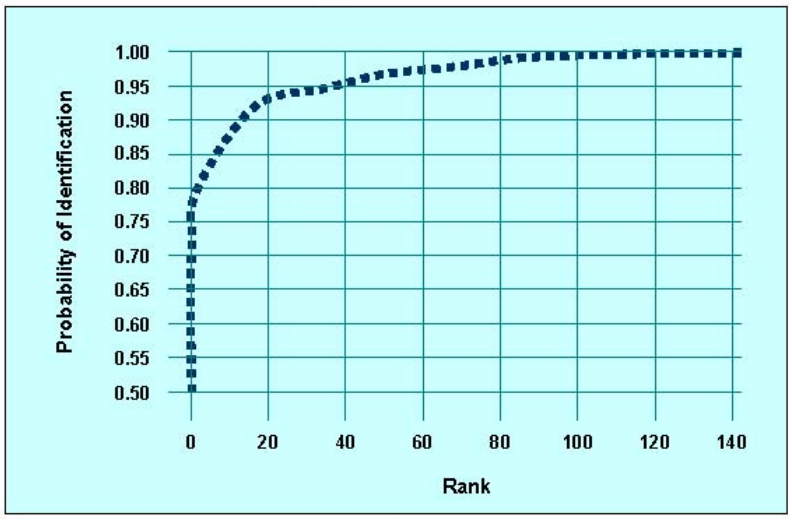


Figure 2-5: Cumulative Match Characteristic (zvetcobiometric, 2012)

In addition, there are other metrics that could be used in the evaluation of the biometric system, such as speed, number of templates, and cost. The time consumed by any biometric system to identify an individual is critical. The competitive criteria are looking for an identification system that computes with a short time and high accuracy. Moreover, the number of templates is considered in any biometric system and the limitation in memory can add another complexity in the performance of a biometric system.

2.3 Multibiometrics

A multibiometric system seeks to overcome some of the issues surrounding the use of single modalities. For example, if the poor person's fingerprint quality prevents him from enrolling in the system, then the use of other biometric attributes, such as the face, would help in the recognition task. Furthermore, multibiometric systems increased the security for some systems by making it difficult for unauthorised persons to spoof multiple biometric traits. These barriers can be solved by fusing and consolidating the resulting information from multiple biometric sources into a system referred to as a multibiometric system (de Oliveira et al., 2010). Numerous researchers have demonstrated that

multibiometric systems outperform single biometric systems in performance matching, globalisation, and resistance in front of unauthorised attacks (Ross et al., 2006, Jain et al., 2005), so the overall system accuracy is improved.

There are different approaches of multibiometric systems which combination of two or more biometric techniques or attributes. These approaches that are possible in a multimodal biometric system include (Ross et al., 2006):

- Multi-modal: using more than one biometric trait (e.g., face, fingerprint, and iris).
- Multi-sensor: employing more than one sensor to capture a single biometric trait (e.g., optical and capacitive fingerprint sensors).
- Multi-instance: using more than one subtype of the same biometric trait (e.g., the left index finger and the right index finger).
- Multiple sample: using more than one sample of the same biometric trait (e.g., multiple face pictures of a person acquired under different pose/illumination conditions)
- Multiple algorithmic: using more than one matcher algorithm in the classification process (e.g., multiple fingerprint matches based on minutiae or filtering)

2.4 Multibiometric Fusion Techniques

This diversity of multibiometric approaches seeks to improve the recognition decision in the verification and identification tasks. The combination process is called fusion. In general, the fusion process could occur at different levels, as follows (Ross et al., 2008).

- Sensor fusion level: Initial biometric data is collected before extracting the feature. This data is captured by multiple sensors or a single sensor

obtaining multiple samples. For example, fusing different face images from one or different cameras.

- Feature fusion level: After collecting multiple samples from one or different biometric traits, the feature vector is extracted from each sample by using different algorithms. Then these vectors are combined for use in the next matching phase. For example, fusing facial and fingerprint features.
- Matching score fusion level: The output results from each biometric classifier are combined at this level to generate a new matching score result to be used then for the decision process. It is believed to be the most accurate type of fusion and is, therefore, most commonly used (Ross et al., 2006).
- Decision fusion level: This type of fusion occurs when each individual biometric system presents its own decision to enable the final decision of the recognition system.

2.5 Biometrics in Law Environment

Biometrics are mainly used in three areas, commercial, such as e-commerce (bank logins), computer network logins, and medical records management; governmental, such as ID cards, driver's licences, passport control, and social security; and forensics or law enforcement, such as parenthood determination, criminal investigation, and terrorist identification (Prabhakar et al., 2003). Each area could use a biometric system either in verification or identification schemes depending on need.

The correlation between biometrics science and forensics is based on the identification of people individuals. Biometric evidence is important in digital forensics to provide valuable information to identify people (Jain and Ross, 2015). The identity could be determined by using personal characteristics found in some

traces on the crime scene such as fingerprints, ear prints, or any digital data recorded, such as the face and voice (Anthony Ho and Li, 2015). Law enforcement could use and analyse biometric traces in different tasks, such as proving the existence of the crime, investigating an offence, and identifying a perpetrator's identity (Arbab-Zavar et al., 2015).

Biometrics now play an important role in the investigation process, especially as CCTV cameras, mobile phones, digital cameras, and many social media applications are now in worldwide use (Perner, 2014). Therefore, biometrics can be easily produced every day and shared through the social network. One of the first conferences in biometrics and forensics was held in 2013 by IEEE/IAPR International Workshop on Biometrics and Forensics (IWBF) and was supported by the EU-sponsored ICT COST Action IC1106 on Integrating Biometrics and Forensics for the Digital Age. This and other workshops confirm the emergence of new topics for research, which is biometric science (i.e., face, voice, and behavioural) within the forensic field (Anthony Ho and Li, 2015).

However, biometric data may suffer from alteration attacks from malicious individuals who try to change or hide the biometric appearance, such as using masks, glasses, hats, makeup for facial recognition, changing the voice for voice verification, and wearing contact lens for iris recognition (Prabhakar et al., 2003). These issues could lead to limiting the use of biometrics in the investigation process. For example, some illegal immigrants get across international borders by faking their faces or fingers (i.e., based on which biometric is required). Therefore, one of the current solutions for this issue is using multimodal biometrics to overcome the limitation of using a single biometric trait (Saini and Kapoor, 2016). For example, several countries, such as Hong Kong, USA, Japan,

and Australia, will use a multimodal biometric border control system that includes airline check in and check out.

2.6 Conclusion

Biometrics offer a huge opportunity for law enforcement across a range of modalities. However, there are several issues that exist to limit their use in practice. Given the huge increase in photography-based content, biometric trait recognition is likely to play an important role in the law enforcement investigation and this importance is increased if the limitations can be overcome.

This chapter provides an overview of critical knowledge about biometric systems. The approach is referred to as measuring the unique physical and behavioural attributes of people. This approach is used either for authentication and to access system security or for identifying individuals or people. For example, fingerprint attributes are considered a unique characteristic for each person. The same theory can be applied to extract other unique features from the face, iris, or ear. In addition, this chapter highlights measurement metrics for calculating the performance of biometric systems and to look for multibiometric techniques and the number of fusion methods to improve the accuracy of the entire system. This previous description led to seeking how to use the biometric system in the field of law. Accordingly, the important biometric objective in digital forensics is the identification of people. One of these biometric components is facial recognition, which helps identify criminals. Therefore, the next chapter will present the definition of digital forensics terms and explain the process and some of the current tools used in the investigative process. Moreover, it will go through identifying facial recognition issues in digital forensics investigations.

3 Digital Forensics and Facial Recognition

3.1 Introduction

This chapter presents a comprehensive analysis of digital forensics and facial recognition. The process and components will be described. This chapter also explains the relationship between a forensic investigator's job and facial recognition. In addition, it focuses on different types of forensic facial recognition challenges. These challenges are considered from several viewpoints, including the system, environment, and user.

3.2 Digital Forensic Science

In the past, forensic science mainly involved the analysis of fingerprints and DNA from the physical crime scene and the questioning of witnesses in order to solve a case. In the last century, owing to the digital information revolution, computer systems have been widely adopted for forensic investigations. A new branch of forensic science, digital forensics arose that specialises in examining digital evidence. Digital forensics dates back to 1984 when the American Federal Bureau of Investigation (FBI) established the Computer Analysis and Response Team (CART) to investigate criminal cases that included digital evidence. In 1990, CART began to co-operate with the Department of Defence Computer Forensics Laboratory (DCFL) for the purpose of training and researching in the digital forensic domain and the DCFL became a base for much of the early curriculum in digital forensics (Nelson et al., 2015).

At the first Digital Forensic Research Workshop (DFRWS), held in Utica, New York, in 2001, digital forensics was defined as "*The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence*

derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorised actions shown to be disruptive to planned operations" (Palmer, 2001). This definition can be used to include all types of digital sources, such as computing systems, CCTV, mobile phones, and any digital technology that may be discovered in the future. Despite the development of technology in the digital investigation field, the DFRWS' definition has been widely accepted since its formation.

Digital evidence requires specialised investigators who know how to deal with digital materials, such as video, text, picture, voice, and email as digital data can easily be modified and/or destroyed. Therefore, the integrity of digital data needs to be fully protected in order to be able to use them as legal evidence in a reliable and trustworthy manner (Yadav, 2011).

3.2.1 The Digital Forensics Lifecycle

Digital forensics includes a multi-stage investigative process starting from the identification of digital media on a crime scene to a stage where digital media is presented as evidence in the court. Since 1984, numerous investigative methodologies have been proposed to illustrate the digital forensics process or lifecycle. Some models are more generic while others are more detailed. Therefore, the investigation's stages may vary in terms of the number of phases according to its author's priorities (Yusoff et al., 2011). For instance, the first model that was proposed in 1984 includes four stages: acquisition, identification, evaluation, and admissibility as evidence (Pollitt, 2007). The DFRWS investigation model consists of seven phases: preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation. Carrier and Spafford (2004) proposed the Integrated Digital Investigation Process (IDIP)

model, which consists of five main phases and each has a number of sub-phases. The five phases are:

- **Readiness**: includes two phases: operation and infrastructural readiness, which prepare the appropriate people and tools for investigative processes.
- **Deployment**: also includes two sub-phases: detection and notification and confirmation and authorisation. This phase gives alerts to the investigator that the incident has been detected and authorises him/her to conduct the analysis.
- **Physical crime scene investigation**: collects physical objects on the crime scene and analyses them. If digital evidence is found, the digital investigation will start.
- **Digital crime scene investigation**: examines the digital data for evidence.
- **Review**: presents the results.

All proposed models aimed to draw an investigatory map to resolve the ambiguities of the crime. This process should be flexible to be applied to any criminal cases. Figure 3-1 shows the main steps of the digital forensic lifecycle or process and its closeness to the DFRWS' model. These six steps are as follows:

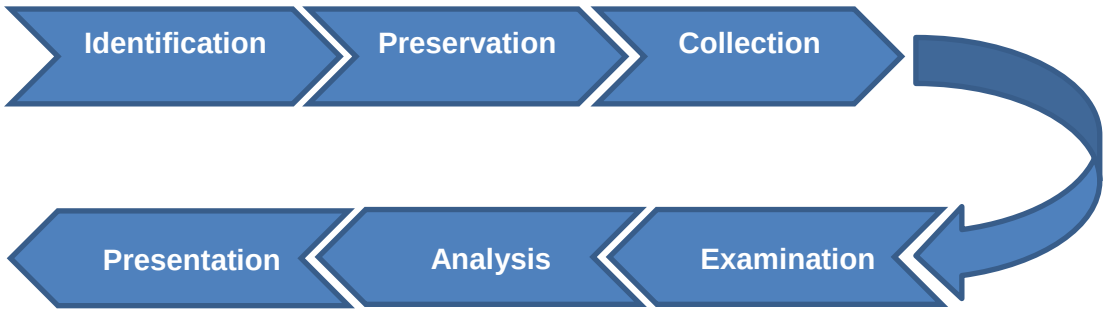


Figure 3-1: The main lifecycle steps of digital forensics

1. **Identification**: this is the first stage and involves the identification of any evidence at the crime's location. In digital investigation, this includes the identification of digital devices and physical objects, such as computer systems, hard disks, USBs, external storage, mobiles, smart watches, sticker notes, and manuals (Raghavan, 2013).
2. **Preservation**: in this phase, all digital and physical evidence is preserved by disconnection from any usage or electrical power. In addition, the digital information inside these sources has to be imaged and isolated (Reith et al., 2002). Moreover, it is essential to protect the evidence's integrity by using a hash function.
3. **Collection**: the content of the digital evidence is acquired from the suspect's device. Also, multiple copies of the original evidence are created to avoid any tampering of the evidence throughout the investigation (Raghavan, 2013). A copy of them is given to the examiner for examination and analysis (Reith et al., 2002).
4. **Examination**: this includes a filtering process for evidence and it is classified into types: 1) the discovery of hidden data and 2) using forensic tools to reduce the amount of obtained data relating to the suspects. Casey (2011) defined this phase as the process that extracts information from digital evidence and prepares it for the analysis phase.
5. **Analysis**: this phase involves rebuilding fragments of data and using statistical and analytical methods to contribute to the result. Relations between the forms of evidence will be sought to build knowledge contributing to the identification of the subject. Moreover, this phase should have had the answer to questions such as who, what, where, when, how, and why. Furthermore, evidence among multiple sources should be correlated (Kent et al., 2006). The analysis can either be dead analysis or

live analysis (Yen et al., 2009). Dead analysis is conducted when a system is in shutdown mode whereas, in the live analysis, data is examined while the system is running.

6. **Presentation**: in this phase, a final report will be created by converting the result into a form that is comprehensible to investigators, who will then make an interpretation of what the evidence shows

3.2.2 Digital Forensics Tools

Digital forensics is an important field to solve crimes that include digital evidence such as computers, mobile, and digital information shared through the internet and social networks, such as images, voice recordings, emails, and documents. Therefore, digital forensics tools have appeared to deal with these types of evidence. The first generation of digital forensic analysis tools was developed in the 1990s and was used in an attempt to acquire and analyse digital evidence (Ayers, 2009). Among a number of the first-generation tools were EnCase (Guidance Software) and Forensic ToolKit (FTK) (AccessData Corp), which have become standard tools for the digital investigation of computers in the acquisition and analysis of evidence. The first-generation tool can be considered as a standard industry tool that enhanced the digital investigation by allowing read-only access to files' content, browse files, keyword searches, and use a range of analysis techniques by using user interfaces based on the Graphical User Interface (GUI) (Ayers, 2009). Although these are abilities of the first-generation tools, some limitations have appeared, including:

- Processing speed: most of the first-generation tools of digital forensics are delayed for many hours or days when analysing the average volume of evidence (Ayers, 2009).

- Evidence-oriented design: the current tools are designed to help investigators find specific pieces of evidence in terabytes of data but it is difficult to reconstruct deep analysis of data that unify time events, correlated between the offender's actions. Instead, examiners perform some of these tasks manually (Garfinkel, 2010).
- Data abstraction: current forensic tools have limited types of data abstraction that are used in a forensic perspective, such as disk image, packet capture files (format to capture network traffic), files (recognise documents and images), file signatures (output of SHA1), and extracted named entities (classifies as ASCII text files or Unicode files) (Garfinkel, 2010).
- Auditability: some first-generation forensic tools have limitations to allow analysts to look inside the tools' working or how evidence has been interpreted and analysed.
- Software error: one concern in forensic tools is errors in the software, such as unexplained crashes. This could lead to disruption or loss of work. Some causes of this software error are errors in the tool, which failed to interpret evidential data or use unsafe programming languages, such as C and C++ where programming errors result in crashes.

In addition, there are two families of forensic tools available: open-source or commercial. While open-source forensic tools are free, their functionalities are somewhat limited; they are also less user-friendly, as many of them are either command line based or Linux based. In comparison, commercial tools are costly: ranging from several hundred to a few thousand pounds. However, they are more thoroughly tested and users get better support. More importantly, these forensic tools must be accepted by courts when presenting evidence and reports.

Therefore, to enhance the first-generation tools, Ayers (2009) tried to specify performance criteria that have to be achieved by second-generation digital forensics tools, such as speed (time for data processing and analysis), parallel processing, accuracy, reliability, audibility, and repeatability. For instance, AccessData has developed FTK to support multiple processing nodes.

Multimedia file (such as image, audio, and video) analysis is important in any criminal case (Perner, 2014). The investigator needs to identify objects, events, and human beings (e.g., face, ear, and fingerprint) from the digital multimedia evidence or analysis of these files. However, there are questions regarding the authenticity of multimedia files (e.g., the history of these files). For example, it has become easy to modify and destroy files, especially by using photo-editing software, such as Photoshop. Therefore, it is necessary firstly to check the authenticity of evidence files. However, most existing forensic tools suffer from including multimedia analysis tasks within their analysis. Therefore, one of the objectives of the FBI's program (FBI, 2015) is to include forensic facial recognition in the next generation of digital forensics tools. Because biometric identification is a reliable approach, it is necessary to develop the available digital forensics tools to determine a suspect's identity from surveillance imagery by including automated biometric recognition technology and attempt to progress it (Jain et al., 2011).

3.3 Forensic Facial Recognition Systems

The history of facial recognition in law enforcement dates back to 1871. At that time, the use of facial identification by means of comparison between two photographs (a daguerreotype and an albumen print) was accepted for the first time by a British court (Porter and Doran, 2000). In 1880, the French criminologist Alphonse Bertillon suggested the classification application and people

photograph search techniques. In 1890, he created a standard form of forensic photography, classifying head descriptions into three categories: nose, forehead, and ear (Arbab-Zavar et al., 2015).

Research then turned towards the use of computers for developing facial recognition techniques. In 1965, Chan and Bledsoe proposed the first semi-automated facial recognition system, depending on humans to extract features from the face (Lohiya and Shah, 2015). In 1973, Takeo Kanade suggested the first fully automated facial recognition system in his PhD thesis (Li and Jain, 2011). In the years that followed, researchers in this field focused on several areas, including the challenges posed by non-cooperative and uncontrolled aspects of facial recognition. This increased interest in the face compared with other human biometric traits for different reasons (Anthony Ho and Li, 2015):

- Biological nature: The face is an easy and convenient biometric feature utilise by humans to identify people. For instance, in access control, it is easy for administrators to track and check an authorised person from any identification card while fingerprints need an expert person with professional skills to check the authorised.
- Non-intrusive: Unlike fingerprints and the iris, facial images can easily be obtained from a distance without physical contact. People feel more comfortable using the face as an introduction to everyday life. The collection of biometric data in the facial recognition system can be done in a user-friendly way.
- Less cooperation: facial recognition needs few requirements of user cooperation. For example, in surveillance applications, a facial recognition system can easily identify a person without the need for the active participation of people.

In addition, forensic facial recognition systems have received a great deal of attention as a result of a number of characteristics (Arbab-Zavar et al., 2015):

- Increasing the database capacity of photos in government agencies and private organisations, such as passport offices, police stations, personal identification documents, bank accounts, and prisoners' records.
- The rapid spread of digital image-capturing devices (e.g., CCTV, smartphones, and digital cameras) allows photos to be easily shared and used between people, especially on social media.

All previous reasons explain why using the face is the first priority in criminal investigations to determine a suspect's identity when most of the evidence is digital.

The technique used by forensic experts was to compare two facial images, one of a suspect, often a "mug shot", and the other from a photographic database. The expert then decided which of the two images was closest in terms of shapes and features. The matching technique in facial recognition was based on four approaches (Ali et al., 2012):

- Holistic comparison: the whole face is considered, using facial comparison methods.
- Morphological analysis: the forensic expert analyses facial features, such as the nose, mouth, eyes, and other "soft" facial features, such as moles and wrinkles. After that, the expert will view the similarities and differences between the two sets of features.
- Photo-anthropometry: the comparison process in this category depends on the quantification measurements of facial landmarks (distance and angles). The condition is that all measurements need to be taken from the

same direction and angle, placing significant limitations on this approach, especially with regard to uncontrolled facial image captures.

- Superimposition: the condition of this approach is that the two images captured have to be from the same angle. This method may, therefore, necessitate facial pre-processing (of pose or orientation) before a comparison is made.

The image fed into the forensic system will determine which approach will be chosen. Some cases require using more than one approach to achieve the recognition aim. Furthermore, the comparison process may vary according to the preferences of the expert involved.

3.4 Facial Recognition Processing

There are four main processes for any digital system that applies facial recognition to identify people, and these processes are shown in Figure 3-2.

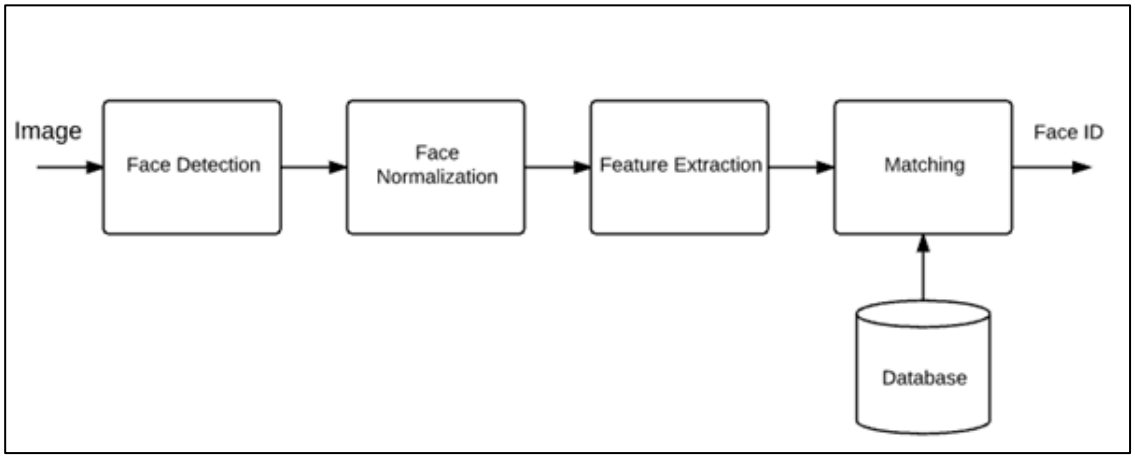


Figure 3-2: The main steps in face recognition process

- **Face detection:** this is the first step in a facial recognition system. It extracts the facial region from the background photo image. In a video file, faces can be tracked and extracted from individual frames. In order to distinguish faces from other objects, such as houses, trees, and cars, a

pattern space that uses different cues (such as facial shape, skin colour, and eigenfaces) to detect the facial area can be applied (Jain and Li, 2005).

- **Face normalisation:** the aim of this step is to locate a facial component or to discern accurately the position of facial landmarks such as mouth, eyes, and nose. The normalisation process is based on geometrical properties (e.g., size and pose) and photometrical properties (e.g., illumination and grey scale) (Li and Jain, 2011). However, variances in pose and illumination in the captured image make the facial recognition process more complex. Figure 3-3 illustrates the facial detection and normalisation techniques.

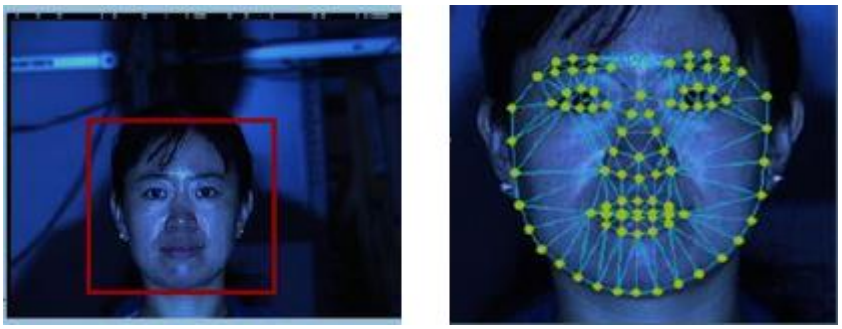


Figure 3-3: The image on the right is facial detection; on the left is facial normalisation (Li and Jain, 2011)

- **Feature extraction**: this step is regarded as the main process in a facial recognition system because significant features will determine the accuracy of the recognition. Therefore, the feature extraction approach is based on face shape, texture, colour, size, or face component details. In addition, there are local facial features or silent areas in the facial skin, such as scars, moles, and freckles, and they have contributed to increased recognition accuracy in recent years (Park and Jain, 2010).
- **Feature matching**: the matching process is based on two schemes of recognition, verification (one to one) and identification (one to many). In the first case, the output will be either yes or no while, in the second case,

the output will give similarity rate for all faces in the database (Li and Jain, 2011).

Generally, these steps will support the two schemes of biometric system verification or identification. After that, the system will evaluate by using some performance metrics that are explained in the previous chapter.

3.5 Forensic Facial Recognition System Barriers

Despite the rapid increase in the use of computers in automated facial recognition by forensic departments, there are still problems to be overcome. Klontz and Jain (2013) studied the Boston Marathon bombings of 2013 and analysed the reasons the automated facial recognition system failed to identify the suspected persons at the time (as mentioned in Chapter One). They concluded that forensic facial recognition is in need of further research, especially when it operates under the unconstrained conditions of people in the presence of digital cameras. The efficiency of facial recognition is affected either by external factors that are unrelated to the user, such as illumination, camera quality or resolution, or more than one person in the same location, which could obscure the subject's face. A second type is internal factors, such as pose variation (uncooperative people), facial expression, and faces occluded by accessories (Jain and Li, 2005). Facial ageing is another issue in the facial recognition system—for example, comparing a recent photo of a subject against a passport photo that is almost 10 years old. This issue is exacerbated with children.

Figure 3-4 illustrates some forensic facial recognition challenges. User cooperation can be divided into two types: (i) cooperative users who present their faces in the right position in front of the camera. For example, passport applications that check and scan faces and authentication systems that require close face access; and. The other type is (ii) the non-cooperative user. In this

scenario the user will be unware of the location of the camera. Also, the distance between the face and the camera in this type can be inconsistent comparing with the first type and therefore the complexity will be increased (Li and Jain, 2011). Some of these issues are highlighted below:

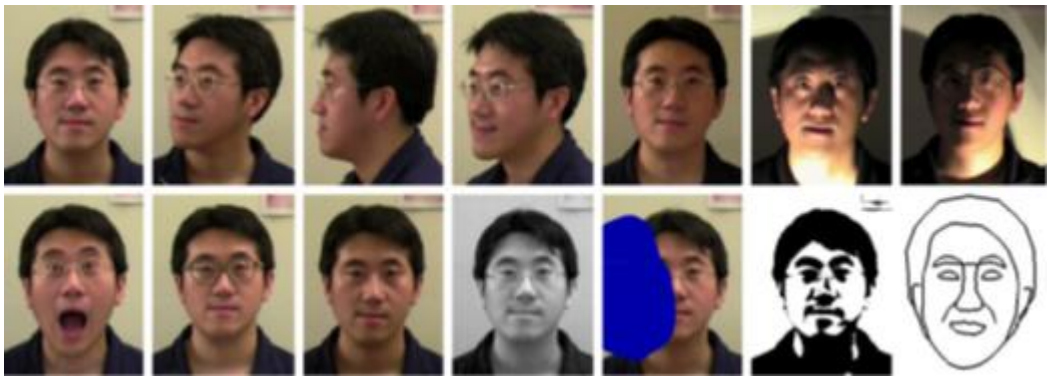


Figure 3-4: Some forensic facial recognition problems (Li and Jain, 2011)

- **Illumination Issues:** changes in light are related to environmental changes and indoor and outdoor environments, and these factors cause a major change in facial appearance (Jain and Li, 2005). The last two images on the right on the top row in Figure 3-4 demonstrate this. In addition, the use of near-infrared (NIR) in some cases has an effect on facial appearance, producing variation in the illumination factor (Jain et al., 2011). However, the direction of the light falling on the face could cause differences in the shading and shadows on the face. These variations in facial appearance could increase the complexity of automatic face recognition, especially if accompanied by another problem, such as facial pose or expression (Zou et al., 2007).

- **Low quality issues:** one of the problems in the field of forensic face detection is the quality of the digital image or video evidence that is acquired. Historically, the surveillance camera video considers low resolution because of the absence of control of distance and environment. In addition, the image captured by these surveillance cameras may suffer from noise, such as poor

lighting and poor camera sensors (Xu et al., 2014). Some researchers have focused on variation in image quality caused by external effects, such as camera sensor quality, illumination conditions, and background noise (Chen and Li, 2011). Another type of low-quality image is a blurred image caused by weather conditions, motion, interruption of image acquisition or image/video zooming (Ghazali et al., 2012).

- **Aging Issues:** recently, researchers have been confronting one major challenge to facial recognition, that of ageing. Geng et al. (2007) identified three characteristics of ageing (1) it is uncontrolled, as no power on earth could halt it; (2) it is personalised: the pattern of ageing is not uniform for all people since it is affected by standards of living, such as diet, lifestyle, and weather; (3) temporal data: the effects of ageing are more apparent in older faces than in younger ones. These characteristics of facial ageing have increased the complexity of the recognition problem. Although the number of studies on solving the ageing issue have increased since 2002, they are still limited in terms of quality and quantity (Juefei-Xu et al., 2011). An important challenge posed by the ageing factor in forensic systems is the difference between the sample in the database sample and recent images of the suspect. Figure 3-5 shows various images of the appearance of the same person over a period of time, ranging from 2 to 40 years. Ageing plays an important role in face recognition cases such as those involving missing children, and in determining the suspect's identity from a database in law

enforcement. Where the face samples have been captured over a potentially long period of time. (Jain et al., 2011).

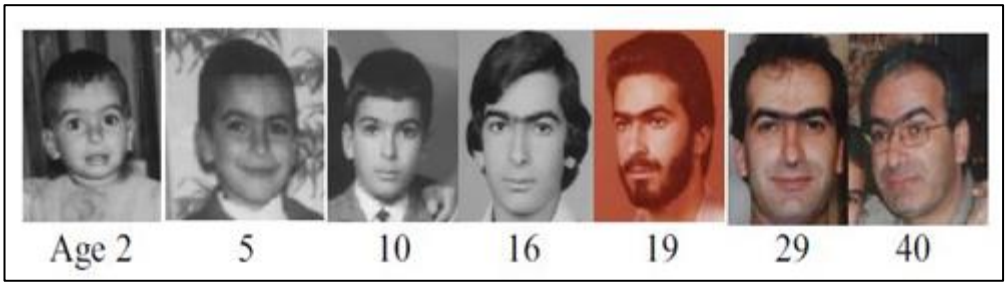


Figure 3-5: Face aging appearance (Jain et al., 2011).

The facial problem can be illustrated in two facial characteristics: shape and texture. Most facial shape changes, i.e., facial growth, happen below the age of 18 years; whereas face texture changes (e.g., wrinkles) are more apparent at ages above 18 years. Therefore, there is a need to distinguish between the two types in order to obtain better results (Li and Jain, 2011). The process of dealing with facial ageing factors in a facial recognition system can be either a generative scheme or a discriminative scheme (Pal and Gautam, 2015). The process in the generative approach is based on transforming the facial shape and texture from that of the current age to that of the targeted age. In the discriminative method, the technique consists of extracting the descriptors and discriminative features from the face to reduce the age gap. Jain et al. (2011) described the two methods (see Figure 3-6).

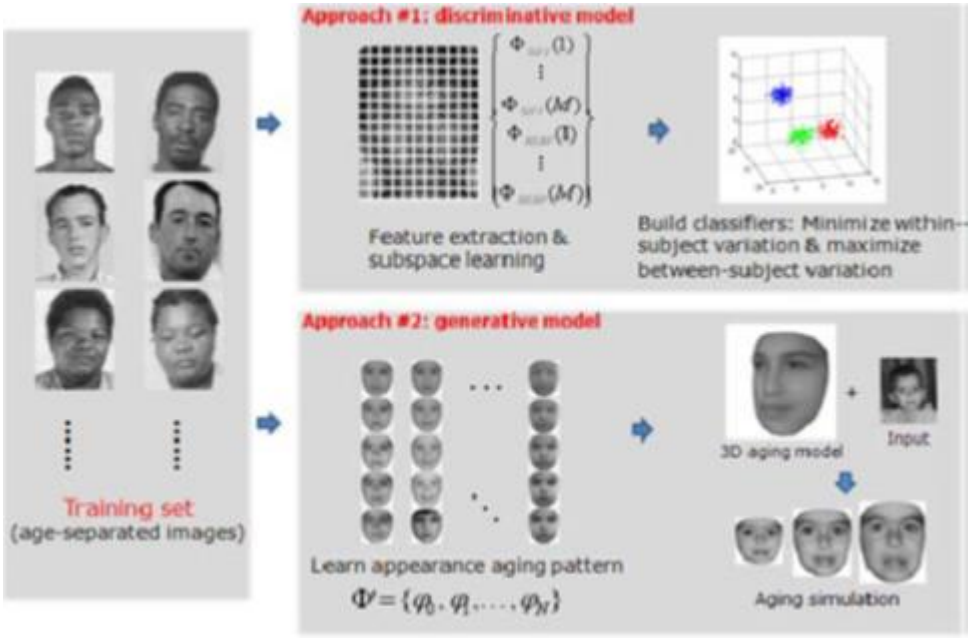


Figure 3-6: The two face-ageing schemes (Jain et al., 2011)

- **Face Pose:** another important challenge in facial recognition systems is how to handle varying face poses in unconstrained facial capture. In most cases, the forensic facial recognition system depends on CCTV cameras or digital cameras to detect a suspect who is considered as non-cooperative in front of the camera. Most facial images stored by police forces or passport offices are frontal poses. There are different degrees of facial poses, such as ±0◦, ±15◦, ±30◦, ±45◦, ±60◦, and ±90◦. The complexity of the facial recognition algorithm raises when the pose degree increases (Zhang and Gao, 2009).
- **Occluded Face:** facial occlusion is a common problem that occurs in the facial recognition system. The face may be obscured by sunglasses, a scarf, or hat and the face may be partially hidden by any object. These are examples of common categories of occlusion in facial recognition systems and are described as dense occlusion (Min and Dugelay, 2012). Min and Dugelay studied another type of occlusion called sparse occlusion, examples of which are facial painting, dirt on the face, and faces behind a fence. Figure 3-7 shows the two types. For the most part, automatic facial recognition fails to identify

suspects' faces in riots, disturbances, and terrorist acts in the street as a result of the increased occlusion of faces in addition to other facial recognition problems (e.g., pose, illumination, and expression) (Klontz and Jain, 2013). The results of law enforcement applications in these cases are negatively affected by poor automatic facial recognition.

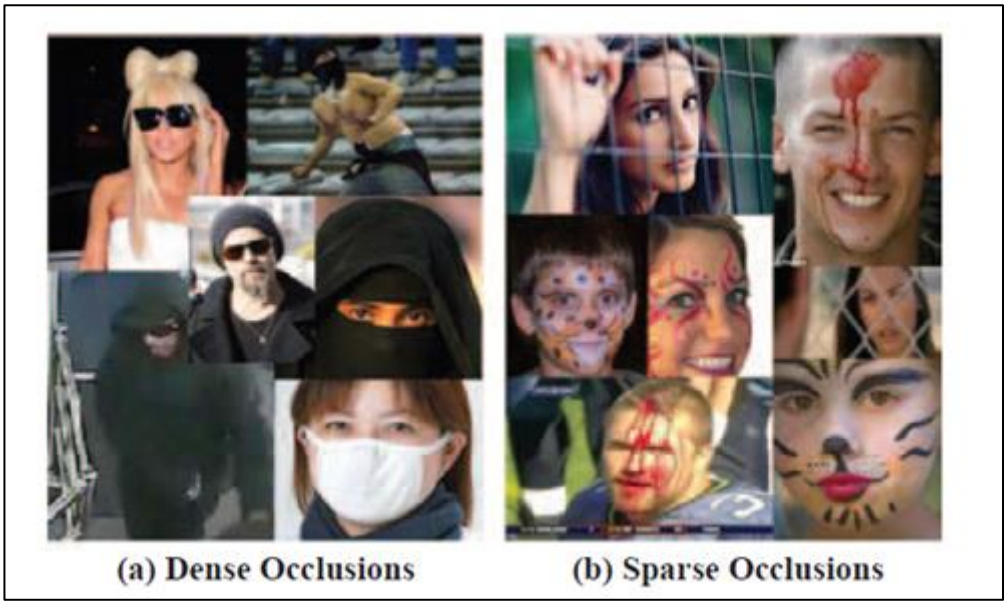


Figure 3-7: Two categories of occlusion: (a) dense occlusion faces, (b) sparse occlusion faces (Min and Dugelay, 2012).

- **Expression Issues:** facial expression consists of outlines of the mouth, eyes, and eyebrows. These contours contribute to categorising facial expressions as happy, sad, disgusted, angry, and surprised (Garg and Choudhary, 2012). Most people can identify the expressions of other people from their faces. However, facial expression variance affects the performance of facial recognition systems when the facial images in the database are without expression.
- **Sketches:** as mentioned before, law enforcement employs biometric technology, such as DNA and fingerprint matches, to determine the identity of criminals. However, there are cases where none of the biometric tools mentioned above can be used in the investigation process but only

eyewitnesses are available. The investigator, therefore, employs a forensic artist to draw a sketch of the suspect's face. Then the forensic sketch will be compared with photos in the database to determine the suspect's identity. There are two types of facial sketches: one drawn from a person's photograph is called a viewed sketch. A second type depends on witnesses who give descriptions of the suspect from which a sketch is drawn. This type is known as a forensic sketch (Klare et al., 2011). There are issues when using forensic sketches in matching. Klare et al. (2011) highlighted these problems as follows: (1) matching across image modalities and (2) performing facial recognition despite possibly inaccurate depictions of the face. Figure 3-8 illustrates the sketch types and matching results.

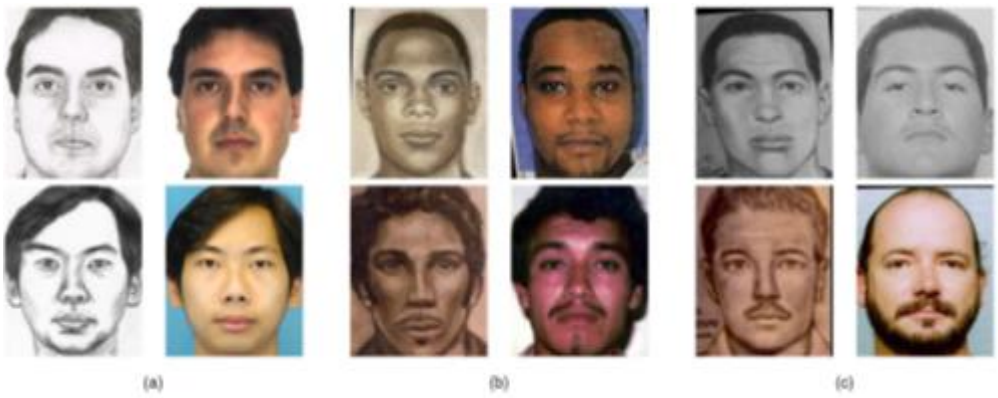


Figure 3-8: The sketch types: (a) viewed sketches and their corresponding photographs, (b) forensic sketches and the corresponding photograph with poor quality, and (c) forensic sketches and the corresponding photograph with poor quality (Klare et al., 2011)

Another challenge is when the system presents a number of candidates for recognition and a final decision has to be made by an expert who decides whether the system has failed or succeeded. This stage requires an expert person that studies all facial images returned by the automated facial recognition system and makes the decision. Finally, the judicial system requires a degree of support for

the decision by an expert person, which could be stated as no support, limited support, or strong support (Ali et al., 2012).

3.6 Conclusion

Facial recognition techniques have been successfully used in identifying suspects within the forensic domain for many years. With the development of digital forensics tools and the availability of mass images (such as online sharing, CCTV, and personal capturing), it is envisaged that the facial recognition technique can be used to assist investigators even better. However, the limited capabilities of existing digital forensic tools in processing multimedia files and the poor quality of some of the images (e.g., the environment night or day, the sensor quality, and the distance of the subject from the camera) present a significant obstacle that prevents facial recognition systems from obtaining a reliable outcome. To improve the accuracy of forensic facial recognition, these barriers should be thoroughly investigated.

4 The Current State of Art

4.1 Introduction

This chapter presents the existing studies on forensic facial recognition. Firstly, the research methodology used to select studies will be explained. Then, a thorough review of the current studies in facial recognition, which start with facial ageing, external issues (lighting and quality), and internal issues (face pose, expression, and occluded) will be presented. It will also describe soft biometrics and how some studies have enhanced the recognition accuracy then will present some commercial systems that give good accuracy in matching. As a result, this chapter focuses on the analysis of each significant issue of facial recognition. Finally, the discussion section will give an analysis of all studies and identify the gaps in forensic facial recognition.

4.2 Research Methodology

To review the current state of the art related to facial recognition challenges in digital investigations, a number of research methodologies have contributed to the formation of the literature review relating to forensic facial recognition barriers. Research has been on many digital libraries and databases. The initial search was started in October 2014. The focus was solely on facial recognition issues in digital forensics. The research methodology used a range of keywords relating to facial recognition challenges and its effect on digital forensics investigations. The literature review methodology was searched for in related publications from various academic databases, such as IEEE, Google Scholar, Science Direct, Springer link, and ACM. Because of huge numbers of the search publication results, the keyword “forensic” was used to filter these results for a more specific set of outcomes. In addition, another strategy was used by following the citations of current significant studies to see where they have been used or which new

ones have been developed. However, the first relevant studies were identified using the search thread and the examination was conducted on the title and abstract of papers.

4.3 Facial Ageing

One of the facial recognition issues is facial ageing and it has added more complexity to the facial recognition process. Pal and Gautam (2015) divided facial recognition into two categories based on age features: generative and discriminative. Generative facial recognition is the prediction of one's facial image according to the age variant while the discriminative category uses various facial feature descriptors to narrow the search environment in the database independent of one's age.

Depending on one's background (e.g., gender, ethnicity, and age), the shape and texture of the face can be affected by the facial ageing process in different variations. For example, young people's facial changes will be affected more by shape (e.g., skull and head size) than texture while, for mature people, the texture properties (e.g., wrinkles and skin texture) have more influence on their facial changes than the shape does (Jain and Li, 2011).

Park et al. (2010) proposed an ageing simulation model for age-invariant facial recognition. Their experiment employed several databases, such as FG-NET, MORPH, and BROWNS. A 3D ageing model was used to correct face orientation and ageing in a model by selecting images from the aforementioned 2D face-ageing databases to build a 3D domain. The team tested and compared three variant models: shape only, separate shape and texture, and combined shape together with texture schemes (e.g., utilising the second level of Principle Component Analysis (PCA) to delete any correlation between shape and texture after the combined process). In addition, they used a commercial face matcher

(i.e., FaceVACS) to evaluate their ageing model. Moreover, the FG-NET database was used as the ageing model and three databases, FG-NET, MORPH-Album1, and BROWNS, were used for the evaluation. In total, 81 facial features were used in the 3D morphable model, where 68 features represented salient features and the other 13 were forehead area lines. Their experimental results are presented in Table 4-1. Although matching results were successful in matching most images and it was a good attempt to build a 3D ageing model from the 2D database, there were some matching attempts that failed because of the large effect of the capture condition of images, such as pose variation in wide angles, or the effect of poor illumination. Further, their work was evaluated by using a single matcher method and the evaluation could be more reliable if other matcher techniques were applied.

Database	Rank-1 Identification Accuracy	
	Original image (%)	After aging (%)
FG-NET (82 probe, 82 gallery)	26.4	37.4
MORPH-Album1 (612 probe, 612 gallery)	57.8	66.4
BROWNS (4 probe, 100 gallery)	15.6	28.1

Table 4-1: Identification accuracy (Park et al., 2010).

Mahalingam and Kambhamettu (2010) proposed a method of age-invariant facial recognition using graph features of the face in the matching. The graph included the appearance and geometry of facial features. The experiment utilised the FG-NET database (containing 1,002 images from 82 subjects with age less than one year to 69 years) and divided it into two equal sets for both training and testing. First, they extracted 68 feature points from images in the FG-NET database and used a generic model to calculate the facial pose and apply the Active Appearance Model (AAM) technique to correct the non-frontal face images. Then, the Local Feature Analysis (LFA) method was performed to extract 150 feature

points for each image. After the identification of the vertex of those feature points, important facial feature descriptions, such as nose, mouth, eye, and face curve, were extracted by using the Local Binary Pattern (LBP) method. Then these face vertices were converted to graph form based on the length of the edges between the vertices and the surrounding points. In addition, ageing models were designed using a graph model of training images and applying the Gaussian Mixture Model (GMM), which involved both changes of shape and texture (Mahalingam and Kambhamettu, 2010). The facial identification was performed in two stages. First, the age model was used to compute posterior solutions to effectively narrow down the environment search and the potential individual was determined for the second stage. In the second stage, the matching was performed using the spatial similarity between the graphs. Two experiments were conducted: the first focused on ages from 18 to 69 years and built the training set by including all images with younger faces while the testing images set contained the older faces. The experiment's performance was evaluated using the CMC curve and the accuracy recognition approximately ranged between 70% and 80% from rank 10 to 20. In the second experiment, the training set included ages from 0 (less than one year) to 30 years and the remainder was grouped in the testing set. The maximum cumulative accuracy achieved approximately 69% in rank 10. The results indicated the large shape variation affects the performance of the second experiments because it considered images of young faces. This is considered an issue in any facial recognition system over age. In addition, this study only employed one dataset for the evaluation while a more robust set of results can be obtained if more datasets are used.

There have been several studies involving face regions that can be exploited to improve facial recognition performance across ageing. One of these attempts

was conducted by (Juefei-Xu et al., 2011), who studied the periocular area of the face, such as eyebrows, eyelids, and eyeballs as they found that eyes are less affected by ageing changes compared to other areas of the face. The FG-NET database was used in their work. Initially, the AAM method and the parallelised anisotropic diffusion model (by running on GPUs programmed with nVidia's CUDA framework) were utilised to correct and normalise images with pose variations and illumination effects, respectively. Then, Walsh-Hadamard transformed encoded Local Binary Patterns (WLBP) were used on the periocular region to obtain different local features. Finally, these features were used to build subspaces by using unsupervised discriminative projection for matching techniques that considered both local and non-local information. Figure 4-1 shows all steps in this study. Their results have demonstrated an identification rate of close to 100% at rank-1 and a 98% verification rate at 0.1% False Accept Rate (FAR). They compared their results with Li et al. (2011) and there was an improvement of about 52.2% on the same database (FG-NET). These regional features are mostly static across ages for the same people. The identification rate of Juefei-Xu et al. (2011) was better than other results of full facial recognition. There are, however, some limitations, as the periocular region is more affected by facial expression and disappears when the subject is wearing glasses or makeup.

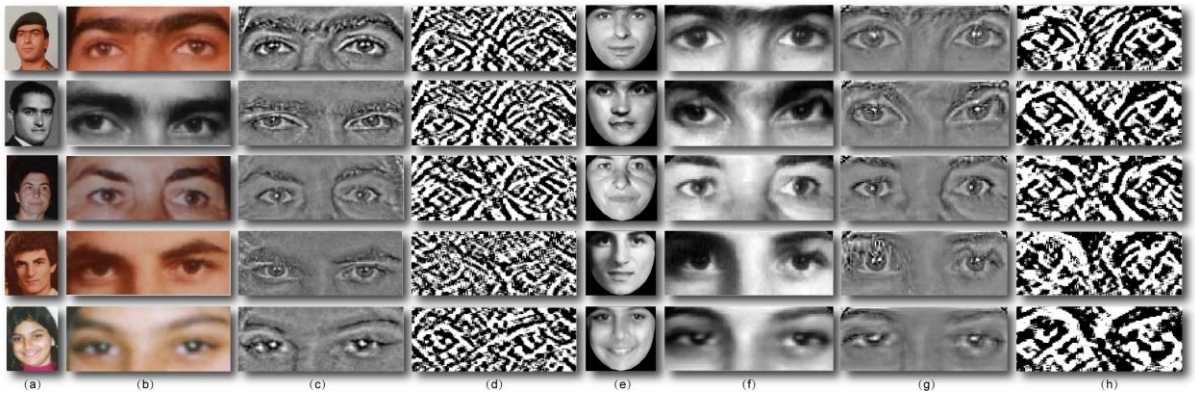


Figure 4-1: (a) full-face image, (b) periocular region normalisation, (c) illumination correction, (d) feature extraction, (e) pose correction for full face, (f) periocular region normalisation, (g) illumination correction, (h) feature extraction

A major issue in criminal cases is time lapses. In most criminal cases that involve children, this is more challenging because the face shape in childhood changes more quickly than in older ages and this makes the search more difficult.

Kemelmacher-Shlizerman et al. (2014) suggested an automatic age-progression application. Their contribution achieved fully automated age progression (i.e., images from the wild) and novel illumination-aware age progression by correcting surface shading without reconstructing the 3D model. The age progression model was based on the original work of (Burt and Perrett, 1995), which dealt with face shape and texture changes in the ageing process. The team collected 40,000 photos from Google search images (i.e., without strong constraints on illumination, pose, and expression changes) to evaluate the system that produces ages between 1 and 80 years. The face age progress process converts the input child face to an old face by applying texture difference computation (shown in Figure 4-2). This study depended on humans' decision to evaluate the age-progression results. The results indicate that people have the ability to recognise adults across age progress while poor at recognising age progress for children.

This technique for adult face ageing could be enhanced by adding other textural changes, such as wrinkles or different hair colours.

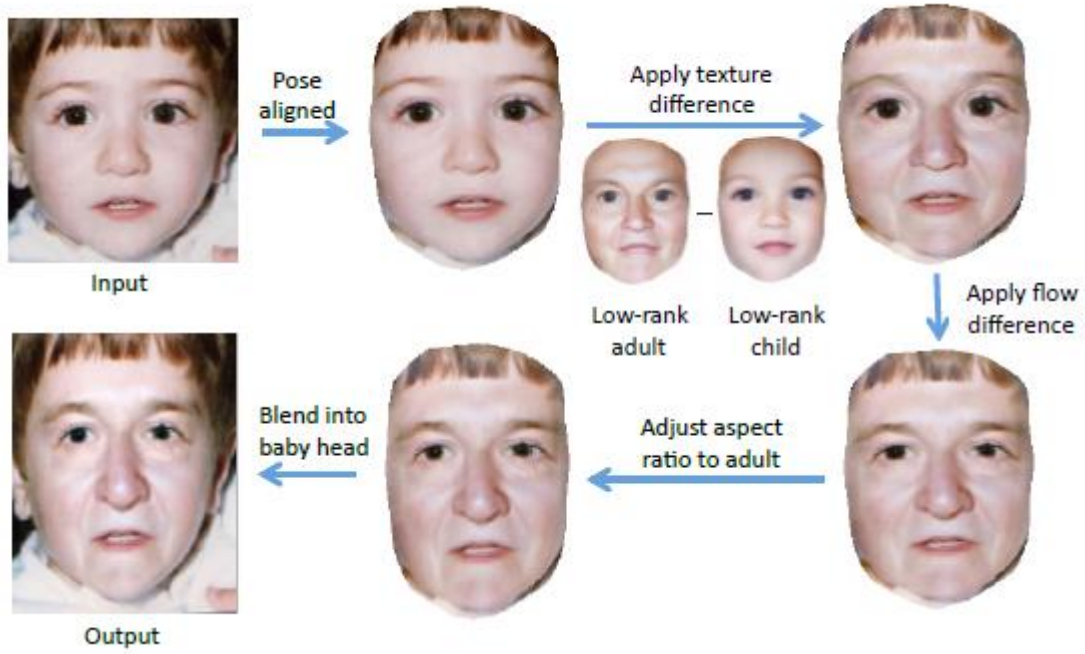


Figure 4-2: Process of illumination-aware age progression (Kemelmacher-Shlizerman et al., 2014)

Ling et al. (2010) proposed a study that adopted a discriminative approach to solving the ageing issue. It aimed to recognise people in terms of estimated age and gender classifications and discriminated between faces by using landmarks or soft biometric traits at a different age. The discriminative approach reduces or filters the search space too. Ling et al. attempted to study face verification for two images at different ages by using discriminative approaches in man ageing this issue without generating a new image to close the gap in time. Firstly, a Gradient Orientation Pyramid (GOP) method was applied for feature extraction because they believed it responded robustly to illumination. A Support Vector Machine (SVM) method was used to classify the framework. Three datasets were used: the FG-NET database (e.g., 82 subjects, and 1,002 images), which is widely used in face ageing analysis. Furthermore, there were two scanned passport

databases (e.g., totally more than 1,800 subjects), which are generally frontal poses with small pose variation. Passport I included 452 intrapersonal images pairs while Passport II contained 1,824 intrapersonal images pairs. The experiments on real passport datasets showed that the EER of the SVM+GO method for Passport I and Passport II was 8.9% and 11.2%, respectively. The EER on the FGNET database was 24.1%. The age average considered in the two experiments was >= 18 years. Ling and his team analyzed their experiments and found that the high error rate was because of the poor quality of years-old images or scanned images compared with those of high quality. There was an issue with images that included spectacles and facial hair, which included a moustache and a beard. In addition, their study found that the error rate increased if the age gap was more than four years. A second experiment was on facial verification across ageing in children, which evaluated the FGNET dataset for age (8-18) called FGnet-18 and (0-8) called FGnet-8. The EERs were 30.5% and 38.6% for FGnet-18 and FGnet-8, respectively. This experiment explained how the verification system for children faces is more complex than for adult faces. Further, the error rate increased for ages from 0 to 8. These results appeared because of the larger face shape variation before age 18.

Attempts have been made to narrow the difference between database facial images and probe facial images in facial recognition applications over time, particularly in the time between childhood and adulthood. Li et al. (2011) depended on the discrimination of faces by local features at different ages rather than generating methods. They designed a new feature descriptor method called Multi-Feature Discriminant Analysis (MFDA), which combines two local features descriptors, Scale Invariant Feature Transform (SIFT) and Multi-Scale Local Binary Pattern (MLBP). Li et al. (2011) believed the discriminative model was

more characterised than the generative model because the eye coordinate area was less affected by ageing. Li et al. observed that the results demonstrated improvements in matching accuracy for generative models (Park et al., 2010) from 79% to 83.9% at rank-1 on the MORPH album 2 database while there was poor accuracy on FG-NET at 47.50%. Similarly, Sungatullina et al. (2013) exploited local facial feature information to discriminate these faces at different years using three descriptor methods for local features: SIF T, Local Binary Pattern (LBP), and GOP. The local features could be extracted by dividing facial images into several patches and applying three descriptive methods on each patch to obtain three vectors of feature information. They then used PCA to convert these vectors from high-dimensional to low-dimensional vectors. Sungatullina et al. suggested the development of a discriminative algorithm called Multi-view Discriminative Learning (MDL). The objective of MDL is to make a concatenation between different feature vectors to produce a new vector that has the feature correlations for the same face then perform a discriminative learning algorithm. Figure 4-3 shows the process aspect of the MDL method with other approaches. The top row illustrates the face samples in the SIFT, LBP, and GOP feature space while the bottom row illustrates the latent space or correlation features by the CCA, MDL, and the ideal approach. FG-NET and MORPH (e.g., 78,000 facial images; 13000 subjects) databases were used to test this study. This study selected images with few changes in expression, pose, and illumination. The experimental results demonstrated that MDL achieved accuracy in recognition at rank-1 in the first dataset of 91.8% while, in the second dataset, the accuracy was 65.2%. The significant point in this study is how to use the local features that considered robust to age variation. In addition, the correlation technique, which adds more discriminative information, could be used to improve the performance of facial recognition systems.

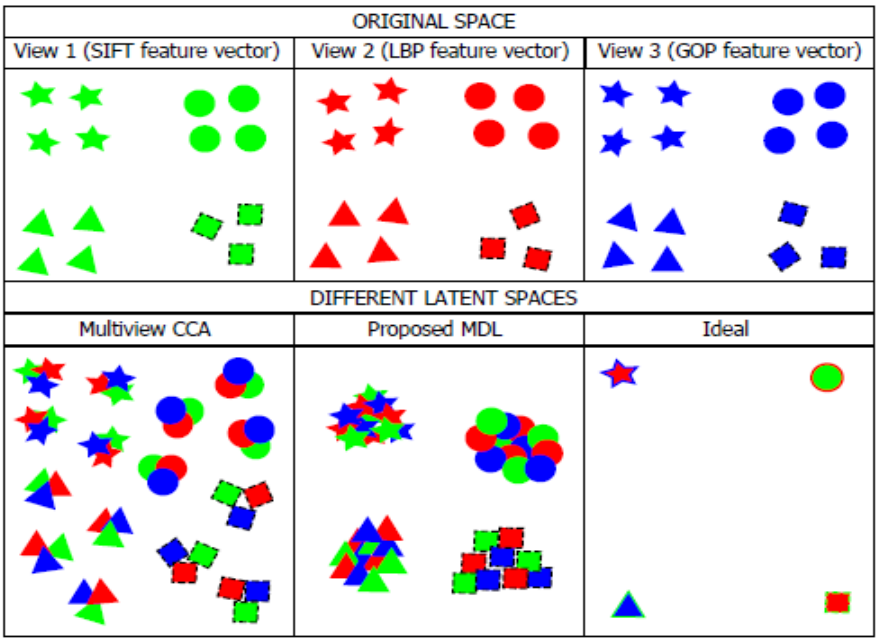


Figure 4-3: The idea of MDL (Sungatullina et al., 2013)

Pal and Gautam (2015) developed the previous work of Li et al. (2011) by correcting facial pose changes, using the Active Appearance Model (AAM), extracting local facial features by applying SIFT and LBP, then verifying the test image utilising a multiclass Support Vector Machine (SVM). This technique depends on the discriminative approach to obtain age-invariant facial recognition. Pal and Gautam used an FG-NET database and achieved 76.6% facial recognition accuracy at rank 1. This approach is limited to subjects aged under 18 because it depends on texture rather than shape change. Further pose correction results were poor when compared to other approaches.

Han et al. (2013) designed hierarchical approaches for automatic age estimation. Their proposed approach consists of four steps: i) image pre-processing (converting colour images to grayscale), ii) facial localisation (localise individual facial components such as forehead, eyebrows, eyes, nose, and mouth), iii) feature extraction, and iv) hierarchical age estimation by classifying each facial component in one of four disjointed age groups using a binary decision tree based on SVM (SVM-BDT). Further, they compared the performance of this approach

with human perception to estimate the age by using the results of crowd-sourced (the Amazon Mechanical Turk service (AMT)) on the FG-NET database and a part of the PCSO database. The training of automatic age estimation was applied on FG-NET, MORPH Album2, and PCSO databases and the results showed that Mean Absolute Error (MAE) was 4.6% for FG-NET, 4.2% for MORPH2, and 5.1% for PCSO. Han and his team observed in their experiments that eyes and noses are more useful in age estimation than other facial parts. The limitation of this study is that it depends on the front of the face and does not pay attention to illumination changes. In 2015, (Han et al.) tried to develop their previous work (Han et al., 2013):

- Adding gender and race estimations.
- Ignoring low-quality face images in the demographic estimation system.
- Studying human estimation for gender and race (black or white).
- Adding two additional databases, FERET and LFW.
- Designing demographic features.

Their experiments achieved fewer errors in age estimation than the previous study (Han et al., 2013), such as that the MAE was 3.8% (FG-NET), 3.6% (MORPH II), 4.1% (PCSO) and 7.8% (LFW). The accuracy of gender classification was 97.6% for MORPH II, 97.1% for PSCO, 96.8% for FERET, and 94% for LFW. The race classification accuracy was 99.1% for MORPH II, 98.7% for PSCO, and 90% for LFW. The FG-NET database was not used in gender and race classification experiments (Han et al., 2015). These studies on estimating age help law enforcement by filtering the gallery database depending on the criminal's age.

Recently, there have been several studies focusing on including commercial algorithms, such as COTS systems, in their research. Best-Rowden and Jain (2018) studied the ability of state-of-the-art commercial facial recognition systems to recognise query face images that are different in ageing with enrolled face images. They used the longitudinal analysis of two of the largest longitudinal databases of repeat criminal offenders. These databases are LEO_LS, which contains 31,852 images of 5,636 subjects, and PCSO_LS, which contains 147,784 images of 18,007 subjects, where the average time span between subjects' multiple image acquisitions is 6.1 and 8.5 years, respectively. Also, in this study, the authors evaluated the performance of a number of COTS system face matchers, which were considered among the top-ranked performers in the FRVT 2013 facial recognition evaluation. The methodology scenario in this study was the verification scenario (one to one) and the results showed that despite decreasing genuine scores, 99% of subjects could still be recognised at 0.01% FAR up to about 6 years and 5.5 years elapsed time for the LEO_LS and PCSO_LS databases, respectively.

Best-Rowden et al. (2016) evaluated the performance of a COTS algorithm against the newborns, infants, and toddlers (NITL) dataset and studied the effects of age variation over one year on the performance of the algorithm. The NITL dataset was collected by the authors over one year. It consists of facial images of 314 children between the ages of 0 and 4 years. The images were captured with variations in lighting, pose, and expression. The experiment results are shown despite that the facial recognition performance for the same session had high accuracy (TAR > 93% at 0.1% FAR) and the cross-session performance degraded significantly to 47.93% TAR at 0.1% FAR at six months age for the children. Furthermore, the age at enrollment (children aged less than 1 year vs.

older than 1 year old) had more effect on the performance of facial recognition than time lapses of 6 or 12 months. The study indicates the COTS algorithm's performance decreases when the age time lapse between the gallery and the probe image increases.

Judging from the aforementioned studies, facial ageing is an unavoidable natural process. Further, it cannot be controlled during facial image acquisition like other sources of face appearance variation, such as lighting, pose, and expression. Therefore, the age issue in facial recognition still has shortcomings and needs to be overcome in law and security systems affected not only by internal factors but by external factors as well.

4.4 External Factors

Low resolution and illumination variation are two external factors obstructing the facial recognition process. Numerous researchers have attempted to minimise their effects on images to increase the accuracy of facial identification, including enhancing the quality of images captured from CCTV cameras or other digital resources. Details of such research will be discussed fully as follows.

Buciu (2010) explored the problematic effects of the illumination factor on facial recognition and designed two correction methods to be applied either to a specific image or applied to any image. Five techniques were analysed for the second type by dealing with light direction and intensity to normalise facial images. These five techniques were the Self Quotient Image (SQI), the Morphological Quotient Image (MQI), the Morphological and Dynamic Quotient Image (DMQI), the Sub-Image Homomorphic Filtering (SHF), and Recombined Multi-Scale Retinex (RMSR). Frontal pose images of the Extended Yale B Database (38 subjects, 2432 samples images) were utilised for the evaluation. After correcting the illumination condition, PCA was applied to reduce data dimension. The various

facial recognition rates used up to five illumination correction methods, as shown in Figure 4-4. The recognition rates increased when the principal components' (PCs) numbers increased. Moreover, the results show the recognition rate only improved slightly after 40 PCs. However, the best performance of two methods, SQI and DMQI, in PCs 90 was 91.67% and 88.82%, respectively. The total processing time was less than one second for all methods and this is critical for real facial recognition systems.

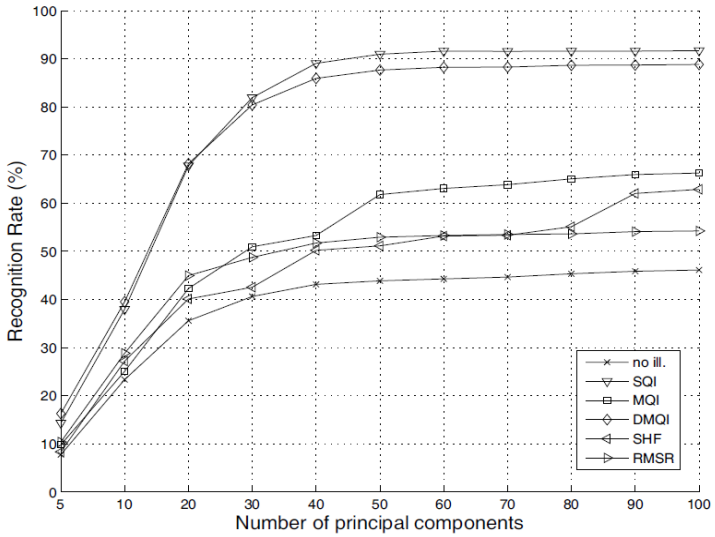


Figure 4-4: Facial recognition performance versus the number of PCs (Buciu, 2010)

Nabatchian et al. (2010) proposed a method of filtering images with illumination variation to obtain smooth images for facial recognition. They designed a technique that could be applied to any single image without the requirement for information about face models or illumination by using a maximum filter. This filter assumed the low-frequency part of the image contained illumination while the high-frequency part contained the illumination reflectance. By applying a logarithm to each image, pixels in the dark domain were expanded and pixels in the bright domain were compressed. Two databases were used for their experiment–Yale B (10 subjects and 5760 images) and extended Yale B (38

subjects)—and images in these databases were divided into five groups according to the light angle of source direction:

- Subset 1 (Θ up to 12°).
- Subset 2 (Θ up to 25°).
- Subset 3 (Θ up to 50°).
- Subset 4 (Θ up to 77°).
- Subset 5 (Θ up to 78°).

The average recognition rates are presented in Table 4-2 with only frontal facial poses. Subset 1 is not presented in the table because it was taken under small illumination variation. This technique is simple and achieved good results in the recognition system. Moreover, it is considered fast as a result of the simplicity of its computation.

Database	Subset 2 (%)	Subset 3 (%)	Subset 4 (%)	Subset 5 (%)
Yale B (10 subjects)	100	100	98.6	98.9
Extended Yale B (38 subjects)	100	99.78	95.44	94.68

Table 4-2: Recognition rates for two database.

Different techniques were used to enhance the facial recognition performance in terms of illumination and pose variation. Choi et al. (2011) proposed a novel technique by adding a weighted average intensity to light angles instead of shadow variations on the facial image. This was called the shadow compensated technique. Figure 4-5 shows an example of the shadow compensated results. In addition, they estimated facial poses to classify images into pose classes in accordance with the edge of the facial orientation. The recognition rate achieved about 99% on the CMU-PIE database and about 92.3% on the Yale B database. Their results were good but the compensated image still had noise pixels that had an effect in extracting discriminative features and if this noise was removed, the performance of the facial recognition system could be improved (Choi, 2012).

(a) Raw images

(b) Shadow compensated images

Figure 4-5: Example of the shadow compensated images in each pose (Choi et al., 2011)

Luan et al. (2014) proposed a method that recognised the frontal pose face of humans under varying levels of illumination and occlusion. This technique did not require any previous knowledge of the illumination or face occlusion. They adopted Robust Principal Component Analysis (RPCA) and represented characteristics of the spare error component by performing sparsity and smoothness descriptors and applying them to facial recognition. In terms of classification, two methods were used: the weighted-based method and the ratio-based method. The approach was evaluated using the Extended Yale B database. The database was divided into five subsets. As a result of the large light variation in them, subset 4 (14 images per subjects) and subset 5 (19 images per subjects) were considered as more important than subsets 2 and 3 that each including 12 images per subject. Subset 1 (seven images per subject) had normal illumination so it was chosen as the training set. The experiments with the

weighted method achieved a 95.06% recognition rate on subset 4 while with the ratio method achieved 54.18%. Subset 5 achieved 49.38% and 38.12%, respectively, for the two methods (Luan et al., 2014). The low performance in subset 5 was as a result of poor light around the face area which cause an extracting error in discriminative. Therefore, the recognition rate could be improved by enhancing the illumination of dark areas.

Super Resolution (SR) is a technique or process that creates high-resolution images of low-resolution images (Baker and Kanade, 2002). Fookes et al. (2012) studied the effect of image resolution on facial recognition performance by applying three types of (SR) methods (i.e., Lin et al. (2005), Schultz and Stevenson (1996), and Baker and Kanade (2002)). PCA or Eigenface and Elastic Bunch Graph Matching (EBGM) were used as the matching facial recognition techniques. The experiments were implemented on the XM2VTS database of 295 subjects with a head rotation shot. They observed an improvement in recognition accuracy variation in accordance with three super-resolution studies as shown:

1 The improvement achieved was about 19% when they used the Lin et al. (2005) method that reconstructed images from low resolution by using the optical flow.
2 The improvement achieved was about 30% when they applied the Schultz and Stevenson (1996) method. They adopted the Bayesian maximum in the super-resolution technique.
3 The least improvement was obtained when they used (Baker and Kanade, 2002) method, referred to as the “hallucination method”. The performance was lower than the interpolation level.

Zeng and Huang (2012) used radial base function (RBF) to design nonlinear mapping from non-frontal low-resolution (NFL) image features to front high-

resolution (FH) image features. This technique attempts to solve the issue of facial recognition in video surveillance, in which most recognition has only one frontal high-resolution face in the gallery for testing. The FERET database was used in the experiments and various SR methods' performance were compared with their own. These methods were GLR (Chai et al., 2007), HGLR (Sharma et al., 2010), Jia's (Jia and Gong, 2005), CRBF (Huang and He, 2011), and Li's (Li et al., 2009). The results were divided into two parts based on whether the input face pose was known or not. Firstly, when the input image pose was known, Zeng and Huang's method had a better facial recognition rate than the others did and achieved 80% at rank 1. The rate decreased to 49% when the pose angle was large (one side of the face). Secondly, in the result when the input face pose was unknown, the results here were similar to the known pose in the first experiment. The main weakness of this study was caused by the error in facial pose estimation.

Despite two previous studies, Ren et al. (2012) criticised the SR method, finding that SR is not always in agreement for improving recognition accuracy and that it takes time, which is not appropriate for real-time techniques. In addition, there was a conclusion made by Xu et al. (2014), who analysed three factors in surveillance systems that affect recognition accuracy. These are 1) camera types, 2) the distance between the human face and the camera, and 3) the facial image resolution. This study depends on indoor surveillance cameras where the motion is mostly slow and pose changes are few. They focused on low-quality images with normal variance lighting conditions. Xu et al. appraised their study by using four datasets, FRGC, AR, ScFace, and Curtin Faces. They ascertained three important points. First of all, when using low-resolution images (when distance to the camera is increased), the recognition rate drops sharply. Secondly, where

distance was fixed for all cameras, they observed a significant change in the recognition rate because of the camera's resolution variations. Finally, facial recognition was largely improved when pre-processing was applied to enhance resolution (Xu et al., 2014).

On the other hand, other external factors could affect the image quality, such as capturing the image near infrared light. This issue was investigated by Guo et al. (2017), who proposed a deep network model that studies the effect of both visible light images and near-infrared images on the performance of facial recognition. They used two datasets in the evaluation: the LFW database (more than13.000 facial images collected from the internet) and YouTube Face Database (YTF) (3425 videos of 1595 different people, all videos are downloaded from YouTube's website). The experimental results demonstrate the performance of the facial recognition algorithm based on the deep network achieved 98.95% and 97.3% for the LFW and YTF datasets, respectively. The deep network model is robust to illumination variation and could be effective in real-world scenarios as it was tested on two wild facial datasets.

The problems of illumination in previous facial recognition studies were attributable to limitations such as facial pose, light angle, and the capture environment (indoors, outdoors, night, etc.). Sometimes, the solution was found without considering the facial image noise and its effect on recognition accuracy. Similarly, the low quality of images is determined by camera resolution, capture environment, and the distance of the object's face from the camera. Despite the fact that all the best studies have tried to enhance the image quality to improve facial recognition accuracy, not all methods were successful and some results had no effect on recognition rate because of the source images being of bad

quality. In addition, in most studies, the probe image had low resolution while the database was of high resolution and this affected the recognition accuracy.

4.5 Internal Issues

Human interaction is unconstrained in front of surveillance cameras; as a result, other facial recognition problems can be related to human interaction, such as facial expressions (e.g., happy, sad, and angry), head motions (e.g., frontal face or not), and partial or occluded face (e.g., face hidden by glasses, hat, or scarf). Therefore, researchers have sought to overcome these challenges and have tried to process images before using them in the recognition system. One of the suggestions for processing the pose of face images was using a 3D technique. Ishimoto and Chen (2009) used 2D facial images to build 3D shape models by using a factorisation method (Tomasi and Kanade, 1992). They extracted 90 features, mapping them to 2D feature points. The maximum facial poses angle of 30 degrees (either right or left) was considered and the experiment was performed on the images of 20 persons. These conditions imposed limitations on the project. Ishimoto and Chen used the 3D shape reconstruction method for new pose angles 45, -15, 15, and 45 in the recognition system and improved system accuracy, as shown in Figure 4-6.

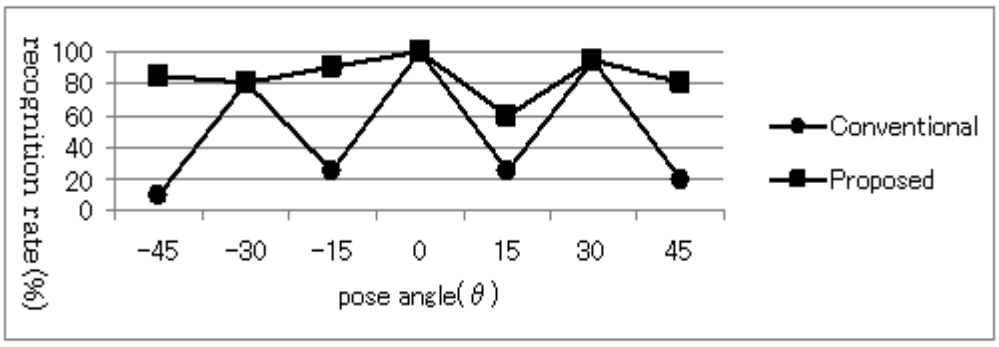


Figure 4-6: The recognition rates results (Ishimoto and Chen, 2009)

Asthana et al. (2011) designed a fully automated facial recognition system by normalising 3D pose variation up to ±45°. First, an Active Appearance Model

(AAM) was used to determine 2D landmark points. View-Based AAM (VAAM) approaches were found to be more robust for pose changes than correction poses using a 3D transformation model. Asthana et al. used the Local Gabor Binary Patterns (LGBP) method for face matching between two faces. Moreover, they used many databases for their experiments. Firstly, two databases, CMU-PIE (68 subjects, 86 images in the gallery, and 408 images as probes) and FERET (200 subjects, 200 images in the gallery, and 1200 images as probes), were used. These were found to be more useful in recognition systems with pose variation and the system achieved an overall recognition rate at rank-1 of 99% and 95.6%, respectively. In addition, they tested the system on three other datasets: USF Human ID 3D (94 subjects, 94 images as gallery, and 18612 images as probes), Multi-PIE (137 subjects, 137 images as gallery, and 1963 images as probes), and FacePix (30 subjects, 30 images as gallery, and 2700 images as probes). The results demonstrated the overall rank-1 recognition rates were 98.8%, 87.7%, and 87%, respectively. These outstanding results may have been achieved because of the small face rotation angle, which was 45 degrees maximum for all databases.

Because the 3D model may need more time to convert 2D faces into 3D models, Yi et al. (2013) sought to use a filter transformation method as a 3D model. They built a 3D features model and fitted it to the 2D image to get pose and shape. Finally, they produced a 3D feature points plane of the 2D image by applying the Gabor filter to extract robust pose features. Yi et al. evaluated their system on the FERET database, limited to a pose angle <= 45 degrees. The mean recognition rate was 95.31% in 12 poses.

Moeini and Moeini (2015) proposed a facial recognition system with pose and expression variations from real-world 2D face images. This system combined two

types of 3D methods, pose synthesis and filter transformation, using gallery images to generate a Feature Library Matrix (FLM) that used the 3D filter transformation method. Each subject in the gallery images, therefore, had an FLM based on face pose triplet angles. Pose synthesis was then used to extract features from real-world images. The authors matched the FLM matrix with feature extraction in a classification technique by using an SVM. The mean recognition rate from experiments achieved on the CMU-PIE database (68 subjects, 68 frontal images as the gallery, and 816 images as the probe) was 98.24%. FERET (200 subjects, 200 frontal images as the gallery, and 1600 images as the probe) achieved about 99.09% mean recognition rate, and there was a 93.16% mean recognition rate for the LFW database, which had 13,233 real-world facial images from 5,749 subjects of various face ages, facial poses, and illumination. Slowness/delay could be a problem with this system because additional time is required for 3D synthesis and the matching process.

In addition, other studies have been proposed to manage the issue of pose variation in facial recognition. Singh et al. (2007) created face mosaics as a form of panoramic view to enhance the face-matching system's performance. The mosaic scheme is illustrated in Figure 4-7. Singh et al. determined face coarse affine alignment for different poses then used phase correlation to detect blocks of 8×8 pixels for image segmentation, which supported the two views' pose to stitch. After three views, the images were connected and multi-resolution was applied on the connection boundary to generate a final face mosaic. Singh et al. treated face mosaics as a gallery and they were matched with probe face images with unconstrained poses by combining log-Gabor transform, C2 feature extraction, and a 2v-SVM classifier. The experiments were conducted with two datasets. The first was a CMU PIE face database (68 subjects), which used

images with a neutral expression as a gallery while images with slight variation in light and expression were used as probes. The WVU Multispectral face was the second database of 40 subjects, which were divided into two sets, visible (WVU visible-light) and short-wave infrared (WVU SWIR). The experiments showed different points:

- When gallery and probe images are mosaics, identification accuracy was achieved 100%.
- The matching performance of mosaic images as a gallery with non-mosaic images was better when a mosaic scheme was not used.
- The results obtained from the WVU SWIR database was better than WVU visible light. This may have been because of the conditions of illumination.
- The range of proposed system performance for mosaic images was between 96.85% and 100% in terms of identification accuracy.

The advantage of using mosaic schemes in gallery datasets is that the storage space can be saved as, instead of using multi images in matching, only a single image was used. The main weakness with mosaic schemes is that a minimum of three poses, including frontal, left and right poses, is required; and if one of these is missing, the system may not work. In addition, the pose direction in this system is the horizontal face pose and it ignores vertical face pose views.

Figure 4-7: (a), (b), and (c) are input images. (d) Mosaiced face generation. (e) The final image (Singh et al., 2007)

Cament et al. (2015) focused on Gabor features in face identification because it gives high accuracy and good results. Figure 4-8 shows their proposal, which includes three stages. The first stage consists of image alignments, using Active Shape Models (ASM) with local normalisation (LN) for illumination compensation. The second stage involves feature extraction using Gabor jet computation, which depends on eye positions as a central location to deform the grid of the face so it is similar to a frontal pose; and finally, a classification part, using a Borda count method and determining the pose variation by using a local statistical model. Cament et al. evaluated the approach by testing on the FERET with maximum pose rotation (-60^{o} to +60^{o}). The results show the mean recognition accuracy for the FERET database with face pose variation was 93% (i.e., a mean high performance with the frontal pose, 15^{o}, 25^{o} and 40^{o} degrees was 93% while the accuracy with the 60^{o} face pose angle was 76%). This study was an effective attempt to deal with pose variation of the face as a feature and to attempt to get a frontal pose without using 3D face representation, although facial recognition performance was still needed for more enhancement of large pose changes or vertical poses.

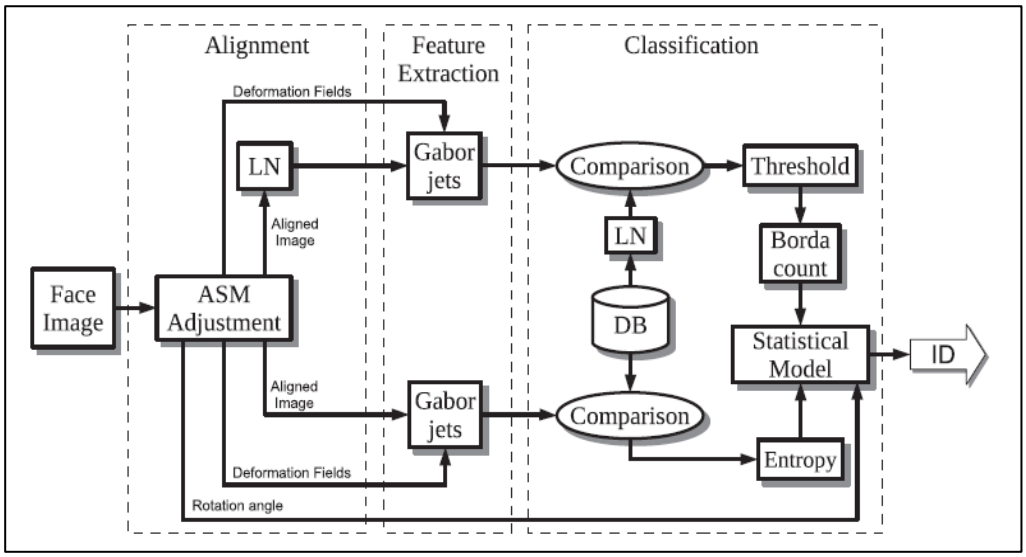


Figure 4-8: The three stages of the proposal system (Cament et al., 2015)

Kim et al. (2013) estimated head poses based on five multiple face templates (ranging from frontal pose to full left pose). They removed the illumination effects with unimportant details by applying the difference of Gaussian (DoG) filtering then divided the face into 10 X 10 grids and extracted the feature vector for each region by using a Local Binary Pattern (LBP). The close template to test the image was discovered by computing the distance between them then determining the corresponding local parts in the test image compared with the template and, finally, mapping each local part in the test image to the strict-frontal position. Kim et al. studied 10 facial parts in the strict-frontal template pose with the left side only because they supposed the right side was the symmetry side of the left. However, that could be considered as a limitation in their study because maybe there are better features on the right side. The matching technique in this proposal depends on finding the distance between two component images, so it needs to compute the same process for two images (Du et al., 2014). This approach was tested on the Multi-PIE public database and it achieved the highest overall accuracy of 72.2%. Further, this approach gave a good result with the verification system but the weakness was found when the pose angle was large, and then it did not discover all facial orientations with the five templates.

Lee et al. (2012) provided a sound feature extraction method that attempted to resolve the issue of matching unconstrained facial poses for probe images with a frontal pose database. They decomposed the facial appearance for each pose by using the Embedded Hidden Markov Model (EHMM) to extract the Subject Specific and Pose Oriented (SSPO) component, which included intrapersonal facial characteristics. The Adaboost weighting scheme was used as a classification technique to combine the SSPO features with the component classifiers. Figure 4-9 shows this approach. This approach could overcome the limitations of the previous study and could work with any facial pose. The evaluation of this approach used four databases. Table 4-3 shows the databases' descriptions and their features. The overall performance rate for close-set identification of three databases PIE, ORL, and IVLAB was 96.01% while for HONDA/USCD it was 95.27%. The results showed that by using the SSPO method to determine facial components, a good level of recognition accuracy can be achieved.

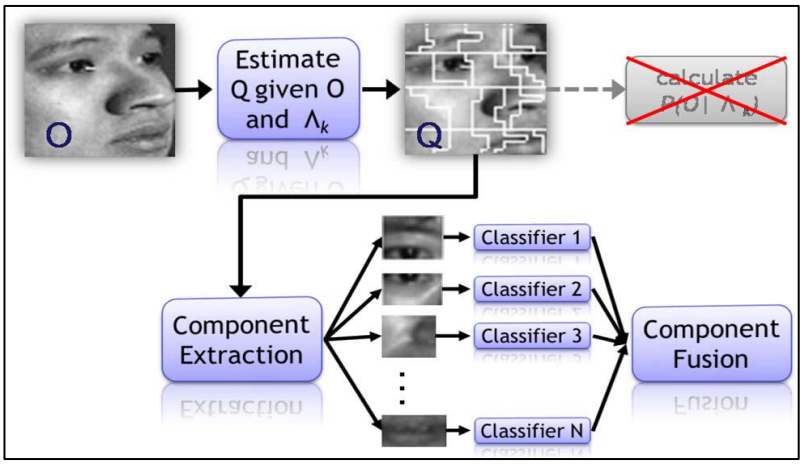


Figure 4-9: The full system approach (Lee et al., 2012)

Database	Subjects	Enrollment	Probe	Var
ORL	40	5	5	~ (+30°, -30°)
IVALB	15	60	30	~ (+60°, -60°)
PIE	68	8	5	~ (+90°, -90°)
Honda/UCSD	20	60	153	~ (+90°, -90°)

Table 4-3: Databases description (Lee et al., 2012)

Other challenges in facial recognition related to uncontrolled human action are facial expression and occlusion. Researchers tried to resolve more than one issue of facial images by combining two or more facial challenges in the study. Sultana et al. (2014) provided a study for recognising faces across facial expression changes further to invariant pose and illumination. They used two methods as feature vectors: the lower order Pseudo Zernike Moments (PZMs) method (Teh and Chin, 1988) and Daubechies Discrete Wavelet Transform (DWT) (Shen and Strang, 1998) with K-NN classifier to develop the performance of the facial recognition system across expression and pose invariants. First, the illumination and shadow effect was eliminated using an improved Weber-face method as the normalisation technique while preserving enough details for recognition purposes. Then the extraction feature vector and, finally, applying the k-NN classifier, which depends on the distance between the features to matching. Sultana et al. evaluated the proposal by testing on different datasets, as shown in Table 4-4. However, they disregarded variations in expression, pose, and illumination by using a small feature number in recognising facial images and applied a feature vector with classifiers on three final databases DB1, DB2, and DB3, as shown in Table 4-4. The results showed that the best recognition rate was on DB2: 98% with little or no expression changes and large illumination changes while a recognition rate of 97% was achieved on DB1 with large pose and expression changes and with little or no illumination changes. The authors maintained that this proposal was better than the PCA method when using the

same databases (Sultana et al., 2014). This system constituted an effective attempt to ignore uncontrolled facial image issues.

Database	Subjects
AT&T	40
AR	70 male, 56 females
Yale	15
Sheffield	20
Final databases	
DB1	40 subjects randomly chosen from AT&T, AR and Sheffield databases.
DB2	40 images randomly chosen from Yale and AR databases.
DB3	Contains all images in DB1 and DB2.

Table 4-4: Facial image databases description (Sultana et al., 2014)

Bhat and Wani (2015) adopted an Elastic Bunch Graph Matching (EBGM) algorithm as the facial recognition approach, with changes in facial expression, pose, and lighting. There are five steps that have to be followed in the EBGM method to identify a face:

- Normalisation: in this step, eye location is needed to rescale the image to 128 pixels on an edge and to normalise image brightness.
- Landmark Localisation: the user selects different points on the facial image to create a bunch graph.
- Face graph creation: using landmark points and creating the structure of the face graph.
- Distance measurement: computing the similarity graph vectors for each pair in the database.
- Identification: choosing the close distance to identify faces.

They depended on three databases for their study: the ORL dataset (40 subjects, 400 images) was used to evaluate the system with facial expression variation; the Yale B dataset (65 subjects, 2470 images) was used for illumination condition evaluation; and, part of the FERET dataset (20 subjects, 200 images) was selected for pose variation. The experiment was divided into two parts: in the first, the authors used the training set, including all the same images from the testing set so, in this case, the recognition rates reached 100% for three face conditions. This is a normal result because all images could be found in the two sets. In the second experiment, the authors divided the subject images into two, half for the testing set and the other half for the training set; in this case, the recognition rates achieved 91.5% for facial expression variation, 65.78% for the illumination condition, and 77% for pose changes (Bhat and Wani, 2015). Generally, this performance achieved good results but each issue process separated from the other with their dataset, so there was no experiment to collect all issues in one system.

Liao et al. (2013) studied how to identify any suspect in a large crowd of people with uncontrolled captured images, such as pose changes, illumination variation, partial occlusion, and disguise. They proposed a partial face-matching technique that included various categories of partial images, such as facial accessories (e.g., hat, sunglasses, scarf, and mask), out of the camera, occlusion by other subjects, non-frontal pose, etc. The Multi-Keypoint Descriptor (MKD) method was used for representing probe image and gallery database features. Further, face probes were achieved by using multitask sparse representation. In addition, the Sparse Representation-based Classification (SRC) technique was used for face matching. This proposal works on both holistic and partial faces, so it automatically determined the type of image from the descriptor details that give

the length size. However, the size of the holistic face descriptor will be greater than that of the partial face descriptor. An additional descriptor technique, the Gabor Ternary Pattern (GTP), was used. It gave more definition to local facial components and it was robust with regard to illumination changes. The evaluation of this proposal was applied to four databases FRGCv2.0, AR, LFW, and PubFig, as illustrated in Table 4-5. The proposed method, MKD-SRC-GTP, was compared with different facial recognition approaches to evaluate the performance by using a ROC curve (i.e., Detection and Identification Rate (DIR) versus FAR). However, the overall results of four databases prove the MKD-SRC-GTP method performs well for the general partial facial recognition issue when compared with other commercial techniques. The ROC curve was about 95%, 90%, 98%, and 16% of 20% for FRGCv2.0, AR, LFW, and PubFig datasets, respectively. This approach was computationally dense but the results showed improvements in partial facial recognition issues with different partial categories.

Datatsets	Scenario	Subjects	Probe	Gallery	Characteristic
FRGCv2.0	Partial patch	20466	25562	10466	Synthesized partial faces.
AR	Occlusion	20135	11530	10135	Occluded holistic faces.
PubFig	Pose & occlusion	5140	8027	5083	Occluded or nonfrontal faces collected in unconstrained condition.
LFW	Pose & occlusion	5749	6000	6000	

Table 4-5: Database description and methods used in experiments of (Liao et al., 2013)

Similarly, Weng et al. (2013) attempted to match partial faces with full faces and, as with (Liao et al., 2013), these did not require manual alignment. They extracted local feature key points instead of holistic features using the Scale-Invariant Feature Transform (SIFT) then concatenated with the speeded up robust features

(SURF) for both probe and gallery images. They designed a new approach for matching, called Metric Learned Extended Robust Point Set Matching (MLERPM), which was based on matching two feature vectors geometry features and texture features for probe images and gallery images. There was an error ratio when matching key points in the probe image that did not appear in the gallery image. To evaluate the performance of their proposal, Weng et al. adopted three datasets: the LWF dataset (5749 subjects; 13233 images, with only 1680 subjects having more than two images), and with variation in lighting, expression, resolution, and makeup; the AR dataset (126 subjects, 70 males and 56 females), including different illumination, expressions and facial disguises (sunglasses and scarf); and the extended Yale B dataset (38 subjects, 2414 frontal faces). The system results demonstrated recognition accuracy for the LFW dataset 50.72% at Rank 1 and 72.75% at Rank 20. The AR dataset for sunglasses and scarf conditions achieved 97.5% while it achieved a different facial recognition accuracy on Extended Yale B up to 98.3% for occlusion rate 30% and 30.2% accuracy for 50% occlusion rate (Weng et al., 2013). Although these results are normal for these conditions, it was observed that the performance was poor on high occlusion images. Furthermore, this method was robust when discriminative facial components were available.

Min and Dugelay (2012) studied another type of occlusion called sparse occlusion, in which faces occluded by stains, text, orthogonal grid, and diagonal grid were differentiated from dense occlusion faces, such as those with scarves and sunglasses. This study suggested using the Robust Principle Component Analysis (RPCA) method to detect automated sparse occlusion parts on faces, after which the Field-of-Experts (FoE) model was applied to inpainting the occluded parts. Min and Dugelay utilised this approach to improve recognition

rates in video surveillance. They evaluated their approach by using the AR database (126 subjects, 4000 images, 70 men and 56 women) with various facial conditions such as expressions, illumination, and occlusion. They tested the proposal by applying three facial recognition algorithms: PCA, Scale Invariant feature transform (SIFT), and Local Binary Patterns (LBP). The results showed improved recognition rates for all three algorithms. The improvement rate for the SIFT technique was more than 75% for all four types of occlusion faces. The results of the LBP technique were closer, because it is more robust for occlusion faces, although, for diagonal grid face images, it improved by 74%. This study was a good attempt for specific types of face occlusion images.

Recently, most researchers are interested in recognising faces from the wild to find faces from a million photos, which is considered a difficult challenge for facial recognition. Wang et al. (2017) proposed a face search system by fusing the number of COTS matchers. Firstly, they filtered a large gallery by using features learned by a convolutional neural network to find the top-k most similar faces. These top-k features were recognised by the COTS matcher. After that, the researchers fused the deep features with the COTS matcher to improve the overall performance. This study used one mugshot dataset and four web face datasets in its experiment, which are PCSO, LFW, IJB-A, CASIA-WebFace, and built a private dataset called “Web-Face”. Web-Face was built by downloading millions of web images that were filtered to ignore all images without faces. Finally, 80 million facial images were collected. The fusion technique used in this study improved the overall performance to 99:5 percent TAR@FAR of 0.01 percent. This study was considered a good attempt to recognise and search for faces among millions of images and this scenario helped the investigation in digital forensics. Additionally, Wang et al. (2017) were using real criminal case

images of the Tsarnaev brothers (the Boston Marathon bombing was described in Section 1.1). They found the younger brother's (Dzhokhar Tsarnaev) photo at rank 1 in one second on a 5Mgallery and at rank 8 in seven seconds on an 80Mgallery.

4.6 Soft Biometric Attributes

Many researchers have used multi-biometric systems to enhance biometric recognition applications. For the facial recognition technique, the researcher added specific details to increase the matching accuracy. Soft biometric traits, such as tattoos, facial marks (e.g., scars, moles, and freckles), gender, height, and eye colour, are considered helpful in enhancing the performance of the facial recognition system (Flynn et al., 2008). Moreover, in face matching, the soft biometric attributes could contribute to the minimisation of large databases by filtering data according to its features and could be used to differentiate between identical twins or as visual evidence in court (Park and Jain, 2010). However, the limitation of soft biometrics is that the traits could not be considered as primary features for recognising people because they could be shared by people and so it serves to support other recognition systems in making a decision.

Park and Jain (2010) suggested using facial marks, gender and ethnicity to increase face matching and retrieval performance. They studied occluded or partial face image cases where facial marks support these categories and they filtered databases by using gender (i.e., male, female, and unknown) and ethnicity (e.g., Caucasian, African-American, and unknown) as demographic information that does not change over time. This study adopted an automated method to determine facial marks by applying the Active Appearance Model (AAM) to determine and remove primary facial components, such as eyes, nose, and mouth. Then, they detected the silent facial marks by using the Laplacian-of-

Gaussian (LoG) operator. Finally, the authors combined two face-matcher models, the mark-based matcher and a commercial face matcher (FaceVACS), to improve the matching process. Park and Jain built two databases DB1 (213 subjects and 213 images for the probe, 10213 subjects and 10213 images for the gallery) from mugshot faces (probe) and the FERET (gallery) database while DB2 (554 subjects with 554 images for the probe, 671 subjects with 671 images for the gallery) consisted of a mugshot face database. They also built two additional databases. DB3 was collected from a video for five subjects and DB4 included five identical twins. Because of the low resolution of images in DB2, the authors used it for the statistical analysis of marks. This framework was tested by several experiments, as follows:

- They used DB1 to match and retrieve faces, so the recognition accuracy at rank-1 improved from 90.61% (only used Face VACS matcher) to 91.08% (adding marks with the FaceVACS matcher). Moreover, the accuracy rate of applying marks with the FaceVACS matcher on gender and ethnicity increased from 91.55% to 92.02%, demonstrating there is a slight improvement in their approach.
- Face occlusion from video frames (DB3) was examined by applying the soft biometric matcher after the failed the commercial face matcher to retrieve five probes face occluded images. So, the proposed soft biometric traits-based matcher was successful at ranks 3, 4, 6, 7, and 8.
- The authors tried to distinguish between identical twins, so they applied the soft biometric matcher on DB4. After the FaceVACS failed to recognise identical twins, the soft biometric marks helped the mark-based matcher to distinguish five pairs of identical twins. However, gender and ethnicity did not help in this case because they were shared between identical twins

within the dataset. In addition, the AAM method did not work well because of the light condition, so the authors had to detect primary features manually.

These results support a forensic application because they depended on automatic extraction for facial marks. The limitation of this work is that it depended on good quality images with normal facial condition and so, when images in DB4 were with strong light, the manual face component detection technique was chosen instead of the automatic technique.

Tiwari et al. (2012) attempted to identify 210 newborns by using the face as a primary biometric while gender, height, weight, and blood were used as secondary biometric traits. The challenge was uncontrolled on newborn facial conditions, so the database images that were collected for one year included variations in expression, pose, and illumination. The vector feature of the newborn face or the posterior probability were extracted and secondary biometrics were computed; then this system was evaluated by implementing four algorithms PCA, Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), and Local Binary Pattern (LBP). The images of newborns (10 images per infant) were divided into a training database (six images per subject and 1260 images totally) and a testing database (four images per subject and 840 totally). The experiments showed that four biometric trails improved recognition by 6% when using just faces. Gender was the worst feature for improving facial recognition accuracy while the increased accuracy rate was 1.5% for blood group, 3.1% for height, and 2.1% for weight, compared with using only the facial recognition, as shown in Table 4-6. This study was a good attempt to analyse infant facial recognition issues. Infants' faces change quickly and any feature vector for an infant's face will change. In addition, most secondary biometrics are not unique, could change

(for example, weight or height except blood group), and are not easily found within a standard image dataset.

Procedure	F%	F+G%	F+H%	F+W %	F+B%	F+G+H+W+B%
Identification Accuracy (Rank-1)	80.42	82.40	83.12	82.10	83.60	86.80

Table 4-6: Accuracy of the face and biometric identification (Tiwari et al., 2012)

Tome et al. (2015) tried to translate a set of facial features into information that could be understood by judges or forensic investigators who need a supporting tool to distinguish between people from their facial images. They took forensic methodology aspects into consideration to produce a study that would be useful in real criminal investigation cases. They divided the facial features into two groups: continuous and discrete. The continuous features represented facial landmark measurements, such as height and width of the nose and mouth, while the discrete features represented the facial landmark shapes, such as the eyebrow shape, such as arched and rectilinear. The facial landmarks were extracted manually from frontal faces. The authors studied and analysed the correlation, stability, and discriminative facial soft biometrics and found that these traits improve the forensic facial recognition accuracy system. In addition, they divided faces into regions according to their discriminative power. They adopted two frontal face databases to extract soft features: the ATVS forensic database (50 subjects, eight samples per subject) and a subset of the MORPH database (130 subjects, six samples per subject). The identification performance of ATVS rank 1 was 48.75% and achieved 100% rate at rank 5. For the second database, MORPH DB, the identification rate at rank 1 was 23.84%, 62.05% for rank 5, and 75.13% for rank 10. For the overall results, the results for the first database were better than the second because the latter one had more unconstrained conditions

and low-quality images. Also, the research team found that the nose and forehead areas have more discriminative information than the eye and eyebrow areas do.

Gonzalez-Sosa et al. (2018) used soft biometric such as gender, ethnicity, age, glasses, beard, and moustache to improve the facial recognition system in unconstrained scenarios. They explored improving two COTS facial recognition systems based on deep learning by fusing soft biometric data. Their experiments evaluated by using the labelled faces in the wild (LFW) database. Firstly, they considered two suggestions to estimate soft biometric information: manual and automatic. The experiments' results showed that soft biometrics are considered valuable data and relate to the face in unconstrained scenarios. There were improvements up to 40% and 15% in the performance of the verification when using manual and automatic soft biometric data, respectively.

4.7 Commercial Face Identification Systems

Recently, many companies have been interested in publishing facial matcher software as a result of the increased interest in this field. Therefore, these commercial software aim to overcome the drawbacks of existing studies and identify unconstrained face images. For this reason, today's many researchers are used to these types of facial-matching systems in their studies. For example, Klontz and Jain (2013) conducted a study of the Boston Marathon bombings of 2013 and analysed the reasons why the automated facial recognition system failed to identify the suspected persons at that time. They used three commercial matchers: NEC NeoFace 3.1, Cognitec FaceVACE 8.6 (they chose these two systems based on their top performance in the National Institute of Standards and Technology (NIST) Multiple Biometrics Evaluation), and PittPatt 5.2.2 (acquired by Google). Their study concluded that forensic facial recognition systems

operate under unconstrained faces of people in the presence of digital surveillance cameras and need more progress to overcome such issues.

Wang et al. (2017) used a few commercial off-the-shelf (COTS) systems to search for persons within large-scale photos by using deep features while Best-Rowden et al. (2016) evaluated one of the COTS system's face matchers on their Newborns, Infants, and Toddlers Longitudinal face image database to explore the ability of face identification on children faces. Their result showed that facial recognition technology still has complexity to recognise young children's faces. In addition, Juefei-Xu et al. (2015) studied the performance of COTS facial recognition systems on partial faces or occluded facial parts. However, they found inconsistencies existed in the COTS systems depending on image sources, especially occluded faces.

As demonstrated above, existing studies have attempted to deal with different commercial systems of facial recognition to identify suspects. There are some advantages of focusing on commercial facial recognition systems in digital forensics, such as they could release the forensic team from building their own algorithm to matching faces in criminal investigations. In addition, investigators could benefit from using updated facial recognition systems by the state of the art in underlying recognition performance.

Nowadays, many companies, such as Amazon, Microsoft, NEC NeoFace, and Google, provide facial matching systems. However, there are some limitations that researchers could encounter when using these types of algorithms in their studies, such as high costs (e.g., NEC NeoFace) and requiring consent from governments and law enforcement organisations (e.g., Cognitec, faceVACS). Furthermore, there is a variance between commercial facial recognition systems in their ability to accept all image issues such as quality, light degree, and face

orientation degree. Therefore, it is essential in the criminal investigation field that researchers check the characteristics of the commercial facial recognition systems that suit their aims.

4.8 Discussion

As mentioned previously, several problems in facial recognition systems need to be addressed to improve their performance, permitting them to offer more reliable assistance to the digital forensic investigation field. Table 4-7 summarises all studies presented in this chapter. Amongst these issues, facial ageing plays a role in the forensic facial identification application. Several studies adopted generating new faces at different ages to minimise the age gap in the face-matching technique. For example, the work of Kemelmacher-Shlizerman et al. (2014) achieved a good level of performance in the age-progression area although they used human decisions instead of an automated identification system. In comparison, others preferred to use the discriminative approach to solve the ageing issue in the facial recognition system, including Li et al. (2011), Sungatullina et al. (2013), and Pal and Gautam (2015). They used the local features of faces as a method to achieve good results in the identification system. In addition, soft biometric traits (e.g., scars, moles, and freckles) were also investigated to improve the accuracy of age estimation. For instance, the experimental results of Ling et al. (2010) demonstrate the recognition performance is improved by using soft features. In particular, these methods could be effective in dealing with adult faces where the face shape normally does not change but where identification is limited by accessories (glasses, hat, scarf, etc.).

Other factors that could play a key role in the face-matching process are external factors (i.e., illumination and resolution). As demonstrated in Section 4.4, several

methods were investigated for enhancing recognition accuracy, such as illumination filter enhancements, the shadow compensated technique, and the super-resolution method. Despite all current studies, external factors still pose significant challenges in the performance of facial recognition systems, especially in digital forensics because most evidence images come from CCTV, which can be located indoors or outdoors and can be affected by changes in light and camera quality.

Another problem in forensic facial recognition is face pose and expression changes. Most studies on face pose variation have limitations in the degree of facial pose angle, with the complexity increasing when the angle increases. However, this issue, and the issue of expression changes, can be managed by extracting soft biometric features or landmarks of the face and further dividing the face into regions and matching each part with the same region in database images, such as in studies by Singh et al. (2007) and Cament et al. (2015). These suggestions could serve to filter a large database and allow for focus on the best face region in matching. Other studies preferred using a 3D model to solve change in the facial pose, such as Asthana et al. (2011), which converted frontal pose by creating 3D face viewing from a 2D image. In some cases, the using 3D model made the system more robust because of the high discriminative information and less sensitive to environmental variations, such as illumination. In contrast, the main drawback of the 3D model is requiring all the elements of the system to be well-calibrated and synchronised to acquire accurate 3D data (texture and depth maps). In addition, costs of the set-up and additional time needed for processing data should also be considered when using a 3D module. Nevertheless, the decision on whether to use a 3D model to improve the

performance depends on several factors, such as case background, the overall cost involved, and the data type available.

In comparison to dealing with a single issue, numerous researchers have tried to solve multiple challenges within the system. For instance, Bhat and Wani (2015) studied face expression, face pose, and illumination. However, each evaluated system performed two stages: firstly, for large lighting changes and, secondly, for large pose and expression change. Therefore, their system was limited to collecting all issues in one experiment. Another good attempt was by Liao et al. (2013) who identified suspect persons' images with uncontrolled images captures, such as partial face and variation in lighting and face pose. They focused on partial and disguised faces in images and achieved good accuracy but, still, the commercial technique as higher accuracy.

Several studies have sought out all previous issues and found that when the face's soft biometric features are used, the accuracy of recognition improves. This feature is able to filter a large database for quick and easy matching and could help with various cases, such as distinguishing identical twins and dealing with partial faces, occluded faces, and changes in facial posture. A number of 2D databases are used in facial ageing experiments but the most popular ones are FG-NET and Morph Album2 because they include various ages for each sample. Furthermore, the popular databases that support illumination studies are the Extended Yale B and CMU-PIE databases. They are characterised by containing samples with different light angles. The FERET and CMU PIE databases are mostly used for experiments on face pose changes while the AR dataset is suitable for partial face experiments.

There are new studies that utilise commercial facial recognition systems in their research. These types of algorithms or systems (commercial) could save time for

researchers to build or improve new face-matching algorithms. In addition, they do not need to have deep knowledge of algorithms. Commercial systems have continuously enhanced their face-matching algorithms and have tried to integrate any new method that has emerged in their systems as a result of the extreme competition between companies. Therefore, most researchers in digital forensics are moving toward the use of these types of facial recognition systems and have included them in their studies (as shown in Section 4.7). For instance, Table 4-7 shows a number of studies have solved facial recognition issues by using COTS algorithms, such as Park et al. (2010), Best-Rowden and Jain (2018), and Best-Rowden et al. (2016), which implemented several COTS algorithms to solve the ageing issue. Also, Wang et al. (2017), Park and Jain (2010), and Gonzalez-Sosa et al. (2018) implemented COTS algorithms to solve issues in forensic facial recognition scenarios.

As has been shown, numerous investigations have been conducted into the effects of different applications of facial recognition. Although many researchers have attempted to solve more than one issue in their studies, few attempts have been made to combine all techniques involved in criminal investigation cases. The major challenges of digital forensics investigation with facial recognition are accuracy, processing more than one issue in the system, limited time, and size of the data that need to be managed. It is, therefore, necessary to seek the best method of solving facial recognition problems and to construct a framework that includes most of these methods in the forensic application.

4.9 Conclusion

This chapter presented a comprehensive analysis of existing work in forensic facial recognition and highlighted the challenges within this field. These challenges are facial ageing, external issues (change lighting and image quality),

and internal issues (face pose, facial expression, and partial face). Accordingly, various studies have been published that provide incremental improvements to solving issues individually. Further, a few researchers have suggested solutions to fusing some of these issues into one system so they can be studied. Furthermore, in the last few years, the research on facial recognition has included deep learning techniques or commercial algorithms to improve the previous studies' results. It is noticed that this type of algorithm makes using the facial recognition technique easier (there is no need for deep experts in recognition algorithms) and promise improved accuracy.

Despite previous literature review, more effort is required to optimise solutions in an integrated framework that deals with the facial recognition issues holistically, in terms of digital forensics, and understand how well they can perform on a range of real-world datasets.

Author/Year	Matcher Algorithm	Dataset	Result	Measure types	Age or Face Pose
Facial Aging					
Park et al. (2010)	COTS	FG-NET	37.4	Rank1	-
		MORPH-Album1	66.4		
		BROWNS	28.1		
Mahalingam and Kambhamettu (2010)	Age Model	FG-NET	70%-80%	Rank10-Rank20	0-69
	The potential individual		69%	Rank10	
(Juefei-Xu et al., 2011)	unsupervised discriminative projection	FG-NET	100%	Rank1	-
Kemelmacher-Shlizerman et al. (2014)	Human decision	(40,000 photos) Google search images	-	-	1-80
Ling et al. (2010)	A Support Vector Machine (SVM)	FG-NET	30.5%	EER	8-18
			38.6%		0-8
		Passport I	8.9%		>=18
		Passport II	11.2%		
Li et al. (2011)	MLBP	MORPH album 2	83.9%	Rank1	-
		FG-NET	47.50%		
Sungatullina et al. (2013)	MDL	FG-NET	91.8%	Rank 1	-
		MORPH album 2	65.2%		
Pal and Gautam (2015)	SVM	FG-NET	76.6%	Rank 1	-
Best-Rowden and Jain (2018)	COTS	LEO_LS	~99%	0.01%FAR	0-17
		PCSO_LS			
Best-Rowden et al. (2016)	COTS	NITL (same session)	93%	TAR at 0.1% FAR	0-4
		NITL (cross session)	47.93%		
External Factors (illumination and resolution)					
		Yale B	98.9%	R1	-

Nabatchian et al. (2010)	SVM and k-nearest Neighbors rule	extended Yale B	94.68%		
Choi et al. (2011)	Nearest Neighbor rule	CMU-PIE	99%	Rank 1	-
		Yale B	92.3%		
Luan et al. (2014)	the weighted based method	Extended Yale B(subset4	95.06%	Rank 1	-
	the ratio-based method	Extended Yale B(subset4)	54.18%		
Guo et al. (2017)	Deep Network Model	LFW	98.95%	Rank 1	-
		YouTube Face	97.3%		
Internal Issues (pose, partial, expression)					
Asthana et al. (2011)	Local Gabor Binary Patterns (LGBP)	CMU-PIE	99%	Rank 1	-
		FERET	95.6%		
Moeini and Moeini (2015)	SVM	CMU-PIE	98.24%.	Rank 1	Pose
		FERET	99.09%		
		LFW	93.16%		
Singh et al. (2007)	2v-SVM	CMU PIE+WVU Multispectral face	96.85%	Rank 1	Pose
Cament et al. (2015)	Borda count method	FERET	93%	Rank 1	Pose
Kim et al. (2013)	LBP (distention measure)	Multi-PIE	72.2%	Rank 1	Pose
Lee et al. (2012)	Adaboost weighting scheme	ORL	96.01%	Rank 1	Pose
		IVALB			
		PIE			
		Honda/UCSD	95.27%		
Sultana et al. (2014)	k-NN	DB1(AT&T, AR and Sheffield)	97%	Rank 1	Expression, pose,illumination
		DB2(Yale and AR)	98%		

		DB3(DB1 and DB2)	96.5		
Bhat and Wani (2015)	Distance Measurement	ORL	91.5%	Rank 1	expression, pose, and lighting
		Yale B	65.7%		
		FERET	77%		
Liao et al. (2013)	MKD-SRC-GTP	FRGCv2.0	95%,	ROC	large crowd of people
		AR	90%,		
		LFW	98%,		
		PubFig	16% of 20%		
Weng et al. (2013)	Metric Learned Extended Robust Point Set Matching (MLERPM)	LWF	50.72%	Rank 1	partial faces
		AR	97.5%		
		extended Yale B	98.3%		
Min and Dugelay (2012)	PCA, SIFT and LBP	AR	~>74	Rank 1	sparse occlusion
Wang et al. (2017)	COTS	PCSO, LFW, IJB-A, CASIA-WebFace and Web-Face	99:5	TAR@FAR of 0.01	Web images
Soft Biometric Attributes					
Park and Jain (2010)	COTS	DB1	91.08%		gender and ethnicity
Tiwari et al. (2012)	PCA	Private	74.34%	Rank 1	new-borns, gender, height, weight, and blood
	ICA		78.12%		
	LDA		80.15%		
	LBP		82.76%		
Tome et al. (2015)	Continuous Discrete Fusion	ATVS forensic	48.75%	Rank 1	
		subset of MORPH	62.05%	Rank 10	
Gonzalez-Sosa et al. (2018)	multi COTS	LFW	88.66 and 93.30	Rank 1	

Table 4-7: Summarise the current state of art

5 An Investigation Into Forensic Facial Recognition

5.1 Introduction

As illustrated in the previous chapter, the major issue concerning forensic facial recognition is the need to overcome a group of facial recognition problems within a single holistic system. The ability to address issues regarding image resolution, lighting, face pose, facial expression, and occluded faces is considered an essential requirement in a forensic facial identification system, as this has a major impact on the system's performance. Moreover, existing forensic tools seek to involve the aforementioned facial identification challenges, especially in the analysis of digital multimedia evidence and image files. By using facial recognition technologies, valuable information used in identifying culprits can be extracted from photographs or videos that are taken at crime scenes (Peacock et al., 2004). As a result, automating the process of suspect recognition can save forensic investigators an immense amount of time when compared with search tasks carried out manually by watching videos.

Previous studies have suggested a solution to facial recognition issues but, to the best of the author's knowledge, these studies rely on including a single issue in their experiments, rather than all issues. In addition, some of the studies presented in the previous chapter did not conduct experiments with a real facial dataset, which leaves concerns regarding accuracy. As a result, there is still a need to investigate a new facial recognition approach that will meet forensic investigation requirements.

With the aim of overcoming the above-mentioned facial recognition challenges, this investigation derived several research questions that need to be identified in the forensic facial recognition field. This research investigates:

- How to determine a baseline performance of a number of commercial facial recognition algorithms individually by using a facial dataset with varying issues such as pose, light, and expression.
- How the performance of the previously examined algorithms changes when using a realistic dataset that is close to or simulates real-world digital forensics scenarios.
- Whether a multi-algorithmic approach would improve underlying classification performance by using a fusion mechanism.

The remaining sections of this chapter describe the methodology of the experiments, details of the facial datasets, and present the results. The chapter then presents an overall discussion of the three experiments conducted.

5.2 Experimental Methodology

5.2.1 Recognition Algorithms

This research used three commercial facial recognition algorithms: Neurotechnology, Microsoft, and Amazon Rekognition. To the best of author's knowledge, these algorithms are not implemented by studies in the field of forensic facial recognition, so it is considered a good challenge for this research. In addition, the research focused on using and evaluating commercial algorithms because this could bring several advantages in practice. It was clear from the literature review that an extensive volume of literature has been undertaken with a good number of commercial systems in place. Further, it would relieve forensic investigators of having to design, implement, and manage facial recognition systems. Leaving specialists to manage an application-independent facial recognition system would also lead to algorithmic improvements and updates beyond a forensic team's capability. Also, commercial algorithms are built based on years of research and experts in the biometric field (facial recognition) and

they adopt biometrics standards, such as data exchange and interoperability (i.e., Internationational Standards Organisation (ISO)). In addition, some of the commercial algorithms provide cloud sources that could be used to speed up the process, eliminate the storage capacity, and allow for collaboration from different locations.

The first commercial system selected in this research was Neurotechnology. Neurotechnology was one of ten algorithms evaluated in the Multiple Biometrics Evaluation (MBE) test report of 2010 by the National Institute of Standards and Technology (NIST) (Grother et al., 2010). In addition, a Neurotechnology software development kit (SDK), called the VeriLook SDK, was designed to be PC-based and supports different programming languages, allowing further rapid development of a biometric application with the help of libraries and functions. It also includes programming samples and tutorials that show how to use the SDK's components. The VeriLook is selected by several national-scale customers, such as Lenovo computer is selected VeriLook to be PC user authentication for some webcam-enabled notebook computers. Border control in a Spanish airport used VeriLook to access the border system for European citizens quickly. More information about this software can be found at https://www.neurotechnology.com.

Microsoft and Amazon are two well-known companies that provide an application programming interface (API) that facilitates the process of facial recognition by interacting with their cloud-based developed algorithms. APIs are the second generation of SDKs and consist of a set of routines or blocks that make programming easier and release users from having to understand how to use libraries and functions. The API service supports different programming languages or platforms, which allow code to be written to use the API's services.

Table 5-1 shows full details of the three algorithms selected for this research to explain their abilities and limitations and enable comparison between them.

Features	Neurotechnology	Microsoft	Amazon Rekognition
Face detection	√	√	√
Multiple face detection	√	√	√
Face recognition (image)	√	√	√
Face recognition (video)	√	-	√
SDK	√	-	-
API	-	√	√
Gender determination	√	√	√
Age estimation	√	√	√
Emotion recognition	√	√	√
Deep learning recognition	√	√	√
Time of open source	1 month	1 year	1 year

Table 5-1: Comparison between the three systems used in the research

5.2.2 Facial Image Datasets

This study proposes a means to determine criminal identity when required in a forensic investigation based on the face. Thus, choosing a suitable dataset was considered as an essential step in this research to ensure that various face-related challenges would be examined. This research chose two facial datasets for its experiments: one is available publicly, CAS-PEAL-R1, and the author collected the second set of data. The methodology for creating the two databases is described in the following sections.

A. CAS-PEAL-R1 Dataset

A large number of facial datasets have been collected and are publicly available because of the importance of facial recognition in various emerging fields, such

as computer vision, security, and digital forensics. The choice of an appropriate dataset was aimed at being able to study and evaluate the performance of algorithms with variant facial issues, such as facial expression, lighting, and pose. Table 5-2 illustrates an overview of some public facial databases with their face conditions. It can be noticed there is some limitation either in missing some facial issues that must be included in this study, such as FERET, AR, and CMU PIE, or the number of subjects is considered small and does not satisfy the practical requirements, such as Yale B, and E-Yale B. This research is deemed important to include as many general forensic facial recognition challenges in one dataset as possible with accepted subjects.

Database	#Subjects	Pose	Expression	Accessory	Lighting	Distance
Yale B (Belhumeur et al., 1997)	10	√			√	
E-Yale B (Georghiades et al., 2001)	28	√			√	
CMU PIE (Sim et al., 2002)	68	√	√		√	
AR (Martinez, 1998)	126		√	√	√	
FERET (Phillips et al., 1998)	1199	√	√		√	√
CAS-PEAL-R1	1040	√	√	√	√	√

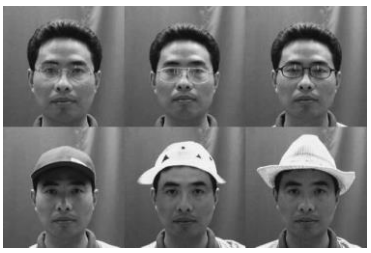
Table 5-2: Overview of some current facial databases

Therefore, the baseline experiment employed one of the current public datasets: the CAS-PEAL-R1 Chinese face dataset (Gao et al., 2008). The dataset was collected by the Chinese Academy of Sciences (CAS) between August 2002 and April 2003. It consists of 30,900 images across 1,040 subjects (595 men and 445 women) for seven categories: pose, expression, accessory, lighting, background, distance, and time. Table 5-3 summarises the CAS-PEAL-R1 face dataset content, which includes two main subsets: frontal and pose. Figure 5-1 shows examples of each set.

1. In the frontal subset, all the images show people looking towards the camera. The subjects have six different expressions between them (normal, smile, frown, surprise, closed eyes, and open mouth), six different accessories (three types of hat and three types of glasses), and at least nine lighting angles. Furthermore, there are other subsets which are background, distance, and ageing (two sessions at a six-month interval).
2. In the pose subset, all the images are taken across 20 different poses. The poses are divided into upward (up to 30°), right (up to 90°), left (up to 90°), and downward (up to 30°).

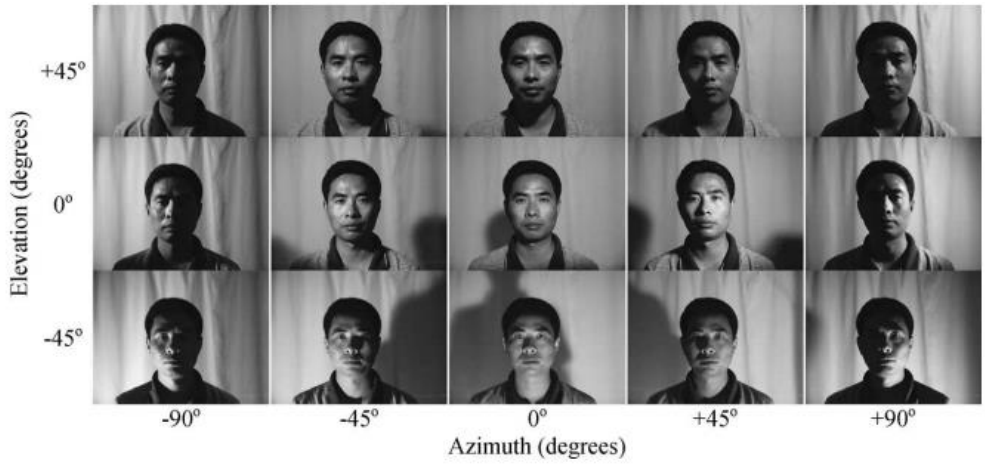
Subset		#Sample	#Subjects	#Images
Frontal	Normal	1	1,040	1,040
	Expression	5	377	1,884
	Lighting	≥ 9	233	2,450
	Accessory	6	438	2,646
	Background	2-4	297	650
	Distance	1-2	296	324
	Aging	1	66	66
Pose		20	1,040	20,800

Table 5-3: Details of the CAS-PEAL-R1 dataset (Gao et al., 2008)



(a)

(b)



(c)

(d)

Figure 5-1: Example images from the CAS-PEAL-R1 dataset: (a) Accessory subset, (b) Expression subset, (c) Lighting subset, and (d) Pose subset (Gao et al., 2008)

The methodological reason for this study using a subset of the CAS-PEAL-R1 dataset (95 subjects only) was that the 95 subjects met all the main conditions (Pose, Lighting, Expression, and Accessory) while the other subsets missed at least one of the conditions. This decision was reached after studying the nature of CAS-PEAL-R1 and viewing the statistics for all the subjects' (1,040) photographs. The remaining subsets (timing, background, and distance) were excluded from the study because the period was just 6 months and was not a large enough difference in time for facial features to change. The background subset changed the unicolour blanket background and this was similar to the Lighting set. Finally, the maximum distance used in the dataset was 1.2 meters (this was considered a small distance). Table 5-4 shows a breakdown of the CAS-PEAL-R1 dataset used in this study.

Subset	#Subjects	Image/subject	#Images
Normal	95	1	95
Accessory	95	6	569
Expression	95	5	475
Lighting	95	≥ 9	1,203
Pose	95	20	1,900
Total			4,242

Table 5-4: Subsets of the CAS-PEAL-R1 dataset used in Experiment 1

The final subject's names statistics for the 95 substances used in this study can be a good result to assist other researchers in the future when using the CAS-PEAL-R1 data set. These individuals were the first 100 subjects in the dataset (from subject 000001 to subject 000100). Five subjects (000030, 000031, 000037, 000043, and 000074) were excluded because they did not satisfy the condition (four facial recognition issues, expression, lighting, pose, and accessories).

B. Collection of a Realistic Dataset

As shown in the previous section, the first dataset (CAS-PEAL-R1) was the most variant face dataset in its ability, in terms of transparent so it is used initially to evaluate the algorithm's performance. However, this dataset is not reflect a forensic problem (i.e., forensic face images). Forensic images (as shown in Chapter 1) could suffer from more issues than standard images such as bad quality, distance from the camera, uncooperative subject, different background, complex accessories (beard, mask), bad illumination, and face orientation. To fulfil the requirement of the second experiment, a more realistic face dataset was required to examine the performance of algorithms with a dataset that simulates law enforcement evidence requirements.

Although a number of face datasets were collected from the internet, such as Labelled Faces in the Wild (LFW) (Huang et al., 2007) and CelebFace (Sun et al., 2013), they do not include enough samples for each subject under unconstrained facial capture challenges (e.g., lighting, pose variation, and expression) that this study has attempted to address.

As a result, the author collected a second facial dataset for this investigation from the web based on celebrities, because of the ease of collecting these images. The criteria used in choosing the images depended on unconstrained facial capture, such as different environments (e.g., day and night), a variety of face poses and differently obstructed faces, distance from the camera, the wearing of accessories (e.g., glasses and hats), and different periods (i.e. different ageing). The collection method aimed to include most of the facial recognition challenges determined in the previous chapter. Another reason for using this method was to simulate the forensic requirements that are involved when examining real case

evidence, so most photographs were taken outdoors (e.g., in a street, airport, or garden) and people were not looking at the camera.

Initially, 4,001 images were gathered manually from 100 subjects, with each subject having at least 30 images. In addition, 100 frontal images (one image per subject) were collected and used for reference (enrolment dataset) while all the other images (4,001) were used as test images (testing dataset). Most of the images only contained a single face. Figure 5-2 illustrates samples for two of the celebrities used in this study as a testing set.

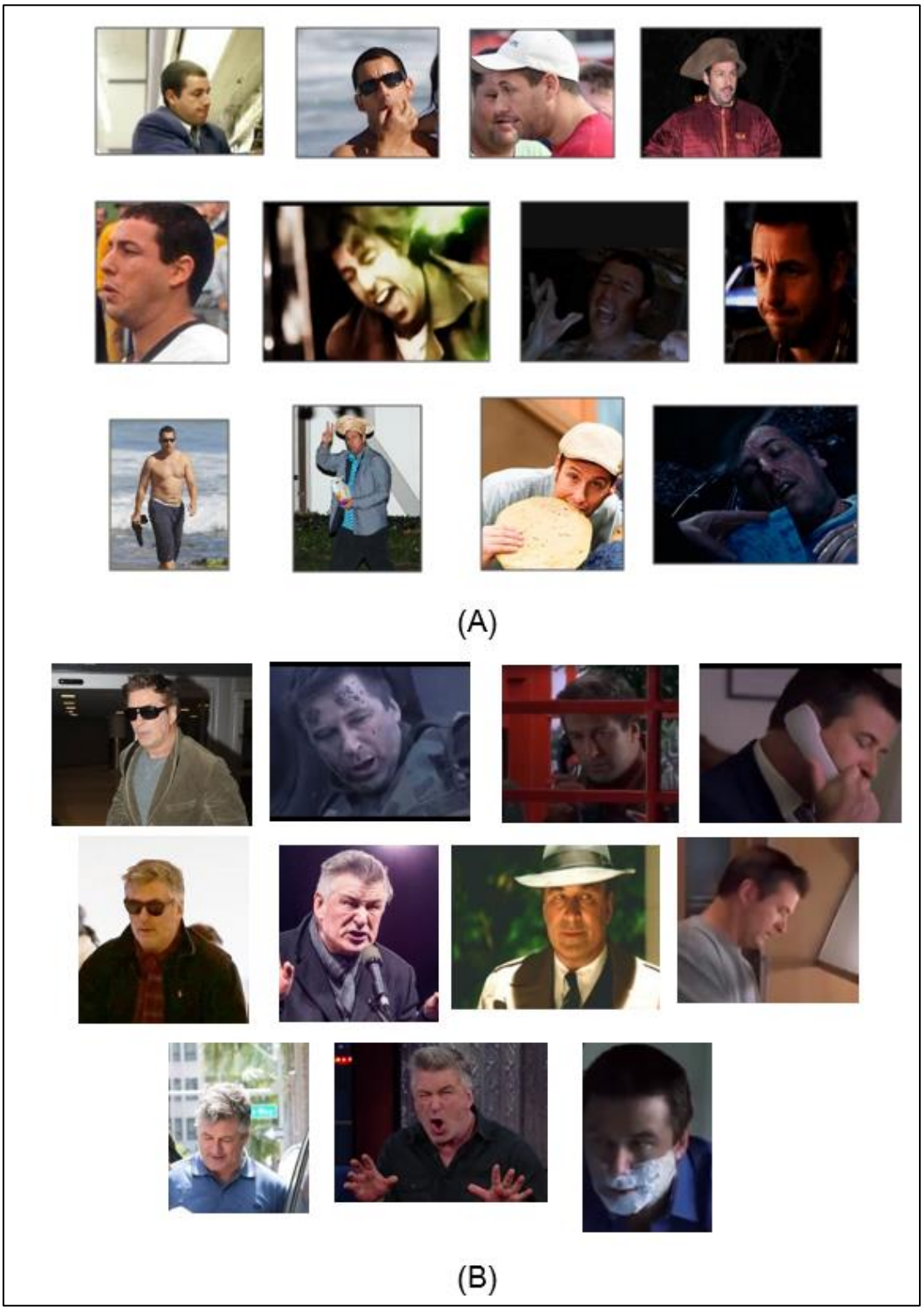


Figure 5-2: Samples of the Realistic dataset used in the test set. A) Photo samples for Adam Sandler and B) Photo samples of Alec Baldwin

The samples illustrate the nature of the imagery and show the reality of the variety and complexity of the issues present in the photographs, such as poor lighting, low resolution, variation in facial expression and pose, the different accessories

worn, and images showing different ages. The issues included in the dataset make it a useful facial dataset that meets researchers' needs in the field of criminal investigation in the future.

Figure 5-3 shows samples of frontal pose face photographs used in the enrolment set. The photographs were chosen because they show no expression and are well lit; the features should be clear to help the matching process.

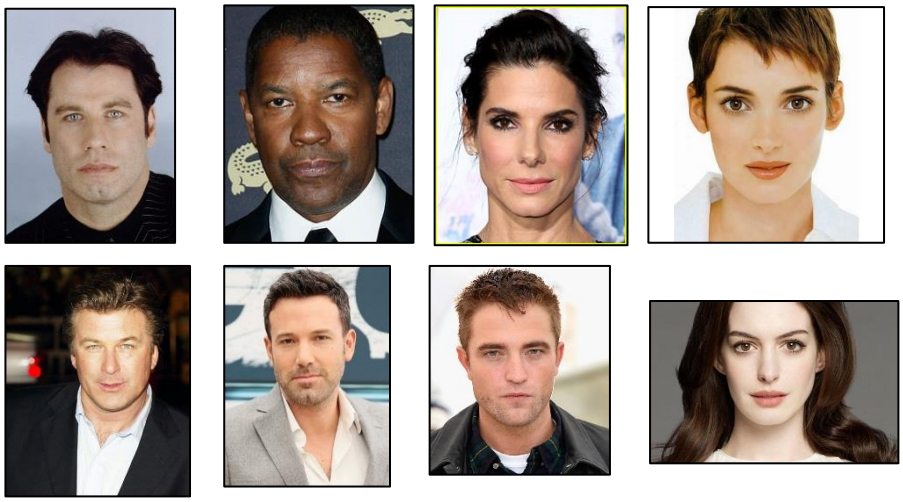


Figure 5-3: Samples of the enrolment dataset collected for the study

The collection process took two weeks, including the time taken to filter the images and check if there were any errors in the files or repetition. After that, the photographs were indexed, so they could be easily checked when using them in the experiments. For example, a number from 1 to 100 was used to identify each subject and all the photographs were numbered as belonging to a particular subject, from 1 to N (i.e., subject 1's photographs were labelled 1_00001, 1_00002, 1_00003, etc.).

5.2.3 Experimental Methodology

A set of experiments was conducted to investigate how to achieve solutions to the three previous questions identified in Section 5.1 and are illustrated below:

- **Experiment 1 – An evaluation of the three chosen algorithms (control experiment):** This provides a basis for understanding how well the three commercial facial recognition systems perform against a standardised facial dataset from publicly available facial datasets.
- **Experiment 2 – Using a realistic dataset:** A replication of the previous experiment to study the accuracy of the previous algorithms by using a facial dataset with greater variance.
- **Experiment 3 – Multi-algorithmic fusion approach:** This experiment seeks to develop a model for use in multi-algorithmic fusion that aims to enhance the previous experimental results.

The results from the first experiment can then be directly compared with those of the second experiment, which focuses on using more forensically realistic images—where numerous facial recognition challenges are likely to co-exist simultaneously. In addition, an analysis of facial recognition research identified several different routes for facial matchers, with different algorithms focused on different aspects of facial image issues. It was this analysis that gave rise to the question of whether a multi-algorithmic fusion approach to facial identification might improve the results (i.e., the strengths of one algorithm overcoming the weaknesses of the others) and, thus, the third experiment was conducted.

The first and second experimental investigations were conducted to determine the performance and nature of the three contributing facial recognition algorithms against certain facial recognition issues but with two different facial datasets. The experimental study in the third investigation was intended to fuse the individual algorithm results. Therefore, the following methodology description is divided into two parts: baseline evaluation and the muti-algorithmic experiment.

- **Baseline Evaluation of Facial Recognition Algorithms Using Two Facial Datasets (Experiments 1 and 2)**

The three baseline facial recognition algorithms used in this investigation are Neurotechnology, Microsoft, and Amazon Rekognition, as explained in Section 5.2.1. These algorithms can be contrasted in terms of their facial matching accuracy. Therefore, there is a fundamental need to understand their individual matching accuracy.

The first software examined was Neurotechnology and this experiment used the tutorials in VeriLook SDK 10.0. The numbers of the script coding in C# were modified so they would be suitable when implementing the software. The main processes required in this software are illustrated in the flowchart in Figure 5-4.

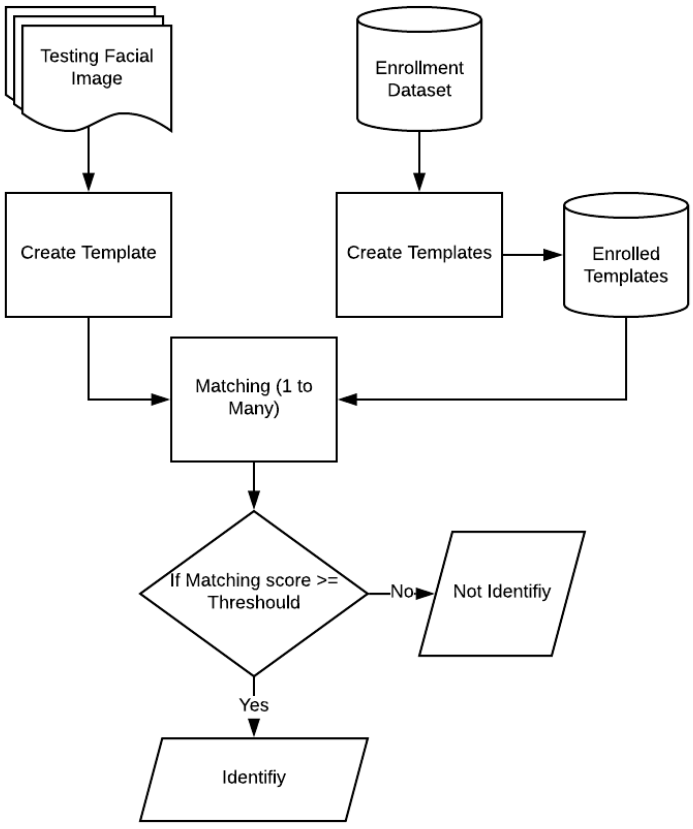


Figure 5-4: Flowchart of the Neurotechnology facial recognition algorithm

As shown in the above flowchart, the matching process requires the face template for matching. The created template process detects faces and extracts facial feature details, such as eyes, nose, and mouth attributes, based on the Neurotechnology feature-extraction algorithm. Finally, the generated template is compressed and stored as a binary file.

The matching process in Neurotechnology also uses the stored templates to identify faces. The test template is compared against all the stored enrolled templates to check whether it belongs to one of them. The result of each comparison is the similarity score and a higher score represents a higher probability that the features belong to the same person. After that, the score is mapped to the checking process (a yes/no answer) by comparing it with the matching threshold. The matching threshold is the minimum score that the identification function will accept to assume the faces being compared are similar. The results give an N-rank result with a score and N depending on the number of subjects.

The second facial recognition software comes from Microsoft. In this investigation, the Windows Presentation Framework (WPF) application was used through the C# environment. As this software is based on a cloud API, a subscription key is required to run any sample. In this experiment, a one-year free trial key from Microsoft was used along with a FaceAPI algorithm. The Microsoft FaceAPI algorithm includes face detection, facial attribute extraction, facial recognition, and face grouping. The main process of the FaceAPI is shown in Figure 5-5.

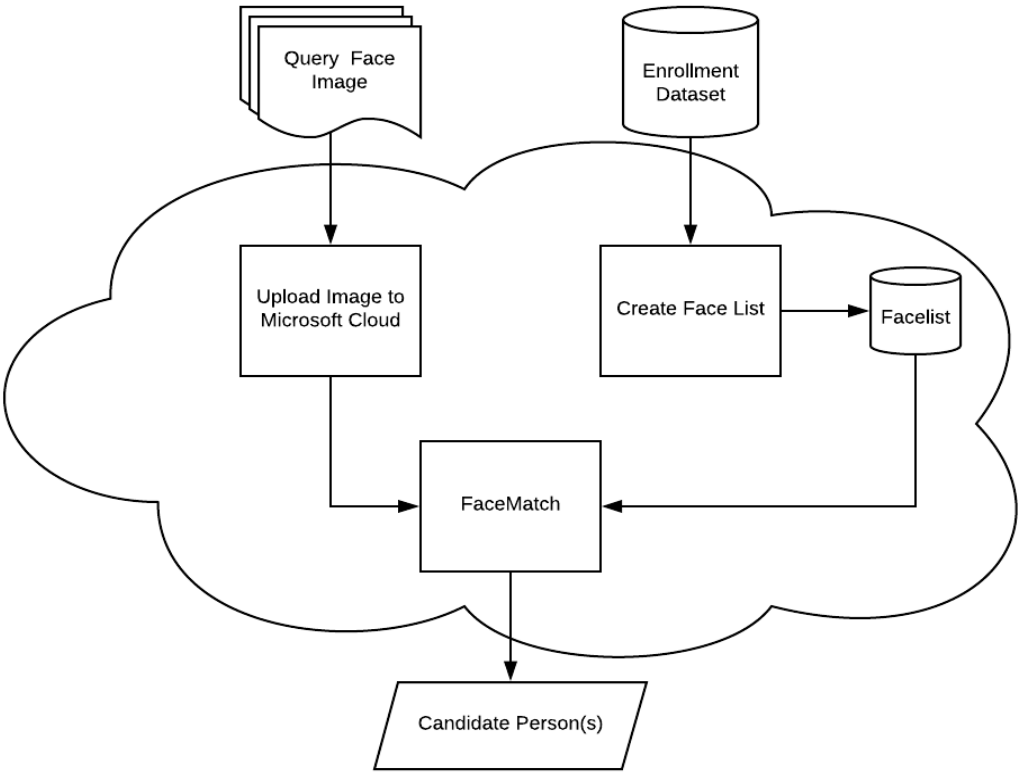


Figure 5-5: Flowchart of the Microsoft Face API process

As shown in Figure 5-5, all the images for testing and enrolment will be uploaded to the Microsoft cloud. In the cloud, a face list is created from the enrolment images that will include all the faces detected from the uploaded images, their attributes, and FaceId. The list is saved in the server until deleted by the user. Microsoft allows a Facelist of up to 1,000 faces to be created.

After creating a Facelist, all the faces are imported from the Facelist for use in a matching or 'find similar' process. One of FaceAPI operations is called Face Identify, which is a match based on (one-to-many) identification methods to find the closest matches to query face images from the Facelist set. Face Identify computes similarities by using an internal threshold and returns a similar ranked face with a confidence value for each match. The confidence value represents the similarity between faces in numerical terms.

The third software used in the research is Amazon Rekognition, which is a cloud-based API, the same as the Microsoft software employed. The Amazon Rekognition API operates by using an Amazon Web Services (AWS) SDK and supports different language environments. Python was chosen as the environment in which to run the Amazon API. Amazon provides an easy way to use the API that does not require expertise in computer vision or machine learning and was released to build complex facial recognition algorithms. The API just needs to be provided with an image and the available services can then detect, analyse, and recognise faces.

Amazon Rekognition provides two types of API operation: non-storage and storage. In non-storage operations, Amazon Rekognition does not keep any information about the input image (no input image bytes persist in the Amazon cloud). Hence, all images are stored in the user repository and imported to Amazon operations when called. Alternatively, in storage API operations, Amazon creates a face collection in one of the user's AWS regions and stores facial feature information for all the input images. It is worth noting that the service does not store actual image bytes but stores extracts of facial features for facial detection in images then saves them in a vector in the facial collection. Amazon then uses features vectors when performing face matching. In this experiment, the storage operations were used because they support the facial identification aspect and can be compared with face collection more than once until deleted by the user. Figure 5-6 summarises the Amazon Rekognition process for its facial recognition service.

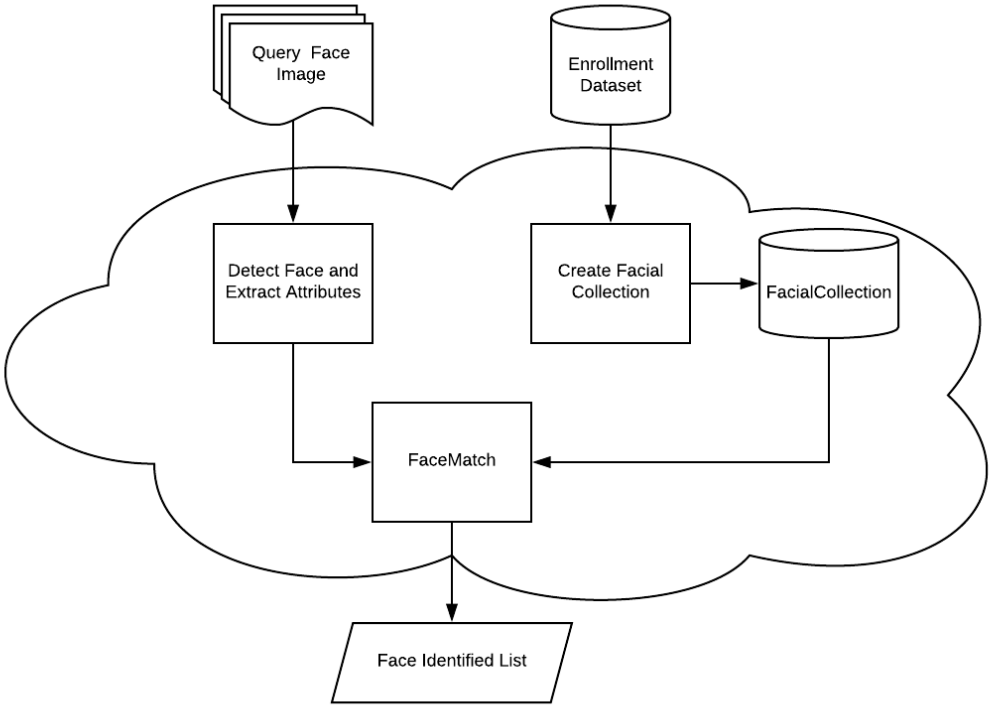


Figure 5-6: Flowchart of the Amazon Rekognition API process

The service returns a confidence score for every result it identifies based on a similarity threshold. The threshold value controls how many results are returned based on the similarity between faces. Therefore, when using a high threshold value, the degree of misidentification will be reduced. If the aim of identification is to return more results, it is better to use a low threshold value, which would be more suitable for forensic investigations. Hence, Amazon's suggestion for a low threshold is 80%.

The CAS-PEAL-R1 dataset was used in the first experiment and the realistic dataset was in the second setup. Each of the two facial datasets was divided into two sets: an enrolment set and a testing set. The two sets are defined and described below:

1. Enrolment set: A collection of known individuals' images used as a source to check identity in the matching process. In the evaluation experiments 1 and 2, this set contained one image per subject under normal conditions.

This assumption is one of the most challenging hypotheses in facial recognition technology because whenever there are more source images per subject, the probability of identifying an individual will be higher. Nevertheless, this assumption was adopted in these experiments to suppose the hardest available solutions in criminal investigations.

2. Testing set: This is a collection of images of unknown individuals that need to be recognised. Therefore, all the sets of the CAS-PEAL-R1 dataset, except for the normal set, and all the Realistic datasets, also except for the normal set, were considered as part of the testing set.

Table 5-5 shows the subdivision details for the CAS-PEAL-R1 facial dataset used in the first experiment.

Subset	#Subjects	Enrolment set		Testing set	
		#Samples	#Images	#Samples	#Images
Accessory	95	1	95	6	569
Expression	95	1	95	5	475
Lighting	95	1	95	9+	1,203
Pose	95	1	95	20	1,900

Table 5-5: Splitting of the CAS-PEAL-R1 dataset

Table 5-6 shows the splitting of the Realistic facial dataset used in the second experiment.

Subset	#Subjects	Enrolment set		Testing set	
		#Samples	#Images	#Samples	#Images
	100	1	100	30+	4,001

Table 5-6: Division of the Realistic facial dataset

- Multi-Algorithmic Fusion Approach (Experiment 3)

In light of potential limitations in the individual facial matcher algorithms in the previous two experiments, the third investigation assessed the use of a fusion

technique to improve the overall performance of the selected matcher systems. Prior biometric research has shown that multi-algorithmic fusion has resulted in improved performance; thus, it seemed prudent to explore this aspect (!!! INVALID CITATION !!! (Mitra et al., 2016, Ross and Jain, 2003)). The use of a multi-algorithmic fusion technique enables an identification system to depend on more than one identification algorithm decision. It also increases overall system reliability because of the existence of multiple identification results for the same input. Consequently, a multi-algorithmic fusion approach was adopted in this experiment by leveraging the knowledge of the three identification systems: Neurotechnology, Microsoft, and Amazon Rekognition.

As referred to in Section 2.4 (in Chapter 2), a different level of fusion method could occur at any process step within a multi-biometric system, such as the sensor level, feature level, matching score level, and decision level (Ross et al., 2006). In this experiment, the decision-level fusion approach was adopted so that it could be applied to three Commercial Off-The-Shelf (COTS) facial matcher results (Neurotechnology, Microsoft, and Amazon). It was considered the most suitable approach because most current commercial biometric systems (i.e., facial recognition) provide access only to the final matching results. In addition, these systems sometimes provide limited access to the feature sets or classifier algorithms used (Ross et al., 2008). The decision-level fusion approach can also be used for a number of different facial matchers without the need to train the system or determine the best features to use and/or modify. The approach is also considered more suitable for digital forensic investigation because it will be scalable and flexible so new robust facial recognition technology can be added in the future. Figure 5-7 shows the main process of the third experiment.

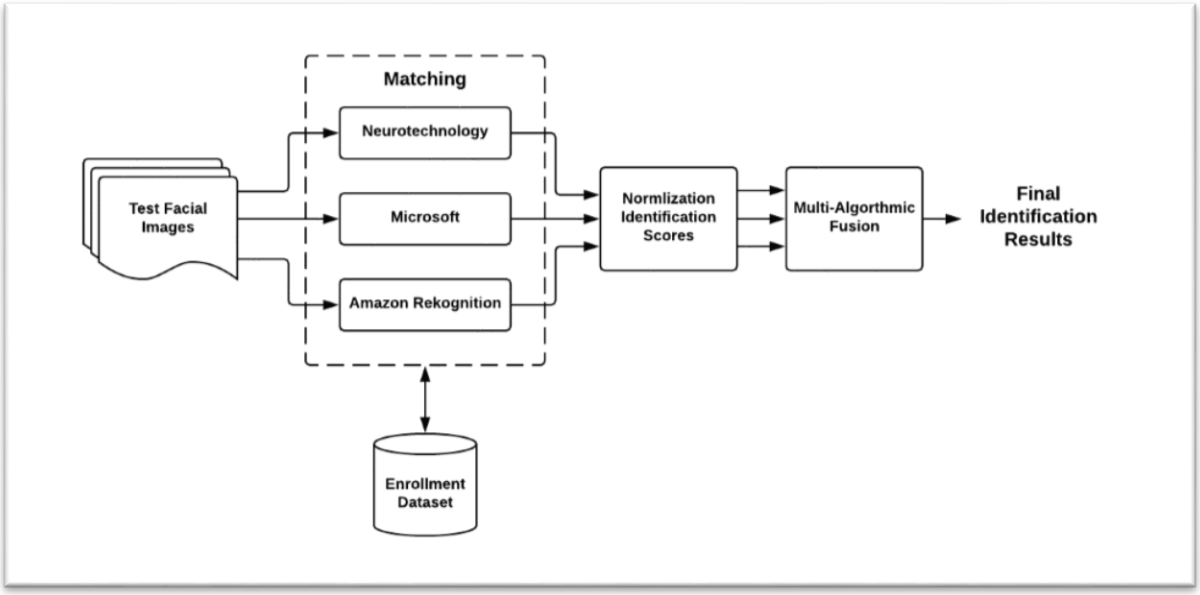


Figure 5-7: Multi-algorithmic fusion approach flow diagram

Before the fusion process, the various score ranges that were obtained from the different matching algorithms (in the previous two experiments) were normalised. The scores were mapped from multiple domains into the public domain. Normalisation is a crucial characteristic of any multi-algorithmic fusion approach. One of the normalisation techniques available is Min-Max, which is considered more suitable when the score bounds produced by the matcher systems are known (Jain et al., 2005). Therefore, all the scores of the three systems were transformed into a common range (0 to 1) so the minimum and maximum scores would range between 0 and 1. The normalisation equation is given as:

$$S\sim = \frac{S_K - min}{max - min}$$

Where:

S_K is the matching score; where K = 1…N (N is the number of samples).

$S\sim$ is the normalised score.

Min and Max are the bounds of the scores.

The decision-level fusion approach has different methods that could be used in this experiment, such as "AND" and "OR" rules (Daugman, 2000), majority voting (Lam and Suen, 1997), and weighted majority voting (Kuncheva, 2004). Facial matching is an identification technique that requires a return for any similar photographs, so this research adopted weighted majority voting in the fusion decision. Therefore, depending on the rank-1 results each system produced, the decision-level fusion approach demonstrated the best result by adding weight to each score level then using majority voting to choose the highest score to make the final decision.

When the three facial recognition systems provided an identity for any testing sample, it was reasonable to consider the more accurate facial recognition system by giving it a higher weight (a weighted majority voting function). The weight value is gained by studying the accuracy or identification rate (IR) of each system in the previous baseline experiments. After that, the IR value is normalised for all systems to be in the same domain. Then all the scores in each system are multiplied by the system weight to produce a new score value. The discriminant equation for a new score computed using weighted voting is:

$$SW_K = W_R S_K$$

Where:

SW_K is the new score after weighting.

W_R is the weight of the R^{th} facial recognition system.

S_K is the matching score where K =1…N (N is the number of samples).

The final decision will take the highest score value among the three identification scores by using majority voting, which leads to the greatest matcher accuracy.

In this investigation, the proposed fusion technique was used to compare the performance between the fusion results and those achieved in experiments 1 and 2.

5.3 Experimental Results

Three experiments were conducted with the aim of studying the performance of a number of facial recognition systems when facing facial image challenges. The evaluation used in the experiments was based upon calculating the IR for rank-1 (one-to-many matching) of the results. The IR is computed as follows:

$$IR = \frac{Correctly\ Matching\ Facial\ Images}{Total\ Testing\ Images\ Number} * 100$$

There were two reasons for only adopting rank-1 in the evaluation. First, the accuracy of the commercial systems mostly gives the best matcher in the first rank. Moreover, they do not return a matching value for N samples in a facial recognition approach, although they can do this when using a verification approach. Second, the adopted setup of the facial dataset (one sample per user in the enrolment set) meant that there would only be one correct matcher (i.e., a complex division of the dataset in the facial recognition technique).

Furthermore, the false positive identification rate (FPIR) and the failure to acquire (FTA) rate were implemented for further data analysis. The FPIR is the ratio of test samples that are classified as true when they are actually false, whereas the FTA represents the rate of failure to create face templates in the testing dataset.

5.3.1 Results of Experiment 1

To evaluate and test the efficiency of the three facial identification systems, each image in the testing set was compared with all the images in the enrolment set in the CAS-PEAL-R1 dataset. The facial dataset was split into an enrolment set and

a testing set, as shown in Table 5-5. The accuracy of the systems was then compared and the strengths and weaknesses of each system identified.

The facial recognition accuracy result for each system is illustrated in Table 5-7. The table shows that Microsoft and Amazon identifiers achieved the same IR for rank-1 in the Accessories category (over 98%). Their Neurotechnology counterpart gained a slightly lower result (by 6%). As can be seen from the results, the IRs for the Expression set for the three systems are all over 98%. The high accuracy of the two sets (Accessories and Expression) reflects the good performance of the three systems in facial recognition in these conditions.

Regarding the Lighting and Pose conditions, the overall performance of the three algorithms decreased. However, Microsoft outperformed the other two algorithms in the Lighting dataset and achieved 86.69%. The Neurotechnology system achieved a low accuracy rate in the Lighting set (63%), 20% less than the accuracy for Microsoft and Amazon Rekognition.

For the final testing set, which is the Pose set, it should be noted that Neurotechnology performance dropped significantly to 31.31%. This could have been caused by the large face pose angle problem in some samples and the limitation of these systems in managing it. The highest facial matching accuracy in the Pose set was achieved by the Amazon system (slightly more than 85%).

Subset	IR at Rank-1 (%)		
	Neurotechnology	Microsoft	Amazon Rekognition
Accessories	92.61	98.76	98.76
Expression	98.31	99.57	99.78
Lighting	63.42	86.69	83.95
Pose	31.31	74.47	85.73

Table 5-7: Experiment 1 results (using CAS-PEAL-R1 dataset)

In order to conduct further analysis, Table 5-8 demonstrates that the FPIR and FTA rates of all the systems for the Lighting and Pose conditions are significantly higher than those presented in the Accessories and Expression conditions. Incorrect matching rates (i.e., FPIR) of the three systems in all face conditions are less than the FTA rates. Some images in the testing subset could have caused the high FTA rates, as they were acquired with low resolution, darkness, poor exposure, and an angled pose. This means the number of templates implemented in the matching process was fewer than the total input images; hence, the high failure in identification rates. For example, the Neurotechnology system achieved the highest FTA rates: 66.1% for Pose and 25.27% for Lighting. This suggests only 34% for Pose and 75% for Lighting face templates were implemented in the matching processing in this system. This weakness in creating templates from images was as a result of a large pose degree and a high rate of darkness and there was just one sample facial image in the enrolment set (a front pose of good quality).

Subset	#Images	Neurotechnology		Microsoft		Amazon Rekognition	
		FPIR (%)	FTA (%)	FPIR (%)	FTA (%)	FPIR (%)	FTA (%)
Accessories	569	0.52	6.85	0.52	0.70	1.05	0.17
Expression	475	0.21	1.47	0	0.42	0	0.21
Lighting	1,203	11.13	25.27	2.66	10.64	5.90	10.14
Pose	1,900	2.05	66.10	0.78	24.73	4.26	10

Table 5-8: FPIR and FTA rates for Experiment 1

Overall, the results of this experiment demonstrate the facial identification ability of the three systems for four facial image issues: accessories, expression, lighting, and pose. It is clear the accuracy of all the commercial systems examined suffers when the darkness of the image is increased, as well as changes in facial orientation. Furthermore, Microsoft's performance was better for the lighting issue than the facial pose issue, whereas Amazon was better in terms of facial pose than the lighting issue.

5.3.2 Results of Experiment 2

The aim of this investigation was to study how the performance of the previous algorithms would change when using a dataset from real-world use. The results of the previous experiment were simulated with a controlled facial dataset, which arguably did not resemble the situation when using images in samples close to current real images of suspects. Moreover, the performance of the three algorithms would differ from Experiment 1 because it dealt with all the facial recognition issues in one group set (not spilt into sets), or it could even contain more than one issue in one sample image. Furthermore, the realistic dataset reflects issues in the forensic investigation as shown in Section 5.2.2, so it represented a good challenge for three algorithms. Therefore, it was felt that it would be useful to conduct another evaluation to enable more investigation in this study. The realistic dataset collected and used in this experiment was split into

two sets: an enrolment set (one image per user) and a testing set (more than 30 images per user with varied issues), as described in Table 5-6.

As demonstrated in Table 5-9, with an IR at rank 1 of 67.23%, the Microsoft facial identification system achieved the best performance against the Realistic dataset while the Neurotechnology system obtained the lowest performance with just 6.6%. In comparison with the results obtained from the first experiment, this experiment showed that the performance of all three systems dropped significantly because of the complexity of realistic facial images in the Realistic dataset, highlighting the challenge that a digital forensics investigator has to face when dealing with real-life scenarios.

Testing Set	IR at rank-1 (%)		
	Neurotechnology	Microsoft	Amazon Rekognition
100	6.60	67.23	48.24

Table 5-9: Experiment 2 results (performance of the three commercial systems with the Realistic dataset)

To understand the reason for the results regarding the accuracy of the systems, Table 5-10 shows more evaluation by presenting the incorrect matching (FPIR) and FTA rates for the three algorithms. The highest FPIR of 11.47% was obtained by the Amazon system, whereas the highest FTA rate (87.35%) was achieved by the Neurotechnology system. Therefore, it is clear the Neurotechnology system failed to create templates for most of the testing photos and this means there is some limitation in its algorithm to manage photo issues such as resolution, illumination, and orientation. Whilst, Microsoft and Amazon's ability to create templates from the testing photos were close to each other.

Testing Set	Neurotechnology		Microsoft		Amazon Rekognition	
	FPIR (%)	FTA (%)	FPIR (%)	FTA (%)	FPIR (%)	FTA (%)
100	6.04	87.35	6.37	36.39	11.47	40.28

Table 5-10: FPIR and FTA rates for experiment 2 for the Realistic dataset

This experiment shows most photos that failed to be acquired by the three systems suffered from big issues such as illumination, quality, distance, and orientation. Further, more than one issue may appear in a photo. For example, Figure 5-8 shows some photo samples from the testing set where the three systems failed to create templates. In comparison with the incorrect matching (FPIR), there were some testing photo samples that represented a challenge to find correct matches. Figure 5-9 shows some examples of the difference between the enrollment and testing photos that matched incorrectly and it is clear that this issue was caused by accessories, face orientation, ageing, and expression. These problems simulate some current issues in forensic investigation discussed in Chapter 1, such as the Boston Marathon bombing and Belgium's Airport attack, and reflect the complexity of the realistic dataset used in this experiment.

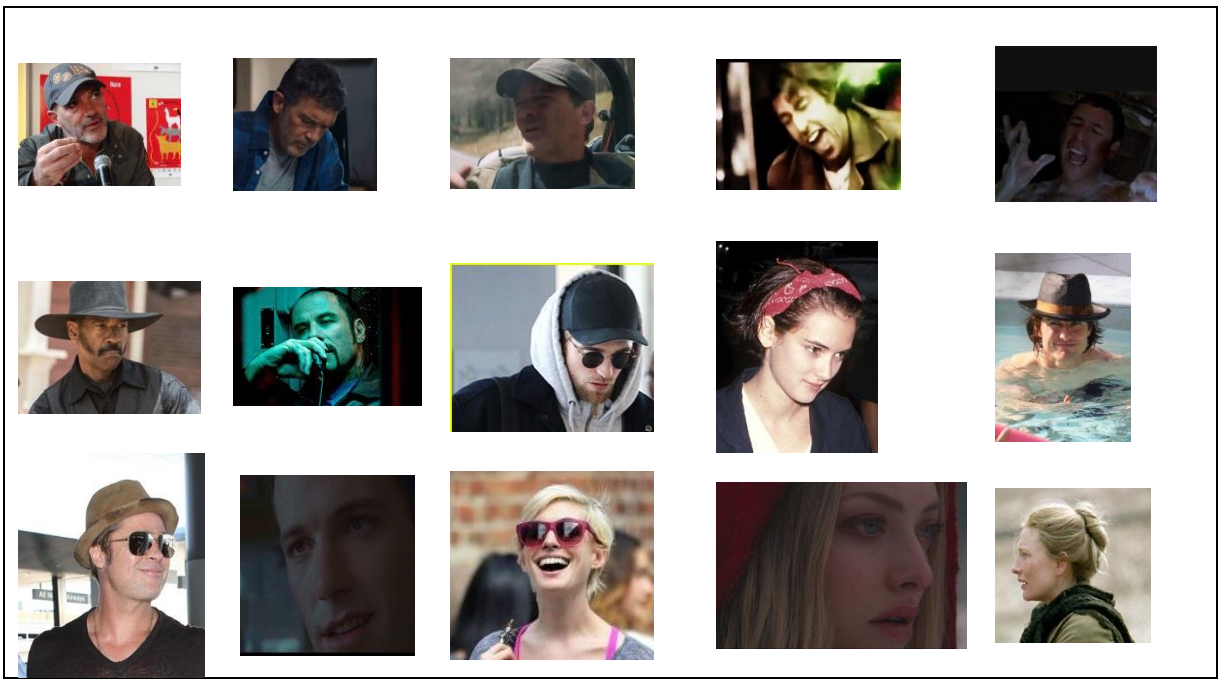


Figure 5-8: Some of testing photos that failed in acquired templates

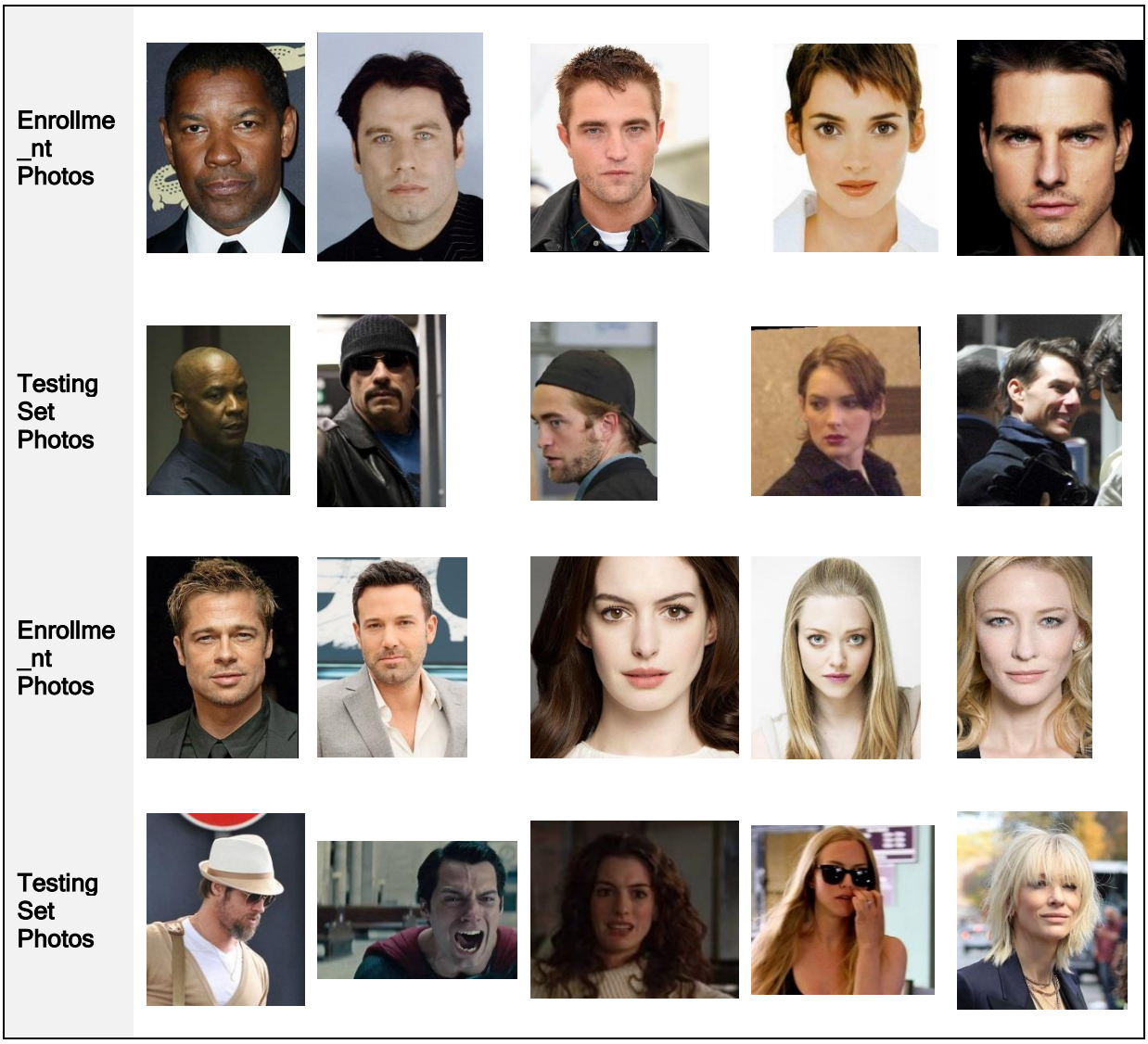


Figure 5-9: Some of testing photos that matched incorrectly

Overall, the FTA rates obtained in this experiment are significantly higher than those presented for the first experiment. As mentioned previously, an increased FTA rate will decrease the number of face templates that can be sent to the next stage (i.e., the matching process). This higher rate in FTA is as a result of the complexity of photos that add more challenges for the three algorithms to discover their features and read them as described.

Indeed, the overall results of this set of experiments demonstrate how the photographs from the realistic dataset affected the performance of three top commercial facial recognition algorithms and show the complexity of the problem this study is trying to solve.

5.3.3 Results of Experiment 3

As mentioned earlier, the facial identification accuracy for the above three systems varied when they were compared. Moreover, sample results in more than one system were assigned the same identity. Therefore, to enhance the overall performance, the multi-algorithmic fusion system was applied using the same dataset setting as the previous experiments. This investigation anticipated that the identification accuracy obtained by the fusion system would improve in comparison with the results of the first and second experiments.

In this experiment, the more accurate system among three identification systems is found by choosing the highest weight (a weighted majority voting function). This first involves multiplying all the scores in each system by the system weight to produce a new score value that is used in the voting. Table 5-11 Illustrates the weight value for each system for two datasets (the calculation method described in Section 5.2.3 – point 2). The final decision will take the highest score value among the three identification scores by using majority voting, which leads to the greatest matcher accuracy.

Subset	Weight Value		
	Neurotechnology	Microsoft	Amazon Rekognition
The CAS-PEAL-R1 dataset			
Accessories	0.9261	0.9876	0.9876
Expression	0.9831	0.9957	0.9978
Lighting	0.6342	0.8669	0.8395
Pose	0.3131	0.7447	0.8573
The Realistic dataset			
Testing set	0.065	0.672	0.482

Table 5-11: The weight values for three systems for two datasets

As illustrated in Figure 5-10, when applying the proposed fusion technique to the results of Experiment 1, which used the CAS-PEAL-R1 dataset, the improved performance can be observed for all four chosen sets. The improvement rates vary among the four sets: Accessories, Expression, Lighting, and Pose. The high identification accuracy that had previously been achieved in the Accessories set by Microsoft and Amazon led to a low improvement rate when using the fusion technique (0.71%). A higher degree of enhancement in terms of Neurotechnology accuracy (6.86%) was observed. Similarly, in the Expression set, enhancement from using the fusion method was low by reason of the high accuracy in the single-algorithm approach for the three systems.

In particular, the identification accuracy for Lighting improved from 63.42% when using the Neurotechnology facial identification system alone to 90.44% when using the fusion method (i.e., an improvement of 27.02%). This enhancement is considered the greatest compared with the improvement in the Microsoft (3.75%) and Amazon (6.49%) systems.

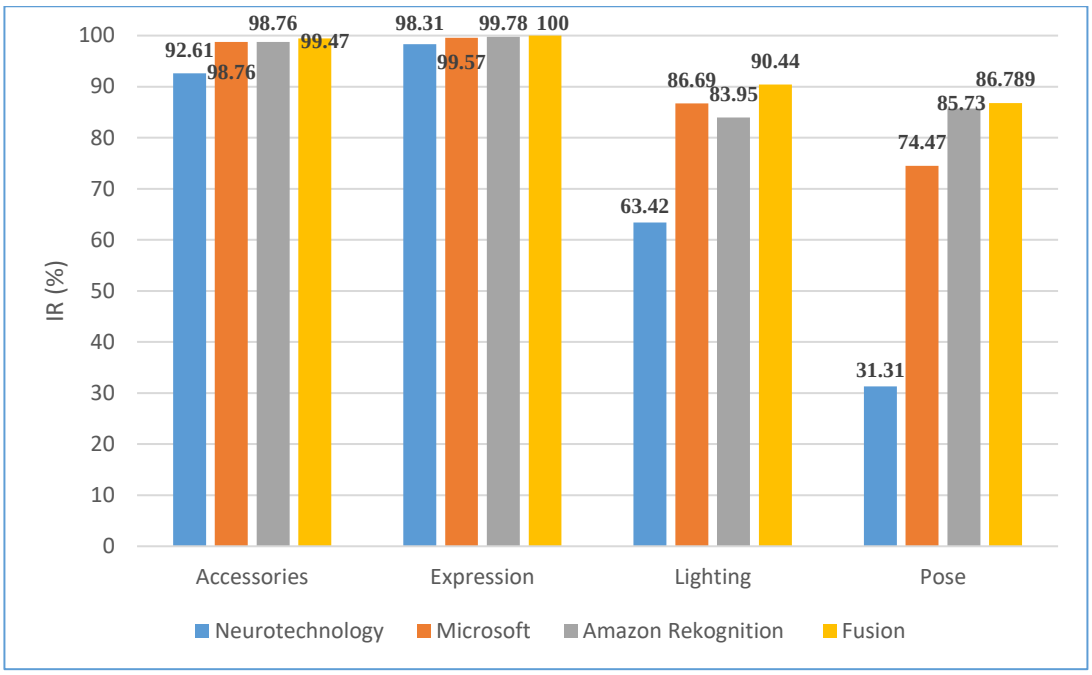


Figure 5-10: Performance comparison between the fusion method and the other systems when using the CAS-PEAL-R1 dataset

In the facial Pose set, the results clearly show the use of the fusion method enhanced performance significantly for the Neurotechnology and Microsoft systems, which were 55.47% and 12.32%, respectively, while the enhancement for the Amazon system was 1.05%. This rate was low compared to the two other systems because Amazon Rekognition had already achieved the highest accuracy in the first experiment.

The results above confirm the proposed fusion technique (which used majority and weighted majority voting) outperformed the three individual facial identification systems in Experiment 1.

Figure 5-11 shows a comparison of the results for the proposed fusion method and the three chosen facial identification systems using the realistic facial dataset. Again, the proposed fusion system achieved the highest performance of 71.6% of IR; this result is over 4% better than the best individual system (i.e., Microsoft) obtained. The considerable variation in the accuracy of the individual algorithms in this study (i.e., the difference between the accuracy of

Neurotechnology compared with Microsoft and Amazon) influenced the outcome of the accuracy of the decision-level fusion technique.

As a result, this experiment shows that improved performance can be obtained using the proposed fusion approach, particularly for the unconstrained dataset. This highlights the potential impact the proposed fusion system could have on the forensic investigation field.

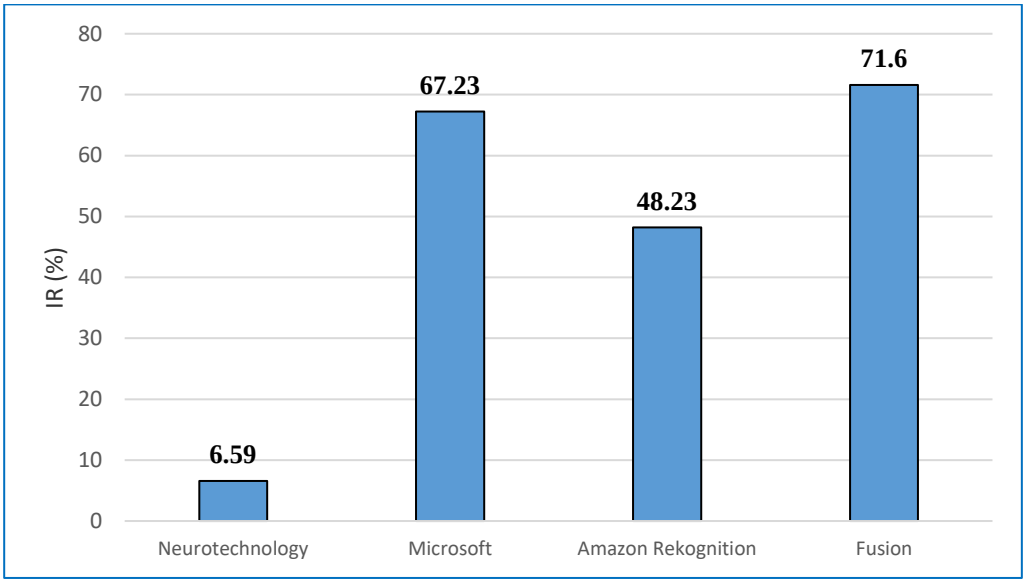


Figure 5-11: Performance comparison between the fusion method and the other systems when using the Realistic dataset

5.4 Discussion

The observations from the first experiment show there is contrast in the performance of the three commercial facial recognition systems (Neurotechnology, Microsoft, and Amazon Rekognition) when using the CAS-PEAL-R1 dataset. In this experiment, the results indicate the accuracy of these systems in matching a front face position with hat, glasses, and some expression. The reason for this is that all face samples were tested with a front pose with good resolution. However, the three systems managed lower facial matching rates with regard to lighting and the lowest for facial pose. The main reason for the drop in system performance was the failure to acquire templates from the input face

images because of poor image quality, low lighting, and a high pose angle. As template generation is considered the primary step in facial recognition systems, this had a negative impact on the number of acceptance templates that could be sent to the matching process. Overall, it can be concluded from this experiment that the Microsoft system performed better than the other systems for lighting, whereas Amazon achieved significant performance in pose challenges. All three systems showed the same high accuracy in the expression and accessories issues. The results give a clear impression of each system when faced with four facial recognition issues.

The second experiment can be used to conclude there was a significant drop in facial matching accuracy for the three systems when compared with the results for Experiment 1. This drop is by reason of the nature of the dataset of the images that were collected for this experiment, which simulated the real world (realistic dataset) to reflect the images of criminal cases. Moreover, the FPIR slightly increased in this dataset when compared with CAS-PEAL-R1, whereas there were significantly increased FTA rates for all systems when compared with the first experiment. This difference between two experiments' results was because of the complexity of the Realistic dataset that simulated the current facial recognition issues in the forensic investigation, such as poor quality, orientation, bad illumination, and accessories. This substantiates the notion that unconstrained facial imagery is a significant challenge in comparison with standard datasets and that commercial systems still struggle to achieve reliable performance.

As in previous experiments, the performance of the systems varied and was not stable for each system under different image issues. Therefore, the hypotheses of the multi-algorithm fusion approach support better performance of the holistic

system. Therefore, the multi-algorithmic fusion approach should be executed in a constructive rather than destructive way. The fusion method improved all the results for the control experiment (experiment 1) by 100% for the Expression issue. In addition, this approach improved the facial matching rate at rank-1 from about 67% for the Microsoft system (which showed the best performance) to 71% for the realistic dataset. The last enhancement is considered an essential effect in the forensic field. The fusion method is needed to propose a more robust technique than the single algorithm and it fuses the good performance for each system in one approach.

This study supports the proposition that a multi-algorithmic approach in forensic facial recognition would lead to improvement of the final accuracy rather than an individual algorithm. Although the use of commercial systems has several advantages, most notably, a degree of specialisation that should see performance rates maximised, there is the issue of privacy. For example, Microsoft uses cloud services in its recognition process; however, in doing so, it saves copies of the submitted images so subsequent algorithmic improvements can be made. From a forensic data privacy perspective, this would be a significant barrier to adoption. It is notable that other systems, such as Amazon and Neurotechnology, do not do this, so it is far from being a standardised approach. In addition, most commercial systems do not provide a feature vector for facial images. Therefore, there is a limitation to doing more research about fusing features or using feature vectors in further analysis.

In addition, through this research, a number of challenges were observed that need further research:

- It is important to explore how to enhance images to increase the rates of generation of face templates to improve identification accuracy, such as

through the enhancement of images by better lighting, quality, and, if possible, correct pose angles by using 3D technology.

- The way in which face grouping could enhance identification system accuracy should be investigated. A face grouping approach divides a facial dataset into several groups based on similar classifications. For example, age and gender (i.e., age estimate and gender determination) are considered two important classification types that need to be investigated.
- The scalability of the dataset should be assessed by increasing the number of users (the number of subjects in the enrolment set). This leads to the question of whether facial identification systems could achieve the same results if data size increased. In addition, study one of the standard face datasets specific to ageing to investigate how the accuracy of the ageing issue.
- The investigation could establish the possibility of evaluating an additional facial identification system or algorithm (i.e., a fourth algorithm) that could have a positive effect on the overall results.
- The ability to use video files instead of static images could also be examined. Further investigation is needed of hypotheses regarding facial identification using a multi-algorithmic fusion approach and whether video would offer the same performance as static images.

5.5 Conclusion

This chapter introduced three experiments to evaluate the performance of three current commercial facial recognition systems (Neurotechnology, Microsoft, and Amazon Rekognition) and how multi-algorithmic fusion could improve accuracy. The multi-algorithmic fusion approach showed high accuracy regarding using facial recognition for two selected datasets: CAS-PEAL-R1 and Realistic. When

using an unconstrained face dataset, the proposed system improved facial identification accuracy when compared with the highest identification accuracy for the commercial systems considered. The results demonstrated the fusion method outperformed the accuracy of the individual identification systems in two types of dataset.

6 A Novel Architecture for Facial-Forensic Recognition

6.1 Introduction

The previous chapter presented a set of experiments that were carried out to improve the accuracy of a facial recognition system by using a multi-algorithmic fusion approach. A few limitations in experiments were also observed in the previous chapter. The prevailing digital forensic investigation experiences intend to incorporate these facial identification challenges, specifically in the analysis of digital multimedia evidence. For example, a child abduction case might require identifying faces (child or suspect) from CCTV but the question posed might be where is the individual going with the child?. Therefore, a correlation of that individual who has been identified from the CCTV streams of the local area and plotted onto a map would be far more useful to an investigator than merely the recognition output. With the aim of overcoming these challenges in the investigation process, the approach does not merely concern the identification of an individual but an understanding of the context, acknowledging the relationships between individuals, and formulating timelines. With the advancement and evolution of digital forensic tools in the facial recognition field, numerous hindrances also exist, must be addressed. Some of those limitations are as follows:

- Manual investigations are still being conducted in practice to investigate specific criminal individuals.
- Facial recognition procedures face major issues in forensic investigation, such as image quality, facial poses and expression, and the effects of lighting and ageing.
- Insufficient ability to analyse photo content in an automated way to extract evidence.

- Lack of capability to answer questions regarding the identification of an individual, appreciation of the relationships between individuals from their facial appearance in photographs, and tracking people on a map.

Therefore, this research seeks to add further analysis of facial recognition results through the ability to answer questions that have aroused the interest of investigators. To achieve the task of addressing and resolving the above-mentioned issues, a novel architecture has been designed and developed which is known as the Facial-Forensic Analysis System (F-FAS). This chapter sheds light on the holistic system architecture and its functions.

6.2 Facial-Forensic Analysis System (F-FAS) Requirements

The objective behind creating the F-FAS is to overcome the aforementioned limitation points (in Section 6.1) relating to current forensic tools in multimedia analysis. F-FAS is an efficient design that analyses the content of photo evidence to identify criminal individuals. Furthermore, it aims to provide deep analysis of the evidence to allow investigators to find answers about different questions, such as individual identification and evidence correlation. To be successful, the F-FAS must meet several essential requirements:

- Acquisition of various data collection from different sources, such as CCTV, mobiles, and computers to create forensic images. As described in Chapter 3, the first part of the digital forensics process is collecting evidence from the crime scene and saving it in a secure manner. Therefore, any proposed system in digital forensics analysis needs this step.
- Ability to pre-process all sources by filtering and indexing photos from multimedia files based on face detection in order to extract the necessary

metadata. The second step in the digital forensics process (Section 3.2.1) starts with analysing the data collected and isolating interesting files. This action will allow investigators to search in depth inside the collected data.

- Identification of suspects by matching their face with existing dataset images. This requirement is considered the base the F-FAS has suggested and it allowed for searching for faces among the millions of images collected. Furthermore, it minimises the effort and time compared to manual searching.
- Provision of a range of forensic analysis techniques, such as facial social network analysis (faces that are associated with others), geolocation, and the ability to adjust a photograph of a suspect's face to assess whether it matches additional evidence. This allows for in-depth analysis of data and enables finding answers to advanced questions to an understanding of the context. This requirement identifies associations between artefacts and presents them in a usable and visual form to draw a wider picture of a crime.
- Data integrity, which aims to ensure there are no unintentional changes to information. The F-FAS must, therefore, ensure the data retrieved are the same as those recorded earlier. As mentioned in Chapter 3, the data integrity is necessary for any system related to criminal investigations. This action increases the integrity of the evidence found.
- The system must be flexible enough to be upgradeable with new technology. For example, the system should be flexible enough to add or update any of the facial recognition technology that can be used in a multi-algorithmic fusion approach.

- System management with a function for managing the overall system process, such as investigators' management, case management, and configuration setting. Authentication and authorisation are also important prerequisites that should be considered in this system.
- Communication, collaboration and knowledge sharing over a public utility that allows the F-FAS to work more efficiently between groups or over networks. It would also allow users to take advantage of the software service without in-depth knowledge of the process required to build this technology. In addition, it avoids repeating the same analyses or searching between users because most actions are shared.

6.3 The F-FAS Architecture

The F-FAS architecture in this research aims to provide tools for investigators, in terms of facial recognition, and to minimise the task of searching a massive amount of data comprising videos and images collected from CCTV cameras, computers, and mobiles. The proposed architecture is a holistic system developed to collect, examine, and analyse multimedia evidence (photo and video) in an effective manner and then produce a reporting document. These objectives could be achieved by using the engines of the main components of the proposed F-FAS framework, as illustrated in Figure 6-1.

The system model depicted in Figure 6-1 is composed of seven major components that represent the main digital forensic methodology levels, from the collection to the reporting stages.

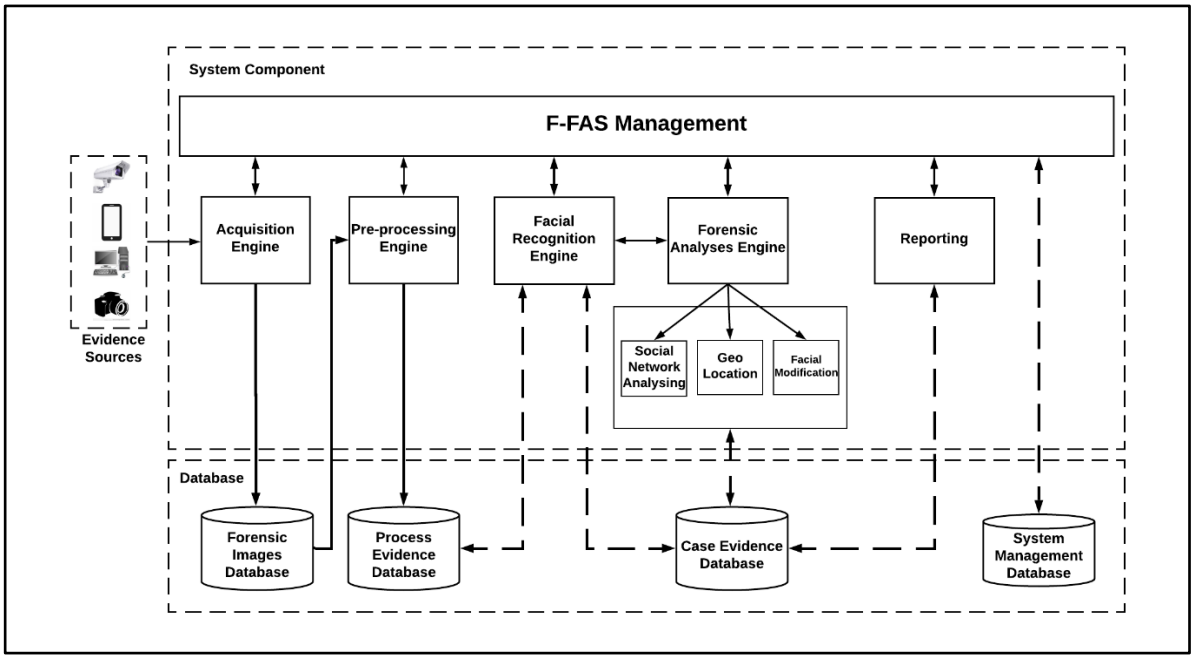


Figure 6-1: A Novel architecture of the F-FAS

The main classes are:

- Acquisition: evidence acquired from different input resources.
- Pre-processing: transforming the evidence acquired into searchable resources, such as identifying videos and photos, indexing, and face detection.
- Facial recognition using the multi-algorithmic fusion: to increase the reliability and accuracy of facial recognition in the facial recognition engine.
- Analysis: enabling quicker answers to queries by using three types of evidence analysis: geolocation, facial modification, and social networks.
- Presentation (reporting and documentation).

Moreover, the above system offers a bookmarking technique that enables investigators to find and refer quickly to important data that have been identified

from the evidence in a case. Bookmarks can be included in the reporting at any stage, whether during investigation or analysis.

The F-FAS Manager has a vital role in the F-FAS architecture, as it is responsible for managing the overall system. Furthermore, the F-FAS Manager provides robust authentication when logging into the system and offers users various authorisation levels that determine the nature of the access to data. Thus, the system manager allows users to achieve different levels of management functionality, such as the ability to create, edit, and delete cases or forensic images. In addition, it manages the system configuration and global setting for each function in the F-FAS and ensures data integrity. It should be noted that the F-FAS is not restricted by the number of facial recognition algorithms in the multi-model approach. Therefore, the system has the ability to adopt a new matching algorithm by offering administrators the scope to add, remove, or update the algorithm. Furthermore, the databases used in the F-FAS architecture are classified into the following: Forensic Images, Process Evidence, Case Evidence and System Management. Each database stores the different types of data relevant to each case.

As discussed earlier, the framework aims to save the time and effort of the investigators involved in a criminal investigation process. Details of each part of the system architecture are provided in the following sections.

6.3.1 Acquisition Engine

The evidence acquisition stage is the first step in solving any criminal case. During this procedure, a digital device, whether related to an incident or used by one of the individuals involved, would, normally, be considered one of the artefacts' suspected resource. Although this device does not play a vital role in the case, it could initially be a foundation for providing a clue or could lead to

further information to facilitate answering several of the questions posed during the investigation. Therefore, to examine digital devices as part of the chain of custody and to protect their integrity, there will be a master copy of the devices, which is essential for their protection. Afterwards, processed data (copies) should be generated for further work throughout the investigation process, rather than using the originals (Marshall, 2009). Figure 6-2 shows the workflow processes of the Acquisition Engine.

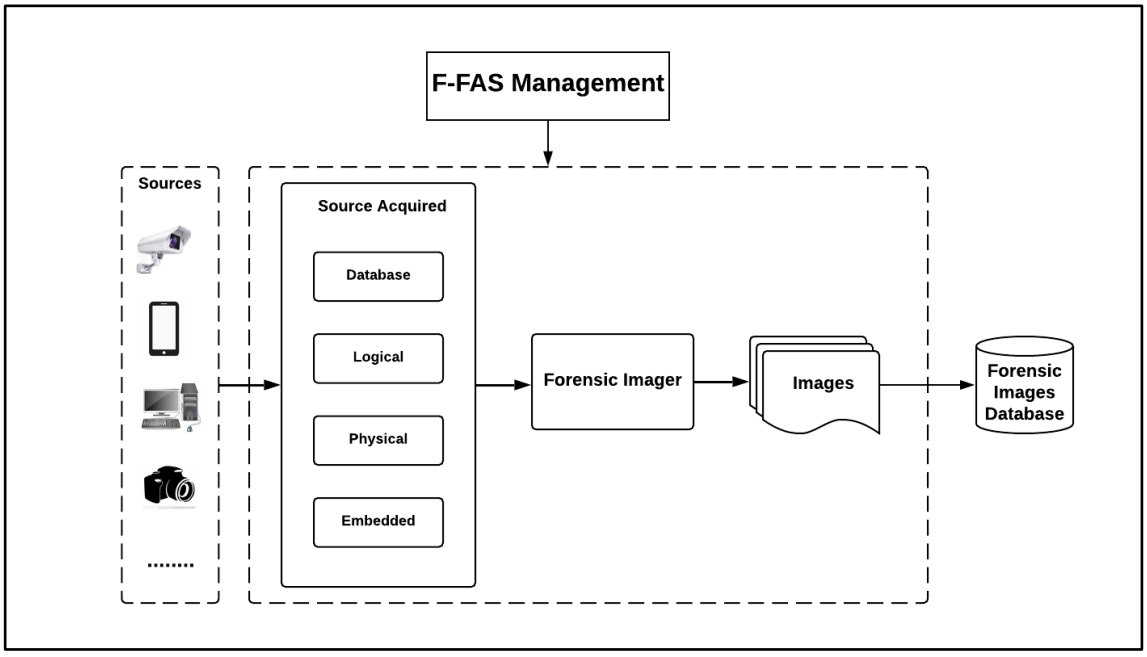


Figure 6-2: Acquisition Engine

The Acquisition Engine enables investigators to collect evidential artifacts from different electronic media sources. For example, CCTV systems are used almost everywhere—inside and outside of buildings—, because they have now become a vital source of evidence in the digital forensics domain. Moreover, with the wide range of technologies available, people are now using various types of devices, such as mobile phones, computers, digital cameras, and tablets. These devices may also contain related devices for storing additional data, such as external hardware, Secure Digital (SD) cards and other types of removable media. Different techniques are, therefore, required to acquire data from these

resources. Interestingly, this engine has the ability to deal with various types of devices and technologies.

First, it is crucial to identify the basic information related to a case in order to gain artifacts. The information gathered, such as case name, incident date, location, number of devices and types, is then fed to the system, which enables the acquisition process to start. Ideally, the data will be acquired in a forensically sound manner by creating a bit-by-bit copy of the data without making any changes or deletions. In addition, to certify the original evidence has not been subject to modification or transferred into unauthorised hands, a hashing function should be used, such as Secure Hash Algorithm 2 (SHA-2) (NIST, 2008) to generate a (unique) digital signature for any digital data. This also allows examiners to check data integrity throughout the investigation process. Some forensic tools that generate forensic images use a hash function as a sub-process of their procedures.

The entire data for the Acquisition Engine stage is then stored in the system database i.e., the Forensic Images database. Table 6-1 presents information that describes all the cases in the F-FAS, such as case ID, name, creation time, the creator of the case, and the location of the crime. Subsequently, this information could be entered by the investigator who creates the case (the investigator ID is used to indicate the person's record in the investigator table). This table is directly connected to the forensic table (Table 6-2) to seek reference to the corresponding forensic images for each case. The table includes all the information that is associated with forensic images after their creation, such as image name, resource type, such as CCTV, mobile or computer, the size of the data acquired, a timestamp for the image, processing, as well as the start and finish time. The table also includes the hash value of each forensic image. These two tables are

considered as integral sources throughout the investigation process until the final report is produced. A copy of the data collection is examined in the next stage, the Pre-Processing Engine, in order to extract the relevant evidence.

Case ID	Case Name	Crime Time	Crime Location	Case Creation Time	Investigator ID	Case description
1	Case1	2017-05-10 16:35:55	Plymouth	2017-05-10 18:00:55	1	Case details
2	Case2	2018-04-14 08:45:05	London	2018-04-14 15:00:15	5	Case details
3	Case3	2018-01-30 11:55:45	Cardiff	2018-01-31 07:20:44	2	Case details
.....	.....	.....	.....	.....	.....	.....

Table 6-1: Case table

Case ID	Image ID	Image Name	Resou _rce Type	Size	Acquisition Time Starting	Acquisition Time Ending	Longitude	Latitude	Hash	Inves. ID
1	1	CCTV _PI1	CCTV	2GB	2017-05-10 18:30:55	2017-05-10 18:45:44	-4.143841	50.376289	5455F06E…	1
1	2	CCTV _PI1	CCTV	3GB	2017-05-10 18:56:57	2017-05-10 19:30:22	-4.143842	50.376290	3456GT23E…	1
1	3	Compu ter_12	CCTV	1GB	2017-05-10 19:40:33	2017-05-10 20:05:07	-4.143845	50.376292	122E87F…	1
.....	.....	.....	.....	.....	.....	.....			.....	.....

Table 6-2: Forensic images table

6.3.2 The Pre-Processing Engine

As already mentioned, it is in the Acquisition Engine phase that the system acquires data from different devices. In the pre-processing stage, the engine tries to answer what the data means and what could be considered related information. The F-FAS architecture provides a Pre-Processing Engine that can automate digital forensics examination and make it easier to complete the investigation process. This engine examines the content of forensic images piece by piece in order to filter the data that are relevant to the case.

The Pre-Processing Engine allows investigators to take a copy of forensic images from the Forensic Images database to ensure the original images are not used. Investigators can also ensure the image copy that has been received is undamaged and unaltered. This strategy enables the investigation process to be accomplished with integrity while guaranteeing the safety of the original evidence. There are different levels of evidence examination in this engine, as shown in Figure 6-3. The first level is forensic pre-processing, which aims to filter and extract relevant files from forensic images.

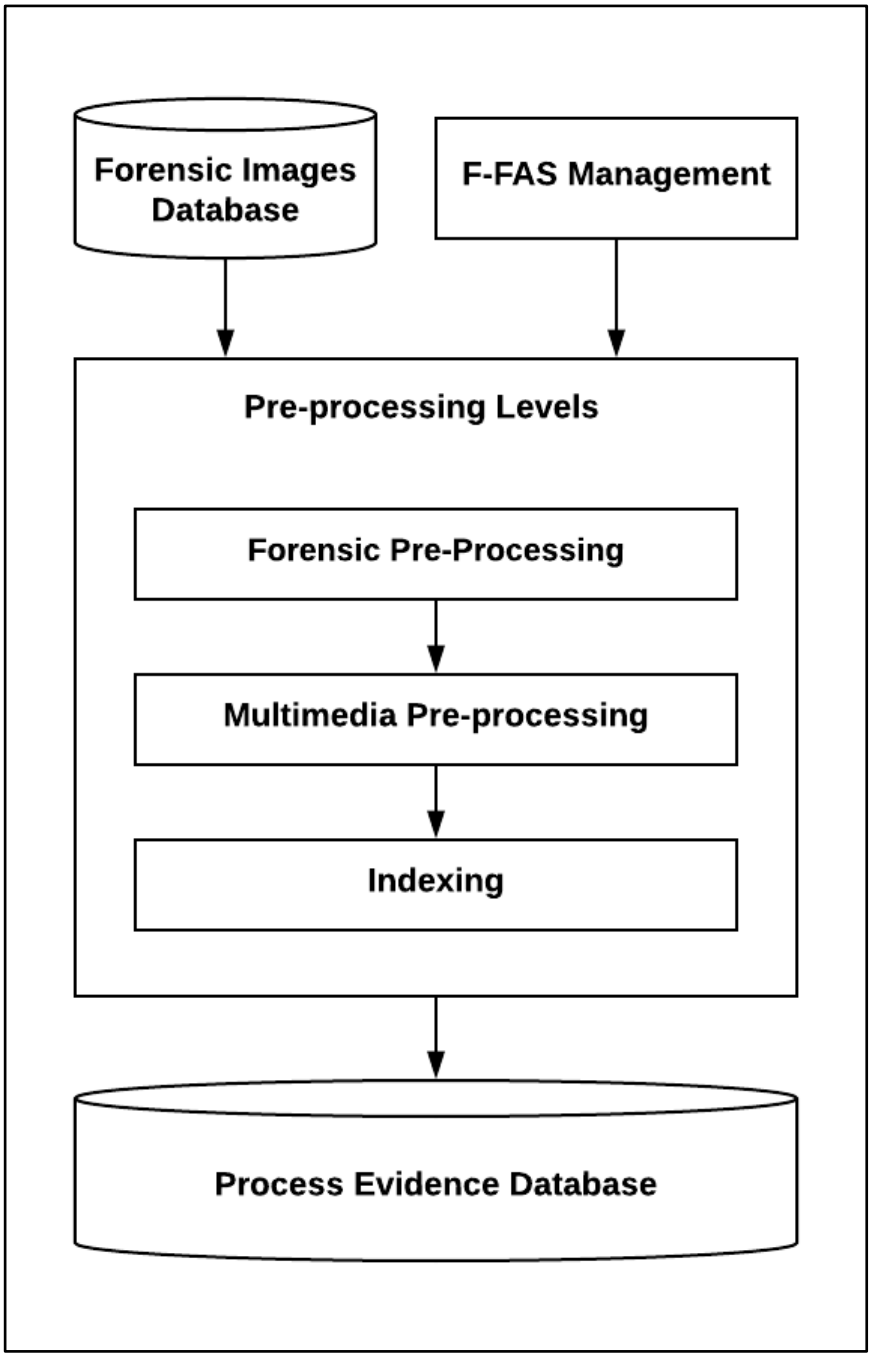


Figure 6-3: Pre-Processing Engine levels

The forensic image has to pass through different forensic examination functions for filtering. Currently, most forensic tools use these functions for examination and analysis. Therefore, there is a pressing need to include some of these functions in the F-FAS architecture. The inclusion of this task could be accomplished by expanding a compound file to extract the child files that could be contained within it, such as RAR or ZIP files. This action can extract any piece of data and will regard it as an individual file to investigate whether it could be considered as

evidence. It can also extract files from unallocated file system space by using data carving, which facilitates the finding of hidden or deleted files. Another task that has to be included in the forensic pre-process level is decryption file or password recovery (to access password-protected files or folders). In some examples, encryption software creates a hash value for each password and then stores it, instead of text, so the data become unreadable. Therefore, it is essential that the F-FAS includes a password-cracking technique. The Pre-Processing Engine could use one or more types of attack scenarios, such as rainbow, dictionary, and rule-based, to recover passwords or decrypt files. However, this technique would use in the F-FAS in seeking to design an integrated forensic tool to solve different issues in a reasonable amount of time. Moreover, one of the Forensic Pre-Processing Engine role is to calculate a hash value for each file content to use later to verify file integrity and identify known files. Known files are then referred to standard system files, which it might ignore, such as operating system or application files. It also refers to illegal materials, which it would be inclined to create an alert for, such as malware. Thus, data reduction could be used, which is based on the comparison of each hash file in a forensic image with the Known File Filter (KFF) database, which contains libraries of the hash values of known files.

The second level of the Pre-Processing Engine is the multimedia stage, as shown in Figure 6-3. As the system focuses on multimedia files (e.g., videos and photos) that contain suspects' faces, the primary task of this pre-processing level is to search for diverse sources that could include this category of file. The multimedia level could use data reduction but normally depends on file types, such as the retrieval of only multimedia files in a forensic image. The file name or extension could be referred to as no meaning file. Instead, as a prerequisite, it is essential

to use a sequence of bytes that represent file types, which are available in the header of the files. Multimedia pre-processing will filter files using a file signature table to isolate multimedia files from the rest. A file extension that has a different format, or Hex signature, needs to be checked with all formats.

In addition, it is a requirement that the F-FAS architecture, which is dependent on still photos, can extract frames from a video. The number of frames per second can be determined in the system as a default value and can be changed in the future via the system settings. Duplicate frame removal is another technique that has to be added to the multimedia examination engine to reduce the number of photo files considered in the investigation process.

Once still images are filtered, a facial detection process is applied to extract photos that include faces to achieve final data reduction by isolating these photos files from others. This process is usually achieved by using an automated face-detection approach that looks for key facial features, such as the eyes, nose, and mouth, to detect faces in an input photo. It then extracts a facial features vector and transforms it into a face template in a standard form. It must be ensured the feature extraction techniques are robust for unconstrained issues, such as changes in pose, lighting, distance, and quality. As the face template is considered to contain the unique characteristics of an individual person, the sample template will be employed for comparison throughout the investigation process in the Facial Recognition Engine. Some photos contain more than one face; therefore, it is important to extract all face templates and make references to the source photo. In addition, for every single face, the face detection operation returns face attributes that are included in the machine learning-based prediction of facial features. For instance, a set of facial attributes will include gender, age range, and the face having a beard, glasses, or a hat. Facial attribute values can

be of numerous types, such as Boolean (if a person has a beard) and string (for gender). However, any photo file that is successful in face detection will be considered as resource evidence. In addition, as part of the suggested F-FAS architecture in this engine, it is vital to extract the metadata of every file, such as name, size, and GPS, which would provide assistance in describing each file.

As a final step in the structure of the Pre-Processing Engine, an indexing component that is stored in such a manner that it enables file retrieval to work efficiently in the future. The results of this engine will be stored directly in the Process Evidence Database. As a result of the size of the data required to deal with this, the indexing approach will save considerable time and effort when performing analysis or searching for any file in the database. It is clear that, in terms of the aim of the F-FAS architecture, the most substantial data will be the photo files that contain faces. It is essential to store these file types and their details on the system storage to use them in the investigation process with the relevant forensic images, as illustrated in Table 6-3. The system also stores the metadata of each file (e.g., capture device, device model, and capture location) by connecting to the metadata table, as shown in Table 6-4. Moreover, Table 6-5 contains all the face templates and attributes that have been generated and will act as a reference for any face in a forensic image.

Image ID	Photo ID	File Location	TimeStamp	Photo Size	Photo Type	Hash	Inves. ID
1	1	\\image1\photoevidence\photo1	2017-05-10 17:30:55	300KB	JPG	234E66H9..	1
1	2	\\image1\photoevidence\photo2	2017-05-10 17:45:10	900Kb	JPG	896F98E2..	1
1	3	\\image1\photoevidence\photo3	2017-05-10 17:55:00	100KB	JPG	654F79E1..	1
.....	.....	.....	.....	.....	.....	.....	.....

Table 6-3: Photo evidence table

Photo ID	Capture Source	Device Model	Model No.	Longitude	Latitude
1	CCTV	Dome camera	WFB-100A	-4.143841	50.376289
2	CCTV	Dome camera	WFB-100A	-4.143841	50.376289
3	CCTV	Dome camera	WFB-100A	-4.143841	50.376289
.....	.....		.....	.....	.....

Table 6-4: Photo metadata table

Photo ID	Face ID	File Location	Template Storage	Gender	Age	Glasses	Hat	Beard
1	1	\\image1\photoevidence\face1-1	\\image1\photoevidence\photo1\template\face11	Male	30	No	No	No
1	2	\\image1\photoevidence\face1-2	\\image1\photoevidence\photo1\template\face12	Male	25	No	Yes	No
2	1	\\image1\photoevidence\face2-1	\\image1\photoevidence\photo1\template\face21	Male	26	Yes	Yes	Yes
.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 6-5: Face photo table

6.3.3 Facial Recognition Engine

This engine is considered as the core component of the F-FAS, as it helps to examine criminal evidence in terms of facial identification and minimises the

scope of a search. Another feature of this engine is that it saves time and effort when examining significant amounts of evidence, which will further facilitate investigators in making a final decision regarding identity by narrowing down the matching results. Following the promising validation results obtained from the experimental packages referred to in the previous chapter, the Facial Recognition Engine uses a multi-algorithmic matching approach by combining them into one engine with the aim of improving system performance. As a multi-algorithmic approach is taken with this engine, the decision to choose a better facial recognition algorithm or commercial system for each case will not be an easy task. Therefore, this system architecture is flexible and can be adapted to any new facial recognition algorithm (whether private or commercial) in the future so the system can be updated.

The question that could affect the F-FAS architecture is what is the cost of using multi-algorithmic facial recognition instead of a single algorithm? Will it cost more? It is suggested that the proposed F-FAS use some commercial algorithms, such as Amazon and Microsoft (Chapter 5), which are based on providing cloud services to customers. The payment system for these algorithms involves paying only for the resources used at any time. For instance, Amazon Rekognition charges customers only for the number of images analysed (e.g., face detection, face matching) or cloud storage. The price is calculated per image analysed and it is minimised when the number of images is increased, as shown in Table 6-6. It is noticed from these prices that this type of system provides cheap prices than buying a full facial recognition system. Further, customers will benefit from using the newest technology developed by computer vision scientists and daily updates of the service's features.

Image Analysis	Price per 1,000 Images Processed
First 1 million images processed per month	$1.00
Next 9 million images processed per month	$0.80
Next 90 million images processed per month	$0.60
Over 100 million images processed per month	$0.40

Table 6-6: Examples of Amazon Rekognition's prices.

To perform facial recognition, this engine contains a controller, which manages the data flow and processes. When a request is received from an investigator (e.g., for a suspect's face), the Facial Recognition Engine performs facial identification by drawing comparisons (one to many) between the input query and the mugshot data in the Process Evidence Database, as shown in Figure 6-4.

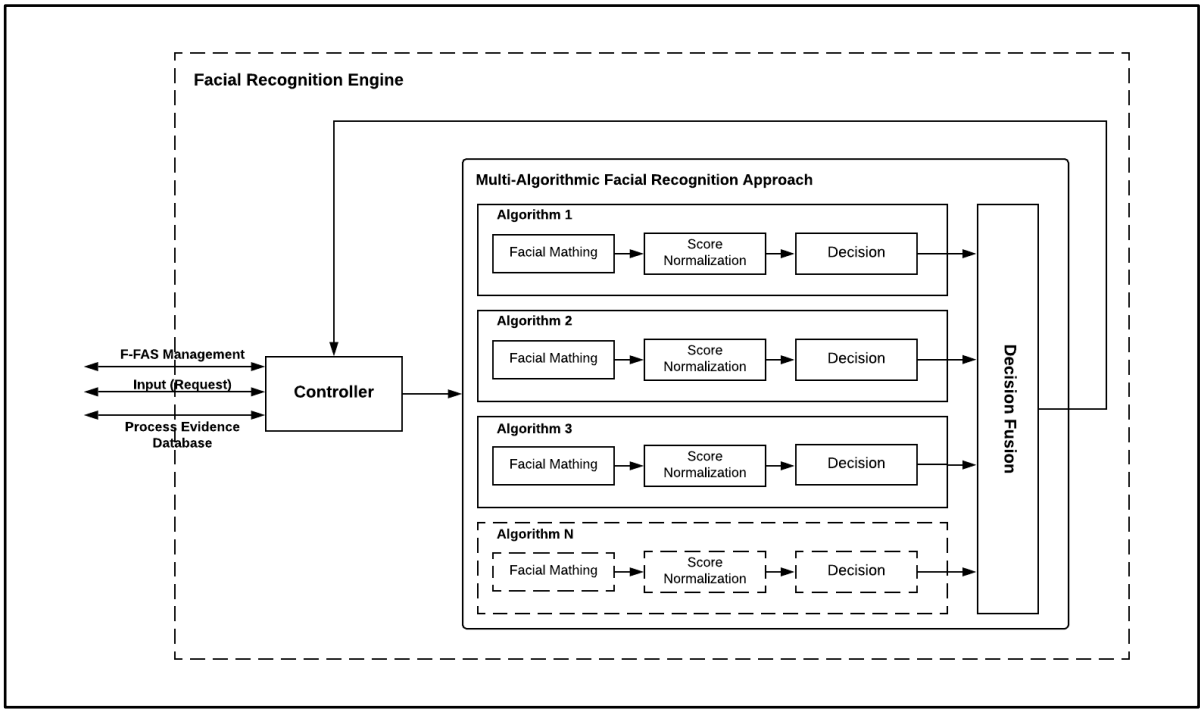


Figure 6-4: Facial recognition Engine

The input query (unknown identity), which represents the suspect's description information, will be in position to check the identity. This could appear in two ways:

1- Photo file: this type is provided to the system when the investigator has any photograph of a suspect's face and then needs to return all the

photos that are matched with it. The photo source could either be from external resources, such as having been collected from a relevant witness or chosen from the data acquired.

2- Soft-biometric information: this type of query arises when the investigator does not have a photograph of a suspect's face but has some information collected directly from a witness at the crime scene. For example, this information could be a victim's or culprit's soft-biometric features, such as age, gender, any accessories, or a beard. This information will help minimise the search scope.

Moreover, the F-FAS structure allows the inspector (optional) to specify a specific narrower search to reduce the amount of time involved. For instance, the search scope time, which signifies the initial start and end time, ultimately relies on the crime case time. Another example is determining the GPS (both longitude and latitude) coordinates that are sought.

The Facial Recognition Engine calculates a match by exploring the similarity between the input query and the face templates available in the Process Evidence Database. In the case of the input query being a photograph, this engine will generate a template with which to draw a comparison while, in a soft-biometric case, the engine will directly draw comparisons with information already saved in the database (Table 6-5). The match score result will then normalise every algorithm to be ready for the decision fusion to generate the suspect's final identity. Decision fusion is essential to give weight to every individual contributing matching algorithm used. High weights are assigned to an individual algorithm based on its features (i.e., the ability to deal with face photo issues, such as illumination and pose) and identification accuracy, which could be tested before being added to the F-FAS structure. Hence, the fusion objective is intended to

provide a robust identification decision. Once the similarity comparison has been made, all relevant face photos will be displayed to the investigator ordered by high matching scores.

All the identity outcomes of the decision fusion will be returned to the controller, which will store them in the Case Evidence Database. This will provide a search ID for every request, the search time, and the ID of the investigator who sent the query. An input query value, for either the photo source or text, represents the soft-biometric information, as shown in Table 6-7. Moreover, any additional search criteria will be stored in Table 6-8. The results of the photo ID that correspond to the search ID will be stored in Table 6-9.

Bookmarking is now a convenient strategy for identifying relevant information because it provides a facility to search for specific bookmarked files rather than looking into all the existing files. Therefore, by using this facility, an investigator is permitted by the architecture to select any particular photo to keep as bookmarked evidence and set a value (1), as shown in Table 6-9.

Search ID	Query Type	Value	Investigator ID	Search Time
1	Photo	\\case1\Queryphoto\photo1	1	2017-05-11 10:30:55
2	Text	male, 35, 1, 1, 0	1	2017-05-11 14:40:11
3	Photo	\\case1\Queryphoto\photo2	6	2017-05-11 15:08:33
.....	.....	.....	.....	.....

Table 6-7: Query details of Facial Recognition Engine

Search ID	Scope Time Start	Scope Time End	Longitude	Latitude
1	2017-05-01 01:00:00	2017-05-10 17:00:00	-4.143830	50.37695
2	2017-04-11 02:00:00	2017-05-11 15:00:00	-1.143830	52.37725
3	2017-05-12 09:00:00	2017-05-30 03:00:00	-3.143830	51.33455
.....	.....	.....	.....	.....

Table 6-8: Additional Filter information for facial recognition process

Search ID	Photo ID	Face ID	Bookmark
1	1	1	1
1	20	2	1
1	35	3	0
.....	.....	.....	.....

Table 6-9: Facial recognition results

At the end of this stage, the investigator will have found most of the photos that are related or are similar to a particular suspect's face. Moreover, the resulting photos will be ready for further analysis to find any correlation between them.

6.3.4 Forensic Analyses Engine

The most significant part of any digital investigation is the ability to analyse the evidence that has been found. Because there is a huge amount of information that needs to be to analysed and evidence found for correlations between data.

This will help reduce the effort and time in the investigation process. The F-FAS architecture provides a wide range of analysis capabilities that will allow investigators to perform in-depth analyses that can lead to cases being solved. Therefore, the Forensic Analyses Engine enables investigators to find answers to those questions that are often not merely about the identification of an individual but are also related to an understanding of the context. The engine identifies associations between artifacts and presents them in a usable and visual form to draw a wider picture of a crime. To help illustrate how the proposed framework would operate, a child abduction example is presented. In this example, it is assumed a child has been kidnapped. Intelligence provides a rough last location of the child and the latest picture of the child. In order to solve a child abduction case, an investigator starts by collecting all preliminary evidence that may help find the child as quickly as possible, such as narrowing the timeframe of abduction, determining the location of the abduction, and any information about the suspect (e.g., face description, age, and gender). The next step of the investigative process would involve collecting all available imagery (e.g., videos from surveillance cameras at the crime scene and from nearby surveillance systems). The current solution would involve teams of investigators manually trawling through the footage from possibly dozens of evidence sources. The use of a manual human-matching process is laborious and time-consuming, resulting in examining large volumes of image data. Further, given the pressurised nature of the task, it is likely to result in a high proportion of human error. The proposed system will permit an investigator to select the necessary evidence sources and automatically process all of the footage. The system will then perform facial recognition across the evidence sources, providing an investigator with a prioritised set of results with which to interact. The investigator will be able to target images from the retrieval results and the forensic analysis engine will

provide further correlation between faces in images and enable the target face to be tracked across different evidence sources. The resulting visualisation would provide the graphical map of the resulting journey alongside the image sources used to identify the victim's path.

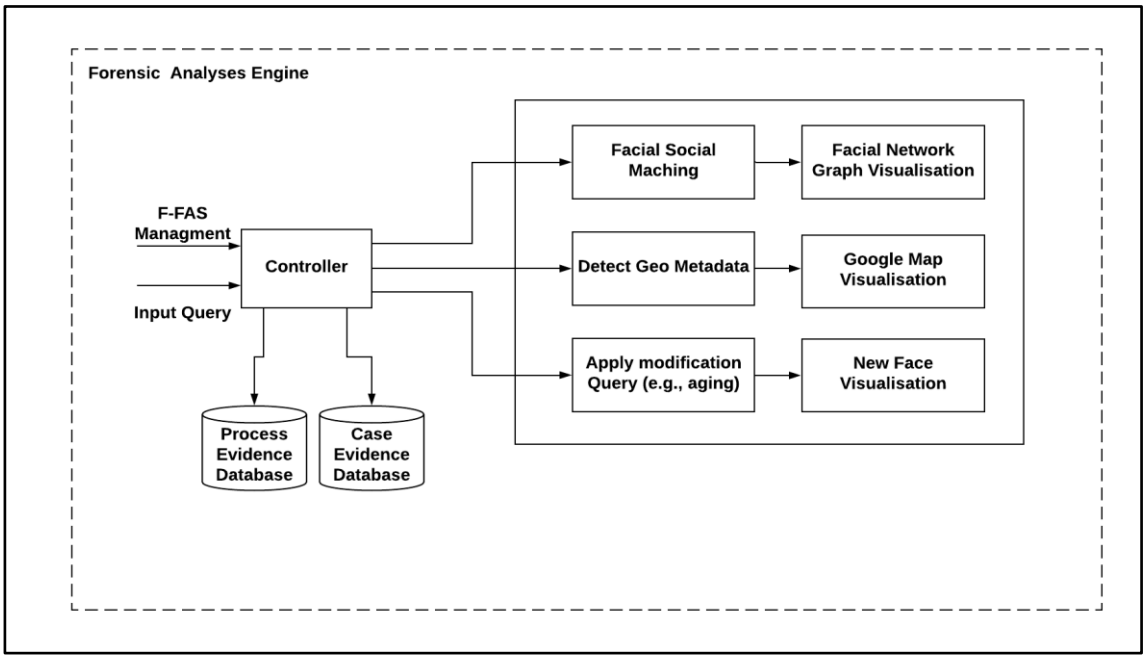


Figure 6-5: Forensic Analysis Engine

The Forensic Analyses Engine carries out its work based on previous Facial Recognition Engine results. The Forensic Analyses Engine provides three different types of forensic analysis processes: Facial Social Network (FSN), Geo Location (GL), and Facial Modification (FM). These three types of analysis aim to identify any relevant correlation between found artifacts based on different input queries or requests and to reduce the outcome evidence in a short time. Figure 6-5 shows the process of the Forensic Analyses Engine.

- **Facial Social Network Analysis**

Facial social network analysis is a type of forensic analysis suggested in this design. This technique is based on searching people's faces using facial matching and returning all photos other than those of the suspected person. Interestingly, it also has the ability to return photos of the faces of the people who

appear in the same image along with the suspect. This form of forensic analysis uses graph theory to depict a network structure. Graph modelling is an approach that describes a connected graph of nodes and relationships. Nodes are utilised to represent a source photo of a suspect's face or faces of people who are closely related. A relationship is represented by the connection between two nodes, which represents the facial similarity between the nodes and the closely related relationship.

In a facial social network, the investigator first selects a face photo of interest from the retrieved results (Facial Recognition Engine) in the Case Evidence Database. At the first level, the Facial Recognition Engine is used to return all possible photos of the selected input. At the next levels, and while building a facial graph network model, a recursive process is repeated for each face that appears closely associated with a photo node (using Table 6-5) and repeats the retrieval process (face matching) to return all possible faces.

This approach appears to reduce the time required to analyse large numbers of photos and can build complete tree tracking for suspects, as well as their relatives. Furthermore, filtering options are provided to aid the investigator, depending on the investigation requirements. For example, only the photos of a suspect that are found in the databases could be displayed or photos of all the faces that seem, by their appearance, to be associated with the suspect. It is noteworthy that visualisation of a facial social network is extremely significant to understanding the results proposed by the investigator. A version of a graph visualisation platform could be used to represent the results as a graph, which the investigator would more easily understand.

- Geo Location

An increasing number of photos are captured on devices to support location-determining technology. Geolocation data are often incorporated into metadata,

thereby providing the potential to analyse artifacts further. From a forensic point of view, location data are valuable because of the ability to indicate the last location of a suspect and could provide an accurate account of that person's movements. Therefore, a geolocation analysis objective could be to:

- refine outcome analysis by location;
- determine the location of photo matches onto maps; and
- track the movements of individuals of interest by location and time.

An investigator can refine photos of interest from the retrieval (facial recognition) results and metadata provide useful information that can assist investigators in determining the exact location of a captured image. The results can present further visual reorientation by using Google Maps. Forensic investigators can then track photos of suspects based on metadata information (Table 6-4). This can be helpful in solving criminal cases and reducing the time that is normally required in manual tracking. In addition, this technique provides an overview of the directions of persons of interest to assist in tracing their whereabouts.

- Facial Modification

Facial modification analysis allows investigators to adjust a photo of a suspect's face to establish if it can be matched with additional evidence. In some cases, there is a difference between the appearance of faces that were saved in a database and the input face query. Therefore, the F-FAS augments the ability to make changes to the shape of a face, such as adding glasses, a beard, pose correction, changing the expression, and employing ageing effects. This technique is intended to increase the probability of retrieving more faces that resemble those changes. In addition, it could enhance facial identification performance by trying to solve some of the forensic facial recognition problems discussed in Chapter 3.

When an investigator selects this type of analysis, it requires choosing a face photo, which needs changes to the face shape. This selection depends either on the facial recognition results from the previous engine or the entry of a new photo. The modification types that need to be applied are then selected and the system allows the simultaneous application of more than one face modification, such as adding glasses *and* a beard. This could use algorithms with a high degree of accuracy that specialise in facial modification suggestions and can be added to the F-FAS architecture. A new facial photo could be re-sent to the Facial Recognition Engine to verify the match results.

6.3.5 Reporting

One of the most important steps in the digital forensic investigation process is generating forensic reporting. Forensic investigators spend much of their time in the examination and analysis of evidence, so their findings can be reported to stakeholders in such a way the results can be understood. Some stakeholders will not have a relevant technical background or in-depth knowledge of facial recognition techniques. Therefore, it is helpful to provide multiple options for viewing digital forensic reporting so it is available in a simple and understandable viewer. The report content could contain tables, facial photos (match results), Google Maps screenshots, and any data associated with a case and a summary.

The F-FAS organises all artifacts found in the reporting viewer, such as bookmarking, search details, and analysis. However, it also allows investigators to select which evidence they prefer to include in the final report or to choose the option to include everything automatically. All the evidence found is stored in the Case Evidence Database and is ready to export for reporting. In addition, the F-FAS manager can feed the reporting of any information that is related to the case details (Forensic Images Database) or the investigators' management details

(System Management Database). Once the reporting is complete, it can be saved as a different file type, such as Word or PDF, and can be shared with other stakeholders.

6.3.6 The F-FAS Management

The F-FAS Management is the central controller of the system processing and provides an interface between the investigator and the underlying processes, which helps investigators manage the overall system. The system management also controls the monitoring of the dataflow through the chain of custody. The F-FAS Management provides an optimal forensic environment that ensures that all stages, from data collection to final reporting, work securely and with integrity. The F-FAS Management's responsibilities can be stated as follows:

- System Security (AAA Functionality)

This task manages the authentication, authorisation, and accountability (AAA) of the overall system in order to fulfil privacy requirements. The first process, authentication is aimed at verifying who will have access to the system by checking credentials such as user ID and password. Once the user is logged in, the authorisation process determines what types of tasks, activities, resources or services are allowed to the user. However, the authorisation of tasks is based on the context of the authentication, so an investigator's account roles are authorised once he/she has been authenticated to log into the system. The final aspect of the AAA function is accountability, which involves tracking, recording, and measuring the operations and activities undertaken by the investigators while they access the F-FAS. The purpose of this process is to record the answers to questions of "Who did", "What did" and "When did", which circumvents the possibility of any conflict between investigators in the future. This process begins once an investigator starts logging into the F-FAS, so the system management

starts recording but is not restricted to point and time of access, current services, or data usage, and updates the user status. Once an investigator's access has ceased, the system management will stop recording, the investigator's status will be updated to inactive, and all the information will be stored in the investigator's records in the database. It is worth mentioning that all the AAA security aspects come under the responsibility of the System Administrator.

- System Tasks

The purpose of system tasks is to distribute the roles between the system members to ensure the smooth running of the whole system throughout the F-FAS, such as system security, case management, investigators' management, and digital investigation steps. The F-FAS task is assigned three different roles: System Administrator, Case Leader, and Forensic Examiner. The System Administrator is responsible for all levels of system security, such as authentication and investigators' account configuration, including adding/deleting users, and giving roles, permissions and authorisations to others, such as the Case Leader and Examiner. In addition, it maintains the system databases to ensure each is running correctly by checking its structure and size, mainly to monitor data archives. It is also responsible for any software configuration (install, modify, remove, turn on/off) the F-FAS needs; for example, the facial recognition systems and analysis tools would be part of the System Administrator's main duties.

The Case Leader is responsible for case management, such as creating cases, acquiring images, examining analysis, managing case data, and reporting further to ensure the investigation process follows the rules as mentioned in ACPO principle 4 (ACPO, 2012). In addition, all cases that are no longer necessary will be considered archived by the System Administrator and Case Leader. Archived

cases are deleted from the case database and will, instead, be stored in the case folder (external disk). The Case Leader could allocate a case to a specific examiner for investigation to identify evidence and perform all tasks (Forensic Examine). The Case Leader's duties could function at two levels: either globally, across all cases in the F-FAS, or through taking specific roles that vary from case to case, as determined by the System Administrator.

The Forensic Examiner is responsible for pre-processing images and examining multimedia files by using the facial recognition function. Moreover, the results can be analysed using a facial social network, geolocation, and facial modification tools, which further extract evidence and verify its integrity by producing a report. However, every role is determined and does not overlap with others, as illustrated in Table 6-10. The F-FAS provides a System Management Database that stores all the information about stakeholder users (i.e., the System Administrator, Case leader, and Forensic Examiner), as shown in Table 6-11.

Tasks	Role		
	System Administrator	Case Leader	Forensic Examiner
Authentication	√		
Authorisation	√		
Accounting	√		
Software Configuration	√		
Investigator Management	√		
Create Case		√	
Acquire Image		√	
Examine Image		√	√
Analyse image		√	√
Extract Evidence		√	√
Reporting		√	√
Delete Image		√	
Archive Case	√	√	
Archive Image	√	√	
Delete Case	√	√	

Table 6-10: The F-FAS' responsibilities and roles

Investgator ID	Investgator Name	Role	Registration Time	Address	Telephone
1	Hiba Mo	Admin	2015-06-18 12:00:10	86 Camden Street….	07747…
2	Alex Jo	Leader	2015-06-18 17:45:10	25 Armada Street ….	07757…
3	Dany John	Examiner	2015-06-18 18:45:10	14 Northhill Road…	07748…
……	……	……	……	……	……

Table 6-11: Investigators' details

6.4 The F-FAS Implementation

F-FAS implementation aims to reflect on how the F-FAS prototype would help in visualising and understanding how the architecture would work in practice. The F-FAS architecture can be designed according to two types of prototype: a web-

based service and a local desktop tool. This implementation chose a web-based design to give an example of how the architecture would function in the web environment. There are also the following advantages of using web-based tools:

1. Cross-platform: this would be identical on any operating system, such as Windows and Mac.
2. User interaction: this uses multiple tabs in a browser to perform several tasks, while a desktop tool would need several windows.
3. Maintenance: a web browser is needed to use it and it would be updated once on the server while a desktop tool would need to be downloaded and updated on a desktop computer.
4. Community: this would provide greater scalability and could be used by worldwide organisations. In addition, the cloud allows using any files from any computer from different locations, so this is a good example that provides more scalability in the investigation process or local private servers (for more security).
5. Usability: this could be accessed by users from any location, who would simply need the internet while a desktop tool is confined to a physical location.

Despite the above advantages, there are some risks that could be faced by using a web-based tool, such as security. For example, the tool would be open to hackers but a cryptography technique could be used to cipher data when moving between the client and server sides. Another risk is that a web tool depends on internet connectivity being live; when this is absent, the tool cannot be accessed. Moreover, the cost could be higher for a web tool that needs to be continuously maintained.

The prototype of the F-FAS was developed with a front-end design to show the visualisation of each engine in the system architecture. The web development used Bootstrap, HTML, and a JavaScript framework to develop the front-end design. The F-FAS prototype provides a responsive design tool that allows it to be adapted to various screen sizes, such as for mobiles, tablets, and desktops. In addition, the F-FAS architecture suggested for future use with private secured local servers to save all files and database. The F-FAS prototype will give good insight into how architecture is designed in the real world.

6.5 The F-FAS Prototype Samples

Each user has a specific task to perform based on its role (Admin, Case Leader, Forensic Examiner), such as the System Administrator having full access to the system, including administrative tasks. When a user is authenticated, all running systems would show up based on the user's authorisation. Figure 6-6 shows the Forensic Examiner dashboard, which includes the range of available cases in the investigator's account that are ready for investigation. There is information that describes each case (case metadata) and four actions that could be taken for each one: open, edit, archive, and print reporting. A new case could also be created and added to the number of current cases.

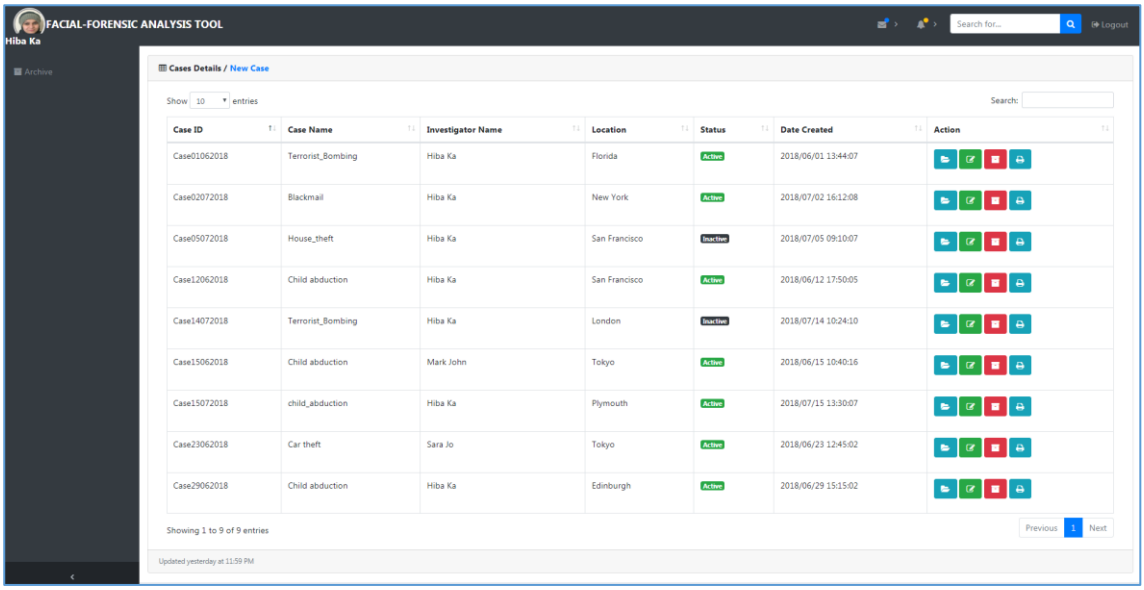


Figure 6-6: Forensic Examiner dashboard

Once the case to be investigated is specified by the investigator and the “open” action has been selected, the case dashboard will appear, as shown in Figure 6-7. This dashboard will show all the digital images acquired and display their metadata, such as type, name, and size. Moreover, a new image could be added by clicking on the “New Image” link; a new window then appears that allows information to be entered about the new digital image, as illustrated in Figure 6-8. The left-hand side of the case dashboard contains a range of actions that are available to aid examination and analysis and create further case reporting.

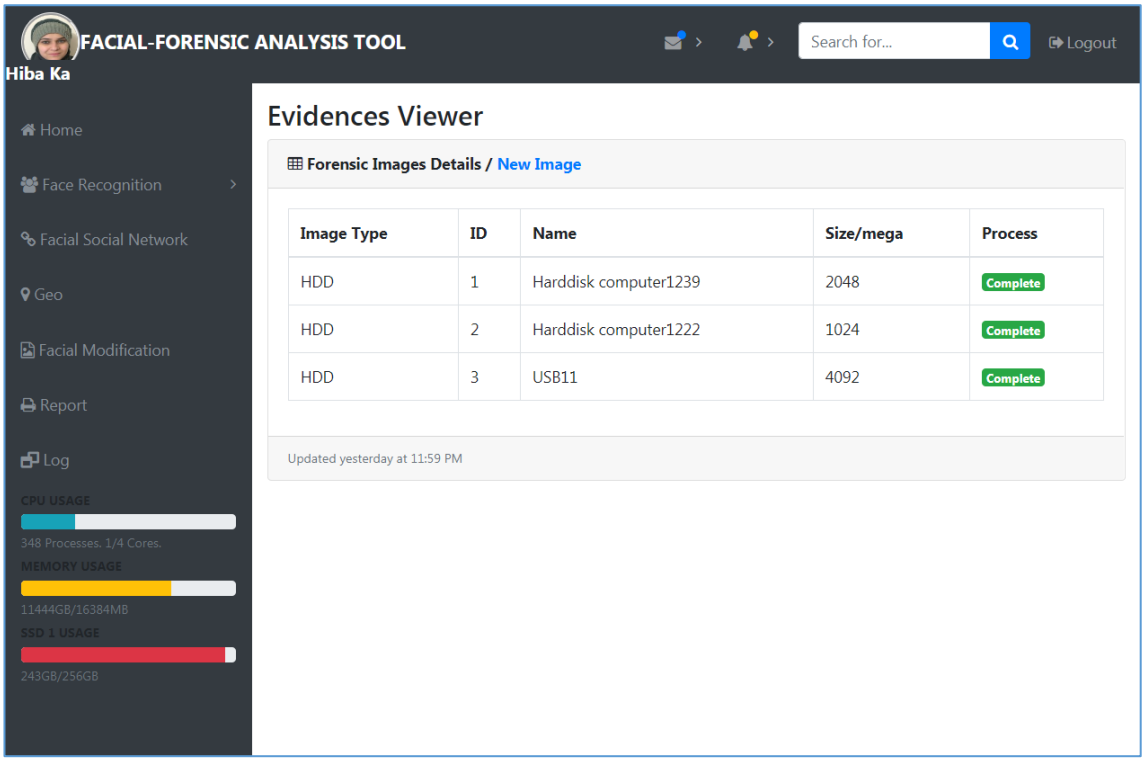


Figure 6-7: The case dashboard

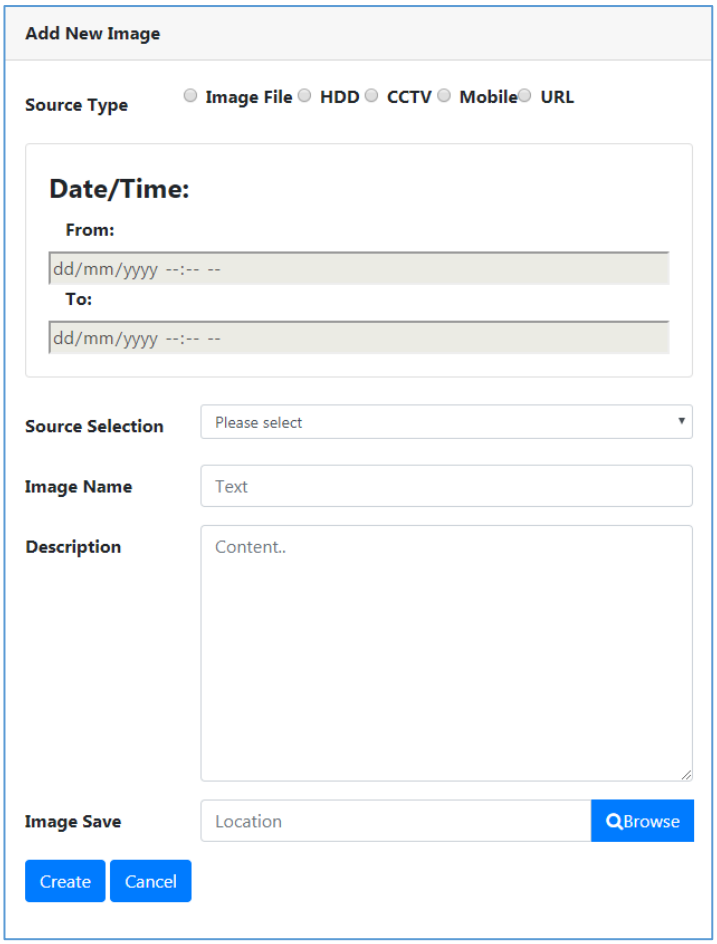


Figure 6-8: The ‘add new forensic image’ window

When all images are uploaded to the servers, they will automatically undergo pre-processing operations (described in Section 6.3.2) to filter the data and generate photo files. The forensic examiner can examine all the photo results using a facial recognition process to determine a suspect's identity. The facial recognition option, as illustrated in Figure 6-9, provides three methods for uploading a suspect's details. The first option is uploading the suspect's digital photo to find a match, if available.

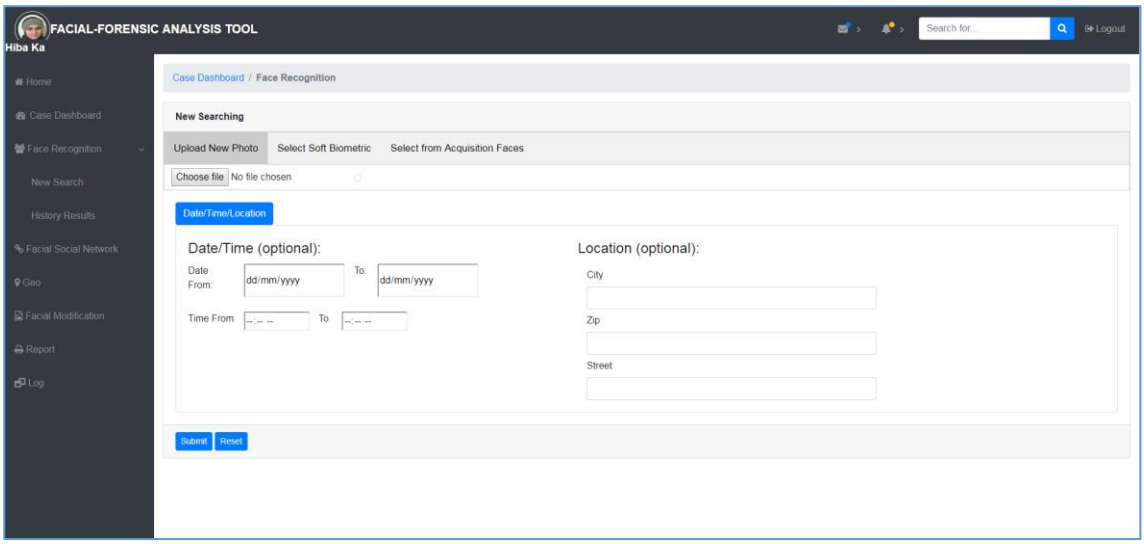


Figure 6-9: The facial recognition options window

For the second option, if the information about a suspect comes from a witness (no photo available), then choose "Select Soft Biometric", as shown in Figure 6-10. For the third option, if the investigator would like to choose from the database of acquired forensic images, choose "Select from Acquisition Faces".

There are also additional (optional) search criteria, such as date, time, and location. These search limitations could help decrease the scope of the search by matching photos that have the specified metadata limitations.

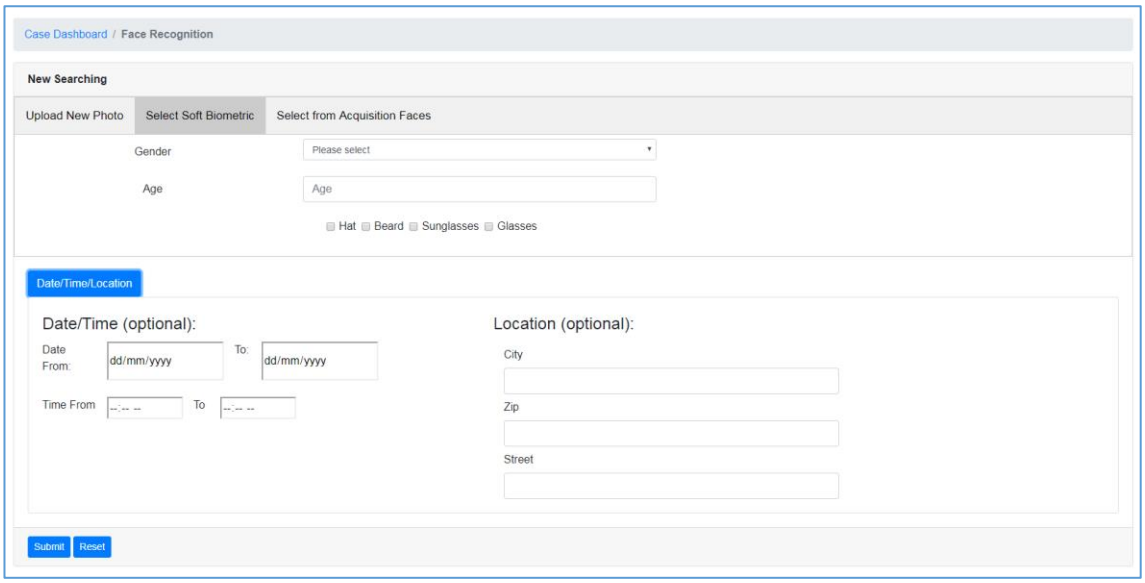


Figure 6-10: Soft biometric option of facial recognition searching

The facial recognition results will be returned after being processed using the multi-algorithmic fusion approach and displayed as a list for the investigator, as shown in Figure 6-11. The investigator can select photos of interest either to add them to the bookmarked evidence or to send them for analysis options.

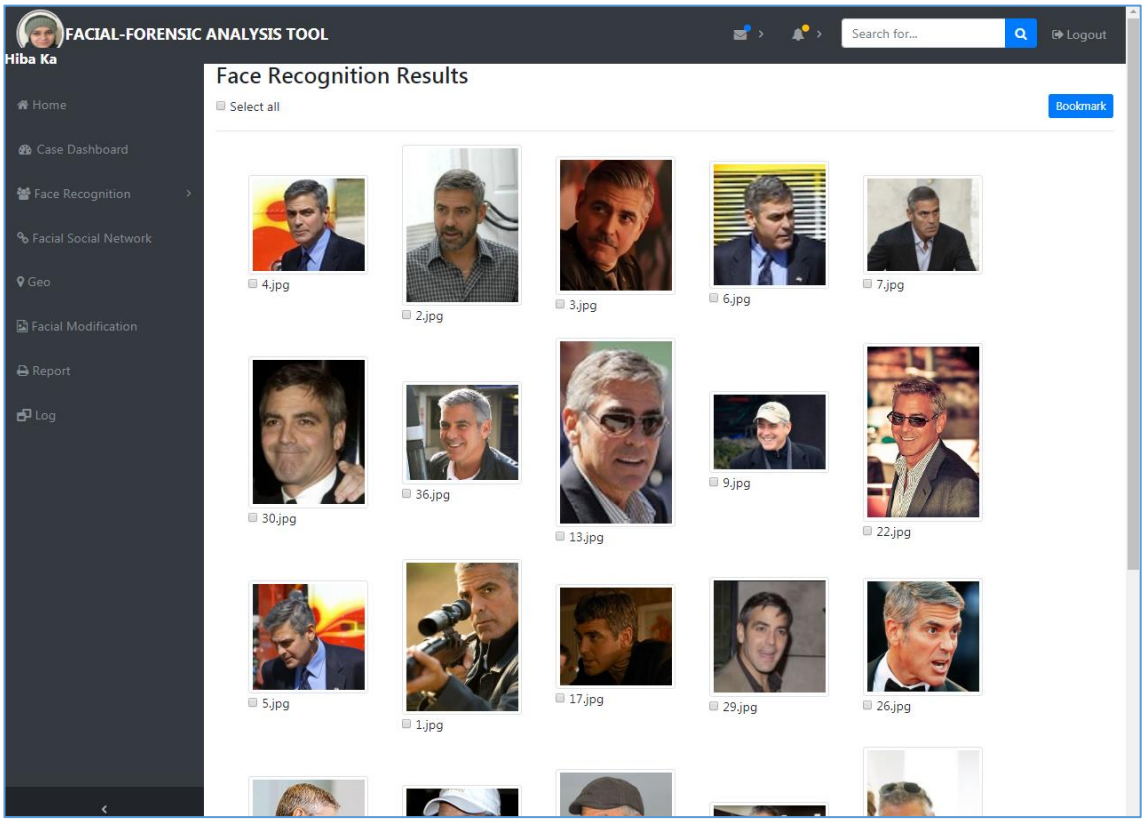


Figure 6-11: Facial recognition results window

To help the investigator avoid repeating any facial recognition searches, the F-FAS tool saves the whole search history and allows the investigator to check it, as illustrated in Figure 6-12.

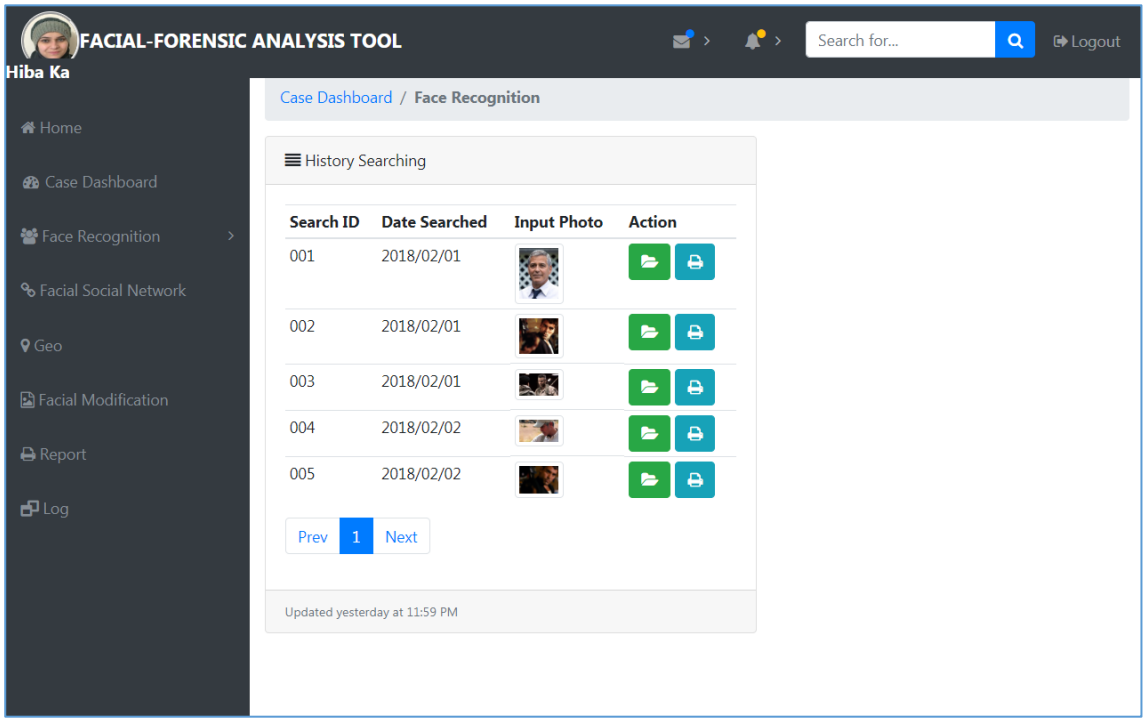


Figure 6-12: History facial recognition searching

The forensic investigator can conduct more analysis of the facial recognition results by using the three forms of analysis provided by the F-FAS tool. The first analysis type is a facial social network, which builds a visualisation network of related faces (based on appearance) shown in a visual format. Graph visualisation provides additional value for data analysis by presenting connections and quick access to photos of interest. Figure 6-13 presents an example of facial social network analysis results.

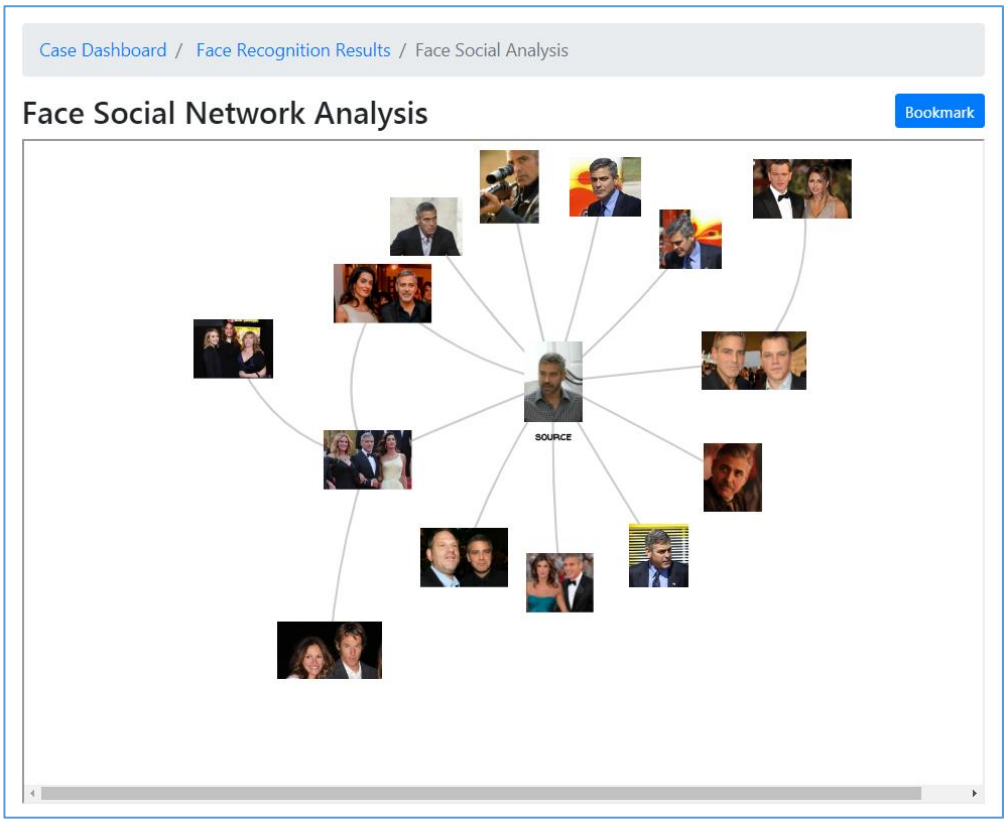


Figure 6-13: Facial Social Network Visualisation

The second type of analysis is to determine the location (geolocation) of photos using their metadata (if available) and showing their location on Google Maps, as shown in Figure 6-14.

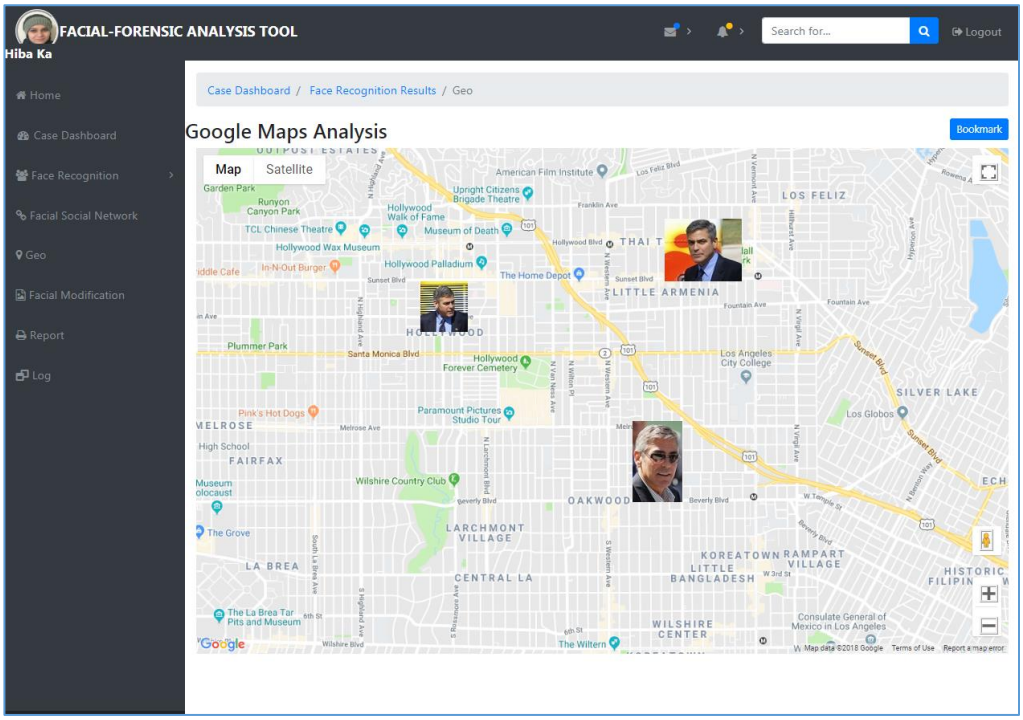


Figure 6-14: Geolocation Visualisation

The final analysis type offered by the F-FAS tool is facial modification. This analysis technique enables the investigator to modify any photo selected and uploaded into the tool. Some facial modifications that are possible are to add a beard, glasses, sunglasses, and effects of ageing. The web design option for this analysis type is shown in Figure 6-15.

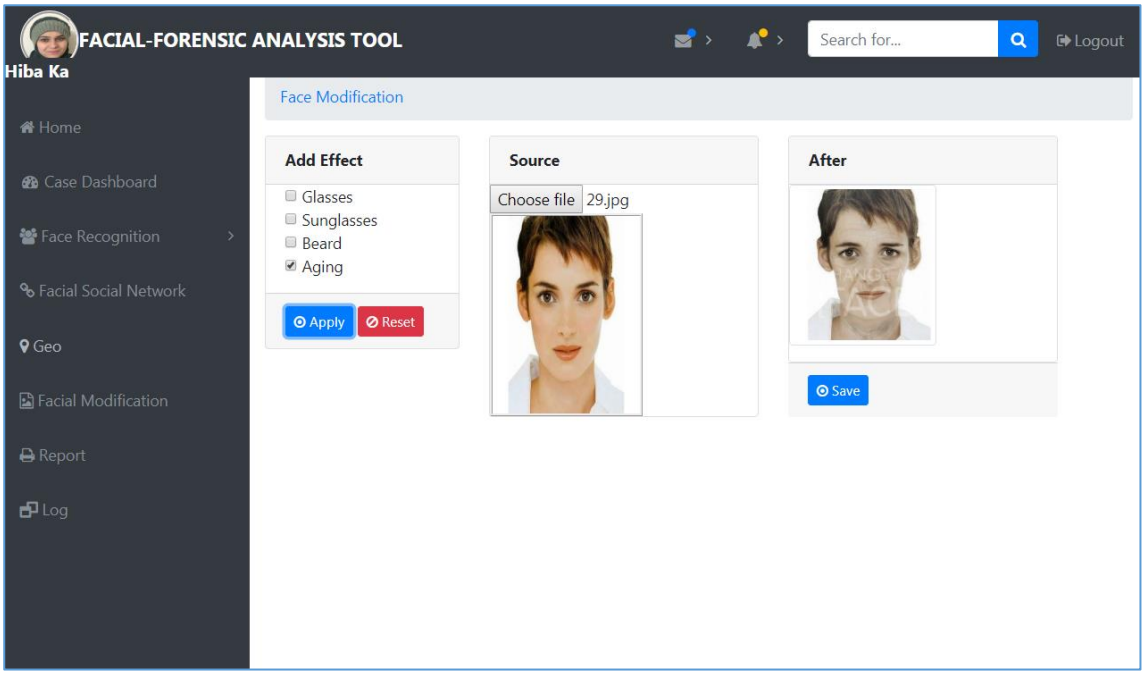


Figure 6-15: Facial Modification window (ageing example)

Once the case has been analysed, the forensic examiner can generate a report and determine the content of that report, as seen in Figure 6-16. The final report is created and can be saved and printed. Figure 6-17 shows an example of a reporting design suggested by the F-FAS tool.

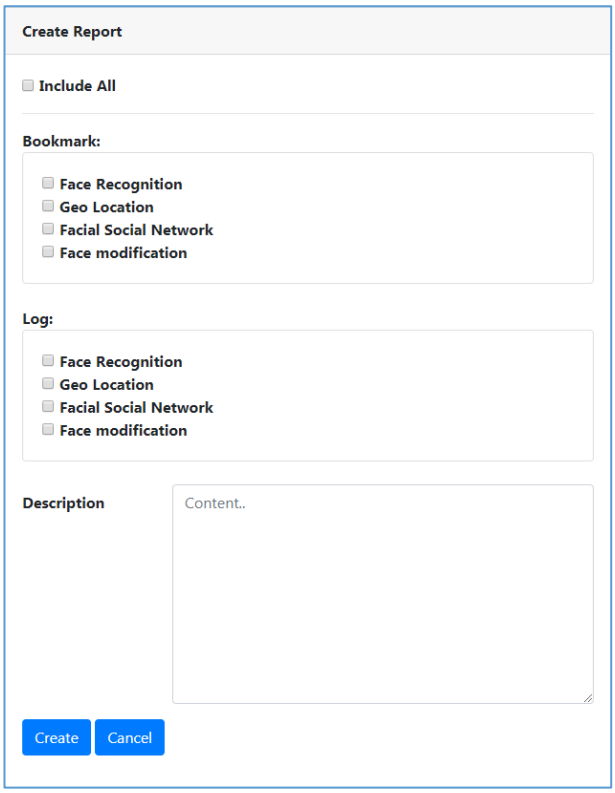


Figure 6-16: Reporting creation window

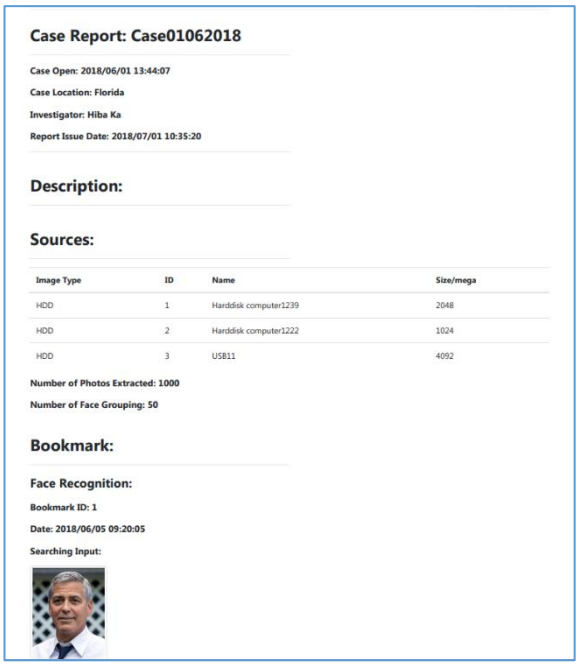


Figure 6-17: An example of reporting design

The System administration dashboard is different from the Examiner and Investigator dashboards, as shown in Figure 6-18. The Administrator can monitor cases and investigators by compiling statistics regarding the numbers of cases, users, and their statuses. The F-FAS tool also offers the Administrator the ability to manage investigators' accounts and add or delete users (Figure 6-19).

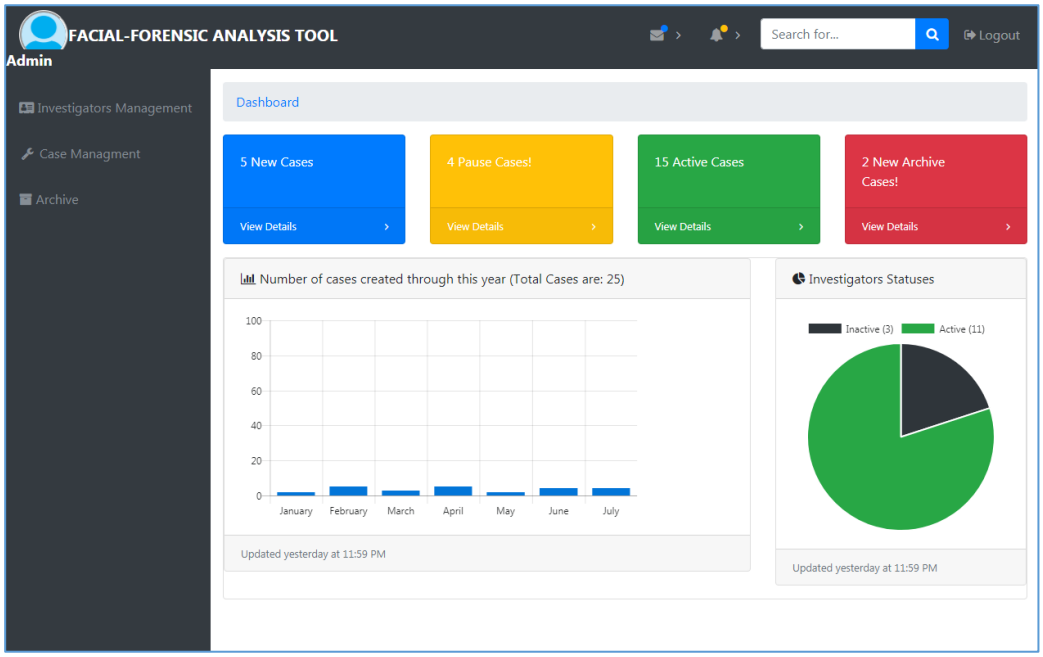


Figure 6-18: Admin dashboard

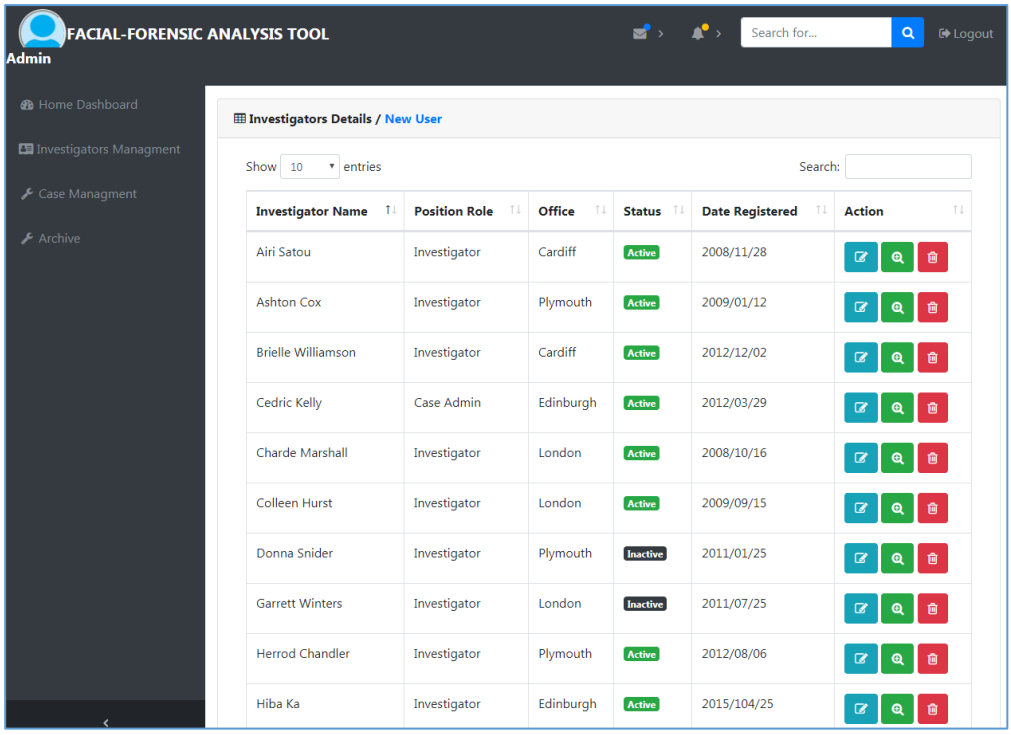


Figure 6-19: Investigator's accounts window

6.6 Conclusion

The architecture characteristics of the F-FAS have been designed in a modular and robust manner to allow investigators' a fundamentally novel approach to facial forensic analysis. The mechanism has been designed on the principle of allowing a suspect's identity to be found by reconstructing digital images. In addition, the system employs various analysis techniques for correlating between facial photos, geolocation, and facial modification to reduce the amount of investigation time required.

The proposed architecture meets most forensic technique requirements with the aim of maintaining the chain of custody, system security (the AAA function), and data monitoring. Further, the system is updatable. The proposed F-FAS architecture has been designed according to a pattern that enables forensic organisations to meet specific requirements by providing the flexibility and convenience of predefined composite factors. All investigation steps are integrated and managed within one system in the F-FAS architectural model.

This chapter has demonstrated that a complete case can be tracked from the point of image acquisition through image examination (facial recognition) and analysis (facial network, geolocation, and facial modification) to the reporting process, all performed within a single system. Based on this architecture, a prototype (web-based service) has been designed and presented to enable the visualisation of the proposed system.

7 System Evaluation

7.1 Introduction

The system evaluation aims to provide an estimation of the performance and to recognise the limitation of the F-FAS. Therefore, selection of the appropriate evaluation methodology is a significant factor that facilitates measuring the concepts of the entire project such as problem, requirements, suggested solution, and architecture. One of the evaluation methods is participative research that evaluates the stakeholders' community (digital forensics experts) in order to review the project, its unique functionality and to identify its strengths, weaknesses, and limitations. The system evaluation is envisaged to have these evaluations, which could be from academic and specialist participators in digital forensics to support the results performed during the PhD research work. Based on the academic context of the research, people with an academic background are considered the best selection to evaluate the project.

For that sake, the system evaluation requires preparatory steps to find the appropriate answers for evaluation. The first step in our system evaluation methodology was to create a video podcast that provided a brief description about the research problem in the facial forensic system by using case example, the model system requirements and architecture (Chapter 6), the experiment methodology and analysis of multi-algorithmic fusion (Chapter 5), and screenshots of interfaces of the developed prototype. The video is approximately 20 minutes long[1]. Afterward, the questions were created to collect feedback that

[1] Available at *https://vimeo.com/289571247* and a password is required to open it, which is "h#17m05*hh12"

evaluates and measures the capabilities of the project. The questions were designed to accomplish the project journey.

The second step is to identify people with various kinds of experiences in the digital forensic field, which assists in seeking feedback to cover different research dimensions. After that, it will invite these academic experts formally by sending individual e-mails. Once the invitation is accepted, the ethical approval consent form will be sent to him/her to be read and signed.

The interview session will be conducted to ask the evaluation questions. Possibly, it will be done over the internet (i.e. Skype) or face to face (if applicable). The communication language will be English and the total amount of time required for the entire interview will be between 30 to 40 minutes (this depends purely on the discussion time).

7.2 Research Methodology

7.2.1 Evaluations Questions

The total questions prepared for the evaluation interview were 10 questions and they are listed as follows:

1. What are your thoughts regarding the research problem?
2. What are your thoughts about the current facial recognition systems?
3. What are your thoughts about taking a multi-algorithmic fusion approach?
4. What are your thoughts about the proposed F-FAS architecture? Is it realisable?
5. The purpose of the F-FAS is to reduce the time taken for data analysis and to reduce the cognitive effort required to understand the relationship between artifacts. Do you feel it does this?

6. What do you think about the effectiveness, reliability, and usability of the following functionalities offered by the F-FAS tool:
    a) Facial recognition
    b) Facial social analysis
    c) Geo analysis
    d) Facial modification
    e) Reporting
7. What is your opinion about case management that is provided by the F-FAS?
8. What do you think are the particular strengths and weaknesses of the F-FAS tool?
9. Do you suggest any other functions could be added to the system to improve its effectiveness?
10. Do you have any further comments?

As described above, these questions were designed precisely, in terms of the project, to explore the problem validity, and the reliability, efficiency, and utility of the suggested tool. Additionally, they inquired about the effectiveness of the experimental methodology, results of the multi-algorithmic fusion technique, and about how this suggestion will minimise the investigator's time and effort throughout the investigation. Moreover, they were used to collect various perspectives to appraise the strengths and weaknesses of the proposed approach. Finally, the open questions were aimed at investigating options for the development of the F-FAS.

7.2.2 The Participants

This section describes the participants selected for the evaluation and provides a short background on them. Because this research focuses on digital forensics, cybersecurity, and digital facial investigation fields, people with adequate

knowledge, background, and experience in this field were chosen. The evaluation methodology focused on academic people more than technical people because it was aimed at evaluating the research problem, experimental results, the novelty and effectiveness of the proposed F-FAS, and the weaknesses and strengths in the suggested system. Furthermore, technical people prefer to test tools or full programming system to give their opinions while academic people focus on theoretical and scientific concepts. Therefore, the suitable method was to search for people via the internet by looking into scientific conferences, committees, scientific journals, and professors or lecturers in educational institutions.

However, there was a list of candidates who were determined. There were about 30 people and the research aim was looking for 10 responses. The time needed for finishing this evaluation was two months from the beginning of September 2018 until the end of November 2018 process (sending e-mails and waiting for responses). Finally, only seven participants responded positively to the e-mail and confirmed their availability. These final participants provided valuable information, sufficient for the evaluation objective. The following list is of the final participants' details, which provides brief information about them and their research interests.

- **Professor Iain Sutherland – Noroff - Norway**

Professor of digital forensics at Noroff educational institution in Kristiansand, Norway. Email lain.sutherland@noroff.no and the Skype interview was on September 17, 2018.

Professor Sutherland is a recognised expert in the area of computer forensics and data recovery. He has authored numerous articles ranging from forensics practice and procedures to forensic tool development. In addition to being actively

involved in research in this area, Professor Sutherland has taught computer forensics at both undergraduate and postgraduate levels. He has advised on forensic problems to police forces and commercial organisations. In addition, he has many papers in the aforementioned field and has supervised students' theses.

- **Dr Robert Hegarty – Manchester Metropolitan University – UK**

Senior lecturer at Manchester Metropolitan University in Manchester, UK. Email: r.hegarty@mmu.ac.uk and the Skype interview was on September 21, 2018.

Dr Hegarty is a senior lecturer in cyber security and digital forensics from the School of Computing, Mathematics, and Digital Technology at Manchester Metropolitan University (UK). He has taught computer forensics, security fundamentals, and file system forensics and analysis at an undergraduate level and further advanced network security at a postgraduate level. In addition, Dr Hegarty has interest in signature detection, cloud computing, and the Internet of Things. He has published several papers in the aforementioned fields and presented related research outcomes at various international conferences.

- **Dr Mo Adda - University of Portsmouth – UK**

Principal lecturer at Portsmouth University in Portsmouth, UK. Email: mo.adda@port.ac.uk and the Skype interview was on September 25, 2018.

Dr Mo Adda has been a principal lecturer in advanced networks, digital forensics, and mobile forensics at the University of Portsmouth since 2002. In addition, he is a course leader of forensics information technology. Dr Adda is interested in different research areas such as multithreaded architectures, mobile networks, and business process modelling, parallel and distributed processing, wireless

networks and sensor networks, network security, embedded systems, simulation and modelling, and mobile intelligent agent technology. Therefore, he has published many papers in the aforementioned fields, presented related research outcomes in various international conferences, and supervised on four of doctoral theses.

- **Professor Yin Pan - Rochester Institute Of technology – USA**

Professor at Rochester Institute of Technology in Rochester, USA. Email: yin.pan@rit.edu and the Skype interview was on September 27, 2018.

Yin Pan is a professor in the Computing Security Department. Dr Pan teaches both undergraduate and graduate courses in digital forensics. She holds four US patents in the areas of network quality of services, voice over IP, and artificial intelligence. Dr Pan has been actively involved in the IT security area, especially in security audits and computer forensics. Her current research interests include game-based digital forensics and memory-based malware detection using machine learning. She has published over 45 papers and presentations in research conferences and journals.

- **Professor Andrew Jones - University of Hertfordshire – UK**

Principal lecturer at the University of Hertfordshire in Hertfordshire, UK. Email: a.jones26@herts.ac.uk and the face-to-face interview was on October 3, 2018.

Andrew Jones is currently a professor at the Centre for Computer Science and Information Research/School of Computer Science at the University of Hertfordshire. Since 2002, after leaving the defence environment, he has had experience working as a manager, a researcher, as well as an analyst in the area of information warfare and computer crime at a defence research establishment

at the University of South Wales. Prof. Jones gave lectures during his scientific life on different subjects such as network security and computer crime, further researching on the threats to information systems and computer forensics. He has authored seven books on topics including information warfare, risk management, and digital forensics and cybercrime, and has published numerous papers on the aforementioned subjects.

- **Professor Bill Stackpole - Rochester Institute of Technology – USA**

Professor at Rochester Institute of Technology in Rochester, USA. Email: bill.stackpole@rit.edu and the Skype interview was on October 12, 2018.

Bill Stackpole is a professor in the Computing Security Department. He teaches various undergraduate and graduate courses in digital forensics and security. Since joining RIT in 2001, Prof. Stackpole has been actively involved in the IT security area, especially in computer forensics, penetration testing, and security competitions. He has interests in different areas including mobile security and forensics, attack trees, and mobile malware mitigation. He has published papers in research conferences and journals.

- **Professor Golden G Richard III - Louisiana State University, USA**

Professor at Louisiana State University (LSU) in Louisiana-USA. Email: golden@cct.lsu.edu and the Skype interview was on October 18, 2018.

Golden G. Richard III is a professor of computer science and engineering at the Louisiana State University and associate director for cybersecurity at the Centre for Computation and Technology (CCT). He teaches different subjects, such as memory forensics, reverse engineering and malware analysis, and operating systems. Prof. Richard is a computer security expert and a fellow of the American

Academy of Forensic Sciences with over 35 years of practical experience in computer systems and computer security. His primary research interests are malware analysis, reverse engineering digital forensics, memory forensics, and operating systems internals. Hence, he has published numbers of papers, books, and chapters in the aforementioned areas and supervised numbers of MSc and PhD students' theses.

7.3 Interviewees' Responses' Evaluation

The evaluation questions were designed in such a manner that intended to explore facial recognition in digital forensics in terms of its problem validity, and suggestion solution efficiency, reliability, and usability. Moreover, it determines the system's limitations and illustrates the strengths of the proposal and further asks an open question that aims to improve the work.

This section presents and discusses the feedback for the questions from experts in detail. To make comparisons between experts' feedback, the answers for the same question were discussed and analysed (question by question). This way of presentation permitted more comprehensive methods to make the evaluation.

7.3.1 Thoughts on the Research Problem

This section of the questionnaire requires that respondents give thoughts on the validity of the research problem. Whether experts believe that facial recognition considers one of the forensic investigation challenges? All the experts agreed on the validity of the research problem.

Professor Sutherland agreed the project identified an appropriate problem that is required to be investigated in terms of digital forensics. He believes this research area still needs more work because of an increasing amount of photo footage that is captured from different electronic devices such as CCTV and mobiles. In

addition, the issue of photos, such as low quality and poor lighting (that have been identified in the literature review chapter) still present a challenge for investigators.

Moreover, Professor Pan believes the research problem is useful because there are case studies that still face complexities in solving it using the technology of facial recognition in digital forensics (i.e., some cases mentioned in Chapter 1). Dr Rob saw the research problem was defined and described clearly and he absolutely agreed the subject problem is valid.

Dr Adda stated that many facial recognition tools have been generated and used in investigation processes, but it is still considered an interesting area to explore and develop.

Professor Stackpole explained that unconstrained facial recognition, especially in crowds, is considered as one law enforcement challenge. For instance, terrorist attacks like the Boston bomber (described in Chapter One),. Therefore, there is still a need for facial recognition that covers some of these issues.

Finally, Professor Richard certainly agreed the project's problem is an important challenge in terms of the current terrorist threat posed to the world. In addition, he added that he observed some recognition systems a long time ago but the result was quite unpromising. He liked that this research suggests addressing these limitations.

7.3.2 Thoughts on Current Facial Recognition Systems

When the participants were asked about their opinions on the current facial recognition system in the forensic environment, the response was varied.

Some of the experts did not have any experience with current facial recognition tools but they found this subject interesting to explore, as noted by specialists in the digital forensic field, Professor Pan and Professor Sutherland.

Dr Adda was wondering about the proposed F-FAS and how it differs from other available facial recognition tools. He suggested that it is better to draw comparisons to make this point understandable.

Dr Hegarty is of the opinion that facial recognition tools have a great effect on different applications, such as mobile authentication (i.e., Apple) and in football stadiums (by the police in the UK), but that it needs more improvement. Moreover, Professor Jones agreed that many facial recognition tools are used currently by law enforcement agencies but are not adequate and there is a lot of work required in this field. Furthermore, he found the suggested tool promises combining more than one investigation requirement.

Professor Stackpole was happy with the results of some facial recognition tools that he used, such as “tagging people on Facebook” and “Google”, and he found they were effective but he showed concern in the aims of using these tools in forensic investigations. Similarly, Professor Richard’s experience in forensic video recognition made him believe available tools could work well under ideal circumstances like one or two facial recognition challenges but it is still facing some limitations.

7.3.3 Thoughts on the Multi-Algorithmic Fusion Experiment

It is interesting to note that all the experts agreed that the multi-algorithmic fusion approach is not only novel but also an interesting contribution in facial identification techniques. Additionally, they were optimistic about the experimental results, which looked promising.

Moreover, a few of the experts, such as Professors Pan and Jones, were impressed by the experimental results and were surprised to notice how the accuracy was improved. Similarly, Professor Sutherland observed the use of multiple algorithms for the sake of facial identification was comparatively better than using the individual algorithm. Furthermore, this will lead to the adoption of new facial recognition techniques to seek better outcomes during the investigation phase. Dr R Hegarty's point of view endorsed Professor Sutherland's ratification. In addition, he supplemented the previous argument by noting that the technique used in this research considers smart technology to bring the best attributes of the system with the inclusion of Amazon, Microsoft, and others to save time to perform similar research carried out by these commercial systems.

Although Dr Adda believes in the good contribution of this approach, he also suggested another technique, known as the hybrid technique, which could be used instead of the fusion technique. He recommends delivering the output of the first algorithms to the second ones and continuing this process until the end to get better results. This recommendation would be more suitable if the multi-algorithmic approach adopted features used rather than a decision in facial recognition and it could be investigated in the future.

Moving further, Professor Stackpole argued that the experimental results, which appeared in the fusion approach, were more effective than any other individual algorithms (Neurotechnology, Amazon, and Microsoft) for both databases that were used (CAS-PAL-R1 and Collection Realistic). Moreover, he stated that Google's facial recognition system could be used for future research because he liked how it works and Professor Jones suggested this.

Professor Richard believes fusing multiple algorithms in facial recognition is a vital approach because it is not obvious that any one of these approaches works perfectly in all cases. In addition to this, he also pointed out that probably, there are many critics who assume using multiple algorithms in one system costs more money. His opinion on this concern was: if a forensic organisation saves one-third of the money and then the terrorist or the kidnapper runs away, will this be considered as a good outcome?

7.3.4 The Architecture of The F-FAS realisable in Term of Digital Forensics

This question asks experts their opinion in the proposed F-FAS architecture and whether it is reliable to use it in forensic investigation. Interestingly, almost all experts were optimistic regarding this and the overall response to this question was positive.

All the experts answered that the F-FAS architecture is considered a valuable contribution to the domain of digital forensics investigation (i.e., in the facial identification field). Professor Pan agreed that this architecture from a computer forensic perspective covers most forensic investigations' fundamental requirements from the acquisition phase through analysis to reporting. Professor Lain thinks this architecture solves an interesting problem in digital forensics challenges. In addition, it has numerous useful features and functions. Furthermore, it appears to be quite a sensible approach in reality.

Dr Adda expressed that the F-FAS architecture appears to be a good system that intends to facilitate investigators in all essential things required to identify suspects' identities. In addition, Professor Jones did not have any specific issue with it. According to Professor Stackpole, as the proposed system includes law enforcement aspects, therefore, it would have realistic usage. He also explains his concerns that probably this proposed system does not work when it has a task

to detect faces from a large crowd of people, which represents a critical issue in this field. Although, this tool or most forensic analysis tools will be able to assist investigators in making decisions, ultimately, the final decision requires human interaction. Similarly, Professor Richard found this system architecture very promising, usable, and better than some projects developed by graduate students that he saw earlier but who had poor utility.

7.3.5 The F-FAS Capabilities to Reduce Effort and Time for Data Analysis

According to one of the F-FAS objectives that focuses on time reduction, as well as a cognitive effort for data analysis and finds the relationship between artifacts; most of the experts believed this aim would be achievable when this tool was used. Professor Lain believed the tool provides adequate information to investigators to further verify data analysis functions, for instance, a social network that permits testing and analyzing massive data for the sake of generating a final decision. Likewise, Dr Hegarty points out that how this tool recognises faces and relationships between people, as well as tracks them on the map in a way that it would assist in terms of time and cognitive effort in the investigation. As a matter of fact, currently, the investigation process looks manually at the artifact, therefore, Professor Jones considered it better to say this tool provides sufficient capabilities to the investigator and absolutely reduces the effort. But, on the other hand, he was not certain about a reduction in time. Contrary to this, Professor Stackpole agrees this tool would expedite the process of identifying individuals but he was also concerned about whether this tool could find the relevant evidence. Professor Richard appreciates the analysis ability of the F-FAS, which allows for immediately understanding the social network between suspects.

7.3.6 Effectiveness, Reliability, and Usability of the F-FAS Functions

While answering this question, the experts provided their views regarding the effectiveness, reliability, and usability of each function included in the F-FAS. All experts acknowledged this tool is a great effort to fill the gap of using facial recognition functions within the digital forensic domain and to do a thorough analysis of the results. Professor Pan agreed on the utility of this function but believed that, probably, the reliability depends on a database that can evaluate this tool. Dr Hegarty and Dr Adda were of the view there is no issue regarding the effectiveness of the function and its flexibility because it continues to work if there is no photo matching in the available information (using soft biometrics, such as age and gender). Because of the facial recognition based on the multi-algorithm fusion approach, Professor Jones and Richard considered it better if it presents the algorithm's results individually and allows the investigator to make a decision.

Regarding the facial social network analysis function, all experts consider this function as the most significant of the F-FAS and they were impressed with this function's ability. Professors Stackpole, Jones, and Richard thought this part was the most powerful and they had no doubt in its ability to be an effective function in the application. In Professor Pan's view, this function would be useful in digital forensics investigations because it will build a relationship network between available pictures but that the reliability depends on the available artifact resources fed to it. Dr Hegarty's perspective in this function is that it appears to give excellent utility demonstration just like the reliability sound. Dr Adda suggested this function should be developed to consider the habit while building the relationship, such as how many times the picture has been taken and why it has been taken (i.e., party, holiday) and this could be, then, investigated in the future.

Additionally, geolocation was another function the F-FAS provided to investigators. According to Dr Hegarty, this function is beneficial across a large dataset as it allows investigators to track suspects' movements. Another utility in Professor Pan's view is that this function could be more effective if the suspect's face is known but CCTV data could be searched daily to find any matches in order to determine the suspect's current location. According to Professors Jones and Stackpole, since geo location depends on the metadata information of the picture file but some pictures do not have this information type, the effectiveness, usability, and reliability of this function would be stopped. On the other side, Professor Richard believes this function faces a challenge in the forensic investigation and depends on the data collected and Professor Sutherland considers it could allow an investigator to manually enter the geolocation, which could assist in improving this function's utility.

Another function provided by the F-FAS is the facial modification function. Professor Pan represented this function like a fuzzy hash signature so it could be modified a little according to the face shape (i.e., add accessories and ageing) to improve the matching performance but it would not work in some cases. Therefore, it depends on the ability of methods and the accuracy of the facial modification. Dr Hegarty and Profesor Jones consider this function's feature would be effective in the investigation process and would be valuable to facilitate investigators to imagine what the suspect's face would look like if some modifications were made to it. Moreover, Professor Richard, Sutherland, and Stackpole agree this function is interesting but they have concerns about the effectiveness and reliability of the function and how could it facilitate the forensic investigation in terms of improving the facial matching accuracy. Hence, they proposed the function could be implemented after further testing.

The reporting is a key function in any forensic tool, therefore, all experts agreed on the utility and reliability of this function. Professor Sutherland explained that the reporting allowed investigators to go deep in the data or evidence found through the investigation process, so the effectiveness of this function depends on its features. According to Dr Hegarty and Professor Stackpole, the F-FAS generates the report with various details, such as to include some images of face matching, relationships on the social networks, and other things, which adds power to the tool's utility. In addition, Professor Pan appreciated the tool's ability to allow investigators to select the data required for the report, such as Bookmark data, rather than to present everything. The proposed F-FAS designs the reporting function to print or save the file as a .pdf but it is better to add another file format to save the report such as .csv or .doc (this is Professor Jones' point of view). Professor Richard did not have strong feelings about how the report looked and he preferred to add some probability or error rate about the matching results and regarding any results that give full details about how evidence is found.

7.3.7 Thoughts on the System Case Management

This question asked the experts about their opinion in the case management of the F-FAS. All the experts were happy with this suggestion. This area will, probably, forget some forensic tools or some organisations that did not design this in their tool. The main reason seems that it could cost them money if the management features are added but still Professor Sutherland considers the ability to monitor the case overview (i.e., active, inactive, and number) a very useful feature of the forensic tool. In addition, Dr Hegarty and Professor Pan saw it vital for forensic tools to provide role-based access control to check the authorisation and authentication of users. In addition, Professor Stackpole

considered it a good approach to provide different roles, such as administrator and investigator, which would be useful when used by law enforcement groups.

7.3.8 Strengths and Weaknesses of the F-FAS Approach

The most significant part of the evaluation was to determine the strengths and weaknesses of the research based on the participating perspectives. In terms of strengths, the overall view for the research was positive and it is considered a good suggestion tool that would be promising in the digital forensics field. This response has been noted from all the experts. Moreover, the architecture is clearly designed and provides several novel and strong components, such as the visualisation aspect of facial social networks and reporting (Dr Hegarty and Professor Stackpole) and the multiple-algorithm technique (Professor Pan and Professor Richard). Professor Jones considered the proposed F-FAS is one of the few tools that aims to cover the forensic investigation requirements and facial recognition analysis. Nevertheless, there were several limitations, as the experts called it, rather than weaknesses of this research. The limitations of some algorithms that would be used in the F-FAS to identify faces led to failure to find the terrorists but this weakness is not in the research's contribution (Professor Richard). In addition, the proposed system should have some determinants about the image quality that are also accepted by the system and a threshold for the facial matching technique. This could be modifiable based on the input requirements (Professor Stackpole). Professor Jones' view on this aspect is that the proposed system (F-FAS) should be fully developed and integrated. In this way, digital forensics investigators can take full advantage of the research and not add new functions until the current system's implementation is finished. Dr Hegarty found there is some deficiency of sufficient information about how the data is stored securely on the server or any container.

7.3.9 Suggested Enhancement

This section summarises some of the significant points suggested by the expert participants. Professor Pan suggested developing the forensic part by adding a data-carving function and finding deleted files like current forensic tools do, such as FTK and Encase. Professor Jones argues that the focus on the current component of the F-FAS tool is developed by it and makes it work perfectly to achieve exhaustive results and it is better to add a new function. In addition, another suggestion from Dr Hegarty is to increase the current cognitive algorithms to counter more challenges such as facial recognition algorithms. Professor Sutherland suggested permitting investigators to go through the raw data and allowing manual selection, which could be a positive addition to the research. The suggestion of Dr Adda was to add some statistics about how many times any particular picture was taken in the same place and this could be useful to present in the reporting. Professor Stackpole saw that the tools work in real-time and that the offline case is considered a big challenge. In this area, further research could be conducted in the future.

7.3.10 Further comments

This is the final question in the research evaluation, which asks the experts whether they have other comments. Therefore, this section presents their opinions on the overall system.

- Professor Sutherland: “Very nice PhD project. That’s excellent. You’re obviously in the process of completing this and writing it up. I think you have a very nice tool that you’ve developed. That’s excellent.”
- Dr Hegarty: “I like the way it brings together things and techniques and I particularly like the visualisations in the results. That demonstrates the

relationship between the different faces on the social networks. I thought that was very good."

- Dr Adda: "I think it is a good job—a good piece of work. You have done it."
- Dr Pan: "Actually. It's quite an impressive piece."
- Dr Jones: "That's an interesting piece of work and I like it. I think that's a good evaluation within working as forensically found and specifically facial recognition. Further, I advise that you do not lose this value when writing the thesis."
- Dr Stackpole: "I think it is a really interesting project. I think that integrating this into a forensic toolkit would be a very powerful addition and thank you for giving me the opportunity to look at it. It's not something that I would have seen prior to this."
- Dr Richard: "I think it's very impressive work."

7.4 Conclusion

It was imperative to evaluate the research work and to receive unbiased and objective feedback from different specialists and academics in the digital forensics field. This evaluation aims to evaluate the research study in terms of the research problem, experimental results, F-FAS architecture and prototype design, system capabilities and limitations, and any suggestion enhancement that could improve the system's usability. However, this chapter describes the entire evaluation process, which begins with generating a video demo, farming queries that aim to extract the evaluation's relevance from the experts. Natural questions are prepared in a manner to cover academic research, law enforcement, and technical requirements. All experts who responded to the evaluation query were contacted and the interviews were conducted over Skype.

The general overview of the experts interviewed was quite positive and interesting and it was worth the effort and time to perform this task. A definite need for the F-FAS has been presented and it promises to fill the gap of using facial identification in forensic investigations. Most of the participants were impressed by the suggested data analysis functions, such as a facial social network. However, some experts found it difficult to judge the proposed tool's reliability unless its concept could be further tested in large and real data environments. In addition, some of the experts suggested focusing on the development of the F-FAS' components to make an integrated forensic approach. Therefore, this research should be taken further based on the feedback received to enhance the capabilities of the F-FAS.

8 Conclusion and Future Work

This chapter illustrates an overview of the entire research work by providing highlights of the achievements, as well as summarising the limitations, which pave the way to conduct further research in this particular area. This research is intended to contribute to the development of techniques that facilitate the automatic facial recognition of persons within the forensic investigation context. To attain this objective, a detailed examination of contemporary state-of-the-art techniques was conducted to determine the existing gap required to be addressed and the most appropriate approaches were investigated to tackle the issue. In addition to this, extensive experiments were considered by using different methodologies to prove the defined concept and, finally, specialists in the field evaluated the result.

8.1 Contributions and Achievements of the Research

The main achievements of this research are:

- Investigating the facial recognition domain within digital forensic investigation from various aspects, such as components of the system, techniques, performance measures, challenges, and studies that have already published their issues. Based on this, the research provides a comprehensive background of this domain.
- Developing an understanding of the context of the current state of the art in facial recognition in terms of finding a problem and then suggested algorithms accordingly. This helped explore the gap in this domain, which has been investigated in the literature. This investigation included the method to solve the facial recognition problems by focusing on the techniques or algorithms used to recognise faces and database types that are, specifically, suitable to study these issues. In addition, the research

identified the challenges posed to the digital forensic investigator in the facial recognition domain.

- Undertaking and modelling a baseline set of experiments to understand the nature of a number of facial recognition systems by using one of the public facial databases to determine their potential contribution to solving facial recognition issues.
- Replication of the previous set of experiments with a more challenging dataset, which was considered using realistic forensic scenario images acquired in uncontrolled environments (i.e., light, and resolution) and uncontrolled face actions (i.e., pose, expression, and accessories). This dataset was collected from the internet to validate what the actual facial recognition system's performance would be in practice.
- Modelling and developing a model using the multi-algorithmic fusion method aimed at investigating whether employing a fusion mechanism that encompasses all available facial recognition algorithms improves the performance of the individual algorithm.
- Proposing a novel Facial-Forensic Analysis System (F-FAS) architecture in order to address the gap in this domain. The proposed tool consists of several components such as acquisition and data examination, which include a series of pre-processing techniques to isolate multimedia files and pictures. By using the integrated components, the proposed tool allows investigators to perform a facial recognition approach to identify a suspect's identity. Furthermore, it provides analyses of numerous functions, such as geolocation, facial modification, and a novel facial social network function. As the objective of this proposed system is to provide a novel forensic tool, using the reporting and case management (AAA

function) were preferred. The F-FAS is promising for the next generation of the digital multimedia forensic tool.

- Developing a functional demonstration of the prototype that reflects the newly proposed novel F-FAS framework to seek a concrete understanding of how this approach works in practice and an illustration of its functional working.
- Evaluation of the whole PhD research by seeking valuable feedback and suggestions from various experts in the field.

8.2 Limitations of the Research

Although this research has several achievements, numerous issues were observed, which must be taken into account. The research limitations are described below.

- The experimental dataset was limited in terms of the number of users (subjects). In an ideal situation, a larger number of subjects would provide more measures of the algorithm's performance that could be actually achieved in practice.
- The experiments, which aimed to recognise faces, did not have enhanced facial images before the images were sent to the recognition system. While this has the capability of improving the matching performance, which can be used for the facial recognition system, this research intends to study the ability to match algorithms to recognise faces with image issues.
- The multi-algorithmic fusion experiment did not use another fusion approach (i.e., matching level fusion). However, the used approach (decision level fusion) was the most appropriate one to consider commercial algorithms, which were used because the system's retraining

was not a prerequisite for them. Therefore, the results' accuracy could be further improved if extra information from matching systems is used.

- The focus in the final prototype has been on front-end design. Developing an integrated trust platform for the fully operational facial forensic analysis prototype would have provided better insight about the effectiveness of using the F-FAS and would have allowed for evaluating specialised operational aspects required for a successful investigation process' scalability and reliability.
- To obtain fair and precise judgment on the validity of the research, as well as on the proposed approach to deal with the problem discussed, observations were only taken into account from experts in the field. This fact contributes to further restrictions in terms of a large number of participants. Additionally, it was relatively difficult to approach these experts because of their busy schedules.

It is noteworthy that despite the limitations mentioned above, it is believed that research has made valid contributions to existing knowledge and has also become a source to provide sufficient evidence related to the concept of proposed ideas.

8.3 Scope for Future Work

This research contribution has advanced the field of facial recognition within digital forensics. However, various areas still exist for future work, specifically related to this research. These suggestions are listed below.

- Development of the matching algorithms in terms of identifying the suspect's identity using additional biometric samples such as the iris and voice to enhance the ability of the F-FAS.

- Further research could be investigated to develop the proposed system's architecture to consider the object recognition in pictures, such as to recognise a specific car and analyse the background of pictures to further seek assistance in predicting the location in which the photo was taken.
- Developing the suggested F-FAS to work online and consider video analyses instead of offline (static image) analyses to add more flexibility to the research.
- Develop the suggested facial social network approach to consider objects (as suggested in the second point) and build a network of query objects.

8.4 The Future of Facial Recognition in Digital Forensics

With the mounting number of images and videos available, multimedia evidence has become a key part of criminal investigations. The large increase in the volume of image and video data has a direct impact on the time and cost of investigations, unlike manual methods where immense effort is required to identify faces. In addition, forensic investigators require a range of forensic analyses to enable them to identify relevant evidence more efficiently. Notwithstanding, there is an ample number of researchers who have investigated the challenges of facial recognition in digital forensics. Furthermore, some facial recognition tools have been used in digital investigations to solve these issues. Moreover, these researches are professionally inadequate in digital investigation areas, such as automatic facial recognition and evidence extraction and analysing and correlating the derived data. As a result, this research has suggested a novel approach (F-FAS) as a new branch of digital forensics that allows investigators to identify faces and analyse the results effectively and accurately while dealing with a large number of images in an automated and proper way.

However, there are many challenges that are still posed for future research. For instance, further research is needed to develop the prototype and evaluate the performance and effectiveness of the F-FAS to verify the ability to meet all the key requirements. Furthermore, forensically, little work has been undertaken using object and facial recognition to better understand the context of images.

References

!!! INVALID CITATION !!! (Mitra et al., 2016, Ross and Jain, 2003).

ABCNEWS. 2013. *Missing Children in America: Unsolved Cases* [Online]. abc News. Available: http://abcnews.go.com/US/missing-children-america-unsolved-cases/story?id=19126967# [Accessed 10 May 2015].

ACPO 2012. ACPO Good Practice Guide for Digital Evidence. *Metropolitan Police Service, Association of chief police officers, GB.*

ALI, T., SPREEUWERS, L. & VELDHUIS, R. 2012. Forensic face recognition: A survey. *Face Recognition: Methods, Applications and Technology.* 9.

ANTHONY HO & LI, S. 2015. *Handbook of digital forensics of multimedia data and devices,* UK, John Wiley & Sons.

ARBAB-ZAVAR, B., XINGJIE, W., BUSTARD, J., NIXON, M. S. & LI, C.-T. 2015. On forensic use of biometrics with face and ear recognition. *Handbook of Digital Forensics of Multimedia Data and Devices.*

ASTHANA, A., MARKS, T. K., JONES, M. J., TIEU, K. H. & ROHITH, M. Fully automatic pose-invariant face recognition via 3D pose normalization. Computer Vision (ICCV), 2011 IEEE International Conference on, 2011. IEEE, 937-944.

AYERS, D. 2009. A second generation computer forensic analysis system. *digital investigation,* 6**,** S34-S42.

BAKER, S. & KANADE, T. 2002. Limits on super-resolution and how to break them. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 24**,** 1167-1183.

BARRY, E. 2018. *From Mountain of CCTV Footage, Pay Dirt: 2 Russians Are Named in Spy Poisoning* [Online]. The New York Times. Available: https://www.nytimes.com/2018/09/05/world/europe/salisbury-novichok-poisoning.html [Accessed 2018].

BBCHOME. 2007. *Madeleine McCann: Timeline* [Online]. UK: BBC. Available: http://www.bbc.co.uk/leicester/content/articles/2007/05/10/madeleine_mccann_round_up_feature.shtml [Accessed 14 May 2015].

BELHUMEUR, P. N., HESPANHA, J. P. & KRIEGMAN, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Yale University New Haven United States.

BEST-ROWDEN, L., HOOLE, Y. & JAIN, A. Automatic face recognition of newborns, infants, and toddlers: A longitudinal evaluation. 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), 2016. IEEE, 1-8.

BEST-ROWDEN, L. & JAIN, A. K. 2018. Longitudinal study of automatic face recognition. *IEEE transactions on pattern analysis and machine intelligence,* 40**,** 148-162.

BHAT, F. & WANI, M. A. 2015. Elastic Bunch Graph Matching Based Face Recognition Under Varying Lighting, Pose, and Expression Conditions. *Elastic,* 1.

BUCIU, I. Efficiency analysis of illumination correction methods for face recognition performance. Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference on, 2010. IEEE, 211-216.

BURT, D. M. & PERRETT, D. I. 1995. Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London B: Biological Sciences,* 259**,** 137-143.

CAMENT, L. A., GALDAMES, F. J., BOWYER, K. W. & PEREZ, C. A. 2015. Face recognition under pose variation with local Gabor features enhanced by Active Shape and Statistical Models. *Pattern Recognition,* 48**,** 3371-3384.

CARRIER, B. 2003. Defining digital forensic examination and analysis tools using abstraction layers. *International Journal of digital evidence,* 1**,** 1-12.

CARRIER, B. & SPAFFORD, E. H. An event-based digital forensic investigation framework. Digital forensic research workshop, 2004. 11-13.

CASEY, E. 2011. *Digital evidence and computer crime: Forensic science, computers, and the internet*, Academic press.

CHAI, X., SHAN, S., CHEN, X. & GAO, W. 2007. Locally linear regression for pose-invariant face recognition. *IEEE Transactions on Image Processing,* 16**,** 1716-1725.

CHEN, X.-H. & LI, C.-Z. Image quality assessment model based on features and applications in face recognition. Signal Processing, Communications and Computing (ICSPCC), 2011 IEEE International Conference on, 2011. IEEE, 1-4.

CHOI, S.-I. 2012. Face Recognition Under Illumination Variation Using Shadow Compensation and Pixel Selection. *Int J Adv Robotic Sy,* 9.

CHOI, S.-I., CHOI, C.-H. & KWAK, N. 2011. Face recognition based on 2D images under illumination and pose variations. *Pattern Recognition Letters,* 32**,** 561-571.

CLARKE, N. 2011. *Transparent user authentication: biometrics, RFID and behavioural profiling*, Springer Science & Business Media.

DE OLIVEIRA, A. E., MOTTA, G. H. M. B. & BATISTA, L. V. A multibiometric access control architecture for continuous authentication. Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on, 2010. IEEE, 171-171.

DECANN, B. & ROSS, A. Relating roc and cmc curves via the biometric menagerie. Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, 2013. IEEE, 1-8.

DU, M., SANKARANARAYANAN, A. C. & CHELLAPPA, R. 2014. Robust face recognition from multi-view videos. *Image Processing, IEEE Transactions on,* 23**,** 1105-1117.

DUGGAN, M. 2013. *Photo and Video Sharing Grow Online* [Online]. Pew Research Internet Project, Internet Science & Tech. Available: http://www.pewinternet.org/2013/10/28/photo-and-video-sharing-grow-online/ [Accessed 24/07/2016 2016].

EL-ABED, M. & CHARRIER, C. 2012. Evaluation of biometric systems. *New Trends and developments in biometrics***,** pp. 149-169.

FBI. 2015. *NEXT GENERATION IDENTIFICATION* [Online]. the FBI's Criminal Justice Information Services (CJIS). Available: https://www.fbi.gov/services/cjis/fingerprints-and-other-biometrics/ngi [Accessed 8-8-2016 2016].

FLYNN, P. J., JAIN, A. K. & ROSS, A. A. 2008. *Handbook of biometrics*, Springer.

FOOKES, C., LIN, F., CHANDRAN, V. & SRIDHARAN, S. 2012. Evaluation of image resolution and super-resolution on face recognition performance. *Journal of Visual Communication and Image Representation,* 23**,** 75-93.

GALLAGHER, S. 2013. *Why facial recognition tech failed in the Boston bombing manhunt* [Online]. UK: ars technica UK. Available: http://arstechnica.com/information-technology/2013/05/why-facial-recognition-tech-failed-in-the-boston-bombing-manhunt/1/ [Accessed 10 May 2015].

GAO, W., CAO, B., SHAN, S., CHEN, X., ZHOU, D., ZHANG, X. & ZHAO, D. 2008. The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* 38**,** 149-161.

GARFINKEL, S. L. 2010. Digital forensics research: The next 10 years. *digital investigation,* 7**,** S64-S73.

GARG, A. & CHOUDHARY, V. 2012. Facial expression recognition using principal component analysis. *Int. J. Sci. Eng. Res. Technol*.

GENG, X., ZHOU, Z.-H. & SMITH-MILES, K. 2007. Automatic age estimation based on facial aging patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 29**,** 2234-2240.

GEORGHIADES, A. S., BELHUMEUR, P. N. & KRIEGMAN, D. J. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence,* 23**,** 643-660.

GHAZALI, N. N. A. N., ZAMANI, N. A., ABDULLAH, S. N. H. S. & JAMESON, J. Super resolution combination methods for CCTV forensic interpretation. Intelligent

Systems Design and Applications (ISDA), 2012 12th International Conference on, 2012. IEEE, 853-858.
GONZALEZ-SOSA, E., FIERREZ, J., VERA-RODRIGUEZ, R. & ALONSO-FERNANDEZ, F. 2018. Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation. *IEEE Transactions on Information Forensics and Security,* 13**,** 2001-2014.
GROTHER, P. J., QUINN, G. W. & PHILLIPS, P. J. 2010. Report on the evaluation of 2D still-image face recognition algorithms. *NIST interagency report,* 7709**,** 106.
GUO, K., WU, S. & XU, Y. 2017. Face recognition using both visible light image and near-infrared image and a deep network. *CAAI Transactions on Intelligence Technology,* 2**,** 39-47.
HAN, H., OTTO, C. & JAIN, A. K. Age estimation from face images: Human vs. machine performance. Biometrics (ICB), 2013 International Conference on, 2013. IEEE, 1-8.
HAN, H., OTTO, C., LIU, X. & JAIN, A. 2015. Demographic estimation from face images: Human vs. machine performance. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* 37(6).
HILL, K. 2011. *Will The London Riots Be The Turning Point For Facial Recognition As A Crime-Fighting Tool?* [Online]. Forbes / Tech. Available: http://www.forbes.com/sites/kashmirhill/2011/08/09/will-london-riots-be-the-turning-point-for-facial-recognition-as-crime-fighting-tool/ [Accessed 11 May 2015].
HUANG, G. B., RAMESH, M., BERG, T. & LEARNED-MILLER, E. 2007. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.
HUANG, H. & HE, H. 2011. Super-resolution method for face recognition using nonlinear mappings on coherent features. *IEEE Transactions on Neural Networks,* 22**,** 121-130.
ISHIMOTO, M. & CHEN, Y.-W. Pose-robust face recognition based on 3D shape reconstruction. Natural Computation, 2009. ICNC'09. Fifth International Conference on, 2009. IEEE, 40-43.
JAIN, A., FLYNN, P. & ROSS, A. A. 2007. *Handbook of biometrics*, Springer Science & Business Media.
JAIN, A., NANDAKUMAR, K. & ROSS, A. 2005. Score normalization in multimodal biometric systems. *Pattern recognition,* 38**,** 2270-2285.
JAIN, A. K., FLYNN, P. & ROSS, A. A. 2008. Handbook of Biometrics. Springer.
JAIN, A. K., KLARE, B. & PARK, U. Face recognition: Some challenges in forensics. Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, 2011. IEEE, 726-733.
JAIN, A. K., KLARE, B. & PARK, U. 2012. Face matching and retrieval in forensics applications. *IEEE multimedia,* 19**,** 20.
JAIN, A. K. & LI, S. Z. 2005. *Handbook of face recognition*, Springer.
JAIN, A. K. & LI, S. Z. 2011. *Handbook of Face Recognition*, Springer.
JAIN, A. K. & ROSS, A. 2015. Bridging the gap: from biometrics to forensics. *Phil. Trans. R. Soc. B,* 370**,** 20140254.
JAIN, A. K., ROSS, A. & PRABHAKAR, S. 2004. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology,* 14**,** 4-20.
JENKINS, N. 2015. *245 million video surveillance cameras installed globally in 2014* [Online]. IHS Technology. Available: https://technology.ihs.com/532501/245-million-video-surveillance-cameras-installed-globally-in-2014 [Accessed 27/07/2016 2016].
JIA, K. & GONG, S. Multi-modal tensor face for simultaneous super-resolution and recognition. Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2005. IEEE, 1683-1690.
JUEFEI-XU, F., LUU, K., SAVVIDES, M., BUI, T. D. & SUEN, C. Y. Investigating age invariant face recognition based on periocular biometrics. Biometrics (IJCB), 2011 International Joint Conference on, 2011. IEEE, 1-7.

KEMELMACHER-SHLIZERMAN, I., SUWAJANAKORN, S. & SEITZ, S. M. Illumination-aware age progression. Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014. IEEE, 3334-3341.
KENT, K., CHEVALIER, S., GRANCE, T. & DANG, H. 2006. Guide to integrating forensic techniques into incident response. *NIST Special Publication*, 800-86.
KIM, K.-H., ZHANG, C., ZHANG, Z. & CHOI, S. Robust part-based face matching with multiple templates. Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, 2013. IEEE, 1-7.
KLARE, B. F., LI, Z. & JAIN, A. K. 2011. Matching forensic sketches to mug shot photos. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 33, 639-646.
KLONTZ, J. C. & JAIN, A. K. 2013. A case study on unconstrained facial recognition using the boston marathon bombings suspects. *Michigan State University, Tech. Rep,* 119, 120.
LEE, P.-H., HSU, G.-S., WANG, Y.-W. & HUNG, Y.-P. 2012. Subject-specific and pose-oriented facial features for face recognition across poses. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* 42, 1357-1368.
LI, A., SHAN, S., CHEN, X. & GAO, W. Maximizing intra-individual correlations for face recognition across pose differences. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009. IEEE, 605-611.
LI, S. Z. & JAIN, A. 2011. *Handbook of Face Recognition*, Springer Science & Business Media.
LI, Z., PARK, U. & JAIN, A. K. 2011. A discriminative model for age invariant face recognition. *Information Forensics and Security, IEEE Transactions on,* 6, 1028-1037.
LIAO, S., JAIN, A. K. & LI, S. Z. 2013. Partial face recognition: Alignment-free approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 35, 1193-1205.
LIN, F. C., FOOKES, C. B., CHANDRAN, V. & SRIDHARAN, S. 2005. Investigation into optical flow super-resolution for surveillance applications.
LING, H., SOATTO, S., RAMANATHAN, N. & JACOBS, D. W. 2010. Face verification across age progression using discriminative methods. *Information Forensics and Security, IEEE Transactions on,* 5, 82-91.
LOHIYA, R. & SHAH, P. 2015. Face Recognition Techniques: A Survey for Forensic Applications. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* 4.
LUAN, X., FANG, B., LIU, L., YANG, W. & QIAN, J. 2014. Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion. *Pattern Recognition,* 47, 495-508.
MAHALINGAM, G. & KAMBHAMETTU, C. Age invariant face recognition using graph matching. Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, 2010. IEEE, 1-7.
MARSHALL, A. M. 2009. *Digital forensics: Digital evidence in criminal investigations*, John Wiley & Sons.
MARTINEZ, A. M. 1998. The AR face database. *CVC Technical Report24*.
MEDEIROS, J. 2018. *How police super-recognisers cracked the Russian novichok case* [Online]. WIRED. Available: https://www.wired.co.uk/article/salisbury-novichok-poisoning-russia-suspects [Accessed 2018].
MIN, R. & DUGELAY, J.-L. Inpainting of sparse occlusion in face recognition. Image Processing (ICIP), 2012 19th IEEE International Conference on, 2012. IEEE, 1425-1428.
MOEINI, A. & MOEINI, H. 2015. Real-world and rapid face recognition toward pose and expression variations via feature library matrix. *Information Forensics and Security, IEEE Transactions on,* 10, 969-984.
NABATCHIAN, A., ABDEL-RAHEEM, E. & AHMADI, M. An efficient method for face recognition under illumination variations. High Performance Computing and Simulation (HPCS), 2010 International Conference on, 2010. IEEE, 432-435.

NELSON, B., PHILLIPS, A. & STEUART, C. 2015. *Guide to computer forensics and investigations*, Cengage Learning.
NIST. 2008. *Secure Hash Standard (SHS)* [Online]. The National Institute of Standards. Available: http://www.nist.gov/manuscript-publication-search.cfm?pub_id=901372] [Accessed 2018].
PAL, R. & GAUTAM, A. K. Age Invariant Face Recognition using multiclass SVM. Electrical, Electronics, Signals, Communication and Optimization (EESCO), 2015 International Conference on, 2015. IEEE, 1-6.
PALMER, G. A road map for digital forensic research. First Digital Forensic Research Workshop, Utica, New York, 2001. 27-30.
PARK, U. & JAIN, A. K. 2010. Face matching and retrieval using soft biometrics. *Information Forensics and Security, IEEE Transactions on,* 5**,** 406-415.
PARK, U., TONG, Y. & JAIN, A. K. 2010. Age-invariant face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 32**,** 947-954.
PERNER, P. NPRF—Our view to novel forensic multimedia data analysis. 2014 22nd Signal Processing and Communications Applications Conference (SIU), 2014. IEEE, 2285-2290.
PHILLIPS, P. J., WECHSLER, H., HUANG, J. & RAUSS, P. J. 1998. The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing,* 16**,** 295-306.
POH, N., CHAN, C., KITTLER, J., UAM, J. F. & UAM, J. G. 2011. D3. 3: Description of Metrics For the Evaluation of Biometric Performance. *Evaluation,* 1.
POLLITT, M. M. An ad hoc review of digital forensic models. Systematic Approaches to Digital Forensic Engineering, 2007. SADFE 2007. Second International Workshop on, 2007. IEEE, 43-54.
PORTER, G. & DORAN, G. 2000. An anatomical and photographic technique for forensic facial identification. *Forensic science international,* 114**,** 97-105.
PRABHAKAR, S., PANKANTI, S. & JAIN, A. K. 2003. Biometric recognition: Security and privacy concerns. *IEEE Security & Privacy***,** 33-42.
RAGHAVAN, S. 2013. Digital forensic research: current state of the art. *CSI Transactions on ICT,* 1**,** 91-114.
REITH, M., CARR, C. & GUNSCH, G. 2002. An examination of digital forensic models. *International Journal of Digital Evidence,* 1**,** 1-12.
REN, C.-X., DAI, D.-Q. & YAN, H. 2012. Coupled kernel embedding for low-resolution face image recognition. *Image Processing, IEEE Transactions on,* 21**,** 3770-3783.
ROSS, A., NANDAKUMAR, K. & JAIN, A. K. 2008. Introduction to multibiometrics. *Handbook of biometrics.* Springer.
ROSS, A. A., NANDAKUMAR, K. & JAIN, A. K. 2006. *Handbook of multibiometrics*, Springer Science & Business Media.
SAINI, M. & KAPOOR, A. 2016. Biometrics in forensic identification: Application s and challenges. *J Forensic Med,* 1**,** 2.
SCHULTZ, R. R. & STEVENSON, R. L. 1996. Extraction of high-resolution frames from video sequences. *Image Processing, IEEE Transactions on,* 5**,** 996-1011.
SHARMA, A., DUBEY, A., JAGANNATHA, A. & ANAND, R. 2010. Pose invariant face recognition based on hybrid-global linear regression. *Neural Computing and Applications,* 19**,** 1227-1235.
SHEN, J. & STRANG, G. 1998. Asymptotics of daubechies filters, scaling functions, and wavelets. *Applied and Computational Harmonic Analysis,* 5**,** 312-331.
SHOICHET, C., BOTELHO, G. & BERLINGER, J. 2016. *Brothers ID'd as suicide bombers in Belgium, 1 suspect 'on the run'* [Online]. CNN. Available: http://edition.cnn.com/2016/03/23/europe/brussels-investigation/ [Accessed 26/07/2016 2016].
SIM, T., BAKER, S. & BSAT, M. The CMU pose, illumination, and expression (PIE) database. Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on, 2002. IEEE, 53-58.

SINGH, R., VATSA, M., ROSS, A. & NOORE, A. 2007. A mosaicing scheme for pose-invariant face recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* 37**,** 1212-1225.

SULTANA, M., GAVRILOVA, M. & YANUSHKEVICH, S. Expression, pose, and illumination invariant face recognition using lower order pseudo Zernike moments. Computer Vision Theory and Applications (VISAPP), 2014 International Conference on, 2014. IEEE, 216-221.

SUN, Y., WANG, X. & TANG, X. Hybrid deep learning for face verification. Proceedings of the IEEE International Conference on Computer Vision, 2013. 1489-1496.

SUNGATULLINA, D., LU, J., WANG, G. & MOULIN, P. Multiview discriminative learning for age-invariant face recognition. Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, 2013. IEEE, 1-6.

TEH, C.-H. & CHIN, R. T. 1988. On image analysis by the methods of moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 10**,** 496-513.

TELEGRAPH, T. 2016. *Madeleine McCann: Police release new "age progression" image* [Online]. The Telegraph. Available: http://www.telegraph.co.uk/news/uknews/crime/9226178/Madeleine-McCann-Police-release-new-age-progression-image.html?frame=2202507 [Accessed 14/7/2016 2016].

TIWARI, S., SINGH, A. & SINGH, S. K. Can face and soft-biometric traits assist in recognition of newborn? Recent Advances in Information Technology (RAIT), 2012 1st International Conference on, 2012. IEEE, 74-79.

TOMASI, C. & KANADE, T. 1992. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision,* 9**,** 137-154.

TOME, P., VERA-RODRIGUEZ, R., FIERREZ, J. & ORTEGA-GARCIA, J. 2015. Facial soft biometric features for forensic face recognition. *Forensic science international,* 257**,** 271-284.

WANG, D., OTTO, C. & JAIN, A. K. 2017. Face search at scale. *IEEE transactions on pattern analysis and machine intelligence,* 39**,** 1122-1136.

WENG, R., LU, J., HU, J., YANG, G. & TAN, Y.-P. Robust feature set matching for partial face recognition. Proceedings of the IEEE International Conference on Computer Vision, 2013. 601-608.

XU, X., LIU, W. & LI, L. 2014. Low Resolution Face Recognition in Surveillance Systems. *Journal of Computer and Communications,* 2**,** 70.

YADAV, S. 2011. Analysis of Digital Forensic and Investigation. *International Journal of Computer Science & Information Technology,* 1**,** 171-178.

YEN, P.-H., YANG, C.-H. & AHN, T.-N. Design and implementation of a live-analysis digital forensic system. Proceedings of the 2009 international Conference on Hybrid information Technology, 2009. ACM, 239-243.

YI, D., LEI, Z. & LI, S. Z. Towards pose robust face recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013. 3539-3545.

YUSOFF, Y., ISMAIL, R. & HASSAN, Z. 2011. Common phases of computer forensics investigation models. *International Journal of Computer Science & Information Technology (IJCSIT),* 3**,** 17-31.

ZENG, X. & HUANG, H. 2012. Super-resolution method for multiview face recognition from a single image per person using nonlinear mappings on coherent features. *Signal Processing Letters, IEEE,* 19**,** 195-198.

ZHANG, X. & GAO, Y. 2009. Face recognition across pose: A review. *Pattern Recognition,* 42**,** 2876-2896.

ZOU, X., KITTLER, J. & MESSER, K. Illumination invariant face recognition: A survey. Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on, 2007. IEEE, 1-8.

ZVETCOBIOMETRIC. 2012. *Biometric Definitions* [Online]. ZVETCO biometric. Available: http://www.zvetcobiometrics.com/Support/definitions.php [Accessed].