3D Face Reconstruction from Limited Images Based on Differential Evolution

Qun Wang
Old Dominion University

Jiang Li
Old Dominion University, jli@odu.edu

Vijayan K. Asari

Mohammad A. Karim
Old Dominion University

Andrew G. Tescher (Ed.)

Original Publication Citation
Wang, Q., Li, J., Asari, V. K., & Karim, M. A. (2011) 3D face reconstruction from limited images based on differential evolution. In A. G. Tescher (Ed.), *Applications of Digital Image Processing XXXIV, Proceedings of SPIE Vol. 8135* (813518). SPIE of Bellingham, WA. https://doi.org/10.1117/12.894253

3D Face Reconstruction from Limited Images based on Differential Evolution

Qun Wang[a], Jiang Li[a], Vijayan K. Asari[b], Mohammad A. Karim[a]
[a]Dept. of Electrical and Computer Engineering, Old Dominion University, VA, 23508;
[b]Dept. of Electrical and Computer Engineering, University of Dayton, OH, 45469

ABSTRACT

3D face modeling has been one of the greatest challenges for researchers in computer graphics for many years. Various methods have been used to model the shape and texture of faces under varying illumination and pose conditions from a single given image. In this paper, we propose a novel method for the 3D face synthesis and reconstruction by using a simple and efficient global optimizer. A 3D-2D matching algorithm which employs the integration of the 3D morphable model (3DMM) and the differential evolution (DE) algorithm is addressed. In 3DMM, the estimation process of fitting shape and texture information into 2D images is considered as the problem of searching for the global minimum in a high dimensional feature space, in which optimization is apt to have local convergence. Unlike the traditional scheme used in 3DMM, DE appears to be robust against stagnation in local minima and sensitiveness to initial values in face reconstruction. Benefitting from DE's successful performance, 3D face models can be created based on a single 2D image with respect to various illuminating and pose contexts. Preliminary results demonstrate that we are able to automatically create a virtual 3D face from a single 2D image with high performance. The validation process shows that there is only an insignificant difference between the input image and the 2D face image projected by the 3D model.

Keywords: 3D face modeling, 3D morphable model, differential evolution, face recognition

1. INTRODUCTION

3D face modeling has been one of the greatest challenges for researchers in computer graphics for many years. Various methods have been used to model the shape and texture of faces under varying illumination and pose conditions from a single given image. Researchers introduced Shape-From-Shade (SFS) [1] [2] to reconstruct the 3D surface of human faces. Unfortunately, SFS shows rapid decrement in performance, such as biased calculations and improper estimates of surface normals caused by varied lighting conditions and cast shadows on the 2D image. Active appearance model (AAM) [3], proposed by Cootes *et al.*, is a statistical deformable technique which has been widely used in computer vision. However, AAM only allows a small range of out-of-plane rotation and displays inadaptation to directed light sources.

In this paper, we propose a novel method for the 3D face synthesis and reconstruction by using a simple and efficient global optimizer. A 3D-2D matching algorithm which employs the integration of 3D morphable model (3DMM) and differential evolution (DE) is addressed. 3DMM is a generic modeling technique based on parametric representation of shape and texture information, which could be used to synthesize individual 3D faces from 2D images. The morphable model is advantaged by its no restriction on the requirement of illumination or reflectance functions though it has additional computational complexity. In 3DMM, the estimation process of fitting shape and texture to 2D images is considered as the problem of approximating global minima in a high dimensional feature space. To this end, DE is integrated in our model, which demonstrates ease of manipulation and excellent convergence properties of moving away from local optima[9] [10].

The remaining sections of the paper are organized as follows: The concept of morphable model is introduced in Section 2. Model matching method based on DE is presented in Section 3. The final experimental results are outlined in Section 4 and we conclude the paper in Section 5.

Applications of Digital Image Processing XXXIV, edited by Andrew G. Tescher, Proc. of SPIE Vol. 8135, 813518 · © 2011 SPIE · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.894253

2. 3D MORPHABLE MODEL

2.1 Introduction

In 1999, T. Vetter and T. Poggio proposed 3DMM [6], a realistic modeling method for 3D faces synthesis by representing the linear combination of exemplar faces. The reconstruction procedure is regarded as conducting iterations of the analysis-by-synthesis process, which are driven by fitting the 3D model to 2D images. Meanwhile, the parameters with respect to 3D environment such as focal length of the camera, illumination and color contrast, can also be modeled explicitly and estimated automatically.

2.2 Model Construction

The prototypical 3D faces are acquired by 3D laser scanners, whose range and texture data are digitized with high precision. Preprocessed through registration and texture extraction, each face is represented in the form of a shape vector and a texture vector as:

$$S = (X_1, Y_1, Z_1, X_2, ..., Y_n, Z_n)^T \tag{1}$$

$$T = (R_1, G_1, B_1, R_2, ..., G_n, B_n)^T \tag{2}$$

where n is the number of vertexes on the 3D face and $(B_j\ G_j\ R_j)$ are the corresponding R, G, B color values of the vertex $(X_j\ Y_j\ Z_j)$. Therefore, a morphable model can be generated by using the linear combination of shape vectors S_i and texture vectors T_i of 3D training faces as [6]:

$$S_{model} = \sum_{i=1}^{m} a_i S_i \qquad T_{model} = \sum_{i=1}^{m} b_i T_i \tag{3}$$

$$\sum_{i}^{m} a_i = \sum_{i}^{m} b_i = 1$$

in which m is the number of training faces, Si and Ti are shape and texture of training faces and a_i and b_i are their corresponding weights contributed to the new face with $0 < a, b < 1$.

In the practical consideration of computational effectiveness, a common technique as PCA (Principal Component Analysis) is employed to reduce the high dimensionality of 3D face data without the loss of potential face information. In particular, PCA performs a transformation of the original cloud data to an orthogonal coordinate system formed by the eigenvectors s_i and t_i of the covariance matrices.

$$\mathbf{S}_{model} = S_{mean} + \sum_{i=1}^{m-1} \alpha_i \mathbf{s}_i \qquad T_{model} = T_{mean} + \sum_{i=1}^{m-1} \beta_i T_i \tag{4}$$

where S_{mean} and T_{mean} are the average shape and texture vectors. S_i and T_i are principal components. $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$ and $\beta = (\beta_1, \beta_2, ..., \beta_n)$ are shape and texture combination coefficients, and α and β obeys Gaussian distribution as:

$$p(\boldsymbol{\alpha}) = \exp\left(-\frac{1}{2}\sum_{i=1}^{m-1}\left(\frac{\alpha_i}{\sigma_{S,i}}\right)^2\right) \qquad p(\beta) = \exp\left(-\frac{1}{2}\sum_{i=1}^{m-1}\left(\frac{\beta_i}{\sigma_{T,i}}\right)^2\right) \tag{5}$$

2.3 Model Matching

Matching the 3D face morphable model to the given face images is a process of model parameter estimation, in which a number of coefficients are required to be determined. For example, camera and illumination model is adopted in the projection of the 3D face model into the image plane since 3D face model and 2D input facial images cannot be measured directly. Aiming at retrieving a 3D face the closest projective image to the input facial image, the error function between 3D model projective image I_{mod} and input image I_{input} is described as:

$$E_I = \sum_{x,y} \left\| I_{input}(x,y) - I_{\mathrm{mod}}(x,y) \right\|^2 \tag{6}$$

In order to create a realistic 2D output which is close enough to the target face image, we make use of the perspective projection and Phong illumination model in the rendering process. Given the k^{th} vertex at (*X,Y,Z*) with texture value (*R,G,B*), the perspective projection on the image plane is represented as:

$$I_k(x,y) = \left(I_{r,k}(x,y), I_{g,k}(x,y), I_{b,k}(x,y) \right)^T \tag{7}$$

where $I_{c,k}(x,y)$ is computed under the Phong illumination model as:

$$I_{c,k}(x,y) = R\left(I_{a,c} + I_{dir,c}(L \cdot N) \right) + K_s\, I_{dir,c}(F \cdot V)^n \tag{8}$$

$I_{a,c}$ and $I_{dir,c}$ are separately intensity of ambient light and direct light of the c^{th} color component. K_s is the reflectance, *L, N, F, V* are light direction, normal, reflective direction and direction of viewer respectively and *n* is the mirror reflectance index.

3. GLOBAL OPTIMIZATION

The morphable model provides an approach in solving face modeling problems under different illumination and pose conditions. However, one of the issues lies in the process of minimizing the cost function (6) that performs error evaluation in the pixel-level measurement. This involves algorithms of image matching and a large-scale optimization. In 3DMM, fitting shape and texture into 2D images is equal to searching for the global minimum in a high dimensional feature space, in which optimization is apt to have local convergence. Stochastic gradient descent [6] and Levenberg-Marquardt [11] method are used to evaluate the residual and global error as well as objective function optimization. Differential Evaluation (DE) appears to be robust against stagnation in local minima and sensitiveness to initial values in face reconstruction. Considering its successful performance, we tentatively introduce DE to tackle the problem in 3D-2D matching.

3.1 Differential Evolution (DE)

Differential Evolution (DE) is a "parallel direct search method" [9], which was first proposed by Storn and Price in 1995 [10]. It is characterized as a stochastic and population-based optimization that is simple and effective for implementation. DE repeatedly processes through operations which are, in turn, "mutation, crossover and selection" until an optimal solution to the objective function $f(x)$ is reached (Figure 1).

The classic version of DE is defined as follows. Suppose we have *N D*-dimensional parameter vectors

$$x_{i,G} = \left[x_1, x_2, ..., x_D \right]^T, i = 1, 2, ..., N \tag{9}$$

representing the population for generation *G*. The algorithm starts by randomly initializing the vector populations with, as the author suggested, a uniform probability distribution [9]. We use a different distribution in our experiment due to the special feature of 3DMM, which we will present later in section 4.

3.1.1 Mutation

For each individual x_i, a corresponding mutation vector v_i is produced according to the equation:

$$v_{i,G} = x_{r1,G} + F * \left(x_{r2,G} - x_{r3,G} \right) \tag{10}$$

in which random index $r1, r2, r3 \in \{1, 2, ..., N\}$ and $r1 \neq r2 \neq r3$. F is a real amplifier designed to control the offset of $v_{i,G}$ to $x_{r1,G}$ by scaling the differential variation $\left(x_{r2,G} - x_{r3,G} \right)$.

3.1.2 Crossover

Trial vectors are introduced in the phase of crossover to expand the range of global search. It is defined in the form as:

$$u_{i,G} = \left(u_{1i,G}, u_{2i,G}, ..., u_{Di,G} \right) \tag{11}$$

in which

$$u_{ji,G} = \begin{cases} v_{ji,G}, if \left(randb(j) \leq CR \right) \vee j = I_{br} \\ x_{ji,G}, if \left(randb(j) > CR \right) \wedge j \neq I_{br} \end{cases} \qquad j = 1, 2, ..., D. \tag{12}$$

In equation (9), $randb(\cdot)$ is a random generator with uniform distribution. I_{br} is an integer randomly chosen from *{1,2,...,D}*, which prevents $x_{i,G}$ from being equal to $u_{i,G}$.

3.1.3 Selection

DE utilizes pair-wise comparison between $u_{i,G}$ and $x_{i,G}$ to survive the vectors with fewer objectives function values to the next generation.

$$x_{i,G+1} = \begin{cases} u_{i,G}, if \left(f\left(u_{i,G}\right) < f\left(x_{i,G}\right) \right) \\ x_{i,G}, otherwise \end{cases} \qquad i = 1, 2, ..., N \tag{13}$$

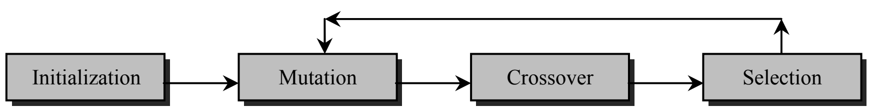


Figure 1. Canonical Differential Evolution Procedure

4. EXPERIMENTAL RESULTS

4.1 3D Face Modeling

Various 3D face databases have been established during the last decades [15]. In our experiments, we use the 3D Basel Face Model (BFM) database [8] to derive the morphable model for 3D shape and texture. The database collects 200 3D faces from 100 male and 100 female subjects, each of which keeps neutral expression, without makeup, accessories and glasses. The registered 3D faces are parameterized as triangular meshes with 53490 vertices [8]. Figure 2 shows the mean face of the 200 faces in the database, which is represented as S_{mean} and T_{mean} in (4).

Figure 2. 3D mean face from BMF database consists of 100 male faces and 100 female faces

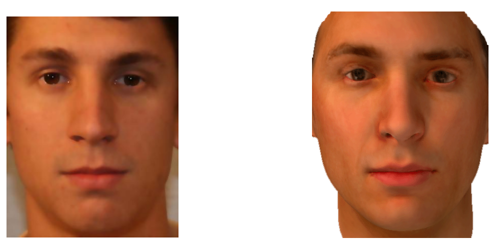
4.2 3D Face Model Training

As aforementioned, our framework for 3D model reconstruction is intuitively led by 3D morphable model equipped with DE. In our experiment, DE populations are initialized by Gaussian distribution instead of uniform distribution. The reason for this is, according to the equation (5), shape and texture combination coefficients, α and β, both obey Gaussian distribution.

In BFM database, both shape and texture are represented by 200 principal components. Due to the consideration of computational efficiency, we only use the first 70 components of S_i and T_i for face training. Even then, final results still show that these components are competent for sound outcome.

4.3 Learning 3D Face Models

In this section, we show our experimental results obtained over lighting and pose variations. We use partial images from the ODU-VL face database as the training image. Figure 3 shows the original image and synthetic image of a test individual



(a) (b)

Figure 3. original image and synthesized image, (a) original image; (b) synthesized image

4.3.1 3D Scene Re-rendering

Once the 3D model is reconstructed, it is ready to re-render the 3D scene by imposing different lighting sources and 3D transforms. Figure 4 depicts composited scenarios for both circumstances.

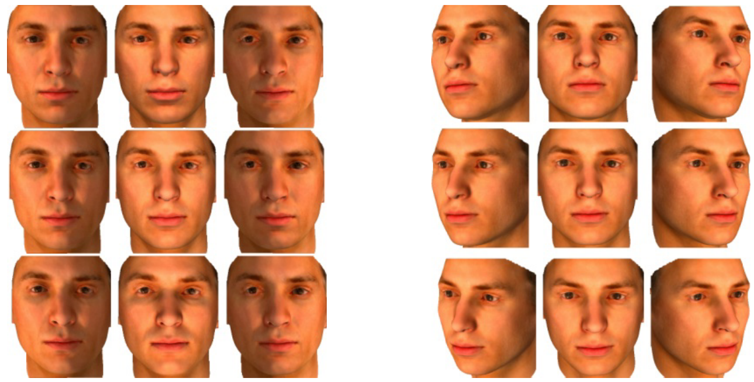


Figure 4. 3D faces rendered under different lighting and pose conditions.

4.3.2 Validation for Synthesized Faces

To validate our algorithm, we apply our model on CMU-PIE database whose image gallery is collected from 68 subjects across 13 poses and under 43 illumination conditions[12]. We select images of 6 subjects which are taken under 3 different lighting conditions and with 5 different pose orientations. We use the frontal image as training data for the 3D model generation while the rest are included in the test dataset for validation. Figure 5 shows images used in our preliminary experiment.

Figure 5. 6 subjects with frontal images for training (top) and one of the pose views for test (bottom).

(Courtesy of CMU-PIE Database)

We utilize Principal Component Analysis (PCA) [14] to evaluate the synthesized images rendered per individual. Each test image is associated with a cluster of 9 synthesized 2D images that are included in the PCA training dataset for recognition purposes. An example of the test image and its corresponding training images are showed in Figure 6.

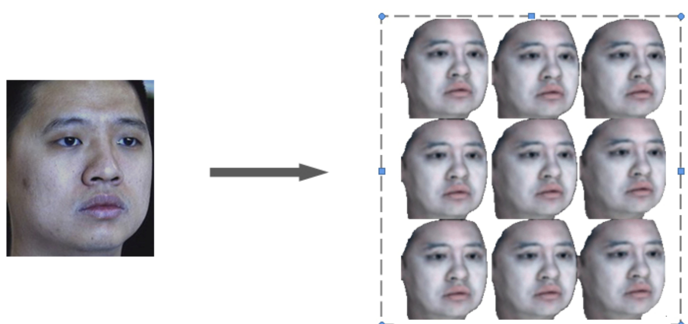


Figure 6. Example test image and corresponding training images. The image in the center of clustered images (right) is imposed with same orientation as the real image (left). The surrounding images (right) are created by minor orientation offsets to the corresponding pose view, which are varied by azimuth and elevation angles ranging from -3 to 3 degrees.

Figure 7 shows the evaluation results from PCA with respect to different training datasets. The top curve (star) is obtained by using test images as training images. The bottom curve (diamond) informs the FR accuracy merely from frontal images training. The curve in the middle (circle) indicates the performance of training images projected by the synthesized 3D model. The comparison between these curves shows that the generated images, to some extent, achieve similarities to the real images. An illustration of the camera, flash lighting positions as well as head positions are plotted in Figure 8.

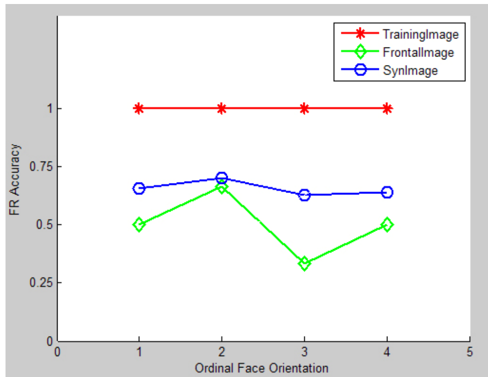


Figure 7. Face orientations v.s. face recognition accuracy w.r.t different sets of training images.

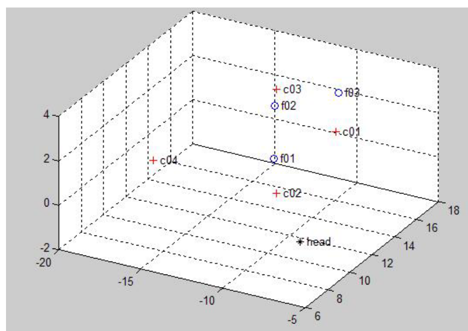


Figure 8. The head position, cameras and flash positions plotted in the 3D Cartesian Coordinate.

(f* stands for flash and c* cameras)

5. CONCLUSIONS

In this paper, we propose a novel framework for 3D face model synthesis, which employs 3D Morphable Model and Differential Evolution. DE is easy to handle and shows strong ability for global optimization.The experimental results show that the new approach is plausible for 3D face model synthesis based on the single 2D image. Future research will be focused on the computation efficiency of DE optimization.

6. ACKNOWLEDGEMENT

The authors would like to express thanks to Dr. T. Vetter, Department of Computer Science, and the University of Basel, for providing the 3D Basel Face Model database for our experiments. The authors also acknowledge the individual whose image is used in our experiments.

REFERENCES

[1] Atick, J. J., Griffin, P. A. and Redlich, A. N., “Statistical approach to shape from shading: reconstruction of three-dimensional face surfaces from single two-dimensional Images,” Neural Computation, 8(6), 1321-1340 (1996).

[2] Georghiades, A. S., Belhumeur, P. N. and Kriegman, D. J., "From few to many: illumination cone models for face recognition under variable lighting and pose," Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23, 643-660 (2001).

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23(6), 681-685 (2001).

[4] Blanz, V. and Vetter, T., "A morphable model for the synthesis of 3D faces," SIGGRAPH '99: Proc. of the 26th annual conference on Computer graphics and interactive techniques, 187-194 (1999).

[5] Parke, F. I., "A parametric model of human faces", PhD thesis, Salt Lake City: University of Utah (1974).

[6] Vetter, T. and Poggio, T., "Linear object classes and image synthesis from a single example image," Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7), 733-742 (1997).

[7] Atick, Joseph J., Griffin, Paul A. and Redlich, A. Norman, "Statistical approach to shape from shading: reconstruction of three-dimensional face surfaces from single two-dimensional images", Neural Computation, 8(6), 1321-1340 (1996).

[8] Paysan, P., Knothe, R., Amberg, B., Romdhani, S., and Vetter, T., "A 3D face model for pose and illumination invariant face recognition," Advanced Video and Signal Based Surveillance, AVSS '09. Sixth IEEE International Conference on, 296-301, (2009).

[9] Storn, R. and Price, K., "Differential Evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces," Technical Report TR-95-012, ICSI, (1995).

[10] Storn, R. and Price, K., "Differential Evolution – a simple and efficient heuristic for global Optimization over continuous spaces", Journal of Global Optimization, 11(4), 341-359 (1997).

[11] Romdhani, S. and Vetter, T., "Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior," Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 986- 993 (2005).

[12] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) database," Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on, 46-51(2002).

[13] M. Turk, and A. Pentland, “Eigenfaces for recognition,” J. Cognitive Neuroscience, 3(1), 71-86 (1991).

[14] http://www.face-rec.org/databases/