JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, 2015
doi: 10.1111/1556-4029.12800
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

DIGITAL & MULTIMEDIA SCIENCES

Pedro Tome,[1] Ph.D.; Julian Fierrez,[1] Ph.D.; Ruben Vera-Rodriguez,[1] Ph.D.; and Javier Ortega-Garcia,[1] Ph.D.

Combination of Face Regions in Forensic Scenarios*

ABSTRACT: This article presents an experimental analysis of the combination of different regions of the human face on various forensic scenarios to generate scientific knowledge useful for the forensic experts. Three scenarios of interest at different distances are considered comparing mugshot and CCTV face images using MORPH and SC face databases. One of the main findings is that inner facial regions combine better in mugshot and close CCTV scenarios and outer facial regions combine better in far CCTV scenarios. This means, that depending of the acquisition distance, the discriminative power of the facial regions change, having in some cases better performance than the full face. This effect can be exploited by considering the fusion of facial regions which results in a very significant improvement of the discriminative performance compared to just using the full face.

KEYWORDS: forensic science, forensic face, face recognition, facial regions, biometrics, facial components, biometrics at a distance

Automatic face recognition systems are generally designed to match images of full faces. However, in practice, the full face is not always available, e.g., due to occlusions and other variability factors. On the other hand, in forensics, the examiners usually carry out a manual inspection of the face images, focusing their attention not only on the full face but also on individual traits. They carry out an exhaustive morphological comparison, analysing the face region by region (e.g., nose, mouth, eyebrows, etc.), even examining traits such as marks, moles, wrinkles, etc.

For these reasons, there are some previous works where facial region-based recognition is studied [1,2] but none of them focus their attention in the regions normally considered by forensic experts apart from our previous works [3,4]. In this work, we have extracted facial components (called from now on facial regions) following forensic protocols from law enforcement laboratories, allowing us to study individually the different facial regions from a human face. In particular, we address in this letter the problem of combining the most discriminative areas of the face for recognition on different acquisition scenarios.

Understanding how different facial regions are combined on different forensic scenarios has some remarkable benefits, for example: (i) allowing investigators to work only with particular regions of the face or (ii) preventing that incomplete, noisy, and missing regions degrade the recognition accuracy. Furthermore, a better understanding of the combination of facial regions should facilitate the study of facial regions-based face recognition. Therefore, the fusion of the different facial regions is performed achieving significant improvements of performance compared to a traditional face recognition system based only on the face as a whole.

Facial Regions Extraction

The proposed facial regions extraction framework is described in detail in [3] and extended in [4]. In this framework, two kinds of region extraction are defined: (i) based on human facial proportions and (ii) based on facial landmarks.

The algorithm for extracting facial regions is an iterative method that takes advantage of the facial landmarks tagged by a human examiner or an automatic system to extract a number of $R = 15$ facial regions based on the forensic practice. The main difference with other extraction techniques is that in this case the extraction can be done on controlled and uncontrolled images thanks to the use of facial landmarks and proportions. This fact allows the algorithm to be suitable to be used in face biometric systems at a distance working with facial landmarks easily tagged with an automatic system (regardless of the facial landmarks tagging systems, or the type of matcher being used). In this work, the facial regions extraction is carried out just considering the facial landmark extractor using $L = 21$ facial landmarks manually tagged by a human examiner.

The extractor based on facial landmarks allows extracting the facial regions with high precision by the correctly located facial landmarks (manually in this case). A facial region is extracted by estimating the center between each one of two facial landmarks per facial trait and by applying a vertical and horizontal offset to generate a bounding box that contains the facial region.

The final region extraction result is the set of 15 facial regions based on forensic laboratories protocols (Spanish Guardia Civil

[1]Biometric Recognition Group - ATVS, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco, Madrid 28049, Spain.

*Partially supported by Spanish Guardia Civil and projects BBfor2 (FP7-ITN-238803), Bio-Challenge (TEC2009-11186), Bio-Shield (TEC2012-34881), Contexts (S2009/TIC-1485), and TeraSense (CSD2008-00068).

Received 11 Dec. 2013; and in revised form 10 April 2014; accepted 2 June 2014.

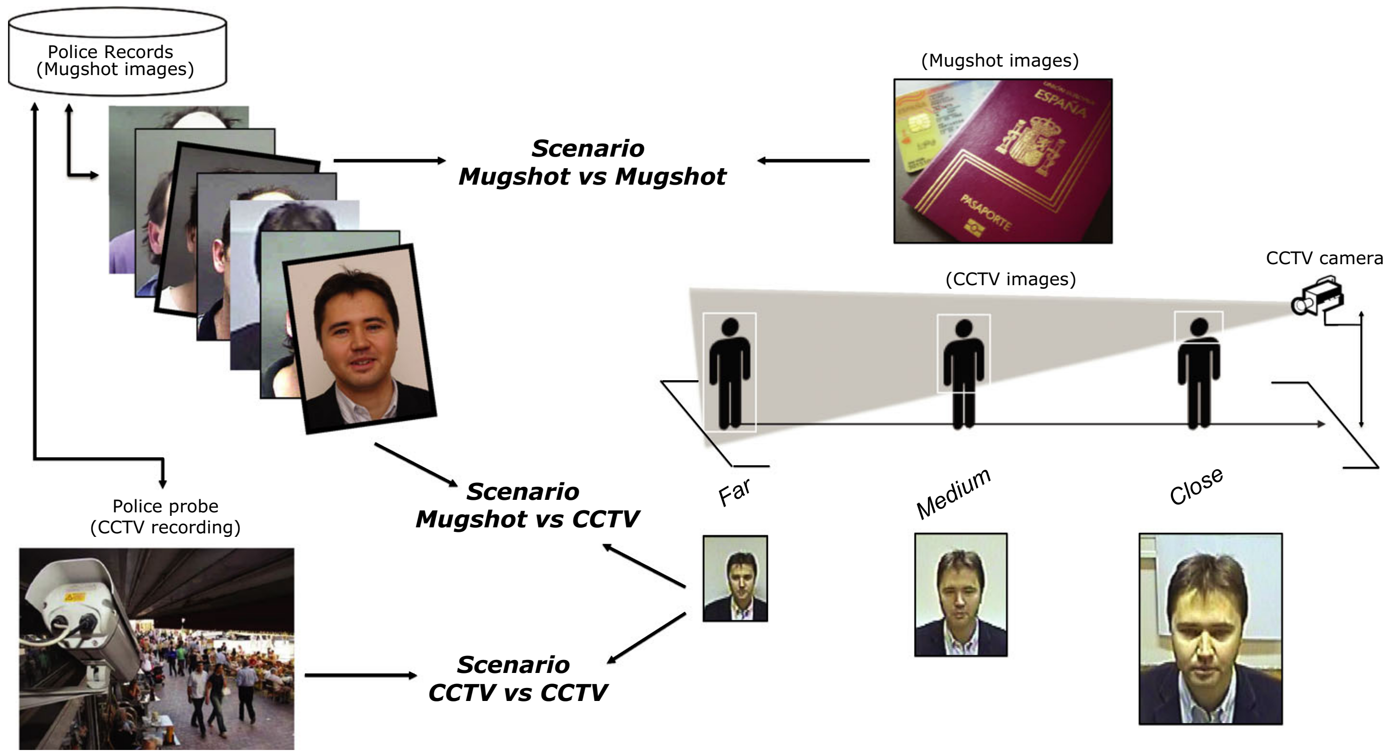

FIG. 1—*Experimental framework describing the three forensic scenarios analysed.*

– DGGC and Netherlands Forensic Institute – NFI)[1,2] as shown in Fig. 2.

Experimental Protocol

Once each facial region has been extracted, Eigen-regions (Principal Component Analysis, PCA) from each facial region are computed. Then, similarity scores are computed in this PCA vector space (dimension 200, retaining 98% of the energy of the original Eigen-region space) using a Support Vector Machine (SVM) classifier with a linear kernel. The experimental protocol followed is described with more detail in [4].

Both databases used in our experiments (MORPH [5] and SC face [6], see some examples in Fig. 3), were divided into three subsets based on the subject ID: development (1–43), SVM training (44–87), and test (88–130). These three subsets were used for training the PCA features, as impostors in the training of SVMs, and for testing the final system performance respectively.

This work studies the three main scenarios considered in the regular practice of a forensic examiner (see Fig. 1), where images of different quality might be compared: (i) mugshot versus mugshot, (ii) mugshot versus CCTV, and (iii) CCTV versus CCTV. In addition, three distances between subject and camera typical in practical applications are analysed: *close*, *medium,* and *far* distance (see Figs 1 and 3 top).

In the experimental protocol followed in this work, first the face is divided in 15 regions, and recognition results based on the Equal Error Rate (EER) are achieved for each facial region individually. Then, the individual facial regions are combined based on their individual results of the EER to improve the final recognition rate.

[1]DGGC – http://www.guardiacivil.es/.
[2]NFI – http://www.forensicinstitute.nl.

Facial Regions Fusion

The fusion of the 15 forensic facial regions in comparison with the performance of the full *face* region is performed. The goal of these fusion experiments is to analyse the recognition performance when combining different information available inside of a human face.

The performance results of each individual facial region can be seen in detail in the previous work [4]. Table 1 also shows the EER result for the best individual region in each scenario.

The fusion is carried out at the score-level for various combinations of regions obtained via sequential search. In particular, the 15 facial regions are fused using a parallel fusion approach based on the sum rule [7,8], starting from the most discriminative region individually, then fusing this region with all the rest and keeping the best fusion of two regions, and continuing this process until all the regions are fused. Figs 4, 5, and 6 show the results, where the order of the facial regions is different in each scenario. In this work, the fusion results are reported only for the case of manual landmark tagging with an extractor based on facial landmarks. Results with similar trends were obtained for the other configurations described in [4].

Before carrying out the fusion, scores of the different facial regions are first normalized to the [0, 1] range using the tan h-estimators described in [9], with $C = 0.01$, and μ_{SD} and σ_{SD} are, respectively, the estimated mean and standard deviation of the genuine score distribution using the development and SVM training sets.

Next, fusion results for the three scenarios considered are presented.

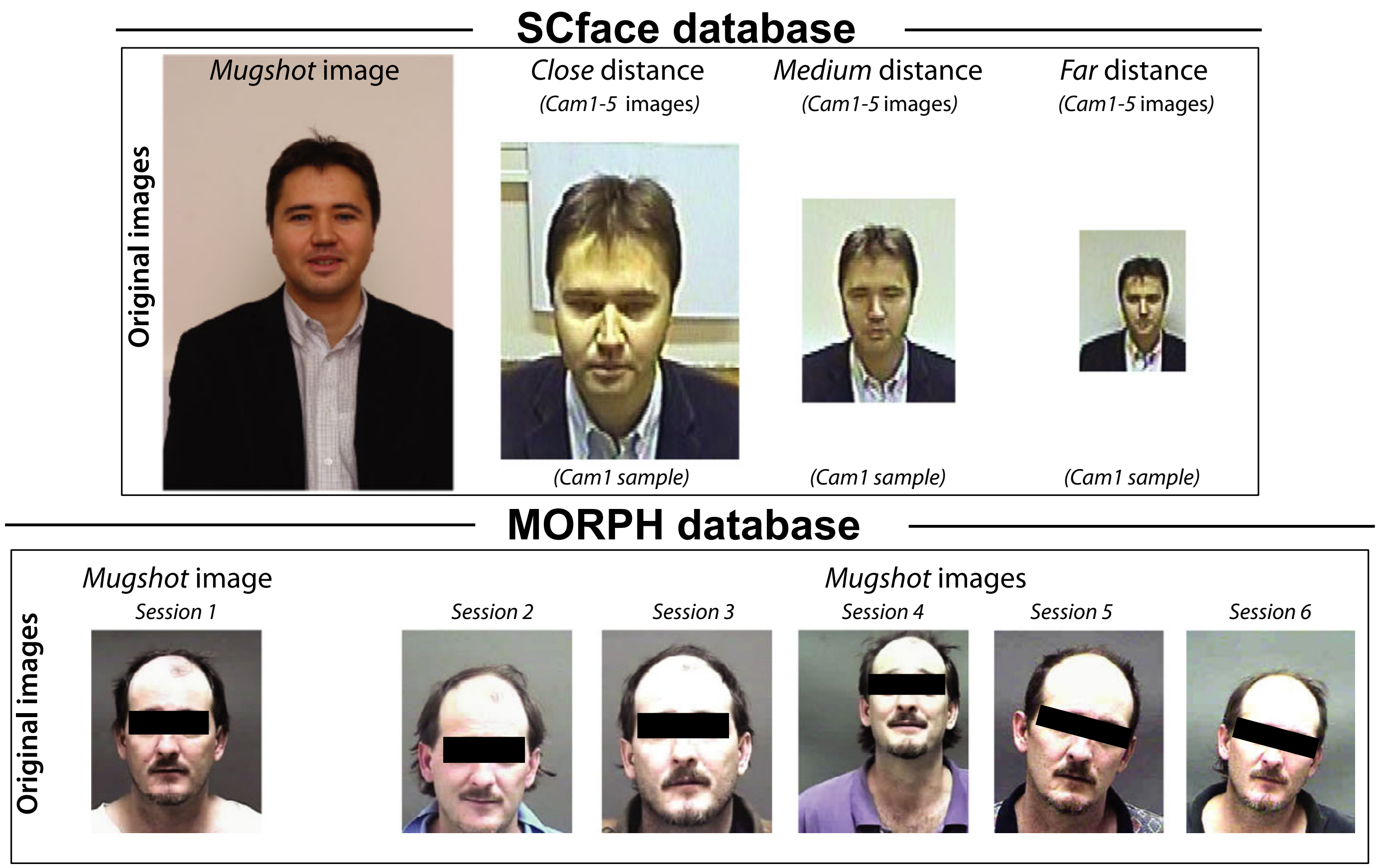

FIG. 2—*(Top) SC face image samples of each datasets for mugshot and Cam1 images for the close, medium, and far distance. (Bottom) MORPH image samples of each session.*

Mugshot versus Mugshot

This experiment presents the fusion results in the mugshot versus mugshot scenario using MORPH database [5]. Results of this fusion process can be seen in Table 1 and Fig. 4 where the full sequence of combination is shown.

The best fusion is reached using the *full* face and the following six facial regions: inner facial traits (*eyebrows*, *nose*, and *left* and *right eye*) and the two *middle faces*. A relative improvement of 51.5% in the EER is obtained with the fusion (from 13.5% to 6.61% EER) compared to only using the *face* region. The main reason of the performance degradation when considering a large number of facial regions is that some of them have very low discriminative power (e.g., *left* and *right ear* regions, very unstable due to variable hair occlusions). When incorporating such low performance regions to the fusion, the result is an overall performance worse than not having considered them.

In previous works [3,4] we showed that inner facial traits provided the best individual performance in the mugshot versus mugshot scenario. Hence, it is reasonable that the fusion of the full *face* with the inner facial regions produces the best recognition performance in this case.

Mugshot versus CCTV

This scenario is analysed using the SC face database [6]. The fusion results obtained for the three distances are shown in Table 1 and Fig. 5. Similar to the previous case, the system performance improves fusing several facial regions compared to just using the full *face* region. This is the most challenging scenario considered, as there is a large difference between images tested (mugshot vs. CCTV).

Close and *medium* distance scenarios combine seven facial regions to achieve the best result, but the *far* scenario needs to combine a total of 10 facial regions to obtain it.

It is interesting to note that in the *close* scenario the best result is obtained with the fusion of inner and outer facial traits together with the full *face* (relative improvement of 56.7% in the EER with respect to using only the full *face* is achieved). Similarly, in the two other distances considered, the best fusion includes inner and outer parts of the face, and relative improvements of over 40% in the EER are obtained with the fusion of regions compared to using only the full *face*.

As can be seen in the EER results of the fusion, this scenario results in significantly worse performance compared to the previous and following scenarios. This is mainly due to the differences between gallery (mugshot) and probe images (CCTV).

Due to that this is the most challenging scenario, in (10) we studied how to further improve the recognition performance in this case by using the color information of the images. Specifically, a combination of the color information of the different facial regions in different color spaces was studied, obtaining best results of 10.8%, 10.8%, and 14.5% of EER (for close, medium, and far distances respectively).

CCTV versus CCTV

Table 1 and Fig. 6 show the fusion results obtained for the three distances analysed for the CCTV versus CCTV scenario (using the SC face database [6]). As can be seen, when the

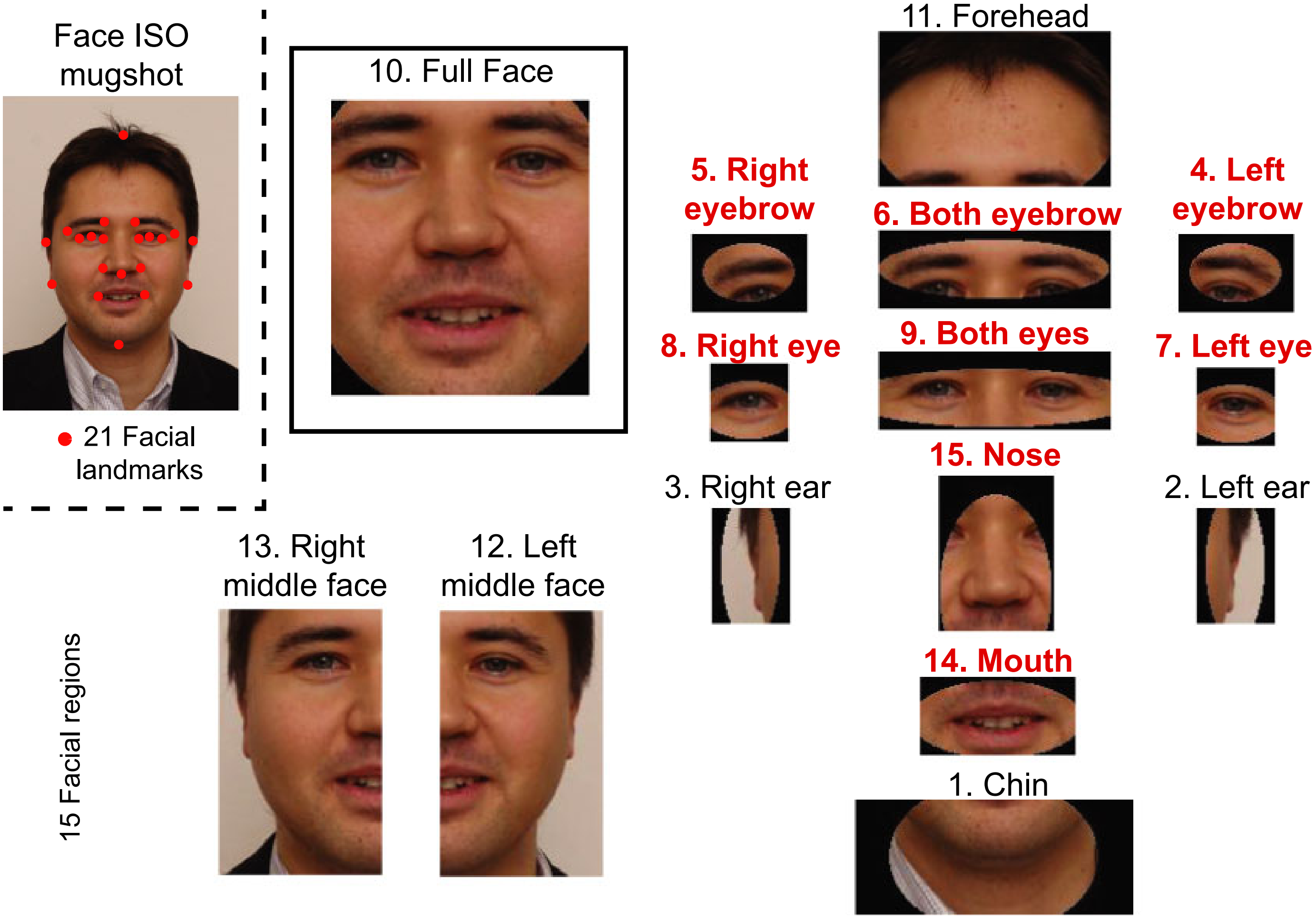

FIG. 3—*The 15 facial regions obtained with the extractor based on facial landmarks manually tagged (red dots). Inner regions (4–9, 14, and 15) are highlighted in red.*

TABLE 1—*Overview of EER results obtained for the full face, the best individual facial region, and the proposed fusion. This is given for the three scenarios considered: Mugshot versus Mugshot, Mugshot versus CCTV, and CCTV versus CCTV scenarios. Fig. 2 shows the facial regions with their corresponding id number (e.g. the id numbers: 10, 6, 12, correspond to full face, eyebrows, and left middle face respectively). Inner facial regions are highlighted in bold.*

				Best Combination		
Scenarios	Full Face EER	Best Individual EER (Region Id)	Facial Regions Fused via Sequential Search (Best Combination)	# Regions Fused	Fusion EER	Relative Improvement Over Full Face
Mugshot vs. Mugshot	13.50%	13.50% (10)	(**10**, 6, 12, 15, 13, 8, 7), 5, 3, 9, 11, 14, 4, 2,1	7	6.61%	51.5%
Mugshot vs. CCTV						
Close	33.10%	22.89%(15)	(15, 14, 8, 2, 3, 11, **10**), 13, 1, 5, 12, 9, 6, 7, 4	7	14.30%	56.7%
Medium	31.20%	27.08%(14)	(14, 11, 12, 2, 15, 3, 1), **10**, 8, 4, 13, 7, 6, 9, 5	7	12.90%	58.6%
Far	28.90%	27.49%(11)	(11, 2, **10**, 1, 3, 5, 12, 6, 14, 15), 13, 7, 4, 8, 9	10	16.80%	41.8%
CCTV vs. CCTV						
Close	8.24%	8.24%(10)	(**10**, 14, 11, 5, 15, 1, 13), 3, 12, 9, 4, 2, 6, 7, 8	7	2.42%	70.6%
Medium	15.20%	15.20%(10)	(**10**, 11, 14, 15, 1, 3, 5), 12, 4, 2, 13, 6, 9, 7, 8	7	2.52%	83.4%
Far	20.40%	17.25%(11)	(11, **10**, 1, 12, 2, 6, 14, 15, 5), 13, 4, 8, 7, 9, 3	9	7.07%	65.3%

acquisition distance increases more facial regions need to be fused with the full *face* region to achieve the best performance. Thus, an increment of variability and complexity involves more information to be fused, as could be expected.

A combination of inner (*mouth*, *nose*, and *right eyebrow*) and outer (*forehead*, *chin*, and *right ear*) facial regions is the best combination in this case. *Close* and *medium* scenarios just need seven facial regions to achieve the best performance combining first the inner facial regions. On the other hand, the *far* scenario again needs a higher number of facial regions to reach the best fusion result as also happened in the previous section in *far* distance. As can be observed the outer facial regions have more discriminant information than the inner regions. In this case, relative improvements of 70.6%, 83.4%, and 65.3% in the EER

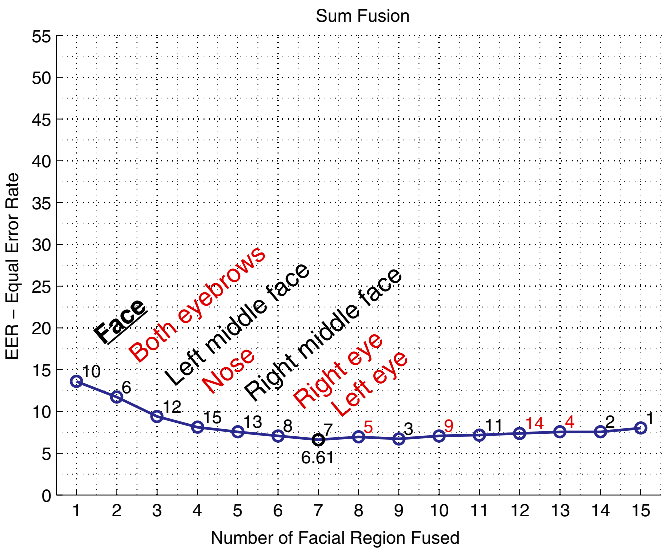

FIG. 4—*EER for sum fusion of the best combination of different facial regions for the Mugshot versus Mugshot scenario. Inner facial regions are highlighted in red. Fig. 2 shows the facial regions with their corresponding id number.*

for the *close*, *medium*, and *far* scenarios are achieved, respectively, for the proposed fusion of regions compared to only using the full *face* for recognition. In this case the EER results after the fusion of facial regions are better as the images compared are more similar than in the previous scenario.

Discussion

This article reports a study of the combination of 15 human facial regions on various forensic scenarios. The best fused performance of facial regions is compared with the full face region, which is the normal case in face recognition. Preliminary results show that a combination of a set of facial regions can significantly improve the system performance by total average improvement of 51.5%, 52.3%, and 73.1% in the three scenarios considered, namely: mugshot versus mugshot, mugshot versus CCTV, and CCTV versus CCTV. The potential of fusion of facial regions on these scenarios has been demonstrated to significantly improve a traditional full *face* based recognition system.

One of the main conclusions extracted with this work is that depending on the scenario considered, the best combination of the facial regions is different. Based on the results obtained, inner facial regions combine better in mugshot and close CCTV scenarios and outer facial regions combine better in far CCTV scenarios. Also, it is interesting to note that a larger number of facial regions is required to achieve the optimal results for far distance comparisons.

In addition to being useful background information that can guide and help experts to interpret and evaluate face evidences, these findings can have a significant impact on the design of

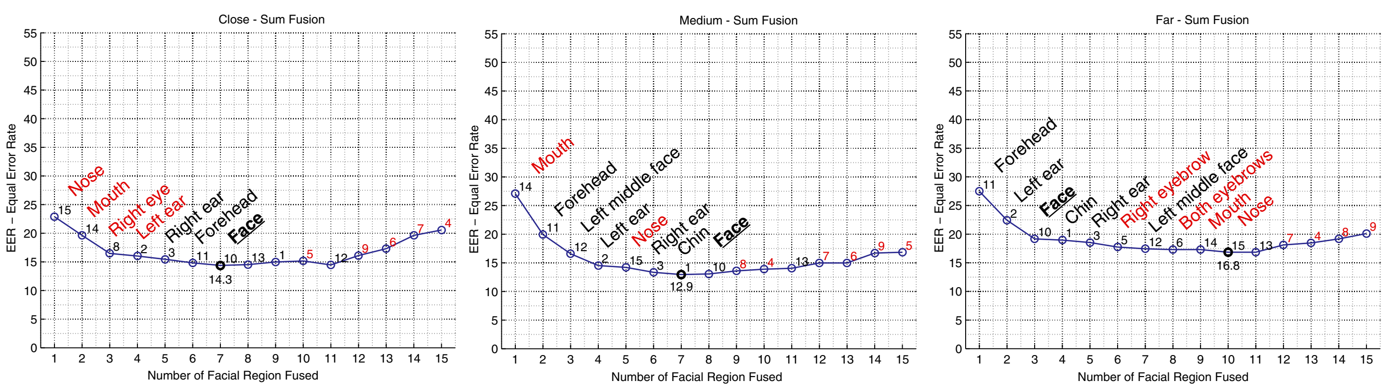

FIG. 5—*EER for sum fusion of the best combination of different facial regions for the three distance scenarios: close, medium, and far for the Mugshot versus CCTV scenario. Inner facial regions are highlighted in red. Fig. 2 shows the facial regions with their corresponding id number.*

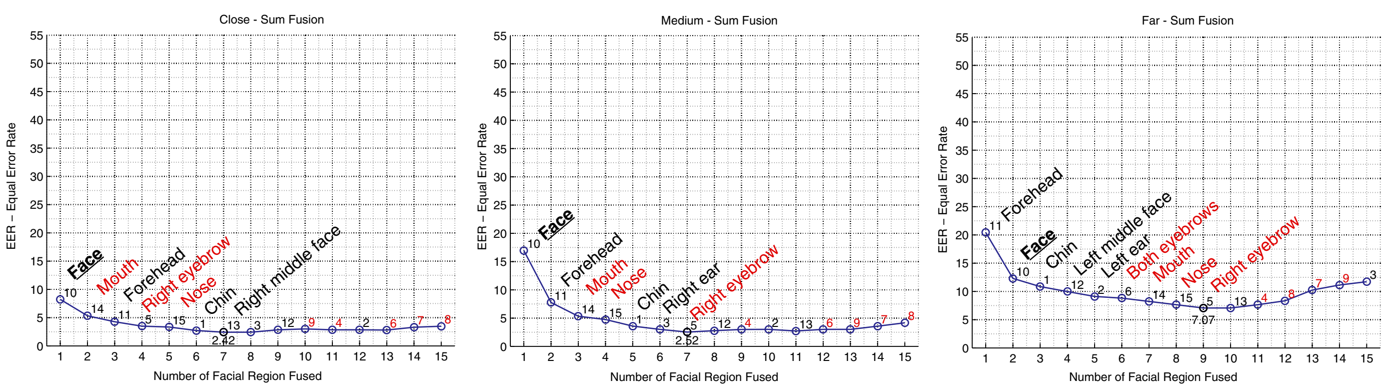

FIG. 6—*EER for sum fusion of the best combination of different facial regions for the three distance scenarios: close, medium, and far for the CCTV versus CCTV scenario. Inner facial regions are highlighted in red. Fig. 2 shows the facial regions with their corresponding id number.*

face recognition algorithms. In particular, the approach followed for combining the information provided by the different regions can be significantly improved using more sophisticated fusion approaches (e.g., quality-based [11], user-dependent [12]), and using more robust facial feature descriptors.

References

1. Ocegueda O, Shah SK, Kakadiaris IA. Which parts of the face give out your identity? Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2011 June 20-25; Providence, RI. Piscataway, NJ: IEEE, 2011;641–8.
2. Bonnen K, Klare B, Jain AK. Component-based representation in automated face recognition. IEEE Trans Inf Forensic Sec 2013;8(1):239–53.
3. Tome P, Blazquez L, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J, Exposito N, et al. Understanding the discrimination power of facial regions in forensic casework. Proceedings of International Workshop on Biometrics and Forensics (IWBF); 2003 April 4-5; Lison, France. Piscataway, NJ: IEEE, 2013;1–4.
4. Tome P, Fierrez J, Vera-Rodriguez R, Ramos D. Identification using face regions: application and assessment in forensic scenarios. Forensic Sci Int 2013;233:75–83.
5. Ricanek K, Tesafaye T. MORPH: a longitudinal image database of normal adult age-progression. Proceedings of the 7th International Conference on Automatic Face Gesture Recognition (FGR); 2006 April 2-6; Southampton, U.K. Piscataway, NJ: IEEE, 2006.
6. Grgic M, Delac K, Grgic S. SCface – surveillance cameras face database. Multimed Tools Appl 2011;51(3):863–79.
7. Fierrez J. Adapted fusion schemes for multimodal biometric authentication [Ph.D. thesis]. Mardrid, Spain: Univ. Politecnica de Madrid, 2006.
8. Kittler J, Hatef M, Duin RP, Matas J. On combining classifiers. IEEE Trans Pattern Anal Mach Intell 1998;20:226–39.
9. Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems. Pattern Recogn 2005;38(12):2270–85.
10. Tome P, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J. Fusion of facial regions using color information in a forensic scenario. In: Ruiz-Shulcloper J, Sanniti di Baja G, editors. Proceedings of 18th IberoamericanGongrees, CIARP, LNCSVol8259; 2013 Nov 20-23; Havana, Cuba. Berlin, Germany: Springer-Verlag Berlin Heidelberg, 2013;399–406.
11. Nandakumar K, Chen Y, Jain AK, Dass SC. Quality-based score level fusion in multibiometric systems. Proceedings of the 18thInternational Conference on Pattern Recognition (ICPR’06); 2006 Aug 20-24; Hong Kong. Piscataway, NJ: IEEE, 2006;473–6.
12. Fierrez-Aguilar J, Garcia-Romero D, Ortega-Garcia J, Gonzalez-Rodriguez J. Adapted user-dependent multimodal biometric authentication exploiting general information. Pattern Recogn Lett 2005;26(16):2628–39.

Additional information and reprint requests:
Pedro Tome, Ph.D.
BiometricRecognitionGroup– ATVS
Escuela Politecnica Superior
Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente
11 - Campus de Cantoblanco
28049 Madrid
Spain
E-mail: pedro.tome@uam.es