Communicated by Paul Viola

Statistical Approach to Shape from Shading: Reconstruction of Three-Dimensional Face Surfaces from Single Two-Dimensional Images

Joseph J. Atick
Paul A. Griffin
A. Norman Redlich
Computational Neuroscience Laboratory, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399 USA

The human visual system is proficient in perceiving three-dimensional shape from the shading patterns in a two-dimensional image. How it does this is not well understood and continues to be a question of fundamental and practical interest. In this paper we present a new quantitative approach to shape-from-shading that may provide some answers. We suggest that the brain, through evolution or prior experience, has discovered that objects can be classified into lower-dimensional object-classes as to their shape. Extraction of shape from shading is then equivalent to the much simpler problem of parameter estimation in a low-dimensional space. We carry out this proposal for an important class of three-dimensional (3D) objects: human heads. From an ensemble of several hundred laser-scanned 3D heads, we use principal component analysis to derive a low-dimensional parameterization of head shape space. An algorithm for solving shape-from-shading using this representation is presented. It works well even on real images where it is able to recover the 3D surface for a given person, maintaining facial detail and identity, from a single 2D image of his face. This algorithm has applications in face recognition and animation.

1 Introduction

Our brain is remarkable in its ability to perceive a three-dimensional world from the two-dimensional images projected on the retina (Ramachandran 1988; Todd and Mingolla 1983; Mingolla and Todd 1986; Gulick and Lawson 1976). How it achieves this task is poorly understood and continues to be an active topic within the neuroscience and machine vision communities. What is well understood are the cues it uses to extract a three-dimensional (3D) interpretation.

Many of these cues have been exploited by artists over the years to add realism to their work. For example, *shading* patterns, first used in

the fourteenth century by Renaissance painters, create a vivid impression of 3D shape in two-dimensional (2D) paintings. While there are other important cues that contribute to our perception of 3D space—such as binocular disparity and motion parallax—in this paper we are interested in the problem of recovering 3D shape from a single 2D image using *shading* information only: the so-called shape-from-shading problem (for an excellent explication of this problem, see Horn and Brooks 1989). Shading is the variation in brightness from one point to another in an image. It carries information about shape because the amount of light a surface patch reflects depends on its orientation (surface normal) relative to the incident light. So, in the absence of variability in surface reflectance properties (surface material), the variability in brightness can be due only to changes in local surface orientation and hence conveys strong information about shape.

At the outset we should emphasize that shape-from-shading is fundamentally a very difficult mathematical problem. The brightness at a given point fixes only the projection of the surface normal onto the incident light vector (see next section), and hence one cannot associate a unique normal to each point. In fact, if normals are assigned independently at each point, then there is an infinite number of normal vector fields that are consistent with the image brightness data, so the problem appears to be ill-posed. Of course, normals cannot be independent since the vector field of a true smooth surface satisfies constraints such as integrability (Guggenheimer 1977). These constraints imply that normals are strongly coupled across the surface and cannot be determined by purely local or point by point analysis. It is this coupling that is at the heart of the complexity of this problem. To make things worse, in typical situations the light source and reflectance properties of the surface are not known and have to be estimated simultaneously with shape.

Most algorithms proposed thus far in the literature attempt to make shape-from-shading well-posed by imposing some smoothness constraints that cut the infinite number of solutions consistent with the image data down to the few that satisfy the constraints. This approach, while theoretically promising, in practice suffers from a number of problems. These include sensitivity to smoothness parameters, multiple false solutions, nonrobustness against noise, and generally poor reconstruction for real world images.

We should note that most previous shape-from-shading algorithms are intended to be applicable to images of smooth but still otherwise arbitrary objects. Technically speaking, this means that one attempts to estimate shape in a space with an excessively large number of degrees of freedom[1] from the limited information contained in the image. The

[1]For a typical image size, the depth function $z(x, y)$ of the surface represents nearly one hundred thousand degrees of freedom. The generic constraints of smoothness are not likely to lower the dimensionality enough if one still insists on being able to represent arbitrary shapes.

difficulties common to such algorithms directly stem from working in this higher dimensional space.

While generality may be a noble aim from a mathematical point of view, it is neither clear that it is practically achievable nor is it obvious that the brain solves its shape-from-shading problems in such a way. It is well known that expectation and prior knowledge of the world can influence our interpretation of sensory data (Gregory 1970; Ramachandran 1990; Anstis 1991). What if it were true that the brain, through evolution or interaction with its environment, has discovered that objects can be classified into object-classes as to their shape? Shape space within each class may then be easy to parameterize and may even be very low-dimensional. If this is true, then shape-from-shading becomes equivalent to estimating a small number of parameters given an image and knowledge of what class the object belongs to—certainly a more tractable problem.

One way to explore this idea is to start with an ensemble of shapes of related 3D objects, use standard statistical techniques such as *principal component* analysis (Karhunen 1946; Loève 1955; Joliffe 1986), and derive a dimensionally reduced representation for shapes in this class. From a database of several hundred laser-scanned heads[2] we carry this out for the class of 3D human heads. We show that principal components provide an excellent low-dimensional parameterization of head shape that maintains facial detail and identity of the person. We use this representation to solve the shape-from-shading problem for any human head; we are able to recover an accurate 3D surface of the head/face of any person from a single 2D image of his face.

The organization of this paper is as follows: in Section 2 we define mathematically the problem of shape-from-shading. In Section 3 we derive a statistical parameterization of head-space that we use in Section 4 to solve shape-from-shading and apply it to real images of faces. Some details about the database used to extract the statistical regularities of human heads are given in Appendix A. In Appendix B we describe an algorithm for determining the light source from the image. Technical details related to the derivation of principal components of human heads are relegated to Appendix C.

For other approaches to shape-from-shading see Horn and Brooks (1989, and references therein), Horn (1970), Ikeuchi and Horn (1981), Brooks (1982), Pentland (1984, 1990), Lee and Rosenfeld (1989), Zheng and Chellapa (1991); Oliensis (1991), and Lehky and Sejnowski (1988). For other approaches to the construction of low-dimensional parameterizations of shape space see Cutzu and Edelman (1995), Edelman (1995), and more specifically for face shape see Vetter and Poggio (1995, and references therein).

[2]This database was made available to us by the Human Engineering Division of the Wright-Patterson Air Force Base; see Appendix A for further information about the database.

2 The Shape-from-Shading Problem

Mathematically speaking, shape-from-shading is equivalent to an inverse rendering problem. As such we will formulate it within a given rendering model—the so-called Lambertian model of surface reflectance. We should keep in mind that the algorithm presented in Section 4 holds for any other rendering model. With the assumption of orthographic projection and Lambertian surfaces, the rendering equation for a single light source is given by[3]

$$\begin{aligned} I(x,y) &= \eta(x,y)\ \mathbf{L}\cdot\mathbf{n}^S(x,y) \\ &\equiv \eta(x,y)\ R(\mathbf{L},\mathbf{n}^S) \end{aligned} \tag{2.1}$$

where $\mathbf{L} \equiv (L_x, L_y, L_z)$ is a vector representing the incident light and $\mathbf{n}^S(x,y)$ is the normal to the surface. $\eta(x,y)$ is called the albedo, and it represents the deviation in reflectance properties due to pigmentation or markings on the surface. Finally, $R(\mathbf{L},\mathbf{n}^S) = \mathbf{L}\cdot\mathbf{n}^S$ is known as the reflectance map. Strictly speaking this model of image formation includes a hard nonlinearity that sets I to zero at points where the reflectivity is negative; those are the points of self-shadowing. In the rest of this paper we will continue to suppress this nonlinearity from our displayed equations, but the reader should keep in mind that it is implicitly taken into account.

We will describe surfaces parametrically. For example, human heads can be described by the function $r(\theta,\ell)$ in cylindrical coordinates, where r is the radius and ℓ and θ are the height and angular coordinates, respectively. The Euclidean (x,y,z) coordinates of each point on the surface are related to these through

$$\begin{aligned} \mathbf{V}(\theta,l) &\equiv [x(\theta,\ell), y(\theta,\ell), z(\theta,\ell)] \\ &= [x_0 + r(\theta,\ell)\sin\theta, y_0 + \ell, z_0 + r(\theta,\ell)\cos\theta] \end{aligned} \tag{2.2}$$

for some shift x_0, y_0, z_0 relating the position of the origin in the two coordinate systems. The local tangent plane to the surface is spanned by the vectors $\partial\mathbf{V}/\partial\theta$ and $\partial\mathbf{V}/\partial\ell$. Thus the direction of the normal $\mathbf{n}^S$ is given by the vector cross-product of these vectors:

$$\mathbf{n}^S(\theta,\ell) \propto \frac{\partial\mathbf{V}}{\partial\theta}\times\frac{\partial\mathbf{V}}{\partial\ell} \tag{2.3}$$

It is not difficult to show that the unit normal is

$$\begin{aligned} \mathbf{n}^S(\theta,\ell) &= \frac{1}{\sqrt{r^2 + (\frac{\partial r}{\partial\theta})^2 + r^2(\frac{\partial r}{\partial\theta})^2}} \\ &\quad\times\left(-\frac{\partial r}{\partial\theta}\cos\theta + r\sin\theta,\ -r\frac{\partial r}{\partial\ell},\ \frac{\partial r}{\partial\theta}\sin\theta + r\cos\theta\right) \end{aligned} \tag{2.4}$$

[3]The coordinates x, y are the 2D projections of the Euclidean 3D coordinates (x,y,z) in which the surface, $z(x,y)$, is embedded; the z axis is along the optical axis.

The procedure for rendering is to first compute $I(\theta, \ell)$ using equations 2.4 and 2.1; then use the coordinate transform 2.2 to recover $I(x, y)$. In general this will evaluate $I(x, y)$ at nonuniform values of x, y; we recover the image on a uniform coordinate grid through standard interpolation techniques (see Wolberg 1992).

Given the image $I(x, y)$ the problem is to find the surface S, the albedo η, and the light $\mathbf{L}$ that satisfy equation 2.1. In general, these three are unknown and have to be determined simultaneously. In practice, however, one may be able to estimate each of the three quantities separately. For example, one could assume initially the albedo is constant, estimate the light direction, and then use that estimate to find the surface. One may even use iterative schemes that alternate between estimating the light direction and the surface shape (Brooks and Horn 1985).

Estimating light source direction turns out not to be a major problem. There are now many successful algorithms for doing so (Pentland 1982; Brooks and Horn 1985; Lee and Rosenfeld 1989; Zheng and Chellappa 1991). In Appendix B we present our algorithm, which is able to determine light direction with an accuracy of better than 5 degrees for images of faces under nonextreme illumination directions. Henceforth, we will not consider this problem and will assume that the light direction has been determined. We will also set the albedo to a constant, and we will return to it in a future publication. Actually, as we will see in Section 4, ignoring albedo turns out to be a very good first approximation for faces.

Equation 2.1 can be viewed as a nonlinear partial differential equation for the surface function $r(\theta, \ell)$. Unfortunately, thinking of it this way is not very useful for real images: Standard methods of numerical integration of differential equations (e.g., characteristic strip method) in practice fail miserably. These methods are inherently too sensitive to noise and they require knowledge of boundary conditions.

An alternative formulation is to think of shape-from-shading as an optimization problem where one attempts to minimize the average error

$$E = \int dxdy \left[I(x, y) - R(\mathbf{L}, \mathbf{n}^S) \right]^2 \tag{2.5}$$

with respect to the surface shape. As stated in the introduction, without imposing constraints, this problem is ill-defined: there is an infinite number of solutions. The constraints are typically chosen to ensure smoothness of the recovered surface and to allow one solution to be favored. There is an entire mathematical discipline known as *regularization theory* that attempts to do this (Tikhonov and Arsenin 1977). The problem is that again in practice, standard regularization does not work satisfactorily with shape-from-shading. Among its problems are sensitivity to choice of constraints and a proliferation of local minima. To avoid these problems, we give up generality and focus on solving shape-from-shading within one object class at a time where knowledge of properties of the class can be used to severely limit the shape degrees of freedom. In this paper we implement this approach for the class of human heads.

3 Low-Dimensional Representation of Human Head Shape

Human head shape is amazingly consistent across billions of people. The gross structure is invariably the same; all people have protrusions we call noses, depressions we call eye sockets, and flatter regions for foreheads and cheeks. Anthropometric surveys (Hursh 1976) have examined the extent of this similarity and have quantitatively confirmed that the variability from one head to another is, in fact, relatively small. Nevertheless, it is these small deviations (on average on the order of a centimeter) that give a face its unique identity. Actually, face shapes are much like imprints on coins, they are generated through small deviations from the large scale structure.

This suggests a hierarchical representation for human head shape: In cylindrical coordinates we can describe any given face $r(\theta, \ell)$ as a perturbation about a function common to all faces, $r_0(\theta, \ell)$—the so-called "mean-head":

$$r(\theta, \ell) = r_0(\theta, \ell) + \rho(\theta, \ell) \tag{3.1}$$

where $\rho(\theta, \ell)$ are small fluctuations $\rho/r_0 < 1$ that capture the identity of the person.

In this paper we take $r_0(\theta, \ell)$ to be the average of the first 200 heads from the USAF database (see Appendix A), *i.e.*,

$$r_0 = \frac{1}{N} \sum_{t=1}^{N} r^t(\theta, \ell) \tag{3.2}$$

In general, however, one may use several r_0s corresponding to the averages of a few clusters. For example, adult males may be described by one cluster while females by another. Furthermore, we should keep in mind that we are free to perform global 3D transformations such as scaling on a given head if necessary to maintain the validity of the expansion 3.1 even for unusually small or large heads.

To represent $\rho(\theta, \ell)$ we use principal component analysis (Karhunen 1946; Loève 1955; Joliffe 1986) and expand the fluctuations in terms of the set of eigenfunctions Ψ_i:

$$r(\theta, l) = r_0(\theta, l) + \sum_i a_i \Psi_i(\theta, \ell) \tag{3.3}$$

The eigenfunctions Ψ_i are derived from the first 200 heads in the USAF database using the procedure described in Appendix C. They represent an empirically derived basis set that captures the statistical regularities in the database and creates a low-dimensional representation of the data. Among their desirable properties is the fact that the mean square error introduced by truncating the summation in 3.3 is minimal.

In Figure 1 we show the mean-head and the first 15 eigenmodes rendered after adding each mode to $r_0(\theta, \ell)$. These pictures are very revealing; many modes seem to predominantly represent features that we can

identify as face components. The claim is that by taking appropriate linear combinations of these modes and adding them to the mean r_0, we can generate with good accuracy the shape of any human head. In fact it is this *generalization* ability of these modes that is crucial to this approach. We have tested this using the remaining 147 surfaces in the database (the ones not used to compute the mean or the modes). Since the eigenmodes are orthogonal, given a surface $r^t(\theta, \ell)$ for some t, its eigenmode expansion coefficients are simply

$$a_i = \frac{1}{\lambda_i} \int d\theta d\ell \Psi_i(\theta, \ell) \left[r^t(\theta, \ell) - r_0(\theta, \ell) \right] \tag{3.4}$$

where λ_i are the eigenvalues. From these coefficients the eigenmode representation of this surface is given by equation 3.3. We have computed the error $|r^{\text{actual}} - r^{\text{eigenmode}}|/r^{\text{actual}}$ for the 147 out-of-sample heads as a function of the number of modes used in the representation. The average over the 147 surfaces and over all points is shown in Figure 2, where we find that the error drops to about 1% by the time we use 40 modes.[4] However, this should be taken only as a rough indication of the quality of the representation since the error measure does not capture the dependence of the error on spatial position and also it has no perceptual meaning. Although, we can say that when the error is less than 1%, the reconstructed surface and the original are perceptually almost identical. This should be compared with the error when no modes are used, i.e., the distance to the mean-head, which ranges from 10 to 27%; ten percent error is perceptually very significant.

We should point out that principal components were used in Sirovich and Kirby (1987) and Kirby and Sirovich (1990) to derive a representation of images of human faces. There, it was shown that eigenmodes (so called eigenfaces) provided an excellent low-dimensional characterization of face images. In this paper we rely on the same principle that makes eigenfaces successful, namely, the fact that human faces whether imaged or as surfaces have few degrees of freedom and thus can be represented with a relatively small number of parameters. Here, of course, we compute eigenmodes for surfaces and not images, so these functions have a different interpretation and utility from those obtained by Sirovich and Kirby. In analogy with eigenfaces one may use the term eigenheads to refer to these modes.

[4]Notice what we have examined here is the out-of-sample generalization error, which is to be contrasted with the in-sample truncation error. The dependence of the latter on the number of modes is known theoretically (rms error is given by the sum of the eigenvalues of the modes that are dropped) and is not of interest in this context. The fact that the out-of-sample error turns out to be very small is a strong indication that the representation has captured the true characteristics of head shape to the point that it can represent any head.

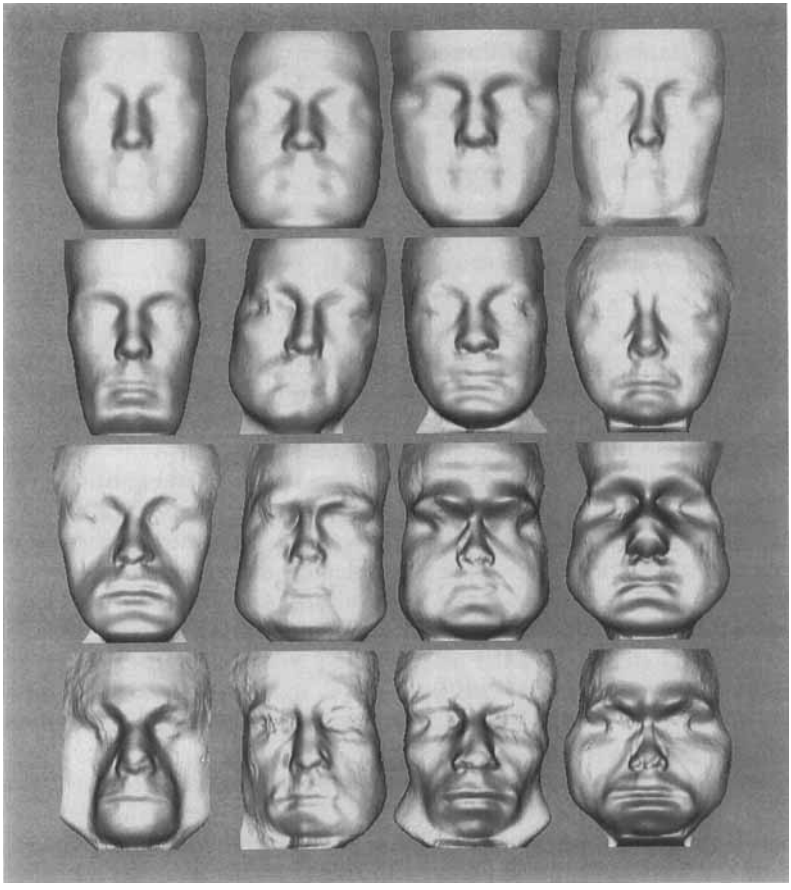

Figure 1: The mean-head surface, r_0 (*upper left-most corner*) and the 15 most significant eigenheads Ψ_i. To display them, the modes Ψ_i have been added to r_0 and then rendered straight-on (see Appendix A for details about database used to extract these modes).

4 Solving Shape-from-Shading

With the space of head shapes parameterized, any individual's head is given by specifying the coefficients $a_i, i = 1, \ldots, N$. For $N = 200$, the number of degrees of freedom in this space is much smaller than those available in a general $r(\theta, \ell)$ [at the resolution we are using $r(\theta, \ell)$ has

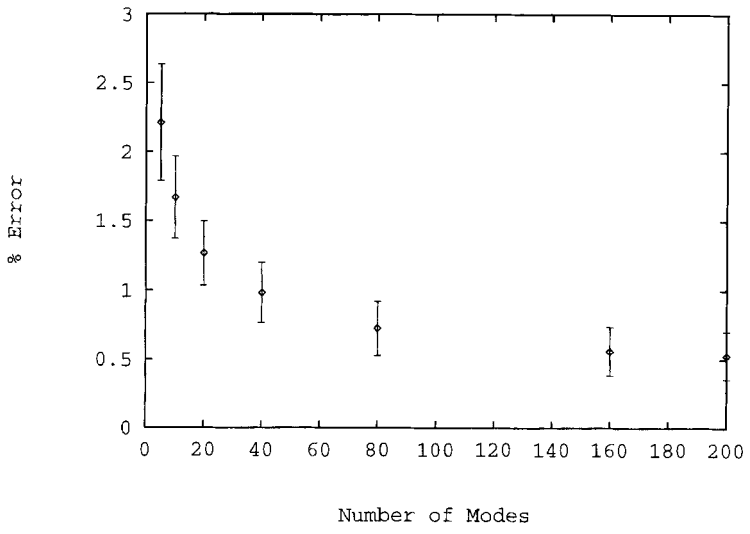

Figure 2: The dependence of the reconstruction error on the number of modes used in the representation. The error is defined as $|r^{\text{actual}} - r^{\text{eigenmode}}|/r^{\text{actual}}$ averaged over all points on the surface and over 147 out-of-sample heads (those not used to construct the eigenmodes) and is displayed as a percentage.

$256 \times 200 = 51{,}200$ real degrees of freedom representing visible surface]. In this section we consider only the problem of determining the surface; we assume that the algorithm in Appendix B was used to determine the source. We will also set the albedo $\eta = 1$; we will return to the problem of albedo in a future publication. Thus the only unknowns in the shape-from-shading problem are the eigenhead coefficients a_i.

To determine a_i we can try to minimize the error function

$$E = \int d\theta\, d\ell [I(x_0 + r\sin\theta, \ell) - R(\mathbf{L}, \mathbf{a})]^2 \tag{4.1}$$

with respect to the coefficients a_i, where

$$r(\theta, \ell) = r_0(\theta, \ell) + \sum_{i=1}^{200} a_i \Psi_i(\theta, \ell) \tag{4.2}$$

and $R(\mathbf{L}, \mathbf{a})$ is the reflectance map computed using 2.4 and 4.2. $I(x_0 + r\sin\theta, \ell)$ is the image $I(x, y)$ sampled in cylindrical coordinates using the radius function r. Notice in this formulation as we vary a_i looking for a minimum both $R(\mathbf{L}, \mathbf{a})$ and $I(x_0 + r\sin\theta, \ell)$ vary. This has the undesirable effect that very often the algorithm can find a trivial solution that will

minimize 4.1, for example, $r \to 0$ or ∞. To avoid this problem, we take advantage of the perturbative nature of human heads in cylindrical coordinates and series expand the image I as follows:

$$I(x_0 + r\sin\theta, \ell) = I(x_0 + r_0\sin\theta, \ell) + \partial_x I(x_0 + r_0\sin\theta, \ell) \times \left[\sum_i a_i \Psi_i(\theta, \ell)\right] + \cdots \tag{4.3}$$

Generally, we find the first-order term is sufficient, but slightly better results can be achieved by keeping terms quadratic in fluctuation. This formulation is also computationally efficient since one computes $I(x_0 + r_0\sin\theta, \ell)$ and $\partial_x I(x_0 + r_0\sin\theta, \ell)$ from $I(x, y)$ once and for all before attempting to determine the coefficients a_i. We compute these from $I(x, y)$ and $\partial_x I(x, y)$ through linear interpolation.

We have used two minimization techniques to determine a_i: conjugate gradient and gradient descent (Press *et al.* 1992), and they both converge to the same solution. The algorithm was tested on about a dozen images generated by Lambertian rendering of surfaces from the out-of-sample portion of the USAF database. We have also tested it on real images taken from a video camera. In Figure 3 three typical results for the purely Lambertian images are shown. In the first column we show the images $I(x, y)$ used as input to the algorithm; in the second column we show the rendering of the reconstructed surface after the minimization algorithm converged. The third and fourth columns give a comparison between the image of the original surface and the image of the reconstructed surface at 90°.

One can verify again and again that the algorithm is capable of reproducing the original face shape for synthetic Lambertian images.[5] Of course, there is limited interest in such images beyond academic considerations, and the true test of this shape-from-shading algorithm is how well it works for real images, images taken by a camera. In Figure 4 we show one such test. The face in the center is a TIFF image taken by a video camera and cropped. The algorithm in Appendix B was used to determine the light direction and its strength (the light direction was found to a very good approximation to be straight-on with overall strength 124). For this image the shape-from-shading algorithm converged in 24 iterations of the conjugate gradient minimization. The surface that it extracts from the image is shown from four different points of view in Figure 4. Since the 3D laser scan of this person is not available we leave it to the reader to decide on the quality of the reconstruction. However we should point out that we found this reconstruction is excellent for designing total contact burn masks from a photograph,[6] an application this algorithm is currently being developed for.

[5] Although we have not done systematic quantification of the error in recovery, from the dozen examples that we considered the average error is on the order of 2%.

[6] These are facial masks that are worn by burn victims during an extended rehabilitation period.

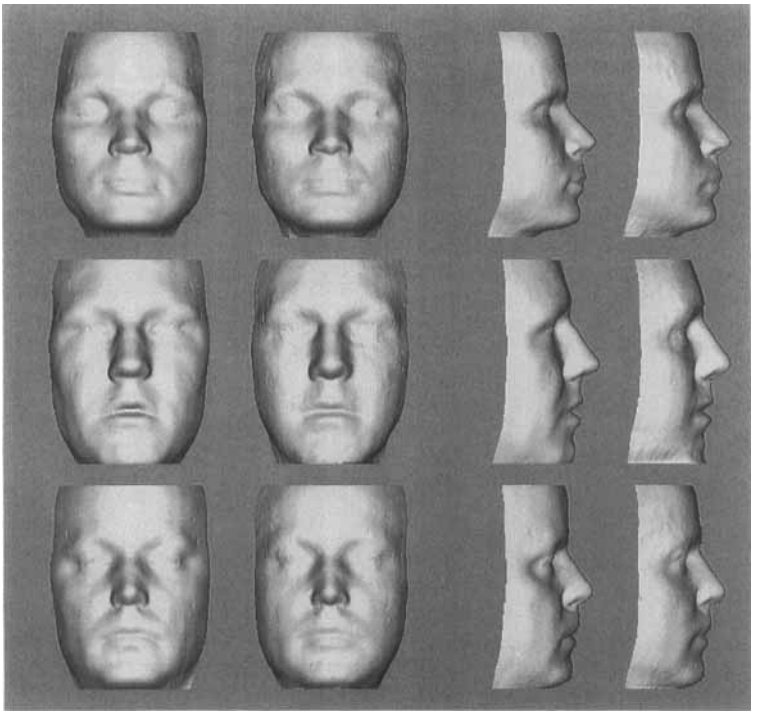

Figure 3: Three comparisons between surfaces extracted by the shape-from-shading algorithm with the true surfaces. The first column from the left are the 2D images used as input to the algorithm. The second column are the recovered surfaces rendered with a straight-on light. The third and fourth columns give the same comparison but viewed after rotation by 90°.

5 Comments and Relevance to the Brain

There are other practical applications to being able to construct a 3D model of a given face from a 2D image. For example, in face recognition, it is often necessary to change the pose of a person in an image to bring it to a canonical frontal pose just as it is necessary to compensate for the size of the head by scaling it to a canonical size (Atick *et al.* 1995). In Figure 5 we use the 3D model extracted by shape-from-shading in the previous section to generate the different poses from a single image. Of course there are other ways to synthesize different views from a single image. Among these, for example, is the technique of *Vetter and Poggio*

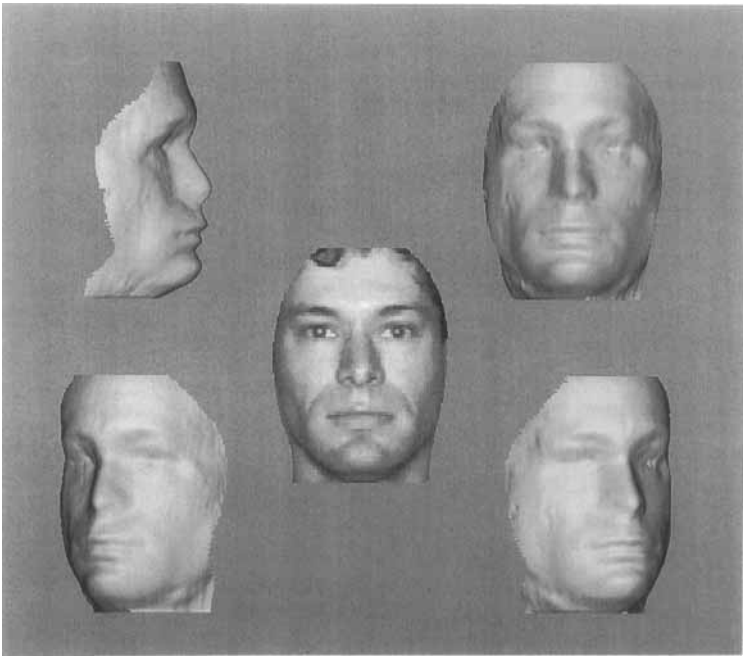

Figure 4: An example of a surface reconstructed from a real 2D image (*center*). Four views of the reconstructed surface are shown surrounding the original 2D image.

(1995) that exploits image transformations that are specific to the object class and learnable from example views of other objects of the same class. This technique, however, requires the derivation of 2D flow fields that map a person's image into a reference image, which, in general, is very difficult. For other techniques that achieve this see Beymer and Poggio (1995) and Lando and Edelman (1995).

We must emphasize that the current algorithm was not optimized for speed or output quality. It is preliminary in the sense that many of its engineering details can be better implemented. Also, within this approach, how well the algorithm works depends on the quality of the eigenhead surfaces, which in turn is a function of the database used. One can go back and rederive the eigenheads from larger and less noisy ensembles of heads (currently there are databases with 50,000 human heads available). Another problem of interest at a more fundamental

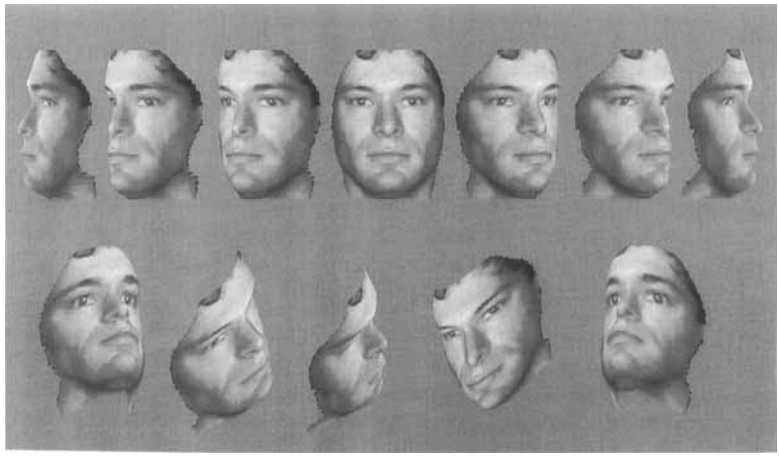

Figure 5: With knowledge of the 3D model of a face and a single image (same as the one in center of Fig. 6) we can generate through texture mapping what this person will look like in any 3D pose.

level is that of determining albedo. We intend to come back to this issue in a later publication.

Finally, one should explore whether the work presented here has any implications to the way the brain solves its shape perception problems. One may be tempted to argue that algorithms of shape-from-shading may not have any relevance to the brain since it is unlikely that the goal of vision is to reconstruct the outside world. This is a somewhat naive interpretation of what shape-from-shading algorithms are supposed to do. Shape is an intrinsic property of objects independent of the viewing and imaging conditions, and the brain could use some shape-from-shading algorithm to extract an invariant representation of objects, one that does not change as, say, the light changes. The eigenheads are precisely that; they provide an invariant representation whose elements are computable from image data. The output of the elements can then be used in cognitive tasks such as recognition and discrimination. In fact, preliminary work in our laboratory shows that shape information, even rudimentary, provides an additional strong signal over and above the albedo signal that improves discriminability in face recognition tasks.

The eigenheads have advantanges, in addition to their computability. For example, being low dimensional they yield better generalization. They also have certain testable implications. For example, since they are specific to the class of heads, one should expect semantic categorization

to affect the perception of shape-from-shading. There is evidence that this is indeed the case in the brain (Churchland *et al.* 1994). In the famous experiments of Helmholtz and in the more recent ones by Ramachandran, human masks viewed from the inside (concave view) invariably appeared convex with the nose protruding toward the observer.[7] The effect disappears when the masks are presented upside down as one would expect if the brain's shape perception mechanism probed by the experiment is specific to the category of heads (upside down heads are not normally encountered and are not expected to belong to the same category as heads).

The eigenheads can be used as a tool for probing the neural response even if one does not accept them as the elements of head representation in the brain. For example, in the temporal lobe it is well-known that neurons exhibit strong object specificity and the so called face-cells respond selectively to human faces and heads. What is not clear yet is what properties of faces these neurons are partial to. One can use the eigenheads as stimuli to systematically measure the response of face cells. One can even apply albedo maps to the shapes generated or render them under different lighting to determine the relative effect of 3D shape, color, and lighting on the neural response.

Appendix A: The Database

To derive the eigenheads we used a database of laser-scanned human heads. The database was made available to us by the Human Engineering Division of the Wright-Patterson USAF base and it is known as the "mini-survey." It contains 347 scanned heads of adult males. Each surface in the database is given as $r(\theta, \ell)$ with 512 units of resolution for θ and 256 units of resolution for ℓ. The data were generated using a CyberWare Laser scanner.

The data provided are unprocessed, which means they contain significant noise and gaps in the surfaces. We performed some simple preprocessing to fill in the gaps and to smooth some of the noise. We also cropped the data down to 256×200 (angular $\times$ height) around the nose since we are interested in the reconstruction of only the facial portion of heads. The back of the head was not used in the analysis presented in this paper. Finally, we have aligned the data using some 3D rigid transformations.

It is important to emphasize that we have split the database into two parts: the first, containing 200 surfaces, was used to compute the mean

[7] Another experiment that could reveal the specificity of the brain's shape perception mechanisms is a variant of the experiments pioneered by Koenderink *et al.* (1992). In these experiments subjects view images of rendered surfaces and adjust a gauge to reflect the perceived local surface normal at many different points on the surface. It would be interesting to see if human performance is quantitatively different when the surfaces are actually 3D human heads.

r_0 and the eigenmodes Ψ_i. The second, containing the remaining 147 surfaces, was used for out-of-sample testing (see Section 3).

Appendix B: Determining the Source

In this appendix we present an algorithm for determining the light source, $\mathbf{L}$, from an image of the face $I(x, y)$. The algorithm is very simple; it uses the mean head surface to model how light and shadow vary over a face as a function of incident light direction $\mathbf{L}$ and determines $\mathbf{L}$ by minimizing the error function:

$$E = \int d\theta\, d\ell \{I[x_0 + r_0(\theta, \ell) \sin\theta, \ell] - \mathbf{L} \cdot \mathbf{n}^0\}^2 \tag{B.1}$$

where $\mathbf{n}^0$ is the normal vector to the mean surface $r_0(\theta, \ell)$ calculated from equation 2.4. The image $I[x_0 + r^0(\theta, \ell) \sin\theta, \ell]$ is simply the input image $I(x, y)$ mapped to cylindrical coordinates using the mean surface r_0. We use linear interpolation to evaluate the image at nonintegral positions.

Since the energy is quadratic in $\mathbf{L}$, the minimum can be evaluated explicitly; $\delta E / \delta L_i = 0$ leads to

$$\sum_{j=1}^{3} L_j \langle n_j^0 n_i^0 \rangle = \langle I n_i^0 \rangle \tag{B.2}$$

whose solution is

$$L_j = \sum_{i=1}^{3} \langle I n_i^0 \rangle (\langle n_i^0 n_j^0 \rangle)^{-1} \tag{B.3}$$

with brackets indicating integration over (θ, ℓ). To avoid the rectification nonlinearity in the rendering equations (see discussion below equation 2.1) this integration is restricted to points where I is greater than zero.

We can test how well this works by rendering out-of-sample faces at different lights, using them as input to the algorithm, and estimating $\mathbf{L}^{\text{est}}$ by solving the above 3×3 linear equation. Figure 6 shows a typical graph of the estimation error, defined as the angle between the estimated vector $\mathbf{L}^{\text{est}}$ and the true vector used to generate the image $\mathbf{L}$, *i.e.*,

$$\Delta\Phi = \cos^{-1}\left(\frac{\mathbf{L} \cdot \mathbf{L}^{\text{est}}}{\|\mathbf{L}\|\, \|\mathbf{L}^{\text{est}}\|}\right) \tag{B.4}$$

The figure is a plot of $\Delta\Phi$ as a function of the standard illumination angles α and γ. These angles are related to the light direction by

$$\begin{aligned} \frac{L_x}{\|\mathbf{L}\|} &= \sin\gamma \cos\alpha \\ \frac{L_y}{\|\mathbf{L}\|} &= \sin\gamma \sin\alpha \\ \frac{L_z}{\|\mathbf{L}\|} &= \cos\gamma \end{aligned} \tag{B.5}$$

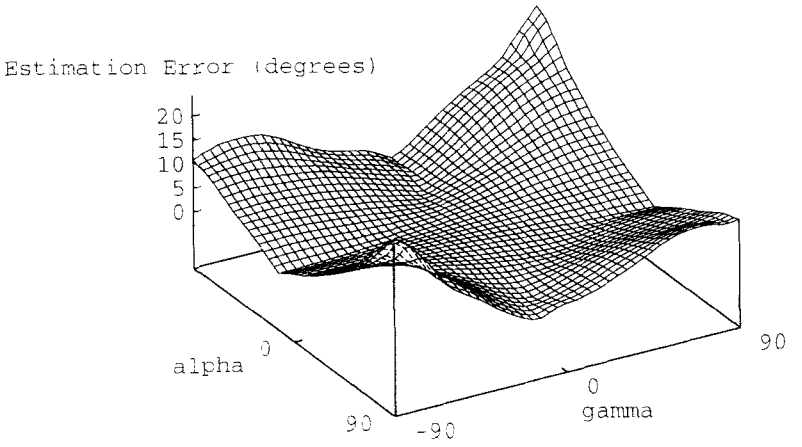

Figure 6: The angular error in estimating light source, $\Delta\Phi$, using the algorithm of Appendix B, as a function of the incident light direction angles, α and γ. For α and $\gamma < 75^\circ$ $\Delta\Phi$ is actually less than 5°.

We find, in the regime $|\alpha| < 75^\circ$, $|\gamma| < 75^\circ$, $\Delta\Phi < 5^\circ$ with the typical value for $\Delta\Phi$ less than 2°. The estimation error goes up to 20° for the extreme lighting conditions of $\alpha = 90^\circ$ and $\gamma = 90^\circ$. This is not surprising considering that under this lighting very little of the image is visible. Thus, for all lighting conditions of interest, this algorithm works exceedingly well.

The reason the algorithm works very well is that incident light is the same for all points on the face and thus can be determined from the large scale shading patterns. Large scale shading for the most part is independent of the identity of the person, and its dependence on the light direction is well captured by the renderings of the mean head, r_0. The algorithm is also robust to noise since $\mathbf{L}$ is estimated from moments $\langle In_i \rangle$ and $\langle n_i n_j \rangle$ that are averaged over a very large number of points. (This number varies since the average is carried out only over the points where I is nonvanishing but for nonextreme lighting conditions it is on the order of 40–50 thousand points.)

Other algorithms for estimating light source in the literature include Pentland (1982), Brooks and Horn (1985), Lee and Rosenfeld (1989), and Zheng and Chellappa (1991).

Appendix C: Principal Components of Human Heads

In this appendix we give some mathematical details relevant to the computation of the principal components of human heads. For an excellent exposition on principal components as applied to the analysis of large datasets see Sirovich and Everson (1992).

To start let $r^t(\mathbf{c}) \equiv [r^t(\theta, \ell) - r_0(\theta, \ell)]$ denote the surface deviation of the tth head surface in the database from the mean, $t = 1, \ldots, N$. The principal components or the eigenmodes are derived by solving the following set of linear equations:

$$\int d\mathbf{c}' R(\mathbf{c}, \mathbf{c}') \Psi_i(\mathbf{c}') = \lambda_i \Psi_i(\mathbf{c}) \tag{C.1}$$

where

$$R(\mathbf{c}, \mathbf{c}') = \frac{1}{N} \sum_{t=1}^{N} r^t(\mathbf{c}) r^t(\mathbf{c}') \tag{C.2}$$

is the covariance matrix and λ_i is the ith eigenvalue. Since $\mathbf{c} \equiv (\theta, \ell)$ ranges over $256 \times 200 = 51{,}200$ (see Appendix A), solving C.1 is equivalent to diagonalizing a $51{,}200 \times 51{,}200$ matrix $R(\mathbf{c}, \mathbf{c}')$, which is practically impossible even by today's computing standards. Luckily this diagonalizing turns out to be unnecessary, because of a technique called the *snap-shot* method invented by Sirovich (1987).

The method asserts that the eigenmodes are linear combinations of the original surfaces and can be written as

$$\Psi_i(\mathbf{c}) = \frac{1}{N} \sum_{p=1}^{N} \alpha_i^p r^p(\mathbf{c}) \tag{C.3}$$

Substituting C.3 into equation C.1 and exchanging the order of summation and integration, equation C.1 can be rewritten as

$$\sum_{n,p} r^n(\mathbf{c}) R_{np} \alpha_i^p = \lambda_i \sum_p \alpha_i^p r^p(\mathbf{c}) \tag{C.4}$$

where

$$R_{np} = \frac{1}{N} \int d\mathbf{c} r^n(\mathbf{c}) r^p(\mathbf{c}) \tag{C.5}$$

is an $N \times N$ matrix. This implies that the coefficients α_i^p satisfy

$$\sum_p R_{np} \alpha_i^p = \lambda_i \alpha_i^n \tag{C.6}$$

Solving this equation is equivalent to diagonalizing an $N \times N$ matrix. For $N = 200$ this can be very easily done. The snap-shot method has turned the nontractable problem of diagonalizing a $51{,}200 \times 51{,}200$ matrix into an equivalent but much more tractable problem involving the diagonalization of a 200×200 matrix.

To recover the original eigenmodes Ψ_i one first solves C.6 for the coefficients α_i^p. These define the linear combinations of the original surfaces $r^t(\mathbf{c})$ that must be combined in C.3 to recover Ψ_i.

Acknowledgments

We wish to thank Kathleen Robinette, Jennifer Whitestone, Barbara McQuiston, and Glen Geisen from the Human Engineering Division of the Wright-Patterson Air Force Base for their cooperation in making the "USAF mini-survey database" available to us. We also wish to thank Marc Rioux for supplying us with an initial set of laser scanned heads from the Canadian National Research Council, Mitch Feigenbaum for very helpful discussions, Penio Penev for his assistance in aligning the database, and Christina Nargolwala for useful comments on the manuscript. This work is supported in part by a grant from the Office of Naval Research contract number N00014-95-1-0381.

References

Anstis, S. 1991. Hidden assumptions in seeing shape from shading and apparent motion. In *Representation and Vision*, A. Gorea, ed., Cambridge University Press, Cambridge, UK.

Atick, J. J., Griffin, P. A., and Redlich, A. N. 1995. Face recognition from live video for real-world applications. *Adv. Imaging* **May**, 58–62.

Beymer, D., and Poggio, T. 1995. Face recognition from one model view. In *ICCV Proceedings*.

Brooks, M. J. 1982. Shape from shading discretely. Ph.D. thesis, Department of Computer Science, Essex University, Colchester, England.

Brooks, M. J., and Horn, B. K. P. 1985. Shape and source from shading. In *Proc. Int. Joint Conf. Artificial Intell.* Los Angeles, 932–936.

Churchland, P. S., Ramachandran, V. S., and Sejnowski, T. J. 1994. A critique of pure vision. In *Large-Scale Neuronal Theories of the Brain*, C. Koch and J. L. Davis, eds. MIT Press, Cambridge, MA.

Cutzu, F., and Edelman, S. 1995. *Exploration of shape space*. Weizman Institute Tech. Rep. CS-TR 95-01.

Edelman, S. 1995. Representation of similarity in 3D object discrimination. *Neural Comp.* 7, 407–422.

Foley, J., van Dam, A., Feiner, S., and Hughes, J. 1990. *Computer Graphics, Principles and Practice*. Addison-Wesley, Reading, MA.

Gregory, R. L. 1970. *The Intelligent Eye*. McGraw-Hill, New York.

Guggenheimer, H. W. 1977. *Differential Geometry*. Dover Publications, New York.

Gulick, W. L., and Lawson, R. B. 1976. *Human Stereopsis: A Psychophysical Analysis*. Oxford University Press, New York.

Horn, B.K.P. 1970. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Ph.D. Thesis, Department of Electrical Engineering, MIT.

Horn, B. K. P., and Brooks, M. J. 1989.*Shape from Shading*. MIT Press, Cambridge, MA.

Hursh, T. M. 1976. The study of cranial form: Measurement techniques and analytical methods. In *The Measures of Man*, E. Giles and J. Fiedlaender, eds. Peabody Museum Press, Cambridge, MA.

Ikeuchi, K., and Horn, B. K. 1981. Numerical shape from shading and occluding boundaries. *Artificial Intelligence* **15**, 141–186.

Jolliffe, I. T. 1986. *Principal Component Analysis*. Springer-Verlag, New York.

Karhunen, K. 1946. Zur spektraltheorie stochasticher. *Prozesse Ann. Acad. Sci. Fennicae* **37**.

Kirby, M., and Sirovich, L. 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transact. Pattern Analysis and Machine Intelligence* **12**, 103–108.

Koenderink, J. J., van Doorn, A. J., and Kappers, A. M. L. 1992. Surface perception in pictures. *Percept. Psychophys.* **52**, 18–36.

Lambert, J. H. 1760. *Photometria sive de Mensura et Gratibus Luminis, Colorum et Umbrae*. Eberhard Klett, Augsburg. Translation in W. Engleman (1982), *Lambert's Photometrie*. Leipzig.

Lando, M., and Edelman, S. 1995. Generalization from a single view in face recognition. *Network* **6**, 551–576.

Lee, C. H., and Rosenfeld, A. 1989. Improved methods of estimating shape from shading using the light source coordinates system. In *Shape from Shading*, B. K. P. Horn and M. J. Brooks, eds., pp. 323–569. MIT Press, Cambridge, MA.

Lehky, S. R., and Sejnowski, T. J. 1988. Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature (London)* **333**, 452–454.

Loève, M. M. 1955. *Probability Theory*. Van Nostrand, Princeton, NJ.

Mingolla, E., and Todd, J. 1986. Perception of solid shape from shading. *Biol. Cybern.* **53**, 137–151.

Oliensis, J. 1991. Shape from shading as a partially well-constrained problem. *Computer Vision, Graphics, Image Process. Image Understand.* **54**, 163–183.

Pentland, A. P. 1982. Finding the illuminant direction. *J. Opt. Soc. Am.* **72**, 448–455.

Pentland, A. P. 1984. Local shading analysis. *IEEE Trans. Pattern. Anal. Machine Itell.* **PAMI-16**, 170–187.

Pentland, A. P. 1990. Linear shape from shading. *Int. J. Computer Vision* **4**, 153–162.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. 1992. *Numerical Recipes: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, Cambridge.

Ramachandran, V. S. 1988. Perception of shape from shading. *Nature (London)*, **331**, 163–166.

Ramachandran, V. S. 1990. Visual perception in people and machines. In *AI and the Eye*, A. Blake and T. Troscianko, eds., pp. 21–77, Wiley, New York.

Simmons, L. 1975. Diffuse reflectance spectroscopy: A comparison of the theories. *Appl. Optics* **15**, 603–604.

Sirovich, L. 1987. Turbulence and the dynamics of coherent structures. *Quart. Appl. Math.* **XLV**, 561–590.

Sirovich, L., and Everson, R. 1992. Management and analysis of large scientific datasets. *Int. J. Supercomputer Appl.* **6**, 50–68.

Sirovich, L., and Kirby, M. 1987. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am.* **a 4**, 519–524.

Tikhonov, A. N., and Arsenin, V. Y. 1977. *Solutions of Ill-posed Problems*. W. H. Winston, Washington, D.C.

Todd, J. T., and Mingolla, E. 1983. Perception of surface curvature and direction of illumination from patterns of shading. *J. Exp. Psychol. Human Percept. Perform.* **9**, 583–595.

Vetter, T., and Poggio, T. 1995. Linear object classes and image synthesis from a single example image. A.I. Memo No. 1531.

Wolberg, G., 1992. *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, CA.

Woodham, R. J. 1981. Analyzing images of curved surfaces. *Artificial Intelligence* **17**, 117–140.

Zheng, Q., and Chellappa, R. 1991. Estimation of illuminant direction albedo, and shape from shading. *IEEE Trans. Pattern Anal. Machine Intelligence* **13**, 680–702.

Received August 8, 1995; accepted February 2, 1996.