

# 3D Face Morphable Models "In-the-Wild"

Chaonan Song

May 25, 2018

## 1 Model Training

In this paper, they make a number of contributions that enable the use of 3DMMs for "in-the-wild" face reconstruction (Fig. 1).



Figure 1: Our "in-the-wild" Morphable Model is capable of recovering accurate 3D facial shape for a wide variety of images.

A 3DMM consists of three parametric models: the *shape*, *camera* and *texture* models.

### 1.1 Shape Model

They denote the 3D mesh (shape) of an object with  $N$  vertexes as a  $3N \times 1$  vector shown in Eq. 1.

$$s = [X_1^T, \dots, X_N^T] = [x_1, y_1, z_1, \dots, x_N, y_N, z_N] \quad (1)$$

where  $x_i = [x_i, y_i, z_i]^T$  are the object-centered Cartesian coordinates of the  $i$ -th vertex. This model can be used to generate novel 3D shape instances using

the function  $\mathcal{S} : \mathbb{R}^{n_s} \rightarrow \mathbb{R}^{3N}$  as Eq. 2.

$$\mathcal{S}(P) = \hat{s} + U_s P \quad (2)$$

where  $p = [p_1, \dots, p_{n_s}]^T$  are the  $n_s$  shape parameters.

### 1.2 Camera Model

The purpose of the camera model is to map (project) Cartesian coordinates centered on the object of the 3D mesh to 2D Cartesian coordinates on the image plane. In this work, they used a pinhole camera model that uses a perspective transformation.

**Perspective projection.** The projection of a 3D point into its 2D location include two steps. First, the 3D point is rotated and translated using a linear view transformation, under the assumption that the camera is still as Eq. 3

$$[v_x, v_y, v_z]^T = \mathbf{R}_v \mathbf{x} + \mathbf{t}_v \quad (3)$$

where  $\mathbf{R}_v \in \mathbb{R}^{3 \times 3}$  are the 3D rotation and translation components, respectively. Then, a nonlinear *perspective* transformation is applied as Eq. 4

$$X' = \frac{f}{v_z} \begin{bmatrix} v_x \\ v_y \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \quad (4)$$

where  $f$  is the focal length in pixel units (they assume that the  $x$  and  $y$  components of the focal length are equal) and  $[c_x, c_y]^T$  is the principal point that is set to the image center.

**Camera function.** The projection operation performed by the camera model of the 3DMM can be expressed with the function  $\mathcal{P}(s, c) : \mathbb{R}^{3N} \rightarrow \mathbb{R}^{2N}$  which applies the transformations of Eqs. 3 and 4 on the points of provided 3D mesh  $s$  with Eq. 5

$$\mathbf{c} = [f, q_1, q_2, q_3, t_x, t_y, t_z]^T \quad (5)$$

being the vector of camera parameters with length  $n_c = 7$ . For abbreviation purposes, we represent the camera model of the 3DMM with the function  $\mathcal{W} : \mathbb{R}^{n_s, n_c} \rightarrow \mathbb{R}^{2N}$  as Eq. 6.

$$\mathcal{W}(\mathbf{p}, \mathbf{c}) \equiv \mathcal{P}(\mathcal{S}(\mathbf{p}), c) \quad (6)$$

where  $\mathcal{S}(\mathbf{p})$  is a 3D mesh instance using Eq. 2.

### 1.3 “In-the-Wild” Feature-Based Texture Model

The generation of an “in-the-wild” texture model is a key component of the proposed 3DMM. To this end, we take advantage of the existing large facial “in-the-wild” databases that are annotated in terms of sparse landmarks. They also define a dense feature extraction function as Eq. 7.

$$\mathcal{F} : \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{H \times W \times C} \quad (7)$$

where  $C$  is the number of channels of the feature-based image. For each image, they first compute its feature-based representation as  $\mathbf{F}_i = \mathcal{F}(\mathbf{I}_i)$  and then use Eq. 6. to sample it at each vertex location to build back a vectorized texture sample  $\mathbf{t}_i = \mathbf{F}_i(\mathcal{W}(\mathbf{p}_i, \mathbf{c}_i)) \in \mathbb{R}^{CN}$ . This texture sample will be nonsensical for some regions mainly due to self-occlusions present in the mesh projected in the image space  $\mathcal{W}(\mathbf{p}_i, \mathbf{c}_i)$ . Finally, an iterative procedure can be used to refine the texture. That is, starting with the 3D fits provided by using only the 2D landmarks [1], a texture model can be learned using the above procedure. This texture model used with the proposed 3DMM fitting algorithm on the same data establishes improved correspondences, refining the texture model.

## References

- [1] Amin Jourabloo and Xiaoming Liu. Large-pose face alignment via cnn-based dense 3d model fitting. In *Computer Vision and Pattern Recognition*, 2016.