Shape Estimation in Natural Illumination

Micah K. Johnson Edward H. Adelson
Massachusetts Institute of Technology
{kimo, adelson}@csail.mit.edu

Abstract

The traditional shape-from-shading problem, with a single light source and Lambertian reflectance, is challenging since the constraints implied by the illumination are not sufficient to specify local orientation. Photometric stereo algorithms, a variant of shape-from-shading, simplify the problem by controlling the illumination to obtain additional constraints. In this paper, we demonstrate that many natural lighting environments already have sufficient variability to constrain local shape. We describe a novel optimization scheme that exploits this variability to estimate surface normals from a single image of a diffuse object in natural illumination. We demonstrate the effectiveness of our method on both simulated and real images.

1. Introduction

The problem of estimating shape from shading has a long history in computer vision. While there are many techniques, most seem to follow one of two basic approaches: classical shape-from-shading (SFS) or photometric stereo. Classical SFS, first formalized by Horn [10], typically assumes a single image with known illumination and reflectance conditions (often point light sources and Lambertian reflectance). Photometric stereo, first described by Woodham [22], uses multiple images with controlled illumination. Both approaches have a long lineage of publications that relax or change the basic assumptions, but a fundamental distinction is the number of lighting conditions: one for SFS and multiple for photometric stereo.

While progress is made every year on SFS, even with strong assumptions on the imaging conditions, the problem is notoriously difficult to solve [24]. For example, consider the shape reconstructions shown in Fig. 1(a). These were computed by two algorithms from a recent survey paper [6] using the image of the Mozart bust as input (upper left of Fig. 1(a)). Shape-from-shading algorithms typically perform well on simple inputs but have difficulty on more complex inputs. The root of the difficulty is local ambiguity—the fact that multiple surface orientations can lead to the same observed intensity. This ambiguity and its effect on local shape representation has been studied in both human and computer vision [4, 11, 14].

Woodham observed that the ambiguity in determining local surface orientation from intensity measurements is removed by varying the direction of illumination between successive images [22]. This technique is called photometric stereo. With three images, the problem of estimating orientations from intensities, assuming constant albedo, is simplified to the point that the mapping can be stored in a lookup table [23]. As an example, in Fig. 1(b) we show a simulated three-color photometric stereo rendering of the Mozart bust along with renderings of the estimated surface, which is indistinguishable from the ground truth.

The assumptions of classical SFS (i.e., distant point source and Lambertian reflectance) are imposed in order to make the problem mathematically tractable. However, we argue that these assumptions actually complicate the problem and that the inherent complexity of natural illumination is beneficial for shape estimation. In effect, the color variation in natural illumination is a form of photometric stereo. We exploit this property with a novel optimization scheme that can estimate surface normals from a single image.

Our technique assumes a known reflectance map, which implies in practice that we must calibrate against a sphere with the same BRDF and in the same illumination as the object of interest. This is restrictive, but it is less restrictive than the assumptions of many SFS and PS algorithms. In all three techniques (ours, SFS, and PS), the BRDF is assumed to be known, either by assumption or by measurement. In SFS, the lighting is typically assumed to be simple (e.g., point source) and from a known direction. In PS, the lighting is designed and controlled. Our technique is the only approach that uses natural illumination that is both complex and uncontrolled. Natural illumination provides a new constraint for SFS and we demonstrate the benefit of this constraint on both synthetic and real images.

2. Related work

We briefly review recent and related work on shape-from-shading and photometric stereo. For a broad overview

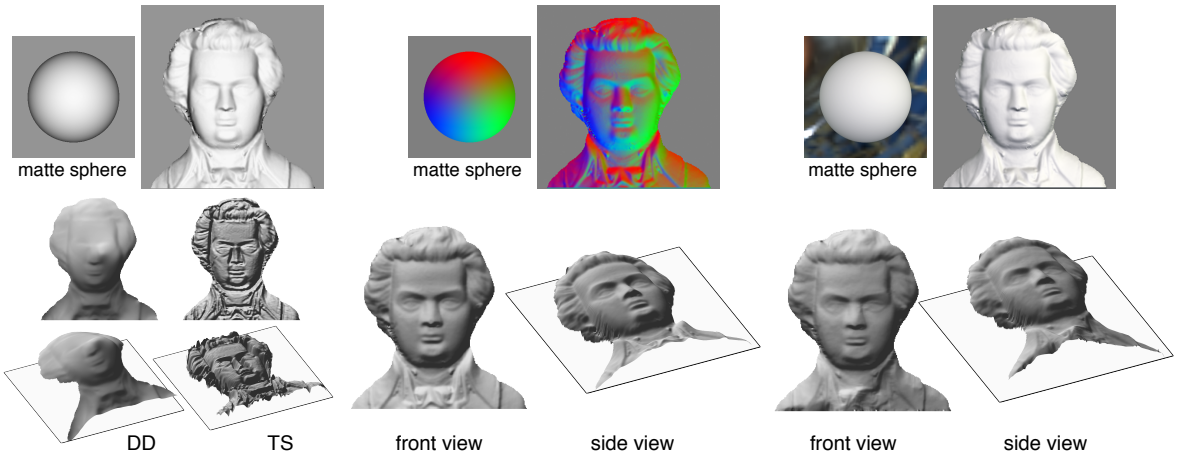


Figure 1. Three approaches to shape reconstruction from shaded images. Along the top row, we show the Mozart bust and a diffuse sphere in three illumination conditions: (a) a point light source, (b) RGB photometric stereo (three point lights), and (c) natural illumination (rendered using an environment map). Along the bottom row, we show results of shape reconstruction algorithms: (a) standard shape-from-shading algorithms by Daniel and Durou [6] (DD) and Tsai and Shah (TS) [21], (b) photometric stereo [22], and (c) our algorithm that assumes natural illumination.

of problems that have been explored, we refer the reader to surveys and recent journal papers [2, 6, 17, 24].

Work on shape-from-shading has focused on a variety of different themes over the last forty years. There have been many iterative techniques and numerical schemes for solving partial differential equations [12], theoretical analyses of ambiguity of solutions under various assumptions [14], and variations on the imaging and reflectance models, such as perspective projection, non-Lambertian reflectance, and local illumination [13, 17].

Of the works that consider the question of uniqueness and ambiguity, most have considered restricted versions of the problem, such as illumination from the camera direction [14], or images of simple shapes [11]. In general, uniqueness is rare in shape-from-shading and if both lighting and albedo are unknown, a family of surfaces exist that can generate the same image [4]. Although uniqueness is rare, we demonstrate empirically that the reduction in ambiguity due to natural illumination is sufficient to find convincing estimates of shape from a single image.

Our approach to shape-from-shading exploits lighting variability to reduce ambiguity in surface orientation, an idea closely related to photometric stereo [22]. Our method is not meant to compete with photometric stereo, however, since PS methods are active. They typically use controlled illumination, while our method assumes uncontrolled natural illumination.

Although most photometric stereo techniques assume controlled lighting, Basri et al. explored photometric stereo of Lambertian surfaces in arbitrary lighting and were able to reconstruct surfaces using 32 to 64 images with unknown lighting [2]. We employ a similar mathematical framework and show that we can estimate shape in an uncontrolled, but calibrated, lighting environment from a single image.

3. Methods

We model the relationship between the observed image intensity and surface orientation through the brightness equation, first proposed by Horn [10]:

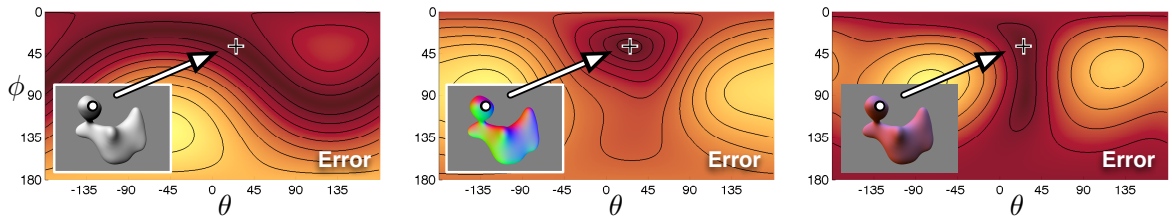
$$I(\mathbf{x}) = s(\mathbf{n}(\mathbf{x})) \,. \tag{1}$$

The observed intensity at pixel $\mathbf{x}$ is the result of a shading function s (or reflectance map) applied to the surface normal $\mathbf{n}$ at pixel $\mathbf{x}$. There are several assumptions in this model: distant lighting, spatially invariant reflectance, constant albedo, no local illumination effects such as cast shadows or interreflections, and a fixed viewpoint.

While lighting can be arbitrarily complex, the appearance of a diffuse object can be described by a low-dimensional model [3, 19]. Informally, the Lambertian reflectance function acts as a low-pass filter on the lighting environment, thus only low-frequency lighting components contribute to appearance. Under these assumptions, the shading function s for Lambertian reflectance can be modeled as a quadratic function of the surface normal [18]:

$$s(\mathbf{n}) = \mathbf{n}^T A \mathbf{n} + \mathbf{b}^T \mathbf{n} + c \,. \tag{2}$$

Note that the quadratic term helps the model account for attached shadows. An example of an object rendered according to this model is shown in the inset of Fig. 2(a).



(a) Point light source (b) RGB photometric stereo (c) Natural illumination

Figure 2. Local ambiguity in estimating a surface normal from shading information. (a) Under a single point light source, the intensity at a point on a surface can be determined by a family of surface normals. The error plot shows the error of estimating the intensity from all normals on the sphere—the normals along dark band all have the same minimal error. (b) With three light sources, as in photometric stereo, there is a single location of minimal error and no ambiguity. (c) Under natural illumination, the ambiguous region is more localized than with the point light source. The problem of estimating shape in natural illumination should be simpler than in classical shape-from-shading.

3.1. Local ambiguity

Given an intensity measurement $I(\mathbf{x})$ at position $\mathbf{x}$, a SFS algorithm needs to recover the surface normal (or height) at position $\mathbf{x}$. With only local information, the problem can be modeled as the minimization of an error function:

$$E(\mathbf{n}) = \|f(\mathbf{n})\|^2 = \|s(\mathbf{n}) - I(\mathbf{x})\|^2 \,. \tag{3}$$

As a visualization, we measure the error according to Eqn. 3 at all surface normals[1] (points on a sphere) using the reflectance map for the object in Fig. 2(a). For the chosen point $\mathbf{x}$ (dot on the surface of the object), the error function shows a large (dark) region of ambiguity; such ambiguities are well-known in SFS [10].

In general, photometric stereo techniques use L lighting conditions rather than one. Mathematically, the shading function $\mathbf{s}$ can be represented as a vector-valued function of the surface normal:

$$\mathbf{s}(\mathbf{n}) = \begin{bmatrix} s_1(\mathbf{n}) \\ \vdots \\ s_L(\mathbf{n}) \end{bmatrix} = \begin{bmatrix} \mathbf{n}^T A_1 \mathbf{n} + \mathbf{b}_1^T \mathbf{n} + c_1 \\ \vdots \\ \mathbf{n}^T A_L \mathbf{n} + \mathbf{b}_L^T \mathbf{n} + c_L \end{bmatrix} . \tag{4}$$

Continuing with our example, we render the same object using simulated three-color photometric stereo, Fig. 2(b). We also show the error function, Eqn. 3, across all surface normals for matching the color vector $\mathbf{I}(\mathbf{x})$ at position $\mathbf{x}$. Note that there is now a single global minimum, removing the ambiguity seen in Fig. 2(a). This is, of course, the motivation for photometric stereo [22].

In natural lighting environments, the shading function is the same as Eqn. 4 with $L = 3$ since we assume three color channels. In Fig. 2(c), we show a Lambertian object rendered in the Grace Cathedral lighting environment. We also show the error function from the same position $\mathbf{x}$ as the previous plots. The reduction in size of the ambiguous region in Fig. 2(c) as compared to Fig. 2(a) demonstrates that the variability in this natural lighting environment provides additional constraints for shape estimation.

[1] In practice, the space of surface normals could be restricted by considering visibility from the camera.

3.2. Nonlinear optimization

Given an observed color vector, we minimize the error function, Eqn. 3, with respect to the surface normal $\mathbf{n}$. Since our shading function is quadratic in $\mathbf{n}$, this is a nonlinear least-squares problem that can be minimized with an iterative technique, such as the Gauss-Newton method. Suppose at the i-th iteration, the estimate of the surface normal that minimizes Eqn. 3 is $\mathbf{n}_i$. The Gauss-Newton method computes an update vector $\mathbf{h}$ that satisfies the following equation:

$$J(\mathbf{n}_i)^T J(\mathbf{n}_i)\mathbf{h} = -J(\mathbf{n}_i)^T \mathbf{f}(\mathbf{n}_i) \,, \tag{5}$$

where $J(\mathbf{n}_i)$ is the Jacobian matrix (i.e., the matrix of partial derivatives) of the function $\mathbf{f}$ at the current estimate $\mathbf{n}_i$:

$$J(\mathbf{n}_i) = \frac{\partial \mathbf{f}}{\partial \mathbf{n}_i} = \begin{bmatrix} 2\mathbf{n}_i^T A_1 + \mathbf{b}_1^T \\ \vdots \\ 2\mathbf{n}_i^T A_L + \mathbf{b}_L^T \end{bmatrix} . \tag{6}$$

The update vector $\mathbf{h}$ satisfying Eqn. 5 is added to the current estimate: $\mathbf{n}_{i+1} = \mathbf{n}_i + \mathbf{h}$.

However, there are two problems with using a standard Gauss-Newton iteration to minimize Eqn. 3. While the surface normals $\mathbf{n}$ are being represented in R^3, they are actually unit vectors that are constrained to the surface of the sphere. The update vector $\mathbf{h}$ may move the current estimate away from the surface of the sphere and this deviation will need to be corrected before the next iteration. This process increases the number of iterations until convergence.

We solve this problem by defining a local frame around the initial surface normal estimate $\mathbf{n}_0$. The frame is parameterized by coordinates u and v, such that surface normals near $\mathbf{n}_0$ are defined by the following function:

$$\mathbf{n}(u, v) = R_0 \begin{bmatrix} u & v & r \end{bmatrix}^T \,, \tag{7}$$

where $r = \sqrt{1-u^2-v^2}$ and R_0 is a rotation matrix such that R_0^{-1} maps the initial estimate $\mathbf{n}_0$ to $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$.

Within the local frame, the shading function in Eqn. 2 can be expressed in terms of coordinates u and v through Eqn. 7. The Jacobian with respect to coordinates u and v is obtained by the chain rule:

$$\begin{bmatrix} \frac{\partial J}{\partial u} & \frac{\partial J}{\partial v} \end{bmatrix} = \frac{\partial \mathbf{f}}{\partial \mathbf{n}} R_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -u/r & -v/r \end{bmatrix} . \quad (8)$$

Note that the particular coordinate frame is only defined for the hemisphere about the original surface normal estimate $\mathbf{n}_0$ and that the Jacobian will be undefined if $u^2+v^2 > 1$. Therefore, we reset the frame according to the current surface normal estimate once $u^2 + v^2 \geq 1/2$.

Additionally, the Jacobian matrix J becomes ill-conditioned as the lighting variation between the color channels decreases. As a result, the Gauss-Newton iteration converges slowly or may even fail to converge. To address this problem, we use a variant of the Gauss-Newton scheme known as Powell's dog-leg [16]. This method is less sensitive to the conditioning of the Jacobian matrix J.

3.3. Local refinement

Due to image noise and local ambiguity, a surface normal cannot be reliably estimated from a single pixel. In this section, we describe a novel optimization scheme that combines local refinement via nonlinear least-squares with a multi-scale propagation technique.

The local refinement stage is a modification of the iterative technique described in the previous section. Rather than optimize a single surface normal, we optimize a patch of k adjacent surface normals, $\mathbf{n}_1$ to $\mathbf{n}_k$. The modified error function is:

$$E(\mathbf{n}_1, \ldots, \mathbf{n}_k) = \|\mathbf{g}(\mathbf{n}_1, \ldots, \mathbf{n}_k)\|^2, \quad (9)$$

where:

$$\mathbf{g}(\mathbf{n}_1, \ldots, \mathbf{n}_k) = \begin{bmatrix} \mathbf{f}(\mathbf{n}_1) \\ \vdots \\ \mathbf{f}(\mathbf{n}_k) \\ \lambda_1 \mathbf{c}_1(\mathbf{n}_1, \ldots, \mathbf{n}_k) \\ \lambda_2 \mathbf{c}_2(\mathbf{n}_1, \ldots, \mathbf{n}_k) \end{bmatrix} . \quad (10)$$

The functions $\mathbf{c}_1$ and $\mathbf{c}_2$ are two different constraints on the surface normals: integrability and smoothness.

The integrability constraint reflects the fact that the surface normals are not an arbitrary vector field—they are local orientations of an unknown surface. We enforce integrability through a penalty on the curl within the patch. In the local frame, at patch coordinate (i, j), the curl can be approximated as $c_y - c_x$ with:

$$\begin{aligned} c_y &= r_{11}\,(u_{i+1,j} - u_{i,j}) + r_{12}\,(v_{i+1,j} - v_{i,j}) \ , \\ c_x &= r_{21}\,(u_{i,j+1} - u_{i,j}) + r_{22}\,(v_{i,j+1} - v_{i,j}) \quad (11) \end{aligned}$$

where r_{ij} is the entry of the rotation matrix, R_0 at row i and column j. The full integrability constraint $\mathbf{c}_1$ is a vector-valued function with the curl for each surface normal, $\mathbf{n}_1$ to $\mathbf{n}_k$, in separate rows.

The smoothness constraint is derived from a generic viewpoint principle [8]: when no change is observed across a region in the image, we assume the underlying surface does not change. In other words, we consider it unlikely for the surface and illumination to change in opposite ways such that no intensity variation is visible. Therefore, our smoothness constraint penalizes surface variation along contours of minimal image change (i.e., along isophotes). Constraints implied by isophotes, and how they influence human perception of shape, have been explored by other authors [7]. Here we show how isophotes can be used as constraints within our optimization.

In Sec. 3.5 we describe a simple method for computing isophotes on shaded images of objects with constant albedo. Assume that method gives θ as the local orientation of shading for this patch. We constrain the surface variation by applying a second-derivative of Gaussian filter, G^2_θ, oriented in direction θ. This filter is defined by a linear combination of three base filters, G_{2a}, G_{2b} and G_{2c}. Due to space considerations, we omit the formulas, which can be found in Table III of [9].

The second-derivative of Gaussian filter responds more strongly to variation along the direction θ than along the orthogonal direction. We use this property as a constraint across the patch:

$$\mathbf{c}_2 = \begin{bmatrix} \sum_{i,j} G^2_\theta(i,j)\, u(i,j) \\ \sum_{i,j} G^2_\theta(i,j)\, v(i,j) \end{bmatrix} . \quad (12)$$

The constraints $\mathbf{c}_1$ and $\mathbf{c}_2$ are weighted by scalars λ_1 and λ_2 in Eqn. 10. We find that convergence is robust across the range 0.01 to 0.5 for both parameters.

The local refinement stage is run for a small number of iterations, typically 5 to 10, for each patch. In the next section, we describe how the local estimates from each patch are propagated to adjacent patches and then across scales.

3.4. Multi-scale propagation

The propagation stage uses surface normal estimates from neighboring patches to provide initial estimates for the current patch. Our approach borrows ideas from patch-based image processing (e.g., [1]) and applies them within a continuous optimization framework.

We build a pyramid from the input image and begin processing at the lowest scale, from the upper left pixel to the lower right. For each pixel, we consider the surrounding patch of surface normals and perform local refinement on the patch, as described in Sec. 3.3. Next, we consider the patch centered one pixel to the left and use it as an initial

condition for the refinement at the current location. We repeat the refinement a third time starting with the patch centered one pixel above the current location. At this point we have three refined patches that explain the image data. To choose the best patch, we consider a region around the patch, which we call the local context.

The local context P is a larger region that encloses the current patch. We define the context error as:

$$T(P) = E(P) + \omega_1 E_1(P) + \omega_2 E_2(P) \,, \qquad (13)$$

where E is the shading error, Eqn. 9, and E_1 and E_2 are the integrability and curl constraints evaluated over the local context. The scalars ω_1 and ω_2 weigh the contributions of the constraints to the error function.

Discretized surface normals often have small but non-zero curl and imposing a penalty on these values can prevent the normals from attaining the correct shape. Instead, we put the curl constraint within a smooth step function, a sigmoid, that only penalizes large values. Thus, the integrability term E_1 of the context error is:

$$E_1(P) = \sum_{p,q \in P} \frac{1}{1 + \exp\left(-\alpha(|p_y - q_x| - \tau)\right)} \,, \qquad (14)$$

where the parameter α controls the rise of the sigmoid and τ controls the location where the sigmoid crosses 0.5. The variables p and q are the estimates of surface gradient in the region, $p = \mathbf{n}_x/\mathbf{n}_z$ and $q = \mathbf{n}_y/\mathbf{n}_z$, and p_y and q_x denote the partial derivatives of these values. We approximate the partial derivatives with forward differences.

The smoothness constraint uses a second-derivative of Gaussian filter that is sized to cover the local context. We choose the L_1 norm to be more robust to large values, since the context error does not need to be differentiable:

$$E_2(P) = \sum_k \left| \sum_{i,j} G^2_\theta(i,j) \mathbf{n}_k(i,j) \right| \,, \qquad (15)$$

where $\mathbf{n}_k$ is an individual component of the surface normal.

The patch that minimizes the context error, Eqn. 13 is selected as the local estimate and then the next patch is processed. On odd iterations, we traverse the image backwards, similar to [1].

At the end of a fixed number of iterations, we upsample the current estimate and process the next scale. The algorithm continues until the finest level of the pyramid has been processed. Since we are processing within an image pyramid, we can also take advantage of multigrid techniques to speed up convergence [5]. We find that the w-cycle, i.e., coarse-to-fine passes beginning at every scale, is particularly effective.

Our initial condition for the iteration is a random array of unit-length vectors, with the z-component fixed to be positive. This initial condition will have very large values for the two constraints E_1 and E_2 and the optimization can get quickly stuck in a local minimum if the parameters ω_1 and ω_2 are too large. But in cases of high image noise, we want the parameters to be large to prevent unwanted surface variation. We have found that increasing ω_1 and ω_2 each iteration, from 0 to their specified values effectively solves both problems.

3.5. Local orientation

We measure the local orientation of intensity variation using the structure tensor. The structure tensor can be computed from three component images:

$$G_x = g_\sigma \star I_x^2 \,, \;\; G_{xy} = g_\sigma \star (I_x I_y) \,, \;\; G_y = g_\sigma \star I_y^2 \qquad (16)$$

where I_x and I_y are the x and y gradients of the image I, g_σ is a Gaussian kernel with standard deviation σ and the symbol '$\star$' denotes convolution.

Let $\hat{C} = (G_x - G_y)/(G_x + G_y)$ and $\hat{S} = (2G_{xy})(G_x + G_y)$. The local orientation is $\theta = \frac{1}{2}\tan^{-1}(\hat{S}/\hat{C})$.

Image gradients at the finest scale are sensitive to noise and therefore the local orientation can be difficult to estimate, especially in flat regions. To improve the estimate, we blend the component images, Eqn. 16, with upsampled components from the previous scale, $G'_i = G_i + G_{i-1}$. By this process, the finest scale obtains pooled estimates from all the previous scales.

4. Results

We evaluate our algorithm on both synthetic and real images. The parameters of the algorithm were kept constant across all experiments. We used 3×3 patches in a 5×5 context, 3 pyramid levels, 5 iterations per level, local refinement weights $\lambda_1 = 1.0$ and $\lambda_2 = 0.01$, context error weights $\omega_1 = 1.0$ and $\omega_2 = 0.1$.

Our optimization algorithm assumes that a model of the lighting environment is known. For both the real and synthetic images, we fit the model using a diffuse calibration sphere. Since the shading function, Eqn. 2, is linear in the lighting environment coefficients, we use the known surface normals of the sphere to solve for the coefficients using least-squares.

4.1. Synthetic images

To help us develop and test our algorithm, we found it useful to work with a set of synthetic images. Our test set consists of 100 images that are generated by rendering 10 shapes in 10 different lighting environments using a physically-based renderer, pbrt [15]. The 10 shapes are shown across the top row of Fig. 3. Our shapes have varying

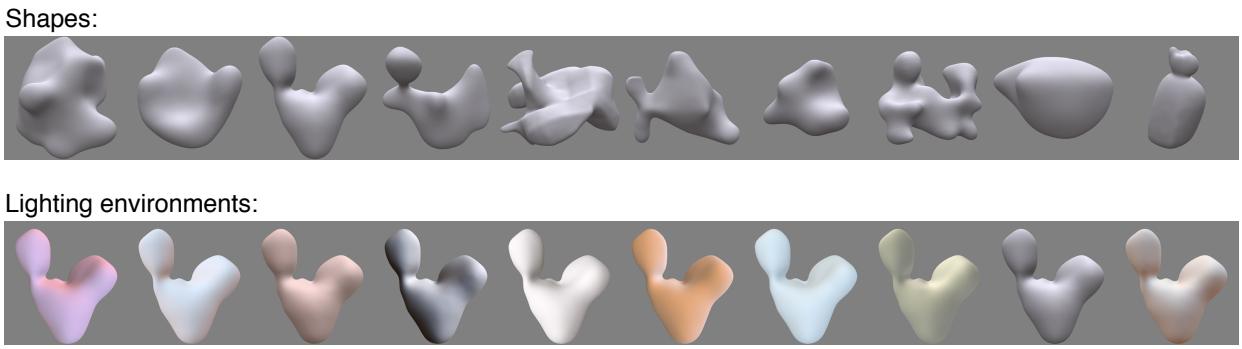


Figure 3. Our test set consists of 100 images, 10 shapes rendered in each of 10 lighting environments. Top: The 10 random shapes rendered in the same lighting environment. Bottom: One shape rendered in each of the 10 lighting environments.

levels of complexity, including large smooth regions, concave regions, self-occlusions, large depth discontinuities, creases and protrusions. To encourage future work on this problem, we have made our dataset available online.[2]

The 10 lighting environments are represented in the bottom of Fig. 3, by rendering one of the shapes in each lighting environment. We used the following lighting environments: Grace Cathedral, Eucalyptus Grove, Uffizi Gallery, Galileo's Tomb, Ennis-Brown Dining Room, Pisa Courtyard, Doge's Palace, Inside Tunnel Machine, At the Window, and Distant Evening Sun. The light probes used for rendering are available online.[3]

To ensure that the rendering model conformed exactly to our assumptions (i.e., no cast shadows or interreflections), we rendered calibration spheres and fit the lighting environment model to the spheres. Before optimization we add Gaussian random noise with standard deviation 0.001 to avoid the existence of an exact numerical solution. This amount of noise for our synthetic images (at 256 pixels across) is comparable to 2% noise in our real images, before downsampling.

For each of the 100 rendered images, we have image masks and ground-truth surface normals (rotated into the camera coordinate system). The image masks identify the background pixels, which are not considered in the optimization or evaluation. To evaluate performance, we compute the angular error between the ground-truth surface normal and the estimated surface normal. Across all surface normals from all 100 images, 90% have an angular error lower than 10 degrees.

To enable comparisons with previous work, we tested our algorithm on two standard surfaces: the Mozart bust and the analytic "vase" [24]. We rendered these surfaces in all 10 lighting environments at an image size of 256×256 pixels. Across all 10 Mozart images, 86% of the surface normals have an angular error lower than 10 degrees. For the vase images, 94% of the normals are within 10 degrees from ground truth.

We also reconstructed depth by integrating the surface normals. We used the L_1 Poisson approach, via iteratively reweighted least squares to minimize the influence of noisy estimates [20]. Since we use Neumann boundary constraints (i.e., we do not specify any known depth values), the reconstructed surface will have an overall depth ambiguity. We resolve the ambiguity by computing an offset δ that best aligns the depth estimate to the ground-truth:

$$A(\delta) = \sum_{\mathbf{x} \in \Omega} \left| z(\mathbf{x}) + \delta - \hat{z}(\mathbf{x}) \right| , \tag{17}$$

where Ω is a mask identifying foreground pixels, $z(\mathbf{x})$ is the estimated depth and $\hat{z}(\mathbf{x})$ is the ground-truth depth at position $\mathbf{x}$.

In Fig. 4 and Fig. 5, we show reconstructions of both objects under two different illumination conditions, Distant Evening Sun and Inside Tunnel Machine. For the vase images, the RMS errors are 0.55 and 0.62 pixels for upper and lower results. For the Mozart images, the RMS errors are 2.7 and 1.1 pixels for the upper and lower results.

4.2. Real images

To evaluate the performance of our algorithm in a more realistic setting, we captured images of objects in natural lighting environments. Since our algorithm does not account for reflectance variation, the objects and calibration target were painted with a diffuse paint. We used an 18-megapixel Canon EOS 550D camera equipped with a 100-mm lens. The camera was mounted on a tripod and set to capture in RAW mode. The illumination of the scene was not modified in any way (i.e., no additional lights). The RAW images were converted to a 16-bit format using Adobe Photoshop. For each image, we also created binary image masks by tracing the boundary of the objects in the image.

[2] http://people.csail.mit.edu/kimo/blobs

[3] http://ict.debevec.org/~debevec/Probes and http://dativ.at/lightprobes

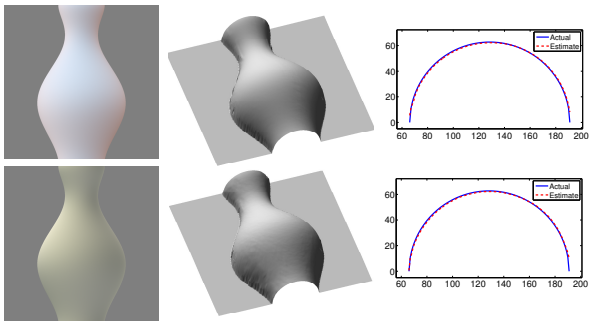


Figure 4. (a) Two renderings of the analytic vase, [24]. (b) Renderings of the reconstructed surface. The RMS error on the depth estimate is 0.55 pixels for the upper result and 0.62 pixels for the lower result. (c) 1D profiles of the center scanline of each estimated depth map, compared to the ground truth.

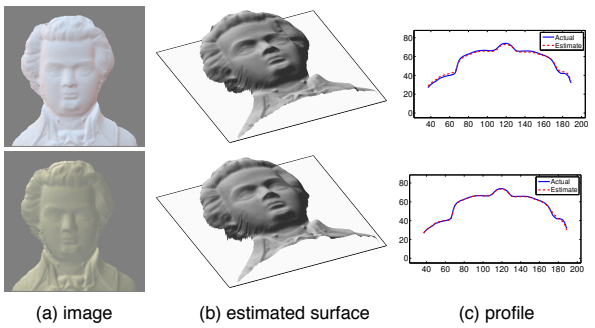


Figure 5. (a) Two renderings of the Mozart bust. (b) Renderings of the estimated surface. The RMS error is 2.7 pixels for the upper result and 1.1 pixels for the lower result. (c) 1D profiles of the center scanline of each estimated depth map, compared to the ground truth.

In Fig. 7, we show the result of our optimization on three objects in the same scene. The objects were photographed in an office with natural light from a window and fluorescent lighting overhead. The inset plots for the real images show the calibration target and the inset plots in the normal maps show a visualization of surface normals on the sphere. In Fig. 6, we show a close-up view of the y-component of the surface normal for each object in Fig. 7. Our optimization achieves a high level of detail from a single image.

Fig. 8 shows the images of the calibration target and a toy frog. The objects were photographed in a hallway with recessed lighting overhead and painted walls. We fit a model of the lighting environment to the calibration target then ran our algorithm to estimate the normal map. We integrated the normals using our L_1 Poisson solver to obtain a depth estimate. To improve the visualization we clipped depth values below the 1st and above 99th percentiles. We render the estimated surface from two different viewpoints.

We find that our optimization algorithm performs well when the lighting environment is sufficiently rich to constrain the solution space, but can fail under modest amounts of image noise if the lighting is similar to a single light source. Under this lighting condition, the Jacobian of the error function, Eqn. 6, will be rank deficient. The optimization will still produce a surface, but it is generally flat along the isophote directions.

Figure 6. Surface normal detail. Close-up views of the y-component of the estimated surface normal for the objects in Fig 7. Our algorithm is able to reconstruct a high level of detail from a single image.

5. Conclusion

Shape estimation is difficult when the illumination consists of a single light direction. The difficulty stems from the local ambiguity between the intensity value and the range of surface orientations that could have produced that value. But under natural illumination, this ambiguity is often reduced. Based on this observation, we have described an algorithm that can estimate the surface normals of a diffuse object, with constant albedo, from a single image under uncontrolled but known illumination.

Acknowledgments

This work was supported in part by the NSF under Grant No. 0739255, NIH contract 1-R01-EY019292-01, and by a grant from the NTT-MIT Research Collaboration. The authors thank P. Debevec and B. Vogl for providing light probes for academic use.

References

[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, 28(3), 2009.

[2] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.

[3] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003.

[4] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999.

[5] W. L. Briggs, V. E. Hensen, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, 2 edition, 2000.

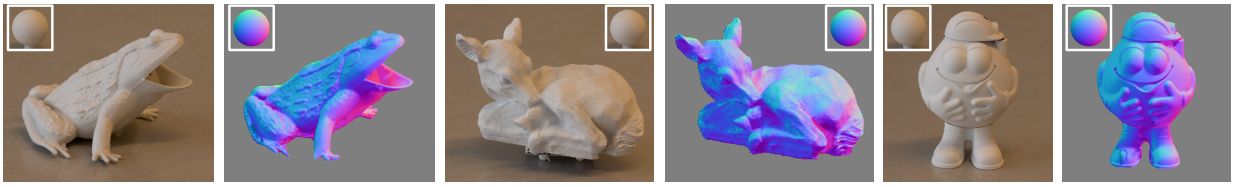


Figure 7. Surface normal estimation. Three objects and a calibration target were photographed in the same lighting environment. All objects were painted to have the same reflectance. Using the calibration target to model the lighting environment, we ran our optimization to solve for the surface normals from each image. The inset plots in the real images show the calibration sphere. The inset plots for the normal maps show a visualization of surface normals on the sphere. Close-up views of the normals for each image are shown in Fig. 6.

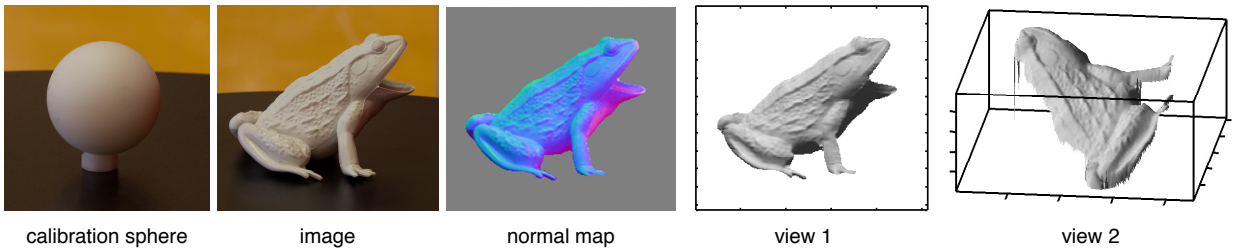


Figure 8. Results of surface reconstruction. An image of a calibration target and toy frog were captured in a natural lighting environment. A model of the lighting environment is fit to the calibration target. From the image of the frog, our algorithm estimates a normal map. We integrate the normals to obtain a depth estimate, which we render from two viewpoints.

[6] J.-D. Durou, M. Falcone, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1), 2008.

[7] R. W. Fleming, A. Torralba, and E. H. Adelson. Specular reflections and the perception of shape. *Journal of Vision*, 4(0):798–820, 2004.

[8] W. T. Freeman. The generic viewpoint assumption in a framework for visual perception. *Nature*, 368(6471):542–545, 2004.

[9] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(9):891–906, 1991.

[10] B. K. P. Horn. *Shape from shading; a method for obtaining the shape of a smooth opaque object from one view*. PhD thesis, Massachusetts Institute of Technology, 1970.

[11] R. Kozera. Uniqueness in shape from shading revisited. *Journal of Mathematical Imaging and Vision*, 7(2):123–138, 1997.

[12] K. M. Lee and C.-C. J. Kuo. Shape from shading with a linear triangular element surface model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(8):815–822, 1993.

[13] K. M. Lee and C.-C. J. Kuo. Shape from shading with a generalized reflectance map model. *Journal of Computer Vision and Image Understanding*, 67(2):143–160, 1997.

[14] J. Oliensis. Uniqueness in shape from shading. *International Journal of Computer Vision*, 6(2):75–104, 1991.

[15] M. Pharr and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann, 2 edition, 2010.

[16] M. J. D. Powell. A hybrid method for nonlinear equations. In P. Rabinowitz, editor, *Numerical Methods for Nonlinear Algebraic Equations*. Routledge, 1970.

[17] E. Prados and O. Faugeras. A generic and provably convergent shape-from-shading method for orthographic and pinhole cameras. *International Journal of Computer Vision*, 65(1):97–125, 2005.

[18] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proc. of SIGGRAPH*, pages 497–500, 2001.

[19] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object. *Journal of the Optical Society of America A*, 18:2448–2559, 2001.

[20] D. Reddy, A. Agrawal, and R. Chellappa. Enforcing integrability by error correction using l_1-minimization. In *Computer Vision and Pattern Recognition*, 2009.

[21] P.-S. Tsai and M. Shah. Shape from shading using linear approximation. *Image and Vision Computing*, 12(8):487–498, 1994.

[22] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[23] R. J. Woodham. Gradient and curvature from photometric stereo including local confidence estimation. *Journal of the Optical Society of America A*, 11:3050–3068, 1994.

[24] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(8):690–706, 1999.