3D Face Reconstruction from a Single Image using a Single Reference Face Shape

Ira Kemelmacher-Shlizerman, Ronen Basri, *Member, IEEE*

Abstract—Human faces are remarkably similar in global properties, including size, aspect ratio, and location of main features, but can vary considerably in details across individuals, gender, race, or due to facial expression. We propose a novel method for 3D shape recovery of faces that exploits the similarity of faces. Our method obtains as input a single image and uses a mere single 3D reference model of a different person's face. Classical reconstruction methods from single images, i.e. shape-from-shading, require knowledge of the reflectance properties and lighting as well as depth values for boundary conditions. Recent methods circumvent these requirements by representing input faces as combinations (of hundreds) of stored 3D models. We propose instead to use the input image as a guide to ”mold” a single reference model to reach a reconstruction of the sought 3D shape. Our method assumes Lambertian reflectance and uses harmonic representations of lighting. It has been tested on images taken under controlled viewing conditions as well as on uncontrolled images downloaded from the internet, demonstrating its accuracy and robustness under a variety of imaging conditions and overcoming significant differences in shape between the input and reference individuals including differences in facial expressions, gender and race.

Index Terms—Computer vision, photometry, shape from shading, 3D reconstruction, lighting, single images, face, depth reconstruction.

1 Introduction

THREE-DIMENSIONAL shape and reflectance provide properties of objects that are invariant to the changes caused by the imaging process including viewpoint, illumination, background clutter and occlusion by other objects. Knowledge of these properties can simplify recognition, allow prediction of appearance under novel viewing conditions, and assist in a variety of applications including graphical animation, medical applications, and more. A major challenge in computer vision is to extract this information directly from the images available to us, and in particular, when possible, from a mere *single* image. In this paper, we use the shading information (the pattern of intensities in the image) along with rough prior shape knowledge to accurately recover the 3-dimensional shape of a novel face from a single image.

In a global sense, different faces are highly similar [17]. Faces of different individuals share the same main features (eyes, nose, mouth) in roughly the same locations, and their sizes and aspect ratio do not vary much. However, locally, face shapes can vary considerably across individuals, gender, race, or as a result of facial expression. The global similarity of faces is exploited, for example, in face recognition methods, to estimate the pose of novel faces by aligning a face image to a generic face model. In this paper we demonstrate how a similar idea can be exploited to obtain a detailed 3D shape reconstruction of novel faces.

We introduce a novel method for shape recovery of a face from a *single image* that uses only *a single reference 3D face model* of either a different individual or a generic face. Intuitively, our method uses the input image as a guide to ”mold” the reference model to reach a desired reconstruction. We use the shading information to recover the 3D shape of a face while using the reference shape and albedo to extract information essential for the recovery process that is unknown a priori, such as lighting and pose. Specifically, we cast the problem as an image irradiance equation [15] with unknown lighting, albedo, and surface normals. We assume Lambertian reflectance, light sources at infinity, and rough alignment between the input image and the reference model. To model reflectance we use a spherical harmonic approximation (following [2], [27]), which allows for multiple unknown light sources and attached shadows.

We begin by using the reference model to estimate lighting and pose, and provide an initial estimate of albedo. Consequently, the reflectance function becomes only a function of the unknown surface normals and the irradiance equation becomes a partial differential equation which is then solved for depth. For this we also employ appropriate boundary conditions. Since in general the recovery of shape and albedo from an image is ill-posed, we further introduce regularization terms to seek solutions that preserve the rough shape and albedo of the reference model. These terms will smooth the *difference* in shape and albedo between the reference model and the sought face. We provide experiments that demonstrate that our method can achieve accurate

- *Ira Kemelmacher-Shlizerman is with the Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195-2350. E-mail: kemelmi@cs.washington.edu*
- *Ronen Basri is with the Department of Computer Science and Applied Mathematics, the Weizmann Institute of Science, Rehovot, 76100 Israel. E-mail: ronen.basri@weizmann.ac.il.*

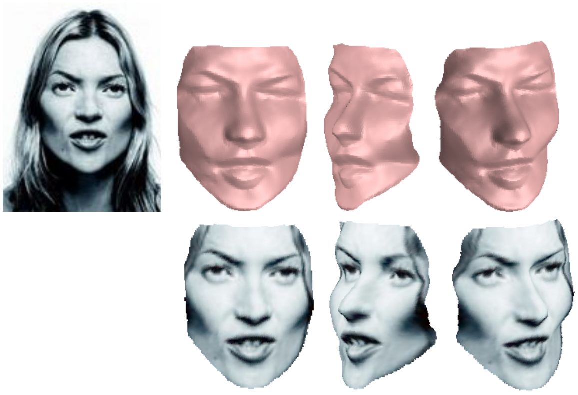

Fig. 1. 3D reconstruction of a face from an image downloaded from the internet using our algorithm. We present the input image (top left), the reconstructed shape viewed from three viewpoints (top right), and the image overlay of the reconstructed shape (bottom right).

reconstructions of novel input faces overcoming significant differences in shape between the input and reference individuals including differences in gender, race, and facial expression. These experiments demonstrate that our method can potentially overcome some of the most critical problems in recovering the 3D models of unknown individuals. In Figure 1 we show an example of a result obtained by applying our algorithm to a real image downloaded from the internet.

The paper is divided as follows. Section 2 describes related work. Section 3 defines the reconstruction problem and the optimization functional. Section 4 describes the reconstruction algorithm. Experimental evaluations are presented in Section 5 and conclusions in Section 6. A preliminary version of this paper appeared in [19].

2 Previous work

Shape from shading (SFS), the problem of three-dimensional reconstruction from a single image using shading information is classically defined as solving the Irradiance Equation, which for Lambertian surfaces is given by $I(x,y) = \rho\vec{l}^{\,T}\vec{n}$. Here $\vec{l}$ is a three component vector representing the direction and intensity of a single point light source placed at infinity, $\vec{n}(x,y)$ is a three component vector representing the surface normal at each surface point and $\rho(x,y)$ is the surface albedo at each point $(x,y) \in \Omega \subset \mathbb{R}^2$. The objective in SFS is to recover the surface $z(x,y)$ whose normal vectors are specified by the unknown $\vec{n}(x,y)$. A regularization term is added in some studies to enforce the smoothness of $z(x,y)$. In general, the SFS problem is ill-posed and its solution requires knowledge of the lighting conditions, the reflectance properties (albedo) of the object (in many studies albedo is assumed to be constant), and boundary conditions (i.e., the depth values at the occluding contours and the extremal points of the underlying shape). Such information is part of the sought 3D shape and is usually unavailable. This limits the applicability of SFS methods to restricted setups. Methods for solving SFS were first introduced by Horn [15], [14]. More recent solutions can be found, e.g., in [9], [20], [28], [35]. Methods for estimating lighting and relaxing the constant albedo assumption were proposed in [24], [37], [33]. RANSAC-based robust methods for estimating lighting were proposed in [13] in the context of multi-view photometric stereo.

In spite of the limitations of SFS, people appear to have a remarkable perception of three-dimensional shapes already from single two-dimensional pictures. Ramachandran [25] proposed that simple shapes are perceived through the assumption that the image is illuminated by a single light source. More complex 3D shapes (like faces) can be perceived through the help of prior knowledge [26]. Furthermore, it was shown that people can successfully recognize faces from novel images overcoming significant viewpoint and lighting variations, while they seem to achieve significantly inferior performance with images of unfamiliar objects, such as inverted faces. This ability is often attributed to familiarity with faces as a class [23].

Indeed many computational studies attempt to use prior knowledge of class information to approach the 3D reconstruction problem. One approach attempts to exploit the symmetry of faces [30], [36]. The advantage of using symmetry is that reconstruction can rely on a mere single image without the need for additional examples of face models. The disadvantage is that point-wise correspondence between the two symmetric portions must be established, and this task is generally difficult. Another method [29] renders faces in novel views by making the restrictive assumption that different faces share the exact same shape while they differ only in albedo.

A widely used approach is to learn the set of allowable reconstructions from a large number of 3D laser-scanned faces. This can be achieved by embedding all 3D faces in a linear space [1], [4], [38], [32] or by using a training set to determine a density function for faces [31], [34]. Similarly, Active Shape Models [7], [10], [21] seek to construct Image-based, linear 2D representations of faces by exploiting large datasets of prototype faces for face recognition and image coding. These methods can achieve accurate reconstructions, but they require a large number (typically hundreds) of face models and a detailed and accurate point-wise correspondence between all the models, as well as expensive parameter fitting. Variations of this approach combine this method with symmetry [8]. Others simplify the parameter fitting procedure [5] by combining surface shape and brightness of all the models into a single coupled statistical model. While this model was shown to generally produce accurate surfaces, it however does not model lighting explicitly, and so it cannot extrapolate to handle novel lighting conditions. Expressing novel faces as combinations of stored 3D faces seems to work very well when the difference between the shape of the novel faces and the stored faces is small. However, in case of larger differences the database needs to be adjusted to fit the particular shapes of the reconstructed faces. For example in case the input face has a smiling expression, the database should include various smiling face shapes.

Unlike previous work we combine *shading* information along with prior knowledge of a *single* reference model to recover the three-dimensional shape of a novel face from a single image. Our method does not use symmetry in the reconstruction process, and it does not require correspondence between many models in a database since it uses a mere single model as a reference. At its core our method solves a shape from shading problem, but it does not assume knowledge of part of the sought 3D model. Our algorithm works with general unknown lighting by representing reflectance using spherical harmonics. We let the gradient in the direction of the normal vanish on the boundaries, and we exploit the reference model to linearize the problem (which leads to increased robustness) and to fill in the missing information – an initial estimate of the albedo and for recovery of the illumination and pose. Finally we regularize the difference between the reference model and the sought 3D shape, instead of smoothing directly the sought shape, which increases the accuracy of the reconstruction.

Most face reconstruction methods assume that faces can accurately be modeled as Lambertian. It was shown in [22] that in many common situations the human skin indeed exhibits nearly Lambertian reflectance properties. Specifically, a face surface was photographed from a sequence of positions and with different lighting directions. Then, by assuming that the face shape and the light source position are known, the photographs were analyzed to determine the bidirectional reflectance function (BRDF). The analysis showed that at incident lighting angles around 30° the BRDF was close to Lambertian. Deviations from the Lambertian reflectance occurred at larger incident angles (above 60°). Specular effects, however, may exist, e.g., when the skin is oily.

3 Problem Formulation

Consider an image $I(x, y)$ of a face defined on a compact domain $\Omega \subset \Re^2$, whose corresponding surface is given by $z(x, y)$. The surface normal at every point is denoted $\vec{n}(x, y) = (n_x, n_y, n_z)^T$ with

$$\vec{n}(x, y) = \frac{1}{\sqrt{p^2 + q^2 + 1}}(p, q, -1)^T, \tag{1}$$

where $p(x, y) = \partial z/\partial x$ and $q(x, y) = \partial z/\partial y$. We assume that the surface of the face is Lambertian with albedo $\rho(x, y)$, and that lighting can be an arbitrary combination of point sources, extended sources and diffuse lighting that need not be known ahead of time. We allow for attached shadows, but ignore the effect of cast shadows and inter-reflections. Under these assumptions it has been shown [2], [27] that Lambertian surfaces reflect only the low frequencies of lighting. Consequently, to an Nth order of approximation, the light reflected by a Lambertian surface (referred to as the *reflectance function*) can be expressed in terms of spherical harmonics as

$$R(x, y) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} l_{nm} \alpha_n Y_{nm}(x, y), \tag{2}$$

where l_{nm} are the coefficients of the harmonic expansion of the lighting, α_n are factors that depend only on n and capture the effect of the Lambertian kernel acting as a low pass filter, so α_n becomes very small for large values of N, and $Y_{nm}(x, y)$ are the surface spherical harmonic functions evaluated at the surface normal. Because the reflectance of Lambertian objects under arbitrary lighting is in general very smooth, this approximation is highly accurate already when a low order (first or second) harmonic approximation is used. Specifically, it has been shown analytically that a first order harmonic approximation (including four harmonic functions, $N = 1$) captures on average at least 87.5% of the energy in an image, while in practice, owing to the fact that only normals facing the camera (the normals with $n_z \geq 0$) are observed, the accuracy seems to approach 95% [11]. A second order harmonic approximation (including nine harmonic functions, $N = 2$) captures on average at least 99.2% of the energy in an image.

For our general formulation, we model below the reflectance function using a second order harmonic approximation, although throughout the text we discuss how it can be modeled also using a first order of approximation and its advantages. In all our experiments both orders produced very similar results. However, the use of a first order approximation results in a significant speedup.

We write the reflectance function in vector notation as

$$R(\vec{n}(x, y); \rho(x, y), \vec{l}\,) \approx \vec{l}\,^T \vec{Y}(\vec{n}(x, y)), \tag{3}$$

with[1]

$$\begin{aligned} \vec{Y}(\vec{n}) = (1, n_x, n_y, n_z, n_x n_y, n_x n_z, n_y n_z, \\ n_x^2 - n_y^2, 3n_z^2 - 1)^T \end{aligned} \tag{4}$$

where n_x, n_y, n_z are the components of the surface normal $\vec{n}$. The image irradiance equation is then expressed as

$$I(x, y) = \rho(x, y) R(x, y). \tag{5}$$

In the first order approximation only the first four components of $\vec{Y}(\vec{n})$ are included and consequently $\vec{l}$ is a four component vector. We will show in Section 4 that using the first order approximation we can transform the reconstruction problem to be *linear* in the depth

[1] Formally in (2) the values of α_n should be set to $\alpha_0 = \pi, \alpha_1 = \frac{2\pi}{\sqrt{3}}, \alpha_2 = \frac{2\pi}{\sqrt{8}}$, and the spherical harmonics functions are $Y = (c_0, c_1 n_x, c_1 n_y, c_1 n_z, c_2 n_x n_y, c_2 n_x n_z, c_2 n_y n_z, \frac{c_2}{2}(n_x^2 - n_y^2), \frac{c_2}{2\sqrt{3}}(3n_z^2 - 1))^T$, where $c_0 = \frac{1}{\sqrt{4\pi}}, c_1 = \frac{\sqrt{3}}{\sqrt{4\pi}}$ and $c_2 = \frac{3\sqrt{5}}{\sqrt{12\pi}}$. For simplicity of notation we omit these constant factors and rescale the lighting coefficients to account for the omitted factors.

variables, yielding a solution that is both fast and very robust.

Note that while the first four harmonics resemble in form to the reflectance obtained by illuminating a surface by a point source and ambient lighting, it is still providing an accurate approximation for a variety of other lighting conditions. Consequently, while in the former case surface patches whose normal are 90° or more from the source are only exposed to the ambient component, and so their surface orientation cannot be recovered (since intensity in these points is independent of normal direction), with a different underlying lighting (possibly with multiple sources) the surface can in principle be reconstructed beyond 90° of the mode of the reflectance function.

To supply the missing information we use either a reference model of a face of a different individual or by a generic face model. Let $z_{\mathrm{ref}}(x,y)$ denote the surface of the reference face with $\vec{n}_{\mathrm{ref}}(x,y)$ denoting the normal to the surface and $\rho_{\mathrm{ref}}(x,y)$ denote its albedo. We use this information to determine the lighting and provide initial guess for the sought albedo.

We further use the reference model to regularize the problem. To that end we define the difference shape as

$$d_z(x,y) = z(x,y) - z_{\mathrm{ref}}(x,y) \tag{6}$$

and the difference albedo as

$$d_\rho(x,y) = \rho(x,y) - \rho_{\mathrm{ref}}(x,y) \tag{7}$$

and require these differences to be smooth. We are now ready to define our optimization function:

$$\min_{\vec{l},\rho,z} \int_\Omega (I - \rho \vec{l}^{\,T} \vec{Y}(\vec{n}))^2 + \lambda_1 (\triangle G * d_z)^2 + \lambda_2 (\triangle G * d_\rho)^2 \; dxdy, \tag{8}$$

where $\triangle G*$ denotes convolution with the Laplacian of a Gaussian, and λ_1 and λ_2 are positive constants. Below we will refer to the first term in this integral as the "data term" and the other two terms as the "regularization terms". Evidently, without regularization the optimization functional (8) is ill-posed. Specifically, for every choice of depth $z(x,y)$ and lighting vector $\vec{l}$ it is possible to prescribe albedo $\rho(x,y)$ to make the data term vanish. With regularization and appropriate boundary conditions the problem becomes well-posed. Note that we chose to regularize d_z and d_ρ rather than z and ρ in order to preserve the discontinuities in z_{ref} and ρ_{ref}. (This regularization is a much weaker constraint than requiring that the sought shape is smooth.)

4 Reconstruction Steps

We assume that the input image is roughly aligned to the reference model and approach this optimization by solving for lighting, depth, and albedo separately. First, we recover the spherical harmonic coefficients $\vec{l}$ by finding the best coefficients that fit the reference model to the image. This is analogous to solving for pose by matching the features of a generic face model to the features extracted from an image of a different face. Next we solve for depth $z(x,y)$. For this we use the recovered coefficients along with the albedo of the reference model, and prescribe appropriate boundary conditions. Finally, we use the spherical harmonics coefficients and the recovered depth to estimate the albedo $\rho(x,y)$. This procedure can be repeated iteratively, although in our experiments one iteration seemed to suffice. These steps are described in detail in the remainder of this section.

The use of the albedo of the reference model in the reconstruction step may seem restrictive since different people may vary significantly in skin color. Nevertheless, it can be readily verified that scaling the albedo (i.e., $\beta\rho(x,y)$, with a scalar constant β) can be compensated for by scaling appropriately the light intensity. Our albedo recovery, consequently, will be subject to this ambiguity. Also to make sure that marks on the reference face would not influence much the reconstruction we first smooth the albedo of the reference model by a Gaussian.

4.1 Step 1: Recovery of lighting coefficients

In the first step we attempt to recover the lighting coefficients $\vec{l}$, by fitting the reference model to the image. To this end, we substitute in (8) $\rho \rightarrow \rho_{\mathrm{ref}}$ and $z \rightarrow z_{\mathrm{ref}}$ (and consequently $\vec{n} \rightarrow \vec{n}_{\mathrm{ref}}$). At this stage both regularization terms vanish, and only the data term remains:

$$\min_{\vec{l}} \int_\Omega (I - \rho_{\mathrm{ref}} \, \vec{l}^{\,T} \vec{Y}(\vec{n}_{\mathrm{ref}}))^2 \; dxdy. \tag{9}$$

Discretizing the integral we obtain

$$\min_{\vec{l}} \sum_{(x,y)\in\Omega} \Big(I(x,y) - \rho_{\mathrm{ref}}(x,y) \vec{l}^{\,T} \vec{n}_{\mathrm{ref}}(x,y) \Big)^2 . \tag{10}$$

This is a highly over-constrained linear least squares optimization with only nine or four unknowns (the components of $\vec{l}$; the dimension of this vector depends on the order of approximation used) and can be solved simply using the pseudo-inverse. The coefficients recovered with this procedure will be used subsequently to recover the depth and albedo of the face. It should be noted that to avoid degeneracies the input face must be lit by non-ambient light, since under ambient light intensities are independent of surface orientation. In Section 5 we show that, in practice, the error of recovering lighting by using the 3D face model (shape and albedo) of a different individual is sufficiently small (the mean angle is 4.9° with standard deviation of 1.2°).

4.2 Step 2: Depth recovery

At this stage we have obtained an estimate for $\vec{l}$. We continue using $\rho_{\mathrm{ref}}(x,y)$ for the albedo and turn to recovering $z(x,y)$. Below we will further exploit the reference face to simplify the data term. We start by writing explicitly the expression for $\vec{Y}(\vec{n})$ using the second order approximation to reflectance, and by representing the surface normal $\vec{n}(x, y)$ using partial derivatives $p(x, y), q(x, y)$ as in Eq. (1)

$$\vec{Y}(\vec{n}) = (1,\ \frac{1}{N}p,\ \frac{1}{N}q,\ \frac{-1}{N},\ \frac{1}{N^2}pq,\ \frac{-1}{N^2}p,\ \frac{-1}{N^2}q, \\ \frac{1}{N^2}(p^2 - q^2),\ \frac{3}{N^2} - 1)^T \tag{11}$$

where $N(x, y) = \sqrt{p^2 + q^2 + 1}$. We will assume that $N(x, y) \approx N_{\text{ref}}(x, y)$. The data term then minimizes the squared difference between the two sides of the following system of equations

$$\begin{aligned} I = \rho_{\text{ref}} l_0 &+ \frac{\rho_{\text{ref}}}{N_{\text{ref}}}(l_1 p + l_2 q - l_3) \\ &+ \frac{\rho_{\text{ref}}}{N_{\text{ref}}^2}(l_4 pq - l_5 p - l_6 q + l_7 p^2 - l_7 q^2 + 3l_8) \\ &- \rho_{\text{ref}} l_8, \end{aligned} \tag{12}$$

with p and q as the only unknowns for each $(x, y) \in \Omega$. In discretizing this system of equations we will use $z(x, y)$ as our unknowns, and replace p and q by the forward differences:

$$\begin{aligned} p &= z(x + 1, y) - z(x, y) \\ q &= z(x, y + 1) - z(x, y). \end{aligned} \tag{13}$$

The data term thus provides one equation for every unknown (except for the pixels on the boundary of Ω). Note that by solving directly for $z(x, y)$ we in fact enforce consistency of the surface normals ("integrability"). Let us now investigate equation (12). In case we consider the first order of approximation to reflectance, this equation becomes

$$I = \rho_{\text{ref}}\ l_0 + \frac{\rho_{\text{ref}}}{N_{\text{ref}}}(l_1 p + l_2 q - l_3). \tag{14}$$

By substituting (14) for p and q (14) we can see that this equation is linear in $z(x, y)$

$$\begin{aligned} I = \rho_{\text{ref}}\ l_0 + \frac{\rho_{\text{ref}}}{N_{\text{ref}}}\,&(l_1 z(x + 1, y) - l_1 z(x, y) \\ &+ l_2 z(x, y + 1) - l_2 z(x, y) - l_3, \end{aligned} \tag{15}$$

and so it can be solved using linear least squares optimization. In case we use the second order of approximation, the data equation (12) is non-linear in $z(x, y)$ and therefore requires a nonlinear optimization procedure.

Next we consider the regularization term $\lambda_1 \triangle G * d_z$. (The second regularization term vanishes at this stage since we have substituted ρ_{ref} for ρ.) We implement this term as the difference between $d_z(x, y)$ and the average of d_z around (x, y) obtained by applying a Gaussian function to d_z. Consequently, this term minimizes the difference between the two sides of the following system of equations

$$\lambda_1(z(x, y) - G * z(x, y)) = \lambda_1(z_{\text{ref}}(x, y) - G * z_{\text{ref}}(x, y)). \tag{16}$$

This system too is linear in $z(x, y)$.

4.3 Boundary conditions for depth recovery

The equations for the data and regularization terms provide two equations for every unknown $z(x, y)$. In the case of first order approximation both equations, (14) and (16), are linear. In the case of a second order approximation one equation is linear (16) while the other is nonlinear (12). This system of equations however is still ill-posed and we need to add boundary conditions.

Shape from shading methods typically use Dirichlet boundary conditions, which require prior knowledge of the the depth values $z(x, y)$ along the boundary of the surface. In addition, these methods require knowledge of the depth values at all the local extremal points inside the bounded surface (e.g., in case of a face these can include the centers of the eyes, the tip of the nose and the center of the mouth). Due to the use of a reference shape our algorithm does not require knowledge of the inner extremal points. However, since our data term includes partial derivatives of $z(x, y)$, we do need a constraint for the exterior boundary of the surface. Since we have a reference model a sensible approach is to use its depth values $z_{\text{ref}}(x, y)$ as Dirichlet boundary conditions, or, alternatively, the derivatives of the reference along the boundaries as Neumann boundary conditions. These constraints however are too restrictive since the depth values of the reference model and their derivatives may be incompatible with the sought solution.

Instead, to obtain boundary conditions we assume in our algorithm that the gradient of the surface in the direction perpendicular to the exterior boundary vanishes (i.e., the surface is planar near the boundaries; note that this does not imply that the entire bounding contour is planar). Specifically, we add for each boundary point the following constraint

$$\nabla z(x, y) \cdot \vec{n_c}(x, y) = 0. \tag{17}$$

where $\vec{n_c}(x, y)$ is a two-dimensional vector representing the normal to the bounding contour. These constraints will be roughly satisfied if the boundaries are placed in slowly changing parts of the face. They will be satisfied for example when the boundaries are placed along the cheeks and the forehead, but will not be satisfied when the boundaries are placed along the eyebrows, where the surface orientation changes rapidly. Similar boundary conditions were used in [6] in the context of photometric stereo.

Finally, since all the equations we use for the data term, the regularization term, and the boundary conditions involve only partial derivatives of $z(x, y)$, while $z(x, y)$ itself is absent from these equations, the solution can be obtained only up to an additive factor. We will rectify this by arbitrarily setting one point to $z(x_0, y_0) = z_{\text{ref}}(x_0, y_0)$.

4.4 Step 3: Estimating albedo

Once both the lighting and depths are recovered, we may turn to estimating the albedo. Using the data term the

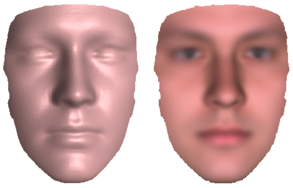

Fig. 2. The generic face model obtained by taking the mean shape over the entire USF database.

albedo $\rho(x, y)$ is found by solving the following equation

$$I(x, y) = \rho(x, y)\vec{l}^{\,T}\vec{Y}(\vec{n}). \tag{18}$$

The first regularization term in the optimization functional (8) is independent of ρ, and so it can be ignored. The second term optimizes the following set of equations

$$\lambda_2 \triangle G * \rho = \lambda_2 \triangle G * \rho_{\text{ref}}. \tag{19}$$

These provide a linear set of equations, in which the first set determines the albedo values, and the second set smoothes these values. We avoid the need to determine boundary conditions simply by terminating the smoothing process at the boundaries.

5 Experiments

We tested our algorithm on images taken under controlled viewing conditions by rendering images of faces from the USF face database [16]. We further tested our algorithm on images taken from the YaleB face database [12], on images of celebrities downloaded from the internet, and on images photographed by us.

5.1 Experimental setup

For the experiment with the USF face database we used 77 face models. These models include depth and texture maps of real faces (male and female adult faces with a mixture of race and ages) obtained with a laser scanner. We used the provided texture maps as albedos. These contain noticeable effects of the lighting conditions, and hence could possibly introduce some errors to our reconstructions. To render an image we illuminated a model simultaneously by three point sources from directions $\vec{l_i} \in \mathbb{R}^3$ and with intensity L_i. According to the Lambertian Law the intensities reflected by the surface due to this light are given by $\rho \sum_i L_i \max(\vec{n}^T \vec{l_i}, 0)$. We also used the 3D faces from this database as reference faces, either by using each of the faces as a reference or by using a generic face obtained by taking the mean shape over the entire database (Fig. 2).

The YaleB face database includes images of faces taken under different viewing conditions (lighting and pose), which we used as input images. To evaluate our reconstructions we also reconstructed each face shape using a photometric stereo method. This was possible since the YaleB database includes many images of each face taken from the same viewpoint but illuminated with varying point source lightings, and the lighting directions are provided. We used 10 such images for photometric stereo reconstruction of each face. The rest of the experiments were made with input images that were downloaded from the internet or that were photographed by us and hence we did not have a laser scan or a photometric stereo reconstruction available.

For the cases when a laser scan or a photometric stereo reconstruction is available, the accuracy of our reconstructions is demonstrated by the presented error maps. The error maps were calculated per pixel as $|z(x, y) - z_{\text{gt}}(x, y)|/z_{\text{gt}}(x, y)$, where $z_{\text{gt}}(x, y)$ denotes the "ground-truth" depth values (laser scan or photometric stereo). In addition, under each error map we present the overall mean and standard deviation values each multiplied by 100 to indicate percents. In the cases that a laser scan or photometric stereo are not available the accuracy of our reconstruction can be evaluated only visually.

We assume alignment between the face in the input image and the reference model. With misalignment the reconstruction results degrade, mainly when the boundaries (occluding contours) of the face in the input image and the reference face are not aligned. To achieve alignment we first use marked points to determine a rigid transformation between the reference model and the input image. We then refine this alignment by further applying an optical flow algorithm. We begin by marking five corresponding points on the input face and on the reference face, two at the centers of the eyes, one on the tip of the nose, one at the center of the mouth and one at the bottom of the chin. In the case of frontal faces we then use these correspondences to determine a 2D rigid transformation to fit the image to the reference model.

For non-frontal faces we apply an additional procedure to recover a rough approximation of the 3D rigid transformation that transforms the reference model to the orientation of the face in the input image using the same marked five points. Specifically, let the 2×5 matrix p' denote the points marked on the input image, and let the 3×5 matrix P' denote the points marked on the reference model. We first subtract the mean from each of the matrices to get $p = p' - \bar{p}'$ and $P = P' - \bar{P}'$. We then let $A = pP^T(PP^T)^{-1}$ and $t = \bar{p}' - A\bar{P}'$, where A is 2×3 and t is a 2-vector. To apply this transformation in 3D we augment A and t by fitting a 3×3 scaled rotation matrix to A, and setting the third component of t be 0. Such a transformation may cause occlusion of parts of the reference shape model. To remove the occluded points we use the approximate z-buffer technique for point sets by [18]. This method examines groups of points that project to the same neighborhood of the depth map, and identifies those points that are closer to the viewer.

Finally to achieve finer alignment we apply the optical flow algorithm of [3] (with default parameters). In particular, after finding the spherical harmonics coefficients $\vec{l}$ we produce a reference image using these coefficients and the reference model. We then find the flow between the reference and input images. This flow is applied to the reference model that is used in the reconstruction algorithm. After the alignment procedure all the images are of size 360×480 pixels.

The following parameters were used throughout all our experiments. The reference albedo was kept in the range between 0 and 255. Both λ_1 and λ_2 were set to 30. For the regularization we used a 2-D Gaussian with $\sigma_x = \sigma_y = 3$ for images downloaded from the web and $\sigma_x = \sigma_y = 2$ for all the rest of the images. Our MATLAB implementation of the algorithm takes about 9 seconds on a quad-code AMD processor 2354 1100Mhz Linux workstation, and the optical flow implementation takes another 20 seconds.

This concludes the general setup. In the remainder of this section we describe the experiments made on images rendered from the USF database as well as images from the YaleB database, images downloaded from the internet, and images photographed by us.

5.2 Images rendered from the USF database

In Figure 3 we present the results of the reconstruction obtained for images rendered from the USF models. We show eight examples. In each example we present the input image and two triplets (two viewpoints) of shapes: the reference shape, the ground truth (laser scan) shape and our reconstructed shape. We also present two error maps: between the laser scan and the reference shape (on the left) and between the laser scan and the reconstructed shape (on the right). The numbers under each of the error maps indicate the mean error and the standard deviation (in percents). In this experiment in all reconstructions we used the same reference face – a generic face obtained by taking the mean shape over the entire USF database.

We observe that the recovered shapes are consistently very similar to the laser scan shapes (visually and also by examining the error values). By comparing the two error maps in each example we can see how the reference shape was modified by the algorithm to fit the image. In most cases the reference face was molded to fit the correct shape very closely overcoming in some cases significant differences in shape (see,e.g., the second, fifth and eighth rows) and in facial expression (see the seventh row). Occasional errors, however, remain in some of the cases particularly near facial features.

In Figure 4 we show the overall mean reconstruction error for each of the 77 faces in the USF database when we use as reference the mean face (upper plot), and when each face is reconstructed with each of the rest of the models in the database serving as a reference model (bottom plot). The red squares mark the difference between the reference and ground truth shapes, and the blue diamonds mark the errors between the reconstruction and ground truth. We can see that in all cases the reconstruction errors are smaller than the differences between the reference model and the ground-truth scans.

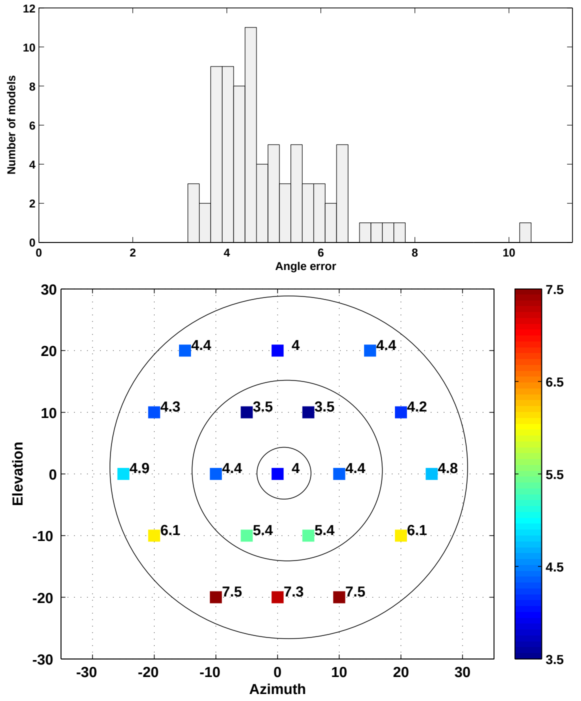

Fig. 5. The mean angle between the true lighting and the lighting recovered with different reference models (shape and albedo). The error was calculated over a database of 77 shapes, over 19 different lightings. The histogram shows the number of models that produced each value of error. The plot shows the average error for each ground truth lighting direction (the color code goes from blue to red representing low to high values). The total mean and standard deviation of this error is $4.9^o \pm 1.2^o$. The Azimuth, Elevation and the error values are in degrees.

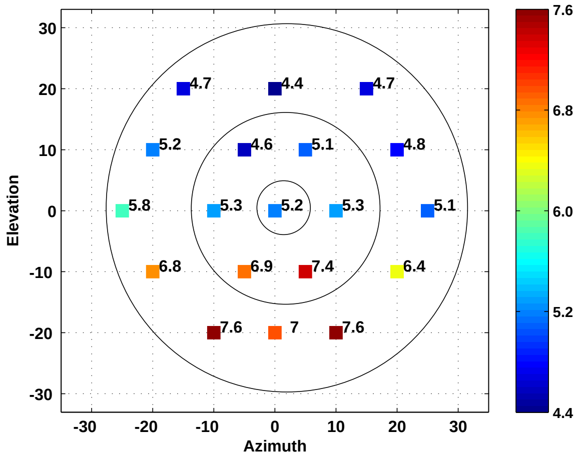

Fig. 6. Mean reconstruction error with 19 different point source lightings (computed over 13 different face shapes). Each point marks a lighting direction and the color and number next to it represent the reconstruction error (in percents). Azimuth and Elevation are in degrees.

Indeed the overall means and standard deviations of the reconstruction error are 4.2 ± 1.2 and 6.5 ± 1.4 in the top

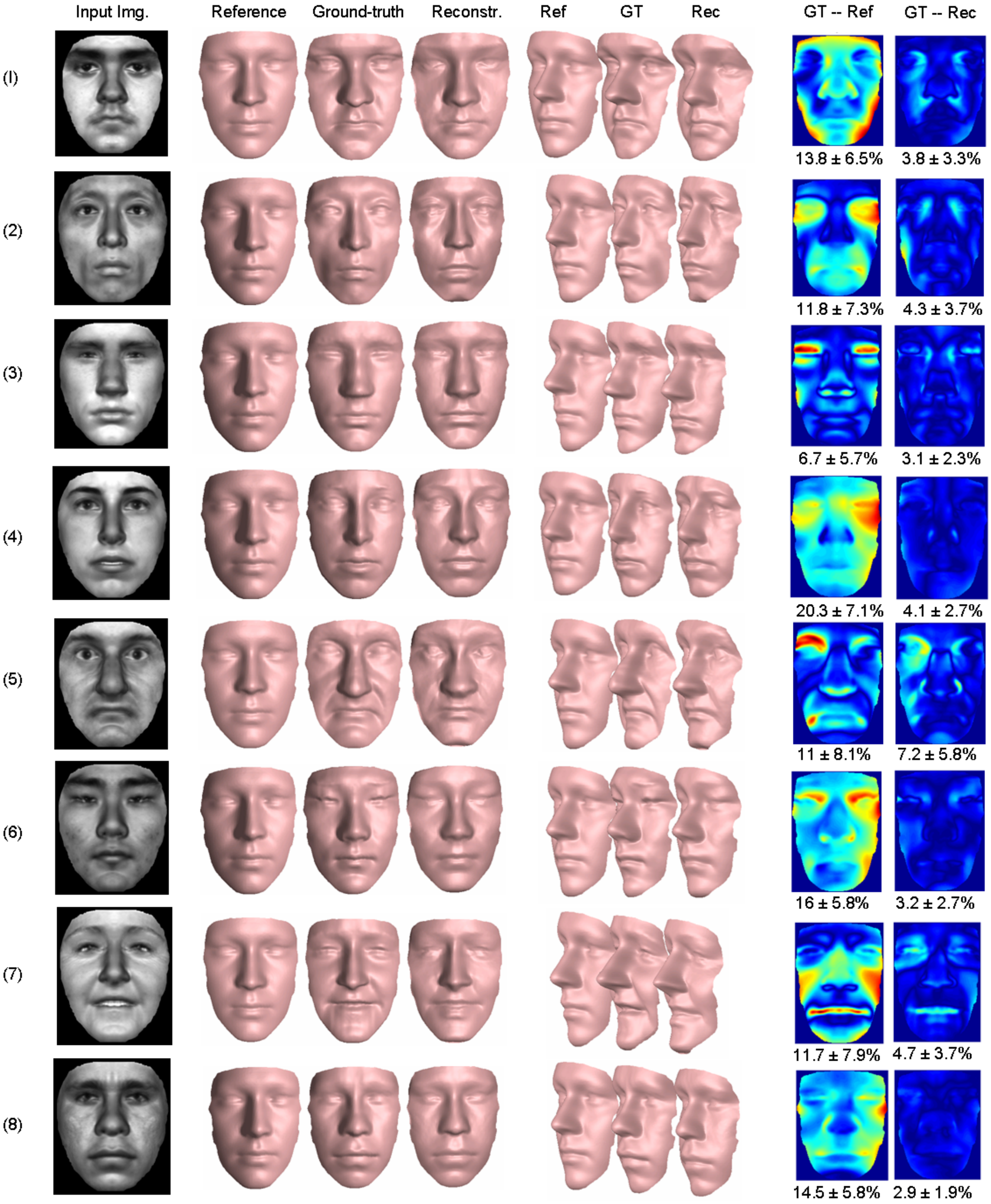

Fig. 3. Eight examples of typical reconstructions from the USF database. In each example we show from left to right the input image, a triplet of shapes (the reference model, the ground truth which is the laser scan and our reconstruction) in two different viewpoints and depth error maps ($100 \cdot |z(x,y) - z_{\mathrm{GT}}(x,y)|/z_{\mathrm{GT}}(x,y)$) between the reference model and the ground truth (left) and between our reconstruction and ground truth (right). The colormap goes from dark blue to dark red (corresponding to an error between 0 and 40). The numbers under each of the error maps represent mean and standard deviation values in percents.

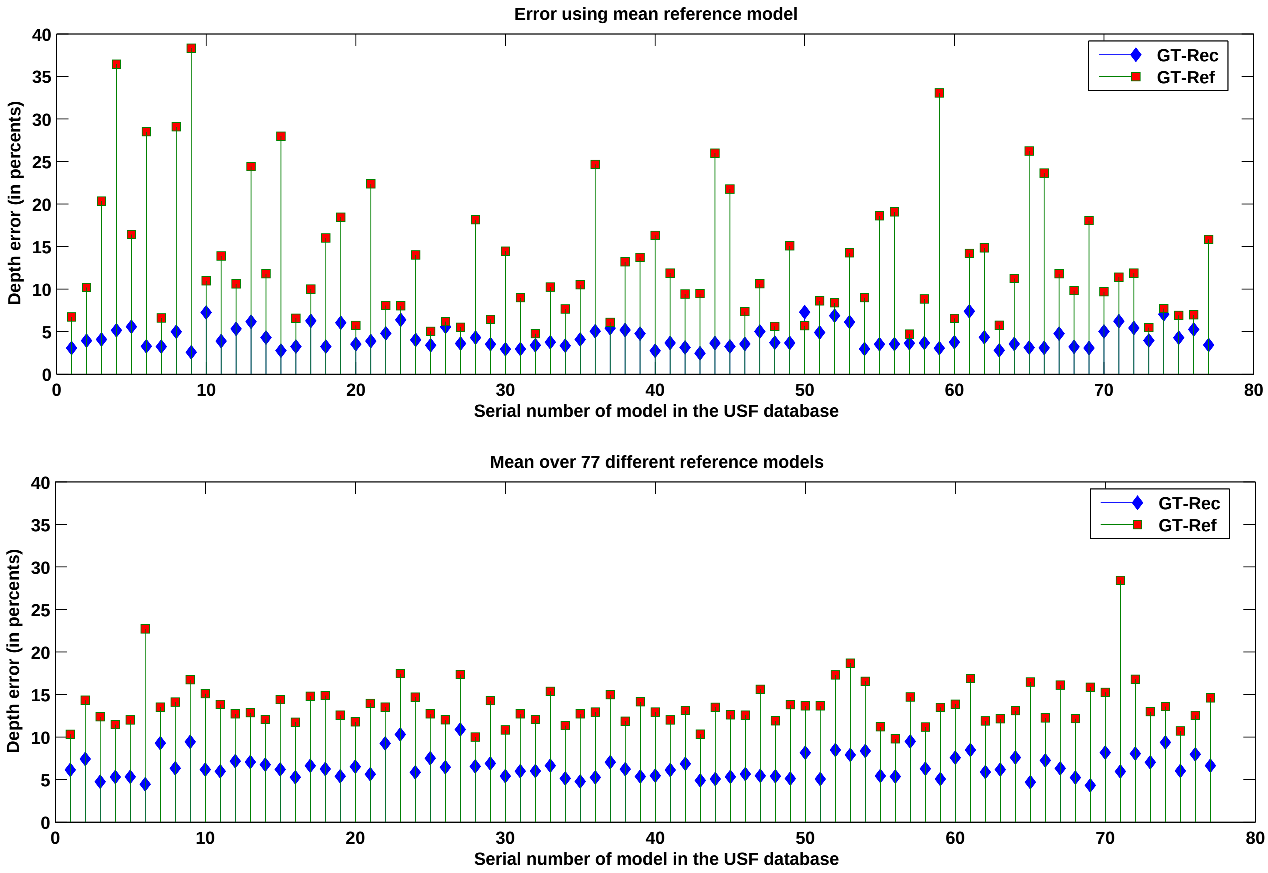

Fig. 4. Mean depth error calculated for each of the 77 models from the USF database with the generic model used as reference (top plot) and with each of the USF models used as reference (average over all the models, bottom plot). In each plot two measures are displayed: the error between the reference and the ground-truth shapes (red squares) and the error between the reconstructed and the ground-truth shapes (blue rhombuses). The overall means and standard deviations in the top plot are 4.2 ± 1.2 and 12.9 ± 7.9 for the GT-Rec and GT-Ref errors respectively. In the bottom plot the overall means and standard deviations are 6.5 ± 1.4 and 13.8 ± 2.8 for the GT-Rec and GT-Ref errors respectively.

and bottom plots respectively, whereas the overall means and standard deviations of the difference between the reference model and the ground-truth are 12.9 ± 7.9 and 13.8 ± 2.8.

We further examined the accuracy of the process by fitting a 2D image to a 3D reference face model (shape and albedo) of a different individual. We have run the following experiment. We first rendered 19 images of each of the face models in the database, each was rendered with a single point light source. We then recovered the lighting from each image by comparing it to all the other 3D models in the database. We calculated for each such pair the angle between the true lighting and the recovered one; this represents the error in lighting recovery. The result of the experiment is shown in Fig. 5. Azimuth and elevation are the angular displacements in degrees from the y-axis and from the y-z plane, respectively. We observe from the histogram that the mean angle is $4.9°$ with standard deviation of $1.2°$, which is sufficiently small. The plot in this figure shows how the error changes for different point source lightings. It appears that negative elevation values tend to produce higher errors. Below we also observe that the reconstruction error experiences similar behaviour.

Figure 6 shows how the reconstruction error varies with a different choice of lighting direction. For this experiment we rendered images of 13 different face models with 19 point source lightings, and calculated the mean reconstruction error for each lighting direction. We observe that the error is higher for negative elevation values. This may be attributed to the errors in the light recovery in Step 1 of the algorithm, which exhibit a similar pattern (see Figure 5).

5.3 Images outside the USF database

In Figure 7 we present six example reconstructions of faces from the YaleB database. To evaluate our reconstructions we additionally reconstructed each face shape using a photometric stereo algorithm. For each example we present the input image, and two triplets (two viewpoints) of shapes: the reference shape, the photometric stereo reconstruction and our reconstruction. We further present the two error maps, between the reference model and the photometric stereo reconstruction and between the reconstruction with our algorithm and the photometric stereo reconstruction. The numbers under each of the difference maps indicate the overall mean and standard deviation. We observe that our algorithm achieved good reconstructions, overcoming significant differences between individuals. Examples one and six also demonstrate that our algorithm is robust to moderate amounts of facial hair, although it is not designed to handle facial hair.

In Figure 8 we further apply our algorithm to images of non-frontal faces from the YaleB database. We present reconstructions of three faces viewed in five viewpoints. For each face and pose we show the reconstructed shape and image overlay of the reconstruction. Under each reconstruction we further present the overall mean and standard deviation of the difference between our reconstruction and the photometric stereo. In each case the face is reconstructed from a single image. In principle, such reconstructions can be combined together using 3D alignment techniques to produce a fuller 3D shape of a person's face and by this overcome the visibility issues that arise in single image reconstruction (e.g., part of the face can be occluded in a single view).

Finally in Figure 9 we present results of our algorithm on eight images that were downloaded from the internet and two more images that were photographed by us. An additional result is shown in Figure 1. While we do not have the ground truth shapes in these experiments, we can still see that convincing reconstructions are obtained. Note especially the reconstruction of Tom Hanks' face obtained with a smile (top right), the wrinkles present in the reconstructions of Clint Eastwood and Samuel Beckett (2nd row on the right and 3rd row on the left), the reconstruction of the painted Mona Lisa (2nd row on the left), the shape details of Gerard Depardieu, and finally the two reconstructions from images we photographed that include two facial expressions of the same person.

6 Conclusion

In this paper we have presented a novel method for 3D shape reconstruction of faces from a *single* image by using only a *single* reference model. Our method exploits the global similarity of faces by combining shading information with generic shape information inferred from a single reference model and by this overcomes some of the main difficulties in previous reconstruction methods.

Unlike existing methods, our method does not need to establish correspondence between symmetric portions of a face, nor does it require to store a database of many faces with dense correspondences across the faces. Nevertheless, although this paper emphasizes the use of a single model of a face to reconstruct another face, we note that our method can supplement methods that make use of multiple models in a database. In particular, we may select to ”mold” the model from a database

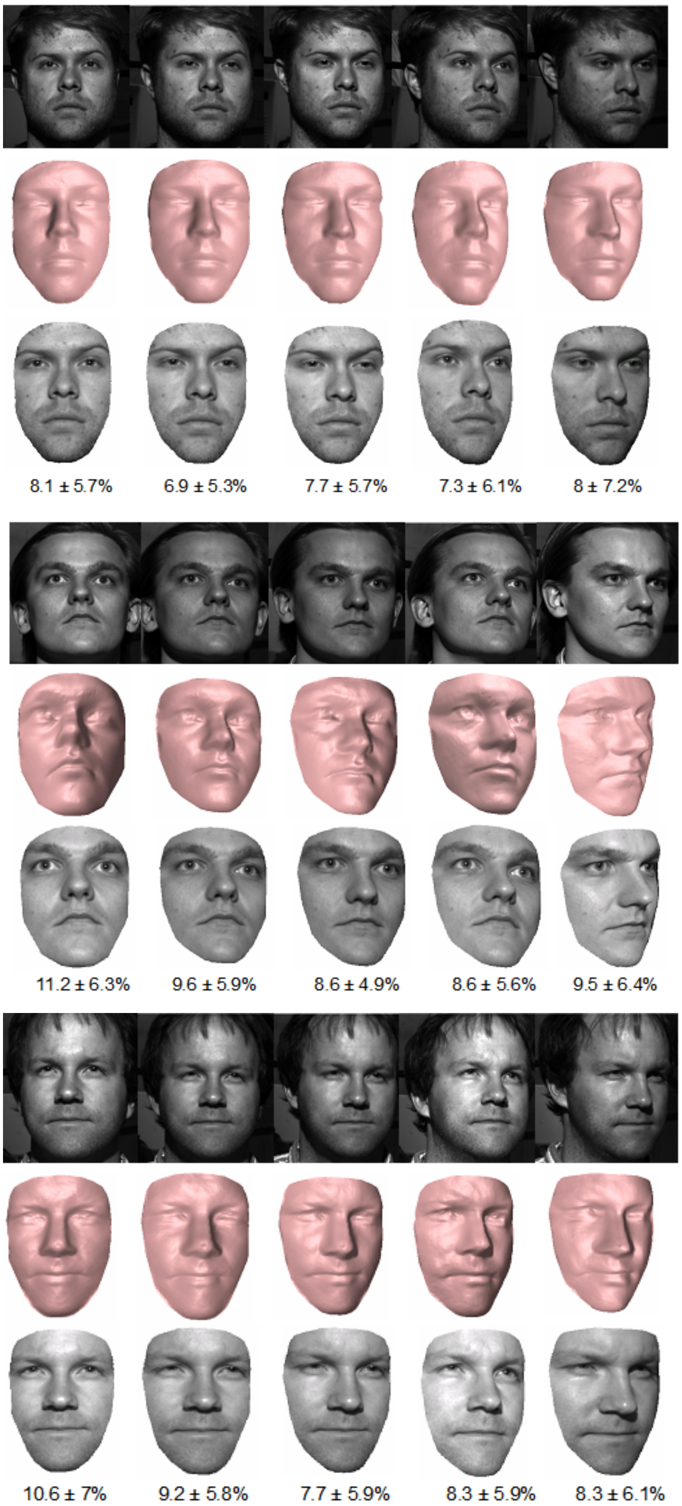

Fig. 8. Reconstruction of three faces from the YaleB database, from 5 different poses. In each example we present from top to bottom the input image, the reconstructed shape and an image overlay on the reconstructed shape. The numbers below each reconstruction show the mean and standard deviation of the depth difference between our reconstruction and a photometric stereo reconstruction.

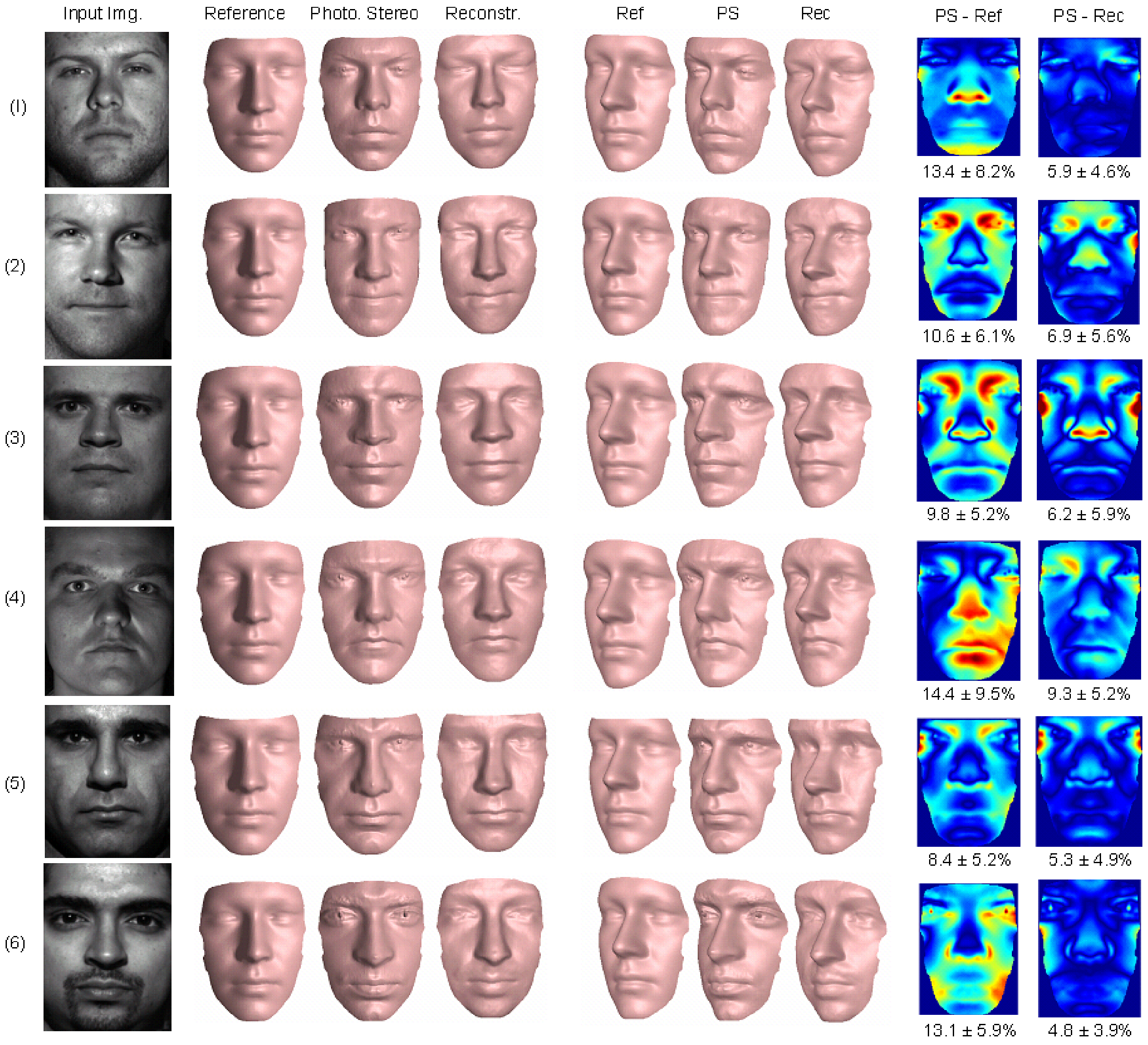

Fig. 7. Six reconstruction examples from the YaleB database. In each example we show from left to right the input image, a triplet of shapes (reference model, reconstruction by photometric stereo and our reconstruction) in two different viewpoints, and depth error maps ($100 \cdot |z(x,y) - z_{\rm PS}(x,y)|/z_{\rm PS}(x,y)$) between the reference model and the photometric stereo reconstruction (left) and between our reconstruction and the photometric stereo reconstruction (right). The colormap goes from dark blue to dark red (corresponding to an error between 0 and 40). The numbers under each of the error images represent the means and standard deviations of these differences in percents.

shape that best fits the input image. Alternatively, we may choose the best fit model from a linear subspace spanned by the database, or we may choose a model based on some probabilistic criterion. In all cases our method will try to improve the reconstruction by relying on the selected model.

Our method handles unknown lighting, possibly coming from combination of multiple unknown light sources. We allow for attached shadows, however we ignore cast shadows. In theory estimation of cast shadows can be incorporated in our method, by finding an initial estimate of shadows using the surface geometry of the reference model and then by iterating our reconstruction procedure to solve for the unknown depth and locations of cast shadows concurrently.

We tested our method on a variety of images, and our experiments demonstrate that the method was able to accurately recover the shape of faces overcoming significant differences across individuals including differences in race, gender and even variations in expressions. Furthermore we showed that the method can handle a variety of uncontrolled lighting conditions, and that

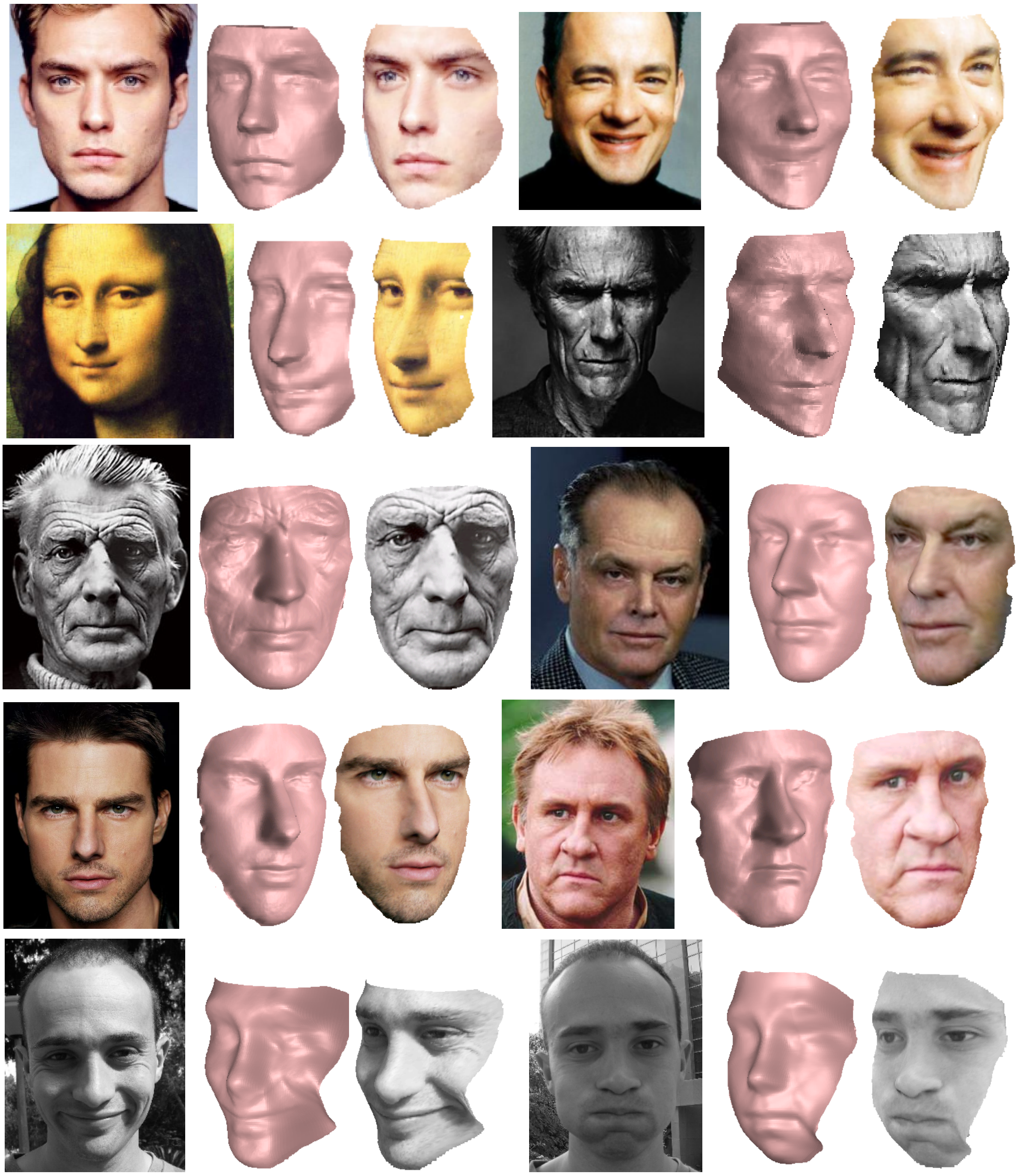

Fig. 9. Reconstruction results on images of celebrities downloaded from the internet and two images photographed by us (bottom row). In each example we present the input image, our 3D shape reconstruction, and an image overlay on the reconstructed shape.

it can achieve consistent reconstructions with different reference models.

As our method uses a single reference model for reconstruction, it would be interesting to see if a similar approach can be constructed for objects other than faces, in which a 3D prototype object is provided and used for reconstruction of similar, yet novel shapes from single images.

Acknowledgment

Research was supported in part by the Israel Science Foundation grant number 266/02 and by the European Commission Project IST-2002-506766 Aim Shape. The vision group at the Weizmann Institute is supported in part by the Moross Laboratory for Vision Research and Robotics.

References

[1] J.J. Atick, P.A. Griffin, and A.N. Redlich. Statistical approach to shape from shading: Reconstruction of 3d face surfaces from single 2d images. *Neural Comp.*, 8(6):1321–1340, 1996.

[2] R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, 2003.

[3] M.J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding (CVIU)*, 63(1):75–104, 1996.

[4] V. Blanz and T.A. Vetter. A morphable model for the synthesis of 3d faces. *SIGGRAPH*, I:187–194, 1999.

[5] M. Castelan, W.A.P. Smith, and E.R. Hancock. A coupled statistical model for face shape recovery from brightness images. *Trans. on Image Processing*, 16(4):1139–1151, 2007.

[6] M.K. Chandraker, S. Agarwal, and D.J. Kriegman. Shadowcuts: Photometric stereo with shadows. *CVPR*, 2007.

[7] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *In Burkhardt and Neumann, editors, Computer Vision ECCV98, Springer, Lecture Notes in Computer Science 1407*, 2, 1998.

[8] R. Dovgard and R. Basri. Statistical symmetric shape from shading for 3d structure recovery of faces. *ECCV*, 2004.

[9] P. Dupuis and J. Oliensis. An optimal control formulation and related numerical methods for a problem in shape reconstruction. *The Annals of Applied Probability*, 4(2):287–346, 1994.

[10] G.J. Edwards, A. Lanitis, C.J. Taylor, and T.F. Cootes. Modelling the variability in face images. *In Proc. of the 2nd Int. Conf. on Automatic Face and Gesture Recognition, IEEE Comp. Soc.*, 2, 1996.

[11] D. Frolova, D. Simakov, and R. Basri. Accuracy of spherical harmonic approximations for images of lambertian objects under far and near lighting. *ECCV*, pages 574–587, 2004.

[12] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 23(6):643–660, 2001.

[13] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multi-view photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008.

[14] B.K.P. Horn. *Obtaining Shape from Shading Information*. The Psychology of Computer Vision. McGraw-Hill, New York, 1975.

[15] B.K.P. Horn and M.J. Brooks, editors. *Shape from Shading*. MIT Press: Cambridge, MA, 1989.

[16] http://www.csee.usf.edu/sarkar. *USF DARPA Human-ID 3D Face Database, Courtesy of Prof. Sudeep Sarkar, University of South Florida, Tampa, FL.*

[17] T.M Hursh. The study of cranial form: Measurement techniques and analytical methods. *The Measures of Man. E. Giles and J. Fiedlaender, eds.*, 1976.

[18] S. Katz, A. Tal, and R. Basri. Direct visibility of point sets. *ACM Transactions on Graphics (SIGGRAPH)*, 26(3):24/1–11, 2007.

[19] I. Kemelmacher and R. Basri. Molding face shapes by example. *ECCV*, 1:277–288, 2006.

[20] R. Kimmel and J.A. Sethian. Optimal algorithm for shape from shading and path planning. *Journal of Mathematical Imaging and Vision*, 14(3):237–244, 2001.

[21] A. Lanitis, C.J. Taylor, and T.F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19.

[22] S.R. Marschner, S.H. Westin, E.P.F Lafortune, K.E. Torrance, and D.P.Greenberg. Image-based brdf measurement including human skin. *In 10thEurographics Workshop on Rendering*, pages 139–152, 1999.

[23] Y. Moses, S. Edelman, and S. Ullman. Generalization to novel images in upright and inverted faces. *Perception*, 25:443–461, 1996.

[24] A.P. Pentland. Finding the illuminant direction. *Journal Optical Society of America*, pages 448–455, 1982.

[25] V. Ramachandran. Perception of shape from shading. *Nature*, 331:163–166, 1988.

[26] V. S. Ramachandran. Visual perception in people and machines. *AI and the Eye, A.Blake and T. Troscianko, eds.*, pages 21–77, 1990.

[27] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: Determining the illumination from images of a convex lambertian object. *JOSA*, 18(10):2448–2459, 2001.

[28] E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM Journal of Numerical Analysis*, 29(3):867–884, June 1992.

[29] A. Shashua and T. Riklin-Raviv. The quotient image: Class based re-rendering and recognition with varying illuminations. *PAMI*, 23(2):129–139, 2001.

[30] I. Shimshoni, Y. Moses, and M. Lindenbaum. Shape reconstruction of 3d bilaterally symmetric surfaces. *IJCV*, 39(2):97–100, 2000.

[31] T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. *CVPR Workshop on Models versus Exemplars*, 2001.

[32] W.A.P Smith and E.R. Hancock. Recovering facial shape and albedo using a statistical model of surface normal direction. *ICCV*, 2005.

[33] P.S. Tsai and M. Shah. Shape from shading with variable albedo. *Optical Engineering*, pages 121–1220, April 1998.

[34] L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. *CVPR*, 2003.

[35] R. Zhang, P.S. Tsai, J.E. Cryer, and M. Shah. Shape from shading: A survey. *PAMI*, 21(8):690–706, 1999.

[36] W. Zhao and R. Chellappa. Symmetric shape-from-shading using self-ratio image. *IJCV*, 45:55–75, 2001.

[37] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *PAMI*, 13(7):680–702, 1991.

[38] S.K. Zhou, R. Chellappa, and D.W. Jacobs. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints. *ECCV*, 2004.

Ira Kemelmacher-Shlizerman received the B.Sc. with honors in Computer Science and Mathematics from Bar-Ilan University in 2001 and M.Sc. and Ph.D. in Computer Science and Applied Mathematics from the Weizmann Institute of Science in 2004 and 2009 respectively. During summer 2006, she was a visiting research student in the Department of Computer Science at Columbia University, New York. Currently she is a Postdoctoral Researcher in the Department of Computer Science and Engineering at the University of Washington. Her research interests lie in the area of computer vision and computer graphics, in particular she has worked on 3D shape reconstruction and recognition under unknown illumination and pose, analysis of impoverished data, 3D deformations and analysis of visual appearance of objects under arbitrary illumination.

Ronen Basri Ronen Basri received the BSc degree in mathematics and computer science from Tel Aviv University in 1985 and the PhD degree from the Weizmann Institute of Science in 1991. From 1990 to 1992 he was a postdoctoral fellow at the Massachusetts Institute of Technology in the Department of Brain and Cognitive Science and the Artificial Intelligence Laboratory under the McDonnell-Pew and the Rothchild programs. Since then, he has been affiliated with the Weizmann Institute of Science in the Department of Computer Science and Applied Mathematics, where he currently holds the position of Professor and the Elaine and Bram Goldsmith Chair of Applied Mathematics. In 2007 he served as Chair for the Department of Computer Science and Applied Mathematics. He further held visiting positions at NEC Research Institute in Princeton, New Jersey, Toyota Technological Institute at Chicago, the University of Chicago, and the Janelia Farm Campus of the Howard Hughes Medical Institute. Dr. Basri's research has focused on computer vision, especially in the areas of object recognition, shape reconstruction, lighting analysis, and image segmentation. His work deals with the development of algorithms, analysis, and implications to human vision. He is a member of the IEEE and the IEEE Computer Society.