

# Efficient and Effective Face Frontalization for Face Recognition in the Wild

Nelson Méndez<sup>(✉)</sup>, Luis A. Bouza, Leonardo Chang,  
and Heydi Méndez-Vázquez

Advanced Technologies Application Center, 7ma A # 21406, Playa, Havana, Cuba  
{nllanes,lbouza,lchang,hmendez}@cenatav.co.cu

**Abstract.** Face image alignment is one of the most important steps in a face recognition system, being directly linked to its accuracy. In this work we propose a method for face frontalization based on the use of 3D models obtained from 2D images. We first extend the 3D Generic Elastic Model method in order to make it suitable for real applications, and once we have the 3D dense model of a face image, we obtain its frontal projection, introducing a new method for the synthesis of occluded regions. We evaluate the proposal by frontalizing the face images on LFW database and compare it with other frontalization techniques using different face recognition methods. We show that the proposed method allows to effectively align the images in an efficient way.

**Keywords:** Face frontalization · 3D face modeling · Face recognition

## 1 Introduction

Face recognition is one of the most used biometric techniques because of its potential applications. Despite the advances in this area, the accuracy of face recognition systems usually degrades under non-controlled scenarios. Among other factors, pose variation is considered one of the most challenging problems and different approaches have been developed for facing it. Pose-invariant face recognition methods can be roughly divided into four groups [4]: pose-robust feature extraction, multi-view subspace learning, face synthesis based on 2D methods and face synthesis based on 3D methods. Among them, frontal face image synthesis based on 3D methods, has the advantage of effectively reducing the differences on pose between 2D images and the real 3D domain, by using only one image, and without needing a large amount of multi-pose training data.

Existing 3D-based face synthesis approaches are still far from perfect, and new methods are still needed [4]. Two specific problems that are in the center of research is to develop more efficient approaches and to deal with self-occluded regions when frontalizing.

In this paper a new method for face image frontalization is presented. The proposal is based on the 3D Generic Elastic Model (3DGEM) approach, aiming

at synthesize a 3D dense face model more efficiently. Besides, we introduce a new method for effectively recover the information of occluded regions. In order to provide a logical and coherent exposition, this paper is divided as follow: Sect. 2 reviews related work; Sect. 3 makes a general description of 3DGEM approach and describes the proposal; later, in Sect. 4, several experimental results are given, in order to prove the accuracy and efficiency of the method; and finally, conclusions and future work ideas are presented in Sect. 5.

## 2 Related Work

The process of synthesizing a frontal image by using a 3D model has three main steps: face landmarks detection and correspondence, 3D face modeling and 2D frontal face image rendering. In this work we focus on the last two problems.

### 2.1 3D Face Modeling from a Single 2D Image

Existing methods for face frontalization using a single 2D image can be classified as 2D [10, 18] or 3D based [3, 6]. Very good results have been obtained in both categories, but since head rotations occur in the 3D space, 3D methods seems to be more effective [4]. In particular, 3D Morphable Models (3DMM) [2] has reported the synthesis of high quality 3D face models using a single input image. In the last years, several works have been proposed in order to improve this technique. Among them, the 3D Generic Elastic Models (3DGEM) [7], is one of the methods that achieves acceptable visual quality with a high performance. Different works have been then developed in order to improve 3DGEM's synthesis quality [9] and its expression-robust property [15]. The improvement of its performing time is another research topic for this technique. A synthesis process between one or two seconds is the best synthesis time reported in the literature for 3DGEM [9], which is still inappropriate for real time applications.

Recently, other approaches different from 3DGEM have been proposed, but in general they are also time consuming. For example, in [14] a person-specific method is proposed by combining the simplified 3DMM and the Structure-from-Motion methods to improve reconstruction quality. However, the proposal incurs a high computational cost and requires large training data. Yin et al. [21] combine elements of 3DMM with convolutional neural networks, but as any deep learning approach, requires a large amount of training images from different persons in different poses.

### 2.2 Filling of Self-occluded Regions

Different strategies have been proposed for filling of self-occluded parts when frontalizing. The most straightforward approach is to interpolate by using neighboring regions. One of the most popular strategy is to use the face symmetry and used mirrored pixels [5], but this can produce incoherent face texture when there are some differences on both sides of the face. Aiming at reducing the effect

caused by occlusions, Hassner et al. [6] design a “soft symmetry” method that takes into account an occlusion degree estimation. But there are other affecting factors like illumination that should be taken into account. The lighting-normalized face frontalization (LNFF) method [3] estimates the illumination invariant quotient image from the visible parts, and uses it for rendering the estimated lighting on the self-occluded part. The method reports very good results but the process of rendering based on the Quotient Image is time consuming and can not model specular lights or cast shadows in an effective way [3].

### 3 Proposal

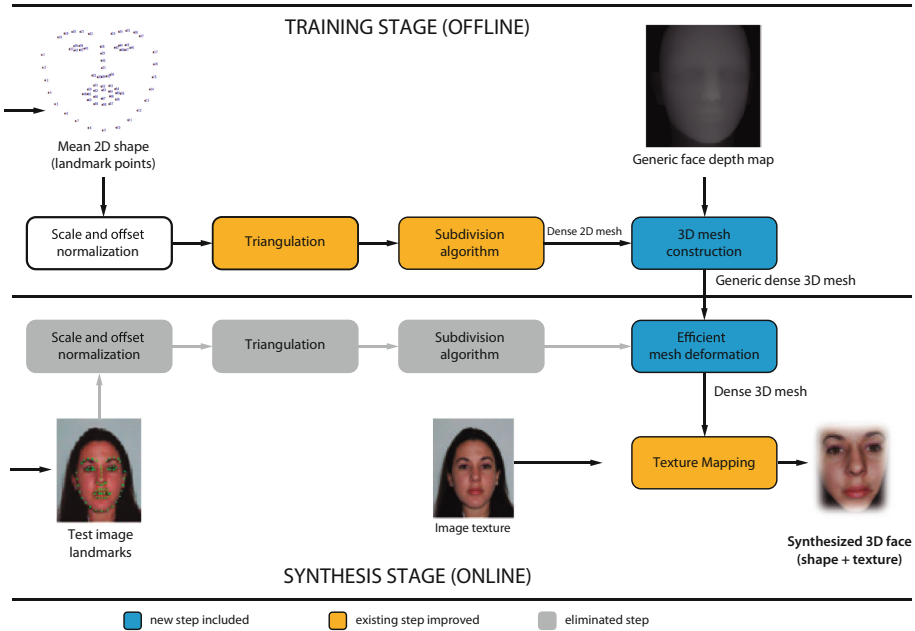
Our method is based on the 3D Generic Elastic Models (3DGEM) [17] approach. An overview of the synthesis process of 3DGEM can be observed in Fig. 1. In the offline or training stage the mean landmark points of the training images are computed. Then these landmarks are transformed in such way to be aligned with their corresponding depth-map. Later, a Delaunay triangulation is performed in order to create a sparse 2D mesh, which finally is refined by some subdivision algorithm, resulting in a dense 2D mesh. Heo [8] proposed the Loop subdivision algorithm to be used. The overall result of the offline stage is a mesh in which, every vertex has its corresponding depth value at the depth-map as a result of to the alignment step. On the other hand, the online or synthesis stage starts with face and landmarks detection on the input image. Then, facial landmarks are spatially transformed to be in the same scale and position of the trained 2D mesh. Afterwards, triangulation and subdivision are performed in the same way as they were executed in the training phase, so trained and subject 2D dense meshes have a 1-on-1 relationship between their vertices. This relationship is used to provide z-coordinates to subject’s mesh, obtaining a 3D dense mesh to which, finally, subject’s texture is applied.

In this work we aim at maintaining the visual quality of the 3D face models obtained by the 3DGEM approach, but improving its facial synthesis processing time. We propose to improve the efficiency of the dense 3D mesh construction process in the synthesis stage and also to enhance the quality of the filling of self-occluded regions.

#### 3.1 Efficient Dense 3D Mesh Construction

One of the most time consuming steps in the original 3DGEM approach is the mesh subdivision process that is followed by the computation of the correspondences between the dense 2D mesh of a given image and the generic depth map. These correspondences must be computed for every image during training and testing, given that the original used Delaunay triangulation cannot guarantee a mesh with the same configuration.

In order to gain in efficiency, we propose to remove the mesh subdivision and correspondence process in the online stage. To do this, it is first necessary to modify the offline stage. The Delaunay triangulation is replaced by a fixed one,



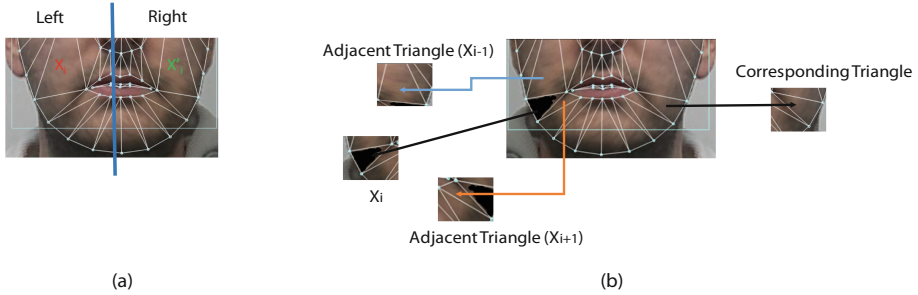
**Fig. 1.** 3DGEN synthesis process diagram. In blue the new steps included in our proposal are depicted, i.e., generic 3D mesh construction in the offline stage and an efficient mesh deformation algorithm in the syntesis stage. These new steps replace the steps filled in gray in order to speed up the synthesis process. In orange, the improved steps of 3DGEN are also illustrated. This figure is based on a diagram exposed by Prabhu et al. in [17]. (Color figure online)

and instead of using a raw subdivision algorithm, an adaptive mesh subdivision technique is applied, which iteratively obtains six triangles on every step, leading to a higher quality and a less complex 3D mesh. Then, the 3D dense mesh is generated in the training stage by using the aligned 2D dense mesh (resulting from the subdivision step) and its corresponding depth-map. Finally, in order to obtain the dense 3D mesh on the online stage, we introduce an efficient 3D mesh deformation algorithm described in [13]. This method uses bounded biharmonic deformations to efficiently and accurately deform a 3D mesh by using a few reference 2D points. In our case the dense 3D mesh learned in the offline stage is deformed taking as input 14 points belonging to the eyes, mouth and nose from the automatic landmarks detected on a given image.

These changes to the 3DGEN pipeline are the key-components of our proposal that allow a faster synthesis process. The differences between 3DGEN and our proposal are highlighted in Fig. 1.

### 3.2 Symmetric Interpolation for Self-occlusion

Once we have the dense 3D mesh for the test image and their corresponding 2D points, every triangular region in the mesh is filled with the texture values



**Fig. 2.** Use of symmetric interpolation for self-occluded regions. In (a) the symmetric regions of the face and a sample triangulation are shown. In (b) the adjacent and corresponding triangles according to face symmetry are shown.

according the corresponding region in the image. However, because of pose variations, there are some regions in which no texture information is available. In 3DGEM, this information is completed by interpolating the texture in the corresponding region based on the symmetry of the face. In order to avoid interpolation artifacts in the resulting texture, in this paper we propose to include the adjacent triangles of the occluded regions in the interpolation process.

For better understanding we divided the face into two regions, left  $X$  and right  $X'$ , as is depicted in Fig. 2(a). As can be seen, in these regions every triangle formed by the points of the face has its corresponding triangle on the opposite side of the face. A given triangle in  $X$  is denoted as  $X_i$  and its corresponding triangle in  $X'$  as  $X'_i$ .

Let  $X_{i+1}$  and  $X_{i-1}$  be the triangles adjacent to triangle  $X_i$ , respectively. Let  $x_{int} = \Upsilon(x_1, x_2)$  be a linear interpolation function between two given pixel values. For every pixel value  $x_i \in X_i$  exist two pixels values  $x_{i+1} \in X_{i+1}$  and  $x_{i-1} \in X_{i-1}$  given by an assignment function and its corresponding pixel  $x'_i \in X'_i$ . Then, our interpolation function is defined as:

$$x_i = \Upsilon\left(\frac{x_{i-1} + x_{i+1}}{2}, x'_i\right). \quad (1)$$

This process is illustrated on Fig. 2(b), and allows us to obtain the estimation of the values of the occluded regions in a simple but effective way.

## 4 Experimental Evaluation

In order of evaluating our approach we conducted experiments on the popular Labeled Faces in the Wild (LFW) database [12]. It contains 13233 images of 5749 people downloaded from the Web. For verification evaluation, the data is divided into 10 disjoint splits, which contain different identities and come with a list of 600 pre-defined image pairs for evaluation: 300 genuine comparisons and 300 impostors.

We use our method as a preprocessing step applied to the face images before the verification process and compare it with the provide images with different alignment techniques: LFW-a, aligned with a commercial software [20]; with funneling [11]; and with deep funneling [10]. In order to evaluate the performance of our proposal for different face descriptors, we use three state-of-the-art methods with available implementations: the Convolutional Neural Network (CNN) provided by the DLib library [1]; the VGG-face deep network [16]; and Fisher Vector method [19]. DLib is also used for face and landmarks detection.

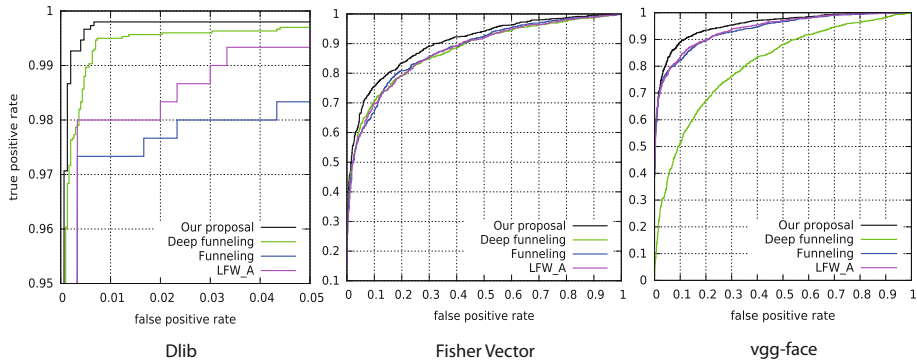
We follow the LFW evaluation protocol, and the Receiver Operating Characteristic (ROC) curves and Area Under ROC Curve (AUC) values are used as evaluation metrics. In Table 1, the AUC values obtained for the three face recognition methods with the different alignment techniques are provided. The obtained ROC curves are shown in Fig. 3.

As can be seen from the Fig. 3 and Table 1, in general the proposed frontalization technique achieves the best results for the three different tested methods, showing its capabilities and benefits in the face recognition process.

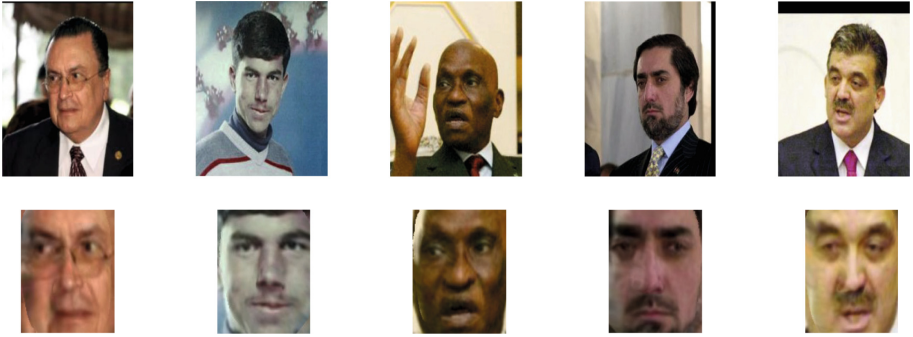
Figure 4 shows some example results of our frontalization method applied to several images from the LFW database. It can be seen how the method is able to correct different levels of pose variations and effectively reconstruct the self-occluded regions.

**Table 1.** AUC values in LFW for different alignment methods.

Alignment method	DLib	Fisher vector	vgg-face
Deep funneling	<b>0.9991</b>	0.8636	0.8061
Funneling	0.9924	0.8736	0.8507
LFW-a	0.9960	0.8743	0.9048
Our proposal	<b>0.9991</b>	<b>0.9024</b>	<b>0.9125</b>



**Fig. 3.** ROC curves for the three face recognition methods with different alignments.



**Fig. 4.** Example results of our proposed frontalization method in some images of the LFW dataset. First row: original images. Second row: images frontalized with the proposed method.

#### 4.1 Computational Efficiency Experiments

The computational efficiency of the proposed method in the synthesis stage has been also evaluated and compared to other methods. We recorded the computational processing time over 200 randomly selected images from the LFW dataset and reported the average processing time for an image. Compared to the 75 ms exhibited by LFW3D [6] and the 1276 ms exhibited by the original 3DGEM [17] for frontalizing an image, our proposal took 33 ms (i.e., 2.3x and 38x faster, respectively). Both LFW3D and 3DGEM were evaluated using the publicly available codes provided by their respective authors. Our proposed method was implemented in C++ using OpenCV. All the experiments were conducted on a single thread of an Intel i7-4770 CPU @3.40 GHz, 8 GB RAM.

## 5 Conclusions

In this paper we present a new method for face image frontalization. The proposal is based on the 3DGEM approach but introducing several changes that allows to perform efficiently the online stage. Our proposal exhibited a speed-up of 38x with respect to the original 3DGEM method, making it more feasible for real applications. Besides, we designed a simple method for filling the occluded regions using adjacent and opposite-side regions, that shows to recover the missing facial information in an effective way. The proposed method was compared in the challenging LFW dataset with three different alignment techniques, with three state-of-the-art recognition methods, and achieved the best results in all the experiments.

## References

1. DLib C++ library kernel description. <http://dlib.net/>. Accessed 02 May 2017
2. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. Technical Report SIGGRAPH (1999)

3. Deng, W., Hu, J., Wu, Z., Guo, J.: Lighting-aware face frontalization for unconstrained face recognition. *Pattern Recogn.* **68**, 260–271 (2017)
4. Ding, C., Tao, D.: A comprehensive survey on pose-invariant face recognition. *ACM Trans. Intell. Syst. Technol. (TIST)* **7**(3), 37 (2016)
5. Ding, L., Ding, X., Fang, C.: Continuous pose normalization for pose-robust face recognition. *IEEE Sig. Process. Lett.* **19**(11), 721–724 (2012)
6. Hassner, T., Harel, S., Paz, E., Enbar, R.: Effective face frontalization in unconstrained images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4295–4304 (2015)
7. Heo, J.: 3D generic elastic models for 2D pose synthesis and face recognition. Technical Report (2009)
8. Heo, J., Savvides, M.: 3-D generic elastic models for fast and texture preserving 2-D novel pose synthesis. *Trans. Inf. Forensics Secur.* **7**(2), 563–576 (2012)
9. Heo, J., Savvides, M.: Gender and ethnicity specific generic elastic models from a single 2D image for novel 2D pose face synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(12), 2341–2350 (2012)
10. Huang, G., Mattar, M., Lee, H., Learned-Miller, E.G.: Learning to align from scratch. In: *NIPS*, pp. 764–772 (2012)
11. Huang, G.B., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: *ICCV*, pp. 1–8. IEEE (2007)
12. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts, Amherst (2007)
13. Jacobson, A.: Algorithms and interfaces for real-time deformation of 2D and 3D shapes. Ph.D. Thesis, ETH Zurich (2013)
14. Jo, J., Choi, H., Kim, I.J., Kim, J.: Single-view-based 3D facial reconstruction method robust against pose variations. *Pattern Recogn.* **48**(1), 73–85 (2015)
15. Moeini, A., Moeini, H., Faez, K.: Pose-invariant facial expression recognition based on 3D face reconstruction and synthesis from a single 2D image. In: *ICPR*, pp. 1746–1751 (2014)
16. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: *BMVC*, vol. 1, p. 6 (2015)
17. Prabhu, U., Heo, J., Savvides, M.: Unconstrained pose-invariant face recognition using 3D generic elastic models. *Trans. Pattern Anal. Mach. Intell.* **33**(10), 1952–1961 (2011)
18. Sagonas, C., Panagakis, Y., Zafeiriou, S., Pantic, M.: Robust statistical face frontalization. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3871–3879 (2015)
19. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild. In: *BMVC*, vol. 2, p. 4 (2013)
20. Wolf, L., Hassner, T., Taigman, Y.: Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(10), 1978–1990 (2011)
21. Yin, X., Yu, X., Sohn, K., Liu, X., Chandraker, M.: Towards large-pose face frontalization in the wild. *arXiv preprint [arXiv:1704.06244](https://arxiv.org/abs/1704.06244)* (2017)