In Between 3D Active Appearance Models and 3D Morphable Models

Jingu Heo and Marios Savvides
Biometrics Lab, CyLab
Carnegie Mellon University
Pittsburgh, PA 15213
jheo@cmu.edu, msavvid@ri.cmu.edu

Abstract

In this paper we propose a novel method of generating 3D Morphable Models (3DMMs) from 2D images. We develop algorithms of 3D face reconstruction from a sparse set of points acquired from 2D images. In order to establish correspondence between images precisely, we Combined Active Shape Models (ASMs) and Active Appearance Models (AAMs)(CASAAMs) in an intelligent way, showing improved performance on pixel-level accuracy and generalization to unseen faces. The CASAAMs are applied to the images of different views of the same person to extract facial shapes across pose. These 2D shapes are combined for reconstructing a sparse 3D model. The point density of the model is increased by the Loop subdivision method, which generates new vertices by a weighted sum of the existing vertices. Then, the depth of the dense 3D model is modified with an average 3D depth-map in order to preserve facial structure more realistically. Finally, all 249 3D models with expression changes are combined to generate a 3DMM for a compact representation. The first session of the Multi-PIE database, consisting of 249 persons with expression and illumination changes, is used for the modeling. Unlike typical 3DMMs, our model can generate 3D human faces more realistically and efficiently (2-3 seconds on P4 machine) under diverse illumination conditions.

I. Introduction

Modeling 3D faces is one of the most difficult and challenging tasks because of large variations occurred in human faces. Due to the rapid growth of 3D sensing techniques, the acquisition of 3D face images becomes easier with reasonable quality. These dense 3D geometry of human faces can be used for face recognition [5], showing the potential to overcome the conventional problems of 2D face recognition, such as pose, and illumination. However, without relying on 3D sensors, estimating 3D facial shape from single or multiple images is one of the fundamental problems of computer vision. Research fields include shape-from-shading (SFS), shape-from-stereo, shape-from-motion (SFM), and shape-from-texture. Since the earliest SFS attempted by Horn [17], numerous contributions have been made to recover shape from a gradual variations of shading in images [11] [26]. However, realistic face modeling is still difficult and the quality of the 3D faces, generated from these methods, is often unsatisfactory. In order to generate realistic 3D faces from one or several 2D images, it is often necessary to use prior knowledge of a statistical 3D face model [1] [2]. 3D Morphable Model(3DMMs) [3] have become popular in a variety of research areas including computer graphics, human computer interaction and face recognition. However, these methods are computational expensive and need manual control points. More efficient, fully automatic, realistic, and robust face modeling is desired in numerous applications, such as access control, entertainment, and video conferences, which require less user's cooperation and real-time computation. Although there are several approaches to modeling 3D face efficiently [23] [12], most of the methods require user's manual interaction, camera calibration, and strong prior knowledge of a 3D model, which might have a limited ability to model any face under various conditions.

In this paper, we address the problem of generating more realistic, efficient and robust 3D models from 2D images in a fully automated way. No prior knowledge on 3D models is required except an average 3D shape. Given a set of 2D images, taken by multiple cameras or a sequence of images taken by a single camera, our goal is to generate a smooth 3D face. To achieve this, it is important to identify three challenging problems. Our algorithms are designed to solve each sub-problem accordingly. First, automatic facial alignment is the most difficult. Since SFM

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

requires precise correspondence points in order to generate model accurately, an automated and accurate landmark point detector (79 points) is designed to solve this problem. Although it is desirable to establish dense correspondence between 2D images, feature detection in a detailed level is a much harder problem. Second, the reconstructed shape has a set of sparse points due to the feature detection problem in 2D images. Dense surface modeling is required in order to generate a smooth human face. The density of point is increased by using the Loop subdivision method, which generates new vertices by a weighted sum of the existing vertices. Finally, generating realistic 3DMMs is a challenging problem. The dense surface needs correction to preserve facial structure. An average shape, acquired from the USF Human-ID database [3], is used to reflect facial structure more realistically.

We use a gigantic 2D image database([14]) for finding shapes in 2D images and reconstructing sparse 3D faces. Then, a 3DMM, which reflects more diverse facial variations, is modeled. This paper is organized in the following manner. Section II contains a brief review of related work. In Section III, we propose Combining Active Shape Models (ASMs) and Active Appearance Models (AAMs)(CASAAMs), which improve fitting on unseen faces as well as handling illumination, pose, and expression changes. Section IV explains the SFM method for 3D reconstruction. In Section VI, the generation of the 3DMM is addressed. Finally, we conclude our proposed work.

II. Background

Model-based approaches to analyzing shapes and appearances are popular for face modeling. Active Contours (Snakes) [18], Active Shape Models [7], Active Appearance Models [6], and 3D Morphable Models(3DMMs) have been proven to be effective tools for representing shapes or appearances in a compact form. ASMs have been shown to be an effective way for analyzing the shape of objects. This shape analysis process is known as the Point Distribution Model (PDM), which is a method for representing with the mean geometry of a shape and some statistical modes of geometric variation inferred from a training set. By utilizing the shape model, ASMs iteratively deform to fit an example of shapes in new images. ASMs use local appearance models for updating shapes, while AAMs use a global appearance model, although both use the same shape model. An AAM is compromised of a shape model and an shape-free appearance (pixel intensity) model. A compact representation of training images and new images can be achieved through training and model fitting. AAMs interpret new images by minimizing the error between input images and reconstructed images using the model instance. Although ASMs and AAMs are widely used for localizing shapes, the fitting accuracy needs improvement for unseen faces, which are not shown during the training stage. We will detail this problem in Section III.

Once facial feature points are established across images for the same person, a 3D face can be generated from SFM. SFM can reconstruct 3D faces defined by camera models. Orthographic, affine, weak perspective, perspective camera models are typically used in practice. For 3D faces, we assume orthographic or weak-perspective models are valid due to their efficiency. The goal of 3D reconstruction is to estimate 3D shape from a set of image point correspondences between images defined by these camera models, by minimizing the geometric error in image coordinate measurements. The Factorization algorithm achieves a Maximum Likelihood (ML) affine reconstruction under isotropic zero-mean Gaussian noise [4]. Bundle adjustment can be used for the refinement of the reconstruction which provides a true ML estimate [16].

The reconstruction based on SFM is often too sparse, since it is difficult to establish dense correspondence between images. Although corners or distinct features [21] can exist in images, typically only a few of points are well localized by current existing methods. Therefore, it is desirable to increase the density of point directly from the sparse 3D faces. Since it is widely used to represent 3D faces by triangle meshes, the density of point can be easily increased by operating on the vertices of each triangle. This scheme is known as Subdivision. Subdivision is a method in computer graphics for the approximation of a smooth surface given a set of sparse triangle meshes. Loop [20] triangle splines generalized the recurrent relations for box splines to irregular meshes. Boor et al. proposed that a smooth surface can be defined using box splines for regular meshes [10]. However, these points, generated by the subdivision methods, often do not preserve facial structure realistically. An average dense shape model, acquired from 3D scanners, is needed to correct these points. In this way, 3D faces can be synthesized by the observed depth information and prototypical facial structure at the same time. These individually generated 3D faces are combined to form a Morphable Model.

AAMs and 3DMMs are quite similar in many respects. The representation space of an AAM is 2D whereas that of an 3DMM is 3D. 3D Active Appearance Models (3DAAMs) can achieve similar representations that a 3DMM can. However, 3DAAMS have difficulty in representing appearance changes due to their sparse 3D shape representations. Only 3D shape information is typically used for constraining 2D AAM shape vectors, presented in a combined 2D+3D AAM approach [24]. Since the appearance of a 3DMM is defined per each 3D vertex (point), it allows to understand image formulation of faces under

various lighting and pose variations. It may be effective to use 2D images for modeling diverse appearance changes and utilize the dense shape information at the same time. We achieve this by mapping 2D frontal images on the 3D face for each person, so that texture values are also defined in 3D. Therefore, our approach lies in between 3DAAMs and 3DMMs, which can model various appearance changes with dense 3D shape changes.

III. CASAAMs

From the developments of AAMs and ASMs, we have noticed that it is the major concern of researchers [9] to find relationships among shape, texture and residual error. Although AAMs are computationally more demanding than ASMs, AAMs often tend to fit more accurately than ASMs if test images are similarly taken with a training set. If test images are deviated from training images, AAMs tend to fit poorly. This problem is addressed by several researchers [13] [19]. It is difficult to verify that which algorithms outperform others under various conditions, since the previous evaluations on ASMs and AAMs use a small database under small variations [8]. Unlike conventional methods for focusing on finding relationships for modeling, we emphasize on finding experts for dealing with different image conditions. ASMs and AAMs are compared under the same conditions, especially focusing on the generalization power with a challenging database (MPIE [14]). Only the first session is used for the experiment out of four sessions with totaling 149,400, consisting of 249 persons with 2 expressions and 15 poses under 20 different illumination conditions. View based approaches are used to identify the number of necessary views for both models. For each view, randomly chosen 5 images of the first 100 persons (5,000) are trained, all images of the rest of 149 persons are used for testing. Figure 1 shows the performance comparison of the ASMs and AAMs for testing. The ASMs have better pixel accuracy over the AAMs. Similar results are obtained for profile views. We also evaluated global ASMs vs. view based ASMs and global AAMs vs. view based AAMs by changing the number of training images. Although the global ASM is trained with 5000 images by randomly sampling different images under pose changes, no significance performance degradation is observed. Therefore, the number of the view-specific ASMs can be reduced. Only two ASMs (Frontal, Profile) are enough for handling all views. One profile ASM can handle left or right profile faces if we assume faces are symmetric. However, if the fitting accuracy is a concern, each view based ASM is preferred. On the other hand, the AAMs need more view AAMs. The global AAM cannot handle more views effectively, showing that it has difficulty in modeling appearances associated with non-linearities

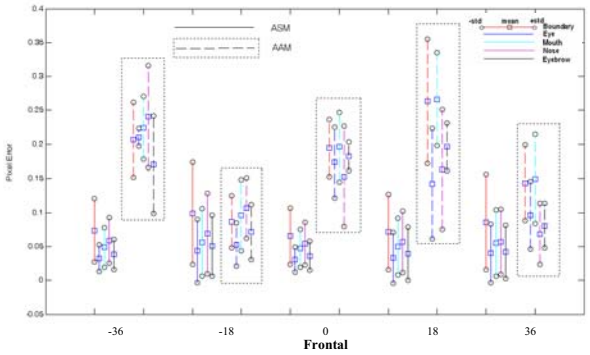

Fig. 1. The ASMs vs. AAMs (dashed rectangle). The ASMs achieve better performance on fitting over the AAMs. 5 facial features, such as boundary, eyes, mouth, nose, and eyebrows are evaluated separately with the mean and standard deviation respectively.

caused by pose changes.

In summary, the ASMs achieved a better generalization ability to unseen people, need less view based models and small training size compared to the AAMs. However, the AAMs can assist the ASMs since typical ASMs lack the ability to verify the fitting performance.

Figure 2 shows an overview of the proposed approach for combining both ASMs and AAMs. By using the ASMs as an initial shape, a corresponding view-based AAM is activated based on estimated poses.

The global ASM is considered an expert on locating shapes for handing view changes with less computational burden. Based on the returned shape, the goal to activate the corresponding view based AAM. Figure 3 shows the performance comparison of the CASAAMS and AAMs, in terms of pixel accuracy for unseen people in the system. Generally, the AAMs are preferred since a precise positioning of the landmark points is important for various applications, such as expression analysis and pose correction. It has been reported that AAMs need good initialization in order to converge into a good solution. If the initial shape is far deviated from the target shape, the solution tends to be stuck in a local minimum. This motivates us to develop a combined approach since the

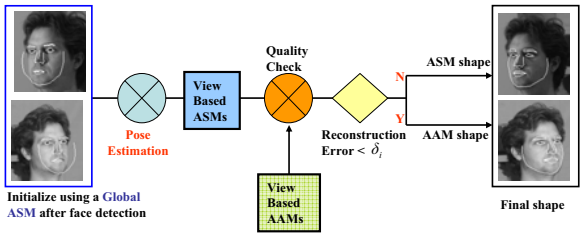

Fig. 2. The proposed work for combining the ASMs and AAMs. The Global ASM is used for better shape initialization for view based ASMs, then the AAMs are applied for final fitting.

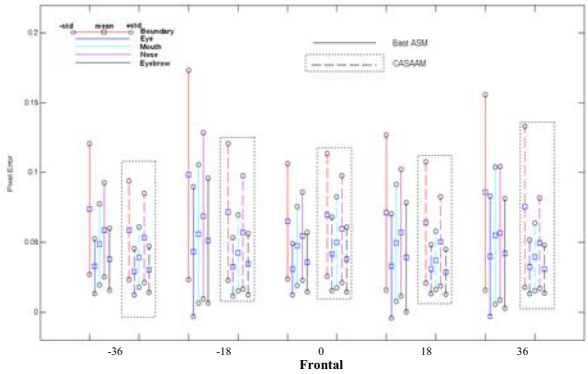

Fig. 3. Performance comparison of the CASAAMs (dashed rectangle) vs ASMs; The ASMs are already proven to be effective enough. However, AAMs can improve the fitting and reduce the variance.

shapes, returned by the ASMs, are appropriate for the initialization of the shapes for the AAMs. By removing unnecessary update steps while focusing on the spaces where the optimal solutions reside, the combined approach may improve the AAMs fitting accuracy as well.

IV. Rapid 3D Reconstruction

It is well known that the 3D reconstruction problem is much easier and reliable with robust feature detectors in 2D images. An accurate and real-time facial alignment method is introduced to solve this problem in the previous section. In this section, we emphasize on how to utilize 2D observations for 3D reconstruction. A 2D or 3D shape $(\mathbf{S})$ is denoted by a linear combination of the shape base vectors $(\mathbf{S}_i)$ through Principal Component Analysis (PCA):

$$\mathbf{S} = \bar{\mathbf{S}} + \sum_{i=1}^{m} p_i \mathbf{S}_i \tag{1}$$

where the coefficients p_i are the shape parameters. We use $\mathbf{S_{2xn}}$ for 2D and $\mathbf{S_{3xn}}$ for 3D, where n indicates the number of vertices. In this paper, we use $n = 79$ and all points are seen from 3D to 2D projection, thus eliminate the visibility problems [15].

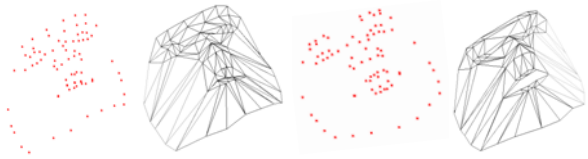

Fig. 4. The reconstructed sparse 3D shapes (in points) and their triangle meshes. Rigid and non-rigid 3D shapes are recovered from our approach. 5 images of the same person under pose variations with the same expression are used for each reconstruction. Total 249 3D sparse faces are reconstructed with expressions changes and these models are used for generating 3D shape base vectors for handing a variety of shape changes.

We define the 2D shape matrix $\mathbf{S_{2xn}}$ as the 2D coordinates (x, y) of the n vertices:

$$\mathbf{S_{2xn}} = \begin{pmatrix} x_1 & x_2 & \dots x_n \\ y_1 & y_2 & \dots y_n \end{pmatrix}. \tag{2}$$

Similarly, the 3D shape matrix can be represented by the 3D coordinates (x, y, z):

$$\mathbf{S_{3xn}} = \begin{pmatrix} x_1 & x_2 & \dots x_n \\ y_1 & y_2 & \dots y_n \\ z_1 & z_2 & \dots z_n \end{pmatrix}. \tag{3}$$

The 3D reconstruction problem from multiple 2D observations is to recover the 3D shape information under noises. The 2D shape is an instance of a projection of 3D with the 2D translation vector $\mathbf{t}$. This can be represented by:

$$\mathbf{s}_{2d} = \mathbf{P}\mathbf{s}_{3d} + \mathbf{t} \tag{4}$$

where $\mathbf{s}_{2d} = (x, y)^T$ and $\mathbf{s}_{3d} = (x, y, z)^T$ indicate each point in 2D and 3D respectively. For multiple(i) camera observations and multiple points (n), the goal of reconstruction is to minimize the overall error

$$\underset{\mathbf{P}^i, \mathbf{t}^i, \mathbf{s}^i_{3dn}}{\arg\min} \sum_{n,i=1} ||\mathbf{s}^i_{2dn} - (\mathbf{P}^i \mathbf{s}_{3dn} + \mathbf{t}^i)||^2. \tag{5}$$

The factorization algorithm achieves the above minimization under Gaussian noise. The following measurement matrix $\mathbf{W}$ can be obtained by stacking the 2D observations $(\mathbf{S_{2xn}})$:

$$\mathbf{W} = \begin{pmatrix} x^1_1 & x^1_2 & \dots x^1_n \\ y^1_1 & y^1_2 & \dots y^1_n \\ \vdots & \vdots & \ddots \vdots \\ x^i_1 & x^i_2 & \dots x^i_n \\ y^i_1 & y^i_2 & \dots y^i_n \end{pmatrix} \tag{6}$$

where each point is represented by object centered coordinate systems. The factorization algorithm can estimate $\mathbf{S_{3xn}}$ by decomposing $\mathbf{W} = \widetilde{\mathbf{P}}_{2i\mathbf{x}3}\widetilde{\mathbf{S}}_{\mathbf{3xn}}$ through Singular Value Decomposition (SVD), since the rank of $\mathbf{W}$ is at most 3. In addition, metric upgrade needs to be performed with additional constraints:

$$\mathbf{P}_{2i\mathbf{x}3} = \widetilde{\mathbf{P}}\mathbf{G}, \mathbf{S_{3xn}} = \mathbf{G}^{-1}\widetilde{\mathbf{S}} \tag{7}$$

where $\mathbf{G}$ is a matrix for correcting the shape into an Euclidean structure since the reconstruction is defined up-to any arbitrary affine transformations. There are a few non-rigid structure-from-motion algorithms [25] [22] for tracking purposes. However, given a fixed pose with multiple cameras, the orthogonality or weak-perspective reconstruction is a still efficient and valid camera model. Furthermore, in our cases, it is desirable to reconstruct 3D shapes from rigid and non-rigid separately, then later

Fig. 6. Odd row images are original MPIE images under different poses. Even row images are the corresponding poses generated from our 3D modeling approach.

Fig. 5. The reconstructed 3D faces from our proposed approach.

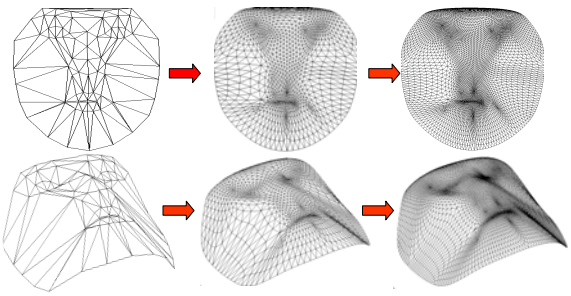

Fig. 7. The density of points is increased by the Loop subdivision method. At each iteration, new points are generated from the sparse 3D shapes (left) to more denser shapes (right).

these several models are combined to perform non-rigid behaviors.

However, the dense surface, after applying subdivision, needs correction to preserve facial structure. An average shape, acquired from the USF Human-ID database [3], is used to reflect facial structure more realistically. Then, the depth of the dense 3D model is modified with an average 3D depth-map in order to preserve facial structure more realistically. Figure 5 shows the reconstructed 3D faces. The second and third column images indicate the rotated images from the images in the same row in the first column respectively. Figure 6 shows the comparison between original MPIE images and synthesized images under different pose changes.

V. In Between 3DAAMs and 3DMMs

From the previous section, we focused on 3D reconstruction from a precise set of correspondence. In this section, we describe how to combine all reconstructed 3D shapes in order to represent compactly and allow non-rigid shape changes. Since each shape needs to be aligned to compute the basis, a modified version of 3D procrustes analysis is used to align different shapes under translation, rotation, and scale transformations. This aligning process is the most important step for representing all trained 3D

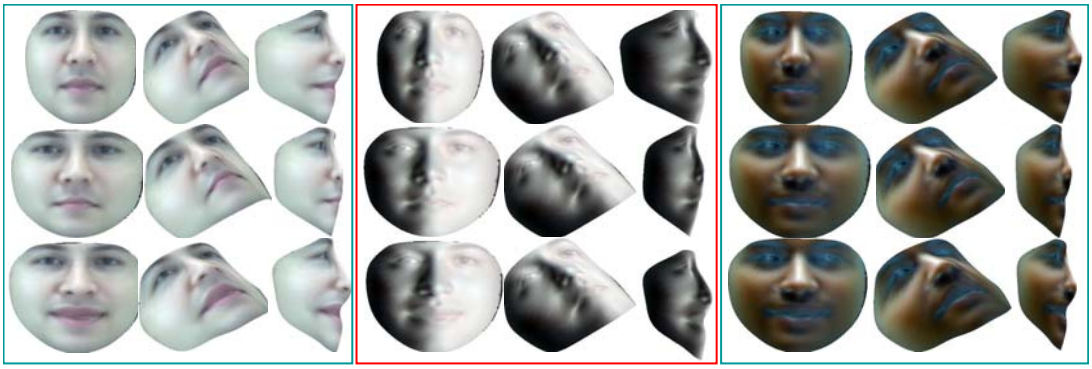

Fig. 8. The 3D shape and 3D texture base vectors. The upper-left image for each rectangle box indicates three largest appearance base vectors. The first images, in the second and third rows for each box, are obtained by changing the shape basis. The rotated images of the base images are shown in the second and third columns according to the first image of each box respectively. From a frontal image and a sparse 3D shape, new images under different poses and illumination can be synthesized from these base vectors. Therefore, we can achieve more compact representations of the MPIE database without modeling of non-frontal images.

shapes compactly. The translational components can be simply removed by subtracting the mean of a 3D face from all the points so that the mean of all the points lies at the origin. The scale component can be removed by dividing by the norm of all the points. In addition, we use 4 feature points (the eyes (2), the center of the forehead (1), and the upper center of the lip (1)) for performing the rotation alignment. This allows us to align the 3D shapes along the the x, y, and z axis respectively. Once the 3D shapes are aligned, each shape can be represented by a linear combination of the base vectors from the eq. 1.

Since the reconstructed 3D shapes are too sparse, dense surface modeling is a necessary step in order to generate a smooth human face. These dense shapes are obtained by increasing the density of points by using the Loop subdivision method. The main idea of the Loop subdivision is to add more control points iteratively. It subdivides each triangle into smaller triangles for the better approximation of the smooth surface based on a set of rules. An illustration of the subdivision method is shown in Figure 7. See [20] for details.

In order to model diverse appearance changes, we use frontal images under different illumination conditions per each person. Then each image is being warped onto the 2D mean shape. Then, an appearance image can be represented by the mean and a linear combination of the base vector through PCA:

$$\mathbf{A}(\bar{\mathbf{x}}) = \bar{\mathbf{A}}(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i \mathbf{A}_i(\bar{\mathbf{x}}). \quad (8)$$

where $\mathbf{A}(\bar{x})$ is the shape free image, warped onto the mean shape $\bar{\mathbf{x}}$ and $\bar{\mathbf{A}}(\bar{\mathbf{x}})$ is the mean appearance image. λ_i are the appearance parameters. Finally, each 2D texture base vector is texture-mapped onto the basis of 3D shapes after applying the subdivision method. This representation is typically used in 3DMMs, while 3DAAMs have difficulty in modeling appearance changes. Figure 8 shows three largest eigenvectors (including the mean appearance) for each 3D shape and 3D texture, where the 2D texture basis are mapped onto the 3D shape basis. By allowing more diverse changes in appearance, the 3DMM, generated from our approach, can be used for understanding of face image formulation.

VI. Conclusion

In this paper, we demonstrate 3D human face reconstruction from multiple images. The CASAAMs are designed in order to establish precise correspondence between 2D images automatically. The proposed 3DMM method, generated from 2D images, can model more diverse appearance changes. Therefore, we claim that our approach lies in between 3DAAMs and 3DMMs since we can model various appearance changes with dense 3D shape changes. The computation time for our proposed 3D face method, together with feature detection, takes only 2-3 seconds, while the standard technology (3DMMs) takes 4-5 minutes with manual feature annotation.

References

[1] J. Atick, P. Griffin, and A. Redlich. Statistical approach to shape from shading: Reconstruction of 3d face surfaces from single 2d images. *Computation in Neurological Systems*, 7(1), 1996.

[2] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illuminations with a 3d morphable model. *Proc. Fifth Int'l Conf. on Automatic Face and Gesture Recognition*, pages 202–207, 2002.

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

[4] C. T. C and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int. Journal of Computer Vision*, 9(2):137–154, 1992.

[5] K. Chang, K. Bowyer, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 2d+3d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.

[6] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *roc. of the European Conf. on Computer Vision*, 2:484–498, 1998.

[7] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models: Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[8] T. F. Cootes, G. Edwards, and C. J. Taylor. Comparing active shape models with active appearance models. *British Machine Vision Conf.*, 1:173–182, 1999.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor. A comparative evaluation of active appearance model algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2:680–689, 1998.

[10] C. de Boor, K. Hollig, and S. D. Riemenschneider. Box splines, 1993. Springer Verlag.

[11] J. D. Durou, M. Falcone, and M. Sagona. survey of numerical methods for shape from shading, 2004. Technical Report 2004-2-R, IRIT.

[12] D. Fidaleo and G. Medioni. Model-assisted 3d face reconstruction from video. *IEEE Intl' Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pages 124–138, 2007.

[13] R. Gross, I. Matthews, , and S. Baker. Appearance-based face recognition and light-fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(4):449–465, 2004.

[14] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition*, 2008.

[15] L. Gu and T. Kanade. 3d alighment of face in a single image. *In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2006.

[16] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. *Cambridge University Press*, 2004.

[17] B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view, 1970. PhD thesis, MIT.

[18] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int'l Conf. on Computer Vision*, pages 259–268, 1987.

[19] X. Liu. Generic face alignment using boosted appearance model. *n Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 2007.

[20] C. T. Loop. Smooth subdivision surfaces based on triangles, 1987. M.S. Thesis, Department of Mathematics, University of Utah.

[21] J. Shi and C. Tomasi. Good features to track. *IEEE Conf. Computer Vision and Pattern Recognition*, 1994.

[22] L. Torresani, A. Hertzmann, and C. Bregler. Learning non-rigid 3d shape from 2d motion. *In: Proc. Neural Information Processing Systems (NIPS)*, 16, 2003.

[23] S. Wang and S. H. Lai. Efficient 3d face reconstruction from a single 2d image by combining statistical and geometrical information. *Asian Conference on Computer Vision*, pages 427–436, 2006.

[24] J. Xiao, S. Baker, I. Matthews, R. Gross, and T. Kanade. Real-time combined 2d+3d active appearance model. *In Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, pages 535–542, 2004.

[25] J. Xiao, J. Chai, and T. Kanade. Closed-form solution to non-rigid shape and motion recovery. *In Proc. European Conference on Computer Vision*, pages 668–675, 2004.

[26] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.