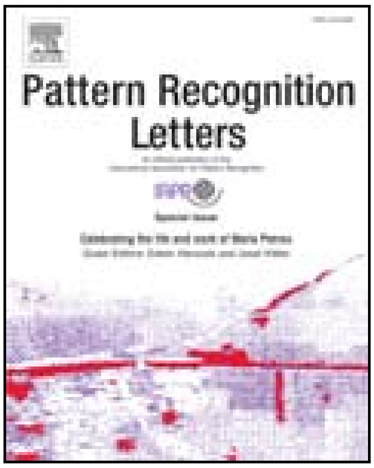
Accepted Manuscript



Facial Marks for Improving Face Recognition

Fabiola Becerra-Riera, Annette Morales-González,
Heydi Méndez-Vázquez

PII: S0167-8655(17)30142-3
DOI: 10.1016/j.patrec.2017.05.005
Reference: PATREC 6811

To appear in: *Pattern Recognition Letters*

Received date: 2 December 2016
Revised date: 28 March 2017
Accepted date: 1 May 2017

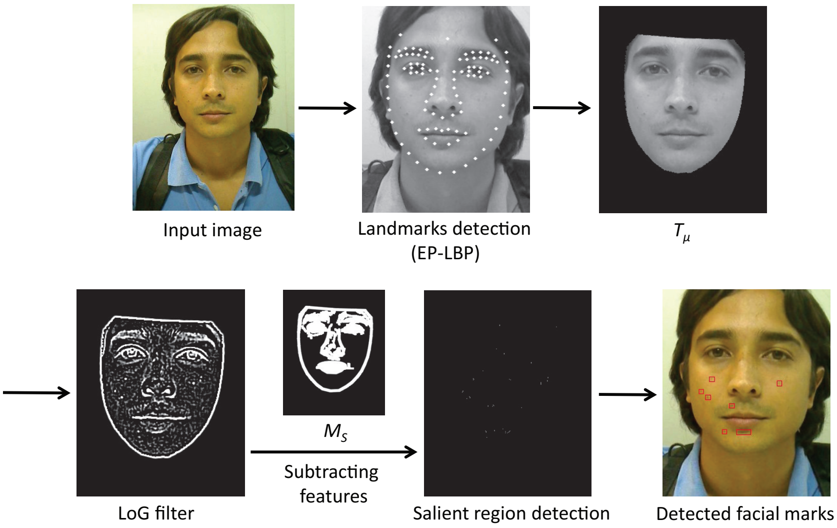
Please cite this article as: Fabiola Becerra-Riera, Annette Morales-González, Heydi Méndez-Vázquez, Facial Marks for Improving Face Recognition, *Pattern Recognition Letters* (2017), doi: 10.1016/j.patrec.2017.05.005

Research Highlights (Required)

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- A new algorithm for the automatic detection of facial marks is proposed.
- A new face dataset with manual annotations for facial mark detection is released.
- A very compact representation of facial marks based on HoG is presented.
- A facial mark similarity measure for comparing two face images is presented.
- The proposal combined with existing face recognition methods improves the accuracy.

Graphical Abstract



Facial mark detection process.

Pattern Recognition Letters
journal homepage: www.elsevier.com

Facial Marks for Improving Face Recognition

Fabiola Becerra-Riera[a], Annette Morales-González[a], Heydi Méndez-Vázquez[a]

[a]Advanced Technologies Application Center (CENATAV). 7a No. 21812, Siboney, Playa, CP 12200, Havana, Cuba

ABSTRACT

Recent studies have shown that the use of soft biometrics (e.g. gender, ethnicity, facial marks) as supplementary information in face images, can increase the accuracy of the recognition process of individuals. Facial marks (e.g. moles, freckles, warts) have shown to be useful, particularly, in this regard. In this paper we propose a new method for combining existing face recognition systems with the information obtained from facial marks in order to improve their performance. We first introduce an algorithm for automatically detecting facial marks, which are then represented using Histograms of Oriented Gradients (HoG), and are matched taking into account their position in the face image. Extensive experiments are conducted in order to show the effectiveness of the proposed facial mark detection algorithm, and to corroborate the benefits of using the information of facial marks on top of traditional face recognition systems. Due to the lack of proper public benchmarks to validate facial mark detection, we also present and make available a dataset with manual annotations for this purpose.

1. Introduction

One of the more popular and used biometric technique is face recognition, which is an active research area that covers various disciplines such as image processing, pattern recognition and computer vision.

Several types of features extracted from images can be employed to perform facial recognition, including soft biometrics such as gender, ethnicity and facial marks (Arigbabu et al. (2015); Jain and Kumar (2012)). According to recent studies, soft biometric traits, despite not being fully discriminative by themselves, may be properly combined with classical facial recognition techniques to increase the accuracy in the verification or identification process. In particular, facial marks have proven extremely useful: despite not uniquely identifying an individual, they can be used to narrow down the search of his identity (Park and Jain (2010)).

Facial marks (e.g. moles, freckles, pimples, scars) can be defined as skin areas (sometimes prominent) that differs in texture, shape and color with respect to the rest of the surrounding skin. Despite some of these marks are not permanent and may disappear even in the short period of a few days, they continue to be widely used in face recognition techniques with the following three main objectives: (1) to complement the result obtained by a conventional face matcher in order to improve the identification accuracy, (2) to reduce the search and enable fast face image retrieval in case of large databases, and (3) to recover partial or off-frontal face images.

Several approaches for automatically detecting and matching facial marks are proposed in the literature. Gogoi et al. (2015) and Pierrard and Vetter (2007) present techniques for the detection of moles prominent enough to be used (by themselves) in the process of identification. Despite their methods can reduce the adverse effect of illumination in face recognition they did not consider other types of facial marks, besides moles. In a similar way, Batool and Chellappa (2014) focus only on the detection of wrinkles/imperfections on their professional software for facial retouching. Lee et al. (2008) introduce scars, marks and tattoos in their tattoo image retrieval system. While tattoos can exist on any part of the body and are more descriptive than other marks, they are not the most representative facial marks; we focus our work on marks appearing exclusively on the face which typically show simple morphologies. Choudhury and Mehata (2012) fully orient their work to the detection of marks covered by cosmetics, while other authors focus on the automatic classification of acne scars (Chantharaphaichi et al. (2015); Ramli et al. (2011)). Unlike most of these methods, that are based on the identification of only one particular type of facial mark, the work of Park and Jain (2010) presents for consideration a method which differs significantly from other studies. It detects all types of facial marks that appear as locally prominent regions, and focuses on the detection of semantically significant facial marks. They combined the facial marks with demographic information and used them for improving face image matching and retrieval performance. This work was later extended by Choudhury and Mehata (2013), in order to process face images which contain salt and pepper noise. This work is also the basis of the proposal of Heflin et al. (2012), in which not only the facial marks but also scars and tattoos are detected and classified in images captured in non-controlled scenarios.

Most of the above methods focus on facial mark detection and classification. A few works study the problem of face recognition considering the marks information. Zhi et al. (2009) combine the traditional Eigenfaces method with a skin mark matcher that showed promising results. Srinivas et al. (2011) represent face images based on the geometric distribution of the annotated facial marks, and establish the similarity between two face images based on a weighted bipartite graph. In the works of Park and Jain (2010); Park et al. (2011) each mark is represented by its (detected) morphology and color, and encoded together with the demographic information into a 50-bin histogram; histogram intersection is then used to calculate the matching scores.

One of the main problems of this field is the lack of a proper public benchmark to validate the detection (and subsequently matching) of facial marks in face images. All of the works presented above either use private datasets with their own annotations, or public datasets, but with no public ground truth for facial mark localization.

Recently we introduced a new algorithm for automatic detection of facial marks (Becerra and Morales-Gonzlez (2016)). The present paper is an extension of that work, which is partly based on the work of Park and Jain (2010). The main novelties of our proposal are: (1) we introduce some important changes in the facial mark detection pipeline that improve the general process; (2) we just describe the texture of the marks regions with a compact representation (without classifying them by their color or morphology); (3) our proposal can be combined with different face recognition systems, showing to improve their accuracy. In order to extend the results presented in Becerra and Morales-Gonzlez (2016), we introduced additional contributions: (4) we extend the scope of our proposal to face identification (not only verification as before), conducting additional experiments on a more difficult dataset; (5) in our proposed combination with traditional face recognition methods, we include, besides the Local Binary Patterns (LBP) descriptor(Ahonen et al. (2004)), the Fisher Vector (FV) encoding (Simonyan et al. (2013), which belong to the top state-of-the-art results for face recognition; and (6) we make public a face dataset with manual facial mark annotations (location and classification) in order to promote the comparison of facial mark detection with state-of-the-art methods.

Our validation on face images containing manually annotated and automatically detected facial marks shows the effectiveness of our facial mark detector and that including facial marks for face recognition reduces the final error in both process: verification and identification.

Particularly for face identification, our method shows its effectiveness in refining the candidate lists obtained using a classical face matcher, which is one of the major applications of the facial marks in forensic scenarios.

The remaining of this paper is structured as follows. First, we present the algorithm for the detection of facial marks, followed by the description of our facial mark matching process between two face images. Next, we describe the experimental process and the results obtained. Finally we present the conclusions and recommendations of our investigation.

2. Facial Mark Detection Algorithm

Facial marks are usually manifested as locally prominent regions, thus a second-order derivative edge detector can be used for their detection. However, the direct application of this type of detector on a face image can generate a considerable number of false positives due mainly to the presence of primary facial features (e.g. eyes, nose, mouth). The location of such features for their subsequent extraction of the facial area, is a necessary step for the successful detection of facial marks. We developed a facial mark detection algorithm based on the work of Park and Jain (2010), but with substantial differences since we improved and replaced some of the methods proposed by the authors in their detection pipeline. This pipeline is depicted in Fig. 1 and it is further explained in the following subsections.

2.1. *Primary Facial Feature Detection and Mapping to Mean Shape*

The first step for the automatic detection of facial marks, as proposed by Park and Jain (2010), is to locate the primary facial features, such as eyes, eyebrows, nose and mouth. To accomplish this, in our approach we apply the EP-LBP model (Méndez et al. (2013)) to each image, replacing the original Active Appearance Model (AAM) used by Park and Jain (2010), with the aim of detecting 112 points outlining the contour of the face and primary facial features. We decided to use EP-LBP because it is faster and it is more robust to illumination problems due to the use of the LBP descriptor in each point's surrounding area. An example of this point detection in a face image is shown in Fig. 1b. The application of this model ensures the normalization of the images in terms of scale and rotation, which allows the representation of each facial mark in a common face-centered coordinate system. Each normalized image is reduced to a facial template T_μ (Fig. 1c), which preserves only the face region and eliminates hair, body and background from the image. This is done with the purpose of simplifying and focusing the process of detecting and matching facial marks only within the face area. The 112 landmarks are used, in addition, for the construction of masks that will eliminate a large number of potential false positives in the following steps.

2.2. *Mask Construction*

Once we have the facial template T_μ of an image (enclosing only the face), we also need to eliminate other facial features that may interfere with the detection of facial marks. Our main goal in this step is to keep only the skin areas where facial marks

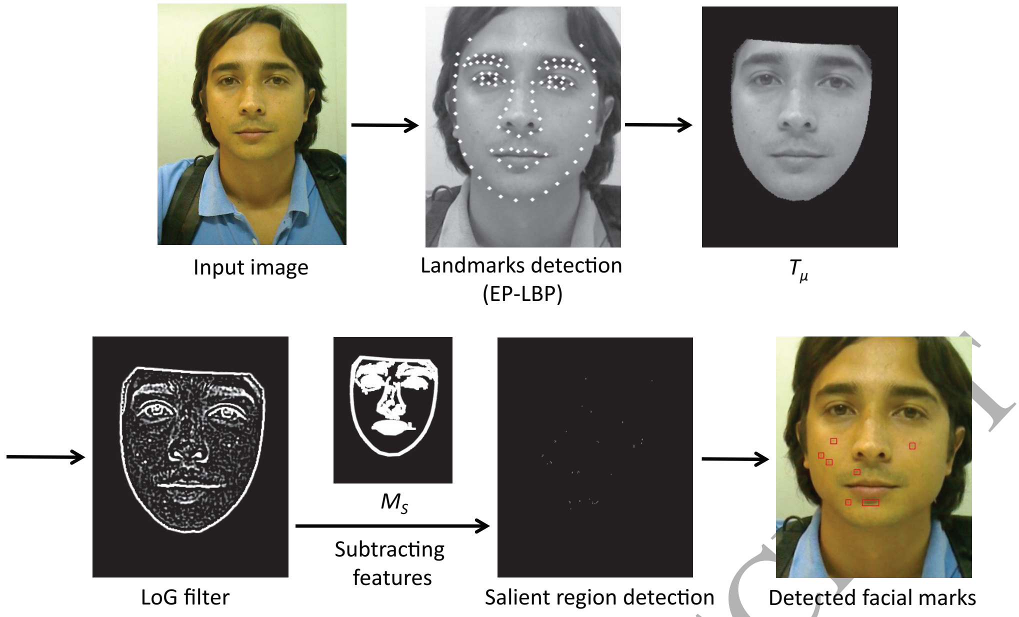


Fig. 1: Steps of the facial mark detection process: (a) Input image, (b) Landmarks detection (EP-LBP), (c) T_μ, (d) LoG filter, (e) M_s, (f) Salient region detection, (g) Detected facial marks.

can appear, and discard all other face areas. To suppress possible false positives generated by primary facial features during the process of detecting candidate facial mark regions, a general mask M_g (Fig. 2 (a)) was built from T_μ, following the procedure described by Park and Jain (2010). M_g is obtained directly from the outlining of the 112 detected facial points and it describes the primary facial features (i.e. eyes, eyebrows, nose and mouth). Since M_g does not cover the particular characteristics of each individual (e.g. beard, wrinkles around eyes and mouth), which are potential false positives, a user-specific mask M_s (Fig. 2 (c)) was constructed as the sum of M_g and the edges that are connected to M_g. For the detection of these peculiarities (Fig. 2 (b)) we employed the Canny edge detector (Canny (1986)) instead of the Sobel operator proposed by Park and Jain (2010). Characterized by its good immunity to noise and the ability to detect true edge points with less error, Canny has shown better results than the Sobel operator. As can be seen in Fig. 2, the use of M_s in (e) largely reduces the false positives that appear near the eyes and mouth in (d), where only M_g was employed.

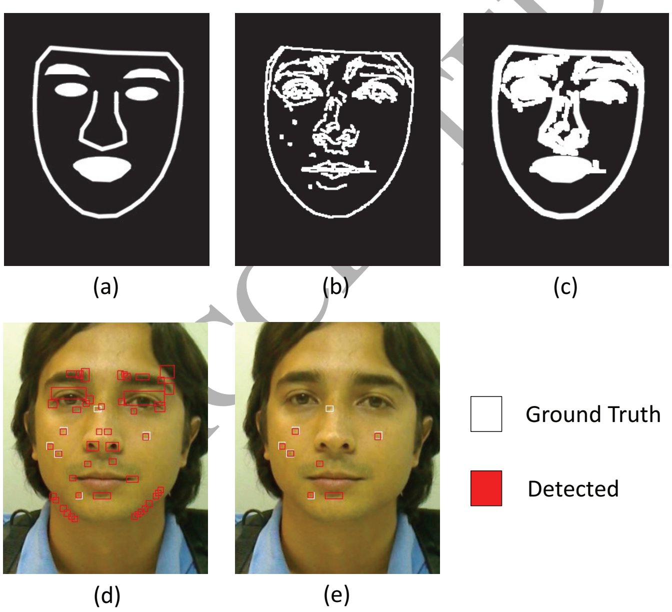


Fig. 2: Effect of M_g and M_s in the detection of facial marks: (a) general mask M_g, (b) edge map obtained by applying the Canny algorithm, (c) user-specific mask M_s, (d) facial marks detected using M_g, (e) facial marks detected using M_s

2.3. Detection of Facial Marks

Facial marks usually appear as isolated and salient regions corresponding to notable changes in intensity. Then, a second-order derivative edge detector was selected for their detection, in this case the Laplacian-of-Gaussian (LoG) filter (Marr and Hildreth (1980)).The LoG filtered image subtracted with the user-specific mask (M_s) undergoes a binarization process with a series of threshold values t_i ($i = 1, ..., K$) in a decreasing order. The threshold t_i is successively applied until the number of resulting connected components is greater than a pre-set value cc. After the detection of a minimum $cc = 10$, the facial mark candidates whose size do not exceed two pixels of width and height are discarded, in order to eliminate pixels or noisy areas that are false positives. Finally, the detected marks are identified by means of a bounding rectangle, indicating its position and size.

3. Facial mark matching process

One of the main objectives proposed in this research was the development of an algorithm to determine the similarity of two face images, based on their facial marks. Consequently, it is necessary to establish a representation of the marks as a first step before starting the matching process.

3.1. Facial mark representation

To obtain a representation of facial marks, each facial mark detected in an image is encoded using a Histogram of Oriented Gradients (HoG) descriptor (Navneet and Bill (2005)) with the goal of obtaining a representation based on the distribution of its intensity gradients. In this case, each histogram is constructed considering a single block of dimension 8 x 8, consisting of a single cell of equal size and 8 bins.

Despite the fact that a representation of the appearance of the marks is necessary, it is not discriminative enough by itself: two marks can be very similar in appearance and, however, they can be located in different regions of the face. It is essential that, for example, a mark located in the forehead is not wrongly considered similar to one located on the chin. The spatial distribution of the marks within the face is an important factor to consider.

Instead of classifying the face regions, for the spatial representation of a mark we employed the x and y central coordinates of its bounding box, normalizing their values in the range $[0, 1]$ through the division between the width and height of the image, respectively. Finally, a facial mark is represented by the vector resulting from concatenating its central coordinates with the appearance features represented by the 8-dimensional HoG. This is a very compact representation that will allow later for a faster comparison between marks.

3.2. Matching facial marks

Given two images I_1 and I_2, and given N_1 and N_2 as the sets of their detected facial marks, respectively, the similarity between I_1 and I_2 (FMM) can be determined by Eq. 1,

$$FMM = \frac{\sum_{i=0}^{|N_1|} \min D(n_i, n_j)}{|N_1|}, \forall n_j \in N_2 | (x_j, y_j) \in R_i \quad (1)$$

where, for each mark $n_j \in N_2$, x_j and y_j are its spatial central coordinates. For each mark $n_i \in N_1$, a rectangular region R_i was built around its central coordinates in I_2, as an area of potential matching, so that only those $n_j \in N_2$, contained in R_i, were considered in the process of matching. This ensures a spatial coherence in the matching of the marks, i.e, very distant facial marks are not verified.

D is a measure of the distance between the marks, in this case computed with the Bhattacharyya distance (Bhattacharyya (1943)), which showed the best results for comparing the obtained histograms during the experimental process. For the score using Bhattacharyya, low values indicate good correlation while high values correspond to little similarity. Being d_{12} the distance between two representative vectors v_1 and v_2, the final score corresponds to $1 - d_{12}$, so higher values respond to greater similarity.

Since our ultimate goal is to improve the accuracy of face recognition, we combine the facial mark matching with traditional face recognition systems. The combination is performed with the output scores of each process. This is possible as long as the similarities scores are in the same range and they share the same behavior (i.e. larger scores represent more similar subjects). The combination is established by a weighted sum of the scores obtained in the following way: $SC = w_{fr} * FR + w_{fmm} * FMM$, where w_{fr} and w_{fmm} ($w_{fr} + w_{fmm} = 1$) represent the weights associated with the score achieved by making use of the face recognition method (FR) and the proposed facial mark matching algorithm (FMM), respectively. In Section 4.3 we will give considerations regarding the selection of these weights.

4. Experiments and results

The process of experimentation was divided into three fundamental parts: the evaluation of the detection of facial marks and experiments of face image verification and identification using facial marks.

We were not able to find in the related literature references to international public databases to perform validation of facial mark detection. Then, in order to validate our results, two subsets of images were created. The first one (DB1), contains 530 images from 265 subjects (two images per person) each of them containing an average of 5 manually annotated facial marks. Moles, freckles, pimples and warts were the most frequently found marks; few images of faces were found with birthmarks, enlightened, darkened areas and pockmarks; and there were no scars or tattoos (See Fig 3). These images belong to an authentication biometric system, and they are taken under controlled conditions (frontal pose, overall good illumination, good image resolution, etc.). We cannot disclose the content of DB1, therefore, we decided to create another dataset, that we can share publicly, in order to contribute to the evaluation of state-of-the-art methods for facial mark detection.

4.1. Celebrities Facial Marks dataset

We present a new dataset, that will be named Celebrities Facial Marks (CFM) [1], that was created with images of celebrities taken from the Internet. There are 30 celebrities with particular moles, scars or tattoos. For each person, there are around 3 and 9 images (5 as average). We have a total of 164 images with variations in pose, aging, expressions and illuminations from which automatic and manually annotated marks were obtained (See Fig. 4). We provide the images along with their facial mark manual annotations, which include their location and classification (moles, scars, tattoos, etc.), and the experimental protocol used for each experiment is detailed, allowing researchers to evaluate their own algorithms in this dataset.

The main difference between DB1 and the CFM datasets is that the former contains images under controlled conditions

[1] This dataset is available at https://www.researchgate.net/project/Visual-Attributes-from-Face-Images.

while the latter is more difficult because the images are taken under uncontrolled situations as described above.

Fig. 3: Examples of two face pairs images from DB1 with manually annotated marks.

Fig. 4: Examples of images from two celebrities in the CFM dataset. First row: Cindy Crawford. Second row: The Game (rapper).

4.2. *Evaluation of the Facial Mark Detection*

Our first experiment was conducted to validate the accuracy of the proposed facial mark detection algorithm (we will call it FM_Canny). This experiment was conducted on DB1 and CFM datasets taking into account the measures of precision, recall and their harmonic mean F-measure.

We extrapolated the standard criterion to evaluate face detection (Jain and Learned-Miller (2010)) to the problem of facial mark detection, in order to establish the similarity between a detected facial mark and an annotated one. Given an image I with A annotated marks, the detection of a facial mark n_i was considered correct if $\exists n_a \in A$ such that:

$$\frac{area(n_i) \cap area(n_a)}{area(n_i) \cup area(n_a)} \geq t_0 \tag{2}$$

The threshold t_0 was established empirically as 0.4.

It was not possible to find available implementations of existing facial mark detectors, therefore, in order to compare our detection results with those of other facial mark detection algorithms on the same image sets, we employed a variant of our detection pipeline by replacing the Canny edge detector by the Sobel operator in step two (we will call it FM_Sobel), which is the one used by Park and Jain (2010). We implemented the algorithm of Choudhury and Mehata (2012) (we will call it FM_SURF), which combines Canny edge detector and SURF features in order to detect facial marks. We also compare our results with the Viola and Jones object detector (Viola and Jones (2001)) as a reference (we will call if FM_V& J). This detector, although initially created for face detection, can be trained to detect a variety of objects in images. In order to detect facial marks, it was trained with a set of 150 positive samples (images of facial marks) and 500 negative samples (images of areas of skin without marks). For all these algorithms we tested several parameter configurations and we show the versions that obtained better results for each case. The results obtained for the DB1 dataset are shown in Table 1 and the results for the CFM dataset are presented in Table 2.

Table 1: Comparison between our detection algorithm FM_Canny with other facial mark detection algorithms in DB1 dataset

Algorithm	Precision (%)	Recall (%)	F-measure (%)
FM_Canny	**73.13**	**57.02**	**64.08**
FM_Sobel	12.50	72.19	21.31
FM_SURF	0.02	0.04	0.03
FM_V& J	26.72	7.58	11.81

Table 2: Comparison between our detection algorithm FM_Canny with other facial mark detection algorithms in the Celebrities Facial Marks dataset.

Algorithm	Precision (%)	Recall (%)	F-measure (%)
FM_Canny	**13.62**	**16.24**	**14.82**
FM_Sobel	2.25	6.22	3.30
FM_SURF	1.06	7.31	1.85
FM_V& J	2.92	7.71	4.24

As can be seen, the proposed detection algorithm (FM_Canny) achieves higher values of precision and recall than the Viola and Jones detector and the SURF detector, which means that it is able to detect more correct marks (superiority in recall) and less false positives (superiority in precision). In the case of the FM_Sobel, for DB1 dataset, the recall is superior to that obtained by our detector, however, the precision is considerably less than ours, so their number of false positives is too high. This result is, in fact, consistent with the reason given in subsection 2.2 for using Canny edge detector instead of Sobel. Sobel detects more edges than Canny (hence the higher recall), but the fraction of them that are relevant is smaller (hence the lower precision). For our approach, the harmonic mean of precision and recall, F-measure, is considerably larger than the others in both datasets. We want to clarify that, for the case of FM_SURF, it is designed to detect bigger facial marks (like tattoos, big scars or burns) and we have few of those in our datasets. That is the reason for its low performance in this context. With these results we can conclude that our variant is in general more consistent in the detection of correct marks. Nevertheless, the overall poor results in the CFM dataset (which is more difficult than DB1 in terms of uncontrolled scenarios and conditions), show that more accurate algorithms for facial mark detection are needed.

4.3. Validation on Face Verification

The most common evaluation metrics for face verification are: False Acceptance Rate (FAR), which represents the number of impostors that are classified as genuines, divided by the total of impostor comparisons; False Recognition Rate (FRR), which is the number of genuines that are classified as impostors, divided by the total of genuine comparisons; and the Equal Error Rate (EER) which is the error when $FAR(i) = FRR(i)$. Another commonly used meatric in biometric authentication systems is the Operating Point (OP) which is set according to the security level of the recognition system, in this case, measuring the amount of genuines wrongly classified when very few impostors are allowed. In our case we use the OP by fixing the FAR to 0.1 %, thus we present the FRR (OP_FRR) for this case.

First of all, we conducted a verification experiment on the CFM dataset only using facial mark matching, in order to compare the facial mark detectors from the previous section in terms of verification accuracy. We chose this dataset because it has a larger variability of marks. The results in terms of EER and OP_FRR are shown in Table 3. In this table we included FM_Manual, which is the matching of the manually annotated facial marks. It can be seen that our detector achieves the best results compared to the other automatic detectors. It is worth noting that the manual annotations obtained much lower error rates than the automatic methods, which is an indication that better results can be obtained if the detection process is improved.

Table 3: Comparison between our detection/matching algorithm FM_Canny with other facial mark detection algorithms in the Celebrities Facial Marks dataset, in terms of verification accuracy.

Algorithm	**EER(%)**	**OP_FRR(%)**
FM_Manual	**28.30**	**99.29**
FM_Canny	**35.65**	**99.37**
FM_Sobel	36.75	99.08
FM_SURF	42.00	100.0
FM_V& J	37.21	99.59

In order to corroborate the benefits of using facial marks to improve face recognition performance (on both face verification and identification) we combined our facial mark matching method with two popular methods: 1) the nearest neighbor classifier with Local Binary Patterns (LBP) operator (Ahonen et al. (2004)), and 2) the Fisher Vector (FV) Faces proposed by Simonyan et al. (2013).

In the verification evaluation(1 vs. 1) we used both DB1 and CFM datasets to conduct the experiments. The facial mark matching with automatic detection (FaceMarks_Auto) is evaluated and compared with the results of LBP and FV methods, as well as their combinations. Also, we decided to perform experiments under the same conditions, but using the manual annotation of facial marks (FaceMarks_Manual). This will give an idea of the benefit of our proposal for verification, if the facial marks were detected perfectly.

Tables 4 and 5 show the results in terms of EER and OP_FRR for DB1 and CFM datasets respectively. As can be seen in the first two rows of both tables, facial marks (as any soft biometric trait) are not discriminative enough to be used in a verification process, however, when they are combined with other classical face recognition techniques, such as LBP (rows 4 and 5) or FV (rows 7 and 8) in this case, the accuracy in the verification process is improved. In table 4, the improvement in the EER is substantial (in the case of FV + FaceMarks_Manual, the EER is reduced by more than half w.r.t. FV alone), which means that genuine pairs that were not correctly matched with the original face descriptors are more similar when the facial mark information is introduced. For the case of the combination with LBP, the error made in the automatic detection of marks is not relevant to the final result (given the small difference between LBP + FaceMarks_Auto and LBP + FaceMarks_Manual), but for the case of the combination with FV, improving the facial mark detection step may have a grater impact in the final result. In Table 5 the improvement of the automatically detected facial marks is less evident, but this is due to the difficulty of the CFM and the poor general detection results. Nevertheless, the improvement when using the manual annotations is important in such uncontrolled conditions. This can be helpful in forensic scenarios where manual annotations can be done.

Table 4: Comparison between LBP, FV and the proposed variants on verification experiments in DB1 dataset.

Algorithm	**EER(%)**	**OP_FRR(%)**
FaceMarks_Auto	29.81	98.86
FaceMarks_Manual	11.60	96.98
LBP	3.00	6.79
LBP + FaceMarks_Auto	1.88	6.41
LBP + FaceMarks_Manual	**1.79**	**6.40**
FV	1.5	3.77
FV + FaceMarks_Auto	1.00	3.70
FV + FaceMarks_Manual	**0.70**	**2.26**

Table 5: Comparison between LBP, FV and the proposed variants on verification experiments in CFM dataset.

Algorithm	**EER(%)**	**OP_FRR(%)**
FaceMarks_Auto	35.65	99.37
FaceMarks_Manual	28.30	99.29
LBP	21.90	100.0
LBP + FaceMarks_Auto	21.78	92.69
LBP + FaceMarks_Manual	**18.13**	**90.99**
FV	8.88	68.14
FV + FaceMarks_Auto	8.74	66.57
FV + FaceMarks_Manual	**7.41**	**48.95**

Another crucial step related to this performance is the selection of the weights w_{fr} and w_{fmm} for the combination. As can be deduced from the results in Table 4, the contribution of the facial marks on their own to the recognition process is very small. The same behavior applies to the combination, where the best results were obtained with higher values of w_{fr} (0.8 and 0.9) and lower values of w_{fmm} (0.1 and 0.2). Although the value of w_{fmm} is small, the verification results of the combination show that they are relevant in our approach.

4.4. Validation on Face Identification

In an initial identification experiment we used the CFM dataset and conducted a five-fold cross validation using each time one image of every person as gallery and the rest of them as probe. We compare the results of the face matching with the manual and automatic marks detection and the combination with FV method, which showed the best results in previous section. Table 6 shows the obtained results in terms of average Recognition Rate at Rank 1 (RR_R1) and Rank 5 (RR_R5).

Table 6: Comparison between FV and the proposed combinations on identification experiments.

Algorithm	RR_R1(%)	RR_R5(%)
FaceMarks_Auto	16.00	33.33
FaceMarks_Manual	34.66	52.66
FV	86.00	98.66
FV + FaceMarks_Auto	87.33	**99.33**
FV + FaceMarks_Manual	**90.00**	**99.33**

In Table 6, as it was discussed before, we can see that the use of only facial marks is not feasible, but its combination with a face recognition method improves the results for identification. In rank 1, using only the FV method an 86% of Recognition Rate (RR) is obtained, which is improved when the facial mark matching is considered, on both manual or automatic variants. Fig. 5 facilitates the visualization of these results until Rank 10. It is worth noting that starting at Rank 3, the results with both automatic and manual marks are similar.

In an additional experiment we evaluated how the method scales up when an open larger set of images is considered. This experiment was performed in the Labeled Faces in the Wild (LFW) dataset Huang et al. (2007) which contains 13,233 target face images. This dataset is used as gallery and we added to this gallery 92 images from our CFM dataset (about half of the images of every subject). The remaining 72 images from CFM dataset are used as query for identification (more details of this protocol are explained in the online documentation that we shared for this dataset). We compare the FV method and the combination with the automatic marks detection in this case. It was not possible to compare with the manual annotations because we do not have this information for the LFW dataset. The cumulative Recognition Rates obtained until Rank 10 are listed on Table 7.

Table 7: Cumulative Recognition Rates (%) for face identification using FV and its combination with FaceMarks_Auto in a larger dataset (LFW + CFM datasets) of more than 13 000 images.

Position	FV	FV + FaceMarks_Auto
1	63.89	63.89
2	72.22	**73.61**
3	79.16	79.16
4	80.55	**83.33**
5	80.55	**87.50**
6	81.94	**87.50**
7	81.94	**87.50**
8	81.94	**88.89**
9	84.72	**88.89**
10	86.11	**88.89**

As can be seen in Tab. 7, for the first position, the combination FV + FaceMarks_Auto does not improve the results of FV alone, but after position 4, the improvement is more clear. This can be seen also in Figure 6 where we show graphically the same results.

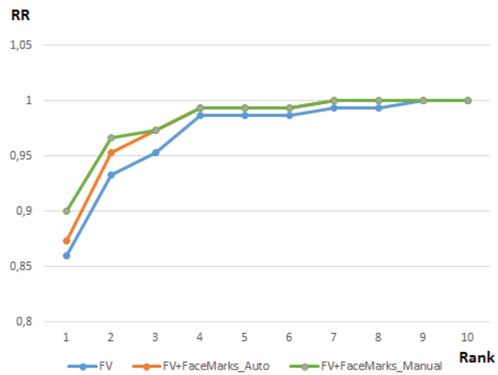


Fig. 5: Cumulative Recognition Rate curve (%) of the FV method and its combination with facial mark matching, using both automatic detection (FaceMarks_Auto) and manual annotation (FaceMarks_Manual)

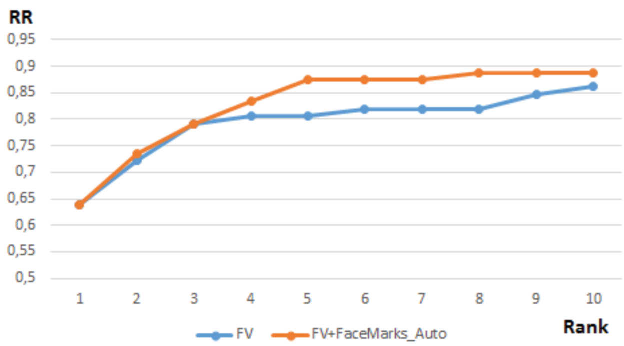


Fig. 6: Cumulative Recognition Rates (%) for face identification using FV and its combination with FaceMarks_Auto in a larger dataset (LFW + CFM datasets) of more than 13 000 images.

It is important to emphasize that, since the images involved in this experiment are under uncontrolled conditions, the extraction of facial marks is affected, as shown in the detection experiments in Table 2. We believe that, if a better detection is accomplished for these images, the improvement of FV + FaceMarks_Auto could be more noticeable. In forensic environments, where facial marks can be manually annotated or edited, these results can have greater impact.

Conclusions

In this work we introduced a facial mark detector that obtained overall good results with respect to similar state-of-the-art methods. We also presented an algorithm for matching the detected facial marks, that can be combined with traditional face recognition methods. By making use of the proposed algorithms for facial mark detection and matching we were able to confirm that facial marks, used on their own for recognition, are not discriminative enough, as it was already stated in the literature. However, by combining the algorithms of facial mark matching with classical techniques for face recognition, we were able to achieve lower error rates in our face verification and identification experiments, which shows the usefulness of this type of trait as complementary information.

The proposed marks representation based on HoG descriptor is very compact and allows fast comparison between marks, which is crucial in large scale face recognition applications. On the other hand the proposal shows very good results without the necessity of classifying the detected facial marks.

In this work we introduced a new public dataset for facial mark detection, providing manually annotated ground truth for each facial mark. We believe that this may encourage further research in this area.

As future work we plan to combine the facial marks with other soft biometrics traits to assist forensic queries based on facial images. Also, we plan to explore other types of combinations of facial marks and traditional recognition systems, such as feature level combination.

References

Ahonen, T., Hadid, A., Pietikãinen, M., 2004. Face recognition with local binary patterns. Lecture Notes in Computer Science : Computer Vision - ECCV 2004 , 469–481.

Arigbabu, O., Ahmad, S., Adnan, W., Yussof, S., 2015. Recent advances in facial soft biometrics. The Visual Computer 31, 513–525.

Batool, N., Chellappa, R., 2014. Detection and inpainting of facial wrinkles using texture orientation fields and markov random field modeling. Image Processing, IEEE Transactions on 23, 3773–3788.

Becerra, F., Morales-Gonzlez, A., 2016. Detection and matching of facial marks in face images. Revista Cubana de Ciencias Informticas 10, 172–181.

Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of Cal. Math. Soc. 35, 99–109.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 8, 679–698.

Chantharaphaichi, T., Uyyanonvara, B., Sinthanayothin, C., Nishihara, A., 2015. Automatic acne detection for medical treatment, in: 6th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), Hua-Hin, pp. 1–6.

Choudhury, Z.H., Mehata, K., 2012. Robust facial marks detection method using AAM and SURF. International Journal of Engineering Research and Applications (IJERA) 2, 708–715.

Choudhury, Z.H., Mehata, K.M., 2013. Biometrics security: Facial marks detection from the low quality images. International Journal of Computer Applications 66, 33–41.

Gogoi, U.R., Bhowmik, M.K., Saha, P., Bhattacharjee, D., De, B.K., 2015. Facial mole detection: An approach towards face identification. Procedia Computer Science 46, 1546 – 1553.

Heflin, B., Scheirer, W., Boult, T.E., 2012. Detecting and classifying scars, marks, and tattoos found in the wild, in: 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 31–38.

Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E., 2007. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49. University of Massachusetts, Amherst.

Jain, A., Kumar, A., 2012. Biometric recognition: An overview, in: Mordini, E., Tzovaras, D. (Eds.), Second Generation Biometrics: The Ethical, Legal and Social Context. Springer Netherlands. volume 11 of *The International Library of Ethics, Law and Technology*, pp. 49–79.

Jain, V., Learned-Miller, E., 2010. FDDB: A benchmark for face detection in unconstrained settings. Tech. Rep. UM-CS-2010-009. Dept. Comp. Science, Univ. Massachusetts, Amherst.

Lee, J.E., Jain, A.K., Jin, R., 2008. Scars, marks and tattoos (SMT): Soft biometric for suspect and victim identification, in: Proceedings of Biometric Symposium, Biometric Consortium Conference, pp. 1–8.

Marr, D., Hildreth, E., 1980. Theory of edge detection, in: Proceedings of the Royal Society of London. Series B, Biological Sciences, pp. 187–217.

Méndez, N., Chang, L., Plasencia, Y., Méndez, H., 2013. Facial landmarks detection using extended profile lbp-based active shape models, in: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 18th Iberoamerican Congress, CIARP 2013, Havana, Cuba.

Navneet, D., Bill, T., 2005. Histograms of oriented gradients for human detection, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society. pp. 886–893.

Park, U., Jain, A.K., 2010. Face matching and retrieval using soft biometrics. IEEE Transactions on Information Forensics and Security 5, 406–415.

Park, U., Liao, S., Klare, B., Voss, J., Jain, A.K., 2011. Face Finder: Filtering a Large Face Database using Scars, Marks and Tattoos. Tech. Rep. MSU-CSE-11-15. Dept. Comp. Science, Michigan State Univ.

Pierrard, J.S., Vetter, T., 2007. Skin detail analysis for face recognition, in: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, pp. 1–8.

Ramli, R., Malik, A., Hani, A., Yap, F., 2011. Segmentation of acne vulgaris lesions, in: Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on, pp. 335–339. doi:10.1109/DICTA.2011.63.

Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A., 2013. Fisher Vector Faces in the Wild, in: British Machine Vision Conference.

Srinivas, N., Aggarwal, G., Flynn, P.J., Bruegge, R.W.V., 2011. Facial marks as biometric signatures to distinguish between identical twins, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Viola, P., Jones, M., 2001. Robust real-time object detection, in: International Journal of Computer Vision.

Zhi, Z., Sergey, T., Venu, G., 2009. Combining facial skin mark and eigenfaces for face recognition, in: ICB, Springer.