UNIVERSITY *of York*

This is a repository copy of *Recovering facial shape using a statistical model of surface normal direction*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/1996/

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Recovering Facial Shape Using a Statistical Model of Surface Normal Direction

William A.P. Smith, *Student Member*, *IEEE*, and Edwin R. Hancock

Abstract—In this paper, we show how a statistical model of facial shape can be embedded within a shape-from-shading algorithm. We describe how facial shape can be captured using a statistical model of variations in surface normal direction. To construct this model, we make use of the azimuthal equidistant projection to map the distribution of surface normals from the polar representation on a unit sphere to Cartesian points on a local tangent plane. The distribution of surface normal directions is captured using the covariance matrix for the projected point positions. The eigenvectors of the covariance matrix define the modes of shape-variation in the fields of transformed surface normals. We show how this model can be trained using surface normal data acquired from range images and how to fit the model to intensity images of faces using constraints on the surface normal direction provided by Lambert's law. We demonstrate that the combination of a global statistical constraint and local irradiance constraint yields an efficient and accurate approach to facial shape recovery and is capable of recovering fine local surface details. We assess the accuracy of the technique on a variety of images with ground truth and real-world images.

Index Terms—Shape-from-shading, albedo estimation, directional statistics, illumination, face modeling.

✦

---

1 INTRODUCTION

SHAPE-FROM-SHADING provides an alluring yet somewhat elusive route to recovering 3D surface shape from single 2D intensity images [41]. Unfortunately, the method has proved ineffective in recovering realistic 3D face shape because of real world albedo variations and the local convexity-concavity instability due to the bas-relief ambiguity. This is, of course, a well-known effect which is responsible for a number of illusions, including Gregory's famous inverted mask [16]. The main problem is that the nose becomes imploded and the cheeks exaggerated [7]. It is for this reason that methods such as photometric stereo [15] have proven more effective.

One way of overcoming this problem with single view shape-from-shading is to use domain specific constraints. Several authors [2], [7], [30], [32], [42] have shown that, at the expense of generality, the accuracy of the recovered shape information can be greatly enhanced by restricting a shape-from-shading algorithm to a particular class of objects. For instance, both Prados and Faugeras [30] and Castelán and Hancock [7] use the location of singular points to enforce convexity on the recovered surface. Zhao and Chellappa [42], on the other hand, have introduced a geometric constraint which exploits the approximate bilateral symmetry of faces. This "symmetric shape-from-shading" method was used to correct for variations in illumination and has been employed for recognition by synthesis. However, the recovered surfaces were of insufficient quality to synthesise novel facial views. Atick et al. [2] proposed a statistical shape-from-shading method based on a low-dimensional parameterization of

facial surfaces. Principal components analysis (PCA) was used to derive a set of "eigenheads" which compactly capture 3D facial shape. Unfortunately, it is surface orientation and not depth which is conveyed by image intensity. Therefore, fitting the model to an image equates to a computationally expensive parameter search which attempts to minimize the error between the rendered surface and the observed intensity. This is similar to the approach adopted by Samaras and Metaxas [32] who incorporate reflectance constraints derived from shape-from-shading into a deformable model. Dovgard and Basri [12] combined the statistical constraint of Atick et al. [2] and the geometric constraint of Zhao and Chellappa [42] into a single shape-from-shading algorithm. However, asymmetry in real face images results in errors in the recovered surfaces. Nandy and Ben-Arie [27] attempt to learn the relationship between 3D shape and image intensity for a number of face parts. Their shape-from-recognition framework helps to constrain the space of solutions to the image irradiance equation, but relies on statistical methods to learn the effects of illumination variation.

1.1 Statistical Methods

There has been considerable effort in the computer vision literature aimed at learning appearance-based models of faces [8], [10], [36], [40]. This is, of course, a complex problem where the devil resides in the detail. The appearance of a face is determined by a number of complex factors. The first of these is the three-dimensional shape of the face [6]. The second is the albedo map which captures effects such as local variations in skin pigmentation and the distribution of facial hair [24]. Finally, there is the process by which light is reflected from the skin. Skin reflectance is itself a complex process, which is thought to be governed by subsurface scattering processes and is affected by blood flow beneath the skin [37]. In particular, the pattern of reflectance is strongly non-Lambertian [25].

● *The authors are with the Department of Computer Science, University of York, York YO10 5DD, UK. E-mail: {wsmith, erh}@cs.york.ac.uk.*

For these reasons, much of the effort here has been aimed at developing generative statistical models that can be used to learn the modes of appearance variation [8], [40], and then subsequently used to synthesize face appearance. Statistical methods are used since they are usually robust, have well-understood parameter estimates and can be applied in a flexible manner [21], [26]. Moreover, they allow the complex physical processes which give rise to variable face appearance to be subsumed into a parametrically efficient model [10]. Although this can be viewed as a merit from the computational standpoint, it circumvents the direct modeling of the underlying shape and reflectance effects that give rise to variable appearance.

Recently, Blanz and Vetter [4], inspired by the statistical shape-from-shading approach of Atick et al. [2], have used a complete 3D facial appearance model for face shape recovery. Their approach is considered state-of-the-art in the face reconstruction literature. They align a training set of facial shape and texture pairs using a modified optical flow algorithm. They then use PCA to derive separate linear models for face shape and texture. As with Atick's [2] approach, depth cannot be inferred from image intensity and, hence, the parameters of the model cannot be recovered from an image directly. Recovering facial shape from an image therefore amounts to a complex optimization process which aims to minimize the error between observed and reconstructed appearance by simultaneously estimating the intrinsic (face shape and texture) and extrinsic (pose, illumination, and camera properties) parameters of the imaged face. This is not "shape-from-shading" in the traditional sense and is more accurately described as analysis by synthesis. The strict statistical constraint provided by a 3D appearance model ensures robust performance under widely varying illumination and pose conditions, and the technique has been used successfully for recognition [5] and facial reanimation [3]. However, it is unclear how data-closeness constraints can be strictly enforced and non-Lambertian, dynamic effects are difficult to capture. The result of this is that the recovered models, while certainly "face-like," lack photorealism. Furthermore, there has also been no attempt to quantify the accuracy of the recovered shape or test its sensitivity to illumination conditions.

1.2 Contribution

Our aim in this paper is to show how a statistical model of face shape may be embedded within a shape-from-shading framework. The aim here is to recover a field of surface normals (a needle-map) from a single intensity image, by exploiting the direct relationship between surface orientation and measured intensity. Unfortunately, the construction of a statistical model for the distribution of facial needle-maps is not a straightforward task. The statistical representation of directional data has proven to be considerably more difficult than that for Cartesian data [17]. Surface normals can be viewed as points residing on a unit sphere and may be specified in terms of the elevation and azimuth angles. This representation makes the computation of distance difficult. For instance, if we consider a short walk across one of the poles of the unit sphere, then although the distance traversed is small, the change in azimuth angle is large. Hence, constructing a statistical

model that can capture the statistical distribution of directional data is not a straightforward task.

To overcome the problem, in this paper, we draw on ideas from cartography. Our starting point is the *azimuthal equidistant* or Postel projection [34]. This projection has the important property that it preserves the distances between locations on the sphere. It is used in cartography for path planning tasks. Another useful property of this projection is that straight lines on the projected plane through the centre of projection correspond to great circles on the sphere. The projection is constructed by selecting a reference point on the sphere and constructing the tangent plane to the reference point. Locations on the sphere are projected onto the tangent plane in a manner that preserves arc-length.

We exploit this property to generate a local representation of the field of surface normals. We commence with a set of needle-maps, i.e., fields of surface normals which, in practice, are obtained from range images. We begin by computing the mean field of surface normals. The surface normals are represented using elevation and azimuth angles on a unit sphere. At each image location, the mean-surface normal defines a reference direction. We use this reference direction to construct an azimuthal equidistant projection for the distribution of surface normals at each image location. The distribution of points on the projection plane preserves the distances of the surfaces normals on the unit sphere with respect to the mean surface normal. We then construct a deformable model over the set of surface normals by applying the Cootes et al. [9] point distribution model to the coordinates that result from transforming the surface normals from the unit sphere to the tangent plane under azimuthal equidistant projection. On the tangent projection plane, the points associated with the surface normals are allowed to move in a manner which is determined by the principal component directions of the covariance matrix for the point-distribution. Once we have computed the allowed deformation movement on the tangent plane, we recover surface normal directions by using the inverse transformation onto the unit sphere.

The general idea of projecting from the Gauss map of a surface to a plane in such a way as to preserve geodesic distance is of course central to the exponential map in differential geometry [11]. In fact, Pennec [28] has recently developed a framework for the analysis of statistical data on manifolds using the exponential map and has applied the method to the analysis of medical images. Our idea could clearly be formulated in this way too, but for simplicity, we adhere to the cartographic analogy.

We use the model to provide a statistical constraint within a geometric shape-from-shading framework. When the surface reflectance follows Lambert's law, the surface normal is constrained to fall on a cone whose axis is in the light source direction and whose opening angle is the inverse cosine of the normalized image brightness. This method commences from an initial configuration in which the surface normals reside on the irradiance cone and point in the direction of the local image gradient. The statistical model is fitted to recover a revised estimate of the surface normal directions. The best-fit surface normals are projected onto the nearest location on the irradiance cones. This process is iterated to convergence and

the height map for the surface recovered by integrating the final field of surface normals. We show how albedo maps can be recovered using the difference between observed and reconstructed image intensity. With the albedo maps to hand, we explore how faces can be realistically reilluminated from different lighting and viewing directions.

1.3 Psychological Motivation

The psychological plausibility behind the approach presented in this paper provides part of our motivation. The most convincing evidence to suggest that shading information plays a major role in human face perception is the severe degradation in performance when faces are illuminated from below. Hill and Bruce [19] studied this phenomenon in depth and concluded that the results were consistent with a process in which lighting from above helps the derivation of 3D shape. In other words, the human shape-from-shading process incorporates an assumption of supraluminance. In the case of faces, additional constraints are brought to bear which may override assumptions made in recovering shape-from-shading for general surfaces. For example, facial surfaces are always perceived as convex [16], even if this interpretation results in a proposed illumination direction which breaches the supraluminance assumption. This suggests that, in some cases, human shape-from-shading makes use of prior class knowledge to constrain the shape recovery process.

However, the lack of complete viewpoint invariance in the human face recognition system [35] suggests that an abstract 3D model of the face is not recovered from a training view and stored for subsequent matching. Instead, Marr's [23] $2\frac{1}{2}$-D sketch appears to provide a plausible representation. This comprises a viewer-centered representation which contains surface primitives, for example, the field of surface orientation information provided by shape-from-shading. The approach presented in this paper for recovering and representing facial shape is partially motivated by this observation.

1.4 Paper Outline

The outline of the paper is as follows: In Section 2, we commence by detailing the azimuthal equidistant projection and explaining how the coordinates resulting from the transformed surface normals can be used to construct a statistical model for needle-map variation. Section 3 focusses on how the resulting statistical model can be fitted to image data. Here, we explore two approaches: The first of these involves fitting the statistical model to surface normal data delivered by shape-from-shading. The second method is an iterative one which aims to find the recovered field of surface normals that best fits the observed image brightness under Lambert's law. Section 4 provides experiments using the 3DFS and Yale-B databases. Finally, Section 5 offers some conclusions and suggests some directions for future research.

2 A STATISTICAL MODEL FOR SURFACE NORMALS

Previous work has shown that both images of faces [36] and facial surfaces [2] can be modeled in a low-dimensional space, derived by applying principal components analysis (PCA) to a training set of images or surfaces. However, fields of surface normals also provide an important source of information

from which a statistical model of face-shape can be constructed. A field of surface normals, or needle-map, provides a more detailed description of an object than a corresponding brightness image. Surface normals are invariant to changes in illumination and surface reflectance. Moreover, topographic information such as surface curvature can be computed from a field of surface normals [39]. Using shape-from-shading [38], the field of surface normals is also more easily recovered from an image than the underlying surface height function since it is orientation and not depth information which is conveyed by variations in image intensity. In the remainder of this section, we show how to build a statistical model of surface normal variation by applying PCA to a training sample of fields of surface normals.

2.1 Azimuthal Equidistant Projection

A *needle-map* describes a surface $z(x,y)$ as a set of local surface normals $\mathbf{n}(x,y)$ projected onto the view plane. Let $\mathbf{n}_k(i,j) = (n_k(i,j)_x, n_k(i,j)_y, n_k(i,j)_z)^T$ be the unit surface normal at the pixel indexed (i,j) in the kth training image. If there are K images in the training set, then at the location (i,j), the mean-surface normal direction is

$$\hat{\mathbf{n}}(i,j) = \frac{\bar{\mathbf{n}}(i,j)}{\|\bar{\mathbf{n}}(i,j)\|}, \qquad (1)$$

where

$$\bar{\mathbf{n}}(i,j) = \frac{1}{K}\sum_{k=1}^{K}\mathbf{n}_k(i,j). \qquad (2)$$

On the unit sphere, the surface normal $\mathbf{n}_k(i,j)$ has elevation angle $\theta_k(i,j) = \frac{\pi}{2} - \arcsin n_k(i,j)_z$ and azimuth angle

$$\phi_k(i,j) = \arctan\frac{n_k(i,j)_y}{n_k(i,j)_x},$$

while the mean surface normal at the location (i,j) has elevation angles $\hat{\theta}(i,j) = \frac{\pi}{2} - \arcsin \hat{n}(i,j)_z$ and azimuth angle

$$\hat{\phi}(i,j) = \arctan\frac{\hat{n}(i,j)_y}{\hat{n}(i,j)_x}.$$

The intrinsic mean of a distribution of points lying on a spherical manifold is, in fact, the *spherical median* [22]. However, we found that aligned facial needle-maps produce surface normal distributions which are highly Fisherian [13] (i.e., have a high concentration parameter) and, hence, the mean direction is a very good approximation to the spherical median. Because of the computational ease of calculating the mean direction, we use this as our definition of the average surface normal.

To construct the azimuthal equidistant projection, we proceed as follows: We commence by constructing the tangent plane to the unit-sphere at the location corresponding to the mean-surface normal. We establish a local coordinate system on this tangent plane. The origin is at the point of contact between the tangent plane and the unit sphere. The x-axis is aligned parallel to the local circle of latitude on the unit-sphere.

Under the equidistant azimuthal projection at the location (i,j), the surface normal $\mathbf{n}_k(i,j)$ maps to the point with coordinate vector $\mathbf{v}_k(i,j) = (x_k(i,j), y_k(i,j))^T$. The
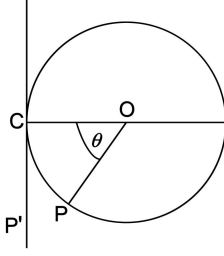
Fig. 1. The azimuthal equidistant projection.

transformation equations between the unit-sphere and the tangent-plane coordinate systems are

$$x_k(i,j) = k' \cos \theta_k(i,j) \sin[\phi_k(i,j) - \hat{\phi}(i,j)]$$

$$y_k(i,j) = k' \bigg(\cos \hat{\theta}(i,j) \sin \phi_k(i,j)$$

$$- \sin \hat{\theta}(i,j) \cos \theta_k(i,j) \cos[\phi_k(i,j) - \hat{\phi}(i,j)] \bigg), \tag{3}$$

where $k' = \frac{c}{\sin c}$ and

$$\cos c = \sin \hat{\theta}(i,j) \sin \theta_k(i,j)$$

$$+ \cos \hat{\theta}(i,j) \cos \theta_k(i,j) \cos[\phi_k(i,j) - \hat{\phi}(i,j)].$$

Thus, in Fig. 1, CP' is made equal to the arc CP for all values of θ. The projected position of P, namely, P', therefore, lies at a distance θ from the centre of projection and the direction of P' from the center of the projection is true. The equations for the inverse transformation from the tangent plane to the unit-sphere are

$$\theta_k(i,j) = \arcsin\bigg(\cos c \sin \hat{\theta}(i,j) - \frac{1}{c} y_k(i,j) \sin c \cos \hat{\theta}(i,j) \bigg)$$

$$\phi_k(i,j) = \hat{\phi}(i,j) + \tan^{-1} \psi(i,j), \tag{4}$$

where

$$\psi(i,j) = \begin{cases} \frac{x_k(i,j) \sin c}{c \cos \hat{\theta}(i,j) \cos c - y_k(i,j) \sin \hat{\theta}(i,j) \sin c} & \text{if } \hat{\theta}(i,j) \neq \pm\frac{\pi}{2} \\ -\frac{x_k(i,j)}{y_k(i,j)} & \text{if } \hat{\theta}(i,j) = \frac{\pi}{2} \\ \frac{x_k(i,j)}{y_k(i,j)} & \text{if } \hat{\theta}(i,j) = -\frac{\pi}{2} \end{cases} \tag{5}$$

and $c = \sqrt{x_k(i,j)^2 + y_k(i,j)^2}$.

2.2 Point Distribution Model

Suppose that each training example is a range image which consists of an array of depth data. For the pixel indexed (i,j) in the kth training sample, the depth is $z_k(i,j)$. Using the range data, we estimate the surface normal directions and the surface normal at the pixel location (i,j) for the kth training image is $\mathbf{n}_k(i,j)$. The components of the vector are transformed into the coordinates $(x_k(i,j), y_k(i,j))$ using the azimuthal equidistant projection. If the range images have M rows and N columns the surface normal coordinates of each training sample may be represented by the long vector:

$$\mathbf{U}^k = [x_k(1,1), , x_k(M,N), y_k(1,1), , y_k(M,N)]^T \tag{6}$$

ordered according to the raster scan (left-to-right and top-to-bottom). Hence, $\mathbf{U}$ is a vector of length $2MN$ that represents an observed field of surface normals. The first MN components of the vector contain the x-coordinates obtained by applying the azimuthal equidistant projection to the surface normals stacked in column order. The second MN components of the vector contain the y-coordinates. Since the azimuthal equidistant projection involves centering the local coordinate system, the coordinates corresponding to the mean direction are $(0,0)$ at each image location. Hence, the long-vector corresponding to the mean direction at each image location is zero.

The K training samples can be used to form the $(2MN) \times K$ data-matrix

$$\mathbf{D} = [\mathbf{U}^1 || \mathbf{U}^K]. \tag{7}$$

The $(2MN) \times (2MN)$ covariance matrix is therefore given by

$$\mathbf{L} = \frac{1}{K} \mathbf{D}\mathbf{D}^T. \tag{8}$$

Since, in practice, $2MN \gg K$, we use the numerically efficient *snap-shot* method of Sirovich [33] to compute the eigenvectors of $\mathbf{L}$. Accordingly, we construct the matrix $\hat{\mathbf{L}} = \frac{1}{K}\mathbf{D}^T\mathbf{D}$ and find the eigenvalues and eigenvectors. The ith eigenvector $\hat{\mathbf{e}}_i$ of $\hat{\mathbf{L}}$ can be used to find the ith eigenvector $\mathbf{e}_i$ of $\mathbf{L}$ using $\mathbf{e}_i = \mathbf{D}\hat{\mathbf{e}}_i$. The eigenvalues of $\mathbf{L}$ may also be found using the eigen-decomposition of $\hat{\mathbf{L}}$. Let λ_i be the ith eigenvalue of $\mathbf{L}$ and $\hat{\lambda}_i$ the ith eigenvalue of $\hat{\mathbf{L}}$. For $i \leq K, \lambda_i = \hat{\lambda}_i$. For $i > K, \lambda_i = 0$.

We may consider small scale variation as noise. Hence, we need only retain S eigenmodes to retain p percent of the model variance. We choose S to be the smallest integer that satisfies the condition: $\sum_{i=1}^{S} \lambda_i \geq \frac{p}{100} \sum_{i=1}^{K} \lambda_i$. We deform the azimuthal equidistant point projections in the directions defined by the matrix $\mathbf{P} = (\mathbf{e}_1 | \mathbf{e}_2 || \mathbf{e}_S)$ formed from the leading S principal eigenvectors. Suppose that $\mathbf{v}_o$ is the vector of coordinates obtained by performing the azimuthal equidistant projection on an observed field of surface normals. We seek the parameter vector $\mathbf{b}$ that minimizes the squared error $\mathcal{E}(\mathbf{b}) = (\mathbf{v}_o - \mathbf{P}\mathbf{b})^T(\mathbf{v}_o - \mathbf{P}\mathbf{b})$. The solution to this least-squares estimation problem is $\mathbf{b}^* = \arg\min_{\mathbf{b}} \mathcal{E}(\mathbf{b}) = \mathbf{P}^T\mathbf{v}_o$. The best fit field of surface normals allowed by the model is $\mathbf{v}_o^* = \mathbf{P}\mathbf{P}^T\mathbf{v}_o$. The deformed vector of azimuthal equidistant projection co-ordinates can be transformed back into a surface normal on the unit sphere using the inverse azimuthal equidistant projection equations given above.

Fig. 2 illustrates this process. Fig. 2a is a distribution of surface normals at one pixel in a model is shown as points on the unit sphere. Fig. 2b shows the azimuthal equidistant projection of the points is shown with the mean point as the centre of projection. The first PCA axis is shown by the line labelled PCA_1. This line corresponds to a great circle on the sphere through the mean direction which minimizes the spherical distance to each point.

3 FITTING THE MODEL TO INTENSITY IMAGES

The ability to fit the model to single intensity images is clearly attractive. The representation of a needle map in the model parameter space provides a compact description of a face
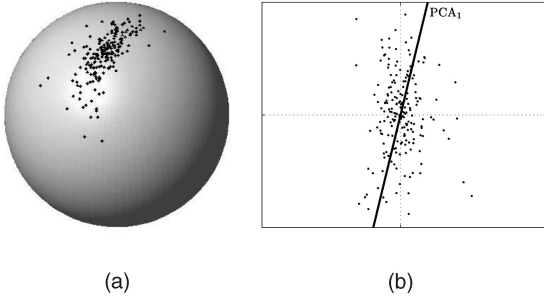
Fig. 2. Projection of points on the unit sphere (a) to points on the tangent plane at the mean point (b).

which is invariant to changes in illumination and surface reflectance. In addition, using the model to help constrain the fitting process results in an improvement in the shape information recovered from an image and allows an estimate of the albedo map to be made. In this section, we present two approaches to fitting the model intensity images. The first of these involves projecting a measured field of surface normals delivered by a geometric shape-from-shading algorithm [38] onto the nearest model configuration in the eigenspace of the training data. The second approach is an iterative one which attempts to ensure that the recovered field of surface normals satisfies Lambert's law. In other words, the model provides a statistical constraint which guides the geometric shape-from-shading algorithm. We begin by introducing Worthington and Hancock's [38] geometric shape-from-shading algorithm which underpins both of these model fitting methods.

3.1 Geometric Shape-from-Shading

If $I(i, j)$ is the measured image brightness at position (i, j), then according to Lambert's law $I(i, j) = \mathbf{n}(i, j).\mathbf{s}$, where $\mathbf{s}$ is the light source direction. In general, the surface normal $\mathbf{n}$ cannot be recovered from a single brightness measurement since it has two degrees of freedom corresponding to the elevation and azimuth angles on the unit sphere. In the Worthington and Hancock [38] iterative shape-from-shading method, data-closeness is ensured by constraining the recovered surface normal to lie on the reflectance cone whose axis is aligned with the light-source vector $\mathbf{s}$ and whose opening angle is $\theta(i, j) = \arccos I(i, j)$. At each iteration, the surface normal is free to move to an off-cone position subject to smoothness or curvature consistency constraints. However, the hard irradiance constraint is reimposed by rotating each surface normal back to its closest on-cone position. This process ensures that the recovered field of surface normals satisfies the image irradiance equation after every iteration.

Suppose that $\mathbf{n}'^{(l)}(i, j)$ is an off-cone surface normal at iteration l of the algorithm, then the update equation is

$$\mathbf{n}^{(l+1)}(i, j) = \Theta \mathbf{n}'^{(l)}(i, j), \qquad (9)$$

where Θ is a rotation matrix computed from the apex angle α and the angle between $\mathbf{n}'^{(l)}(i, j)$ and the light source direction $\mathbf{s}$. To restore the surface normal to the closest on-cone position, it must be rotated by an angle $\alpha = \theta(i, j) - \arccos[\mathbf{n}'^{(l)}(i, j).\mathbf{s}]$ about the axis $(u, v, w)^T = \mathbf{n}'^{(l)}(i, j) \times \mathbf{s}$. Hence, the rotation matrix is

$$\Theta = \begin{pmatrix} c + u^2 c' & -ws + uvc' & vs + uwc' \\ ws + uvc' & c + v^2 c' & -us + vwc' \\ -vs + uwc' & us + vwc' & c + w^2 c' \end{pmatrix}, \qquad (10)$$

where $c = \cos(\alpha)$, $c' = 1 - c$ and $s = \sin(\alpha)$.

The method is initialized by placing the surface normals on their reflectance cones such that they are aligned in the direction opposite to that of the local image gradient. This initialization is consistent with the assumption that the object under study is globally convex [38]. The polar angle $\phi(i, j)$ of the local negative image gradient is given by:

$$\phi(i, j) = \arctan \left(\frac{\left(-\frac{\partial I}{\partial y} \right)_{i,j}}{\left(-\frac{\partial I}{\partial x} \right)_{i,j}} \right) \qquad (11)$$

and, hence, each normal is initialized as follows:

$$\mathbf{n}^{(0)}(i, j) = \begin{pmatrix} \sin \theta(i, j) \cos \phi(i, j) \\ \sin \theta(i, j) \sin \phi(i, j) \\ \cos \theta(i, j) \end{pmatrix}. \qquad (12)$$

3.2 Projection onto the Model Eigenspace

Worthington and Hancock used a local smoothness constraint regulated by a robust kernel to update the field of surface normals within the geometric shape-from-shading framework. We use this constraint to recover a field of surface normals from an intensity image and then fit the model to the estimated normals using the matrix multiplication $\mathbf{b} = \mathbf{P}^T \mathbf{v}_o$, where $\mathbf{v}_o$ is the needle map under azimuthal equidistant projection and $\mathbf{P}$ the eigenvectors of the model. The algorithm can therefore be described as follows:

1. Calculate an initial estimate of the field of surface normals $\mathbf{n}$ using (12).
2. Each normal in the estimated field $\mathbf{n}$ is updated using the robust regularizer and rotated back to its closest on-cone position using $\mathbf{n}^{(l+1)} = \Theta(\tau(\mathbf{n}^{(l)}))$, where τ is a robust smoothing kernel.
3. Compare $\mathbf{n}^{(l+1)}$ and $\mathbf{n}^{(l)}$. Continue if difference indicates convergence, otherwise return to 2.
4. Each normal in the estimated field $\mathbf{n}$ undergoes an azimuthal equidistant projection ((3)) to give a vector of transformed coordinates $\mathbf{v}_o$.
5. The vector of best fit model parameters is given by $\mathbf{b} = \mathbf{P}^T \mathbf{v}_o$. The best fit needle-map $\mathbf{n}'$ is given by the inverse azimuthal equidistant projection ((4)) of the best fit vector of transformed coordinates $\mathbf{v} = \mathbf{P}\mathbf{P}^T \mathbf{v}_o$.

Existing shape-from-shading algorithms are not capable of reliably recovering accurate needle-maps from real-world images [41]. Nevertheless, we investigate this approach in our experimental section to contrast the performance of existing shape-from-shading algorithms with the method we propose below.

3.3 Combining the Statistical Model and Geometric SFS

A more attractive alternative is to use the statistical constraint provided by the model itself in the process of fitting the model to an intensity image. Once trained, the statistical model represents the space of valid face shapes. We can exploit this prior knowledge in order to help resolve the ambiguity in the

shape-from-shading process. We do this using an iterative approach which can be posed as that of recovering the best-fit field of surface normals from the statistical model, subject to constraints provided by the image irradiance equation. As noted above, when the surface reflectance follows Lambert's law, then the surface normal is constrained to fall on a cone whose axis is in the light source direction and whose opening angle is the inverse cosine of the normalized image brightness. This method commences from an initial configuration in which the surface normals reside on the irradiance cone and point in the direction of the local image gradient. The statistical model is fitted to recover a revised estimate of the surface normal directions. The best-fit surface normals are then projected onto the nearest location on the irradiance cones. Our approach to fitting the model to intensity images is hence an iterative process in which we interleave the process of fitting the statistical model to the current field of estimated surface normals and then reenforcing the data-closeness constraint provided by Lambert's law by mapping the surface normals back onto their reflectance cones. The height map for the surface is recovered by integrating the final field of surface normals using the method of Frankot and Chellappa [14]. The algorithm can be summarized as follows:

1. Calculate an initial estimate of the field of surface normals $\mathbf{n}$ using (12).
2. Each normal in the estimated field $\mathbf{n}$ undergoes an azimuthal equidistant projection ((3)) to give a vector of transformed coordinates $\mathbf{v}_o$.
3. The vector of best fit model parameters is given by $\mathbf{b} = \mathbf{P}^T \mathbf{v}_o$.
4. The vector of transformed coordinates corresponding to the best-fit parameters is given by $\mathbf{v}' = (\mathbf{P}\mathbf{P}^T)\mathbf{v}_o$.
5. Using the inverse azimuthal equidistant projection ((4)), find the off-cone best fit surface normal $\mathbf{n}'$ from $\mathbf{v}'$.
6. Find the on-cone surface normal $\mathbf{n}''$ by rotating the off-cone surface normal $\mathbf{n}'$ using $\mathbf{n}''(i,j) = \Theta \mathbf{n}'(i,j)$.
7. Test for convergence. If $\sum_{i,j} \arccos[\mathbf{n}(i,j).\mathbf{n}''(i,j)] < \epsilon$, where ϵ is a predetermined threshold, then stop and return $\mathbf{b}$ as the estimated model parameters and $\mathbf{n}''$ as the recovered needle map.
8. Make $\mathbf{n}^{(l+1)}(i,j) = \mathbf{n}''(i,j)$ and return to Step 2.

3.3.1 Initialization

We have experimented with a number of different initializations for the fitting process given above and found that the final solution is fairly insensitive to the choice of initialization, even to the extent that every normal can be initialized to point in the same direction on their reflectance cones. However, the smallest starting error and minimum number of iterations required for convergence was achieved by incorporating the model into the initialization. To do this, we place each normal on their reflectance cone at the point closest to the local average normal direction. This initialization can be expressed:

$$\mathbf{n}^{(0)}(i,j) = \begin{pmatrix} \sin\theta(i,j)\cos\hat{\phi}(i,j) \\ \sin\theta(i,j)\sin\hat{\phi}(i,j) \\ \cos\theta(i,j) \end{pmatrix}. \qquad (13)$$

Fig. 3. Angular difference between final n' and n''.

3.4 Practical Considerations

Upon convergence, we have a choice between two solutions. On the one hand, the "on-cone" surface normal $\mathbf{n}''$ provides a strict solution of the image irradiance equation within the framework of geometric shape-from-shading. This represents a data-driven solution in that reillumination of the recovered normals with the input lighting configuration will exactly reproduce the original image. On the other hand, the "best-fit" surface normal $\mathbf{n}'$ provides a least squares fit of the statistical model to the shape-from-shading normals. This represents a model-driven solution in which the field of surface normals is globally constrained to correspond to a valid face shape.

In the ideal Lambertian, unit albedo case, the strict on-cone solution is likely to provide a more accurate solution. It is also likely that this solution will recover more of the fine surface detail and discriminating features which are not captured by the model. However, real-world face images contain albedo variations caused by skin pigmentation and facial hair. By enforcing data-closeness, pixels of low albedo will be interpreted as having large incident angles. In Fig. 3, we show the angular change as data-closeness is restored to a typical final best fit needle map, i.e., the angular difference between the best-fit surface normal $\mathbf{n}'$ and the corresponding on-cone surface normal $\mathbf{n}''$. From the plot, it is clear that the changes are almost solely due to the variation in albedo at the eyes, eyebrows, and lips. Aside from these regions, there is very little change in surface normal direction, indicating that the needle-map has converged to a solution which satisfies the data-closeness constraint except in regions of actual variation in albedo. In this case, the best-fit of the statistical model may, in fact, provide a more accurate estimate of the underlying facial shape. We provide empirical evidence of this observation in our experimental results.

3.4.1 Albedo and Data-Closeness

For a real-world image, should we choose the best-fit normals $\mathbf{n}'$ as our estimate of the underlying facial shape, we can still ensure satisfaction of the data-closeness constraint in some sense. To do so, we allow albedo to vary. In other words, we relax the data-closeness constraint at the final iteration and account for differences between predicated and observed image brightness by variations in albedo. The imposition of data-closeness in previous iterations ensures the model-fit has been encouraged towards a solution which closely satisfies the constraint. If the final best-fit field of surface normals is reilluminated using a Lambertian reflectance model, then the predicted image brightness is given by $I(i,j) = \rho(i,j) [\mathbf{s}.\mathbf{n}'(i,j)]$, where $\rho(i,j)$ is the albedo at position (i,j). Since

I, $\mathbf{s}$, and $\mathbf{n}'$ are all known, we can estimate the albedo at each pixel using the formula:

$$\rho(i,j) = \frac{I(i,j)}{\mathbf{s}.\mathbf{n}'(i,j)}. \tag{14}$$

The combination of the best-fit needle-map and corresponding estimated albedo map will still exactly recreate the input image when reilluminated with the original lighting. Additionally, albedo estimated in this manner may vary arbitrarily. This means distinguishing pigmentation or facial hair is accurately recovered.

In Section 4, we demonstrate how the strict solution within the geometric shape-from-shading framework represents a significant improvement over the needle maps estimated using the original curvature consistency constraints proposed by Worthington and Hancock [38]. Moreover, we show how the best-fit needle-map and estimated albedo result in realistic synthesised images under novel illumination and viewpoint, of comparable accuracy to the more computationally intensive analysis-by-synthesis approach of Blanz and Vetter [4].

3.4.2 Alignment and Reflectance Normalization

In the above analysis, we have assumed that during training the sample facial surfaces have been aligned, i.e., that a pixel (i,j) corresponds to the same point on all sample faces. Furthermore, our shape-from-shading algorithm assumes that the input image has been aligned with the model. Alignment is not the focus of this paper and we have therefore used a combination of existing methods and manual alignment which we describe below. Clearly, there is scope for integrating a fine-scale alignment process into the iterative model fitting. However, this would serve to confuse our main contribution of a method for accurate facial shape recovery. Moreover, the fact that good results are obtainable with an approximate alignment is a strength of the method.

During training, we use a database of facial range images in which the surfaces have been aligned using an existing registration algorithm [4]. Prior to applying our shape-from-shading algorithm, the input images were manually cropped to remove background, hair, and clothing. Finally, the images were aligned and scaled to be brought into correspondence with the model using a number of manually marked feature points.

As is common in the vision literature, we have also assumed that skin reflects light according to Lambert's law. However, to improve the real-world applicability of the approach we use a preprocessing step to correct for deviations from Lambertian reflectance. Rather than incorporate a reflectance model that tries to capture some of the properties of skin reflectance, we exploit a recently reported image-based reflectance estimation method [31].

This is a computationally efficient process of complexity $O(MN)$, for an $M \times N$ pixel image, which provides a good estimate of an isotropic, monotonic BRDF from a single image. By assuming the surface under study is locally spherical, the method uses the local image gradient to find correspondences between points on the surface and points on the reflectance sphere. An estimate may then be made of the radiance function which relates the incidence angle to measured intensity. Lambertian correction is then simply a case of inverting the radiance function for each pixel to recover the incident angle, from which the Lambertian intensity can be calculated.

4   EXPERIMENTS

In this section, we present experiments with our method. There are three elements to this study: We commence by examining the model when trained on fields of surface normals extracted from range data. Second, we show the results of fitting the model to intensity data and show the surface height data that can be reconstructed from the fitted fields of surface normals. Third, we illustrate how the fitted models can be used synthesize realistic novel facial views.

4.1 Model Training

In this section, we describe how our model is constructed from real-world data. We commence by building a "ground truth" model using fields of surface normals extracted from range data. This allows us to show the utility of the model in capturing facial shape in a compact manner when trained on relatively "clean" data. We used the 3DFS data set [1] which consists of 100 high-resolution scans of subjects in a neutral expression. The scans were collected using a *Cyberware*TM 3030PS laser scanner. The database is pre-aligned, registration being performed using the optical flow correspondence algorithm of Blanz and Vetter [4]. Fields of surface normals were extracted by orthographically projecting the three surface normal components onto a view plane positioned fronto-parallel to the aligned faces.

We begin by examining the principal modes of variation for a model trained on fields of surface normals derived from range images of faces. In Fig. 4, we show the first five modes of variation of this model. In each case, we deform the points under azimuthal equidistant projection by ± 3 standard deviations along each of the first five principal axes. We then perform the inverse azimuthal equidistant projection and visualize the resulting needle-maps as fields of vectors (first three columns) and rendered with Lambertian reflectance and frontal illumination (second three columns). The modes encode shape only since the needle-maps are invariant to illumination conditions and the training set contained no variation in expression. The modes clearly capture distinct facial characteristics. For example, mode 1 encodes head size and also seems to be correlated with gender. This is manifested in the broader jaw, brow, and nose in the negative direction, all of which are masculine features. The third mode encodes the difference between long, narrow faces and short, wide faces, whereas the second mode encodes the difference between a pointy and a rounded chin.

In Fig. 5, we provide a visualization of the variability in the shape of different facial regions. Since it is the natural model for multivariate directional data we fit a von Mises-Fisher (vMF) distribution to the distribution of surface normals at each location. The vMF distribution is analogous to the multivariate Gaussian distribution in $\Re^q$ [13]. The distribution is specified by two parameters: the mean direction vector, which, in our case, is $\hat{\mathbf{n}}$, and the concentration parameter κ. For $q = 3$, the distribution over the vector $\mathbf{n}$ is described by the following probability density function:
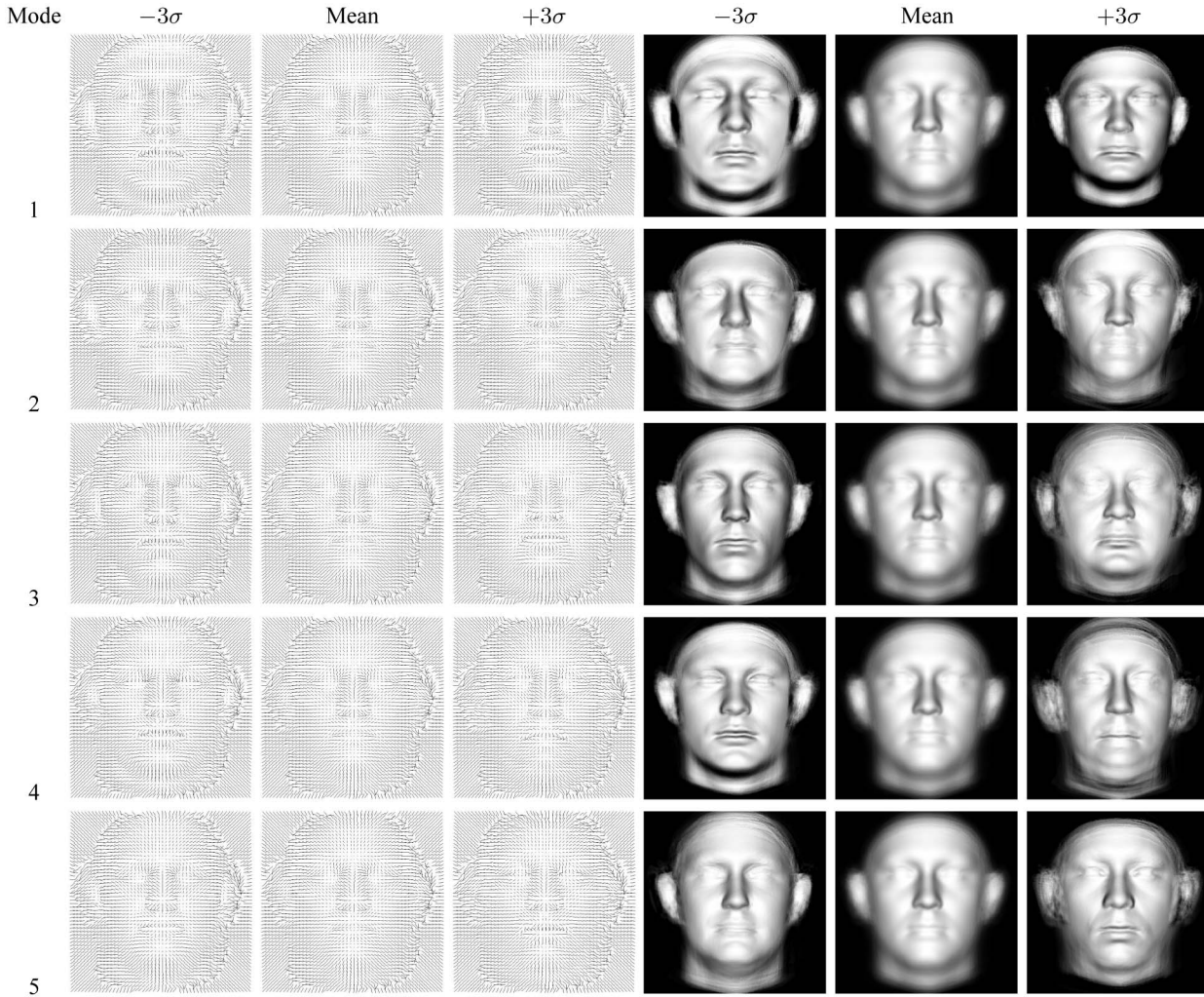
Fig. 4. The first five modes of variation of a statistical surface normal model trained on a set of facial needle-maps extracted from range data. The deformed needle-maps are visualized as: Fields of vectors (first three columns) and rendered with Lambertian reflectance and frontal illumination (second three columns). The mean face is shown in the central column and ± 3 standard deviations along each of the first five principal modes of variation are shown in the first and third columns.

$$f(\mathbf{n}; \hat{\mathbf{n}}, \kappa) = \frac{\kappa^{\frac{1}{2}}}{2\pi^{\frac{3}{2}} I_{\frac{1}{2}}(\kappa)} e^{\kappa \hat{\mathbf{n}}^T \mathbf{n}}, \qquad (15)$$

where $I_{\frac{1}{2}}$ is a modified Bessel function of the first kind and order $\frac{1}{2}$. Using the algorithm in [18], we estimate κ at each location and plot $\log(\kappa)$ in Fig. 5. The larger the value of κ, the greater the density of the distribution of surface normal directions around the mean direction. Thus, white corresponds to an area of very low variance and black to an area of very high variance. The plot demonstrates that much of the variation in facial surface orientation is confined to the

Fig. 5. Plot of model concentration shown as $log(\kappa)$, where κ is the von Mises-Fisher concentration parameter.

eyes, nose, lips, and chin, while the cheeks and forehead vary less between subjects.

4.2 Fitting the Model to Data

In this section, we explore the fitting of the model to data using the two methods outlined in Section 3. We examine the influence of illumination direction, the number of iterations, the number of model dimensions, and nonconstant albedo on the accuracy of the recovered surface normals. We use a larger database of range scans for this study in order to provide ground truth data. In this case, the model is trained on 180 examples, with 20 retained as ground truth to measure out-of-sample fitting accuracy.

4.2.1 Fitting the Model to Needle-Maps

We begin by fitting the model to fields of surface normals extracted from an image using shape-from-shading, as described in Section 3.2. We use the shape-from-shading technique of Worthington and Hancock [38]. In Fig. 6, we show an example of fitting to a field of surface normals extracted from a real-world image. The input image is shown on the left, followed by the needle-map recovered using
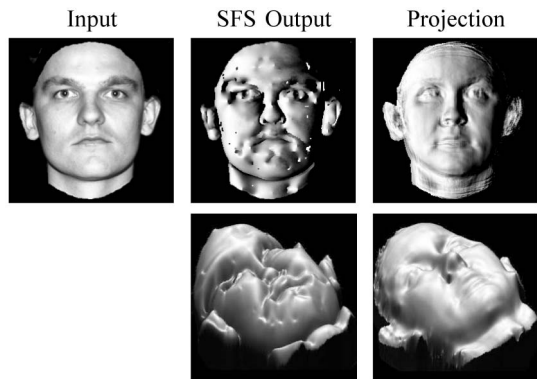
Fig. 6. Behavior of the projection fitting process. The input image is shown on the top left. The first row shows the needle map delivered by SFS followed by its projection onto the model eigenspace. The needle maps are reilluminated by a light source with direction $\mathbf{s} = (-1, 0, 1)^T$. The second row shows the surfaces recovered from the corresponding needle maps.

shape-from-shading. The projection of the needle-map onto the model eigenspace is shown on the right. Since the recovered needle-map is guaranteed to satisfy data-closeness, it would appear identical to the input image when rendered with a light source from the original direction $(\mathbf{s} = (0, 0, 1)^T)$. For this reason, in the top row, we show the needle-maps reilluminated with a light source moved along the negative x-axis to subtend an angle of 45 degrees with the viewing direction. From the reilluminations, it is clear that there are severe flaws in the initial needle-map recovered using shape-from-shading. Although the gross distribution of intensity appears correct, the resulting image is noisy and

not realistic and suffers from convex/concave errors and feature implosions. This is made explicit in the corresponding recovered surface shown in the second row. The projection onto the model eigenspace helps resolve some of these errors, resulting in a more plausible image. From the corresponding recovered surface, it is clear that many of the feature implosions are corrected and noise reduced. However, given the gross errors in the initial needle-map, it seems unlikely that this represents the closest model fit of the underlying facial shape of the input face.

4.2.2 Fitting the Model to Intensity Images

We now show how the statistical model may be fitted to intensity data using the method outlined in Section 3.3. We commence by considering the iterative behavior of the algorithm. The top row of Fig. 7 shows how a needle map develops over 25 iterations of the algorithm. In this figure, the input face is shown on the top left. Since the needle maps satisfy data-closeness at every iteration, they would all appear identical when rendered with a light source from the original direction $(\mathbf{s} = (0, 0, 1)^T)$. For this reason, in the top row, we show the needle maps reilluminated with a light source moved along the negative x-axis to subtend an angle of 45 degrees with the viewing direction. After one iteration, there is a significant global improvement in the recovered needle map. Subsequent iterations make more subtle improvements, helping to resolve convex/concave errors and sharpening defining features. For comparison, the second row shows the corresponding needle maps recovered using the original curvature consistency constraint of Worthington and Hancock [38] reilluminated in the same manner. Although there is a steady improvement in the quality of
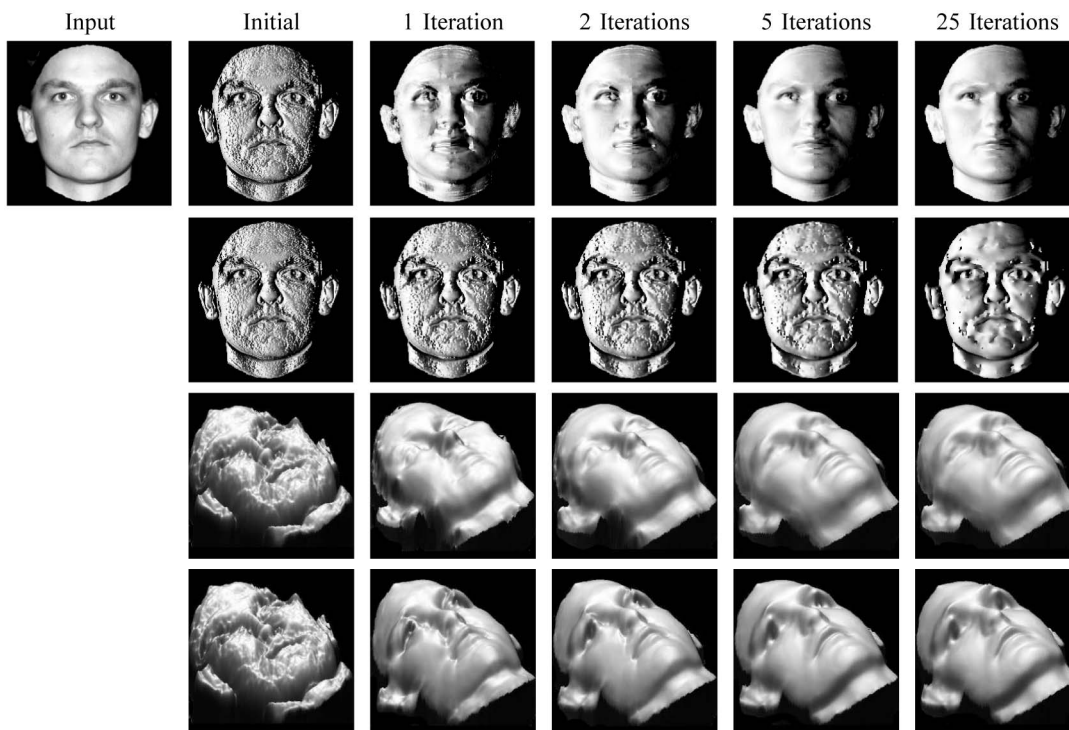

Fig. 7. Behavior of the iterative fitting process over 25 iterations. The input image is shown on the top left. The first row shows the recovered needle maps reilluminated by a light source with direction $\mathbf{s} = (-1, 0, 1)^T$. For comparison, the second row shows similarly reilluminated needle maps recovered by the Worthington and Hancock algorithm. The third and fourth rows show the surfaces recovered from $\mathbf{n}'$ (third row) and $\mathbf{n}''$ (fourth row).
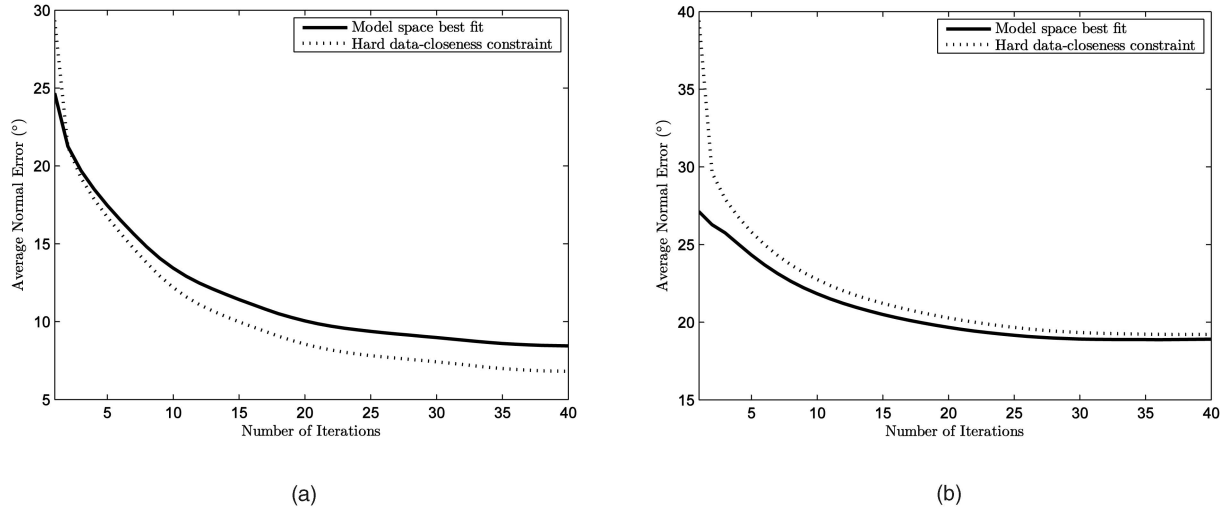
(a)

(b)

Fig. 8. Average normal error versus number of iterations of the algorithm when fitting to synthetic data with: (a) constant albedo and (b) varying albedo.
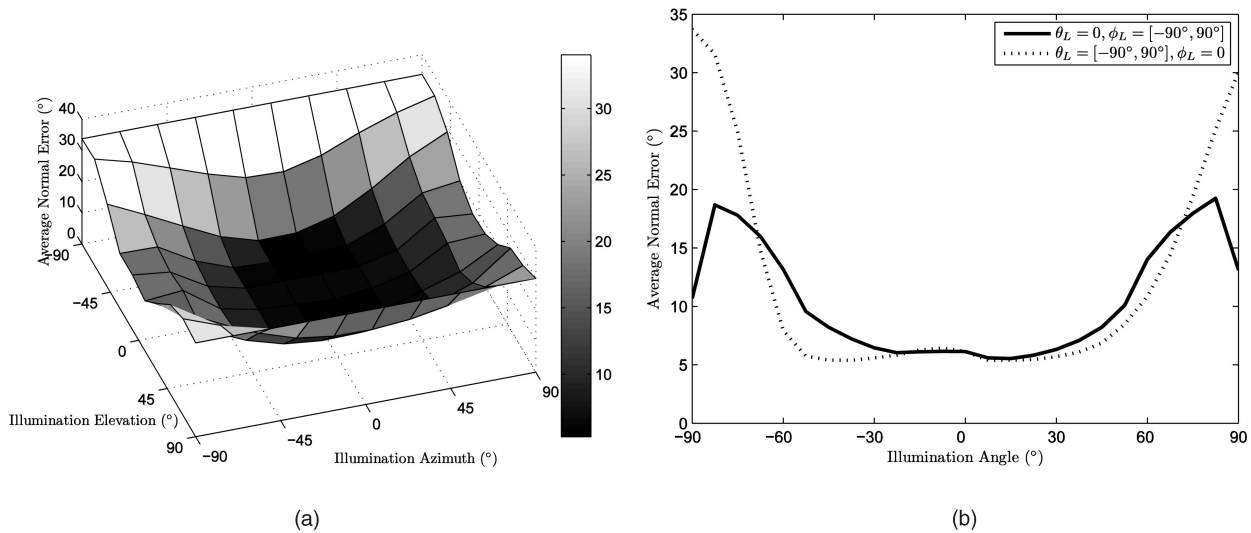

(a)

(b)

Fig. 9. Analysis of the accuracy of the fitting process under varying illumination. (a) Average error in the estimated surface normal direction as both the azimuth and elevation of the light source are varied. (b) Contours from (a) in which the elevation of the light source is set to 0 and the azimuth varied (solid line) and the azimuth is set to 0 and the elevation varied (broken line).

the recovered normals, there are gross global errors as well as feature implosions around features such as the nose.

In Fig. 7, we also show the surfaces recovered from the current best fit needle maps, $\mathbf{n}'$, (third row) and the needle maps which satisfy data-closeness, $\mathbf{n}''$, (bottom row) as the algorithm iterates. Surface recovery is effected using the method of Frankot and Chellappa [14]. As one would expect, the imposition of data-closeness results in errors in the recovered surface where there is variation in albedo, most notably around the eyes and eyebrows. In both sets, there is a clear improvement in the recovered surface as the algorithm iterates. The implosion of the nose is corrected, the surface becomes smoother and finer details become evident, for example, around the lips.

In Fig. 8, we provide a quantitative analysis of the iterative behavior of the fitting process using ground truth data. For each of the 20 out-of-sample faces, we render the needle-map with a light source situated at the viewpoint to yield an image to which we apply the fitting process. In Fig. 8a, we render the needle-maps with Lambertian reflectance and unit albedo,

whereas in Fig. 8b, we render the needle-maps with the albedo recorded by the Cyberware scanner. We plot the average normal error across all needle-maps in degrees against the number of iterations. The solid curve shows the error for the best-fit needle map in the model space, $\mathbf{n}'$, and the broken curve for the needle-map with data-closeness enforced, $\mathbf{n}''$. From both plots, it is clear that the algorithm converges rapidly, particularly, in the first few iterations. As would be expected, the introduction of varying albedo reduces the accuracy of the recovered normals. An interesting feature to note is that in Fig. 8a when fitting to images with constant albedo, the imposition of data-closeness results in higher accuracy. However, in Fig. 8b when fitting to images with varying albedo, this is reversed and the best-fit to the needle-map in the model space gives higher accuracy and, hence, is closer to the underlying shape. This adds quantitative evidence to our observation that on real-world images, $\mathbf{n}'$ is a better estimate of the true facial shape.

In Fig. 9, we examine the influence of illumination direction on the accuracy of the recovered needle-map.
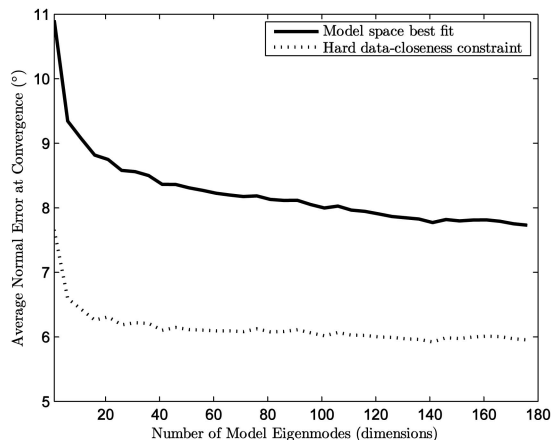
Fig. 10. Average surface normal error at convergence versus the number of model eigenmodes retained.
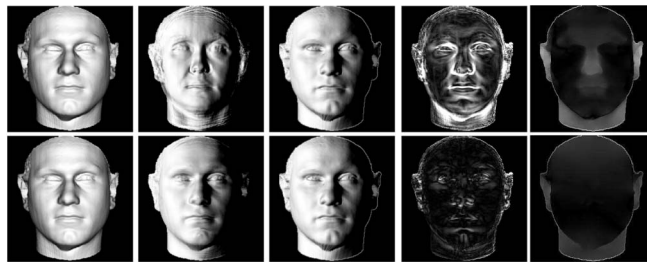

Fig. 11. Accuracy of fitting process on ground truth data. From left to right: the input image, the recovered needle-map reilluminated by a light source with direction $s = (-1, 0, 1)^T$, the ground truth needle map similarly reilluminated, the angular error, and the height error. The projection fitting method is shown on the top row, the iterative fitting method is shown on the bottom row.

Again, we render the out-of-sample ground truth needle-maps with Lambertian reflectance and unit albedo but we vary the elevation (θ_L) and azimuth (ϕ_L) of the light source through $(-90°, 90°)$. When $\theta_L = \phi_L = 0$, the light source direction is coincident with the viewing direction, when $\phi_L < 0$ it is from the left, when $\theta_L < 0$ it is from below and vice versa. The plot in Fig. 9a shows the average normal error over all out-of-sample needle-maps in degrees as both θ_L and ϕ_L are varied. For clarity, in Fig. 9b, we show contours from Fig. 9a in which only the azimuth is varied (solid line) and only the elevation is varied (broken line). From both plots, it is clear that the process is robust to variations in illumination of ±45 degrees of both azimuth and elevation (note the plateau of values under 10 degrees). Another feature evident from the plots is that extreme elevation values have a greater detrimental effect than extreme azimuth angles.

Since faces are approximately bilaterally symmetric, the errors introduced by varying ϕ_L (i.e., moving the light source left and right) are approximately symmetric. A more interesting feature is visible when θ_L is varied (i.e., the light source is moved up and down). In both Figs. 9a and 9b, it is clear that greater errors are introduced when θ_L is large and negative than when large and positive. In other words, using our algorithm, it is harder to recover facial shape when the light source is moved below the face than above. This potentially has some interesting psychological implications. Typically, poor human performance, when perceiving images of faces illuminated from below, has been attributed to an illumination direction estimation process in which the light source is constrained to come from above [20]. However, with perfect knowledge of the illumination direction, our algorithm recovers less accurate facial shape information when the illumination is from below compared to above. This suggests that there is something implicit in human face shape which makes it harder to recover shape accurately when illuminated from below. Hence, at least part of the degradation noted in human performance may related to this fact.

Finally, in Fig. 10, we investigate the effect of the number of eigenmodes (or dimensions) retained in the model on the accuracy of the recovered facial shape. We plot the average normal error at convergence versus the number of eigenmodes retained. Increasing the number of eigenmodes retained has the effect of increasing the accuracy. However, the effect is more evident for the best-fit needle map than

the on-cone surface normal. Moreover, only the first 40 or so dimensions have a significant impact on the accuracy. In other words, even a limited number of dimensions provides a statistical constraint which is sufficient to guide the shape-from-shading process to an accurate solution when combined with the hard data-closeness constraint.

4.2.3 Comparing Iterative Fitting and Fitting to Shape-from-Shading Normals

In Fig. 11, we quantitatively compare the two approaches to fitting the model using ground truth data. In the top row, we fit the model to surface normals recovered using shape-from-shading and in the bottom row we use the iterative fitting process. The estimated needle-map is shown in column 2, reilluminated by a light source with direction $s = (-1, 0, 1)^T$. For comparison, the ground truth needle map is shown similarly reilluminated in the third column. It is clear that using shape-from-shading normals results in major structural differences between the best fit and ground truth needle map, in particular, the exploded cheeks and inaccurate jaw line. In the fourth column, the angular error for each surface normal is shown. There are clearly large errors around the nose, jaw, and eyes as well as evidence of general noise. The overall average angular error of the fitted needle-map was 13.25 degrees. In contrast, the needle-map recovered using the iterative fitting process appears very close to the reilluminated ground truth needle-map with no obvious errors. An interesting exception is the dimple in the chin which is not visible in the input image, but which becomes visible when the ground truth needle map is reilluminated from a different direction. The fitting process fails to recover this structure. The angular errors are much reduced, mainly limited to the occluding boundary. The overall average angular error of the fitted needle map was 3.93 degrees. This represents more than three times greater accuracy than the projection fitting method. In the final column, we also show an error plot of the reconstructed height. Similar results were obtained with a variety of ground truth data.

4.3 Synthesizing Novel Views

In this section, we focus on how the fitted models can be used for the purposes of novel view synthesis. We focus on the results obtained using the iterative method outlined in Section 3.3 since from the previous section this appears to outperform the projection method. The data used in this study comes from the Yale-B database [15]. In the images, the faces are in a frontal pose and were illuminated by a point

Fig. 12. Column 1 shows the input images of the 10 subjects in the Yale B database. Column 2 shows the estimated albedo maps. Columns 3-6 show synthesized views of the subjects under novel illumination. The light source directions are $\mathbf{s} = (-1, 0, 1)^T$, $(1, 0, 1)^T$, $(0, 1, 1)^T$, and $(0, -1, 1)^T$, respectively.

light source situated approximately at the viewpoint, i.e., in direction $\mathbf{s} = (0, 0, 1)^T$.

In Fig. 12, we present the results of the albedo estimation process and show synthesized images under novel illumination. In the first column, we show the input images of the 10 subjects from the Yale-B database. In the

second column, we show the albedo maps estimated using (14). The results of the albedo estimation process appear intuitively convincing. For instance, the albedo map identifies the eyes, eyebrows, facial hair, nostrils, and lips. Moreover, there are no residual shading effects in the albedo map, for example, the nose is given constant
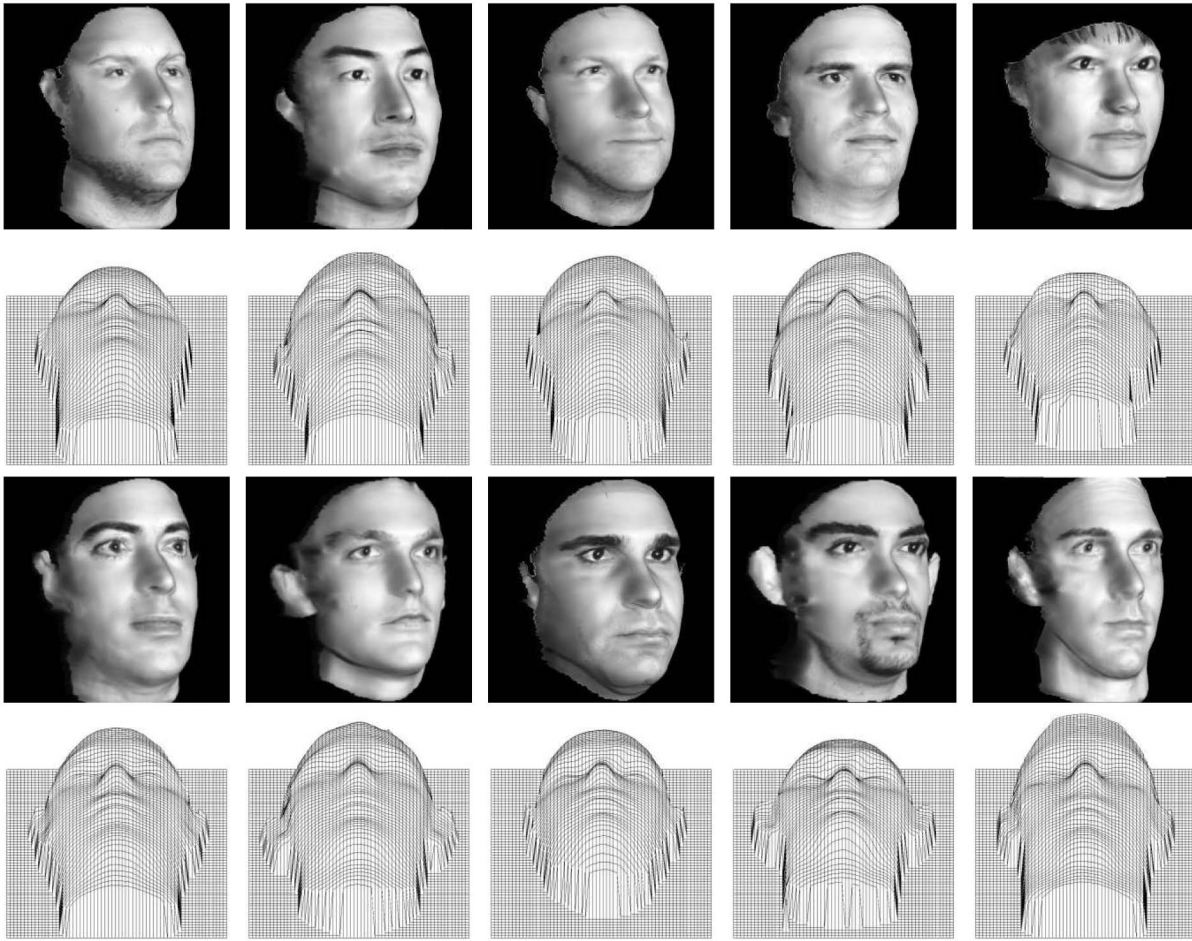
Fig. 13. Surfaces recovered from the 10 subjects in the Yale B database. In the first and third rows, the surfaces are rendered with their estimated albedo maps, Lambertian reflectance, and are shown rotated 30 degrees about the vertical axis. The light source remains fronto-parallel with respect to the face. In the second and fourth rows, the surface meshes are shown rotated 40 degrees about the horizontal axis.

Fig. 14. Surface recovered from subject 8 of the Yale B database. The surface is again rendered with the estimated albedo map, Lambertian reflectance, and rotated 30 degrees about the vertical axis. The light source is circled from full left profile to full right profile with respect to the face.

albedo. The method works well because albedo is approximately constant across much of a face's surface and, hence, the gross structure of the needle map is not overly disrupted by the imposition of data-closeness at each iteration.

The remaining four columns of Fig. 12 show the best fit needle maps reilluminated with Lambertian reflectance and the estimated albedo maps. The light source is moved to subtend an angle of 45 degrees with the view direction along the positive and negative x and y-axes. The needle maps show considerable stability under large changes in illumination direction and result in near photorealistic synthesized images.

In Fig. 13, we show synthesized images of the input faces in novel pose. In the first and third rows, the surfaces are shown rotated 30 degrees about the vertical axis. As with the previous images, the surfaces are rendered with Lambertian reflectance and the estimated albedo maps. The light source remains fronto-parallel with respect to the face (i.e., from the original direction). The resulting synthesized images are near photorealistic under a large change in viewpoint. Certainly, the results are comparable with those of Georghiades et al. [15] in which seven input images were used per subject. Rows 2 and 4 of Fig. 13 show the meshes of the recovered surfaces to allow inspection of the recovered shape alone. In Fig. 14, we demonstrate that the recovered surface and albedo map are sufficiently stable to synthesize images in both novel pose and novel illumination. We show the surface of subject 8 rendered as in the previous figure, except that the light source is circled from left profile to right profile.

(a)

(b)

Fig. 15. (a) Synthesized views of famous faces. Input images are shown in the central column, synthesized poses of -24 degrees, -12 degrees, 12 degrees, and 24 degrees from frontal are shown in columns 1, 2, 4, and 5, respectively. (b) Recovered shape shown rendered with Lambertian reflectance and rotated 30 degrees from frontal.

Finally, in Fig. 15, we demonstrate the use of our method in a practical application. In Fig. 15a, we have synthesized novel poses of famous faces from frontal images for use as stimuli in a neuropsychological study. The input images have been texture mapped onto the recovered shape and, so, in effect, the illumination remains stationary with respect to the face. In this case, little control was exercised over the conditions present in the input images, which include varying illumination direction and facial expression, yet we are still able to synthesize useful images of the subjects in novel poses. In Fig. 15b, we show the recovered surfaces, rendered with Lambertian reflectance and rotated 30 degrees from frontal. Distinguishing features of the input face shape have been clearly recovered.

4.4 Shape-from-Shading versus Analysis-by-Synthesis

Our method provides a novel statistical constraint which can be efficiently implemented within the framework of geometric shape-from-shading. The fields of surface normals recovered using our constraint represent a significant improvement over previous attempts to apply shape-from-shading to images of faces [41], [42]. If we relax the data-closeness constraint and take the best-fit of the model to the data, our method provides shape and albedo estimates which allow convincing view synthesis under novel pose and illumination. However, in both cases, its application is limited to a particular class of objects (faces). Therefore, the method must be judged with respect to state-of-the-art face reconstruction approaches as well as generic shape-from-shading algorithms. For this reason, in this section, we provide a comparison between our approach and the morphable model used by Blanz and Vetter [4] to recover facial shape.

In contrast to Blanz and Vetter's model, our approach requires a more restrictive setup (frontal face view and known illumination). However, it offers a number of potential advantages. Since a field of surface normals may be estimated directly from an image using shape-from-shading, the model parameters are recoverable using a matrix multiplication. This avoids posing the fitting process as a minimization problem and we therefore avoid the pitfalls of local minima and slow convergence. In addition, data-closeness can be restored locally by rotating each surface normal in the best fit needle-map onto the irradiance cone. The advantage brought to bear by the imposition of this local constraint is clearly demonstrated in Fig. 16. On the right, we show the surface recovered by our method from the highlighted region of the image on the left. The surface is displayed as both a wire frame and rendered surface. Discriminating local surface features (in this case, wrinkles) not captured by the model have been accurately recovered. The shape of the recovered surface is in good agreement with the highlighted intensity. This would not be possible using a morphable model unless the training set contained examples of similarly positioned wrinkles.
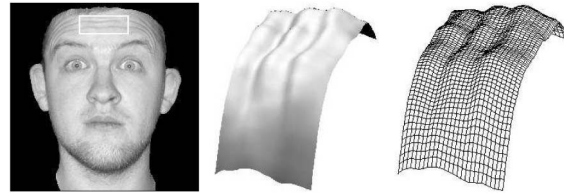

Fig. 16. Demonstration of recovery of local shape features. The (exaggerated) surface is shown on the right, recovered from the highlighted region of the image on the left.

In Table 1, we compare the strengths and weaknesses of the two approaches. The potential advantages of our method include: an efficient implementation (each step of the algorithm in Section 3.3 is implemented as a matrix multiplication), faster convergence (we obtain a stable solution from a variety of initialisations in 25-30 iterations), and the potential for higher accuracy (enforcing data-closeness as a hard constraint avoids model dominance and allows accurate recovery of atypical surface features). Finally, our model is consistent with psychological observations regarding human face perception. Our approach suggests that ideas from shape-from-shading could be incorporated into the morphable model framework in order to drive convergence and improve the recovery of local shape features.

5 CONCLUSIONS

We have shown how a statistical model of shape may be constructed from fields of surface normals using the azimuthal equidistant projection. We demonstrated that such a model trained on facial needle maps extracted from range data captures facial shape in a compact manner and is capable of generalizing to out-of-sample faces.

We presented two methods for fitting the model to image brightness data. The first of these involves fitting the model to a field of surface normals extracted from the image using shape-from-shading. The second method is an iterative one in which the model is used as a statistical constraint in a geometric shape-from-shading framework. This process can

TABLE 1

Comparison of Face Reconstruction Methods		
	Vetter and Blanz	SFS
Illumination:	**Unknown, single point**	Known, single point (could be estimated [29])
Pose:	**Arbitrary**	Frontal
Minimum number of images:	1	1
Albedo:	Linear Statistical Constraint	**Arbitrary**
Reflectance:	Phong	**Any isotropic, monotonic BRDF, corrected to Lambertian**
Satisfaction of Data-closeness:	Weak	**Strict**
Alignment:	**Manually initialised with sparse feature points, fitting improves**	Manual
Approximate shape recovery time:	4.5 minutes	**10 seconds**

be posed as that of recovering the best-fit field of surface normals from the statistical model, subject to constraints provided by the image irradiance equation. The method proves rapid to converge, is robust to variations in illumination direction of up to 45 degrees from the viewing direction, and delivers realistic surfaces when the fields of surface normals are integrated. We then showed how the fitted models could be used to improve the reilluminations of fields of surface normals recovered using shape-from-shading and to estimate an albedo map of the face.

Our future plans revolve around placing the iterative process in a statistical setting using the EM algorithm and a von-Mises distribution to model the likelihood for the surface normal data. We also plan to develop ways of aligning the model with images which are not in a frontal pose. Finally, we intend to investigate using the fitted models for the purposes of face recognition. This can be achieved in two ways: first, by using them as a generative model from which the pose and illumination space can be sampled, and second, the model parameters may provide an illumination insensitive description of the face which can be used for recognition.

RREFERENCES

[1] "USF HumanID 3D Face Database," Courtesy of Sudeep Sarkar, Univ. of South Florida, Tampa, 2004.
[2] J.J. Atick, P.A. Griffin, and A.N. Redlich, "Statistical Approach to SFS: Reconstruction of 3D Face Surfaces from Single 2D Images," *Neural Computation,* vol. 8, no. 6, pp. 1321-1340, 1996.
[3] V. Blanz, K. Scherbaum, T. Vetter, and H.P. Seidel, "Exchanging Faces in Images," *Proc. EUROGRAPHICS,* pp. 669-676, 2004.
[4] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces," *Proc. Computer Graphics SIGGRAPH,* pp. 187-194, 1999.
[5] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 9, pp. 1063-1074, Sept. 2003.
[6] V. Bruce, A. Coombes, and R. Richards, "Describing the Shapes of Faces Using Surface Primitives," *Image and Vision Computing,* vol. 11, no. 6, pp. 353-363, 1993.
[7] M. Castelán and E.R. Hancock, "Acquiring Height Maps of Faces from a Single Image," *Proc. Int'l Symp. 3D Data Processing Visualization and Transmission,* pp. 183-190, 2004.
[8] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *Proc. European Conf. Computer Vision,* pp. 484-498, 1998.
[9] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham, "Training Models of Shape from Sets of Examples," *Proc. British Machine Vision Conf.,* pp. 9-18, 1992.
[10] N.P. Costen, T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Automatic Extraction of the Face Identity-Subspace," *Image and Vision Computing,* vol. 20, pp. 319-329, 2002.
[11] M.P. Do Carmo, *Differential Geometry of Curves and Surfaces.* Prentice-Hall, 1976.
[12] R. Dovgard and R. Basri, "Statistical Symmetric Shape from Shading for 3D Structure Recovery of Faces," *Proc. European Conf. Computer Vision,* vol. 2, pp. 99-113, 2004.
[13] R.A. Fisher, "Dispersion on a Sphere," *Proc. Royal Soc., Ser. A,* vol. 217, pp. 295-305, 1953.
[14] R.T. Frankot and R. Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 4, pp. 439-451, Apr. 1988.
[15] A. Georghiades, P. Belhumeur, and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, pp. June 2001.
[16] R.L. Gregory, "Knowledge in Perception and Illusion," *Philosophical Trans. Royal Soc. London B,* vol. 352, pp. 1121-1128, 1997.
[17] T. Heap and D. Hogg, "Extending the Point Distribution Model Using Polar Coordinates," *Image and Vision Computing,* vol. 14, pp. 589-599, 1996.
[18] G.W. Hill, "Algorithm 571: Statistics for von Mises' and Fisher's Distributions of Directions: $I_1(x)/I_0(x)$, $I_{1.5}(x)/I_{0.5}(x)$ and Their Inverses," *ACM Trans. Math. Software,* vol. 7, no. 2, pp. 233-238, 1981.
[19] H. Hill and V. Bruce, "Effects of Lighting on the Perception of Facial Surfaces," *J. Experimental Psychology: Human Perception and Performance,* vol. 22, no. 4, pp. 986-1004, 1996.
[20] A. Johnston, H. Hill, and N. Carman, "Recognising Faces: Effects of Lighting Direction, Inversion, and Brightness Reversal," *Perception,* vol. 21, pp. 365-375, 1992.
[21] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Towards Automatic Simulation of Ageing Effects on Face Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 4, pp. 442-455, Apr. 2002.
[22] K.V. Mardia and P.E. Jupp, *Directional Statistics.* John Wiley and Sons Ltd., 2000.
[23] D. Marr, *Vision.* W.H. Freeman and Co., 1982.
[24] S.R. Marschner, B. Guenter, and S. Raghupathy, "Modeling and Rendering for Realistic Facial Animation," *Proc. 11th Eurographics Workshop Rendering,* pp. 231-242, 2000.
[25] S. Marschner, S. Westin, E. Lafortune, K. Torrance, and D. Greenberg, "Reflectance Measurements of Human Skin," Technical Report PCG-99-2, Program of Computer Graphics, Cornell Univ., 1999.
[26] I. Matthews, T.F. Cootes, and J.A. Bangham, "Extraction of Visual Features for Lipreading," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 2, pp. 198-213, Feb. 2002.
[27] D. Nandy and J. Ben-Arie, "Shape from Recognition: A Novel Approach for 3-D Face Shape Recovery," *IEEE Trans. Image Processing,* vol. 10, no. 2, pp. 206-217, 2001.
[28] X. Pennec, "Probabilities and Statistics on Riemannian Manifolds: A Geometric Approach," Technical Report RR-5093, INRIA, 2004.
[29] A.P. Pentland, "Finding the Illuminant Direction," *J. Optical Soc. Am.,* vol. 72, no. 4, pp. 448-455, 1982.
[30] E. Prados and O. Faugeras, "A Rigorous and Realistic Shape from Shading Method and Some of Its Applications," Technical Report RR-5133, INRIA, 2004.
[31] A. Robles-Kelly and E.R. Hancock, "Estimating the Surface Radiance Function from Single Images," *Graphical Models,* vol. 67, no. 6, pp. 518-548, 2005.
[32] D. Samaras and D. Metaxas, "Illumination Constraints in Deformable Models for Shape and Light Direction Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 247-264, Feb. 2003.
[33] L. Sirovich, "Turbulence and the Dynamics of Coherent Structures," *Quartly Applied Math.,* vol. XLV, no. 3, pp. 561-590, 1987.
[34] J.P. Snyder, *Map Projections—A Working Manual, USGS Professional Paper 1395.* US Government Printing Office, 1987.
[35] N.F. Troje and H.H. Bülthoff, "Face Recognition under Varying Poses: The Role of Texture and Shape," *Vision Research,* vol. 36, no. 12, pp. 1761-1771, 1996.
[36] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 586-591, 1991.
[37] M.J.C. van Gemert, S.L. Jacques, H.J.C.M. Sterenborg, and W.M. Star, "Skin Optics," *IEEE Trans. Biomedical Eng.,* vol. 36, no. 12, pp. 1146-1154, 1989.
[38] P.L. Worthington and E.R. Hancock, "New Constraints on Data-Closeness and Needle Map Consistency for Shape-from-Shading," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 12, pp. 1250-1267, Dec. 1999.
[39] P.L. Worthington and E.R. Hancock, "Surface Topography Using Shape-from-Shading," *Pattern Recognition,* vol. 34, no. 4, pp. 823-840, 2001.
[40] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-Time Combined 2D + 3D Active Appearance Models," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 535-542, 2004.
[41] R. Zhang, P.S. Tsai, J.E. Cryer, and M. Shah, "Shape-from-Shading: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 8, pp. 690-706, Aug. 1999.
[42] W.Y. Zhao and R. Chellappa, "Symmetric Shape-from-Shading Using Self-Ratio Image," *Int'l J. Computer Vision,* vol. 45, pp. 55-75, 2001.

William A.P. Smith received the BSc degree in computer science from the University of York, United Kingdom in 2002. He is currently studying toward the PhD degree in computer vision, also at the University of York, under the supervision of Professor E.R. Hancock. His research interests include shape-from-shading, face recognition, estimation of reflectance properties, statistical and geometric shape modeling, and the psychology of face perception. He has published more than 25 papers in refereed conferences and journals. He is a student member of the IEEE and the British Machine Vision Association.

Edwin R. Hancock studied physics as an undergraduate at the University of Durham and graduated with honors in 1977. He remained at Durham to complete the PhD degree in the area of high energy physics in 1981. Following this, he worked for 10 years as researcher in the fields of high-energy nuclear physics and pattern recognition at the Rutherford-Appleton Laboratory (now the Central Research Laboratory of the Research Councils). During this period, he also held adjunct teaching posts at the University of Surrey and the Open University. In 1991, he moved to the University of York as a lecturer in the Department of Computer Science. He was promoted to senior lecturer in 1997 and to reader in 1998. In 1998, he was appointed to a chair in computer vision. Professor Hancock now leads a group of some 15 faculty, research staff, and PhD students working in the areas of computer vision and pattern recognition. His main research interests are in the use of optimization and probabilistic methods for high and intermediate level vision. He is also interested in the methodology of structural and statistical pattern recognition. He is currently working on graph-matching, shape-from-X, image databases, and statistical learning theory. His work has found applications in areas such as radar terrain analysis, seismic section analysis, remote sensing, and medical imaging. He has published more than 100 journal papers and 400 refereed conference publications. He was awarded the Pattern Recognition Society medal in 1991 and an outstanding paper award in 1997 by the journal *Pattern Recognition*. In 1998, he became a fellow of the International Association for Pattern Recognition. He has been a member of the Editorial Boards of the journals *IEEE Transactions on Pattern Analysis and Machine Intelligence* and *Pattern Recognition*. He has also been a guest editor for special editions of the journals *Image and Vision Computing* and *Pattern Recognition*. He has been on the program committees for numerous national and international meetings. In 1997 with Marcello Pelillo, he established a new series of international meetings on energy minimization methods in computer vision and pattern recognition.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.