

Jiang Du · Dan Song · Yanlong Tang · Ruofeng Tong · Min Tang

“Edutainment 2017” a visual and semantic representation of 3D face model for reshaping face in images

Received: 15 September 2017 / Revised: 20 December 2017 / Accepted: 22 January 2018 / Published online: 16 February 2018
© The Visualization Society of Japan 2018

Abstract Reshaping faces is an interesting thing and a common demand in today’s media industry and cosmetic industry. Especially, the reshaped face images contribute a lot to special effects industry and plastic-beauty industry. In this paper, we provide a novel parametric representation of 3D face model that describes the face shape as a linear combination of semantic and non-semantic bases. The semantic (local) bases correspond to each individual face part and they are used to semantically edit each face part, while the non-semantic bases explain other shape variations to span the subspace of the face shape. First, we build a sparse and spatially localized parametric face model from a dataset of 3D face models by sparse principal component analysis. Second, to define the semantic bases, we train a regression model to correlate semantically significant values like nose height and mouth width. Finally, the bases of the resulting parametric face model is orthogonalized to all defined semantic bases by the Gram–Schmidt algorithm for generating the novel parametric representation of 3D face model. The novel representation can be applied in reshaping face in the images. The experiment results demonstrate that our representation not only retains the accuracy for 3D face reconstruction but also provides users a user-friendly tool to edit facial parts for desired facial shapes.

Keywords Image manipulation · Image editing · Semantic parameters · Face image reshaping · Deformation

1 Introduction

The editing of human face shape has always been one of the most important components in modeling and designing face model and plays a great role in many applications ranging from medical service and digital forensics to digital education and entertainment. For example, in medical beauty, it is helpful to provide a vivid preview of the client’s edited ideal face before plastic surgery (Chou et al. 2012). In forensics, it is significant to compose the suspect’s face according to the description of victims and witnesses (Klum et al. 2013). In early childhood education, when parents tell their children about the fable like the story of Pinocchio, it is vivid to show that the liar’s nose gets longer. In entertainment, it is interesting to show personal or friends’ edited funny face (Nirkin et al. 2017).

In recent years, there existed some methods to help users to edit the face shape at least to some extent. These existing methods can be roughly classified into two groups: 2D-image-based and 3D-face-model-based methods.

The 2D-image-based methods derive from image-processing techniques and image analyses (Suryanarayana and Dubey 2017), like image morphing, seamless cloning and face detection in images. For example, Chou et al. (2012) proposed a swapping method to transfer face parts between the paired images. Fan et al. (2016) proposed an image-morphing-based method to beautify the face and Zhang et al. (2014) proposed an image-warping-based method to clone face expression. Because such methods only edit face in the 2D image, they can often deal with frontal faces in the image.

The 3D face-model-based methods utilize the 3D face shape estimated from the image and edit the face on the 3D model. Blanz and Vetter (1999) first used a 3D morphable face model which is a parametric representation of 3D face to edit the face shape. Afterwards, the multilinear face model (Vlasic et al. 2005) and bilinear face model (Cao et al. 2014) were proposed to edit the face shape. For example, Li et al. (2013) and Thies et al. (2016) used the 3D face model to reanimate the face expression. The 3D-model-based methods can introduce more information than the 2D-images-based methods such as a lot of work using 3D models (Alvarez et al. 2017). However, these 3D face models are strongly limited because of the global control of their parameters, which cannot control on plausible local effects in an intuitive manner (Neumann et al. 2013).

Some commercial software such as Adobe Photoshop, Autodesk 3D Max and Maya can also edit the face shape on 2D or 3D condition. However, these professional editing tools are mainly designed for low-level editing tasks and demand professional skills (Zhou et al. 2010). Moreover, using these tools to do the editing task is also a time-consuming process and still a tedious task even for skillful users.

A possible solution for intuitively editing the face shape is to operate at the level of the face parts, and change each part with respect to semantical manner such as making nose high or eyes big. However, it is challenging due to the following reasons. First, a realistic reshaping effect demands a spatially varying deformation within individual face parts. Second, it is unclear how to make the changes introduced to individual parts globally coherent.

To solve these problems, in this paper, we provide a novel parametric representation of 3D face model that describes the face shape as a linear combination of semantic and non-semantic bases. The semantic (local) bases correspond to each face part and are used to semantically edit each face part, while the non-semantic bases explain other shape variations to span the subspace of the face shape. More specifically, inspired by Neumann et al. (2013) who built a sparse and spatially localized model from an animated mesh sequence, we first build a sparse and spatially localized parametric face model in which shape variations are learned from a dataset of 3D face models by sparse principal component analysis (SPCA). Second, to define the semantic bases from the resulting parametric face model, similar to Blanz and Vetter (1999), we train a regression model to correlate the resulting shape variations to semantically significant values like nose height and mouth width. Finally, the Gram-Schmidt algorithm (Golub and Vanloan 1996) is used to orthogonalize the bases of the resulting parametric face model to all defined semantic bases, which generates the novel parametric representation of 3D face model.

Semantically editing the face shape can be easily completed by manipulating the coefficients of the semantic bases. Moreover, it is also convenient to reshape the face in the image relying on our novel parametric representation of 3D face model. The 3D face shape can be recovered from the image by the proposed representation and edited by changing the semantic coefficients. Then, the edited 3D face is rendered into a new image with all relevant scene parameters estimated from the image including face pose, face texture, albedo and illumination, similar to Blanz et al. (2004).

The main contributions of this paper are summarized in two aspects. First, we provide a novel semantic parameterized face model in which the semantic bases correspond to the semantically meaningful deformation modes of each face part. Second, relying on the proposed face model, we provide a framework to intuitively and semantically edit the face shape in the image.

2 Related work

In this section, we review related work on the editing of the face shape following the categories mentioned before: 2D-image-based- and 3D-face-model-based methods. We first discuss the 2D-image-based methods including image-warping-based- and image-compositing-based methods. We second discuss the 3D-face-model-based methods in which many 3D parametric face models are used to edit the face shape. We then discuss its appealing application called face reanimating which also motivates us to focus on the editing of the face shape.

2.1 Image warping

Many existed image warping techniques allow users to interactively manipulate the image. Relying on image-warping methods, Leyvand et al. (2008) optimize the position of the facial landmarks to guide the warping process for enhancing the attractiveness of faces in frontal images. Then, Fan et al. (2016) proposed a similar method on mobile devices which focused on smooth transition between the warped shape and background in the images. In fact, to some extent, all image-warping methods can edit the face shape even if they do not aim at the face (Shamir and Sorkine 2009). However, they need the time-consuming editing processes which is not parametric. Moreover, the editing of the 2D face shape is often lamented to its lack of one-dimensional face information (Chalas et al. 2017).

2.2 Image compositing

It is common to synthesize the face image by composing the face parts according to the description of the victims and witnesses in the digital forensics (Klum et al. 2013). For the requirement of cosmetic plastic surgeries, Chou et al. (2012) proposed an alternation method of face parts between the paired images. The selected face parts from the source image could be transferred to the corresponding region by Poisson image editing. However, unlike Kemelmacher-Shlizerman (2016) who used the filtering mechanism to choose the candidate images as similar face viewpoint as the input image, Chou’s method can only handle the frontal face in images. Afterwards, Chalas et al. (2017) proposed a similar compositing method on the 3D mesh to generate 3D face. Though 3D compositing face has enough geometric information, the compositing methods are limited in the fixed shape of the selected face part which cannot be edited dynamically. Thus, these compositing methods cannot be suitable for the applications like making eyes bigger or nose higher.

2.3 3D-parametric face model

In the past decades, many 3D facial models have been proposed for various purposes. The most widely used model is the linear model based on principal components analysis (PCA). Researchers usually fit the PCA shape model to the face in the image for estimating the 3D face shape (Paysan et al. 2009 and Li et al. 2013). However, these models do not provide the semantic controls. Besides the morphable face model, Blanz and Vetter (1999) pioneered the investigation of relationship between attributes and face shapes. They construct 1D vectors describing the variation along attribute direction, such as expressions, weight and gender. Nevertheless, this method still contains correlation among face parts and could not perform local deformation when we have face-editing tasks like “big eyes”, “high nose” and “wide mouth”. Unlike Streuber et al. (2016) who learned a mapping between semantic attributes and the face shape, Liao et al. (2012) relied on the Laplacian deformation method to edit the 3D face shape for enhancing the symmetry and proportion of face and changing weight of the face which still lack the support of the local editing.

2.4 Face reanimating

Recently, face reanimating is an appealing application in computer graphics and vision. Blanz et al. (2003) first proposed a face-reanimating method based on the 3D-parametric face model to edit the face expression in images and videos. Afterward, this method was extended into fully automatic techniques for face reenactment or digital avatar performance in video (Thies et al. 2016 and Li et al. 2013). These methods are also called face expression transfer through transferring the expression parameters between the 3D parametric face models. When transferring the expression between images, unlike Vlasic et al. (2005) who directly transferred the parameters of source face image to the target image, Thies et al. (2016) used a subspace deformation transfer technique to preserve personalized information in each actor’s expression. Although the editing of the face expression has been paid a lot of attention, it is quite different from the problem what we focus on in this paper. Our proposed face reshaping method can result in the change of personal identity, while the expression editing needs to preserve the user’s identity (Blanz and Vetter 1999; Thies et al. 2016). Inspired by the recent research that studied the editing of local expression (Lewis et al. 2014), we find that there are a lot of potential applications that require the local editing of the face shape.

3 The semantic parametric 3D face representation

In this section, we describe how to find a group of basis vectors whose linear combinations could faithfully represent 3D faces. Moreover, some basis vectors can correspond to the semantical deformation trend. Similar to all data-driven methods, before a face shape model can be learned, it is necessary to produce a training dataset which contains all information what we need such as semantic parameters. Then, we begin with the dataset to learn the shape variations by SPCA. Finally, the semantic bases can be defined by training a regression model to correlate the shape variations to semantically significant values.

3.1 Dataset preparation

There are many excellent 3D face databases for various purposes such as Blanz’s 3D Morphable Model (Blanz and Vetter 1999) and Cao’s FaceWareHouse (Cao et al. 2014). However, they are not suitable for our purpose due to the lack of semantic parameters that describe the attributes of face like nose height or mouth width. Thus, we establish a dataset by reconstructing 3D face models from Chicago Face Database (Ma et al. 2015) used for psychological studies.

We adopt an off-the-shelf method (Tran et al. 2016) to reconstruct the 3D face models. So far, it is the only method which can reconstruct invariant and accurate 3D face shapes (Nirkin et al. 2017). Then, these reconstructed models compose our dataset including 200 face models (100 males and 100 females).

Because all face meshes shared the same topology, we trim the face mesh to remove the part which we do not care. Each face mesh f containing $N(39,503)$ vertices $\mathbf{v}_i^{(f)}$ is assembled a single “sample matrix” $\mathbf{X} \in \mathbb{R}^{3N \times F}$ by stacking the vertices in a column-wise fashion. F denotes the number of face meshes and $\mathbf{v}_i^{(f)}$ contains the x, y, z -coordinates of i th vertex of f th facial mesh.

Similar to the preparation of other matrix analysis method for convenience, we assume the vertex coordinates in the sample matrix are expressed as the residual displacements to the mean shape $\bar{\mathbf{x}}$ of the facial meshes. We also assume that rigid alignment through global translation and rotation is performed before assembling the sample matrix. Formally, from now on, we assume $\mathbf{X} \leftarrow \mathbf{X} - \bar{\mathbf{x}}$ and accordingly $\mathbf{v}_i^{(f)} \leftarrow \mathbf{v}_i^{(f)} - \bar{\mathbf{v}}$. The final sample matrix is shown as follows:

$$\mathbf{X} = \begin{bmatrix} (\mathbf{v}_1^{(1)})^T & (\mathbf{v}_1^{(2)})^T & \dots & (\mathbf{v}_1^{(F)})^T \\ (\mathbf{v}_2^{(1)})^T & (\mathbf{v}_2^{(2)})^T & \dots & (\mathbf{v}_2^{(F)})^T \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{v}_N^{(1)})^T & (\mathbf{v}_N^{(2)})^T & \dots & (\mathbf{v}_N^{(F)})^T \end{bmatrix}_{(3N \times F)}. \quad (1)$$

3.2 The 3D face representation with local support

Once the 3D face dataset is obtained, we are looking for an appropriate group of basis vectors \mathbf{B} which can faithfully represent 3D faces. Moreover, we require that some basis vectors have local support which can describe the deformation trend of each face part. To this end, we look for an appropriate matrix factorization of the sample matrix \mathbf{X} into K basis vectors $\mathbf{B} \in \mathbb{R}^{3N \times K}$ with weights $\mathbf{W} \in \mathbb{R}^{K \times F}$ as following:

$$\mathbf{X} = \mathbf{B} \cdot \mathbf{W}. \quad (2)$$

It is the common PCA, if the space of solutions to Eq. (2) is constrained on the basis vectors \mathbf{B} by orthogonality $\mathbf{B} \cdot \mathbf{B}^T = \mathbf{I}$. However, the solution obtained by PCA is often called principal components (here we call them basis vectors) that usually have global support on the whole mesh, i.e., each basis vector can affect the deformation of the whole mesh, which is difficult to define the semantic bases for our purpose. Thus, sparse and spatially localized representation is necessary. To this end, inspired by Neumann et al. (2013) who used sparsity on the basis vectors, the l_1 norm can be introduced to regularize the basis vectors $\Omega(\mathbf{B})$ (Bach et al. 2012). The problem of looking for basis vectors can be formulated as a minimization problem as follows:

$$\arg \min_{\mathbf{B}, \mathbf{W}} \|\mathbf{X} - \mathbf{B} \cdot \mathbf{W}\|_F^2 + \Omega(\mathbf{B}), \quad \text{s.t. } \max(|\mathbf{W}_{:,j}|) = 1, \forall j, \quad (3)$$

where $\mathbf{W}_{:,j}$ denotes the j th column of the weights \mathbf{W} and the constraints on the weights can avoid the weights getting too large and the basis vectors getting too small. Moreover, the absolute term allows negative weights that can allow two directions of deformation on the mesh.

To find a suitable regularizer for the dataset, through observing the basis vectors \mathbf{B} , we find each triplet in each column is a three-dimensional vector $\mathbf{b}_k^i = [x, y, z]_k^{(i)}$ corresponding to the displacement of the vertex i in the basis vector k . While regularizing the basis vectors \mathbf{B} will induce sparsity, each dimension of the displacement vector will vanish (Neumann et al. 2013). To make the displacement sparsity (or vanish), we can induce the l_1 norm to penalize the displacement vectors as following:

$$\Omega(\mathbf{B}) = \sum_{k=1}^K \sum_{i=1}^N \lambda_{ki} \|\mathbf{b}_k^i\|_2 \tag{4}$$

which is called l_1/l_2 norm and is a representation of group sparsity (more details are related to Bach et al. 2012). The parameters λ_{ki} are exploited to enforce local representation for the basis vectors. The human face can be segmented into several individual parts according to the semantic information. In our experiments, we manually segment the face into five parts as the colored regions shown in Fig. 1.

The parameters λ_{ki} can be one when the vertex i belongs to the control region of the basis vector k , otherwise it is zero. The optimization of Eq. (3) can be solved using the alternating direction method of multipliers (ADMM) (more details are related to Neumann et al. 2013).

3.3 The semantic representation bases

Once obtaining the 3D face representation with local support, we can represent the dataset \mathbf{X} by the new representation form. However, these bases cannot semantically describe meaningful deformation. Thus, we introduce a powerful tool to incorporate the semantic parameters to enable the representation meaningful. In our experiments, we use 20 semantic parameters as enumerated in Table 1.

Similar to Streuber et al. (2016), we learn a mapping between the face shape space and semantic parameter space as following:

$$\operatorname{argmin}_{o,s} [1|\mathbf{X}] \begin{bmatrix} \mathbf{o} \\ \mathbf{A} \end{bmatrix} - \mathbf{L}, \tag{5}$$

where $\mathbf{L}_{i,h} = [l_{i,1}, \dots, l_{i,20}]$ denotes semantic vector to the face i . The regression coefficients \mathbf{A} and the corresponding offset \mathbf{o} can be computed in a least-squares sense. Then, we can use the magnitude of each row of matrix \mathbf{A} to find which basis vector can represent the defined semantic parameters. After that, we reorder these base vectors by placing semantically related base vectors in front. However, these base vectors still have a drawback which is not orthogonality, which will result in the case that increasing nose height normally results in additional nose width. Thus, inspired by Hasler et al. (2009) who used Gram–Schmidt algorithm, we make the base vectors orthogonal to all the semantic base vectors. Both the semantic and non-semantic bases compose our final semantic parametric 3D face representation.

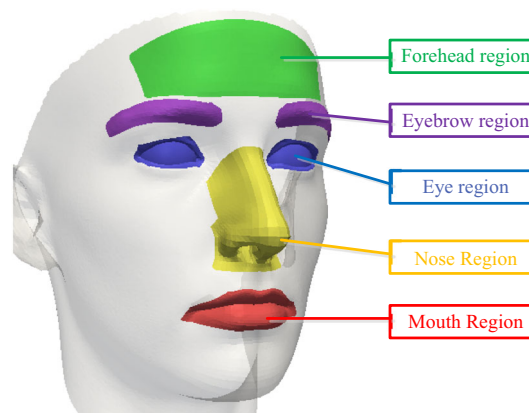


Fig. 1 The individual face part regions

Table 1 The semantic parameters

1	Mouth wide/narrow	11	Eyes long/short
2	Mouth convex/concave	12	Eyes up/down
3	Mouth up/down	13	Forehead wide/narrow
4	Mouth left/right	14	Forehead convex/concave
5	Nose long/short	15	Forehead large/small
6	Nose wide/narrow	16	Eyes interval distance long/short
7	Nose high/low	17	Right eye up/down
8	Nose left/right	18	Right eye width/narrow
9	Nose up/down	19	Left eye up/down
10	Eyes large/small	20	Left eye width/narrow

4 Visually reshaping the face in images

Reshaping the face in images is an interesting application, in particular, the reshaped results can be visually presented in the image at the same time. Relying on our proposed 3D face representation, it can be completed to intuitively and semantically reshape the face due to the semantic bases which corresponds to the meaningful deformation trend. To reshape the face in the image, similar to Lu et al. (2016) who parameterized an image, we estimate the parameters from an image, including 3D face shape F , image shading $s(x)$, albedo $\rho(x)$ and illumination θ . These parameters can be rendered into the original image by any reasonable modern 3D-rendering process with the following rule:

$$I(x) = \text{Render}(\rho(x), F, \theta, s(x)). \quad (6)$$

Through manipulating the face shape F , we can easily obtain the reshaped face image. Thus, we need to represent the face in image with our proposed 3D face representation.

4.1 Recovering 3D face geometry

Recovering 3D face geometry from image has been a hot field since Blanz and Vetter (1999) used their 3D morphable face model to reconstruct the 3D face from images. We follow their method to estimate the 3D face geometry relying on our proposed representation. We use the detected facial landmarks by an off-the-shelf detector (Kazemi and Sullivan 2014) including 68 landmarks. Then, through minimizing the distance between the landmarks on 3D face and corresponding ones on the image, the parameters and pose of 3D face can be computed with the following equation:

$$E(\chi) = \omega_c E_c(\chi) + \omega_{\text{lan}} E_{\text{lan}}(\chi) + \omega_{\text{reg}} E_{\text{reg}}(\chi), \quad (7)$$

where the unknowns $\chi = \{\mathbf{w}, \mathbf{R}, \mathbf{t}\}$ denotes the parameters of 3D face, $F(\mathbf{w})$ represents a 3D face mesh and (\mathbf{R}, \mathbf{t}) denotes the pose of the face in the image. ω_c , ω_{lan} and ω_{reg} are energy term weights. The term E_c minimizes the distance between the synthetic face and the input image, the landmark term E_{lan} minimizes the distance between the landmarks, and the regularization term penalizes the deviation of the face from the normal distribution:

$$E_c(\chi) = \frac{1}{|M|} \sum_{p \in M} \|C_{\text{input}}(p) - C_{\text{render}}(p)\|_2, \quad (8)$$

where C_{input} is the input image and C_{render} is the rendered image. $p \in M$, which is a visible pixel computed from face segmentation that is introduced in the next section. The landmark fitting term E_{lan} and the regularization term E_{reg} are defined as:

$$E_{\text{lan}} = \frac{1}{|F|} \sum_{f_i \in F} \|f_i - \prod \Phi_P(\mathbf{R}\mathbf{v}_i + \mathbf{t})\|_2^2, \quad (9)$$

$$E_{\text{reg}} = \sum_{i=1}^n \left(\frac{\alpha_{id,i}}{\sigma_{id,i}} \right)^2, \quad (10)$$

where $f_i \in F$ is a 2D landmark in the image, and the objective function is optimized using a Gauss–Newton solver based on iteratively reweighted least squares with three levels of image pyramids (more details are

related to Blanz and Vetter 1999). Figure 2 shows an example of recovering 3D face geometry and we fit the recovered 3D face on the face image.

4.2 Recovering other parameters and image rendering

After recovering 3D face geometry, we also need to recover other relevant scene parameters mentioned before rendering the new face image. First, to preserve all local characteristics of the face in the image, we segment the image into the face region and background following the recent state-of-the-art, deep method (Nirkin et al. 2017). Figure 3 gives an example of face segmentation to demonstrate that this method can satisfy our purpose. The excellent result is very helpful for the extraction of face texture.

For the illumination parameters, we follow Lu et al. (2016)’s method that used a spherical harmonic (SH) model. The illumination coefficients θ can be estimated by the following equation:

$$\arg \min_{\theta} \sum_{x \in \Omega_f} (\mathcal{S}(x) - \mathbf{SH}(n(x), \theta))^2, \tag{11}$$

where $\mathcal{S}(x)$ denotes the shading from intrinsic method (Ramamoorthi and Hanrahan 2001), $\mathbf{SH}()$ denotes the SH-based rendering of the normal $n(x)$, and $x \in \Omega_f$ restricts this optimization to the face pixels as shown in Fig. 3c.

Once obtaining all the relevant scene parameters, we can render the new face image with reshaped face model. We compose the final image with the new face shape by blending the new rendered face image and the background image. Figure 4 shows an example of final face image blending.

5 Experiments

To evaluate our 3D face representation and the face reshape method, we perform quantitative evaluation to assess reconstruction quality on unseen data and respectively show some reshaped face result on 3D models and 2D images.

5.1 Quantitative evaluation

It is a common manner to assess the 3D face representation through the reconstruction quality from the unseen dataset. For this, we randomly split 100 face meshes as the training dataset from our reconstructed dataset introduced in the Sect. 3.1, including (50 males and 50 females) and the rest 100 face meshes are used as the testing dataset. We train the PCA face model and our proposed face model with varying number of the basis vectors K . Given the fixed number of basis vectors, the PCA model has the absolute advantage to compress the shape space with high proportion. To demonstrate that our method also keep almost comparable compressed proportion in some number of basis vectors, we invert Eq. (2) for representing the training set and measure the difference in vertex coordinates with the error metric introduced by Kavan et al. (2010) (root mean squared error). The average reconstruction errors over the sample matrix (i.e., the training dataset) with different number of the basis vectors are reported in Fig. 5.

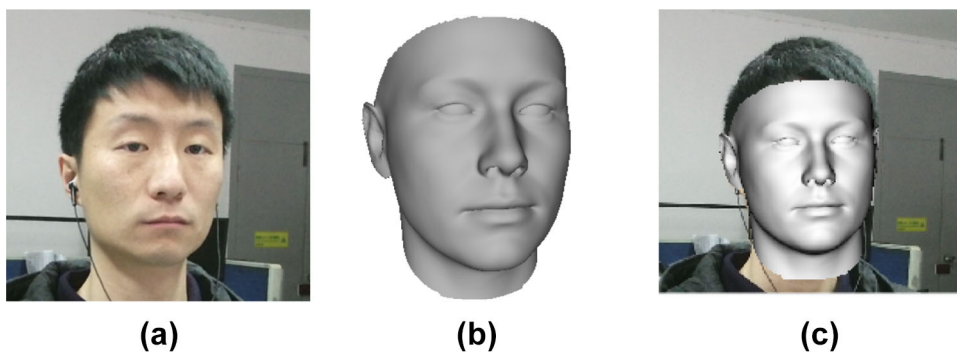


Fig. 2 An example of recovering 3D face geometry. **a** The input image, **b** the recovering 3D face geometry, **c** fitting image with the 3D face

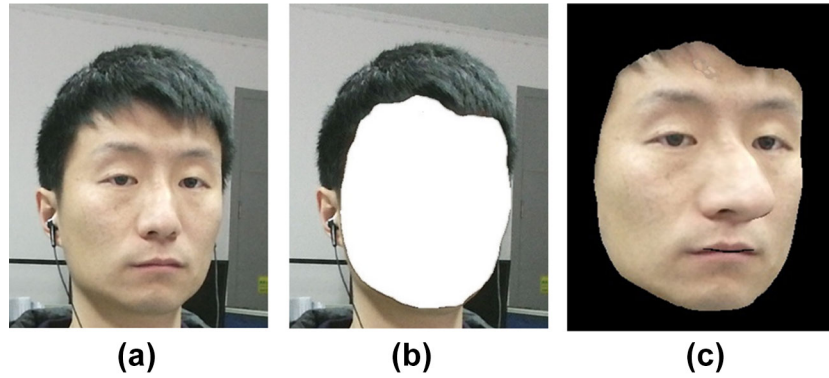


Fig. 3 An example of face segmentation, **a** the input images, **b** segmented background images, **c** facial regions

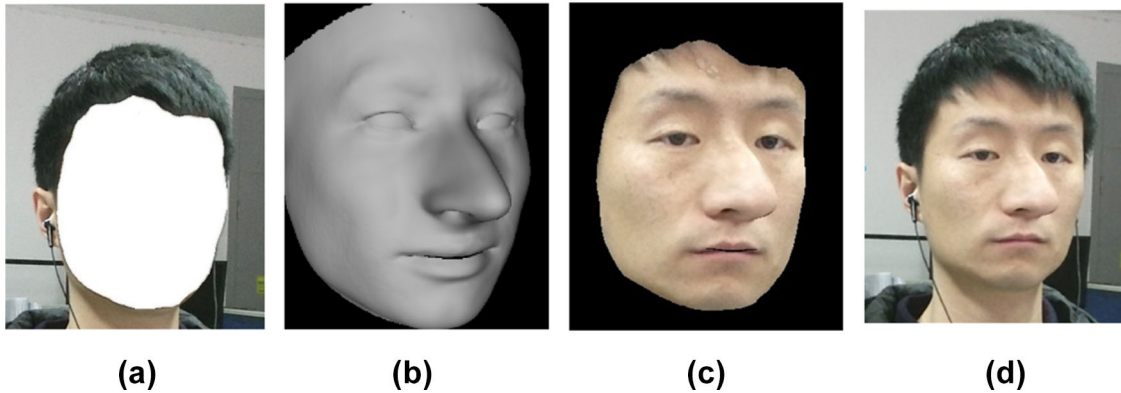


Fig. 4 An example of final blending, **a** the background image, **b** the reshaped 3D face, **c** rendered facial image, **d** blending result

The results in Fig. 5 indicate that our face representation has comparable compressed proportion with PCA model in more than about 35 bases. Moreover, the error under about 45 bases is nearly equal to the one under about 40 bases of PCA model.

For the testing dataset, we select varying number of bases to represent the testing set to show the generalization to unseen data and the results are shown in Fig. 6. The results in Fig. 6 show that our 3D face representation with localized bases generalizes better to new face shapes than PCA model inside the testing dataset. To understand why our representation generalizes so well to unseen data, considering the case that there are two models, one of which has big eyes and a small mouth, while a second one has small eyes and a big mouth. PCA model uses a global linear combination, so it is not possible to produce a new face with small eyes and a small mouth. In contrast, our representation applies localized basis vectors corresponding to facial parts for eyes and mouth, and can reproduce their desirable shapes separately, even though this is unseen in the dataset. We select 65 bases as the basis vectors to represent the 3D face in the next applications.

5.2 Face reshaping results

After representing the 3D face geometry with the proposed semantic bases, we can change the coefficients of the semantic bases to easily reshape the 3D face models, and then the reshaped face models can be rendered into new images. In Figs. 7 and 8, we show several results of the reshaped 3D face models and their corresponding rendered images.

Figures 7a and 8a give the input images and the corresponding recovered 3D face models. The results from Figs. 7b and 8b to 7i and 8i show the reshaped 3D face models and their rendered effects following the semantic parameters in Table 1 which are mouth wide/narrow, mouth convex/concave, nose long/short, nose wide/narrow, eyes large/small, eyes long/short, forehead convex/concave and forehead large/small, respectively. The small pictures in the corner show the deformation area of the face which is guided by the 3D face model and colored with red. The editing direction in Fig. 7 is inverse of that in Fig. 8.

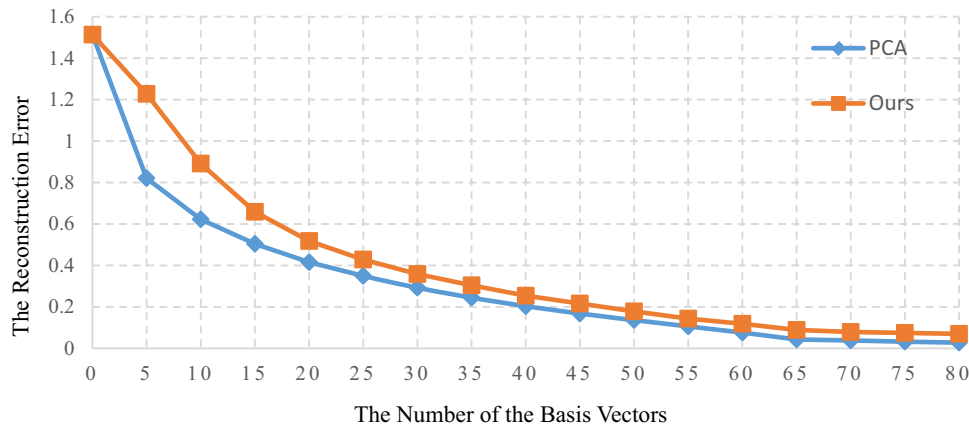


Fig. 5 Generalization to unseen data, we plot reconstruction error (y-axis) with respect to the number of the basis vectors on the training dataset. Our 3D face representation has comparable compressed proportion with more than about 35 bases

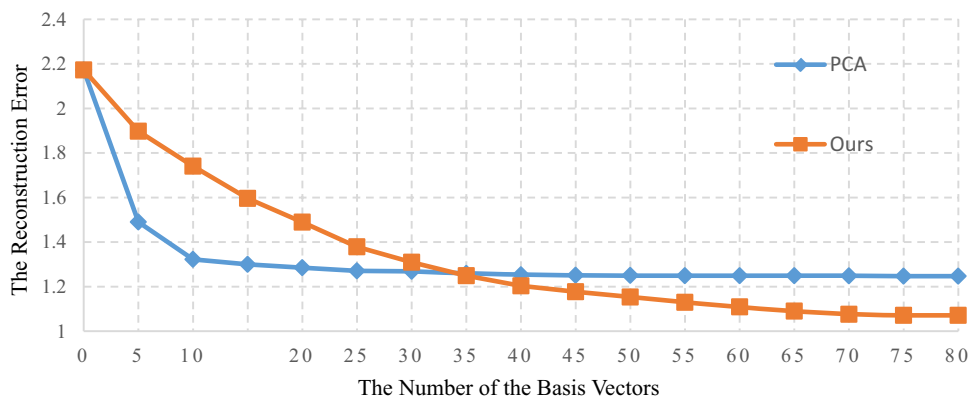


Fig. 6 Generalization to unseen data, we plot reconstruction error (y-axis) with respect to the number of the basis vectors on the training dataset. Our 3D face representation has comparable compressed proportion in more than about 35 bases

The (c), (d) and (h) columns in Figs. 7 and 8 show the editing results with the depth direction. For example, Figs. 7d and 8d show the long and short noses. These results demonstrate the recovered 3D face model helps to improve the editable ability, which can make up for the limitation of traditional 2D image processing methods.

Figure 9 shows a deformation sequence of the nose when we change the semantic parameters of nose parts. With the deformation of the face model, the face in the image also occurred the same deformation. If you continue to increase the parameters, the result is like Pinocchio’s lying which is also a potential application for teaching children such as e-book (Kang et al. 2017). This result cannot be achieved by the image-based method or the PCA face model.

We also edit the face shape following the semantic meanings. Figure 10 shows the reshaped results which demonstrate that our proposed representation can reshape the face locally as the reshaped nose in Fig. 10b and right eye in Fig. 10c. Moreover, These reshaping trends are semantic. These semantic bases can also be combined to reshape the face as shown in Fig. 10d.

6 Conclusion and future work

In this paper, we provide a novel parametric representation of 3D face model that describes the face shape as a linear combination of semantic and non-semantic bases. The semantic bases correspond to the meaningful

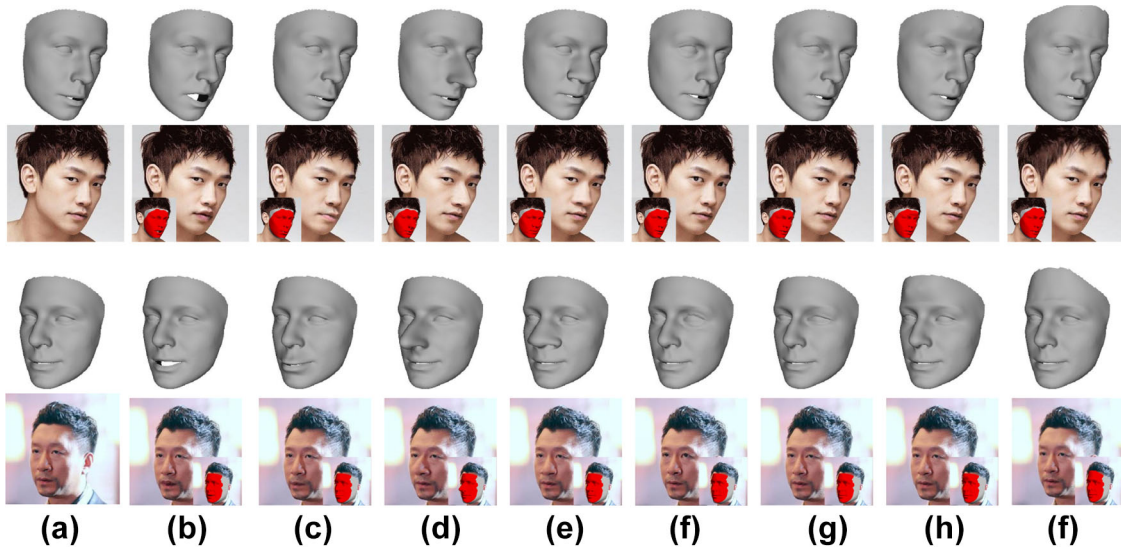


Fig. 7 Results of reshaped 3D face models and their corresponding reshaped images. **a** Original images and 3D face model, **b** making the mouth wide, **c** making the mouth convex, **d** making the nose long, **e** making the nose wide, **f** making the eyes large, **g** making the eyes long, **h** making the forehead convex, **i** making the forehead large

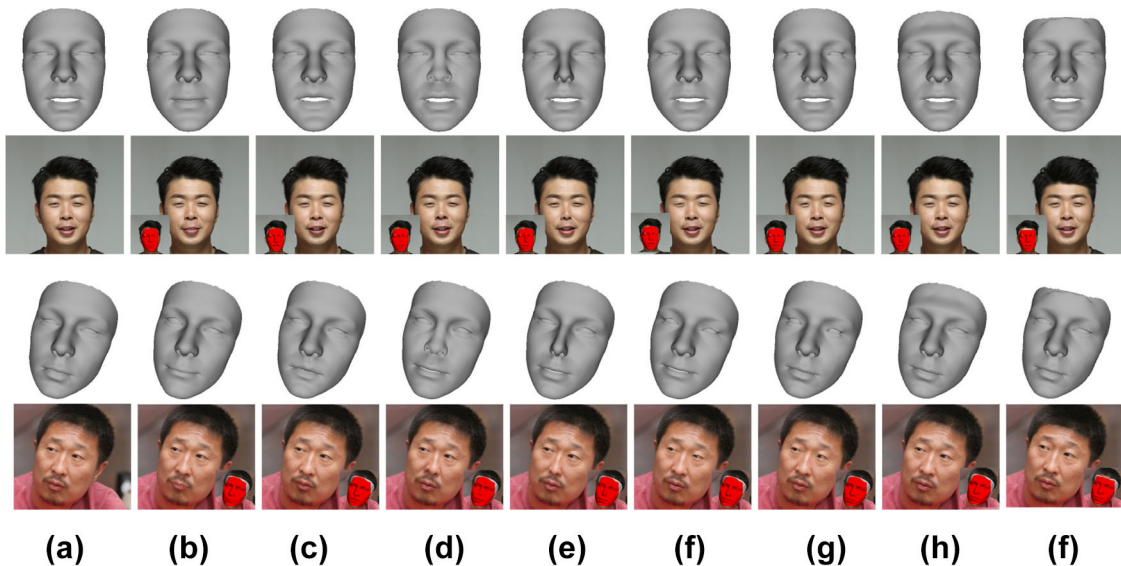


Fig. 8 Results of reshaped 3D face models and their corresponding reshaped images. **a** Original images and 3D face model, **b** making the mouth narrow, **c** making the mouth concave, **d** making the nose short, **e** making the nose narrow, **f** making the eyes small, **g** making the eyes short, **h** making the forehead concave, **i** making the forehead small

deformation trend of each face parts. Thus, this representation can edit the face shape semantically and intuitively. We apply the proposed face representation in the editing of face image. The results demonstrate how our face representation provides novel capabilities for intuitive and localized face shape editing. Moreover, our representation builds a rich space of deformations that generalized better to unseen data due to the local support. The reshaping of face image is also a powerful tool for face shape editing, as well as for data exploration and statistical shape processing, and we are eager to see interesting applications of our 3D face representation.

Because our approach adopts several previous works, their limitations might also be our obstacles. For example, the deep method of face segmentation is sensitive to illumination, which might lead to an incomplete facial texture. A worthwhile solution is to use the segmentation of 3D mesh (Mansouri and

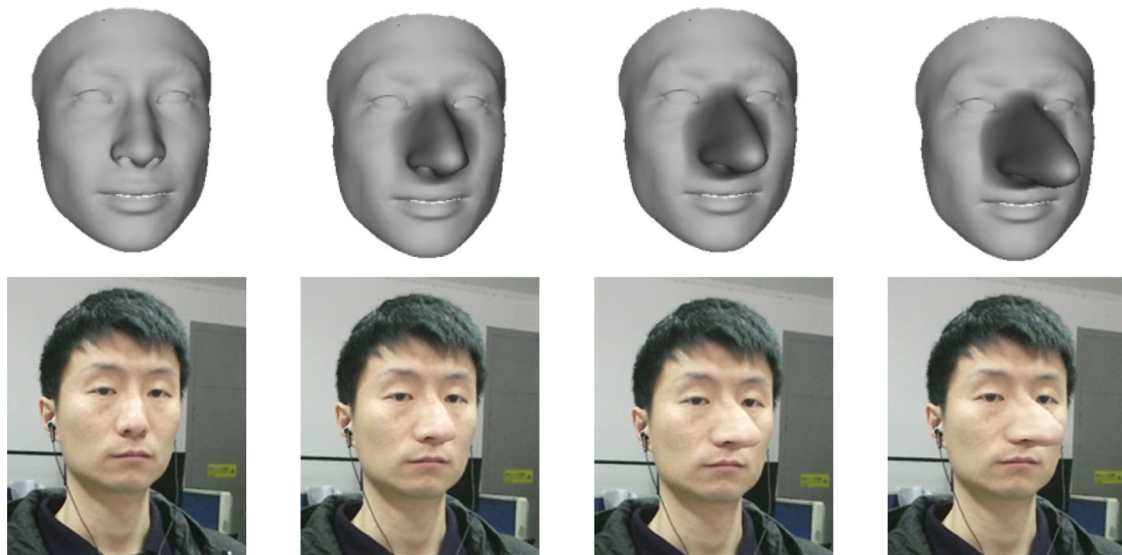


Fig. 9 The deformation sequence of the nose. From the left to left we gradually increase the nose-length parameters

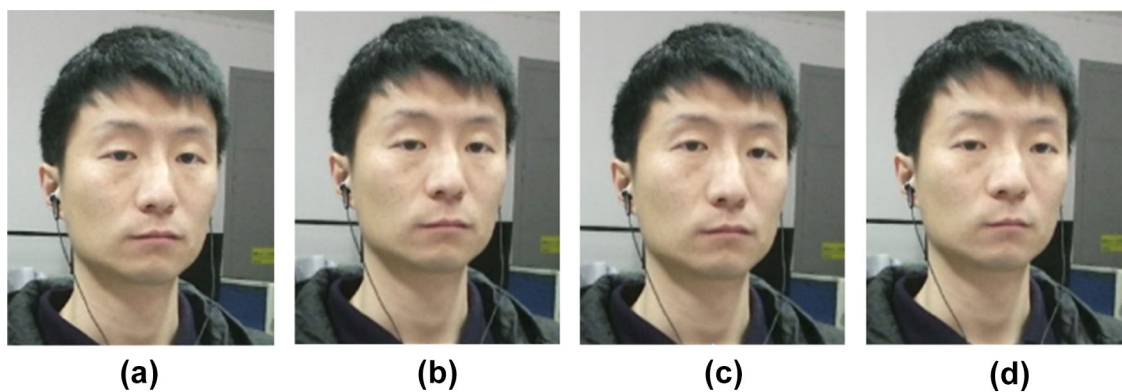


Fig. 10 The semantic editing of face in images, **a** the original image, **b** increasing the height of the nose, **c** increasing the height of the right eye, **d** increasing the distance of the eyes, the height of the nose and reducing the width of the mouth

Ebrahimnezhad (2016) reconstructed from image to guide the segmentation of the facial region in the image. Another limitation is that the most time-consuming part is to solve the nonlinear optimization, although it takes about 4 s in our experimental environment. As a result, the user needs to wait for the reconstruction to finish before the real-time face editing. Recently, the excellent convolution neural networks (CNN)-based method (Guo et al. 2017) uses a data-driven approach to avoid repeatedly solving the nonlinear optimization problems. In the future, we will follow their method to train the CNN to replace the traditional optimization solution, which will present a complete real-time system to edit the face images. We will also improve our method to deal with facial expressions in the images. Cao et al. (2014) proposed the facial database called Facewarehouse including a wider range of faces and expressions which can help to deal with facial expressions in the images. We also leave it to our future work. We are also trying to firstly retrieve face image (Lu et al. 2017), and then semantically edit the image for interesting applications.

Acknowledgements The research is supported in part by NSFC (61572424) and the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme FP7 (2007-2013) under REA Grant Agreement no. 612627-”AniNex”. Min Tang is supported in part by NSFC (61572423) and Zhejiang Provincial NSFC (LZ16F020003).

References

- Alvarez FJA, Parra EBB, Tubio FM (2017) Improving graphic expression training with 3D models. *J Vis* 20(4):889–904
- Bach F, Jenatton R, Mairal J, Obozinski G (2012) Optimization with sparsity-inducing penalties. *Found Trends Mach Learn* 4(1):1–106
- Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. In: *Proceedings of the 26th annual conference on computer graphics and interactive techniques*. ACM, pp 187–194
- Blanz V, Basso C, Poggio T, Vetter T (2003) Reanimating faces in images and video. *Comput Gr Forum* (Blackwell Publishing, Inc) 22(3):641–650
- Blanz V, Scherbaum K, Vetter T, Seidel H-P (2004) Exchanging faces in images. *Comput Gr Forum* (Blackwell Publishing, Inc) 23(3):669–676
- Cao C, Weng Y, Zhou S, Tong Y, Zhou K (2014) Facewarehouse: a 3d facial expression database for visual computing. *Vis Comput Gr IEEE Trans* 20(3):413–425
- Chalas I, Urbanova P, Jurik V, Ferkova Z, Jandova M, Sochor J, Kozlikova B (2017) Generating various composite human faces from real 3D facial images. *V Comput* 33(4):443–458
- Chou J-K, Yang C-K, Gong S-D (2012) Face-off: automatic alteration of facial features. *Multimed Tools Appl* 56(3):569–596
- Fan X, Chai Z, Feng Y, Wang Y, Wang S, Lou Z (2016) An efficient mesh-based face beautifier on mobile devices. *Neurocomputing*. 172:134–142
- Golub GH, Vanloan CF (1996) *Matrix computations* (Johns Hopkins studies in mathematical sciences). The Johns Hopkins University Press, Baltimore
- Guo Y, Zhang J, Cai J, Jiang B, Zheng J (2017) 3DFaceNet: real-time dense face reconstruction via synthesizing photo-realistic face images. [arXiv:1708.00980](https://arxiv.org/abs/1708.00980)
- Hasler N, Stoll C, Sunkel M, Rosenhahn B, Seidel H-P (2009) A statistical model of human pose and body shape. *Comput Gr Forum* 28(2):337–346
- Kang D, Tian F, Sahandi R (2017) Interactive illustration of collage for children with folktale e-book. *J Vis* 20(3):639–650
- Kavan L, Sloan PP, O’Sullivan C (2010) Fast and efficient skinning of animated meshes. *Comput Gr Forum* (Blackwell Publishing Ltd) 29(2):327–336
- Kazemi V, Sullivan J (2014) One millisecond face alignment with an ensemble of regression trees. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1867–1874
- Kemelmacher-Shlizerman I (2016) Transfiguring portraits. *ACM Trans Gr (TOG)* 35(4):94 (1–94:8)
- Klum S, Han H, Jain AK, Klara B (2013) Sketch based face recognition: forensic vs. composite sketches. In: *Proceedings of the 6th international conference on biometrics*. IEEE, pp 1–8
- Lewis JP, Anjyo K, Rhee T, Zhang M, Pighin FH, Deng Z (2014) Practice and theory of blendshape facial models. *Eurographics* (State Art Rep) 1(8)
- Leyvand T, Cohen-Or D, Dror G, Lischinski D (2008) Data-driven enhancement of facial attractiveness. *ACM Trans Gr (TOG)* 27(3):38
- Li H, Yu J, Ye Y, Bregler C (2013) Realtime facial animation with on-the-fly correctives. *ACM Trans Gr (TOG)* 32(4):42 (1–42:10)
- Liao Q, Jin X, Zeng W (2012) Enhancing the symmetry and proportion of 3D face geometry. *Vis Comput Gr IEEE Trans* 18(10):1704–1716
- Lu J, Sunkavalli K, Carr N, Hadap S, Forsyth D (2016) A visual representation for editing face images. [arXiv:1612.00522](https://arxiv.org/abs/1612.00522)
- Lu Z, Yang J, Liu Q (2017) Face image retrieval based on shape and texture feature fusion. *Comput Visual Media* 3(4):359–368
- Ma DS, Correll J, Wittenbrink B (2015) The Chicago face database: a free stimulus set of faces and norming data. *Behav Res Methods* 47(4):1122–1135
- Mansouri S, Ebrahimnezhad H (2016) Segmentation-based semi-regular remeshing of 3D models using curvature-adapted subdivision surface fitting. *J Vis* 19(1):141–155
- Neumann T, Varanasi K, Wenger S, Wacker M, Magnor M, Theobalt C (2013) Sparse localized deformation components. *ACM Trans Gr (TOG)* 32(6):179
- Nirkin Y, Masi I, Tran AT, Hassner T, Medioni G (2017) On face segmentation, face swapping, and face perception. [arXiv:1704.06729](https://arxiv.org/abs/1704.06729)
- Paysan P, Knothe R, Amberg B (2009) A 3D face model for pose and illumination invariant face recognition. In: *6th IEEE international conference on advanced video and signal based surveillance (AVSS’09)*, IEEE, Genova, Italy, pp 296–301
- Ramamoorthi R, Hanrahan P (2001) An efficient representation for irradiance environment maps. In: *Proceedings of the 28th annual conference on computer graphics and interactive techniques*. ACM, pp 497–500
- Shamir A, Sorkine O (2009) Visual media retargeting. *ACM SIGGRAPH ASIA 2009 courses*. ACM, p 11
- Streuber S, Quiros-Ramirez MA, Hill MQ, Hahn CA, Zuffi S, O’Toole A, Black MJ (2016) Body talk: crowdshaping realistic 3D avatars with words. *ACM Trans Gr (TOG)* 35(4):54
- Suryanarayana GK, Dubey R (2017) Image analyses of supersonic air-intake buzz and control by natural ventilation. *J Vis* 1–17
- Thies J, Zollhofer M, Stamminger M, Theobalt C, Niebner M (2016) Face2face: real-time face capture and reenactment of RGB videos. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, pp 2387–2395
- Tran AT, Hassner T, Masi I, Medioni G (2016) Regressing robust and discriminative 3D morphable models with a very deep neural network. [arXiv:1612.04904](https://arxiv.org/abs/1612.04904)
- Vlasic D, Brand M, Pfister H, Popovic J (2005) Face transfer with multilinear models. *ACM Trans Gr (TOG)* 24(3):426–433
- Zhang Y, Lin W, Zhou B, Chen Z, Sheng B, Wu J (2014) Facial expression cloning with elastic and muscle models. *J Vis Commun Image Represent* 25(5):916–927
- Zhou S, Fu H, Liu L, Cohen-Or D, Han X (2010) Parametric reshaping of human bodies in images. *ACM Trans Gr (TOG)* 29(4):126