# 3D Face Reconstruction Using A Single or Multiple Views

Jongmoo Choi, Gérard Medioni, Yuping Lin
*University of Southern California - USA*
{*jongmoo, medioni, yuping*}*@usc.edu*

Luciano Silva, Olga Bellon, Mauricio Pamplona
*Universidade Federal do Parana - Brazil*
{*luciano, olga, pamplona*}*@ufpr.br*

Timothy C. Faltemier
*Progeny Systems Corporation*
*tfaltemier@progeny.net*

## Abstract

*We present a 3D face reconstruction system that takes as input either only one single view or several different views. Given a facial image, we first classify the facial pose into one of five predefined poses, and then detect two anchor points that are then used to detect a set of predefined facial landmarks. Based on these initial steps, for a single view we apply a warping process using a generic 3D face model to build a 3D face. For multiple views, we apply sparse bundle adjustment to reconstruct 3D landmarks which are used to deform the generic 3D face model. Experimental results on the Color FERET database confirm our framework is effective in creating realistic 3D face models that can be used in many computer vision applications, such as 3D face recognition at a distance.*

## 1. Introduction

We propose a method to generate realistic 3D faces using a single or multiple facial images. Structure-from-motion techniques for multiple facial images do not apply directly [1]. Accurate estimation of head-camera motion depends on a number of accurate correspondences. However, outliers that we cannot avoid in low-resolution facial images may lead to large errors. Since the accurate head-camera motion is a prerequisite for accurate 3D reconstruction, building a reasonable 3D face model from low-resolution images is a very difficult problem. Statistical model based methods [2, 3] have been applied for single or multiple images, which require a large training database, a critical initialization step and many parameters needed to be tuned carefully.

Our framework generates a 3D face efficiently using facial features. We first classify the facial pose into one of five predefined poses, and then detect two anchor points that are then used to detect facial landmarks. Based on the detected set of landmarks, we apply a warping process to a generic 3D face model, in order to build a 3D face for a given single view. In case of multiple views, we use sparse bundle adjustment (SBA) [4] to reconstruct 3D landmarks from a set of 2D landmarks. The generic 3D face model is deformed by the reconstructed 3D landmarks and pre-defined 3D landmarks. The warp module, following the deformation, produces a realistic 3D faces which are consistent with the input images.

One of the key issues is finding facial features in the presence of large pose variations. We apply a "divide and conquer" strategy to solve this difficult problem: we classify the facial pose in a first step, then apply a view-based approach for the landmark detection step. This is similar to [5] but we detect two anchor points for each view. These anchors provide a strong constraint for the landmark detection, which produces reliable results.

We evaluate the performance of individual modules, including pose/anchor detector and landmark detector, with the Color FERET and CMU multi-PIE databases [6, 7]. Visual assessment on the reconstruction results with single and multiple views confirms the appropriateness our methods.

## 2. System Overview

Our system consists of three parts as shown in Fig. 1: (1) landmark extraction modules, (2) 3D reconstruction for single images, and (3) 3D reconstruction for multiple images. The landmark extraction module consists of pose classifier, anchor detector, and view-based landmark detectors. Both reconstruction parts use the warping module and the 3D reconstruction module for multiple images includes the sparse reconstruction and the

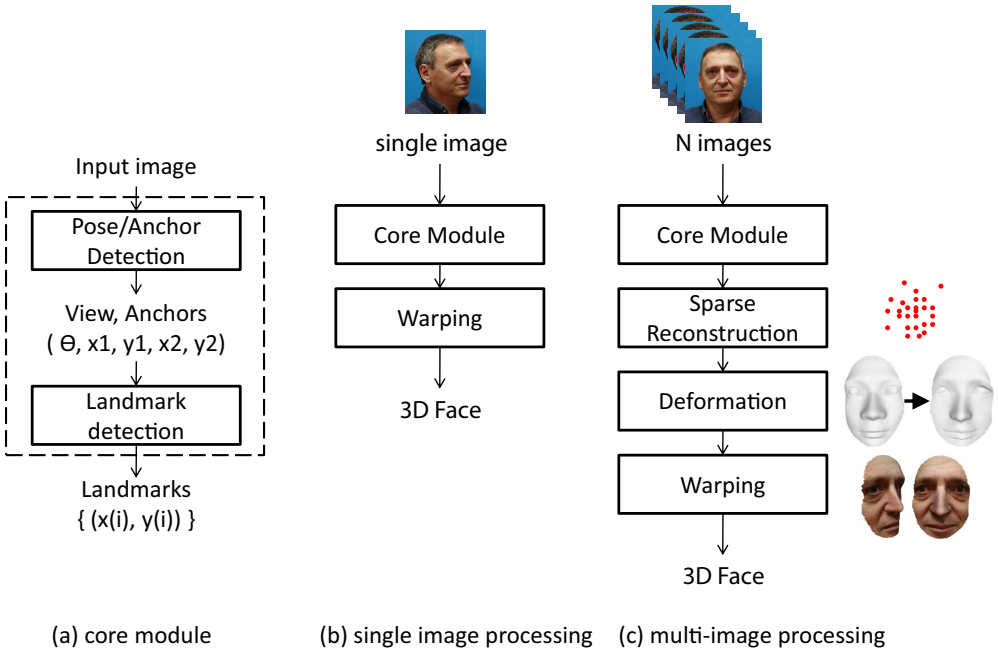


**Figure 1. System Overview.**

deformation module. The output is a 3D face model which can be rendered from any point of view.

(a) frontal (b) half-frontal (c) profile

**Figure 2. Pose/anchor detection results.**

### 2.1 Pose Classification and Anchor Points Detection

To classify faces across different poses, we use five boosted cascade face detectors: two standard classifiers available in OpenCV (haarcascad_frontalface _alt2.xml and haarcascade_profileface.xml) and three classifiers we designed to detect specific poses (frontal, half-frontal, profile). The specific classifiers were obtained by training Haar cascades using 3019 background images and a set of face images for each pose (2409 for frontal, 505 for half-frontal, 2007 for profile). The specific classifiers are more efficient than the generic classifiers for pose classification, but the generic classifiers are better for face location purposes. In this first stage, the detection results for these five detectors have the following output: 1) face location and pose, if the face was detected by a specific classifier; 2) face location, if the face was detected only by a standard classifier; 3) nothing, if the face was not found. When the face is detected but its pose cannot be defined in the first stage, the face is then submitted to another appearance recognition technique to obtain its pose. The appearance recognition technique employed in this stage is the Principal Component Analysis (PCA) [10]. In this approach, a set of face images that have their poses already known are used in the training stage and each probe image is compared to all images in the training set; the pose of the closest training image is assigned to the probe image. Due to the dimensionality reduction provided by PCA, this can be performed in real time for large training sets.

Two different training sets were created by using the generic classifiers employed in the first stage. Face images of the ColorFERET database were extracted using the bounding square obtained by the generic classifier response, and the resulting images were resized to $20 \times 20$ pixels and used for training. The first generic

**Figure 3. Landmark detection results.**

classifier extracted 2716 frontal faces, 242 faces of half-frontal and 9 profile faces. The second one extracted 601 frontal faces, 478 faces of half-frontal and 1121 profile faces.

The PCA dimensionality reduction was set to map 90% of the data variation, resulting in 32 axes for the first training set and 21 axes for the second one. This amount of data variation is typically employed [10] because it provides a good cost-benefit between dimensionality reduction and recognition results. The final result is 97% of correct pose classification, i.e., 5110 images for the entire ColorFERET database (5234 images). Boosted cascade classifiers were also employed to locate landmarks in faces already detected and classified (Fig. 2).

For each desired landmark, a region involving its location is extracted from the face image and used as positive instance for training. The region extracted is also removed from the input image and the resulting image is used as background for training. With these images, we create classifiers able to extract the desired anchors. This procedure was applied to all anchors shown in Fig. 2 producing 99% of correct landmark detection.

### 2.2 Facial landmark detection

We use a view-based approach to handle pose variations. The range of rotation angle (yaw) is 90 to +90. Facial views are defined as frontal, half-frontal, and profile. For each classified view, we apply a landmark detector. The individual detector consists of a shape model and a texture model [9]. The shape model is defined a linear combination of some basis shapes which are trained by principal component analysis and a set of training shape database. The texture model includes a set of individual detectors that respond to special locations of facial features such as lip corner and eye corner. The module finds the optimal location of landmarks in terms of a combination of shape information and texture information. It finds the best shape that has the minimum distance between shape model and individual texture feature point detector.

Fig. 3 shows a set of landmark detection results on five views from $+90$ degree to $-90$ degree view. The markers (plus) show the detected landmark locations. Each view has a definition of the landmarks. The total number of landmarks is 47 (frontal: 42, profile: 18).

We evaluated our landmark detection modules. The error is defined as the average Euclidean distance between ground truth points and estimated points, whose values are normalized by a reference eye distance (32 pixel). The average errors obtained in our experiments are: frontal 1.95 pixels; half-frontal 3.5 pixels; and profile 2.2 pixels.

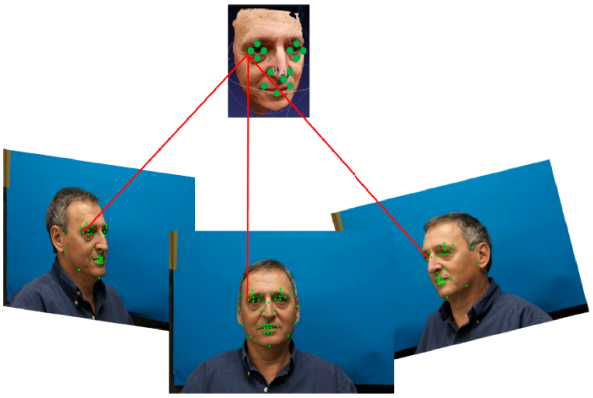

**Figure 4. Sparse 3D Reconstruction of Landmarks.**

## 3. 3D Face Reconstruction

### 3.1 Warping

To obtain a precise texture warping, a generic model is aligned to the probe face using the landmarks previously obtained. To do that, we use the Levenberg-

**Figure 5. A warping result using a single input image (bottom) and a 3D reconstruction result using the sparse reconstruction and deformation (top)**

Marquardt iterative minimization approach [8], to obtain the best transformation between landmarks of the generic model and probe image. A pre-alignment can be obtained by using center of mass and the result of the pose classification. After that, the points of the generic model are projected into the probe image using the obtained transformation, and occlusions are overcome through the use of symmetry.

### 3.2 3D Landmarks Reconstruction

Given a set of facial images, including some pose variations, we can compute 3D facial landmarks by triangulating the 2D landmarks extracted from multiple images. This is not accurate enough in practice, we therefore use a sparse bundle adjustment (SBA) algorithm [4] to refine the 3D landmark points and the camera parameters as shown in Fig. 4.

We evaluated the performance of our sparse reconstruction using a subset of multi-PIE database [7]. We selected 80 subjects with 5 different views as the same as our view definitions. The total number of images is 400. The average re-projection error across all landmarks for all 80 subjects was 1.34 pixels.

### 3.3 Deformation

In this stage, the deformation between landmarks of a 3D generic model and the same landmarks of an input face is mapped through Thin Plate Splines [11]. Both landmark sets must be pre-aligned to avoid mapping variations in translation, rotation and scale as deformation. After mapping the deformation, the obtained transformation is applied to all points of the the generic model to obtain a final deformed model of the input face. Fig. 5 shows 3D reconstruction results using a single input image and 5 images. The nose areas clearly show the difference of quality between the two approaches. Some distortion can be seen in the warping result using a single image.

## 4. Conclusion

We presented a framework to build a realistic 3D face from a single or multiple facial images. We first classify the facial pose, detect two anchor points, and then detect facial landmarks. For a single view, we use warping to build a 3D face. For multiple views, we reconstruct 3D landmarks, deform the generic 3D face model, and then apply warping. Our approach has been validated experimentally.

## References


[1] G. Medioni, J. Choi, C. H. Kuo, D. Fidaleo, Identifying Noncooperative Subjects at a Distance Using Face Images and Inferred Three-Dimensional Face Models, IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans 39(1): 12–24, 2009.

[2] B. Amberg, A. Blake, A. W. Fitzgibbon, S. Romdhani, T. Vetter, Reconstructing High Quality Face-Surfaces using Model Based Stereo. ICCV: 1–8, 2007.

[3] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3D Face Model for Pose and Illumination Invariant Face Recognition, AVSS: 296–301, 2009.

[4] M. I. A. Lourakis, A. A. Argyros, SBA: A Software Package for Generic Sparse Bundle Adjustment, ACM Trans. Math. Software 36(1): 1–30, 2009.

[5] J. Heo, M. Savvides, Face Recognition Across Pose Using View Based Active Appearance Models (VBAAMs) on CMU Multi-PIE Dataset, ICVS: 527–535, 2008.

[6] Color FERET database (http://face.nist.gov/colorferet/)

[7] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, S. Baker, Multi-PIE. Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 2008.

[8] J. Salvi, X. Armangu, J. Batlle, A comparative review of camera calibrating methods with accuracy evaluation, Pattern Recognition 35(7):1617–1635, 2002.

[9] L. Zhang, et. al., Robust Face Alignment Based on Local Texture Classifiers, Proc. IEEE ICIP: 354–357, 2005.

[10] W. S. Yambor, B. A. Draper and J. R. Beveridge, Analyzing PCA-based face recognition algorithm: Eigenvector' selection and distance measures. Empirical Evaluation Methods in Computer Vision. World Scientific Press, 2002.

[11] F. L. Bookstein, Principal warps: thin-plate splines and the decomposition of deformations, IEEE Transactions on Pattern Analysis and Machine Intelligence 11(6): 567–585, 1989.