Soft Biometrics and Their Application in Person Recognition at a Distance

Pedro Tome, Julian Fierrez, Ruben Vera-Rodriguez, and Mark S. Nixon

Abstract—**Soft biometric information extracted from a human body (e.g., height, gender, skin color, hair color, and so on) is ancillary information 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. One significant example is visual surveillance, where face images are usually captured in poor quality conditions with high variability and automatic face recognition systems do not work properly. In this scenario, the soft biometric information can provide very valuable information for person recognition. This paper presents an experimental study of the benefits of soft biometric labels as ancillary information based on the description of human physical features to improve challenging person recognition scenarios at a distance. In addition, we analyze the available soft biometric information in scenarios of varying distance between camera and subject. Experimental results based on the Southampton multibiometric tunnel database show that the use of soft biometric traits is able to improve the performance of face recognition based on sparse representation on real and ideal scenarios by adaptive fusion rules.**

Index Terms—**Soft biometrics, labels, primary biometrics, face recognition, at a distance, on the move.**

I. INTRODUCTION

A WIDE variety of biometric systems have been developed for automatic recognition of individuals based on their physiological/behavioural characteristics. These systems make use of a single or a combination of traits like face, gait, iris, etc., for recognizing a person. On the other hand, the use of other ancillary information based on the description of human physical features for face recognition [1] has not been explored in much depth.

Biometric systems at a distance have an outstanding advantage: they can be used when images are acquired non-intrusively at a distance and other biometric modes such as iris cannot be acquired properly. Given such situations, some biometrics may have a severe degradation of performance due to variability factors caused by the acquisition at a distance but they can still be perceived semantically using human vision. In this paper we analyze how these semantic annotations (labels) are usable as soft biometric signatures, useful for identification tasks.

A research line growing in popularity is focused on using this ancillary information (soft biometrics) in less constrained scenarios in a non-intrusive way, including acquisition "on the move" and "at a distance" [2]. These scenarios are still in their infancy, and much research and development is needed in order to achieve the levels of precision and performance that certain applications require.

As a result of the interest in these biometric applications at a distance, there is a growing number of research works studying how to compensate for the main degradations found in uncontrolled scenarios [3]. Here, the ancillary information such as soft biometrics can contribute to improve and compensate the degraded performance of systems at a distance.

The main contribution of the present paper is an experimental study of the benefits of soft biometric labels as ancillary information for challenging person recognition scenarios at a distance. In particular, we provide experimental evidence on how the soft labels of individuals witnessed at a distance can be used to improve their identification and help to reduce the effects of variability factors in these scenarios. Additionally, we propose a new adaptive method for incorporating soft biometrics information to this kind of challenging scenarios considering face recognition.

In order to do so, the largest and most comprehensive set of soft biometrics available in the literature is first described. These soft biometrics labels (called from now on soft labels) are manually annotated by several experts. These soft labels have been grouped considering three physical categories: global, body and head. The stability of the annotations of the different experts and their discriminative power are also studied and analyzed.

Finally, the available soft biometric information in scenarios of varying distance between camera and subject (*close*, *medium* and *far*) have been analyzed. The rationale behind this study is that depending on the particular scenario, some labels may not be visually present and others may be occluded.
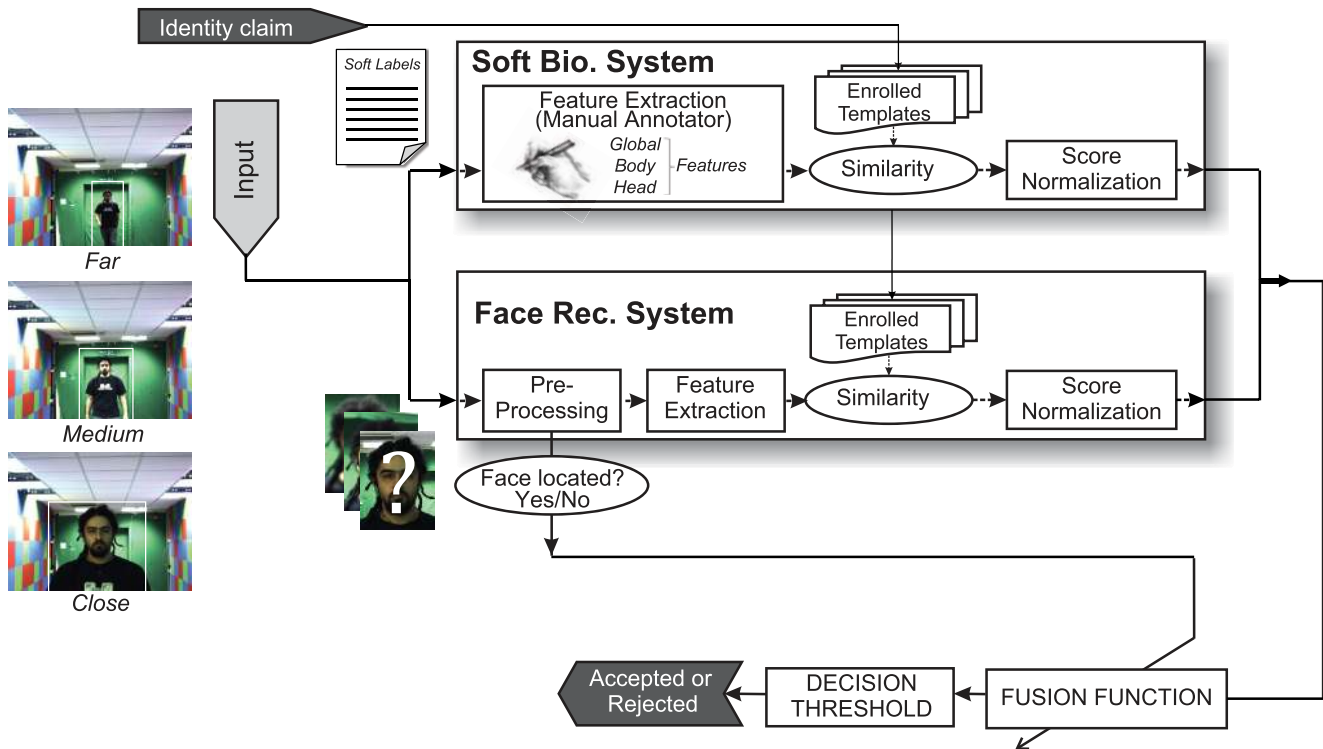
Fig. 1. Experimental framework. Two biometric systems are used, one based on soft labels and another based on face images. A final adaptive fusion is carried out at the score level.

As a result, the discriminant information of soft labels will vary depending on the distance.

The experimental framework used in this paper is shown in Fig. 1. This figure shows how from a video at a distance of a person walking, soft labels and faces from a subject are extracted. In this case, soft labels are extracted manually by human annotators because this process is still far from being implemented by an automatic system.

To date, this is the first publication showing the relation between the distance and the performance of soft biometrics for recognition at a distance.

The rest of this paper is organized as follows: Section II summarizes the related works, Section III reports an analysis of the soft biometrics obtained in this work. Section IV presents the experimental framework, scenario definition, and experimental protocol. Section V describes the recognition systems, and Section VI provides the experimental results and discussions. Finally, Section VII summarizes the contributions of this work.

II. RELATED WORK

First works in soft biometrics [4]–[6] tried to use demographic information (e.g., gender and ethnicity) and soft attributes like eye color, height, weight and other visible marks like scars [1], [7] and tattoos [8] 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) and can also be useful as a source of evidence in courts of law because they are more descriptive than the numerical matching 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.* [9] explored comparative facial attributes in the LFW Face Database [10] for face verification. In this case the proposed soft labels were extracted automatically based on still face images using trained binary classifiers.

Other works like [12]–[14] are focused on the automatic extraction of soft biometrics from video datasets. They proposed some soft labels based on height and color from the human body that can be easily extracted using automatic methods. Dantcheva *et al.* [15] proposed a group of soft labels based on nine semantic traits, mainly focusing on facial soft biometrics (e.g., beard, glasses, skin color, hair color, length, etc.), some body measures based on the torso and legs, and the clothes color.

On the other hand, D. Adjeroh *et al.* [16] studied the correlation and imputation in human appearance analysis of using automatic continuous data focusing on measurements of the human body. This study was carried out on the CAESAR anthropometric dataset, which is comprised of 45 human measurements or attributes for 2369 subjects. They analyzed these soft labels grouped in clusters and concluded that some of them inside each cluster can be predicted.

The latest works such as D. Reid and M. Nixon [17] introduce the use of comparative human descriptions for facial identification. They use twenty-seven comparative traits extracted manually from mugshot images to accurately describe facial features, which are determined by the Elo rating system from multiple comparative descriptions.

The present work involves the application of an extensive set of labels that can be visually described by humans at a distance and they are quantifiable in a discrete way. The soft labels considered here are based on head, global and body anthropometric measures and while previous works try to extract automatically them, here the soft labels have been tagged by human experts; this is another important difference. Thanks to it, we can analyze how humans understand and describe human body and face features visually at a distance.

The integration of soft biometric information to improve the accuracy of primary biometric systems has previously been studied in the literature following a probabilistic approach [4], [16]. In contrast in the present work, we have exploited the idea of the inclusion of soft biometrics with the primary biometric mode (face in this case), following an adaptive fusion scheme at the score level.

III. SOFT BIOMETRICS DATA ANALYSIS

In this paper a set of soft biometrics has been used, whose main value is that it is discernible by humans at a distance.

These physical trait labels are obtained from the Southampton Multibiometric Tunnel Database (TunnelDB) [18] which contains biometric samples from 227 subjects for which 10 gait sample videos from between 8 to 12 viewpoints are taken simultaneously. The TunnelDB database also contains high-resolution frontal videos to extract face information and high-resolution still images taken to extract ear biometrics. There are roughly 10 of such sets of information gathered for each subject.

The TunnelDB datasets were annotated against recordings taken of the individuals in laboratory conditions [19]. The annotation process was as follows: an annotator visualized the full video of a subject walking toward the camera and then generated one set of soft labels per each video. It is important to note that the process followed here is independent of the distance. A range of discrete values is given to each trait label, e.g. "Arm length" marked as 1 (very short), 2 (short), 3 (average), 4 (long), and 5 (very long). The annotation process of each label is described in detail in [20]. A summary of these trait labels and their associated discrete semantic terms is provided in Table I.

The labels and the labelling process were largely inspired by an earlier study in Psychology which generated a list of 23 traits, each formulated as a bipolar five-point scale, and the reliability and descriptive capability of these traits was gauged [21]. The 13 most reliable terms, the most representative of the principal components, were incorporated into the final trait set with the same scale [20].

These labels were designed based on which traits humans are able to consistently and accurately use when describing people at a distance. The traits were grouped in 3 classes, namely:

- *Global* traits (age, ethnicity and sex). The demographic information as the gender and ethnicity of a person does not typically change over the lifetime, so it can be used to filter the database to narrow down the number of candidates. On the other hand, age is easily estimated

TABLE I
PHYSICAL SOFT LABELS AND THEIR ASSOCIATED SEMANTIC TERMS. EXTRACTED FROM [20]

Body

Trait	Range of Values
1. Arm Length	Very Short, Short, Average, Long and Very Long
2. Arm Thickness	Very Thin, Thin, Average, Thick and Very Thick
3. Chest	Very Slim, Slim, Average, Large and Very Large
4. Figure	Very Small, Small, Average, Large and Very Large
5. Height	Very Short, Short, Average, Tall and Very Tall
6. Hips	Very Narrow, Narrow, Average, Broad and Very Broad
7. Leg Length	Very Short, Short, Average, Long and Very Long
8. Leg Direction	Very Bowed, Bowed, Straight, Knock Kneed and Very Knock Kneed
9. Leg Thickness	Very Thin, Thin, Average, Thick and Very Thick
10. Muscle Build	Very Lean, Lean, Average, Muscly and Very Muscly
11. Proportions	Average and Unusual
12. Shoulder Shape	Very Rounded, Rounded, Average, Square and Very Square
13. Weight	Very Thin, Thin, Average, Big and Very Big

Global

Trait	Range of Values
14. Age	Infant, Pre Adolescence, Adolescence, Young Adult, Adult, Middle Aged, Senior
15. Ethnicity	European, Middle Eastern, Indian/Pakistan, Far Eastern, Black, Mixed, Other
16. Sex	Female, Male

Head

Trait	Range of Values
17. Skin Color	White, Tanned, Oriental and Black
18. Facial Hair Color	None, Black, Brown, Red, Blond and Grey
19. Facial Hair Length	None, Stubble, Moustache, Goatee and Full Beard
20. Hair Color	Black, Brown, Red, Blond, Grey and Dyed
21. Hair Length	None, Shaven, Short, Medium and Long
22. Neck Length	Very Short, Short, Medium and Long
23. Neck Thickness	Very Thin, Thin, Average, Thick and Very Thick

by physical traits at a distance and it can also be used to filter suspects.

- *Body* features that describe the target's perceived somatotype [22] (height, weight, etc.) These traits have a close correlation between the style and kind of clothes that the subject is wearing in the annotation process. For example, tight clothes will allow to obtain more stable labels than loose clothes.

- *Head* features, an area of the body humans pay great attention to if it is visible [23] (hair color, beards, etc.) These are very interesting soft biometrics to be fused with face recognition systems.

To understand the role of soft labels and their application to biometrics at a distance, the internal correlation, the stability, and the discrimination power of the different labels with semantic annotations is studied and analyzed in the next Section. In this paper, a total of 13.340 labels from 58 subjects annotated by 10 different experts[1] are used in the experiments reported in Section VI. The remaining subjects in TunnelDB were annotated only by just 1 or 2 different experts and were rejected for this analysis.

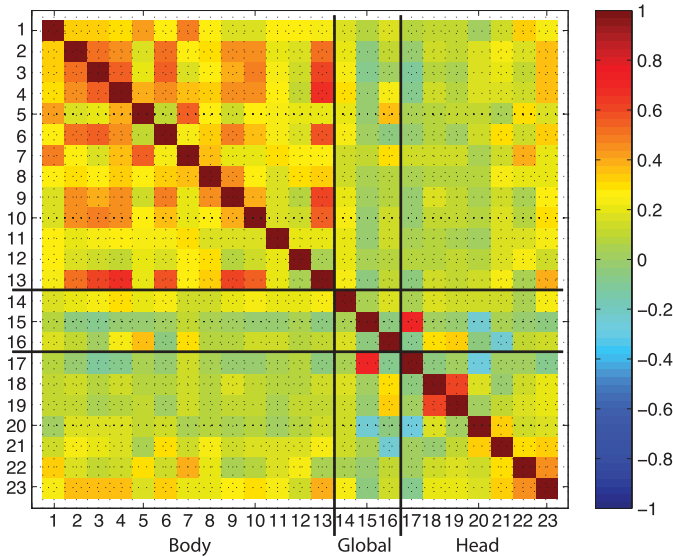[1]Available at http://atvs.ii.uam.es/tsb_db.html

Fig. 2.   Correlation between labels of the 58 subjects considered based on Pearson's coefficient r (see Eq. 1).

Fig. 3.   Annotators' stability for the 23 soft labels considered (see Table I).

A. Correlation Between Labels

This section reports an analysis of the correlation between the labels defined. For this purpose the correlation between all pairs of labels of the three groups defined (*global*, *body* and *head*) 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 the pairs of semantic terms at hand. Here each semantic term was converted to numerical values in the range 1 to 5 if the annotation contains the semantic term (e.g. very short, short, average, long and very long) and 0 if the annotation was left empty by the annotator (they were not sure what to annotate). X_i and Y_i are the label values across all individuals and annotators, therefore $N = 580$ annotations (58 subjects × 10 annotators). 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.

The correlation matrix containing the correlation between all labels is represented graphically in Fig. 2. Colors in the red range represent correlation coefficients close to 1.0 and thus a positive correlation, while colors in the blue range represent correlation coefficients close to −1.0 and thus a negative correlation. Pale green represents no correlation between labels.

Similarly to the previous work [20], the 58 subjects selected for the experiments follow the same tendencies regarding correlation between labels. As a novelty with respect to [20],
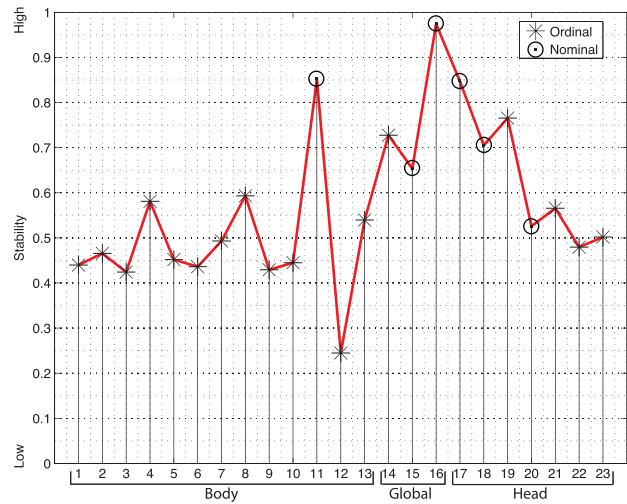
here the correlation has been studied grouping the labels in 3 categories: *body*, *global*, and *head*.

Focusing our attention in the *global* labels, very small correlation between these 3 features and all the remaining ones is observed in the graph as could be expected. On the other hand, some *body* labels are very correlated between them mainly due to the proportion relationships of the human body (e.g., the larger the arms the larger the legs). This means that physical characteristics like the chest (3), and the figure (4) are very correlated. Therefore if we try to recognize people just by using these correlated features the success rate will not be very high.

Head features do not present the same correlation between them compared to *body* traits (except e.g. facial hair color (18) and facial hair length (19) or neck length (22) and neck thickness (23) which are highly correlated).

Fig. 2 also shows some strong relationships between demographic traits such as ethnicity (15) and skin color (17), or hair color (20), as was expected.

As observed in [16] the human body measurements are often correlated. In the same way, our experimental results also show correlations between body measurements.

B. Stability Analysis of Annotations

This section reports an analysis of the stability of the human annotations for all soft labels. This is done by calculating the stability coefficient, defined for label X as:

$$\text{Stability}_X = 1 - \frac{1}{SA}\sum_{i=1}^{S}\sum_{a=1}^{A}|X_{ia} - \text{mode}_a(X_{ia})| \quad (2)$$

where X_{ia} is the annotated value for subject i by annotator a, $A = 10$ is the total number of annotators, $S = 58$ is the total number of subjects, and $\text{mode}_a(X_{ia})$ is the statistical mode across annotators (i.e., the value most often annotated for subject i).

The resulting stability coefficients for all labels are depicted in Fig. 3. Using the definitions in chapter 11 of [24], we can
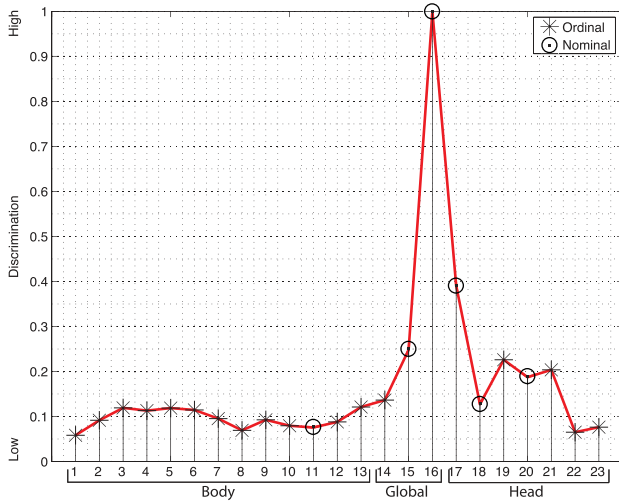
Fig. 4. Discrimination power of the 23 soft labels considered (see Table I).

see that some of the features are *nominal*, i.e., their values cannot be ordered meaningfully (e.g., ethnicity (15), sex (16), skin (17), facial hair (18) and hair color (20)) whereas others are *ordinal*, i.e., their values can be meaningfully ordered (e.g., arm length (1), arm thickness (2), height (4), weight (13), and hair length (21)).

In Fig. 3 we can see that sex (16) (a *nominal* label that has just two terms, male and female), is the most stable label due to the low variability. Other *nominal* features such as body proportions (11) and skin color (17) have also high stability. On the other hand, the stability of *ordinal* features such as arm length (1), height (5), hips (6), or shoulder shape (12) is lower due to the high variability and the different point of view of the annotators.

Although these two types of features (*nominal* and *ordinal*) may be processed differently (e.g., using different similarity measures), here in this paper we have processed them in the same way as an initial approach.

C. Discrimination Power Analysis

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

$$\text{Discrimination}_X = \frac{\frac{1}{S(S-1)} \sum_{i=1, i \neq j}^{S} \sum_{j=1}^{S} |\mu_i - \mu_j|}{\sigma} \quad (3)$$

$$\mu_i = \text{mean}_a(X_{ia}), \quad \mu_j = \text{mean}_a(X_{ja}), \quad \sigma = \frac{1}{S} \sum_{i=1}^{S} \sigma_i \quad (4)$$

where $\sigma_i = \text{std}_a(X_{ia})$, i and j index subjects, and a indexes annotators.

The discrimination coefficient for the X^k labels ($k = \{1, \ldots, K = 23\}$) is depicted in Fig. 4. There we can see that the *body* features (IDs 1-13) are less discriminant than the *global* (IDs 14-16) and *head* (IDs 17-23) features. The least discriminant features are the arm length (1) and neck length (22) followed by leg direction (8) and neck

thickness (23). These are *ordinal* features and therefore the majority of the subjects share similar annotations.

Eq. 3 gives an idea of the discrimination power of each label, given that $\sigma > 0$. If $\sigma = 0$, i.e., there is no variation across annotators, then this measure is not reliable. This is the case for the label *sex* (16). Fig. 3 showed that sex is the most stable label (i.e., the annotators give always a correct decision), hence the intra-variability will be 0 and consequently Discrimination$_X \rightarrow 1$. Therefore in the case where we have a label without annotation mistakes (where the annotators always select the correct value) Eq. 3 cannot predict correctly the discrimination power. When gathering larger data sets we anticipate that there are more likely to be more errors in the labelling of sex than have been experienced here.

Better results are reached for the *nominal* features such as ethnicity (15), or skin color (17), and the most discriminative is the sex (16) due to the clear identification by the human annotators in the TunnelDB database. Consequently, we can predict that *global* and *head* features will provide better person recognition results than *body* features.

IV. EXPERIMENTAL FRAMEWORK

A. Scenario Definition

The annotation process in [18] was as follows: an annotator visualized the full video of a subject walking toward the camera and then generated a set of the soft labels defined in Table I per each video, hence the labels are unique for the whole set of three distances.

In our case using those sets of labels, three different challenging scenarios, varying the distance between camera and subject, have been defined and used in our experiments in order to understand the behaviour of soft biometric labels and their best application to biometrics at a distance. For this purpose, high resolution frontal face sample videos from the TunnelDB database [18] have been used together with their corresponding physical soft labels analyzed in the previous sections. A summary of this process is shown in Fig. 5. The three scenarios are defined as follows:

- **Close** distance (~ 1.5m). Includes both the face and the shoulders.
- **Medium** distance (~ 4.5m). Includes the upper half of the body.
- **Far** distance (~ 7.5m). Includes the full body.

The rationale behind this study is the fact that depending on the particular scenario, some labels may not be visually present and others may be occluded. As a result, the discriminative information of the soft biometrics will vary depending on the distance. Table II shows the soft labels available for each of the scenarios defined.

B. Experimental Protocol

The same dataset selected for the soft labels from the TunnelDB was used for the face recognition system. Each user has 10 sessions, so 580 images per scenario from high-resolution frontal face sample videos have been used. For each of the 10 sessions of a subject, the first frame (*close* distance), the middle frame (*medium* distance) and the last
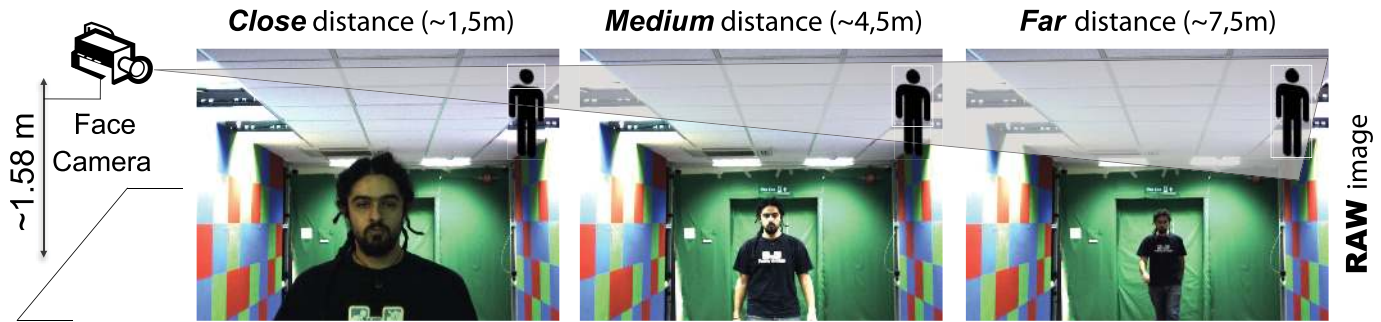
Fig. 5. Scenario defined based on the TunnelDB [18]: *close*, *medium*, and *far* distance images used in the experimental work. *Body* region visible at the three distances considered. A person walking frontal to the camera is captured by a high-resolution video camera (10 fps and resolution of 1600 × 1200).

TABLE II
SOFT LABELS AVAILABLE VISUALLY IN EACH SCENARIO
USING NUMBERING FROM TABLE I

Scenario	Body	Global	Head
Close	11,12	all	all
Medium	1,2,3,4,10,11,12,13	all	all
Far	all	all	all

frame (*far* distance) from the frontal videos have been selected to generate the image samples used in the experiments, having in total 1740 images (58 subjects × 10 sessions × 3 distances).

The database was divided into gallery and testing sets. For each subject 9 face images and 9 sets of soft labels were used for the training and the remaining session was used for testing following a leave-one-out approach [24] generating this way 580 similarity target scores and 33640 similarity non-target scores.

V. RECOGNITION SYSTEMS

A. Verification Based on Soft Biometrics

This section describes a person verification system based only on soft biometrics. First, each label in numeric form (see Section III) is normalised to the range [0, 1] using the tanh-estimators described in [25]:

$$X'^k = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{X^k - \mu_{X^k}}{\sigma_{X^k}}\right)\right) + 1\right\} \qquad (5)$$

where X^k is the $k = \{1, \ldots, K\}$ soft label ($K = 23$), X'^k denotes the normalized label, and μ_{X^k} and σ_{X^k} are respectively the estimated mean and standard deviation of the label under consideration (see Table I for the list of the labels).

Note that, depending on the scenario considered (*close*, *medium*, and *far*), there are $K = 12$, 17, or 23 labels, respectively (see Table II).

Similarity scores $s(\mathcal{C}, \mathbf{x})$ are computed using the Mahalanobis distance [24] between the test vector with K labels $\mathbf{x} = \{X'^1, \ldots, X'^k\}$ and a statistical model $\mathcal{C}$ of the client, obtained using a number of gallery labels (9 examples per label in our experiments), as follows:

$$s(\mathcal{C}, \mathbf{x}) = \frac{1}{\left(\left(\mathbf{x} - \mu^{\mathcal{C}}\right)^T \left(\Sigma^{\mathcal{C}}\right)^{-1} \left(\mathbf{x} - \mu^{\mathcal{C}}\right)\right)^{1/2}} \qquad (6)$$

where $\mu^{\mathcal{C}}$ and $\Sigma^{\mathcal{C}}$ are respectively the mean vector and covariance matrix obtained from the gallery labels, which form the statistical model of the client $\mathcal{C} = \{\mu^{\mathcal{C}}, \Sigma^{\mathcal{C}}\}$.

B. Verification Based on Face Biometrics

For the face recognition experiments, two different systems have been used and compared (one commercial and one proprietary): *i*) Luxand FaceSDK 4.0, and two face recognition systems based on SRC [26], *ii*) VJ-SRC, using automatic face detection based on Viola Jones [27], and *iii*) ID-SRC using ideal face detection marked manually.

FaceSDK by Luxand[2] is a high-performance and multi-platform face recognition solution based on facial fiducial feature recognition.

A proprietary **VJ-SRC** face recognition system based on Viola Jones to detect faces and using a matcher based on SRC [26], [28] is also used. Face segmentation and location of the eyes are two of the main problems in face recognition systems at a distance. For our experiments, we have also manually tagged the eyes' coordinates which allows us to consider an ideal case of face detection in the **ID-SRC** face recognition system. This way, we can compare the behaviour of soft labels when fused with face images on real (VJ-SRC) and ideal (ID-SRC) scenarios at a distance free of segmentation errors.

The SRC matcher is a state-of-the-art system based on recent works in sparse representation for classification purposes. Essentially, this kind of systems spans a face subspace using all known gallery face images, and for an unknown face image they try to reconstruct the image sparsely. The motivation of this model is that given sufficient gallery samples of each person, any new test sample for this same person will approximately lie in the linear span of the gallery samples associated with the person.

VI. EXPERIMENTS

This section describes the experimental analysis of the discrimination power of individual and grouped soft labels and the performance of the considered face recognition systems in the three scenarios defined. Then, a fusion of the two modalities in different conditions is studied. Results are
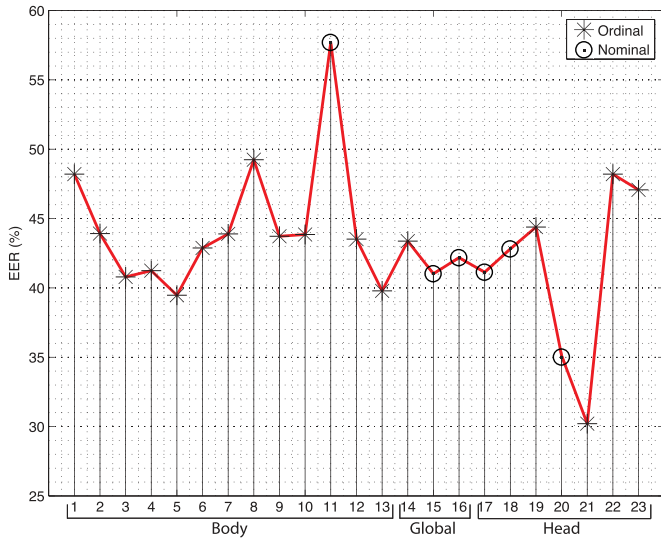
[2]http://www.luxand.com/facesdk/

Fig. 6.   EER (%) obtained for each individual soft label defined in Table I .
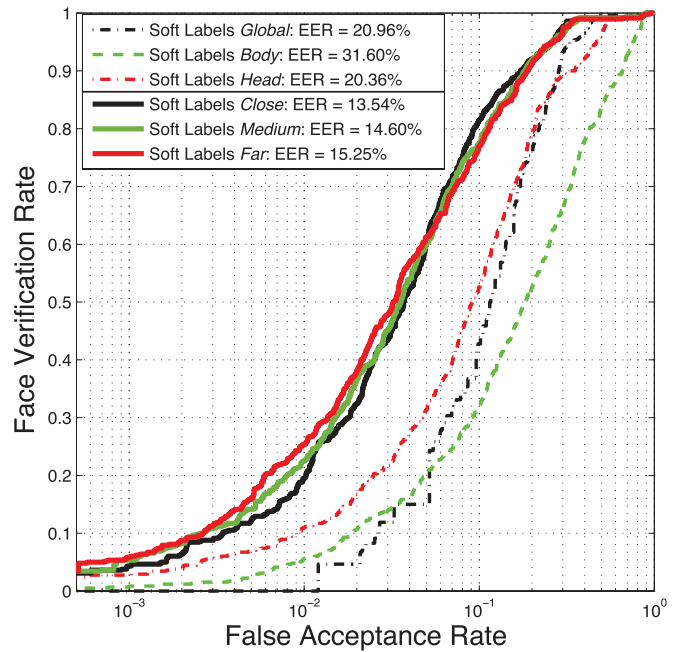

Fig. 7.   ROC curves obtained for the physical labels sets (*global*, *body*, and *head*) grouped following the definition in Table I and for the three defined scenarios in Table II (*close*, *medium*, and *far*), i.e., the soft labels that would be visible at these distances.

reported using ROC curves, with EERs and verification rates (VR) working at a different FAR points (FAR = 0.1%, 1%, and 10%).

A. Soft Labels

1) Analysis of Individual Soft Labels: This section presents the discrimination power of each individual soft label following the leave-one-out experimental protocol described in Section VI. As shown in Fig. 6, hair length (21) achieves the best results (EER = 30.27%) but it is worth noting that this was not the most discriminative feature regarding the initial experiments shown in Fig. 4. Another relevant label with a high performance and discrimination power is hair color (20) with an EER = 35.11%. The rest of soft labels achieve similar performance, with better results in general for *head* labels compared to *body* labels, as anticipated in Section III-A As can be seen, individual labels are not very discriminative on their own.

2) Analysis of Grouped Soft Labels: The aim of this experiment is to study the discriminative power of the three groups of soft labels considered in the different scenarios at a distance defined in Section IV-A. Fig. 7 shows the performance of each set of labels considered. Here, dashed lines represent the sets: *global*, *body* and *head*, while solid lines represent all the available labels in each scenario at a distance as defined in Table II.

There is a significant difference between *global*, *head* and *body* regarding the performance as can be observed. The performance of *body* labels is clearly lower compared to *global* and *head* sets as predicted in Sections III-B and III-C through the stability and discrimination analysis.

Regarding the other 3 groups of labels that take into account the labels visible at the 3 distances defined the difference of performance is not that significant as can be seen in Fig. 7. *Far* scenario is comprised of all available labels including *body* labels, therefore it experiences a decrease in EER performance compared to the other scenarios in some regions of the plot

(e.g., around FAR = 0.1 = 10%). On the other hand, the other two scenarios have a lower number of soft labels available but result in better EER performance.

It is important to note that although soft labels provide low recognition performance when used as a stand alone system, they can help to improve hard biometric systems as we will show in Sect. 6.3.

3) Analysis of Gallery Set Size for Soft Labels: An important parameter to be considered in soft labels systems is the size of the gallery set. For this purpose, we have evaluated the system with different number of gallery samples (varying between 1 to 9 samples) following a leave-one-out methodology.

Fig. 8 shows the different configurations analyzed for the six sets of soft labels defined in the previous section. As can be seen, all soft label sets follow the same trend, the system recognition performance (EER) improves significantly when more samples are used in the training stage. For *global*, *body*, and *head* sets using more than 5 gallery samples the system performance saturates. On the other hand, for *close*, *medium*, and *far* sets, the performance saturates for more than 7 samples.

As it was expected the more features are included in the set (e.g., for *far* labels which include all 23 labels) the larger the performance improvement for increasing gallery samples until saturation. The relative performance improvement before the saturation for small datasets (e.g., *global* with only 3 labels) is much smaller.

As Fig. 8 shows, the *head* labels achieve better performance than the *global* when more than 5 gallery samples are considered in the training stage. This effect can be explained by the
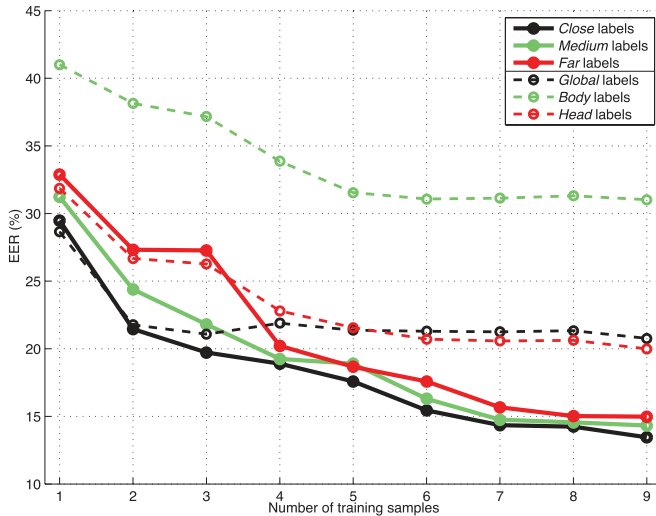
Fig. 8. EER (%) obtained when varying the number of gallery samples.

TABLE III

FACE DETECTION ERRORS IN THE THREE SCENARIOS AT A DISTANCE
FOR VIOLA JONES AND FACESDK SYSTEMS. FTA AND FTD
ERROR PERCENTAGES ARE CALCULATED FOR THE
TOTAL NUMBER OF FACE IMAGES (N = 580)

	Scenario	FTA	FTD	Total$_E$=FTA+FTD
Viola Jones	Close	11 (1.89%)	11 (1.89%)	22 (3.79%)
	Medium	2 (0.34%)	116 (20.00%)	118 (20.34%)
	Far	96 (16.55%)	304 (52.40%)	400 (68.96%)
FaceSDK	Close	5 (0.86%)	20 (3.44%)	25 (4.31%)
	Medium	365 (62.93%)	66 (11.37%)	431 (74.31%)
	Far	571 (98.44%)	9 (1.55%)	580 (100%)

Fig. 9. ROC curves of SRC systems obtained using two configurations: automatic (VJ-SRC, dashed lines) and manual (ID-SRC, solid lines, FTA = 0%, FTD = 0%).

different number of labels that comprises both sets: 3 labels for *global* and 7 for *head* (see Table I). In other words, the higher number of degrees of freedom for the *head* set leads to improved performance compared to the *global* set if the training set is large enough.
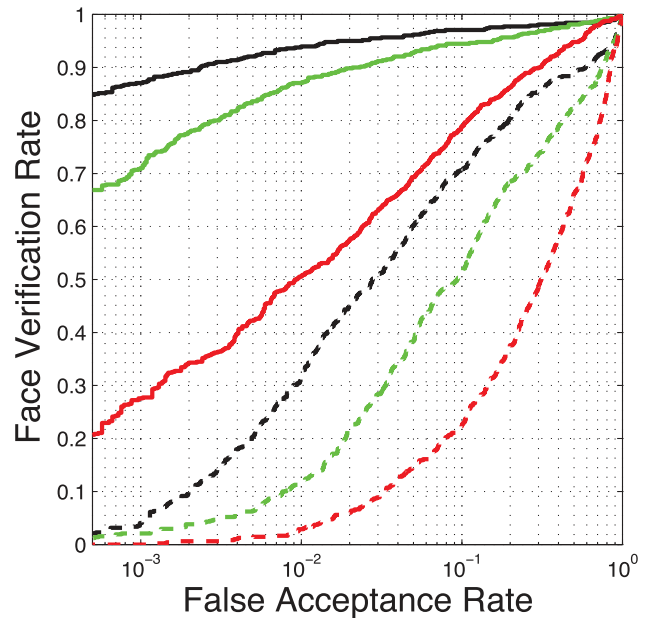
B. Face Recognition

1) Analysis of Face Detection Errors: This section presents an analysis of the three scenarios considered: *close*, *medium*, and *far*. Two face detection systems have been evaluated: *i)* proprietary based on Viola Jones, and *ii)* a commercial system (FaceSDK) based on facial landmarks.

Two different detection errors have been defined and analyzed:

- Fail To Acquire (FTA): when there is a face in the image, but it is not detected.
- Fail To Detect (FTD): when the face detector finds an object in the image, but it is not a face.

The first error FTA will be a feedback report for the systems but the second error FTD has to be analyzed manually by an operator or automatically by an error detector system. In this paper FTD error was evaluated manually observing the faces detected by both systems.

Table III shows the detection errors for the two systems evaluated. Firstly, Viola Jones approach achieves less FTA errors than FaceSDK system, but introduces a high number of FTD errors which will affect the system recognition performance. The FTA errors in *close* scenario are due to short people whose middle part of the face is outside of the vision plane of the camera.

As can be seen, the scenarios at a distance analyzed are very challenging. Analyzing the results both systems work poorly at *medium* and *far* distances due to the high variability and the low quality of face images. The Viola Jones approach achieves a reasonable FTA error in these distances but a large number of detections are not faces (FTD error is very high). On the other hand, the FaceSDK system has a higher FTA with lower FTD. The total error is so large for FaceSDK (73.31% and 100%) that it was discarded for the following experiments.

2) Analysis of Face Recognition Systems: The results achieved for VJ-SRC and ID-SRC systems with automatic and manual (FTA = 0% and FTD = 0%) face detection are presented in Fig. 9. As can be seen in the manual face detection (ID-SRC system, solid lines), the database analyzed is very challenging and the system performance decreases quickly when the acquisition distance increases. On the other hand, poor results are achieved for the case of using the automatic Viola Jones face detector (VJ-SRC) due to the high number of FTD errors but also because in this case there is no pose
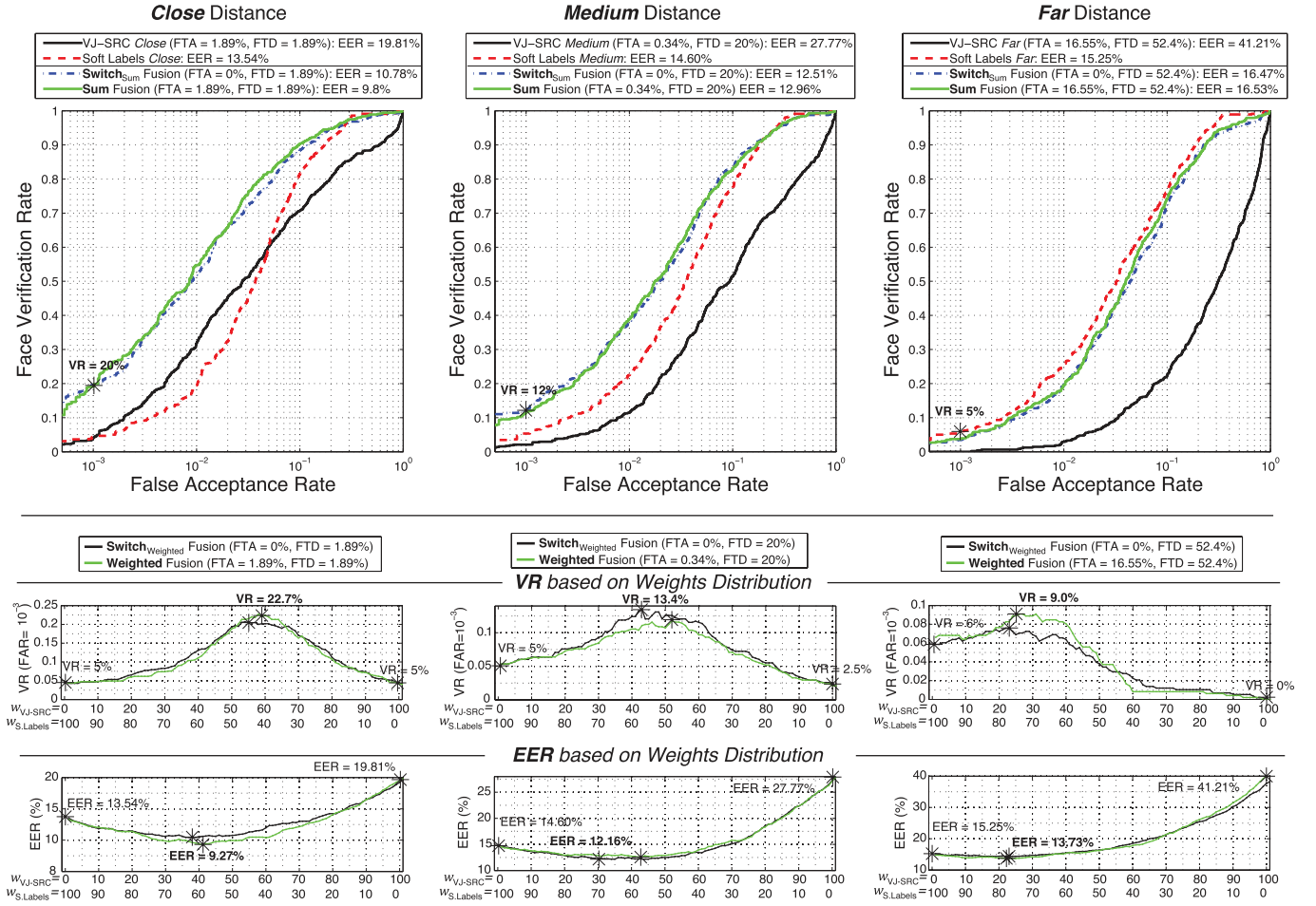
Fig. 10. ROC curves for the VJ-SRC system (automatic face detection errors) together with the corresponding improvement by sum and switch fusion for the three scenarios defined: *close* (left), *medium* (center), and *far* (right). Best configuration of weights for each fusion (VR and EER performance) is in bold in bottom graphs.

compensation and normalisation regarding the position of the eyes as in the ideal case.

Therefore, a large improvement in the EER is achieved for all distances by considering manual face detection compared to Viola Jones in the SRC system. On the other hand, the system performance with automatic face detection is very poor in a FAR = 0.001 = 0.1% with Verification Rates (VR) lower than 5%. It is important to note that for *far* scenario with ideal face detection (ID-SRC system) the VR is lower than 30%, which shows the complexity of the database analyzed.

C. Fusion of Face and Soft Biometrics

Soft biometrics offer several benefits over other forms of identification at a distance as they can be acquired from low resolution and low frame rate videos, and have great invariant attributes such as to camera viewpoint, sensor ageing and scene illumination. This allows for the use of soft biometrics when primary biometric identifiers cannot be obtained or when only a description of the person is available.

This section analyzes how soft labels can improve the face recognition system performance through the fusion of both biometric systems. The fusion method used is based on

the combination of the systems at the score-level following different fusion approaches [29], [30]: *i*) the sum rule, *ii*) an adaptive switch fusion rule, and *iii*) a weighted fusion rule. As indicated in Fig. 1, the switch fusion rule uses only the soft labels for recognition in the cases where no face images are detected, and sum or weighted fusion is applied if both scores are available. This helps the real automatic systems to achieve better performance dealing with low resolution images.

To carry out the fusion stage of the two biometric modalities, scores of the different systems were first normalized to the [0, 1] range using the tanh-estimators described in [25]. This simple method is demonstrated to give good results for the biometric authentication problem.

Experiments are carried out by fusing the soft labels with VJ-SRC and ID-SRC face recognition systems over the three acquisition distances: *close*, *medium* and *far*. First, we consider the case of the fusion of soft labels with the automatic face detection errors, and then the case of their fusion with an ideal face recognition using manual face detection.

1) Fusion With Automatic Face Detection Errors: This experiment studies the fusion of soft labels with the VJ-SRC system with automatic face detection carried out
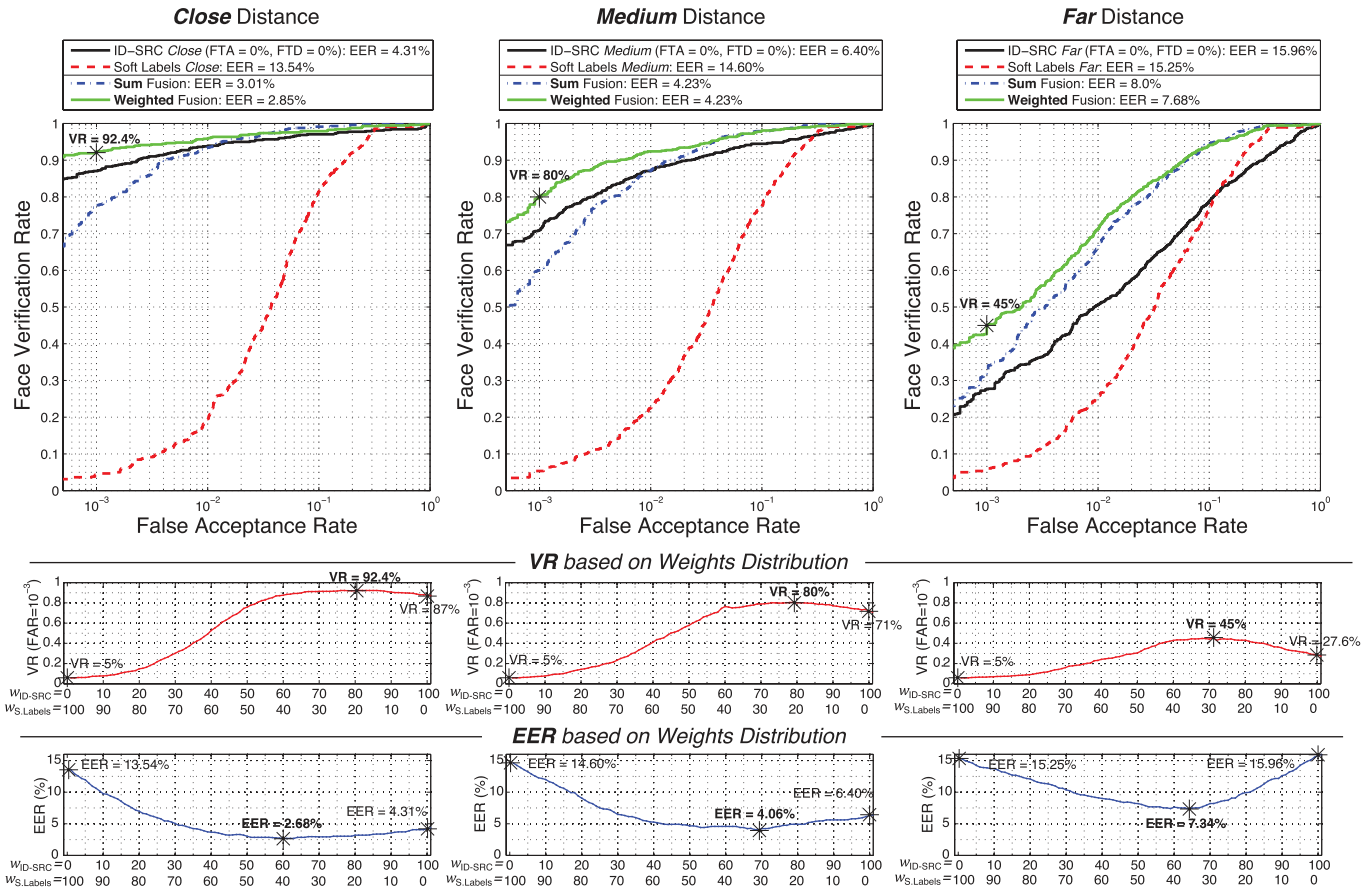
Fig. 11. ROC curves for the ID-SRC system (manual face detection) and its corresponding improvement by sum and weighted fusion rule for the three scenarios defined. Best configuration of weights for weighted fusion (VR and EER performance) is in bold in bottom graphs.

using a switch fusion. In case the face recognition system fails to acquire (FTA) a face due to variability factors, soft labels can help to improve the system performance.

In video surveillance systems (at a distance), in most cases you the presence of the person can be detected but the faces do not always have enough quality to be useful. In that case, the automatic systems are going to produce a FTA error and this switch fusion allows us to use a soft biometric system where traditional systems do not work. This is case also happens in forensic scenarios when criminals cannot be identified in surveillance videos by their faces (due to occlusions or low quality) but the soft information (clothes, body and head information, etc) could be very useful.

Fig. 10 shows 4 ROC profiles in each graph: the VJ-SRC face recognition system, the soft labels system and two fusions. The first fusion applies a sum rule of the scores from the two systems only if both of them are available, otherwise it emits a FTA. As a result using this sum fusion FTA is non-zero. On the other hand, the switch fusion always results in an output score as described above, reducing the FTA error to 0 in this case. Detection errors showed in Table III show the cases in which the switch fusion selects only the soft labels for the three scenarios defined.

The sum fusion of the two systems achieves absolute improvements of 10.0%, 14.8%, 24.6%, and relative improve-

ments of 50.1%, 53.3%, and 59.9% of EER for *close*, *medium*, and *far* scenarios, respectively compared to the VJ-SRC face recognition system. As shown, soft labels improve the system performance and allow the system to maintain robustness in *far* scenario. The same conclusion is confirmed for the switch fusion of the systems, which achieves absolute improvements of 9.0%, 15.2%, 24.7%, and relative improvements of 45.0%, 54.9%, and 60.0% of EER for *close*, *medium*, and *far* scenarios, respectively, compared to the VJ-SRC face recognition system.

As can be seen, the EERs for sum and switch sum fusion are similar, with the advantage of switch fusion of eliminating all FTA errors. In these scenarios a weighted fusion rule has been also evaluated. Fig. 10 (bottom) shows the VR and EER for varying weights in the weighted and switch weighted fusion. Based on these results, we have fixed the following weights: $w^{face} = 0.6$ and $w^{soft} = 0.4$ for *close*, and *medium* distance, and finally $w^{face} = 0.25$ and $w^{soft} = 0.75$ for *far* distance. Using this configuration we achieve an absolute increment in VR of around 2% for all the distances.

Therefore, as the results show, a real face recognition system which do not have a good performance due to the variability factors derived from acquisition at a distance, could be improved using soft biometric labels visually available in the scene.

2) Fusion With Manual Face Detection: This experiment focuses on use of the soft labels in order to improve the ID-SRC system with ideal face detection (FTA = 0% and FTD = 0%). Fig. 11 shows the ROC curves of both systems and two fusions (sum and weighted fusion rules) for different FAR points.

In this case the incorporation of soft labels improves the face recognition system performance. The sum fusion achieves significant relative improvements of 30.1%, 33.9%, and 49.8% in the EER for *close*, *medium*, and *far* scenarios respectively. On the other hand analyzing the Verification Rate (VR) in a high security point such as FAR = 0.001 (0.1%), the system performance deteriorates. A relative decrement of about 10% in the VR for *close* and *medium* scenarios is obtained but in *far* scenario the VR increases moderately. These results are due to the poor performance of soft labels in a high security working point.

A weighted fusion has been proposed in order to solve the problem of the VR deterioration. The fusion gives more weight to the most robust system which is the face recognition system in FAR = 0.1%. Different weights have been tuned for the 3 distances based on the EER performance of the systems. Fig. 11 (bottom) shows the VR and EER for varying weights. Based on these results, we have fixed the following weights: $w^{face} = 0.8$ and $w^{soft} = 0.2$ for *close* and *medium* distance, and finally $w^{face} = 0.7$ and $w^{soft} = 0.3$ for *far* distance. Using this configuration we achieve an absolute increment in VR of 5.3%, 8.9%, 20.4%, and a relative increment in VR of 92.4%, 80.0%, and 45.0%, for *close*, *medium*, and *far* scenarios, respectively.

Therefore, the usage of soft labels can still help to improve the systems in these better conditions. The face detection stage is a key factor in order to achieve good results in scenarios at a distance. Consequently a single weighted fusion rule combining soft biometrics allows to improve the system performance where the primary biometrics are not working due to variability factors in the scenarios at a distance.

VII. Conclusion

This work reports a study of how the usage of soft labels can help to improve a biometric system for challenging person recognition scenarios at a distance. It is important to emphasize that the use of this ancillary information is very interesting in scenarios suffering from very high variability conditions. These soft labels can be visually identified at a distance by humans (or an automatic system) and fused with hard biometrics (as e.g., face recognition). It is important to note that this kind of soft information is still a developing field in relation to its automatic extraction.

First, the stability and discriminative power of the largest and most comprehensive set of soft labels available from the literature, has been studied and analyzed. The discriminative information of these labels grouped by physical categories (*body*, *global* and *head*) has also been studied.

Moreover, the available soft biometric information in scenarios of varying distance between camera and subject (*close*, *medium* and *far*) has been analyzed. The rationale behind this study is that depending on the scenario, some labels may not be visually present and others may be occluded. Thus, the discriminative information of soft biometrics will vary depending on the distance. To the best of our knowledge, this is the first publication to date showing the relation between scenarios at a distance and the performance of soft biometrics for person recognition.

Finally, some fusion rules have been proposed and studied to incorporate soft biometrics to these challenging scenarios at a distance considering a state-of-the-art face recognition system. Experiments are carried out considering both automatic and manual face detection. Results have shown the benefits of the soft biometrics information maintaining robustness of the face recognition performance and also improving the performance on a high security level. We have shown how this visually-available ancillary information can be fused with traditional biometric systems and improve their performance in scenarios at a distance.

References

[1] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 406–415, Sep. 2010.

[2] S. Z. Li, B. Schouten, and M. Tistarelli, *Handbook of Remote Biometrics for Surveillance and Security*. New York, NY, USA: Springer-Verlag, 2009, pp. 3–21.

[3] Robust, Riyadh, Saudi Arabia. (2008). *Robust Biometrics: Understanding Science & Technology* [Online]. Available: http://biometrics.cylab.cmu.edu/ROBUST2008

[4] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, "Integrating faces, fingerprints and soft biometric traits for user recognition," in *Proc. Biometric Authentication Workshop, LNCS*, 2004, pp. 259–269.

[5] A. K. Jain, S. C. Dass, K. Nandakumar, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Proc. Int. Conf. Biometric Authentication*, 2004, pp. 731–738.

[6] D. Heckathorn, R. Broadhead, and B. Sergeyev, "A methodology for reducing respondent duplication and impersonation in samples of hidden populations," in *Proc. Annu. Meeting Amer. Sociol. Assoc.*, 1997, pp. 543–564.

[7] A. K. Jain and U. Park, "Facial marks: Soft biometric for face recognition," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 37–40.

[8] J. Eun Lee, A. K. Jain, and R. Jin, "Scars, marks and tattoos: Soft biometric for suspect and victim identification," in *Proc. Biometric Symp., Biometric Consortium Conf.*, 2008, pp. 1–8.

[9] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE 12th ICCV*, Oct. 2009, pp. 365–372.

[10] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Karlsruhe Inst. Technol., Univ. Massachusetts, Boston, MA, USA, Tech. Rep. 07-49, Oct. 2007.

[11] A. Gupta and L. S. Davis, "Beyond nouns: Exploiting prepositions and comparative adjectives for learning visual classifiers," in *Proc. ECCV*, 2008, pp. 16–29.

[12] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft-biometrics: Unconstrained authentication in a surveillance environment," in *Proc. DICTA*, 2009, pp. 196–203.

[13] D. Vaquero, R. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk, "Attribute-based people search in surveillance environments," in *Proc. IEEE WACV*, Snowbird, UT, USA, Dec. 2009, pp. 1–3.

[14] Y. Fu, G. Guo, and T. S. Huang, "Soft Biometrics for Video Surveillance," in *Intelligent Video Surveillance: Systems and Technology*, Y. Ma and G. Qian, Eds. Cleveland, OH, USA: CRC Press, 2009, pp. 407–432, ch. 15.

[15] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification: New trends and challenges," *Mutimedia Tools Appl.*, vol. 10, pp. 1–36, Oct. 2010.

[16] D. Adjeroh, D. Cao, M. Piccirilli, and A. Ross, "Predictability and correlation in human metrology," in *Proc. IEEE Int. WIFS*, Dec. 2010, pp. 1–6.

[17] D. Reid and M. Nixon, "Human identification using facial comparative descriptions," in *Proc. ICB*, Jun. 2013, pp. 1–7.

[18] R. D. Seely, S. Samangooei, L. Middleton, J. Carter, and M. Nixon, "The University of southampton multi-biometric tunnel and introducing a novel 3D gait dataset," in *Proc. IEEE Biometrics, Theory, Appl. Syst.*, Sep. 2008, pp. 1–6.

[19] R. D. Seely, "On a three-dimensional gait recognition system," Ph.D. dissertation, School Electron. Comput. Sci., Univ. Southampton, Southampton, U.K., 2010.

[20] S. Samangooei, M. Nixon, and B. Guo, "The use of semantic human description as a soft biometric," in *Proc. 2nd IEEE Biometrics, Theory, Appl. Syst.*, Oct. 2008, pp. 1–7.

[21] M. D. MacLeod, J. N. Frowley, and J. W. Shepherd, "Whole body information: Its relevance to eyewitnesses," in *Adult Eyewitness Testimony: Current Trends and Developments*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[22] C. N. Macrae and G. V. Bodenhausen, "Social cognition: Thinking categorically about others," *Annu. Rev. Psychol.*, vol. 51, no. 1, pp. 93–120, 2000.

[23] J. Hewig, R. H. Trippe, H. Hecht, T. Straube, and W. H. R. Miltner, "Gender differences for specific body regions when looking at men and women," *J. Nonverbal Behavior*, vol. 32, no. 2, pp. 67–78, 2008.

[24] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. New York, NY, USA: Academic, 2008.

[25] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.

[26] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[27] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[28] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *Proc. NIPS*, 2006, pp. 609–616.

[29] P. Tome, J. Fierrez, F. Alonso-Fernandez, and J. Ortega-Garcia, "Scenario-based score fusion for face recognition at a distance," in *Proc. IEEE CVPRW*, Jun. 2010, pp. 67–73.

[30] J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative multimodal biometric authentication based on quality measures," *Pattern Recognit.*, vol. 38, no. 5, pp. 777–779, May 2005.

Julian Fierrez received the M.Sc. and the Ph.D. degrees in telecommunications engineering from Universidad Politecnica de Madrid, Madrid, Spain, in 2001 and 2006, respectively. Since 2002, he has been with the Biometric Recognition Group, first at Universidad Politecnica de Madrid, and since 2004 at Universidad Autonoma de Madrid, where he is currently an Associate Professor. From 2007 to 2009, he was a Visiting Researcher with Michigan State University, USA, under a Marie Curie fellowship. His research interests and areas of expertise include signal and image processing, pattern recognition, and biometrics, with emphasis on signature and fingerprint verification, multi-biometrics, biometric databases, system security, and forensic applications of biometrics. He has been and is actively involved in European projects focused on biometrics (e.g., TABULA RASA and BEAT), and is a recipient of a number of distinctions for his research, including Best Ph.D. Thesis in Computer Vision and Pattern Recognition from 2005 to 2007 by the IAPR Spanish liaison, Motorola Best Student Paper at ICB 2006, the EBF European Biometric Industry Award 2006, the IBM Best Student Paper at ICPR 2008, and EURASIP Best Ph.D. Award 2012.

Ruben Vera-Rodriguez received the M.Sc. degree in telecommunications engineering from Universidad de Sevilla, Spain, in 2006, and the Ph.D. degree in electrical and electronic engineering from Swansea University, U.K., in 2010. Since 2010, he has been with the Biometric Recognition Group - ATVS, Universidad Autonoma de Madrid, Spain, first as the recipient of a Juan de la Cierva postdoctoral fellowship from the Spanish Ministry of Innovation and Sciences, and is currently an Assistant Professor. His research interests include signal and image processing, pattern recognition, and biometrics. In 2007, he received the Best Paper Award at the Fourth International Summer School on Biometrics, Alghero, Italy, by top international researchers in the field.

Pedro Tome received the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical engineering from Universidad Autonoma de Madrid (UAM), Spain, in 2008 and 2013, respectively. Since 2007, he has been with the Biometric Recognition Group - ATVS, UAM, where he is currently a Postdoctoral Researcher. He has carried out different research internships in worldwide leading groups in biometric recognition such as Image and Information Engineering Laboratory, Kent University, Canterbury U.K., CSPC - Communication Signal Processing and Control Group from Southampton University, U.K., and Security and Surveillance Research Group - SAS from University of Queensland, Australia. His research interests include signal and image processing, pattern recognition, computer vision, and biometrics. His current research is focused on biometrics at a distance and video-surveillance, using face and iris recognition and he is actively involved in forensic face evaluation.

Mark S. Nixon is a Professor of computer vision with the University of Southampton, U.K. His research interests are in image processing and computer vision. His team develops new techniques for static and moving shape extraction which have found application in automatic face and automatic gait recognition and in medical image analysis. His team were early workers in face recognition, later came to pioneer gait recognition and more recently joined the pioneers of ear biometrics. Amongst research contracts, he was a Principal Investigator with John Carter on the DARPA supported project Automatic Gait Recognition for Human ID at a Distance and he was previously with the FP7 Scovis project and is currently with the EU-funded Tabula Rasa project. His vision textbook, with A. Aguado, *Feature Extraction and Image Processing* (Academic Press) reached 3rd Edition in 2012 and has become a standard text in computer vision. With T. Tan and R. Chellappa, their book *Human ID Based on Gait* is part of the Springer Series on Biometrics and was published in 2005. He has been a chair or program chair of many conferences (BMVC 98, AVBPA 03, IEEE Face and Gesture FG06, ICPR 04, ICB 09, IEEE BTAS 2010) and given many invited talks. He is a member of the IAPR TC4 Biometrics and the IEEE Biometrics Council. He is a fellow of IET and FIAPR.