On Combining Face Local Appearance and Geometrical Features for Race Classification

Fabiola Becerra-Riera[1(✉)], Nelson Méndez Llanes[1],
Annette Morales-González[1], Heydi Méndez-Vázquez[1], and Massimo Tistarelli[2]

[1] Advanced Technologies Application Center (CENATAV),
7A #21406 Siboney, Playa, 12200 Havana, Cuba
{fbecerra,nllanes,amorales,hmendez}@cenatav.co.cu
[2] Computer Vision Laboratory PolComing, University of Sassari,
Viale Mancini 5, 07100 Sassari, Italy
tista@uniss.it

Abstract. In the field of demographic attribute classification, race estimation is perhaps the least studied topic in the literature. CNN-based approaches report the best results to the day, but they are computational expensive for practical applications. We propose a simpler approach by combining local appearance and geometrical features to describe face images, and to exploit the race information from different face parts by means of a component-based methodology. Experimental results obtained in the FERET subset from EGA database, with traditional but effective classifiers like Random Forest and Support Vector Machines, are very close to those achieved with a recent deep learning proposal.

Keywords: Soft-biometrics · Race classification ·
Face appearance representation · Face anthropometric representation

1 Introduction

Visual attributes are one of the most intuitive and natural ways of describing a face. They can range from soft biometrics, which include demographic information (gender, age, race), facial marks and certain physical characteristics of the face; to other environmental related aspects. The estimation of visual attributes has been an active research topic in recent years because of their multiple applications in domains such as biometric authentication, access control, video surveillance and security systems. Soft biometrics can be useful in different ways: to perform recognition by means of a bag of attributes, to reduce the search space of a hard biometric system by restricting comparisons to those matching a certain soft biometric profile, and to complement the evidence from hard biometric traits [9].

Among the demographic attributes, race (ethnicity) is perhaps the least studied soft biometric in the literature. In particular, in the recognition of race, the

somatic traits of some populations are not well defined; within the same population, people may exhibit certain characteristics to a greater or lesser extent [5]: in the case of Caucasians, for example, the skin tone and the geometry of some facial features can vary from one individual to another. Taking these factors into account, it is clear that the accuracy of a race classifier is intrinsically linked to the robustness in the estimation of other attributes that characterize it by definition, such as skin color and face shape.

Different approaches have been proposed for race classification [7]. They range from using global [13] to local features [12], as well as other visual information such as skin and lips color, and forehead area [16]. The combination of local descriptors have also shown to be effective [4,17], despite recently the best state-of-the-art results in this topic have been achieved using Convolutional Neural Networks (CNN) [2,20]. However, deep learning approaches continue to be extremely demanding in terms of computing time and memory consuming during training and deployment, among different aspects that remains to be addressed to make them practical [6].

In this work we propose a simple but accurate method to race estimation. We first exploit the effectiveness of component-based approaches for attribute classification [11] and analyze the influence of different face regions for the specific problem of race estimation. Besides, we incorporate anthropometric information directly linked to the race definition itself [8]. We then evaluate different strategies for the fusion of both local appearance descriptors and geometrical features. Traditional classifiers are employed to obtain the best feature combination for the final prediction. The rest of this paper is organized as follows. Section 2 introduces the proposal. Section 3 presents the evaluation protocol and Sect. 4 the experimental analysis. Finally, Sect. 5 concludes the paper.

2 Proposed Approach

Since the definition of race from face includes the consideration of intrinsic attributes such as skin color and shape of the facial features, we base our face representation on both appearance and geometric characteristics. We analyze the impact of using color and texture features, and anthropometric measures separately and explore the best way of combining them in a more robust descriptor through the use of two different classifiers: Support Vector Machine (SVM) and Random Forest (RF). In the following subsections we explain in details how the face image is represented taking into account this two different features and the strategies used to combine them.

2.1 Appearance Features

The texton-like features [18] incorporate 17 filters (filterbank) to extract color and texture information that we exploit to obtain an appearance-based face representation for the estimation of race.

We subdivided a face image into 10 interest regions (see Fig. 1) to explore its influence separately for race estimation. We follow the procedure defined in [3] for extracting the regions and we include hair and contour components since visual information surrounding the face has proven to be important for attribute classification in the literature [1]. The filterbank features were employed to codify the face parts because of their good results for the classification of irregular regions [10]. For each region we consider only a set of sparse points to avoid redundant and expensive calculations, and extract color and texture information for each of them. The appearance representation of a single region is a 34-component descriptor where the mean and variance of the extracted filterbank features are concatenated; SVM and RF classifiers are used to find the best combination of regions in such a way that information provided by each one complements the others. The final representation of a face image is conceived as the concatenation of the best region feature vectors, which can result in a 340-component descriptor (34 region descriptor $*$ 10 regions) if the complete set of regions is used.
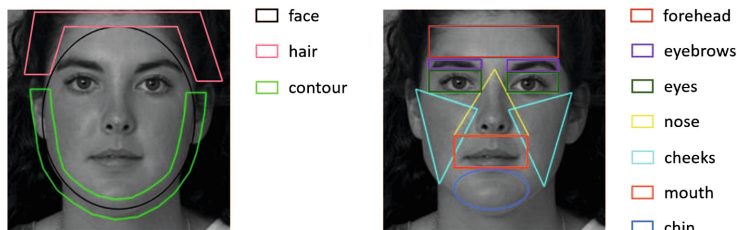

Fig. 1. Face image subdivision into 10 regions of interest: face, hair, contour, forehead, eyebrows, eyes, nose, cheeks, mouth and chin.

2.2 Geometrical Features

Anthropometric or shape-based methodologies have been widely used in literature to tackle race estimation problem [7]. Most of these approaches use 3D anthropometric statistics for race categorization. Hence, they recover the facial geometrical structure by using 3D face models. Obtaining these 3D models can be computational costly, so we have decided to use some distances between landmark points in 2D images that can be seen as geometric invariants in 3D models. This geometric representation was inspired in the work of [14], that explored multiple 2D/3D geometric invariants for face recognition.

We use 68 landmark points (control points) distributed around the face in the following way: 17 points for face contour, 12 for the eyes, 10 for the eyebrows, 9 corresponding to the nose and 20 to the mouth. Following the 2D/3D invariant measures described in [14] for the case of 2D images, we computed the ratio of distances of all possible combination of four and five non-coplanar or collinear control points. This led us to a high-dimensional vector that is reduced applying Principal Components Analysis (PCA). It is evident that some configurations

(distances or ratios) are more significant than others, and some of them can be redundant, so this allows us to obtain the best invariants for our problem. Some of the selected distances are illustrated in Fig. 2.

Fig. 2. Invariant distances for geometric face representation

2.3   Combination Strategies

Two different strategies for feature combination were explored as well as the influence of the selected classifiers (SVM, RF) in the final accuracy results. By means of the first strategy we concatenate appearance and geometric features in a single descriptor and validate its effectiveness in the estimation of race. The second strategy was inspired in the work of [19], in which different late fusion procedures were analyzed. In particular, we employed the geometric mean of the probabilistic outputs, that showed the best performance results in [19], outperforming even the more sophisticated fusion techniques. This second strategy has the additional advantage that different classifiers can be used for different features, therefore allowing the combination of their best performance individually.

3   Evaluation Protocol

Although there are several works in this topic, there is not a direct comparison among different approaches with a fixed protocol or database. Most researches use commonly accepted representative databases for face recognition, such as FERET [2]. However, these databases are usually race ill-balanced. For that reason we have decided to use the EGA database [15] which integrates different single race datasets to create a more heterogeneous and representative collection for race recognition (see Fig. 3).

The EGA dataset contains 2 345 images taken from CASIA-Face V5, FEI, FERET, FRGC, JAFFE and Indian Face Database. Images are labeled in terms of gender, 3 age groups (young, adult and middle age people) and 5 racial groups

(African-American, Asian, Caucasian, Indian, and Latin). Most of the images are frontal and they do not present illumination problems, nor occlusions (except for some cases using eyeglasses) nor facial expression variation. Since this dataset does not have a standard protocol for attribute classification, we designed our own. We split the total images into 5 folds balanced in terms of age, gender and race and performed a 5-fold cross validation. In the next section we explain in detail how the experiments were conducted.
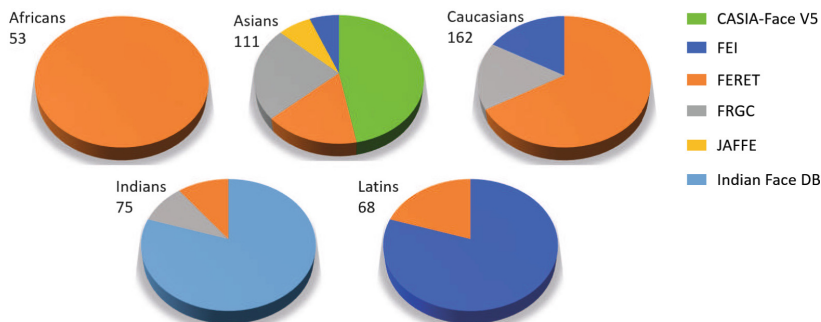

Fig. 3. Contribution (number of subjects) by race of each source dataset to EGA.

4   Results and Discussion

With the aim of showing the effectiveness of the proposed descriptors for race estimation, we performed several experiments in the EGA dataset.

First, we evaluated multiple region combinations codified with filterbank features, to find the subset that contributes in a most significant way for race estimation. Our experiments showed that the eyes-cheeks-chin-face-hair combination (using a 170-dimensional vector after the concatenation) achieved the best general performance with a good balance between classes. In Table 1, each reported result that use the appearance filterbank features was obtained with this best region combination.

Before exploring feature combination strategies, we evaluated the appearance and geometric features separately for the race estimation task. In the case of geometric features we selected 150 components after applying PCA, in order to be similar in dimension to the appearance-based descriptor. As can be seen in Table 1, filterbank features showed the higher accuracy by means of a RF classifier (row 1), while in the case of the geometric measures SVM achieved a better separation between classes (row 4). In general, the best results with individual features was obtained with the SVM classifier (83.7% for geometrics).

We conducted a second group of experiments taking into account the different strategies to combine the face feature representations. Although the SVM classifier achieved superior results with single features, its general performance taking into account both appearance and geometric descriptors was similar to

the one obtained with the RF classifier which, in the correct classification of Latin people, was more accurate in all the experiments. This previous result reinforce the fact that, by using the first combination strategy, denoted as FB + Geom in Table 1, RF classifier obtained the most accurate race estimation (83.1%), with a better balance between the 5 classes, and the highest results in the classification of Latins (80.2%), which makes a great difference in relation to the 66.8% achieved by SVM.

Table 1. Race classification accuracy on EGA dataset

Features	Classifier	Accuracy					
		General	African	Asian	Caucasian	Indian	Latin
filterbank (FB)	RF	82.9%	87.9%	78.7%	85.2%	84.3%	77.9%
filterbank (FB)	SVM	75.6%	74.7%	77.1%	78.4%	81.3%	60.6%
geometrics (Geom)	RF	75.8%	78.1%	69.7%	81.9%	75.5%	69.7%
geometrics (Geom)	SVM	83.7%	89.8%	85.7%	89.1%	82.9%	63.8%
FB + Geom	RF	83.1%	87.9%	78.1%	84.6%	85.9%	**80.2%**
FB + Geom	SVM	81.4%	84.5%	78.9%	85.3%	86.7%	66.8%
gMean FB + Geom	RF + SVM	**87%**	**90.9%**	**87.5%**	**90.4%**	**87.5%**	74.7%

With the second strategy we fused appearance and anthropometric measures by employing the geometric mean (gMean) of their probabilistic results. We used RF classifier with the filterbank representation and SVM with geometric features, according to the results obtained in the first set of experiments. This late fusion strategy reported the best results in Table 1, achieving a general performance of 87% of accuracy; with around 91% of effectiveness in the estimation of Africans and Caucasians and an improvement to 87.5% in the case of Asians and Indians.

We also compared our proposal with a recent CNN approach based on the VGG-architecture [2], in the FERET subset from the EGA dataset. In this case, similar to previous works, we focused on the estimation of only 3 classes: Black, Asian and White people. In Table 2 we show our results by employing single descriptors and their combinations.

It can be seen that the 93.7% of accuracy achieved by using our late fusion of local descriptors is very close to the 94% obtained with the Anwar and Islam deep learning solution, with only 0.3% of average accuracy difference. However, our individual classification of Black and White people is superior to the one reported for the network: 1.3% more accurate in the case of Whites and almost 8% for Black people. Asian estimation is, according to the overall obtained results, the weak-spot of our proposal. Once again the geometric mean of appearance and geometric features achieved superior results compared to single descriptors and their concatenation.

Table 2. Race classification accuracy on FERET subset from the EGA database

Features	Classifier	Accuracy			
		General	Black	Asian	White
filterbank (FB)	RF	85.6%	87.2%	20.9%	97.3%
geometrics (Geom)	SVM	92.9%	92.5%	74.3%	96.9%
FB + Geom	SVM	87.8%	91.3%	43.8%	94.5%
gMean FB + Geom	RF + SVM	93.7%	**93.9%**	68.6%	**98.5%**
Anwar and Islam CNN [2]	-	**94%**	86%	**94.3%**	97.8%

5   Conclusions

In this work we tackle the race estimation problem by means of a component-based approach and geometric descriptors. We exploit the information provided by different face regions, and codify them with both appearance and geometric characteristics to achieved a description of attributes that distinguish race by definition, such skin color and shape of the facial features. We explore two different feature combination strategies and employ traditional classifiers like SVM and RF to obtain a final race prediction. Our late fusion strategy, based on the geometric mean of both appearance and anthropometric probabilistic results, achieved accuracy values very close to those obtained by a recent deep learning proposal (only 0.3% less accurate than the CNN approach), in the FERET subset from EGA database. These results show that there are still some promising alternatives to the use of expensive CNN approaches for the estimation of attributes.

References

1. Afifi, M., Abdelhamed, A.: Afif4: Deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces. arXiv preprint arXiv:1706.04277 (2017)
2. Anwar, I., Islam, N.U.: Learned features are better for ethnicity classification. Cybern. Inf. Technol. **17**(3), 152–164 (2017)
3. Becerra-Riera, F., Méndez-Vázquez, H., Morales-González, A., Tistarelli, M.: Age and gender classification using local appearance descriptors from facial components. In: International Joint Conference on Biometrics (IJCB), pp. 799–804 (2017)
4. Bekhouche, S.E., Ouafi, A., Dornaika, F., Taleb-Ahmed, A., Hadid, A.: Pyramid multi-level features for facial demographic estimation. Expert. Syst. Appl. **80**(C), 297–310 (2017)
5. Carcagnì, P., Coco, M.D., Cazzato, D., Leo, M., Distante, C.: A study on different experimental configurations for age, race, and gender estimation problems. EURASIP J. Image Video Process. **2015**(1), 37 (2015)

6. Cheng, J., Wang, P., Li, G., Hu, Q., Lu, H.: Recent advances in efficient computation of deep convolutional neural networks. Front. Inf. Technol. Electron. Eng. **19**(1), 64–77 (2018)
7. Fu, S., He, H., Hou, Z.G.: Learning race from face: a survey. Trans. Pattern Anal. Mach. Intell. (TPAMI) **36**(12), 2483–2509 (2014)
8. Gill, G., Hughes, S., Bennett, S., Miles Gilbert, B.: Racial identification from the midfacial skeleton with special reference to american indians and whites. J. Forensic Sci. **33**(1), 92–99 (1988)
9. González-Sosa, E., Fiérrez, J., Vera-Rodríguez, R., Alonso-Fernández, F.: Facial soft biometrics for recognition in the wild: recent works, annotation and cots evaluation. IEEE Trans. Inf. Forensics Secur. **13**(7), 2001–2014 (2018)
10. Gould, S., Fulton, R., Koller, D.: Decomposing a scene into geometric and semantically consistent regions. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8, September 2009
11. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Describable visual attributes for face verification and image search. Trans. Pattern Anal. Mach. Intell. (TPAMI) **33**(10), 1962–1977 (2011)
12. Manesh, F.S., Ghahramani, M., Tan, Y.P.: Facial part displacement effect on template-based gender and ethnicity classification. In: 11th International Conference on Control Automation Robotics Vision, pp. 1644–1649. IEEE (2010)
13. Ou, Y., Wu, X., Qian, H., Xu, Y.: A real time race classification system. In: International Conference on Information Acquisition. IEEE (2005)
14. Riccio, D., Dugelay, J.L.: Geometric invariants for 2D/3D face recognition. Pattern Recognit. Lett. **28**(14), 1907–1914 (2007)
15. Riccio, D., Tortora, G., Marsico, M.D., Wechsler, H.: EGA - ethnicity, gender and age, a pre-annotated face database. In: Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS), pp. 1–8. IEEE (2012)
16. Roomi, S.M.M., Virasundarii, S.L., Selvamegala, S., Jeevanandhame, S., Hariharasudhan, D.: Race classification based on facial features. In: 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 54–57. IEEE (2011)
17. Salah, S.H., Du, H., Al-Jawad, N.: Fusing local binary patterns with wavelet features for ethnicity identification. Int. J. Comput. Inf. Syst. Control. Eng. **7**, 330–336 (2013)
18. Shotton, J., Winn, J., Rother, C., Criminisi, A.: *TextonBoost*: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_1
19. Tamrakar, A., et al.: Evaluation of low-level features and their combinations for complex event detection in open source videos. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3681–3688 (2012)
20. Wang, W., He, F., Zhao, Q.: Facial ethnicity classification with deep convolutional neural networks. In: You, Z., et al. (eds.) CCBR 2016. LNCS, vol. 9967, pp. 176–185. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46654-5_20