Cascaded Regressor based 3D Face Reconstruction from a Single Arbitrary View Image

Feng Liu, Dan Zeng, Jing Li, Qijun Zhao

College of Computer Science, Sichuan University, Chengdu, China

qjzhao@scu.edu.cn

Abstract

State-of-the-art methods reconstruct three-dimensional (3D) face shapes from a single image by fitting 3D face models to input images or by directly learning mapping functions between two-dimensional (2D) images and 3D faces. However, they are often difficult to use in real-world applications due to expensive online optimization or to the requirement of frontal face images. This paper approaches the 3D face reconstruction problem as a regression problem rather than a model fitting problem. Given an input face image along with some pre-defined facial landmarks on it, a series of shape adjustments to the initial 3D face shape are computed through cascaded regressors based on the deviations between the input landmarks and the landmarks obtained from the reconstructed 3D faces. The cascaded regressors are offline learned from a set of 3D faces and their corresponding 2D face images in various views. By treating the landmarks that are invisible in large view angles as missing data, the proposed method can handle arbitrary view face images in a unified way with the same regressors. Experiments on the BFM and Bosphorus databases demonstrate that the proposed method can reconstruct 3D faces from arbitrary view images more efficiently and more accurately than existing methods.
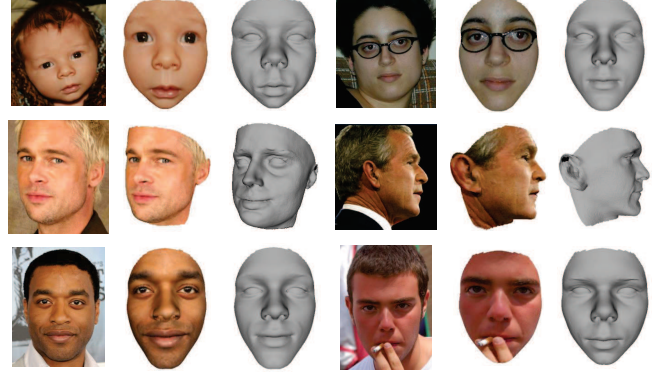
Figure 1. Reconstruction results of our proposed method on some images downloaded from the internet. The three columns show from left to right input images, reconstructed 3D faces with texture, and the 3D face shapes.
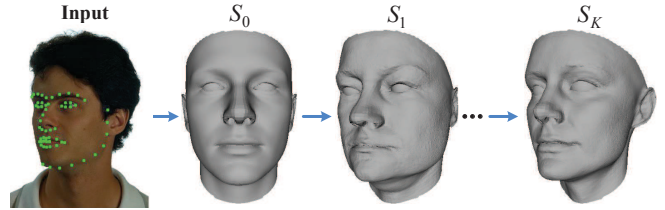

Figure 2. Illustration of the proposed cascaded regressor based 3D face shape reconstruction. The initial 3D face shape S_0 is gradually adjusted until the final shape S_K for the input face image.

1. Introduction

As a fundamental problem in computer vision, reconstructing three dimensional (3D) shapes from two dimensional (2D) images has recently gained increasing attention because of its benefits to many real-world applications, for example, automated face recognition [5, 23]. Using 3D face data to recognize identities is believed to be more robust and more accurate than using only 2D face images [1]. Despite its high recognition accuracy, fast acquisition of high resolution and high precision 3D face data is still difficult, especially under varying conditions or at a distance. On the other hand, 2D face images can be much easier captured with the widespread cameras, and there are already plenty of 2D face image databases. It is thus highly demanded to develop efficient 3D face reconstruction algorithms based on 2D face images such that the rich resources of 2D face images and facilities can be better utilized. This paper focuses on the problem of reconstructing 3D faces from a single 2D face image that can be of an arbitrary view.

Human face is a type of 3D objects consisting of shape and texture components, whereas a 2D face image is a projection of the 3D human face onto a 2D plane. Let $M =$

Methods	Online Optimization	Prior Knowledge	Training Data	Robust to		Real-Time
				Illumination	View	
Shape from X [9, 2, 21, 13, 27, 3]	Yes	Reflectance properties, etc.	No	No	Yes	Yes
3DMM based [4, 5, 17, 12, 24, 15]	Yes	Morphable model	Yes	Yes	Yes	No
Correlation Analysis based [16, 8, 25, 20]	No	None	Yes	No	No	No
Our Proposed	No	None	Yes	Yes	Yes	Yes

Table 1. Comparison of the proposed method and existing methods

(S, T) be a 3D human face, where $S = \{(x_i, y_i, z_i)|i = 1, 2, \cdots, m\}$ and $T = \{t_i|i = 1, 2, \cdots, m\}$ are, respectively, its shape and texture components, m is the total number of points, (x_i, y_i, z_i) the coordinates and t_i the texture value of the i_{th} point in M. I denotes the projected 2D face image of M. $I(u, v)$ is the texture value at the pixel (u, v) in I. The purpose of 3D face reconstruction is to compute $\hat{M} = (\hat{S}, \hat{T})$, an estimate of M, based on the input 2D face image I. This is obviously an ill-posed problem [4].

In order to solve the ill-posed 3D face reconstruction problem, different priors or constraints have been introduced, resulting in the shape-from-X methods and statistical 3D face model based methods. Shape-from-X methods recover 3D shapes via analyzing certain clues in the 2D texture images, say shading [9, 2, 21] or motion information [22, 14]. While classical shape-from-X methods [13] are initially designed for generic 3D shape reconstruction, their performance in recovering 3D face shapes can be further improved by using some reference 3D face models as additional constraints [27, 3]. These methods usually have limited accuracy because (i) their assumed connection between 2D texture clues and 3D shape information is too weak to discriminate between different human faces and (ii) they do not fully exploit the prior knowledge of 3D faces.

3D Morphable Model (3DMM) [4, 17, 12, 24], as a typical statistical 3D face model, explicitly learns the prior knowledge of 3D faces in a statistical analysis way. It represents a 3D face as a linear combination of basis 3D faces, which are obtained by applying principal component analysis (PCA) on a set of densely aligned 3D faces. The 3D face reconstruction problem is treated as a model fitting problem, in which the model parameters (i.e., linear combination coefficients and camera parameters) are optimized so that the resulted 2D projection of the 3D face is best consistent with the input 2D face image in terms of the locations (and textures) of a set of annotated facial landmarks [15] (e.g., eye centers, mouth corners, and nose tip). 3DMM-based methods usually require online optimization and are thus computationally intensive [5]. The reconstruction accuracy is

highly affected by the used landmarks because the reconstruction cost is completely determined by the landmarks.

Correlation analysis based methods [16, 8, 20] have recently been proposed to establish mapping between 3D faces and 2D face images based on a training set of pairing 3D faces and 2D face images. They first learn respective transformations for 3D faces and 2D face images to transform them into a common feature space, in which the 3D faces and 2D face images in pairs have maximum correlation/covariance. They then construct a mapping between 3D faces and 2D face images within the common feature space. These methods are generally more computationally efficient than 3DMM-based methods because no online optimization is needed. However, existing correlation analysis based methods learn the mapping functions based on the texture features [20]. As a result, they are sensitive to illumination changes in 2D face images. Moreover, they consider only frontal view face images, but the mapping function learned for frontal view is usually not suitable for other different views.

This paper aims to develop a real-time 3D face reconstruction algorithm that can cope with arbitrary view face images. To this end, we start from an initial 3D face shape, and gradually adjust it according to some annotated landmarks in the input 2D face image and a series of regressors that are off-line learned. The set of used landmarks is consistent for all different views, and whenever a landmark is not visible due to large view angles, it is marked as missing data. In this way, the proposed cascaded regressor based method avoids complicated online optimization, and can effectively handle arbitrary view face images in a unified way. During the off-line training of the regressors, the optimization objective is to minimize the total error between the estimated 3D face shape $\hat{S}$ and the true 3D face shape S. This ensures a higher reconstruction accuracy than the 3DMM-based methods which consider only the fitness of the landmarks instead of the whole 3D face. As we know, cascaded regressor-based methods have been very widely used to address 2D face alignment [26, 6] and

3D face alignment [7, 10], because its intrinsic accuracy and efficiency. However, we are the first to solve the problem of reconstructing 3D face shapes from single arbitrary view images as a regression problem in the shape space. Table 1 highlights the differences between our proposed method and existing methods. See Fig. 1 for example results of our proposed method on some photos found on the internet.

The rest of this paper is organized as follows. Section 2 discusses the relationship of the proposed method with existing methods and the motivation of this work. Sections 3 and 4 present in detail our proposed method and its implementation. Section 5 reports the experimental results. Section 6 finally concludes the paper.

2. Related Work and Motivation

As mentioned above, the formation of a 2D face image is determined not only by the source 3D face but also by the camera parameters and imaging conditions. Given a 2D face image, however, its camera parameters or imaging conditions are usually unknown. To recover the source 3D face from the 2D face image, previous 3DMM-based methods iteratively seek for a solution in a parametric space that is defined by the statistical 3D face model and the imaging model. It starts from a set of initial parameter values, and iteratively updates them according to the cost in terms of some annotated facial landmarks. The reconstructed 3D face is obtained by the final optimal parameter values. Instead of searching for solutions in the parametric space, correlation analysis based methods exploit the correlation between 3D faces and 2D face images in transformed feature spaces, and learn mapping functions from 2D face features to 3D face features in the feature spaces. To summarize, these existing methods all involve an extra step to transform 3D faces to some latent space. This makes the methods more complicated and prone to error.

We thus propose to reconstruct 3D faces directly in the 3D face shape space. This is fulfilled also in an iterative way by gradually adjusting the 3D face shape with some pre-trained regressors (see Fig. 2). One key problem in our approach is how to compute the shape adjustment. We solve this by establishing (linear) regressions between the landmark displacement and the shape adjustment. These regressors are off-line trained to minimize the reconstruction error of the entire 3D face shapes. We use landmark displacement rather than texture difference for the regressors because landmarks are more directly related to shape information and they are more robust to illumination changes.

Another problem arises with landmarks when we deal with arbitrary view face images, i.e., some of the landmarks might be invisible in some view angles (see Fig. 3). Some previous work [11] proposed multi-view methods to process different views separately. Others introduced additional label for each landmark to indicate whether it is visible or

not [15]. Instead, in this paper, we treat the invisible landmarks as missing data, and use the same regression functions for all different views. This way, we can reconstruct 3D face shapes from arbitrary view images in a unified way.

3. The Proposed Approach

3.1. The Reconstruction Process

Given a 2D face image, we assume that a set of facial landmarks have been annotated on it. Let $P = (u_1, v_1, u_2, v_2, \cdots, u_l, v_l)^t$ ('t' denotes the transpose operator) be the coordinate vector of the landmarks, (u_i, v_i) the coordinates of the i_{th} landmark on the 2D image, and l the total number of landmarks. For the landmarks that are invisible in the image, their coordinates are set to be a constant value (e.g., '0' in our experiments), which means missing data.

To reconstruct its 3D face shape, the initialization step first picks a shape S_0 as the initial value. In order to compute the adjustment to S_{k-1} at the k_{th} iteration, S_{k-1} is projected onto 2D plane by a pre-defined transform F, which maps a point from 3D space to 2D plane. The projections of the landmarks in S_{k-1} corresponding to the annotated landmarks on the 2D image are computed as follows,

$$\hat{P}_{k-1} = F(h(S_{k-1})), \tag{1}$$

where $h(\cdot)$ is an operator which extracts the landmarks in the 3D face shape [1]. The landmark displacement is then defined as

$$\Delta P_{k-1} = P - \hat{P}_{k-1}. \tag{2}$$

The shape adjustment can be computed via a regression over this landmark displacement, i.e.,

$$\Delta S_{k-1} = f_k \left(\Delta P_{k-1} \right), \tag{3}$$

where $f_k(\cdot)$ is the regression function. If a linear regressor is employed, Eqn. (3) can be simplified as

$$\Delta S_{k-1} = R_k \times \Delta P_{k-1}, \tag{4}$$

where $R_k \in \mathcal{R}^{(3m) \times (2l)}$ serves as the regressor for the k_{th} iteration. The reconstructed 3D face shape can be then updated by

$$S_k = S_{k-1} + \Delta S_{k-1}. \tag{5}$$

After applying all the K cascaded regressors, the reconstructed 3D face shape is obtained, $\hat{S} = S_K$. For the texture component, the texture values of the pixels on the input 2D face image are directly assigned to their corresponding vertices in the 3D face shape.

---

[1]In this work, we assume that all the 3D face shapes have been densely aligned. Hence, the same landmarks on all the 3D faces share the same index in their point clouds. The operator $h(\cdot)$ simply returns the corresponding vextices as the landmarks.
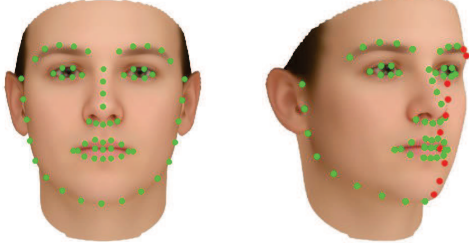
Figure 3. 68 facial landmarks are used in the proposed method. Some of them (e.g. red points on the right-side face image) might become invisible due to self-occlusion in non-frontal views.

3.2. The Training Phase

To apply the above 3D face reconstruction process, we should have the K regressors. They are obtained via an off-line training on a set of N training samples. Each training sample consists of a true 3D face shape S^i, a 2D face image I^i of the 3D face, and its facial landmarks P^i ($i \in \{1, 2, \cdots, N\}$). Based on these training samples, we learn the regressors $\{R_1, R_2, \cdots, R_K\}$ one by one.

Suppose the reconstructed 3D face shapes at the k_{th} iteration are $\{S_k^i | i = 1, 2, \cdots, N\}$. The landmark displacement is computed for each training sample according to Eqns. (1) and (2). The following objective function is optimized over all the training samples to obtain the k_{th} regressor R_k:

$$\underset{R_k}{\arg \min} \sum_{i=1}^{N} \| \left(S^i - S_{k-1}^i \right) - R_k \, \Delta P_{k-1}^i \|_2^2 . \quad (6)$$

Obviously, this optimization looks for a regressor which can minimize the total error of the reconstructed 3D face shapes, but not just the error at landmarks. This optimization can be effectively solved by using least squares methods [26]. Algorithm 1 summarizes the training and reconstruction steps in our proposed method.

4. Implementation Details

4.1. Training Data

A number of densely aligned 3D faces are needed to train the regressors in the proposed method. In this work, we used the Basel Face Model (BFM) [18] to construct 120 synthetic 3D faces (50% female). From each of them, 2D face images were generated at 21 different views with yaw rotations of ($0°$, $\pm10°$, $\pm15°$, $\pm20°$, $\pm30°$, $\pm40°$, $\pm50°$, $\pm60°$, $\pm70°$, $\pm80°$ and $\pm90°$). The resolution of the projected 2D face images was 1024×768 pixels. On each 2D face image, 68 facial landmarks (see Fig. 3) were recorded, whose coordinates were computed according to the source 3D face and the projection function used to generate the 2D face image.

Algorithm 1 Cascaded Regressor based 3D Face Reconstruction:

I. Training phase

  Input: A set of 3D face shape $\{S^i\}_{i=1}^N$
        Landmarks on the corresponding images $\{P^i\}_{i=1}^N$
        Initial shape S_0
  Output: Regressors $\{R_k\}_{k=1}^K$
    $k = 1$
    while $k < K$
      1: Compute 2D landmarks $\hat{P}_k$ from current 3D shape S_k
      2: Learn R_k using Eqn. (6)
      3: Update estimated 3D shape using Eqn. (5)
      4: $k = k + 1$
    end while

II. Reconstruction phase

  Input: Face image I and its landmarks P
        Regressors $\{R^k\}_{k=1}^K$
        Initial shape S_0
  Output: Reconstructed 3D face shape S_K
    $k = 1$
    while $k < K$
      1: Compute the 3D face shape increment using Eqn. (3)
      2: Update 3D face shape using Eqn. (5)
      3: $k = k + 1$
    end while

4.2. Initialization

Initialization of the proposed method includes the selection of initial 3D face shape (S_0) and the setting of 3D to 2D mapping (F). In this work, we used the average 3D face shape of the training samples as the initial 3D face shape. But this is not necessary, because the proposed method updates the reconstructed 3D face shape based on the landmark displacement, rather than on the shape itself. We will show in the experiments that our proposed method can achieve good results with different initial 3D face shapes.

As for the 3D to 2D mapping, we decompose it into two parts: a rigid transformation to the 3D face, and a weak perspective projection to map the transformed 3D face to 2D plane. The parameters involved in the mapping were estimated based on the average 3D faces and the average 2D landmarks in the training data such that they can well correspond via the mapping. Once this mapping function is obtained, it is used for all data in both training and testing.

5. Experimental Results

5.1. Databases and Protocols

To evaluate the performance of the proposed method, we use two datasets, namely BFM [18] and Bosphorus [2]. The BFM dataset contains ten 3D face scans with nine different poses/views, including one frontal view and eight other views heading ±15, ±30, ±50 and ±70 degrees around Y axis. The Bosphorus dataset has the 3D faces of 105 subjects. In our experiments, we consider their frontal and left/right profile views.

The proposed method is compared with several state-of-the-art methods, including the improved Shape-from-Shading method (SFS) [13], MFF [19] and SSF [28]. The latter two methods both are based on 3DMM. They differ from each other in the features used for evaluating the model fitting cost.

Two metrics are used to assess the reconstruction accuracy of different methods, i.e., Rooted Mean Square Error (RMSE) and Per-pixel Depth Error Map (PDEM). The RMSE between a true 3D face shape S and its reconstructed one $\hat{S}$ is defined as [11]

$$E = \frac{1}{m} \sum_{i=1}^{m} \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (z_i - \hat{z}_i)^2}. \quad (7)$$

Their PDEM is computed by

$$E_z(x_i, y_i) = |\hat{z}_i(x_i, y_i) - z_i(x_i, y_i)|/z_i(x_i, y_i). \quad (8)$$

5.2. Quantitative Evaluation

Figure 4 shows the average RMSE of different methods on the BFM dataset with respect to different views of input face images. As can be seen, the average RMSE of our proposed method is obviously lower than that of the counterpart methods. More importantly, its accuracy is very stable across different views. This proves the effectiveness of the proposed method in handling arbitrary view face images. Figure 5 shows the reconstruction results of our method and the SSF method on one subject in the BFM dataset. The PDEMs of these two methods clearly show the superiority of the proposed method.

The average depth error for each of the 105 Borphorus subjects is shown in Fig. 6. By comparing the proposed method to the SFS method, we can see that the proposed method successfully reduces the average error from 6.69% to 2.24% for frontal faces. Note that for profile faces, because the ground truth data are partial faces, we compute the error for the available landmarks only, and our method achieves average error of 2.95% and 2.91% for left and right profile faces, respectively.
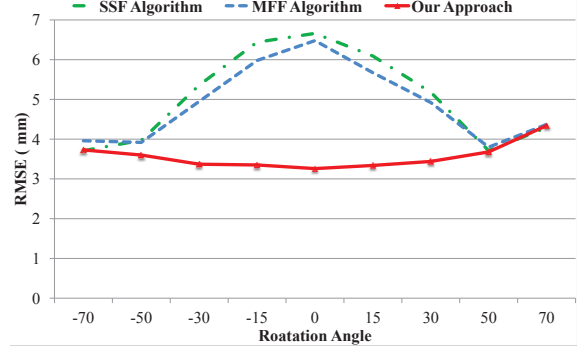
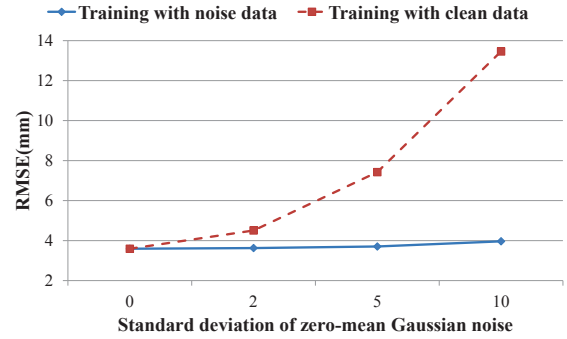Figure 4. Average RMSE of different methods under different view angles on the BFM dataset.

Figure 7. The average RMSE over the BFM dataset of our method when different noisy data are used for testing.

5.3. Results on Noisy Data

The proposed method requires facial landmarks as the input. In the above evaluation experiments, the ground truth or manually marked landmarks are used. In real-world applications, when landmarks are automatically detected, there will be always some errors in the detected landmarks' locations. In this experiment, we assess the impact of noisy landmarks on the 3D face reconstruction accuracy. Specifically, we disturb the ground truth landmarks (called clean landmarks) with some additive zero-mean Gaussian noise, resulting in noisy landmarks. We then conduct two series of experiments: (i) training using data with clean landmarks, and testing using data with noisy landmarks, and (ii) training and testing both using data with noisy landmarks. Figure 7 presents the RMSE of the proposed method under noise of different standard deviations (the larger the noise's standard deviation, the stronger the noise is. As shown by the results, if the regressors are trained with clean data, they will be very sensitive to the noise in landmarks. To improve the robustness of the regressors to noisy data, it is highly recommended to augment the training data by adding some noise.
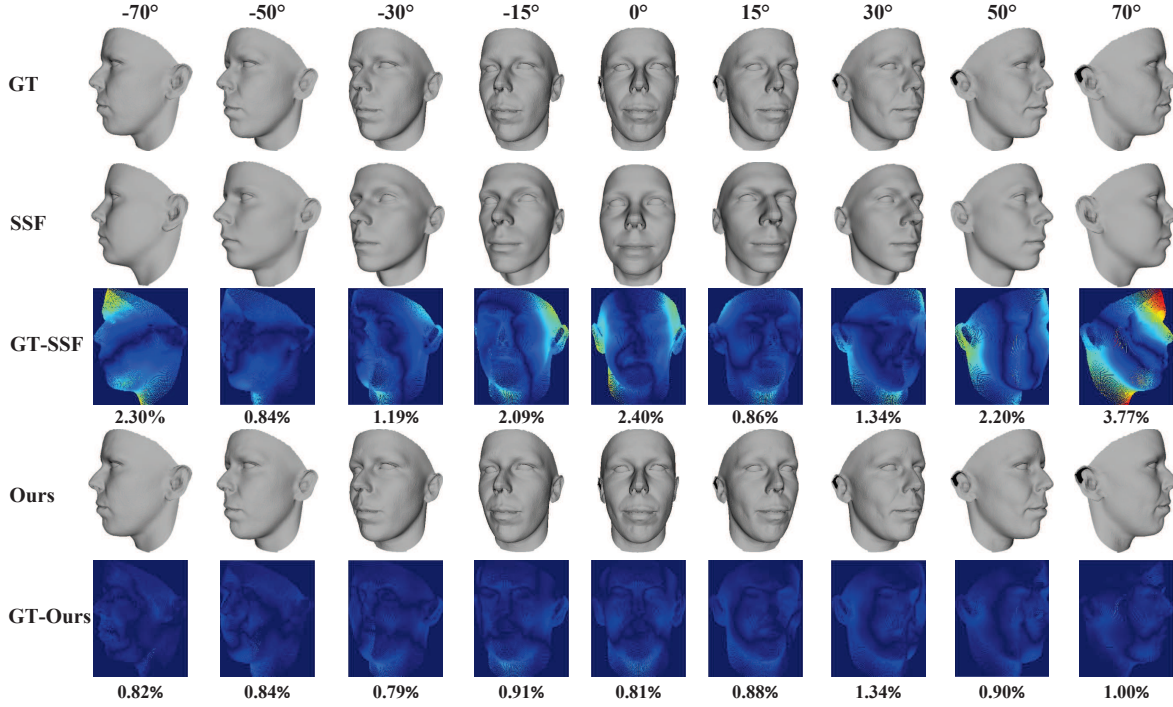
Figure 5. Reconstruction results of different methods for one BFM sample at 9 different views. First row: The ground truth 3D face shapes. Second and forth rows: The reconstructed 3D face shapes by the SSF method [28] and our proposed method. Third and fifth rows: Their corresponding PDEMs. Brighter values in PDEMs indicate larger errors.
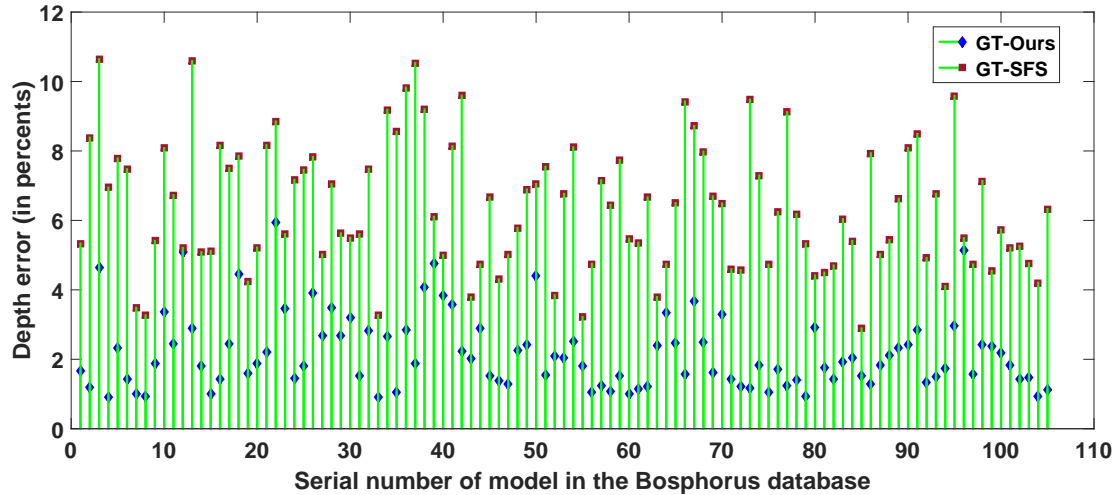

Figure 6. Mean depth error obtained by the SFS [13] and our proposed methods for each of the 105 subjects in the Bosphorus database. The overall average error of our method is 2.24%, while that of SFS is 6.69%.

5.4. Training Data Diversity

As a data-driven method, the obtained regressors' generalization ability is likely to be affected by the diversity or completeness of the training data. In the following ex-periments, we will evaluate the proposed method from two aspects: (i) generalization to view angles that are not seen in the training data, and (ii) generalization to novel subjects, particularly subjects of different ages.
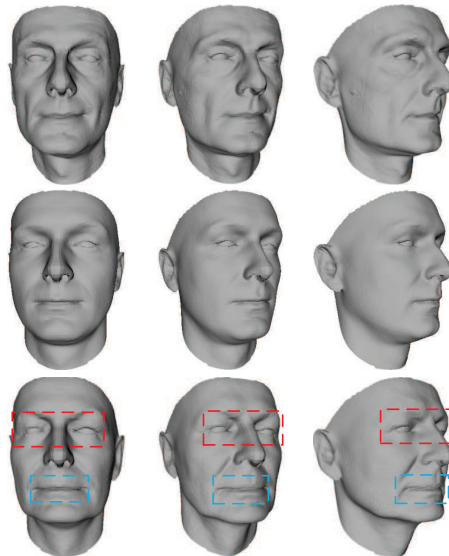
Figure 9. Evaluation of generalization ability to novel subjects of different ages. First row: The ground-truth 3D face shapes. Second row: The reconstruction results when training data do not include subjects of similar ages. Third row: The reconstruction results when subjects of similar ages are used for training.

Figure 8. Reconstruction results for two BFM subjects under novel view angles. Row 1 and Row 4 show the input images at different views. Row 2 and Row 5 show the 3D faces reconstructed by our method. Row 3 and Row 6 are the corresponding error maps.

Generalization to Novel Views. In the above experiments, we trained the regressors with data of 21 different views, i.e., $(0°, ±10°, ±15°, ±20°, ±30°, ±40°, ±50°, ±60°, ±70°, ±80°$ and $±90°)$. In this experiment, we apply them to face images of view angles $(−85°, −55°, −35°, 25°$ and $75°)$, which are not included in the training data. The results are shown in Fig. 8. The reconstructed faces are visually plausible, and the corresponding PDEMs also show low reconstruction error. The average RMSE across the 10 subjects in the BFM dataset is 3.59 mm (milimeter) and 3.65 mm for testing data of seen views and novel views, respectively. This proves that the trained regressors can be well generalized to novel views if they are trained with sufficient data of well-sampled view angles.

Generalization to Novel Subjects. The training data used in the above experiments contain subjects aging between 20 and 40 years. To evaluate the generalization ability of the obtained regressors to novel subjects, we apply them to a face image of a subject whose age is about 70 years. The results are shown in the second row in Fig. 9. It is not surprising that the reconstructed 3D faces look very different from the true ones (shown in the first row in Fig. 9). In order to see the impact of training data, we randomly replace 10 of the 120 training subjects with another 10 subjects whose ages are more close to the testing subject. This time, the reconstructed 3D faces become much better as shown in the third row in Fig. 9. The newly trained regressors can achieve a more than 25% improvement in terms of RMSE. These results suggest the importance of training data. To simply put, for better generalization ability of the regressors, more diverse training data should be used, for example, data of subjects from different ethnicities and a wide range of ages.

5.5. Computational Efficiency

The proposed method can reconstruct 3D faces in real time. During training, according to our experiments, it usually converges within three to five iterations. This means that only three to five regressors are needed. During testing, it takes only about 43 ms (milisecond) on average to reconstruct one 3D face [3]. As a comparison, the C-RBF method in [20] needs about 590 ms to reconstruct one 3D face according to the authors' results.

6. Conclusion and future work

This paper has for the first time proposed a regression based approach to the reconstruction of 3D faces from sin-

---

[3]All experiments in this paper were conducted using MATLAB on a $64bit$ windows workstation with Intel $i3$ CPU and $4G$ memory.

gle arbitrary view images. It directly searches in the shape space for the optimal 3D face shape for an input 2D face image. The searching process is done by gradually adjusting the 3D face shape with a series of regressors, which compute the shape adjustment based on the landmark displacement. These cascaded regressors are off-line learned from a set of training data, which include pairs of 3D faces and 2D face images of different views with annotated landmarks. Compared with previous 3D face reconstruction methods, the proposed method has the following advantages: (i) it does not require on-line optimization, but directly does linear regressions over the 3D face shape, and hence has high computational efficiency, (ii) it better exploits 3D face shape prior by explicitly optimizing the total 3D face shape reconstruction error, and thus can more accurately recover 3D faces, and (iii) it treats the invisible landmarks as missing data, and provides a unified and efficient way to handle arbitrary view images. Experiments on the BFM and Bosphorus databases have demonstrated the superiority of the proposed method over existing state-of-the-art methods.

In this paper, we assume that landmarks on 2D face images are available; and in the experiments, manually annotated landmarks are used. In our future work, we are going to couple landmark localization with 3D face reconstruction so that they can be automatically done simultaneously. Besides, we will extend the proposed 3D face reconstruction method to cope with a set of 2D face images of multiple views.

References

[1] R. Abiantun, U. Prabhu, and M. Savvides. Sparse feature extraction for pose-tolerant face recognition. *IEEE PAMI*, 36(10):2061–2073, 2014.

[2] J. T. Barron and J. Malik. Shape, albedo, and illumination from a single image of an unknown object. In *CVPR*, pages 334–341, 2012.

[3] S. Biswas, G. Aggarwal, and R. Chellappa. Robust estimation of albedo for illumination-invariant matching and shape recovery. *IEEE PAMI*, 31(5):884–899, 2009.

[4] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proc. SIGGRAPH*, pages 187–194, 1999.

[5] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE PAMI*, 25(9):1063–1074, 2003.

[6] X. P. Burgos-Artizzu, P. Perona, and P. Dollár. Robust face landmark estimation under occlusion. In *ICCV*, pages 1513–1520, 2013.

[7] C. Cao, Y. Weng, S. Lin, and K. Zhou. 3d shape regression for real-time facial animation. *ACM Transactions on Graphics (TOG)*, 32(4):41, 2013.

[8] M. Castelán and J. Van Horebeek. 3d face shape approximation from intensities using partial least squares. In *CVPR Workshops*, pages 1–8, 2008.

[9] B. K. Horn and M. J. Brooks. *Shape from shading*. MIT press, 1989.

[10] L. A. Jeni, J. F. Cohn, and T. Kanade. Dense 3d face alignment from 2d videos in real-time. In *FG*, 2015.

[11] J. Jo, H. Choi, I.-J. Kim, and J. Kim. Single-view-based 3d facial reconstruction method robust against pose variations. *Pattern Recognition*, 48(1):73–85, 2015.

[12] I. Kemelmacher-Shlizerman. Internet based morphable model. In *ICCV*, pages 3256–3263, 2013.

[13] I. Kemelmacher-Shlizerman and R. Basri. 3d face reconstruction from a single image using a single reference face shape. *IEEE PAMI*, 33(2):394–405, 2011.

[14] S. J. Lee, K. R. Park, and J. Kim. A sfm-based 3d face reconstruction method robust to self-occlusion by using a shape conversion matrix. *Pattern recognition*, 44(7):1470–1486, 2011.

[15] Y. J. Lee, S. J. Lee, K. R. Park, J. Jo, and J. Kim. Single view-based 3d face reconstruction robust to self-occlusion. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–20, 2012.

[16] Z. Lei, Q. Bai, R. He, and S. Z. Li. Face shape recovery from a single image using cca mapping between tensor spaces. In *CVPR*, pages 1–7, 2008.

[17] A. Patel and W. A. Smith. Simplification of 3d morphable models. In *ICCV*, pages 271–278, 2011.

[18] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3d face model for pose and illumination invariant face recognition. In *IEEE AVSS*, pages 296–301, 2009.

[19] S. Romdhani and T. Vetter. Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *CVPR*, volume 2, pages 986–993, 2005.

[20] M. Song, D. Tao, X. Huang, C. Chen, and J. Bu. Three-dimensional face reconstruction from a single image by a coupled rbf network. *IEEE IP*, 21(5):2887–2897, 2012.

[21] S. Suwajanakorn, I. Kemelmacher-Shlizerman, and S. M. Seitz. Total moving face reconstruction. In *ECCV*, pages 796–812. 2014.

[22] S. Ullman. The interpretation of structure from motion. *Proc. of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426, 1979.

[23] K. Wang, X. Wang, Z. Pan, and K. Liu. A two-stage framework for 3d facereconstruction from rgbd images. *IEEE PAMI*, 36(8):1493–1504, 2014.

[24] L. Wang, B. Liu, S. Su, Y. Cheng, and S. Li. An improved 3d bilinear multidimensional morphable models used in 3d face recognition. In *ISEEE*, volume 3, pages 2052–2056, 2014.

[25] Q. Xiao, L. Han, and P. Liu. 3d face reconstruction via feature point depth estimation and shape deformation. In *ICPR*, pages 2257–2262, 2014.

[26] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, pages 532–539, 2013.

[27] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE PAMI*, 21(8):690–706, 1999.

[28] X. Zhu, D. Yi, Z. Lei, and S. Z. Li. Robust 3d morphable model fitting by sparse sift flow. In *ICPR*, pages 4044–4049, 2014.