Shape and Spatially-Varying BRDFs from Photometric Stereo

Dan B Goldman, Brian Curless, *Member, IEEE*,
Aaron Hertzmann, *Member*, *IEEE*, and Steven M. Seitz, *Senior Member*, *IEEE*

Abstract—This paper describes a photometric stereo method designed for surfaces with spatially-varying BRDFs, including surfaces with both varying diffuse and specular properties. Our optimization-based method builds on the observation that most objects are composed of a small number of *fundamental materials* by constraining each pixel to be representable by a combination of at most two such materials. This approach recovers not only the shape but also material BRDFs and weight maps, yielding accurate rerenderings under novel lighting conditions for a wide variety of objects. We demonstrate examples of interactive editing operations made possible by our approach.

Index Terms—Shape/scene analysis, reflectance digitization and image capture.

---◆---

1 INTRODUCTION

RECONSTRUCTING scenes with realistic materials from images is a challenging, open research problem. In many real-world scenarios, both the shape *and* the materials are unknown and need to be recovered. Although the problem is very difficult in this general setting, good solutions exist when either the materials or shape are known. In particular, if the shape is known, one of several BRDF estimation methods can be applied [15], [17], [19], [23], [36] to estimate material properties. On the other hand, when the BRDF is arbitrary but known or can be measured using reference objects, *example-based* photometric stereo methods [9], [24] enable reconstruction of shape models. When both the shape and materials are unknown, the problem becomes significantly more challenging. The problem has many, many intertwined variables, resulting in a high-dimensional optimization.

In this paper, we address the problem of computing both shape and spatially-varying BRDFs of objects using a novel photometric stereo approach. We seek to achieve much of the generality of the example-based approach to photometric stereo of Hertzmann and Seitz [9], while removing the need for reference objects. Our approach also acquires the unknown target materials, thereby enabling novel editing operations, i.e., we solve for the reference objects as part of the reconstruction process. Our approach

alternates between solving for shape given a model for the BRDFs, and solving for the spatially-varying BRDFs given a shape model.

Our problem of shape and material reconstruction is highly underconstrained if the BRDF at each pixel is arbitrary. Moreover, many typical regularization approaches are not applicable. For example, a simple smoothness term would prevent reconstruction of high-frequency texture. We believe the prior assumptions made about the BRDF are at the crux of the problem, and our model is both narrow enough to eliminate many ambiguities of interpretation, and flexible enough to produce a visually plausible approximation of a wide variety of real-world objects.

Researchers have observed that many objects, whether man-made or natural, can be decomposed into a small number of *fundamental materials* (e.g., [15]). We contribute the additional observation that, even for surfaces composed of many distinct materials, most pixels can be well represented by mixtures of just one or two of those materials. Therefore, we impose the novel constraint that each pixel has a mixture of pairwise combinations of the materials. This additional constraint eliminates many sources of ambiguity that would occur with arbitrary linear combinations of multiple materials, and thereby allows us to resolve surface shape using photometric stereo.

Given these constraints, we can now solve for a set of global material parameters, per-pixel normals, and per-pixel material weights for each material. Our approach is to alternate between optimizing global parameters and optimizing per-pixel weights and normals. First, we hold per-pixel normal and material weight estimates fixed and optimize over BRDF parameters using nonlinear least-squares optimization. Then, we hold the BRDF parameters fixed, and jointly optimize normals and material weights for each pixel using a combination of brute force search and linear projection. With this scheme, we can minimize a highly nonlinear objective function with tens of thousands of unknowns.

- *D.B Goldman is with Adobe Systems, Inc., 801 N. 34th St., Seattle, WA 98103. E-mail: dgoldman@adobe.com.*
- *B. Curless and S.M. Seitz are with the Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195-2350. E-mail: {curless, seitz}@cs.washingto.edu.*
- *A. Hertzmann is with the Department of Computer Science, University of Toronto, Bahen Centre, BA5268, 40 St. George St., Toronto, ON M5S 2E4, Canada. E-mail: hertzman@dgp.toronto.edu.*

Using our reconstructions, many novel object editing techniques are made possible. For example, in addition to changes of lighting and limited changes of viewpoint, the fundamental materials provide a convenient representation for editing materials globally over a surface. The materials can be painted or adjusted using operations similar to those available in commercial image editing software. Finally, a number of surface property transfer operations are made possible by our representation.

2 RELATED WORK

In the seminal work on photometric stereo, the surface materials are assumed to be Lambertian or spatially uniform. Silver [24] used images of a wooden sphere to reconstruct other wooden objects imaged under the same lighting, based on a proposal by Woodham [32]. Nayar et al. [20] recovered normals along with the contributions of both pure specular and Lambertian diffuse reflection for each pixel, using extended area light sources.

More recent work has begun to extend the range of this approach to surfaces with more complex, spatially-varying BRDFs. The example-based method of Hertzmann and Seitz [9] demonstrated that reference objects can be used even in the case of varying materials. In contrast, we recover shape and materials without requiring sample objects composed of the same materials as the target. This allows us to reconstruct natural objects for which samples of known shape are not always available. Furthermore, our approach reconstructs BRDFs as well as shape, which allows rendering from other illuminants and viewing directions. Georghiades [6] described a method that accommodates variation of diffuse reflectance across a surface, but modeled the specular properties as being constant, whereas our linear combination model allows for variation of specular properties across the surface. Our approach permits reconstruction of more complex surface properties, such as natural objects comprised of materials with different surface roughness or mixtures of man-made materials such as metallic and nonmetallic paints. However, Georghiades also recovers lighting directions, whereas we assume they are known. Mallick et al. [16] illustrated that a rotation in color space can eliminate specularities on the surfaces of colored objects, subsequently allowing classical photometric stereo methods to be applied.

BRDF acquisition from photographic data has been widely researched in the computer graphics community. Most methods assume the scene geometry is known [17], [19], [23], [36]. Sato et al. [23] and Marschner et al. [17] captured shape using a laser range scanner and separately captured varying BRDF measurements over the surface from photographs. A BRDF estimation approach that accounts for interreflections is described by Yu et al. [36]. More recent works attempt to improve upon this process or capture more BRDF variation across surfaces, but still assume that the geometry is known a priori. For example, Matusik et al. [19] determine the best viewpoints for BRDF recovery. Lensch et al. [15] cluster material estimates over a known surface. In this work, they also refine the scanned geometry using extracted normal maps, as in our present work. However, they require approximate scanned geometry as input, whereas our method generates a surface model solely from the input images.

Classical work in physics-based vision includes segmentation based on physical models of surface reflectance [2], [7], [12]. We use a more complete reflectance model, making it possible to recover smoothly varying materials from natural objects. In addition, we take advantage of multiple images of the same objects, enabling segmentation of pixels that may be saturated in one or more images.

Recent works by Lawrence et al. [14] and Alldrin et al. [1] have made significant progress toward material separation and nonparametric BRDF estimation using a unified approach. Alldrin et al. also recover shape in their process. However, both of these works use a very large set of input images for reconstruction, and it is unclear how they would perform with much smaller input data sets. In this paper, we use a highly constrained material model, which allows us to successfully acquire models using only a dozen or fewer input images.

We also note the success of Helmholtz stereopsis in shape reconstruction on arbitrary surfaces [28], [38]. These methods avoid reconstructing BRDFs explicitly by exploiting the reciprocity property of BRDFs. However, our approach recovers not only shape but also a BRDF representation that can be edited, relit, and viewed from different viewpoints. Our method also uses a simpler capture mechanism, requiring only a single uncalibrated camera position.

Although our method uses only a single viewpoint, some recent work utilizes information in multiple views to reconstruct both shape and materials from images. Jin et al. [10], [11] utilized level-set optimization along with region-based segmentation to recover shape and material regions, whereas Yu et al. [35] recovered uniform materials and shape using a coarse-to-fine triangle mesh optimization. However, these works use only a single lighting condition as input, and therefore, cannot recover BRDFs. Paterson et al. [21] use a light source attached to a moving camera at a fixed angle to recover per-pixel BRDFs and geometry, but this method cannot recover specular properties of pixels that were never observed in specular highlight. Our approach shares specular properties between distant regions of the surface. Their method is also limited to "nearly planar" samples since the fixed angle between view and lighting direction limits the surface normals at which specular highlights can be observed. Treuille et al. [26] use a voxel representation to carve away regions of space which are not consistent with the images. In contrast, our approach requires only a single viewpoint, and does not require reference objects.

Finally, Vogiatzis et al. [30] obtain excellent shape reconstruction using both multiple lighting conditions and multiple views. Whereas their approach uses strong assumptions about near-Lambertian reflectance with constant albedo, our algorithm is constructed explicitly to recover both shape and materials of objects with spatially-varying reflectance.

3 PROBLEM STATEMENT

The input to our system is a set of images of a static target object taken from a fixed distant camera under a different distant illuminant in each image. We assume the lighting is
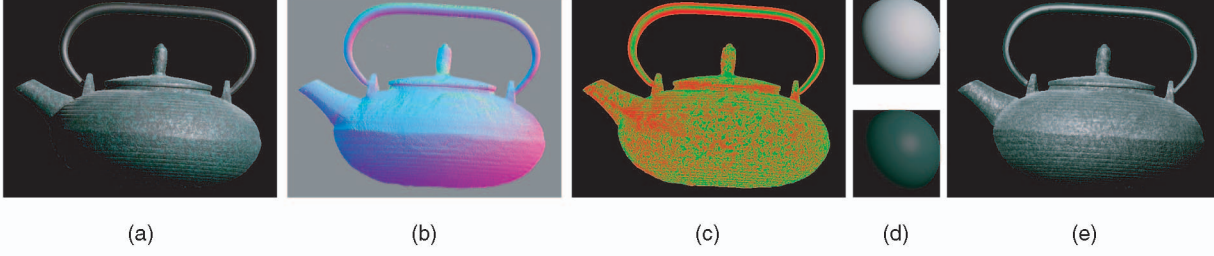
Fig. 1. From (a) photographs of an object taken under varying illumination (one of 10 photographs is shown here), we reconstruct (b) its normals and materials, represented as (c) a material weight map controlling a mixture of (d) fundamental materials. Using this representation we can (e) rerender the object under novel lighting.

known, and we do not model the effects of cast shadows, interreflections, transparency, or translucency. From these inputs, we seek to reconstruct shape and BRDFs with a constrained material model.

Our material model is motivated by the observation that real world variations in BRDF across a surface are often a result of the surface's composition from a small number of substances. For example, a block of wood with light and dark grain can be viewed as having two different substances that are blended in different amounts across the surface. We call these substances *fundamental materials*, and the mixtures of these materials at each pixel are specified by *material weight maps*.

To reconstruct the surface and materials using this model, the user provides the number of fundamental materials for which to solve. The output is a set of BRDF parameters for each of the fundamental materials, and a surface normal and material weights at each pixel. We can then reconstruct the surface by integrating the normal field. (However, as with all surface reconstructions using only normals, our approach cannot determine absolute depths, or relative depths between disconnected objects, and depth discontinuities must be explicitly marked.)

3.1 Model

We model the lights as distant directional sources, so the lighting direction $\mathbf{L}_i$ is constant over each image (indexed by i). We also model the camera as orthographic, so the view direction $\mathbf{V}$ is constant for all samples (and is therefore elided in the formulation which follows). We model the color at pixel p as if generated from a convex combination of fundamental materials:

$$I_{i,p,c} = \sum_m \gamma_{p,m} f_c(\mathbf{n}_p, \mathbf{L}_i, \alpha_m). \tag{1}$$

We use the symbol $I_{i,p,c}$ to represent the intensity of channel c of pixel p in image i. The function f_c represents color channel c of the parameterized lighting model with normal $\mathbf{n}_p$, lighting condition $\mathbf{L}_i$, and BRDF parameter vector α_m; there is one α_m for each fundamental material.

Although the fundamental materials are constant over the image, the material weight maps γ vary spatially. For example, Fig. 1a shows a cast iron teapot with speckled green paint. Fig. 1c shows two material weight maps (encoded in the red and green channels), and Figs. 1d and 1e are renderings of the two fundamental materials, corresponding to the cast iron and green paint, respectively.

To keep the solution space highly constrained, we use the isotropic Ward model [13] as our parametric reflectance model:

$$\rho_{bd,iso}(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\rho_d}{\pi} + \frac{\rho_s}{\sqrt{\cos \theta_i \cos \phi_r}} \frac{\exp[-\tan^2 \delta / \beta^2]}{4\pi\beta^2}, \tag{2}$$

where ρ_d and ρ_s are the diffuse and specular reflectance coefficients, β is a measure of roughness, and δ is the angle between $\mathbf{n}$ and the halfway vector $\mathbf{h} = (\mathbf{V} + \mathbf{L})/\|\mathbf{V} + \mathbf{L}\|$. This model has only seven parameters (in our system, ρ_d and ρ_s are RGB vectors), and is thus well suited to our problem. Each fundamental material thus has a parameter vector α_m comprised of $\rho_{d,m}$, $\rho_{s,m}$, and β_m.

Since our light sources are directional, f_c is the product of the BRDF ρ and the light intensity, with the viewing direction held constant. We normalize the input images by dividing each by its light intensity. This normalization avoids giving undue weight to some images over others. Note that this also means that we drop the intensity factor when computing f_c.

Using these imaging and reflectance models alone, the problem still remains underconstrained since there are many combinations of materials that can accurately describe most pixels. For example, a Lambertian gray pixel can be described by one gray material, or a combination of a black material and a white material, or combinations of various shades of gray, or even mixtures of several colored materials. In order to eliminate some of these sources of ambiguity, we will enforce constraints representing priors on material combinations.

The first constraint limits the range of valid material weights γ. For each additional fundamental material, the dimensionality of the mixture increases, but, in practice, we observe that only one or two materials predominate at most pixels. In the case of surfaces with discrete materials, this can be shown geometrically: Most pixels can be represented using a single material, and pixels on the border between two discrete regions can be represented using two materials, blended to account for the spatial filtering of the sensor. Thus, the only locations where more than two materials would be required are (rare) point junctions where three or more discrete regions meet. We have found in practice that using only two materials at each pixel suffices to produce plausible results, even for surfaces with continuous materials, for which no such geometric argument exists. Therefore, we use pairwise convexity constraints for these material weights, which mitigates overfitting:
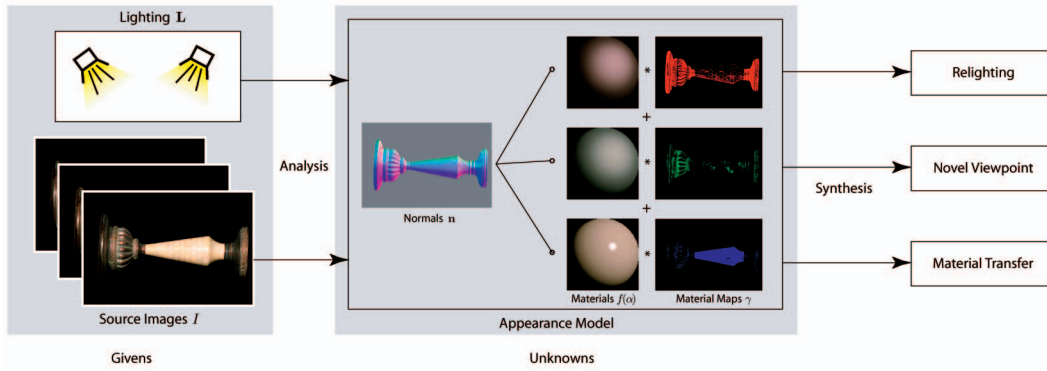
Fig. 2. A schematic illustration of our system. The normal map $\mathbf{n}$ is used to remap the fundamental materials $f(\alpha)$, and the results are modulated by the material maps γ and summed together. This representation enables relighting, change of viewpoint, and material editing operations. Source images I and variable lighting $\mathbf{L}$ are givens, and all of the other variables are unknown a priori.

$$\gamma_{p,m} \geq 0,$$
$$\sum_m \gamma_{p,m} = 1, \tag{3}$$
$$\exists m_1, m_2 \quad \text{such that} \quad \gamma_{p,m_1} + \gamma_{p,m_2} = 1.$$

Another problem is that, when fitting to the model described above, it is still possible for the fundamental materials to extrapolate far beyond any observed materials in the scene. For example, consider the case of a surface with gray matte paint with albedos in the range 0.4-0.6. We would like to restrict the individual albedo estimates for the surface to lie in this range, to prevent errors such as setting a fundamental material albedo to 0 (black). However, we don't know the range of materials in advance, and, without an additional penalty term, this surface could easily be described by basis materials with albedos 0 and 1. But we would like our vertices—the fundamental materials—to lie close to the space of observed pixels. Therefore, we also include a small spring term between each pair of fundamental materials that works together with the pairwise convexity constraint to limit the range of materials to those observed in the data. This term need not be very strong as it is intended only to disambiguate between solutions which would have the same or nearly the same energy under the original objective function (see Fig. 3).

For the Ward model, we use the following spring term:

$$Q_{sp}(\alpha) = \sum_{i \neq j} \varepsilon_d \|\rho_{d,i} - \rho_{d,j}\|^2 + \varepsilon_s \|\rho_{s,i} - \rho_{s,j}\|^2 + \varepsilon_\beta \|\beta_i - \beta_j\|^2, \tag{4}$$

where (i,j) are all pairs of distinct materials. In practice, ρ_d and ρ_s tend to vary over a similar range, whereas the roughness β typically varies an order of magnitude more. Accordingly, we set $\varepsilon_d = \varepsilon_s = 1.0$ and $\varepsilon_\beta = 0.1$ for all of our examples. Note that these values are very small relative to the data term, which sums over many pixels, so that the spring term has an effect only when the data term alone does not constrain the solution. Although this spring term has the same form as a regularizer, we do not expect the fundamental materials to be similar to each other. Rather, this term is applied as a simple method to constrain the fundamental materials to lie close to the space of observed pixels.

Bringing together the imaging model of (1), the reflectance model of (2), and the constraints of (3) and (4), we formulate the following objective function to solve for shape and materials:

$$Q(\mathbf{n}, \alpha, \gamma) = \sum_{i,p,c} \left(I_{i,p,c} - \sum_m \gamma_{p,m} f_c(\mathbf{n}_p, \mathbf{L}_i, \alpha_m) \right)^2 + Q_{sp}(\alpha), \tag{5}$$

where Q is minimized with respect to the normals, material parameters, and material weight maps (denoted by $\mathbf{n}$, α, and γ, respectively). Here, $I_{i,p,c}$ refers to normalized measured pixel values. Not all normal fields correspond to real surfaces, but we constrain the normals $\mathbf{n}$ to be an integrable field. A schematic overview of our system's inputs, model, and outputs is depicted in Fig. 2.

4 ALGORITHM

Optimization of (5) is nontrivial; many optimization algorithms become easily trapped in local minima due to the nonlinear terms of f_c and integrability constraints for the normal field $\mathbf{n}$. Our optimization approach is designed to avoid these problems. This approach has five components, each of which is described in detail in the sections which follow.

Light calibration. The light source direction and intensity are estimated using diffuse gray and chrome spheres captured under the same illumination as the target object. The calibration method is described in Section 4.1.

Initialization. Our normal maps are initialized using Lambertian photometric stereo, with thresholds to reject
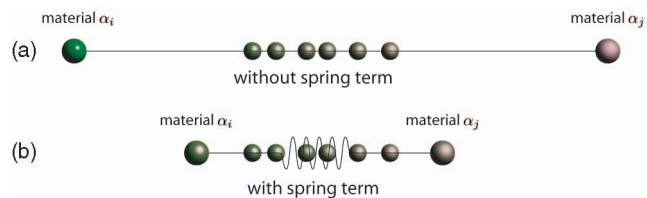

Fig. 3. The spring term acts to keep the fundamental materials close to the observed data. (a) Without the spring term, the fundamental materials may drift arbitrarily far from the data, as long as the data can be represented by a linear combination. (b) With the spring term, the fundamental materials are drawn closer to each other and to the observed data.

specular highlights. This also gives us an initial estimate of diffuse albedo, which is used to initialize the material weight maps (Section 4.2).

After initialization, the system optimizes the objective function iteratively by repeating the following three steps:

1. **Optimize BRDF parameters.** The BRDF parameters are optimized while holding the normals and material weights constant (Section 4.3).
2. **Compute surface normals and material weight maps.** To compute the normals and material weight maps, while holding the BRDF parameters constant, (5) is optimized jointly over normals and material weights. The normal optimization is performed as a discrete search, and the material weights are optimized by linear projection (Section 4.4).
3. **Enforce integrability.** The normals generated in the previous step are not guaranteed to be consistent with a 3D surface, so integrability is enforced by solving a Poisson equation to obtain a least-squares surface reconstruction, and subsequently, the normals are recomputed (Section 4.5).

Termination. Steps 1-3 are iterated until the objective function no longer decreases in successive iterations of the outer loop. Each step in our algorithm is guaranteed to monotonically decrease the objective function, except the enforcement of integrability (Section 4.5). However, in our experience, this projection step makes only very minor changes to the surface. Therefore, our optimization approach is likely—although not guaranteed—to find a solution near a local optimum.

4.1 Light Calibration

We begin by calibrating the light source directions and relative intensities by photographing calibration objects under the various lighting conditions.

We determine the illumination direction of each light source from an image of a chrome sphere, by reflecting the viewing vector about the normal at the point of greatest brightness. In principle, we could use this same object to recover light intensity as well: The relative light intensity can be found by integrating the measured radiance over all the observed pixels. However, we have found that a large number of exposures may be necessary to obtain an intensity estimate with low variance from an image of a chrome sphere alone. This is because all of the intensity is concentrated in a small number of pixels, so even a small amount of pixel noise gives an integral estimate with high variance.

Instead, we determine the light intensity using a diffuse sphere as an additional calibration object. Thus, only one low-dynamic-range exposure for each calibration object is required to reconstruct both intensity and direction.

Given the correct lighting direction $\mathbf{L}_i$, the intensity of a diffuse sphere in image i is $I_{i,p} = \ell_i \rho \mathbf{n}_p^T \mathbf{L}_i$, where ρ is the diffuse albedo of the sphere and ℓ_i is the intensity of the light in image i. So, given the known normals, we solve for the relative light intensity $\ell_i \rho = \sum_p I_{i,p} / \sum_p \mathbf{n}_p^T \mathbf{L}_i$. This is performed independently for each color channel.

If ρ is known, the absolute intensity can be recovered. However, in this paper, we have used the relative
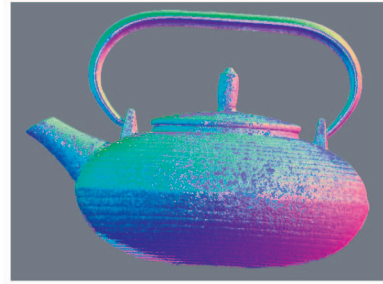

Fig. 4. RGB-encoded normals acquired using the Lambertian photometric stereo method, with threshholding to exclude shadows and highlights.

intensities $\ell_i \rho$ to solve for materials, so all of the BRDFs we have recovered are actually scaled by the unknown constant scale factor ρ.

These calibration objects are similar to those used by Hertzmann and Seitz [9], but they are used in a different way. In particular, we do not require or expect that the BRDFs of our target objects are composed of a linear combination of the calibration materials.

4.2 Initialization

We initialize normals using Lambertian photometric stereo [32]. Since this method fails in the presence of specular highlights, we employ manually selected intensity thresholds to reject shadow and specular highlight pixels from consideration. For particular configurations of lights and shiny objects, some pixels may have fewer than four inlier samples with which to estimate a normal and, for these pixels, we simply choose a normal pointing toward the camera as an arbitrary plausible normal. Although the resulting normal map is quite poor (Fig. 4), it suffices as an initial guess.

Lambertian photometric stereo also provides an estimate of the diffuse albedo. To compute our initial material weight maps, we first transform the diffuse albedo into HSV colorspace [2] and discard the V channel. This transformation reduces distortions of the diffuse albedo estimate due to specular highlights and shadows. We then cluster the pixels of the image using the parameterization $(\cos(2\pi H), \sin(2\pi H), S)$. We use an EM optimization for mixtures-of-Gaussians to segment this image into separate regions [33].

4.3 BRDF Parameter Optimization

The BRDFs of the fundamental materials, denoted as f_c in (5), are generally nonlinear functions of their parameters α. We optimize the objective function over all α simultaneously, using the Levenberg-Marquardt nonlinear optimization algorithm [22].

4.4 Computing Normals and Material Weight Maps

Next, we jointly optimize the normals ($\mathbf{n}_p$) and material weights ($\gamma_{p,m}$) of (5). We first precompute the function f_c over a discrete sampling of normals $\mathbf{n}$ for each of the lighting samples $\mathbf{L}_i$ and fundamental material parameters α_m. In practice, this simply means rendering a small "virtual sphere" (Figs. 1d, 1e, and 8d) of each fundamental material under each lighting condition of the input set.
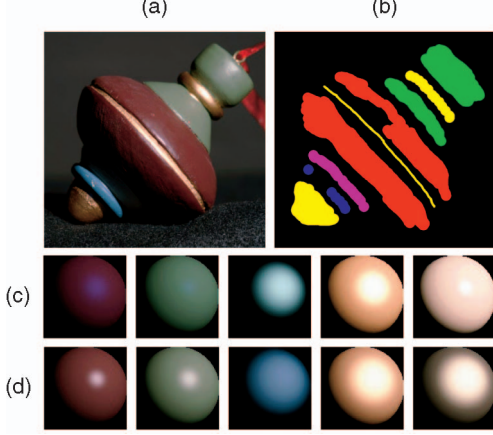
Fig. 5. (a) A Christmas ornament. (b) Manually specified constraints. (c) Fundamental materials recovered for the ornament model without weight constraints, and (d) with weight constraints.

Given these samples of the appearance functions f and the pairwise convex combination constraint, weights $\gamma_{p,m}$ are computed by linear projection and brute force search over all normals and all pairwise combinations of fundamental materials. Specifically, let $\phi_{m,i}^c(\mathbf{n}_p) = f_c(\mathbf{n}_p, \mathbf{L}_i, \alpha_m)$ denote the virtual sphere images for material m. For a given choice of normal $\mathbf{n}_p$ and pair of fundamental materials m_1, m_2, the objective function reduces to

$$Q_p = \sum_{i,c} \left(I_{i,p,c} - \gamma_{p,m_1} \phi_{m_1,i}^c(\mathbf{n}_p) - \gamma_{p,m_2} \phi_{m_2,i}^c(\mathbf{n}_p) \right)^2. \quad (6)$$

This is minimized by substituting in the constraint $\gamma_{p,m_2} = 1 - \gamma_{p,m_1}$, and solving $dQ_p/d\gamma = 0$. After some rearrangement, we have:

$$\gamma_{p,m_1} \leftarrow \frac{\sum_{i,c} \left(I_{i,p,c} - \phi_{m_2,i}^c(\mathbf{n}_p) \right) \left(\phi_{m_2,i}^c(\mathbf{n}_p) - \phi_{m_1,i}^c(\mathbf{n}_p) \right)}{\sum_{i,c} \left(\phi_{m_2,i}^c(\mathbf{n}_p) - \phi_{m_1,i}^c(\mathbf{n}_p) \right)^2}. \quad (7)$$

Since we wish to constrain the solution to convex combinations, we also clamp γ_{p,m_1} to lie between 0 and 1. We solve for these optimal convex weights for each normal and pair of materials, and select the pair and normal with the lowest objective Q_p.

For the first several iterations of the algorithm, the material weights are held at their initial values. Once the other parameters have converged, they are allowed to vary freely in this step.

Note that if three or more different materials in the scene are not linearly independent, the fundamental materials are not uniquely defined. For example, the object in Fig. 5a has several materials that can be plausibly described by many linear combinations. In such cases, we prefer the fundamental materials to correspond to specific materials visible in the scene, i.e., these pixels should have weights of 1 for one material and 0 for all others. To enforce this preference, the user may optionally paint constraint maps that fix the weights of particular pixels throughout the optimization procedure. At such pixels, only the normal is optimized, again by brute force search over the normal samples. Figs. 5b, 5c, and 5d illustrate the painted constraints for this object and the fundamental

materials solved with and without such material weight constraints, respectively. Note that this user intervention is not required by our algorithm, though it does improve the results in certain cases. (In this paper, only the result shown in Fig. 5 was obtained using such constraint maps.)

Depending on the resolution of the virtual sphere, the full brute force normal search described in Section 4.4 can be quite slow; a single pass over the image may take several hours to complete. In order to accelerate the computation, we limit the brute-force search to normals that lie close to the previously estimated normal for each pixel. This can result in small areas that become "trapped" at the wrong normal values, so, after the algorithm converges, we perform a final pass of normal/weight optimization using the full global normal search. Although, in principle, this strategy could cause the algorithm to converge to a suboptimal solution, in practice we have found it gives good results with dramatically less computation than the full normal search.

4.5 Enforcing Integrability

To compute a 3D surface from the estimated surface orientations, given the normal $\{n_x, n_y, n_z\}$ for each point, we solve for the height field $z(x, y)$ that minimizes

$$\Psi(z) = \sum_{x,y} \left(n_z \frac{\partial z(x,y)}{\partial x} + n_x \right)^2 + \left(n_z \frac{\partial z(x,y)}{\partial y} + n_y \right)^2 \quad (8)$$

using the approximations $\frac{\partial z(x,y)}{\partial x} = (z(x+1, y) - z(x, y))$, $\frac{\partial z(x,y)}{\partial y} = (z(x, y+1) - z(x, y))$ [5], [27]. This amounts to integrating the normal field. The minimization gives rise to a large but sparse system of linear equations which we solve using the conjugate gradient method [22].

The normals are then recomputed from this surface approximation. This step can be viewed as projecting the normal field into the subspace of feasible normal fields.

5 RESULTS AND APPLICATIONS

We evaluated the performance of our algorithm both by comparing its performance on synthetic data sets to ground truth, and by reconstructing novel illumination conditions for acquired data sets.

5.1 Synthetic Examples

To compare our reconstructions against ground truth, we applied several different material models to a synthetic object. Our model is the "Mozart's bust" depth map often used to evaluate shape-from-shading algorithms [37], and we used densely sampled reflectance functions from the MERL BRDF database [18]. The models were rendered using a local lighting model (i.e., without cast shadows or interreflections) under distant illumination and orthographic projection, and multiple materials with varying texture maps. (We rendered the objects under the same lighting direction vectors acquired with our real-world objects, described in the next section and illustrated in Fig. 11.) Fig. 6 illustrates some images from this synthetic data set.

We compared the reconstructions of both normals and material weight maps for these examples. (Since our reconstruction method does not handle discontinuities,
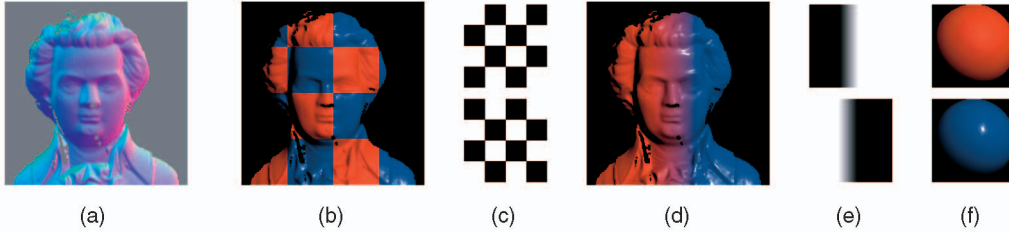
Fig. 6. Examples of the synthetic images used for reconstruction, with associated material maps used to blend the two materials "dark-red-paint" and "blue-acrylic" from the MERL BRDF database [18]. (a) Ground truth normals. (b) "Checker" input. (c) "Checker" weight maps. (d) "Ramp" input. (e) "Ramp" weight maps. (f) Spheres rendered using the two test materials.

and has no method of resolving absolute depth, we do not compare against ground truth depths.) The RMS error is reported in Table 1.

Fig. 7 illustrates the estimated models. Notice that even though the Ward model is not flexible enough to accurately represent the sharp specular peak of the acrylic BRDF, we still obtain a reasonable solution for both shape and materials. Also, the "ramp" example posed a special challenge because our color segmentation-based approach for initializing the material weight maps failed dramatically in this case: The left and right sides were grouped together in the first cluster, and the central ramp region in the second cluster. Nonetheless, our algorithm partially recovered from this poor initialization to produce reasonable results for both shape and material maps.

5.2 Acquired Examples

Although our ground truth evaluation demonstrates the effectiveness of the algorithm under ideal conditions, we also evaluated our algorithm using acquired images.

To capture our source images, we programmed a Lutron lighting control system and Canon 10D camera with a 400 mm telephoto lens to automatically capture multiple exposures of each lighting direction. (Although our algorithm assumes an orthographic camera and parallel light rays, we have obtained good results with distances from target to camera and lights of only 5 feet. The targets were all 6 inches in diameter or smaller.) Our images are captured at the full resolution of the camera ($3,072 \times 2,048$), but most of the examples in this paper were computed at a downsampled resolution of 768×512. A typical capture session with 12 light sources takes about 25 minutes, most of which is simply the time to download the high-resolution images to disk via the camera's USB 1.0 interface. The multiple exposures of each lighting direction are then combined into high dynamic range images using the technique of Debevec and Malik [4]. Since we use multiple fixed light sources, we only calibrate the light sources once each capture session.

In our model, the number of fundamental materials (indexed by m) is not specified, and is chosen by the user. For man-made objects with discrete materials, this number is chosen by simply counting different materials or paints. For natural objects with continuous materials, we find that within homogeneous regions (i.e., the skin of a fruit or the surface of a leaf), two materials typically sufficed to describe the range of surface properties. Although this will not hold true for more colorful flora or fauna, we found these rules of thumb worked well for all of the objects we tried.

Most of the examples shown in this paper converged after 10-20 iterations of the outer loop of the algorithm (steps 1, 2, and 3 from Section 3), in about 5-10 hours on a 2.8 GHz Xeon processor.

A few of our reconstructions, with different viewpoints and illuminations, are shown in Figs. 1, 8, 9, and 10. Note that for each object, the algorithm estimates a detailed normal map, plausibly segments the surface materials, estimates the reflectance properties of the materials, and reproduces the input imagery fairly accurately. A few artifacts still occur in regions of the surface that had highlights in most of the views, such as the frontal portion of the candlestick and the lower part of the leaf. The presence of highlights in all of the images causes the algorithm to overestimate the diffuse component in such regions. These artifacts do not have a significant impact on rerendering and relighting of these objects.

Since our goal is to produce a visually plausible model, we compare to ground truth by considering the reconstruction error for novel lighting conditions. In Fig. 10, we show our relighting results compared with real photos taken

TABLE 1
RMS Error for Synthetic Data Sets Illustrated in Fig. 6

	recovered normals	weight maps
checker	0.0383	0.0279
ramp	0.0740	0.0685

For normals, the error is computed over normalized vectors with range $[-1, 1]$ in each dimension. For weight maps, the error is computed over the estimated weights, with range $[0, 1]$.
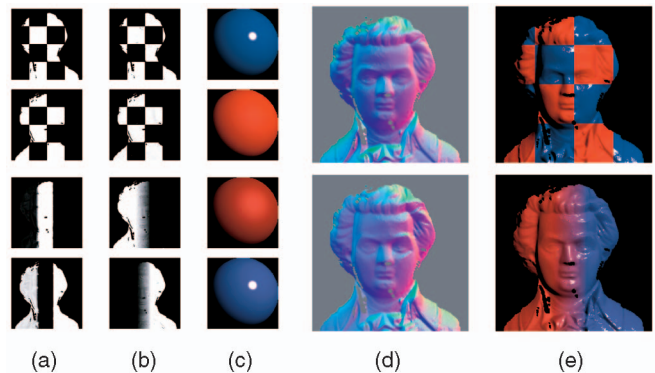

Fig. 7. Reconstruction results for the synthetic data sets shown in Fig. 6. (a) Initial weight maps. (b) Final recovered weight maps. (c) Recovered materials. (d) Recovered normals. (e) Rerendered model.
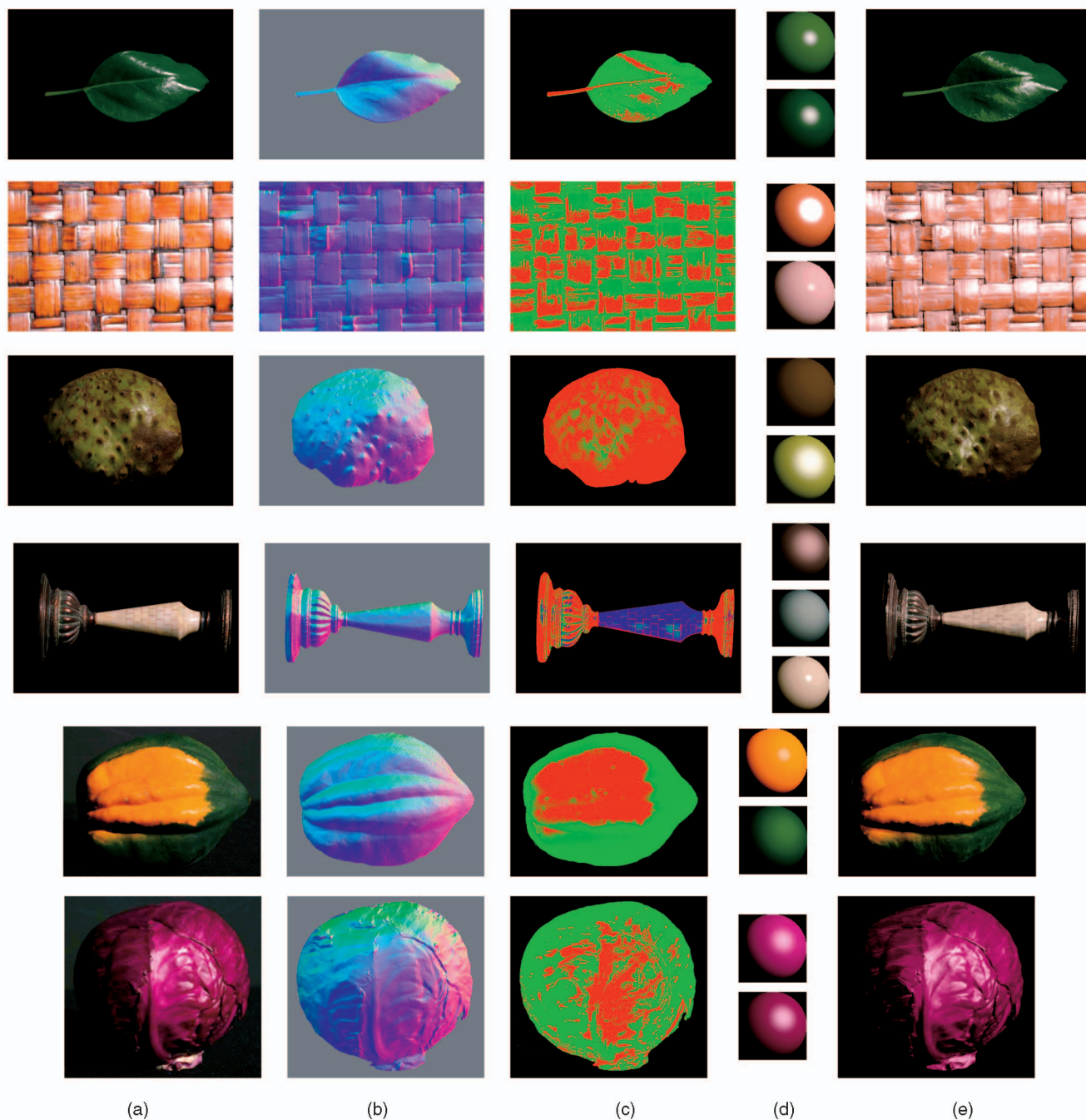
(a) (b) (c) (d) (e)

Fig. 8. Input, model, and reconstructions for leaf, woodweave, cherimoya, candlestick, squash, and cabbage. (a) Source image (1 of 10). (b) Recovered normal map, RGB-encoded. (c) Recovered material weights, in false color. (d) Virtual spheres for fundamental materials. (e) Model rendered under original lighting condition (compare with (a)).

under lighting conditions that were held out from the input data set. The RMS error is 0.0290 for the leaf image pair and 0.0303 for the teapot image pair (with source images scaled to a maximum pixel value of 1.0). Additional relighting results can be seen in the supplementary video materials, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.102. Fig. 11 illustrates the lighting directions used for acquisition and reconstruction in Fig. 10 and in the video. Note that we can produce plausible relightings of objects even when the lighting directions are well outside the convex hull of lighting directions used for acquisition.

5.3 Editing Operations

Direct BRDF Manipulation. The Ward BRDF model has a small number of parameters that can be directly manipulated to change one or more of the fundamental materials, without modifying the others. For example, in Fig. 12a, we have manually edited the BRDF parameters of the green paint of our teapot to appear as gold leaf.

BRDF Transfer. BRDF parameters captured from one object can easily be transferred to another. Fig. 12b illustrates materials captured from our cherimoya, applied to the rerendering of our teapot.
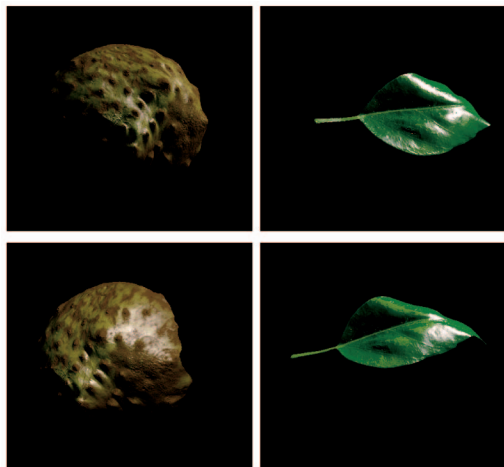
Fig. 9. Synthesized views of our reconstructions of a cherimoya (a tropical fruit) and a leaf.
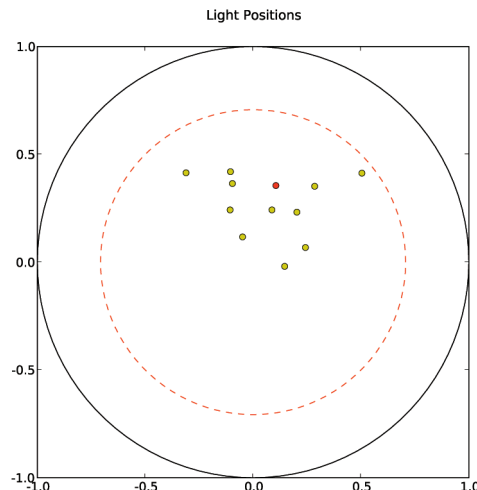

Fig. 11. This diagram illustrates the lighting directions used for acquisition (yellow dots), one light held out for ground truth comparison in Fig. 10 (red dot), and the path used to construct the relighting animations in the supplementary video materials, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.102 (dashed red line). Normalized direction vectors are projected into the xy plane: All vectors have positive z (i.e., all lighting directions lie on the camera side of the acquired object).

Material/shape extraction and transfer. One application of our method is to extract surface texture and detail for transfer to new objects; we show an example in which texture and normals were extracted from a portion of the cherimoya in Fig. 8 and reapplied to a computer graphic model of a head. To achieve this, we extract planar texture and normal maps from a small crop of the recovered cherimoya shape and materials. Then, we use texture synthesis to generate more of these textures and remap the resulting textures onto the head surface.

Our first goal was to remove large-scale variations in shape, while keeping the fine-scale "textural" detail. We applied Gaussian smoothing to the depth map $z(x, y)$ to produce a *base surface* $z'(x, y)$. We additionally defined a target surface, in our case a flat plane $z''(x, y) = 0$. Since the source patch is relatively flat, we did not need to reparameterize it, although reparameterization could be used for more highly curved surfaces. A conventional approach would be to transfer displacement vectors from their local coordinate frames on the base mesh to the corresponding coordinate frames on the target mesh (e.g., [3]); this change-of-basis amounts to a rotation for each displacement vector. Instead, we transferred surface normals, i.e., applying to each surface normal a change-of-basis from the base surface to the target surface. This gave us a new normal field with global shape variations removed; a new depth map was obtained by surface integration as in Section 4.5 (the target surface was discarded). In this example, we found that our approach preserved more

high-frequency detail than transferring displacements. The material weight map was unchanged by this process.

This gave us a planar patch of surface texture that may be rendered, or used as source data for texture synthesis. We used texture synthesis to produce a much larger patch from just the normal map (Fig. 13f), and then used Image Analogies [8] to produce a corresponding weight map from the normal map (Fig. 13g). (Alternatively, texture synthesis could be performed by directly comparing pixel neighborhoods between the source and target surfaces, similar to [25], [29], [31], [34].)

Finally, the synthesized normals were integrated again to form a depth map that can be used for bump or displacement mapping. In Fig. 14, we applied much larger maps synthesized in this fashion to a polygon mesh obtained from a Cyberware scan, and rendered the resulting model using Pixar's Photorealistic RenderMan with a 70-line surface shader.

6 DISCUSSION AND FUTURE WORK

We have demonstrated a method that acquires both shape and spatially-varying BRDFs from a set of photographs
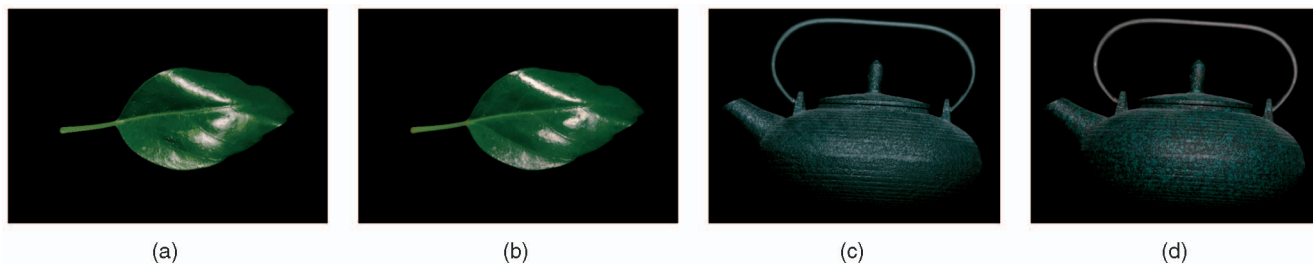

Fig. 10. Comparison to ground truth. (a) Leaf rendered under a novel lighting condition not in the training set. (b) Real leaf under the same lighting condition. (c) Teapot rendered under a novel lighting condition not in the training set. (d) Real teapot under the same lighting condition.
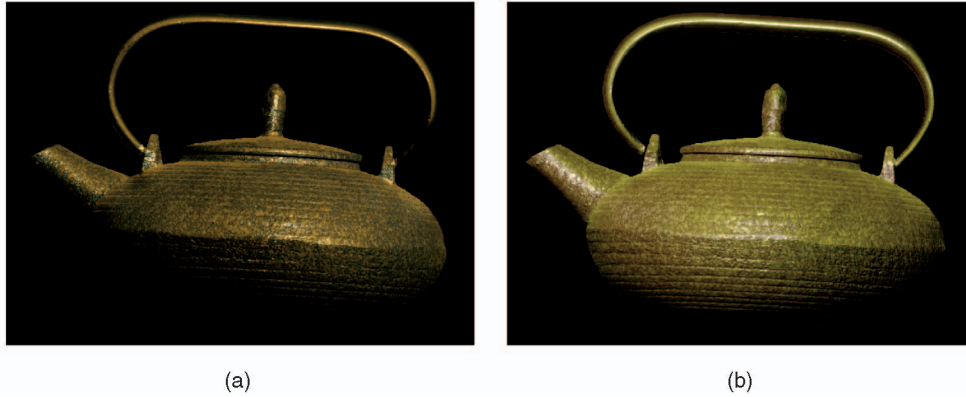
Fig. 12. Teapot edited to appear (a) as gold leaf by direct manipulation of BRDF parameters and (b) as a cherimoya by transfer from another model.
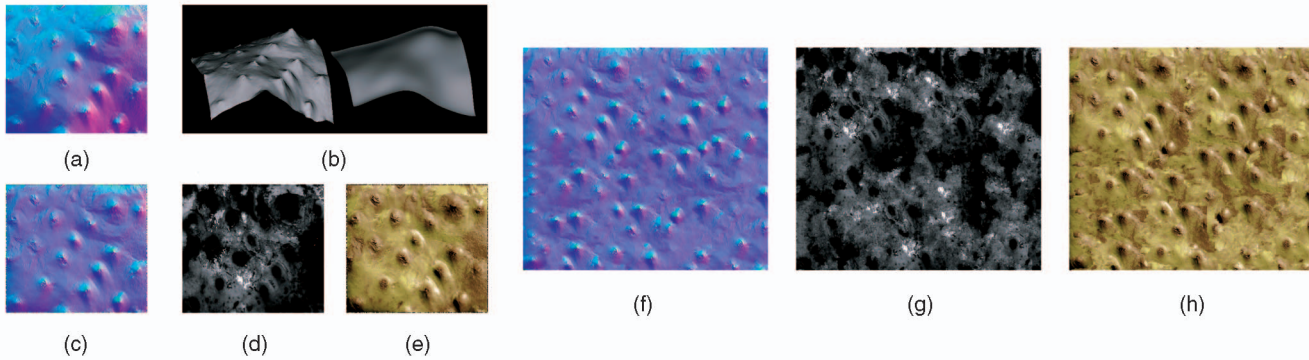

Fig. 13. Intermediate steps in the construction of Fig. 14. (a) Normals extracted using our methods are integrated to construct (b) depths, and smoothed. The smoothed normals are used to create (c) a "flattened" normal map which retains high-frequency texture. This is combined with (d), the recovered weight map, to render (e) a flattened swatch of cherimoya skin. (f) Larger normal maps and (g) weight maps can be constructed using texture synthesis to create (h) novel renderings.

captured under varying illumination. Although our shape and material reconstructions are of lower fidelity than those attainable using methods that assume one or the other is given and use large numbers of observations, we can nonetheless acquire a wide range of models, which can be reused under various lighting and viewing conditions. The spatially varying BRDFs that we acquire enable a set of useful and interesting editing operations. We believe this

Fig. 14. A head model textured using material maps and bump maps synthesized from a cherimoya skin (as shown in Fig. 13). 3D model provided by Domi Piturro.

approach represents an important step toward acquisition and reconstruction of both shape and material from a single set of photographic data.

In particular, the increasing resolution of consumer-grade digital cameras suggests that very high-resolution shape reconstruction may be attainable at low cost using photometric stereo methods, especially when one considers that we have used only one-sixteenth of the available pixels in our images. To illustrate this point, Fig. 15 shows a submillimeter surface reconstruction for one of our data sets. Although this reconstruction has extraordinary resolution, it has some geometric distortions due primarily to the inaccuracy of the distant light source and orthographic camera assumptions. We would like to remove these assumptions in future work.

Our approach is able to capture shape and BRDFs of reflective objects using a small number of photos and without repainting the objects gray, as is typically required for high-quality laser range scanning. However, the objective function in (5) is still subject to some overfitting, usually in the case where pixels appear in specular highlight in most or all of the image samples. Such pixels may be assigned a bright diffuse material instead of a dark specular material because there is too little data to distinguish diffuse from specular color. This can be seen in Fig. 6, where a handful of pixels on the forehead and right eyelid are mislabeled with the red paint material. This occurs because neither material has exactly the right
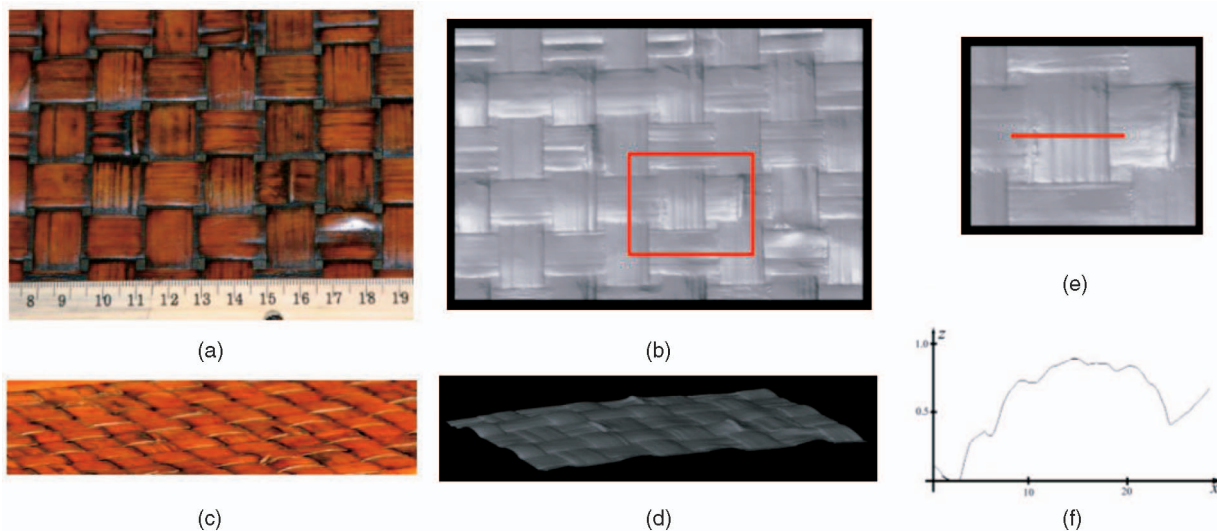
Fig. 15. (a) Wooden tabletop with ruler for scale (in centimeters). (b) Plastic shaded reconstruction. (c) Oblique view photograph. (d) Oblique view reconstruction. (e) Closeup of inset. (f) Depth profile of red line in inset, with z-axis scaled up for clarity (axes units are millimeters).

specular properties to represent this pixel, but the red material is somewhat brighter and is therefore a better fit. However, these pixels often have neighbors which are assigned properly. It is therefore possible that adding a smoothness term for material weight maps to our objective function will ameliorate these artifacts. Such an objective function will require a modified optimization approach.

Also, because we employ a local reflectance model, our algorithm does not properly account for shadows, interreflections, and subsurface scattering. Few real-world objects are free of these effects; therefore, future work must address techniques to compensate for their appearance.

Despite these limitations, we believe our method will enable more rapid acquisition of computer models suitable for applications in computer graphics and other domains, from source material at a range of scales that are inaccessible to laser range scanners.

REFERENCES

[1] N. Alldrin, T. Zickler, and D. Kriegman, "Photometric Stereo with Non-Parametric and Spatially-Varying Reflectance," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2008.
[2] R. Bajcsy, S. Lee, and A. Leonardis, "Color Image Segmentation with Detection of Highlights and Local Illumination Induced by Inter-Reflections," *Proc. IEEE Int'l Conf. Pattern Recognition,* pp. 785-790, 1990.
[3] H. Biermann, I. Martin, F. Bernardini, and D. Zorin, "Cut-and-Paste Editing of Multiresolution Surfaces," *ACM Trans. Graphics,* vol. 21, no. 3, pp. 312-321, July 2002.
[4] P.E. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," *Proc. ACM SIGGRAPH '97,* pp. 369-378, 1997.
[5] D.A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach.* Prentice Hall, 2003.
[6] A.S. Georghiades, "Recovering 3-D Shape and Reflectance from a Small Number of Photographs," *Proc. Eurographics/SIGGRAPH Workshop Rendering,* pp. 230-240, 2003.
[7] G. Healey, "Segmenting Images Using Normalized Color," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 22, no. 1, pp. 64-73, Jan./Feb. 1992.
[8] A. Hertzmann, C.E. Jacobs, N. Oliver, B. Curless, and D.H. Salesin, "Image Analogies," *Proc. ACM SIGGRAPH,* pp. 327-340, 2001.
[9] A. Hertzmann and S.M. Seitz, "Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 8, pp. 1254-1264, Aug. 2005.
[10] H. Jin, S. Soatto, and A.J. Yezzi, "Multi-View Stereo beyond Lambert," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 171-178, 2003.
[11] H. Jin, A.J. Yezzi, and S. Soatto, "Region-Based Segmentation on Evolving Surfaces with Application to 3D Shape and Radiance Estimation," *Proc. European Conf. Computer Vision,* pp. 114-125, 2004.
[12] G. Klinker, S. Shafer, and T. Kanade, "A Physical Approach to Color Image Understanding," *Int'l J. Computer Vision,* vol. 4, no. 1, pp. 7-38, 1990.
[13] G.J.W. Larson, "Measuring and Modeling Anisotropic Reflection," *Proc. ACM SIGGRAPH,* pp. 265-272, 1992.
[14] J. Lawrence, C. DeCoro, S. Rusinkiewicz, A. Ben-Artzi, R. Ramamoorthi, W. Matusik, and H. Pfister, "Inverse Shade Trees for Non-Parametric Material Representation and Editing," *ACM Trans. Graphics,* vol. 25, no. 3, pp. 735-745, 2006.
[15] H.P.A. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel, "Image-Based Reconstruction of Spatial Appearance and Geometric Detail," *ACM Trans. Graphics,* vol. 22, no. 2, pp. 234-257, 2003.
[16] S.P. Mallick, T.E. Zickler, D.J. Kriegman, and P.N. Belhumeur, "Beyond Lambert: Reconstructing Specular Surfaces Using Color," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 619-626, 2005.
[17] S.R. Marschner, S.H. Westin, E.P.F. Lafortune, K.E. Torrance, and D.P. Greenberg, "Image-Based BRDF Measurement Including Human Skin," *Proc. Eurographics/SIGGRAPH Workshop Rendering,* pp. 139-152, 1999.
[18] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A Data-Driven Reflectance Model," *ACM Trans. Graphics,* vol. 22, no. 3, pp. 759-769, July 2003.
[19] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "Efficient Isotropic BRDF Measurement," *Proc. Eurographics/SIGGRAPH Workshop Rendering,* pp. 241-248, 2003.
[20] S.K. Nayar, K. Ikeuchi, and T. Kanade, "Determining Shape and Reflectance of Hybrid Surfaces by Photometric Sampling," *IEEE Trans. Robotics and Automation,* vol. 6, no. 4, pp. 418-431, Aug. 1990.

[21] J.A. Paterson, D. Claus, and A.W. Fitzgibbon, "BRDF and Geometry Capture from Extended Inhomogeneous Samples Using Flash Photography," *Computer Graphics Forum,* vol. 24, no. 3, pp. 383-391, 2005.

[22] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing,* second ed. Cambridge Univ. Press, 1992.

[23] Y. Sato, M.D. Wheeler, and K. Ikeuchi, "Object Shape and Reflectance Modeling from Observation," *Proc. ACM SIGGRAPH '97,* pp. 379-387, 1997.

[24] W.M. Silver, "Determining Shape and Reflectance Using Multiple Images," master's thesis, Massachusetts Inst. of Technology, 1980.

[25] X. Tong, J. Zhang, L. Liu, X. Wang, B. Guo, and H.-Y. Shum, "Synthesis of Bidirectional Texture Functions on Arbitrary Surfaces," *ACM Trans. Graphics,* vol. 21, no. 3, pp. 665-672, July 2002.

[26] A. Treuille, A. Hertzmann, and S.M. Seitz, "Example-Based Stereo with General BRDFs," *Proc. European Conf. Computer Vision,* 2004.

[27] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision.* Prentice Hall, 1998.

[28] P. Tu and P.R.S. Mendonça, "Surface Reconstruction via Helmholtz Reciprocity with a Single Image Pair," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 541-547, 2003.

[29] G. Turk, "Texture Synthesis on Surfaces," *Proc. ACM SIGGRAPH '01,* pp. 347-354, Aug. 2001.

[30] G. Vogiatzis, C. Hernandez, and R. Cipolla, "Reconstruction in the Round Using Photometric Normals and Silhouettes," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1847-1854, 2006.

[31] L.-Y. Wei and M. Levoy, "Texture Synthesis over Arbitrary Manifold Surfaces," *Proc. ACM SIGGRAPH '01,* pp. 355-360, Aug. 2001.

[32] R.J. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Eng.,* vol. 19, no. 1, pp. 139-144, 1980.

[33] T. Yamazaki, "Introduction of EM Algorithm into Color Image Segmentation," *Proc. IEEE Int'l Conf. Intelligent Processing Systems,* pp. 368-371, 1998.

[34] L. Ying, A. Hertzmann, H. Biermann, and D. Zorin, "Texture and Shape Synthesis on Surfaces," *Proc. Rendering Techniques 2001: 12th Eurographics Workshop Rendering,* pp. 301-312, June 2001.

[35] T. Yu, N. Xu, and N. Ahuja, "Recovering Shape and Reflectance Model of Non-Lambertian Objects from Multiple Views," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 226-233, 2004.

[36] Y. Yu, P. Debevec, J. Malik, and T. Hawkins, "Inverse Global Illumination: Recovering Reflectance Models of Real Scenes from Photographs," *Proc. ACM SIGGRAPH,* pp. 215-224, 1999.

[37] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah, "Shape-from-Shading: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 8, pp. 690-706, Aug. 1999.

[38] T. Zickler, P.N. Belhumeur, and D.J. Kriegman, "Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction," *Int'l J. Computer Vision,* vol. 49, nos. 2/3, pp. 215-227, 2002.

Dan B Goldman received the BS and MS degrees in computer science from Stanford University in 1994 and 1995, respectively, and the PhD degree in computer science and engineering from the University of Washington in 2007. He is a senior research scientist at Adobe Systems in Seattle. Before joining Adobe, he was in both research and production at Industrial Light and Magic, Digital Domain, and Electronic Arts. He received an ARCS Fellowship in 2002. His research interests include computer vision, computer graphics, and human-computer interaction.

Brian Curless received the BS degree in electrical engineering from the University of Texas at Austin in 1988 and the PhD degree from Stanford University in 1997. He is an associate professor in the Department of Computer Science and Engineering at the University of Washington. Following his PhD, he helped cofound the Digital Michelangelo Project as a research associate at Stanford. He joined the faculty at the University of Washington in January 1998 and spent his 2004-2005 sabbatical at Columbia University. He received the Gores Teaching Award and the Arthur Samuels Computer Science Thesis Award while at Stanford, and a US National Science Foundation Career Award, a Sloan Fellowship, and an ACM Student Chapter Teaching Award while at the University of Washington. He is also a coeditor-in-chief of the journal *Foundations and Trends in Computer Graphics and Vision*. He is a member of the Graphics and Imaging Laboratory (GRAIL), which is dedicated to innovation in computer graphics and computer vision. His research is focused on computational and 3D photography. He is a member of the IEEE.

Aaron Hertzmann received the BA degree in computer science and art and art history from Rice University in 1996, and the MS and PhD degrees in computer science from New York University in 1998 and 2001, respectively. He is an associate professor of computer science at the University of Toronto. In the past, he has been with the University of Washington, Microsoft Research, the Mitsubishi Electric Research Lab, Interval Research Corporation, and the NEC Research Institute. He serves as an associate editor for the *ACM Transactions on Graphics*. His awards include an MIT TR100 in 2004, an Ontario Early Researcher Award in 2005, a Sloan Foundation Fellowship in 2006, a Microsoft New Faculty Fellowship in 2007, and a University of Toronto Teaching Award in 2008. His research interests include computer vision, computer graphics, and machine learning. He is a member of the IEEE.

Steven M. Seitz received the BA degree in computer science and mathematics from the University of California, Berkeley in 1991 and the PhD degree in computer sciences at the University of Wisconsin, Madison in 1997. He is a professor in the Department of Computer Science and Engineering at the University of Washington. Following his doctoral work, he spent one year visiting the Vision Technology Group at Microsoft Research, and subsequently two years as an assistant professor in the Robotics Institute at Carnegie Mellon University. He joined the faculty at the University of Washington in July 2000. He was twice awarded the David Marr Prize for the best paper at the International Conference of Computer Vision, and has received a US National Science Foundation Career Award, a US Office of Naval Research Young Investigator Award, and an Alfred P. Sloan Fellowship. Seitz, Noah Snavely, and Richard Szeliski developed the Photo Tourism technology underlying Microsoft's Photosynth. He is a senior member of the IEEE.

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