

# Manifold Estimation in View-based Feature Space for Face Synthesis across Poses

Xinyu Huang, Jizhou Gao, Sen-ching S. Cheung, and Ruigang Yang

Center for Visualization and Virtual Environments  
University of Kentucky, USA

**Abstract.** This paper presents a new approach to synthesize face images under different pose changes given a single input image. The approach is based on two observations: 1. a series of face images of a single person under different poses could be mapped to a smooth manifold in the unified feature space. 2. the manifolds from different faces are separated from each other by their dissimilarities. The new manifold estimation is formulated as an energy minimization problem with smoothness constraints. The experiments show that face images under different poses can be robustly synthesized from one input image, even with large pose variations.

## 1 Introduction

Face synthesis has been an active research topic in computer vision since the 1990s. Synthesizing an unseen view of a face accurately and efficiently can definitely improve the quality of face recognition and face reconstruction. However, it remains to be a challenging problem due to pose, illumination, facial expression variations, and occlusions. It is also known that changes caused by pose and illumination are much larger than the changes of personal appearances [1].

In this paper, we propose a novel approach to synthesize unseen views across various poses given one input image. We assume that face relighting, pose estimation, and face alignment are solved using existing technologies. For example, in order to change the illumination condition of a face image (i.e., face re-lighting), lots of approaches have been developed recently (e.g., [2, 3]). Most of them can even deal with cast shadows and saturated areas. Another important problem, face alignment that usually fits a 2D/3D face model to one input image could be solved across poses, illumination, and even partial occlusions (e.g., [4–7]). However, fitting a model to one input image accurately does not guarantee that this model is suitable for all other conditions since face fitting and synthesis from one 2D input itself is an ill-posed problem. For example, a face model generated from one image could fail to represent the same face at other poses, especially at large pose changes. In addition, though face models can be used to estimate poses, it turns out that appearance-based methods using non-linear dimensionality reduction technologies could obtain more accurate pose estimation, e.g., less than 2 degree in [8, 9].

In our approach, we first build multiple view-based Active Appearance Models (AAMs) [10] for a database of face images from different persons under various poses. Each AAM extracts shape and appearance parameters corresponding to a certain pose. After transforming the parameters from all AAMs into the same feature space, a face image can be mapped to a high-dimensional point in the unified feature space. Furthermore, we observe that a series of face images of a single person under different poses form a smooth manifold and the shape of manifolds from similar faces resemble each other. Given a new input face image, i.e., a new point in the unified feature space, the ultimate goal, therefore, is to estimate a new manifold corresponding to the input face at different poses. We test our method on the images obtained from 3D face Morphable Model (3DMM) [11] and CMU-PIE database [12]. Experiments show that this approach is able to synthesize faces effectively even with large pose changes.

The rest of the paper is organized as followings. Section 2 reviews the related works. Section 3 provides our algorithm of manifold estimation. Section 4 shows our experimental results for face synthesis. Section 5 gives the conclusion and future work.

## 2 Related Works

There are many existing methods for face synthesis across poses. They can be divided into two categories, model-based methods [13, 4–6], and appearance-based methods [14–17].

In [13], Vetter *et al.* propose that a 2D view of an object could be represented by a weighted sum of 2D views of other objects. The weights remain the same on the transformed views when the object belongs to a linear class. Similar to this idea, the morphable model [6] is formulated as a vector space of 3D shapes and surface textures that are spanned by a training set that contains 200 laser scanned 3D face models. 3D shape reconstruction is a fitting problem by minimizing the difference between the rendered model image and the input image. This method generates face models with good qualities and is widely used in many face applications. However, it is computationally expensive and needs a careful manual initialization and segmentation. Cootes *et al.* [10] first propose AAM to solve the face fitting problem. Romdhani *et al.* [5] extend it to a nonlinear model using kernel PCA to align faces across poses. A more general method using the Gaussian mixture model of shape-and-texture appearance is proposed in [18] and fitting is solved by the EM algorithm. In [19], a combined 2D+3D AAM is proposed for real-time face fitting. In [4], three separate AAMs are trained at different face poses. Assuming all the modeled features are visible, a linear model is used to represent the correlations between appearance in two views and appearance from a new pose can then be predicted given an input face image.

In [14], Gross *et al.* estimate the Fisher Light-Field of the subject’s head and then perform the matching between the probe and gallery using the Fisher Light-Fields. In [16], Tenenbaum *et al.* first propose the bilinear model to separate

the content (intra-class) and style (inter-class). This model is extended to multi-linear model in [17] to separate face poses, expressions, and appearances. In [20], Li *et al.* further extend the model to the nonlinear case in which an input space is transformed to a feature space using the Gaussian kernel.

In [15], Turk *et al.* first compare the parametric eigenspace and view-based eigenspace. Inspired by the view-based eigenspace, [21, 22] show that appearance of one object forms a smoothly varying manifold in the eigenspace. The pose estimation and object recognition can be performed by distance computation between the projection of a new input and manifolds in an object database. This appearance model is further applied in visual tracking and recognition in video sequences [23].

### 3 Algorithm

#### 3.1 Manifolds of Parameters in Feature Space

In an AAM, a shape  $s$ , defined by a set of 2D facial markers, can be represented by a mean shape  $\bar{s}$  and a set of shape bases  $s_i$ :

$$s = \bar{s} + \sum_{i=1}^d \alpha_i s_i \quad (1)$$

where  $\alpha = (\alpha_1, \dots, \alpha_d)^T$  is the shape parameter. Similarly, the appearance  $t$  of a face can be represented by a mean appearance  $\bar{t}$  and a set of appearance bases  $t_i$ :

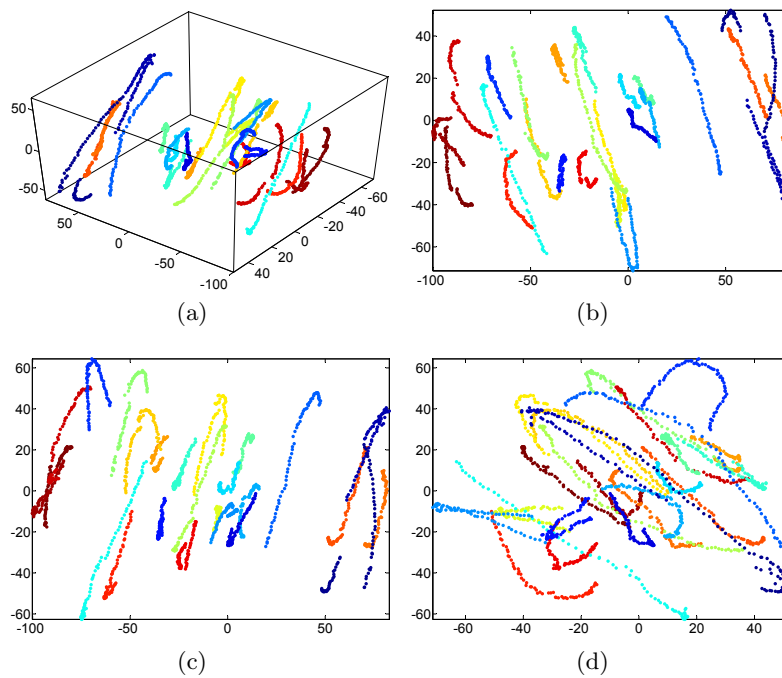
$$t = \bar{t} + \sum_{i=1}^d \beta_i t_i \quad (2)$$

where  $\beta = (\beta_1, \dots, \beta_d)^T$  is the appearance parameter. A mean shape  $\bar{s}$ , a mean appearance  $\bar{t}$ , shape bases  $s_i$ , and appearance bases  $t_i$  are obtained by applying Principal Component Analysis (PCA) on a face database. Moreover, appearances defined by  $t$  in Equation (2) are shape-free textures by applying piece-wise affine transformations to the mean shape  $\bar{s}$ . When new parameters  $\alpha_{new}$  and  $\beta_{new}$  are given, we are able to compute  $s_{new}$  and  $t_{new}$  separately and then warp the  $t_{new}$  back to its shape  $s_{new}$  to obtain a new face image.

Given one AAM, a face image can be uniquely defined by two high dimensional points,  $\alpha \in \mathbb{R}^d$  in the shape parameter space and  $\beta \in \mathbb{R}^d$  in the appearance parameter space, where  $d$  is the number of PCA bases. However, due to the non-linearity of pose changes, shape and appearance variations can not be modeled by a single AAM. Similar to the approach in [4], we build a mixture of AAMs that cover the whole pose changes.

Furthermore, after  $n$  view-based feature spaces are transformed to the reference feature space (e.g.,  $90^\circ$  at the frontal view) by multiplying  $n$  projection matrices, we observe that the transformed parameters form smooth manifolds. Each manifold corresponds to a series of face images of a single person under

different poses. Figure 1 shows the distribution of shape parameters of 25 faces. Only the first three principal components are displayed for visualization purpose. The manifolds of appearance parameters have the similar distributions as Figure 1. The manifolds of parameters have two important properties, smoothness and separateness. Smoothness means that the shape of a manifold changes smoothly as poses vary. Separateness means that manifolds are separated from each other in the feature space. The more the two faces are similar, the closer the two corresponding manifolds are to each other. The manifolds used in our approach are quite different from those estimated in the input space directly using nonlinear dimensionality reduction technologies [24, 25]. In their works, manifolds of all faces are grouped together so that pose changes are significantly larger than personal facial features. This is useful for pose estimation. However, it also makes the synthesis of new faces intractable.



**Fig. 1.** Manifolds of shape parameters of 25 faces. Each color represents one face horizontally varying from  $0^\circ$  to  $180^\circ$ . (a) First three principal components. (b) First and second principal components. (c) First and third principal components. (d) Second and third principal components.

### 3.2 Estimation of A New Manifold with One Input

Assuming only one  $d$ -dimensional shape parameter  $u^p \in \mathbb{R}^d$  at the  $p_{\text{th}}$  pose is given, the first step of our approach is to find a good approximation of this point from other database points  $\mathbf{x}^p = (x_1^p, \dots, x_k^p)$  at  $p_{\text{th}}$  pose in the shape parameter space. The similar approximation can also be applied in the appearance space. This could be formulated as an energy minimization problem,

$$\begin{aligned} \arg \min_{\pi_i} \|u^p - \sum_{i=1}^k \pi_i x_i^p\|_2, \\ \text{s.t. } \sum_{i=1}^k \pi_i = 1 \end{aligned} \quad (3)$$

where  $\pi = (\pi_1, \dots, \pi_k)$  is the reconstruction weights. Let  $Z = [\dots, x_i^p - u^p, \dots]$  be a matrix with  $k$  column vectors, the linear system  $Z^T Z \pi = 1$  is solved for the weights  $\pi$ . However, this system needs to be regularized if  $Z^T Z$  is singular or nearly singular that happens when  $k > d$ . This makes the solution of weights  $\pi$  unstable [26].

When the number of different faces in the database  $k$  is much larger than the number of principal components  $d$ , we adopt Matching Pursuit (MP) [27] to compute the reconstruction weights. This is based on two reasons. First, MP is a good choice to compute weights on over-complete bases. Second, a better face synthesis could be obtained by a few nearest neighbors. For example, a young female's face could be better synthesized by similar faces instead of all the faces including ones with beard. At each iteration of MP, the principal component  $x_\gamma^p$  is found such that it results in the maximal inner product with residual  $R^i u^p$ :

$$\begin{aligned} R^i u^p &= \langle R^i u^p, x_\gamma^p \rangle x_\gamma^p + R^{i+1} u^p \\ x_\gamma^p &= \arg \max_{\substack{x_\gamma^p \\ 1 \leq \gamma \leq k}} |\langle R^i u^p, x_\gamma^p \rangle| \end{aligned} \quad (4)$$

The  $u^p$  is approximated by  $m$  basis after convergence,

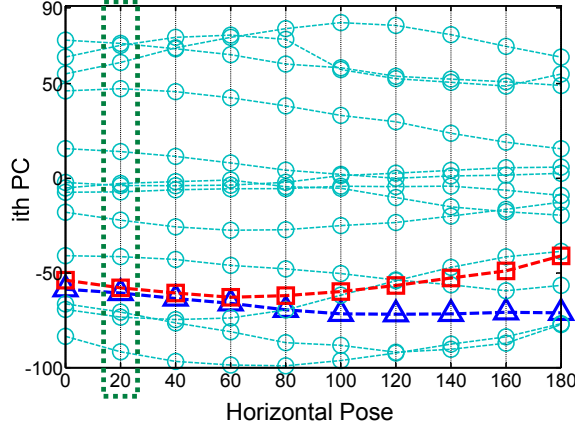
$$\begin{aligned} u^p &\approx \sum_{i=1}^m \pi_i x_{\gamma_i}^p \\ \pi_i &= \langle R^i u^p, x_{\gamma_i}^p \rangle \end{aligned} \quad (5)$$

Our second step is to estimate  $\mathbf{u}$  at other poses. Since principal components are orthogonal to each other, synthesis could be done along each direction  $l$ . Given the weights  $\pi$  and  $m$  face bases that best reconstruct  $u^p$ , we assume, for simplicity, the weights and face bases remain the same at other poses. Thus, the estimation is formulated as a minimization problem of the following energy

function  $E$ ,

$$\begin{aligned}
E &= E_1 + \lambda E_2 \\
E_1 &= \sum_{j=1}^n \left\| \sum_{i=1}^m \pi_i x_{i,l}^j - u_l^j \right\|_2 \\
E_2 &= \sum_{j=2}^{n-1} w_j \left\| 2y_l^j - y_l^{j+1} - y_l^{j-1} \right\|_2 + w_1 \left\| y_l^2 - y_l^1 \right\|_2 + w_n \left\| y_l^n - y_l^{n-1} \right\|_2 \\
w_j &= \exp\left(-\frac{1}{m} \sum_{i=1}^m \left\| 2y_{i,l}^j - y_{i,l}^{j+1} - y_{i,l}^{j-1} \right\|_2\right)
\end{aligned} \tag{6}$$

where  $E_1$  is the prior knowledge of  $u_l^j$  at  $j_{\text{th}}$  pose and  $l_{\text{th}}$  principal component,  $E_2$  is the first-order smoothness constraint for the estimated manifold,  $y$  is the projection of  $u$  in the reference feature space, and  $w_j$  is a weight associated to the average smoothness in the face database at  $j_{\text{th}}$  pose. Figure 2 shows the estimation framework. The initial guess is obtained by translation of a manifold on which its point at  $p_{\text{th}}$  pose is the nearest neighbor of the input point. Our approach is summarized in algorithm 1.



**Fig. 2.** Estimation of a new manifold at  $i_{\text{th}}$  principal component given one point at 20 degree. Light green manifolds are the ones in the database. The dark blue manifold is the initial guess, and the red manifold is estimated by minimizing the energy function  $E$ .

A further intuitive attempt to obtain a more accurate manifold estimation is to model changes of reconstruction weights  $\pi$  instead of using fixed values. However, it turns out difficult and is not very useful from our experiments. The reason is that the variance of weights  $\pi$  is small and their distribution

---

**Algorithm 1** Manifold Estimation

---

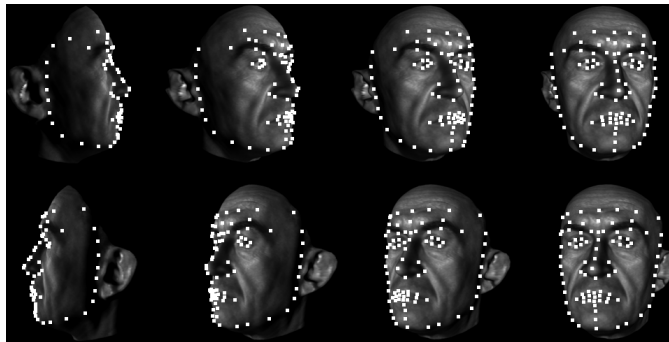
- 1: Compute shape and appearance parameters in the reference feature spaces using Equation (1)- (2).
  - 2: Compute reconstruction weights  $\pi$  at  $p_{\text{th}}$  pose using Equation (3) or Matching Pursuit.
  - 3: For each principal component, minimize energy function  $E$  in Equation (6) to obtain a new manifold.
- 

is arbitrary and highly person-dependent although changes of weights  $\pi$  are smooth. Therefore, it is a reasonable approximation to apply the same weights at other poses.

## 4 Experiments

### 4.1 Database

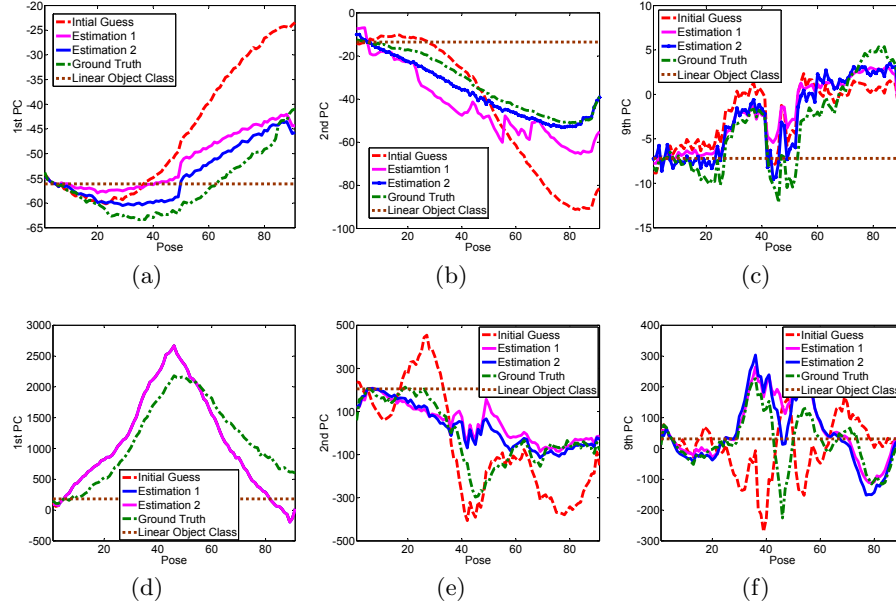
Face images in the database are generated semi-automatically based on 25 3D face models in [11]. These 25 3D models include persons from different ages, races, and genders. On each 3D model, 92 feature points are labeled. 2D locations of feature points are computed automatically while faces with rotations are rendered using 3ds Max<sup>®</sup>. Poses are changed horizontally from  $0^\circ$  to  $180^\circ$  with  $2^\circ$  increments. 2D locations of occluded feature points and feature points on face outlines are further refined by pushing them to the face boundaries. A set of examples is shown in Figure 3.



**Fig. 3.** Examples in face database. Each face has 92 feature points.

### 4.2 Results of Manifold Estimation

Figure 4 shows a typical manifold estimation given the pose at 20 degree. Only 20 faces are used in the face database for training currently. Other 5 faces are testing



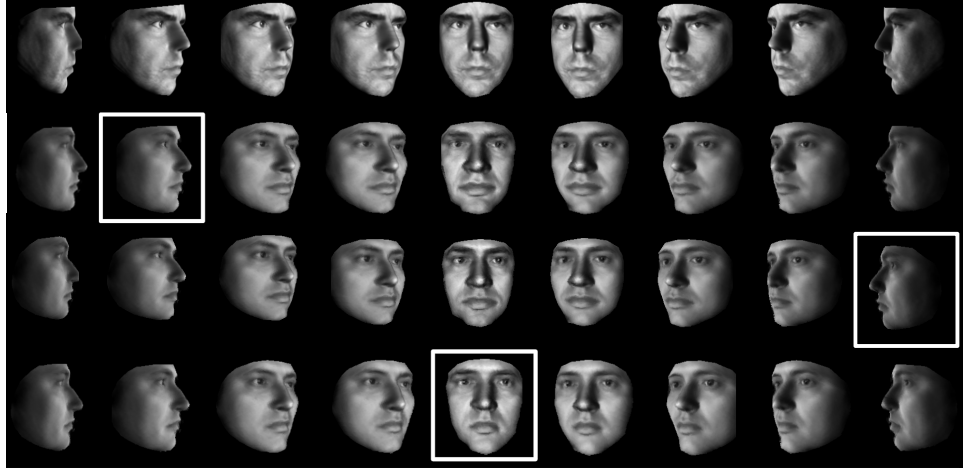
**Fig. 4.** Manifold estimation results from  $0^\circ$  to  $180^\circ$  with an input image at  $20^\circ$ . Estimation 1 is generated using Equation (3). Estimation 2 is generated using Matching Pursuit. (a)  $1_{\text{st}}$  principal component of shape parameters. (b)  $2_{\text{nd}}$  principal component of shape parameters. (c)  $9_{\text{th}}$  principal component of shape parameters. (d)  $1_{\text{st}}$  principal component of appearance parameters. Notice that the initial guess, estimation 1, and estimation 2 are overlapped. (e)  $2_{\text{nd}}$  principal component of appearance parameters. (f)  $9_{\text{th}}$  principal component of appearance parameters.

cases. Nine principal components are used for shape and appearance parameters. Our estimation is close to the ground truth. In [13], the parameters  $\alpha$  and  $\beta$  are fixed assuming a face belongs to a linear object class. It turns out that this is valid within small pose changes (around  $\pm 10$  degrees). The estimation using MP is slightly better than the estimation using Equation (3). The improvements from MP should increase when more faces are used in the database. Notice that a manifold becomes less smooth when the variance of principal component getting smaller, which also contributes less to the face synthesis.

### 4.3 Results of Face Synthesis

Figure 5 shows our results of face synthesis. Our approach is robust to synthesize faces from  $0^\circ$  to  $180^\circ$  given an input at any pose. Furthermore, the appearances of all the synthesized faces are consistent even with input at different poses (as shown in Figure 5, faces are synthesized from inputs at  $20^\circ$ ,  $90^\circ$ , and  $170^\circ$ , respectively).

Figure 6 shows the results from one input from CMU-PIE database. The illumination of input image is normalized roughly such that it is similar to the



**Fig. 5.** Synthesis results. The images with white rectangles are input images. The first row is the ground truth generated from 3D models in [11]. The second row is the synthesis results by given an input at  $20^\circ$ . The third row is generated by given an input at  $170^\circ$ . The fourth row is generated by given an input at  $90^\circ$ .

illumination condition in our database. This also can be done by existing face re-lighting algorithms.

Notice that the input images in Figure 5 and Figure 6 (with white rectangles) are not close to the ground truth since shape and appearance parameters can not be computed very accurately using only 20 faces in the database. We plan to use a larger database in the future.



**Fig. 6.** Synthesis results from one input in CMU-PIE database [12]. The first row is generated by our approach with an input at 90 degrees. The second row is the ground truth.

## 5 Conclusion

In this paper, we propose a novel approach for face synthesis across poses. By applying manifold estimation in the unified view-based feature spaces, this approach is able to synthesize unseen views, even for large pose changes. Moreover, it is straightforward to extend this approach to handle multiple inputs that can improve the estimation accuracy. In the future, we will test and refine our approach on a large face database that should be able to improve the synthesis quality and apply this method in different areas, such as face recognition and 3D face reconstruction.

## References

1. Romdhani, S., Ho, J., Vetter, T., Kriegmann, D.: Face Recognition Using 3D Models: Pose and Illumination. *Proceedings of the IEEE* **94**(11) (2006)
2. Wang, Y., Liu, Z., Hua, G., Wen, Z., Zhang, Z., Samaras, D.: Face Re-Lighting from a Single Image under Harsh Lighting Conditions. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (2007)
3. Luong, Q., Fua, P., Leclerc, Y.: Recovery of reflectances and varying illuminants from multiple views. In: *European Conference on Computer Vision*. (2002)
4. Cootes, T.F., Walker, K., Taylor, C.J.: View-Based Active Appearance Models. In: *IEEE International Conference on Automatic Face and Gesture Recognition*. (2000) 227–232
5. Romdhani, S., Psarrou, A., Gong, S.: On Utilising Template and Feature-Based Correspondence in Multi-view Appearance Models. In: *European Conference on Computer Vision*. (2000) 799–813
6. Blanz, V., Vetter, T.: "face recognition based on fitting a 3d morphable model". *IEEE Transaction on Pattern Analysis and Machine Intelligence* **25**(9) (2003) 1063–1074
7. Gu, L., Kanade, T.: 3D Alignment of Face in a Single Image. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2006)
8. Balasubramanian, V.N., Ye, J., Panchanathan, S.: Biased Manifold Embedding: A Framework for Person-Independent Head Pose Estimation. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (2007) 1–7
9. Fu, Y., Huang, T.S.: Graph Embedded Analysis for Head Pose Estimation. In: *International Conference on Automatic Face and Gesture Recognition*. (2006) 3–8
10. Cootes, T.F., Edwards, G., Taylor, C.J.: Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(6) (2001) 681–685
11. Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces. In: *SIGGRAPH*. (1999) 187–194
12. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression (pie) database of human faces. Technical Report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (January 2001)
13. Vetter, T., Poggio, T.: Linear Object Classes and Image Synthesis From a Single Example Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7) (1997) 733–742
14. Gross, R., Matthews, I., Baker, S.: Fisher Light-Fields for Face Recognition Across Pose and Illumination. (2002)

15. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (1991) 586–591
16. Tenenbaum, J.B., Freeman, W.T.: Separating Style and Content with Bilinear Models. *Neural Computation* **12**(6) (2000) 1247–1283
17. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear Analysis of Image Ensembles: TensorFaces. In: European Conference on Computer Vision. (2002) 447–460
18. Christoudias, C., Darrell, T.: On modelling nonlinear shape-and-texture appearance manifolds. In: IEEE Conference on Computer Vision and Pattern Recognition. (2005)
19. Xiao, J., Baker, S., Matthews, I., Kanade, T.: Real-Time Combined 2D+3D Active Appearance Models. In: IEEE Conference on Computer Vision and Pattern Recognition. (2004)
20. Li, Y., Du, Y., Lin, X.: Kernel-Based Multifactor Analysis for Image Synthesis and Recognition. In: IEEE International Conference on Computer Vision. (2005) 114–119
21. Murase, H., Nayar, S.K.: Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision* **14**(1) (1995) 5–24
22. Graham, D., Allinson, N.: Face recognition from unfamiliar views: subspace methods and pose dependency. In: IEEE International Conference on Automatic Face and Gesture Recognition. (1998)
23. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: Video-based face recognition using probabilistic appearance manifolds. In: IEEE Conference on Computer Vision and Pattern Recognition. (2003)
24. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **290**(5500) (2000) 2319–2323
25. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* **290**(5500) (2000) 2323–2326
26. Zhang, Z., Wang, J.: MLLE: Modified Locally Linear Embedding Using Multiple Weights. In: Neural Information Processing Systems (NIPS). (2006)
27. Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. In: IEEE Transactions on Acoustics, Speech, and Signal Processing. (1993) 3397–3415