3D-aided face recognition robust to expression and pose variations

B.Chu[1,2], S.Romdhani[1] and L.Chen[2]

[1] Morpho, SAFRAN Group
11 Boulevard Galliéni 92130 Issy-Les-Moulineaux - France
{baptiste.chu, sami.romdhani}@morpho.com
[2] Université de Lyon, CNRS
Ecole Centrale de Lyon, LIRIS UMR5205, F-69134 Lyon, France
{baptiste.chu, liming.chen}@ec-lyon.fr

Abstract

Expression and pose variations are major challenges for reliable face recognition (FR) in 2D. In this paper, we aim to endow state of the art face recognition SDKs with robustness to simultaneous facial expression variations and pose changes by using an extended 3D Morphable Model (3DMM) which isolates identity variations from those due to facial expressions. Specifically, given a probe with expression, a novel view of the face is generated where the pose is rectified and the expression neutralized. We present two methods of expression neutralization. The first one uses prior knowledge to infer the neutral expression image from an input image. The second method, specifically designed for verification, is based on the transfer of the gallery face expression to the probe. Experiments using rectified and neutralized view with a standard commercial FR SDK on two 2D face databases, namely Multi-PIE and AR, show significant performance improvement of the commercial SDK to deal with expression and pose variations and demonstrates the effectiveness of the proposed approach.

1. Introduction

The last evaluation of face recognition algorithms performed by the NIST in 2010, MBE [11], showed that high accuracy can be obtained on frontal face images : a verification rate higher than 95% is obtained for a false alarm rate of 0.1%. However, when one of the images to compare is non-frontal, this verification rate drops to 20% and the 2009 NIST MBGC report [16] concludes that : "Cross pose matching is still very difficult."

These evaluations do not report results when the expression of the subjects changes. Independent assessment of face recognition softwares under systematic expression variations is lacking. However, recently, expression robust face recognition has been attracting more and more the attention of the academic world [8] [9] [13] [20] [23] [26]. This is owing to the availability of face database with systematic expression variations such as the AR dataset [14], PIE [19] and Multi-PIE [10].

Interestingly, the PIE dataset, and even more so, Multi-PIE, provide face images with *combined* variations of expression and pose. However, despite the availability of these datasets since ten years now, the problem of robustness of face recognition towards combined variations of pose and expression has been left untouched. The aim of this paper is to present a method able to compensate for expression and pose variations. We present also the first results, to the best of our knowledge, on the non-frontal and non-neutral portion of Multi-PIE dataset.

Another lesson of the latest NIST evaluations, MBE and MBGC, is the fact that the highest accuracy on frontal and neutral face recognition has been obtained by commercial SDKs [11]. Hence, to construct our pose and expression tolerant face recognition system, it would be interesting to leverage these results. Therefore, we address this problem as a pre-processing : how to modify the pose and the expression of a face image, such that it can then be used by a commercial SDK ?

The pose normalisation of an input face image is usually done by synthesizing the face at a reference pose [5]. This reference pose needs to be the same for all face images. A natural choice is the frontal pose, as the face recognition algorithms are tuned to work at this pose. Similarly, they expect expressionless face images, *i.e*. neutral faces. But, what is the definition of a neutral face ? How to modify a face image such as it is neutral ? Naturally these pose and expression "modifications" need to leave the identity part of the person intact. In this paper, we propose two meth-

ods to perform these modifications. The first method, "expression neutralization", can be used in any recognition scenario but assumes that, given enough generic prior knowledge, one example of an input face image is sufficient to infer its expression deformations (*i.e*. given enough generic prior knowledge, the shape of an individual is sufficient to estimate the shape of this person for any expression). The second method, "expression transfer", is only applicable in a verification scenario but its assumption is less restrictive : it assumes that the deformations due to expressions, for an ensemble of individuals, lie on a linear subspace.

The pose normalisation is usually addressed by a warping [6] [12] [2] [3] that utilizes the input image as a texture map. This warping can be done in 2D or 3D. Owing to the fact that a human head is a 3D object, we believe that it is easier to model the variation of its shape (for an ensemble of individuals) and of its pose in 3D, free of planar projection non-linearities (leaving the resolution of these non-linearities when fitting an input image). One of the most successful method for 3D face modelling method is the 3D Morphable Model [6]. As it happens, this model has been extended with expression variations [4]. Here, we revisit this method, and show how it can be applied to face recognition robust to pose and expression changes. After the review of the state of the art, Section 2 briefly describes how the Expression 3D Morphable Model was constructed and how it is used to estimate the 3D shape and the pose of an input face image. The "expression neutralization" and the "expression transfer" methods are then described in Section 3. These methods are then empirically compared to each other and to the state of the art on the Multi-PIE and AR datasets in Section 4.

1.1. Related work

As mentioned earlier, there is little state of the art on combined pose and expression normalisation for face recognition. One of the notable recent approach is the one of Berg *et al*. [3]. Similarly to the method presented here, it aims to warp a face image to a novel view that is frontal and neutral without loosing the identity information. This is done by locating as many as 95 feature points in the input image. Then, for each individual of a training set, the closest set of feature points are searched across pose and expression variations. These feature points positions are then averaged across training individuals, thereby finding the location of the feature points of an "average person" with the same pose and expression as the ones of the input image. A new image is then obtained by warping the positions of these points to the canonical positions of an average individual at a frontal pose and neutral expression. Hence, if the input image shows a person with a fat nose, it will still have a fat nose after warping. As it is a 2D warping method, it copes with moderate pose variations.

Other methods have been designed to deal either with pose or expression variations. As far as pose is concerned, pose rectification is usually treated as an alignment pre-processing : Blanz *et al*. [5] use a 3DMM to generate a frontal view from an input image. This pre-processing was tested in conjunction with off the shelf commercial face recognition systems during the NIST FRVT-2002. The conclusion is that it was effective in compensating for pose : Matching 45° views with frontal views increased the verification rate for 1% false alarm from 17% to 79%. As this method is difficult to implement, due to the fact that it requires a large set of training 3D scans that are registered and its usage on an input image is computationally expensive, researchers proposed simpler 2D methods : Huang *et al*. [12] perform an in-plane alignment using a small 2D training set by minimizing entropy. Asthana *et al*. [2] use a feature point detector to locate a dozen landmark points and estimate 3D pose by fitting an average 3D face. The novel view is then obtained by using the input image as a texture map and the 3D shape of the average face.

On the other hand, methods designed to address the expression variability of FR systems, usually, incorporate the expression robustness directly in the FR algorithm and do not treat it as a pre-processing. For instance, they assume that the expression variations are localized and these regions are treated as outliers and are hence not used in the comparison [20] [23] [26].

2. 3D face morphable model

3D Morphable Model, proposed by Blanz *et al*. [6], is one of the most successful methods to represent the space of human faces. Learned from 3D scans, a 3DMM proposes to approximate any individual face using a linear combination of limited modes. Thanks to the third dimension, the 3DMM is much more accurate with respect to its 2D counterpart and can deal with out-plane pose variations, *e.g*., yaw or pitch rotations. Given an input face image in 2D, the 3DMM fitting algorithm estimates the shape deformation coefficients, as well as those for the pose and the texture. A novel view of the input 2D face image can then be generated where the pose is rectified. A standard FR SDK can then be invoked, thereby making it capable to deal with pose changes.

In this section, we briefly describe an extension of 3D morphable model with expression variations as proposed by [4]. An expression neutralization can thus be performed by modifying the expression coefficients while keeping the identity coefficients.

2.1. Training set

The construction of a 3D morphable model is based on a statistical analysis of 3D face scans in full correspondence. There exists several public 3D face datasets with expression

variations, *e.g*. BU-3DFE [28], Bosphorus [18] or D3DFacs [7].

Given its significant set of expressive scans, the BU-3DFE database has been chosen as training database of our morphable model. In this database, 100 individuals are available. Each of the 100 individuals in the dataset was asked to perform the six universal expressions in addition to the neutral one : Angry, Happy, Fear, Disgust, Sad and Surprise. For the six non-neutral expressions, four stages of intensity were recorded. In this work, we use the highest intensity expression scans. Then, our training dataset contains 700 3D face scans.

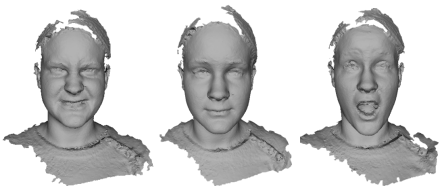

Figure 1. Angry, neutral and surprise scans from the BU-3DFE database

2.2. Modelling identity and expression variations

Before any statistical analysis, a registration step, as proposed by Amberg [1], is performed. Once densely aligned, principal component analysis (PCA) can then be applied to those registered 3D face scans to extract the principal modes of variation.

Each 3D face shape (with n_{vert} vertex) is represented by a $3n_{vert} \times 1$ dimensional vector, $S_{i,e} = (X_1, Y_1, Z_1, X_2, ..., Y_{n_{vert}}, Z_{n_{vert}})$, where e refers to a given expression (0 for neutral and 1-6 for expressions) whereas i in $0...n_{id} - 1$ is the identity index. First, each face is translated by removing the mean shape $\overline{S}$. We separate the subset into two subsets (one with neutral faces and the second for the expressive scans). Two PCA are thus computed (one on each set).

The first one, computed on neutral scans $S_{i,0}$, gives an identity morphable model whose principal axis of variations are in the matrix A_{id}. New neutral faces can be generated by varying the identity coefficient vector $\alpha_{id} : (n_{id} - 1) \times 1$

$$S = \overline{S} + A_{id}\alpha_{id} \tag{1}$$

The second PCA models the deformations due to expressions. A PCA is thus performed on the offsets between expressive scans and neutral scans : $\Delta S_{i,e} = S_{i,e} - S_{i,0}$ for e=1..6 yielding the axis of deformations due to expression in A_{exp}. Face deformations due to expressions can hence be generated by varying the expression coefficients vector $\alpha_{exp} : (6n_{id} - 1) \times 1$

$$\Delta_{exp} = A_{exp}\alpha_{exp} \tag{2}$$

Combining equation (1) and (2), we can generate any face with identity and expression variations :

$$S = \overline{S} + \begin{bmatrix} A_{id} & A_{exp} \end{bmatrix} \begin{bmatrix} \alpha_{id} \\ \alpha_{exp} \end{bmatrix} \tag{3}$$

Some examples of face generated using this morphable model are shown in Figure 2.

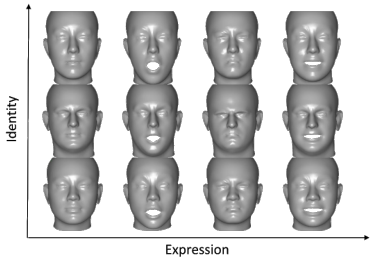

Figure 2. Faces generated with identity and expression variations

2.3. Fitting the morphable model to a 2D face image

This 3D morphable model can be fitted to a 2D image as proposed by Matthews in [15]. The coefficients of the morphable model are computed to minimize the distance between the 2D image and the projection of the 3D face. The fitting process is initialized using the texture information of the input 2D face image, *e.g*., contour, feature points (as defined in the mpeg4 norm [17]) and the silhouette. Then, the Levenberg-Marquart method is used to solve the minimization problem.

Given a 2D image, the fitting computes the different model parameters : Identity parameters ($\alpha_{id} \in \mathbb{R}^{n_{id}-1}$), expression parameters ($\alpha_{exp} \in \mathbb{R}^{n_{exp}-1}$) and pose parameters ($\alpha_{pose} \in \mathbb{R}^6$).

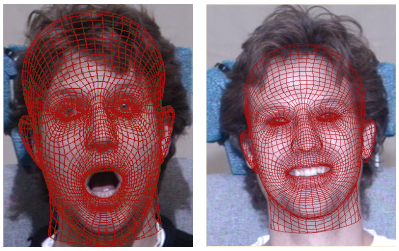

Figure 3. 3DMM fitting on face images with expression

3. Expression neutralization

Expression variations are sources of non-rigid deformations of facial shapes, causing changes in the appearance of faces both in 2D and 3D. These appearance changes are major sources of the FR accuracy drop for standard FR SDKs mostly optimized for frontal and neutral 2D face images.

The method presented in this paper is based on synthesizing of a novel view with a neutral expression while preserving the identity. Specifically, a novel frontalized and neutralized view of an input probe is generated using the 3D morphable model previously presented. Given a 2D face image, this novel view is synthesized using the framework described in Figure 4.

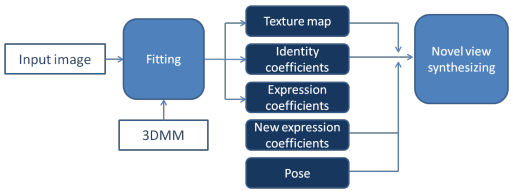

Figure 4. The framework for synthesizing a novel view

To generate a frontal view, Blanz *et al*. [5] extracts the parameters of the 3D morphable model (Texture map, identity coefficients and pose parameters) and perform a rendering with new pose parameters. In this work, we make use of the previously introduced extended 3D morphable model which isolates the identity variations from those due to the expression changes. Thus, the fitting of the 3DMM leads to a set of expression coefficients in addition to the parameters in a standard 3DMM, *i.e*., identity, pose. A rendering can then be performed by changing the expression coefficients and the pose parameters to generate a frontalized and neutralized view of the face.

In this section, we present two methods to choose the expression coefficients for synthesizing neutralized face images. The first one is based on a mean expression extracted from neutral images. Then, a second method is proposed to render an image with an expression closest to the gallery one while keeping a same set of identity coefficients. This method is based on a expression transfer : The expression computed on the gallery image is transferred to the probe image.

3.1. Mean neutral expression

The 3DMM presented in the previous section can generate any individual face with any expression. A standard neutral expression can be used to render all faces with the same neutral expression.

We determine this neutral expression on a training set of N *neutral* images. The 3DMM is fitted to the i^{th} neutral image of the training set. Expression coefficients α^i_{exp} corresponding to this image can thus be extracted.The mean neutral expression coefficients are then computed as the mean of all extracted expression coefficients on the neutral training set.

$$\alpha_{exp\_neutral} = \frac{1}{N}\sum_{i=1}^{N}\alpha^i_{exp}$$

Given these coefficients, a novel view of the probe image can be generated with the mean neutral expression.

Specifically, the 3DMM is fitted to an input 2D face image to compute the identity coefficients, the expression coefficients and the pose parameters. A novel mean neutral view is then generated using the same identity coefficients but with the mean neutral expression coefficients and the input image as texture map (Figure 6).

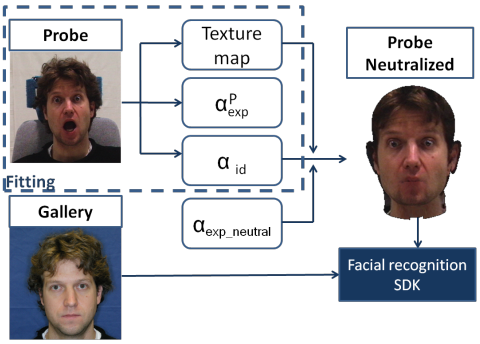

Figure 6. Expression neutralization framework

With this method, each face image is neutralized with the same mean neutral expression.

In this method, each face image is processed independently. This process can thus be performed during the enrolment. The main drawback of this method is that the separation of identity and expression may be inaccurate. Facial deformations related to expression can be assigned to identity part or the other way round. In the next section, we propose a method where the 3DMM is simultaneously fitted to a gallery and a probe pair to better separate identity and expression.

3.2. Expression transfer for verification context

This method is specifically designed for verification ("Am I the person I claim to be ?"). In this context, both the probe and gallery faces are available during the matching process. *In making the assumption that the two face images are of the same identity, their appearance difference under the same pose and roughly the same lighting conditions is thus mainly due to the facial expression variations.* Under such a hypothesis, the 3DMM can be simultaneously fitted to the probe image and the gallery image, using a same set of identity parameters and two different sets of expression coefficients.

Specifically, a unique set of identity coefficients is computed on the two images along with two sets of expression parameters (one for each image). The probe image is then used as texture map (Figure 7).

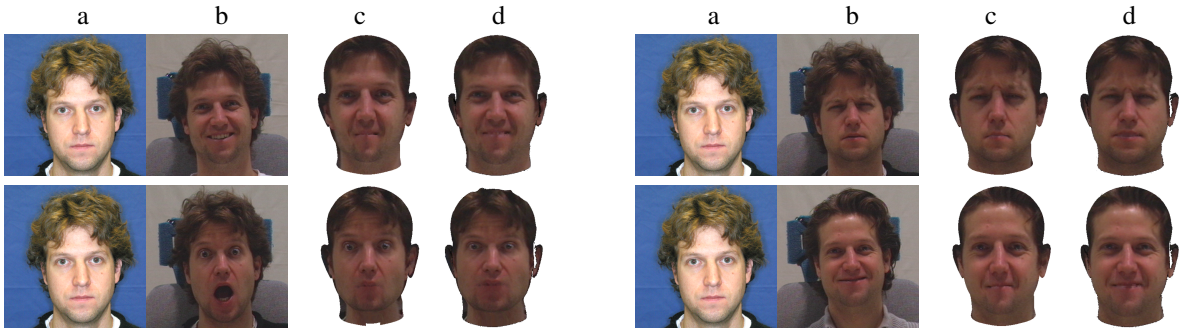

Figure 5. Examples of expression neutralization on the CMU MultiPIE database. Each row contains the gallery image (a), the probe image (b), the result of the expression neutralization using the mean neutral expression (c) and the result of the expression transfer from the gallery to the probe (d).

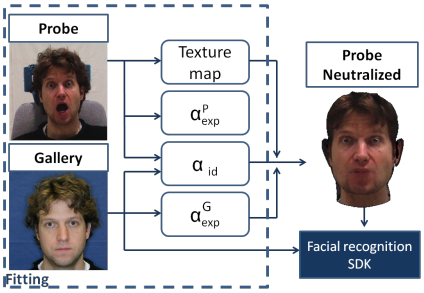

Figure 7. Expression transfer framework

Then, the novel view rendering is performed using the identity coefficients, the expression coefficients extracted from the gallery image and the texture map extracted from the probe.

The key points of this method are this simultaneous fitting of the identity coefficients and the transfer of the gallery expression to the probe. Thanks to this simultaneous fitting, deformations of the face are better separated between identity and expression. Furthermore, compared with the previous mean neutral expression method, expression transfer generates a novel view of the probe with an expression which is the closest to the gallery one. However, given the simultaneous fitting of the 3DMM to the probe and the galley faces, expression transfer cannot be computed offline. As a result, verification is the most suitable applicative scenario of this method.

4. Experimental evaluations

In this section, we demonstrate the robustness of our two methods to expression variations and pose changes using two popular face databases, namely Multie-PIE [10] and AR [14]. For these experiments, mug-shot images are used as reference whereas images with variations in expression, illumination and pose as probe. Results are shown in rank-1 recognition accuracy.

As presented in section 2.3, the fitting of the extended 3D morphable model is initialized using the texture information. To concentrate on the effectiveness of the proposed approach, manually labeled images are used at the initialization step to exclude possible disturbances due to inaccurate landmarks.

To evaluate our work, we use a commercial FR SDK. Different pre-processing configurations are tested. First, the SDK is used without pre-processing. Secondly, we preprocess each probe image using the mean neutral expression method. Finally, the expression transfer method is evaluated : for each query, the expression of the gallery face is transferred to the probe both for matching and non-matching pairs (Figure 8).

Figure 8. Expression transfer with different gallery identities

4.1. CMU-Multi PIE

The CMU Multi-PIE database [10] contains more than 750000 images with variations in pose, illumination and expression of 337 people. Each subject depicts various facial expressions (Smile, surprise, squint, disgust and scream), and 15 poses under 19 illuminations.

Expression variations As reference, neutral faces with ambient illumination and frontal pose are used. To evaluate our work, we use four subsets of this database. As in [26], each subset is related to a specific expression (Smile in session 1, squint and surprise in session 2 and smile in session 3) with frontal pose and different illumination conditions $\{0,2,7,13\}$ for a more challenging recognition.

Table 1 shows the recognition rates for each subsets. We can see that the mean neutral expression method im-

	Sur-S2	Sqi-S2	Smi-S1	Smi-S3	Average	Standard deviation
SRC[25]	51.4%	58.1%	93.7%	60.3%	65.9%	18.9
LLC[22]	52.3%	64.0%	95.6%	62.5%	68.6%	18.7
RRC_L$_2$[27]	59.2%	58.1%	96.1%	70.2%	70.9%	17.7
RRC_L$_1$[27]	68.8%	65.8%	**97.8%**	76.0%	77.1%	14.4
No pre-processing with the commercial SDK	83.7%	89.4%	94.6%	91.5%	89.8%	4.6
Mean neutral expression with the commercial SDK	89.4%	87.0%	94.2%	92.5%	90.8%	3.2
Expression transfer with the commercial SDK	**99.1%**	**95.9%**	**97.8%**	**98.6%**	**97.9%**	1.4

Table 1. Recognition rates on the CMU MultiPIE database on different expression subsets with illuminations variations.

	Commercial SDK with											
	No pre-processing				Mean neutral expression				Expression transfer			
camera	Sur-S2	Sqi-S2	Smi-S1	Smi-S3	Sur-S2	Sqi-S2	Smi-S1	Smi-S3	Sur-S2	Sqi-S2	Smi-S1	Smi-S3
13_0 (30°)	46.6%	58.7%	73.3%	58.7%	62.0%	52.2%	73.5%	66.1%	**82.3%**	**71.9%**	**90.8%**	**86.1%**
14_0 (15°)	66.5%	79.5%	89.5%	79.5%	86.7%	75.1%	90.7%	85.8%	**95.6%**	**89.2%**	**97.2%**	**98.7%**
05_1 (0°)	85.0%	89.7%	94.3%	89.7%	93.5%	87.7%	94.8%	89.4%	**99.5%**	**96.6%**	**99.2%**	**97.8%**

Table 2. Recognition rates on CMU MultiPIE with combined pose and expression variations. To the best of our knowledge, there is no results in the state-of-the-art on these subsets.

proves the overall recognition accuracy of the commercial FR SDK, in particular with the subsets with strong expression deformations (Smile-S3, Surprise-S2). In the two other subsets, this method slightly decreases the performance of the commercial FR SDK. In Squint-S2, the main deformations of the faces are related to the closed eyes and it is hard to the fitting algorithm to affect these deformations to the expression part (closed eyes) or to the identity part (epicanthal fold).
The last row of the table shows the results with the expression transfer method. This method improves the recognition rate of the commercial FR SDK in all the four experiments. The simultaneous fitting of the 3DMM to both the gallery and the probe images makes the problem of fitting more constraint. A better separation between the identity and the expression can thus be achieved.

The expression transfer method clearly outperforms the other methods with respect to expression variations. The last column of Table 1 shows an important decrease of the recognition rate variations with this method.

Expression and pose variations In this section, we present some results of simultaneous expression neutralization and pose normalization on different subsets. Given the lack of such experiments in the state-of-the-art, we designed the following experimental protocol. For each expression (Smile in session 1, Smile in session 3, Surprise in session 2 and Squint in session 2), three subsets with different poses are used (Camera 05_1, 14_0 and 13_0 approximately at 0°, 15° and 30°). Figure 9 shows two examples of simultaneous expression neutralization and pose normalization.

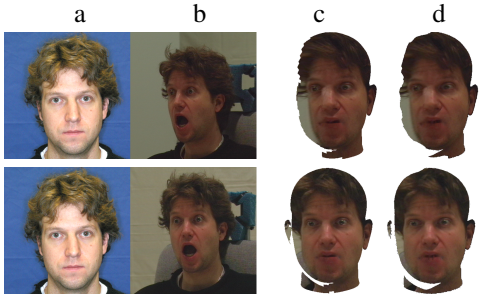

Figure 9. Simultaneous pose normalization and expression neutralization on the CMU Multi-PIE database with surprise expression and moderate pose variations (About 30° (*first row*) and 15° (*second row*)). Each row contains the gallery image (a), the probe image (b), the result of the mean neutral expression (c) and the result of the expression transfer from the gallery to the probe (d).

Table 2 presents the corresponding recognition rates. As it can be seen, the performance of the standard commercial FR SDK is significantly improved when the probe faces are pose normalized and their facial expressions neutralized, either using the mean neutral expression method or the expression transfer method. This improvement is particularly impressive when the expression transfer method is used, leading to a recognition rate increase as high as 36 points for the subset surprise in session 2 (Sur-S2) with a yaw angle of 30°.

4.2. AR database

The AR database contains more than 4000 frontal images of 126 subjects with variation in expressions, illuminations and occlusions. As in [23], we choose a subset (50 male and 50 female subjects) for probes. Since identities

	Smi-S1	Ang-S1	Scr-S1	Smi-S2	Ang-S2	Scr-S2
SRC [25]	98.0%	89.0%	55.0%	79.0%	78.0%	31.0%
PD [20]	**100.0%**	97.0%	93.0%	88.0%	86.0%	63.0%
SOM [21]	**100.0%**	98.0%	88.0%	88.0%	90.0%	64.0%
DICW [24]	**100.0%**	99.0%	84.0%	91.0%	92.0%	44.0%
CTSDP [9]	**100.0%**	**100.0%**	95.5%	98.2%	**99.1%**	**86.4%**
FS [23]	**100.0%**	**100.0%**	91.4%	94.5%	98.0%	58.6%
No pre-processing with the commercial SDK	**100.0%**	**100.0%**	94.0%	**99.0%**	99.0%	76.0%
Mean neutral expression with the commercial SDK	**100.0%**	**100.0%**	96.0%	**99.0%**	99.0%	85.0%
Expression transfer with the commercial SDK	**100.0%**	**100.0%**	**97.0%**	**99.0%**	99.0%	82.0%

Table 3. Recognition rates on the AR database on different expression subsets.

are not specified in the previous works, we randomly chose the subset. The gallery images are neutral faces recorded during session 1. To evaluate our work, six subsets with different expressions are tested.

Table 3 compares the recognition rates of the proposed methods with the state-of-the-art. As it can be seen, face recognition with the scream expression appears to be the most difficult task whereas other expression subsets generate a recognition rate between 99% and 100%. The large deformations of the face when screaming are related to a widely opened mouth and closed eyes. Once more, the proposed two pre-processing methods, in pose normalizing and neutralizing facial expressions, improve the accuracy of the commercial FR SDK. They both outperform all local-based approaches, in particular with the more challenging expressions. In the case of the scream expression in session 2, the deformations of the faces are not only due to the expression but also the temporal variation. In such a case, our methods achieve a comparable performance to CTSP which is a method also based on the warping of face images.

5. Conclusion

We proposed two novel methods to improve the robustness of standard FR SDKs with respect to expression and pose variations. Given the fact that standard FR SDKs are mostly optimized to perform 2D FR with high reliability using frontal and neutral face images, we proposed to synthesize a novel view of the probe where the expression is neutralized and the pose rectified. For this purpose, an extended 3D morphable model, which isolates the identity variations from those due to facial expressions, is used.

First, we presented a method to generate an image with a mean neutral expression. To this end, novel view of the probe is synthesized with expression coefficients extracted on training neutral images. The main advantage of this method is that only the probe is needed during the expression neutralization. This pre-processing can thus be performed during the enrolment.

In order to improve the recognition rate with expressions where identity and expression are difficult to separate, we proposed a second method which transfers the expression of the gallery image to the probe image in making the assumption that they are of the same identity. As both gallery and probe are needed, this method is better suited to verification.

The experiments that we conducted on both Multi-PIE and AR dataset showed that the proposed methods significantly improved the robustness of a standard commercial FR SDK towards expression and pose variations and demonstrated the effectiveness of the proposed methods. We also presented the experimental results of the proposed methods on a challenging experimental protocol using Multi-PIE where pose and facial expression are simultaneously observed on face probe images.

In our future work, we want to improve the extended 3DMM to better handle the facial deformations.

Acknowledgment

We acknowledge support from the French National Research Agency projects Physionomie (ANR-12-SECU-0005-01), Secular (ANR-12-CORD-0014) and Biofence (ANR-13-INSE-0004-02).

References

[1] B. Amberg. *Editing faces in videos*. PhD thesis, University of Basel, 2011. 3

[2] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith. Fully automatic pose-invariant face recognition via 3d pose normalization. In *ICCV 2011*. IEEE. 2

[3] T. Berg and P. N. Belhumeur. Tom-vs-pete classifiers and identity-preserving alignment for face verification. In *BMVC*, 2012. 2

[4] V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. *Comput. Graph. Forum*, 2003. 2

[5] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter. Face recognition based on frontal views generated from non-frontal images. In *CVPR IEEE*, volume 2, 2005. 1, 2, 4

[6] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Siggraph'99. Proceedings*, 1999. 2

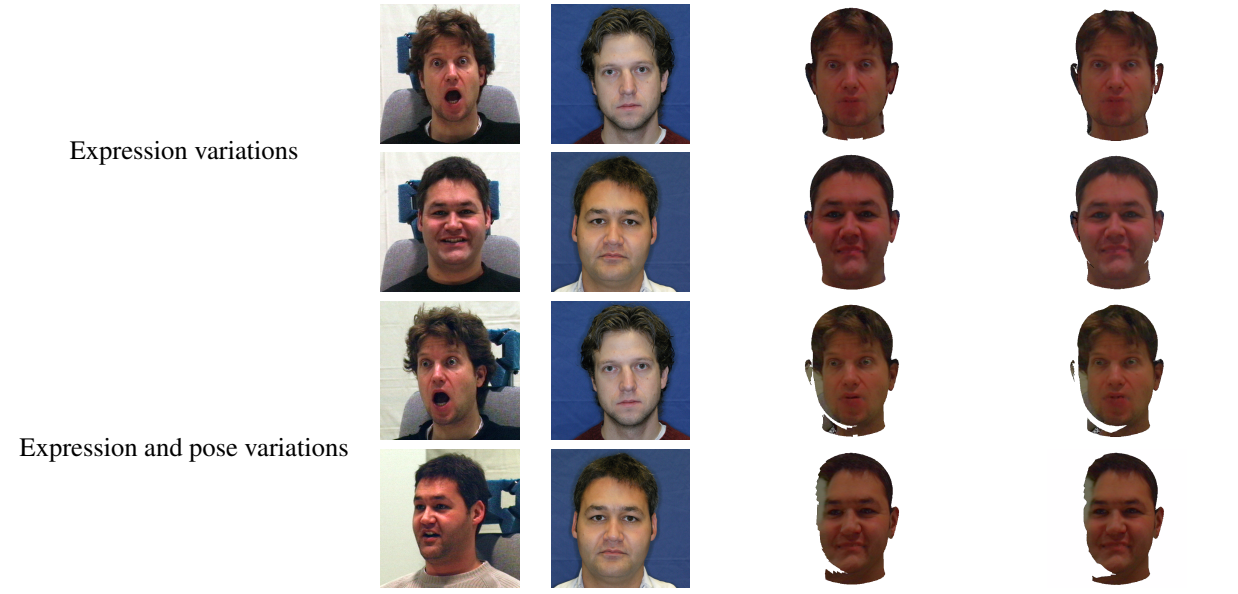

Figure 10. Qualitative evaluation of the proposed methods

[7] D. Cosker, E. Krumhuber, and A. Hilton. A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In *IEEE ICCV 2011*. 3

[8] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3d face recognition under expressions, occlusions and pose variations. *IEEE transactions on pattern analysis and machine intelligence*, 2013. 1

[9] T. Gass, L. Pishchulin, P. Dreuw, and H. Ney. Warp that smile on your face: optimal and smooth deformations for face recognition. In *FG 2011)*. 1, 7

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 2010. 1, 5

[11] P. J. Grother, G. W. Quinn, and P. J. Phillips. Report on the evaluation of 2d still-image face recognition algorithms. *NIST Interagency Rep*, (7709), 2010. 1

[12] G. B. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *IEEE ICCV 2007*. 2

[13] H. Li, D. Huang, P. Lemaire, J.-M. Morvan, and L. Chen. Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities. In *18th IEEE ICIP*, Sept. 2011. 1

[14] A. M. Martinez and R. Benavente. The AR face database. Technical report, CVC, June 1998. 1, 5

[15] I. Matthews, J. Xiao, and S. Baker. 2d vs. 3d deformable face models: Representational power, construction, and real-time fitting. *International journal of computer vision*, 2007. 3

[16] NIST. Video Challenge Problem Multiple Biometric Grand Challenge Preliminary Results of Version 2, 2009. 1

[17] J. Ostermann. Animation of synthetic faces in MPEG-4. In *Computer Animation 98. Proceedings*. 3

[18] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gkberk, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. In *Biometrics and Identity Management*. Springer, 2008. 3

[19] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition*, 2002. 1

[20] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu. Face recognition under occlusions and variant expressions with partial similarity. *Information Forensics and Security*, 2009. 1, 2, 7

[21] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Recognizing partially occluded, expression variant faces from single training image per person with som and soft k-nn ensemble. *IEEE Transactions on Neural Networks*, 2005. 7

[22] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *IEEE CVPR 2010*. 6

[23] X. Wei and C.-T. Li. Fixation and saccade based face recognition from single image per person with various occlusions and expressions. 1, 2, 6, 7

[24] X. Wei, C.-T. Li, and Y. Hu. Face recognition with occlusion using dynamic image-to-class warping (dicw). In *FG13*. 7

[25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence*, 2009. 6, 7

[26] M. Yang, L. Zhang, J. Yang, and D. Zhang. Robust sparse coding for face recognition. In *IEEE CVPR 2011*. 1, 2, 5

[27] M. Yang, L. Zhang, J. Yang, and D. Zhang. Regularized robust coding for face recognition. *IEEE Transactions on Image Processing*, 2013. 6

[28] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In *FG08*. 3