Fusion of Facial Regions using Color Information in a Forensic Scenario

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

Biometric Recognition Group-ATVS, EPS, Universidad Autonoma de Madrid
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
{pedro.tome, ruben.vera, julian.fierrez, javier.ortega}@uam.es

Abstract. This paper reports an analysis of the benefits of using color information on a region-based face recognition system. Three different color spaces are analysed (RGB, YC_bC_r, $l\alpha\beta$) in a very challenging scenario matching good quality mugshot images against video surveillance images. This scenario is of special interest for forensics, where examiners carry out a comparison of two face images using the global information of the faces, but paying special attention to each individual facial region (eyes, nose, mouth, etc.). This work analyses the discriminative power of 15 facial regions comparing both the grayscale and color information. Results show a significant improvement of performance when fusing several regions of the face compared to just using the whole face image. A further improvement of performance is achieved when color information is considered.

Keywords: Face recognition, facial regions, forensics, color information, facial components, video surveillance, at a distance.

1 Introduction

Automatic face recognition systems are generally designed to match grayscale 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 color face images, focussing their attention not only on the grayscale full face but also on individual traits and color information. 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.

There are some previous works where grayscale facial region-based recognition is studied [1–3] but non of them focus their attention in the color regions normally considered by forensic experts. 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 normally considered in current practice of forensic examiners. In particular, we address in this paper the problem of combining the most discriminative areas of the face for recognition using the available color information on a very challenging video surveillance scenario.
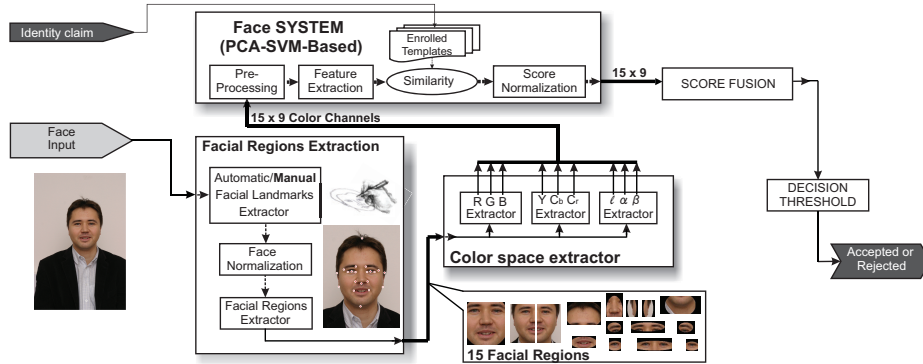
Fig. 1. Experimental framework diagram description.

In contrast to traditional grayscale systems, this paper studies the discriminative power of each facial region using three color spaces: RGB, YC_bC_r, and $l\alpha\beta$. Fig. 1 summarizes the experimental framework followed.

Understanding how different facial regions from different color spaces are combined on a very challenging scenario has some remarkable benefits, for example: *i*) allowing investigators to work only with particular regions of the face in different color spaces, *ii*) improving the face recognition performance using all available information from color images, or *iii*) preventing that incomplete, noisy, and missing regions degrade the recognition performance. Further, a better understanding of the combination of facial regions in different color spaces should facilitate the study of facial regions-based face recognition. Therefore, the fusion of the different facial regions from different color spaces is performed achieving significant improvements of performance compared to a traditional face recognition system based only on the grayscale face as a whole.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of the automatic facial region extraction procedure and presents the color spaces analysed. Section 3 describes the experimental protocol followed, the database and the verification system adopted for the experiments. Section 4 reports an experimental fusion of different facial regions using different color spaces. Finally, Section 5 draws some conclusions of our work.

2 Facial Regions Extraction and Color Methodology

The proposed facial regions extraction framework is described in detail in [4] and extended in [3]. In this framework, two kinds of regions extraction are defined: *i*) based on human facial proportions, and *ii*) based on facial landmarks. For this work, the second extractor based on facial landmarks has been adopted. This extractor, based on facial landmarks manually located, allows to extract the facial regions with high precision.
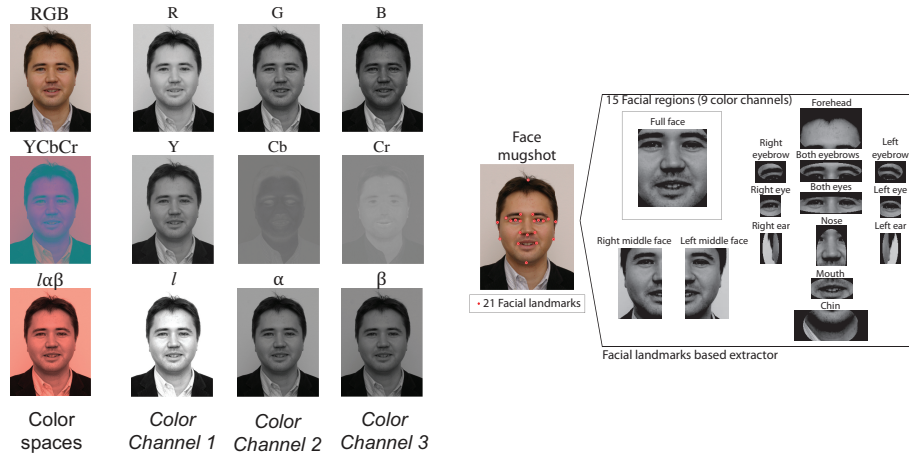
Fig. 2. (Left) Grayscale intensity values of faces for each color space analysed. (Right) Facial regions extraction based on facial landmarks extractor. The regions are extracted for the 9 color channels considered here.

Color Channel 1 Id Num.	Color Channel 2 Id Num.	Color Channel 3 Id Num.	Facial Region	Facial Region Size (h × w)
1	16	31	Chin	75x181
2	17	32	Left ear	75x51
3	18	33	Right ear	75x51
4	19	34	Left eyebrow	51x75
5	20	35	Right eyebrow	51x75
6	21	36	Both eyebrows	51x151
7	22	37	Left eye	51x51
8	23	38	Right eye	51x51
9	24	39	Both eyes	51x151
10	**25**	**40**	**Full face**	**192x168**
11	26	41	Forehead	101x151
12	27	42	Left middle face	173x106
13	28	43	Right middle face	173x106
14	29	44	Mouth	51x101
15	30	45	Nose	101x75

Table 1. Facial regions id for each color channel and their sizes for extractor based on facial landmarks (height × width in pixels).

The final region extraction result is the set of 15 facial regions (see Table 1) based on forensic laboratories protocols[1] as shown in Fig. 2.

There are some previous works where color spaces such as RGB or YC_bC_r have been used for face recognition [5, 6]. But, to the best of our knowledge, this is the first work where color information is used for face recognition using 15 facial regions.

When dealing with color images, the RGB color space is commonly used. This color space is composed by three channels (red, green, and blue), which are correlated among them. The components that form the second color space considered YC_bC_r

---

[1] Spanish Guardia Civil DGGC), http://www.guardiacivil.es/ and
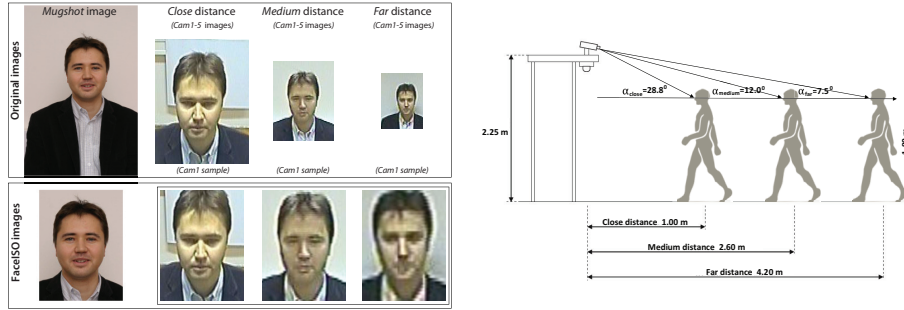Netherlands Forensic Institute (NFI), http://www.forensicinstitute.nl.

Fig. 3. (Left) SCface image samples of each datasets for *mugshot* and *Cam*1 images, and their corresponding normalized face for *close*, *medium*, and *far* distances. (Right) The three acquisitions distances: *close*, *medium* and *far*. Acquisition angle of each distance calculated for a subject with mean height of 1.80 meters.

are as follows: Y, luminance component, C_b, blue component $(B - Y)$, and C_r, red component $(R - Y)$ [7].

Both RGB and YC_bC_r color spaces have correlated color channels among them. We also consider the $l\alpha\beta$ color space [8], which minimizes the perceptual correlation among the channels of an image. The parameter l represents the luminance or brightness of the image and α and β represent the chromatic content, i.e., the color information. Fig. 2 (left) shows an example of each color channel for these three color spaces considered in the experiments.

3 Experimental Protocol

Once each facial region has been extracted from each color channel, Principal Component Analysis (PCA) is computed obtaining eigen-regions. Then, similarity scores are computed in this PCA vector space (dimension 200, retaining an average of 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 in more detail in [3].

The database used in our experiments SCface [9], (see Fig. 3 (left)), was divided into 3 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. The procedure followed is summarized in Table 2.

Fig. 3 (left) shows an example of a mughost image, and the images acquired by one of the surveillance cameras. As can be seen there is a considerable scenario variation in terms of quality, pose and illumination. The change in the pose is specially important due to the different angles between the person and the cameras as shown in Fig. 3 (right). In this work a very challenge scenario of videosurveillance is studied considering a common case that a forensic examiner can find in practice: *mugshot vs CCTV* images.

SCface DB (130 Subjects) - Mugshot vs CCTV protocol			
Subsets	1…43 Subject (43 Subjects)	44…87 Subject (44 Subjects)	88…130 Subject (43 Subjects)
Mugshot			SVM Training (Clients)
Cam 1			
Cam 2	Development set (PCA subspace)	SVM Training (Impostors)	Test
Cam 3			
Cam 4			(Clients/Impostors)
Cam 5			

Table 2. Partitioning of the SCface DB according to the *Mugshot vs CCTV images* protocol.

In addition, three distances between subject and camera typical in practical applications are analysed: *close*, *medium* and *far* distances (see Fig. 3 (right)).

4 Facial Regions Fusion

This section describes the fusion of the 15 forensic facial regions extracted from a human face in comparison with the performance of the whole *face* region normally used in face recognition systems. The fusion is carried out at score–level combining the facial regions for the color channels considered here.

Before carrying out the fusion, scores of the different facial regions are first normalized to the $[0, 1]$ range using the tanh-estimators described in [10].

For this paper three different experiments were defined in order to analyse the potential of color information in a face recognition system: *i) Exp.1* Grayscale baseline system, where the grayscale facial regions are fused as the traditional face recognition systems. *ii) Exp.2* Fusion of color channels from each color space, (e.g. for RGB color space, the channels $\{R, G, B\}$ are fused for each facial region considered). *iii) Exp.3* Fusion of all color channels, where all 9 available color channels are fused for each face region.

4.1 *Exp.1* Grayscale (Baseline System)

The fusion is carried out at the score–level for various combinations of grayscale regions. In particular, the 15 facial regions are fused using a parallel fusion approach based on the sum rule [11], starting from the most discriminative, then fusing this trait with the rest and keeping the best fusion of two regions, and continuing this process until all the regions are fused.

The fusion results obtained for the three distances are shown in Table 3 (*Exp.1*). As can be seen the system performance improves fusing several facial regions compared to just using the full *face* region.

Close and *medium* distance scenarios combine 7 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).

	Color Space	Close Scenario Fusion (# Regions) – Full face	Medium Scenario Fusion (# Regions) – Full face	Far Scenario Fusion (# Regions) – Full face
Exp.1	*Grayscale*	14.30% (7) – 33.10%	12.90% (7) – 31.20%	16.80% (10) – 28.90%
	RGB	11.58% (12) – 32.19%	**10.79% (13)** – 30.21%	14.61% (15) – 29.96%
Exp.2	YC_bC_r	12.89% (16) – 29.50%	12.65% (8) – 33.35%	16.37% (21) – 31.72%
	$l\,\alpha\,\beta$	**10.79% (12)** – 31.82%	11.20% (16) – 31.09%	**14.50% (18)** – 28.93%
Exp.3	*ALL*	**9.03% (27)** – 29.96%	**10.33% (22)** – 30.33%	**13.12% (39)** – 28.93%

Table 3. EER results for the score–level fusion obtained for sequential region fusion and the full face for the color channels of the three color spaces. In brackets is indicated the number of regions fused.

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.

4.2 *Exp.2* Fusion of Three Color Channels

For the *Exp.2*, the score–level fusion is carried out fusing the three channels in a color space, i.e., $15\times3 = 45$ facial regions (as Table 1 shows) using a parallel fusion approach as in the previous experiment.

Table 3 (*Exp.2*) shows the fusion results for the three distances analysed. Fig. 4 shows the sequential fusion results obtained for the three distances and their corresponding color space with best performance ($l\alpha\beta$ for *close* and *far* distance, and RGB for *medium* distance). Similar to the previous case the system performance improves fusing several facial regions compared to just using the *full face* region. It is interesting to note that the number of regions fused to obtain the best performance increases with the distance between the subject and the camera.

Comparing the fusion results with the baseline system based on grayscale facial regions, relative improvements of performance of 24.5%, 16.3%, and 13.7% for *close*, *medium* and *far* distance, are achieved respectively. These results support the utility of color information using facial regions to improve the performance of traditional face recognition systems.

4.3 *Exp.3* Fusion of All Color Channels

In this case, all facial regions from all color channels are combined following the same fusion methodology. In this case, we combine the 3 sets of 45 facial regions considered in the previous experiment, i.e., 135 facial regions in total.

Table 3 (*Exp.3*) shows the fusion results for this experiment. As can be seen this experiment achieves the best EER results for the three distances compared to the previous experiment. However this case needs to fuse more facial regions to achieve the best performance (approximately double than *Exp.2*), and just around 1% EER of improvement is achieved compared to *Exp.2*. Again, the increment of the acquisition distance increases the number of facial regions to be combined to achieve the best performance.
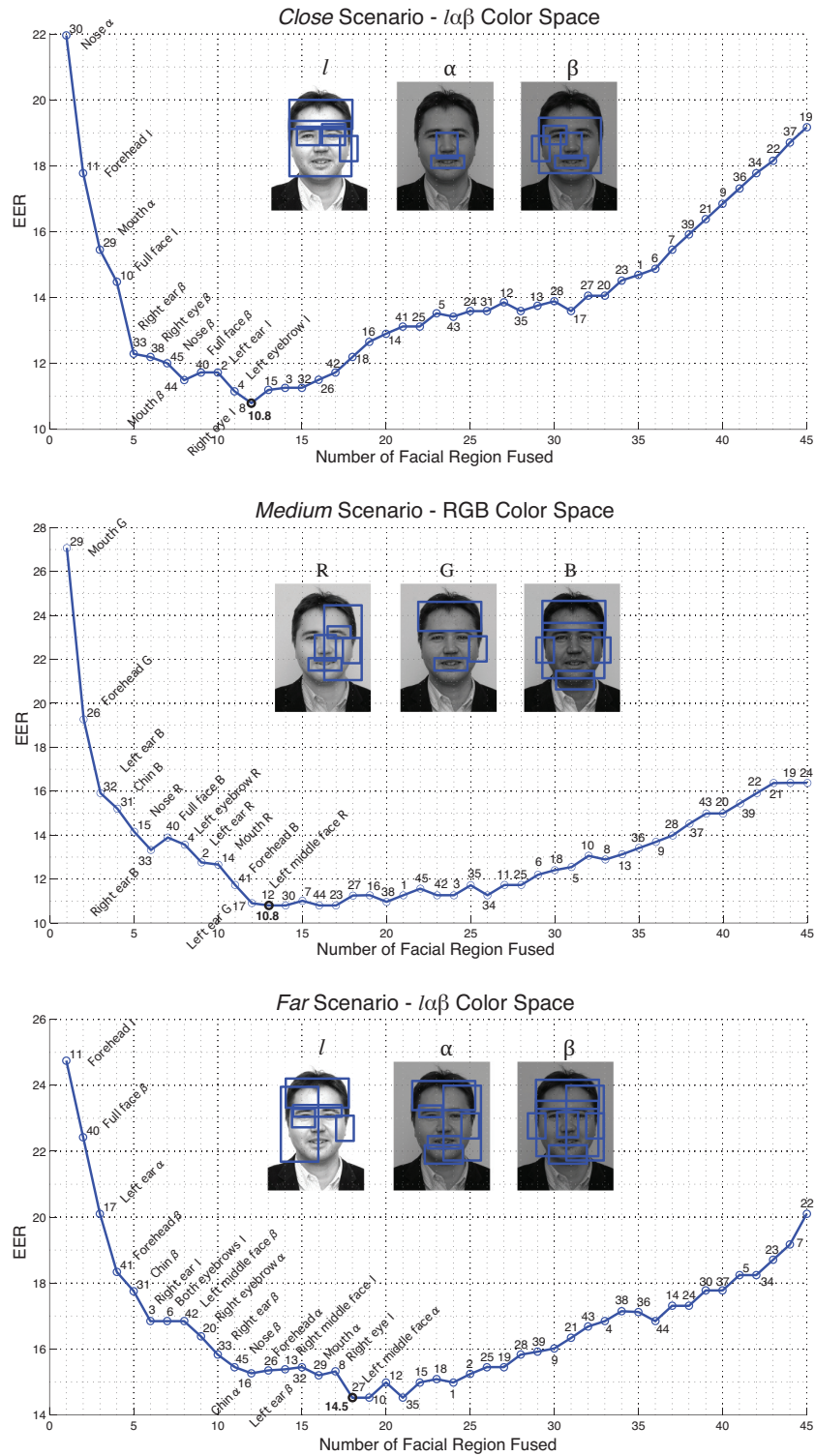
Fig. 4. EER for sum sequential fusion of the best combination of different facial regions for the best individual color space in each distance scenario: *close* ($l\alpha\beta$), *medium* (RGB) and *far* ($l\alpha\beta$).

Similarly, in the three distances considered, the best fusion includes inner and outer parts of the face, and relative improvements of over 66% in the EER are obtained with the regions fusion compared to only using the *full face*.

5   Conclusions

This paper reports an study of the combination of 15 human facial regions extracted from three different color spaces on a very challenging scenario comparing mugshot versus CCTV images. 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 in different color spaces can significantly improve the system performance by a relative average improvement of over 66% for the three distances considered. The combination of facial regions with color information allows to improve the system performance with a relative improvement of over 20% comparing with the traditional face recognition systems using only grayscale information. The potential of fusion of facial regions on these scenarios has been demonstrated to significantly improve a traditional full face recognition system performance.

Acknowledgements

References

1. Ocegueda, O., Shah, S.K., Kakadiaris, I.A.: Which parts of the face give out your identity? In: Proccedings of CVPR, IEEE (2011) 641–648
2. Bonnen, K., Klare, B., Jain, A.K.: Component-based representation in automated face recognition. IEEE Transactions on Information Forensics and Security **8**(1) (2013) 239–253
3. Tome, P., Fierrez, J., Vera-Rodriguez, R., Ramos, D.: Identification using face regions: Application and assessment in forensic scenarios. Forensic Science International, 2013 (2013) (Submitted).
4. Tome, P., Blazquez, L., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J., Exposito, N., Leston, P.: Understanding the discrimination power of facial regions in forensic casework. In: International Workshop on Biometrics and Forensics, Lisboa, Portugal (April 2013)
5. Singh, S.K., Chauhan, D.S., Vatsa, M., Singh, R.: A robust skin color based face detection algorithm, tamkang. Journal of Science and Engineering **6** (2003) 227–234
6. Liu, Z., Liu, C.: Robust face recognition using color information. In Tistarelli, M., Nixon, M., eds.: Advances in Biometrics. Volume 5558 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2009) 122–131
7. Gonzalez, R.C., Woods, R.E.: Digital Image Processing (3rd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA (2006)
8. Ruderman, D.L., Cronin, T.W., Chiao, C.C.: Statistics of coneresponses to natural images: implications for visual coding. Journal Optical Society Of America **15** (1998) 2036–2045
9. Grgic, M., Delac, K., Grgic, S.: Scface - surveillance cameras face database. Multimedia Tools Appl. **51**(3) (February 2011) 863–879
10. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. Pattern Recognition **38**(12) (December 2005) 2270–2285
11. Fierrez, J.: Adapted Fusion Schemes for Multimodal Biometric Authentication. PhD thesis, Univ. Politecnica de Madrid (2006)