High-Fidelity Pose and Expression Normalization for Face Recognition in the Wild

Xiangyu Zhu, Zhen Lei, Junjie Yan, Dong Yi, Stan Z. Li
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

Pose and expression normalization is a crucial step to recover the canonical view of faces under arbitrary conditions, so as to improve the face recognition performance. Most normalization algorithms can be divided in to 2D and 3D methods. 2D methods either estimate a flow to simulate the 3D geometry transformation or learn appearance transformations between different poses. 3D methods estimate the depth information with a face model and normalize faces through 3D transformations.

An ideal normalization is desired to preserve the face appearance with little artifact and information loss, which we call high-fidelity. However, most previous methods fail to satisfy that. In this paper, we present a 3D pose and expression normalization method to recover the canonical-view, expression-free image with high fidelity. It contains three components: pose adaptive 3D Morphable Model (3DMM) fitting, identity preserving normalization and invisible region filling, which is briefly summarized in Fig. 1.

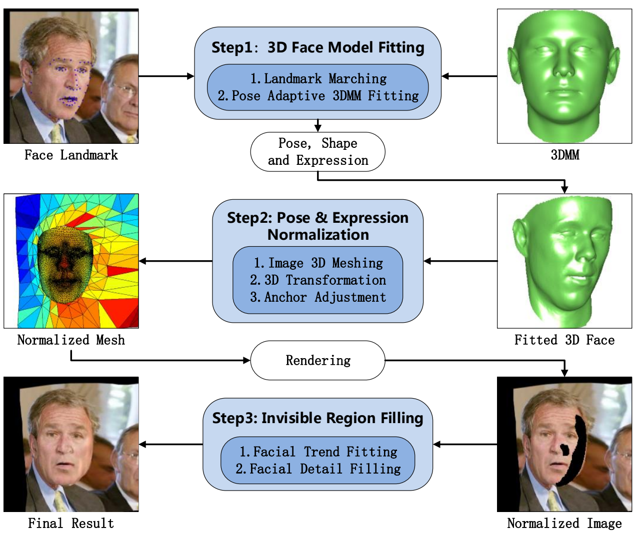

Figure 1: Overview of the High-Fidelity Pose and Expression Normalization (HPEN) method

With an input image, the landmarks are detected with the face alignment algorithm and we mark the corresponding 3D landmarks on the face model. Then the 3DMM can be fitted by minimizing the distance between the 2D landmarks and projected 3D landmarks:

$$\arg \min_{f,R,t_{3d},\alpha_{id},\alpha_{exp}} \|s_{2d} - fPR(\overline{S} + A_{id}\alpha_{id} + A_{exp}\alpha_{exp} + t_{3d})\| \quad (1)$$

where α_{id} is the shape parameter, α_{exp} is the expression parameter. f, R, t_{3d} are pose parameters. However, when faces deviate from the frontal pose, the correspondence between 2D and 3D landmarks will be broken, which we model as "landmark marching": when pose changes, the contour landmarks move along the parallel to the visibility boundary, see Fig. 2(a). To deal with the phenomenon we propose an approximation method to adjust contour landmarks during 3DMM fitting. The 3D model are firstly projected with only yaw and pitch to eliminate in-plane rotation. Then for each parallel, the point with extreme x coordinate will be chosen as the marching destimation, see Fig. 2(b).

With the fitted 3DMM, The face can be normalized through 3D transformations. In this paper we also normalize the external face region which contains discriminative information as well. Firstly we mark three groups of anchors which are located on the face boundary, face surrounding and image contour, see Fig. 3(a). Then their depth are estimated by enlarging the fitted

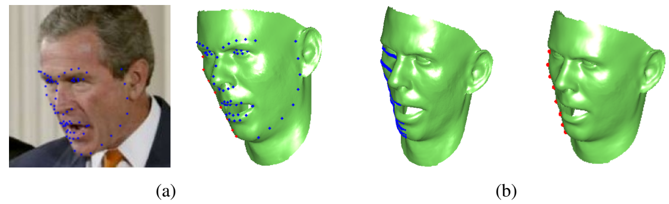

Figure 2: (a) The landmark marching phenomenon. The red points are the moved contour landmarks. (b) Project 3D face with only yaw and pitch and get adjusted landmark positions.

3D face and the whole image is turned into a 3D object through triangulation, see Fig. 3(b). Finally, the 3D meshed face object are normalized with inverse rotation and expression resetting, see Fig. 3(c).

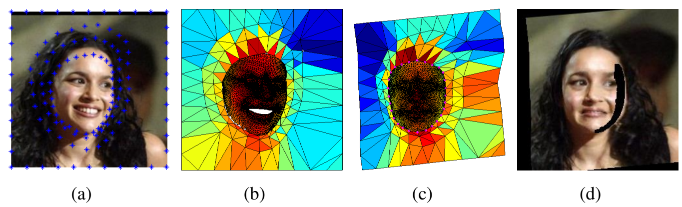

Figure 3: (a) The anchors. (b) The meshed face object (c) The normalized face object (d) The normalization result.

If the yaw angle of face is too large, there may be some regions become invisible due to self-occlusion. The basic idea of dealing with self-occlusion is utilizing the facial symmetry. However, due to the existence of illumination, facial symmetry cannot always hold. Directly copying pixels will lead to non-smoothness and weird illumination. In this paper, we propose a new way to deal with the invisibility: Fitting the trend and filling the detail, which deals with illumination and texture separately.

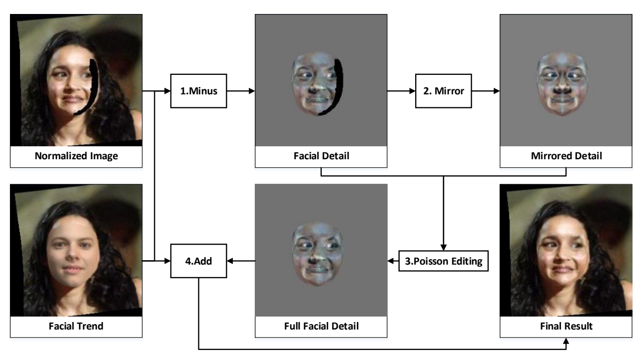

Figure 4: The flowchart of facial detail filling.

We define the facial trend as the illuminated mean face texture, which can be estimated with the spherical harmonic reflectance bases constructed from the fitted 3D model and the mean face texture. Then the difference between the image pixel and the facial trend can be seen as the illumination-free facial detail, which roughly satisfies the symmetry assumption. In order to keep the smoothness of filling boundary, we adopt the Poisson editing method to insert the mirrored facial detail into the invisible region. Finally, the facial trend and facial detail are added to form the final result. Fig. 4 demonstrates the process of facial detail filling.

The implementation details are described in the paper and the code is released. The experiments demonstrate that we achieve the state-of-the-art on both LFW and MultiPIE.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.