Eurographics Symposium on Rendering 2003
Per Christensen and Daniel Cohen-Or (Editors)

Recovering 3-D Shape and Reflectance From a Small Number of Photographs

Athinodoros S. Georghiades

Department of Electrical Engineering, Yale University, New Haven, CT 06520, U.S.A.

Abstract

There are computer graphics applications for which the shape and reflectance of complex objects, such as faces, cannot be obtained using specialized equipment due to cost and practical considerations. We present an image-based technique that uses only a small number of example images, and assumes a parametric model of reflectance, to simultaneously and reliably recover the Bidirectional Reflectance Distribution Function (BRDF) and the 3-D shape of non-Lambertian objects. No information about the position and intensity of the light-sources or the position of the camera is required. We successfully apply this approach to human faces, accurately recovering their 3-D shape and BRDF. We use the recovered information to efficiently and accurately render photorealistic images of the faces under novel illumination conditions in which the rendered image intensity closely matches the intensity in real images. The accuracy of our technique is further demonstrated by the close resemblance of the skin BRDF recovered using our method, to the one measured with a method presented in the literature and in which a 3-D scanner was used.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Physically based modeling, I.4.8 [Image Processing and Computer Vision]: Shading, Shape.

1. Introduction

One can envision a home computer user wanting to insert a 3-D model of his or her face into an interactive application to be rendered under variable lighting and viewpoint. This should ideally be done easily and cost-effectively using off-the-shelf equipment. This would require a way to recover the 3-D shape and reflectance without using a 3-D scanner to estimate the shape, or a gonioreflectometer to measure its BRDF.

Towards that end, we present an image-based technique that uses a small number of photographs of non-convex objects, possibly captured with a home camera, to recover both their 3-D shape and BRDF. No knowledge is assumed about the camera position or the light sources, except that the object in question is illuminated by a single light source placed at a different (but still unknown) position in each photograph. In other words, there is no need to perform cumbersome photometric or geometric calibration; the photographs could be easily captured by waving a light source in front of the object. The speed with which the images could be captured, and the small number required, assures that the subjects do not move in any appreciable way, and hence the required alignment can be easily performed.

We have developed a reconstruction technique that incorporates a parametric model of reflectance and uses the image brightness values to recover the surface shape (along with the surface normal field), the albedo (the ratio of incoming to outgoing radiance), the light source directions and intensities, and finally the parameters of the parametric reflectance model. We chose to use the Torrance and Sparrow (T-S) model[26] of reflectance in our reconstruction algorithm because it can capture the major effects of the BRDF of a large class of surfaces, including human skin. We argue in favor of the T-S model in more detail in Section 3. The experimental results in this paper also support its use in our reconstruction algorithm.

Section 5 shows reconstruction results both for faces as well as for an inanimate object. They demonstrate that our algorithm has the potential to handle a large variety of surfaces with distinctly different reflectance properties. Using our method, the estimated skin reflectance has been shown to closely resemble the measured reflectance function reported by Marschner, et al.[16]. In that work, they made use of the 3-D shape which was recovered using a 3-D scanner. Finally, in Section 6, we demonstrate the increased photorealism of synthetic images of human faces created using

the reflectance properties and 3-D shape recovered with our technique. Rendering these images was done under novel illumination conditions, extrapolating from those in the input (or training) images. The rendered image intensities closely match those in real images.

2. Previous Work

Traditionally, the measurement or estimation of the BRDF has been completely separated from the shape estimation process. If the shape was needed, it was usually recovered using a 3-D scanner, and was mostly utilized in image-based techniques for estimating the BRDF. In fact, non-image based methods did not use any shape information; the sample surface was planar. In these, specialized and expensive equipment known as gonioreflectometers were used to measure the BRDF[18, 26]. These devices accurately place the camera and light source at different positions with respect to the planar sample; each camera and light source position provides a single measurement of the BRDF.

Since the BRDF can have three or four degrees of freedom, measuring even a sparse set of its values can be very time consuming. The requirement of a flat sample, and the meticulous geometric and photometric calibration makes this method unsuitable for many kinds of surfaces, like the skin of living subjects, which cannot be rolled into a plane. Furthermore, such techniques provide only the BRDF; for rendering images under variable lighting an viewpoint, the shape of the surface would need to be acquired separately.

Various imaged-based methods have been developed over the years that speed-up the process of BRDF measurement; their common characteristic being the use of a 2-D sensor. Most of these previous methods utilize shape variation, usually determined with a 3-D range scanner, to measure multiple samples of the BRDF simultaneously. Because of the shape variation, and hence of the variation in surface normals, this results in a collection of different incidence and viewing direction samples. This collection of samples usually corresponds to a 2-D range of angles (or degrees of freedom) for the BRDF, leaving only one or two degrees of freedom to be measured in a series of measurements (or image acquisitions). Image-based methods that use shape in one form or another have been developed by Marschner et al.[16], Lu et al.[15], Ikeuchi and Sato[11], and Sato et al.[21].

Ikeuchi and Sato[11] use the shape of the surface, recovered using a 3-D range finder, and a single image to fit the parameters of the T-S reflectance model. Using these parameters, they predicted the values of the BRDF beyond the range of incident and emittance angles determined by the light source and viewing directions in the single image. This method has been extended by Sato et al.[21] to capture spatial variations in the BRDF using multiple images. They again fit reflectance model parameters to the acquired data. Our method also uses the T-S model, but unlike their methods, we do not assume any knowledge of the shape.

In the method by Lu et al.[15] the full angular range in the incidence plane is measured without using low parametric models. To accomplish that, they use a cylindrical sample and multiple images with different light source positions. This method has been extended by Marschner et al.[16], where the 3-D shape of the surface, recovered using a range finder, is used in conjunction with multiple images, where the camera moves but the light source is kept fixed, to measure the full angular range of the BRDF extending well beyond the incidence plane. With their method, they have been able to measure the BRDF of human skin on living subjects.

Unlike other image-based methods, our technique *simultaneously* determines the BRDF and the 3-D shape of the surface. The presented algorithm falls under the category of techniques that recover shape of objects from multiple images. These techniques are collectively known as *photometric stereo* methods, and they are related to inverse-rendering. When it was originally introduced in the early 1980's[28, 23, 10], photometric stereo was constrained to work under the assumption that the surface in question was Lambertian[13] (i.e., purely diffuse) and that the light sources were known. The latter assumption introduces the need of cumbersome calibration to determine the direction and intensity of light sources, which can be difficult to do accurately in practice. Although researchers have over the years applied non-Lambertian reflectance models to photometric stereo, they again assumed a known reflectance map and hence knowledge of the light source directions and strengths[25, 2, 14, 17, 24].

Our surface reconstruction and BRDF estimation algorithm is in essence an extended *uncalibrated* photometric stereo technique. It is uncalibrated because it assumes no knowledge of the reflectance map (i.e., the light sources and the parameters of the reflectance model are unknown), and it is extended because it does not assume that the surfaces are purely diffuse. Because of the latter attribute, it provides an alternative imaged-based way of estimating both the BRDF and the 3-D shape of surfaces.

Using only image brightness values to recover the surface shape, the albedo, the light source directions and intensities, and the BRDF is a difficult task. To make the problem tractable, we have assumed a parametric model of reflectance, namely the Torrance and Sparrow (T-S) model. This is a reasonable assumption because the T-S model of reflectance is quite expressive and can capture the reflectance properties of a large number of different surfaces. Other parametric models of reflectance have recently been developed which may be more accurate in modeling the reflectance of human skin, such as those by Hanrahan and Krueger[8] and Jensen et al.[12], but they are more complicated and difficult to use. The T-S model of reflectance is simpler in comparison and the promising experimental results presented in this paper justify its use.

3. Surface Reflectance Functions

The surface BRDF is usually represented by a four parameter function. Those four parameters are the two incoming light source direction angles, θ_i, ϕ_i, and the two viewing direction angles, θ_r, ϕ_r.

The BRDF is defined as follows:

$$r(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{dL_r(\theta_r, \phi_r; \theta_i, \phi_i)}{dE_i(\theta_i, \phi_i)}, \qquad (1)$$

where $dL_r(\theta_r, \phi_r; \theta_i, \phi_i)$ is the outgoing irradiance from an infinitesimal patch on the surface, and dE_i is the incident radiance from an infinitesimal source. Due to Helmholtz reciprocity[9], $r(\theta_i, \phi_i; \theta_r, \phi_r) = r(\theta_r, \phi_r; \theta_i, \phi_i)$. This means the BRDF is the same if the light source and the camera are interchanged.

The above representation assumes that the light is monochromatic, and that it arrives at and bounces away from the same point on the surface. This precludes translucency and phosphorescence. Under the isotropy assumption, a common simplification and not an unreasonable one, $r(\theta_i, \phi_i; \theta_r, \phi_r) = r(\theta_i, \theta_r, \phi_r - \phi_i) = r(\theta_i, \theta_r, \Delta\phi)$.

As it can be surmised by looking around the room, the BRDF of surfaces can be quite complicated. It can change dramatically not only across object boundaries but also within objects. To simplify things and to make them more tractable in different applications, such as photometric stereo, binocular stereo, tracking, and so on, the surface reflectance has usually been assumed to be Lambertian. Under the Lambertian assumption, the appearance of an object surface is the same as the viewing direction changes, and is only proportional to the cosine of the angle between the local surface normal and the light source direction. Invoking the Helmholtz reciprocity, it can be shown that the BRDF is constant, i.e.,

$$r_{\text{Lamb}} = a_d \tag{2}$$

where a_d is the Lambertian (diffuse) "albedo". A constant BRDF implies a perfectly diffuse surface, and although this is never true in reality, the Lambertian assumption can nonetheless be a good approximation to the reflectance function of many real surfaces.

On the other extreme is the perfectly specular reflectance exhibited by a perfect mirror. In this case, the BRDF of the surface can be shown to be a Dirac delta function with an infinite magnitude only when the incoming light source direction is the reflection of the viewing direction about the surface normal and within the incident plane. Like the Lambertian model, this is also an approximation of real mirror surfaces.

Over the years various non-Lambertian parametric models have been proposed for the reflectance of real-world surfaces. These parametric models can be divided into physically-based and empirically-based models. One issue that has bedeviled the computer graphics community is the complexity of these different models. The most well-known empirical model is the Phong model[20], and it has been popular because of its simplicity and fairly decent rendering results. Nevertheless, the Phong model has no physical basis because there are important reflectance effects it cannot capture. These include the significant increase in the BRDF values and the off-specular forward scattering when the incidence angle becomes large. It has been noted that many materials, including metals, oxides[26], and human skin[16], exhibit these properties.

The Torrance and Sparrow model[26] is a more physically-based model which can capture those two effects. It assumes that reflectance consists of two components. The first is associated with bulk material effects and it is assumed to lead to a Lambertian lobe colored by the albedo at a particular position on the surface. The other component is a specular lobe assumed to be purely related to surface scatter.

With the T-S model, the surface is modeled as a large collection of perfectly specular micro-facets whose surface normal deviation from the average surface normal is assumed to be a zero-mean Gaussian—the higher the variance of deviation the rougher the surface, and hence the duller its appearance. This surface scatter leads to a specular lobe in the forward direction which is unaffected by the bulk material properties. Combining these two terms gives the following BRDF for the T-S model:

$$r_{\text{TS}} = a_d + a_s Q F(\theta', \eta) \frac{\exp(-\nu^2 \theta_a^2)}{\cos\theta_i \cos\theta_r}, \tag{3}$$

where a_d is the Lambertian (diffuse) "albedo", a_s is the specular "albedo", ν is the surface roughness (the lower its value the higher the roughness), and θ_a is the angle between the surface normal and the bisector of the (incoming) light source direction and the (outgoing) viewing direction. Q is the bistatic shadowing (also known as the geometric attenuation) factor, $F(\theta', \eta)$ is the Fresnel reflectivity[22], where θ' is the phase angle, the bisecting angle between the incoming and outgoing directions, and η is the index of refraction. Note that the T-S model BRDF is isotropic and satisfies the Helmholtz reciprocity[9].

Furthermore, we make two more assumptions. First, while the Lambertian albedo a_d is allowed to vary spatially, the parameters of the specular term, a_s and ν, are assumed to be constant across the surface. (Note that θ_a in Equation 3 is still a function of (x, y) because it is a function of the surface normal field.) The spatial invariance of a_s and ν is, of course, not entirely true in real surfaces, but it is not an unreasonable assumption as demonstrated by the experimental results, while it makes the problem more tractable. Allowing a_s and ν to vary across the surface is an interesting problem, but it can be very difficult to solve reliably when one is also estimating the 3-D shape of the surface using only image intensities and no knowledge about the light sources. The other assumption is that this paper does not consider color. It is of course important, but this would require to estimate the wavelength dependence of the Fresnel reflectivity (and of the albedos) for different materials, which can be very cumbersome and is beyond the scope of this paper.

The image intensity derived with the T-S model when a single point light source $\mathbf{s}$ illuminates the object is given by

$$I_{\text{TS}} = a_d ||\mathbf{s}|| \cos\theta_i + a_s ||\mathbf{s}|| Q F(\theta', \eta) \frac{\exp(-\nu^2 \theta_a^2)}{\cos\theta_r}, \tag{4}$$

where $||\mathbf{s}||$ is the intensity of a point light source. Note that we assume the light source to be very far away from the object and hence its distance to each point on the surface is assumed to be the same. Therefore, in the equation above, the inverse-square distance term on the right-hand side has been absorbed into the $||\mathbf{s}||$ term.

The T-S model deviates from the Phong model on two

counts. First, there is the exponential function that comes from the Gaussian assumption of surface roughness mentioned above. It should be noted that the cosine term in the Phong model has no physical basis. It just provides a faster way of approximating the exponential function. The second difference derives from the existence of the $1/\cos\theta_r$ and the $1/\cos\theta_i$ terms in the expression for the T-S model in Equation 3.

These two cosine terms in the denominator can affect the BRDF considerably. They lead to a significant increase in the BRDF values along with (increasingly) off-specular maxima as the incident angle increases (See Figure 1 for an example.) As noted above, many materials, including metals, dielectrics (e.g., oxides)[26], and human skin[16], exhibit this behavior, and therefore makes the T-S model a good candidate for modeling the reflectance of these materials.

One aspect of reflectance not captured by the T-S model is the backscatter lobe. This can be observed in metals as well as in painted surfaces. In this case, the model by Oren and Nayar[19], which is based on the same surface roughness model and assumptions as the T-S model, is more suitable for modeling the backscatter lobe. Nevertheless, the Oren and Nayar model does not capture the forward scatter, a more common and arguably more important effect. Notably, human skin as well as many dielectrics (e.g., oxides) do not exhibit any significant back-scattering properties[16, 26] and the T-S model can capture their reflectance properties quite well.

4. Surface Reconstruction and Recovery of Reflectance Properties

The reconstruction algorithm we present here uses only image intensities to recover the 3-D shape and surface reflectance properties. As mentioned before, it is in essence an extended uncalibrated photometric stereo technique. It is uncalibrated because it assumes no knowledge of the reflectance map (i.e., the light sources and the parameters of the reflectance model are unknown), and it is extended because it incorporates an non-Lambertian reflectance model, namely the T-S model of reflectance shown in Equation 3.

In our current implementation, the bistatic scattering (or geometric attenuation), Q, and the Fresnel reflectivity, F, are ignored due to their very insignificant effect when the phase angle is quite small (i.e., when the light source and viewing directions are close together[25]), which was the case in the training images of the example reconstructions shown in the next section. Note that the Fresnel reflectivity becomes more significant as the phase angle increases and it was re-introduced in the image synthesis process presented in Section 6. Note also that this implementation employs gray-scale images, since with color images the wavelength dependence of the Fresnel reflectivity would need to be estimated. Nevertheless, it has been reported in the literature that for human skin the index of refraction, $\eta \sim 1.37 - 1.5$[27]. In Section 6 we have assumed that $\eta = 1.5$.

While $a_d(x,y)$, the Lambertian albedo in Equation 3, is allowed to vary spatially, the parameters of the specular term in the T-S model are assumed not to be a function of (x,y), i.e., a_s and ν are constant over (x,y). This assumption makes the overall problem significantly more tractable and is not an unreasonable one.

In this exposition, let the surface of the object viewed by an orthographic camera be modeled by a height function $z(x,y)$, where (x,y) denotes a Euclidean coordinate system in the image plane of the camera. Note that the requirement of an orthographic camera is not very stringent. The algorithm still works even when the face or object is only 1-2 feet away from the camera. The general rule of thumb is that the distance from the camera should be at least 3-4 times the maximum range of depth of the object surface.

Furthermore, assume that the surface reflectance of the object is given by the T-S model shown in Equation 3. For such a surface, its surface normal field is given by

$$\mathbf{n}(x,y) = \frac{(z_x(x,y), z_y(x,y), -1)}{\sqrt{z_x^2(x,y) + z_y^2(x,y) + 1}}, \tag{5}$$

where $z_x(x,y)$ and $z_y(x,y)$ are the $x-$ and $y-$derivatives of the object's surface. Although we assume that‘ the surface $z(x,y)$ is continuous, when estimating the surface normals continuity is not guaranteed—the estimates $\mathbf{n}^*(x,y)$ may not be an integrable field due to noise, i.e., the resulting field may not correspond to a continuous surface. To enforce continuity, we expand the surface $z(x,y)$ using basis surfaces (i.e., height functions):

$$z(x,y;c(\mathbf{w})) = \sum c(\mathbf{w})\phi(x,y;\mathbf{w}), \tag{6}$$

where $\mathbf{w} = (u,v)$ is a two dimensional index over which the sum is performed, and $\{\phi(x,y;\mathbf{w})\}$ is a finite set of basis functions which are not necessarily orthogonal. We chose the discrete cosine basis so that $\{c(\mathbf{w})\}$ is exactly the *full* set of discrete cosine transform (DCT) coefficients of $z(x,y)$ (which are equal to the number of pixels in the image). Note that the discrete cosine basis can fully describe any continuous surface $z(x,y)$, including non-convex surfaces. If the partial derivatives of each basis function, $\phi_x(x,y;\mathbf{w})$ and $\phi_y(x,y;\mathbf{w})$, are integrable (i.e., continuous), then the partial derivatives of $z(x,y)$ are guaranteed to be integrable as well; that is, $z_{xy}(x,y) = z_{yx}(x,y)$.

Note that the partial derivatives of $z(x,y)$ can also be expressed in terms of this expansion, giving

$$z_x(x,y;c(\mathbf{w})) = \sum c(\mathbf{w})\phi_x(x,y;\mathbf{w}) \tag{7}$$

$$z_y(x,y;c(\mathbf{w})) = \sum c(\mathbf{w})\phi_y(x,y;\mathbf{w}). \tag{8}$$

Since the $\phi_x(x,y;\mathbf{w})$ and $\phi_y(x,y;\mathbf{w})$ are integrable (continuous) and the expansions of $z_x(x,y)$ and $z_y(x,y)$ share the same coefficients $c(\mathbf{w})$, it is easy to see that $z_{xy}(x,y) = z_{yx}(x,y)$. Note that the surface normals $\mathbf{n}(x,y)$ are functions of $c(\mathbf{w})$ and are fully determined by them.

Let the surface $z(x,y)$ be illuminated by a single point light source $\mathbf{s}_i$, and let the camera acquire an n pixel image of this surface. Let the image be denoted by the vector $\mathbf{I} \in \mathbb{R}^n$. Note that the image is written as a vector with each pixel corresponding to an element I_j of the vector $\mathbf{I}$. A pixel

I_j samples the image irradiance at some point in the image plane as specified by its Euclidean coordinates (x,y). Under the T-S model, the intensity given in Equation 4 can also be expressed as

$$I_{\text{TS}}(a_d(x,y),\mathbf{n}(x,y;c(\mathbf{w})),a_s,\nu,\mathbf{s}_i) = a_d\mathbf{n}^T\mathbf{s}_i + \frac{a_s\exp(-\nu^2\left[\arccos\left(\frac{\mathbf{n}^T(\frac{\mathbf{s}_i}{||\mathbf{s}_i||}+\hat{\mathbf{v}})}{||\frac{\mathbf{s}_i}{||\mathbf{s}_i||}+\hat{\mathbf{v}}||}\right)\right]^2)}{\mathbf{n}^T\hat{\mathbf{v}}} \quad (9)$$

where $\hat{\mathbf{v}}$ is the unit-length viewing direction.

We acquire k images of the object by fixing the object's position relative to the camera and moving the light source to some unknown position before each acquisition. (One could use a video camera to quickly capture the necessary images as the light source moves. This would minimize any movements by the subject, which would significantly simplify the necessary alignment.) Let the acquired set of images be denoted by the matrix $X=[\mathbf{I}_1,\ldots,\mathbf{I}_k]$ with I_{ij} being the value of the j-th pixel in the i-th image. The unknown light source directions scaled by their corresponding source intensities are denoted by $\mathbf{s}_i$, for all k images.

Note that the images can have shadows (both cast and attached), and possibly saturations—these do not satisfy the T-S model of reflectance and should not be used in the estimation process. We therefore need to determine which pixels are observing either a saturation or a shadow and mark them as being invalid. Unlike saturations, which can be simply determined, finding shadows is more involved. In our implementation, a pixel is labeled as observing a shadow if its intensity divided by its corresponding Lambertian albedo is below a threshold—as an initial estimate of the albedo, we use the average of the training images. A conservative threshold is then chosen (for each collection of images) to determine shadows, making it almost certain that no invalid data is included in the estimation process, at the small expense of throwing away a few valid measurements. The only requirement we have is that there are at least three (and preferably more) valid measurements at each pixel position. Any invalid data (both shadows and saturations) are treated as missing measurements by our estimation method.

To determine the shape and reflectance properties of the object, we solve the following minimization problem:

$$\min_{c(\mathbf{w}),\,a_d(x,y),\,a_s,\,\nu,\,\mathbf{s}_i} O(c(\mathbf{w}),\,a_d(x,y),\,a_s,\,\nu,\,\mathbf{s}_i) = \sum_{ij} m_{ij}|I_{ij}-I_{\text{TS}}(a_d(x,y),\mathbf{n}(x,y;c),a_s,\nu,\mathbf{s}_i)|^2 \quad (10)$$

where,

$$m_{ij}=\begin{cases}1 & I_{ij}\text{ valid pixel measurement,}\\ 0 & \text{otherwise.}\end{cases}$$

I_{TS} is given in Equation 9, and the Cartesian position (x,y) corresponds to the j-th pixel. We solve this minimization using steepest descent in an iterative scheme. Note that all the parameters of I_{TS} are unknown and minimizing over all them at once can be formidable. We use an iterative scheme where in each iteration the optimization parameters are updated one after the other independently, i.e., optimization is performed over one parameter, e.g. $c(\mathbf{w})$, while the other parameters are kept fixed, then over $a_d(x,y)$, and so on.

After the initialization, we let the algorithm run for a number iterations with only the Lambertian model (where we keep a_s fixed to zero). We empirically chose to run the algorithm for 30 iterations after which the full T-S model is introduced. This is to prevent the solution from falling into a local minimum during the initial iterations and when the surface normals, albedos, and light sources are very far from their respective solutions. The full algorithm is given below:

1. Find the average of the training images and use it as an initial estimate of the albedo, $a_d(x,y)$. Set $a_s=0$ and $\nu=2$, and let fixed for the first 30 iterations. Initialize the DCT coefficients so that $c(\mathbf{w})=0,\ \forall\mathbf{w}$.
2. Without doing any row or column permutations, sift out all of the full rows (with no missing or invalid measurements) of matrix X to form a full sub-matrix $\tilde{X}$. The number of rows in $\tilde{X}$ is almost always larger than its number of columns, k. This is because the number of rows of X is equal to the number of pixels in an image, which can be in the thousands.
3. Perform SVD on $\tilde{X}$ to find an initial estimate of matrix $S\in\mathbb{R}^{3\times k}$ which best spans the row space of $\tilde{X}$. Each column of S corresponds to an initial estimate of light source $\mathbf{s}_i$. Note that under the Lambertian assumption and when there are no shadows, the image formation model is linear and could be decomposed (factorized) into its constituent parts using SVD. Of course, in this case, because of missing elements, we can only perform SVD on the full submatrix $\tilde{X}$ and can only get an initial estimate of S, but not of the surface normals. Because of the required sifting, this results in gaps in the normal field over the range of (x,y), which prevents us from finding an initial estimate of $c(\mathbf{w})$ using the projection technique of Frankot and Chellapa[6].
4. Update the values of $c(\mathbf{w})$ using steepest descent, by moving from its current estimate in the opposite direction of $\frac{\partial O(c(\mathbf{w}))}{\partial c(\mathbf{w})}$. The values of all other optimization parameters are kept fixed during this step. The partial derivatives with respect to each $c(\mathbf{w})$ are determined analytically by taking advantage of the known analytical structure of the objective function. To avoid overshooting during each step, each partial derivative with respect to $c(\mathbf{w})$ is divided by a factor proportional to $(P_x(\mathbf{w})+P_y(\mathbf{w}))$, where $P_x(\mathbf{w})=\sum_{x,y}|\phi_x(x,y;\mathbf{w})|^2$ and $P_y(\mathbf{w})=\sum_{x,y}|\phi_y(x,y;\mathbf{w})|^2$. This step is repeated multiple times; it is non-optimal since the objective function is not linear w.r.t. $c(\mathbf{w})$. In our implementation, we repeat it 12 times for better convergence of $c(\mathbf{w})$ before moving to the next step.
5. Update the Lambertian albedo $a_d(x,y)$ while all the other optimization parameters are kept fixed. The albedo $a_d(x,y)$ is allowed to vary across the object surface, unlike the specular albedo, a_s. Note that the optimization function $O(c(\mathbf{w}),a_d(x,y),a_s,\nu,\mathbf{s}_i)$ is linear with respect to each albedo value $a_d(x,y)$. Note also that each albedo value $a_d(x,y)$ is decoupled from all the rest and hence we

can easily solve for each independently using linear least squares.

6. Update each of the light source directions and strengths $\mathbf{s}_i$. Note that the light sources are not coupled with one another, hence they can be updated independently. Since each vector $\mathbf{s}_i$ is only a three-vector, we use Newton's method (instead of Steepest descent). Note that we could find a closed form solution of the partial derivatives $\nabla O(\mathbf{s}_i)$ and the Hessian $\nabla^2 O(\mathbf{s}_i)$, but it proved easier to estimate them numerically.
7. Update specular albedo a_s and surface roughness ν if the iteration count is more than 30. Note that the optimization function O is linear with respect to a_s. Hence we can solve for a_s using linear least squares. The roughness parameter ν is solved using Newton's method. (Note that a_s and ν are constant over the whole surface and hence they are not functions of (x, y).)
8. Repeat steps 4-7 until the estimates converge, as judged by the rate of change of the objective function.
9. Perform inverse DCT on the coefficients $c(\mathbf{w})$ to get the surface $z(x, y)$.

Allowing the algorithm to run for a few iterations (30 in this case) with only the Lambertian model before introducing the full T-S model, prevents the algorithm from falling in a local minimum in almost all cases. The algorithm converges within 200-2000 iterations depending on the conditioning of the data. By conditioning, we mean the strength of the specular lobe that drives the solution to find the 3-D shape. It is well known that when there is no specular component (i.e., under the Lambertian assumption) and when no knowledge about the light sources is assumed, then there is an inherent ambiguity in the recovery of the shape of a continuous surface. This ambiguity is called the Generalized Bas Relief (GBR) ambiguity and has been reported in the literature of *uncalibrated* photometric stereo[29, 5, 1, 7]. When the Lambertian assumption is relaxed this ambiguity is removed[4], and the stronger the component of the specular lobe the higher the conditioning and hence the faster the convergence. In the case of the globe example reconstruction in the next section, the algorithm converged within 200 iterations, while we let the algorithm run for up to 2000 iterations in the face examples. Nevertheless, we need to stress that the algorithm has no outside help, such as 3-D laser scanner or any knowledge about the light sources, and it thus tries to solve a fairly general problem with few restrictions.

5. Results

In this section we show reconstruction results for faces as well as for a part of a globe. The globe was made from a material whose reflectance was expected to be well approximated by the T-S model. The reflectance of the faces on the other hand can be more complicated. The results demonstrate that the proposed reconstruction algorithm can recover the 3-D shape and the reflectance of objects that do not obey the Lambertian reflectance assumption. In the case of faces, the recovered reflectance function closely resembles the previously measured reflectance function of human skin[16]. With the recovered 3-D shape and reflectance properties, synthetic images can be more photorealistic. Examples of image syn-

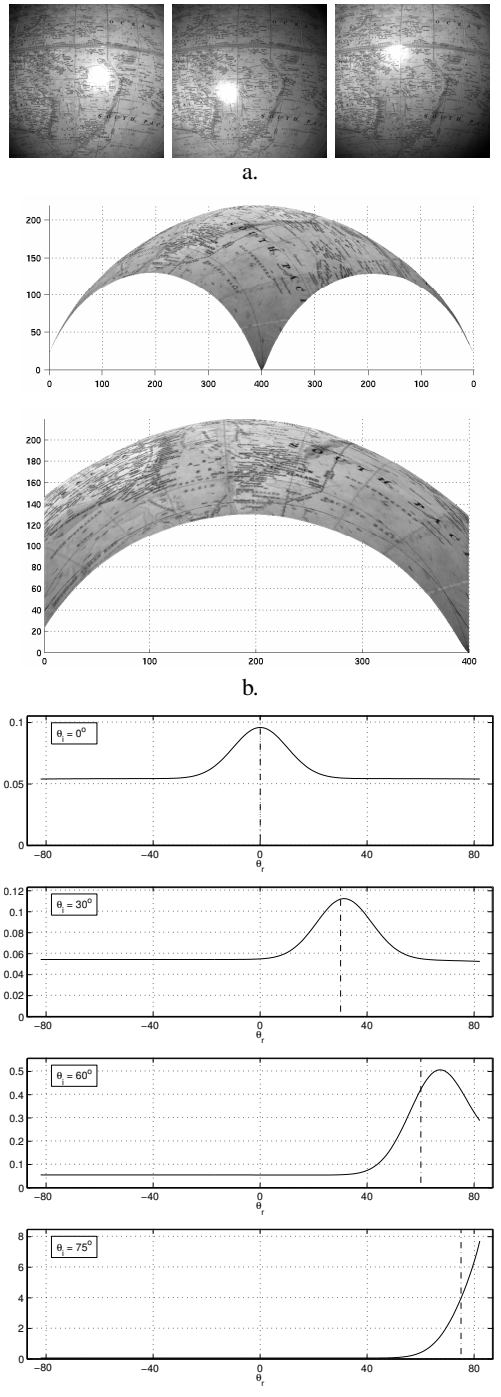

Figure 1: *The surface reconstruction of part of a globe: a. Three of the 12 input images used in the reconstructions. b. The reconstructed surface shown along the xy−direction (top) and along the y−direction (bottom). Units of the vertical axes are in pixels. c. The BRDF of the globe at different incidence angles, θ_i. The plots show the BRDF values in the incidence plane as the viewing angle θ_r varies.*

thesis and comparisons with real images can be found in the next section.

Figure 1 shows the reconstruction of part of the globe. Twelve input single light source images (of size 400×400 pixels) were used (three of which are shown in Figure 1.a), where the source directions can vary up to 24^o from the optical axis. They were selected such that the light sources were not centered around the optical axis. This helps remove any accidental symmetry that can influence the experimental results. Note also that no attempt was made to threshold out specularities in the input images (Fig. 1.a); only the saturated pixels were removed. Enough non-saturated specular pixels were included as valid measurements in the estimation process.

The reconstructed shape, shown in Fig. 1.b, is very close to a sphere, the expected shape. There is no perceived distortion due to the presence of the specularities. The specularities have been discounted by the T-S model of reflectance. Figure 1.c shows the recovered BRDF of the globe at different incidence angles, θ_i, as the exit angle θ_r varies. The parameters of the T-S model recovered during the reconstruction were: $a_d = 0.0541$ (the average Lambertian albedo over the whole surface), $a_s = 0.0415$, and $\nu = 8.1255$. (The figure shows a representative example of the estimated BRDF where the average of the estimated Lambertian albedo over the whole surface is used as the value of a_d in Equation 3. In reality a_d is a function of (x, y) and hence the BRDF is different at each position (x, y) on the surface.)

Observe that, as expected, the BRDF of the globe deviates significantly from that of a diffuse surface, as demonstrated by the pronounced specular lobe at all incidence angles. The specular lobe becomes larger and exhibits increasing off-specular behavior as the incidence angle becomes larger, a key property of the T-S model of reflectance not shared by the Phong model. This property is mostly attributed to the $1/\cos\theta_i$ and the $1/\cos\theta_r$ terms in the expression of the T-S model in Equation 3, and to a lesser extend to the Fresnel reflectivity. Note that the Fresnel reflectivity $F(\theta', \eta)$ in Equation 3, which was ignored during the reconstruction process due to its insignificant effect at low phase angles, has been re-introduced in these BRDF plots because the phase angle can be much larger than 30^o. The index of refraction, η, for a lacquered surface was assumed to be 1.5.

Figure 2 shows the reconstruction of a human face. As was the case with the globe reconstruction in Figure 1, the 12 training images of the face (six of which are shown in Figure 2.a) were selected so that they were not centered around the optical axis. This helps remove any accidental symmetry that can influence the experimental results. Figure 2.b shows the reconstructed 3-D shape and Figure 2.c (left-side) shows the BRDF of the face demonstrating the ability of our algorithm to recover both.

Significantly, incorporating the T-S model in uncalibrated photometric stereo can recover the non-Lambertian nature of skin reflectance. The left of Figure 2.c shows the skin BRDF recovered using our algorithm. Notice its similarity (up to a global scale factor) with the previously *measured* BRDF of human skin shown on the right[16]. In our method, and unlike that previous technique, both the 3-D shape *and* the BRDF of human skin were recovered simultaneously—there was no need for a 3-D scanner.

The skin BRDF is close to Lambertian at small incidence angles, but exhibits strongly increasing off-specular behavior as the incidence angle becomes larger. Notice also how the scale increases by almost 40 times from top to bottom. This almost Lambertian behavior when the incidence angle is small is what justified the use of the Lambertian model by researchers in the past for reconstructing human faces when the light sources are close to the camera. In our algorithm, the deviations from the Lambertian assumption provided enough information to recover the parameters of the T-S model of reflectance. With these, the skin reflectance can be accurately predicted for a wide variety of incidence and viewing angles significantly extrapolating from those in the training images.

As noted before, the Fresnel reflectivity was ignored during the reconstruction process due to its insignificant effect at low phase angles (less than 30^o). Nevertheless, as with the globe, it was re-introduced in the BRDF plots (shown on the left-side of Figure 2.c) because the phase angle can be much larger than 30^o. The index of refraction, η, was assumed to be 1.5—the index of refraction of human skin reported in the literature is usually 1.37-1.5 [27].

Figure 3 demonstrates the ability of our method to recover the BRDF of a face with and without sweat. For this experiment, the subject's face on the right has been rinsed with water, artificially simulating sweat. As expected, the recovered BRDF of the wet skin exhibits a noticeably more pronounced specular lobe, even at low incidence angles, in contrast to the case on the left.

6. Image Synthesis

In this section, we use the recovered 3-D surface and reflectance properties of the human face shown in Figure 2 to create synthetic images of the face under novel lighting conditions, extrapolating from those in the training images. The image synthesis process is in essence the sum of a Lambertian component and a specular component as shown in Equation 4 and demonstrated in Figure 4. Note that the extent of the specular component over the face, created with only one set of estimated parameters for the whole face and shown in Figure 4, is qualitatively similar to the extent of the specular component of a face separated using polarized light[3]. This provides further support for using the T-S model for capturing the reflectance of human skin.

To be fair, we would like to point out that assuming the parameters of the specular lobe, a_s and ν, are constant over the surface has its shortcomings. Note that the Lambertian component of the T-S model in Figure 4 has specularities on the nose and in the eyes. This is because the sharper than average specularities present in the training images at those locations could not be accounted for by the broader specular lobe defined by a single set of specular parameters which average the specular sharpness across the whole surface. This

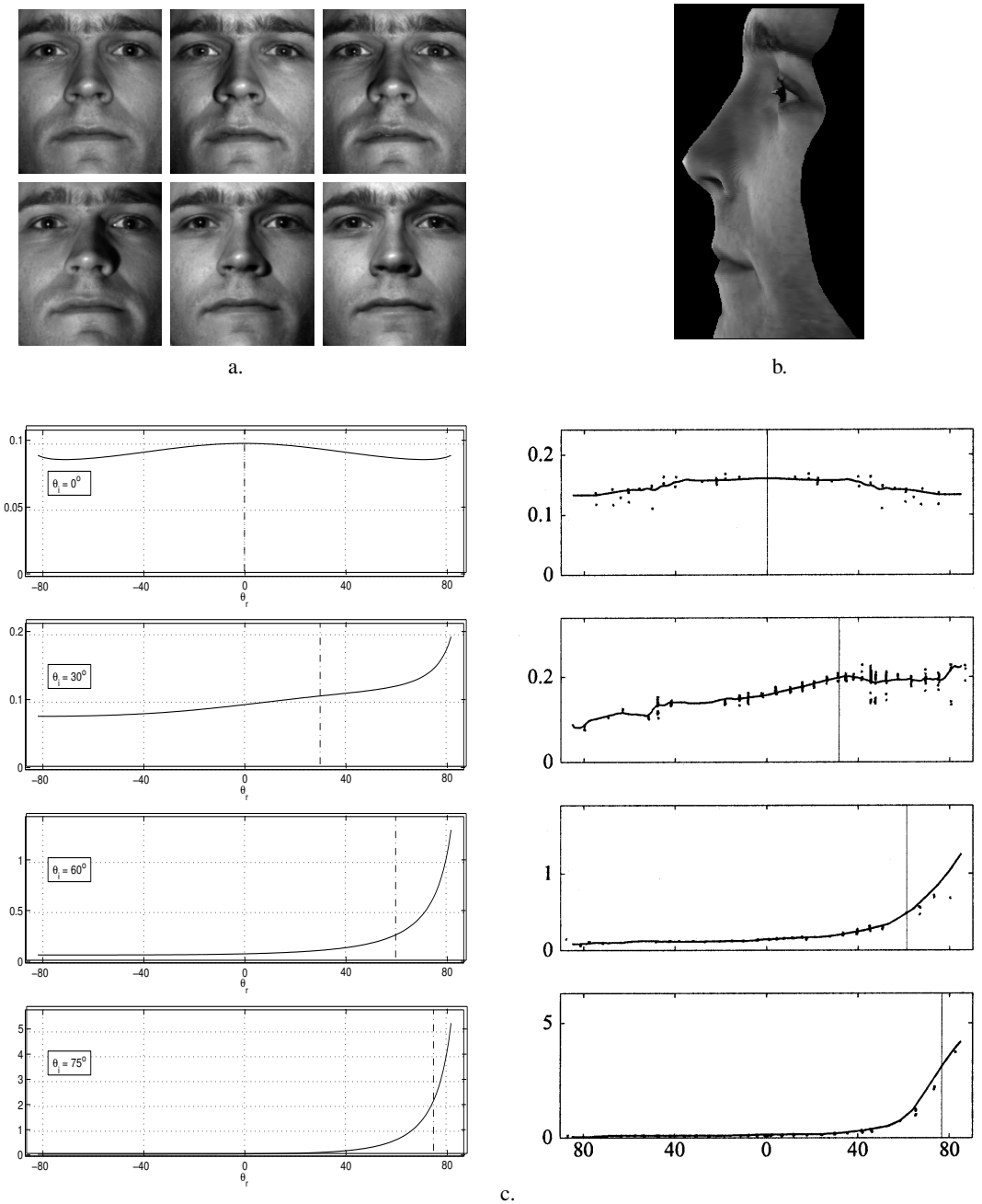

Figure 2: *The surface reconstruction of a face: a. Six out of the 12 images used in the reconstruction where the light source direction can vary up to* 24^o *from the optical axis. b. Profile view of the T-S model reconstruction. c. On the left, the estimated BRDF of the face, recovered using our method, at different incidence angles,* θ_i*. The plots show the BRDF values in the incidence plane as* θ_r *varies. The parameters of the T-S model recovered during the reconstruction were:* $a_d = 0.0776$ *(the average Lambertian albedo over the whole face),* $a_s = 0.0229$*, and* $\nu = 2.2483$*. Note that the BRDF is close to Lambertian (i.e., almost constant) at small incidence angles, but exhibits increasing off-specular behavior as the incidence angle becomes larger. Observe that the scale increases by almost 40 times from the top plot to the bottom. Furthermore, the specular lobe seems significantly removed from the perfectly specular direction, exhibiting a maximum at almost glancing* θ_r *for most of the range of* θ_i*. Notice the similarities (up to a global scale factor) of the estimated BRDF on the left with the* measured *BRDF of human skin shown on the right. (The plot on the right is courtesy of Stephen R. Marschner*[16]*.)*

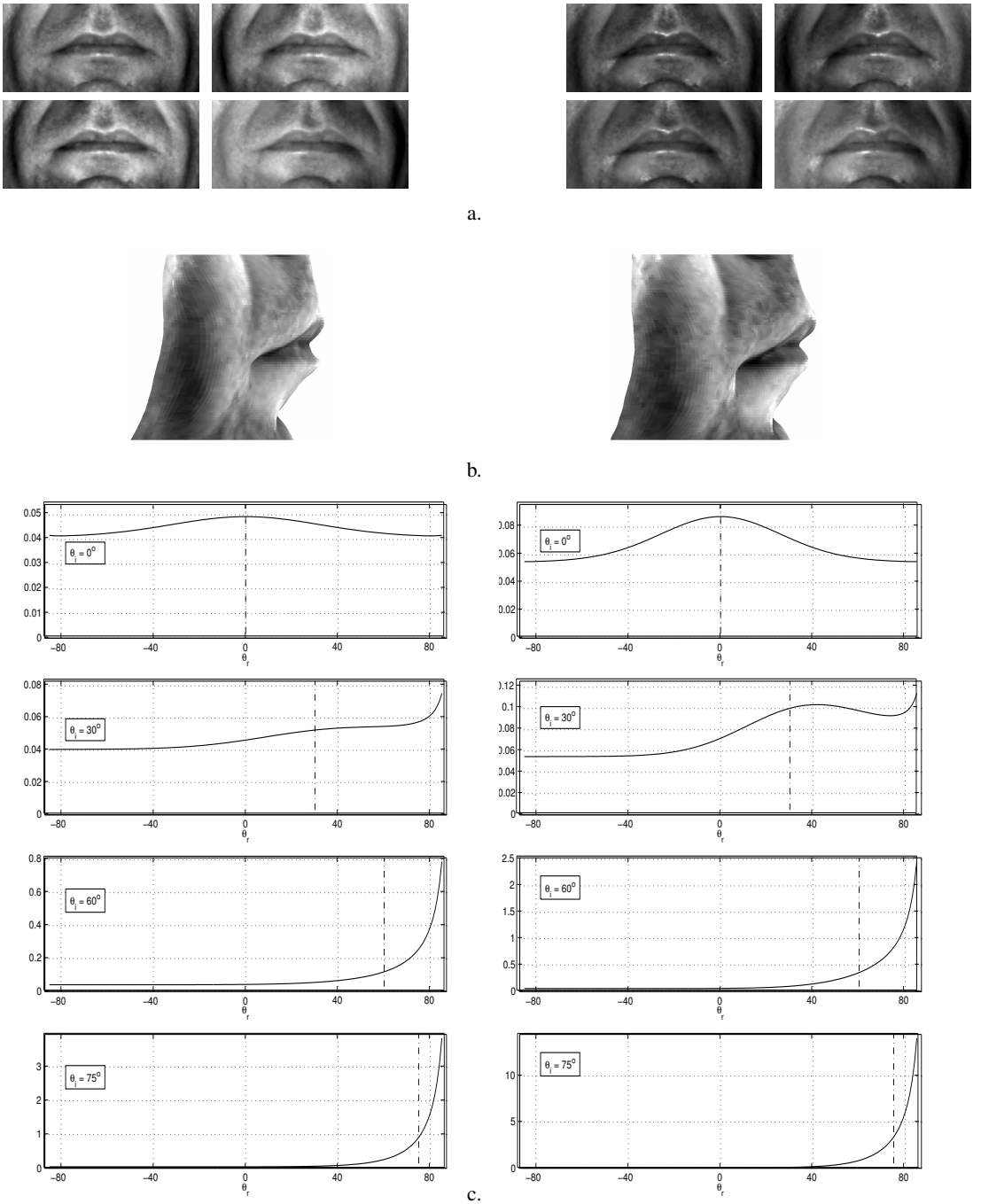

Figure 3: *Reconstructions of part of a human face using the T-S model reconstruction algorithm. The face on the right is the same as on the left, except the subject's face has been rinsed with water, simulating sweat: a. Four out of the 12 images used in each of the reconstruction where the light source direction was within* 24^o *from the optical axis. Note the higher "glossiness" in the right-side images. b. The surface reconstructions. c. The estimated BRDFs. In the case without water, on the left, the recovered parameters of the T-S model were:* $a_d = 0.0406$ *(the average albedo over the whole face region),* $a_s = 0.0087$*, and* $\nu = 2.8414$*. As in Figure 2.c, the estimated BRDF is close to Lambertian at small incidence angles, but exhibits increasing off-specular behavior as the incidence angle becomes larger. The recovered parameters for the case with water (right) were:* $a_d = 0.0548$*,* $a_s = 0.0328$*, and* $\nu = 3.3981$*. Notice how in this case the estimated BRDF exhibits a fairly large specular lobe even at low incidence angles. It still exhibits the strong off-specular behavior as the incidence angle becomes larger.*

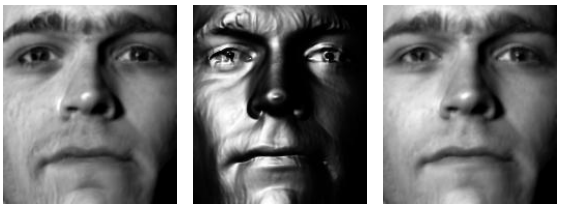

Figure 4: *Image synthesis using the 3-D model reconstruction and the estimated BRDF shown in Figure 2. LEFT: The Lambertian component of the T-S model (see Equations 3 and 4) ; MIDDLE: the specular component (Its extent over the face is qualitatively similar to the extent of the specular component of a face separated using polarized light[3].); RIGHT: the combination of the Lambertian and specular components for the completed synthesized single light source image. Creating multiple light source images is simply a superposition of single light source images.*

led to some distortion in the spatially varying Lambertian albedo a_d which appears in the Lambertian component of the T-S model in Figure 4.

Synthetic images are shown in Figure 5. In each of the four examples, we compare the image synthesized using the 3-D surface and reflectance properties recovered using the T-S model, to an image synthesized using a reconstruction of the face while only Lambertian reflectance was assumed. The images were created using an in-house rendering program to estimate both the shading and the shadows. (Note that the Lambertian rendering is a different image not associated with the Lambertian component of the T-S model. In the T-S model case the Lambertian component is added to the specular component to form the T-S model image.) Both of these synthetic images are then compared to a real single light source image. To create the synthetic images, a simulated light source was moved around so that they were as closely matched to their corresponding real image as possible. (Note that the intensities in all images are relative because of the unknown light source intensity in the real images.) As shown in Figures 5, using the recovered 3-D shape and skin BRDF of Figure 2 leads to more accurate synthetic images. This is also demonstrated in Figure 6 which shows that the T-S model intensities follow the real image intensities much more closely.

7. Conclusions

In this paper, we have described an algorithm that successfully recovers the 3-D shape and reflectance properties of a surface using only a small number of photographs. It requires no knowledge about the camera position or the light source positions and strengths, and it also obviates the use of a 3-D scanner.

An issue we have not handled is the possibility of allowing spatial variation in the parameters of the T-S model. As mentioned before, allowing the parameters to vary across the surface is an important issue. Nevertheless, it can be very difficult to reliably solve in practice because estimating the 3-D shape and reflectance parameters at every position using only image intensities from a small number of images is generally ill-conditioned. This means that some form of regularization would be required to make the problem tractable. Another important issue is the presence of a back-scatter lobe in materials such as metals. The presented reflectance model could be extended to recover the shape and the reflectance properties of such materials.

In the case of color images—not considered here—the wavelength dependence of the Fresnel reflectivity (as well as of the albedos) would need to be determined. For the Fresnel reflectivity, this is usually performed using specialized equipment that measure the normal reflectivity of a surface as a function of wavelength which can then provide an estimate of η as a function of wavelength, and consequently of the Fresnel reflectivity. As mentioned before, measuring this wavelength dependence is quite cumbersome, and, although admittedly necessary in the case of color images, it is beyond the scope of this paper.

Despite these issues, our method is able to reliably recover the shape and surface reflectance properties for a variety of non-Lambertian objects. In the case of human faces, the recovered BRDF was shown to closely resemble the measured skin reflectance reported in the literature[16]. Our method's ability to recover the skin BRDF was also corroborated by synthesizing photorealistic images of the face under novel and fairly extreme lighting.

References

1. P.N. Belhumeur, D.J. Kriegman, and A.L. Yuille. The bas-relief ambiguity. *Int. Journal of Computer Vision*, 35(1):33–44, November 1999.
2. E.N. Coleman, Jr. and R. Jain. Obtaining 3-Dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18(4):309–328, April 1982.
3. P. Debevec, T. Hawkins, C. Tchou, H. Duiker, and M. Sagar. Acquiring the reflectance field of a human face. In *Computer Graphics (SIGGRAPH)*, pages 145–156, 2000.
4. O. Drbohlav and R. Sara. Specularities reduce ambiguity of uncalibrated photometric stereo. In *Proc. European Conf. on Computer Vision*, page II: 46 ff., 2002.
5. J. Fan and L.B. Wolff. Surface curvature and shape reconstruction from unknown multiple illumination and integrability. *Computer Vision and Image Understanding*, 65(2):347–359, February 1997.
6. Robert T. Frankot and Rama Chellapa. A method for enforcing integrabilty in shape from shading algorithms. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 10(4):439–451, 1988.
7. A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, June 2001.
8. P. Hanrahan and W. Krueger. Reflection from layered surfaces due to subsurface scattering. In *Computer Graphics (SIGGRAPH)*, pages 165–174, August 1993.
9. H. von Helmholtz. *Treatise on Physiological Optics*. Dover, New York, 1925.
10. B.K.P. Horn. *Computer Vision*. MIT Press, Cambridge, Mass., 1986.

Figure 5: *Each part, from a. through d., shows comparisons between real and synthetic single light-source images demonstrating the increased photorealism when using the recovered 3-D shape and skin reflectance shown in Figure 2. Each image triplet shows (from left to right): a real image; a synthetic image rendered using the T-S model reconstruction; a Lambertian rendering using only a Lambertian model reconstruction of the face. To create the synthetic images, a simulated point light source was moved around so that they were as closely matched to their corresponding real image as possible. Notice how the synthetic images using the T-S model reconstruction more accurately capture the real skin reflectance. For example in b., the lighting is quite extreme coming from behind the face. Despite this, the T-S model captures the specularity on the nose fairly well, unlike the Lambertian case. This is also demonstrated by the intensity profiles shown in Figure 6.*

11. K. Ikeuchi and K. Sato. Determining reflectance properties of an object using range and brightness images. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 13(11):1139–1153, November 1991.

12. H.W. Jensen, S.R. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Computer Graphics (SIGGRAPH)*, pages 511–518, August 2001.

13. J.H. Lambert. *Photometria Sive de Mensura et Gradibus Luminus, Colorum et Umbrae*. Eberhard Klett, 1760.

14. S. Lin and S. Lee. Estimation of diffuse and specular appearance. In *Int. Conf. on Comp. Vis.*, pages 855–860, 1999.

15. R. Lu, J.J. Koenderink, and A.M.L. Kappers. Optical properties (bidirectional reflection distribution functions) of velvet. *Applied Optics*, 37(25):5974–5984, 1998.

16. S.R. Marschner, S.H. Westin, E.P.F Lafortune, K.E. Torrance, and D.P. Greenberg. Image-based BRDF measurement including human skin. In *10th Eurographics Workshop on Rendering*, pages 139–152, Grenada, Spain, June 1999.

17. S.K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Journal of Robotics and Automation*, 6(4):418–431, August 1990.

18. F.E. Nicodemus, J.C. Richmond, J.J. Hsia, I.W. Ginsberg, and T. Limberis. Geometric considerations and nomenclature for reflectance. Monograph 161, National Bureau of Standards (US), October 1977.

19. M. Oren and S.K. Nayar. Generalization of the Lambertian model and implications for machine vision. *Int. J. Computer Vision*, 14:227–251, 1996.

20. B.T. Phong. Illumination for computer generated images. *Communications of the ACM*, 18(6):311–317, June 1975.

21. Y. Sato, M.D. Wheeler, and K. Ikeuchi. Object shape and reflectance modeling from observation. In *Computer Graphics (SIGGRAPH)*, pages 379–388, August 1997.

22. R. Siegel and J.R. Howell. *Thermal Radiation Heat Transfer*. McGraw-Hill, New York, 1972.

23. W.M. Silver. *Determining Shape and Reflectance Using Multiple Images*. PhD thesis, MIT, Cambridge, MA, 1980.

24. F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 18(4):449–454, April 1996.

25. H.D. Tagare and R.J.P. de Figueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 13(2):133–152, February 1991.

26. K.E. Torrance and E.M. Sparrow. Theory for off-specular reflection from roughened surfaces. *J. Opt. Soc. Am.*, 57:1105–1114, 1967.

27. M.J.C. van Gemert, S.L. Jacques, H.J.C.M. Sterenborg, and W.M. Star. Skin optics. *IEEE Trans. on Biomedical Engineering*, 36(12):1146–1154, December 1989.

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

29. A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 158–164, 1997.

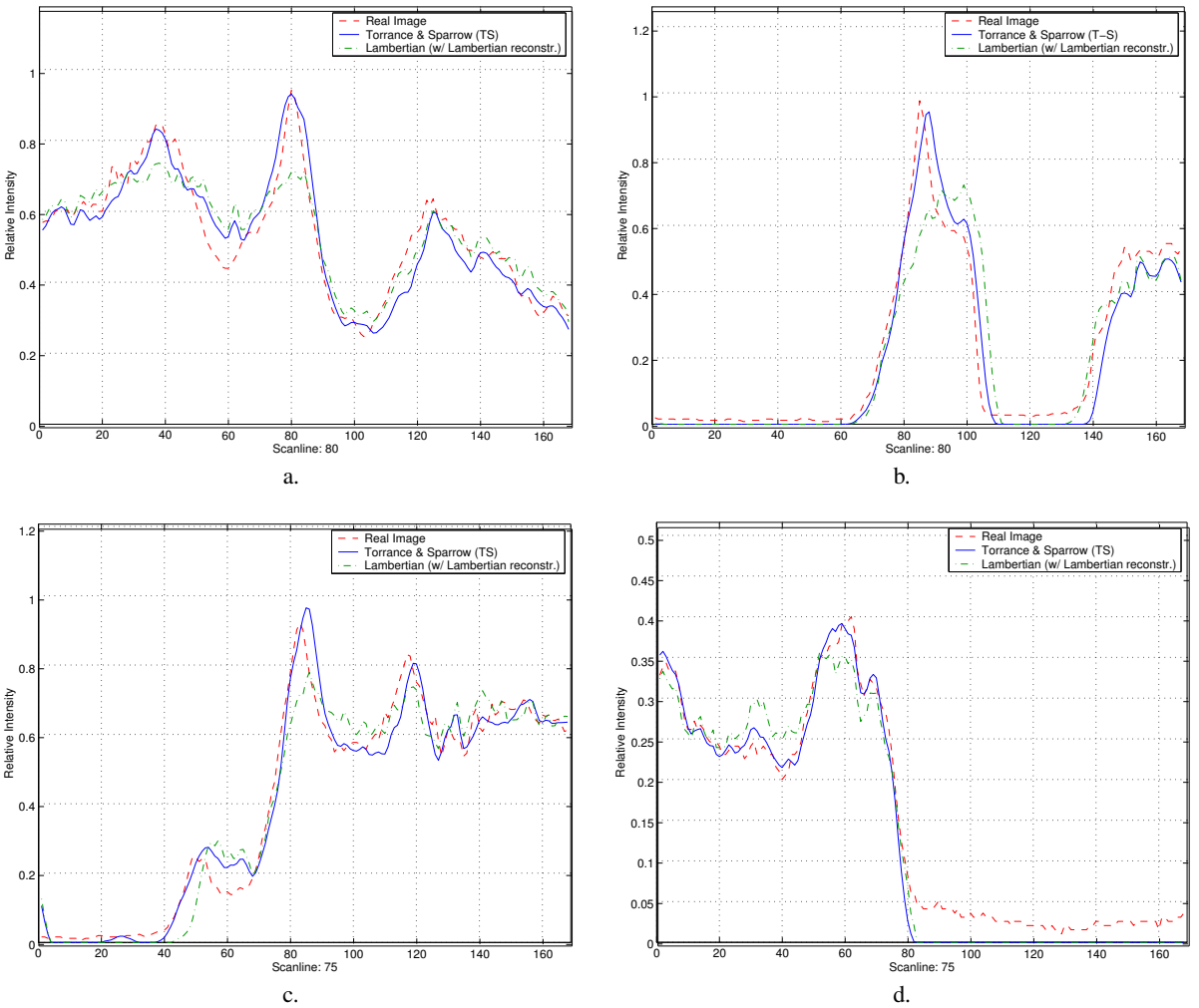

Figure 6: *Each plot, from a. through d., shows the intensity profiles from the three images shown in the corresponding part of Figure 5. The profiles are along a horizontal scan-line traversing the images approximately half-way down the nose ridge. Note how in all cases the T-S model intensity profiles more closely match the real intensity profiles. In d., where the light originates from the lower-left direction, there appears to be very little specular component in the real image. (This was evident from the small specular component in the synthetic image using the T-S model.) Nevertheless, the T-S profile follows closely the intensity profile of the real image, while the Lambertian profile (using the Lambertian model reconstruction) deviates from the real intensity both on the nose and on the left cheek. (Note that since the intensities are relative, the Lambertian profile could have been scaled so as to coincide with the intensity of the real image around the nose region, but this would have widened the gap around the left cheek even further.) The deviation in the Lambertian intensity is most likely attributed to the bias in the Lambertian reconstruction that tried to account for the presence of some specular component in the training images. (Note that on the right side of the face, the intensity profile of the real image deviates from that in the synthetic images because of the existence of ambient lighting, which we have not included in the synthetic images.)*