UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

Electronics and Computer Science

Facial Profile Recognition Using Comparative Soft Biometrics

by

Malak Zayed Alamri

Thesis for the degree of Doctor of Philosophy

7th September 2024

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES
Electronics and Computer Science

Doctor of Philosophy

FACIAL PROFILE RECOGNITION USING COMPARATIVE SOFT BIOMETRICS

by Malak Zayed Alamri

The identification of suspects in surveillance footage is crucial for maintaining public safety, preventing crime, conserving police resources, and aiding forensic investigations. Although eyewitness testimonies are valuable assets in numerous criminal cases, it is presently rather challenging to identify individuals in real-world closed-circuit television (CCTV) footage solely based on eyewitness descriptions. As a result, there has been a significant rise in the interest of using soft biometrics, which are physical and behavioural attributes that are used to semantically describe people under adverse surveillance conditions. Traditional biometrics are used when images or videos are available. Nevertheless, in certain instances, only eyewitness testimonies are available. In such scenarios, soft biometrics are applied to transform the eyewitness testimony into a collection of features that can be utilised for automated recognition. Furthermore, when images, videos and eyewitness testimonies are available, the fusion of soft biometrics with traditional biometrics becomes essential.

The objective of this thesis is to investigate the integration of soft biometrics with traditional biometrics, enabling the search of video footage and biometric data based on descriptive information to identify suspects. The existing literature on facial soft biometrics mainly focuses on the frontal face, However this approach fails to acknowledge the importance of facial profiles, which have been demonstrated to be highly accurate. It is crucial to consider facial profiles because there are situations in which only these profiles are captured in images and videos from surveillance and security cameras. In such instances, existing facial recognition algorithms designed for frontal views are ineffective, emphasising the necessity for recognition systems specifically tailored for profile faces.

This thesis builds upon previous research on using soft biometrics for human recognition, with a specific emphasis on the potential of soft biometrics in identifying facial profiles. Soft biometrics involves crowdsourcing human annotations through ordered and similarity comparisons. Advanced machine learning techniques are also used to estimate comparative attributes from images. In addition, we analyse the attribute's correspondence between the traditional biometric and soft biometric based on facial profiles. Therefore, we have bridged the gap between human perception and computer vision for facial profile biometric. In comparison to prior work on facial profiles, the developed approaches have demonstrated a higher level of performance. Our findings indicate that the performance of the system further improves after fusing the semantic and visual spaces. Furthermore, this thesis examines the bilateral symmetry of human facial profiles and develops a method for extracting features from facial profiles, inspired by the few-shot learning framework. Our algorithm, which is based on few-shot learning, achieves an impressive level of accuracy even when working with datasets containing a large number of subjects and low number of samples per subject.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Malak Zayed Alamri, declare that the thesis entitled *Facial Profile Recognition Using Comparative Soft Biometrics* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;

- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

- where I have consulted the published work of others, this is always clearly attributed;

- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

- I have acknowledged all main sources of help;

- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

- parts of this work have been published, listed as [Publications P1-P4] under Section 1.3.

Signed:......*Malak Alamri*...

Date:.........07/09/2024...

Acknowledgements

I am deeply grateful to everyone who invested their time and effort in discussing the wide range of topics that supported the completion of this thesis. First and foremost, I would like to express my deepest gratitude to my advisor, Dr.Sasan Mahmoodi, whose expertise, guidance, and unwavering support have been instrumental in the completion of this dissertation. Your insightful feedback and encouragement have greatly contributed to my academic and personal growth throughout this journey.

I am also profoundly grateful to my examiners, Dr.Iain Martin and Dr.Xiaohao Cai, for their invaluable feedback and constructive criticism. Your thorough evaluation and challenging questions have significantly enhanced the quality of my research. Further, I would like to thank the Jouf University in Saudi Arabia and the Saudi Arabian Cultural Bureau in London for financially supporting my PhD scholarship.

To my family, your constant love and support have been my foundation throughout this process. I am especially thankful to my parents and siblings, whose belief in me provided the motivation I needed to persevere. Finally, I would like to extend my heartfelt thanks to my son, Nawaf. Your patience, understanding, and unconditional love have been my greatest source of strength. Your presence in my life is a constant reminder of the importance of resilience and determination.

Chapter 1

Context and Contributions

1.1 Context

The most popular authentication methods in various security systems involve verification based on individual tokens or recognition based on biometrics. However, the reliability of tokens, such as identity cards or passwords, is decreasing due to duplication and the limited capacity of human memory. As a result, the use of biometrics, including facial recognition and fingerprints, is gaining popularity worldwide. According to a survey conducted by (Abdelwhab et al. 2018) in 2018, unlike tokens used in other methods, biometrics are unique to each individual, cannot be easily transferred, and are readily available. Additionally, they cannot be borrowed, forgotten, shared, stolen, or observed. Facial recognition technology is rapidly expanding within the security sector due to its efficiency. Biometric systems are also becoming popular because of their various applications. For instance, numerous organisations and agencies, such as border control, law enforcement, and the telecommunications and healthcare industries use facial recognition to enhance security and control measures. The complexity of closed-circuit television (CCTV) systems has increased to include data recordings from different angles, including side views of people. As a result, the growing difficulty of identifying individuals under different surveillance conditions has significantly spurred interest in facial profile biometric systems.

Many government institutions and private organisations face various challenges regarding system security and individual identification, leading to a reliance on traditional biometrics for facial verification (Abdelwhab et al. 2018). However, in real-time situations, traditional facial recognition systems often struggle to identify individuals due to challenging data characteristics, including low resolution images and images taken from different angles. In these cases, it is crucial to consider facial profile recognition algorithms and design facial profile recognition systems specifically for recognition and identification purposes. It is worth noting that the existing literature on biometrics

FIGURE 1.1: Closed-circuit television (CCTV) surveillance footage released in connection with a stabbing at Block 1 nightclub, Warrington, in 2019 (Everett 17-4-2019).

primarily focuses on the front-facing aspect of the face, while facial profiles have received less attention, despite their importance. The increase in global security threats has led to the widespread use of technology in the biometric field. There is a pressing need for more advanced biometric technologies, especially in facial recognition. For example, the images extracted from CCTV footage of a crime scene, as shown in Figure 1.1, highlight the importance of a system that effectively recognises faces in low quality surveillance images and in different poses. In certain images, such as the one depicted in Figure 1.2, the facial profile may be the only identifiable biometric modality, making it the preferred choice for identification purposes.

The objective of this thesis encompasses three main aspects: 1) The proposal of a facial profile soft biometric system based on descriptions obtained from eyewitness testimonies. 2) The development of a novel facial profile biometric system through automated analysis of facial profile images using traditional machine learning techniques. 3) The integration of traditional biometric and soft biometric systems to create a more robust facial profile biometric system with enhanced performance. Recognition algorithms utilised in such systems face significant challenges due to their increased object dynamism. Factors such as poor lighting, blurring and poor illumination conditions, or variations in poses can notably diminish the effectiveness and accuracy of recognition algorithms (Rajesh et al. 2016). Despite efforts, achieving heightened accuracy, certainty, and effectiveness remains a persistent challenge for most algorithms. This study aims to evaluate the efficacy of facial profile soft biometrics in enhancing reliability and efficiency. The innovative approaches examined herein are expected to improve the accuracy and reliability of traditional biometrics by considering the dynamic features that characterise individuals.

FIGURE 1.2: Closed-circuit television (CCTV) surveillance footage released in connection with an incident in Southampton city centre in 2024 (Ramos 23/02/2024).

1.2 Contributions

The primary contribution of this research is the establishment of a soft biometric system that focuses on facial profiles. We highlight the following contributions:

1. **Development of a novel facial profile recognition systems:** Facial profile recognition lags behind frontal face recognition in terms of advancement. Consequently, facial recognition algorithms often struggle when faced with changes in pose (Klare, Klum et al. 2014), as these algorithms are typically only designed for frontal face recognition. The main contribution of this work is the development of a recognition and identification system for facial profiles, utilising machine learning techniques. It is worth noting that existing facial profile biometric systems often rely on small datasets with a limited number of subjects. In this thesis, we have devised both traditional machine learning and deep learning-based methods for facial profile recognition, using a large dataset that includes numerous subjects (Chapters 5 and 6).

2. **Introducing a new set of facial attributes for semantic profiling:** Typically, facial profiles provide less information for identification compared to frontal views. Expanding on previous studies on relative and fine-grained attributes, we present a fresh set of facial profile attributes, including facial profile length, philtrum size, and ear-to-chin distance, which facilitate object retrieval. This research uses a subset of the XM2VTSDB database that meets the research criteria, consisting of 920 images appropriate for identifying distinct individuals (Chapter 3).

3. **Identifying the features required for recognition:** This thesis presents a scheme for evaluating the importance, consistency, and distinctiveness of soft biometrics in comparative facial profiles. It also investigates the role of attributes in the processes of recognition and identification (Chapter 4).

4. **Introducing majority voting for soft biometrics:** To improve recognition accuracy, this thesis introduces the novel technique of majority vote threshold.

The majority vote is an average of responses from all of the labellers. The detailed impact of this threshold is extensively explored (Chapter 4).

5. **Automatic retrieval of comparative facial profile soft biometrics:** This aspect of the study explores the automatic estimation of comparative facial profile soft biometrics from images and aims to bridge the gap between humans and computers in understanding comparative facial profile attributes. The objective is to establish a connection between comparative facial attributes and the corresponding facial segments in facial profile images. Biometric signatures can be extracted from facial profile images using the comparative label estimation method (Nawaf Yousef Almudhahka et al. 2017a). The study also investigates the relationship between soft and traditional biometrics for facial profile components in the image (Chapter 5).

6. **Fusion of soft and traditional biometrics:** This study aims to evaluate the effect of combining soft biometric and traditional biometric methods on recognition accuracy. We argue that incorporating soft biometrics with traditional biometrics is essential for addressing limitations in security and control systems when it comes to identifying individuals based on their facial profiles. We employ various methodologies to implement fusion (Chapter 5, 6).

7. **Few-shot learning for traditional facial profile recognition:** This thesis introduces a method within the few-shot learning framework to extract facial profile features. Leveraging domain adaptation and reverse validation, we propose a technique called reverse learning (RL) to explore bilateral symmetry in facial profiles (Chapter 6).

1.3 Publications

The following papers have been the outcomes of this research so far:

P1. M. Alamri and S. Mahmoodi, "Facial Profile Recognition Using Comparative Facial Soft Biometrics," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, 2020, pp.1-4.

P2. M. Alamri and S. Mahmoodi, "Face Profile Biometric Enhanced by Eyewitness Testimonies," in *26th International Conference on Pattern Recognition (ICPR)*, Montréal, Canada, 2022, pp.1127-1133.

P3. M. Alamri and S. Mahmoodi, "Face Profile Biometric Systems: An Overview," in *Face Recognition across the Imaging Spectrum*, Second Edition, T. Bourlai, Ed. Netherlands: Springer, 2024, https://www.barnesandnoble.com/w/face-recognition-across-the-imaging-spectrum-thirimachos-bourlai/1123059650.

P4. M. Alamri and S. Mahmoodi, "Facial Profile Biometrics Based on Domain Adaptation and Reverse Learning." [Submitted to *IEEE T-BIOM*].

1.4 Soft Biometric Processes and Definitions

Biometrics applications, such as fingerprint recognition, face recognition, and ear recognition can utilise both invasive or non-invasive methods. Invasive methods require the subject's cooperation and a sensor that collects data, while non-invasive methods do not require cooperation (Singh et al. 2013). A biometric system assesses the physical or behavioural traits of an individual. However, traditional biometrics, which rely on human cooperation with systems, may not always be feasible in security breach investigations. As a result, those using biometrics face numerous challenges when identifying individuals, including those authorised for remote system access, those attempting to gain physical access or egress from specific locations, or those involved in criminal activities. To address these challenges, soft biometrics can provide additional insights for improved recognition by using descriptions similar to eyewitness statements from crime scenes. However, many recognition systems struggle to identify individuals in real-world conditions, which are characterised by challenging data features, such as viewpoint variance and low resolution (Masi et al. 2016). Jain defines soft biometric traits as 'characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals' (Jain et al. 2004). Nixon et al. define soft biometrics as 'the estimation or use of personal characteristics describable by humans for use as an aid or to effect person recognition' (Nixon, Correia et al. 2015).

The limitations of traditional biometrics when combining eyewitness testimonies with evidence obtained from images, as well as when creating completely reliable recognition systems, highlight the importance of strengthening security by incorporating soft biometrics. These approaches could include an additional layer designed to reduce the weaknesses and limitations of the current systems. Soft biometrics encompass dynamic individual characteristics, such as age, height, skin colour, ethnicity, and facial dimensions, which provide valuable supplementary information and improve the precision and dependability of traditional biometrics (Abdelwhab et al. 2018). A comprehensive description of an individual can instil certainty regarding their identity when multiple factors are combined to characterise the individual at a given moment. Therefore, soft biometrics have emerged as critical tools that should complement other biometrics rather than operate independently to improve recognition accuracy. A recent study (Abdelwhab et al. 2018) has reported that such tools enhance the accuracy and speed of primary biometrics, eliminate the need for individuals' cooperation with sensors, improve effectiveness in scenarios where collecting primary biometric traits is challenging, facilitate offline system training, and alleviate security concerns by collecting and providing ancillary descriptions. By functioning as supplementary tools for primary biometrics

rather than standalone systems, soft biometrics bring substantial enhancements to these systems, as they enrich individual descriptions, thereby enhancing accuracy, reliability, and efficiency. The growing integration of soft biometrics into security systems consequently enhances the recognition of authorised individuals and individuals suspected of posing a significant threat to businesses or the public.

The growing need to improve security and control measures highlights the importance of implementing soft biometrics. These systems can convert standard human descriptions into characteristics that are easier for machines to handle. Soft biometrics usually involves first defining a set of traits and assigning a label to each one. These traits can then be evaluated in either a categorical or comparative context, with the labels indicating the presence of a particular trait in one subject compared to another. For instance, one individual may have darker skin than another. The efficient utilisation of biometrics requires the identification of multiple traits that can be seen from different angles. However, many of these facial traits can be observed in face images and facial movements, indicating that using a minimal number of features can improve the performance of the recognition system. Therefore, the gathering of facial traits is an important step in accurately recognising or identifying a subject based on specific information.

To compute a soft biometric identification, collaboration between humans and machines is essential to derive visual descriptions of an individual's facial profile image using a set of semantic attributes. This process typically involves two steps: 1) soft biometric annotation, where human annotators provide neutral descriptions of the subject's facial profile image, and 2) automatic soft biometric recognition, where machine learning algorithms are employed to effectively estimate ground-truth labels from the facial profile images. Figure 1.3 illustrates the integration of semantics, computer vision, and soft biometrics in this research. Some of the soft biometric terms used in this thesis include:

Soft biometrics: Visual descriptions of individuals consist of semantic characteristics.

Soft biometric traits: Specific human characteristics, such as age, described using a semantic property, such as 'younger'.

Soft biometric annotation: Human annotators provide objective visual descriptions of subjects' images and generate semantic labels.

Soft biometric recognition: Automatic recognition based on semantic attribute labels using subjects' images with target ground-truth descriptions.

Soft biometric retrieval: Uses target ground-truth and estimated labels to automatically identify subjects based on their soft biometric characteristics.
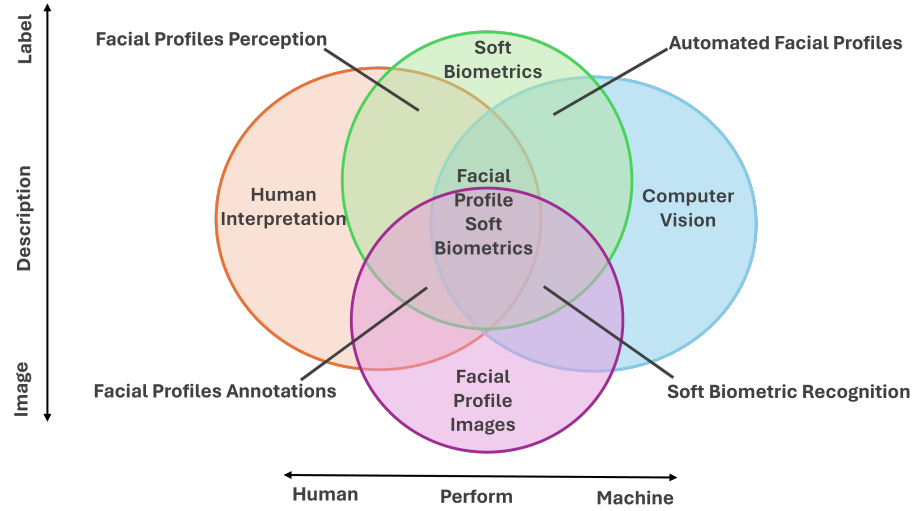
FIGURE 1.3: An overview of how subject descriptions, from images to soft biometric labels, can be transmitted between humans and machines. Each individual sphere corresponds to one of the following four subject areas: computer vision, soft biometrics, human perception, and facial profile images. The overlaps between spheres indicate the points where multiple topic areas converge. The vertical axis depicts the mapping from low-dimensional labels to high-dimensional images, while the horizontal axis reflects the interaction between humans and machines.

Semantic descriptions: Words or phrases that directly convey interpretable meaning to annotators when describing visually discernible physical traits of a subject (Samangooei et al. 2008).

1.5 Thesis Overview

This thesis is organised as follows:

Chapter 2 [*Publication* P3]: This chapter introduces soft biometrics and discusses their advantages over traditional biometrics, as well as their overall benefits. It also covers different types of soft biometrics and their formats, such as categorical and comparative. Additionally, it includes a literature review to identify the significant findings and gaps in previous research. Furthermore, it provides a summary of datasets suitable for facial profile identification and recognition.

Chapter 3 [*Publication* P1]. This chapter explores comparative attributes and their use in recognition through soft biometrics. It introduces a new set of soft biometric attributes and explains how researchers can use crowdsourced comparative facial profile labels to create virtual ratings using a rating system. Additionally, it presents the XM2VTSDB dataset and explains how this research uses it to identify individual facial profiles through comparative soft biometrics. Furthermore, the chapter provides a comprehensive analysis of potential ranking methods for comparative soft biometrics. It also

includes a summary of existing algorithms, highlighting their potential problems based on previous studies that have evaluated these algorithms.

Chapter 4 [*Publication* P1]. The chapter highlights the importance of introducing a new technique for the majority vote threshold and its impact on the accuracy rate. It also reviews the contributions of different soft biometric attributes and their ability to discriminate. Additionally, the chapter highlights potential areas for improvement in the recognition process based on the observed results. Finally, it summarises the findings and discusses their potential implications.

Chapter 5 [*Publication* P2]. This chapter examines the automatic retrieval of biometric signatures from facial profile images. It investigates the estimation of relative attributes, which form the foundation for generating biometric signatures. Moreover, the chapter presents a technique for generating automatic biometric signatures using multiple linear regression (MLR). Furthermore, it suggests a method for enhancing facial profile recognition biometric systems by integrating comparative soft biometrics with traditional biometrics.

Chapter 6 [*Publication* P4]. This chapter explores the concept of bilateral symmetry in human facial profiles. It introduces a new biometric approach within the few-shot learning framework to extract facial profile data. The method, known as reverse learning (RL), employs domain adaptation and reverse validation to analyse bilateral symmetry.

Chapter 7. This chapter presents the conclusions derived from the study and discusses the potential implications of its findings for comparative soft biometrics. Furthermore, it offers recommendations for avenues for future research.

Chapter 2

Facial Profiling as a Biometric and Identity Science

2.1 Introduction

As evidenced by the increasing body of literature on the topic, researchers have begun to understand the potential value of using facial profiles in human recognition. This chapter presents a literature review of the existing research on facial profile biometrics, with the goal of establishing the foundations for evaluating how the use of facial profile images could contribute to facial recognition. Generally, descriptions of the human face are inconsistent and elusive, particularly when given in eyewitness descriptions. To enable automatic identification based on these descriptions, it is necessary to examine the correspondence between computer vision and human perception in order to obtain objective and consistent interpretations and then combine them. However, there is a semantic gap between human perception and the computer vision in the process of image understanding (Smeulders et al. 2000; Hare et al. 2006). Thus, this chapter also explores previous work undertaken by other researchers on facial soft biometrics in recognition and verification.

Section 2.2 investigates facial profiling in the literature. Section 2.3 analyses datasets that are suitable for facial profiling identification and recognition. Section 2.4 discusses the significance of facial profiling in the classification and recognition of individuals' identity, and Section 2.5 establishes the foundation of this work by focusing on soft biometrics. It documents the shift from using traditional characteristics to relying on visually perceived descriptions for identification purposes. Finally, Section 2.6 provides an overview of the study findings.

2.2 The History of Facial Profiling

Figure 2.1 shows a brief history of the development of facial profiling methods in the literature. The first attempt to use facial profiles as a biometric entity can be traced back to Sir Francis Galton in the 1889s, who standardised profile traces with typical lengths and angles for identification and description (Galton 1889). Over the course of the ensuing years, a number of tests and adjustments were conducted with the aim of enhancing the precision of facial profile biometrics. The attempts Kaufman et al. (1976), Leon Harmon and Hunt (1977) and Leon Harmon, Kuo et al. (1978) made in the 1970s to improve the efficiency of an earlier version of their work stand out as particularly noteworthy. These researchers further developed the precursor to automatic biometric face profile recognition by calculating the distances and angles between facial landmark points. The purpose of including these components was primarily to make the procedure more inclusive of facial profile aspects, which would ultimately result in an improvement in the accuracy of the technology. The geometrics of the facial features were extracted in order to compose feature vectors that were 10 dimensions long. This feature vector was constructed with the positioning of nine characteristic points on the profile, serving as the basis to compute these feature vectors. Following that line of investigation, the number of facial landmark points was increased to 11. Adding more landmark points then resulted in the generation of a 17-dimensional feature vector for each profile (LD Harmon et al. 1981). According to the findings of Harmon study, correct recognition was achieved in 96% of cases.

In recent years, much research has been carried out to study the method of identifying facial profiles utilising landmark points. This technique makes use of numerous facial landmarks as points of reference, including the chin, nose, forehead, nose bridge and mouth. Features are determined by measuring the distances that separate landmark points, the angles that connect those points and the areas of specific triangles that are formed by the points. The next important step forward was taken in the 1990s, when Aibara, Ohue and Oshita (1993), Aibara, Ohue and Matsuoka (1991) and Somaie et al. (1995) presented their work on the application of P-Fourier descriptors (PFDs) to low-frequency range signals. Expanding upon this work, Somaie et al. (1994a) and Somaie et al. (1994b) constructed a system using eigenvectors and a novel linear dynamic model to conclude that feature analysis of many facial photos might yield a collection of essential basic elements.

In a previous study (Wu et al. 1990), facial profile outlines were extracted automatically. Furthermore, B-splines were employed in that work to extract six interest points, that is, the nostril, nose peak, forehead, chin, eye and mouth points. A 24-dimensional feature vector was then computed by calculating the intervals between two adjacent points along with the angle between the curved sections linking two neighbouring points. Feature vectors in the probe were then compared with feature vectors in the gallery

1970s
Harmon et al and Kaufman et al investigated how to improve the accuracy of facial profile identification by measuring the distances and the angles between the fiducial points

1999
Lipoščak and Loncaric proposed a new approach for face recognition from profile images

2002
Gao and Leung proposed a string matching method for face profile recognition

2005
Zhou and Bhanu introduced a practical approach for human recognition based on face profiles curvature matching

2013
Ding et al provided a novel side-view face authentication method based on discrete wavelet transform.

1889
Francis Galton's normalised face profile traces with characteristic lengths and angles

1990s
Aibara et al and Somaie et al reported on the use of Fourier descriptors in the low-frequency range (PFD)

2001
Wallhoff et al e presented a novel hybrid ANN/HMM approach to recognize a person from that person's profile view

2004
Bhanu and Zhou proposed a method using dynamic time warping

2005
Pan et al investigated histogram comparing for human facial profile recognition

FIGURE 2.1: Major milestones in the history of facial profile biometrics.

based on a successive search technique. In (Wu et al. 1990), the gallery and probe contained three samples each. In the first attempt, 17 out of 18 face profile subjects were correctly identified. The unsuccessful image feature vector within the training sets was then integrated in the gallery to ensure successful retraining. In the second attempt, the datasets were expanded so that all of the individuals were correctly recognised.

Lipoščak et al. (1999) suggested a scale-space filtering method based on facial landmarks which resulted in an identification rate of 90% when it was evaluated using data from a variety of different subjects (30 in total). Furthermore, Gao et al. (2002) developed the concept of employing string matching as a technique for the purposes of facial profile recognition. For this technique, the facial profile is converted into a series of line segments at the beginning of the process. After that, the characteristics of each line segment, such as its length, direction and midpoint, are utilised as the primary components of the representation of the line segments. The facial profiles are then aligned, and a string matching approach is employed to calculate the distance score, which measures the similarity between the probing profiles and the gallery profiles. However, this technique is flawed, as it makes the erroneous assumption that the two profiles being compared have identical curve features. As a result, it is ineffective in the presence of occlusion. It is important to note that the two aforementioned methods only make use of two-dimensional (2D) profiles. Therefore, they completely ignore the potential for various

head poses to affect a portion of an individual's facial profile. Consequently, neither the 2D nor the 3D profile comparisons in these studies yield any significant results.

Wallhoff and Rigoll (2001) and Wallhoff, Muller et al. (2001) described a novel hybrid of artificial neural networks (ANNs) and hidden Markov models (HMMs) to recognise a person using a facial profile view (90°). This was accomplished in spite of the fact that the recognition system only used a single frontal image of individuals as training data in their study. Yet, this approach has the potential to achieve remarkable identification rates of up to approximately 60%. Bhanu et al. (2004) employed the dynamic time warping (DTW) technique, which compares and contrasts different facial profiles. The pronasale and nasion points are found by computing the profile's overall average curvature and using that information to locate the landmarks. The similarity score that compares the profiles in the probe with the relevant profile in the gallery is determined by the DTW using curvature as an important measurement. In their study, this method achieved an identification rate of roughly 90% in two different datasets (Bhanu et al. 2004).

This approach was expanded to video-based facial profile identification by Zhou et al. (2005) who aligned numerous low-resolution facial profile images. As this method is dependent on the pre-processing of scale space, the calculation of the landmark point placement in approaches based on curvature is considered to be of the utmost importance. This is weighted more heavily towards profiles with larger variances in the Gaussian used to estimate the required scale. When applied to a wide variety of profiles, the scale-space technique does not provide any assurance that the resulting curvature patterns will be the same in every single instance. The highest accuracy achieved in their study was 78.6%, while the lowest accuracy recorded was 64.3%. On average, almost 70% of individuals were accurately identified by facial profile (Zhou et al. 2005).

In the study that Zhou et al. (2007) conducted on the recognition of side-view faces, two methodologies were put forward, namely, principal component analysis (PCA) and multiple discriminant analysis (MDA). To match the side-view face photos, a method based on curvature of the face was utilised to extract landmark points, and these points were then employed in both the training and the test sets. The highest achieved recognition rate was 91.1% (four errors out of 45 persons). Pan et al. (2005) reported the results of an experimental assessment of some alternative solutions to the problem of facial profile recognition. They investigated and evaluated various approaches to profile alignment, including tangent-based standardisation, 2D iterative closest point orientation (2D ICP) and simulated annealing (SA) synchronisation. In order to assess the three methods, the researchers used databases that contained pictures of faces observed from the side, as well as information about facial profiles that had been extracted from three-dimensional (3D) models of faces. Their study concluded that the SA and 2D ICP techniques yielded the best results in terms of facial profile recognition.

Ding et al. (2013) described a novel method for authenticating faces viewed in profile. The approach used makes use of wavelet modification in conjunction with a random forest classifier. The method of extracting the side-view face contours begins with the profile image. Following that, discrete wavelet transform (DWT) analysis is performed on the features that are retrieved from the data. Subsequently, a feature vector is computed using a super Fourier descriptor for each face based on the associated wavelet coefficients. During the evaluation, the suggested strategy was shown to have the highest feasible level of accuracy at 92.50%.

The inability to capture the entire frontal face lowers the performance of recognition algorithms. For example, Park et al. (2010) noted that frontal, partial and non-frontal face images are unsuitable for use with conventional face matching systems, which rely on numeric matching scores. However, the use of facial-landmark-based matching algorithms allows for facial profile identification, either in combination with any matching face or independently during face obstruction (Park et al. 2010). In addition, a classifier can use face matching systems to extract traits from the sides of faces to match subjects. Furthermore, a revolutionary method was provided in the study conducted by Jia et al. (1994) to identify distinctive features of a facial profile based on a single frontal image to facilitate automatic facial recognition. The results of this research show that the analysis of the Walsh power spectrum of a profile feature is an effective way to positively identify and distinguish one person from another.

There are several deep learning-based methods that have explicitly investigated facial profile recognition and proposed methods for extracting discriminative features from facial profiles. Sengupta et al. (2016) illustrated how several algorithms performed when using a restricted protocol and how each one degraded from frontal–frontal to frontal–profile. In their study, the frontal-frontal and frontal-profile experiment achieved classification accuracy of 96.40% and 84.91%, respectively, using a deep features-based method. Moreover, in recent years, generative models, such as generative adversarial network (GAN)-based methods, have been widely used to synthesise the frontal view from the profile view in order to improve facial profile recognition systems (J. Zhao et al. 2018; P. Li et al. 2019; Yin et al. 2020). In addition, deep-learning-based methods have also demonstrated high levels of performance. Facial recognition has been greatly improved by deep learning techniques, which are trained on a large-scale dataset and demonstrate high-level recognition rates under challenging conditions (Parkhi et al. 2015; Deng et al. 2019; Meng et al. 2021).

In a recent study (Romeo et al. 2019), a method for automatically normalising facial profiles to improve the performance of deep-learning-based profile face matching was introduced. Given a non-normalised image set with at least one pre-corrected image per individual, the method presented in (Romeo et al. 2019) generate a normalised image set. This method has also been tested on the NIST and the MIT-CBCL face recognition database. In their study, a deep learning recognition system trained on

facial profiles yielded a promising accuracy rate, with a 7.2% improvement over the baseline (i.e. non-normalised facial profiles). In this thesis, we use a high-performance deep learning technique (few-shot learning) on datasets with a small number of samples (eight samples per subject) to study both the facial profile recognition and the bilateral symmetry on the facial profiles. The results are presented in Chapter 6.

Depending on the approach that is taken to extract the features from the face, facial profile recognition methods can be broken down into one of five distinct subclasses: approaches based on landmark points (geometric features); approaches based on PFD; approaches based on DWT; approaches based on HMM; and approaches based on DTW (see Figure 2.2). Table 2.1 provides a summary of the most important facial profile recognition approaches.

TABLE 2.1: Summary of facial profile recognition approaches.

Publication	Technique	Database	Accuracy
Galton 1889	Variable and independent feature	-	-
Kaufman et al. 1976	Geometric features	-	90%
Harmon et al. 1977	Geometric features	-	-
Harmon et al. 1978	Geometric features	-	82%
Harmon et al. 1981	Geometric features	-	96%
Wu et al. 1990	Geometric features	OWN database	100%
Aibara et al. 1991	Fourier descriptor	OWN database	94.49%
Aibara et al. 1993	Fourier coefficients	OWN database	93.1%
Somaie and Ipson 1994	Fourier transformation	OWN database	100%
Jia and Nixon 1994	Walsh power spectrum	OWN database	-
Lipoščak and Loncaric 1999	Geometric Features	OWN database	90%
Wallhoff et al. 2001	HMM-approach	MUGSHOT database	56%
Gao and Leung 2002	String matching	Face database, University of Bern	96.7%
Bhanu and Zhou 2004	Dynamic time warping technique (DTW)	Data from University of Bern	90%
Zhou and Bhanu 2005	Dynamic time warping technique (DTW)	OWN database	78.6%
Pan et al 2005	Geometric features	Data from University of Bern and FERET database	-
Zhou and Bhanu 2007	PCA and MDA	OWN database	91.1%
Park and Jain 2010	Active appearance models	FERET database	92.02%
Ding et al 2013	Discrete wavelet transform (DWT)	The GTAV face database	92.50%

2.3 Facial Profile Databases

The availability of more diverse facial datasets has increased over the past 10 years, which is one reason why the performance of facial recognition algorithms has improved so dramatically. The growing number of facial databases also allows for the evaluation of the performance of facial recognition systems under different conditions (e.g. blurring, noise and turbulence) (Bourlai 2016). This section will discuss the databases that contain facial profiles. Although some of these datasets do not primarily focus on facial profiles, they do contain some facial profile images among their diverse collections. Sample images from different facial profile databases are depicted in Figure 6.4, and a summary of the databases is provided in Table 6.2.

The Sheffield Face Database (previously UMIST) (Graham 1998): This contains 564 images of 20 subjects of mixed ethnicity, gender and appearance.
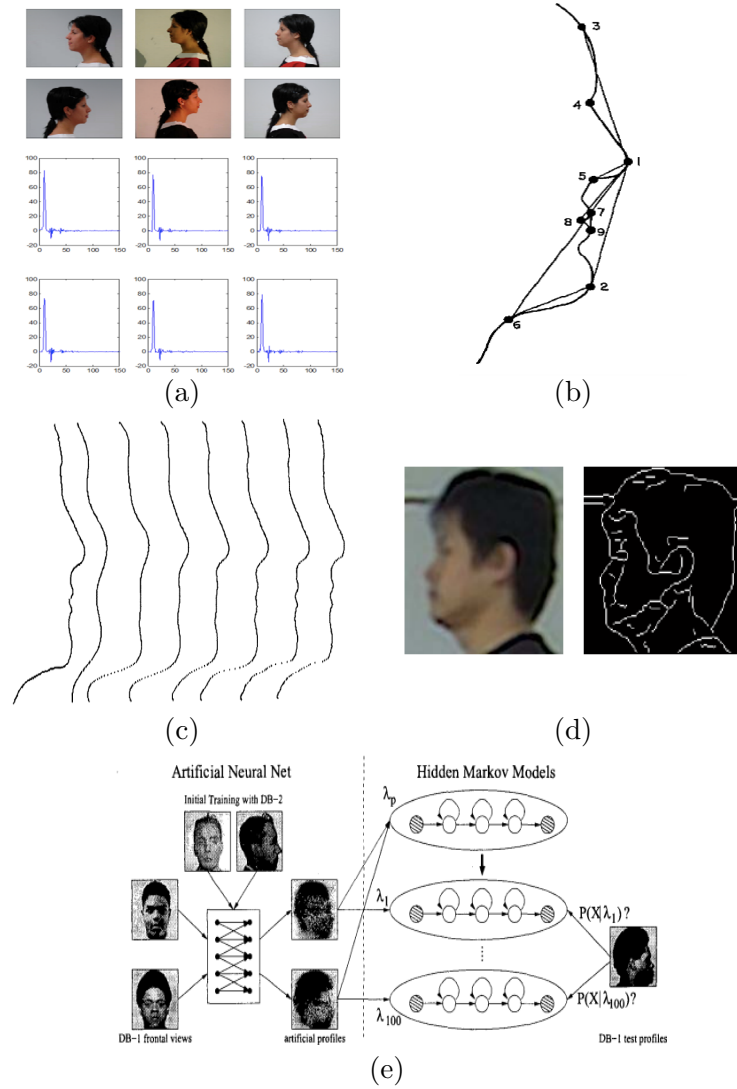
FIGURE 2.2: Examples of different approaches used to extract features from a facial profile; (a) Discrete wavelet transform (Ding et al. 2013), (b) Landmark points (Leon Harmon and Hunt 1977), (c) P-Fourier descriptors (Somaie et al. 1995), (d) Dynamic time warping technique (Bhanu et al. 2004) and (e) Hidden Markov model (Wallhoff, Muller et al. 2001)

Each individual is shown from different angles, from the profile to the front. The images are all in PGM format, with dimensions of around 220 × 220 pixels and 256-bit grey scale.

XM2VTSDB (CDS002) (Messer et al. 1999): The XM2VTSDB contains 295 subjects; each subject is shown in 8 images rotated from –90 to +90 (4 left side images and 4 right side images) constituting a total of 2,360 images. The images are at a resolution of 720 × 576 pixels.

Face Recognition Technology (FERET) Database (Phillips et al. 2000): This is a colour database that is mostly used for face recognition. It has 11,338

colour images with a resolution of 512×768 pixels captured in a semi-controlled environment, with 13 different poses from 994 individuals.

ND-Collection (UND-F and UND-J2) (Flynn et al. 2003; Yan et al. 2007): This collection of biometric images is owned by the University of Notre Dame (UND) and serves as the repository for the UND Biometrics Database. There are numerous categories in this dataset, including UND-F and UND-J2. The UND-F category has 942 3D (and corresponding 2D) pictures of the profiles of 302 subjects collected between 2003 and 2004; while UND-J2 has 1,800 3D (and corresponding 2D) profile images from 415 subjects collected between 2003 and 2005.

The FEI Face Database (Amaral et al. 2008): This is a collection of images of faces, which is part of a Brazilian face database, taken between June 2005 and March 2006. It includes 14 images for each of the 200 subjects, making up a total of 2,800 images. All of the pictures are in colour and were taken in front of a plain white background with the face rotated up to about 180 degrees. The scale varies by up to 10%, and the original dimensions of each image are 640×480 pixels.

Labeled Faces in the Wild (LFW) (Huang et al. 2008): This is a collection of face images created with the intention of researching the issue of unrestricted face recognition. The Viola Jones face detector was used to detect and centre 13,233 pictures of 5,749 people on the web. There are 1,680 subjects with two or more distinct images.

NIST Mugshot Identification Database (MID) (NIST Special Database 18) (Watson 2008): This database has been made available for use in the development and evaluation of automated mugshot identification systems. There are a total of 1,573 individuals in this database. It includes both front and profile views. The dataset consists of 89 cases that contain two or more facial profiles, as well as a single facial profile of 1,268 individuals.

CMU Multi-PIE (Multi Pose, Illumination, Expressions) (Sim et al. 2001): This contains over 750,000 images of 337 subjects recorded in up to four sessions over the span of five months. The images capture different poses, degrees of illumination, and expressions. The pose range contains 15 different viewpoints, including facial profiles.

MUCT (Maastricht University Compression Test) (S. Milborrow et al. 2010): The MUCT database consists of 3,755 images of 276 subjects. Compared to other databases, it was specifically created to offer a diverse range of images with different lighting conditions, ages, and ethnic backgrounds. Every image is 640×480 pixels.

CelebA (CelebFaces Attributes) (Liu et al. 2015): This dataset contains over 202,599 images sized 178 × 218 pixels of 10,177 celebrities, including facial profiles, each annotated with 40 binary labels indicating facial attributes, such as hair colour, gender and age.

SiblingsDB Database (HQ-fp and HQ-fps) (Vieira et al. 2014): The Siblings DB database is divided into three individual datasets, namely HQf, HQfp and HQfps. Each individual in HQfp is represented by an expressionless frontal and profile image. Meanwhile, HQfps contains a total of 112 subjects, with a set of four images per subject. These images consist of two expressionless images (one frontal and one profile), as well as two smiling images (one frontal and one profile).

IARPA Janus Benchmark A (IJB-A) (Klare, Klein et al. 2015): This database was created with the objective of introducing an additional challenge to the task of face recognition. This is achieved by gathering facial images that exhibit a diverse range of variances in terms of position, illumination, expression, resolution and occlusion. The IJB-A database is composed of 5,712 images and 2,085 videos collected from 500 subjects, with an average of 11.4 images and 4.2 videos per subject.

VGGFace2 Dataset (Cao et al. 2018): VGGFace2 aims to provide a comprehensive dataset that may be utilised for the purpose of facial recognition across various poses and age groups. The dataset includes 3.31 million images of 9,131 subjects, with an average of 362.6 images per subject. The average resolution of the images is 137 × 180 pixels.

GTAV Face Database (Tarrés 2012): This is one of the more current face databases, and its primary goal is to test how well face recognition algorithms work when dealing with strong poses and lighting changes. There are a total of 44 subjects in this database, and each subject has 27 images which show various pose views under three distinct lighting conditions. The resolution of the images is 240 × 320 pixels.

MegaFace (Kemelmacher-Shlizerman et al. 2016): This is a publicly accessible dataset used to measure the performance of face recognition algorithms. MegaFace contains 4.7 million photos that represent 672,057 unique subjects, with an average of 7 images per subject. The average resolution of the images is 137 × 180 pixels.

Celebrities in Frontal-Profile in the Wild (CFPW) (Sengupta et al. 2016): The CFPW has 100 famous individuals. There are a total of 14 images of each individual, with 10 frontal and 4 side view. The images were obtained from the internet, recorded in uncontrolled situations, and from individuals of varying ages.

Datasets	#Subjects	#Images	#Img/subj	#Views	Year
Sheffield Face Database	20	564	-	10	1998
XM2VTSDB (CDS002)	295	2,360	8	2	1999
FERET	1199	11,338	2	7	2000
ND-Collection (UND-F)	302	942	3 or 4	1	2004
ND-Collection (UND-J2)	415	1800	4 or 5	1	2005
FEI Face Database	200	2800	2	14	2006
Labeled Faces in the Wild	5,749	13,233	1/2.3/530	-	2007
NIST special database 18	1,573	3,248	1 or more	-	2008
CMU Multi-PIE	337	750,000	6	15	2009
MUCT	276	3755	10 to 15	1	2010
CelebA	10,177	202,599	-	-	2015
SiblingsDB Database (HQ-fp)	158	316	1	2	2014
SiblingsDB Database (HQ-fps)	112	448	2	2	2014
IARPA Janus Benchmark A	500	5,712	11.4	3	2015
VGGFace2 Dataset	9,131	3.31	1	-	2018
GTAV Face Database	44	1,628	10	2	2016
MegaFace	690,572	4.7 M	3/7/2469	-	2016
CFPW	100	1400	14	2	2016
IARPA Janus Benchmark B	1,845	21,798	6.37	3	2017
300W-LP/LPA	59,439	366,564	-	-	2019

TABLE 2.2: Summary of facial profile databases.

IARPA Janus Benchmark B (IJB-B) (Whitelam et al. 2017): The IJB-B dataset is a template-based face dataset containing 21,708 images and 7,011 videos of 1,845 subjects. These images and videos were gathered from the internet and are completely unrestricted in terms of pose, lighting, image quality, etc.

300W-LP/LPA (Large Pose Augmented) (Zhu et al. 2016): The 300W-LPA dataset expands upon the 300W-LP dataset by rotating each image by 5 degrees in both directions up to a maximum of 25 degrees. It comprises 366,564 images of 59,439 subjects. It is a large-scale face alignment dataset that includes face profiles.

Several other datasets also contain some profile images, such as the UCD Colour Face Image Database (Sharma et al. 2003) and PaSC (Beveridge et al. 2013), however, they are more suitable for face detection than face recognition. After an extensive search for a dataset that includes poses at 90-degree angles to support the purpose of this research, a number of relevant datasets for use in the analysis, including Multi-PIE, VGGFace2, LFW and XM2VTSDB, were initially considered for this study. The main goal during the search process was to find a dataset that had sets of four or more facial profile images of the same subject taken over a period of time and entirely obscuring the other side of the face, to ensure that it was not in the camera's view and, therefore, the image did

not reveal information about the other side of the face. In the end, the XM2VTSDB dataset proved to be the best option.

(a) XM2VTSDB Database   (b) MUCT Database

(c) GTAV Face Database   (d) Sheffield Face Database

FIGURE 2.3: Sample images of facial profile databases in a constrained environment: a) XM2VTSDB database, b) MUCT database, c) GTAV Face database and d) Sheffield Face database.

2.4 The Role of Facial Profiles in Biometric and Classification Systems

Recent growth in biometrics has made it necessary to use facial profiles as a type of biometric, and recent research has demonstrated that facial profile images might be helpful in various biometric and recognition contexts. This section will provide a description of the various methods that use facial profiles for recognition and classification. Here, we illustrate that the role of facial profiles in recognition and classification can be grouped into three categories based on the type of recognition/classification, specifically: gender and age classifications, human recognition and facial profile cross-recognition. In Section 2.4.1, gender and age classification using facial profile images are discussed. Human recognition is presented in Section 2.4.2. Finally, facial profile cross-recognition is explained in Section 2.4.3, based on facial profile bilateral symmetry. Table 2.3 summarises the different classifications and recognition methods using facial profile images.

2.4.1 Gender and Age Classifications from Facial Profile Images

Ghaffary et al. (2011) introduced a novel profile-based facial recognition system that makes use of the geometric information contained in the profile outline curve. A cascaded Adaptive Boosting (AdaBoost) method is used at the beginning of the process to

determine the location of the ear, the tip of the nose and an area of the skin. The face region is extracted using the H and S colour channels and a 2D histogram back-projection technique. Furthermore, the angular sampling procedure is undertaken using a matrix that moves from the midpoint of the ear and the end of the nose, in order to obtain the normalised classification model from the profile curve of the face. The method was found to achieve a detection rate of 96.67% when it was tested on a set of 150 different profile face images. Moreover, Tariq et al. (2009) achieved 71.20% gender and 71.66% ethnicity classifications using silhouetted facial profiles of 441 images.

Using convolutional neural network (CNN) models, Yaman, Eyiokur, Sezgin et al. (2018) achieved accuracy rates of 94% and 52% for the classification of gender and age, respectively. In this study, a large-scale ear dataset was constructed from the profile and near-profile photos that were included in the Multi-PIE face dataset. Before completing the task, the CNN models were optimised using a vast ear dataset. Moreover, the results presented in Bukar et al. (2017) work demonstrated that images of the ear, as well as of the profile of the face, can be helpful for age estimation. The findings were based on an aggregate of four distinct datasets that were used in their experiment. ResNet-50, ResNet-101 and ResNet-152 were used in order to extract features from the data. It should be noted that each of these datasets came with its own set of benefits and drawbacks. In addition, a sparsely populated version of the partial least squares regression technique was also employed with the features that were extracted. The findings of the tests on the FERET dataset revealed that an optimal outcome may be achieved when utilising ResNet-152 features with a mean absolute error of 5.50. In another study, support vector classification (SVC) was used in both single and multimodal experiments in Zhang's lab to determine gender based on ears and profiles ()zhang2011hierarchical. For the purposes of the experiments, the UND-F dataset was utilised. While the multimodal system were found to have an accuracy of 97.65%, the ear-only and profile-only methods achieved accuracy rates of 91.78% and 95.42%, respectively.

Yaman, Irem Eyiokur et al. (2019) also illustrated how combining ear and facial profile photos yields a high percentage of accuracy when classifying age and gender. In their study, several deep CNN frameworks, which were both multimodal and multitask, were developed. In addition, various fusion methods have also been studied by the same researchers, including those that combine data intensities, spatial locations, channels, features, and scores. In on study, two CNN models, namely VGG-16 and ResNet-50, were employed. It was observed that the highest level of accuracy of 65.73% was achieved when using VGG-16 on facial profiles to estimate age. Despite the fact that ResNet-50's accuracy with ear photos was significantly lower than that of VGG-16, the two models obtained nearly the same level of accuracy (60%) when it came to facial profile images. The ResNet-50 CNN model has also been shown to produce the best results for gender classification with ear images, with an accuracy rate of 98.00%. Thus, although the results achieved by VGG-16 are very close to the ResNet-50 results, the best result

possible for gender classification may be obtained from ear images using the ResNet-50 CNN model. However, VGG-16 has been shown to achieve 95.81% accuracy and ResNet-50 94.05% accuracy when used to predict age based on facial features. It has also been demonstrated that images of the ear are more useful for determining gender than images of the facial profile. This is despite the fact that facial profile photos perform significantly better than ear images in terms of age classification. In addition, the VGG-16 and ResNet50 CNN models were both found to be able to achieve a classification accuracy of 99.11% with the use of spatial fusion in their multimodal trials. Based on these results, it is abundantly clear that facial profile photos constitute a rich data source that may be used for age and gender differentiation. Figure 2.4 depicts an overview of the multimodal and multitask framework for analysing and classifying the age and gender of individuals.

The experimental findings presented by Yaman, Eyiokur and Ekenel (2022) indicate that images of the ear and the profile of the face contain features that could be utilised to obtain soft biometric traits. That is, in cases where a frontal view of the subject is unavailable, it has been demonstrated that ear and facial profile images could be used effectively by soft biometric recognition systems. Yaman's study achieved a gender classification accuracy of 98.33% and an age classification accuracy of 65.73% when using facial profile images. When compared to methods that only use frontal face images, the multimodal system presented in that study was found to achieve higher levels of recognition rates. This is also the case when comparing the multimodal method to the unimodal method. As a consequence, using facial profiles continues to be one of the most effective biometric technologies to determine an individual's true gender. In spite of this, however, little research has been conducted to explore the use of facial profiles.

2.4.2 Human Recognition from Facial Profile Images

A significant number of studies on how to recognise faces using limited information have also been published, and several of these have focused on human facial profiles (Aibara, Ohue and Oshita 1993; LD Harmon et al. 1981; Wu et al. 1990). X. Zhang et al. (2008) created a novel approach to facial recognition by combining frontal and side pictures of the face to better handle rotational poses. This was accomplished using a combination of image features. The numerical results from the study presented a reliable face-recognition framework using both frontal and profile face images. The experimental results showed that the strategy of using both frontal and side views to distinguish rotated faces could improve the value of forensic databases in automated human face recognition systems. The recognition rate using a side view only was 73%.

Moreover, Sarangi et al. illustrated that, compared to state-of-the-art cross-media biometrics for the same dataset, systems employing different biometric configuration settings, such as facial profiles and ears, could significantly improve identification performance (Sarangi, Mishra et al. 2018). According to the findings of their study, the recognition rate achieved when using the ear was 90.94%, while the accuracy achieved when using facial profile was 93.57%. However, when combining the two modalities, the recognition rate was 99.12%.

A recent study (Sarangi, Panda et al. 2022) also investigated a multimodal biometric recognition system that utilised ear and profile face images as input. In this system, the discriminative feature vectors were created using three feature extraction methods, including the local directional patterns (LDP), local phase quantisation (LPQ) and binarised statistical image features (BSIF). Furthermore, the experimental findings of two benchmark datasets (UND-E and UND-J2) showed that the proposed strategy exhibited superior performance in comparison with individual modalities (i.e. unimodal ear biometrics). When combining the LDP and LPQ techniques for facial profiles and ears, it was found to achieve a 100% recognition rate on the UND-E dataset. Using only facial profile images, the approach achieved a 93.57% identification rate, whereas the use of ear images only led to a 90.95% identification rate.

Rathore et al. also created a multimodal system based on the ear and facial profile images (Rathore et al. 2013). Three different databases were used to evaluate the experiment, namely IITK Data Set 1, UND-E and UND-J2. To clarify, there are 801 facial profile images in IITK Data Set 1, 464 facial profile images in UND-E and 1,800 facial profile images in UND-J2. The images were improved using three techniques, that is, adaptive histogram equalisation (AdHist), non-local means (NLM) and steerable filter (SF). In addition, both the speeded up robust features (SURF) and scale-invariant feature transform (SIFT) algorithms were utilised to calculate features from the photos. The fusion was implemented at both the feature and score levels. The findings demonstrated that when the ear and facial profile are combined, performance is noticeably better than when they are used separately. In short, when AdHist, NLM and SF were combined for facial profiles, a 99.16% recognition rate in IITK was achieved.

Another recent study investigated (Alqaralleh et al. 2023) masked face recognition by employing BSIF and CNN as the main recognition algorithms. Two distinct datasets, ND-TWINS-2009-2010 and Celebrities in Frontal-Profile in the Wild (CFPW), were utilised to evaluate the accuracy of the proposed method. Since the images in the datasets are maskless, masks were synthetically added to the images. The experimental results demonstrated that this approach could achieve a 98% success recognition rate when using a facial profile. Overall, the findings reported in the aforementioned studies indicate that multimodal and facial profile recognition systems enjoy superior performance compared to unimodal ear or facial profile systems.

TABLE 2.3: Summary of different classification and recognition methods for facial profile images.

Publication	Classification Type	Technique	Database	Accuracy
Zhang et al. 2008	Rotated face recognition	Holistic matching method, PCA, and LBP	MUGSHOT database	73% by using LBP
Tariq et al. 2009	Gender and ethnicity	Shape context based matching	441 silhouetted face profiles	71.20% for gender classification and 71.66% for ethnicity classification
Zhang and Wang 2011	Gender	OWN technique	UND-F	95.43%
Ghaffary et al. 2011	Gender	Geometric information of the outline curve of profile silhouettes	GTAV	96.67%
Bukar and Hassan 2017	Age	Deep learning	FERET	84.92%
Yaman et al. 2018	Age and gender	Deep learning	Own Dataset built from the Multi-PIE face dataset	94% for gender classification and 52% for age classification
Sarangi et al. 2018	Human identification	LDP+BSIF+LPQ, PCA	UND-E	95.32%
Toygar et al. 2018	Face profile bilateral symmetry	BSIF+LPQ+LBP	ND-twin database	81.52%
Yaman et al. 2019	Age and gender	Deep learning	UND-F, UND-J2, and FERET	98.33% for gender classification and 65.73% for age classification
Yaman et al. 2022	Age and gender	Deep learning	FERET	65.73% for age classification and 95.81% for gender classification
Sarangi et al. 2022	Face recognition	SPT-LPQ	UND-E	94.73%
Rathore et al. 2023	Human recognition	AdHist + NLM + SF	IITK	99.16%
Alqaralleh et al. 2023	Masked face recognition	BSIF	ND-twin database	98.00%

(a)   (b)
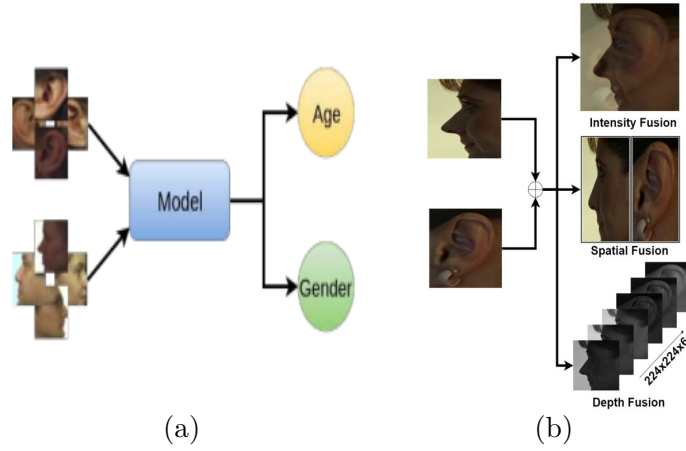
FIGURE 2.4: Examples of gender and age classification from facial profile images by deep learning; (a) Multitask age and gender classification framework (Yaman, Irem Eyiokur et al. 2019) and (b) Combining ear and face profile for age and gender classification (Yaman, Eyiokur and Ekenel 2022)

2.4.3 Facial Profile Cross-Recognition

A recent study (Toygar et al. 2018) found that facial profiles are more powerful than ear biometrics. However, the research recommended the simultaneous use of facial profiles and ear images from both sides to obtain the optimum result. The study employed two distinct biometric features, namely facial profiles and ear biometrics. Experiments were carried out utilising the feature extraction techniques of principal component analysis (PCA), SIFT, local binary pattern (LBP), LPQ and BSIF to represent the facial profile and ear images. Using the ND-Twins-2009-2010 and Universally Better Economical Rides (UBEAR) databases, a number of experiments were conducted. Two fusion schemes were used in the process, that is, feature-level fusion and score-level fusion. As mentioned above, based on the experiment's results, face profile recognition systems were shown to be more effective than ear recognition systems in the presence of symmetry. This is because the right and left facial profiles are more similar and share more discriminative traits than the right and left ears. For instance, when using the UBEAR database and BSIF method to train on the left ear and test the symmetry on the right ear, the accuracy was found to be 71.73%, whereas the accuracy was found to be 81.52% when the training was conducted on the left profile and the test on the right profile.

It is noted that none of the studies described above address bilateral symmetry using deep learning based method. In this study, we use bilateral symmetry to address scientific questions, such as whether we can recognise or identify a right/left facial profile (probe) using a left/right facial profile (gallery) in Chapter 6.

2.5 Soft Biometrics

In criminal investigations, law enforcement organisations must first identify the suspects by gathering eyewitness descriptions of their physical traits, such as their facial features, build and clothing (Sporer 1992). Therefore, an accurate description provided by an eyewitness can affect the identification of the suspect. As a result, these testimonies need to be integrated into a computer vision/machine learning system that analyses facial profile images numerically.

In this section, we introduce soft biometrics for human recognition to illustrate the effectiveness of soft biometrics on facial recognition in general. We will also discuss soft biometric approaches in detail and also distinguish between categorical and comparative soft biometrics.

2.5.1 Human Recognition/Identification Using Soft Biometrics

The identification and recognition of individuals are becoming increasingly important around the world for security and control purposes. Organisations, agencies and individuals are thus creating technologies to improve this recognition in diverse contexts. As part of this development, soft biometric systems are becoming increasingly popular due to their ability to bridge the semantic gap between human descriptions and image features. These systems use characteristics that are easily observable by humans, such as gender, hair colour and height, so that the individuals who have been described can be searched for. Eyewitness descriptions can be quantified using soft biometric methods integrated into computer systems for human recognition (Reid, Samangooei et al. 2013). Soft biometrics do not require numerical image analysis; instead, verbal human descriptions are processed for identification and recognition.

The growing popularity of soft biometrics reflects the fact that these systems possess unique advantages over conventional ones. For instance, face recognition at a distance (FRAD) requires excellent frame rates and image resolution to identify individuals (Ao et al. 2009). In contrast, soft biometrics allow people to identify traits from low-frame rate videos with a low resolution, even when a subject is seen from an arbitrary angle that makes the frontal perspective unclear. The face attributes provided by eyewitnesses are numerically described by soft biometrics for identification and recognition. In this regard, if the human descriptions are provided by eyewitnesses, soft biometric systems can be used for recognition even in cases where traditional biometric identifiers are not available (Reid 2013).

2.5.2   Soft Biometric Approaches

Biometric recognition systems comprise two distinct subsystems that work together to accomplish security and control. Such systems are prevalent in recognition devices due to their ability to make use of important biometric identifiers, such as fingerprints, facial features and hand geometry, to identify people during authentication. Biometric systems based on soft biometric traits, such as gender and height, have also been attracting attention recently (Jain et al. 2004). The combination of soft biometrics and traditional biometrics has created highly reliable systems, and an increasing number of studies support the addition of soft biometrics to face and fingerprint recognition for the formulation of reliable systems to identify individuals (Reid and Nixon 2010; Samangooei et al. 2008; Martinson et al. 2013; Jain et al. 2004).

In addition, some soft biometric approaches are increasingly considering non-facial elements (Nixon, B. H. Guo et al. 2017), such as body type and clothing, whereas others continue to focus on facial traits. In recent studies, traditional biometrics have been combined with eyewitness testimony (soft biometrics) for superior performance. Furthermore, soft biometric and traditional gait biometric techniques have also been fused to improve the performance of gait recognition systems via both holistic and model-based approaches (Martinho-Corbishley, Nixon et al. 2015).

Soft biometrics first emerged in the late 1990s, and various investigations have since improved and complemented biometric systems. For example, a 2004 study by Jain et al. used a database of 160 subjects to assess the identification of individuals using age group, gender and height as the soft biometrics. Soft biometric techniques were shown to improve the performance of traditional biometric fingerprint systems in recognition and identification by 6% (Jain et al. 2004). Heckathorn et al. (Heckathorn et al. 2001) also found that personal features, including eye colour, gender, race, height and visible marks, such as scars and tattoos, could be used to improve the accuracy of traditional systems in the recognition of individuals. Previous studies on the role of human traits in recognition have thus acknowledged the critical role of soft biometrics in improving traditional biometric systems. Overall, the inclusion of soft biometric traits improves the accuracy and reliability of established biometric systems by considering individuals' physical characteristics.

The ubiquity of surveillance camera networks on business premises has not significantly reduced crime rates, and there are still many crimes committed in which an eyewitness account is the only available description of the perpetrator's appearance. Moreover, surveillance camera networks are often only useful when security agents can search through the videos and a face database to match an eyewitness description. The use of soft biometrics would improve the timely resolution of many crimes by allowing security agents to conduct intelligence investigations. Thus, improving the reliability of biometrics depends on the inclusion of soft biometric traits. Doing so would lead to the

accommodation of pattern recognition technology to allow humans to describe the appearance of a subject's face so that security agents can use this information to search their databases. Klare, Klum et al. (2014) provided a reliable approach to the use of verbal facial attributes described by humans in criminal investigations. They noted that an attribute extraction algorithm could be used to compare various traits, such as eyebrow, chin and eye shapes, based on hand-drawn police sketches from target repositories. This is because the identification of frontal face traits is a critical process in improving the forensic identification of suspects.

The use of biological features in identification is becoming increasingly attractive to different agencies and organisations e.g. tech companies and banks. Dantcheva et al. (2011) described biometrics as the use of physical or behavioural traits, such as signature, gait, iris, voice, fingerprints or face, to recognise individuals. However, biometric systems encounter problems due to limited discriminability, overlaps between identities, poor data quality and missing information (Dantcheva et al. 2011).

2.5.3 Categorical versus Comparative Soft Biometrics

For recognition when using soft biometrics, we need a scheme to assign a score to subjects' traits. This scoring scheme could be categorical or comparative. Such scores could then be used as feature vectors for recognition purposes. In a categorical framework, a trait is put in a certain category. For example, height can be classed into 10 categories, from the shortest to the tallest height. Each person needs to be assigned to one of these groups. However, in a comparative framework, the scores are assigned by comparing the traits of two individuals. For example, the heights of two individuals are compared (e.g. subject A is taller than subject B) and ranked using a ranking algorithm, such as Elo [1](Reid, Nixon and Stevenage 2013; Nawaf Yousef Almudhahka et al. 2017b) or RankSVM (Parikh, Kovashka et al. 2012; Martinho-Corbishley, Nixon et al. 2016). The most effective is the comparative method, as no human error can be made by assigning an incorrect score to a trait. However, human errors may occur when assigning an attribute to a certain category.

Reid and Nixon originally introduced comparative soft biometrics, which are based on comparisons of subjects (Reid and Nixon 2011). The development of this approach could enhance the reliability and effectiveness of scoring schemes for attributes in soft biometrics. Comparative soft biometrics generate biometric signatures and feature vectors for subjects to be used in recognition and identification. In addition, comparative descriptions are a powerful method of obtaining reliable human descriptions based on eyewitness information and soft biometric evidence (Reid, Nixon and Stevenage 2013). The use of comparative soft biometrics to describe people also allows for the recording of numerous

---

[1]The Elo rating system is named after the Hungarian-American physics professor (and 8-time state champion chess master) Arpad Elo, who originally devised his rating system around 1960.

differences and an accurate representation of traits for more precise recognition. In 2011, Parikh et al. introduced the concept of 'relative attributes', which provides a way to rank images according to attribute (trait) intensity using class-level ordered pairwise comparisons (e.g. 'bears are furrier than giraffes') (Parikh and Grauman 2011).

Several previous studies have acknowledged that comparative traits could be used for recognition. For instance, a study conducted by Reid in 2013 required 63 contributors to compare one subject with another (the study had 50 subjects in total) and used the comparisons to collect traits. The collected traits included a detailed consideration of facial components and led to a 96.7% identification accuracy (Reid 2013). In another study, Samangooei et al. used global and body soft biometric traits to demonstrate the importance of comparative soft biometrics in human descriptions of low-quality surveillance footage (Samangooei et al. 2008). There is additional documented evidence of the importance and benefits of comparative attributes over categorical attributes in the formulation of reliable recognition systems (Reid, Nixon and Stevenage 2013).

2.5.4 Impact of Facial Soft Biometric on Recognition

The testing phase of face biometric systems has long been completed successfully, and currently, these systems are widely used in practice. The effective use of biometrics requires the identification of multiple traits that are visible from a variety of angles. However, most of these traits can be found in face images and recordings of facial movements, meaning that the use of a minimum number of features can facilitate the completion of recognition. Ensuring there is a collection of frontal face traits can be critical in successfully recognising or identifying a subject based on specific information. However, a facial profile reveals some characteristics of the facial structure that cannot be observed in a frontal face image. These aspects include the nose, chin and cheekbones, which can have an impact on the recognition rate.

Facial verification uses established features that make a subject unique. At the same time, descriptive visual attributes are used as conspicuous labels to describe the appearance of an image (Kumar et al. 2009). There are also features of the face that can be used in different domains, for example, nose size, jaw shape, age and gender. Although descriptive visual attributes obtained from the frontal face are not unique, their dimensions or underlying traits mean they can be used to describe a unique subject. Kumar et al. noted that attribute-based representation has specific benefits for vision tasks because attributes are manifold, meaning they can be used to create different levels of descriptions. A classifier can thus generalise them by learning them once and using the obtained knowledge to identify new categories, objects and subjects without the need for additional training (Kumar et al. 2009).

The frontal face has been used by investigators in the majority of studies due to the rich information it contains, which helps in the detection and recognition of subjects, even in uncontrolled situations (Reid, Samangooei et al. 2013; Nixon, Correia et al. 2015; Gonzalez-Sosa et al. 2018). For example, in one study using frontal face images Almudhahka et al. used a variety of algorithms to assess the role of soft biometric attributes in the performance of the identification process (N. Almudhahka et al. 2016), and Kwon and Lobo attempted to classify subjects as 'babies', 'young adults' or 'senior adults' using age classification from facial images. To do so, Kwon and Lobo analysed the skin wrinkle and feature-position ratios, chiefly those found on the frontal face, to complete their classification (Kwon et al. 1999).

The first methods to face verification used soft biometrics, demonstrating that excellent performance is required for an enhanced understanding of face, age and gender. Consequently, many researchers have also assessed the frontal face as a critical component in soft biometrics. Almudhahka et al. investigated the topic on a large scale in the context of human identification using comparative facial soft biometrics on a dataset with 4,038 subjects (Nawaf Y Almudhahka et al. 2016). They reported that a rank-10 identification rate of 96.98% along with a verification accuracy of 93.66% could be achieved with only 24 soft biometric traits and 10 comparisons. Furthermore, the role of the frontal face in facial identification and verification has received copious attention, leading to an improved understanding of numerous traits that can be obtained from facial movement recordings and images.

In addition, the effectiveness of side-view images has also been assessed in various studies. For example, Ghalleb et al. (2016) investigated the utilisation of facial and body soft information to improve the performance of a face verification system in remote acquisition circumstances. Specifically, the face verification method used the Discrete Wavelet Transformation. Facial and body soft biometric traits including skin colour, shoulder width, head width and height and waist width for side view were used. Adding soft traits to the hard biometric system reduced the equal error rates from 19.30% to 3.00% and 28.00% to 7.48% for frontal and side views, respectively when using 360 images of 18 subjects.

This literature review demonstrates the necessity of introducing and investigating soft biometrics in side-view databases, as discussed further in Chapters 3 and 4. Furthermore, the results also highlight how crucial it is to automatically perform comparative soft biometrics and potentially bridge the semantic gap between human (or eyewitness) testimonies and computer vision for facial profiles. We discuss this further in Chapter 5. Consequently, the objective of this thesis is to improve earlier studies by investigating the use of comparative facial profile soft biometrics for human recognition and automatic retrieval. The main focus is on performance assessments, the development of soft facial profile biometrics for use in novel applications and the identification of facial profile traits that could be employed for advanced retrieval and identification.

2.6 Conclusions

The field of facial profile biometrics is expanding, with recent studies investigating topics such as pose and rotational variability, illumination and occlusion. In addition, the process of determining subtle characteristics from facial profiles is still in its infant stages of development. In this chapter, a review of the biometrics associated with the face profiles was presented. We discussed different methods of identifying facial profiles as well as the role that facial profiles play in other classification areas, such as age and gender. The chapter also presented an outline of soft biometrics and their significant function in investigations based on eyewitness testimonies. This aspect ensured a compelling explanation and evaluation of the existing research concerning soft biometrics. In addition, the literature review compiled a list of the most important studies on biometric facial profile recognition and identification while pointing out the shortcomings of each of these studies. In summary, the review discussed the implications of the existing knowledge gaps identified in the literature to formulate the objectives of this research project.

The literature shows that using comparative comparison on soft biometrics outperforms the categorical one. Furthermore, soft biometric systems perform exceptionally well when used to identify individuals in a variety of domains such as face, clothing, and gait. The literature shows that there is some study on deep learning towards facial profiles recognition. However, it should be emphasised that deep learning-based features have not been utilised before to investigate both facial profile recognition and bilateral symmetry on facial profiles. In addition, no previous studies on evaluating the impact of soft biometrics using facial profile data.

Chapter 3

Comparative Facial Profile Soft Biometrics

3.1 Introduction

This chapter provides an introduction to comparative soft biometrics and outlines the approach used to crowdsource systematic comparative interpretations of soft biometric human facial qualities, along with its potential application in human image recognition. It summarises the acquisition of relative labels through crowdsourcing, which facilitated the collection of data and information from a diverse and international pool of interpreters, accounting for the anticipated response variation in surveillance settings. Additionally, all of the responses from the crowdsourcing are analysed. The chapter concludes with a discussion on ranking algorithms and their application to pairwise comparisons. Moreover, this study builds on earlier research in relative attributes and soft biometrics, as discussed in Sections 2.5.

The remainder of this chapter is organised as follows: Section 3.2 introduces the database used throughout this research, while Section 3.3 provides a more detailed explanation of the concept of relative attributes; Section 3.4 discusses the experimental design for capturing facial profile attributes and defines the key attributes required to construct and collect labels for a dataset from human annotators; Section 3.5 describes the crowdsourcing annotation task using the Appen platform, assesses the stability of crowdsourced facial profile soft biometric traits, analyses their distributions, and determines their level of uncertainty; Section 3.6 discusses different ranking techniques that can be used to obtain the signature features for each subject in the database; Finally, Section 3.7 provides an overview of our findings.

3.2   XM2VTSDB Dataset

After an extensive search for a dataset that included poses at 90-degree angles to support the purpose of this research, we identified several relevant datasets to use in the analysis. These datasets included VGGFace2, LFW, UMDFaces, and XM2VTSDB (as outlined in Section 2.3). The main goal during the search process was to find a dataset that had sets of four or more facial profile images of the same subject taken over a period of time. Additionally, we wanted to make sure that the other side of the face was completely hidden in order to avoid any information related to that from being disclosed. Ultimately, the XM2VTSDB dataset proved to be the best option. This dataset contains one left and one right profile image (sample) per person per session, totalling 2,360 images. Each image is 720×576 pixels.

The XM2VTSDB database [2], a research resource established and maintained by the University of Surrey (see (Messer et al. 1999) for details), is an extended version of the M2VTS database, as it comprises more video recordings of each subject during each session compared to M2VTS. Participants from the XM2VTSDB database attended four sessions and the database was developed over a significant period, enabling a wide range of appearance variations of the individual subjects (see Figure 3.1). For example, there were variations in participants' face shape, facial hair, hairstyles and the use of glasses. These factors made it a challenging dataset to work with for recognition. However, upon examining the dataset, we discovered that 230 out of the 295 subjects met our criteria and could be included in the analysis. The rationale for choosing this dataset for our research study was that the images only show the facial profiles of the subjects to the biometric system, without any other part of the face being exposed to the system or to eyewitnesses. In contrast, other face datasets such as LFW also reveal some parts of the other side of the facial profiles, making the LFW dataset less challenging than the XM2VTSDB dataset in terms of facial profile recognition.

XM2VTSDB includes a fairly significant number of subjects. However, the variety of XM2VTSDB images is extremely limited: the lighting is uniform, the majority of the individuals are white, the facial expressions are mainly neutral, and the positions in some images are not strictly 90 degrees (Stephen Milborrow et al. 2010). Furthermore, it only includes four samples per participant per side view.

3.2.1   Occlusion-based Facial Profile Data

Occlusion presents a significant challenge across various domains, including object detection, crowdsourcing, and soft biometrics. Within face recognition, a primary obstacle is the detection of facial features obscured by occluded areas (Ge et al. 2017). Currently,

---

[2]https://store.surrey.ac.uk/product-catalogue/feps-faculty-of-engineering-physical-sciences/ electronic-engineering/xm2vtsdb-multi-modal-face-database.

FIGURE 3.1: Sample facial profile images from the XM2VTSDB dataset.

researchers are exploring methods for reconstructing images with and without occlusion. An occluded face refers to one partially obscured from view, often by an object such as a mask or due to obstructed camera footage. Notably, the XM2VTSDB dataset encompasses various occlusions that impact identification performance, including ear, neck, and eyebrow occlusions, as well as instances where glasses obscure the eyes (see Figure 3.2).

(a)   (b)   (c)   (d)
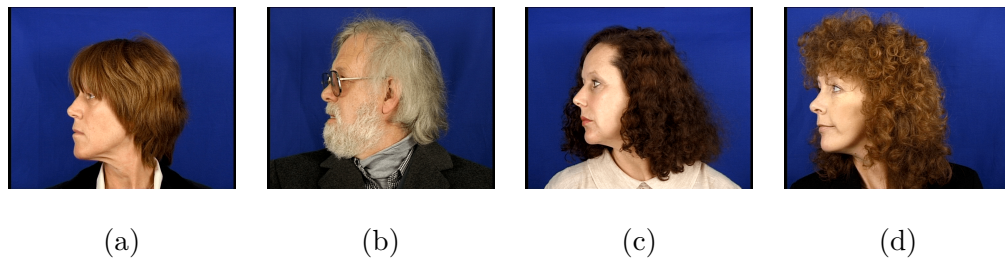
FIGURE 3.2: Sample facial profile images from the XM2VTSDB dataset where the effect of some occlusions can be noted: (a) Ear & Eyebrow; (b) Eye & Neck; (c) Ear & Neck; (d) Ear & Eyebrow.

3.3 Relative Attributes

Recent research has explored leveraging relative information to enrich face description of objects within images. Various strategies have been used to exploit the similarities between objects as a descriptive tool. In 2009, Kumar et al. achieved superior performance compared to traditional binary classifiers by employing pairwise "simile attributes" to describe face regions in relation to a set of reference appearances, for example, *"a mouth like Harrison Ford"* (Kumar et al. 2009), as depicted in Figure 3.3. In addition, G. Wang et al. (2010) identified objects based on similarities, with few or no examples. Through descriptions such as *"a zebra is comparable to a horse in shape and a crosswalk in texture"*, the technique is able to detect zebras without the need for any training examples. It has also been demonstrated that using similarity descriptions can enhance object recognition in images, even when training data is limited. Although both approaches use relative information to improve descriptions, they differ from our method. While the similarity between reference subjects or objects facilitates description, our approach enables sorting based on specific traits through subject comparison.
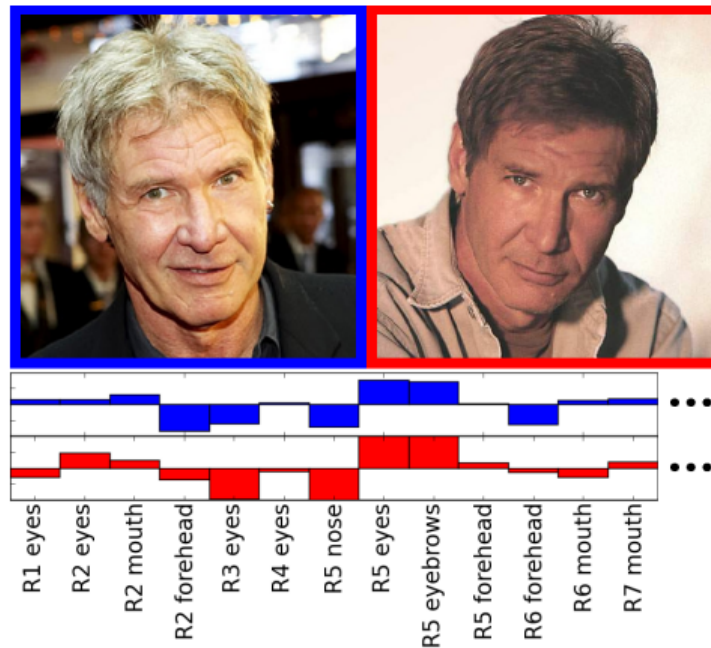
FIGURE 3.3: Using simile attributes to describe visual appearance (Kumar et al. 2009).

To date, there has been a lack of evaluation of facial profile retrieval using comparative soft biometrics based on verbal descriptions. While individuals present their perceptions regarding the inclusion or exclusion of facial profile characteristics in their face descriptions, comparative attributes are crucial for assessing one face relative to others.

As discussed briefly in 2.5.3, the application of comparative soft biometrics in the context of human identification has revealed that relative (or comparative) attributes hold

more advantages compared to categorical attributes (Parikh, Kovashka et al. 2012; Reid, Samangooei et al. 2013; Reid, Nixon and Stevenage 2013; Reid and Nixon 2011). It has been found that presenting attributes in a relative manner offers several benefits (Parikh, Kovashka et al. 2012). In general, comparative attributes help overcome the challenges associated with describing nuanced traits. Additionally, creating a comparison provides a reference point. For instance, determining whether the nostril size of Subject A is wide or narrow becomes more challenging without comparing it to an established reference. Introducing Subject B renders the description of the binary attribute more meaningful by establishing a consensus that the nostril of Subject A is, for example, narrower than that of Subject B. Consequently, utilising comparative attributes in soft biometrics improves agreement by bringing clarity to descriptions. Figure 3.4 illustrates how employing relative attributes can enhance descriptive content.

A is a male   **B** is a female

(a) : Categorical labels

A is more masculine than **B**   **B** is more feminine than **A**

(b) : Comparative labels
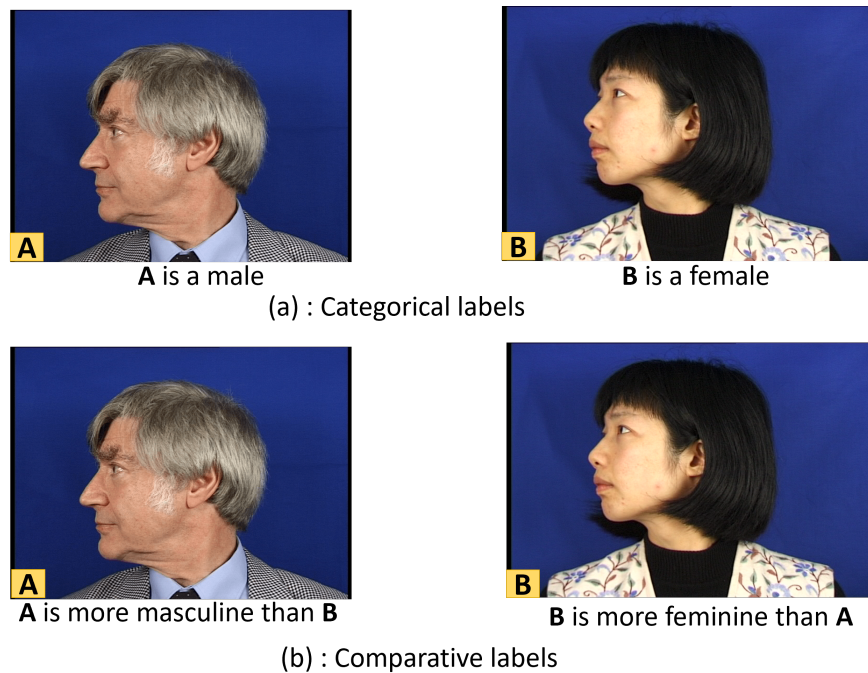
FIGURE 3.4: Gender of samples from the XM2VTSDB database expressed using two methods: (a) categorical labels and (b) comparative labels.

3.3.1 Label Comparisons of Facial Profiles

Modern biometric data analysis machines employ a model that is similar to the one used by people to formulate descriptions. Comparative labels enable both systems and individuals to compare the relative features of a group of people. Consequently, soft biometric attributes should be associated with an established comparative label. In addition, Samangooei et al. observed that researchers are continually endeavouring to delineate the scope of soft biometrics using various comparative labels, such as the 5-point bipolar scale (Samangooei et al. 2008).

The increased popularity of the bipolar scale as a comparative label has sparked mixed opinions regarding its usage. Almudhahka et al. utilised the findings of an experimental study to assert that compressing a 5-point Likert scale into a 3-point scale improved the identification rate, with certain levels being disregarded (Nawaf Yousef Almudhahka et al. 2017b). The researchers also advocated for the elimination of minor details to decrease the variation in data collection and analysis. The current study uses a 4-point bipolar scale for the comparative labels associated with specific attributes as in (Nawaf Yousef Almudhahka et al. 2017b), following a consistent format: 'More A/Less B', 'Same', 'More B/Less A' and 'Cannot see'. The label values are 1 for 'More A/Less B', 0 for 'Same', –1 for 'More B/Less A', and -2 for 'Cannot see'.

3.4 Experiment Design to Capture Facial Profile Attributes

This evaluation utilised comparative facial profile soft biometrics to identify individuals alongside existing soft biometric features. Together, these describe critical traits of the human face (N. Almudhahka et al. 2016; Nawaf Y Almudhahka et al. 2016; Nawaf Yousef Almudhahka et al. 2017b; Reid and Nixon 2011) (e.g., eyebrow, eye, nose, and neck), enabling the definition of 23 attributes relevant when determining the identity of each facial profile. Furthermore, this study suggested incorporating 10 additional attributes to describe a facial profile. These attributes include nostril size, nose tip, nose size, face profile length, face profile width, ear orientation with respect to the head, ear–head ratio, ear–chin distance, ear–nose distance, and skin condition. It is important to note that this study is inspired by the related literature (Martinho-Corbishley, Nixon et al. 2015; N. Almudhahka et al. 2016), which considers skin colour and gender as comparative human traits. The research conducted by Reid et al. was also significantly influenced by police witness evidence forms, which inspired them to define these traits using both multi-class and traditional binary labels (Reid, Nixon and Stevenage 2013). This approach can lead to homogeneous labelling in the case of an unbalanced dataset or if cross-race influences the annotation (Nixon, Correia et al. 2015). Hence, the primary objective of comparative annotation is to mitigate such effects by providing an objective description of the differences between two subjects.

The description of each feature must include the formulation of a suitable set of comparative labels to define the attributes. Table 3.1 presents the attributes derived from the facial profile scale. The measurements were evaluated using the 4-point bipolar scale described in Section 3.3.1. This approach proved to be effective in formulating and describing scores, as a numerical value enabled the creation of an ordered list for each trait as a vector, which could be used to rank subjects.

This experimental evaluation involved selecting a suitable subset of subjects from the dataset. Ultimately, the experiment included a total of 230 subjects, each with four facial profile samples, sourced from the XM2VTSDB dataset (Messer et al. 1999). These samples from the database facilitated the assessment of recognition accuracy by mapping features to images in the dataset.

3.5 Crowdsourcing of Comparative Facial Profile Traits

Many previous studies have used crowdsourcing platforms to annotate datasets because this approach offers a reliable method for analysing characteristics and labels (N. Almudhahka et al. 2016; Nawaf Yousef Almudhahka et al. 2017b; Samangooei et al. 2008). Crowdsourcing generates a vast collection of responses, accompanied by high-quality comparative annotations, as the data are gathered by a large number of annotators with diverse cultural, linguistic, and national backgrounds. Welinder and Perona point out that the advancement of online platforms also enables the assignment of annotation tasks to numerous individuals who are computer-literate, resulting in the production of verifiable and reliable outcomes within a few hours (Welinder et al. 2010). In general, crowdsourcing platforms are reliable systems for acquiring extensive human knowledge and performing tasks based on visual understanding (Welinder et al. 2010).

In this study, we designed and executed a crowdsourced annotation task for the collection of labels using the Appen platform [3]. The Appen platform allowed us to manage the number of tasks each user could handle and enabled us to present 'gold-standard' questions with predetermined responses to test the contributors. By distributing the analysis and encouraging participants to provide a variety of answers, the platform ensured the collection of high-quality annotations while filtering out any fraudulent responses.

3.5.1 Question and Response Design

The crowdsourcing method was used to determine the comparative facial profiles of the samples obtained from the XM2VTSDB dataset. In each task, the contributors on the online platform (referred to as labellers) used the labels listed in Table 3.1 to compare attributes of two facial profile images. Figure 3.5 provides an example of a crowdsourced comparison that was conducted on the Appen platform. During the annotation process, the labeller was required to compare a subject on the left with another subject on the right, using predetermined comparative labels. A total of $33 \times \binom{920}{2}$ unique annotations were collected, with each subject being compared to two other subjects in terms of each trait. The questions used in the crowdsourcing platform followed a psychometric

---

[3]https://appen.com/.

TABLE 3.1: Soft facial profile biometric attributes and possible associated response labels.

No.	Soft Traits	Comparative Labels			
		1	0	-1	-2
1	Eyebrow length	More Long	Same	More Short	Cannot see
2	Eyebrow shape	More Raised	Same	More Low	Cannot see
3	Eyebrow thickness	More Thick	Same	More Thin	Cannot see
4	Spectacles	More Covered	Same	Less Covered	Cannot see
5	Eye-eyebrow distance	More Large	Same	More Small	Cannot see
6	Eye lashes	More Long	Same	More Short	Cannot see
7	Eye size	More Large	Same	More Small	Cannot see
8	Philtrum size	More Long	Same	More Short	Cannot see
9	Nostril size	More Wide	Same	More Narrow	Cannot see
10	Nose tip	More Pointed Down	Same	Less Pointed Down	Cannot see
11	Nose size	More Large	Same	More Small	Cannot see
12	Lips thickness	More Thick	Same	More Thin	Cannot see
13	Face profile length	More Long	Same	More Short	Cannot see
14	Face profile width	More Wide	Same	More Narrow	Cannot see
15	Skin smoothness	More Smooth	Same	Less Smooth	Cannot see
16	Skin condition	More Clear	Same	More Pimples	Cannot see
17	Forehead hair	More Forehead Hair	Same	Less Forehead Hair	Cannot see
18	Ear size	More Large	Same	More Small	Cannot see
19	Ear orientation with respect to head	More Further from head	Same	More Closer to head	Cannot see
20	Ear-head ratio	More Large	Same	More Small	Cannot see
21	Ear-chin distance	More Further	Same	More Closer	Cannot see
22	Ear-nose distance	More Large	Same	More Small	Cannot see
23	Cheek shape	More Flat	Same	More Prominent	Cannot see
24	Cheek size	More Large	Same	More Small	Cannot see
25	Chin & jaw shape	More Receding	Same	More Protruding	Cannot see
26	Double chin	More Large	Same	More Small	Cannot see
27	Chin height	More Large	Same	More Small	Cannot see
28	Neck length	More Long	Same	More Short	Cannot see
29	Neck thickness	More Thick	Same	More Thin	Cannot see
30	Age	More Old	Same	More Young	Cannot see
31	Gender	More Masculine	Same	More Feminine	Cannot see
32	Skin colour	More Dark	Same	More Light	Cannot see
33	Figure (Shape)	More Fat	Same	More Thin	Cannot see

procedure, wherein contributors were presented with two images of different individuals and asked to compare them (Martinho-Corbishley 2018).

FIGURE 3.5: Example of a question used on the Appen platform.

The questions and comparisons on the online platform were intentionally designed to be simple and straightforward, ensuring accuracy and reliability. To improve response speed and encourage completion of the questions, the answer form was designed with vertically aligned radio buttons. For each question, there were 15 different contributors assigned to provide answers. To simplify the questions that were difficult to distinguish, we included a 'Cannot see' option. This option was crucial in reducing the likelihood of inaccurate responses when contributors lacked confidence in their answers. Each contributor was limited to answering up to 30 pages, with each page containing 10 image annotation categories. In order to assess the accuracy of the responses and eliminate false answers, test questions were designed and presented to the labellers. Additionally, each page included a hidden test question and nine annotation tasks to ensure continuous monitoring of the reliability of the respondents. Test questions were tried out in various subsets to evaluate their acceptability. Accordingly, the following points were considered:

- The response rate for 'Cannot see' was limited to 20% for all annotations. Those who exceeded this limit were removed from the process.

- Response distributions that deviated significantly from the average distributions during initial trials were rejected.

- Before participating in the task, the labellers' accuracy levels were tested. Only contributors with a reliable level of accuracy were allowed to perform the comparison task. Failure to pass the 'quiz mode' on the online platform disqualified contributors from participating and prevented them from being paid (Center 2019b). To ensure fairness, we selected questions from the most obvious comparisons and only rejected responses that were fundamentally incorrect. In short, respondents needed to score 80% or higher in 'quiz mode' to proceed.

- The contributors selected were from Level 1 [4], which represented the largest number of participants.

3.5.2 Annotation Response Analysis and Discussion

This study collected a sufficient number of responses from the crowdsourcing platform. The cost of the trial used in the analysis was \$4,554 and yielded 455,400 judgments for the set of 230 subjects (920 images). Each question was answered by 15 different labellers. Labellers were paid \$0.01 USD (1 cent) for each image comparison. Figure 3.6 provides a description of the outcomes of the comparative analysis and a summary of the distribution of annotations after comparing 230 subjects. The number of responses using the term 'Same' as a description was relatively low in comparison to those using 'More' or 'Less'. However, there were a high percentage of 'Same' labels for gender, which could be attributed to its binary nature. Overall, it appeared that gender was rarely straightforward for all contributors, with a growing number of responses indicating 'significantly More Feminine/Masculine' and almost half indicating 'Same'. However, it is worth noting that the annotators tended to refrain from selecting 'Same' when referring to the nose tip, eyebrow thickness, and nose size. Figure 3.6 illustrates the overall distribution of the collected labels. However, it is important to note that the significance and informative value of the attributes may not be accurately represented in this data. Therefore, these aspects are discussed in detail in the next chapter.

The analysis of the distribution of inferred relative measurement stability, along with the collected annotations, demonstrates the suitability of soft biometric traits and the effectiveness of the crowdsourcing task. Figure 3.6 provides a summary of the confidence scores, which are the answer scores given by the contributors. These scores were generated by comparing facial profiles. The confidence score indicates the level of agreement among different contributors and was calculated by weighting their scores. Specifically, equation 3.1 was used to determine the confidence score for response (X) where CS_X is the confidence score, S_X is the sum of the scores of all the contributors who have responded (X) and S_C is the sum of the scores of all contributors:

---

[4]The levels indicate the level of experience the contributors needed to perform the task, with Level 1 being the largest pool of contributors and Level 3 being the smallest.
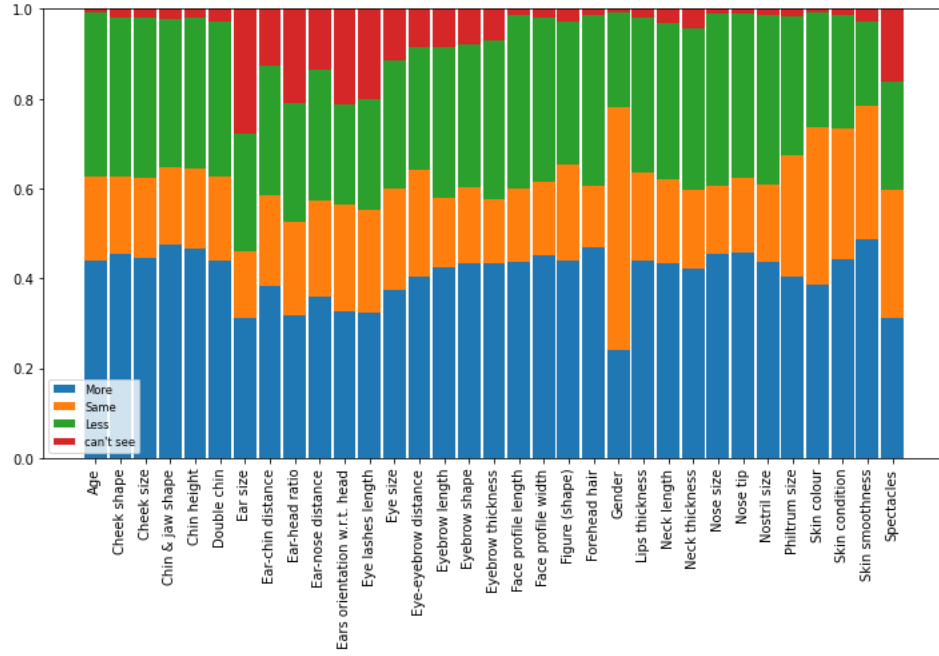
FIGURE 3.6: Distribution of collected comparative facial profile labels for 230 subjects.

$$CS_X = \frac{S_X}{S_C} \tag{3.1}$$

The response X can be categorised as 'More A/Less B', 'Same', or 'Less A/More B'. Formally, we can represent $X \in \Omega_{(A/B)}$ as more A or less B, same, less A or more B. The confidence score indicates the level of confidence in the results obtained through the crowdsourcing platform. The final results were determined by selecting the answer with the highest confidence (Center 2019a). Figure 3.7 displays the confidence scores for the two subjects in relation to the 'average comparison of nostril size' and shows that the two subjects had confidence scores between 0.73 and 0.8 for this question.

Following the completion of the crowdsourced comparison task for the facial profiles, two ranking methods were employed to enhance the ratings of the subjects' facial profile attributes. These rankings were based on the confidence scores. For example, a high score for 'nostril size' suggested that the nostril is wide, and the majority of participants indicated 'More wide'.

3.5.3 Inferring Additional Comparisons

Initially, each image (sample) of a subject was compared 15 times with images of two different subjects in the same gallery for a given trait. From these comparisons, we

---

[5]The aggregated report includes one result for every row. Since multiple contributors answered each row, the aggregated CSV aggegated contributor's judgements based on individual contributors' trust ratings.
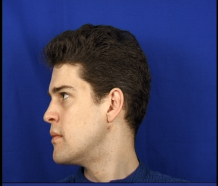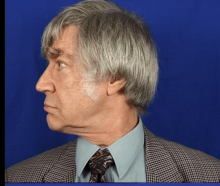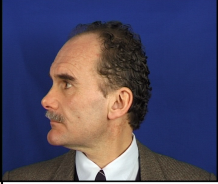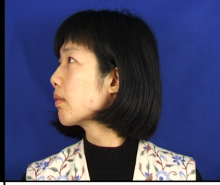
Subject 1		
Subject 2		
Average comparison of nostril size	More Narrow	More Wide
Confidence score	0.73	0.8

FIGURE 3.7: Two examples of the confidence score output for the 'average comparison of nostril size' from a sample aggregated report [5].

calculated an average score ±1, 0 using a threshold. This allowed us to obtain a small set of image pairs that could be used to determine the ranks based on the transitivity of order relations and draws. To illustrate, let us consider three different subjects *A, B, C* , and a trait intensity function *t*. In this case, the following relationship 3.2 holds:

$$
\begin{aligned}
&\text{If } t(A) > t(B) \quad and \quad t(B) > t(C) \quad then \quad t(A) > t(C) \\
&\text{If } t(A) = t(B) \quad then \quad t(B) > t(I) \quad \forall I \, s.t. \quad t(A) > t(I)
\end{aligned} \tag{3.2}
$$

Without loss of generality, we can illustrate these conditions using a specific subject *A* and a comparison score of $+1$ (or the $>$ relation). In our implementation, we applied the aforementioned conditions interchangeably to both *A* and *B* and for scores of ±1.

The implementation of the second condition was straightforward. Specifically, it involved adding new comparisons by replacing *B* with *A* for every comparison that included *B* if $t(A)=t(B)$. However, if implemented as it is, the transitivity calculations can be time-consuming. Therefore, to achieve an efficient implementation, we modelled the comparisons using nodes in a graph and utilised elements of graph theory to locate comparable nodes. The main concepts behind this process are as follows:

- For a chosen trait *q*, an arc (*A*,*B*) with weight 1 exists only if the comparison (*A*,*B*) with outcome $+1$ (or $t(A) > t(B)$) exists in the set of aggregated comparisons.

- If a node *M* is connected to another node *N*, then $t(M) > t(N)$.

Consequently, the process reduces to constructing a graph of similar subjects based on a comparison score of +1 and identifying connected components. The latter can be accomplished by utilising the power properties of the adjacency matrix (Cormen et al. 2001). In our implementation, we transformed negative outcomes (-1) into positive ones by inverting the images in the trait comparison (pairwise comparison). Subsequently, we calculated the powers of the adjacency matrix for values ranging from 2 and the number of subjects in the gallery. The algorithm was terminated once the resulting adjacency matrix no longer updated. Then, we repeated the aforementioned steps, taking into account negative results where nodes A and B were connected with a weight of 1 if there was a comparison (A,B) with an outcome -1. Finally, we gathered the newly created comparisons and their outcomes from the non-zero cells of the power matrix that were initially empty in the adjacency matrix.

The following advantages are associated with the additional comparisons in the dataset:

- The approach is cost-efficient because it starts with a minimal set of comparisons. Furthermore, analysing the power matrix can serve as an indicator of whether more data needs to be collected. In fact, if the aggregated crowdsourcing comparison mostly contains equalities, it will result in gaps or sparse rows in the final power matrix. In such cases, a human agent or algorithm can visually identify a few comparisons that could complete the data. These comparisons can be considered the missing links between connected components.

- Connected components represent pre-ranked subjects. When all nodes belong to the same connected component, it indicates a complete ranking for the data.

Once the comparisons have been computed, we can proceed to rank the subjects in each gallery individually. The ranks will be stored in a matrix, where each column represents a trait and each row corresponds to a subject, storing its ranks for all the traits in the gallery. We consider the ranking step a feature vector with the threshold of averaged scores that serves as a hyper-parameter that needs to be validated.

3.6 Ranking Methods

In practical situations, ranking is a crucial action that involves rating objects in relation to others based on a preset criterion and creating a corresponding order for them. Ranking has various real-life applications, including evaluating products, ranking sports teams (such as football teams participating in world cups) and players in games (such as chess), retrieving documents, and rating movies. Due to the availability of large datasets with limited resources, there is a need for automated ranking. Therefore, prioritising ranking objects over classifying them is essential. Given a database containing pairwise

items, the goal of ranking is to sort the items based on their relative strengths and a given criterion of interest that can be inferred from the comparisons. For example, a pairwise comparison between n items could be whether Person A is older than Person B. However, in an actual database, only a subset of pairwise comparisons is possible, while noise within the dataset remains constant. Human subjectivity in labelling is a common source of noise in databases. Section 3.5, which describes the crowdsourcing of comparative labels, illustrates some of these issues. The evaluation of semantic stability in our study showed discrepancies in the judgements of the annotators, and some of the pairwise comparisons were obtained through different methods. Consequently, the ranking algorithms selected had to address the challenges of missing comparisons and data contamination caused by noise. More importantly, the algorithm needed to demonstrate sufficient computational efficiency to effectively handle big data in real-time applications.

Various machine learning approaches can be utilised to rank objects (Negahban et al. 2012; Yun et al. 2014; Fogel et al. 2014; Hunter et al. 2004; X. Chen et al. 2013; Arasu et al. 2002; Glickman 1995; Joachims 2002; Freund et al. 2003; Fu et al. 2015; Herbrich et al. 2006; Wauthier et al. 2013). Ranking algorithms are generally classified into four groups: score-based ranking, learning to rank, maximum likelihood estimation and spectral ranking (Fogel et al. 2014). As the name suggests, score-based ranking uses existing rating systems to assign scores based on pairwise comparisons and then uses those scores to determine the order of items. The Elo rating system (Glickman 1995) is an example of a score-based ranking algorithm. It assumes a normal distribution of skill levels for players and then calculates rankings based on the difference between actual and expected game outcomes. Initially developed for ranking chess players, this system is now employed in various applications, including soft biometrics identification (Nawaf Y Almudhahka et al. 2016; Nawaf Yousef Almudhahka et al. 2017b; N. Almudhahka et al. 2016; Reid, Nixon and Stevenage 2013). Another score-based ranking algorithm, called TrueSkill (Herbrich et al. 2006), tracks changes in a player's skill and updates their confidence over time. Both the Elo system and TrueSkill use an unsupervised approach, but they require parameter tuning to fit different datasets. Parameter tuning is the process of altering the values of the parameters in a model to improve its performance, which is important for enhancing computational efficiency. For example, the Elo system requires tuning the score adjustment parameter K. The K-factor is a crucial value in the calculation of chess ratings. To clarify, the maximum number of points that can be won or lost in a single chess game is indicated by this number. Therefore, this number also signifies the rate at which the player's rating will either increase or drop. Moreover, the parameter optimisation technique can be used to determine the optimal value of K for a specific problem. Other studies, such as (Wauthier et al. 2013; Huber et al. 1963), have also developed computationally efficient algorithms that use point differences to rank objects using a score-based method.

Learning algorithms for ranking are based on adopting a different approach, namely inferring a scoring function from a training dataset and then using that function to determine rates for unseen data. As a result, the first step is to infer the scoring function. Several ranking algorithms have been developed using this approach, such as RankBoost (Freund et al. 2003), RankSVM (Joachims 2002), and Unified Robust Learning to Rank (URLR) (Fu et al. 2015). These algorithms are examples of supervised learning as they use a scoring function derived from an example pairwise feature comparison. In addition to training the rating models, learning to rank algorithms also requires parameter tuning. For example, RankSVM has multiple parameters which need to be optimised and tuned to address a specific problem e.g. C which is the trade-off constant between maximising the margin and satisfying the pairwise constraints. Applying these algorithms has also yielded positive results. For instance, RankSVM has been used in soft biometric identification to generate accurate biometric signatures (Martinho-Corbishley, Nixon et al. 2015). Furthermore, it has been used to facilitate subject identification and retrieval by deriving relative rates from comparative clothing features (Jaha et al. 2014).

The estimation of rates using the maximum likelihood method can help to rank pairwise comparisons. This approach closely resembles the Bradley–Terry probabilistic model (Hunter et al. 2004). Regarding spectral approaches, linear map theory can be utilised to analyse relationship matrices and facilitate ranking. For example, the PageRank algorithm considers the connections between items in a graph to determine their rank and subsequently employs a random walk to evaluate the probability of a particular node being visited (Arasu et al. 2002). Another recently developed spectral algorithm, SerialRank (Fogel et al. 2014), enables ranking by using similarity matrices created from pairwise comparisons.

The literature review above illustrates the variety of methods proposed for object ranking. It is important to highlight a key finding: the majority of these algorithms necessitate some form of learning. Furthermore, it is clear that algorithms which require parameter tuning have been favoured for ranking soft biometric attributes. Tuning is also utilised in the Elo rating system (Reid, Nixon and Stevenage 2013; Reid and Nixon 2011; N. Almudhahka et al. 2016) and RankSVM (Jaha et al. 2014; Jaha et al. 2015; Martinho-Corbishley, Nixon et al. 2015). Moreover, during the training phase, RankSVM requires pairwise comparisons.

3.7   Conclusions

This chapter has described the use of crowdsourcing to acquire comparative labels, and the results of the response were presented and discussed in detail. Furthermore, we discussed how biometric signatures can be obtained through the relative rating of attributes. Using the Appen platform, we collected a database of facial profile comparisons

from XM2VTSDB for subject comparisons. We also discussed whether additional comparisons could be inferred. Finally, we discussed several potential ranking algorithms that were presented in previous studies.

In the context of human identification, the application of soft biometrics has demonstrated that comparative attributes hold more advantages than categorical attributes. Analysing the distribution of estimated relative measurement stability and the collected annotations shows the suitability of soft biometric features and the effectiveness of the crowdsourcing task. Moreover, inferring additional comparisons helps to fill in the gap between images comparisons and allows for a more reasonable rating of the images. In this chapter, we compared alternative ranking approaches and found that both Elo and RankSVM performed well when rating soft biometric features. The following chapter focuses on how facial profile comparisons can be used in biometrics.

Chapter 4

Comparative Soft Biometric for Facial Profile Recognition

4.1 Introduction

In the preceding chapter, the methodology for ranking was discussed, wherein scores are computed for objects or items within a dataset to denote their relative strength. This chapter introduces two distinct score-based ranking algorithms for use in this study: Elo and RankSVM. The process of arranging a list of subjects based on a single attribute is presumed to be optimal for facilitating successful attribute-based facial profile comparisons for recognition purposes. Accordingly, each subject's attributes may be delineated through relative measures. These measures are derived from the comparison data pertaining to each attribute. An evaluation of the performance of facial profile comparison methodologies can be conducted through the utilisation of RankSVM (Parikh and Grauman 2011) and Elo (Reid, Nixon and Stevenage 2013) ranking systems.

The application of comparative soft biometrics encompasses several distinct processes and activities. In our study, the systematic procedure commenced with the establishment of a dataset facilitated by the Appen platform. Within this platform, a collection of image pairs was provided for comparison based on a single attribute, such as a facial profile trait. Subsequently, the pairwise comparisons were transformed into scores linked with the respective images employing a ranking algorithm. For each image, we utilised the set of generated ranks for each trait to build a trait vector. Each component of the vector denoted the subject's rank for the corresponding trait. For example, a subject characterised by a long nose pointing upwards would attain a higher score for the trait nose tip compared to a subject with a nose pointing downwards. Ultimately, these trait scores served as the basis for selecting the most pertinent traits and discerning the identity of the individual depicted in the image. Given that the analysis revolved around trait comparisons without direct engagement with the images, this recognition

task may have been denoted as identity determination through verbal facial trait comparisons. This approach is also recognised as subject recognition employing comparative soft biometrics (see Figure 4.1).
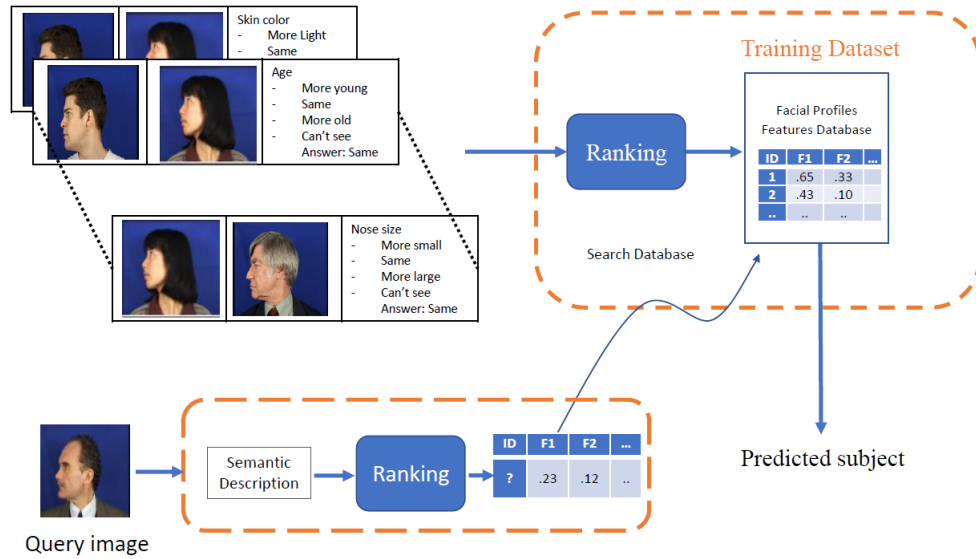
FIGURE 4.1: The subject recognition process using comparative soft biometrics.

The research presented in this study is part of a larger effort to assess whether verbal descriptions can be used instead of or in addition to image pixels in human recognition. This chapter summarises the research findings and discusses the potential implications for identifying individuals using relative facial profiles and soft biometrics. Specifically, we investigated human facial profile recognition within the XM2VTSDB database by considering comparative facial soft biometrics and incorporating additional attributes related to the side-view of the face. Additionally, we aimed to expand on the limited understanding of current research on comparative facial soft biometrics across different traits. Finally, this chapter examines the use of comparative soft biometric traits for recognising a given facial profile.

Section 4.2 begins by outlining the experimental design and describing the distribution of training and testing data. Sections 4.3 and 4.4 explain how the Elo and RankSVM algorithms are used to accurately identify relative subject signatures from new comparative annotations. Subsequently, Section 4.5 examines the stability of the estimated scores based on the chosen ranking, while also illustrating the correlations between traits and their associated labels. Section 4.6 explores the use of the trait selection methodology to identify a subset of traits that are proficient in classifying images based on traits and evaluates the experimental results. Lastly, Section 4.7 provides an overview of our findings.

4.2 Experiment Design to Assess Facial Profile Recognition Performance

4.2.1 Dataset

Initially, crowdsourcing was used to facilitate and document the comparisons between the images. Each comparison was attempted by 15 different labellers, resulting in a total of 15 attempts per comparison. Additionally, our dataset was divided into four separate folds, ensuring that images were only compared if they belonged to the same fold.

The dataset consisted of images of 230 subjects. However, only one image per subject was used in each fold, creating four distinct folds. It is worth noting that the images for each subject were taken at different times, resulting in variations due to factors such as weight changes, ageing, skin condition, or facial feature obstructions caused by changes in clothing, accessories, hairstyles, or facial hair styles. In total, a set of 33 traits was considered. The main aim of this analysis was to reduce this set to a minimal subset of essential traits for efficient facial profile recognition.

4.2.2 Test Set

During the test phase, we employed the leave-one-fold-out approach, wherein images from one fold were reserved for testing purposes. The objective was to train a classifier to become capable of predicting the identity of each test subject based on comparisons within its designated fold, while utilising it for identification across the remaining three training folds. We consider the following scenarios:

- All subjects within the chosen fold were regarded as suspects.

- The witnesses solely observed the suspects.

- The images of suspects in the police database may have differed from those memorised by witnesses.

Subsequently, during the test stage, each test comparison (i.e., a comparison between two test subjects) was treated as a comparison between a test subject and a training subject. This was based on the implicit assumption that images of the same subjects should generally exhibit consistency across the folds. This assumption can be interpreted as positing the absence of a covariate shift (David et al. 2010), implying that the training and test data should share a *similar* distribution to some extent. In scenarios where covariate shift is present, a domain adaptation framework would be more appropriate for evaluating facial recognition outcomes. By employing this specific setting, we reinforced the consistency of the crowdsourcing results by aggregating final responses from each labeller to maximise the cross-validation classification accuracy.

4.2.3 Model Design

The algorithm commenced by consolidating the crowdsourcing outcomes into a collection of comparisons with distinct scores. Then, a ranking technique was applied to convert the discrete trait comparisons for a given image into a vector comprising 33 scalar trait intensities. Following this, a classification algorithm was employed to ascertain the identities of the faces based on the scores obtained.

As depicted in equation 4.1, for a specific trait q, let $s_{q,i,j,l}$ denote the comparison between the image indexed by i and the image indexed by j concerning trait q, as evaluated by labeller $\ell \in \{1, ..., 15\}$. The score is designated as -1 if the trait in image i is deemed less intense than in image j by the labeller. Conversely, the score is assigned as +1 in the opposite scenario and 0 if the trait intensity appears identical in both images.

The aggregated result $s_{q,i,j}$ is defined as a discretised average E of all labellers' responses:

$$\forall\,\theta \in\,]0,1[,\, s_{q,i,j} = \begin{cases} \mp \left(E_\ell\big[s_{q,i,j,\ell}\big] \text{ if } \big|E_\ell\big[s_{q,i,j,\ell}\big]\big| \geq \theta \\ 0\ otherwise \end{cases} \tag{4.1}$$

In essence, the sign of the average response is considered if its magnitude surpasses a certain threshold (either less than $-\theta$ or greater than θ); otherwise, the result is deemed as 0. This formulation closely resembles a majority vote, where the final response corresponds to the label chosen by the largest number of labellers. In fact, the average of all of the responses tends to converge towards the dominant label if one exists. However, in cases where no majority label is apparent, relying solely on a majority vote can lead to considerable instability and ambiguity, especially if all of the labels receive an equal number of votes. In such instances, the average response tends to hover around 0, indicating a more balanced distribution of crowdsourcing outcomes compared to a majority vote. Table 4.1 provides an illustrative example demonstrating the advantage of using the average response over the majority approach.

Number of votes for:			Majority label	Average response
+1	**0**	**-1**		
5	5	5	None	0
7	1	8	-1	-0.06

TABLE 4.1: Example of majority label and average response.

In the ranking phase, we employed Elo and RankSVM as two rating schemes. Detailed explanations of these methods are provided in the following sections. Subsequently, the KNearest Neighbour (KNN) recognition method was used to identify the query image.

Our algorithm's hyper-parameters encompassed the threshold, Elo and RankSVM parameters, as well as the KNN parameters. Testing was conducted using a leave-one-fold-out setting, with three-fold cross-validation employed to determine the optimal parameters. For the descriptive analysis, which involved correlations, statistics, etc., a threshold was selected and used to generate ranks for all of the folds. The threshold selection process was based on the test results obtained in the predictive analysis, similar to considering the test sets during validation or performing a four-fold cross-validation (see Figure 4.2). However, it is important to note that the descriptive analysis was solely comprehensive and did not influence the selection of relevant traits.
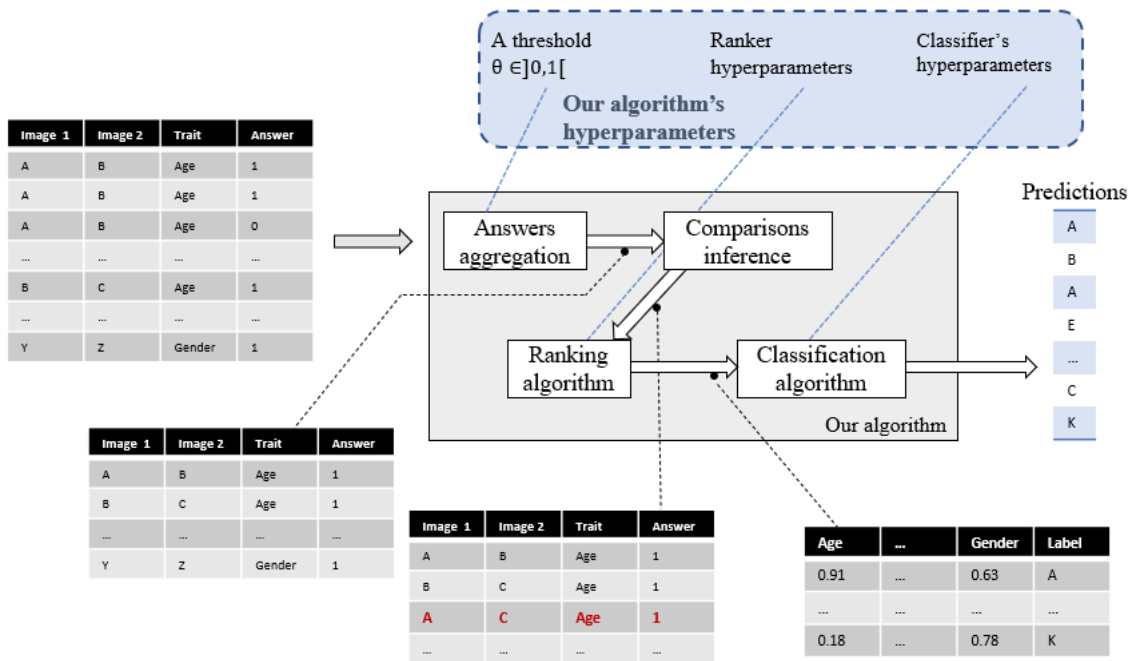
FIGURE 4.2: High-level overview of the model selection process.

4.2.4 K-fold Cross Validation

To assess the classification performance of the comparison model, a single fold was designated for testing. It was observed that some folds pose greater classification challenges than others. Therefore, the selection of the test fold was varied, and the performance metrics were averaged across folds, accompanied by a confidence interval delineated by the minimum and maximum values. Notably, as the hyper-parameters were not optimised through the learning algorithm, one training fold (from the remaining three folds) served as a validation set. This validation set was employed to evaluate the model's performance across a given set of hyper-parameters.

The hyper-parameter yielding the optimal metrics was subsequently selected and applied to the three remaining training folds during the final training session. An iterative

procedure was employed across all validation set options to determine the best hyper-parameters, taking into account the averaged performance metrics and considering the potential variations in validation metrics. This methodology can be likened to a k-fold cross-validation, where each fold represents a distinct validation set. Figure 4.3 illustrates the cross-validation settings.

The cross-validation approach offers several advantages:

- It mirrors the testing conditions, wherein one fold is reserved as a test fold.

- It guarantees the inclusion of all subjects in both the training and validation sets.

- It ensures that validation subjects are not juxtaposed against training subjects during the evaluation process.
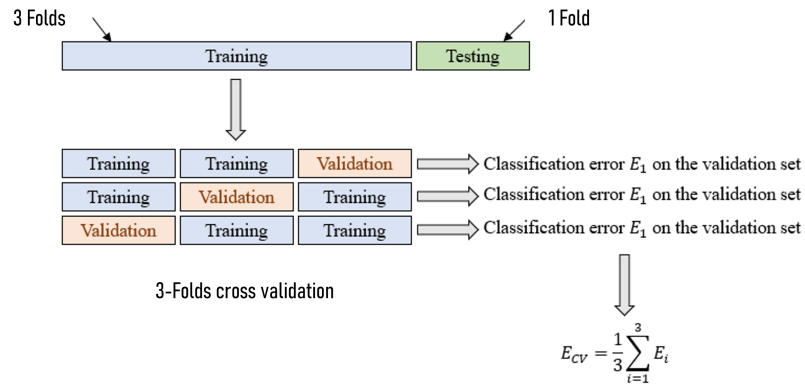
FIGURE 4.3: Cross-validation process.

4.3 Ranking Using Elo

4.3.1 Algorithm Formulation

The Elo rating system, widely utilised in ranking chess players, operates by assessing the disparities between actual game outcomes and the anticipated results. This system can be adapted to compare traits within a soft biometric context, akin to its application in evaluating the skills of chess players. By leveraging the Elo system, subjects can be ranked based on soft biometric traits such as nose size or eyebrow shape. Almudhahka et al. utilised the Elo system to assess the relative rankings between traits derived from biometric signatures and comparative labels (Nawaf Yousef Almudhahka et al. 2017b). Analogous to a chess match, a comparison inputted into the Elo rating system is employed to adjust the intensity of traits for Subject A and Subject B. The comparative

formula used in the system is as follows:

$$\overline{R}_A = R_A + K(S_A - E_A) \tag{4.2}$$

$$\overline{R}_B = R_B + K(S_B - E_B) \tag{4.3}$$

where $\overline{R}_A$ and $\overline{R}_B$ represent the new ratings, while R_A and R_B denote the previous ratings before considering the comparison. S_A and S_B correspond to the comparison scores (-1 for lesser, 0 for identical, and 1 for greater), and K is a parameter determined by the user. Specifically, the parameter K governs the rating's susceptibility to updates, with its value influencing the magnitude of rating adjustments. To determine the optimal value for K, the cross-validation technique is employed. The current ratings of subjects, denoted as E_A and E_B, are computed using the following equations:

$$E_A = \frac{1}{Q_A} \tag{4.4}$$

$$E_B = \frac{1}{Q_B} \tag{4.5}$$

In Equations 4.4 and 4.5, Q_A and Q_B signify the current ratings of the subjects. The inclusion of a constant U in the formulas 4.6 and 4.7 illustrates how the current ratings of the subjects can influence the outcomes.

$$Q_A = 1 + 10^{\frac{R_A - R_B}{U}} \tag{4.6}$$

$$Q_B = 1 + 10^{\frac{R_B - R_A}{U}} \tag{4.7}$$

In this study, the Elo rating system was utilised for comparisons between labels gathered from the sample dataset. Thus, q represented a trait, t_q^{Elo} denoted a function mapping an image to its Elo score for the trait q, and s_q signified a function comparing two images and generating a verbal trait comparison. From the expressions of Elo ranking steps, given a choice of K, and for images A and B such that $s_q(A, B) = 1$ but $t_{q,old}^{Elo}(A) \ll t_{q,old}^{Elo}(B)$, the new score would adjust the ranks of $t_q^{Elo}(A)$ *and* $t_q^{Elo}(B)$ closer to each other without altering their order, as if the score $s_q(A, B)$ were not considered. To maintain consistency, the ranks were continually updated using the same comparisons until new scores were obtained that aligned with them. Furthermore, if $s_q(A, B) = 0$ *but* $t_{q,old}^{Elo}(A) \neq t_{q,old}^{Elo}(B)$, the ranks were adjusted to converge towards a middle value without becoming equal. The updated rule counteracted the effects of previously considered comparisons. Thus, the equalities in trait comparisons were transformed into inequalities by leveraging the set of comparisons of both A and B. Additional details regarding this method can be found in Section 3.5.3.

4.4 Ranking Using RankSVM

4.4.1 Algorithm Formulation

This section introduces the RankSVM algorithm for inferring relative attributes. Although RankSVM is primarily employed to rank web pages (Joachims 2002), it can also be used independently to rank pairwise comparisons. For each trait q, the algorithm learns a function r_q that takes a feature vector and predicts a scalar rank that, ideally, preserves the outcomes of the comparisons (wins, draws, and losses). In essence, the ranking function is optimised to maintain the inherent order of the samples, as represented by the aggregated comparisons.

The comparison trait is computed as the difference between the features[6] of the samples under comparison, with the label representing the outcome of the comparison (-1,0,1). An initial model h is trained to predict the comparison outcomes and is subsequently used to infer the ranking function r_q.

Mathematically, the objective is to achieve the following relationship:

$$\forall\, (x_1, x_2) \in X^2, h(x_1 - x_2) > 0 \Leftrightarrow r_q(x_1) > r_q(x_2) \tag{4.8}$$

A linear model (without a bias term) is a common choice for h, as its linearity ensures that $h(x_1 - x_2) = h(x_1) - h(x_2)$, thus automatically preserving the order. Additionally, h can serve as the ranking function. The function r_q is represented as follows:

$$\forall\, x \in X, r_q(x) = w_q^T x \tag{4.9}$$

In this context, the rank must be inferred solely from verbal descriptions. This implies that the aim is to translate the comparison outcomes for a given query into scores for each image without relying on image pixels. These scores are then used to rank subjects. Consequently, each image is individually represented using a hypothetical feature space. Assuming n denotes the number of compared images, in our representation, each image's feature is a one-hot vector $x_i \in \{0,1\}^n$, where only the i component equals 1. This assumes that each image in the dataset is unique concerning the considered trait. Figure 4.4 illustrates the generation of the hypothetical features of one-hot images.

The method outlined in (Martinho-Corbishley 2018) was used to ascertain the rank of each image. The study employed the widely adopted soft-margin RankSVM, originally

---

[6]A feature should refer to a visual characteristic. As image pixels are not taken into consideration in these experiments, an imaginary ideal feature space is utilised. Specifically, a one-hot encoding of the labels is used. A one-hot encoding is a set of bits where the only valid combinations of values consist of a single high (1) bit and all the others low (0) (Harris et al. 2015). With one-hot encoding, we convert each categorical value into a new categorical column and assign a binary value of 1 or 0 to these columns. Each integer value is represented as a binary vector, where all the values are 0 except for the index marked with a 1. One-hot encoding is a representation of categorical variables as binary vectors.
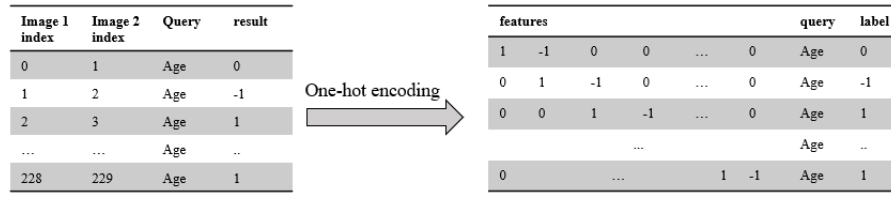
FIGURE 4.4: Representation of images using a hypothetical feature space.

devised by Joachims (2002), with a subsequent extension by Parikh, Kovashka et al. (2012) to accommodate similarity constraints. This augmentation introduced the similarity constraint to ensure that comparisons yielding the same outcome resulted in images possessing ranks that were sufficiently proximate. However, upon implementing the algorithm, it was observed that the similarity constraint often led to identical rankings for similar images, which may not have accurately reflected their comparative distinctions between them and other images.

Populating the dataset with equalities obviates the necessity for a similarity constraint, as images sharing similar traits naturally yield identical comparison outcomes when juxtaposed against other images. Additionally, it is observed that $C \neq 0$ is sufficient to retrieve comparison outcomes accurately within the range of $\{1, -1\}$ without error. Optimal selection of the C value entails minimising the sum of absolute differences among ranks of images possessing comparable trait salience. Stated differently, it is desirable to determine a C value that inherently brings the ranks of similar images closer together than any alternative option (see Figure 4.5).
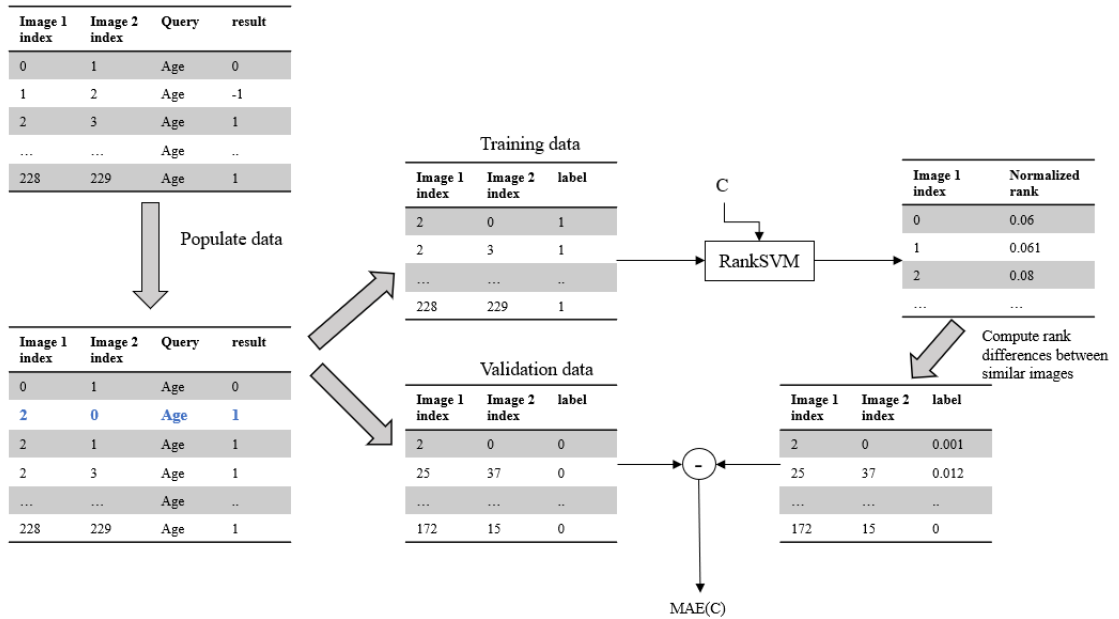
FIGURE 4.5: RankSVM pipeline with C validation using the subset of similar images. Note that the similarity constraint is implicitly implemented by transferring non-zero comparisons to images involved in zero comparisons.

4.5 Attribute Significance

4.5.1 Attribute Correlations

Correlation analysis served as a method to unveil the similarities between attributes and to detect linear dependencies among them. By analysing the correlation matrix across various attributes, associations were determined, and the strength of features based on one another could be predicted. Thus, this approach facilitated the development of a descriptive correlation between feature values. In this study, Pearson's correlation coefficient (r) was computed for all the pairs of attributes listed in Table 3.1 to identify dependencies among the 33 facial profile attributes.

4.5.1.1 Pearson's Correlation Coefficient

Pearson's correlation coefficient r was employed to measure linear relationships between attributes, facilitating the identification of dependencies between variables X and Y, as expressed by the following equation (Tome, Vera-Rodriguez et al. 2015; Gonzalez-Sosa et al. 2018):

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.10}$$

The variables X and Y represent independent variables describing traits from a facial profile. x_i and y_i denote two distinct labels, each representing the ith annotation of a given subject. The computation of r involves the division of covariance between X and Y, denoted by σ_{XY}, by the product of the standard deviations of the two variables.

The resultant r value lies within the range of +1 and –1, with 0 indicating no linear correlation. A value approaching +1 signifies a positive linear correlation between labels, while a value approaching –1 indicates a negative linear correlation between labels.

Our experimental findings suggest a low correlation among the traits, implying that the attributes utilised in this study were predominantly independent. As illustrated in Figure 4.6, the dark brown shading in the cells signifies traits with a high positive correlation, while dark blue represents a strong negative correlation, and grey/white indicates the absence of any linear correlation. Similar correlations were observed between the RankSVM and Elo ranking systems, as illustrated in the figure. The matrix revealed significant correlations among several traits, particularly those associated with width and size, such as face profile length, face profile width, double chin, chin height, and cheek size. Furthermore, it indicated a positive correlation between nose size and nostril size. Notably, there was a correlation between ear size and age, aligning with the findings
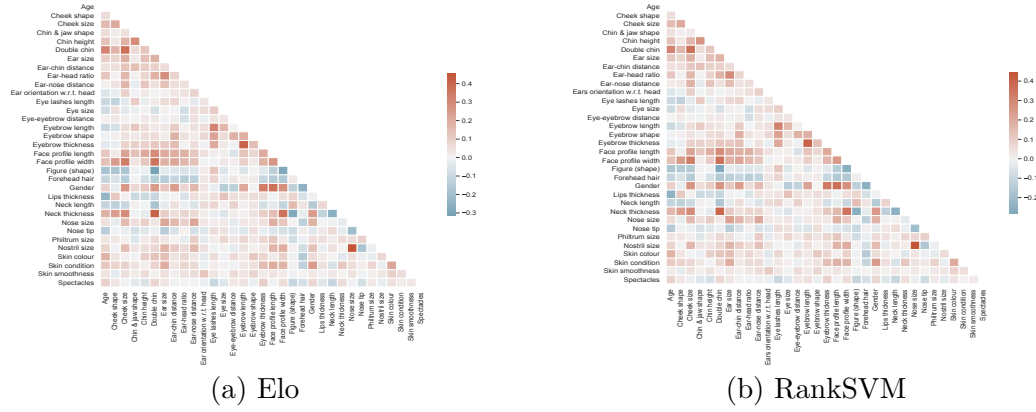
FIGURE 4.6: Correlation matrix for facial profile attributes.

of earlier research (Tan et al. 1997). However, neck thickness and length exhibited a negative correlation.

Positive correlations between traits and labels indicate similarities, while negative correlations illustrate differences or uniqueness. This analysis and exploration of correlations among traits has enriched our understanding of each trait's contribution to recognition and identification. A substantial correlation between two traits suggests that the values of one trait may be accurately predicted using a linear function of the values of the other.

4.5.2 Attribute Contribution to Facial Profile Recognition

Enhancing the efficiency and accuracy of recognition necessitates reducing the number of non-useful traits. The trait selection method can be leveraged to identify the most promising traits that will yield the highest profile recognition accuracy. In this study, mutual information (MI) was used to determine the optimal subset of traits. (B. Guo et al. 2008).

4.5.2.1 Mutual Information Measure

The identification of an attribute's importance provides a solid understanding of its efficacy as a semantic descriptor and its contribution to effective identification and recognition. In this subsection, MI measurement was used to evaluate the significance of facial profile attributes. MI treats each trait as a random variable. For a trait X, let $p(x)$ denote $p(X = x)$, representing the probability of the occurrence of the value x. MI quantifies the overlap between the probability distributions of X and Y. The maximum MI value indicates identical probability distributions of X and Y, implying a unique permutation mapping of the values of X to Y. Conversely, a minimum MI

value, typically 0, suggests complete independence between the random variables. Formally, this independence is expressed as the logarithmic term yielding 0, which implies $p(x, y) = p(x)p(y)$, signifying independence between the variables X and Y. MI is computed using the formula 4.11:

$$MI = I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \ln \left(\frac{p(x, y)}{p(x)p(y)} \right) \tag{4.11}$$

where X denotes a variable representing an attribute, and Y is an attribute representing a label. The computation of MI for two traits involves $p(x, y)$, the probability density function for X and Y, with $P(x)$ and $P(y)$ being marginal probability density functions. The association of MI using Shannon entropy is expressed as:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \tag{4.12}$$

where $H(X)$ (equation 4.13) and $H(Y)$ (equation 4.14) denote the entropy of X and Y, respectively. $H(X|Y)$, as illustrated in equation 4.15, is the conditional entropy of X given Y, and $H(Y|X)$, as shown in equation 4.16, is the conditional entropy of Y given X. Moreover, equation 4.17 represents the joint entropy $H(X, Y)$ of X and Y. The discrete Shannon entropy is thus defined as:

$$H(X) = - \sum_{x \in X} p(x) \ln p(x) \tag{4.13}$$

$$H(Y) = - \sum_{y \in Y} p(y) \ln p(y) \tag{4.14}$$

The joint and conditional entropies can be described as follows:

$$H(X|Y) = - \sum_{x \in X} \sum_{y \in Y} p(x|y) \ln p(x|y) \tag{4.15}$$

$$H(Y|X) = - \sum_{y \in Y} \sum_{x \in X} p(y|x) \ln p(y|x) \tag{4.16}$$

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \ln p(x, y) \tag{4.17}$$

where $p(x|y)$ represents the conditional probability of $X = x$ given $Y = y$.

In this context, X represents an attribute's rank (observation), Y represents the subject's identity (or the label), $H(Y)$ indicates the uncertainty about Y, and $H(Y|X)$ represents

(a) Elo

(b) RankSVM
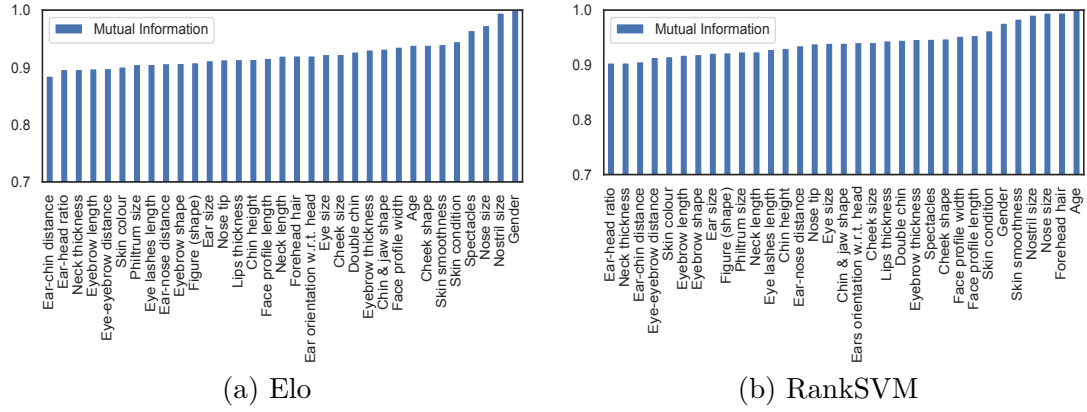
FIGURE 4.7: Normalised mutual information for each of the 33 attributes with the target variable (label).

the uncertainty remaining after knowledge of observation X. Consequently, $I(X,Y)$ measures the amount of information about Y that can be obtained from the measurement of X. MI is used to assess the discriminative power of each facial profile attribute by jointly computing its value with the label. Figure 4.7 displays the normalised entropy and MI associated with each attribute. In this study, equation 4.12 was employed to compute the MI. Subsequently, each of the 33 attributes was normalised using Equation 4.18. Normalised mutual information (NMI) is a method used to normalise the MI score to adjust the results on a scale from 0 (indicating no MI) to 1 (representing perfect correlation). There existed a slight difference between the two ranking methods in terms of the MI results.

$$NMI(X,Y) = \frac{2I(X,Y)}{H(X) + H(Y)} \tag{4.18}$$

Figure 4.7 illustrates a general tendency towards agreement in assigning high scores between both rating systems. The MI suggested that nose size and nostril size were of higher significance compared to attributes such as ear-chin distance, ear-head ratio, and neck thickness, which exhibited lower MI scores.

Based on the Elo rating system, the MI analysis also indicated that gender possessed the highest discriminative power, while attributes like ear-head ratio and ear-chin distance exhibited the lowest discriminative power, as shown in Table 4.2. Surprisingly, the RankSVM approach indicated that general attributes such as age have more discriminative power compared to proportion attributes such as ear-head ratio, which display lower discriminative strength.

TABLE 4.2: Mutual information analysis of soft facial profile traits using the 33 other attributes.

	Attribute	MI I(X,Y)	
		Elo	RankSVM
A1	Eyebrow length	0.8976	0.9179
A2	Eyebrow shape	0.9070	0.9191
A3	Eyebrow thickness	0.9306	0.9467
A4	Spectacles	0.9645	0.9470
A5	Eye-eyebrow distance	0.8981	0.9140
A6	Eyelash length	0.9053	0.9284
A7	Eye size	0.9227	0.9397
A8	Nose-mouth distance	0.9051	0.9239
A9	Nostril size	0.9950	0.9910
A10	Nose tip	0.9134	0.9388
A11	Nose size	0.9732	0.9948
A12	Lips thickness	0.9138	0.9442
A13	Face profile length	0.9159	0.9538
A14	Face profile width	0.9355	0.9524
A15	Skin smoothness	0.9401	0.9838
A16	Skin condition	0.9452	0.9624
A17	Forehead hair	0.9202	0.9950
A18	Ear size	0.9119	0.9215
A19	Ear orientation w.r.t. head	0.9202	0.9410
A20	Ear-head ratio	0.8964	0.9037
A21	Ear-chin distance	0.8849	0.9060
A22	Ear-nose distance	0.9067	0.9353
A23	Cheek shape	0.9388	0.9478
A24	Cheek size	0.9228	0.9410
A25	Chin & jaw shape	0.9320	0.9398
A26	Double chin	0.9271	0.9452
A27	Chin height	0.9140	0.9305
A28	Neck length	0.9200	0.9241
A29	Neck thickness	0.8965	0.9038
A30	Age	0.9388	1
A31	Gender	1	0.9763
A32	Skin colour	0.9009	0.9152
A33	Figure (shape)	0.9083	0.9223

4.6 Performance Evaluation of Facial Profiles

4.6.1 Recognition Rate Measurement

Once cross-validation is employed for tuning, as described in Section 4.2.4, it can no longer reliably be used to assess how well the method performs on out-of-sample data.

Therefore, we used the hold-out/test set for this purpose. Each fold from the original dataset served once as a test set, resulting in four performance measures.

For each test set, the validation sets were used to optimise the threshold, select the most relevant traits for classification using the sequential floating forward selection (SFFS) algorithm, and determine the classifier's parameters. The validation approach outlined in Figure 4.8 was applied to the ranking algorithm. The goal was to minimise the mean absolute error (MAE) of the validation on the validation set, which contained no labels, while maximising the accuracy on the ordered training subset. The ranking was optimised to provide the best translation of crowdsourcing comparisons into scalar ranks, rather than enabling better classification.

The process for measuring the recognition rate, as depicted in Figure 4.8, encompassed the following steps:

- The model was trained using three folds, each containing 230 subjects.

- The model with parameters yielding the best average accuracy across all validation folds was selected.

- The chosen model was refined using the parameters from the three folds, and testing was conducted on the test data contained within the fourth fold (the hold-out fold).
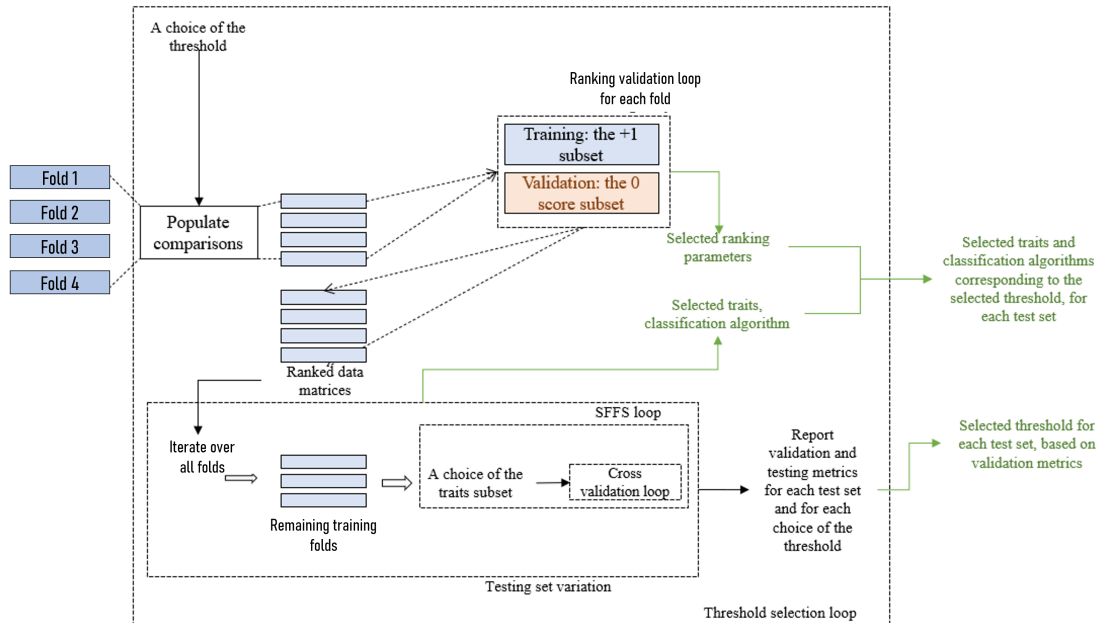

FIGURE 4.8: Efficient training and collection of performance metrics.

The ranking procedure is a time-consuming phase and we found that splitting the data into training and test sets after computing the rankings was more efficient than doing

so as an initial step. For each threshold option, we reported both validation and testing metrics, selecting the optimal threshold for each test set based on the corresponding validation metrics. Finally, we collected the results for each test set, which include ranked data, the selected threshold, attributes, and the classification algorithm used.

The sequential forward floating selection (SFFS) method was used to identify the subset of traits that optimised the classification accuracy in the validation. This technique iteratively updated the selected set of traits by adding or removing traits to achieve maximum accuracy. More explanation of this method can be found in Subsection 4.6.1.1.

When assessing performance within biometric systems, two primary metrics are used: the receiver operating characteristic (ROC) curve, which evaluates verification performance, and the cumulative match characteristic (CMC) curve, which assesses identification performance (DeCann et al. 2013). Additionally, performance evaluation can involve quantifying differences among multiple observations of an individual (intra-class variations) and variations between subjects (inter-class variance). The term "distinctiveness" commonly refers to the effectiveness of a biometric system. For example, low intra-class variations indicate permanence and repeatability, while high inter-class variations indicate biometric traits that facilitate successful differentiation between individuals (Reid, Samangooei et al. 2013). The following sections outline the methodology for measuring performance within each of these assessment tools.

4.6.1.1 The Sequential Floating Forward Selection Algorithm

This section explains how SFFS is used to determine the best set of attributes for achieving the highest classification accuracy. Unlike feature importance-based methods, which evaluate the discriminative power of each feature independently, the SFFS algorithm selects features as a cohesive group of complementary attributes (Yoshida 2010; Shirbani et al. 2013). Although these attributes may have low individual dependency on the target (measured using metrics such as MI or correlation), their combined integration often results in significantly improved predictive effectiveness.

The subset selection method works as a bottom-up search technique, starting with an empty set of features and then determining the optimal features to include or exclude to improve the performance metric or objective function. Each iteration of this process improves the soft biometric verification system by adjusting the set of traits used, as shown in Figure 4.9 (Yoshida 2010). The SFFS procedure begins with an empty set (d = 0). In each iteration, it aims to identify the most advantageous attribute to include in order to improve the performance metric. Simultaneously, within the same iteration, it attempts to eliminate another attribute that was selected in previous iterations to further enhance the metric. This iterative process continues until no further improvement can be made by adding or removing an attribute. At this point, the algorithm concludes and outputs

the final selected set of attributes as the optimal configuration for optimising the metric. The SFFS analysis provides valuable insights into the importance of facial profile traits
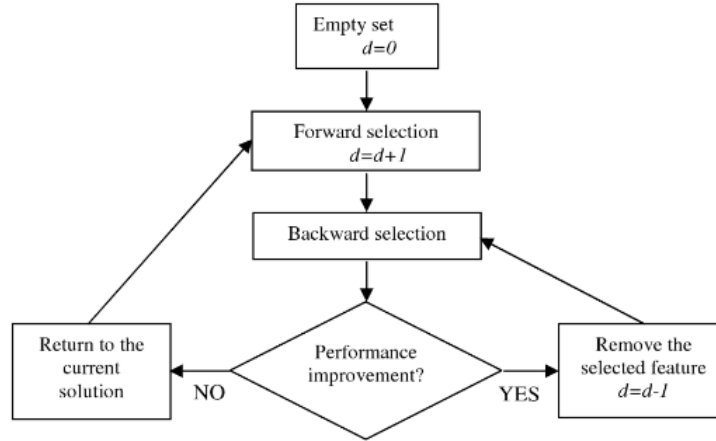
FIGURE 4.9: Flowchart illustrating the procedural steps of the sequential floating forward selection algorithm (Yoshida 2010).

in human face recognition. In Table 4.3, we present the optimal combination of traits for each ranking method separately, resulting in the highest rates. SFFS identified nearly 20 attributes (d = 20) out of the 33 traits, which are considered crucial for enhancing accuracy and reliability, with peak accuracies of 74% for Elo and 77% for RankSVM. It is worth noting that the SFFS algorithm excluded traits such as ear size, ear-nose distance, neck length, and neck thickness. This exclusion may have been due to the concealment of these traits within our dataset, possibly because of factors like hair or other objects such as earrings. As shown in Table 4.3, Elo and RankSVM exhibited similarities in terms of the most significant traits, although they differed in their respective hierarchy of importance. Table 4.4 provides a summary of the recognition results.

4.6.2 Inter and Intra-Class Variations

Biometric recognition performance can be assessed using similarity measurement functions. For example, the sum of absolute differences (norm 1) was employed to compare biometric trait vectors and construct a distance matrix. This selection was informed by cross-validation loops aimed at optimising the choice of KNN distance. The distance is defined by the following equation:

$$d_E(X, Y) = \sum_{f \in t} \left| x(f) - y(f) \right| \tag{4.19}$$

where $d_E(X, Y)$ represents the distance between X and Y, X denotes the trait t vector of the subject in the probe, and Y is the trait t vector of the subject in the gallery. Both the distance calculated using all of the traits and that calculated using the union

TABLE 4.3: List of best comparative soft biometric facial profile traits inferred using sequential floating forward selection for Elo and RankSVM.

Ordering	SFFS Attribute	
	Elo	RankSVM
1	Gender	Forehead hair
2	Chin height	Nose size
3	Age	Age
4	Nostril size	Gender
5	Nose size	Nostril size
6	Lips thickness	Lips thickness
7	Spectacles	Spectacles
8	Figure (shape)	Double chin
9	Forehead hair	Figure (shape)
10	Double chin	Face profile width
11	Skin condition	Nose tip
12	Eyebrow thickness	Chin height
13	Face profile width	Cheek size
14	Skin colour	Skin condition
15	Cheek shape	Philtrum size
16	Nose tip	Cheek shape
17	Chin & jaw shape	Eyebrow thickness
18	Eyebrow length	Chin & jaw shape
19	Cheek size	Face profile length
20	Philtrum size	Eye–eyebrow distance

TABLE 4.4: Summary of recognition performance results for Elo and RankSVM.

Elo				
Test Set	Trait Selection	Number of Traits	Test Accuracy	Threshold
fold1	SFFS Traits	20	0.74	0.256
fold2	SFFS Traits	17	0.69	0.264
fold3	SFFS Traits	15	0.70	0.256
fold4	SFFS Traits	22	0.73	0.256
fold1	All Traits	33	0.59	0.224
fold2	All Traits	33	0.61	0.212
fold3	All Traits	33	0.71	0.218
fold4	All Traits	33	0.58	0.229
RankSVM				
Test Set	Trait Selection	Number of Traits	Test Accuracy	Threshold
fold1	SFFS Traits	20	0.72	0.233
fold2	SFFS Traits	22	0.76	0.264
fold3	SFFS Traits	20	0.77	0.264
fold4	SFFS Traits	21	0.70	0.256
fold1	All Traits	33	0.59	0.224
fold2	All Traits	33	0.61	0.212
fold3	All Traits	33	0.71	0.218
fold4	All Traits	33	0.58	0.229

(a) Elo   (b) RankSVM
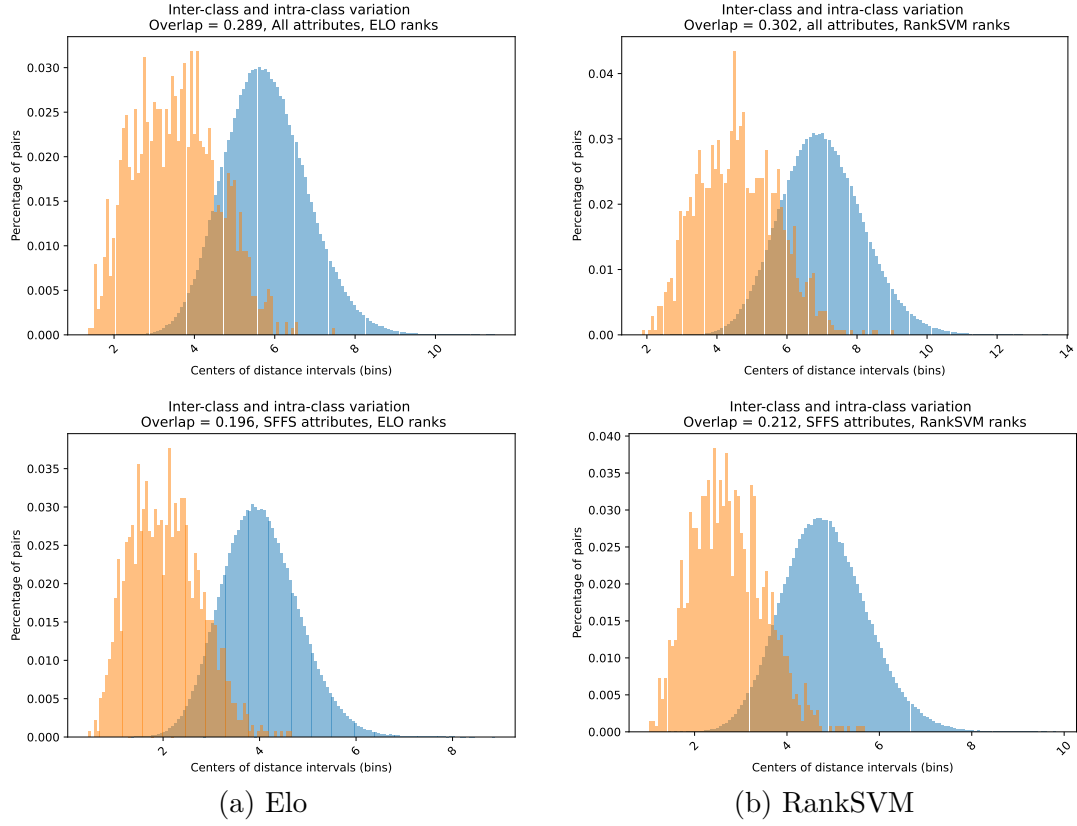
FIGURE 4.10: Inter and intra-class variations with all and sequential floating forward selection traits using absolute differences (norm 1).

of selected SFFS traits t were applied across all test sets. Figure 4.10 illustrates the overlap between inter-class and intra-class results.

When samples of the same class are clustered outside the overlap region, effective distance selection in the k-neighbour search can lead to complete or optimal separation of the classes. In our study, the choice of search radius was closely related to the selection of k-neighbours in the KNN algorithm, with the latter being optimised in the cross-validation loop.

4.6.3 Receiver Operating Characteristics and Detection Error trade-off

The ROC serves as an indicator of binary classification performance. The output of the evaluated classifier typically ranges from ± 1, and hence is termed as either positive or negative. These predictions can be true or false, delineated as follows:

- A true positive (TP) occurs when the classifier accurately predicts $+1$.

- A false positive (FP) arises when the classifier mistakenly predicts $+1$.

- A true negative (TN) manifests when the classifier accurately predicts –1.

- A false negative (FN) occurs when the classifier mistakenly predicts –1.

Initially, the predictions are scalar scores, which can be interpreted as probabilities for the positive labels (e.g., logistic regression) or as confidence scores (e.g., leaves of a decision tree, random forests, etc.). Binary classification is achieved by applying a threshold to the scalar prediction. By varying the threshold, one can adjust the algorithm to optimise precision, recall, or strike a trade-off. The ROC curve illustrates the variation of the true positive rate against the false positive rate. Each point on the curve corresponds to a specific choice of threshold.

In this study, which involved 230 classes, the classification task could be thought of as 230 binary classifications, where the classifier aims to answer the question: 'Does this subject belong to this class?' However, plotting 230 ROC curves at once would make the figure unreadable. Therefore, a multi-label classification approach was used, allowing the classifier to predict multiple positive classes for a subject. To generate ROC data points, a 230 × 230 binary matrix was created by applying a threshold to predicted test scores. This matrix was then compared to the true binary labels, which was also a matrix of the same size but with only one positive label per row.

Since each test set had optimal traits, a threshold, a trained ranking algorithm, and a classifier associated with it, a multilabel scalar prediction was generated for each test set. These predictions were combined across all four matrices, resulting in a scalar matrix of size 920 × 230. Subsequently, this matrix was transformed into a vector, which was compared to the true labels using all possible thresholds to plot the ROC curve (see Figure 4.11).

The ROC curves show consistent variations across different scenarios. The feature selection algorithm significantly improved both ranking algorithms, resulting in an increase in the area under the curve (AUC) for both methods. In particular, the highest AUC value achieved was 0.994 when using Elo with SFFS selection.

The detection error trade-off (DET) curve was employed for assessing system performance and illustrating the trade-off between the false positive rate (FPR) and false negative rate (FNR). The feature selection method improved the performance of both ranking algorithms, as shown in Figure 4.12. As a result, both techniques experienced a decrease in the equal error rate (EER) on the DET curve. In this case, the RankSVM algorithm with SFFS selection produced the lowest EER value of 0.046.

4.6.4 Cumulative Match Characteristic

A CMC curve is a tool used to assess the retrieval performance (DeCann et al. 2013). This metric measures recognition accuracy by employing the KNN method with the
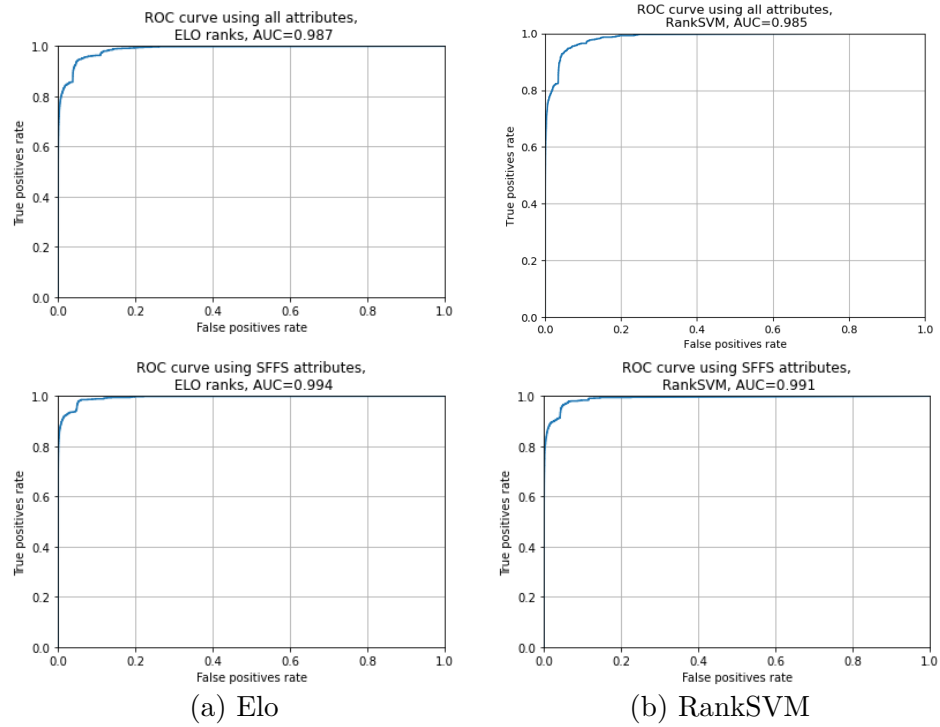
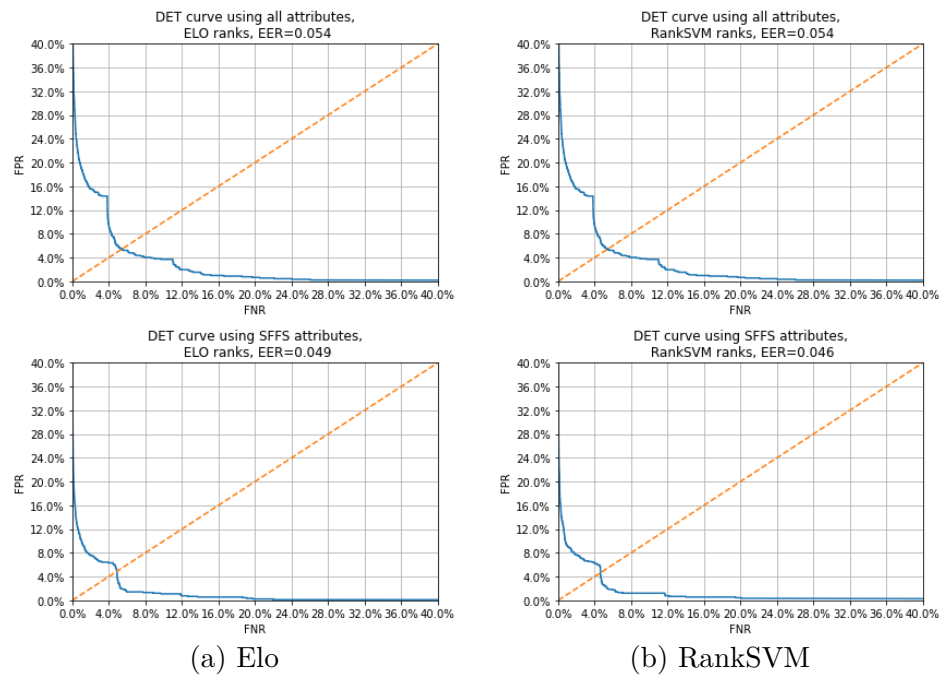FIGURE 4.11: Example receiver operating characteristic curves.

FIGURE 4.12: Examples of detection error trade-off.

SFFS. In our study, it evaluated the presence of correct subjects among k = 1 to 920 (representing all samples in the dataset). Figure 4.13 shows the recognition performance for facial profile traits using soft facial traits. In this curve, the first candidate achieved

72% accuracy when using Elo and 74% accuracy for RankSVM. This level of accuracy increased as the number of candidates progressed, almost reaching 100% at rank-115 for both ranking systems. The reported CMC results were averages. For each test dataset, the algorithm was trained using the remaining three folds. In each iteration, the best threshold, the set of SFFS traits, the optimal number of neighbours (k), and the best distance (p) were determined. This process was conducted for the four test datasets with equally divided validation–test splits. In every experiment, probe subjects were identified by first recognising semantic labels from a sample image and then conducting semantic retrieval against a fold of known descriptions.

(a) Elo	(b) RankSVM
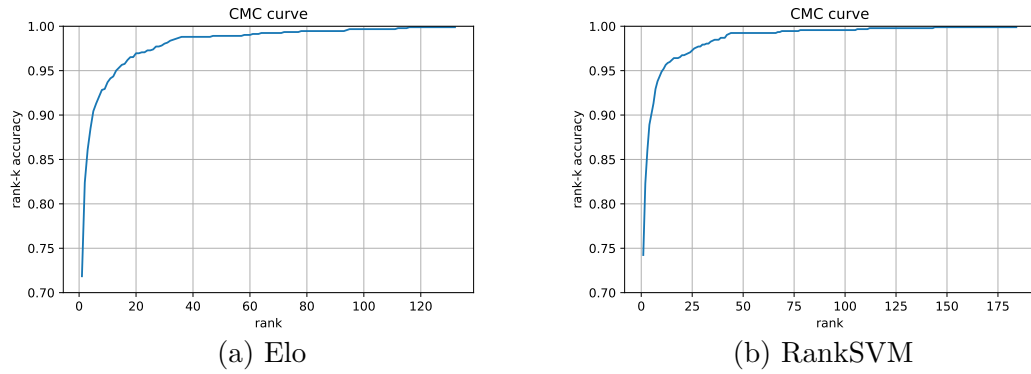
FIGURE 4.13: Recognition via cumulative match characteristic performance for the sequential floating forward selection attributes.

4.6.5 Threshold Impacts on Recognition Performance

As detailed in Section 3.5.3, the computation of ratings using both Elo and RankSVM involved adjusting a threshold to determine the sensitivity of the ranking scores to each majority vote. This threshold was not assigned to a default value and needed to be selected to maximise the recognition rate. Subsequently, the mean recognition performance was computed with different threshold values, as shown in Figure 4.14. The impact of the selected threshold on recognition performance was significant, indicating that test accuracy varies based on the chosen threshold. For example, the accuracy of the Elo ranking improved to 72% when the threshold value was adjusted from 0.15 to 0.24, as demonstrated in Figure 4.14. We depicted the ranges of threshold values representing the highest possible test accuracy.

When producing aggregated comparisons, the initial step involves averaging the outcomes of each comparison, followed by discretising the results into the set -1,0,+1 using a threshold $\theta \in [0, 1]$. Furthermore, the discretisation process entails using a threshold below which result magnitudes deem a comparison a draw. Specifically, if the absolute

(a) Elo   (b) RankSVM
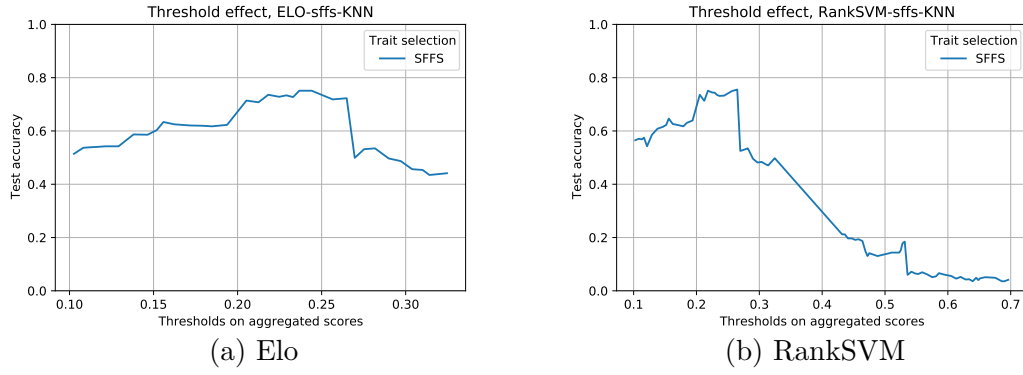
FIGURE 4.14: Effect of threshold value on recognition performance.

value of the average outcome for a given comparison is less than the threshold, the comparison outcome is replaced by zero. Conversely, if the average outcome's absolute value exceeds the threshold, it is considered of sufficient magnitude to be replaced by its sign.

Two trivial cases can be distinguished: when $\theta = 0$ and $\theta = 1$. In the first case, taking the sign of the average outcome serves as the final aggregated comparison outcome, resulting in non-zero discretised outcomes. Conversely, in the second case, all comparisons are replaced by 0 since each averaged comparison outcome has an absolute value lower than 1. This leads to all subjects having the same rank for every trait.

For a given query q, the comparison data $\mathcal{C}$ can be divided into two sets

- The set of similar (unordered) subjects S_q, defined by comparisons with an outcome of 0 (representing pairs of images exhibiting equally intense traits).

- The set of ordered subjects O_q, defined by comparisons with a positive outcome[7].

When $\theta = 0$, the ordered set contains all comparisons ($S_q = \phi$). By varying the threshold from 0 to 1, comparisons associated with low-magnitude averaged outcomes are transferred from the ordered set to the unordered set until the ordered set becomes empty ($O_q = \phi$). By selecting the optimal threshold, the algorithm strikes a balance between the size of S_q and the size of O_q.

The discretisation of averaged comparison outcomes can also serve as a method to stabilise shared comparisons between the folds. In this study, some image pairs had averaged outcomes with small magnitudes across all folds but with varying signs. One instance of these fluctuating results occurred when labellers used the terms 'Less' and 'More' to indicate 'Slightly Less' and 'slightly More', resulting in an almost uniform distribution of non-zero answers. These fluctuating outcomes could have been replaced by the 'Same'

---

[7]Negative outcomes can be converted to positive ones by swapping the two images in the pairwise comparison.

TABLE 4.5: Example of averaged comparison outcomes.

Image1	Image2	Trait	Number of +1 Answers	Number of 0 Answers	Number of -1 Answers	Averaged Outcome	Fold in the Gallery
A	B	Age	5	5	5	0.0	1
A	B	Age	2	13	0	0.13	2
A	B	Age	6	2	7	-0.06	3
A	B	Age	7	0	8	-0.06	4

answer, or a null outcome, by selecting a sufficiently high value for θ. An example of such a case is presented in Table 4.5.

In the example (Table 4.5), when $\theta = 0.14$ was set, the comparison between the ages of Subjects A and B yielded the 'Same' outcome for all folds. The thresholding effect could be viewed as noise cancellation, where small fluctuating averaged outcomes were interpreted as a noisy manifestation of the 'Same' response. This is because the selection of a threshold corresponds to a trade-off between stabilising outcomes and considering their signs.

The curves depicting the effect of the threshold on the test accuracy illustrate this trade-off. Smaller values of θ generally lead to increasing accuracy, indicating effective cancellation of noisy outcomes. Conversely, higher values are associated with decreasing accuracy, suggesting that relevant comparison averages are erroneously regarded as noisy and replaced by 0. The intermediate values of θ where the accuracy remains constant correspond to the trade-off point.

Two simple conditions are employed to discard irrelevant thresholds and to facilitate early stopping of the validation process:

1. When the feature selection method yields a single trait with high variance in validation accuracy.

2. When ranks are identical for all subjects relative to one or more traits and cannot be normalised from 0 to 1.

The conditions mentioned above led to validation curves halting at 0.33 for Elo and 0.69 for RankSVM. However, given the accuracy variations related to threshold values, employing the early stopping method could have aided in selecting the optimal threshold without the need for exhaustively examining all of the possible thresholds. The early stopping procedure involves evaluating the thresholds in ascending order, storing the threshold associated with the best validation accuracy, and periodically verifying (e.g., for each of the 10 tested thresholds) if the accuracy increased, or equivalently, if the optimal threshold has changed its value. If there is no improvement after a certain

number of verifications, the algorithm terminates and returns the best threshold. Subsequently, the validation of the remaining parameters is conducted. In future, we will use the explained early stopping approach to accelerate the validation process in future experiments.

4.7 Conclusions

This chapter explored the use of comparative facial profile soft biometrics in human recognition. The findings have contributed to our understanding of the relative facial attributes and their usefulness in human recognition. Consequently, a new set of comparative facial profile soft biometrics has been proposed, focusing on traits selected by SFFS for their significant role in facial recognition. Additionally, this chapter explained the generation of relative ratings for biometric signatures. The attribute analysis demonstrated the statistical significance of comparative facial profile soft biometrics, allowing for the ranking of attributes based on semantic stability and discriminative strength. The experiments conducted also provide insights into the effects of label compression and the challenges associated with tuning parameters in rating systems, with the goal of improving recognition accuracy. Overall, the research emphasises the potential for ranking soft biometrics in larger datasets and achieving high levels of recognition accuracy. The findings also emphasise the feasibility of using a threshold for rating soft biometrics in larger datasets, providing further evidence of the effectiveness of information theory in machine learning and ranking contexts.

The correlation analysis of trait comparisons has revealed substantial correlations among numerous traits, particularly those associated with width and size, such as cheek size and chin height. Furthermore, findings from the MI measurement showed that nose size and nostril size were of higher significance compared to attributes such as ear-chin distance, and neck thickness, which received lower scores in both ranking methods. In addition to the statistical findings presented in this chapter, we have also demonstrated that out of the 33 traits, SFFS found about 20 traits that are important for improving accuracy and reliability. Lastly, it was found that Elo and RankSVM have the highest accuracy rates, at 74% and 77%, respectively.

Chapter 5

Automatic Biometric Signatures

5.1   Introduction

The preceding chapter focussed on using comparative semantic facial profile attributes for recognition within a semantic space. This involves using a semantic probe within a semantic database. The findings discussed in Chapter 4 demonstrate the efficiency of using comparative facial profile soft biometrics for subject recognition. Previous chapters have explored the task of recognising individuals within soft biometric databases derived from verbal descriptions, such as crowdsourced comparisons. However, in the field of criminal investigations, identification requires querying databases that include various modalities, such as mugshots or surveillance footage (Davis et al. 2015). These databases play a crucial role in linking verbal descriptions provided by eyewitnesses with an unknown suspect. Therefore, there is a need for a framework that can autonomously generate biometric signatures from facial profile images in order to establish a new visual space that bridges the semantic gap between machines and humans in terms of facial profile attributes.

The aims of the chapter are: (1) establishing an association between the traditional facial profile biometric system and the soft biometric system by numerically mapping features extracted from a facial profile images to its soft biometric attributes; and (2) assessing the impact of integrating soft biometric systems with traditional biometric systems to improve facial profile recognition performance. The experiments in this chapter utilised the entire relevant subset of images from the XM2VTSDB dataset, which includes 230 subjects with four facial profile images each. The generated Elo comparisons were used as labels, as detailed in Section 4.3. The objective was to use image features to predict ranks using a ranking system.

Section 5.2 describes the framework that was used to extract visual features from the facial profile images. In Section 5.3, the process for automatically retrieving biometric

signatures from the images is explained. Subsequently, Section 5.4 explores human facial profile recognition using biometric signatures obtained automatically from facial profile images. Finally, Section 5.5 provides an overview of the findings.

5.2 Extracting Visual Features from Facial Profile Images

The previous two chapters primarily focused on evaluating the recognition of human features using annotations provided by humans. Subsequently, the main challenges that remained unresolved was determining the association between traditional and soft facial profile biometric systems and whether we could establish a numerical mapping of features extracted from a facial profile based on its soft biometric attributes. Another question was to what extent comparative soft biometrics can improve traditional biometric performance. This study addressed these questions by examining recognition systems in different scenarios, including human-labelled facial profile traits (soft biometrics), automatically extracted attributes (traditional biometrics), and a fusion of both biometric systems. The main inspiration for this study came from the methodology used in (Nawaf Yousef Almudhahka et al. 2017a), which assessed comparative face labels and created biometric signatures based on the relative rates of traits. The procedure consisted of two steps: (1) detecting facial profile landmarks and (2) mapping traditional facial profile features to their corresponding soft biometric attributes.

5.2.1 Facial Profile Detection
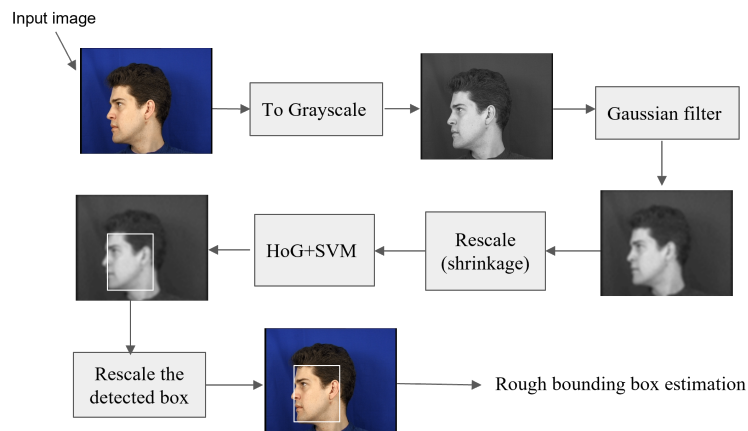
FIGURE 5.1: Facial profile detection process employing HOG and SVM.

To begin with, our focus was on detecting the bounding box of a facial profile in all of the images in the dataset and then identifying the important landmarks within the bounding box. Figure 5.1 illustrates the sequential steps required for accurately detecting facial profiles. To achieve this, we used a pre-trained histogram of oriented gradients (HOG) with a linear support vector machine (SVM) object detector, as introduced in (Dalal

et al. 2005). The study demonstrated that HOG performs better than existing methods like wavelets in detecting humans. HOG is a descriptor used to extract features from an image by analysing its gradients. It is primarily used in facial recognition and object detection and can be extracted from greyscale images. Gradients are typically more prominent around the edges and corners, which facilitates the effective detection of these regions in facial profiles. Figure 5.2 shows an example of using HOG to detect the facial profile in a sample image from the XM2VTSDB dataset. Following this, SVM was used to classify the images using the HOG method.

(a) (b)
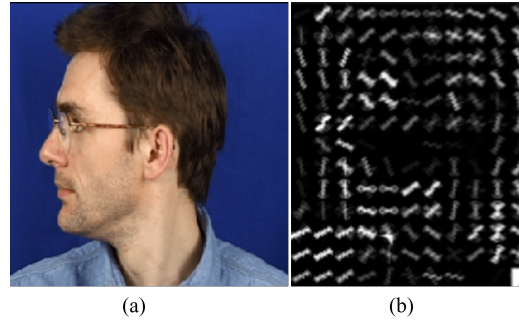
FIGURE 5.2: Using HOG to detect the facial profile for a sample facial profile image from the XM2VTSDB dataset: (a) facial profile image and (b) HOG.

Following this, the objective was to segment the different components of the facial profile, such as the eye, nose, and so on. This segmentation process enabled the automatic generation of features associated with each characteristic, as indicated in Table 3.1. The pose estimation method described in (Kazemi et al. 2014) was used for this purpose.

5.2.1.1 Correcting the Bounding Box

In cases where the facial profile was not accurately identified in the previous step, we used bounding box regression (Y. He et al. 2019) to improve the shape of the box, as shown in Figure 5.3. This method considered the average quadratic error of pixel position for each of the four sides of the detected bounding box.
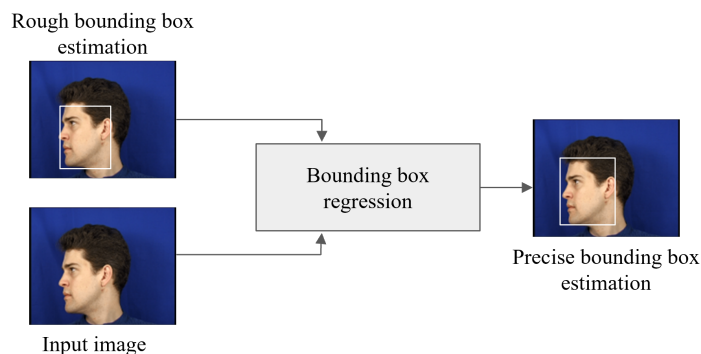
FIGURE 5.3: Correction procedure for the facial profile bounding box.

5.2.2 Landmark Detection

While there are various methods and models available for landmark detection, such as local binary features (LBF) (Ren et al. 2014), ensemble of regression trees (ERT) (Kazemi et al. 2014), and deep alignment network (DAN) (Kowalski et al. 2017), they face challenges including insufficient illumination, extreme head poses, and occlusions. ERT method, also known as fast face alignment methods, rely on cascades of regressors. Face alignment, a crucial component in most face analysis systems, aims to locate several key points on human faces in images or videos. The ensemble of regression trees can directly estimate the positions of face landmarks using a sparse subset of pixel intensities.

The majority of face alignment algorithms rely on a face detection bounding box to initialise the shape. To detect facial landmarks, the face must first be extracted from the image. This extracted region of interest (ROI) is then used to obtain the landmarks. Shape predictors, also known as landmark predictors, are used to predict key (x,y)-coordinates of a given 'shape'. Following the algorithm proposed by Tzimiropoulos et al. (2013), shape predictors are used to locate individual facial structures such as the eyes, eyebrows, nose, lips/mouth, and jawline. They require two inputs: the greyscale version of the image and a rectangle object containing the coordinates of the face area.

5.2.2.1 Handling Missing Labels

In our study, the proposed algorithm did not successfully detect key points for the ear and neck. To solve this issue, we expanded the facial profile bounding box to include the neck area and the ear. Thus, we annotated the bounding box to cover the neck and the expected position of the ear. Following that, we trained the predictor to locate the bounding box for the neck and the ear using accurate points from the training data. Afterwards, we identified a larger bounding box that encompassed the facial profile, ear, and neck. This enlarged bounding box was then inputted into the pose estimator to predict the landmarks. Since our focus was specifically on facial profiles, the resulting landmarks may not have been completely accurate. Therefore, to further enhance the accurate localisation of facial profiles landmarks in an image, we used custom shape predictors, following these steps:

- Annotated the dataset with the main key points.

- Trained the shape predictor using a designated training set.

- Evaluated the performance of the shape predictor using a separate testing set.

The outcome was satisfactory, with 518 images containing all of the landmarks, which was sufficient for constructing a model face with typical landmarks. Once fitted, this

model could be used to estimate the missing points. We divided the 518 images into training and testing sets for the shape predictor. Figure 5.4 displays a sample face image from the XM2VTSDB dataset alongside the estimated shape and the ground truth shape. The resulting model exhibited the following characteristics:

- **Accurate:** Our shape predictor effectively predicted and localised the locations of most facial profile landmarks.

- **Small:** The size of our facial profile landmark predictor was smaller than the original pre-trained face landmark predictor.

- **Fast:** Our model demonstrated a faster performance compared to dlib's pre-trained facial landmark predictor, as it predicted fewer locations.
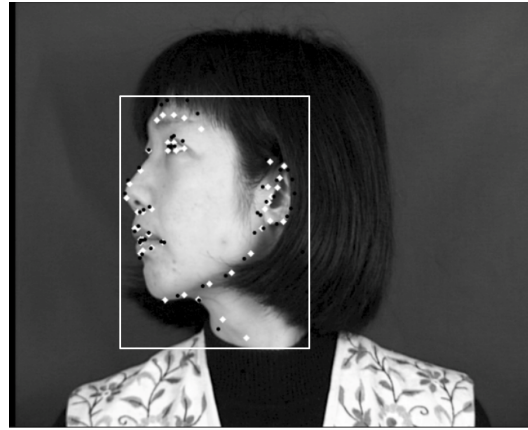
FIGURE 5.4: Localisation of facial profile landmarks for a sample facial profile image from the XM2VTSDB dataset. The black circle represents the estimated shape and the white dots represent the ground truth shape.

5.2.3 Face Alignment Using Rigid Similarity Transformation

The alignment and transformation framework was based on the relevant literature (Zeng et al. 2011; Walker et al. 1991). These transformations involved estimating a combination of rotation, translation, and scale that mapped the key points from one set to another on a template. Since our objective was to align all of the detected faces with a template image to minimise the rotational and scale differences, we considered the transformation as a planar rigid transformation with scaling. This approach ensured consistency when comparing subjects and reduced the impact of pose variations. The optimal parameters of the transformation, including scaling, rotation, and translation, were calculated. Additionally, all of the images in the dataset were normalised to a common reference face.

The algorithm aimed to minimise the quadratic error between the positions of the template key points and the transformed positions of the query. To effectively address this

problem, we used the matching step in the iterative-closest-point algorithm (Horn et al. 1988). This approach demonstrated a high level of efficiency when handling facial profile alignment in the dataset. An example of the transformation is shown in Figure 5.5. The next step involved segmenting the components of the facial profile. We used a total of twelve parts, and Figure 5.6 shows eight of these parts as examples.

(a) reference   (b) new   (c) after transformation

FIGURE 5.5: Applying the transformation, where (a) is a reference image, (b) is the registered image, which rotates around the xy axis, and (c) shows the transformation effect.

5.2.4 Generating Visual Features

After segmenting the main components of the principal facial profile, as shown in Figure 5.6, the goal was to train models to predict the attributes from the images, which will be discussed later in this chapter. To achieve this, the global image feature (GIST) was used to create a visual descriptor vector for each attribute of the facial profile (Oliva et al. 2001; Xie et al. 2018). The extraction of the GIST features was successfully employed by Parikh et al. Parikh and Grauman (2011) to address issues related to face description and recognition of outdoor scenes.

Facial Profile Detection   Facial Profile Landmarks Detection   Facial Profile Component Segmentation
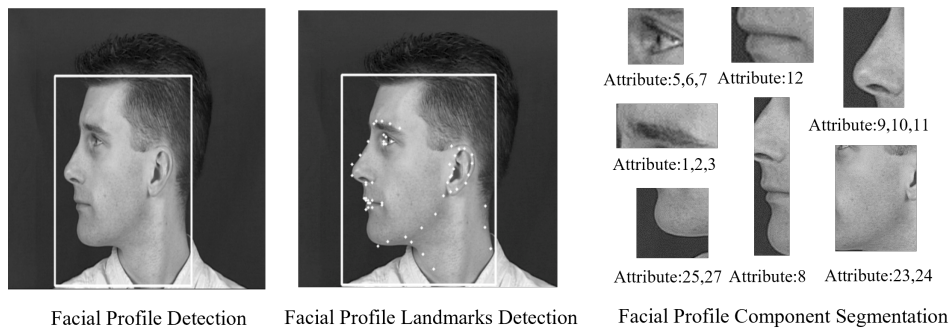
FIGURE 5.6: Facial profile landmark detection and facial profile component segmentation alongside their corresponding traits.

The process began by normalising the intensity of the facial profile component image. Next, a series of Gabor filters were applied at four scales, with eight orientations per scale, to extract the GIST features. These features were then used to create a set of

features. The orientation maps, which consisted of 32 orientations, were divided into 44 grids and the mean intensity for each block in the grid was computed. As a result, each facial profile component produced a vector of 512 features. Later in this section, we will discuss how these features were used to train a model that aimed to generate comparative facial profile labels. By utilising the generated features, we were able to examine the relationship between our traditional facial profile biometric system and the soft biometric system. This was done by numerically mapping the features extracted from a facial profile to its soft biometric attributes.

5.3 Retrieval of Biometric Signatures

In the previous section, we discussed the generation and extraction of visual features used to retrieve biometric signatures from facial profile images. This section now introduces an approach to estimate the relative rates of attributes based on these visual features. Additionally, it outlines and evaluates the effectiveness of this approach in predicting soft biometric features.

5.3.1 Mapping Traditional Facial Profile Features to Soft Biometric Comparative Attributes

The GIST features extracted from images were associated with comparative soft biometric labels. These labels indicated the difference between two subjects as 'Less' or 'More', which corresponded to -1 or 1. To automatically estimate the target soft biometric labels, a supervised regression method was employed to learn a prediction model based on the visual features. We trained 31 sets of linear regressors, where the visual elements were considered independent variables used to predict a dependent variable (i.e., a comparative label). Linear regression was chosen because it explicitly takes into account order, enabling the generation of comparative labels. The regression model was trained using the visual attributes of the training subjects, along with the corresponding normalised relative rates obtained using Elo, with the help of crowdsourced comparative labels (ground truth). Multiple linear regression simulated a linear relationship between a set of independent variables (features) and their corresponding dependent variables (labels), allowing the prediction of the label of a new (unknown) sample based on its features. The objective of training the regressors was to establish a direct correlation between visual characteristics with GIST descriptors and comparative label values (-1 or 1) for the 31 face profile traits (see Table 3.1). At this stage, we decided to exclude neck features from the table as they exhibited less power compared to other features. The dataset was divided randomly into four folds. Three of the folds, each containing m comparisons ($m = 20,700$), were used to train the regressors. The fourth fold, containing n comparisons ($n = 6,900$), was used for testing. This resulted in a total of $m + n$

comparisons for each attribute t, while $1 \leq t \leq 31$. The following model explains the relationship between the visual aspects of each attribute and the comparative labels:

$$y_c^t = \beta_0 + \beta_1 x_{c1}^t + \beta_2 x_{c2}^t + ... + \beta_p x_{cp}^t + \epsilon_t \qquad (5.1)$$

The comparison index, denoted by $1 \leq c \leq m$, y_c^t represents the comparative label value for $y \in \{-1, 1\}$ of the comparison c for the attribute t. x_{ck}^t is the difference in the kth GIST feature between the two individuals comprising the appraisal c for the trait t, where $1 \leq k \leq p$ describes the GIST feature index, and ϵ_t represents the error term. The goal of training is to estimate a vector of weight coefficients $\hat{\beta} = \beta_0, \beta_1, ..., \beta_p$ in order to minimise the sum of squared residuals (equation 5.2):

$$minimise \left(\frac{1}{m} \sum_{c=1}^{m} |y_c^t - \hat{y}_c^t|^2 \right) \qquad (5.2)$$

The value of the comparative label to be predicted by (5.1) is represented by y_c^t, while the predicted comparative label is denoted by $\hat{y}_c^t$. The projected comparative label b for a trait t between individuals (i, j) constituting a comparison in the testing group was derived as shown in equation 5.3 once the weight coefficients vector $\hat{\beta}$ was generated:

$$b_d^t(i, j) = \begin{cases} -1 & \text{if } \hat{y}_d^t < 0 \\ 1 & \text{if } \hat{y}_d^t \geq 0 \end{cases} \qquad (5.3)$$

The comparison's index in the testing group was denoted by $1 \leq d \leq n$. As explained in the previous chapter, the anticipated comparison labels were utilised to derive comparative rates for the 31 traits through the Elo rating. These relative rates were then used to generate biometric signatures for all of the subjects, comprising 31 relative rates constituting soft biometrics used for subject recognition. After the comparative labels were predicted from the facial profile images, we determined the associated rating for each attribute using Elo.

5.3.2 Analysis

The purpose of this investigation was to assess the correlation between predicted relative rates based on visual features (visual space) and those generated through crowdsourced comparative labels using the Elo ranking method. The analysis evaluated the accuracy of the regression model presented in this section and the visual descriptors discussed in Section 5.2.4. Spearman's correlation coefficient ρ was used to quantify the accuracy of the predicted relative rates based on visual features and their correlation with semantically derived ground truth.
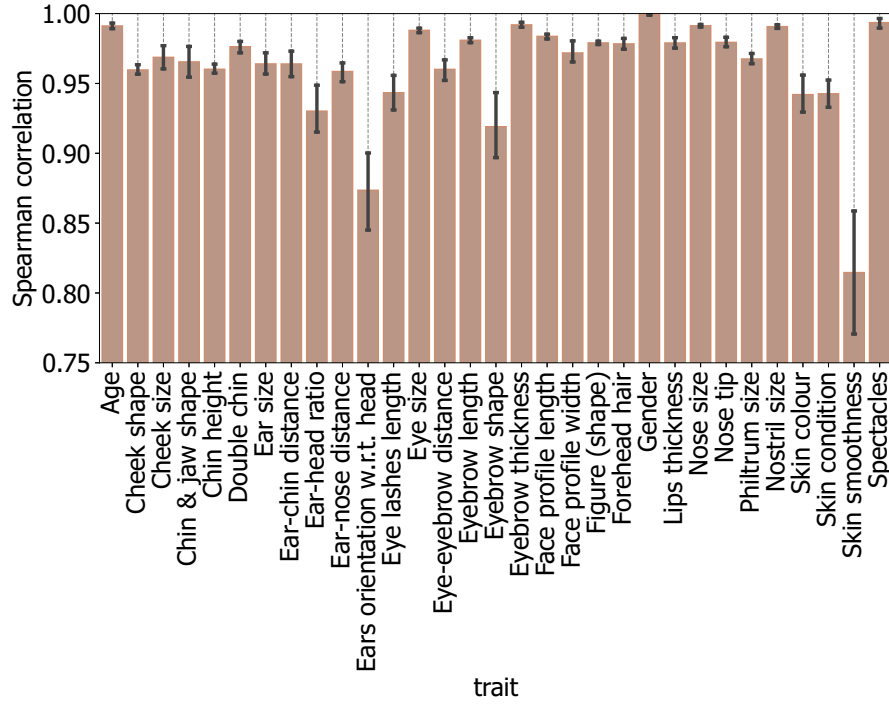
FIGURE 5.7: Spearman's correlation coefficient is used here to average and quantify the correspondence between soft and traditional generated labels.

Figure 5.7 illustrates the relationship, as well as the Spearman's ρ correlation, between soft and traditional comparative facial profile biometric features. These features were evaluated using Equations (5.1) and (5.2). As illustrated in Figure 5.7, clearly indicated a strong relationship between the variables. The strongest relationship was observed for binary-like attributes (e.g., gender), followed by nose size and nostril size. However, most eye and ear characteristics exhibited the least, yet still significant, correspondence. However, the discriminative capability of soft biometric facial profile features significantly influences traditional biometrics, as highlighted in Figure 5.7. Thus, this emphasises the necessity of combining soft biometrics and traditional biometric systems.

5.4 Facial Profile Recognition Performance

5.4.1 Construction of the Features Gallery

In our experiments, we developed and tested three different types of features: human-labelled, automatically labelled, and fusion features, as described below:

- The human features gallery (soft biometric), as explained in Chapter 4, was created by comparing pairs of images based on specific characteristics. This comparison was done using a platform that contained a collection of image pairs representing

facial profile traits. A ranking system was then used to convert the pairwise comparisons into scores that were associated with the images. The set of rankings created for each characteristic was used to construct a feature vector for each facial profile. As a result, each subject had a unique biometric signature.

- The automatic features gallery (traditional biometrics) was built using trained regressors. These regressors generated comparative labels for each subject, based on the visual GIST features. Additionally, a biometric signature was created for every participant in the dataset using the Elo ranking system. To train the regressors, facial profile images were paired and compared.

- Finally, once the feature vectors had been extracted for both annotation systems, a fusion vector was created by concatenating these feature vectors into a single vector for each image.

5.4.2 Facial Profile Recognition

This experiment aimed to assess the effectiveness of identifying an unknown topic (probe) from each gallery described in the previous section. We employed a leave-one-out cross-validation (LOOCV) technique to calculate the recognition rate. Using recursive feature elimination (RFE) (Guyon et al. 2002), we utilised XGBoost (T. Chen et al. 2016) as a classifier. In this investigation, we used the 31 features listed in Table 3.1. With soft biometric feature vectors, we achieved an average recognition rate of 97%, while with GIST-related features, we attained an average recognition rate of 84%. In addition, the fused feature vectors achieved an average recognition rate accuracy of 98% (see Figure 5.8). This result demonstrates the significant contribution of eyewitness testimonies (soft biometrics) in improving the performance of traditional biometrics by 14%.
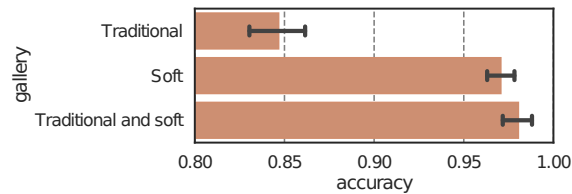
FIGURE 5.8: Average accuracy for each of the features galleries.

The DET curve was used to evaluate the system performance and demonstrate the trade-off between the false positive rate (FPR) and false negative rate (FNR). Figure 5.9 shows that by combining features from eyewitness testimonies and facial profile images, the system achieved an equal error rate (EER) of 0.86%. In comparison, the soft biometric and traditional biometric systems had EERs of 0.95% and 2.72%, respectively. In summary, our findings show that fusing soft and traditional comparative facial profile systems can

greatly improve recognition accuracy. Moreover, combining eyewitness testimonies and computer vision techniques can result in a biometric system with higher performance.
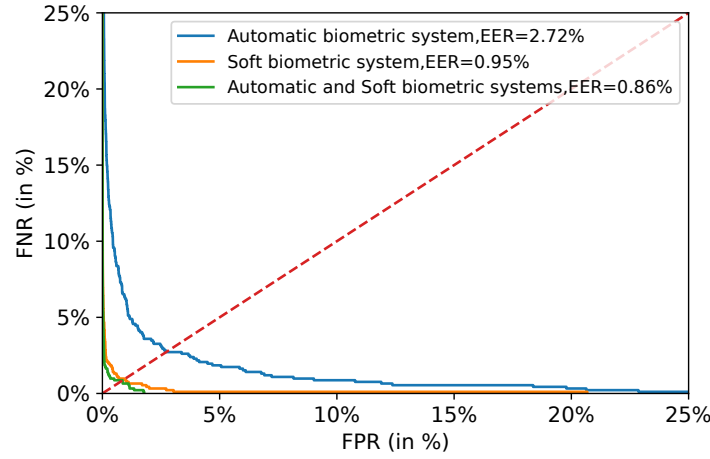
FIGURE 5.9: DET curve for each of the three biometric systems.

XGBoost handles multiclass classification by treating it as a series of one-versus-all binary classification problems. Moreover, the prediction for a specific instance was stored in a vector with K components, where each component j represented the probability (defined by the classifier) of belonging to class j. Furthermore, we evaluated performance using the CMC curve. To construct a CMC curve (DeCann et al. 2013), we needed to determine the k-best class predictions for a given sample. This was achieved using the estimated class probabilities as a measure of similarity: a higher probability of belonging to a class indicates that the sample is more similar to the typical sample from that class. Therefore, we rearranged the indices of the prediction vector in decreasing order of estimated probabilities. A prediction is considered rank-k accurate if the true label appears within the first k reordered indices. Figure 5.10 depicts the recognition performance for facial profile traits, comparing the use of soft, traditional, and fusion facial profile features. As can be seen from Figure 5.10, the system that combined soft and traditional biometrics achieved an accuracy of 98% at rank-1, which further increased to 100% at rank-15.

In a study by Nawaf Yousef Almudhahka et al. (2017b), 24 comparative attributes were used and the identification of 430 subjects from the LFW dataset was evaluated. The study achieved a retrieval rate of 98.37% at rank-5 for human features and a retrieval rate of 90.93% at rank-10 for automatic features. In comparison to Nawaf Yousef Almudhahka et al. (2017b), we achieved a promising level of accuracy using facial profiles. It was expected that the recognition rate would be lower for facial profiles since only half of the face is visible and not all facial information is accessible to observers. However, our experimental results showed a 98% recognition rate for our facial profile recognition

FIGURE 5.10: Recognition through CMC performance.

system, indicating that facial profiles are sufficient to qualify as a separate biometric modality.

5.5 Conclusions

This chapter explored the significance of facial profiles in biometric systems used for security and surveillance. It also investigated the impact of human comparative labels on the performance of facial profile recognition. Furthermore, the chapter examined how the integration of comparative soft biometrics with traditional biometrics can enhance the effectiveness of biometric systems specifically designed for facial profile recognition. To automatically determine comparative facial profile labels, we employed a method that combined computer vision with human descriptions of facial profiles. Additionally, we conducted experiments with three biometric systems using a subset of the XM2VTSDB dataset.

We demonstrated that it is possible to automatically generate facial profile descriptions for human recognition. Our developed approach allows for subject identification and retrieval utilising either an image or a verbal description, reflecting forensic circumstances in which a suspect's image or an eyewitness's descripti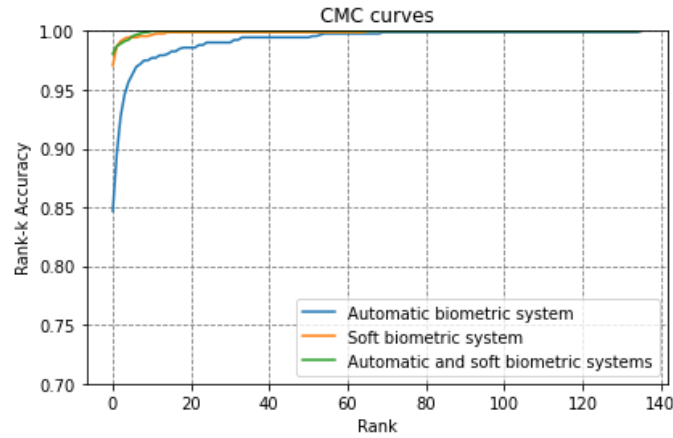on are available. The experimental results obtained a 98% recognition rate for our facial profile recognition system, indicating that facial profiles are sufficient to qualify as a separate biometric modality. Our findings also suggested that combining automatic biometric systems with comparative facial profile soft biometrics in humans leads to a significant improvement in recognition accuracy.

Chapter 6

Facial Profile Biometrics: Domain Adaptation and Deep Learning Approaches

6.1 Introduction

Previous studies have established that human facial profiles are regarded as a biometric modality, and there is evidence of bilateral symmetry in facial profile biometrics (Toygar et al. 2018). This chapter centres on the bilateral symmetry of human facial profiles and offers an analysis of facial profile images for recognition purposes. We put forward a method based on the few-shot learning framework for extracting facial profile features. Through the utilisation of domain adaptation and reverse validation, we also introduce a technique known as reverse learning (RL) to explore bilateral symmetry.

As discussed in Chapter 2, various deep learning methods have been developed for facial profile recognition. These methods propose techniques for extracting the relevant features from facial profiles (Sengupta et al. 2016; J. Zhao et al. 2018; P. Li et al. 2019; Yin et al. 2020). In recent years, these methods have made significant progress, achieving impressive identification rates even in challenging conditions by leveraging extensive training on massive datasets (Parkhi et al. 2015; Deng et al. 2019; Meng et al. 2021). However, it is important to note that all deep learning methods require a large number of samples per class during training to achieve acceptable performance levels. In our dataset, we only had eight samples per class. Few-shot learning (FSL) has emerged as a strategy to address this challenge. It allows pre-trained deep models to be used on new data with only a small number of labelled examples, eliminating the need for retraining. In this chapter, we introduce a few-shot learning approach based on reverse validation

and domain adaptation. The goal was to achieve high accuracy without requiring extensive training data. Moreover, we investigated the impact of soft biometric features when combined with traditional biometric features obtained through few-shot learning.

Section 6.2 focuses on the image preprocessing phases, while Section 6.3 explains the feature extraction stage using two few-shot learning techniques. After that, Section 6.4 describes the experimental design, recognition performance, and subsequent discussions. Section 6.5 examines the impact of fusing comparative soft biometrics (semantic space) obtained from the RankSVM ranking system with traditional biometrics (visual space) obtained from few-shot learning. Finally, Section 6.7 summarises the conclusions drawn from the previous sections and highlights the implications of the study.

6.2   Image Preprocessing and Descriptions

Chapter 5 provided a detailed explanation of the various steps involved in processing the images from the dataset to identify facial profiles and key points. In this process, we used two convolutional neural network (CNN) models to extract the features. It is important to note that each model has specific requirements for the input image sizes. Figure 6.1 illustrates all of the procedures involved in extracting these features.
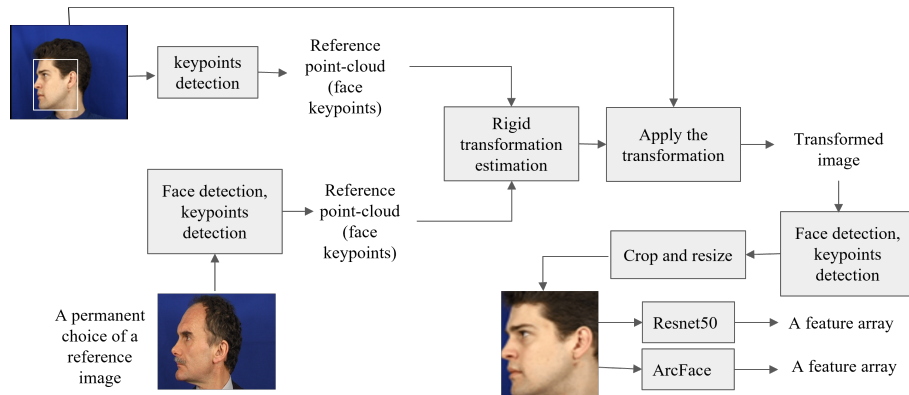
FIGURE 6.1: Image preprocessing procedures.

6.3   Extracting Visual Features from Facial Profiles

Deep neural networks are widely used in computer vision for extracting features. In this section, we propose a structure inspired by previous research (Y. Wang et al. 2020; Parnami et al. 2022; Garcia et al. 2017) to leverage few-shot learning based on the ResNet50 neural network (K. He et al. 2016) and ArcFace (Deng et al. 2019). A typical CNN consists of feature extraction and classification components. During training, the model learns the unique facial features and generates feature embeddings through the feature extraction process. In the FSL strategy, after training is completed, it becomes

possible to skip the classification step and generate feature embeddings for each face image. Figure 6.2 illustrates the CNN architecture and the key stages used for feature extraction.
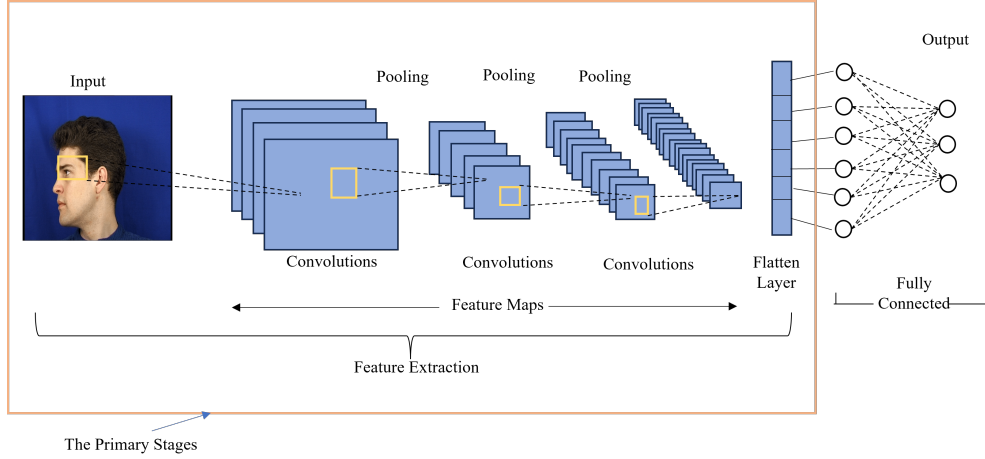
FIGURE 6.2: Convolutional neural network architecture.

- **ResNet50** (K. He et al. 2016), short for residual network, is a specific type of CNN known for its ability to maintain a lower error rate, even in deeper networks. Our network accepts input images of size 244×244 pixels and computes 7×7 grid feature maps in the last layer before the fully connected network. The pretrained model was optimised on the ImageNet dataset, which consisted of multiple classes. Due to the difference between the datasets used to train the networks and our target data, we chose to exclude the output of the last fully connected layer. Our numerical experiments demonstrated that the learned feature space efficiently represented human faces using these $7 \times 7 \times 2048$ feature maps, resulting in 49 2048-dimensional vectors.

  We then used principal component analysis (PCA) (Jolliffe 2002) as the next step to reduce the dimensionality of the feature space. PCA helps in preserving important information patterns while eliminating redundant information. After reducing the dimensionality, we used the feature vectors to train a KNearest Neighbour (KNN) recognition method with $k = 4$. To prevent overfitting, we used approximately $\sim 50\%$ of the eigenvectors that captured the most significant variations. To train the classifier, we used two samples per class. One sample was used for the validation process, and the other sample was used as unseen test data for performance evaluation. Furthermore, in our experiments, we reduced the dimensionality of the feature vectors from 700 to 477 using SFFS and PCA algorithms. Figure 6.3 shows that with 477 components, we achieved a variance of nearly 95%.

- **ArcFace** (Deng et al. 2019) is an innovative deep face recognition algorithm proposed by Jiankang Deng et al. It is considered a discriminative model. The authors proposed the additive angular margin loss function, which has proven to be highly
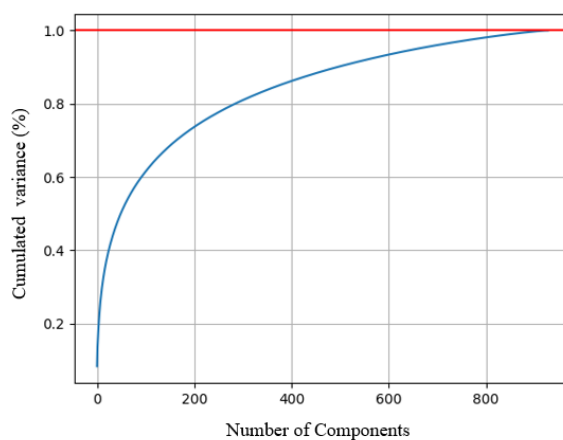
FIGURE 6.3: The number of components needed to explain variance.

effective in obtaining discriminative features for facial recognition. It consistently outperforms other state-of-the-art loss functions. ArcFace addresses the challenge of effectively learning discriminative face features by incorporating an angular margin that pushes the learned features of different classes apart in the angular space. By enhancing the separability between classes, ArcFace significantly improves the model's ability to distinguish between similar faces. This model is trained using the LFW dataset to obtain pre-trained weights. The LFW dataset consists of 13,233 facial photos collected from the web. The ArcFace model expects inputs of size 112×112 and returns 512-dimensional vector representations. Importantly, ArcFace simplifies the process by compressing the extracted features to only 512 components, thereby eliminating the need for PCA.

We selected ResNet50 (Montero et al. 2022) as the backbone for all of the network architectures tested in the ArcFace repository, due to its optimal balance between accuracy and parameter count. In this context, the base model was ResNet50, and we utilised ArcFace as the loss function. Generally, the term backbone refers to the feature-extracting network responsible for processing input data into a specific feature representation. The backbone plays a crucial role in extracting and encoding features from the input data, capturing both low-level and high-level features. After feature extraction, we calculated the ArcFace loss and used it to update the network parameters through backpropagation. We chose ArcFace (Deng et al. 2019) as the baseline for two reasons: it employs a SoftMax-loss-based methodology, eliminating the need for an exhaustive training data preparation stage, and it has been demonstrated to yield the best results for the original face recognition task.

6.4 Facial Profile Recognition Using Data Adaptation and Reverse Validation Methodology

Domain adaptation, a subfield of machine learning, aims to train a model on a source dataset and achieve a high level of accuracy on a target dataset that differs significantly from the source (Farahani et al. 2021). Conversely, if some samples in the target dataset lack labels, reverse validation (Morvant et al. 2011; Y. Wang et al. 2020) is employed to identify the best model that minimises the difference between the conditional distributions of the source and target datasets. In this chapter, we propose a method grounded in domain adaptation and reverse validation, assuming that some facial profile image samples lack labels and exhibit distinct conditional distributions from the source (training) dataset. Inspired by Ganin et al. (2016), we introduced an algorithm called reverse learning (RL) based on domain adaptation and reverse validation. Unlike conventional prediction methods, our approach employed a two-step prediction process. We use SFFS (Shirbani et al. 2013) to select the most optimal features from the distributions of the training and validation data. In our RL algorithm, the two-step prediction helped identify the best model with optimised hyper-parameters and features between the training and validation datasets, addressing issues associated with covariate shift assumptions (David et al. 2010).
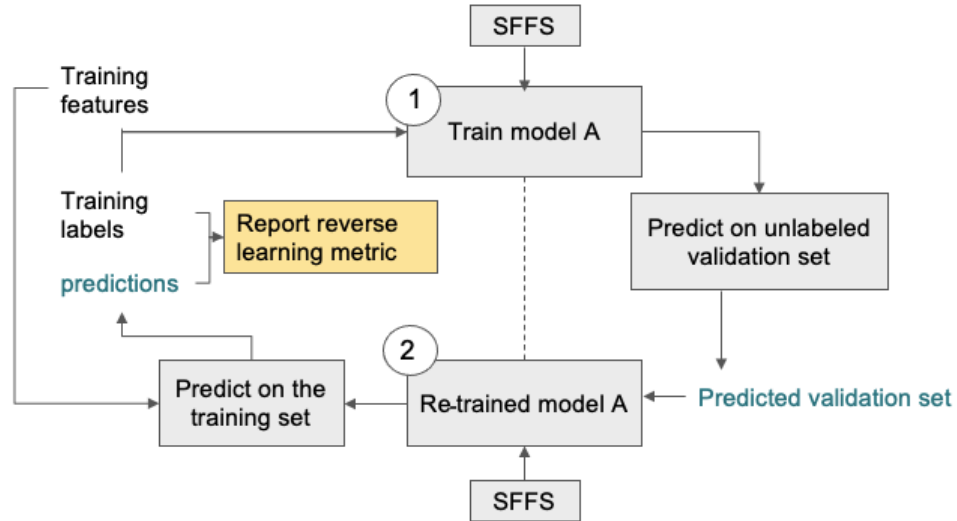
FIGURE 6.4: Training process of our reverse learning (RL) algorithm: (1) training the model on the training set, (2) training the model on the validation set.

In the first step, the process began with a simple training procedure using SFFS, As shown in Figure 6.4. In the second step of the prediction process, after making predictions for the validation set, we trained on the validation set, made predictions for the training set, and aimed to minimise errors using the SFFS algorithm. In the second prediction step, the validation set served as the training set. The second prediction was finally used to quantify the error on the training set, named the reverse learning

metric. Finally, we evaluated the performance of our system by presenting unseen test data from the target dataset. We assessed the performance of the RL method using the XM2VTSDB database, which included 230 subject with four facial profile samples from each side-view.

6.4.1 Performance Analysis

In this section, we introduce a domain adaptation approach that is employed to assess the effect of this strategy on the recognition accuracy of facial profiles from opposite sides of the face. We detected a noticeable shift in data within the same feature space when evaluating the model fitted on the training set. This discrepancy arose from the limited number of samples per class and the variations in subjects' appearances across the four images, as well as between the training and test sets. To address this challenge, we employed a technique that aimed to identify the gap between the training, validation, and test sets.

We conducted a series of experiments on side-view face recognition to assess the effectiveness of the RL method compared to traditional approaches. We then employed a similar technique in two distinct settings:

- *Experiment for same view facial profile (left side)*

  In this setting, we employed the leave-one-out cross-validation (LOOCV) method to measure the recognition rate. In addition, we opted to use KNN as a simple classifier. The model was trained and evaluated from the perspective of the left side, with the test data also originating from the left side. We evaluated the performance of this experiment by analysing four samples from the left side view of each individual in the dataset. For this experiment, 80% of the data on the same side was allocated for training and validation, while the remaining 20% served as the unseen testing set. By utilising ResNet50 features, we achieved an average recognition accuracy of 85% when training KNN with 80% of the samples in the dataset (traditional strategy). Interestingly, we observed an increase in accuracy to 91% when employing the RL strategy. However, when using ArcFace features, the average recognition accuracy dropped to 71%. Notably, this accuracy remained consistent even with the application of the RL approach.

- *Experiment for mirror bilateral symmetry view (left and right side)*

  In this case, the facial profile side images were initially mirrored from right to left, following the way in which previous processes had been applied to detect facial profiles. We evaluated the performance of this experiment by analysing eight samples from the left and right side view of each individual in the dataset. In this experiment, we trained and validated on the left side and measured the

recognition rate on the right side as unseen test data. Our numerical experiments revealed that utilising RL can enhance recognition accuracy to 75%, compared to 70% with the traditional method, based on ResNet50 features.

Additionally, the model was trained on left-side facial profiles and validated on unlabelled right-side profiles. Testing was then conducted using unseen data from the right side. By using right facial profiles as unlabelled data in the validation process, a recognition rate of 82% was achieved. This marked a significant improvement of 12% compared to traditional methods. Interestingly, when ArcFace features were utilised, the accuracy remained consistent across all three techniques, at 56%. However, it is worth noting that the RL approach did not lead to any enhancement in accuracy.
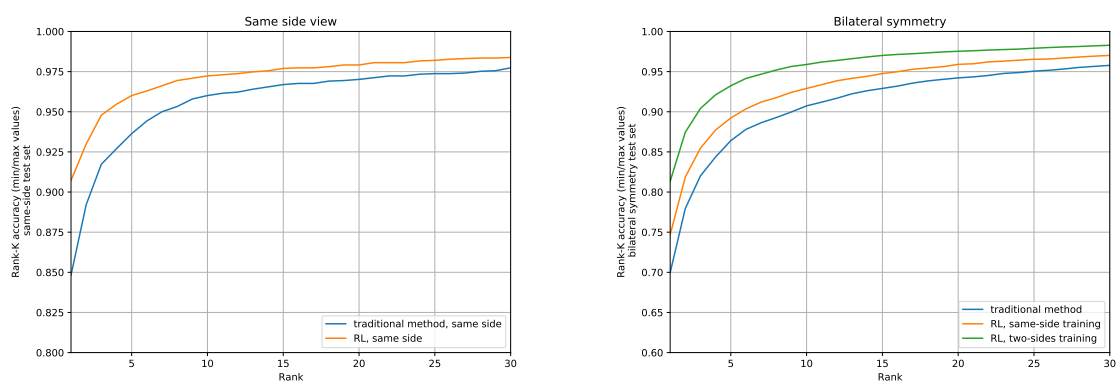
FIGURE 6.5: Recognition via CMC performance for ResNet50 features.
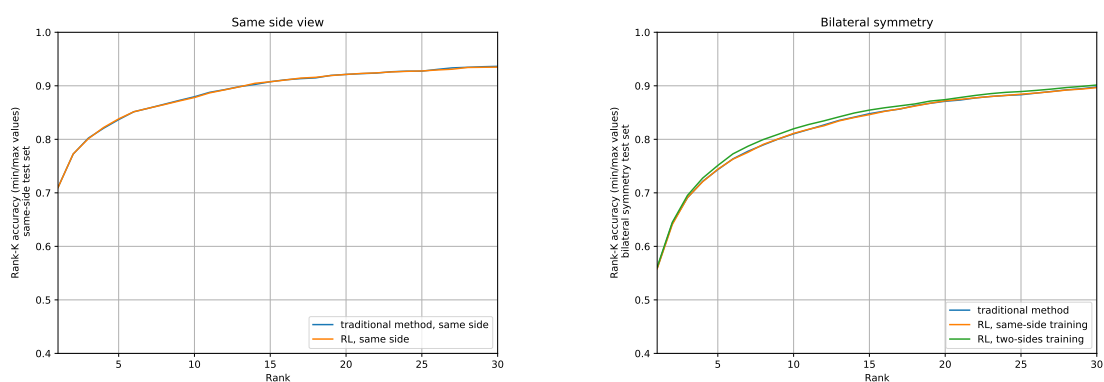
FIGURE 6.6: Recognition via CMC performance for ArcFace features.

One of the most important techniques for evaluating model performance is cumulative matching characteristics (CMC) curves (DeCann et al. 2013). The CMC curve shows the recognition at different ranks, indicating the probability of finding an accurate match at a particular rank. Figures 6.5 and 6.6 present the recognition performance of facial profiles.

In Figure 6.5, which used ResNet50 features, the accuracy for same-side recognition improved from 85% at rank-1 to 96% at rank-10 when employing RL. Additionally, the accuracy for BS increased from 71% at rank-1 to 91% at rank-10 when using only right-side images as test data. However, when right-side images were also included in the validation stage, the accuracy rose from 82% at rank-1 to 96% at rank-10.

In Figure 6.6, which used ArcFace features, the accuracy for RL on the same side improved from 71% at rank-1 to 87% at rank-10. Moreover, the accuracy for BS using only right-side images as test data improved from 56% at rank-1 to 81% at rank-10. Additionally, the accuracy for BS using only right-side images as test data improved from 56% at rank-1 to 81% at rank-10. Similarly, when right-side images were also used in the validation stage, the accuracy increased from 56% at rank-1 to 83% at rank-10. Our numerical experiment results are summarised in Table 6.1 and reveal the following two findings:

TABLE 6.1: Recognition rates with and without domain adaptation based on RL algorithm; (BS) bilateral symmetry, (L) left side, (R) right side, (T) training set, (V) validation set, (TS) test set, (SD) standard deviation.

View	Method	Dataset			Model			
					ResNet50		ArcFace	
		T	V	TS	Accuracy	SD	Accuracy	SD
Same side	Traditional	L	L	L	85%	0.016	71%	0.012
	RL	L	L	L	91%	0.014	71%	0.011
BS	Traditional	L	L	R	70%	0.021	56%	0.009
	RL	L	L	R	75%	0.011	56%	0.008
	RL	L	R	R	82%	0.042	56%	0.041

1) By utilising ResNet50 features, the recognition rate for the same side exceeded 85% with the traditional method and improved by 6% with the RL method. For the opposite side, the recognition rate using the traditional method is 70%, which increased to 75% with the RL method when the left side was included in both the validation and training sets. Notably, the accuracy significantly improved by 12% when right facial profiles were included in the validation process.

2) When employing ArcFace features, the recognition rates for both the same and opposite sides remained unchanged, and the model performance did not alter even after introducing the RL method.

Our numerical experiments utilising ResNet50 indicated that the RL method proposed here outperformed the state-of-the-art for datasets containing more than 200 subjects. The dataset utilised in this chapter differs from the one used in previous work. To clarify, our dataset presented greater challenges due to the following reasons: 1) it comprised 230 subjects, which is substantially larger than all datasets considered in (Ding et al. 2013; Bhanu et al. 2004; X. Xu et al. 2007), 2) some of the facial profiles in our dataset

were occluded, yet our method still achieved a high recognition rate of 91% using our RL method, surpassing all methods presented in (Ding et al. 2013; Bhanu et al. 2004; X. Xu et al. 2007), and 3) in the method proposed here, we assumed that the validation set was unlabelled. Such an assumption rendered our dataset more challenging than those in the literature.

In our study of 230 subjects, we achieved a cross-recognition rate of 82%, which is the highest among our results. This performance can be compared to the work presented in (Toygar et al. 2018), where only 46 subjects with left and right profile images were used. The best cross-recognition rate reported in (Toygar et al. 2018) was 81.52%, when the model was trained on the left side and tested on the right side. However, in our study, facial profiles with a larger number of subjects achieved an 82% cross-recognition rate. This indicates that facial profiles contain sufficient information to be considered an independent and important biometric modality. According to the findings of the experiments in (Toygar et al. 2018), facial profile recognition systems outperform ear recognition systems due to the higher bilateral symmetry of facial profiles. This is because the right and left facial profiles are more similar and share more discriminative traits than the right and left ears. Table 6.2 presents the recognition rates of facial profiles from various methods found in the literature.

TABLE 6.2: Recognition rates of facial profiles in the literature with various methods.

Publication	Dataset	Method	Accuracy
Same side view			
Ding et al. (2013)	44 subjects	DWT + RF	92.50%
Bhanu et al. (2004)	30 subjects	DWT	90%
X. Xu et al. (2007)	38 subjects	PCA	77.63%
Other side view			
Toygar et al. (2018)	46 subjects	BSIF + LPQ + LBP	81.52%

Table 6.1 presents the results which demonstrate that ResNet50 performed better than ArcFace in both same-side and cross-recognition experiments. While ArcFace demonstrated strong performance when dealing with a slightly angled view of faces and excelled in large-scale face identification tasks, its performance was lower when applied to facial profiles. The accuracy decreased from 99% on LFW to 71% in our dataset, indicating that ArcFace struggles to effectively generalise to new unseen data that features highly angled views of faces.

6.5 Fusion of Semantic and Visual features

The fusion method conducted in Chapter 5 has conclusively shown the efficacy of using comparative facial profile soft biometrics for subject identification. The results have

also demonstrated the inherent capacity of the semantic space for the recognition and identification of human faces using comparative soft biometrics. This section examines the extent to which soft biometrics could complete computer vision features. After extracting visual features from facial profile images, as explained in Section 6.3, we developed two fusion techniques: one was based on feature concatenation and the other exploited Adaboost to fuse semantic features obtained by RankSVM with visual features obtained by ResNet50, as shown in Figure 6.7. The first strategy acted at the features's space level and consisted of concatenating semantic features with visual features for each sample, then creating a classifier that took the new feature space as input. The second approach consisted of creating a classifier as a mixture of two classifiers, each one of them using a single type of features, ResNet50 or RankSVM exclusively.

A comparative evaluation of the classification metrics of both methods is conducted at the end of the section. Moreover, for the feature-based fusion method, we examined the impact of the concatenation on inter-class and intra-class disparity. The experiments focussed exclusively on the recognition of the left side of the facial profile, as we lacked verbal pairwise comparisons for the right side.
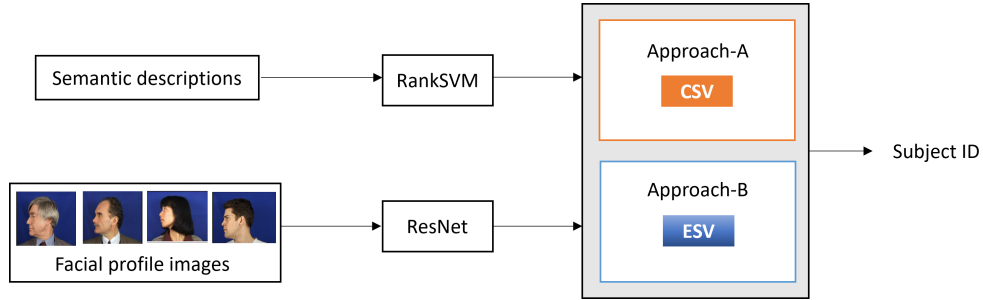
FIGURE 6.7: Two fusion approaches.

6.5.1 Concatenation Semantic and Visual (CSV)

As part of this approach, we concatenated the semantic features and visual features for each image in the dataset into a single feature vector. The used framework can be described as follows: Let $\mathbb{R}^{d1}$ and $\mathbb{R}^{d2}$ be the two vector spaces. The set $\mathbb{R}^{d1+d2}$ is the set of elements $(x : y)$ a concatenated vector of x and y, where $x \in \mathbb{R}^{d1}$ and $y \in \mathbb{R}^{d2}$.

The CSV technique was examined using the KNN classifier. The evaluation utilised the 230 subjects from the XM2VTSDB dataset and employed a 4-fold cross validation approach. After mapping a pair of vectors into a higher dimensional vector for each image in the dataset, we applied PCA after the concatenation to reduce the dimensionality, resulting in a reduction of the total number of features from 950 to the total number of images, which is 920. An SFFS with LOOCV was then used to select the relevant features and the best hyper-parameters. The algorithm selected 280 components. Figure 6.8 shows the method that was developed.

FIGURE 6.8: Framework of concatenation semantic and visual features.

6.5.1.1 Analysis and Evaluation

The objective of this analysis was to investigate the precision of the suggested CSV technique when identifying individuals base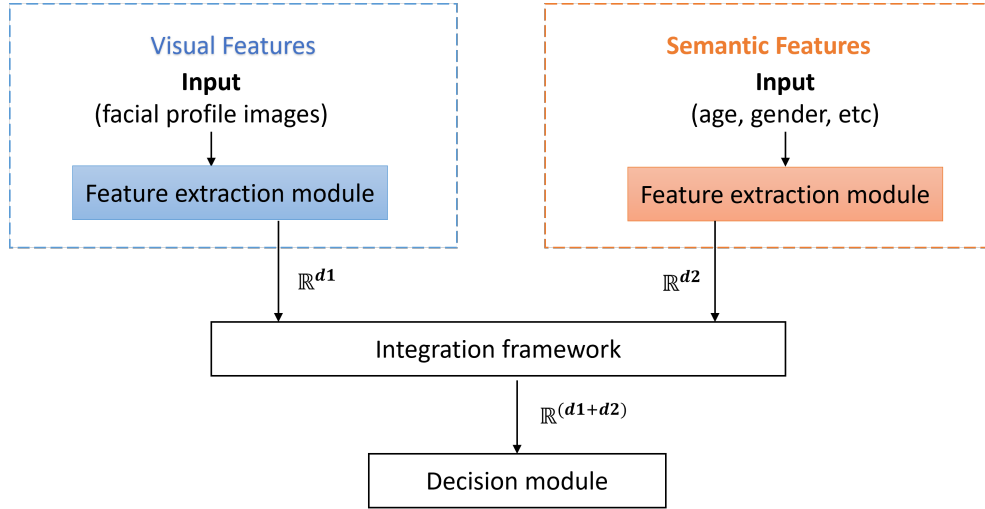d on facial profiles. Moreover, the objective of this analysis was to assess the integration performance between the visual features (representing the visual space) and the semantic space features inferred from the crowd-sourced comparative labels via the RankSVM algorithm. Our experiments demonstrated the benefits of utilising the facial profile soft biometric traits and fusing them with traditional biometric to enhance the recognition accuracy. Biometric recognition performance can be assessed using similarity measurement function as explained in Section 4.6.2. Figure 6.9 illustrates the overlap between inter and intra class variations for concatenation approach and visual features, while Figure 4.10 illustrates the overlap between inter and intra class variations for semantic features which computed previously. Figure 6.9 and 4.10 demonstrated that the concatenation approach and visual features had a similar overlap area between same and different classes, with a value of 0.148. In contrast, the overlap area for RanSVM was 0.212.

6.5.2 Ensemble Method for Fusing Semantic and Visual (ESV)

Ensemble methods are approaches used to increase the accuracy of outcomes in models by mixing several models rather than using a single one. The primary types of ensemble methods include bagging, boosting, and stacking. The model based approach (ensemble method) consists of learning a linear combination of two classifiers, each one of them exclusively using one type of feature. Such a model is called a generalised linear model. Ensemble methods can achieve such a combination. In stacking, the predicted classification probabilities of base classifiers are considered as a feature vector. In that case, a model is trained to use the probabilities efficiently and make final predictions. In
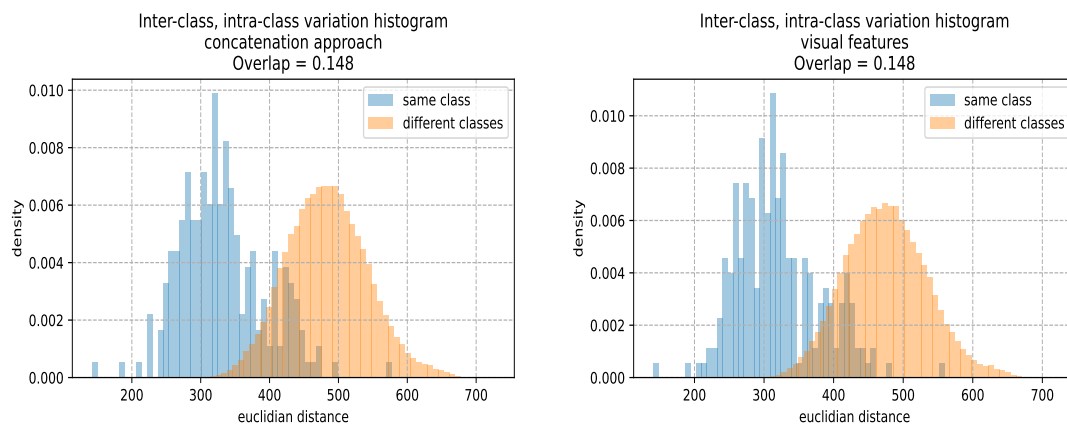
FIGURE 6.9: Inter and intra-class variations.

bagging for example, each individual classifier deals with a different set of features and set of samples from other classifiers in the ensemble. In boosting, each new individual classifier is trained to correct the misclassifications made by the current ensemble.

As we were interested in combinations that perform better than the best classifier in the combination, we chose boosting, specifically Adaboost, to make the generalised linear model. We constructed a model on each feature space, then iteratively selected one of the models and computed its weights in the linear combination. The goal of the training loop was to create a set of complementary classifiers along with the optimal generalised linear weights. We were inspired by the algorithms proposed in (Schapire et al. 1998; Kégl 2013). Note that since each model uses a different set of features, we could also call this method boosting with column sampling. Figure 6.10 describes the developed method. To implement the ensemble approach, multiple steps were required, as follows;

1. We trained KNN+SFFS on each feature space, the results were two classifiers;

2. We precomputed the leave-one-out class probability predictions, which resulted in two matrices m by K; initialising base classifiers weights to zeros;

3. We initialised multiclass one hot labels, and multiclass weights in a way that rewarded true positives more than false positives. In fact a misclassification on the correct label cost $1/(2.m)$, and equaling simultaneously $K-1$ errors on the negative labels;

4. Boosting iteration T

   4.1. Select the best classifier maximising the weighted multiclass margin (edge) sum $(w_i k * y_i k * p_i k)$, where w is the weight, y is the label, p is the probability and i is the number of samples.
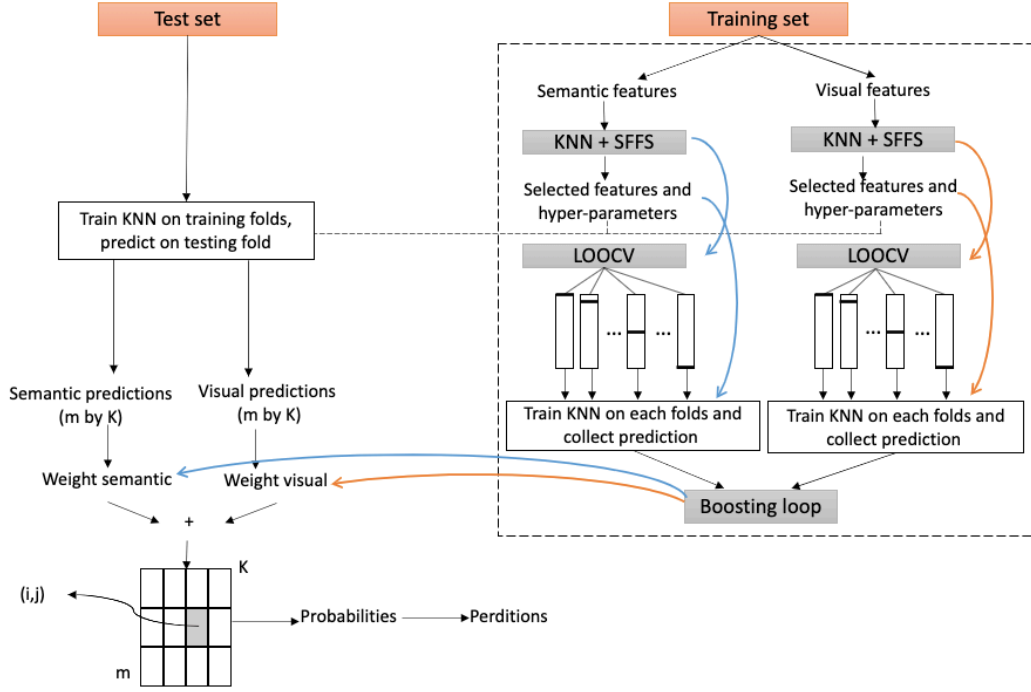
FIGURE 6.10: Framework of fusion using ensemble method.

4.2. Compute the classifier's weight $\alpha_t = \frac{1}{2}log\left(\frac{1+edge}{1-edge}\right)$ and increment the base classifier's weight with α_t.

4.3. Update the sample weight.

4.4. Next iteration, repeat 4.1. until the maximum iteration is achieved or $edge = 1(\alpha = 0)$.

6.5.2.1 Analysis and Evaluation

The cumulative match characteristic (CMC) and receiver operating characteristic (ROC) results are presented as the mean value obtained from a 4-fold cross-validation. Figure 6.11 depicts the CMC and ROC curves obtained from this experiment.

Figure 6.11 demonstrates that the accuracy for same-side recognition was 92% at rank-1 when employing the concatenation method. Furthermore, this accuracy increased to 98% at rank-20. In contrast, the precision of the ensemble method began at 95% at rank-1 and increased to 100% by rank 20. The area under the Roc curve was 0.998 for the ensemble approach and 0.988 for the concatenation approach. Note, the Resnet features and the concatenation method produced nearly identical results.

Figure 6.12 demonstrates that the fusion accuracy reached an equal error rate (EER) of 0.031 and 0.009 using the concatenation and ensemble methods, respectively. Furthermore, the performance result indicates that the ensemble approach exhibited superior accuracy compared to the concatenation approach in the fusion recognition experiment.

(a) CMC Curve

(b) Roc Curve
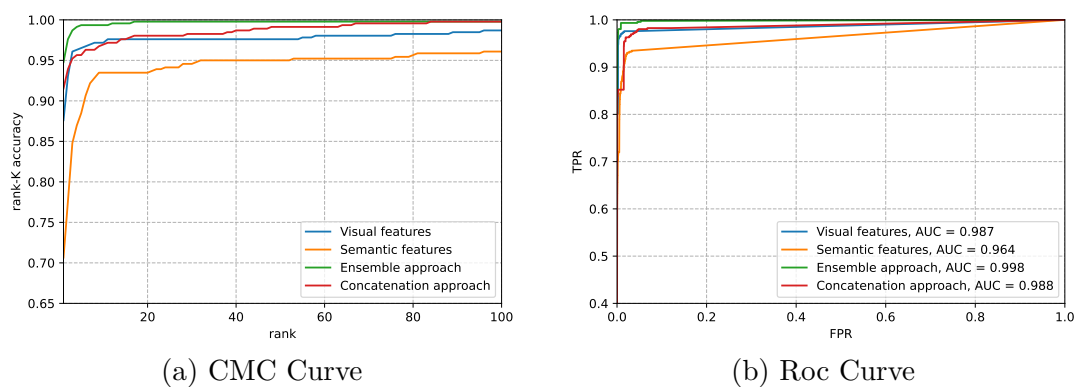
FIGURE 6.11: Recognition via CMC and ROC performance for all approaches: Visual (ResNet features), Semantic (verbal features), Concatenation method and Ensemble method.
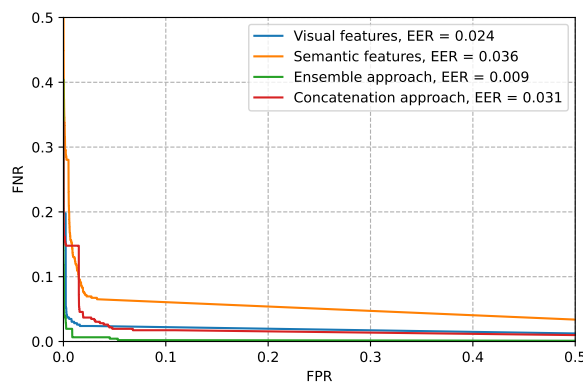
FIGURE 6.12: DET curve for each of the recognition systems.

6.6 Scale to a Larger Dataset

Typically, the use of facial soft biometrics to identify human faces has been investigated utilising datasets with constraints (Reid, Samangooei et al. 2013; Klare, Klum et al. 2014; Tome, Fierrez et al. 2014). However, it has also been tested in real-world identification scenarios with larger populations, large scale data and a diverse set of demographic data, in addition to more challenging visual conditions of surveillance such as high variability in illumination, facial expressions, resolution and pose (Nawaf Yousef Almudhahka et al. 2017a; Nawaf Yousef Almudhahka et al. 2017b). The comparison labels/scores to study comparative facial profile soft biometrics in this thesis, as well as in the existing work on comparative attributes, were collected, while subjects' images were presented to annotators. However, real life scenarios will involve an eyewitness describing a suspect's face from memory, followed by the retrieval of the suspect's identity from a database. The most challenging evaluation is simulating real-world operation, which involves identifying unseen suspects based solely on an eyewitness description.

In contrast to the experiments conducted in Chapter 4, which focused on identifying and verifying individuals in facial profiles soft biometric databases created using labels provided by humans (i.e. crowdsourced comparisons), identification in criminal investigations entails the utilisation of several databases, including mugshots and CCTV images, to ascertain the identity of an unknown suspect based on verbal descriptions provided by eyewitnesses (Davis et al. 2015). As a result, a framework for automatically producing biometric signatures from facial profile photos is required. Chapter 5 introduced a new visual space that reduces the semantic gap between humans and machines in terms of relative facial features. We provide an explanation of two distinct real-life scenarios, as follows:

- In a practical situation, to identify a new facial profile image, we can begin by automatically generating biometric signatures, as explained in Chapter 5. This can be done by employing techniques such as trained RankSVM. Once we obtain the feature vector for each attribute, we can utilise a pre-trained model like Knn to calculate the distances to different classes. If the minimum distances are less than a threshold, the facial profile class is recognised. We can select a threshold defining the maximum distance from any facial profile class and another threshold defining the maximum distance from facial profile space. The acceptability threshold should also vary at low levels of 0, where only images that project very closely to the known face classes are recognised, to result in minimal errors, although many images may be rejected as unknown (Turk et al. 1991). Figure 6.13 shows that the first step in image retrieval and identification is to transform comparative descriptions to relative measures that can be utilised as a biometric signature.
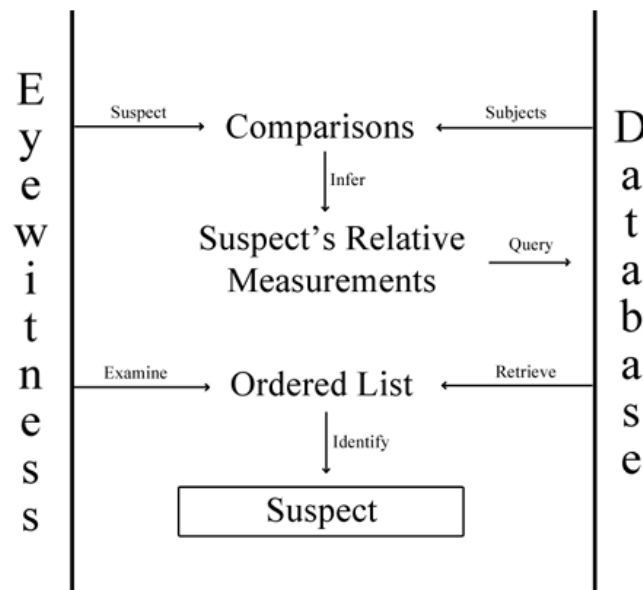

FIGURE 6.13: Verbal identification from soft biometric database (Reid 2013).

- To scale the system up to a larger dataset, our classifier is initially trained on the old portion of the dataset and then the new portion of dataset can be split into the gallery set and the probe set. The gallery set is used to retrain our pre-trainied classifier, while the feature vectors of the probe are utilised to compare with the feature vectors from the gallery view (Chao et al. 2019).

6.7 Conclusion

This chapter introduced a method for facial profile recognition that combines few-shot learning, domain adaptation, and reverse validation techniques. We employed a similarity registration technique to ensure the precise alignment of all facial profile images. By utilising two pre-trained CNN models, ResNet50 and ArcFace, we implemented few-shot learning. Among these models, ResNet50 demonstrated superior performance compared to ArcFace. Specifically, our RL algorithm, which utilised ResNet50, outperformed the traditional training methods discussed in this chapter.

The results obtained from our RL method reveal significant improvements in classification rates. Specifically, the recognition rate for same side (left side) reached 91%, surpassing the state-of-the-art performance achieved with datasets of similar sizes. Additionally, promising outcomes were observed in cross-recognition, with a rate of 82% achieved when right-side images were used in the validation stage. Furthermore, a recognition rate of 75% was attained when left-side images were employed in validation, with right-side images used for testing. Numerical experiments indicated a notable 7% enhancement in cross-recognition accuracy when right-side faces were included in the validation process. Therefore, we can conclude that depending on the application, recognition is viable even when facial profiles are the sole biometric modality available, including scenarios involving bilateral symmetry. Additionally, this chapter examines the effects of integrating soft biometric features with conventional biometric features obtained through few-shot learning in two different approaches. The results indicate that semantic features enhance the performance of traditional biometric systems in both methods.

Chapter 7

Conclusions and Future Work

7.1   Summary and Concluding Remarks

The objective of this thesis is to investigate the process of human identification using comparative facial profiles, specifically focusing on soft biometrics. Towards this objective, a new collection of soft biometric attributes was introduced, and a challenging dataset was used for the identification and recognition experiments.

In Chapter 2 we provided a summary of soft biometrics and their important role in investigations relying on eyewitness testimonies. Furthermore, the literature review presented a comprehensive list of the most significant studies on biometric facial profile recognition and identification. It also highlighted the limitations of the methods and techniques presented in this chapter. Chapter 3 illustrated that using crowdsourcing for comparative soft biometrics leads to precise, relative attributes aiding in subject retrieval. In addition, we presented a range of potential ranking algorithms that have been previously introduced to the research on the subject.

In Chapter 4 we evaluated the proposed semantic human facial profile attributes by employing two ranking systems to generate comparative ratings. An assessment of facial profile analysis using the designated attributes and response evidence identified the potential for considering facial profiles as a soft biometric. Moreover, this thesis makes a significant contribution by introducing the majority vote threshold, and emphasises its influence on accuracy. Chapter 5 emphasised the power of regressing relative attributes and investigated the improvement of biometric systems for facial profile identification by fusing comparative soft biometrics with traditional biometrics.

In Chapter 6 we presented a novel recognition method based on few-shot learning, domain adaptation and reverse validation of facial profile images. We used a few-shot learning based on two pre-trained CNN models, namely ResNet50 and ArcFace. We also examined the bilateral symmetry of facial profiles.

In our study, we found that facial profiles can be used as a biometric modality. In addition, we devised both traditional machine learning and deep learning-based methods for facial profile recognition, using a large dataset that includes numerous subjects. Moreover, this study identified the relative features which are crucial for providing useful information when describing facial profile images. The introduction of a new set of soft biometric features, such as the size of the nostril and the philtrum, has a positive effect on the accuracy of facial profiles recognition. The performance outcomes obtained through the use of semantic facial profile attributes suggest that soft biometric frameworks are highly significant when combined with traditional biometrics. Our system, which fuses soft and traditional biometrics, achieved an accuracy of 98% at rank-1, which further increased to 100% at rank-15.

In short, the analyses of facial profiles through the designated features demonstrated the practical advantages of using facial profiles in soft biometrics. We also studied the correspondence between features extracted with computer vision techniques, and human-derived labels for the improvement of facial profile recognition. This result clearly demonstrates a robust correlation between the variables. The most significant correlation was found between binary-like characteristics (such as gender), followed by nose size and nostril size. Lastly, the majority of eye and ear traits presented minimal but noteworthy correlations.

In conclusion, the results presented in this thesis provide a significant contribution to the existing research on soft biometrics. They demonstrate that utilising comparative facial profile features is effective when handling challenging databases. The findings of the thesis indicate that soft biometrics based on comparative facial profiles can significantly assist in reducing the semantic gap between humans and machines. This enables the retrieval of identities based on semantic descriptions from images database. Finally, the thesis enhances our understanding of how bilateral symmetry might be applied to recognition as a fundamental problem in machine learning.

7.2 Future Work

Our research not only presents novel methods that enhance the recognition of facial profiles, but also highlights numerous areas for further investigation. Some of these possible potential research directions are as follows:

7.2.1 Facial Profile Soft Biometric Identification

Although XM2VTSDB is a challenging dataset that accurately represents the difficult visual conditions of facial profile images, it would be interesting to evaluate the effectiveness of comparative facial profiles as soft biometrics for identification in surveillance

databases that have even more adverse visual circumstances. Further research is needed to understand the elements that impact the performance and completeness of semantic descriptions for relative facial profiles features. Furthermore, in order to enhance the comprehensiveness of the datasets, future research may explore the integration of facial profile attributes with soft biometrics from other domains, including gait, face, ear, and body shape. The main focus of future work should involve the assessment of the level of accuracy and the development of soft profile biometrics for use in novel applications.

7.2.2   The Impact of Occlusion

Extending automatic soft/traditional annotation to footage in uncontrolled contexts can cause occlusion issues. Visual or semantic features can be concealed by the scenery, the subject's body (self occlusion) or covariates such as jewellery, hats and clothing. The presence of these occluded features can impact the labelling of the biometric data, resulting in misleading descriptions. In real scenarios, eyewitness memory is frequently unreliable and can be significantly compromised by stress. This may result in features being described inaccurately. Soft features that are ambiguous or absent can be predicted by taking advantage of the established structure, which improves the description. Missing soft labels can be accurately predicted by understanding the structure between features. For example, if the suspect is of a certain age and has an occluded ear, the suspect's age can be used to predict the size of the ear, considering the fact that there is a significant and positive correlation between age and ear size. Similarly, regarding traditional biometrics, training models may be employed in an attempt to forecast features from occluded facial profiles. This has the potential to enhance the precision of relative measurements automatically computed from facial profile signatures or to optimise search queries in the absence of a comprehensive description.

7.2.3   Additional and Predicted Comparisons

The project has utilised inferred comparisons to address the constraints of the limited data collected by the contributors. Although this has yielded successful results, the inferred process could be removed by collecting more annotations between the subjects images. By gathering more comparisons, there would be no need for inference, therefore the full benefit of comparisons would be observed. Enhanced analysis using soft biometrics may also allow for the derivation of individual labels for recognition using computer vision. The future work should allow for facial profile identification or recognition using computer vision techniques to complement labelled attributes obtained from soft biometrics.

7.2.4 Descriptions from Memory

The comparisons collected for examining the soft biometrics of comparative facial profiles in this thesis, as well as in prior research on comparative traits, were acquired by presenting individuals' photos to annotators. Nevertheless, in practical situations, eyewitnesses may be asked to describe a suspect's facial profile to search a database for their identity. Therefore, it would be interesting to ascertain the precision of facial profiles in remembering the facial characteristics of unknown individuals from memory, in comparison to presented individuals. Additionally, it would be valuable to explore the effectiveness of these comparisons of soft biometrics for the purpose of retrieving identity.

7.2.5 Facial Profile Datasets

There is a limited number of facial profile datasets accessible for analysis and evaluation. Therefore, exploring the possibility of creating more datasets with well-defined ground truths is a highly attractive direction for future research. Creating such datasets for facial profiles analysis involves generating a large number of labelled instances with complex real life circumstances. Additionally, it would be advantageous if the datasets that were created incorporated more challenging modifications such as noise, rotation, translation, etc.

7.2.6 EFIT-V for Facial Profiles

Effective implementation of our research would involve integrating it into established suspect identification systems such as EFIT-V (Wight Constabulary 2024). EFIT-V uses computer software to generate a composite image of the facial characteristics of a suspect as stated by an eyewitness (Figure 7.1). Presently, these systems employ a random selection process to choose initial variables from a large parameter space and then refine them based on user feedback to achieve optimal combinations. Using such techniques to generate comparative facial profile soft biometrics for a suspect is highly effective.

7.2.7 Bilateral Symmetry

Future work will concentrate on annotating the opposite side of the face utilising crowd-sourcing technologies. Following that, it will investigate bilateral symmetry using soft biometric traits. Furthermore, we will be able to combine bilateral symmetry using soft biometric traits with traditional bilateral symmetry and study the effect of adding the soft biometric trait on recognition.
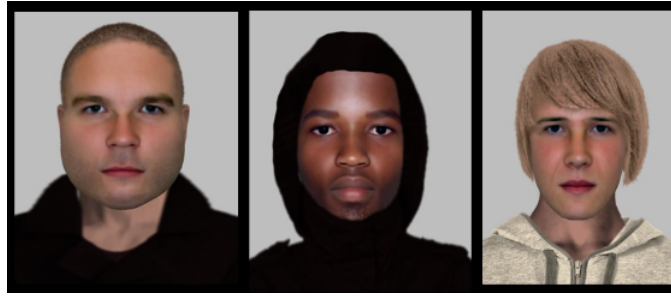
FIGURE 7.1: EFIT images released as part of a burglary investigation in Southampton (Wight Constabulary 2024).

7.2.8 Deep Learning

Future work will also consider the use of an end-to-end deep learning model for higher retrieval accuracy, as well as the scalability of automatic biometric signatures. It would be interesting if future studies could examine the power of regressing relative attributes with a deep learning CNN. Additionally, it would be worth investigating different CNN models as an alternative to the pre-trained ResNet50 and ArcFace models.

References

Abdelwhab, Abdelgader and Serestina Viriri (2018). 'A survey on soft biometrics for human identification'. In: *Machine Learning and Biometrics*, p. 37.

Aibara, Tsunehiro, Kenji Ohue and Yasushi Matsuoka (1991). 'Human face recognition by P-type Fourier descriptor'. In: *Visual Communications and Image Processing'91: Image Processing*. Vol. 1606. SPIE, pp. 198–203.

Aibara, Tsunehiro, Kenji Ohue and Yoshiaki Oshita (1993). 'Human face profile recognition by a P-Fourier descriptor'. In: *Optical Engineering* 32.4, pp. 861–863.

Almudhahka, Nawaf Y, Mark S Nixon and Jonathon S Hare (2016). 'Unconstrained human identification using comparative facial soft biometrics'. In: *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, pp. 1–6.

Almudhahka, Nawaf Yousef, Mark S Nixon and Jonathon S Hare (2017a). 'Automatic semantic face recognition'. In: *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, pp. 180–185.

– (2017b). 'Semantic face signatures: Recognizing and retrieving faces by verbal descriptions'. In: *IEEE Transactions on Information Forensics and Security* 13.3, pp. 706–716.

Almudhahka, Nawaf, Mark Nixon and Jonathon Hare (2016). 'Human face identification via comparative soft biometrics'. In: *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*. IEEE, pp. 1–6.

Alqaralleh, Esraa, Ayman Afaneh and Önsen Toygar (2023). 'Masked face recognition using frontal and profile faces with multiple fusion levels'. In: *Signal, Image and Video Processing* 17.4, pp. 1375–1382.

Amaral, Vagner do and Carlos Eduardo Thomaz (2008). 'Normalizaçao espacial de imagens frontais de face'. In:

Ao, Meng, Dong Yi, Zhen Lei and Stan Z Li (2009). 'Face recognition at a distance: system issues'. In: *Handbook of Remote Biometrics: For Surveillance and Security*, pp. 155–167.

Arasu, Arvind, Jasmine Novak, Andrew Tomkins and John Tomlin (2002). 'PageRank computation and the structure of the web: Experiments and algorithms'. In: *Proceedings of the Eleventh International World Wide Web Conference, Poster Track*, pp. 107–117.

Beveridge, J Ross, P Jonathon Phillips, David S Bolme, Bruce A Draper, Geof H Givens, Yui Man Lui, Mohammad Nayeem Teli, Hao Zhang, W Todd Scruggs, Kevin W Bowyer et al. (2013). 'The challenge of face recognition from digital point-and-shoot cameras'. In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, pp. 1–8.

Bhanu, Bir and Xiaoli Zhou (2004). 'Face recognition from face profile using dynamic time warping'. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* Vol. 4. IEEE, pp. 499–502.

Bourlai, Thirimachos, ed. (2016). *Face recognition across the imaging spectrum.* Switzerland: Springer.

Bukar, Ali Maina and Hassan Ugail (2017). 'Automatic age estimation from facial profile view'. In: *IET Computer Vision* 11.8, pp. 650–655.

Cao, Qiong, Li Shen, Weidi Xie, Omkar M Parkhi and Andrew Zisserman (2018). 'Vggface2: A dataset for recognising faces across pose and age'. In: *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, pp. 67–74.

Center, Appen Success (2019a). *How to Calculate a Confidence Score.* https://success.appen.com/hc/en-us/articles/201855939-How-to-Calculate-a-Confidence-Score. [Online; accessed 14-February-2020].

– (2019b). *Test Question Enabling and Quiz Mode.* https://success.appen.com/hc/en-us/articles/211363686-Test-Question-Enabling-and-Quiz-Mode. [Online; accessed 20-October-2019].

Chao, Hanqing, Yiwei He, Junping Zhang and Jianfeng Feng (2019). 'Gaitset: Regarding gait as a set for cross-view gait recognition'. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01, pp. 8126–8133.

Chen, Tianqi and Carlos Guestrin (2016). 'Xgboost: A scalable tree boosting system'. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Chen, Xi, Paul N Bennett, Kevyn Collins-Thompson and Eric Horvitz (2013). 'Pairwise ranking aggregation in a crowdsourced setting'. In: *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 193–202.

Cormen, Thomas H, Charles E Leiserson, Ronald L Rivest and Clifford Stein (2001). 'Introduction to algorithms second edition'. In: *The Knuth-Morris-Pratt Algorithm*.

Dalal, Navneet and Bill Triggs (2005). 'Histograms of oriented gradients for human detection'. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE, pp. 886–893.

Dantcheva, Antitza, Carmelo Velardo, Angela D'angelo and Jean-Luc Dugelay (2011). 'Bag of soft biometrics for person identification'. In: *Multimedia Tools and Applications* 51.2, pp. 739–777.

David, Shai Ben, Tyler Lu, Teresa Luu and Dávid Pál (2010). 'Impossibility theorems for domain adaptation'. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, pp. 129–136.

Davis, Josh P and Tim Valentine (2015). 'Human verification of identity from photographic images'. In: *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV*, pp. 211–238.

DeCann, Brian and Arun Ross (2013). 'Relating roc and cmc curves via the biometric menagerie'. In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, pp. 1–8.

Deng, Jiankang, Jia Guo, Niannan Xue and Stefanos Zafeiriou (2019). 'Arcface: Additive angular margin loss for deep face recognition'. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690–4699.

Ding, Sihao, Qiang Zhai, Yuan F Zheng and Dong Xuan (2013). 'Side-view face authentication based on wavelet and random forest with subsets'. In: *2013 IEEE International Conference on Intelligence and Security Informatics*. IEEE, pp. 76–81.

Everett, Adam (17-4-2019). *Warrington Guardian-New CCTV images after man stabbed in face in Block 1 club*. https://www.warringtonguardian.co.uk/news/17580449.new-cctv-images-man-stabbed-face-block-1-. [Online; accessed 30-July-2020].

Farahani, Abolfazl, Sahar Voghoei, Khaled Rasheed and Hamid R Arabnia (2021). 'A brief review of domain adaptation'. In: *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pp. 877–894.

Flynn, Patrick J, Kevin W Bowyer and P Jonathon Phillips (2003). 'Assessment of time dependency in face recognition: An initial study'. In: *Audio-and Video-Based Biometric*

Person Authentication: 4th International Conference, AVBPA 2003 Guildford, UK, June 9–11, 2003 Proceedings 4. Springer, pp. 44–51.

Fogel, Fajwel, Alexandre d'Aspremont and Milan Vojnovic (2014). 'Serialrank: Spectral ranking using seriation'. In: *Advances in Neural Information Processing Systems* 27, pp. 900–908.

Freund, Yoav, Raj Iyer, Robert E Schapire and Yoram Singer (2003). 'An efficient boosting algorithm for combining preferences'. In: *Journal of machine learning research* 4.Nov, pp. 933–969.

Fu, Yanwei, Timothy M Hospedales, Tao Xiang, Jiechao Xiong, Shaogang Gong, Yizhou Wang and Yuan Yao (2015). 'Robust subjective visual property prediction from crowdsourced pairwise labels'. In: *IEEE transactions on pattern analysis and machine intelligence* 38.3, pp. 563–577.

Galton, Francis (1889). 'Personal identification and description'. In: *Journal of anthropological institute of Great Britain and Ireland*, pp. 177–191.

Ganin, Yaroslav, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand and Victor Lempitsky (2016). 'Domain-adversarial training of neural networks'. In: *The journal of machine learning research* 17.1, pp. 2096–2030.

Gao, Yongsheng and Maylor KH Leung (2002). 'Human face profile recognition using attributed string'. In: *Pattern Recognition* 35.2, pp. 353–360.

Garcia, Victor and Joan Bruna (2017). 'Few-shot learning with graph neural networks'. In: *arXiv preprint arXiv:1711.04043*.

Ge, Shiming, Jia Li, Qiting Ye and Zhao Luo (2017). 'Detecting masked faces in the wild with lle-cnns'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2682–2690.

Ghaffary, Keramat Allah, Fardin Akhlaghian Tab and Habibollah Danyali (2011). 'Profile-based face recognition using the outline curve of the profile Silhouette'. In: *IJCA special issue on: Artifical Intelligence Techniques-Novel Approaches & Practical Applications" AIT*.

Ghalleb, A El Kissi and N Essoukri Ben Amara (2016). 'Soft and hard biometrics for the authentication of remote people in front and side views'. In: *International Journal of Applied Engineering Research* 11.14, pp. 8120–8127.

Glickman, Mark E (1995). 'A comprehensive guide to chess ratings'. In: *American Chess Journal* 3.1, pp. 59–102.

Gonzalez-Sosa, Ester, Julian Fierrez, Ruben Vera-Rodriguez and Fernando Alonso-Fernandez (2018). 'Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation'. In: *IEEE Transactions on Information Forensics and Security* 13.8, pp. 2001–2014.

Graham, Daniel B (1998). 'Face recognition: From theory to applications'. In: *NATO ASI Series F, Computer and Systems Sciences* 163, pp. 446–456.

Guo, Baofeng and Mark S Nixon (2008). 'Gait feature subset selection by mutual information'. In: *IEEE Transactions on Systems, MAN, and Cybernetics-part a: Systems and Humans* 39.1, pp. 36–46.

Guyon, Isabelle, Jason Weston, Stephen Barnhill and Vladimir Vapnik (2002). 'Gene selection for cancer classification using support vector machines'. In: *Machine learning* 46.1, pp. 389–422.

Hare, Jonathon S, Paul H Lewis, Peter GB Enser and Christine J Sandom (2006). 'Mind the Gap: Another look at the problem of the semantic gap in image retrieval'. In: *Multimedia Content Analysis, Management, and Retrieval 2006*. Vol. 6073. International Society for Optics and Photonics, p. 607309.

Harmon, LD, MK Khan, Richard Lasch and PF Ramig (1981). 'Machine identification of human faces'. In: *Pattern Recognition* 13.2, pp. 97–110.

Harmon, Leon and Willard Hunt (1977). 'Automatic recognition of human face profiles'. In: *Computer Graphics and Image Processing* 6.2, pp. 135–156.

Harmon, Leon, SC Kuo, PF Ramig and U Raudkivi (1978). 'Identification of human face profiles by computer'. In: *Pattern Recognition* 10.5-6, pp. 301–312.

Harris, Sarah and David Harris (2015). *Digital design and computer architecture*. Morgan Kaufmann.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun (2016). 'Deep residual learning for image recognition'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

He, Yihui, Chenchen Zhu, Jianren Wang, Marios Savvides and Xiangyu Zhang (2019). 'Bounding box regression with uncertainty for accurate object detection'. In: *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, pp. 2888–2897.

Heckathorn, Douglas D, Robert S Broadhead and Boris Sergeyev (2001). 'A methodology for reducing respondent duplication and impersonation in samples of hidden populations'. In: *Journal of Drug Issues* 31.2, pp. 543–564.

Herbrich, Ralf, Tom Minka and Thore Graepel (2006). 'TrueSkill™: a Bayesian skill rating system'. In: *Proceedings of the 19th international conference on neural information processing systems*, pp. 569–576.

Horn, Berthold KP, Hugh M Hilden and Shahriar Negahdaripour (1988). 'Closed-form solution of absolute orientation using orthonormal matrices'. In: *JOSA A* 5.7, pp. 1127–1135.

Huang, Gary B, Marwan Mattar, Tamara Berg and Eric Learned-Miller (2008). 'Labeled faces in the wild: A database forstudying face recognition in unconstrained environments'. In: *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.*

Huber, Peter J et al. (1963). 'Pairwise comparison and ranking: optimum properties of the row sum procedure'. In: *Annals of Mathematical Statistics* 34.2, pp. 511–520.

Hunter, David R et al. (2004). 'MM algorithms for generalized Bradley-Terry models'. In: *The annals of statistics* 32.1, pp. 384–406.

Jaha, Emad Sami and Mark S Nixon (2014). 'Soft biometrics for subject identification using clothing attributes'. In: *IEEE international joint conference on biometrics*. IEEE, pp. 1–6.

– (2015). 'Viewpoint invariant subject retrieval via soft clothing biometrics'. In: *2015 International Conference on Biometrics (ICB)*. IEEE, pp. 73–78.

Jain, Anil K, Sarat C Dass and Karthik Nandakumar (2004). 'Can soft biometric traits assist user recognition?' In: *Biometric technology for human identification*. Vol. 5404. International Society for Optics and Photonics, pp. 561–572.

Jia, Xiaoguang and Mark S Nixon (1994). 'Analysing front view face profiles for face recognition via the Walsh transform'. In: *Pattern recognition letters* 15.6, pp. 551–558.

Joachims, Thorsten (2002). 'Optimizing search engines using clickthrough data'. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 133–142.

Jolliffe, Ian T (2002). *Principal component analysis for special types of data*. Springer.

Kaufman, Gerald J and Kenneth J Breeding (1976). 'The automatic recognition of human faces from profile silhouettes'. In: *IEEE Transactions on systems, Man, and Cybernetics* 2, pp. 113–121.

Kazemi, Vahid and Josephine Sullivan (2014). 'One millisecond face alignment with an ensemble of regression trees'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1867–1874.

Kégl, Balázs (2013). 'The return of AdaBoost. MH: multi-class Hamming trees'. In: *arXiv preprint arXiv:1312.6086*.

Kemelmacher-Shlizerman, Ira, Steven M Seitz, Daniel Miller and Evan Brossard (2016). 'The megaface benchmark: 1 million faces for recognition at scale'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4873–4882.

Klare, Brendan F, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah and Anil K Jain (2015). 'Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1931–1939.

Klare, Brendan F, Scott Klum, Joshua C Klontz, Emma Taborsky, Tayfun Akgul and Anil K Jain (2014). 'Suspect identification based on descriptive facial attributes'. In: *IEEE International Joint Conference on Biometrics*. IEEE, pp. 1–8.

Kowalski, Marek, Jacek Naruniec and Tomasz Trzcinski (2017). 'Deep alignment network: A convolutional neural network for robust face alignment'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 88–97.

Kumar, Neeraj, Alexander C Berg, Peter N Belhumeur and Shree K Nayar (2009). 'Attribute and simile classifiers for face verification'. In: *2009 IEEE 12th international conference on computer vision*. IEEE, pp. 365–372.

Kwon, Young H and Niels da Vitoria Lobo (1999). 'Age classification from facial images'. In: *Computer vision and image understanding* 74.1, pp. 1–21.

Li, Peipei, Xiang Wu, Yibo Hu, Ran He and Zhenan Sun (2019). 'M2FPA: A multi-yaw multi-pitch high-quality dataset and benchmark for facial pose analysis'. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10043–10051.

Lipoščak, Zdravko and Sven Loncaric (1999). 'A scale-space approach to face recognition from profiles'. In: *International Conference on Computer Analysis of Images and Patterns*. Springer, pp. 243–250.

Liu, Ziwei, Ping Luo, Xiaogang Wang and Xiaoou Tang (Dec. 2015). 'Deep Learning Face Attributes in the Wild'. In: *Proceedings of International Conference on Computer Vision (ICCV)*.

Martinho-Corbishley, Daniel (2018). 'Discovering human descriptions for ubiquitous visual identification'. PhD thesis. University of Southampton.

Martinho-Corbishley, Daniel, Mark S Nixon and John N Carter (2015). 'Soft biometric recognition from comparative crowdsourced annotations'. In: *6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15)*. IET, pp. 1–6.

Martinho-Corbishley, Daniel, Mark S Nixon and John N Carter (2016). 'Retrieving relative soft biometrics for semantic identification'. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, pp. 3067–3072.

Martinson, Eric, Wallace Lawson and J Gregory Trafton (2013). 'Identifying people with soft-biometrics at fleet week'. In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 49–56.

Masi, Iacopo, Stephen Rawls, Gérard Medioni and Prem Natarajan (2016). 'Pose-aware face recognition in the wild'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4838–4846.

Meng, Qiang, Shichao Zhao, Zhida Huang and Feng Zhou (2021). 'Magface: A universal representation for face recognition and quality assessment'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14225–14234.

Messer, Kieron, Jiri Matas, Josef Kittler, Juergen Luettin and Gilbert Maitre (1999). 'XM2VTSDB: The extended M2VTS database'. In: *Second international conference on audio and video-based biometric person authentication*. Vol. 964. Citeseer, pp. 965–966.

Milborrow, S., J. Morkel and F. Nicolls (2010). 'The MUCT Landmarked Face Database'. In: *Pattern Recognition Association of South Africa*. http://www.milbo.org/muct.

Milborrow, Stephen, John Morkel and Fred Nicolls (2010). 'The MUCT landmarked face database'. In: *Pattern recognition association of South Africa* 201.0, p. 535.

Montero, David, Marcos Nieto, Peter Leskovsky and Naiara Aginako (2022). 'Boosting masked face recognition with multi-task arcface'. In: *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, pp. 184–189.

Morvant, Emilie, Amaury Habrard and Stéphane Ayache (2011). 'Sparse domain adaptation in projection spaces based on good similarity functions'. In: *2011 IEEE 11th International Conference on Data Mining*. IEEE, pp. 457–466.

Negahban, Sahand, Sewoong Oh and Devavrat Shah (2012). 'Iterative ranking from pairwise comparisons'. In: *Advances in neural information processing systems* 25, pp. 2474–2482.

Nixon, Mark S, Paulo L Correia, Kamal Nasrollahi, Thomas B Moeslund, Abdenour Hadid and Massimo Tistarelli (2015). 'On soft biometrics'. In: *Pattern Recognition Letters* 68, pp. 218–230.

Nixon, Mark S, Bingchen H Guo, Sarah V Stevenage, Emad S Jaha, Nawaf Almudhahka and Daniel Martinho-Corbishley (2017). 'Towards automated eyewitness descriptions: describing the face, body and clothing for recognition'. In: *Visual Cognition* 25.4-6, pp. 524–538.

Oliva, Aude and Antonio Torralba (2001). 'Modeling the shape of the scene: A holistic representation of the spatial envelope'. In: *International journal of computer vision* 42.3, pp. 145–175.

Pan, Gang, Lei Zheng and Zhaohui Wu (2005). 'Robust metric and alignment for profile-based face recognition: An experimental comparison'. In: *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1*. Vol. 1. IEEE, pp. 117–122.

Parikh, Devi and Kristen Grauman (2011). 'Relative attributes'. In: *2011 International Conference on Computer Vision*. IEEE, pp. 503–510.

Parikh, Devi, Adriana Kovashka, Amar Parkash and Kristen Grauman (2012). 'Relative attributes for enhanced human-machine communication'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 26. 1, pp. 2153–2159.

Park, Unsang and Anil K Jain (2010). 'Face matching and retrieval using soft biometrics'. In: *IEEE Transactions on Information Forensics and Security* 5.3, pp. 406–415.

Parkhi, Omkar M, Andrea Vedaldi and Andrew Zisserman (2015). 'Deep face recognition'. In:

Parnami, Archit and Minwoo Lee (2022). 'Learning from few examples: A summary of approaches to few-shot learning'. In: *arXiv preprint arXiv:2203.04291*.

Phillips, P Jonathon, Hyeonjoon Moon, Syed A Rizvi and Patrick J Rauss (2000). 'The FERET evaluation methodology for face-recognition algorithms'. In: *IEEE Transactions on pattern analysis and machine intelligence* 22.10, pp. 1090–1104.

Rajesh, KM and M Naveenkumar (2016). 'A robust method for face recognition and face emotion detection system using support vector machines'. In: *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*. IEEE, pp. 1–5.

Ramos, Jose (23/02/2024). *Man 'follows' woman in Ikea and Westquay in Southampton*. https://www.dailyecho.co.uk/news/24139842.man-follows-woman-ikea-westquay-southampton/. [Online; accessed 24-February-2024].

Rathore, Rajeev, Surya Prakash and Phalguni Gupta (2013). 'Efficient human recognition system using ear and profile face'. In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, pp. 1–6.

Reid, Daniel A (2013). 'Human Identification using Soft Biometrics'. PhD thesis. University of Southampton.

Reid, Daniel A and Mark S Nixon (2010). 'Imputing human descriptions in semantic biometrics'. In: *Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence*, pp. 25–30.

– (2011). 'Using comparative human descriptions for soft biometrics'. In: *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, pp. 1–6.

Reid, Daniel A, Mark S Nixon and Sarah V Stevenage (2013). 'Soft biometrics; human identification using comparative descriptions'. In: *IEEE Transactions on pattern analysis and machine intelligence* 36.6, pp. 1216–1228.

Reid, Daniel A, Sina Samangooei, Cunjian Chen, Mark S Nixon and Arun Ross (2013). 'Soft biometrics for surveillance: an overview'. In: *Handbook of statistics*. Vol. 31. Elsevier, pp. 327–352.

Ren, Shaoqing, Xudong Cao, Yichen Wei and Jian Sun (2014). 'Face alignment at 3000 fps via regressing local binary features'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1685–1692.

Romeo, Justin and Thirimachos Bourlai (2019). 'Semi-Automatic Geometric Normalization of Profile Faces'. In: *2019 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, pp. 121–125.

Samangooei, Sina, Baofeng Guo and Mark S Nixon (2008). 'The use of semantic human description as a soft biometric'. In: *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*. IEEE, pp. 1–7.

Sarangi, Partha Pratim, BS P Mishra and Sachidanada Dehuri (2018). 'Multimodal biometric recognition using human ear and profile face'. In: *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*. IEEE, pp. 1–6.

Sarangi, Partha Pratim, Madhumita Panda, Subhashree Mishra and Bhabani Shankar Prasad Mishra (2022). 'Multimodal biometric recognition using human ear and profile face: An improved approach'. In: *Machine Learning for Biometrics*. Elsevier, pp. 47–63.

Schapire, Robert E and Yoram Singer (1998). 'Improved boosting algorithms using confidence-rated predictions'. In: *Proceedings of the eleventh annual conference on Computational learning theory*, pp. 80–91.

Sengupta, Soumyadip, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa and David W Jacobs (2016). 'Frontal to profile face verification in the wild'.

In: *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, pp. 1–9.

Sharma, Prag and Richard B Reilly (2003). 'A colour face image database for benchmarking of automatic face detection algorithms'. In: *Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE Cat. No. 03EX667)*. Vol. 1. IEEE, pp. 423–428.

Shirbani, F and H Soltanian Zadeh (2013). 'Fast SFFS-based algorithm for feature selection in biomedical datasets'. In: *AUT Journal of Electrical Engineering* 45.2, pp. 43–56.

Sim, Terence, Simon Baker and Maan Bsat (2001). 'The CMU Pose, Illumination and Expression database of human faces'. In: *Carnegie Mellon University Technical Report CMU-RI-TR-OI-02*.

Singh, Sukhdeep and Sunil Kumar Singla (2013). 'A review on biometrics and ear recognition techniques'. In: *International Journal of Advanced Research in Computer Science and Software Engineering* 3.

Smeulders, Arnold WM, Marcel Worring, Simone Santini, Amarnath Gupta and Ramesh Jain (2000). 'Content-based image retrieval at the end of the early years'. In: *IEEE Transactions on pattern analysis and machine intelligence* 22.12, pp. 1349–1380.

Somaie, AA and SS Ipson (1994a). 'From Modelling to Human Face Profile Identification'. In: *AMSE Conference*. Vol. 1, pp. 81–91.

– (1994b). 'Recognition of Human Face Profiles Using a Phase Fourier Descriptor in a 1-D System'. In: *AMSE Conference*. Vol. 1, pp. 81–91.

– (1995). 'A human face profile identification system using 1-D real Fourier descriptors'. In: *International journal of infrared and millimeter waves* 16.8, pp. 93–104.

Sporer, Siegfried Ludwig (1992). 'An archival analysis of person descriptions'. In: *Biennial Meeting of the American Psychology-Law Society in San Diego, California*.

Tan, R, V Osman and G Tan (1997). 'Ear size as a predictor of chronological age'. In: *Archives of gerontology and geriatrics* 25.2, pp. 187–191.

Tariq, Usman, Yuxiao Hu and Thomas S Huang (2009). 'Gender and ethnicity identification from silhouetted face profiles'. In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 2441–2444.

Tarrés, F (2012). 'GTAV face database'. In: *http: gps-tsc. upc. es/GTAV/ResearchAreas/GTAVDatabase. htm*.

Tome, Pedro, Julian Fierrez, Ruben Vera-Rodriguez and Mark S Nixon (2014). 'Soft biometrics and their application in person recognition at a distance'. In: *IEEE Transactions on information forensics and security* 9.3, pp. 464–475.

Tome, Pedro, Ruben Vera-Rodriguez, Julian Fierrez and Javier Ortega-Garcia (2015). 'Facial soft biometric features for forensic face recognition'. In: *Forensic science international* 257, pp. 271–284.

Toygar, Önsen, Esraa Alqaralleh and Ayman Afaneh (2018). 'Symmetric ear and profile face fusion for identical twins and non-twins recognition'. In: *Signal, Image and Video Processing* 12, pp. 1157–1164.

Turk, Matthew and Alex Pentland (1991). 'Eigenfaces for recognition'. In: *Journal of cognitive neuroscience* 3.1, pp. 71–86.

Tzimiropoulos, Georgios and Maja Pantic (2013). 'Optimization problems for fast aam fitting in-the-wild'. In: *Proceedings of the IEEE international conference on computer vision*, pp. 593–600.

Vieira, Tiago F, Andrea Bottino, Aldo Laurentini and Matteo De Simone (2014). 'Detecting siblings in image pairs'. In: *The Visual Computer* 30, pp. 1333–1345.

Walker, Michael W, Lejun Shao and Richard A Volz (1991). 'Estimating 3-D location parameters using dual number quaternions'. In: *CVGIP: image understanding* 54.3, pp. 358–367.

Wallhoff, Frank, Stefan Muller and Gerhard Rigoll (2001). 'Recognition of face profiles from the MUGSHOT database using a hybrid connectionist/HMM approach'. In: *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*. Vol. 3. IEEE, pp. 1489–1492.

Wallhoff, Frank and Gerhard Rigoll (2001). 'A novel hybrid face profile recognition system using the FERET and MUGSHOT databases'. In: *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*. Vol. 1. IEEE, pp. 1014–1017.

Wang, Gang, David Forsyth and Derek Hoiem (2010). 'Comparative object similarity for improved recognition with few or no examples'. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 3525–3532.

Wang, Yaqing, Quanming Yao, James T Kwok and Lionel M Ni (2020). 'Generalizing from a few examples: A survey on few-shot learning'. In: *ACM computing surveys (csur)* 53.3, pp. 1–34.

Watson, Craig I. (2008-10-16 14:10:54 2008). 'NIST Special Database 18. NIST Mugshot Identification Database (MID)'. en. In:

Wauthier, Fabian, Michael Jordan and Nebojsa Jojic (2013). 'Efficient ranking from pairwise comparisons'. In: *International Conference on Machine Learning*. PMLR, pp. 109–117.

Welinder, Peter and Pietro Perona (2010). 'Online crowdsourcing: rating annotators and obtaining cost-effective labels'. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, pp. 25–32.

Whitelam, Cameron, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen et al. (2017). 'Iarpa janus benchmark-b face dataset'. In: *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 90–98.

Wight Constabulary, Hampshire & Isle of (2024). *E-fit images released in Southampton burglary investigation*. https://www.hampshire.police.uk/news/hampshire/news/appeals/2024/january/e-fit-images-released-in-southampton-burglary-investigation/. [Online; accessed 22-November-2023].

Wu, Chyuan Jy and Jun S Huang (1990). 'Human face profile recognition by computer'. In: *Pattern recognition* 23.3-4, pp. 255–259.

Xie, Bei, Jiaohua Qin, Xuyu Xiang, Hao Li and Lili Pan (2018). 'An Image Retrieval Algorithm Based on Gist and Sift Features.' In: *Int. J. Netw. Secur.* 20.4, pp. 609–616.

Xu, Xiaona and Zhichun Mu (2007). 'Feature fusion method based on KCCA for ear and profile face based multimodal recognition'. In: *2007 IEEE international conference on automation and logistics*. IEEE, pp. 620–623.

Yaman, Dogucan, Fevziye Irem Eyiokur and Hazım Kemal Ekenel (2022). 'Multimodal soft biometrics: combining ear and face biometrics for age and gender classification'. In: *Multimedia Tools and Applications* 81.16, pp. 22695–22713.

Yaman, Dogucan, Fevziye Irem Eyiokur, Nurdan Sezgin and Hazim Kemal Ekenel (2018). 'Age and gender classification from ear images'. In: *2018 International Workshop on Biometrics and Forensics (IWBF)*. IEEE, pp. 1–7.

Yaman, Dogucan, Fevziye Irem Eyiokur and Hazim Kemal Ekenel (2019). 'Multimodal age and gender classification using ear and profile face images'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0.

Yan, Ping and Kevin W Bowyer (2007). 'Biometric recognition using 3D ear shape'. In: *IEEE Transactions on pattern analysis and machine intelligence* 29.8, pp. 1297–1308.

Yin, Yu, Songyao Jiang, Joseph P Robinson and Yun Fu (2020). 'Dual-attention GAN for large-pose face frontalization'. In: *2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020)*. IEEE, pp. 249–256.

Yoshida, Hiro (2010). *Advanced computational intelligence paradigms in healthcare 5: intelligent decision support systems*. Vol. 5. Springer Science & Business Media.

Yun, Hyokun, Parameswaran Raman and S Vishwanathan (2014). 'Ranking via robust binary classification'. In: *Advances in Neural Information Processing Systems* 27, pp. 2582–2590.

Zeng, Huaien and Qinglin Yi (2011). 'Quaternion-Based Iterative Solution of Three-Dimensional Coordinate Transformation Problem.' In: *J. Comput.* 6.7, pp. 1361–1368.

Zhang, Xiaozheng, Yongsheng Gao and Maylor KH Leung (2008). 'Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases'. In: *IEEE Transactions on Information Forensics and Security* 3.4, pp. 684–697.

Zhao, Jian, Yu Cheng, Yan Xu, Lin Xiong, Jianshu Li, Fang Zhao, Karlekar Jayashree, Sugiri Pranata, Shengmei Shen, Junliang Xing et al. (2018). 'Towards pose invariant face recognition in the wild'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2207–2216.

Zhou, Xiaoli and Bir Bhanu (2005). 'Human recognition based on face profiles in video'. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*. IEEE, pp. 15–15.

– (2007). 'Integrating face and gait for human recognition at a distance in video'. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 37.5, pp. 1119–1137.

Zhu, Xiangyu, Zhen Lei, Xiaoming Liu, Hailin Shi and Stan Z Li (2016). 'Face alignment across large poses: A 3d solution'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 146–155.