Forensic Science International 257 (2015): 271 – 284

DOI: http://dx.doi.org/10.1016/j.forsciint.2015.09.002

Facial Soft Biometric Features for Forensic Face Recognition

Pedro Tome, Ruben Vera-Rodriguez, Julian Fierrez and Javier Ortega-Garcia

Biometric Recognition Group - ATVS, Escuela Politecnica Superior
Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
{pedro.tome, ruben.vera, julian.fierrez, javier.ortega}@uam.es

Abstract

This paper proposes a functional feature-based approach useful for real forensic caseworks, based on the shape, orientation and size of facial traits, which can be considered as a soft biometric approach. The motivation of this work is to provide a set of facial features, which can be understood by non-experts such as judges and support the work of forensic examiners who, in practice, carry out a thorough manual comparison of face images paying special attention to the similarities and differences in shape and size of various facial traits. This new approach constitutes a tool that automatically converts a set of facial landmarks to a set of features (shape and size) corresponding to facial regions of forensic value. These features are furthermore evaluated in a population to generate statistics to support forensic examiners. The proposed features can also be used as additional information that can improve the performance of traditional face recognition systems. These features follow the forensic methodology and are obtained in a continuous and discrete manner from raw images. A statistical analysis is also carried out to study the stability, discrimination power and correlation of the proposed facial features on two realistic databases: MORPH and ATVS Forensic DB. Finally, the performance of both continuous and discrete features is analysed using different similarity measures. Experimental results show high discrimination power and good recognition performance, especially for continuous features. A final fusion of the best systems configurations achieves rank 10 match results of 100% for ATVS database and 75% for MORPH database demonstrating the benefits of using this information in practice.

Keywords:
Forensics, face recognition, anthropometric measures, soft biometrics, biometrics, face traits.

1. Introduction

Traditionally, face recognition approaches have fallen into two main categories: feature-based and holistic [1, 2, 3, 4]. Feature-based approaches first process the input image to identify and extract distinctive facial landmark points corresponding to facial regions such as the eyes, nose, mouth, etc., and then compute some measures describing those regions. On the other hand, holistic approaches attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. Also, there are hybrid methods that detect landmark points and then apply techniques used by holistic methods [5, 6, 7, 8].

On the other hand, the use of other ancillary information based on the description of human physical features for face recognition [9] has not been explored in much depth. Formally, the soft biometric information extracted from a human body (e.g., height, gender, skin color, hair color, etc.) is ancillary information that can be easily distinguished at a distance but it is not fully distinctive by itself in recognition tasks. However, this soft information can be explicitly fused with biometric recognition systems to improve the overall recognition when confronting high variability conditions.

This paper proposes a novel facial feature-based approach (hereafter called facial soft biometric features), which can describe the shape, orientation and size of the facial traits. This work is motivated by the description of the different types of facial traits made by A. Bertillon in the *portrait parle* (speaking picture) [10, 11], which has been improved and is still used with some variations by police agencies around the world [12, 13]. A. Bertillon, a French police officer from the late 1800's, created the first method to identify criminals based on

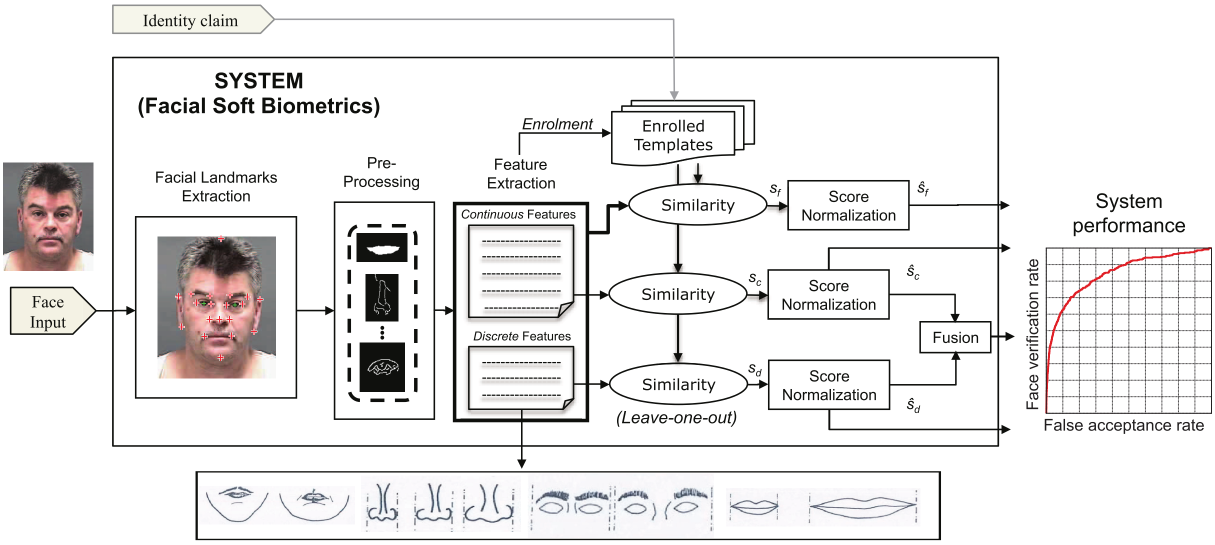


Figure 1: Experimental framework. Four configurations are analysed: *continuous* ($\hat{s}_c$) and *discrete* ($\hat{s}_d$) features individually, feature-level fusion ($\hat{s}_f$) and score-level fusion.

anthropometric measures. For each person arrested he collected in a card details of body measurements, descriptions of the facial traits such as the nose and ears, body markings, date of birth, and criminal history together with full-face and profile mugshots [15].

Nowadays, in practice, when forensic examiners compare two face images, they carry out a detailed visual comparison of the two images and write a morphological report. They focus their attention not only on the full face but also on each individual trait. Concretely, they carry out an exhaustive morphological comparison, analyzing the face, region by region (i.e., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc. In this trait by trait comparison they also consider the shape and size of the traits and use Bertillon specifications for descriptions. It is also very important to notice the difference between the traditional face features from the automatic system perspective [14] and the forensic face comparisons carried out in practice by forensic laboratories [10, 16]. The automatic systems use a computation language to extract features, which generally cannot be understood by non-expert persons.

The biometric system developed in this paper has been designed to help the work of forensic examiners when carrying out morphological comparisons of facial images. As can be seen in Fig. 1, , given an input or unknown face, the first step is to carry out facial landmark extraction. Then, an image pre-processing stage specifically designed for each facial trait is performed to obtain continuous and discrete facial features. The system developed follows a forensic methodology and thus can give an indication of the shape, orientation and size of relevant facial traits in an automatic way. This can help and support the decision of the forensic examiners. This work also carries out a thorough study of the uniqueness and discrimination power of such anthropometric-based facial features, as they are normally used now without a scientific assessment of their error rates in many practical cases [14, 17].

Additionally, the continuous and discrete features proposed here can also be used to perform automatic recognition. In order to do so, the features extracted from the input image are directly compared against the features from reference faces with a known identity (enrolled template) to produce a matching score. This approach can be seen as a type of facial soft biometric information which can complement more standard automatic facial recognition systems to improve their performance and robustness.

In this sense, first works in soft biometrics [18, 19, 20] tried to use demographic information (e.g., gender and ethnicity) and soft attributes like eye color, and other visible marks like scars [9, 21] and tattoos [22] as ancillary information to improve the performance of biometric systems. They showed that soft biometrics can complement the traditional (primary) biometric identifiers (like face recognition) [23], and can also be useful as a source of evidence in courts of law because they are more descriptive than the numerical matching

Facial Trait	*Continuous*	Definition	*Discrete* (Forensic categorisation)
Forehead	1. Height	$d(\mathbf{b}_1, \mathbf{b}_3)$	1. Height (Short, Average, and Large)
	2. Width	$d(\mathbf{b}_1, \mathbf{b}_4)$	2. Width (Small, Average, and Large)
Eyebrows	3. Separation (Dist. between eyebrows)	$d(\mathbf{a}_3, \mathbf{a}_4)$	3. Separation (Near and Distant)
	4. Inner Elevation$_L$ (Dist. eyebrow and eye)	$(1/n)\sum_{i=1}^{n} d(\mathbf{a}_9, \mathbf{p}_i)$	4. Elevation (Low, Average, High, and Asymmetric)
	5. Outer Elevation$_L$ (Dist. eyebrow and eye)	$(1/n)\sum_{i=1}^{n} d(\mathbf{a}_{11}, \mathbf{p}_i)$	
	6. Inner Elevation$_R$ (Dist. eyebrow and eye)	$(1/n)\sum_{i=1}^{n} d(\mathbf{a}_8, \mathbf{p}_i)$	
	7. Outer Elevation$_R$ (Dist. eyebrow and eye)	$(1/n)\sum_{i=1}^{n} d(\mathbf{a}_6, \mathbf{p}_i)$	
	8. Length$_L$	$d(\mathbf{a}_4, \mathbf{a}_5)$	5. Length$_L$ (Short and Long)
	9. Length$_R$	$d(\mathbf{a}_2, \mathbf{a}_3)$	6. Length$_R$ (Short and Long)
	10. Average Width$_L$	$(1/m)\sum_{i=1}^{m} d(\mathbf{r}_i, \mathbf{s}_i), (\mathbf{r}_1 = \mathbf{a}_4, \mathbf{r}_m = \mathbf{a}_5)$	7. Width$_L$ (Narrow, Linear, and Wide)
	11. Average Width$_R$	$(1/m)\sum_{i=1}^{m} d(\mathbf{r}_i, \mathbf{s}_i), (\mathbf{r}_1 = \mathbf{a}_2, \mathbf{r}_m = \mathbf{a}_3)$	8. Width$_R$ (Narrow, Linear, and Wide)
	12. Angles between corners$_L$	$\alpha(\mathbf{a}_4, \mathbf{a}_5)$	9. Direction$_L$ (Horizontal, Oblique Internal, and Oblique External)
	13. Angles between corners$_R$	$\alpha(\mathbf{a}_2, \mathbf{a}_3)$	
			10. Direction$_R$ (Horizontal, Oblique Internal, and Oblique External)
			11. Form$_L$ (Arched, Rectilinear, and Sinuous)
			12. Form$_R$ (Arched, Rectilinear, and Sinuous)
Eyeball and Orbit	14. Horizontal Opening$_L$	$d(\mathbf{a}_9, \mathbf{a}_{11})$	13. Horizontal Opening$_L$ (Small and Large)
	15. Horizontal Opening$_R$	$d(\mathbf{a}_6, \mathbf{a}_8)$	14. Horizontal Opening$_R$ (Small and Large)
	16. Interocular Dist. (inner corners)	$d(\mathbf{a}_8, \mathbf{a}_9)$	15. Interocular Dist. (Small, Normal, and Large)
	17. Angles$_L$ between corners	$\alpha(\mathbf{a}_9, \mathbf{a}_{11})$	
	18. Angles$_R$ between corners	$\alpha(\mathbf{a}_6, \mathbf{a}_8)$	
Nose	19. Width	$d(\mathbf{a}_{12}, \mathbf{a}_{14})$	16. Width (Small, Average, and Large)
	20. Height	$d(\mathbf{a}_1 3, \mathbf{b}_1)$	17. Height (Short, Average, and Large)
	21. Nose Root Width	$2 \cdot d(\mathbf{c}_1, \mathbf{c}_2)$	18. Nose Root Width (Narrow, Average, and Wide)
	22. Naso-Labial Height	$d(\mathbf{a}_{13}, \mathbf{c}_3)$	19. Naso-Labial Height (Short, Average, and Large)
Mouth	23. Length	$d(\mathbf{a}_{15}, \mathbf{a}_{16})$	20. Length (Small, Average, and Large)
	24. Average Height	$(1/m)\sum_{i=1}^{m} d(\mathbf{r}_i, \mathbf{s}_i), (\mathbf{r}_1 = \mathbf{a}_{15}, \mathbf{r}_m = \mathbf{a}_{16})$	21. Orientation (Oblique$_L$, Neutral, and Oblique$_R$)
	25. Angles between corners	$\alpha(\mathbf{a}_{15}, \mathbf{a}_{16})$	22. Particularities (Heart Form)
Chin	26. Width	$(1/t)\sum_{i=1}^{t} d(\mathbf{k}_i, \mathbf{l}_i), (\mathbf{k}_1 = \mathbf{c}_4, \mathbf{k}_t = \mathbf{c}_5)$	23. Width (Small and Large)
	27. Height	$d(\mathbf{b}_5, \mathbf{a}_{17})$	24. Height (Short, Average, and Large)
Ears	28. Length$_L$	$d(\mathbf{a}_{20}, \mathbf{a}_{21})$	
	29. Length$_R$	$d(\mathbf{a}_{18}, \mathbf{a}_{19})$	
	30. Angle$_L$ between corners	$\alpha(\mathbf{a}_{20}, \mathbf{a}_{21})$	
	31. Angle$_R$ between corners	$\alpha(\mathbf{a}_{18}, \mathbf{a}_{19})$	
Contours	32. Average Line Length	$d(\mathbf{a}_{17}, \mathbf{b}_3)$	

Table 1: Facial soft biometrics features grouped in *continuous* and *discrete* values. Central column is the mathematical definition of each continuous value, where d is the Euclidean distance and α the angle with respect to the horizontal line as shown in Fig. 2. All the landmarks marked here in **bold** can be visualised also in Fig. 2.

scores generated by a traditional face matcher. But in most cases, this ancillary information by itself is not sufficient to recognize a user.

More recently, Kumar *et al*. [24] developed a face recognition system based on the extraction of attribute features (soft biometrics) obtaining good results for the challenging LFW Face Database [25]. Niinuma *et al*. [26] proposed a new framework for continuous user authentication that primarily uses soft biometric traits (e.g., color of user's clothing and facial skin), scheme that continuously monitors and authenticates the logged in user. Other recent publications used comparative human descriptions for facial identification, e.g. [27], where twenty-seven comparative traits were used to accurately describe facial features. The latest related work published by Klare *et al*. [28] presented a method for using human describable face attributes to perform face identification in criminal investigations.

On the other hand, the facial landmark detection is a technological challenge widely studied in the literature [29, 30, 31]. The automatic systems can estimate the facial landmarks on a frontal face but in general, the selection of them are not completely useful for forensic analysis. These systems focus their attention on applications that do not require manual intervention in contrast to the forensic practice. For these reasons, the selection of facial landmarks are sometimes far way from the strict requirements followed by forensic examiners to carry out a fair comparison between two faces.

All these previous works focused their attention on the face recognition field from the point of view of the automatic systems. In contrast, the main difference with the previous publications is that this paper follows a forensic methodology in order to be useful to real forensic caseworks.

The paper is structured as follows. Section 2 describes the proposed methods for the extraction of the facial soft biometric features. We propose a set of measures following the practice of some forensic agencies [12, 13] and obtain two sets of features with continuous and discrete values respectively. Continuous values are converted into discrete values in order to obtain a classification of the types of facial traits present, which are used by forensic examiners in their reports and compute population statistics. Section 3 describes the materials, carrying out an analysis of the stability, discrimination power and correlation for the proposed facial features for two databases, MORPH and ATVS Forensic Database, for which we carry out a statistical analysis of the different types of facial traits found in their populations. These statistics can be very useful for forensic practitioners towards a better understanding of the facial information content. Section 4 presents the results and discussion and finally, Section 6 summarizes the contributions of this work.

2. Methods: Soft Biometric Features

This Section describes the proposed methods for the extraction of the facial soft biometric features proposed in this paper from frontal forensic mugshot images. The experimental framework implemented is based on practical protocols currently implemented manually in some international forensic laboratories [12, 13]. The proposed layout is summarized in Fig. 2) and a detailed description is given below.

This Section describes the extraction of the facial soft biometric features proposed in this paper from frontal face images. The experimental framework implemented is based on practical protocols currently implemented manually in some international forensic laboratories [12, 13]. The proposed layout is summarized in Fig. 2) and a detailed description is given below.

The facial features are grouped in 2 classes (as shown in Table 1), namely:

- ***Continuous*** features which are real valued numbers. These features are extracted using distances and angles between facial landmarks (e.g. eyebrows length, nose height and width, etc.)
- ***Discrete*** features which represent a finite number of categories. For example, the shape of the eyebrow can be arched, rectilinear or sinuous. This group of features are divided in classes using a training set and thresholds as explain below.

These *Discrete* features are selected by following a forensic categorization based on the subjective perception of the people and can be considered the main features used by forensic laboratories and in court cases. One of the reasons is that law enforcement works with the subjective information given by the victims. On the other hand *continuous* features are based on objective information directly extracted from the faces, which in the majority of cases is translated to the forensic categorization to write the report to the judges.

The facial feature extraction process can be summarized in three steps:

1. Facial landmarks extraction and normalization.
2. Facial region extraction and pre-processing.
3. Facial soft biometric features generation.

First, a set of facial landmarks ($\mathbf{a}_i$) are extracted manually by a human examiner from the raw image. These landmarks could also be extracted using an automatic system, but in this work only manual landmark extraction has been considered in order to discard the variability introduced by an automatic system in the analysis and to emulate the real task of a forensic examiner. An analysis of the variability of facial landmark extraction was reported in [32] comparing manual versus automatic landmark extraction. Results showed quite similar performance for both human and machine, obtaining even lower variability the automatic system for landmarks placed in the ocular region, but a higher variability for the automatic system for landmarks located in the mouth region.

In this work, a set of 21 facial landmarks was defined following recommendations from the Spanish Guardia Civil [12], the Netherlands Forensic Institute [13] and ENFSI [33], in a similar way as in [34].

After that, face images and landmarks are normalized using the eye centres (IPD = 75 pixels) following the ISO standard [35] obtaining the FaceISO image as shown in Fig. 3. This step eliminates variations in translation, scale and rotation in horizontal plane. After this normalization, facial regions are extracted and pre-processed, as detailed in [7].

In this work we extract three sets of facial landmarks: *i)* facial landmarks ($\mathbf{a}_i$) are the set of 21 landmarks which are manually extracted, and they can be midline landmarks (a_1, a_{13}, and a_{17}) or bilateral (the rest of a_i). Then after the pre-processing stage described above, two more sets of facial landmarks are extracted automatically: *ii)* geometrical landmarks ($\mathbf{b}_i$) and *iii)* estimated landmarks ($\mathbf{c}_i$) as shown in Fig. 2. Geometrical landmarks are automatically extracted by geometry from the facial landmarks ($\mathbf{a}_i$). These geometrical

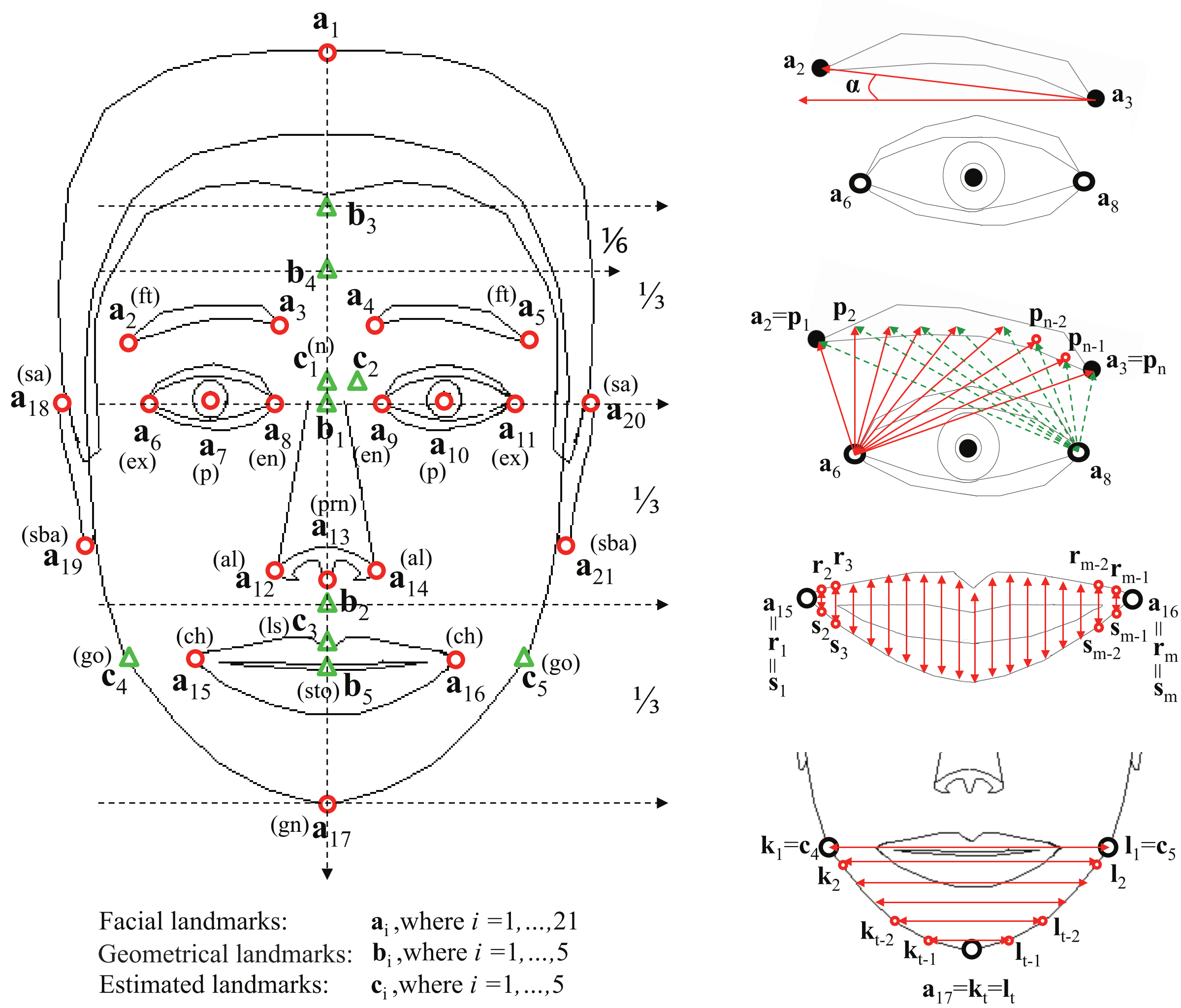


Figure 2: Left: Face layout with landmarks: facial ($\mathbf{a}_i$), geometrical ($\mathbf{b}_i$) and estimated ($\mathbf{c}_i$). Right: Visual representation of distances calculation between landmarks for computing the soft features (see Table 1). Standard anthropometric landmarks names [37] are given in brackets next to the corresponding landmarks used in this study.

landmarks do not need to be placed specifically in the midline due to asymmetries of the faces.

- $\mathbf{b}_1 = \frac{1}{2}(\mathbf{a}_8 + \mathbf{a}_9)$
- $\mathbf{b}_2 = \frac{1}{2}(\mathbf{a}_{17} + \mathbf{b}_1)$
- $\mathbf{b}_3 = \mathbf{b}_1 + (\mathbf{b}_1 - \mathbf{b}_2)$
- $\mathbf{b}_4 = \mathbf{b}_1 + \frac{2}{3}(\mathbf{b}_1 + \mathbf{b}_2)$
- $\mathbf{b}_5 = \frac{1}{2}(\mathbf{a}_{15} + \mathbf{a}_{16})$

In a similar way, the estimated landmarks ($\mathbf{c}_i$), shown in Fig. 2, are calculated automatically by using the extraction of facial regions and using image processing techniques [36] to locate them. This process uses the original facial landmarks ($\mathbf{a}_i$) as reference, but it is important to notice that the estimated landmarks can be also estimated visually by a forensic examiner based on their definition:

- $\mathbf{c}_1$ = Point of sunken part of the nose
- $\mathbf{c}_2$ = Point of half nose root width
- $\mathbf{c}_3$ = Upper point of the mouth
- $\mathbf{c}_4$ = Right limit point of chin
- $\mathbf{c}_5$ = Left limit point of chin

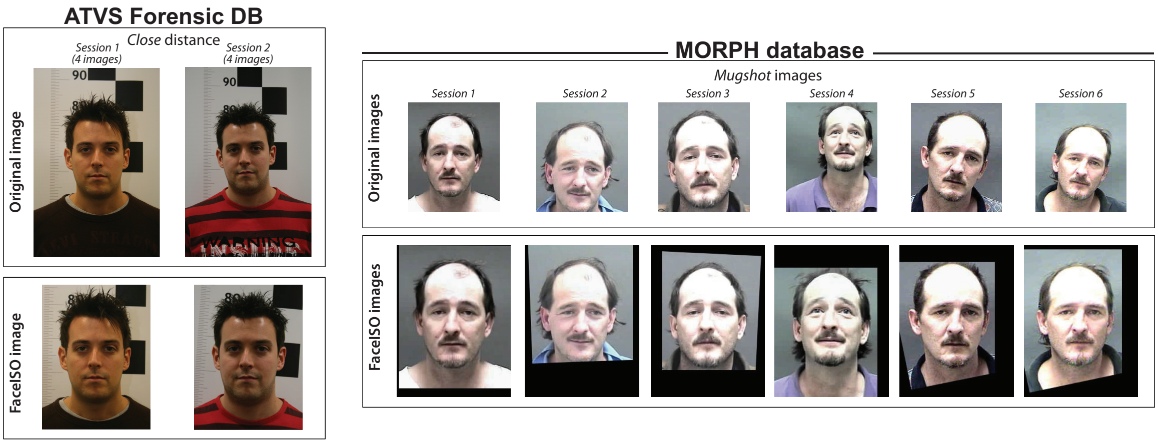


Figure 3: Databases used in the experiments. Left: ATVS Forensic DB shows an example of each session, which is comprised of four images. Right: MORPH database shows an example of all images available for one person (one per session).

Finally, using the positions of all defined landmarks, *continuous* facial features are extracted following the mathematical definitions given in Table 1. The *discrete* features (the forensic categorization) are obtained applying thresholds to each *continuous* feature based on a random selection of approx. 30% of the population of the database at hand. These thresholds are obtained uniformly between the minimum and maximum values of the corresponding continuous feature. It is important to note that *discrete* features such as 9–12, 21, and 22 need the combination of several *continuous* features to be calculated.

Fig. 2 right explains the particular cases of some features. To obtain the elevation of the eyebrows (4–8), the lower contour of each eyebrow is localized and the average distance to several points on this contour is calculated (8 points in our experiments). The average width of each eyebrow (10–11) and the mouth height (22) follow the same concept by the estimation of the upper and bottom contour of each facial trait and the calculation of the average distance across all vertical parallel segments. Finally, the chin width (26), uses the estimated landmarks ($\mathbf{c}_4$ and $\mathbf{c}_5$) for the estimation of the average distance across all the horizontal parallel segments between the right and left limits of the face.

3. Materials: Face Data Analysis

The proposed facial soft biometrics are obtained from two mugshot databases: *i*) the ATVS Forensic DB [34] and *ii*) a subset of MORPH DB [38], which contain frontal face biometrics samples from 50 and 130 subjects respectively. Fig. 3 shows examples of both databases.

The first database, ATVS Forensic DB, is captured with collaboration with DGGC [12], and is comprised of 4 frontal high resolution images (2592×3888) per subject at close distance collected in two different acquisition sessions ($50 \times 4 \times 2 = 400$ images). The second database, MORPH DB, is comprised of low resolution frontal images (200×240) captured in uncontrolled close conditions in 6 different sessions, having a total of $130 \times 6 = 780$ images.

In the next section we study the application of the proposed facial soft features to forensics, their internal correlation, stability, and discriminative power.

3.1. Population statistics

The extraction of *discrete* features allows us the analysis of the population from a statistical point of view. This means that we can automatically analyse the physical traits of the human face across a population, which is very interesting for forensics. In forensics, the examiners usually carry out a manual inspection of the face images, focussing their attention on each individual facial trait. Therefore, population statistics of such traits give the examiners a very useful information towards their decisions.

This section analyses the statistics of the populations of the two databases considered with high and low resolution images. For this purpose the distribution of each facial trait assigned to the *discrete* features across all subjects in the databases are calculated.

Fig. 4 shows the population statistics for the 24 discrete values proposed in both ATVS Forensic and MORPH DBs. The first one, ATVS Forensic DB, represents a population of 50 European subjects (32 male and 18 female), captured in an academic environment with

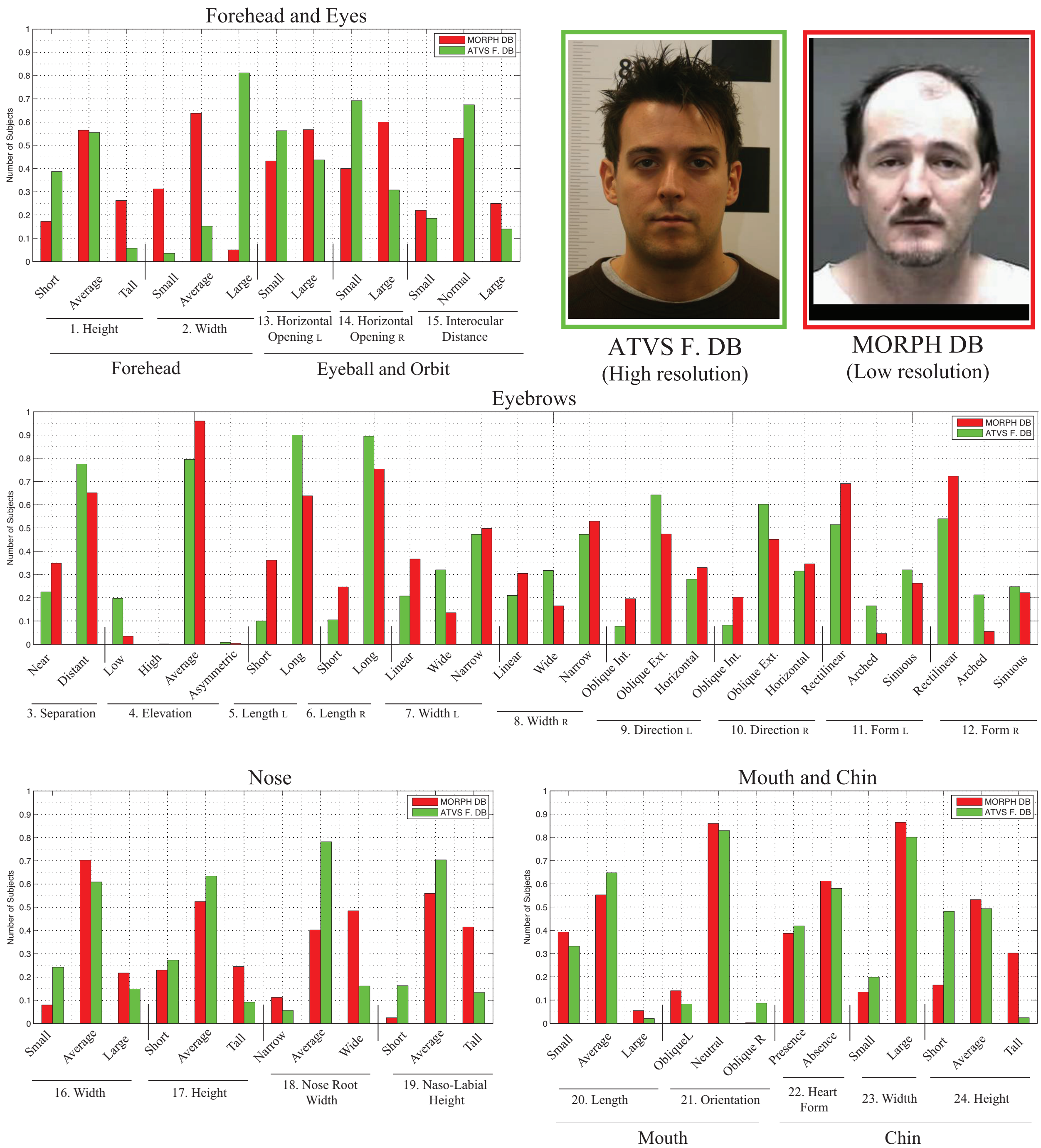


Figure 4: Population statistics comparison for ATVS Forensic DB and MORPH DB based on the *discrete* facial soft biometrics features detailed in Table 1.

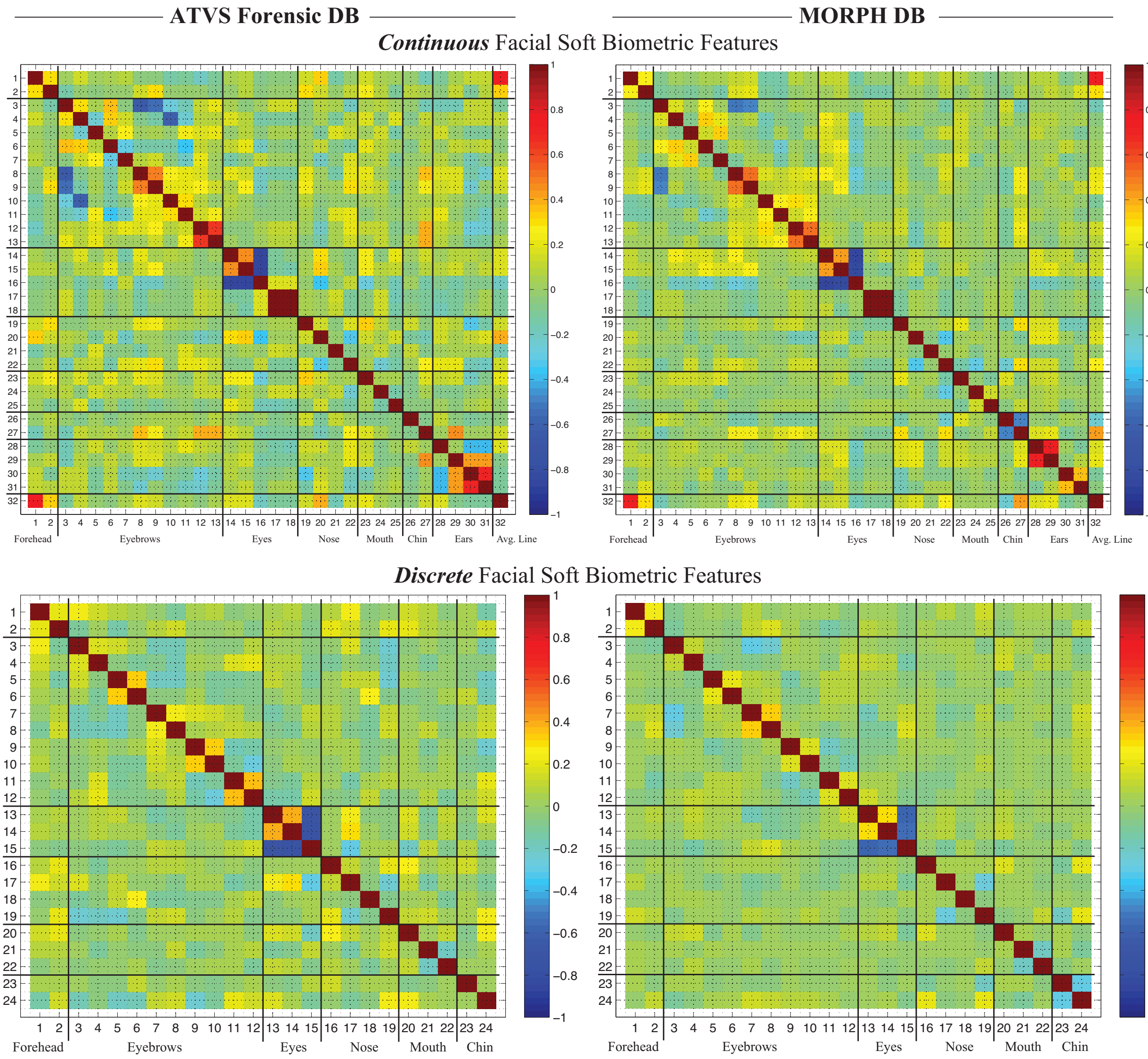


Figure 5: Correlation between *continuous* and *discrete* features based on Pearson's coefficient r (see Eq. 1) for ATVS Forensic DB (left) and MORPH DB (right). Numbering of facial soft biometrics features is detailed in Table 1.

an age range between 19–45 years. The MORPH DB was captured in a criminal environment and is also comprised of European subjects (88 male and 42 female) and the age range is between 16–60 years.

For the generation of the *discrete* features, a different set of subjects was used to obtain the thresholds between categories (approx. 30% of the population). For ATVS Forensic DB, 18 random subjects (12 male and 6 female) were used, while 42 random subjects (21 male and 21 female) were used for MORPH DB, as can be seen using both males and females from each database. We believe this selection was representative of the databases population. Thresholds were calculated uniformly between the max and min of the selected feature values in the population.

As shown in the graphs of both databases the main differences are on the forehead width (2), eyebrows width (7, 8), horizontal opening of the eyes (13, 14), nose root width (18), naso-labial height (19), and chin height (24). The other facial traits have approximately the same distribution. The resolution, population and acquisition environment differences between both databases (see Fig. 3) can explain some of these statistical differences.

It is interesting to note that for the elevation of the eyebrows most values are concentrated in the categories low and average, and the high and asymmetric categories contain very few examples. This is due to the definition of the categories, where asymmetric eyebrows means that one is higher than other, and high eyebrows means that the elevation of the eyebrow is higher than the regular distance. Therefore, in the populations analysed these two cases are practically not present.

Additionally, it is important to remark that these population differences can be useful to improve the identification tasks in forensics and also to increase the performance of face recognition systems. These differences between people into populations are a requirement exploited by forensic examiners, where automatic systems are still in a developing stage.

3.2. Correlation Between Features

This section reports an analysis of the correlation between the facial soft features defined. For this purpose the correlation between all pairs of features of the two groups defined (*continuous* and *discrete*) is computed using the Pearson's correlation coefficient:

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}} \quad (1)$$

where σ_{XY} represents the covariance of the two variables X and Y divided by the product of their standard deviations σ_X and σ_Y. The variables X and Y represent numerical values associated to a pair of facial soft features. For *discrete* features each semantic term was converted to numerical values in the range 1 to P, being P the number of divisions made for each category (e.g. short, average, and long). X_i and Y_i are the feature values across all individuals and images, therefore $N = 400$ images in ATVS Forensic DB and $N = 780$ in MORPH DB. The value r provides the correlation coefficient which ranges from -1.0 to 1.0. A value of 1.0 implies that a linear equation perfectly describes the relationship between X and Y, with all data points lying on a line for which Y increases as X increases. A value of -1.0 implies that all data points lie on a line for which Y decreases as X increases. A value of 0 implies that there is no linear correlation between the variables.

Results of the correlation between *continuous* features are presented in Fig. 5 top, where both databases are compared. Some pairs of features such as forehead height (1) and average line length (32), eyebrows length (8 and 9), eyebrows angles (12 and 13), horizontal opening of eyes (14 and 15), ears length (28 and 29), and ears angle (30 and 31) are clearly correlated in both databases as was expected. This means for example that when the forehead height increases, the average line of the face also increases. Note that there exits a notable difference in the results between both databases for the ears trait (28–31) due to the different characteristics between them.

On the other hand, there is a negative correlation in pairs of features such as eyebrows separation (3) and length (8,9), eyebrows elevation (4) and width (10), and a remarkable negative correlation between horizontal opening of both eyes (14,15) and interocular distance (16). This means that for example when the eyebrows separation increases, their length decreases. Note an important inverse correlation between chin width (26) and height (27) in MORPH DB that denotes a singularity in this population analysed.

In the same way, the correlation between *discrete* soft biometric features has been analysed and is shown in Fig. 5 bottom. Again a positive correlation between some pairs of facial features such as eyebrows length (5 and 6), eyebrows direction (8 and 9), eyebrows form (11, 12), horizontal opening of eyes (13 and 14) and the nose height (17) can be observed.

The highest negative correlations are between horizontal opening of the eyes (13,14) and interocular distance (15), mouth orientation (21) and mouth heart form (22). Another interesting negative correlation in ATVS Forensic DB is between interocular distance (15) and nose height (17), i.e., as the interocular distance increases, the height of the nose decreases.

3.3. Stability Analysis

This section reports an analysis of the stability of the *continuous* and *discrete* features defined. This is done by calculating the stability coefficient:

$$\text{Stability}_X = 1 - \frac{1}{SM} \sum_{i=1}^{S} \sum_{m=1}^{M} |X_{im} - \text{mode}_m(X_{im})| \quad (2)$$

where X_{im} is the feature value for subject i from its sample image m, $M = 8$ or $M = 6$ is the total number of sample images per subject for ATVS Forensic DB and MORPH DB, respectively; $S = 50$ or $S = 130$ is the total number of subjects, also respectively, and $\text{mode}_m(X_{im})$ is the statistical mode across m (i.e., the feature value most often present for subject i). In the case of *continuous* features the statistical mean instead of the mode used: $\text{mean}_m(X_{im})$ across m values (i.e., the mean value for a given subject i).

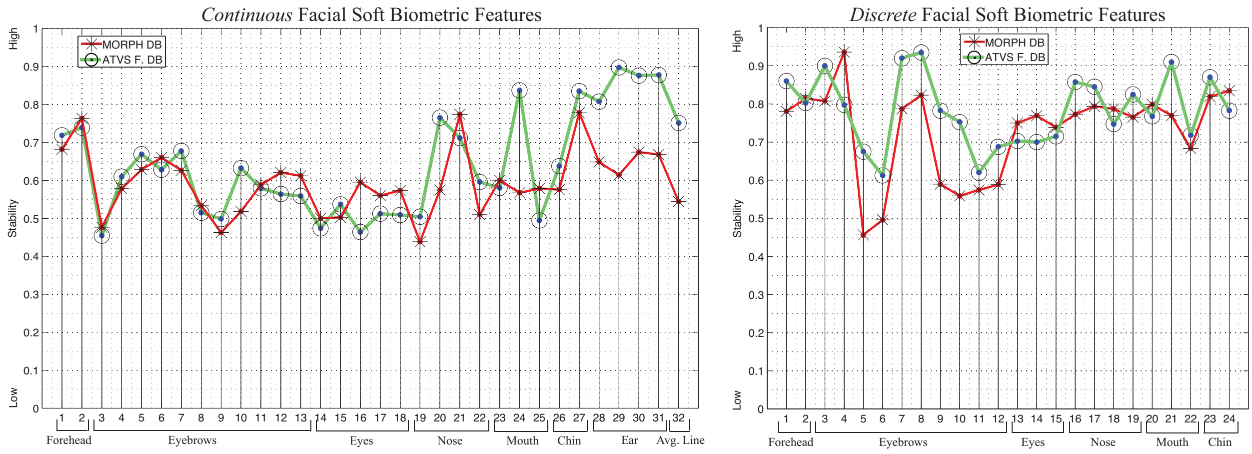


Figure 6: Stability analysis for the 32 *continuous* and 24 *discrete* facial soft biometrics features considered for both databases (see Table 1).

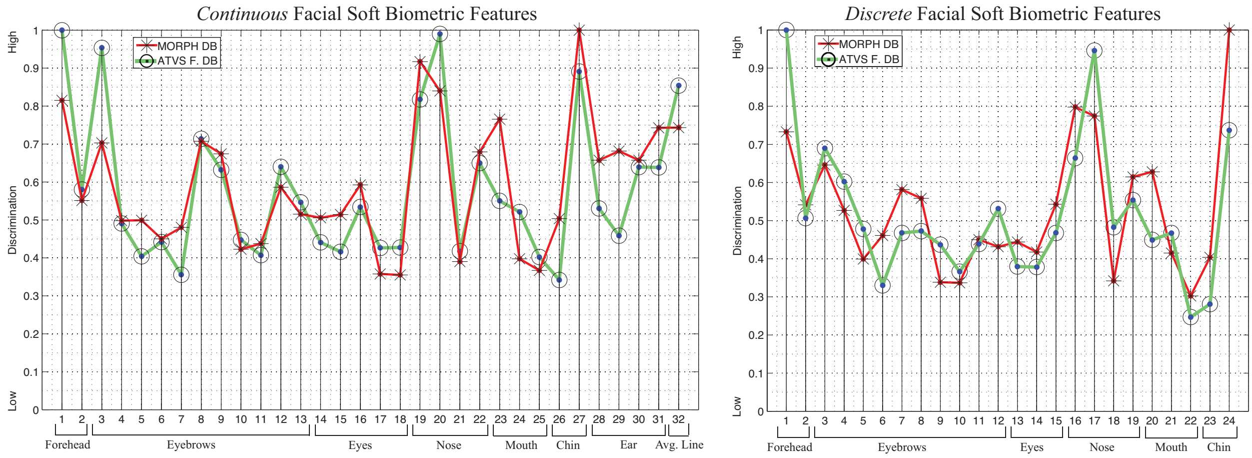


Figure 7: Discrimination power of the 32 *continuous* and 24 *discrete* facial soft biometrics features considered for both databases (see Table 1).

The resulting stability coefficients for all facial soft biometric features are depicted in Fig. 6. In this figure (left) the *continuous* features show that for MORPH DB the nose (21), forehead width (2), and chin height (27) are the most stable features, due mainly to the normalization process. Face images are normalized based on the distance between the eye centres therefore the real horizontal width of some traits is lost. On the other hand, ATVS Forensic DB presents stable features such as the ears features (28–31) as a consequence of the higher quality of the images.

The stability results of *discrete* features are shown in Fig. 6 (right). Again small differences between both databases are observed. In general, forehead height (1), eyebrows traits (3-12), and nose width (16) and height (17) in ATVS Forensic DB have better stability than in MORPH DB.

As observed in Fig. 6, both databases (high and low resolution) present approximately the same stability in the facial soft biometric features extracted. This demonstrates the value of the proposed features.

In this paper these two types of features (*continuous* and *discrete*) will be processed differently (e.g., using different similarity measures), and together in order to study the potential of both of them jointly.

3.4. Discriminative Power Analysis

In order to evaluate the discriminative power of the facial soft biometric feature X, we compute for it the ratio between the inter-subject variability, and the intra-subject variability as follows:

$$\text{Discrimination}_X = \frac{\frac{1}{S(S-1)} \sum_{i=1, i \neq j}^{S} \sum_{j=1}^{S} |\mu_i - \mu_j|}{\sigma},$$
$$\mu_i = \underset{m}{\text{mean}}(X_{im}),\ \mu_j = \underset{m}{\text{mean}}(X_{jm}), \quad (3)$$
$$\sigma = \frac{1}{S} \sum_{i=1}^{S} \sigma_i,\ \sigma_i = \underset{m}{\text{std}}(X_{im})$$

where i and j index subjects, and m indexes images for a given subject.

The discrimination coefficient for the X features is depicted in Fig. 7 ($k = \{1, ..., K\}$, $K = 32$ or $K = 24$, for *continuous* or *discrete* values). There we can see that for *continuous* and *discrete* features the eyebrows and eyes traits are less discriminant than the nose and the forehead, contrary to what is obtained in other face recognition research [7].

The least discriminant *continuous* facial soft features are the right eyebrow outer elevation (7) and the chin width (26) in ATVS Forensic DB, while the eyes angles between corners (17-18) and mouth angles (25) for MORPH DB. In contrast, the least discriminant *discrete* features are the mouth heart form (22) and chin width (23), and the eyebrows direction (9-10), respectively for ATVS Forensic DB and MORPH DB.

4. Results and Discussion

4.1. Experimental Protocol

The facial soft features described before (*continuous* and *discrete*) have been obtained from two databases: ATVS Forensic DB and MORPH DB as described previously. The objective in our experiments with these features is to understand their behaviour and their best application to forensics and face recognition.

The experimental protocol followed is based on a cross-validation method (leave-one-out approach). The leave-one-out approach that we have implemented first divides the data of a given subject by selecting a varying number of training reference samples and one of the remaining samples not used for training is left out for testing. The sets of features extracted from the reference samples form the enrolled template for a given user as per Fig. 1. We then iterate by rotating the selected training samples a number of times equal to the total quantity of samples minus one (i.e., 7 in ATVS Forensic DB and 5 in MORPH DB).

Regarding the person recognition system based only on facial soft biometric features, first, each *continuous* or *discrete* feature in numeric form is normalised to the range [0,1] using the tanh-estimators described in [39]:

$$\hat{X}^k = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{X^k - \mu_{X^k}}{\sigma_{X^k}}\right)\right) + 1\right\} \quad (4)$$

where X^k is the $k = \{1, ..., K\}$ soft feature, $\hat{X}^k$ denotes the normalized feature, and μ_{X^k} and σ_{X^k} are respectively the estimated mean and standard deviation of the feature under consideration of the training set. Note that, depending on the features considered (*continuous* or *discrete*), there are $K = 32$ or 24 facial features, respectively (see Table 1).

In this study three different similarity measures based on various distances [40] are compared: *i*) Mahalanobis, *ii*) Euclidean, and *iii*) Hamming.

Similarity scores based on the Mahalanobis distance between the test vector with K features $\mathbf{x} = [\hat{X}^1, ..., \hat{X}^K]^T$ and a statistical model of the client being compared C are computed as follows:

$$s_M(\mathbf{x}, C) = \frac{1}{\left((\mathbf{x} - \boldsymbol{\mu}^C)^T\, (\boldsymbol{\Sigma}^C)^{-1}\, (\mathbf{x} - \boldsymbol{\mu}^C)\right)^{1/2}} \quad (5)$$

where $\boldsymbol{\mu}^C$ and $\boldsymbol{\Sigma}^C$ are respectively the mean vector and covariance matrix obtained from the training features, which form the statistical model of the client.

Similarity scores based on the Euclidean distance are computed as follows:

$$s_E(\mathbf{x}, C) = -\frac{1}{M} \sum_{i=1}^{M} \left((\mathbf{x} - \mathbf{y}_i)^T\, (\mathbf{x} - \mathbf{y}_i)\right)^{1/2} \quad (6)$$

where $\mathbf{y}_i$ are the M training vectors corresponding to the subject at hand.

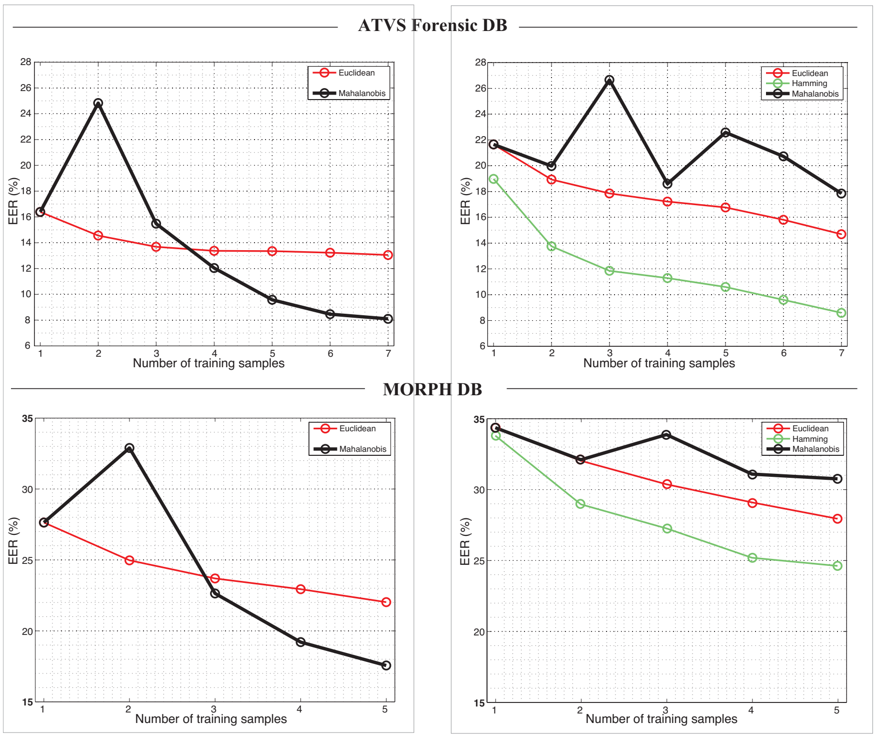
The similarity measure based on Hamming distance is computed as:

$$s_H(\mathbf{x}, C) = -\frac{1}{MK} \sum_{i=1}^{M} \#_k\{\hat{X}^k \neq \hat{Y}_i^{\,k}\} \quad (7)$$

where $\mathbf{x} = [\hat{X}^1, \ldots, \hat{X}^K]^T$, $\mathbf{y}_i = [\hat{Y}_i^{\,1}, \ldots, \hat{Y}_i^{\,K}]^T$ are the M training vectors corresponding to subject C, and $\#_k\{condition\}$ indicates the number of cases across k where the *condition* holds.

It is important to note that the Hamming distance only makes sense with *discrete* features and it will not be applied to *continuous* features.

Finally, system performance results are presented using common measures used by the biometric community [43]. In a biometric authentication system there are two types of errors that can be made: *i)* False Acceptance and *ii)* False Rejection. False Acceptance (or



Continuous Facial Soft Biometric Features   ***Discrete*** Facial Soft Biometric Features

Figure 8: EER (%) obtained when varying the number of training samples for the three sets of features considered: *continuous* and *discrete*. On top ATVS Forensic DB, and on the bottom MORPH DB. Note the EER range is different in both databases.

false match) occurs when the system mistakes the face images under analysis that are from different persons to be from the same person. On the other hand, a False Rejection (or false nonmatch) occurs when the system mistakes the face images under analysis that are from the same person to be from two different persons. There is a tradeoff between False Acceptance Rate (FAR) and False Rejection Rate (FRR) in every biometric system. In fact, both FAR and FRR are functions of the system threshold θ; if θ is decreased to make the system more tolerant to input variations and noise, then FAR increases. On the other hand, if θ is raised to make the system more secure, then FRR increases accordingly. The system performance at all the operating points (thresholds θ) can be depicted in the form of a Receiver Operating Characteristic (ROC) curve. A ROC curve is a plot of the authentication rate (1-FRR) against FAR for various threshold values (see Fig. 9).

A popular performance measure is the Equal Error Rate (EER), which is the error rate at the operation point where the False Acceptance Rate and the False Rejection Rate have the same value. The lower the EER the higher the accuracy of the biometric system.

Additionally, CMC curves (Cumulative Match Curves) are computed for the best systems configurations giving match results for ranks 1, 5 and 10.

4.2. Experimental Results

4.2.1. Analysis of Training Set Size for Soft Features

An important parameter in soft biometric systems is the size of the training set. We have evaluated the system with different number of training samples following a cross-validation methodology.

Fig. 8 shows the Equal Error Rate (EER) for the different configurations analysed for the three sets of features: *i*) *continuous*, *ii*) *discrete*, and *iii*) *fusion* of features, which analyses both of them together. The figure shows results for the three different similarity distances defined: Euclidean, Hamming (only *discrete* features), and Mahalanobis.

As can be seen, all feature sets follow the same trend, i.e., the system recognition performance (EER) for Mahalanobis improves significantly when more samples are used in the training stage as was expected as this similarity distance is based on a statistical model of the client, so the more samples are contained in the model

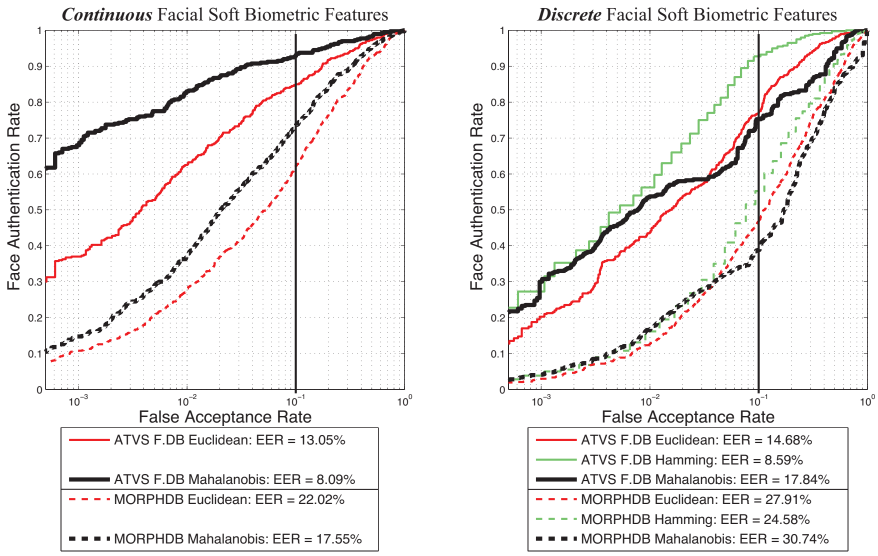


Figure 9: ROC curves obtained for the facial soft biometric features sets: *continuous* and *discrete*.

the better the performance. However, in the case of Euclidean distance in the *continuous* set using more than 3 training samples the system performance saturates in both databases. It is also important to note the increment of EER in the Mahalanobis system when few 3 training samples are considered.

In contrast, the Hamming distance achieves the best results on the *discrete* features in both databases. This Hamming distance achieves a relative improvement of 12-24% and 60-70% for MORPH and ATVS Forensic DBs respectively, compared to the Euclidean and Mahalanobis distances. Therefore the Hamming distance is the most adequate for the *discrete* features.

It is also interesting to note the difference of performance in both databases. This difference could be explained by factors such as the size of the populations, the environmental acquisition (cooperative and non-cooperative), the resolutions of both databases, etc.

Finally, in Fig. 9 we show ROC curves when training with $M = 7$ and 5 samples for each database respectively.

4.2.2. Analysis of Individual Facial Soft Features

This section reports the discrimination power of each individual facial soft feature following the same leave-one-out experimental protocol as before with all sessions available. As shown in Fig. 10 (left), the *continuous* set of soft features in both databases follow the same trend but the system performance of ATVS Forensic DB is around 10% better (in absolute EER) for all features. This is mostly due to the difference of resolution and acquisition environment between both databases, which affects directly the precision of the landmark tagging process.

The forehead height (1) followed by nose height (20) and chin height (27) achieve the best results (EER< 25% and EER< 35% in both databases respectively) and it is worth noting that these were the most discriminative features in the initial experiments shown in Fig. 7. Another relevant feature with a high performance and discrimination power is the nose width (19) with an EER of 28.84% and 35.85%, respectively for ATVS Forensic and MORPH DBs.

Discrete facial features achieve similar performance results than *continuous* features. Again individual facial features on ATVS Forensic DB achieved better performance than features on MORPH DB as expected. The forehead height (1) and nose height (17) continue obtaining the best performance results. The worst performance results are obtained for the eyebrows elevation (4) in MORPH DB and eyebrow left length (5) in the case of ATVS Forensic DB, results previously predicted in Fig. 7. The remaining facial soft features in both feature sets achieve higher performance, with better results in general for forehead and nose features compared to eyebrows or eye features, as anticipated in Sect. 3.4. As can be seen, individual facial features are not very discriminative on their own.

4.2.3. Analysis of Feature Selection: SFFS

In order to find the most discriminative set of facial soft biometrics features, and therefore increase the

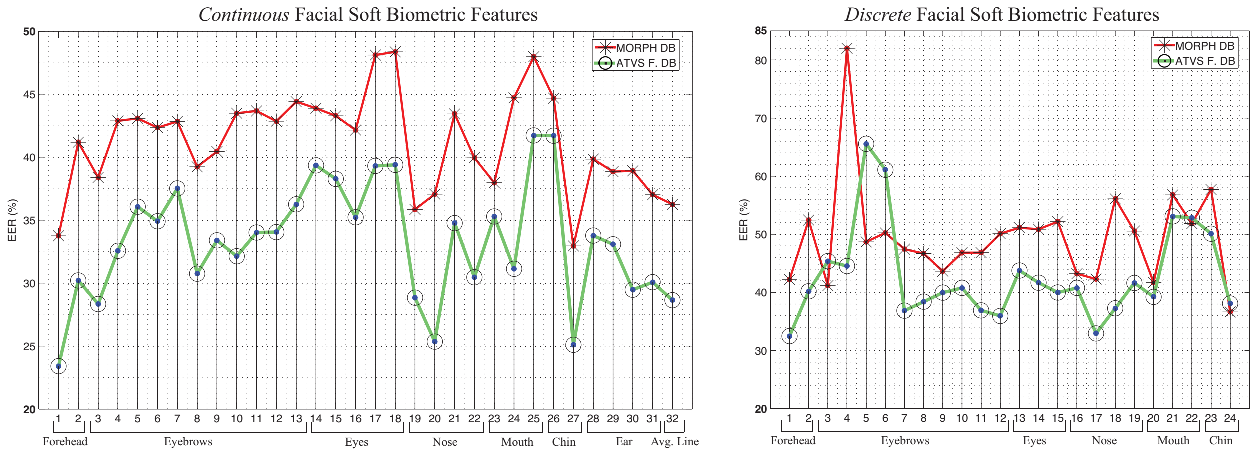


Figure 10: Average EER (%) obtained for each individual facial soft biometric features (32 *continuous* and 24 *discrete*) defined in Table 1. Average EER calculated between the three difference distances considered: Mahalanobis, Hamming, and Euclidean. The Hamming distance is not considered to compute the results of the *continuous* features.

Database	Distance	SFFS Feature Selection	# Selected Features	EER (%)	EER (%) All Features
ATVS F. DB	Euclidean	(**27**,**20**,**3**,31,19,9,30,16,22,23,32,12,25,13)	14	7.38	13.05
	Mahalanobis	(**27**,**3**,31,**20**,30,19,8,23,9,32,12,28,13,25)	14	**3.70**	8.09
MORPH DB	Euclidean	(**19**,**20**,**27**,3,1,31,23,9,30,22,28,4,8,16,32,13)	16	16.50	22.02
	Mahalanobis	(**27**,1,**19**,31,**20**,23,29,30,3,22,9,15,12,32,28,6)	16	**14.01**	17.55

Table 2: SFFS selected **continuous** features (defined in Table 1) for each system analysed. The three most discriminative features in Fig. 7 (left) are in **bold** for each database.

Database	Distance	SFFS Feature Selection	# Selected Features	EER (%)	EER (%) All Features
ATVS F. DB	Euclidean	(**17**,**1**,**24**,18,19,12,3,20,22,16,7,13,4,14)	14	10.34	14.68
	Hamming	(**17**,**1**,15,8,18,9,3,11,16,**24**,20,12,4,6,14,21,7,22,19,2,23,13,10)	23	**7.84**	8.59
	Mahalanobis	(**1**,**17**,**24**,16,19)	5	13.34	17.84
MORPH DB	Euclidean	(**24**,**17**,20,1,**16**,3,7,5,13,15,8)	11	21.78	27.91
	Hamming	(**24**,**17**,3,20,**16**,9,1,10,13,5,18,4,8,7,22)	15	**20.89**	24.58
	Mahalanobis	(**24**,**16**,**17**,1)	4	24.95	30.74

Table 3: SFFS selected **discrete** features (defined in Table 1) for each system analysed. The three most discriminative features in Fig. 7 (right) are in **bold** for each database.

performance of the biometric system, a feature selection process is performed. In addition, the reduction of the number of features decreases the computational cost too.

Among the different feature selection algorithms [41], the one employed in this work is the Sequential Floating Forward Selection (SFFS) [42]. This suboptimal searching technique is an iterative process in which, in each iteration, a new set of features (whose choice is based on the results of previous subsets) is used to compute a certain criterion. This is done until the criterion does not improve. For more details see [42, 41, 40]. In our case the criterion is related to the performance of the system, in particular, it is to minimize the value of the EER.

Once the features are selected, the feature vector has, depending on the experiment, between 4 and 23 components. The SFFS algorithm is able to provide the most discriminative set of features with a dimension specified by the user or with the dimension that gives the best value for the criterion (in that case the dimension is not specified). The latter approach was performed in our system. Tables 2 and 3 summarize the results after applying the SFFS algorithm to each system for both databases.

As shown in Table 2 in boldface, the most frequently selected *continuous* features across the databases are: eyebrows separation (3), nose width (19), nose height (20), and chin height (27), which correspond to some of the most discriminative features individually (see

Database		EER (%) All Features Results			EER (%) SFFS Results		
ATVS F. DB	Continuous	8.09 (32 Feat. s_M)			3.70 (14 Feat. SFFS s_M)		
	Discrete	8.59 (24 Feat. s_H)			7.84 (23 Feat. SFFS s_H)		
	Fusion	F. Sum	F. Prod	F. Weight	F. Sum	F. Prod	F. Weight
		5.59	5.59	**4.90**	3.95	4.00	**3.06**
MORPH DB	Continuous	17.55 (32 Feat. s_M)			14.01 (16 Feat. SFFS s_M)		
	Discrete	24.58 (24 Feat. s_H)			20.89 (15 Feat. SFFS s_H)		
	Fusion	F. Sum	F. Prod	F. Weight	F. Sum	F. Prod	F. Weight
		16.63	16.62	**15.86**	14.20	14.20	**12.27**

Table 4: Fusion results of the best systems in Fig. 9 and SFFS results in Tables 2 and 3 for the continuous ($s_{Mahalanobis}$) and discrete ($s_{Hamming}$) features for ATVS Forensic and MORPH DBs.

Fig. 7 left). The feature selection approach achieves relative improvements of 118.64% and 25.26% of EER for ATVS Forensic and MORPH databases, respectively, when considering the Mahalanobis distance, resulting in a overall performance of 3.70% of EER for a system with 14 *continuous* features on ATVS Forensic DB.

Similar results can be seen for *discrete* features in bold in Table 3: the most discriminative features individually are always in the SFFS selected feature sets: forehead height (1), nose width (16), nose height (17), and chin height (24). The Hamming distance obtained the best overall results with 7.84% of EER on ATVS Forensic DB.

4.2.4. Score Fusion of Continuous and Discrete Features

This section describes the score-level fusion of both *continuous* and *discrete* facial soft biometric features in order to increase the system recognition performance. For these fusions we have selected the best system in each set of features, Mahalanobis-based and Hamming-based system, respectively for *continuous* and *discrete* sets. The comparison between all features performance and the most discriminative features selected by SFFS is also presented.

Three different score-level fusion rules have been evaluated: (i) sum, (ii) product, and (iii) weighted sum fusion. The weighted sum fusion gives more weight to the most robust system, which is the *continuous* system based on the EER of the systems to be fused. For the experiments, weights of 70% and 30% have been used based on the EER performance of the individual systems.

Table 4 shows the fusion results for these three different fusion rules. As we can see for all the fusions the best individual system is improved, thus this demonstrates how different similarity measures applied to different features can improve the system performance.

The best results are achieved using a weighted fusion in both databases using the most discriminant features obtained by SFFS in the previous section. This configuration achieved a final system performance of 3.06% and 12.27% of EER for ATVS Forensic and MORPH databases respectively.

Finally, we show the results obtained in a biometric identification scenario (1 to N comparisons), for the best configuration obtained using the best selected features (SFFS algorithm) and fusion of systems (as per Table 4). Results are shown using CMC curves in Fig. 11, giving also results for ranks 1, 5 and 10. For the case of ATVS Forensic DB the weighted fusion achieves the best results achieving rank 1 results of 48,75% identification rate, and 100% from rank 5. For the case of MORPH DB the weighted fusion also achieves the best results achieving rank 1 results of 23,84% identification rate, 62,05% for rank 5 and 75,13% for rank 10. Therefore the identification power of these proposed facial soft biometric features is confirmed.

Based on the popular *Rule of 30* [44] used for determining the required size of a corpus, which states that "to be 90% confident that the true error rate is within 17% of the observed error rate, there must be at least 30 errors". We can confirm that we comply with this rule in all results reported in this paper for the two databases used.

5. Conclusions

This paper highlights some of the problems and challenges in the field of forensic face recognition. Contrary to standard automated face recognition, forensic face recognition offers a set of tools that can help investigators to narrow the identity of a subject, but not fully perform the identification.

There has been a substantial improvement in the capabilities of forensic face recognition as a result of ongoing studies on facial aging, facial marks, sketch to photo matching, video based face recognition, and NIR image to photo match. However, many challenging problems related to forensic face recognition still exist,

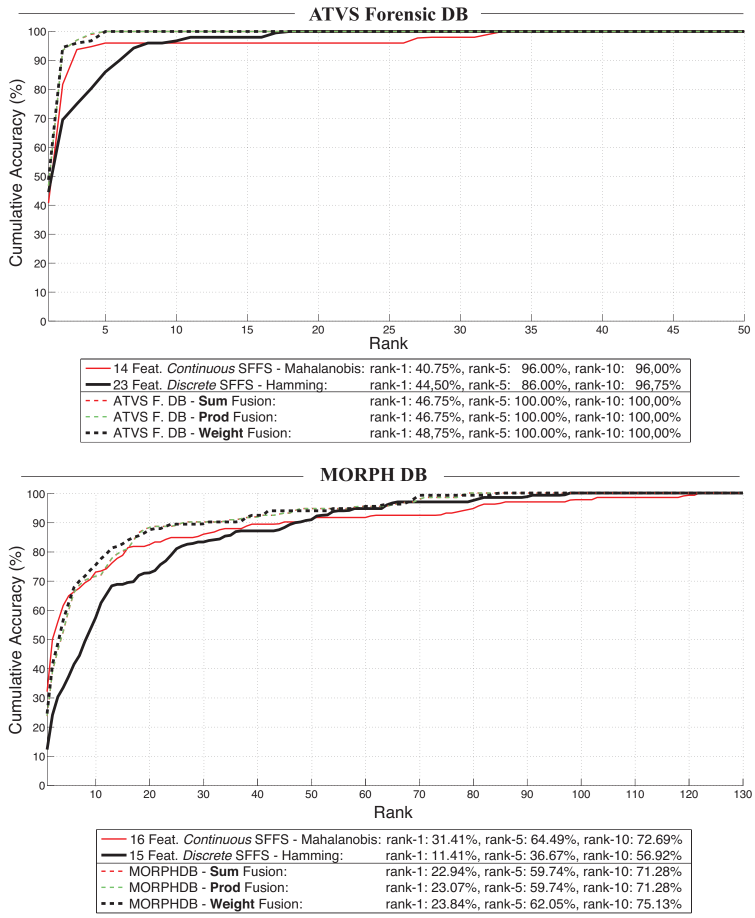


Figure 11: CMC curves obtained for the fusion of the systems after SFFS algorithm for ATVS Forensic DB (top) and MORPH DB (bottom).

which offer excellent opportunities to face recognition researchers.

The motivation of this work is to provide a set of facial features which can be understood by non-experts such as judges and support the work of forensic examiners who, in practice, carry out a thorough manual comparison of face images paying special attention to the similarities and differences in shape and size of various facial traits. In this context, this work proposes a wide set of facial soft biometric anthropometric features for forensics that has been described and evaluated. These features are extracted following forensic protocols based on the forensic anthropometrical and morphological analysis. *Continuous* and *discrete* facial features are extracted and analysed.

The correlation, stability, and discriminative power of the proposed facial soft features have been studied and evaluated. The experimental results have shown that a system completely based on facial soft biometric features for forensics can provide good accuracy in person recognition tasks. Additionally, these facial soft biometric traits can be used to improve the performance of face recognition systems.

Moreover, four different analyses have been carried out: training set size, individual and grouped performance, and feature selection.

One of the benefits of the proposed set of features is that some of them can be extracted even from very low quality images such as those present in forensic scenarios with faces at a distance. An interesting finding is that features extracted from the nose and forehead regions are consistently more discriminative compared to features extracted from the eye and eyebrow regions.

Finally, some fusion rules have been studied to improve the recognition performance of the proposed features. Experiments are carried out considering all features and just the most discriminative ones. Results have shown the benefits of the proposed facial soft biometric features in forensic scenarios.

References

[1] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (10) (1993) 1042–1052.

[2] M. A. Grudin, On internal representations in face recognition systems, Pattern Recognition 33 (7) (2000) 1161–1177.

[3] B. Heisele, P. Ho, J. Wu, T. Poggio, Face recognition: component-based versus global approaches, Computer Vision and Image Understanding 91 (1-2) (2003) 6–21.

[4] R. Jafri, H. R. Arabnia, A survey of face recognition techniques., Journal of Information Processing Systems 5 (2) (2009) 41–68.

[5] K. Bonnen, B. Klare, A. Jain, Component-based representation in automated face recognition, IEEE Transactions on Information Forensics and Security 8 (1) (2013) 239–253.

[6] G. Givens, J. Beveridge, P. Phillips, B. Draper, Y. Lui, D. Bolme, Introduction to face recognition and evaluation of algorithm performance, Computational Statistics and Data Analysis 67 (C) (2013) 236–247.

[7] P. Tome, J. Fierrez, R. Vera-Rodriguez, D. Ramos, Identification using face regions: Application and assessment in forensic scenarios, Forensic Science International (FSI) 233 (2013) 75–83.

[8] P. Tome, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Fusion of facial regions using color information in a forensic scenario, in: 18th Iberoamerican Congress on Pattern Recognition (CIARP2013), LNCS 8259, Habana, Cuba, 2013, pp. 399–406.

[9] U. Park, A. K. Jain, Face matching and retrieval using soft biometrics, Trans. Info. For. Sec. 5 (3) (2010) 406–415. doi:10.1109/TIFS.2010.2049842.

[10] A. Bertillon, Signaletic Instructions including the theory and practice of Anthropometrical Identification, R.W. McClaughry Translation,The Werner Company, 1896.

[11] H. Rhodes, Alphonse Bertillon: Father of Scientific Detection, Abelard-Schuman, New York, 1956.

[12] Spanish Guardia Civil (DGGC), http://www.guardiacivil.es/.

[13] Netherlands Forensic Institute (NFI), http://www.forensicinstitute.nl.

[14] B. Klare, A. Jain, On a taxonomy of facial features, in: Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), 2010, pp. 1–8.

[15] A. Bertillon, The Bertillon System of Identification, Chicago, 1896.

[16] Facial Identification Scientific Working Group, Guidelines for Facial Comparison Methods, Tech. Report, 2012.

[17] G. Edmond, K. Biber, R. Kemp, G. Porter, Laws looking glass: Expert identification evidence derived from photographic and video images, Current Issues in Criminal Justice 20 (3) (2009) 337–377.

[18] A. K. Jain, K. N, X. Lu, U. Park, Integrating faces, fingerprints and soft biometric traits for user recognition, in: In Proceedings of Biometric Authentication Workshop, LNCS 3087, Springer, 2004, pp. 259–269.

[19] A. K. Jain, S. C. Dass, K. Nandakumar, K. N, Soft biometric traits for personal recognition systems, in: Proceedings of International Conference on Biometric Authentication, Hong Kong, 2004, pp. 731–738.

[20] D. Heckathorn, R. Broadhead, B. Sergeyev, A methodology for reducing respondent duplication and impersonation in samples of hidden populations, in: Proc. of Annual Meeting of the American Sociological Association, Toronto, Canada, 1997, pp. 543–564.

[21] A. K. Jain, U. Park, Facial marks: Soft biometric for face recognition, in: IEEE International Conference on Image Processing, 2009, pp. 37–40. doi:10.1109/ICIP.2009.5413921.

[22] J. Lee, A. K. Jain, R. Jin, Scars, marks and tattoos (smt): Soft biometric for suspect and victim identification, in: In Proc. Biometric Symposium, Biometric Consortium Conference, 2008, pp. 1–8.

[23] P. Tome, J. Fierrez, R. Vera-Rodriguez, M. S. Nixon, Soft biometrics and their application in person recognition at a distance, IEEE Transaction on Information Forensics and Security.

[24] N. Kumar, A. C. Berg, P. N. Belhumeur, S. K. Nayar, Attribute and simile classifiers for face verification, in: In IEEE International Conference on Computer Vision (ICCV, 2009.

[25] G. B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007).

[26] K. Niinuma, U. Park, A. K. Jain, Soft biometric traits for continuous user authentication, Transaction on Information, Forensics and Securisty 5 (4) (2010) 771–780.

[27] D. Reid, M. Nixon, Human identification using facial comparative descriptions, in: IAPR International Conference on Biometrics, ICB, 2013.

[28] B. F. Klare, S. Klum, J. Klontz, E. Taborsky, T. Akgul, A. K. Jain, Suspect identification based on descriptive facial attributes, in: International Joint Conference on Biometrics (IJCB), 2014.

[29] J. M. Saragih, S. Lucey, J. F. Cohn, Deformable model fitting by regularized landmark mean-shift, Int. J. Comput. Vision 91 (2) (2011) 200–215.

[30] D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CVPR '12, IEEE Computer Society, Washington, DC, USA, 2012, pp. 2879–2886.

[31] Z. Zhang, P. Luo, C. Loy, X. Tang, Facial landmark detection by deep multi-task learning, in Proc. of European Conference on Computer Vision (ECCV), Vol. 8694 of Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 94–108.

[32] R. Vera-Rodriguez, P. Tome, J. Fierrez, J. Ortega-Garcia, Comparative Analysis of the Variability of Facial Landmarks for Forensics Using CCTV Images, in Proc. 6th Pacific-Rim Symposium on Image and Video Technology (PSIVT), 2013.

[33] European Network of Forensic Science Institutes, http://enfsi.eu/.

[34] R. Vera-Rodriguez, P. Tome, J. Fierrez, N. Exposito, F. J. Vega, Analysis of the variability of facial landmarks in a forensic scenario, in: International Workshop on Biometrics and Forensics, Lisboa, Portugal, 2013.

[35] ISO/IEC 19794-5:2011, Information technology – biometric data interchange formats – part 5: Face image data, in: International Organization for Standardization, 2011.

[36] R. C. Gonzalez, R. E. Woods, Digital Image Processing (3rd Edition), Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[37] L. G. Farkas, Anthropometry of the Head and Face, Raven Press, 2, 1994.

[38] K. Ricanek, T. Tesafaye, Morph: a longitudinal image database of normal adult age-progression, in: Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, 2006, pp. 341 –345.

[39] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (12) (2005) 2270–2285.

[40] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Fourth

Edition, 4th Edition, Academic Press, 2008.
[41] L. C. Molina, L. Belanche, A. Nebot, Feature selection algorithms: A survey and experimental evaluation, in: IEEE International Conference on Data Mining, IEEE Computer Society, 2002, pp. 306–312.
[42] P. Pudil, J. Novovičová, J. Kittler, Floating search methods in feature selection, Pattern Recognition Letters 15 (11) (1994) 1119–1125.
[43] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Transactions on Circuits and Systems for Video Technology, 14 (1) (2004) 4–20.
[44] G. Doddington, M. Przybocki, A. Martin, D. Reynolds, The NIST speaker recognition evaluation - Overview, Methodology, Systems, Results, Perspective, Speech Communication 31 (2-3), (2000) 225–254.