A Survey of Different 3D Face Reconstruction Methods

Amin Jourabloo
jourablo@msu.edu
Department of Computer Science and Engineering

Abstract—3D face reconstruction is a very challenging problem in computer vision, and is used as a prerequisite for many important vision tasks, e.g., face recognition and face frontalization. The importance of 3D face reconstruction is that we can estimate illumination and pose information by using the 3D shape of the face. Since the human face is a non-rigid 3D object, estimating the 2D shape of the face is not sufficient because the 2D shape model can make 2D shapes which are not valid projection of a 3D shape. In this report, we review four different methods for 3D face reconstruction which either add constraints of 3D shapes to the 2D shape model or explicitly estimate the 3D shape of the face. The 3D reconstructed face can be used for 3D face recogntion systems and they can achieve a high successful identification rate on different databases e.g., PIE and FERET.

Index Terms—3D face reconstruction, Face recognition, 3DMM, Shape from shading, Structure from motion, Multiview stereo.

1 Introduction

3D face reconstruction is the process of estimating the dense 3D shape of the face from either single or multiple face images of the same person. Improving the accuracy of 3D face reconstructionn is beneficial for many computer vision tasks e.g., face recognition [1] and face frontalization [2].

The importance of 3D face reconstruction is that we can estimate illumination and pose information by using the 3D shape of the face. This information will help 3D face recognition systems to utilize identity information for learning classifiers and discriminating among subjects. In 2D face recognition systems, all of this information are combined and they are analyzed with different statistical learning method by comparing the intensity of pixels of 2D face images.

Although 2D shapes are very powerful for modeling 2D and 3D objects, the 2D shape model can make 2D shapes which are not a valid projection of a 3D shape. Since the human face is a non-rigid 3D object, for achieving acceptable 2D shapes either incorporating constraints of 3D shapes or explicitly estimating the 3D shape of the face is necessary. Hence, in this report, we review two methods which utilize the constraints of 3D shapes for estimating the 3D face:

- Structure from Motion
- Multiview Stereo

and we review two other methods which explicitly estimate the 3D face by adding some assumptions for illumination and estimating pose of the face:

- 3D Morphable Model (3DMM)
- Shape from Shading

The 3D reconstructed face can be used for 3D face recogntion systems and they can achieve a high successful identification rate on different databases e.g., PIE [3] and FERET [4]. It is shown that the 3D face recognition results are comparable with performance of 2D face recognition systems.

In the rest of this report, we review four different methods for 3D face reconstruction. In section 2, we compare 2D and 3D shape models. We review 3D face reconstruction methods in section 3. We compare 2D and 3D face recognition in section 4. In section 5, we mention some challenges of 3D face reconstruction, and, finally, we conclude in section 6.

2 2D and 3D Shape Model

We represent the 2D shape s which contains n 2D feature points as:

$$s = \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ v_1 & v_2 & \dots & v_n \end{pmatrix}. \quad (1)$$

We assume that the 2D shape s is projection of a 3D shape $\bar{s}$ with weak perspective projection $P = [P_p\ P_t]$ assumption

$$\bar{s} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ z_1 & z_2 & \dots & z_n \end{pmatrix}, \tag{2}$$

$$s = P_p \bar{s} + P_t, \tag{3}$$

$$s = \begin{pmatrix} i_x & i_y & i_z \\ j_x & j_y & j_z \end{pmatrix} \bar{s} + \begin{pmatrix} o_x \\ o_y \end{pmatrix}, \tag{4}$$

where $\mathbf{i} = (i_x, i_y, i_z)$ and $\mathbf{j} = (i_x, i_y, i_z)$ are the projection axes with equal length constraint ($\mathbf{i}.\mathbf{i} = \mathbf{j}.\mathbf{j}$) and orthogonality constraint $\mathbf{i}.\mathbf{j} = 0$; and o_x, o_y are translations. A weak prespective projection matrix has six degrees of freedom which are one scale, three rotation angles (pitch, yaw, roll) and two translations.

2.1 Comparing Representation Power 2D and 3D Shape Model

The 2D shape model is very powerful and it can represent 3D shapes. The projection matrix in equation (4) can be decomposed into six matrices which each one convert the 3D shape to a 2D shape. It means that each 3D shape can be represented with combination of six 2D shapes. Therefore, 2D shapes can represent 3D shapes but they need more parameters. Also, 2D shape models can make 2D shapes which are not valid projection of a 3D shape. We can make such 2D shapes by ignoring the orthogonally constraint ($\mathbf{i}.\mathbf{j} = 0$) of projection matrix in equation (4).

3 3D Face Reconstruction

In this section, we review four different approaches for 3D face reconstruction and compare their advantages and disadvantages. A summary of properties of different methods is shown in Table 1.

3.1 3D Morphable Model

The 3D morphable model (3DMM) [5] is a method for 3D reconstruction of the face from a single image. This method consists of two learned vector spaces from a set 3D scans of heads for representing faces with various poses and illuminations. The first vector space represent the 3D shape of the face $\bar{s}$ and the second one represents texture $\bar{t}$ of the face. For learning the 3D shape and texture bases, which describe shape and texture vector spaces, a set of dense point-to-point aligned 3D scans of heads is needed. A modified optical flow algorithm is utilized for aligning the three-dimensional surface of faces to a pre-selected reference 3D face surface. We apply the Principal Component Analysis (PCA) on the set of shapes and textures of aligned 3D scans separately for learning the shape and texture bases. We can represent each input face image as a combination of these 3D shape and texture bases to reconstruct its 3D surface:

$$\bar{s} = \bar{s}_0 + \sum_{i=1}^{m} \alpha_i \bar{s}_i, \quad \bar{t} = \bar{t}_0 + \sum_{i=1}^{m} \beta_i \bar{t}_i, \tag{5}$$

where $\bar{s}_0$ and $\bar{t}_0$ are mean shape and mean texture, α and β are model coefficients, and $\bar{s}_i$ and $\bar{t}_i$ are bases respectively.

For reconstructing 3D surface of the input face image I, we need to estimate model coefficients α, β. First, we use labeled facial feature points F on the input face image for computing a rigid transformation to fit the mean 3D surface on it with the same pose. Then, we use a cost function to estimat the model coefficients, pose, illumination and color parameters. A maximum a posteriori estimator (MAP) with independence assumption is used for making optimization function:

$$P(\alpha, \beta, \rho | I, F) \sim P(I|\alpha, \beta, \rho) P(F|\alpha, \beta, \rho) P(\alpha) P(\beta) P(\rho), \tag{6}$$

where ρ contains pose and illumination parameter. The optimization function is:

$$E = -2 \log P(\alpha, \beta, \rho | I, F), \tag{7}$$

$$E = \frac{1}{\sigma_I^2} E_I + \frac{1}{\sigma_F^2} E_F + \sum_i \frac{\alpha_i^2}{\sigma_{S,i}^2} + \sum_i \frac{\beta_i^2}{\sigma_{T,i}^2} + \sum_i \frac{(\rho - \bar{\rho}_i)^2}{\sigma_{R,i}^2}, \tag{8}$$

where E_I is error between the input image and the reconstructed image, E_F is fitting error of facial feature points, $\sigma_E^2, \sigma_F^2, \sigma_R^2$ are regularizing weights, $\sigma_{S,i}^2, \sigma_{T,i}^2$ are eigen values of the 3D shape and texture bases and $\bar{\rho}$ is an initialization vector for pose and illumination parameters. A stochastic version of Newton's method is used for making an iterative fitting algorithm to minimize (8) and estimating parameters.

The advantage of this method is that it can estimate the dense 3D shape face from a single input image but 3D scans of heads are needed for making 3D shape and texture bases.

3.2 Shape from Shading

The shape from shading is a method for reconstruction 3D surface with shading information from

TABLE 1
The comparison of different face 3D reconstruction methods.

3D Reconstruction Method	Single or Multiple Images	3D shape	2D initial Facial Feature Points	Object	Initialiation	Note
3D Morphable Model	single	dense	Yes	face	Yes (for illumination)	3D scans of head are needed
Shape from Shading	single	dense	Yes	face	No	a 3D scan is needed
Structure from Motion	multiple	sparse	Yes	arbitary	No	tracked data is needed; offline
Multiview Stereo	multiple	dense	No	arbitary	No	camera views are needed

a single image by assuming a Lambertian surface. Light reflected from a Lambertian surface (reflectance function) can be modeled with the second order spherical harmonics approximation as:

$$R\left(\vec{n}(x,y);\rho(x,y),\vec{l}\right) \approx \vec{l}^T\vec{Y}(\vec{n}(x,y)), \tag{9}$$

$$\begin{aligned}\vec{Y}(\vec{n}) = (1, n_x, n_y, n_z, n_xn_y, n_xn_z, n_yn_z,\\ n_x^2 - n_y^2, 3n_z^2 - 1)^T,\end{aligned} \tag{10}$$

where $\vec{l}$ contains coefficients (with nine elements) of the harmonic expansion of lighting, $\vec{Y}$ is the surface spherical harmonic functions and n_x, n_y and n_z are elements of the normal surface vector $\vec{n}$ of surface z:

$$\vec{n} = (n_x, n_y, n_z)^T = \frac{1}{\sqrt{p^2+q^2+1}}(p, q, -1)^T \tag{11}$$

$$\begin{cases} p(x,y) = z(x+1,y) - z(x,y) \\ q(x,y) = z(x,y+1) - z(x,y) \end{cases} \tag{12}$$

where p, q are partial derivatives in discrete space. The Irradiance equation which is used for 3D reconstruction of a surface with the Lambertian surface assumption can be defined as:

$$I(x,y) = \rho(x,y)R(x,y), \tag{13}$$

where $\rho(x,y)$ is the surface albedo.

In [6], the shape from shading method is extended to use a prior knowledge for 3D reconstruction of face. The prior reference face model is used for estimating lighting, pose and it is used as the initial albedo. The optimization function for estimating the surface of the input image I can be expressed as:

$$\min_{z,\rho,\vec{l}} \int (I - \rho\vec{l}^T\vec{Y}(\vec{n}))^2 + \lambda_1\left(\Delta G \star d_z\right)^2 + \lambda_2(\Delta G \star d_p)^2 dxdy, \tag{14}$$

where $\Delta G\star$ is the Laplacian of Gaussian convolution, λ_1 and λ_2 are regularizer weight, d_z is difference of estimated the surface of the input image and the reference model, and similarly d_p is difference of the albedo. The second and the third terms regularize the difference between the surface of the input image and the reference model which improve precision of the method.

An iterative optimization algorithm is used for reconstructing the 3D surface of the input face image. At the beginning, a rigid transformation is calculated from location of labeled facial feature points for aligning the reference model with the same pose of the input image. In the first step of optimization algorithm, we estimate the lighting coefficients $\vec{l}$ of the input image with substituting the normal surface vectors and the albedo of the reference model in the $\vec{Y}(\vec{n})$ and ρ of (14). In the second step, the depth z of the input face image is estimate by using the estimated lighting coefficients and the albedo of the reference model. The optimization of this step is linear with assuming the first order of approximation of reflectance and it is nonlinear with the second order approximation. In the third step, the albedo ρ of the input face image is estimated by using the Irradiance equation:

$$I(x,y) = \rho(x,y)\vec{l}^T\vec{Y}(\vec{n}). \tag{15}$$

These three steps are repeated until the changes are negligible. The advantages of this method are that it neither needs a set of aligned dense 3D shapes nor needs to use the symmetric portions of the face.

3.3 Structure from Motion

The structure from motion [7] is a method for reconstructing 3D shape of an arbitrary object which is tracked during several frames. This method estimates non-rigid shape, motion and shape bases. It utilizes both the rotation constraints which are the orthogonally constraints on the camera rotation, and the basis constraints which implicitly determine a unique set of 3D bases for a non-rigid shape. The factorization techniques can be used for this method. We consider the 2D location of n feature points in N frames and we stack the 2D points to a matrix W:

$$W = \begin{pmatrix} u_1^0 & u_2^0 & \dots & u_n^0 \\ v_1^0 & v_2^0 & \dots & v_n^0 \\ \vdots & \vdots & \vdots & \vdots \\ u_1^N & u_2^N & \dots & u_n^N \\ v_1^N & v_2^N & \dots & v_n^N \end{pmatrix} \tag{16}$$

The W matrix can be factorized to $M \in \mathbb{R}^{2(N+1)\times 3(\bar{m}+1)}$ matrix which contains scaled projection matrices and $B \in \mathbb{R}^{3(\bar{m}+1)\times n}$ which contains 3D shape bases, $W = MB =$

$$\begin{pmatrix} P^0 & \bar{\alpha}_1^0 P^0 & \dots & \bar{\alpha}_{\bar{m}}^0 P^0 \\ P^1 & \bar{\alpha}_1^1 P^1 & \dots & \bar{\alpha}_{\bar{m}}^1 P^1 \\ \vdots & \vdots & \vdots & \vdots \\ P^N & \bar{\alpha}_1^N P^N & \dots & \bar{\alpha}_{\bar{m}}^N P^N \end{pmatrix} \begin{pmatrix} \bar{s}_0 \\ \vdots \\ \bar{s}_{\bar{m}} \end{pmatrix} \tag{17}$$

where P^i are camera projection matrices with weak perspective model assumption, $\{\bar{s}_i\}_{i=0}^{\bar{m}}$ are estimated 3D bases of face and $\bar{\alpha}_i$ are coefficients of estimated 3D shape bases.

The factorization can be performed with the Singular Value Decomposition (SVD) on W and after computing the corrective matrix for imposing the rotation and the basis constraints, M and B matrixes can be determined uniquely. The matrix B which contains estimated 3D shape bases can be used for estimating the 3D shape of the object. In the next subsections, we review the active appearance model (AAM) which is a 2D face alignment method and the extention of this method which simultaneously estimate the 2D and 3D shapes of the face.

The disadvantage is that this method estimates a set of sparse person specific 3D shape bases and since it needs the tracked data, it should be performed offline for nonrigid objects.

3.3.1 2D Active Appearance Model (AAM)

The active appearance model [8] is a 2D face alignment method which uses two set of bases for representing shape and apperance of face:

$$s = s_0 + \sum_{i=1}^{m} \alpha_i s_i, \quad A(u) = A_0(u) + \sum_{i=1}^{l} \beta_i A_i(u), \tag{18}$$

The appearance $A(u)$ is represented within the mean shape s_0, and u contains the location of pixels inside the mean shape. We perform a 2D similarity transformation $N(s;\mathbf{q})$, which q contains rotation, traslation and scale parameters, for fitting the estimated appearance $A(u)$ with parameter β to the estimated shape s with shape parameter α. It means that we have a piecewise affine warp form the mean shape s_0 to $N(s;\mathbf{q})$ which is $W(u;\alpha;\mathbf{q})$. The optimization function for fitting a face in the input image I is:

$$\min_{\alpha,\mathbf{q},\beta} \sum_{u\in\bar{s}} \left(A_0(u) + \sum_{i=1}^{l} \beta_i A_i(u) - I(W(u;\alpha;\mathbf{q})) \right)^2. \tag{19}$$

3.3.2 2D+3D Active Appearance Model (AAM)

The 2D AAM method for face alignment can be extended [9] to use the 3D shape bases and estimate the 2D shape which is consistent with the 3D shape of the face. For imposing this extension, we use the structure from motion method to estimate a set of 3D bases and we add the constraints to (19) as:

$$\min_{\alpha,\mathbf{q},P,\bar{\alpha},\beta} \sum_{u\in\bar{s}} \left(A_0(u) + \sum_{i=1}^{l} \beta_i A_i(u) - I(W(u;\alpha;\mathbf{q})) \right)^2 + K \left\| N\left(s_0 + \sum_{i=1}^{m} \alpha_i s_i; q\right) - P\left(\bar{s}_0 + \sum_{i=1}^{\bar{m}} \bar{\alpha}_i \bar{s}_i\right) \right\|^2. \tag{20}$$

where K is a regularizer weight for the basis constraints, P is the projection matrix. The optimization function (20) can estimate 2D shapes which are consistent with the 3D shape bases and it estimates both of the 3D shape and the pose of the input face image.

The disadvantage of this method is that the estimated 3D shape bases are person specific.

3.4 Multiview Stereo

The multiview stereo is a method for reconstructing the 3D surface of an arbitrary object from a set of images which are taken with known camera views. In [10], this method is used for reconstructing the 3D surface of the face in two stages. At the first stage, feature extraction and feature matching are performed for multiple pair of selected face images and the dense disparity map for each pair is computed by using the estimated epipolar geometry. In the second stage, a single 3D point cloud is made by combining the estimated pair wise reconstruction of the face. This 3D point cloud contains outliers which is due to the error in camera positions and other noises. For removing outliers, a tensor voting method [11] can be utilized which determines the accuracy of feature matching. The tensor voting uses first-order and second-order tensors which show the roles of a points in space.

This method can estimate the 3D shape of an arbitrary object but it needs multiple images with known camera views.

4 2D AND 3D FACE RECOGNITION

We review two set of experiments for face recognition by using two different 3D face reconstruction methods.

4.1 3D Face Recognition with 3DMM

In [5], a set of experiments are performed for evaluating 3D face recognition with the 3DMM on two public datasets. The first dataset is PIE dataset [3], from this dataset 4488 face images of 68 subject with three viewpoints and 22 different illuminations are selected. The second one is FERET dataset [4], from this dataset 1940 face images of 194 subjects are selected for evaluation. For each face image, the shape parameter α and the texture parameters β are extracted by using the optimization method (8). The estimated parameters are normalized with their standard deviations and each face is represented with a vector $c = [\alpha\ \beta]$. Three similarity measures are used for comparing two faces c_1 and c_2, Mahalanobis distance d_M, cosine of the angle between two vectors d_A and another similarity measure which considers variation of coefficients from different images of the same person:

$$d_W = \frac{\langle c_1, c_2 \rangle_W}{\|c_1\|_W \cdot \|c_2\|_W}, \tag{21}$$

where $\langle c_1, c_2 \rangle_W = \langle c_1, C_W^{-1} c_2 \rangle$ and C_W is the covariance matrix.

One frontal view face image per person is used for gallery set and the rest of images are used for the probe set. The overall percentage of identifications for the two dataset are shown in Table 2.

TABLE 2
The comparison of successful identifications on PIE and FERET datasets with the 3DMM method.

Database	d_M	d_A	d_W
PIE	87.2%	94.2%	95.0%
FERET	80.3%	92.2%	95.9%

According to Table 2, the 3DMM method achieved overall 95% successful identifications on both datasets. The performance of 3DMM method for the verification task on the PIE dataset is 77.5 percent hit rate and on the FERET dataset is 87.9 percent hit rate at 1 percent false alarm rate.

4.2 2D and 3D Face Recognition with Multiview Stereo

Medioni et al. [10] compared 2D-2D and 3D-3D face recognition performances in their proposed system

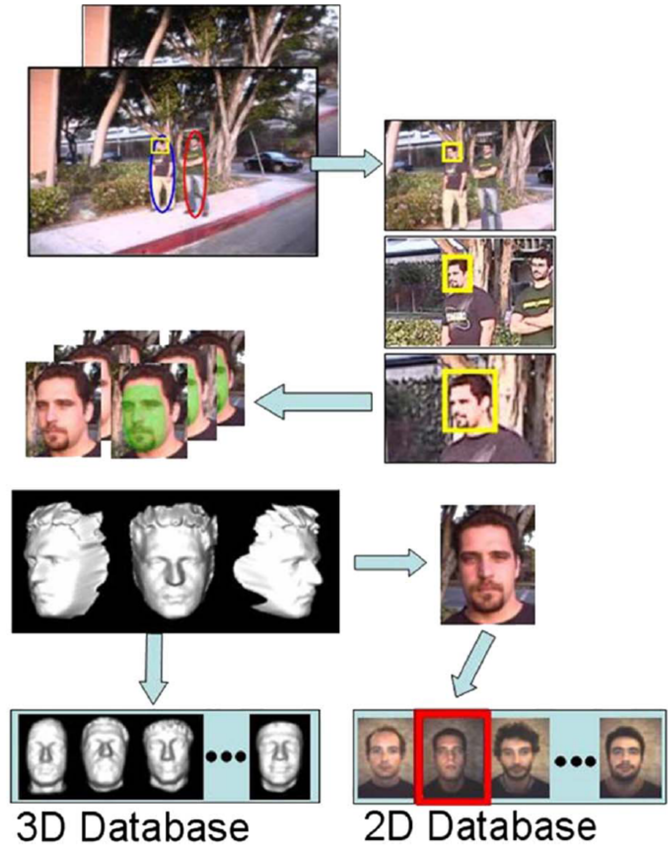

Fig. 1. The overall process of proposed 3D face recognition method in [10].

for 3D face reconstruction at a distance from the camera. In their system a fixed ulterahigh-resolution camera is used for capturing video of a person from a distance. Then, the person and the location of the face are detected. The region of interest is extracted from the face images and a 3D dense model face is generated from the sequence face images. The dense reconstructed 3D face shape is used for 2D and 3D face recognition. The overall process of 3D face recognition is shown in Fig. 1.

A 3D gallery set of 358 3D face model from 358 subjects is used for the experiments. The 3D face models are reconstructed with a stereo cameras. The probe set contains 23 facial sequences of 23 different subjects with different distance from the camera (3, 6 and 9 m). The multiview stereo method is used for performing 3D face reconstruction for each subject. The Geometrix software development kit (SDK) is used for computing the distance between the probe face p and the gallery set $G = \{G(1), G(2), \ldots, G(358)\}$. The Geometrix SDK is based on the registration of two 3D facial shapes which is performed by iterative closest point technique for finding point-to-point correspondence between the two shapes. The identity is assigned to each probe face by

$$ID = arg \min_{i} d\left(s(p), G(i)\right), \tag{22}$$

where $d()$ is the geometrix SDK distance and $s()$ is a scaling function for normalization. The distance between the probe face and the gallery faces are computed with different scales and a identity which has minimum distance is assigned to the probe face.

For 2D-2D face recognition, 358 2D images of the same subjects of the 3D gallery are used for 2D gallery set and the 2D images of 23 facial sequences are used for probe set. Table 3 shows the verification rates of 2D-2D and 3D-3D face recognition at 1 percent false alarm rate with different distance to the camera.

TABLE 3
The comparison of verification percentage with multiview stereo method.

Distance	2D-2D	3D-3D
3m	95.8%	87.0%
6m	94.0%	78.3%
9m	92.3%	69.6%

Both 2D-2D and 3D-3D face recognitions achieved acceptable results. The 2D face recognition (95.8%) performs better than 3D face recognition (87.0%), and the 3D face recognition is more sensitive to the distance to the camera. For comparing with the best 3D reconstruction method as a reference, the ground truth 3D scans are used for 3D face recognition and the result is 99% at 1 percent false alarm rate. This experiment shows that the quality of reconstructed 3D face is very important for face recognition performance.

5 Challenging problems in 3D face Reconstruction

For a 3D face reconstruction system, it is necessary to reconstruct faces with different ages, ethnic groups and facial expressions. To alleviate this problem, we should have bigger 2D and 3D face databases which cover all of the mentioned variation of faces. Also, the system should be robust to the different illuminations and the occlusions made by glasses, beards and hairs, and the occlusion made by pose of the face (self-occlusion). For making the process of 3D reconstruction fully automatic and amending it to be invariant to occlusion, we can use automated initialization and face alignment methods, which are robust to occlusion and pose of the face.

Another challenge is that the most of 3D reconstruction systems assume weak perspective projection model, it is interesting to change this assumption to full perspective projection model.

6 Conclusion

In this report, we reviewed four different 3D face reconstruction methods. The 3DMM and the shape from shading methods explicitly estimate the dense 3D shape of the face and the illumination parameters from a single image but the structure from motion and the multiview stereo methods impose the constraints of 3D shapes to 2D shapes for 3D face reconstruction from multiple face images. The face recognition performances in section 4 show that the quality of 3D reconstructed face is very important and higher performance can be achieved by enhancing 3D face reconstruction methods for uncontrolled environments.

References

[1] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3d face recognition under expressions, occlusions, and pose variations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 9, pp. 2270–2283, 2013.

[2] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 787–796.

[3] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 46–51.

[4] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," *Image and vision computing*, vol. 16, no. 5, pp. 295–306, 1998.

[5] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1063–1074, 2003.

[6] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 2, pp. 394–405, 2011.

[7] J. Xiao, J.-x. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," in *Computer Vision-ECCV 2004*. Springer, 2004, pp. 573–587.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 681–685, 2001.

[9] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-time combined 2d+ 3d active appearance models," in *CVPR (2)*, 2004, pp. 535–542.

[10] G. Medioni, J. Choi, C.-H. Kuo, and D. Fidaleo, "Identifying noncooperative subjects at a distance using face images and inferred three-dimensional face models," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 39, no. 1, pp. 12–24, 2009.

[11] P. Mordohai and G. Medioni, "Perceptual grouping for multiple view stereo using tensor voting," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 3. IEEE, 2002, pp. 639–644.