# Face Reconstruction Across Different Poses and Arbitrary Illumination Conditions

Sen Wang, Lei Zhang, and Dimitris Samaras

Department of Computer Science,
SUNY at Stony Brook, NY, 11794
{swang,lzhang,samaras}@cs.sunysb.edu

**Abstract.** In this paper, we present a novel method for face reconstruction from multi-posed face images taken under arbitrary unknown illumination conditions. Previous work shows that any face image can be represented by a set of low dimensional parameters: shape parameters, spherical harmonic basis (SHB) parameters, pose parameters and illumination coefficients. Thus, face reconstruction can be performed by recovering the set of parameters from the input images. In this paper, we demonstrate that the shape and SHB parameters can be estimated by minimizing the silhouettes errors and image intensity errors in a fast and robust manner. We propose a new algorithm to detect the corresponding points between the 3D face model and the input images by using silhouettes. We also apply a model-based bundle adjustment technique to perform this minimization. We provide a series of experiments on both synthetic and real data and experimental results show that our method can have an accurate face shape and texture reconstruction[1].

## 1 Introduction

Face recognition from images has received significant attention in the past few decades. Although rapid progress has been made in this area during the last few years, the general task of recognition remains unsolved. In general, face appearance does not depend solely on identity. It is also influenced by illumination and viewpoint. Thus, recovery of 3D shape and texture from face images is an important task for an accurate face recognition system. In this paper, we propose a novel method to extract accurate 3D shape and texture from multi-pose face images taken under arbitrary unknown lighting.

Previous work[19][20] has shown that any face image taken under arbitrary unknown lighting and pose can be represented by a set of low dimensional parameters: shape parameters, spherical harmonic basis parameters, pose parameters and illumination parameters. Thus, given input images, 3D face reconstruction can be performed by estimating the shape and spherical harmonic basis parameters of the face. In this paper, we demonstrate that, given a set of multi-posed

[1] We would like to thank Sudeep Sarkar and Simon Baker for providing databases and Thomas Vetter and Sami Romdhani for helpful discussions. This research was supported by grants from U.S. Department of Justice(2004-DD-BX-1224) and National Science Foundation(ACI-0313184)

T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 91–101, 2005.

face images, the shape and texture parameters can be recovered by minimizing the silhouette errors and image intensity errors respectively.

We recover shape by using silhouette images because the silhouette images depend only on the shape and pose of the objects and thus are illumination independent. This reconstruction technique is also called visual hull[10][8] and the accuracy of shape reconstruction depends on the number and location of cameras used to capture images. In general, such methods cannot perform shape recovery accurately for complex objects such as human faces when the visual hull is constructed from a small number of cameras. However, prior knowledge of the object to be reconstructed can help shape recovery by providing an important constraint. In our method, the 3D face model we constructed with separate shape and texture parts provides such prior knowledge and thus facilitates accurate shape recovery.

Our method can be described by the following steps: 1) From a set of 3D faces[2] obtained by laser-based cylindrical scanners, we construct a 3D face Model with separate shape and texture parts; 2) Given a set of multi-pose input images of a human face under unknown lighting, we estimate the pose parameters and shape parameters by minimizing the difference between the silhouette of the face model and the input images. 3) Using the correspondences provided by the recovered 3D shape, we recover the illumination parameters and the spherical harmonic basis parameters by minimizing the image intensity errors. Thus, the texture of the face can be computed from the recovered spherical harmonic basis.

The main contributions of our paper are the following:

- We propose a new and efficient method to recover 3D shape and appearance from multi-pose face images under arbitrary unknown lighting.
- We present a novel algorithm to detect the corresponding points between the 3D face model and the input images by using silhouettes and use model-based bundle adjustment[16] to minimize errors and recover shape and pose parameters.
- We reconstruct appearance by recovering the spherical harmonics basis parameters from multiple input face images under unknown light while texture and illumination information are recovered in tandem.

This paper is organized as follows. In the next section, we will discuss the related work on face reconstruction. In Section 3, we will introduce shape recovery by using silhouette face images. In Section 4, we will explain appearance recovery by using our 3D face model. Experimental results on both synthetic and real data are presented in Section 5. The final Section presents the conclusions and future work directions.

## 2 Related Work

In recent years, there is extensive research on face reconstruction both from a single image and from image sequences. The main approaches are shape from stereo[4], shape from shading[15], shape from structured light[12] and shape from silhouettes[18].

Blanz and Vetter's face recognition system is the closest in spirit to our work. They are the first to reconstruct the shape and texture by using a face morphable model. They also apply the 3D morphable model successfully in both face recognition and synthesis applications [3][2]. In their method, they acquire the necessary point to point correspondences by using a gradient-based optical flow algorithm[2][14]. This method might suffers in situations where the illumination information is general and unknown. Compared with their method, our method determines the correspondences from silhouette which is less sensitive to illumination and texture variations.

Lee et al.[9] proposed a method of silhouette-based 3D face shape recovery by using a morphable model. They used a boundary weight XOR method to optimize the procedure and used a downhill simplex method to solve the minimization problem which is time consuming. Since they fitted a generic face model to silhouette images by marking several feature points by hand, the accuracy of their method depends on the accuracy of these feature points which can not be updated after manually marked in the generic face model. Compared with their work, we apply a model-based bundle adjustment technique to solve the optimization problem and during the optimization, the pose information is also updated thus providing better shape recovery.

Fua[6] used a generic face model to derive shape constrains and used a model-driven bundle adjustment algorithm to compute camera motions. However, the 3D face model by recovered this model-driven bundle adjustment method needs to be refined through an additional step of mesh-based deformable model optimization. In [5], Dimitrijevic et al. also used a 3D morphable model to recovery shape from face image sequences. A simple correlation-based algorithm is used to find feature points whose performance might depend on the accuracy of the correspondences detected by the cross correlation algorithm.

## 3 Shape Recovery

In this section we introduce our new approach to the recovery shape from multi-pose face images by using silhouette images as input to extract correspondence and recover shape parameters.

### 3.1 Shape Part of 3D Face Model

Let $S(\alpha)$ be the shape of an arbitrary face model parameterized by a vector $\alpha = \alpha_1, \alpha_2, ..., \alpha_n$. We want to use the silhouette images to recover this vector $\alpha$. In our method, we used a collection of 3D faces supplied by USF as the bootstrap data set and we applied PCA[2] to register and align the database of 3D faces to get the statistical shape model. This model can be used to reconstruct both a new and existing faces through the linear combination of a bootstrap set of 3D face shapes.

$$s(\alpha) = \overline{s} + \sum_{i}^{M} S_i \alpha_i. \tag{1}$$

where $S_i$ is the $ith$ eigen-vector of the variation shape matrix and $\overline{s}$ is the mean shape of the bootstrap faces.

### 3.2 Silhouette Extraction

We extract face silhouettes from each input image. At the beginning we initialize a 3D face model from the input images and project the face model onto the image plane in order to extract the silhouettes of this model. Because the face model we use is not the whole head model, we do not need the complete head silhouette but only the silhouette of the facial area (in Fig. 1: example of silhouette extraction).

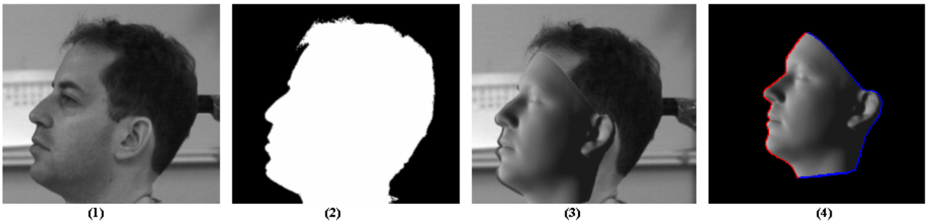


**Fig. 1.** Example of silhouette extraction. (1) is one of the input images, (2) is the face silhouette of this input image, (3) shows the fitting of the generic face model (shaded surface rendering) to input image, (4) is the silhouette of the fitted model (we just use the silhouette of the facial area, the red curve in left. The right blue curve is the silhouette of the omitted head boundary.

### 3.3 Correspondence Detection

Once we have extracted the silhouettes from the input face images and the 3D face model after fitting, we need to find the correspondences between them to update the shape and pose parameters. First, we detect the points with high curvature in the silhouettes of the face model and the input images and match them as initial correspondences by using a shortest Euclidean distance metric. Using these initial correspondences, we detect the correspondences of the remaining points in silhouettes by using a smooth matching function. Given a set of known distance vectors of feature points $u_i = p_i - \hat{p}_i$ at every matched high-curvature point $i$, we construct a function that gives the distance vectors $u_j$ for every unmatched vertex $j$. We attempt to find a smooth vector-valued function $f(p)$ fitted to the known data $u_i = f(p_i)$, from which we can compute $u_j = f(p_j)$. There are several choices for constructing this function [7][11]. Similar to [11], we use a method based on radial basis functions $f(p) = \sum_i w_i \phi(\|p - p_i\|)$ where $\phi(r)$ is radial symmetric basis function. We also use an affine basis as part of our algorithm, so the function has the form: $f(p) = \sum_i w_i \phi(\|p - p_i\|) + Tp + m$.

To determine the coefficients $w_i$ and the affine components $T$ and $m$, we solve a linear equation $u_i = f(p_i)$, with the constraints $\sum_i w_i = 0$ and $\sum_i w_i p_i^T = 0$, which remove the effects of any affine transformation from the radial basis function. Here we have chosen to use $\phi(r) = e^{-r/c}$, where $c$ is a pre-selected constant ($c = 64$ in our experiment).

After we construct the function $u = f(p)$, we can use $\hat{p}_j = p_j - f(p_j)$ and a shortest Euclidean distance metric to find the remaining correspondences in silhouettes of both the face model and the input images.

### 3.4 Shape and Pose Parameters Update

Given a set of multi-pose input images, the shape and pose parameters can be recovered as following:

1) Initialize the shape parameters $\alpha$ as 0 and initialize a number of feature points in the input images. In our experiments, we manually mark 7 feature points in both the first image and the 3D face model. By matching these features across images using the point matching technique in [21], we can acquire the corresponding feature points in the other input images and thus get the initial fitting information.

2) Extract the face contour $c_i$ (image (2) in Fig. 1) in each input image and using the current fitting information, project the face model to the image plane and extract the face model contours $s_i$ (red line in (4) of Fig. 1) as described in section 3.2.

3) From the contours of the face model $\{s_i, i = 1...N\}$, find the corresponding points in the silhouettes of the input images $\{c_i, i = 1...N\}$ by using the methods presented in section 3.3.

4) The contour of the model $s_i$ can be represented as $s_i = C_i^m[P \times M^p(\bar{s} + S\alpha)]$ where $C_i^m(x)$ is the contour extraction operator. $M^p$ is the transformation matrix from the original face model coordinate system to the camera coordinate system. $P$ is the camera projection matrix to project the 3D model to the 2D image.

Thus, the minimization can he written as follows:

$$min \sum_{i=1}^{n} \|c_i - s_i\|^2 = min \sum_{i=1}^{n} \|c_i - C_i^m[P \times M^p(\bar{s} + S\alpha)]\|^2 \qquad (2)$$

For such an expression, we update the shape and pose parameters by using model-based bundle adjustment techniques[16] to solve this minimization problem.

5) After we get the new face model and new fitting parameter values, we reproject the new 3D face model to the input images and perform 2)- 4) iteratively until the change of shape and pose parameters are smaller than $\xi_s$ and $\xi_p$, which are pre-selected thresholds.

## 4 Texture Recovery

In this section we describe a method that recovers texture from multi-pose face images under arbitrary unknown lighting. We use a spherical harmonics illumination representation to recover the spherical harmonic basis which contains texture information.

### 4.1 Texture Component of the 3D Face Model

As described in [1][13], any image under arbitrary illumination conditions can be approximately represented by the linear combination of the spherical harmonic basis as:

$$I \approx b\ell \tag{3}$$

where $b$ is the spherical harmonic basis and $\ell$ is the vector of the illumination coefficients.

The set of images of a convex Lambertian object obtained under a wide variety of lighting conditions can be approximated accurately by a 9 dimensional linear subspace. Since human faces can be treated approximately as Lambertian, we compute a set of 9 spherical harmonic basis images by using a collection of 3D faces similar to [1] as follows:

$$\begin{array}{lll} b_{00}=\frac{1}{\sqrt{4\pi}}\lambda, & b_{10}=\sqrt{\frac{3}{4\pi}}\lambda.*n_z, & b_{20}=\frac{1}{2}\sqrt{\frac{3}{4\pi}}\lambda.*(2n_{z^2}-n_{x^2}-n_{y^2}), \\ b^o_{11}=\sqrt{\frac{3}{4\pi}}\lambda.*n_y, & b^e_{11}=\sqrt{\frac{3}{4\pi}}\lambda.*n_x, & b^o_{22}=3\sqrt{\frac{5}{12\pi}}\lambda.*n_{xy}, \\ b^o_{21}=3\sqrt{\frac{5}{12\pi}}\lambda.*n_{yz}, & b^e_{21}=3\sqrt{\frac{5}{12\pi}}\lambda.*n_{xz}, & b^e_{22}=\frac{3}{2}\sqrt{\frac{5}{12\pi}}\lambda.*(n_{x^2}-n_{y^2}). \end{array} \tag{4}$$

where the superscripts $o$ and $e$ denote the odd and the even components of the harmonics respectively, $\lambda$ denote the vector of the object's albedos, $n_x, n_y, n_z$ denote three vectors of the same length that contain the $x, y$ and $z$ components of the surface normals. Further, $n_{xy}$ denote a vector such that the $ith$ element $n_{xy,i} = n_{x,i}n_{y,i}$.

In recent work [20], the set of spherical harmonic basis images of a new face can be represented by a linear combination of a set of spherical harmonic basis computed from a bootstrap data set of 3D faces.

$$b(\beta) = \bar{b} + \sum_{i}^{M} B_i\beta_i. \tag{5}$$

where $\bar{b}$ is the mean of the spherical harmonic basis and $B_i$ is the $ith$ eigen-vector of the variance matrix.

### 4.2 Texture and Illumination Parameters Update

According to Eq. 3 and 5, using the recovered shape and pose information, a realistic face image can be generated by:

$$I = (\bar{b} + B\beta)\ell \tag{6}$$

where $\beta$ is the spherical harmonic basis parameter to be recovered and $\ell$ is the vector of illumination coefficients. Thus, given a set of $n$ input images $I^i_{input}, i = 1...n$ of a face, the spherical harmonic basis parameters $\beta$ of the face and the illumination coefficients $\ell = (\ell_1, \ell_2, ...\ell_n)$ can be estimated by minimizing the difference between the input images and the rendered images from Eq.6:

$$min_{\beta,\ell} \sum_{i=1}^{n} \|I_{input}^{i} - (\bar{b} + B\beta)\ell_i\|^2 \tag{7}$$

Eq. 7 is similar to Eq. 2, thus, we can solve Eq. 7 similarly. Given input images $I : I_1, I_2, ..., I_n$, we initialize the set of spherical harmonic basis parameters $\beta = 0$ and thus, $b = \bar{b} + B\beta = \bar{b}$. Hence, the set of illumination coefficients $\ell_i$ of each input image $I_i$ can be initially estimated by solving a linear equation: $b\ell_i = I_i$. With the initial illumination coefficients $\ell_i$, we can solve Eq. 7 using the same technique applied to Eq. 2.

The core of the recovery process is the minimization of the image errors as shown in Eq. 7. Thus, the recovery results depend on the initial values of the illumination coefficients. Our experiments on synthetic data showed that the illumination coefficients $\ell$ computed by using the mean spherical harmonic basis $(\bar{b})$ were close to the actual values, which made the whole recovery fast and accurate.

After we estimate the spherical harmonic basis from input images, the texture of a face can be computed as $\lambda = b_{00}\sqrt{4\pi}$ according to Eq. 4.

# 5 Experiments

In this section, we provide experimental results of our method on both synthetic data and real data for face reconstruction.

## 5.1 Synthetic Data

We use synthetic data as ground truth to show the accuracy and robustness of our method. In our experiments, we synthesize 30 face models by randomly assigning different shape and spherical harmonic basis parameters to our 3D face model. For each model we also generate 14 images with different poses and different illuminations (image sequence of one face in Figure 2). We recover the shape and texture from these images and compare them with shape and texture of the original face models.

To quantify the accuracy of our method we compute the errors between recovered models and original synthesized face models. At first, we compute the errors of shape and texture in each vertex between the reconstructed face model and the ground truth face model by: $err_s(i) = \frac{\sqrt{(\tilde{x}_i - x_i)^2 + (\tilde{y}_i - y_i)^2 + (\tilde{z}_i - z_i)^2}}{\sqrt{x_i^2 + y_i^2 + z_i^2}}$ and $err_t(i) = \frac{\|\tilde{I}_i - I_i\|}{I_i}$ where $(x_i, y_i, z_i)$ and $I_i$ are the coordinate and texture of $ith$ vertex of the ground truth face model, and $(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)$ and $\tilde{I}_i$ are the coordinate and texture of $ith$ vertex of the reconstructed face model. Then, we compute the maximum, minimum, mean and standard deviation of the shape and texture errors by comparing all 30 reconstructed 3D face models to the original faces as shown in Table 1. From these experimental results we can see that our method achieves accurate shape and texture recovery from synthetic data. Figure 3 shows the relationship between the reconstructed shape and the number of input images

**Table 1.** Statistical errors of shape and texture recovery all these 30 synthetic faces

| | Max | Min | Mean | Std. dev. |
|---|---|---|---|---|
| Shape | 12.35% | 0.97% | 3.53% | 3.237% |
| Texture | 23.83% | 1.87% | 4.78% | 4.659% |

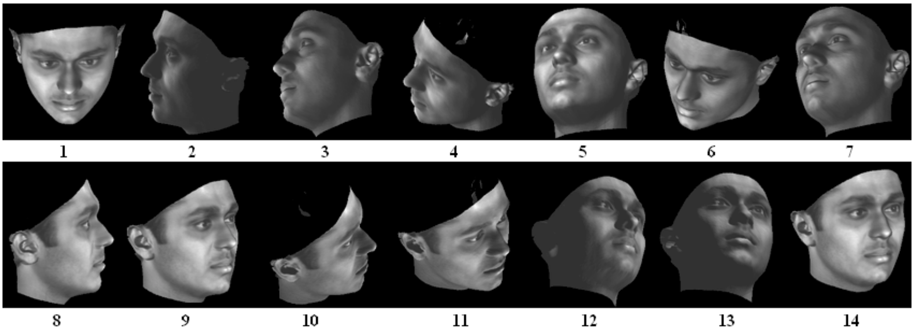


**Fig. 2.** 14 input images synthesized for the same face in different pose and different illumination

as a subset of the input image sequence in Figure 2 and Figure 4 shows the errors between the recovered shape from different numbers input images and the original face shape. With the increase of the number of input images, we get more accurate results of shape recovery and if the input images are more than 6, the improvement of shape reconstruction will be less influenced by the number of input images. Figure 5 shows 2 examples of shape and texture reconstruction results.

### 5.2 Real Data

We use the CMU PIE database [17] for our real data experiments. In the PIE data set, there are 13 different poses and 22 illumination conditions per pose for each subject. The silhouettes of face images can be detected by subtracting the background image from the input images. Figure 6 shows two accurate shape and texture recovery results of our method. The experimental results on the real data demonstrate that our method can recover good shape and texture from multi-pose face images under unknown illumination conditions.

## 6 Conclusions and Future Work

In this paper, we proposed a novel method for face modeling from multi-pose face images taken under arbitrary unknown illumination conditions. We demonstrated that the shape and spherical harmonic basis parameters can be estimated by minimizing the silhouette errors and image intensity errors. We proposed a new algorithm to detect the corresponding points between the model and the

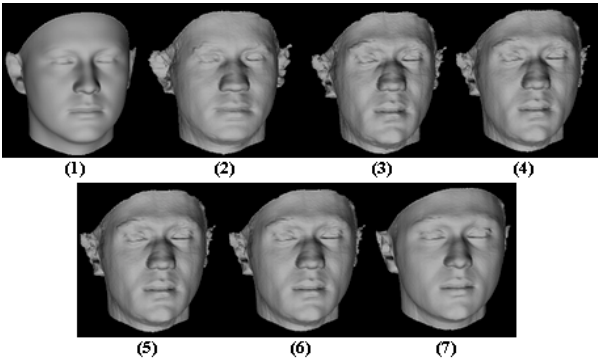


**Fig. 3.** Shape reconstruction using a varying number of input images in Fig 2. (1) is the mean face model, (2) is the reconstructed shape from image 1 to 3, (3) is the reconstructed shape from image 1 to 6, (4) is the reconstructed shape from image from 1 to 9, (5) is the reconstructed shape from image from 1 to 12, (6) is the reconstructed shape from image from 1 to 14 and (7) is the original shape of the face in Fig 2

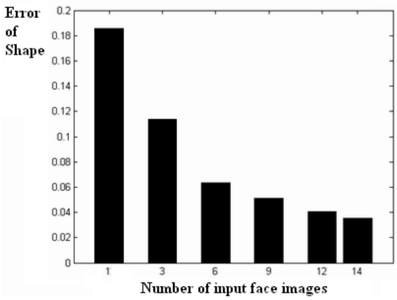


**Fig. 4.** The errors between the reconstructed face shape and the original face shape in Fig. 3

**Fig. 5.** Some reconstruction results from synthetic faces. In each row, the first image is original shape (shaded surface rendering) followed by original texture. The third image is the mean face model which is initially fitted to the input images. The last 2 images are the reconstructed face shape and texture

**Fig. 6.** Reconstruction results for 2 subjects from real images. Original images are in the first row, reconstructed face shapes are in the second row and recovered textures are in the last row

input images by using silhouettes. We also applied a model-based bundle adjustment technique to solve the minimization problems. We provide a series of experiments on both synthetic and real data and experimental results show that our method can reconstruct accurate face shape and texture from multi-pose face images under unknown lighting. In future, in order to extract more robust correspondences for shape recovery, we plan to use both silhouette information and image intensity information after delighting the input face images. At this time, there exist few publicly available sets of face images under arbitrary illumination conditions, so we plan to continue validation of our method on databases with greater variability of light sources as they become available.

## References


1. R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, February 2003.
2. V. Blanz and T. Vetter. A morphable model for the synthesis of 3d-faces. In *SIGGRAPH*, 1999.
3. V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *PAMI*, 25(9):1063–1074, Sept. 2003.
4. F. Devernay and O. D. Faugeras. Computing differential properties of 3d shape from stereoscopic images without 3d models. In *CVPR*, pages 208–213, 1994.
5. M. Dimitrijevic, S. Ilic, and P. Fua. Accurate face model from uncalibrated and ill-lit video sequences. In *CVPR*, pages 1034–1041, 2004.
6. P. Fua. Regularized bundle adjustment to model heads from image sequences without calibration data. *IJCV*, 38(2):153–171, 2000.

7. M. Gregory and Nielson. Scattered data modeling. In *Computer Graphics and Application*, 1993.
8. A. Laurentini. The visual hull concept for silhouette based image understanding. *PAMI*, 16(2):150–162, 1994.
9. J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. Silhouette-based 3d face shape recovery. In *Graphics Interface*, 2003.
10. W. Matusik, C. Buehler, R. Raskar, L. McMillan, and S. J. Gortle. Image-based visual hulls. In *SIGGRAPH*, 2003.
11. F. Pighin, Hecker J., Lischinski D., Szeliski R., and D. Salesin. Synthesizing realistic facial expressions from photographs. In *SIGGRAPH*, pages 75–84, 1998.
12. M. Proesmans, L. Vangool, and A. Osterlinck. Active acquisition fo 3d shape for moving objects. In *ICIP*, 1996.
13. R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *SIGGRAPH*, 2001.
14. S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error function. In *ECCV*, volume 4, 2002.
15. D. Samaras and D. Metaxas. Incorporating illumination constraints in deformable models. In *CVPR*, pages 322–329, 1998.
16. Y. Shan, Z. Liu, and Z. Zhang. Model-based bundle adjustment with application to face modeling. In *ICCV*, 2001.
17. T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination and expression database. *PAMI*, page 1615–1618, 2003.
18. L. Tang and T.S. Huang. Analysis-based facial expression synthesis. In *ICIP*, volume 94, pages 98–102, 1996.
19. L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *CVPR*, volume I, pages 19–25, 2003.
20. L. Zhang, S. Wang, and D. Samaras. Face synthessis and recognition from a single image under arbitrary unknow lighting using a spherival harmonic basis morphable model. In *CVPR*, 2005.
21. Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, 1995.