Chapter 2
Comparative Face Soft Biometrics for Human Identification

Nawaf Yousef Almudhahka, Mark S. Nixon and Jonathon S. Hare

Abstract The recent growth in CCTV systems and the challenges of automatically identifying humans under the adverse visual conditions of surveillance have increased the interest in soft biometrics, which are physical attributes that can be used to describe people semantically. Soft biometrics enable human identification based on verbal descriptions, and they can be captured in conditions where it is impossible to acquire traditional biometrics such as iris and fingerprint. The research on facial soft biometrics has tended to focus on identification using categorical attributes, whereas comparative attributes have shown a better accuracy. Nevertheless, the research in comparative facial soft biometrics has been limited to small constrained databases, while identification in surveillance systems involves unconstrained large databases. In this chapter, we explore human identification through comparative facial soft biometrics in large unconstrained databases using the Labelled Faces in the Wild (LFW) database. We propose a novel set of attributes and investigate their significance. Also, we analyse the reliability of comparative facial soft biometrics for realistic databases and explore identification and verification using comparative facial soft biometrics. The results of the performance analysis show that by comparing an unknown subject to a line up of ten subjects only; a correct match will be found in the top 2.08% retrieved subjects from a database of 4038 subjects.

Face Biometrics and Semantic Face Recognition

The crucial role of surveillance systems in crime prevention and public safety has motivated massive deployments of CCTV networks around the world [1–4].

N.Y. Almudhahka (✉) · M.S. Nixon · J.S. Hare
University of Southampton, Southampton, UK
e-mail: nya1g14@ecs.soton.ac.uk

M.S. Nixon
e-mail: msn@ecs.soton.ac.uk

J.S. Hare
e-mail: jsh2@ecs.soton.ac.uk

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_2

For instance, the number of CCTV cameras deployed in cities and town centres in the UK was estimated between 1.85 [5] and 5.9 million [6]. This expansion in the usage of surveillance systems has increased the reliance on CCTV imagery for suspect identification, which is the first challenge faced by law enforcement agencies in criminal investigations [3]. Thus, the need for identifying suspects from imagery databases (i.e. mugshots or CCTV footage) has motivated research in human identification using semantic descriptions based on eyewitnesses' statements with a view to enabling searching a database of subjects through verbal descriptions [7–9]. These semantic descriptions are based on soft biometrics, which refer to physical and behavioural attributes that can be used to identify people.

The main advantage of soft biometrics as compared with traditional hard biometrics (e.g. iris, DNA, and fingerprint) is that they can be acquired at a distance without individuals' involvement. In addition, soft biometrics enjoy sufficient robustness to the challenging visual conditions of surveillance such as occlusion of features, viewpoint variance, low resolution, and changes in illumination [7, 9, 10]. Therefore, soft biometrics can play a significant role in criminal investigations, where it is required to retrieve the identity of a suspect from an imagery database using a verbal description (i.e. eyewitness statement). Furthermore, soft biometrics bridge the semantic gap resulted from the difference between machines and humans in estimating attributes, which enables retrieval from a database solely by verbal descriptions as illustrated in Fig. 2.1.

Soft biometrics can be categorized according to their group and format. In terms of group, soft biometrics may fall under global, facial, body, or clothing attributes. While in terms of format, soft biometrics can be classified as: categorical, where individual attributes are assigned to specific classes (e.g. *square* versus *round* jaw); or comparative, where attributes of an individual are estimated relative to another individual (e.g. subject *A* has a *more rounded* jaw than subject *B*) [12]. Comparative soft biometrics are discussed further in more details in the next section. Figure 2.2 shows example categorical soft biometric attributes from the four different groups of soft biometrics: facial, body, clothing, and global. Figure 2.3 presents example soft biometric attributes in the comparative format. More highlights on the different soft biometric attribute formats are provided in Table 2.1.

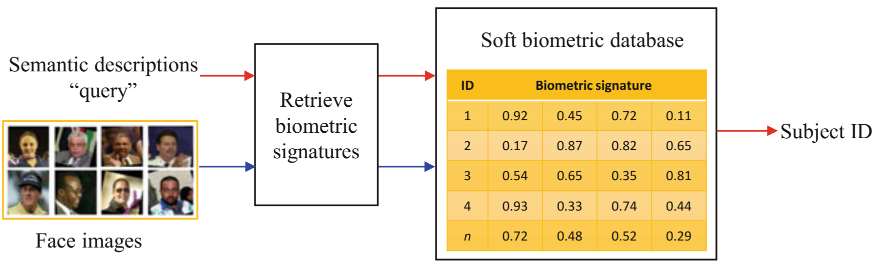

Fig. 2.1 Soft biometric identification using face attributes

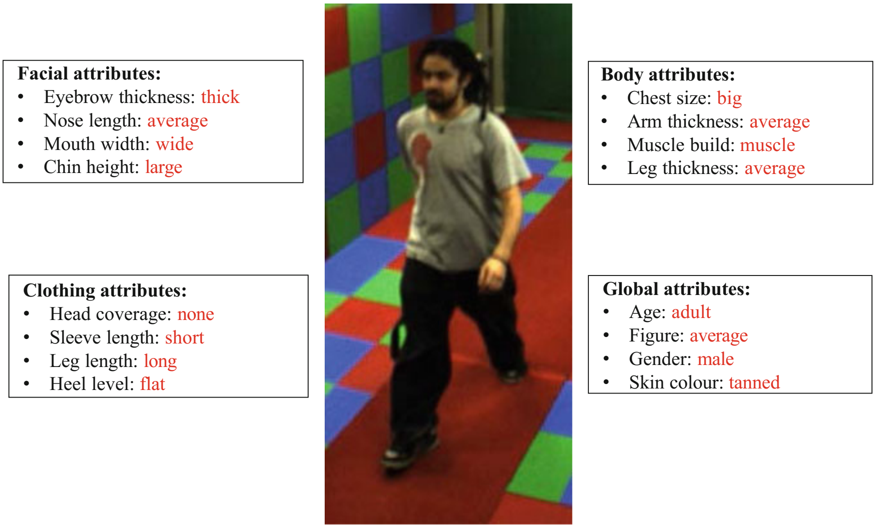

Fig. 2.2 Example categorical soft biometric attributes for a sample from the University of Southampton Biometric Tunnel (BioT) database [11]

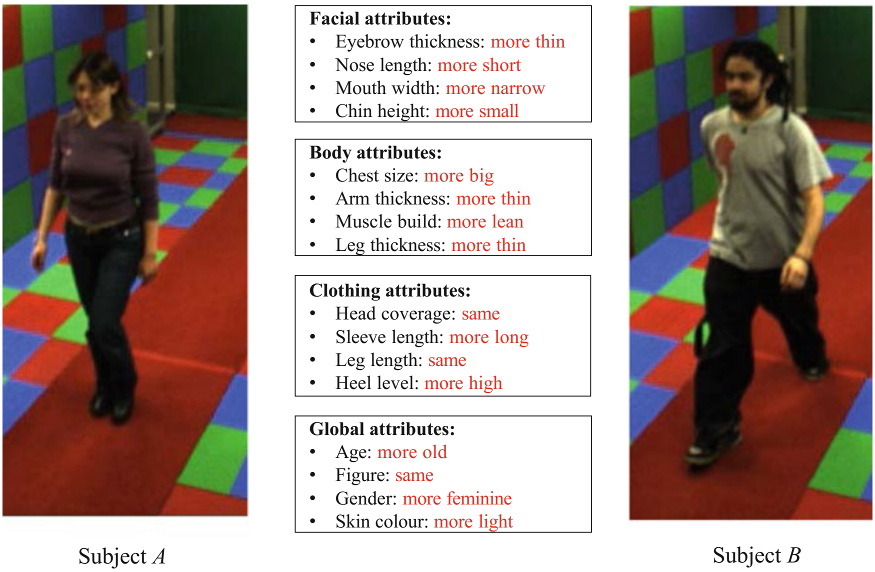

Fig. 2.3 Example comparative soft biometrics for a pair of subjects from the BioT database. The labels describe attributes of subject *A* with respect to subject *B*

Table 2.1 Semantic attribute formats

Format	Type	Nature of labels	Example
Categorical	Presence/absence	Binary (T/F)	*A* has **bushy** eyebrows
	Multiclass	Absolute	*A* has a **very large** mouth
Comparative	Similarity	$=, \neq$	*A* is the **same** age as *B*
	Dominance	$<, >$	*A* is **older** than *B*

Literature Review

Face Recognition in Psychology

Understanding face recognition from a psychological perspective is vital for studying facial identification using soft biometrics, as human identification accuracy is significantly affected by behavioural and physiological factors such as memory and encoding [13, 14]. In addition, human face recognition has substantial implications on automatic face recognition [15, 16], which has a wide range of real-life applications. The existing psychology literature on face recognition by humans is extensive and focuses particularly on understanding how the human visual and neural systems process the facial features and recognise faces. One of the key studies that has addressed face recognition in humans is that of Bruce and Young [17], who provisioned a theoretical model for understanding how humans recognize faces that is based on a set of distinguishable codes for face recognition, which are: label pictorial, structural, identity-specific semantic, visually derived semantic, and name. Bruce and Young's study suggested that regular face recognition involves the structural and identity-specific semantic codes (e.g. a familiar person's occupation and friends), while the other codes are used in lower levels for face recognition and perception. In [18], Hancock et al. outlined the factors that affect humans accuracy of recognizing unfamiliar faces. It has been found that changes in viewpoint and illumination significantly affect the ability of humans to recognize unfamiliar faces from images. Also, the spatial relationship between the facial components (i.e. configuration) has a high impact on recognition accuracy. Furthermore, the study emphasised the effect of face distinctiveness on recognition accuracy, and highlighted the role that can be played by the machines in aiding face recognition performance by humans.

A detailed overview on the human face recognition in a psychological and neural contexts was offered by O'Toole [19]. Her study provided some insights for addressing the problem of automatic face recognition. O'Toole stressed on humans ability to identify faces with the aid of semantic categories such as age, gender, and race, which increases recognition accuracy. Moreover, the study pointed the significance of the observer's as well as the observed person race on the recognition accuracy. Sinha et al. [15] outlined the key findings from previous experimental studies of face recognition in humans that can aid face recognition in computational systems. Their study highlighted the impact of spatial resolution on recognition performance, in addition to highlighting the holistic processing of facial features by the human visual system. Also, the study showed that pigmentation cues (i.e. texture or colour) have at least the same importance as the shape cues in face recognition. Tsao and Livingstone [20] highlighted the differences between computer vision algorithms and humans in face detection and recognition. The study emphasised on the independence of face detection and recognition stages in the human neural system, and outlined the differences in processing faces as compared with other objects in the neural system. Also, the study stressed the norm-based coding nature of face processing in humans, and the interpretation of faces in terms of their distinctiveness. Finally, the study highlighted the holistic processing of faces in the human neural system and stressed on the effect of the whole face on the processing of the individual facial components.

Several studies investigated the significance of facial features in recognition performance. Haig [21] explored the relative importance of facial features in face recognition by manipulating primary facial features of target samples and evaluating the recognition performance of the observers accordingly. The study found that eye-eyebrow, followed by mouth, and nose, have the highest importance for face recognition. Davies et al. [22] assessed the saliency of different facial features, where their experiments were based on manipulating faces using the Photofit Kit and monitoring the identification performance of different alternatives of samples. They have found that forehead and eyes have the highest saliency (i.e. their changes are most likely to be noticed), while chin and nose have the least saliency. Sadr et al. [23] investigated the role of eyes and eyebrows in face recognition and found that eyebrows have a significant role in face recognition that is at least similar to that of the eyes. Furthermore, their experiments reported that the absence of eyebrows results in a larger degradation in the face recognition performance as compared to the absence of eyes.

In summary, the findings of these studies highlight the role of semantic bindings in human face recognition, and show how categorization of facial or personal characteristics can affect the human face recognition performance. In addition, they outline the relative importance of the different facial parts, and show that eye-eyebrow region is the most important in face recognition. The implications of these studies will be noted on the proposed facial soft biometrics in this chapter.

Human Descriptions and Eyewitness Statements

There is a large volume of published studies that explored the reliability of verbal descriptions in eyewitness statements for suspects identification in criminal investigations. Kuehn [24] evaluated the effectiveness of verbal descriptions provided by victims of violent crimes to the police and examined the descriptions completeness using a random sample from the police records. The examined descriptions included the suspects' physical traits which are: age, sex, height, weight, build, skin colour, hair, and eye. The results of the study showed that more than 85% of the eyewitnesses were able to provide at least six attributes in their descriptions, where sex could be identified in 93% of the descriptions, while eye colour was the least to be recognised in the sample. The results also revealed that eyewitnesses cannot recall discrete traits of the suspects, but rather they have a general impression about the suspects.

In an analysis of the content of 139 eyewitnesses' descriptions, Sporer [25] found that 31% of the witnesses reported clothing attributes, 29.6% described facial features, 22% specified global features (i.e. age, and gender) in addition to movement descriptions, and the remaining descriptors used other features (e.g. smell, accent, and accessories). Sporer's analysis of the facial descriptions showed that most of the eyewitnesses described the upper half of the face, and more specifically, the hair of the suspect. Also, the study pointed that although the hair was the most mentioned in eyewitnesses descriptions, it is less reliable for locating the suspects as compared with the inner facial features, since hair style can be easily changed.

Koppen et al. [26] assessed the completeness and accuracy of eyewitnesses' descriptions using 2299 statements of 1313 eyewitnesses for 582 different robbers from official police records and investigated the factors that affect the accuracy and completeness of the statements. The findings emerged from their study revealed that the eyewitnesses tend to provide more information about general descriptions (e.g. age and gender) as compared to facial features. In addition, the study showed that the completeness of the descriptions did not necessarily imply their accuracy, thus although the information provided by the eyewitness was little, it tended to be accurate. In [27], Burton et al. explored subjects' ability to identify target people from surveillance videos. They found that familiarity with target faces has a substantial impact on the accuracy of identification. Thus, face recognition performance with familiar targets is much better than it is with unfamiliar ones. Furthermore, the study revealed that even with the poor quality of surveillance data, the face has a significantly higher impact on identification accuracy compared with the gait and body.

A significant analysis was presented by Meissner et al. [28], which outlined the psychological factors that affect eyewitness descriptions as: (1) encoding-based, which affects a person's perception such as illumination, distance, and stress; (2) person variables, which are age, gender, and race; and (3) the effect of inaccurate information from co-witnesses or investigators. Lee et al. [29] have conducted a detailed examination of the impact of a feature-based approach in suspect identification. Their experiments have shown that using a subjective feature-based approach for retrieving suspects from a database of mugshots is more efficient and accurate than presenting mugshots for an eyewitness in arbitrary order. Their experiments have also revealed that the feature-based approach of identifying suspects is effective for recognising faces in realistic conditions.

Overall, these studies reveal that the accuracy and completeness of eyewitnesses descriptions are determined by multiple factors such as the spatial and temporal settings of the incident, in addition to the eyewitness personal variables (e.g. age and gender). Furthermore, the findings of these studies stress on the tendency of eyewitnesses to describe general physical characteristics such as age and gender (i.e. global soft biometrics) in their statements, whereas facial features were described less likely. Also, the feature-based approach of identifying faces has been investigated and found to have a better impact on the efficiency and accuracy of suspects identification. Taken all together, these outcomes imply that global soft biometrics (e.g. age and gender) are essential for identification. In addition, the findings highlight the inadequacy of facial features for verbal descriptions as compared with other physical features, suggesting the introduction of more effective semantic facial attributes, which is one of the objectives of this chapter.

Facial Soft Biometrics

Due to its richness of features and details, the human face is considered as the most informative source of attributes for identification at a short distance as compared to other soft biometrics such as body and clothing [8, 9, 30]. Also, human face recognition demonstrated great robustness for challenging visual conditions such as low resolution and pose variability [31]. Therefore, a great deal of previous research into soft biometrics has focused on facial attributes either to improve the performance of traditional face recognition systems or to perform identification exclusively based on facial attributes [32].

The earliest exploration of semantic facial attributes emerged in [33], where face verification using the LFW database [34] was examined via attribute classifiers that were trained to recognize the presence or absence of a trait (i.e. categorical soft biometrics). The attributes covered 65 describable visual traits such as eyebrow shape, nose size, and eye width. The approach resulted in lowering the error rates by 23.92% compared to the state-of-the-art reported at that time on the LFW database. In [35], the authors studied the use of facial marks such as scars, moles, and freckles, as categorical soft biometrics to improve face recognition performance using the FERET database [36]. Their experiments demonstrated the improvement that can be achieved by augmenting facial marks (as soft biometrics) with traditional facial hard biometrics.

A key study in facial soft biometrics is that of Reid and Nixon [37], which is the first study to investigate human face identification using comparative soft biometrics. In this study, 27 comparative facial attributes were defined, and annotations were collected for subjects from the University of Southampton Gait Database (SGDB) [38]. The experiments showed that comparative facial soft biometrics outperform categorical facial soft biometrics. Thus, the rank-1 retrieval accuracy achieved using categorical attributes is 59.3%, compared to 74.5% in case of comparative attributes.

The first study that explored the interaction between automatically extracted soft biometrics and human generated soft biometrics is that of Klare et al. [39], which presented a method for using categorical facial attributes to perform identification in criminal investigations. Aiming to capture all persistent facial features, a set of 46 facial attributes were defined, and a model was trained to extract facial features and estimate them using SVM Regression automatically. Identification experiments were performed using the FERET [36] database with all the possible combination of probe-gallery (i.e. human vs. machine). Identification using an automatic probe and gallery resulted in the best recognition accuracy as compared with the other three identification scenarios in which human annotations are used (i.e. for probe, gallery, or both).

The study by Tome et al. [40] considered shape, orientation, and size of facial attributes as soft biometrics that can be utilized for forensic face recognition. The study proposed an approach to automatically convert a set of facial landmarks to a set of facial attributes (shape and size), which can be of continuous or discrete values (i.e. categorical). These features were used to generate statistics that aid forensic examiners in carrying morphological comparisons of facial images. Also, they were used to improve the performance of traditional face recognition system. Using ATVS [30] and MORPH [41] databases, the experiments revealed that the facial soft biometrics improve the accuracy of traditional face recognition systems. Recent work by Samangouei et al. [42] has investigated the use of categorical facial attributes for active authentication on mobile devices using the MOBIO [43] and AA01 [44] unconstrained mobile datasets. Their approach has highlighted the reliability of binary facial attributes for face verification on mobile devices and demonstrated the improvement that can be gained from fusing the scores of low-level features with the attribute-based approach. Table 2.2 summarizes the existing work that studied facial attributes for identification and verification.

In general, it can be seen from this literature review that the use of facial soft biometrics for human face identification has been studied using relatively constrained databases [8, 30, 39], whereas the real identification scenarios involve larger populations with significant demographic diversity, in addition to more challenging visual conditions of surveillance such as high variability in illumination, facial expressions, resolution, and pose. Except for Kumar et al. work [33], no study has so far addressed the use of facial soft biometrics for subject identification in large and unconstrained datasets like Labelled Faces in the Wild (LFW) [34]. In addition, although comparative soft biometrics have shown better identification accuracy as compared to categorical soft biometrics [8, 12], it is still not known whether comparative soft biometrics can scale for large and unconstrained datasets (e.g. LFW), as the studies in comparative soft biometric were performed using small and relatively

Table 2.2 Existing work on identification and verification using facial attributes

Publication	Dataset (no. of subjects)	Attributes	Labels
Kumar et al. [33]	LFW (5749)	65 categorical (binary)	Automatic
Reid and Nixon [37]	SGDB (100)	27 comparative	Human-based
Klare et al. [39]	FERET (1196)	46 categorical	Automatic and human-based
Tome et al. [40]	MORPH (130) and ATVS (50)	32 categorical	Automatic
Samangouei et al. [42]	MOBIO (152) and AA01 (50)	44 categorical (binary)	Automatic

constrained database [11]. Altogether, the findings from this literature review highlight the importance of exploring human identification using comparative facial soft biometrics in large unconstrained databases. Accordingly, this chapter aims to address the inadequacies of previous studies and to explores unconstrained human identification using comparative facial soft biometrics.

Comparative Facial Soft Biometrics

As mentioned in section "Face Biometrics and Semantic Face Recognition", the research in soft biometrics has largely been focused on categorical descriptions of facial attributes [7, 30, 39, 40, 45]. However, describing visual attributes in a relative (comparative) format has several advantages [46]. First, it makes richer semantics for humans (e.g. person *A* is *thinner* than person *B*). Second, it enables comparisons with a reference object (e.g. person *A* is *taller* than Bernie Ecclestone). Third, it improves interactive learning and makes searching based on an attribute more efficient (e.g. search for a *younger* person). Figure 2.4 illustrates the descriptive enrichment that can be gained by using comparative attributes. Besides these advantages of comparative attributes, the application of comparative soft biometrics in human identification

A is adult **B** is senior **C** is senior

(a) Categorical age attributes

A is younger than **B** **B** is younger than **C** **C** is older than **A**

(b) Comparative age attributes

Fig. 2.4 Expressing age of subjects from the LFW database [34, 47] using semantic labels has demonstrated the superiority of relative attributes as compared to the categorical attributes [8, 10, 12].

The aim behind comparative soft biometrics is to create a biometric signature for each individual that embeds the individual's physical attributes, and consequently, allows each individual to be uniquely identified in a database as shown Fig. 2.1. This biometric signature is a vector that is composed of the relative strength of each soft biometric attribute. Relative rates are inferred from pairwise comparisons between the individual being evaluated and other individuals in the database. The generation of relative rates from pairwise comparisons can be achieved using a ranking algorithm that infers a rate for each item in a dataset from its pairwise comparisons. One popular ranking algorithm that has been used in prior work on comparative soft biometrics is the Elo rating system [8, 10, 12, 37], which is a well-known algorithm for rating chess players [48, 49]. Also, RankSVM [50], which is a formulation that learns a rating function from example pairwise comparisons to infer the relative rate of attributes, has been used in some studies on comparative soft biometrics [51, 52] to predict biometric signatures. The relative rating of soft biometric attributes using the Elo rating system is described later in this section.

Attributes and Comparative Labels

The human face has an abundance of features that can be used for identification. However, to consider a facial feature as an effective soft biometric attribute, it needs to be understandable, memorable, and describable. These aspects have governed the selection of facial features to define the soft biometric set used in this chapter, which covered the major facial components (i.e. eyes, eyebrows, mouth, and nose), with an emphasis on eyebrows according to their pivotal role in human face recognition [23]. Table 2.3 lists the proposed comparative soft biometric attributes that are analysed throughout this chapter. Each attribute is associated with a comparative label that is based on three-point bipolar scale, which ranges from -1 to 1, such that -1 is associated with the "*Less*" label, while 1 is associated with the "*More*" label. The normalised value of a comparative label is used to generate the relative rate of the attributes, which is computed using the Elo rating system as described later in section "Relative Rating of Attributes".

Dataset and Label Acquisition

Labelled Faces in the Wild (LFW) [34] is a popular database that is used for unconstrained face recognition, and consists of more than 13000 facial images extracted from the web. The images of LFW have significant variations in pose, lighting, resolution, and facial expressions, which make them suitable to study unconstrained face recognition. The LFW database is composed of two subsets: *View 1*, which is

Table 2.3 Comparative facial soft biometrics defined to analyse unconstrained human identification and verification

No.	Attribute	Label
1	Chin height	[More Small, Same, More Large]
2	Eyebrow hair colour	[More Light, Same, More Dark]
3	Eyebrow length	[More Short, Same, More Long]
4	Eyebrow shape	[More Low, Same, More Raised]
5	Eyebrow thickness	[More Thin, Same, More Thick]
6	Eye-to-eyebrow distance	[More Small, Same, More Large]
7	Eye size	[More Small, Same, More Large]
8	Face length	[More Short, Same, More Long]
9	Face width	[More Narrow, Same, More Wide]
10	Facial hair	[Less Hair, Same, More Hair]
11	Forehead hair	[Less Hair, Same, More Hair]
12	Inter eyebrow distance	[More Small, Same, More Large]
13	Inter pupil distance	[More Small, Same, More Large]
14	Lips thickness	[More Thin, Same, More Thick]
15	Mouth width	[More Narrow, Same, More Wide]
16	Nose length	[More Short, Same, More Long]
17	Nose septum	[More Short, Same, More Long]
18	Nose-mouth distance	[More Short, Same, More Long]
19	Nose width	[More Narrow, Same, More Wide]
20	Spectacles	[Less Covered, Same, More Covered]
21	Age	[More Young, Same, More Old]
22	Figure	[More Thin, Same, More Thick]
23	Gender	[More Feminine, Same, More Masculine]
24	Skin colour	[More Light, Same, More Dark]

dedicated to training and model selection; and *View 2*, which is dedicated to performance analysis. The training subset of *View 1* consists of 9525 sample face images for 4038 subjects, some of these subjects have one sample in the database, while the others have two or more samples.

To explore unconstrained identification using comparative facial soft biometrics, a dataset that includes the 4038 subjects of the training subset of *View 1* from the LFW database was created by selecting one sample face image for each subject, and applying random selection whenever multiple samples exist for a subject. The selected images were all aligned using deep funnelling [47], which is an approach that incorporates unsupervised joint alignment with unsupervised feature learning to align face images, and reduces the effect of pose variability correspondingly. Also, all the images in the dataset were normalized to an inter-pupil distance of 50 pixels to ensure consistent comparisons between subjects.

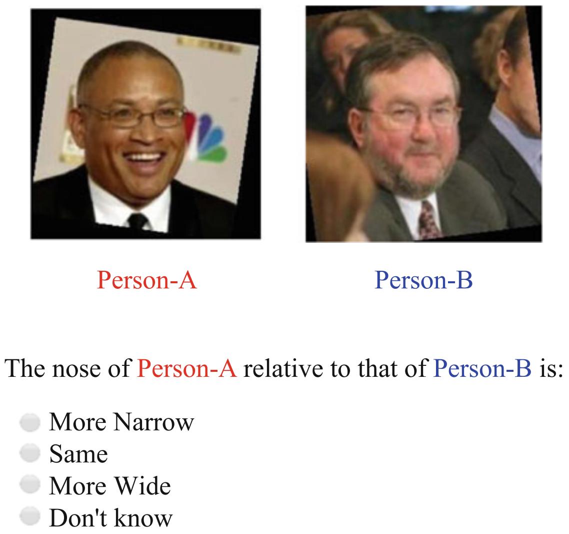

Fig. 2.5 Example crowdsourced comparison

The number of pairwise comparisons that result from a set of n items is $n(n-1)/2$, accordingly, the 4038 subjects in the LFW dataset result in 8.15 million pairwise comparisons, which is a massive number that is infeasible to be crowdsourced. Therefore, a graph that models pairwise relations between the 4038 subjects has been designed using a topology that ensures the involvement of each subject in at least four pairwise comparisons. The graph resulted in 10065 pairwise comparisons that were crowdsourced via the CrowdFlower platform,[1] and each of the 10065 crowdsourced comparisons consists of 24 questions targeting the comparative labelling of the attributes that are listed in Table 2.3. As explained earlier, each attribute is labelled based on a 3-point bipolar scale that represents the difference between the two subjects being compared. Figure 2.5 shows an example crowdsourced comparison. The crowdsourcing of the LFW dataset comparisons resulted in the collection of 241560 comparative labels for the 10065 comparisons as shown in Table 2.4. The labels collected through crowdsourcing were used to infer more comparisons. Thus, given two comparisons that involve subjects A and B with a common subject C, a new comparison between A and B can be inferred according to the rules outlined in Table 2.5. Relation inference results in increasing the coverage of the dataset, and enriches the analysis accordingly.

[1] http://www.crowdflower.com.

Table 2.4 Crowdsourcing job statistics

	Collected	Inferred	Total
Attribute comparisons	241560	132879504	133121064
Subject comparisons	10065	5536646	5546711
Average number of comparisons per subject	4.98	1371.1	N/A
Number of annotators (contributors)	9901	N/A	N/A

Table 2.5 Relation inference rules

(A, C)	(B, C)	$\inf(A, B)$
=	=	=
>	<	>
<	>	<
>	=	>
<	=	<
>	>	N/A
<	<	N/A

Relative Rating of Attributes

Comparative soft biometrics aim to create a biometric signature for each individual that embeds the individual's physical attributes, and consequently, allows each individual to be uniquely identified. This biometric signature is a vector that is composed of the relative strength of each soft biometric attribute. The relative strength (or relative rate) is inferred from the pairwise comparisons between the individual begin evaluated and other individuals in the database. The generation of relative rates from pairwise comparisons can be achieved using a ranking algorithm that infers a rate for each item in a dataset from its pairwise comparisons. One popular ranking algorithm that has been used in prior work on comparative soft biometrics is the Elo rating system [8, 10, 12, 37], which is a well-known algorithm for rating chess players [48, 49]. Also, RankSVM [50], which is a formulation that learns a rating function from example pairwise comparisons to infer the relative rate of attributes, has been used in some studies on comparative soft biometrics [51, 52] to predict biometric signatures.

In this analysis, the Elo rating system is used to generate biometric signatures from comparisons, as its applicability and effectiveness for comparative soft biometrics have been already demonstrated [12, 37]. In addition, the Elo rating does not require training as is the case with RankSVM [50], which has also been proposed for rating comparative soft biometrics [51, 52]. The rating process in the Elo rating system starts by initializing the rates of all players in a tournament to an initial default value. Then, for a game between players A and B with the initial (default) rates R_A and R_B correspondingly, the expected scores, E_A and E_B are calculated as:

$$E_A = \left[1 + 10^{\frac{(R_B - R_A)}{400}}\right]^{-1} \tag{2.1}$$

$$E_B = \left[1 + 10^{\frac{(R_A - R_B)}{400}}\right]^{-1} \tag{2.2}$$

Subsequently, based on the game outcome (i.e. loss, win, or draw), the new rates, $\bar{R}_A$ and $\bar{R}_B$, for players A and B respectively, are:

$$\bar{R}_A = R_A + K(S_A - E_A) \tag{2.3}$$

$$\bar{R}_B = R_B + K(S_B - E_B) \tag{2.4}$$

where S_A and S_B are scores that are set depending on the game outcome as: 0 for loss, 1 for win, and 0.5 for draw, while K is the score adjustment parameter that determines the sensitivity of rate update, and its value is selected through cross validation, as it has a significant impact on the outcomes from the rating process. The Elo rating system can be used to rate facial soft biometrics in a similar scheme to chess rating. Thus, by considering the subjects of a dataset as players in a tournament, and assuming that a comparison between two subjects, A and B, for a particular facial attribute, X, is a game between two players, correspondingly, this comparison can result in one of three possible outcomes for each of the two players as: "*Less X*", "*More X*", or "*Same X*", for example: "subject A has a *more thick* eyebrow than subject B". Accordingly, the rates of the two subjects that make the comparison are updated based on the comparison outcome using Eqs. 2.1–2.4. Figure 2.6 shows examples of the outcomes of the relative rating using the Elo system for selected attributes.

Attribute Analysis

The analysis in this section aim to explore three main aspects: (1) statistical distribution and the correlation between attributes; (2) attribute discriminative power; and (3) attribute semantic stability. The analysis in this section was conducted using the

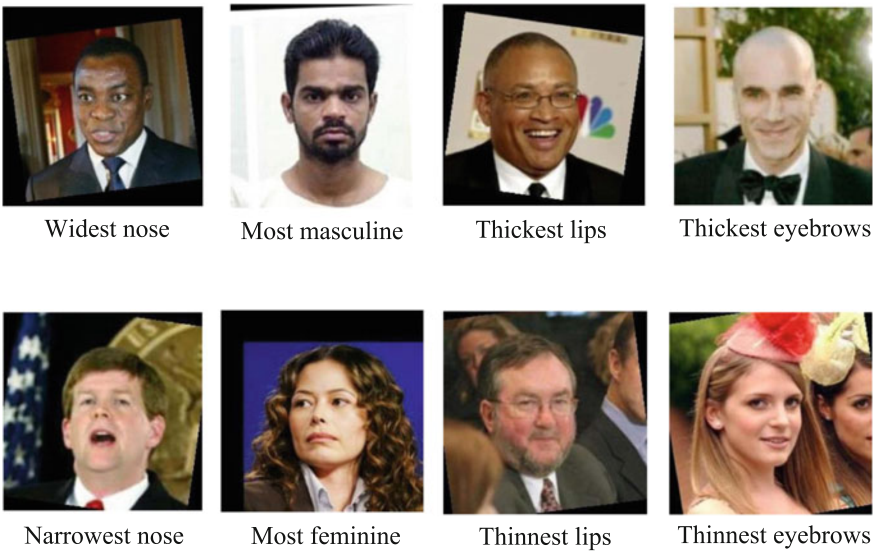

Fig. 2.6 The subjects with the lowest and highest relative rates for selected attributes

relative rates of the attributes, which were generated from the crowdsourced comparative labels using the Elo rating system as explained in the previous section.

Statistical Overview

Fig. 2.7 shows the distribution of the attributes in the LFW dataset summarized using box plot. As can be noted from Fig. 2.7, the attributes significantly differ in their distribution and variation, which reflect the demographic diversity of the LFW dataset. Also, there is a strong presence of outliers with most of the attributes, which indicates the challenging nature of the dataset. Closer inspection of the distribution shows that *facial hair* and *spectacles*, which are more of a binary nature, have the greatest variation. On the contrary, *eyebrow hair colour* has the lowest variation. These findings might suggest the potential discriminative power of binary-like attributes, and the low power of *eyebrow hair colour*. This to be further investigated through the discriminative power analysis.

Investigating correlation between the attributes allows exploring their independence, in addition to the potential contribution of individual attributes in the distinguishing subjects. Figure 2.8 shows the correlation map for the attributes, where it can be seen that all correlations between the attributes are either insignificant or weak, with the exception of the negative correlation between *facial hair* and *skin colour*, which can be attributed to the low contrast between darker *skin colour* and

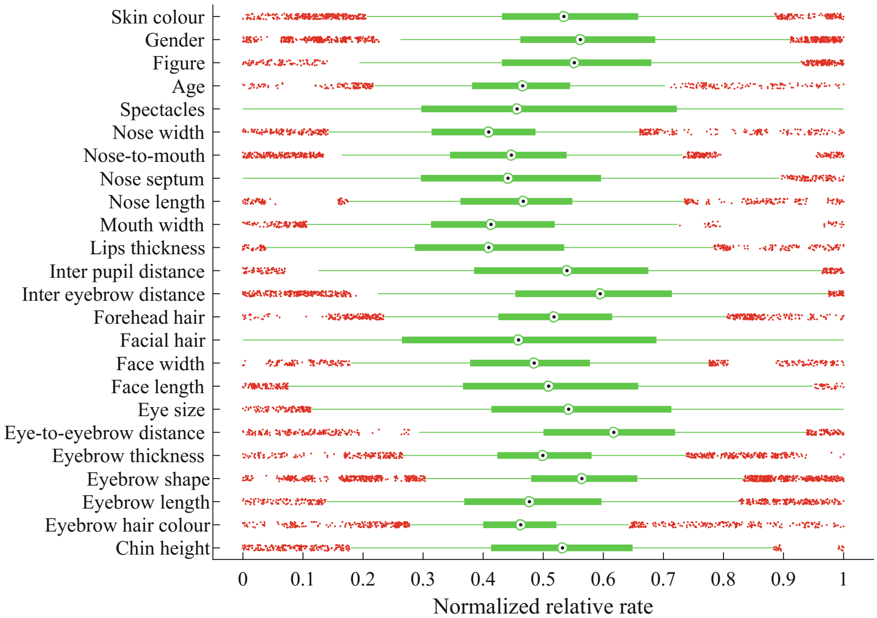

Fig. 2.7 Box plot for relative rates of the attributes

facial hair. In general, the findings from the correlation analysis confirm the independence of the attributes, and reveal the informative value embedded in each attribute.

Attribute Discriminative Power

Discriminative power analysis allows ranking attributes on their capabilities in distinguishing subjects, and contribution in identification accordingly. Discriminative power can be assessed using entropy [39], which is an information theoretic measure that represents the average amount of information contained in a random variable X, and it is calculated as follows:

$$H(X) = -\sum_{x \in X} p(x) log_2 \left[p(x) \right] \tag{2.5}$$

where X is a discrete random variable, and $p(x)$ is the probability distribution function of X. In the context of soft biometrics, it is assumed that the relative rates of attributes are random variables, and thus, entropy can be used to measure the information contained in each attribute, providing us with an indicator of the impact of each attribute in distinguishing between subjects.

Figure 2.9 shows the discriminative power of the attributes in terms of normalized entropy. It can be seen from Fig. 2.9 that *spectacles*, *facial hair*, and *gender*, which

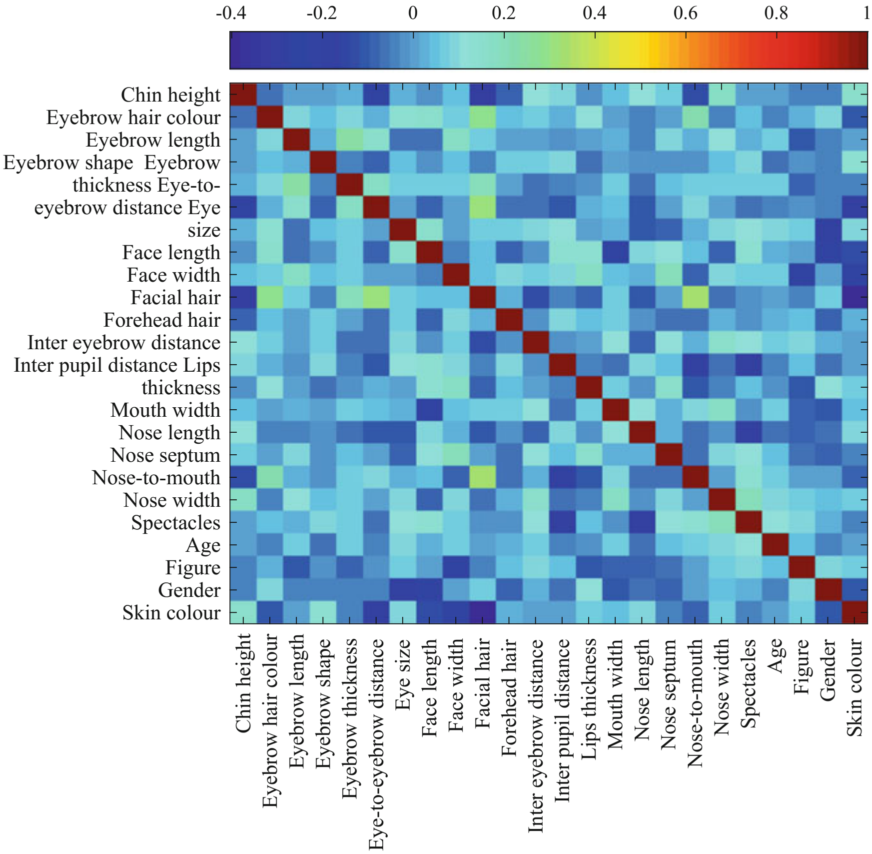

Fig. 2.8 Correlations between the attributes

are binary-like attributes, have relatively high discriminative power. Also, the results show the relatively high discriminative power of *eyebrow shape* and *inter eyebrow distance*, which highlight the role of eyebrows in human face recognition. On the other hand, the analysis shows that *age* has the lowest discriminative power, which can be attributed to the inaccuracy of human estimations for age from face images [53]. Furthermore, *eyebrow hair colour* and *forehead hair* demonstrate low discriminative power, which can be due to the difficulties of estimating these attributes.

Attribute Semantic Stability

Semantic stability can be defined as the consistency of an attribute rate among different annotators, which is substantial in assessing the attribute effectiveness and

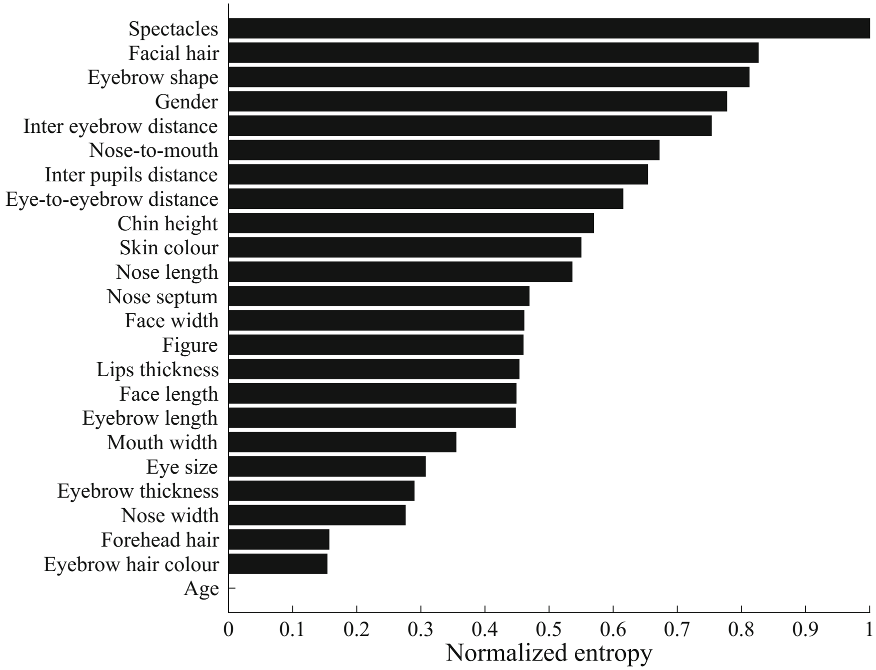

Fig. 2.9 Discriminative power of the attributes

robustness. The semantic stability of the attributes is evaluated by creating two different galleries, each of which consists of the biometric signatures of all the subjects in the dataset, where each biometric signature is composed of the relative rates of the 24 soft biometric attributes (listed in Table 2.3). The relative rates in each gallery are inferred using the Elo rating system based on two mutually exclusive subsets of comparative labels, which represent two different groups of annotators. Then, the semantic stability is measured for each attribute among the two galleries as the Pearson's correlation between the subjects' rates in both galleries. The results of the semantic stability analysis are shown in Fig. 2.10.

The semantic stability analysis demonstrated that all the attributes are statistically significant ($p < 0.05$), regardless of the strength of the correlation (stability) between the two galleries. Moreover, an interesting outcome from the result of the semantic stability analysis is the agreement of its ranking for the attributes with the ranking resulted from the discriminative power analysis in Fig. 2.9. Thus, the binary-like attributes (i.e. *spectacles*, *facial hair*, and *gender*) beside *eyebrow shape* have the highest semantic stability, while *age*, *eyebrow hair colour*, and *forehead hair* have the lowest discriminative power. This correspondence in the findings of semantic stability and discriminative power analysis reveals the robustness of entropy for assessing discriminative power of soft biometric attributes.

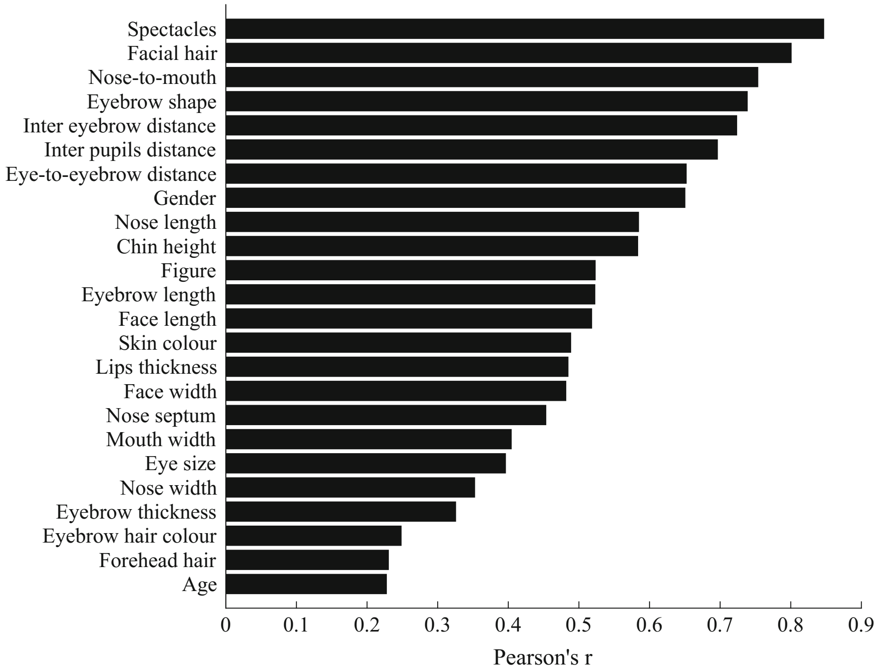

Fig. 2.10 Semantic stability of the attributes

Analysing Facial Comparisons

Identification Using Facial Comparisons

This experiment simulates a realistic scenario in which a semantic database DB_s is searched to identify an unknown subject using a verbal description for the subject's face as illustrated in Fig. 2.1. The identification performance evaluation follows a 6-fold cross validation, where the 4038 subjects of the LFW dataset are randomly divided into six equal subsets, and each subset is used for testing while the remaining five folds are used for training. For each subject in the test set, a probe biometric signature PR_s is generated from the relative rates of the 24 attributes (listed in Table 2.3). The relative rates are computed using the Elo rating system and based on comparisons between the probe and c other randomly selected subjects from the training folds. The remaining comparisons, after excluding those used in generating PR_s, are used to generate a biometric signature for each subject, which makes up the database DB_s to be searched. Then, the distance, d_p, between the probe and each subject in DB_s is calculated using the Pearson correlation coefficient as follows:

$$d_p = 1 - \frac{\sum_{i=1}^{t}(PR_s(i) - \overline{PR}_s)(S_c(i) - \overline{S}_c)}{\sqrt{\sum_{i=1}^{t}(PR_s(i) - \overline{PR}_s)^2}\sqrt{\sum_{i=1}^{t}(S_c(i) - \overline{S}_c)^2}} \tag{2.6}$$

where PR_s is the biometric signature of the probe, $S_c \in DB_s$ is the biometric signature of the counterpart subject in the database DB_s that is compared to the probe, and t is the number of features composing the semantic face signature. The rank of the correct match to the probe is used to report the identification performance via a Cumulative Match Characteristic (CMC) curve. This cross validation runs over the six folds, and repeated till the harmonic mean of identification rates among all the ranks converges. Figure 2.11 shows the CMC curve resulted from this experiment and it can be seen that using ten subject comparisons to generate the probe biometric signature, which is the ideal size of identity parade [10], an identification rate of 92.62% can be achieved. Rank-10 identification rate increases to 98.14% and 99.41% as the number of subject comparisons increase to 15 and 20 respectively.

The relationship between the number of comparisons used in generating biometric signatures and identification performance can be seen in Fig. 2.12 and it shows that the identification performance in terms of mean rank of retrieved match converges starting from eight comparisons. Furthermore, the impact of using facial comparisons on identification performance can also be seen from another interesting perspective, which is the compression of search range in a database. Thus, narrowing down search range becomes vital for efficiency of identification in large database. In addition, when verbal descriptions are not sufficiently accurate, search compression can lead to filtering out a long list of suspects, making subject retrieval more efficient.

Figure 2.13 demonstrates the compression in the search range that can be achieved in the LFW dataset using the comparative facial soft biometrics. It shows that using two comparisons only, the search range can be narrowed down to 10.55% of the total dataset with probability $p = 0.99$ that a correct match with the subject will be found.

To the best of our knowledge, the only study that has investigated human identification using facial soft biometrics for both probe and gallery in a relatively large database is that of Klare et al. [39], which achieved a rank-1 accuracy of 22.5% using

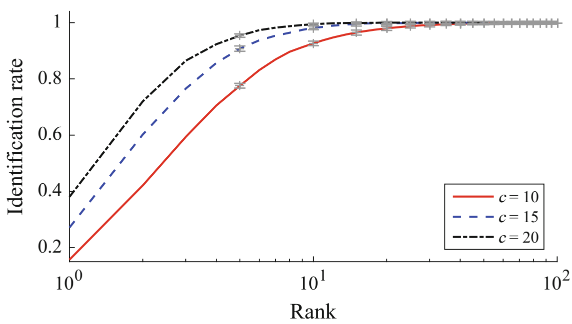

Fig. 2.11 Identification performance using $c = \{10, 15, 20\}$ subject comparisons

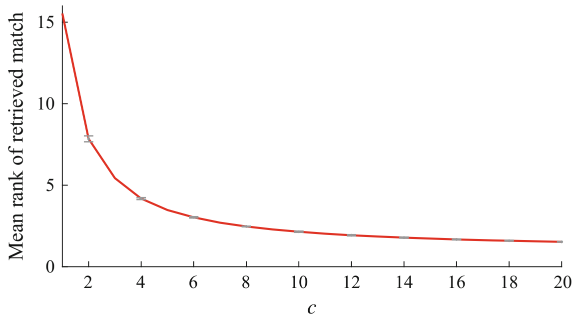

Fig. 2.12 Effect of number of subject comparisons, c, on identification performance

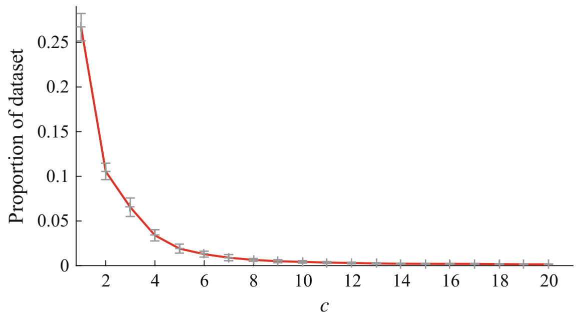

Fig. 2.13 Compression achieved in search range with probability of finding a correct match $p = 0.99$ at different number of comparisons, c

46 attributes (27 categorical and 19 binary) with 1196 subjects from the relatively constrained FERET database. Our approach can reach a rank-1 accuracy of 38.04% with 20 subject comparisons and 24 comparative attributes only, using the larger and more challenging LFW dataset. This performance advantage is due to the use of comparative attributes [10]. Furthermore, the results of the identification experiment demonstrate the power of comparative soft biometrics for unconstrained face identification, and reveal their scalability for relatively large databases.

Verification Using Facial Comparisons

Analysing the verification accuracy of comparative facial soft biometrics is necessary to evaluate the extent of agreement among semantic descriptions collected from different annotators (i.e. eyewitnesses) for the same subject. Each biometric signature for a subject can be considered as a unique sample, provided that it was generated using a unique group of comparisons. Therefore, in this experiment, two samples (biometric signatures) were generated for each of the 4038 subjects of the LFW dataset using randomly selected and mutually exclusive comparisons. The number of comparisons that were selected to generate biometric signatures was set to 10, as it is the average ideal size for an identity parade [10]. The verification was assessed using the two galleries resulted from the generated biometric signatures.

As Fig. 2.14 shows, the comparative facial soft biometrics achieved an equal error rate (EER) of 15.4% using ten subject comparisons only. Also, the attributes achieved an area under the curve (AUC) of 92.32% as can be seen in Fig. 2.15a. Furthermore, it can be seen from Fig. 2.15b that as the number of subjects comparisons used in generating biometric signature increases, the verification performances improves.

Using the BioT database, Tome et al. [30] achieved an EER of 13.54% by utilizing an automatic face detection and recognition system, with the fusion of 23 categorical soft biometrics (13 bodily, 3 global, and 7 head). While the proposed comparative attributes (without face hard biometrics) resulted in a slightly higher EER of 15.4% using ten subject comparisons, which decreases to 11.32% outperforming the approach in [30] with five more subjects comparisons only. Overall, the results of this experiment demonstrate the verification capability of the proposed comparative soft biometrics and show that the attributes can outperform the performance of automatic face recognition with or without fusing categorical soft biometrics.

Approach Limitations

Although the effectiveness and reliability of the proposed approach have been demonstrated in this chapter, some issues need to be considered for the practical application of the approach. First, the subject comparisons have taken place while the images of subject pairs were presented for the annotators, whereas real identification scenarios involve recalling suspects' facial attributes from eyewitnesses' memory. Second, as explained in section "Relative Rating of Attributes", the relative rating of comparative soft biometrics using the Elo rating system involves a cross validation to tune the score adjustment parameter, and this can be costly with large databases. Third, despite its effectiveness for studying comparative soft biometrics, crowdsourcing of comparisons is not scalable for very large databases due to the time and cost associated with it. Accordingly, a framework for the automatic estimation of

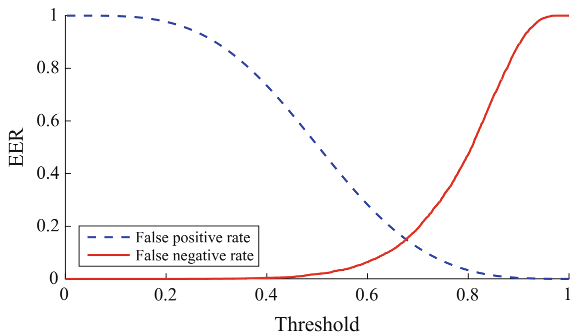

Fig. 2.14 Error curves resulted from verification using facial comparisons

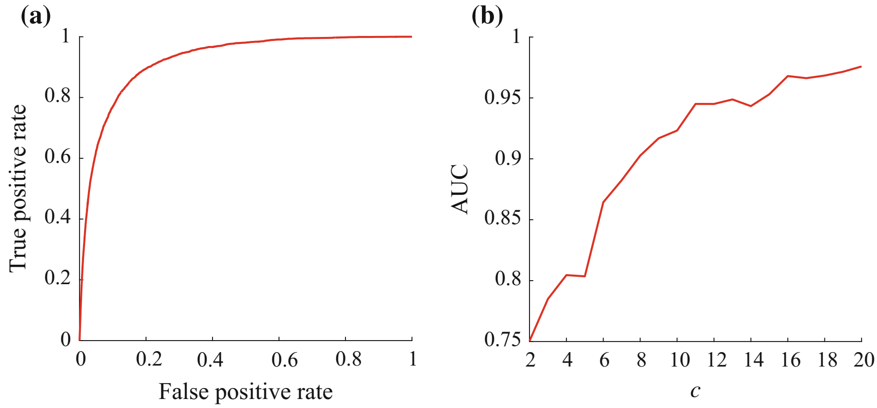

Fig. 2.15 **a** ROC curve for comparative facial attributes using $c = 10$ subject comparisons. **b** Area under the curve (AUC) versus number of subject comparisons, c

comparative facial soft biometrics is required to reduce the dependency on human annotators. Finally, most of the subjects of the LFW database are public figures that are familiar to the annotators, while the real life scenarios involve the description of unfamiliar faces (e.g. eyewitness descriptions of a suspects), which might be more difficult to recall and describe.

Summary

This chapter aims to analyse unconstrained human identification using comparative facial soft biometrics. We have presented a literature review on identification using facial attributes and highlighted the need to bridge the knowledge gap in comparative facial soft biometrics. Also, we have proposed a novel set of comparative facial soft biometrics and demonstrated their significance. The performance analysis conducted in this chapter using the well-known LFW database have revealed the reliability of comparative facial soft biometrics for unconstrained identification and verification, in addition to their scalability for large databases. Thus, by comparing an unknown subject to a line up of ten subjects only, a correct match will be found in the top 84 retrieved subjects from a database of 4038 subjects. Furthermore, the attributes have revealed a notable verification accuracy achieving an EER of 15.4% with biometric signatures generated from ten subject comparisons only. In conclusion, the findings of this chapter extend our knowledge of capabilities of comparative soft biometrics and can serve as a base for future investigations of other comparative soft biometrics such as body and clothing.

In terms of directions for future work, it would be interesting to explore the automatic retrieval of comparative facial soft biometrics from images. Also, as criminal investigations involve the collection of memory-based descriptions from eyewitnesses, future work could determine the accuracy of humans in recalling facial attributes of unknown subjects from memory relative to presented subjects. and exploring the effectiveness of these soft biometric comparisons for identity retrieval.

References

1. Davis JP, Valentine T (2015) Human verification of identity from photographic images. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, pp 209–238
2. Crawford TAM, Evans K (2016) Crime prevention and community safety. Oxford University Press
3. Valentine T, Davis JP (2015) Forensic facial identification. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, pp 323–347
4. Kemp R, White D (2016) Face identification 3. In: An introduction to applied cognitive psychology, pp 39
5. Gerrard G, Thompson R (2011) Two million cameras in the UK. CCTV Image 42(10):e2
6. Barrett D (2013) One surveillance camera for every 11 people in Britain, says CCTV survey. The Telegraph 10
7. Samangooei S, Guo B, Nixon MS (2008) The use of semantic human description as a soft biometric. In: 2008 2nd IEEE international conference on biometrics: theory, applications and systems, BTAS 2008. IEEE, pp 1–7
8. Reid D, Samangooei S, Chen C, Nixon M, Ross A (2013) Soft biometrics for surveillance: an overview. Mach Learn Theory Appl 327–352
9. Nixon MS, Correia PL, Nasrollahi K, Moeslund TB, Hadid A, Tistarelli M (2015) On soft biometrics. Pattern Recognit Lett 68:218–230
10. Reid DA, Nixon MS, Stevenage SV (2014) Soft biometrics; human identification using comparative descriptions. IEEE Trans Pattern Anal Mach Intell 36(6):1216–1228
11. Seely RD, Samangooei S, Lee M, Carter JN, Nixon MS (2008) The University of Southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In 2008 2nd IEEE international conference on biometrics: theory, applications and systems, BTAS 2008. IEEE, pp 1–6
12. Reid DA, Nixon MS (2011) Using comparative human descriptions for soft biometrics. In: 2011 international joint conference on biometrics (IJCB). IEEE, pp 1–6
13. Ryder H, Smith HMJ, Flowe HD (2015) Estimator variables and memory for faces. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, pp 159–183
14. Robbins R, McKone E (2007) No face-like processing for objects-of-expertise in three behavioural tasks. Cognition 103(1):34–79
15. Sinha P, Balas B, Ostrovsky Y, Russell R (2006) Face recognition by humans: nineteen results all computer vision researchers should know about. Proc IEEE 94(11):1948–1962
16. Jonathon P (2014) Phillips and Alice J O'Toole. Comparison of human and computer performance across face recognition experiments. Image Vis Comput 32(1):74–85
17. Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77(3):305–327
18. Hancock PJB, Bruce V, Burton AM (2000) Recognition of unfamiliar faces. Trends Cogn Sci 4(9):330–337
19. O'Toole AJ (2005) Psychological and neural perspectives on human face recognition. In: Handbook of face recognition. Springer, pp 349–369
20. Tsao DY, Livingstone MS (2008) Mechanisms of face perception. Annu Rev Neurosci 31:411–437
21. Nigel D (1986) Haig. Exploring recognition with interchanged facial features. Perception 15(3):235–247

22. Davies G, Ellis H, Shepherd J (1977) Cue saliency in faces as assessed by the 'Photofit' technique. Perception 6(3):263–269
23. Sadr J, Jarudi I, Sinha P (2003) The role of eyebrows in face recognition. Perception 32(3):285–293
24. Lowell L (1974) Kuehn. Looking down a gun barrel: Person perception and violent crime. Percept Mot Skills 39(3):1159–1164
25. Sporer SL (1992) An archival analysis of person descriptions. In: Biennial meeting of the American Psychology-Law Society in San Diego, California
26. PJ van Koppen, Lochun SK (1997) Portraying perpetrators: The validity of offender descriptions by witnesses. Law Hum Behav 21(6):661
27. Burton MA, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor-quality video: Evidence from security surveillance. Psychol Sci 10(3):243–248
28. Meissner CA, Sporer SL, Schooler JW (2013) Person descriptions as eyewitness evidence. In: Handbook of eyewitness psychology: memory for people, pp 1–34
29. Lee E, Whalen T, Sakalauskas J, Baigent G, Bisesar C, McCarthy A, Reid G, Wotton C (2004) Suspect identification by facial features. Ergonomics 47(7):719–747
30. Tome P, Fierrez J, Vera-Rodriguez R, Nixon MS (2014) Soft biometrics and their application in person recognition at a distance. IEEE Trans Inf Forensics Secur 9(3):464–475
31. O'Toole A, Jonathon Phillips P (2015) Evaluating automatic face recognition systems with human benchmarks. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, p 263
32. Arigbabu OA, Ahmad SM, Adnan WA, Yussof S (2015) Sharifah Mumtazah Ahmad, Wan Azizun Adnan, and Salman Yussof. (2015) Recent advances in facial soft biometrics. Vis Comput Int J Comput Graph 31(5):513–525
33. Kumar N, Berg AC, Belhumeur PN, Nayar SK (2009) Attribute and simile classifiers for face verification. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE, pp 365–372
34. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report, Technical report 07–49, University of Massachusetts, Amherst
35. Park U, Jain AK (2010) Face matching and retrieval using soft biometrics. IEEE Trans Inform Forensics Secur 5(3):406–415
36. Phillips JP, Moon H, Rizvi SA, Rauss PJ (2000) The feret evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104
37. Reid DA, Nixon MS (2013) Human identification using facial comparative descriptions. In: 2013 International Conference on Biometrics (ICB). IEEE, pp 1–7
38. Shutler JD, Grant MG, Nixon MS, Carter JN (2004) On a large sequence-based human gait database. In: Applications and science in soft computing. Springer, pp 339–346
39. Klare BF, Klum S, Klontz JC, Taborsky E, Akgul T, Jain AK (2014) Suspect identification based on descriptive facial attributes. In: 2014 IEEE International Joint Conference on Biometrics (IJCB). IEEE, pp 1–8
40. Tome P, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J (2015) Facial soft biometric features for forensic face recognition. Forensic Sci Int 257:271–284
41. Ricanek K, Tesafaye T (2006) Morph: a longitudinal image database of normal adult age-progression. In: 2006 7th international conference on automatic face and gesture recognition, FGR 2006. IEEE, pp 341–345
42. Samangouei P, Patel VM, Chellappa R (2017) Facial attributes for active authentication on mobile devices. Image Vis Comput 58:181–192
43. McCool C, Marcel S, Hadid A, Pietikäinen M, Matejka P, Cernockỳ J, Poh N, Kittler J, Larcher A, Levy C et al (2012) Bi-modal person recognition on a mobile phone: using mobile phone data. In: 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). IEEE, pp 635–640
44. Fathy ME Patel VM, Chellappa R (2015) Face-based active authentication on mobile devices. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp 1687–1691

45. Jain AK, Park U (2009) Facial marks: soft biometric for face recognition. In: 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, pp 37–40
46. Parikh D, Grauman K (2011) Relative attributes. In: 2011 IEEE International conference on computer vision (ICCV). IEEE, pp 503–510
47. Huang G, Mattar M, Lee H, Learned-Miller EG (2012) Learning to align from scratch. In: Advances in neural information processing systems, pp 764–772
48. Glickman ME (1995) A comprehensive guide to chess ratings. Am Chess J 3:59–102
49. Elo AE (1978) The rating of chessplayers, past and present. Arco Publishing
50. Joachims T (2002) Optimizing search engines using clickthrough data. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 133–142
51. Jaha ES, Nixon MS (2014) Soft biometrics for subject identification using clothing attributes. In: 2014 IEEE international joint conference on biometrics (IJCB). IEEE, pp 1–6
52. Martinho-Corbishley D, Nixon MS, Carter JN (2016) Soft biometric recognition from comparative crowdsourced annotations. IET Biometrics 1–16
53. Han H, Jain AK (2014) Age, gender and race estimation from unconstrained face images. Department of Computer Science Engineering, Michigan State University, East Lansing, MI, USA, MSU Technical Report (MSU-CSE-14-5)

http://www.springer.com/978-3-319-68532-8

Surveillance in Action
Technologies for Civilian, Military and Cyber Surveillance
Karampelas, P.; Bourlai, T. (Eds.)
2018, XIV, 412 p. 149 illus., 128 illus. in color.,
Hardcover
ISBN: 978-3-319-68532-8