Dealing with Occlusions in Face Recognition by Region-based Fusion

Ester Gonzalez-Sosa, Ruben Vera-Rodriguez, Julian Fierrez and Javier Ortega-Garcia
Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
{ester.gonzalezs, ruben.vera, julian.fierrez, javier.ortega}@uam.es

***Abstract*—The last research efforts made in the face recognition community have been focusing in improving the robustness of systems under different variability conditions like change of pose, expression, illumination, low resolution and occlusions. Occlusions are also a manner of evading identification, which is commonly used when committing crimes or thefts. In this work we propose an approach based on the fusion of non occluded facial regions that is robust to occlusions in a simple and effective manner. We evaluate the region-based approach in three face recognition systems: Face++ (a commercial software based on CNN) and two advancements over LBP systems, one considering multiple scales and other considering a larger number of facial regions. We report experiments based on the ARFace database and prove the robustness of using only non-occluded facial regions, the effectiveness of a large number of regions and the limitations of the commercial system when dealing with occlusions.**

I. Introduction

Face recognition has been established in the biometric recognition field as one of the least intrusive traits. During the last decade, research efforts have switched from controlled and constrained scenarios to unconstrained and uncontrolled ones. The majority of these new recent efforts have been focused on improving face recognition systems under variability conditions such as illumination [13], pose or expression [16] and low-quality images [18]. In the last few years, the problem of face recognition under occlusions has also started to receive attention from the face recognition community [5], [12].

Occlusions can significantly affect the performance of face recognition systems. There are many different objects that people may wear causing occlusions. The number of reasons for wearing them is also countless. Some people may partially cover their face undeliberately. For instance, there are the veils or burkas for religious or cultural convictions. Others may wear artefacts for safety reasons like mouth masks or head caps for medical staff; mouth masks for people in extreme-pollution cities; helmets and eye masks for extreme adverse work conditions; sunglasses, scarves, caps or hats for adverse weather conditions. People practising any type of sport may wear an accessory that covers part of the face such as swimming caps or eyewears for swimming or helmets for rugby, among others.

There is also the case of criminals, thieves, football hooligans, etc. who tend to wear scarves, sunglasses or even balaclavas on purpose so that they can not be recognized. This variety of potential occlusions is very common to be present in unconstrained and forensic scenarios. As an example of a forensic scenario, the two brothers of the Boston Marathon [7] were wearing sunglasses and caps, making very difficult their identification by both automatic systems and forensic experts. State-of-the-art approaches based on deep learning techniques have reported images showing people with occlusions as common mistakes [17]. Therefore, more research is needed in the search of robust face recognition systems under any type of occlusions.

Previous studies have shown the convenience of using local regions instead of holistic approaches, which tend to be more robust to pose, illumination etc. [2], [14], [15].

The work carried out in [9] explored the problem of face recognition with occlusions under a patch-based approach. They divided the image into 64 blocks and then LGBPHS are computed for each block. The final feature vector is the concatenation of all LGBPHS descriptors of each block. The comparison between two images is obtained by computing the chi-square distance between the non-occluded patches. In [2], a region-based approach is developed. In this case, four regions (eyes, eyebrows, nose and mouth) are considered and each of them is described by a combination of local binary patterns with different radius.

From our point of view, approaches based on patches are not as meaningful as using facial regions, as facial regions are more in compliance with the forensic examiner point of view than facial patches [15]. Typical facial regions configurations in the challenge of face recognition under occlusions usually work with no more than 4 regions [2]. However, there is still facial information that has been left aside. Chin, ear, forehead are examples of regions that are not normally considered in region-based face recognition systems and may be useful to achieve more robust face recognition systems. Based on that, in this work:

- we empirically prove the robustness of region-based approaches under occlusions, showing that the simplicity of fusing non occluded facial regions surpasses state-of-the-art approaches [17]. We confirm also the conclussions achieved in [2], [9].
- we propose a novel and simple approach considering the fusion of 15 facial regions, achieving important relative improvements with respect to the 4-facial region approach. Future extensions of this approach will help to address more uncontrolled types of occlusions.

- we provide an exhaustive analysis of holistic and 4 and 15 facial regions approaches conducting experiments with the Face++ commercial system, a Multiscale LBP system and a LBP system. We also analyse the impact of occlusions in the extraction of facial regions and their individual performance, being them occluded or not.

Three different scenarios are considered: neutral, sunglasses and scarf. The results achieved proves the benefits of discarding the occluded regions and using a large number of facial regions, achieving an average relative improvement of EER of 57.92% and 75.98% for sunglasses and scarf scenarios respectively compared to using only 4 facial regions.

This paper is structured as follows. Section II describes the ARFace database. Section III features the three different face recognition systems considered in this work. Section IV addresses the experimental protocol followed in our experiments and Section V presents the major results obtained in this paper. Finally, Section VI offers some brief conclusions and future work.

II. Database

There are only two reference databases that explicitly deal with occlusions: ARFace database [8] for the 2D domain, and UMB-DB [4] for the 3D domain. ARFace database is the one considered in this work, as it is the most popular benchmark database in the literature that deals with real occlusions (there are other databases that generate artificial occlusions).

ARFace database contains images of 136 subjects (76 men and 60 women). It is comprised of 26 images per subject divided into two sessions of 13 images each. Images present variations regarding expressions (neutral, smile and anger), illumination, and occlusions (sunglasses and scarf), and are acquired under controlled conditions. Sessions are separated by two weeks. There are some subjects with some single missing images and other subjects that only have one of the two sessions. The database is comprised hence of more than 3300 images. Images are 768×576 pixels.

III. System Description

As already mentioned, three different systems are considered: the Face++ (commercial system[1] based on deep neural networks), a Multiscale LBP and a LBP approach.

A. Preprocessing

The preprocessing stage is different according to the specific face recognition system employed. In the case of Face++ system images are divided into the different patches that feed the neural network. For the LBP-based systems, first landmarks are extracted using the Face++ automatic landmark extraction module, and then, the image is gray scaled and aligned according to the position of the eyes.

[1]For more information please visit http://www.faceplusplus.com/api-overview/. We use the Official Matlab SDK For Face++ v2.

B. Face++ commercial system

Face++ is a commercial face recognition system, which has achieved striking performance rates in the LFW competition (achieving the second best rate in the *unrestricted with labelled outside data* protocol with 0.9950 ± 0.0036 of mean accuracy). Face++ is based on a structure of deep network called Pyramid CNN [11]. This approach adopts a greedy-filter-and-down-sample operation enabling the training procedure to be very fast and computation efficient. The structure of the Pyramid CNN can naturally incorporate feature sharing across multi-scale face representations, increasing the discriminative ability of the resulting representation.

In our case, the Face++ API is used in order to obtain verification results. It is worth noting that there is not a public description of the particular implementation used by this API. First, the faces to be compared are individually detected by using the detection module of the API. Once faces are detected a similarity score is obtained by calling the comparison stage. The API Face++ gives similarity scores for four facial regions: eyebrows, eyes, nose and mouth.

C. LBP approaches

Apart from the commercial Face++ system, in this work two LBP approaches are proposed: a Multiscale LBP system and a LBP system using 4 and 15 facial regions respectively. A facial region is extracted by defining a region around specific landmarks. Landmarks are provided by the Face++ API. Three different configurations are available: 5, 25 and 83 landmarks. After several experiments, the set of 25 landmarks is the one chosen for extracting the different regions (see Figure 1).

1) Multiscale LBP system for 4 regions: Our implementation of Multiscale LBP system is inspired by the system proposed in [3], which describes a face through LBP features [1] computed at regions centred in landmarks at different scales. However, as we follow an approach based on facial regions, the approach carried out in this paper computes the LBP features of each facial region at different scales rather than LBP features of regions centred in landmarks. One of the reason that motivate us this change is that with the original approach [3], the different scale versions would include occluded regions. With our approach we are able to isolate the facial regions affected by occlusions. LBP features are extracted using available code by [1].

First the four different face regions are extracted from the original image: eyebrows, eyes, nose and mouth (see Figure 1 right). The use of four regions is done so as to make fair comparisons with the commercial software that also use these four facial regions. Each region is extracted by defining a region around a central landmark. To compute the Multiscale LBP feature vector for a specific region and scale, first the facial region is divided into a grid of 10×10 cells, and then the LBP histogram is computed for each cell. This procedure is done for five different scales: 2, 1, 0.5, 0.25 and 0.125. The output vector for each region is achieved by concatenating the 59-vector LBP histogram of all cells at different scales.

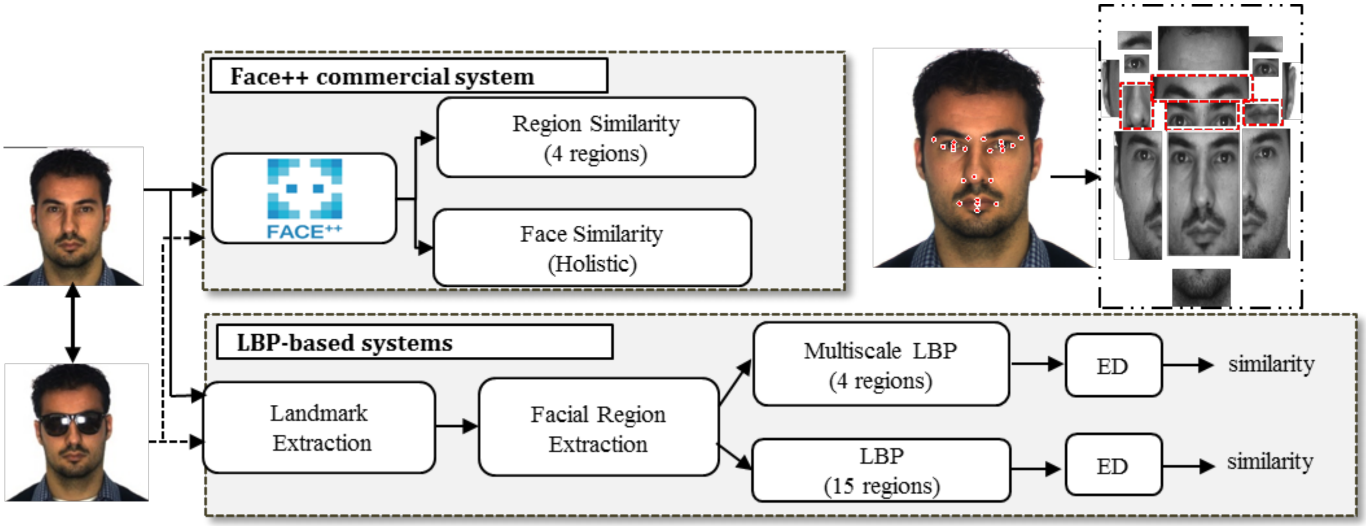

Fig. 1. The figure presents the three systems considered in this work for addressing occlusions in face recognition systems. The upper part of the diagram shows the use of the Face++ commercial system based on CNN, giving a holistic similarity between two faces, and also similarities regarding 4 different facial regions. The lower part presents the local approaches considered: a multiscale LBP system using 4 facial regions and LBP system increasing the number of facial regions up to 15, which is the main contribution of this work. ED stands for Euclidean Distance.

In order to reduce the dimensionality of the features, a PCA projection matrix is estimated for each region using the neutral, smile and anger face images from the development set. In all cases, the LBP feature vector is reduced to a dimensionality of 400 projected components. For the test phase, the Multiscale LBP features associated to each facial region are first computed and then projected into the PCA subspace. Similarity measures are obtained using the Euclidean distance. The fusion between facial regions is done at the score level following the sum rule.

2) LBP system for 15 facial regions: Previous works have shown good results using a higher number of facial regions [15], [14]. As we do not have a restriction of only using 4 regions as is the case with Face++ system, a LBP system based on 15 facial regions is assessed. The regions from the chin, left and right ear, left and right eye, left and right eyebrow, forehead, left middle face, right middle face and the whole face region are also considered apart from the four regions presented in Subsection III-C1. In this case, the LBP features are computed for every region at a single scale $s = 1$ (see Figure 1 right). Similarity measures are obtained using the Euclidean distance.

D. Holistic approaches

With the aim of making fair comparisons with the regions-based systems, a holistic approach is built for each of the three systems considered. For the Face++ commercial system, the API Face++ gives a similarity score for the whole face. With the LBP based approaches, an equivalent scheme is followed but considering the whole face as a unique facial region. In the LBP holistic approach, the image is divided into blocks of 10 by 10 and then the LBP is computed for each block, yielding a feature vector composed of the concatenation of the LBP of all blocks. The same applies to the Multiscale LBP holistic approach but with 5 different scale versions of the face.

TABLE I
DEVELOPMENT AND EVALUATION SETS.

Set	Men Identity	Women Identity
Development	1-49	1-39
Evaluation	50-76	40-60

IV. EXPERIMENTAL PROTOCOL

The whole database is divided into two different sets: development and evaluation according to Table I. First, to estimate the PCA matrix projection of the Multiscale LBP system and LBP system, we use the neutral, smile and anger images with homogeneous illumination of both sessions from subjects of the development set.

Results are reported for three different scenarios (neutral, sunglasses and scarf) in terms of EER. In all cases, images with homogeneous illumination from subjects of the evaluation test are used. For the neutral scenario, neutral images from the first session are compared to neutral images from the second session; for the sunglasses scenario neutral images from both sessions are compared to sunglasses images from both sessions and lastly, for the scarf scenario, neutral images from both sessions are tested against scarf images from both sessions. To the best of our knowledge, there is no standard experimental protocol for this database. As the aim of this work is to study solely the influence of occlusions over the face recognition systems, we decided to discard all the illumination images present in the database. It is worth to mention that the number of trials involved in the neutral scenario is half with respect to the trials in the sunglasses and scarf scenario.

TABLE II
PERFORMANCE OF INDIVIDUAL REGIONS AND FUSION SCHEMES WITH FACE++ COMMERCIAL SOFTWARE AND THE MULTISCALE LBP IN TERMS OF EER%. (*INDICATES THE NON-OCCLUDED REGIONS, BEST OF EACH PAIR OF (SYSTEM,SCENARIO) IS BOLDED)

		Individual regions				Fusion	
System	Scenario	Eyebrows	Eyes	Nose	Mouth	Four Facial Regions	Non-Occluded Facial Regions
Face++	Neutral	7.69	7.69	2.56	7.69	2.56	**2.56**
Face++	Sunglasses	50.71	46.42	33.57*	**12.14***	20	17.14
Face++	Scarf	15.97*	17.85*	**8.12***	54.46	11.6	11.56
MultiscaleLBP	Neutral	15.38	7.69	12.82	10.25	4.58	**4.58**
MultiscaleLBP	Sunglasses	41.46	45.71	22.14*	14.28*	12.85	**11.80**
MultiscaleLBP	Scarf	16.07*	12.5*	27.67*	49.95	10.71	**10.71**

V. EXPERIMENTS

Experiments are carried out on the ARFace database following the aforementioned experimental protocol for the three systems: Face++, Multiscale LBP and LBP systems. For each system and scenario, the performance of the individual facial regions is reported. Performance of the different fusion schemes at score level is reported as well: i) fusion of the whole set of facial regions; ii) fusion of non-occluded regions. For each specific scenario a subset of non-occluded regions is defined. In this work, the subset of non-occluded regions has been defined manually for the two scenarios. As in [2], the subset of non-occluded facial regions is comprised of: nose and mouth region for the sunglasses scenario (2 regions) and eyes, eyebrow and nose for the scarf scenario (3 regions).

Results from Face++ system and Multiscale LBP over 4 regions are shown in Table II. It can be seen that: i) the fusion of the whole set of facial regions is not always better than the performance of the best single region, ii) the best results are achieved when using either the best single region or the fusion of the non-occluded facial regions. When analysing the influence of occlusions over the performance of the single regions, it can be seen that the occluded regions worsen severely their performance while the non-occluded regions worsen slightly their performance comparing with the neutral scenario. The worsening observed in the non-occluded regions is due to several reasons. Firstly, the performance of the nose region is deteriorated in both scenarios and systems as this region is partially affected by the presence of sunglasses or scarves. For instance, the error of the nose region increases from 2.56% of EER to 8.12% of EER for the Face++ system and from 12.82% of EER to 27.67% of EER in the Multiscale LBP system for the neutral and scarf scenarios respectively. The same applies to the nose region in the sunglasses scenario. Also, the performance of the non-occluded regions is also affected by the loss of accuracy of the automatic landmark extraction algorithms when dealing with images with occlusions.

Comparing the performance of the two systems, we see the commercial software works best under the neutral scenario but the Multiscale LBP achieves better results in the presence of any of the occlusions considered. Concretely, the relative improvement of the Multiscale LBP approach with respect to the commercial system is 31% and 7.35% EER for sunglasses and scarf respectively.

The Multiscale LBP was defined using 4 regions in order to make a fair comparison with the Face++ system. However, there is still non-occluded information from the face that may be exploited in order to improve the robustness of the system. Therefore we have developed a LBP based system for 15 facial regions. In Figure 2 the performance of the 15 different regions is plotted for the three different scenarios. It is very noticeable the regions whose performance is severely affected in the presence of occlusions. In the case of the sunglasses scenario those regions are right eyebrow, eyebrows, left eyebrow, right eye, eyes, and the left eye. In the case of the scarf scenario, they are only the mouth and chin regions. Other regions such as the forehead, the nose and right ear are somewhat affected by the presence of sunglasses or scarf. Likewise the Multiscale LBP system over 4 regions, a degradation of performance is also presented in the non-occluded regions. As stated before, this is mainly due to the loss of accuracy of the positioning of landmarks with occluded images.

Among the non occluded facial regions, the best discriminative ones are the whole face and forehead in the neutral scenario (5.12% of EER and 7.69% of EER); the mouth and chin in the sunglasses scenario (13.23% of EER and 14.24% of EER) and the forehead (7.14% of EER), right eye and left eye in the scarf scenario (both with 5.12% of EER). The worst discriminative ones are left ear and right ear for both neutral (23.07% of EER and 20.05% of EER) and sunglasses scenario (27.55% of EER and 36.02% of EER) and the nose and left ear (26.78% of EER and 25.89% of EER) for the scarf scenario. Ear regions are not easy to acquire robustly due to its high dependence on minor variations on the subject pose and the subject appearance (specially with women's hair style).

Table III presents the results obtained with the three different systems as well as the three different scenarios involved, for both local and holistic approaches. From last column of Table III, we conclude that the holistic approach is only the best solution when using the commercial software Face++ compared to the non-occluded region fusion for this system. However, with the LBP approaches, the region-based local approaches outperform greatly either their associated holistic approach, achieving even better results than the holistic approach of the Face++ system.

Also from Table III, it can be observed the striking improvement achieved when increasing the number of facial regions from 4 to 15 regions in both scenarios. From Table III, we

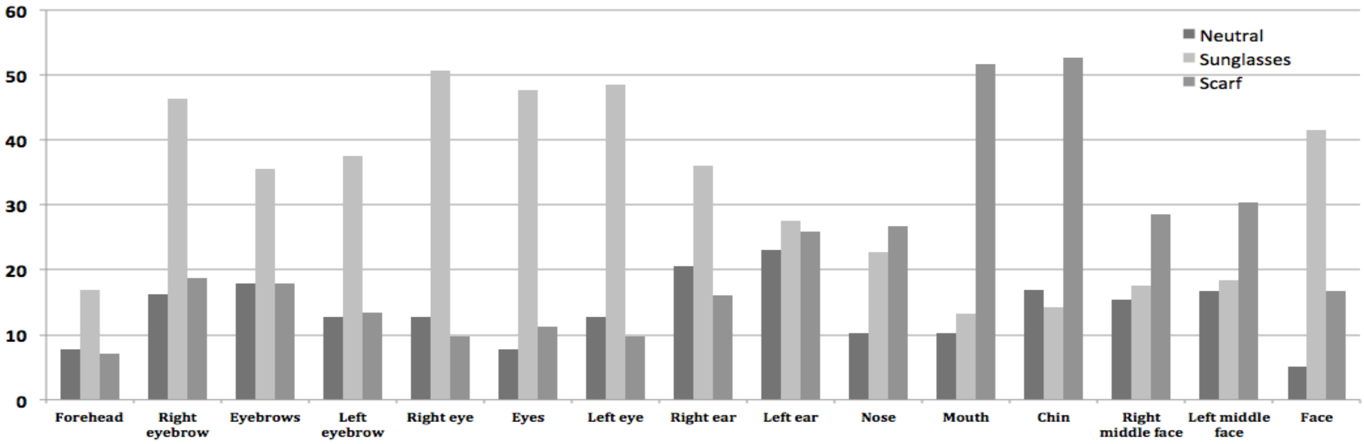

Fig. 2. Performance of the 15 different facial regions for the LBP system in terms of EER (%) for the three different scenarios: neutral, sunglasses and scarf.

TABLE III
COMPARISON BETWEEN LOCAL AND HOLISTIC APPROACHES FOR THE FACE++ , MULTISCALE LBP AND LBP SYSTEMS IN TERMS OF EER%.

System	Scenario	Fusion of Whole Set of Facial Regions	# Regions	Fusion Non-Occluded Facial Regions	# Regions	Holistic
Face++	Neutral	2.56	4	2.56	4	1.29
Face++	Sunglasses	20	4	17.14	2	17.85
Face++	Scarf	11.6	4	11.56	3	7.14
MultiscaleLBP	Neutral	4.58	4	4.58	4	5.10
MultiscaleLBP	Sunglasses	12.85	4	11.80	2	41.31
MultiscaleLBP	Scarf	10.71	4	10.71	3	15.67
LBP	Neutral	1.07	15	**1.07**	15	5.22
LBP	Sunglasses	8.82	15	**5.88**	6	42.32
LBP	Scarf	3.57	15	**2.67**	10	16.96

may see that when fusing the whole set of 15 facial regions for the sunglasses and scarf scenario, EERs of 8.82% and 3.57% are achieved respectively. The average relative improvement with respect to the systems using 4 facial regions (Face++ and Multiscale LBP) is of 43.64% for the sunglasses scenario and 67.94% for the scarf scenario.

A subset of non-occluded regions is also defined for the LBP system. Concretely the non-occluded regions: forehead, right ear, left ear, nose, mouth, chin are considered for the sunglasses scenario while the non-occluded regions: forehead, right eyebrow, eyebrows, left eyebrow, right eye, eyes, left eye, right ear, left ear, nose are considered for the scarf scenario. It is also deduced here that performance improves when considering the fusion of the non-occluded facial regions. The EER is reduced from 8.82% to 5.88% for the sunglasses scenario (using 6 of the 15 facial regions) and from 3.57% to 2.67% for the scarf scenario (using 10 of the 15 facial regions) , yielding further relative improvements of 33.33% and 25.21% respectively. The average relative improvement of the LBP with 15 facial regions systems with respect to the best solution of Face++ and Multiscale system is of 57.92% and 75.98% for the sunglasses and scarf scenarios respectively. Also, not all types of occlusions affect equally to the performance of the system. The performance of the system is decreased according to the specific region occluded. As the eye region is more discriminative than other facial regions, it is logical to find worse performance with sunglasses than with scarves.

VI. CONCLUSION

In this work we give more insights into the problem of face recognition with occlusions under a region-based approach. We evaluate face recognition systems with 4 facial-regions, achieving better results than a state-of-the-art commercial system. Then we propose to increase the number of facial regions considered up to 15 turning out in importance relative improvements.

Experiments have been conducted with three different systems: Face++ commercial system, a Multiscale system and a LBP system under three different scenarios: neutral, sunglasses and scarf scenario. The efectiveness of using a selective and local approach when dealing with occlusions has been empirically proved. Our proposed approach based on 15 facial regions with the fusion of the non-occluded ones outperforms greatly any of the 4 facial regions in the three scenarios considered.

Even if results are presented for two constrained types of occlusions (sunglasses and scarf), this fusion of local regions may be useful to address other occlusions in more real-world conditions. Some more realistic databases such as Remote Face [10] or LFW [6] will be considered for future work. The extraction of facial regions in uncontrolled conditions will be one of the main challenges to overcome.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. on PAMI*, 28(12):2037–2041, 2006.

[2] K. Bonnen, B. F. Klare, and A. K. Jain. Component-based representation in automated face recognition. *IEEE Trans, on IFS*, 8(1):239–253, 2013.

[3] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *IEEE Proc. of CVPR*, pages 3025–3032, 2013.

[4] A. Colombo, C. Cusano, and R. Schettini. Umb-db: A database of partially occluded 3d faces. In *IEEE Proc. of ICCV Workshops*, pages 2113–2119, 2011.

[5] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3d face recognition under expressions, occlusions, and pose variations. *IEEE Trans. on PAMI*, 35(9):2270–2283, 2013.

[6] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Proc. of ECCV Workshop*, 2008.

[7] J. C. Klontz and A. K. Jain. A case study on unconstrained facial recognition using the boston marathon bombings suspects. *Tech. Rep Michigan State University*, 119:120, 2013.

[8] A. M. Martinez. The ARFace database. *CVC Technical Report*, 24, 1998.

[9] R. Min, A. Hadid, and J.-L. Dugelay. Efficient detection of occlusion prior to robust face recognition. *The Scientific World Journal*, id 519158, 2014.

[10] J. Ni and R. Chellappa. Evaluation of state-of-the-art algorithms for remote face recognition. In *17th IEEE Proc of ICIP*, pages 1581–1584, 2010.

[11] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In *IEEE Proc. CVPR*, pages 1891–1898, 2014.

[12] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu. Face recognition under occlusions and variant expressions with partial similarity. *IEEE Trans. on IFS*, 4(2):217–230, 2009.

[13] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. on IP*, 19(6):1635–1650, 2010.

[14] P. Tome, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia. Combination of face regions in forensic scenarios. *Journal of Forensic Sciences*, May 2015.

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

[16] X. Zhang and Y. Gao. Face recognition across pose: A review. *Pattern Recognition*, 42(11):2876–2896, 2009.

[17] E. Zhou, Z. Cao, and Q. Yin. Naive-deep face recognition: Touching the limit of lfw benchmark or not? *arXiv preprint arXiv:1501.04690*, 2015.

[18] W. W. Zou, P. C. Yuen, and R. Chellappa. Low-resolution face tracker robust to illumination variations. *IEEE Trans. on IP*, 22(5):1726–1739, 2013.

ACKNOWLEDGMENT

This work has been partially supported by project Cogni-Metrics TEC2015-70627-R (MINECO/FEDER). E. Gonzalez-Sosa is supported by a PhD scholarship from Universidad Autonoma de Madrid.